

Evolution and Genetics of Floral Color Polymorphisms in *Clarkia gracilis* ssp. *sonomensis*

and *Erythronium umbilicatum*

by

Rong-Chien Lin

Department of Biology
Duke University

Date: _____

Approved:

Mark D. Rausher, Advisor

Xinnian Dong

Kathleen Donohue

Paul S. Manos

Thomas Mitchell-Olds

John H. Willis

Dissertation submitted in partial fulfillment of
the requirements for the degree of Doctor
of Philosophy in the Department of
Biology in the Graduate School
of Duke University

2020

ABSTRACT

Evolution and Genetics of Floral Color Polymorphisms in *Clarkia gracilis* ssp. *sonomensis*

and *Erythronium umbilicatum*

by

Rong-Chien Lin

Department of Biology
Duke University

Date: _____

Approved:

Mark D. Rausher , Advisor

Xinnian Dong

Kathleen Donohue

Paul S. Manos

Thomas Mitchell-Olds

John H. Willis

An abstract of a dissertation submitted in partial
fulfillment of the requirements for the degree
of Doctor of Philosophy in the Department of
Biology in the Graduate School of
Duke University

2020

Copyright by
Rong-Chien Lin
2020

Abstract

Floral color polymorphisms are pervasive in nature. Understanding the genetic basis of such polymorphisms allows us to address major evolutionary questions including: what types of genetic changes contribute to the evolution of phenotypic diversification? Have they arisen from new or pre-existing genetic material? And are certain types of genetic changes disproportionately involved? This dissertation addresses these questions by investigating two polymorphisms involving loss of anthocyanin pigmentation in some of the individuals in the populations. In Chapter 1, I characterized the genetic basis of the “white cup” phenotype in *Clarkia gracilis* ssp. *sonomensis*, where no anthocyanins are produced in the basal region of the petal. I demonstrated that the cup pigmentation is controlled by an *R2R3-MYB* transcription factor using transcriptome analysis, gene expression assays and cosegregation examination. I also found that petal pigmentation requires at least four *R2R3-MYB* genes, each gene exhibiting a spatiotemporal expression pattern that is different from each other, and each color pattern element is controlled by a different *R2R3-MYB* gene. In Chapter 2, I examined the evolution of petal pigmentation patterning by analyzing the phylogenetic relationship of these petal *R2R3-MYB* genes from *C. g.* ssp. *sonomensis*, its closely related subspecies, *C. g.* ssp. *albicaulis* and their progenitor species, *C. amoena* ssp. *huntiana* and *C. lassenensis*. I also compared the expression domains of these *R2R3-MYB* genes in these

four (sub)species. I found that the *R2R3-MYB* genes that have region-specific expression patterns are derived from duplication events occurred before polyploidization of *C. gracilis*. These findings suggest that gene duplication of the *R2R3-MYB* genes plays an important role in the evolution of petal pigmentation patterning in *C. gracilis*. In Chapter 3, I identified an *R2R3-MYB* gene involved in producing purple or yellow anthers in *Erythronium umbilicatum* using transcriptome analysis, gene expression assays and a likelihood-estimation procedure. This *R2R3-MYB* gene regulates the expression of three anthocyanin enzyme-coding genes coordinately. However, this *R2R3-MYB* is present in multiple, highly similar copies. Isolating causal genetic changes from these copies has not been successful, as it would require a better knowledge of the *Erythronium* genome. In sum, this work shows that *R2R3-MYB* transcription factors are the primary determinants of floral-color polymorphisms and the *R2R3-MYB* gene family generated by gene duplication facilitates the evolution of novel characters.

Dedication

To my family.

Contents

Abstract	iv
List of Tables	xii
List of Figures	xiv
Acknowledgements	xvii
Introduction	1
1. Genetics of petal pigmentation patterning in <i>Clarkia gracilis</i> ssp. <i>sonomensis</i> (Onagraceae)	5
1.1 Introduction	5
1.2 Materials and Methods	12
1.2.1 Study system	12
1.2.2 Plant growth and crosses	13
1.2.3 Pigment identification	14
1.2.4 Transcriptome sequencing	16
1.2.5 Bioinformatic analyses	17
1.2.6 Cloning of full-length coding sequences	18
1.2.7 Quantification of gene expression	20
1.2.8 Genotyping F ₂ plants at <i>CgsMYB12</i>	21
1.2.9 Semi-quantitative assessment of gene expression across flower-bud developmental stages	22
1.2.10 Overexpression of <i>C. g. ssp. sonomensis</i> petal <i>R2R3-MYB</i> genes in <i>Arabidopsis</i>	23
1.3 Results	25

1.3.1 Cup color is controlled by a single locus	25
1.3.2 Different color regions contain different types of anthocyanins	26
1.3.3 Anthocyanin enzyme-coding genes are downregulated in white cup	29
1.3.4 Three <i>R2R3-MYB</i> genes are acting differently in petal pigmentation	32
1.3.5 Cup color cosegregates with <i>CgsMYB12</i>	38
1.3.6 These petal <i>R2R3-MYB</i> genes activate anthocyanin production	39
1.4 Discussion.....	40
1.4.1 A functional mutation in <i>CgsMYB12</i> is associated with white-cup formation	40
1.4.2 Gene duplication contributes to evolution of novel pattern elements	45
1.4.3 Petal pigmentation of <i>Clarkia gracilis</i> ssp. <i>sonomensis</i> requires multiple <i>R2R3-MYB</i> genes.....	47
1.4.4 Evolutionary lability of anthocyanin regulatory network.....	48
2. The evolution of petal pigmentation patterning in <i>Clarkia gracilis</i>	51
2.1 Introduction.....	51
2.2 Materials and Methods	56
2.2.1 Plant growth.....	56
2.2.2 Cloning of the <i>R2R3-MYB</i> genes	57
2.2.3 Phylogenetic analysis.....	57
2.2.4 Semi-quantitative assessment of gene expression across flower-bud developmental stages.....	58
2.3 Results	59
2.3.1 The <i>R2R3-MYB</i> genes evolved before polyploidization of <i>C. gracilis</i>	59

2.3.2 The expression domains of these <i>R2R3-MYB</i> genes are conserved between species	61
2.4 Discussion.....	63
2.4.1 The petal <i>R2R3-MYB</i> genes evolved through ancient duplications prior to polyploidization of <i>Clarkia gracilis</i>	63
2.4.2 Model for the evolution of color patterns in <i>C. gracilis</i>	64
2.4.3 Polyploidization, gene loss, and the evolution of white patches	66
3. Downregulation of anthocyanin genes contributes to an anther-color polymorphism in two trout lilies (<i>Erythronium</i>).....	70
3.1 Introduction.....	70
3.2 Materials and Methods	76
3.2.1 Study system	76
3.2.2 Sample collection.....	77
3.2.3 Characterization of anthocyanidins in anthers	78
3.2.4 Transcriptome sequencing	79
3.2.5 Bioinformatic analyses.....	80
3.2.6 Cloning of full-length coding sequences.....	80
3.2.7 Quantification of gene expression in anthers.....	83
3.2.8 Semi-quantification of gene expression in leaves and anthers	85
3.2.9 Likelihood estimation of <i>cis</i> -regulatory mutations causing triple downregulation	85
3.2.10 Construction of <i>EuMYB3</i> gene tree.....	86
3.3 Results	87

3.3.1 Cyanidin-derived anthocyanins occur in purple anthers only	87
3.3.2 Identification of anthocyanin genes.....	88
3.3.3 Three enzyme-coding genes are downregulated simultaneously in the yellow anthers.....	92
3.3.4 <i>Cis</i> -regulatory mutations in the enzyme-coding genes are not likely responsible for triple downregulation.....	93
3.3.5 Downregulated genes are not also nonfunctional	96
3.3.6 A single set of transcription factors is likely responsible for triple downregulation	97
3.3.7 Multiple copies of <i>EuMYB3</i>	102
3.4 Discussion.....	105
3.4.1 Downregulation of transcription factors explains absence of anthocyanins in yellow anthers.....	105
3.4.2 Similar downregulation observed in <i>E. americanum</i> and <i>E. umbilicatum</i>	112
3.4.3 Conclusion.....	114
Conclusions.....	115
Appendix I	117
A1.1 The scripts used to run the bioinformatic analyses.....	119
A1.2 Reference gene selection for quantitative real-time PCR (qPCR).....	123
Appendix II.....	138
Appendix III.....	143
A3.1 The scripts used to run the bioinformatic analyses	146
A3.2 Derivation of equation (1)	151

References 174
Biography 190

List of Tables

Table 1: Anthocyanin genes identified from the transcriptomes through BLASTing and transcript abundance estimation	29
Table 2: Cosegregation of a 1-bp deletion in <i>CgsMYB12</i> and the cup phenotype in <i>Clarkia gracilis</i> ssp. <i>sonomensis</i>	38
Table 3: Gene expression profile in the transcriptomes of purple and yellow anthers of <i>Erythronium umbilicatum</i>	88
Table 4: Gene expression profile in the transcriptomes of red, yellow and orange anthers of <i>Erythronium americanum</i>	90
Table 5: Nucleotide sequence differences in the coding regions of five anthocyanin enzyme-coding genes between purple- and yellow-anthered <i>E. umbilicatum</i>	92
Table 6: Total number of <i>EuMYB3</i> sequences obtained from the purple- and yellow-anthered individuals.	102
Table 7: The number of different <i>EuMYB3</i> sequences obtained from gDNA of each individual.....	103
Table A1.1: Voucher information.....	117
Table A1.2: Query sequences used in BLAST searches.	118
Table A1.3: Primers used in Chapter 1.....	121
Table A1.4: The candidate genes selected for reference gene selection.	126
Table A1.5: Expression stability analyses of the candidate reference genes	128
Table A2.1: Voucher information.....	138
Table A2.2: Primers used in Chapter 2.....	139
Table A2.3: FPKM estimates of <i>CgaMYB6</i> and <i>CgaMYB11</i> in the <i>C. g.</i> ssp. <i>albicaulis</i> transcriptomes.	141
Table A3.1: The <i>Erythronium umbilicatum</i> anther-color polymorphism in 16 sites in North Carolina.	143

Table A3.2: The query sequences used in the BLAST searches.....	145
Table A3.3. Primers used in Chapter 3.....	149

List of Figures

Figure 1: A simplified schematic diagram of the anthocyanin biosynthetic pathway.....	6
Figure 2: Flowers of <i>Clarkia gracilis</i> ssp. <i>sonomensis</i>	10
Figure 3: Anthocyanidins in the different color regions in the F ₂ flowers.....	28
Figure 4: Differences in expression of the enzyme-coding genes between pink background and white cup.....	32
Figure 5: The neighbor-joining phylogenetic tree of R2R3-MYB proteins.....	33
Figure 6: Differences in expression of three R2R3-MYB genes between pink background and white cup.....	34
Figure 7: Expression patterns of the anthocyanin enzyme-coding genes and the R2R3-MYB genes across the flower-bud developmental stages.....	37
Figure 8: Phenotypic and molecular characterization of the <i>Arabidopsis</i> transgenic plants.....	41
Figure 9: Petal pigmentation in <i>Clarkia gracilis</i> ssp. <i>sonomensis</i> during petal development.....	47
Figure 10: Flowers of <i>Clarkia</i> (sub)species.....	54
Figure 11: The maximum likelihood phylogenetic tree of R2R3-MYB genes.....	60
Figure 12: Expression patterns of MYB6, MYB11 and MYB12 across developmental stages.....	63
Figure 13: Model of petal pattern element evolution in <i>Clarkia</i>	67
Figure 14: A simplified schematic diagram of the anthocyanin biosynthetic pathway....	73
Figure 15: Anthocyanidins in the anthers of <i>Erythronium umbilicatum</i>	87
Figure 16: Anthocyanidins in the anthers of <i>Erythronium americanum</i>	88
Figure 17: Expression of five enzyme-coding genes between 10 purple- and 10 yellow-anthered <i>E. umbilicatum</i>	94

Figure 18: Expression of five enzyme-coding genes between 3 purple-, 3 orange and 3 yellow-anthered <i>E. americanum</i>	95
Figure 19: Semi-quantitative real-time PCR of the downregulated genes.....	97
Figure 20: Expression of three anthocyanin-associated transcription factors between 10 purple- and 10 yellow-anthered <i>E. umblicatum</i>	98
Figure 21: The neighbor-joining phylogenetic tree of bHLH proteins.....	99
Figure 22: The neighbor-joining phylogenetic tree of R2R3-MYB proteins.....	100
Figure 23: Semi-quantitative real-time PCR of <i>EuMYB3</i> in the anther and leaf tissues.	101
Figure 24: The maximum-likelihood tree of 88 <i>EuMYB3</i> sequences.....	104
Figure A1.1: Five floral phenotypes in the F ₂ population.	129
Figure A1.2: Differences in expression of <i>CgsMYB6</i> and <i>CgsMYB12</i> between the pink cup and white cup.	130
Figure A1.3: Alignment of the full-length <i>CgsMYB12</i> and <i>CgsMYB12W</i> sequences.....	131
Figure A1.4: Cosegregation of a 1-bp deletion in <i>CgsMYB12</i> and the cup phenotypes..	132
Figure A1.5: Alignment of anthocyanin-regulating R2R3-MYB proteins.....	136
Figure A2.1: Sections dissected from the petal of <i>Clarkia</i>	142
Figure A3.1: Distributions of <i>Erythronium americanum</i> and <i>E. umbilicatum</i>	153
Figure A3.2: Sequence alignment of the <i>EuChs</i> coding region.....	155
Figure A3.3: Sequence alignment of the <i>EuF3h</i> coding region.....	157
Figure A3.4: Sequence alignment of the <i>EuDfr</i> coding region.....	160
Figure A3.5: Sequence alignment of the <i>EuAns</i> coding region.....	162
Figure A3.6: Sequence alignment of the <i>EuUf3gt</i> coding region.....	165
Figure A3.7: Sequence alignment of the <i>EuWDR1</i> coding region.....	167

Figure A3.8: Sequence alignment of the <i>EuWDR2</i> coding region.....	169
Figure A3.9: Alignment of anthocyanin-regulating R2R3-MYB proteins.....	171
Figure A3.10: Sequence alignment of the full-length <i>EuMYB3</i>	173

Acknowledgements

This long journey would not be possible without the help and support of many people throughout these years. I would first and foremost like to thank my advisor, Mark Rausher, for his guidance and patience, and particularly for believing in my abilities and giving me the confidence to face several research challenges.

I would like to thank my committee, Xinnian Dong, Kathleen Donohue, Paul Manos, Tom Mitchell-Olds and John Willis for very useful feedback, discussions and encouragement when I was struggling with failed experiments. I learned how to think broadly and critically about evolutionary questions from each of our meetings.

I would like to thank all the members of the Rausher Lab. I have been extremely lucky to be a member of the Rausher Lab with them. Joanna Rifkin, Talline Martins and Jie Huang were very generous in providing their help and offering their time when I arrived at Duke. The support and companionship of Joanna Rifkin, Irene Liao, Kimmy Stanton, Kate Ostevik and Allan Castillo made my years in graduate school an intriguing and enjoyable experience. Additionally, thank you for making sure we always have cookies or snacks in our office. These are important energy supplies for the long-hour work in the lab.

I am thankful to the Biology Department administration, Anne Lacey, Jo Bernhardt, Jim Tunney and Randy Smith. I really appreciate for their warmth, help and

support. They always solved my problems in a timely fashion.

I thank the greenhouse staff, Michael Barnes, John Mays, Jorge Fidhel Gonzalez and Jenny Gordon, for keeping my plants healthy and watering my plants when I was out of town. Especially, I would like to thank John Mays who shared with me his experience in taking care of the *Clarkia* plants.

I also received help and advice from other labs and outside Duke Biology. I would like to thank Rodolfo Zentella and Jianhong Hu for their help in the *Arabidopsis* transformation experiment. Particularly, Rodolfo kindly provided the *Arabidopsis* seeds, the plant expression vector and the *Agrobacterium* cells, and taught me how to do plant transformation. This experiment would not be possible without his help. I would like to thank George Dubay and Mina Shehata in Duke Chemistry, who helped me solve the issue of drying the anthocyanidin extracts. I would like to thank Fred Nijhout for teaching me how to run HPLC. I would like to thank Emily Austen for providing the samples of *Erythronium americanum* and also for the useful discussions and suggestions. I would like to thank Rancho Santa Ana Botanical Garden for sending the seeds of *Clarkia amoena* ssp. *huntiana*. And I would like to thank Duke Forest, North Carolina Botanical Garden and Eno River State Park for allowing me to do fieldwork and collect samples in these places.

Lastly, I would like to thank my family and friends for understanding and emotional support over these years.

Introduction

Understanding how genetic changes alter the developmental process and generate phenotypic variation has been one of the major goals in evolutionary developmental biology. Floral-color polymorphism exists in a wide range of species. In recent decades, experimental studies of genetic basis of floral-color changes between or within species have been accumulating (e.g., Quattrocchio et al 1999, Streisfeld et al. 2013; Yuan et al. 2013; Wessinger & Rausher 2014a; Hsu et al. 2015; Gates et al. 2018). Several important questions have emerged from these studies including: Are certain types of mutations are disproportionately involved in flower color changes (Streisfeld & Rausher 2010; Wessinger & Rausher 2012)? Have these mutations arisen from new or pre-existing genetic material (e.g., Schwinn et al. 2006; Streisfeld & Rausher 2009a; Martins et al. 2013, 2017)? What role do gene duplications play in the evolution of complex pigment patterns (e.g., Schwinn et al. 2006)?

Streisfeld & Rausher (2010) examined mutations underlying different types of evolutionary transitions in flower color, and found that for the transition from completely pigmented to completely unpigmented (white/yellow) flowers, mutations in R2R3-MYB transcription factors are preferentially fixed. In addition to uniformly pigmented flowers, many flowers have pigmentation patterns that contrast with background coloration, such as venation (stripes associated with veins) or petal spots. In

such types of flowers, *R2R3-MYB* transcription factors are found to be the primary regulators controlling spot or stripe formation in the petal (e.g., Schwinn et al. 2006; Albert et al. 2011; Shang et al. 2011; Martins et al. 2013, 2017; Yuan et al. 2014; Hsu et al. 2015). Overrepresentation of mutations in *R2R3-MYB* genes can be explained by likely variation among anthocyanin genes in the magnitude of pleiotropy caused by mutations. Compared to mutations in anthocyanin enzyme-coding genes, *bHLH* or *WDR* transcription factors, mutations in *R2R3-MYB* genes is often expected to incur relatively little deleterious pleiotropy, because anthocyanin-associated *R2R3-MYB* genes are frequently tissue-specific and also functionally specific in regulating anthocyanin production only (Ramsay & Glover 2005; Quattrocchio et al. 2006; Streisfeld & Rausher 2010). Thus, mutations in *R2R3-MYB* genes usually do not change other physiological or developmental processes or anthocyanin production in other body parts of the plant.

Although these studies have increased our understanding of flower pigmentation, there are still knowledge gaps that need to be bridged with more studies. For example, our knowledge regarding pigmentation of other types of petal patterns, such as patterns involving pigmentation loss in parts of the petal, and pigmentation of other floral, non-petal tissues, such as anthers, remains little.

My dissertation addresses these questions by examining the genetic basis of two floral-color polymorphisms that both involve loss of anthocyanin pigmentation: petal

pigmentation patterning in *Clarkia gracilis* ssp. *sonomensis* (Chapters 1, 2) and an anther-color polymorphism in *Erythronium umbilicatum* (Chapter 3).

Chapter 1 investigates the genetics of petal pigmentation patterning in *C. g.* ssp. *sonomensis*. I particularly use the “white-cup” flower phenotype in this subspecies to determine the type of mutations leading to loss of pigmentation in only the basal region of the petal. I identify the R2R3-MYB transcription factors that regulate anthocyanin production in different regions of the petal, which gives rise to pigmentation patterning. The “white-cup” phenotype is caused by a functional mutation in an *R2R3-MYB* gene that is only expressed in the basal region. These *R2R3-MYB* genes show different spatial and/or temporal expression patterns and seem to be functionally differentiated by activating different anthocyanin enzyme-coding genes. The fact that multiple *R2R3-MYB* genes are required for petal pigmentation suggests gene duplication playing a critical role in the evolution of pigmentation patterning.

Chapter 2 explores the evolution of petal pigmentation patterning. I perform phylogenetic analysis of these *R2R3-MYB* genes from two subspecies of *C. gracilis* and the progenitor species of *C. gracilis*. I also compare the expression domains of these *R2R3-MYB* genes in the four (sub)species. The results reveal that the *R2R3-MYB* genes with region-specific expression patterns are derived from duplication events occurred before polyploidization of *C. gracilis*. After polyploidization that created duplicates of

these petal *R2R3-MYB* genes, these genes have undergone gene loss/silencing, expression domain contracted/moved, and nonfunctionalization. These modifications following duplications contribute to the petal pattern diversification in *C. gracilis*.

Chapter 3 characterizes the genetics of an anther-color polymorphism in *E. umbilicatum*. Based on gene expression differences between purple and yellow anthers and between anther and leaf tissues, I found that downregulation of three enzyme-coding genes and a bHLH transcription factor in the yellow anthers, which is likely controlled by an *R2R3-MYB* gene, is responsible for the polymorphism. However, this *R2R3-MYB* gene is present in multiple, highly similar copies. Despite several attempts, the causal copy to the polymorphism has not been determined successfully. This difficulty comes largely from the limited genomic resources of *E. umbilicatum*. In the near future, when our understanding about the *Erythronium* genome becomes much clearer, it will contribute to the identification of the causal copy, and encourage further investigations on the evolution of the anther-color polymorphism.

1. Genetics of petal pigmentation patterning in *Clarkia gracilis* ssp. *sonomensis* (Onagraceae)

1.1 Introduction

A fundamental goal in evolutionary developmental biology is to understand the types of genetic changes contributing to the morphological diversity that we observe in nature. Identifying these mechanisms would allow us to explore several basic questions about morphological evolution including: are these changes located in a single gene or multiple genes (Shapiro et al. 2004; Hoekstra et al. 2006; Mackay et al. 2009)? Have they arisen from new or pre-existing genetic material (Irish & Litt 2005; Hedrick 2013; Chau & Goodisman 2017; Van de Peer et al. 2017)? And are certain types of the genetic changes disproportionately involved in morphological diversification (Carroll 2005, 2008; Hoekstra & Coyne 2007; Stern & Orgogozo 2008)? Much effort has been devoted to addressing these questions in different organisms or systems. One of the systems that has long fascinated biologists is the variation in color patterns, (Mallet & Joron 1999; Nachman et al. 2003; Whibley et al. 2006; Hopkins & Rausher 2011; Wessinger & Rausher 2014a,b; Nijhout 2017; Stoddard & Hauber 2017). Previous investigations of the evolution of color patterns have demonstrated that *cis*-regulatory modifications of pre-existing developmental programs are often responsible for wing coloration divergence in insects, such as *Drosophila* and butterflies (Gompel et al; 2005; Martin & Reed 2010; Reed et al 2011; Martin et al. 2012; Martin & Reed 2014; Koshikawa et al. 2015; Wallbank

et al. 2016; Martin & Courtier-Orgogozo 2017) and floral color differences in plants (see below).

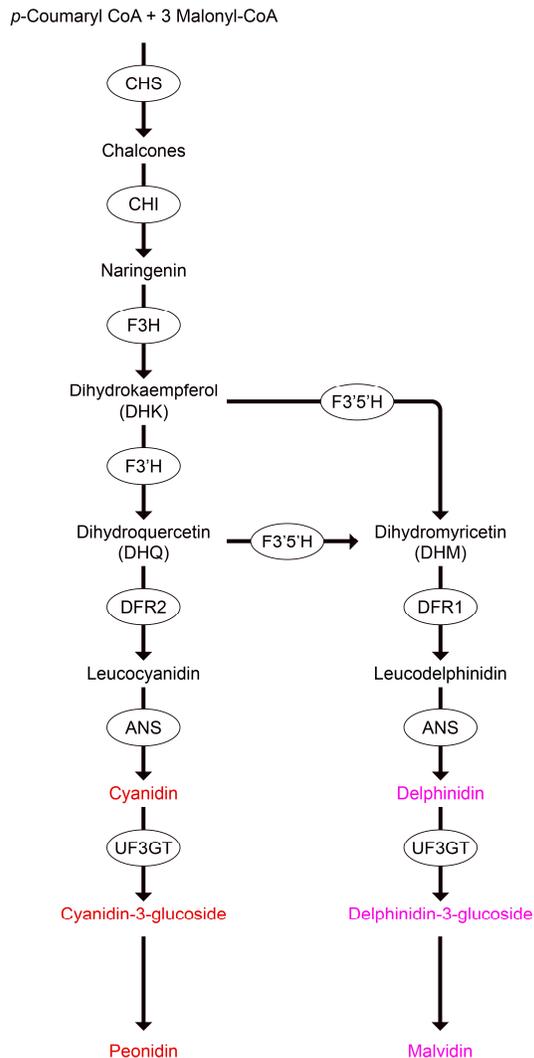


Figure 1: A simplified schematic diagram of the anthocyanin biosynthetic pathway. Enzymes are shown in circles: CHS, chalcone synthase; CHI, chalcone isomerase; F3H, F3H, flavanone-3-hydroxylase; F3'H, flavonoid 3'-hydroxylase; F3'5'H, flavonoid 3',5'-hydroxylase; DFR, dihydroflavonol-4-reductase; ANS, anthocyanidin synthase; UF3GT, UDP-flavonoid-3-O-glucosyl-transferase.

In plants, flower color is evolutionarily labile. The anthocyanin biosynthetic pathway (Figure 1) responsible for much floral pigmentation has been extensively studied and is highly conserved across angiosperms (Holton & Cornish 1995), facilitating the effort to characterize the genetic changes underlying flower-color evolution. Most of the flower-color transitions for which genetic changes have been characterized involve transitions between different colors of completely pigmented flowers (e.g., from blue to red flowers) or completely pigmented flowers to flowers that completely lack pigments (e.g., from blue/red to white/yellow flowers). *Cis*-regulatory changes are frequently associated with these flower color transitions (Streisfeld & Rausher 2009a; Des Marais & Rausher 2010; Dick et al. 2011; Hopkins & Rausher 2011; Streisfeld et al. 2013; Yuan et al. 2013; Wessinger & Rausher 2014a; Gates et al. 2018), while in other cases, coding mutations that abolish the enzyme activity are causal (Habu et al 1998; Quattrocchio et al 1999; Coberly & Rausher 2003; Chang et al. 2005; Smith & Rausher 2011; Wu et al. 2013; Wessinger & Rausher 2014a; Coburn et al. 2015). Although coding and *cis*-regulatory mutations both could be involved in flower color differences, substitutions seem to target different genes, depending on the transition type. The between-species transition to white/yellow (unpigmented) flowers tends to involve mutations (both coding and *cis*-regulatory) in transcription factors, usually anthocyanin-regulating R2R3-MYB proteins, but sometimes also R3-MYB repressors (Quattrocchio et

al 1999; Schwinn et al. 2006; Whittall et al. 2006; Streisfeld et al. 2013; Yuan et al. 2013; Gates et al. 2018). By contrast, the transition from blue to red flowers tends to involve mutations in the enzyme-coding genes (Streisfeld & Rausher 2009a, 2010; Des Marais & Rausher 2010; Smith & Rausher 2011; Wessinger & Rausher 2012, 2014a).

In addition to the plant species that have uniformly colored corollas, many plants have flowers with contrasting color elements that create particular patterns. To date, genetic evaluation of floral pigmentation patterning has focused on venation (pigment stripes associated with the veins) (Schwinn et al. 2006; Albert et al. 2011; Shang et al. 2011; Hsu et al. 2015) and spot formation (Chiou & Yeh 2008; Martins et al. 2013, 2017; Yuan et al. 2014; Hsu et al. 2015; Yamagishi et al. 2010, 2014, 2018; Yamagishi 2018). The results of these studies show a striking similarity: *R2R3-MYB* transcription factors are important in regulating these patterns. On the other hand, little information is available about genetic changes underlying the color patterns involving pigmentation loss in parts of the corolla. Additional studies on different types of patterning will expand our understanding about the genetics of flower pigmentation and will allow us to examine whether the *R2R3-MYB* genes are always the major determinants in pattern formation. Moreover, the fact that multiple *R2R3-MYB* genes expressed in the petal to control background coloration and stripe or spot formation (Schwinn et al. 2006; Albert et al. 2011; Hsu et al. 2015) suggests that gene duplication followed by spatial

subfunctionalization may have contributed to the evolution of these color pattern elements. And again, the extent to which gene duplication has contributed to the evolution of other types of pattern element, such as floral sectors that lack pigmentation, is currently unknown.

In this study, we present the results of experiments that address these issues by examining the biochemical, genetic and developmental basis of loss of pigmentation in a sector of the flower of *Clarkia gracilis* ssp. *sonomensis* (Onagraceae), a distinct pattern element termed “cup” (Figure 2; Gottlieb & Ford 1988). The great diversity in floral pigmentation patterns in *Clarkia* has facilitated examination of the evolution of how floral color patterns affect interactions with pollinators (Eckhart et al. 2006; Eisen & Geber 2018; Jones 1996a,b). This variation constitutes a resource for examining the types of genetic changes responsible for evolutionary changes in these pattern elements. However, between-species crosses in *Clarkia* are often problematic, and thus it is usually not possible to identify the genetic differences associated with the target traits using mapping populations generated by interspecific crosses. Fortunately, *C. g.* ssp. *sonomensis* provides a good opportunity for such a study, because the *C. g.* ssp. *sonomensis* flowers are polymorphic for the both presence of a red spot (Jones 1996a,b) and a white cup (Gottlieb & Ford 1988); in addition, this subspecies can be crossed successfully with other *C. gracilis* subspecies.



Figure 2: Flowers of *Clarkia gracilis* ssp. *sonomensis*. (A) The central-spotted, white-cupped and unspotted, pink-cupped parental plants used in the crosses. (B) The dissection sections of a white-cupped petal.

Absence of anthocyanin pigmentation may result from functional or regulatory mutations in any of the genes associated with the anthocyanin biosynthetic pathway. Loss of pigmentation in only a part of the petal means that while anthocyanin production is eliminated in the unpigmented region, its production is unaffected in the rest of the petal. Here, we propose possible scenarios that could explain loss of pigmentation in the cup region only in the *C. g.* ssp. *sonomensis* petal: (1) mutations disrupt enzyme function: there are two copies of at least one anthocyanin enzyme-coding gene, one expressed in the background and the other expressed in the cup. Functional mutations in the cup copy change the enzyme activity and eliminate anthocyanin production; (2) mutations reduce the expression of enzyme-coding genes: there are multiple, discrete *cis*-regulatory elements (for example, enhancers) in the enzyme-coding genes, and different elements are used in different petal regions by different transcriptional regulators. *Cis*-regulatory mutations in one or more elements

that are specific for cup pigmentation reduce the expression of enzyme-coding genes in the cup, leading to decreased production of anthocyanins; (3) mutations disrupt the function of a transcription factor: functional mutations in a transcription factor that is only expressed in the cup destroy its ability to activate the enzyme-coding genes, leading to decreased production of anthocyanins; (4) mutations reduce the expression of a transcription factor: *cis*-regulatory mutations in a transcription factor or mutations in one of its activators, diminish the expression of the transcription factor and also the enzyme-coding genes in the cup, causing decreased production of anthocyanins.

The goal of the present study was to determine which of the above scenarios can explain the white cup formation in *C. g. ssp. sonomensis*. To address this goal, we identified the enzyme-coding genes and associated transcription factors in the anthocyanin biosynthetic pathway that are expressed in the pigmented and unpigmented regions of the petal. We then addressed the following questions:

- (1) What are the genetic changes responsible for the presence of a white cup?
- (2) Does the background color development differ from the cup color development? If so, are there different genes controlling background and cup pigmentation?

1.2 Materials and Methods

1.2.1 Study system

Clarkia gracilis is composed of four subspecies, and is native to northern California. It is the only polyploid in the section *Rhodanthos*. This allotetraploid species is thought to be derived from two diploid species: *C. amoena* and an extinct species related to *C. lassenensis* and *C. arcuata* (Abdel-Hameed & Snow 1968, 1972). All four subspecies have petals with a pink background, but differ with respect to the presence and position of red petal spots. *C. g. ssp. sonomensis* is the only subspecies that has central spots (although unspotted *C. g. ssp. sonomensis* are occasionally observed). Two other subspecies, *C. g. ssp. albicaulis* and *C. g. ssp. tracyi*, have basal spots. The fourth subspecies, *C. g. ssp. gracilis*, is highly selfing and lacks petal spots (Martins et al. 2013). Spotted *C. g. ssp. sonomensis* plants were shown to have reproductive advantage over the unspotted plants, as mediated by pollinator preference (Jones 1996a,b). The color difference between the petal background and the spot is caused by accumulating different anthocyanin pigments. The pink background contains malvidin-derived anthocyanins and the red spot contains cyanidin-, malvidin-, and peonidin-derived anthocyanins (Soltis 1986; Martins et al. 2013). The presence or absence of the central spot is determined by a single locus (Gottlieb & Ford 1988). Recent investigations have identified this locus as an R2R3-MYB transcription factor (*CgMYB1*) controlling the

enzyme-coding genes in the anthocyanin biosynthetic pathway. The shift in spot position between subspecies has been shown to be due to a *cis*-regulatory change in *CgMYB1* (Martins et al. 2017).

Clarkia g. ssp. sonomensis has one additional character that is also polymorphic: the presence of “white cup”. In the white cup phenotype, the basal region of each petal lacks pigmentation and is white (Figure 2). Gottlieb & Ford (1988) concluded that the presence/absence of cup pigmentation is controlled by a single locus, with the “white” cup allele being recessive. However, the identity of the gene underlying this phenotypic variation remains unknown.

1.2.2 Plant growth and crosses

The seeds of *Clarkia gracilis ssp. sonomensis* (C. L. Hitchc.) H. Lewis & M. Lewis used as parental plants in this study were kindly provided by Talline Martins (University of Florida) (Appendix I, Table A1.1). Seeds were germinated by placing them inside of the folds in the damp paper towels. The paper towels were then placed in the sandwich zip bags to minimize the moisture loss. The bags were placed in the growth room at 15-18°C in the dark. When the radicles appeared, the seeds were transferred to wet Fafred 4P soil (Sun Gro Horticulture) in the 3.1” x 2.2” x 2.3” cells. Germinated seeds were grown in the growth room with 16-hour day length. After 3-4

weeks, seedlings were transferred to 5-inch pots and grown in the Duke greenhouse (20-24 °C).

We performed crosses to characterize the genetic control of white cup segregation. A central-spotted, white-cupped plant was crossed to an unspotted, pink-cupped plant (Figure 2A). The anthers of the pollen recipient flowers were removed several days before the stigmas became receptive to avoid self-pollination. The emasculated flowers were pollinated by touching the stigmas with dehiscing anthers from the desire sire. A single F₁ individual was selfed to generate an F₂ population.

1.2.3 Pigment identification

We characterized the pigments in different color regions of the F₂ flowers (Appendix I, Figure A1.1) using high performance liquid chromatography (HPLC). Anthocyanidins, the aglycone precursors of anthocyanins, were extracted following the method in Harborne (1984) with modifications. Ten flowers from a single plant were used for each flower type. Each petal was dissected into two or three parts, depending on whether the central spot is present or not: (1) the top (background); (2) the cup; (3) the central spot (Figure 2B).

Dissected sections were soaked in 1 mL 2N HCl overnight. The 2N HCl supernatants were heated in a boiling bath for 60 min. The cooled 2N HCl extracts were washed twice with 500 uL ethyl acetate. The ethyl acetate layer and aqueous layer were

separated by centrifugation at 13,000 rpm for 1 min. The ethyl acetate supernatants were removed and evaporated under vacuum, and the remaining aqueous layer was washed once with 150 μ L isoamyl alcohol. After centrifugation, the isoamyl alcohol layer that is enriched with anthocyanidins was collected. The isoamyl alcohol supernatants were dried using a rotary evaporator, resuspended in 100 μ L methanol with 1% (v/v) HCl, and stored in a -20 °C freezer. On the next day, 50 μ L of the elution was injected on a Shimadzu LC-10AT liquid chromatograph with a 4.6 x 150 mm Alltech Prevail reverse phase C18 column (Alltech Associates, Deerfield, IL) at a flow rate of 1 mL/min.

Anthocyanidins were separated by gradient elution at 30 °C using solvents A (HPLC-grad water, 0.1% trifluoroacetic acid) and C (1-propanol, 0.1% trifluoroacetic acid) with the following program: 15% C from 0 to 4 min; linear increase to 20% C from 4 to 10 min; 20% C from 10 to 14 min; linear increase to 22.5% C from 14 to 16 min; instantaneous increase to 27.5% C; 27.5% C from 16 to 18 min; instantaneous decrease to 15% C; 15% C from 18 to 21 min. Peaks were detected at 520 and 540 nm. Anthocyanidins were identified by comparison with standard solutions of delphinidin, petunidin and peonidin from Polyphenols Laboratories (Sandnes, Norway), and cyanidin, malvidin, and pelargonidin from Indofine Chemical Company (Hillsborough, NJ, USA).

1.2.4 Transcriptome sequencing

RNA sequencing was performed for the pink background (top) and white cup (cup) regions of the petal from the white-cupped plants. These two regions were dissected from the flower-buds that were collected when the buds just became erect (approximately 1 day before flowering). Total RNA was extracted from the dissected regions from four plants using Spectrum Plant Total RNA Kit (Sigma). The top and cup RNA samples were prepared by pooling equal amounts of RNA from each plant. The pooled RNA was run with Bioanalyzer Agilent RNA 6000 Nano Kit (Agilent Technologies) to ensure that all samples had good-quality RNA for library construction. We used KAPA Stranded mRNA-Seq Kit (KAPA Biosystems) with 4 and 3 µg of the pooled top and cup RNA, respectively, to construct libraries of each petal region. The libraries were barcoded using NEBNext Multiple Oligos for Illumina (New England BioLabs). The quality of the resulting libraries was examined with Bioanalyzer Agilent High Sensitivity DNA Kit (Agilent Technologies). Equal amounts of barcoded libraries were pooled and sequenced on an Illumina HiSeq 4000 platform performing 150 bp paired-end reads at the Duke University Sequencing & Genomic Technologies Shared Resource Center.

1.2.5 Bioinformatic analyses

Raw Illumina reads were first trimmed to remove adaptor and barcode sequences using Trimmomatic 0.36 (Bolger et al. 2014) bundled in the Trinity 2.4.0 software package (Grabherr et al. 2011). Only the trimmed, paired-end reads were used in transcriptome assembling. The transcriptome references were *de novo* assembled using Trinity. To search for the anthocyanin gene sequences in the resulting transcriptomes, we ran the tblastx program in NCBI-BLAST-2.6.0+ (<ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/2.6.0/>; Camacho et al. 2009), using the sequences from *Arabidopsis thaliana* and *Punica granatum* as queries (Appendix I, Table A1.2). Gene expression levels were estimated as FPKM values (the number of RNAseq fragments per kilobase of transcript effective length per million fragments mapped to all transcripts) using RSEM 1.3.0 (Li & Dewey 2011) built into the Trinity package. Trimmed, paired-end reads of the pink background (top) and white cup (cup) samples were mapped separately to the transcripts in the transcriptome reference of the pink background sample to generate their corresponding RSEM estimates of transcript abundance. The estimates were then normalized with TMM (trimmed mean of M-values) normalization (Robinson & Oshlack 2010) using edgeR (Robinson et al. 2010) included in the Trinity scripts to generate the expression values measured as FPKM. The same procedure was repeated for mapping the reads of the top and cup samples separately to

the white cup transcriptome reference. Scripts used to run these analyses are listed in Appendix I, A1.1. The bitscores of BLAST hits and FPKM values were both considered when selecting the subjects (reference sequences).

1.2.6 Cloning of full-length coding sequences

From the transcriptomes of the two color regions in the petal, we identified eight anthocyanin enzyme-coding genes (*CgsChs*, *CgsChi*, *CgsF3h*, *CgsF3'h*, *CgsF3'5'h*, *CgsDfr1*, *CgsAns* and *CgsUf3gt*), three *R2R3-MYB* genes of subgroup 6 (*CgsMYB6*, *CgsMYB11* and *CgsMYB12*), two *bHLH* genes and two *WDR* genes (Table 1).

To determine in which petal regions these genes were expressed, the full-length coding sequences of the enzyme-coding and *R2R3-MYB* genes were amplified with the cDNA samples from the top and cup regions of the white-cupped parental plant used to create the F₂ progeny. cDNA of each sample was synthesized with total RNA extracted as described above. Before making cDNA, genomic DNA (gDNA) was first removed from RNA using RQ1 RNase-Free DNase (Promega). DNase-treated RNA (0.5 µg) was used to synthesize cDNA in a 20 µL reaction with 200U of M-MuLV Reverse Transcriptase (New England BioLabs) and 1µM of Oligo d(T)₁₈ mRNA Primer (New England BioLabs). PCR primers were designed based on the reference sequences retrieved from the transcriptomes, and listed in Appendix I, Table A1.3. PCR reactions were conducted using Q5 High-Fidelity DNA Polymerase (New England BioLabs) with

the touchdown PCR program: denaturation at 98 °C for 30 sec, followed by 20 cycles of 98 °C for 10 sec, 68-48 °C for 30 sec (decreasing the annealing temperature by 1 °C per cycle), and 72 °C for 1 min, and 20 cycles of 98 °C for 10 sec, 48 °C for 30 sec, and 72 °C for 1 min, and final extension at 72 °C for 2 min. PCR products were gel-purified using QIAquick Gel Extraction Kit (Qiagen), phosphorylated using T4 Polynucleotide Kinase (New England BioLabs), ligated with the pCR-Blunt vector (Invitrogen), and transformed into 5-alpha *Escherichia coli* cells (New England BioLabs). At least five colonies per ligation were sequenced by Sanger Sequencing (Eton Bioscience, San Diego, CA). SEQUENCHER 5.0 (Gene Codes, Ann Arbor, MI) was used to correct basecalling errors and align sequence fragments.

To evaluate whether the identified *R2R3-MYB* genes are homologs to the known anthocyanin regulators, we performed phylogenetic analyses to construct the phylogenetic relationship of these petal *R2R3-MYB* transcription factors and the related *R2R3-MYB* proteins from other species. The coding sequences of the petal *R2R3-MYB* genes were translated into amino acid sequence, and aligned with other *R2R3-MYB* protein sequences using MUSCLE 3.8 (Edgar 2004). A neighbor-joining phylogenetic tree was constructed using MEGA 6.06 (Tamura et al. 2013) with the JTT amino acid substitution model. Clade support was estimated by 1000 bootstrap replicates.

1.2.7 Quantification of gene expression

The expression levels of eight enzyme-coding genes and three *R2R3-MYB* genes in the top (pink background) and cup (white cup) regions were further analyzed using quantitative real-time PCR (qPCR). cDNA samples of the two regions (collected approximately 1 day before flowering) were prepared as described above. cDNA samples were then diluted to 2.5 ng/μL for qPCR. The *bHLH* and *WDR* genes were not examined because our analyses of the other genes allowed us to identify the cause of lack of pigmentation in the white cup.

The qPCR primers used are listed in (Appendix I, Table A1.3). Primer specificity was examined by visualizing the PCR products on 2% agarose gels and confirmed by sequencing the PCR products. Each 20 μL qPCR reaction contained 10 μL of DyNAmo HS SYBR Green qPCR master mix (Thermo Scientific), 0.3 μM of each primer and 2.5 ng cDNA template. Five biological replicates for each petal region and two technical replicates for each sample were performed. Reactions were run on a Roche LightCycler 96 with the following conditions: 95°C for 15 min, followed by 55 cycles of 95°C, 15 sec, 60°C for 30 sec, and 72°C for 30 sec. A melting curve analysis was run at the end of the cycle to verify that a single product was amplified. The reactions were repeated if the threshold (Ct) values of technical replicates differed by greater than 10%. A cDNA sample from the pink background was arbitrarily chosen as a control sample and was

included in each run. PCR efficiency of each gene was calculated according to Peirson et al. (2003). The relative expression ratios of target genes were normalized with the expression levels of *GAPDH* (glyceraldehyde 3-phosphate dehydrogenase) (reference gene selection; Appendix I, A1.2), using equation 1 in Pfaffl (2001). Then the relative expression levels were calculated as the logarithm of the ratios.

1.2.8 Genotyping F₂ plants at *CgsMYB12*

CgsMYB12 was found only expressed in the cup, and therefore, the coding sequences of *CgsMYB12* from the pink-cupped and white-cupped parents were cloned. We then found one 1-bp deletion in the exon 3 of *CgsMYB12* from the white-cupped parent, which generates a premature stop codon (see below). This deletion presumably eliminates the function of *CgsMYB12*. To determine whether this deletion cosegregated with the cup color, we genotyped the parents, an F₁, and 40 F₂ plants (eight from each of the five flower types) to detect the deletion in these plants, using the polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP) method (Ota et al. 2007).

A *CgsMYB12* fragment covering a portion of intron 2 and the whole exon 3 was amplified with the primers cMYB12-7F and cMYB12-4R (Appendix I, Table A1.3) using gDNA as the templates. gDNA was extracted from the petal or leaf tissue using the cetyltrimethylammonium bromide (CTAB) protocol (Doyle & Doyle 1987). PCR products were digested with the restriction enzyme *BbvI* (New England BioLabs). This

enzyme cuts the fragments without the deletion (from the pink-cupped plants) twice, but cuts the fragments with the deletion (from the white-cupped plants) only once (Appendix I, Figure A1.3), which enables distinguishing the genotypes by the lengths of digested fragments. Digested PCR fragments were visualized on 2.5% agarose gel.

1.2.9 Semi-quantitative assessment of gene expression across flower-bud developmental stages

We found that at least two of the *R2R3-MYB* genes were present in the same region of the *C. g. ssp. sonomensis* petal. This raised questions about the transcriptional control of these *R2R3-MYB* genes. For example, are they functionally redundant? Are they expressed at different times? Do they activate different sets of the enzyme-coding genes? To answer these questions, we examined the expression patterns of the *R2R3-MYB* genes and the enzyme-coding genes at different stages during the flower-bud development.

Flower buds were collected from bud at different sizes up to 1 day before flowering, and were categorized into the following stages: (1) < 5 mm, white petal; (2) < 10 mm, white petal with the central spot; (3) > 10 mm, white petal. The petal was dissected into white background and white cup; (4) > 15 mm, background color appeared. The petal was dissected into pink background and white cup; (5) ~ 20 mm, cup color appeared. The petal was dissected into pink background and pink cup. The last stage corresponds to the stage of 1 day before flowering. Three flower types in the F₂

population were used: central-spotted, pink-cupped (Type I), central-spotted, white-cupped (Type III) and unspotted, pink-cupped (Type IV). The buds were collected from three plants each from the three types. Total RNA and cDNA samples were prepared as described above. PCR reactions were conducted using *Taq* DNA Polymerase (New England BioLabs) with the primers used in the qPCR assays. PCR products were visualized on 2% agarose gels. The brightness of PCR bands reflects the expression levels of the tested genes, and was scored as expressed (“+”), weakly expressed (“(+)”) or not expressed (blank).

1.2.10 Overexpression of *C. g. ssp. sonomensis* petal *R2R3-MYB* genes in *Arabidopsis*

The functionality of the petal *R2R3-MYB* genes was examined through stable transformation of *Arabidopsis thaliana*. We generated the transgenic *Arabidopsis* to overexpress *CgsMYB6*, *CgsMYB11*, the functional *CgsMYB12* (from the pink-cupped parent), the nonfunctional *CgsMYB12* (*CgsMYB12W*, from the white-cupped parent) and the *Arabidopsis pap1-D* (the dominant allele of *AtMYB75*, serving as a positive control; Borevitz et al. 2000). If these *R2R3-MYBs* participate the anthocyanin production, we should observe the elevated expression levels of the anthocyanin enzyme-coding genes in the transgenic plants, compared to that in the wild-type (untreated) plants. If not, we would observed the similar expression levels of the enzyme-coding genes between the transgenic plants and the wild-type.

The full-length coding sequences of these *R2R3-MYB* genes were amplified using Q5 High-Fidelity DNA Polymerase with specifically designed primers (Appendix I, Table A1.3) to engineer the *Bam*H1 and *Xho*1 restriction sites at 5' and 3' ends, respectively. The PCR products were digested, gel-purified and cloned into the Gateway entry vector pENTR2B (Invitrogen). The LR recombination reactions using LR Clonase II Enzyme Mix (Invitrogen) were used to transfer the inserted genes from the entry vector to the plant overexpression vector pGWB418 (Nakagawa et al. 2007), where the inserted genes were under the control of a constitutive CaMV (cauliflower mosaic virus) 35S promoter. All clones were sequenced at every step to guarantee the correct sequences and orientation of the inserts. The constructs were then introduced into *Agrobacterium tumefaciens* strain GV3101 by electroporation, and transformed into *Arabidopsis thaliana* Col-0 using the floral-dip method (Clough & Bent 1998). Seeds of the treated plants were collected, sterilized with 20% bleach and 0.1% SDS, vernalized at 4 °C for 3 days in the dark, and screened on 1/2 MS (Murashige and Skoog) medium containing 0.05% MES, 3% sucrose, 50 µg/mL kanamycin and 200 µg/mL cefotaxime. Transformants were selected and then transferred to soil for growth to maturity. After two generations of selection, homozygous T3 plants with a single insertion were obtained. Total RNA of each transgenic line was extracted from the pooled 5-day-old T3 plants using Spectrum Plant Total RNA Kit, and was used for cDNA synthesis as described above. PCRs were

performed to confirm the presence of the inserted genes. For each construct, three independent lines with high expression levels of the inserted genes were selected to examine the expressions of three enzyme-coding genes (*AtChs*, *AtDfr* and *AtAns*) with the normalization with *AtGAPDH*, following the qPCR procedure described above. Differences in gene expressions between the treatments were analyzed with one-way analyses of variance (ANOVA) using JMP 7 (SAS Institute, Cary, NC). Multiple comparisons for all pairs were performed using Tukey-Kramer honestly significant difference (HSD) test and Student's *t*-test.

1.3 Results

1.3.1 Cup color is controlled by a single locus

We created F₂ progeny by crossing a central-spotted, white-cupped to an unspotted, pink-cupped *C. g. ssp. sonomensis*. No white cup is observed in the F₁ plants, indicating the white cup allele is recessive. An F₁ plant was selfed to create a F₂ population ($N = 204$). Assuming the locus controlling spot position is different from the locus that determines the cup color, six phenotypes are expected in F₂ generation: (I) central-spotted, pink-cupped; (Ia) central- and basal-spotted, pink-cupped; (II) central- and basal-spotted, white-cupped, (III) central-spotted, white-cupped; (IV) basal-spot, pink-cupped; and (V) basal-spotted, white-cupped. However, because phenotypes I and Ia were difficult to distinguish, we pooled them into a single phenotype I, producing the

five phenotypic categories shown in Appendix I, Figure A1.1. The number of individuals having ($N = 45$) versus not having ($N = 159$) the white cup does not deviate from the 1:3 ratio ($\chi^2_{(1, N=204)} = 0.941, P = 0.32$), suggesting that the white cup formation involves a single locus, which is consistent with Gottlieb & Ford (1988). Segregation of spot position is also consistent with single-locus variation: the numbers of phenotypes I, II, III, and IV+V ($N = 110, 20, 12$ and 62 , respectively) do not differ significantly from the expected ratio 9:2:1:4 ($\chi^2_{(3, N=204)} = 3.800, P = 0.28$). This result is consistent with that reported by Martins et al. (2013).

1.3.2 Different color regions contain different types of anthocyanins

Colors of the *C. g. ssp. sonomensis* petal are due to accumulation of different anthocyanins in different regions. The pink background has only malvidin-derived anthocyanins. The central red spot has cyanidin-, malvidin- and peonidin-derived anthocyanins, where cyanidin- and peonidin-derived anthocyanins are present in larger amounts, compared to malvidin-derived anthocyanins. These results are consistent with Martins et al. (2013). The anthocyanidin composition in the cup region varies among different cup phenotypes in the F_2 flowers (Figure 3). Phenotypes II and V, which are white-cupped and basal-spotted, produce cyanidin and peonidin in approximately equal amounts, and substantially less malvidin. Phenotypes I and IV, which are pink-cupped and basal-spotted, produce mostly malvidin, but small amounts of cyanidin and

peonidin, which are presumably produced by contamination of the cup sample with parts of the basal spot. Finally, phenotype III, which is purely white-cupped, produces no detectable anthocyanidins. This analysis indicates that the pink cup region produces the same anthocyanidins as the petal background.

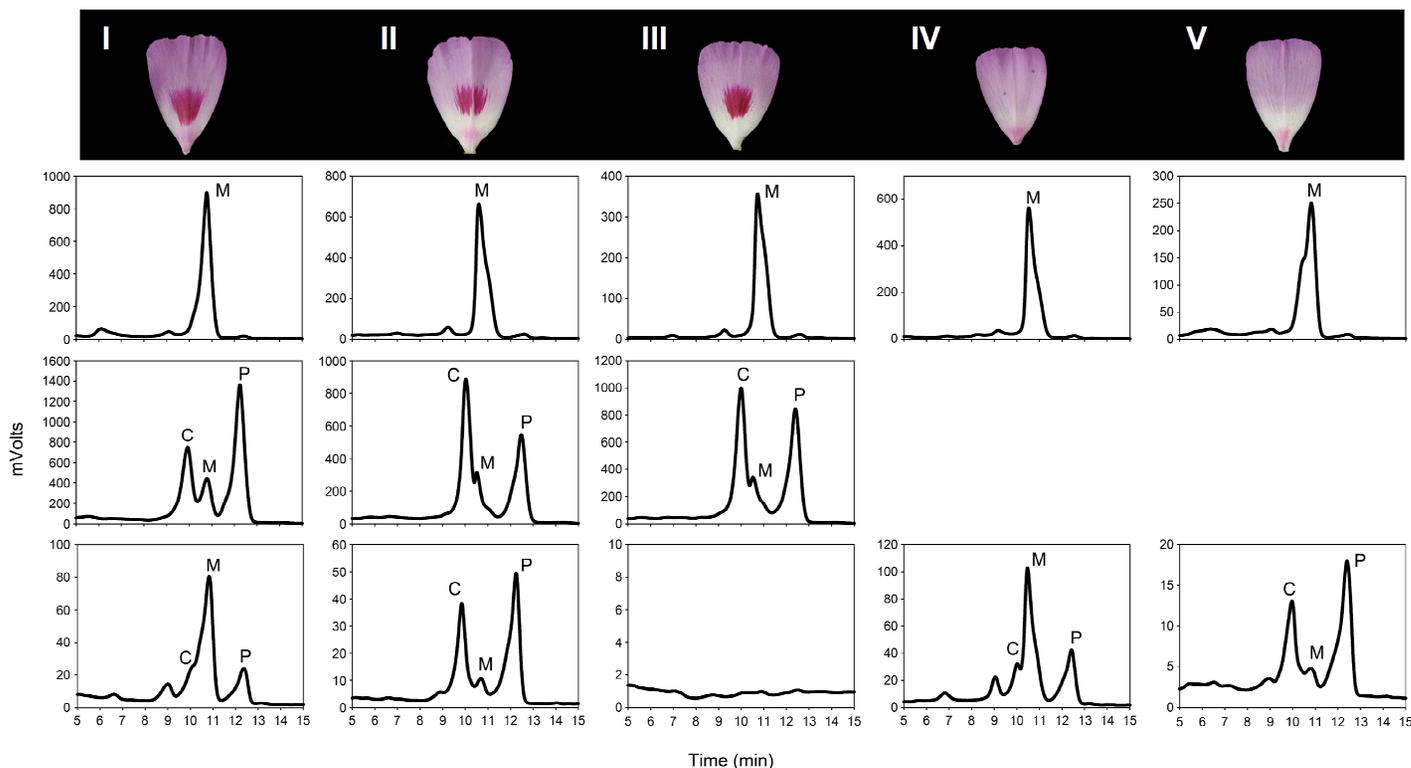


Figure 3: Anthocyanidins in the different color regions in the F₂ flowers. Each column representing a different F₂ phenotype (I-V; see Appendix I, Figure A1.1) has three or two panels showing HPLC traces of the dissected regions: top, pink background (phenotypes I-V); middle, red spot (phenotypes I-III) or no data for unspotted flowers (IV, V); bottom, pink cup (phenotypes I, IV), white cup with basal spot (phenotypes II, V) and pure white cup (phenotype III). Peaks correspond to cyanidin (C), malvidin (M), and peonidin (P).

1.3.3 Anthocyanin enzyme-coding genes are downregulated in white cup

BLAST searches against the transcriptomes of pink background and white cup identified eight anthocyanin enzyme-coding genes (*CgsChs*, *CgsChi*, *CgsF3h*, *CgsF3'h*, *CgsF3'5'h*, *CgsDfr*, *CgsAns* and *CgsUf3gt*) (Table 1). Four out of the eight genes have more than one copy identified (*CgsChs*, *CgsF3h*, *CgsDfr* and *CgsUf3gt*), which may not be surprising since *C. gracilis* is a tetraploid and multiple copies of some anthocyanin enzyme-coding genes in *C. gracilis* have been reported (Martins et al. 2013). In these cases, we chose the copies having the highest FPKM values in the pink background transcriptome as the candidate genes, given the conspicuous differences in the FPKM values among the copies. Because there are at least six-fold differences in the FPKM values, we concluded that the other copies with lower FPKM values probably do not play the primary roles in anthocyanin biosynthesis.

Table 1: Anthocyanin genes identified from the transcriptomes through BLASTing and transcript abundance estimation. Each contig represents a different gene, which has several isoforms determined by Trinity. *The BLAST bitscore of each contig is shown as the highest bitscore among its isoforms. Gene expression levels were estimated as FPKM values. The contigs highlighted in blue are chosen as candidate genes in this study.

(TOP) Pink background transcriptome

Contig ID	Gene	BLAST bitscore*	FPKM	
			Pink background (TOP)	White cup (CUP)
RCL4_23897_c0_g1	<i>CgsChs</i>	815	2249.45	631.2
RCL4_40972_c0_g1	<i>CgsChsA</i>	263	0.7	0.3
RCL4_32248_c0_g1	<i>CgsChsB</i>	228	0.57	0
RCL4_17557_c0_g1	<i>CgsChi</i>	374	326.28	13.32

RCL4_18909_c0_g2	<i>CgsF3h</i>	687	119.76	19.95
RCL4_19013_c0_g1	<i>CgsF3hA</i>	632	18.22	18.35
RCL4_30788_c0_g1	<i>CgsF3hB</i>	305	0.56	0.97
RCL4_16647_c0_g1	<i>CgsF3'h</i>	420	7.87	6.56
RCL4_23202_c0_g1	<i>CgsF3'5'h</i>	520	1401.29	150.85
RCL4_20983_c0_g1	<i>CgsDfr1</i>	660	3166.85	119.53
RCL4_23223_c0_g1	<i>CgsAns</i>	645	2133.83	75.29
RCL4_17638_c0_g1	<i>CgsUf3gt</i>	185	204.48	66.46
RCL4_16934_c1_g2	<i>CgsUf3gtA</i>	167	23.58	15.09
RCL4_17638_c0_g2	<i>CgsUf3gtB</i>	162	2.67	2.87
RCL4_8923_c0_g1	<i>CgsUf3gtC</i>	151	2.45	0.35
RCL4_22269_c1_g1	<i>CgsMYB6</i>	211	617.93	70.76
RCL4_17423_c1_g1	<i>CgsMYB11</i>	213	74.74	2.96
RCL4_23191_c0_g2	<i>CgsbHLH1</i>	492	49.42	15.23
RCL4_20882_c1_g1	<i>CgsbHLH2</i>	237	6.91	6.98
RCL4_16941_c0_g1	<i>CgsWDR1</i>	758	12.41	8.65
RCL4_22922_c0_g1	<i>CgsWDR2</i>	357	11.43	13.52

(CUP) White cup transcriptome

Contig ID	Gene	BLAST bitscore*	FPKM	
			Pink background (TOP)	White cup (CUP)
RCL5_24819_c0_g1	<i>CgsChs</i>	813	2291.01	457.46
RCL5_4024_c0_g1	<i>CgsChsC</i>	204	0	1.96
RCL5_6851_c0_g1	<i>CgsChsD</i>	279	37.25	0.68
RCL5_19109_c0_g1	<i>CgsChi</i>	377	560.54	17.09
RCL5_20018_c0_g3	<i>CgsF3h</i>	691	135.66	15.95
RCL5_20018_c0_g2	<i>CgsF3hA</i>	497	16.59	12.2
RCL5_39396_c0_g1	<i>CgsF3hB</i>	243	0.58	1.03
RCL5_17225_c0_g1	<i>CgsF3'h</i>	420	6.09	3.93
RCL5_25379_c1_g1	<i>CgsF3'5'h</i>	519	2337.86	171.63
RCL5_20386_c0_g2	<i>CgsDfr1</i>	657	4156.96	111.14
RCL5_20386_c0_g1	<i>CgsDfr3</i>	508	1.6	343.3
RCL5_2066_c0_g1	<i>CgsDfrA</i>	325	0.12	0.76
RCL5_32321_c0_g1	<i>CgsDfrB</i>	262	0.47	0.83
RCL5_20731_c0_g2	<i>CgsAns</i>	647	1996.33	53.59
RCL5_16732_c0_g1	<i>CgsUf3gt</i>	185	249.14	56.3
RCL5_18554_c1_g2	<i>CgsUf3gtA</i>	167	27.77	12.63
RCL5_9656_c0_g1	<i>CgsUf3gtB</i>	169	1.91	6.08

RCL5_25197_c1_g1	<i>CgsMYB6</i>	211	979.66	76.49
RCL5_17253_c1_g1	<i>CgsMYB12</i>	204	18.27	88.47
RCL5_23692_c0_g2	<i>CgsbHLH1</i>	492	57.72	12.41
RCL5_21600_c0_g1	<i>CgsbHLH2</i>	231	7.74	6.47
RCL5_16829_c0_g1	<i>CgsWDR1</i>	758	11.74	8.66
RCL5_21338_c0_g1	<i>CgsWDR2</i>	357	13.1	12.82

The *CgsDfr* we identified in the pink background transcriptome is *CgsDfr1*, which has been proposed to involve in petal background pigmentation, in contrast to *CgsDfr2* involving in spot pigmentation (Martins et al. 2013). We also found that there is another *CgsDfr* copy in the white cup transcriptome, which is likely *CgsDfr3* mentioned in Martins et al. 2013. This *CgsDfr* copy is highly expressed in the white cup but not in the pink background. Lack of positive correlation between the *CgsDfr3* expression and the anthocyanin amounts suggests that *CgsDfr3* may not contribute to anthocyanin production. We thus excluded *CgsDfr3* from further analyses.

We cloned and sequenced the full-length coding regions of eight enzyme-coding genes amplified with cDNA samples of pink background and white cup, which revealed identical sequences between the two petal regions. Our results suggest that the same copies of the enzyme-coding genes were expressed in these two regions. However, the qPCR results showed that the enzyme-coding genes are downregulated in the white cup (Student's *t*-test, $P < 0.05$; Figure 4). Coordinate downregulation of all enzyme-coding genes suggests that one or more sets of transcription factors are involved in regulating the expression of enzyme-coding genes.

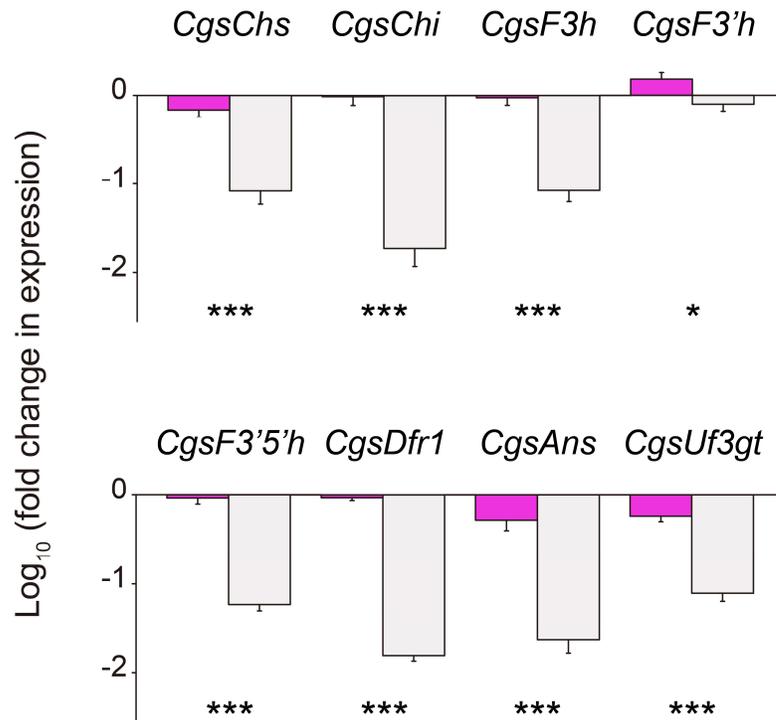


Figure 4: Differences in expression of the enzyme-coding genes between pink background and white cup. Gene expression was assessed using cDNA samples from flower buds collected 1 day before flowering. Vertical bars represent the means of five replicates and error bars indicate standard error. Pink bars represent pink background and light gray bars represent white cup. * $P < 0.05$, * $P < 0.001$.**

1.3.4 Three *R2R3-MYB* genes are acting differently in petal pigmentation

BLAST searches also identified the *R2R3-MYB* transcription factors (*CgsMYB6*, *CgsMYB11*, *CgsMYB12*) that are likely associated with the anthocyanin biosynthetic pathway (Table 1) by containing a short motif “[K/R]P[R/Q]PR” that is conserved in the anthocyanin-regulating subgroup 6 *R2R3-MYB* proteins (Appendix I, Figure A1.5). Phylogenetic analysis shows that these *R2R3-MYB* genes are clustered with the *R2R3-*

MYB proteins that have been characterized for anthocyanin biosynthesis (Figure 5), implying that these three *R2R3*-MYB genes may have similar functionality.

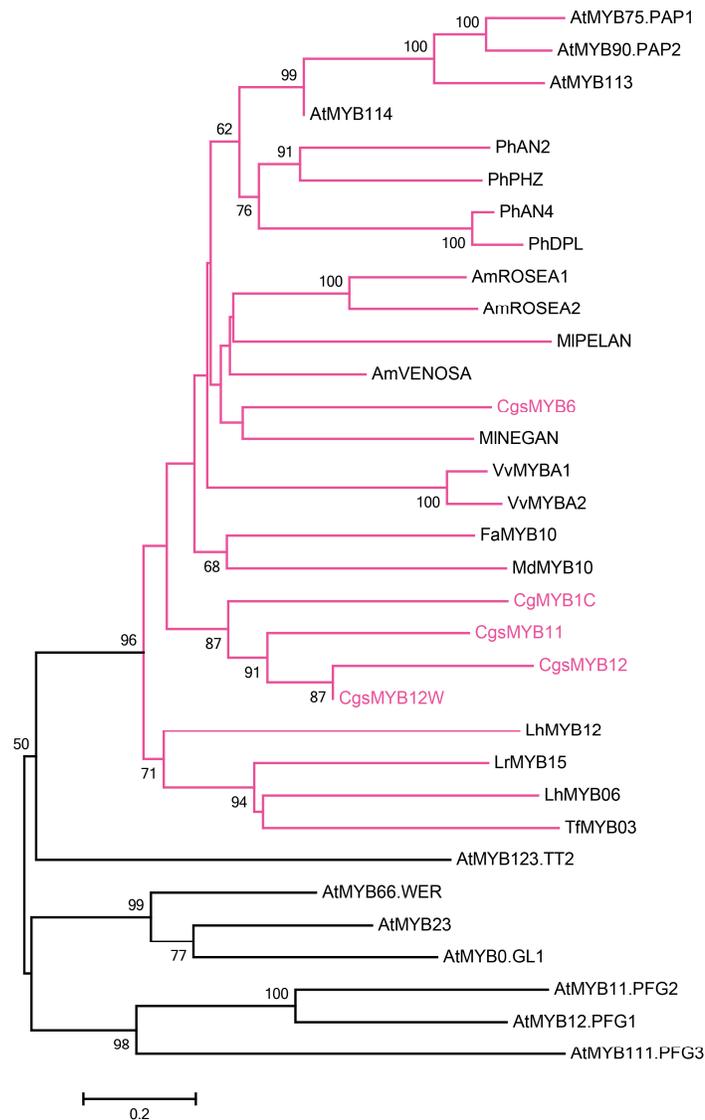


Figure 5: The neighbor-joining phylogenetic tree of R2R3-MYB proteins. The clade containing R2R3-MYBs of subgroup 6 is shown in pink. The R2R3-MYB from the petal of *Clarkia gracilis* ssp. *sonomensis* is shown in pink as well. Bootstrap support values $\geq 50\%$ are shown. The *Arabidopsis thaliana* sequences were retrieved from TAIR (<https://www.arabidopsis.org>): subgroup 5: AtMYB123 (AT5G35550); subgroup 6: AtMYB75 (AT1G56650), AtMYB90 (AT1G66390), AtMYB113 (AT1G66370) and

AtMYB114 (AT1G66380); subgroup 7: AtMYB11 (AT3G62610), AtMYB12 (AT2G47460) and AtMYB111 (AT5G49330); subgroup 15: AtMYB0 (AT3G27920), AtMYB23 (AT5G40330) and AtMYB66 (AT5G14750). Other sequences were retrieved from GenBank: *Antirrhinum majus* AmROSEA1 (ABB83826), AmROSEA2 (ABB83827), AmVENOSA (ABB83828); *CgMYB1C* (AOW41207); *Fragaria x ananassa* FaMYB10 (ABX79947); *Lilium* spp. LhMYB6 (BAJ05399), LhMYB12 (BAJ05398); *Lilium regale* LrMYB15 (BAU29929); *Malus domestica* MdMYB10 (ACQ45201); *Mimulus lewisii* MIPELAN (AHJ80987), MINEGAN (AHJ80988); *Petunia x hybrida* PhAN2 (AAF66727), PhAN4 (ADQ00392), PhDPL (ADW94950), PhPHZ (ADW94951); *Tulipa fosteriana* TfMYB3 (AHY20034); *Vitis vinifera* VvMYBA1 (BAD18977), VvMYBA2 (BAD18978).

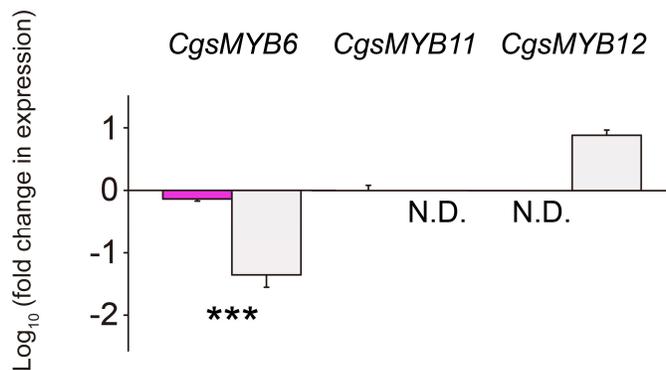


Figure 6: Differences in expression of three *R2R3-MYB* genes between pink background and white cup. Gene expression was assessed using cDNA samples from flower buds collected 1 day before flowering. Vertical bars represent the means of five replicates and error bars indicate standard error. N.D. indicates that the expression was not detectable. Pink bars represent pink background and light gray bars represent white cup. * $P < 0.001$.**

These three *R2R3-MYB* genes all show different expression patterns between pink background and white cup (Figure 6). *CgsMYB6* is expressed in the whole petal and across all flower-bud developmental stages (Figure 7). While its expression level is significantly lower in the white cup, compared to the pink background (Figure 6), *CgsMYB6* is still highly expressed in the white cup (Figure 7B). Actually, the expression

levels of *CgsMYB6* do not differ between the pink and white cups (Appendix I, Figure A1.2).

CgsMYB6 seems to be able to activate all of the enzyme-coding gene, except *CgsAns* (Figure 7). This inference is based on lack of the *CgsAns* expression in the stage 1 petal and in the stage 3 white cup, when neither of the other two *R2R3-MYBs* is expressed. The expression of *CgsAns* in the stage 2 petals, which were not dissected into regions, is explainable by the expression of this gene in the central spot, which is controlled by *CgMYB1* (Martins et al. 2017).

A



Type I

Gene	Stage 1	Stage 2	Stage 3		Stage 4		Stage 5	
	white petal	White petal, central spot	background white	cup white	background pink	cup white	background pink	cup pink
<i>CgsChs</i>	+++	+++	+++	+++	+++	+++	+++	+++
<i>CgsChi</i>	+(+) +	+++	+++	(+) (+)	+++		+ (+) (+)	
<i>CgsF3h</i>	+++	+++	+++	+++	+++	+++	+++	+ (+)
<i>CgsF3'h</i>	+++	+++	+++	++	+++	(+) (+) (+)	++ (+)	(+) (+)
<i>CgsF3'5'h</i>	(+) (+)	+++	++ (+)	+++	+++	+++	+++	+++
<i>CgsDfr1</i>	(+) ++	+++	+++	(+)	+++	+++	+++	+++
<i>CgsAns</i>		+++	+	(+)	+++	++	+++	+++
<i>CgsUf3gt</i>	+++	+++	+++	+++	+++	++	+++	+++
<i>CgsMYB6</i>	+++	+++	+++	+++	+++	+++	+++	+++
<i>CgsMYB11</i>			(+) +		+++		+++	
<i>CgsMYB12</i>						(+) + (+)		+++

B



Type III

Gene	Stage 1	Stage 2	Stage 3		Stage 5	
	White petal	White petal, central spot	Background	Cup	Background	Cup
			White	White	Pink	White
<i>CgsChs</i>	(+) ++	+++	+++	+++	+++	(+)
<i>CgsChi</i>	(+) +	+++	+++	(+) (+)	+	
<i>CgsF3h</i>	+++	+++	+++	+++	+++	(+)
<i>CgsF3'h</i>	+++	+++	+++	+++	+	(+)
<i>CgsF3'5'h</i>	(+)	+	+	++	+++	++ (+)
<i>CgsDfr1</i>	(+)	+++	++		+++	++ (+)
<i>CgsAns</i>		+++	++		+++	
<i>CgsUf3gt</i>	(+) ++	+++	+++	+++	+++	+
<i>CgsMYB6</i>	+++	+++	+++	+++	+++	+++
<i>CgsMYB11</i>					+++	
<i>CgsMYB12</i>				(+)		+++

C



Type IV

Gene	Stage 1	Stage 3		Stage 4		Stage 5	
	white petal	background	cup	background	cup	background	cup
		white	white	pink	white	pink	pink
<i>CgsChs</i>	+++	+++	+++	+++	+++	+++	+++
<i>CgsChi</i>	++	+++	+	(+) ++		(+)	
<i>CgsF3h</i>	+++	+++	+++	+++	+++	+++	+++
<i>CgsF3'h</i>	+++	+++	+++	+	+	+	+
<i>CgsF3'5'h</i>	(+) (+) +	(+)	+++	+++	+++	+++	+++
<i>CgsDfr1</i>	++	+	+	+++	++ (+)	+++	+++
<i>CgsAns</i>		(+)		+++	+	+++	+++
<i>CgsUf3gt</i>	+++	+++	+++	+++	(+) ++	+++	++
<i>CgsMYB6</i>	+++	+++	+++	+++	+++	+++	+++
<i>CgsMYB11</i>		(+)		+++		+++	
<i>CgsMYB12</i>					(+)	(+)	+++

Figure 7: Expression patterns of the anthocyanin enzyme-coding genes and the R2R3-MYB genes across the flower-bud developmental stages. (A) Central-spotted, pink-cupped Type I flowers. (B) Central-spotted, white-cupped Type III flowers. (C) Unspotted, pink-cupped Type IV flowers. Three plants each from the three flower types were examined. Based on the PCR-band brightness on the gels, the expression levels were scored as "+", "(+)", and blank, respectively representing expressed, weakly expressed and not expressed. A "+" or "(+)" sign shows a data point from one plant. Pictures above the columns designate the bud phenotypes. The scale bar indicates 5 mm.

The other two *R2R3-MYB* genes, *CgsMYB11* and *CgsMYB12*, are expressed in the late stage(s), but in different regions: *CgsMYB11* is only expressed in the background region and *CgsMYB12* is only expressed in the cup region (Figures 6, 7). Specifically, *CgsAns* is not expressed in the petal background until *CgsMYB11* has been expressed (Stages 3-5), while it is not expressed in the pink cup until *CgsMYB12* is expressed (Stages 4 and 5). By contrast, the expression of *CgsMYB11* or *CgsMYB12* does not seem to increase the expressions of *CgsChs*, *CgsChi*, *CgsF3h*, *CgsF3'h* and *CgsUf3gt*, although this pattern does not necessarily mean that in the absence of *CgsMYB6*, *CgsMYB11* or *CgsMYB12* would not activate some or all of these genes.

Of the two *bHLH* and two *WDR* genes identified in the transcriptomes, *CgsbHLH2*, *CgsWDR1* and *CgsWDR2* are expressed at similar levels in the pink background and white cup, as judged by estimated FPKM values (Table 1). By contrast, the expression level of *CgsbHLH1* in the white cup is only a quarter to one-third of the level in the petal background (Table 1), suggesting it may be slightly downregulated in

the white cup. Because there was no biological replication of the transcriptomes, however, the statistical significance of this difference cannot be determined.

1.3.5 Cup color cosegregates with *CgsMYB12*

Sequencing of *CgsMYB12* from plants with pink and white cups revealed differences that are correlated with cup color (Appendix I, Figure A1.3). The copy from the white-cupped parental plant, *CgsMYB12W*, contains a 1-bp deletion in exon 3 that generates a premature stop codon, whereas the copy from the pink-cupped parental plant lack this deletion (Table 2A; Appendix I, Figure A1.3).

Table 2: Cosegregation of a 1-bp deletion in *CgsMYB12* and the cup phenotype in *Clarkia gracilis* ssp. *sonomensis*. (A) Correlation of the deletion with cup color in the parental plants and an F₁ plant used in the crosses. (B) Cosegregation in 40 F₂ plants. The alleles at this locus are shown as C and a colon, representing the nucleotide base cytosine and the deletion, respectively.

(A)	Parent1	Parent2	F ₁
	white-cupped	pink-cupped	pink-cupped
<i>CgsMYB12</i> genotype	::	CC	:C

(B)	F ₂				
	Type I	Type II	Type III	Type IV	Type V
<i>CgsMYB12</i> genotype	pink-cupped	white-cupped	white-cupped	pink-cupped	white-cupped
::	0	8	8	0	8
CC	3	0	0	2	0
:C	5	0	0	6	0

The truncated protein *CgsMYB12W* contains 124 amino acids, where 7 of the last 9 amino acids are different from those in the functional copy, *CgsMYB12*, which

contains 220 amino acids. Moreover, this deletion cosegregates with cup color in F₂ plants (Table 2B; Appendix I, Figure A1.4). There was a perfect cosegregation between cup colors and the genotypes of *CgsMYB12*: all plants with white cups are homozygous for the deletion, while heterozygotes and homozygotes for lack of deletion have pink cups.

1.3.6 These petal *R2R3-MYB* genes activate anthocyanin production

To confirm these petal *R2R3-MYB* proteins function as anthocyanin regulators, we created transgenic plants to overexpress these *R2R3-MYB* genes in *Arabidopsis* wild-type Col-0. If an *R2R3-MYB* gene participates the regulation of anthocyanin enzyme-coding genes, we should observe pigment accumulation in the transgenic plants and detect the elevated expression levels of the enzyme-coding genes. We generated transgenic plants with each of the five singly-inserted genes: *CgsMYB6*, *CgsMYB11*, *CgsMYB12*, *CgsMYB12W* (the nonfunctional *CgsMYB12*) and a positive control, *pap1-D*. The transgenic plants inserted with the functional genes are obviously pigmented in hypocotyls (Figures 8B-8E). The qPCR results show that the most upstream enzyme-coding gene in the pathway, *AtChs*, is expressed at similar levels between all transgenic plants and the wild-type. By contrast, *CgsMYB6* and *CgsMYB11* increase the expression of *AtDfr* and *AtAns* (Student's *t*-test, $P < 0.01$). *CgsMYB12* also increase the expression of *AtDfr* (Student's *t*-test, $P = 0.0275$), and seems to increase the expression of *AtAns* as well, as this expression difference is approaching to statistical significance (Student's *t*-test, P

= 0.0723). However, the result of activation of *AtDfr* by *CgsMYB12W* (Student's *t*-test, *P* = 0.0112) is perplexing, but it is clear that *CgsMYB12W* does not increase the expression of *AtAns* (Student's *t*-test, *P* = 0.1201) (Figure 8G).

1.4 Discussion

1.4.1 A functional mutation in *CgsMYB12* is associated with white-cup formation

Our results show that there are two domains for pink coloration in the *C. g. ssp. sonomensis* petal, the background and cup regions. The pigments produced in these domains, malvidin-derived anthocyanins, is identical. Not surprisingly, all enzyme-coding genes are highly expressed in both of the petal background and pink cup, as is *CgsMYB6* (Figures 7A, 7C). The primary difference resides in the expression of *CgsMYB11* and *CgsMYB12*: the former is expressed exclusively in the petal background, whereas the latter is expressed exclusively in the pink cup. The cup pigmentation can thus be considered part of the "background" pigmentation, indicating that what appears to be continuous background pigmentation is in reality similar pigmentation in two different petal regions. Several lines of evidence indicate that lack of pigmentation in the white cup is due to a functional mutation in *CgsMYB12*. First, the functional copy of this gene is present in plants with a pink cup, while the copy in plants with a white cup contains a single base pair deletion, causing a frame shift and a premature stop codon. Because the protein produced by this truncated copy is only about half the length of the

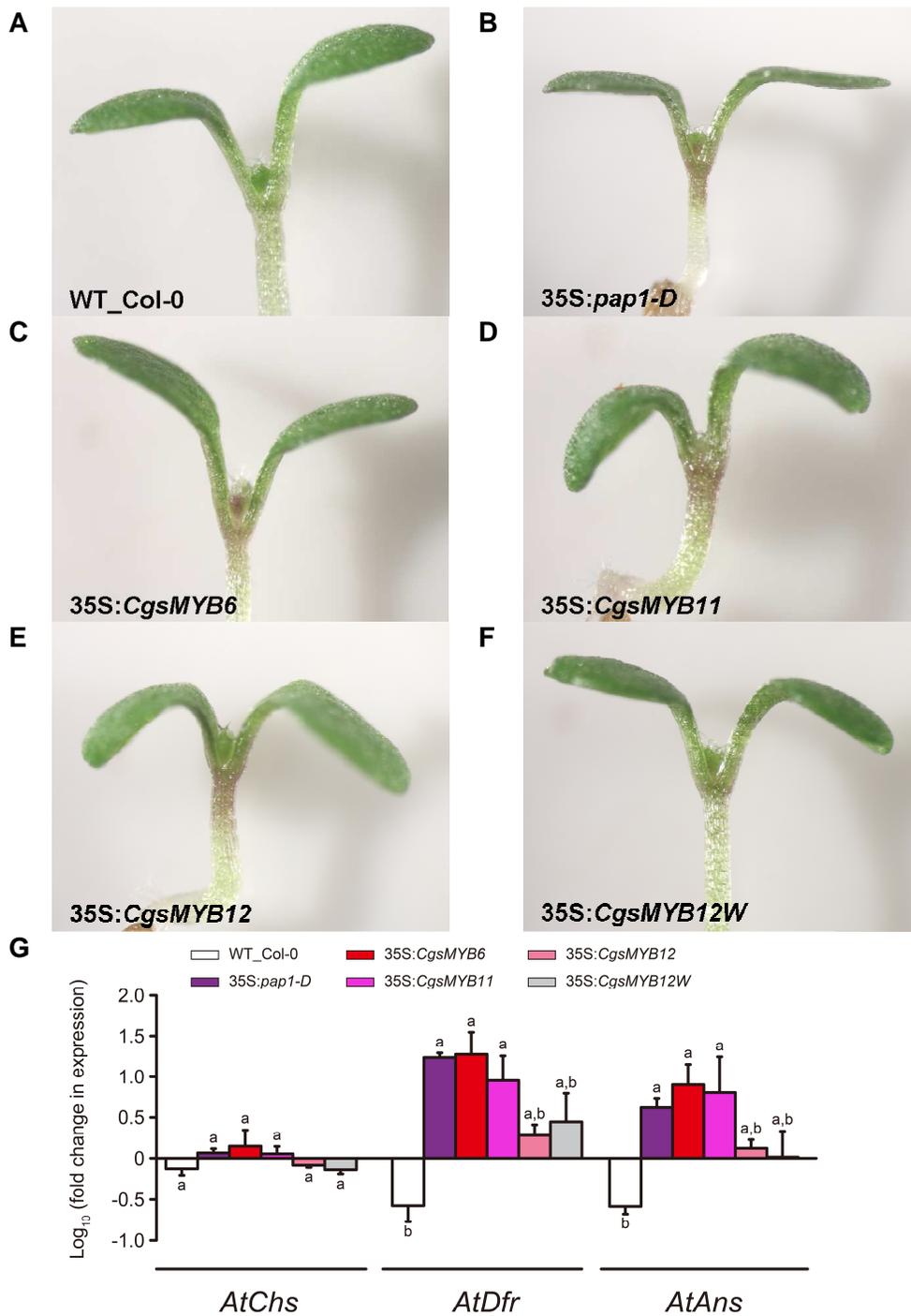


Figure 8: Phenotypic and molecular characterization of the *Arabidopsis* transgenic plants. (A-F) Phenotypes of the 5-day-old *Arabidopsis* transgenic plants. (G) Differences in expression of three enzyme-coding genes in the *Arabidopsis* transgenic

plants. qPCR was conducted using cDNA samples of the pooled 5-day-old T3 seedlings. Vertical bars represent the means of three independent lines and error bars indicate standard error. Statistical significance was based on one-way ANOVA and the Tukey-Kramer HSD test.

untruncated protein, it is most likely non-functional. Second, while *CgsAns* is activated by the functional copy of *CgsMYB12* in pink cups, it is not activated in white cups where the presumed defective copy, *CgsMYB12W*, is expressed. Third, the 1-bp deletion cosegregates with the cup color. And fourth, *Arabidopsis* transformations with *CgsMYB12* activate anthocyanin production but transformation with *CgsMYB12W* does not (Figures 8E, 8F).

Two puzzling results regarding the transformation experiment are that (1) although *CgsMYB12* activated anthocyanin production, it did not increase the expression of *AtAns* significantly. One possible explanation for this pattern is that *AtAns* is actually upregulated by *CgsMYB12*, but at a level that we did not have the power to detect with only three biological replicates. This explanation is supported by the observation that pigments are produced in plants transformed with *CgsMYB12*; (2) *CgsMYB12W* activated the expression of *AtDfr*. The reason of such a activation is unclear, but apparently, this elevated *AtDfr* expression cannot make anthocyanin production, because no pigment accumulated in the transgenic plants of *CgsMYB12W*.

Interestingly, in white cups, all enzyme-coding genes are downregulated compared to the petal background and pink cups, despite the expression of *CgsMYB6*, which normally activates all of these genes except *CgsAns*. A possible explanation for

this pattern is that the *CgsMYB12W* protein interferes with activation by *CgsMYB6*. The inactivating mutation of *CgsMYB12W* is located in exon 3. This means that *CgsMYB12W* contains an intact DNA-binding domain of R2 and R3 repeats, including a bHLH binding motif in the R3 repeat (Appendix I, Figure A1.5), suggesting that *CgsMYB12W* is likely able to bind to the promoter regions of the enzyme-coding genes and interact with the bHLH protein (Zimmermann et al. 2004) or even able to form a MYB-bHLH-WDR transcription protein complex. By contrast, it lacks a large portion of the C-terminal domain, including the whole motif of subgroup 6 that features all anthocyanin-regulating R2R3-MYB proteins (motif 6; Appendix I, Figure A1.5). The C-terminal domain is involved in activation or repression of transcription (Dubos et al. 2010). The C-terminal domains of C1 (from maize) and AN2 (from petunia) have been shown to activate transcription in yeast one-hybrid assays (Goff et al. 1992; Sainz et al. 1997; Quattrocchio et al. 1999). Both of these properties suggest that the *CgsMYB12W* protein may block the action of *CgsMYB6* in activating these genes by competitive inhibition. Further studies will be needed to determine whether this explanation is correct.

The expression level of *CgsbHLH1* also appears to be downregulated in the white cup, compared to the petal background. This could possibly be explained by *CgsMYB12W* being nonfunctional, since, at least in *Arabidopsis*, R2R3-MYB genes contribute to activation of *bHLH* genes (Baudry et al. 2006). This downregulation of

CgshHLH1 could also contribute partially to downregulation of the enzyme-coding genes in the white cup.

Our findings are consistent with the common observation that evolutionary loss of pigmentation typically involves downregulation or loss of function in *R2R3-MYB* genes (Streisfeld & Rausher 2010). The anthocyanin-associated *R2R3-MYB* genes are usually functionally specific in regulating anthocyanin production (Ramsay & Glover 2005; Quattrocchio et al. 2006), and are present in multiple copies, with different copies control the anthocyanin biosynthesis in different tissues (Yamagishi et al. 2010; Albert et al. 2011). Therefore, mutations in a particular *R2R3-MYB* gene would only cause the phenotypic change in a limited spatial domain (e.g. flowers), and would not affect anthocyanin pigmentation in other domains (e.g. stems, leaves). There is thus expected to be minimal deleterious pleiotropy associated with downregulation or loss of function in these genes, and purifying selection is expected to be weak or non-existent on these types of mutation.

By contrast, the anthocyanin enzyme-coding genes are often expressed in the vegetative tissue as well and involved in the synthesis of other secondary metabolites, such as proanthocyanins and flavonols (Koes et al. 2005; Wessinger & Rausher 2012). Similarly, bHLH and WDR transcription factors often regulate not only anthocyanin production but also other physiological or developmental processes, such as proanthocyanidin synthesis, the formation of trichomes or root hairs and the production

of seed coat mucilage (Koes et al. 2005; Ramsay & Glover 2005). Accordingly, mutations in the enzyme-coding genes and bHLH or WDR transcription factors could influence more than one tissue and are expected to have more negative pleiotropic effects. Such deleterious mutations are likely purged by purifying selection during the evolutionary processes before they could be observed in nature.

1.4.2 Gene duplication contributes to evolution of novel pattern elements

Our findings that at least four *R2R3-MYB* genes responsible for petal pigmentation patterning in *C. g. ssp. sonomensis* suggest that gene duplication is central to the evolution of petal pigmentation patterning. The role of gene duplication in evolution has been popularized by Ohno's seminal book (1970). Recently, with accumulating genome sequence data, duplicated genes have been found to be pervasive in all kinds of life forms (Himmelreich et al. 1996; Klenk et al. 1997; Tomb et al. 1997; Rubin et al. 2000; The *Arabidopsis* Genome Initiative 2000; Li et al. 2001; Velasco et al. 2010; Myburg et al. 2014) and the importance of gene duplication in the evolution of phenotypic novelty has been clearly demonstrated (Conant & Wolfe 2008; Freeling et al. 2015; Innan & Kondrashov 2010).

The genetics of petal pigmentation patterning usually involves multiple *R2R3-MYB* genes with each gene controlling a specific pattern element (Schwinn et al. 2006; Albert et al. 2011; Hsu et al. 2015). This is likely facilitated by duplication of anthocyanin-regulating *R2R3-MYB* genes, which appears to be common in flowering

plants. For example, in *Arabidopsis thaliana*, the R2R3-MYB genes of subgroup 6, *AtMYB75*, *AtMYB90*, *AtMYB113* and *AtMYB114*, are phylogenetically clustered (Figure 5) and also physically linked (Stracke et al. 2001), likely the result of recent tandem duplications. Such similar cases include *AmROSEA1* and *AmROSEA2* in *Antirrhinum majus* (Schwinn et al. 2006), and *MgMYB1*, *MgMYB2* and *MgMYB3* in *Mimulus gattatus* (Cooley et al. 2011).

Differential expression domains of the R2R3-MYB genes in *C. g. ssp. sonomensis* may be explained by repeated duplications followed by subfunctionalization, in which duplicated genes are functionally divergent because each retains a different part of the original ancestral function. For example, the single-copy ancestral gene of *CgsMYB11* and *CgsMYB12* may have been expressed across the whole petal, while after duplication, *CgsMYB11* and *CgsMYB12* are expressed in separate parts of the ancestral expression domain. The alternative explanation may be the recruitment of other MYB-like proteins that have been expressed in the specific regions, but ancestrally performed other functions. For example, *CgMYB1* and *CgsMYB12* may have originated from other MYB proteins that did not function as anthocyanin regulators, but have already been expressed in the spot and cup, respectively. Under this scenario, there would have been changes in their coding sequences that allow them to recognize binding sites in the *cis*-regulatory regions of the anthocyanin enzyme-coding genes they now regulate.

Distinguishing between these possibilities will require detailed examination of the phylogenetic relationships between *MYB* genes in *Clarkia*

1.4.3 Petal pigmentation of *Clarkia gracilis* ssp. *sonomensis* requires multiple *R2R3-MYB* genes

In addition to the spot regulator *CgMYB1* that has been documented in Martins et al. (2017), we identified three *R2R3-MYB* genes that appear to be involved in regulating the anthocyanin pigmentation in the *C. g. ssp. sonomensis* petal: *CgsMYB6*, *CgsMYB11* and *CgsMYB12*. These three genes show different temporal and/or spatial expression patterns during the flow-bud development (Figures 6, 7, 9).

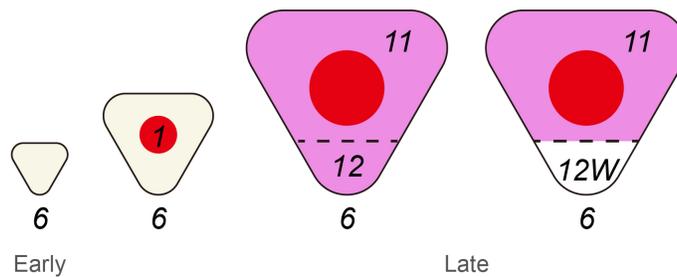


Figure 9: Petal pigmentation in *Clarkia gracilis* ssp. *sonomensis* during petal development. The numbers indicate the *R2R3-MYB* genes: 1: *CgMYB1*, 6: *CgsMYB6*, 11: *CgsMYB11*, 12: *CgsMYB12* and 12W: *CgsMYB12W*, the nonfunctional copy of *CgsMYB12*. The positions of these numbers indicate the expression domains of these *R2R3-MYB* genes. *CgsMYB6* (6) is shown below the petal because it is expressed across the whole petal. At the early stage where only *CgsMYB6* is expressed, most of the enzyme-coding genes are also expressed, except *CgsAns*. *CgsAns* is not expressed until any of the region-specific *R2R3-MYB* genes is expressed.

CgsMYB6 is expressed earliest and stays expressed throughout the development.

At the late developmental stages, *CgsMYB11* and *CgsMYB12* are expressed in the background and cup regions respectively. In addition, these three *R2R3-MYB* genes

appear to have differentiated in their target-gene specificities by activating different sets of the enzyme-coding genes. At the very early developmental stage when the petal is tiny and white, *CgsMYB6* is the only expressed *R2R3-MYB* gene among these petal *R2R3-MYB* genes. At this stage, the expressed enzyme-coding genes include *CgsChs*, *CgsChi*, *CgsF3h*, *CgsF3'h*, *CgsF3'5'h*, *CgsDfr1* and *CgsUf3gt*, although the expression level of *CgsF3'5'h* is lower than other genes. This implies that *CgsMYB6* acts to activate most of the enzyme-coding genes, except *CgsAns*. *CgsAns* is not expressed until region-specific *CgMYB1*, *CgsMYB11* or *CgsMYB12* has been expressed. According to Martins et al. (2013; 2017), *CgMYB1* is expressed before the background color appears and only in the spot region. The spot determinant *CgMYB1* activates *CgDfr2* (the *CgDfr* copy only expressed in the red spot) and *CgAns* strongly. In the background and cup regions, *CgsAns* is activated by *CgsMYB11* and *CgsMYB12* respectively (Figures 7, 8). It is possible that these two genes by themselves could also activate some or all of the enzyme-coding genes, but if so, this is masked by their prior activation by *CgsMYB6*.

1.4.4 Evolutionary lability of anthocyanin regulatory network

Most flowering plants produce anthocyanins and do so using a highly conserved anthocyanin biosynthetic pathway. The enzyme-coding genes of this pathway are recognizably homologous across angiosperms, and genes from one species can rescue function of loss-of-function mutants in taxonomically distant species (Quattrocchio et al 1999; Streisfeld & Rausher 2009a). The basic regulation of the enzyme-coding genes is

also very similar across angiosperms, with R2R3-MYB, bHLH, and WDR proteins forming a complex that activates the enzyme-coding genes (Xu et al. 2015). Nevertheless, the nature of the interactions among these transcription factors, as well as which enzyme-coding genes they target, is known to vary across species, indicating that the regulatory network is evolutionarily labile. Our results have revealed several new ways in which this regulatory network has evolved and diverged from that of other plants that have been examined.

First, our results provide evidence for a new type of partitioning of enzyme-coding genes among one or more R2R3-MYB transcription factors. In the flowering plants in which anthocyanin regulation has been characterized, the enzyme-coding genes usually are controlled coordinately in a given tissue by either one or two R2R3-MYB transcription factors (Mol et al. 1998; Koes et al. 2005; Xu et al. 2015). In some species (e.g., *Lilium* spp., *Zea mays*, *Ipomoea purpurea* and *I. nil*), a single R2R3-MYB protein activates all enzyme-coding genes (Dooner 1983; Tiffin et al. 1998; Morita et al. 2006; Lai et al. 2012). In other species, principally eudicots (e.g., *A. majus*, *Petunia hybrida*, *Mimulus aurantiacus* and *M. lewisii*), one R2R3-MYB protein activates genes that are upstream of the pathway, while a second activates downstream genes (Martin et al 1991; Quattrocchio et al. 1993; Streisfeld & Rausher 2009b; Streisfeld et al. 2013; Yuan et al. 2014).

The transcriptional regulation we found in *C. g. ssp. sonomensis*, however, does not appear to conform to either of these two common patterns. In particular, *CgsMYB6* seems to activate both upstream and downstream genes, except for *CgsAns*, which is activated by *CgsMYB11* and *CgsMYB12*, depending on the petal regions. The full set of enzyme-coding genes these latter two transcription factors activate by themselves is currently unknown, but it is clear that these and *CgsMYB6* partition activation of enzyme-coding genes in a way that is novel. It thus appears that the target gene specificity of R2R3-MYB proteins is more evolutionarily labile than has been hitherto appreciated. Additionally, the results of our transformation experiments reveal other aspects of the regulatory network that have evolved. In particular, the observation that *CgsMYB6* activates *Ans* in *Arabidopsis* but not in *Clarkia* implies that the *cis*-regulatory regions of *Ans* have diverged between these two species.

The picture that emerges from these patterns is that the target specificity of R2R3-MYB transcription factors evolves, often rapidly. This lability appears to be caused by both evolution of target motifs recognized by these transcription factors, as well as evolution of those transcription factors.

2. The evolution of petal pigmentation patterning in *Clarkia gracilis*

2.1 Introduction

In recent decades, evolutionary biologists have debated about the types of genetic changes that are important to the evolution of morphological diversity. Some argues that new genetic material supplied through gene/genome duplication allows duplicates to evolve new functions or new phenotypes (Zhang 2003; Rensing 2014; Van de Peer et al. 2017). For example, duplications of genes involved in regulating transcription, followed by the retention of regulatory gene paralogs, are associated with functional diversification of transcription factors and contribute to the increase of organismal complexity in animals and plants (Schmitz et al. 2016; Lang et al. 2010). On the other hand, genetic changes in the pre-existing genetic material are considered by some to be the main drivers of evolutionary novelty (Carroll 2005; Wray 2007; Stern & Orgogozo 2008). This viewpoint is particularly supported by the prevalence of *cis*-regulatory modifications in regulating animal body plans (Wittkopp et al. 2002; Shapiro et al. 2004; Stern & Frankel 2013), including the butterfly wing color patterns (Reed et al. 2011; Martin & Reed 2014).

In plants, many species have flowers with discrete pigmentation patterns within the corolla that contrast with the background coloration of the flower. Such petal color patterns include petal spots, strips overlaying veins (venation), irregular blotches, white patches or combinations of these. Many color patterns affect plant-pollinator interactions

that may increase the pollinator visit frequency and/or pollinator efficiency in transferring pollen (Waser & Price 1983; Jones 1996a,b; Medel et al. 2003; Heuschen et al. 2005; Lunau et al. 2006). Molecular mechanisms of petal pattern development are thus of interest for understanding the potential effects on reproductive success mediated by pollinators.

Recent genetic and developmental examinations of petal pigmentation patterning have revealed that novel pattern elements are frequently the result of gene duplication, with different anthocyanin-regulating R2R3-MYB transcription factors controlling different pattern elements (Schwinn et al. 2006; Albert et al. 2011; Shang et al. 2011; Hsu et al. 2015; Chiou & Yeh 2008; Martins et al. 2013, 2017; Yuan et al. 2014; Yamagishi et al. 2010, 2014, 2018; Yamagishi 2018). Such a commonality may be explained by transcriptional control of anthocyanin biosynthesis. Pattern pigmentation requires the expression of all three components (R2R3-MYB, bHLH and WDR) of the protein complex that regulates transcription of the enzyme-coding genes in the anthocyanin biosynthetic pathway (Ramsay & Glover 2005; Petroni & Tonelli 2011; Xu et al. 2015). The R2R3-MYB regulators are particularly central to pattern formation, because the anthocyanin-regulating *R2R3-MYB* genes usually exist as a multigene family, with each gene exhibiting a different spatial expression pattern (Schwinn et al. 2006; Albert et al. 2011; Hsu et al. 2015). Moreover, the anthocyanin-regulating *R2R3-MYB* genes are usually functionally specialized in regulating anthocaynin biosynthesis, while the bHLH

and WDR partners often participate in regulating other processes (e.g., trichome formation), along with anthocyanin production (Koes et al. 2005; Ramsay & Glover 2005). Consequently, the *R2R3-MYB* genes are usually the primary determinants of spatially localized anthocyanins.

At the same time, evidence is accumulating that the targeting of anthocyanin enzyme-coding genes by *R2R3-MYB* transcription factors is evolutionarily labile (Rausher 2006, Chapter 1). Moreover, this lability appears to involve both changes in the transcription factors that alter which *cis*-regulatory motifs they recognize, as well as evolutionary change in those motifs (Chapter 1). Thus, both viewpoints on the relative importance of gene duplication and regulatory mutations described above appear to account for some changes in anthocyanin regulation.

Our findings of petal pigmentation in *Clarkia gracilis* ssp. *sonomensis* are in agreement with these studies of petal patterning. *C. g.* ssp. *sonomensis* is one of the four subspecies of *C. gracilis*. These subspecies vary in petal pigmentation patterning. All four subspecies have petals with a pink background, but differ mainly with respect to the presence and position of red petal spots. *C. g.* ssp. *sonomensis* is the only subspecies that has central spots (although unspotted *C. g.* ssp. *sonomensis* are occasionally observed). Two other subspecies, *C. g.* ssp. *albicaulis* and *C. g.* ssp. *tracyi*, have basal spots. The fourth subspecies, *C. g.* ssp. *gracilis* is a selfer and lacks petal spots (Martins et al. 2013). In addition to the petal spot, *C. gracilis* also has a white patch pattern element. *C. g.* ssp.

sonomensis is polymorphic for the presence of a “white cup”, a white sector at the base of the petal, where no anthocyanins are produced. The white cup phenotype is recessive, and the dominant phenotype is pink cup. A similar white patch phenotype can also be found in the subspecies *C. g. ssp. albicaulis*, which has a white “band” in the middle of the petal (Figure 10).

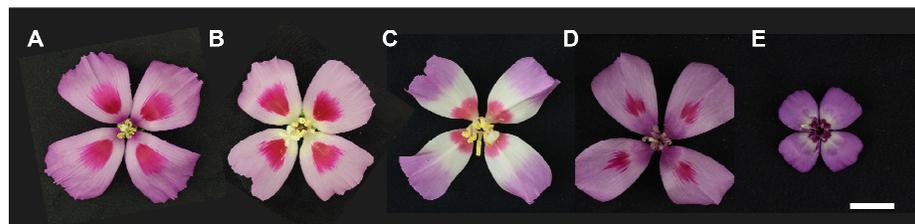


Figure 10: Flowers of *Clarkia* (sub)species. (A) pink-cupped *Clarkia gracilis* ssp. *sonomensis*; (B) white-cupped *Clarkia gracilis* ssp. *sonomensis*; (C) *C. g. ssp. albicaulis*; (D) *C. amoena* ssp. *huntiana*; (E) *C. lassenensis*. The scale bar indicates 15 mm.

In the previous chapter, we described four *R2R3-MYB* genes involved in pigmentation patterning in the *C. g. ssp. sonomensis* petal. *CgMYB1* (Martins et al. 2017), *CgsMYB11* and *CgsMYB12* are region-specific regulators that control pigmentation in red spot, pink background and pink cup, respectively. To activate all enzyme-coding genes in the anthocyanin biosynthetic pathway, the participation of *CgsMYB6* is also required. *CgsMYB6* is expressed throughout the petal and regulates all of the enzyme-coding genes except *CgsAns*. A phylogenetic analysis in Chapter 1 revealed that *CgMYB1*, *CgsMYB11* and *CgsMYB12* cluster together and form a monophyletic clade, while *CgsMYB6* is clustered with the *R2R3-MYB* proteins from other species in a different clade. This suggests that *CgMYB1*, *CgsMYB11* and *CgsMYB12* share a common

ancestor and evolve through duplications of the ancestral gene, while *CgsMYB6* may have a different evolutionary origin.

Clarkia gracilis is an allotetraploid that is thought to be derived from the two diploid species: *C. amoena* ssp. *huntiana* and an extinct species related to *C. lassenensis* and *C. arcuata* (Abdel-Hameed & Snow 1968, 1972). In this study, *C. lassenensis* was chosen to represent the extinct parental species, because a relatively better chromosome pairing was observed in the *C. lassenensis* x *C. gracilis* triploids than that in the *C. arcuata* x *C. gracilis* triploids (Hakansson 1946, cited by Abdel-Hameed & Snow 1972). *C. amoena* ssp. *huntiana* and *C. lassenensis* have similar petal pigmentation patterns as *C. gracilis*. *C. amoena* ssp. *huntiana* has pink petals and red central spots. *C. lassenensis* has purple-pink petals, purple-red basal spots, and narrow white bands above the basal spots (Figure 1; Appendix II, Figure A2.1).

Because *C. gracilis* is an allotetraploid, the *R2R3-MYB* genes expressed in the petal may have arisen in one of two ways. One possibility is that different duplicates represent copies of orthologous genes from the two progenitor species. For example, *CgMYB1* may correspond to a gene inherited from *C. amoena* ssp. *huntiana*, whereas *CgsMYB12* may correspond to a gene inherited from the other progenitor species. A second possibility is that there was duplication within the *C. gracilis* lineage of an *R2R3-MYB* gene inherited from one progenitor, while the orthologous gene from the other progenitor was lost.

The main goal of this chapter is to determine which of these two possibilities is correct. To do so, we investigate whether genetic mechanisms underlying the petal patterns in the species closely related to *C. g. ssp. sonomensis* are the same as those in *C. g. ssp. sonomensis*. Specifically, we aim to address these questions: (1) Do the homologs of *CgMYB1*, *CgsMYB6*, *CgsMYB11* and *CgsMYB12* exist in the closely related species *C. amoena ssp. huntiana* and *C. lassenensis*, and in the subspecies having the white band pattern, *C. g. ssp. albicaulis*? If so, (2) do the *R2R3-MYB* genes exhibit the same spatial expression patterns in these species? (3) What is the phylogenetic relationship among these *R2R3-MYB* genes?

2.2 Materials and Methods

2.2.1 Plant growth

Seeds of *Clarkia gracilis ssp. albicaulis* (Jeps.) H. Lewis & M. Lewis, *C. amoena ssp. huntiana* (Jeps.) H. Lewis & M. Lewis and *C. lassenensis* (Eastw.) H. Lewis & M. Lewis (Appendix II, Table A2.1) were germinated by placing them inside of the folds in the damp paper towels. The paper towels were then placed in the sandwich zip bags to minimize the moisture loss. The bags were placed in the growth room at 15-18°C in the dark. When the radicles appeared, the seeds were transferred to wet Fafred 4P soil (Sun Gro Horticulture) in the 3.1" x 2.2" x 2.3" cells. Germinated seeds were grown in the growth room with 16-hour day length. After 3-4 weeks, seedlings were transferred to 5-inch pots and grown in the Duke greenhouse (20-24 °C).

2.2.2 Cloning of the *R2R3-MYB* genes

The coding regions of *MYB6*, *MYB11* and *MYB12* were amplified from *C. g. ssp. albicaulis*, *C. amoena ssp. huntiana* and *C. lassenensis* with the primers listed in Appendix II, Table A2.2 using cDNA as templates. We also amplified *MYB1* from *C. amoena ssp. huntiana* because the sequence deposited in GenBank is only 271-bp long (GenBank accession no. KX592430). cDNA samples were prepared with total RNA extracted from the dissected petals collected approximately 1 day before the flower-buds open. The petals were dissected into sections based on the colors. For *C. g. ssp. albicaulis*, the petals were dissected into top (pink background), middle (white band) and base (red spot) sections. For *C. amoena ssp. huntiana*, the petals were dissected into top (pink background) and base (pink cup) sections. For *C. lassenensis*, the petals were dissected into top (purple-pink background) and base (narrow white band and purple-red spot) sections (Appendix II, Figure A2.1).

The detailed methods of RNA extraction and cDNA synthesis were described in Chapter 1. The PCR products were gel-purified, cloned and sequenced following the methods described in Chapter 1.

2.2.3 Phylogenetic analysis

The nucleotide sequences of the *R2R3-MYB* genes from *C. g. ssp. sonomensis*, *C. g. ssp. albicaulis*, *C. amoena ssp. huntiana* and *C. lassenensis* and from *Arabidopsis thaliana*, *Antirrhinum majus* and *Mimulus lewisii* were aligned using MUSCLE (Edgar 2004). A

maximum likelihood phylogenetic tree was constructed using PhyML version 20120412 (<http://www.atgc-montpellier.fr/phyml>) (Guindon et al. 2010) under the TN93+I+G substitution model determined with Akaike Information Criterion (AIC) by Smart Model Selection (SMS) version 1.8.1 (Lefort et al. 2017) integrated into PhyML. Clade support was estimated by approximate likelihood ratio tests (aLRT) based on a Shimodaira–Hasegawa-like procedure (Anisimova & Gascuel 2006).

2.2.4 Semi-quantitative assessment of gene expression across flower-bud developmental stages

We collected flower buds from three plants each of *C. g. ssp. albicaulis*, *C. amoena ssp. huntiana* and *C. lassenensis* during the flower-bud development. The flower-buds were collected at different sizes until 1 day before flowering (when buds just became erect), and were categorized into different stages according to the order that the colors appear: (1) white petal; (2) central or basal spot (depending on species) appears; (3) central or basal spot well defined; (4) background and cup colors appear. The larger petals (stages 3, 4) were dissected into sections differing in color.

cDNA samples were prepared as described in Chapter 1 and were diluted to 2.5 ng/μL for PCR. PCR reactions were conducted using *Taq* DNA Polymerase (New England BioLabs) with the primers listed in Appendix II, Table A2.2. PCR products were visualized on 2% agarose gels. The brightness of PCR bands reflects the expression levels of the tested genes, and was scored as expressed (“+”), weakly expressed (“(+)”) or not expressed (blank).

2.3 Results

2.3.1 The *R2R3-MYB* genes evolved before polyploidization of *C. gracilis*

We found that the homologs of *MYB6*, *MYB11* and *MYB12* are present in *C. g. ssp. albicaulis*, *C. amoena ssp. huntiana* and *C. lassenensis*, with one exception that we have not been successful in amplifying *MYB12* from *C. g. ssp. albicaulis* despite several attempts. Full-length or nearly full-length coding sequences of these genes were used in phylogenetic analysis. The maximum-likelihood phylogenetic tree of the *R2R3-MYB* DNA sequences (Figure 11) reveals the same pattern as the neighbor-joining tree of the *R2R3-MYB* proteins reported in Chapter 1 that *MYB6* is distantly related to *MYB1*, *MYB11* and *MYB12*. And again, *MYB1*, *MYB11* and *MYB12* form a monophyletic clade, where the gene homologs from different species form a subclade with high branch support (Figure 11). This pattern suggests that at least two duplication events occurred prior to the speciation of *C. gracilis* and thus created these three genes: one duplication yielded *MYB1* and the progenitor of *MYB11* and *MYB12*, while a second duplication of that progenitor produced *MYB11* and *MYB12*. *MYB6* is in a different, well-supported clade that contains *R2R3-MYBs* from *A. majus* and *M. lewisii*, and like other *R2R3-MYB* genes, the *MYB6* homologs also form a subclade. Within each of their subclades, *MYB1*, *MYB11* and *MYB12* show the same pattern of divergence, in which *C. gracilis* is more closely related to *C. amoena ssp. huntiana* than to *C. lassenensis*, with this relationship showing high support. This pattern suggests that the copies of these genes in *C. gracilis*

are orthologous to the copies inherited from *C. amoena* ssp. *huntiana*. By contrast, *MYB6* appears to show the opposite pattern, in that *C. gracilis* is more closely related to *C. lassenensis* than to *C. amoena* ssp. *huntiana*. This pattern suggests that the *MYB6* copy in *C. gracilis* is orthologous to the copy inherited from *C. lassenensis*.

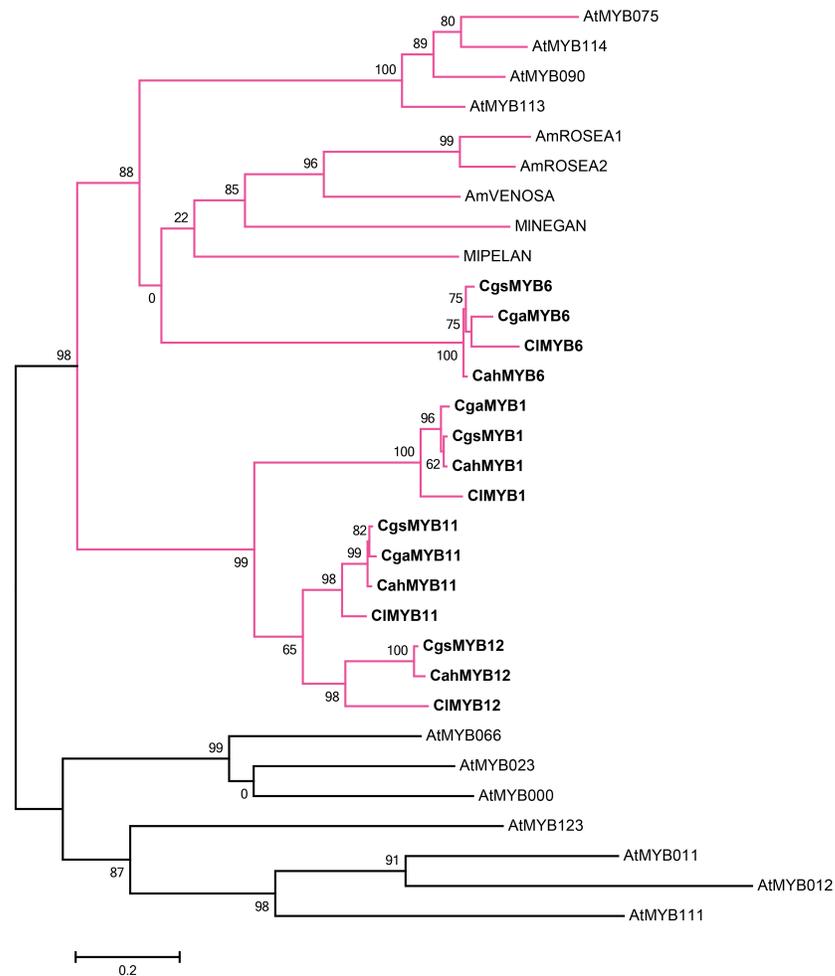


Figure 11: The maximum likelihood phylogenetic tree of *R2R3-MYB* genes. The clades contain the *R2R3-MYB* genes of subgroup 6 are shown in pink. The genes from the *Clarkia* (sub)species are shown in bold. Branch support are shown as aRLT statistics. The *Arabidopsis thaliana* sequences were retrieved from TAIR (<https://www.arabidopsis.org/>); subgroup 5: AtMYB123 (AT5G35550); subgroup 6:

AtMYB75 (AT1G56650), AtMYB90 (AT1G66390), AtMYB113 (AT1G66370) and AtMYB114 (AT1G66380); subgroup 7: AtMYB11 (AT3G62610), AtMYB12 (AT2G47460) and AtMYB111 (AT5G49330); subgroup 15: AtMYB0 (AT3G27920), AtMYB23 (AT5G40330) and AtMYB66 (AT5G14750). Other sequences were retrieved from GenBank: *Antirrhinum majus* AmROSEA1 (DQ275529), AmROSEA2 (DQ275530), AmVENOSA (DQ275531); *Clarkia gracilis* ssp. *albicaulis* CgaMYB1 (CgMYB1B, KX592431); *C. g.* ssp. *sonomensis* CgsMYB1 (CgMYB1C, KX592432); *C. lassenensis* CIMYB1 (KX592428); *Mimulus lewisii* MIPELAN (KJ011144), MINEGAN (KJ011145).

2.3.2 The expression domains of these *R2R3-MYB* genes are conserved between species

The expression patterns of *MYB6* and *MYB11* in *C. g.* ssp. *albicaulis*, *C. amoena* ssp. *huntiana* and *C. lassenensis* are consistent with what we observed in *C. g.* ssp. *sonomensis* (Figure 12). *MYB6* is expressed throughout the development in all sections in the petal. *MYB11* is expressed at the late developmental stage and only expressed in the petal background. However, *MYB12* seems to act differently among these (sub)species. We were unable to amplify *MYB12* from *C. g.* ssp. *albicaulis*, which may be the result of that *MYB12* is not expressed or not present in *C. g.* ssp. *albicaulis*. In *C. amoena* ssp. *huntiana* and *C. lassenensis*, *MYB12* is still expressed at the late stage, but its expression is detectable in the both top (background) and basal (cup/spot) regions, which is different from the cup-specific expression of *MYB12* in *C. g.* ssp. *sonomensis*.

A



Gene	Stage 1	Stage 2	Stage 3			Stage 4		
	White petal	White petal, basal spot	Background	Middle	Base	Background	Middle	Base
<i>CgaMYB6</i>	++ (NA)	+++	+++	(+) ++	+ (+) (+)	+++	+++	(+) (+) (+)
<i>CgaMYB11</i>			(+) +	+		+++		
<i>CgaMYB12</i>								
<i>CgaGAPDH</i>	+++	+++	+++	+++	+ (+) +	+++	+++	(+) ++

B



Gene	Stage 1	Stage 2	Stage 3		Stage 4	
	White petal	White petal, central spot	Background	Base	Background	Base
<i>CahMYB6</i>	+++	+++	+++	+ (+)	+++	+ (+) +
<i>CahMYB11</i>					+++	
<i>CahMYB12</i>					+ (+)	++
<i>CahGAPDH</i>	+++	+++	+++	++ (+)	+++	+++

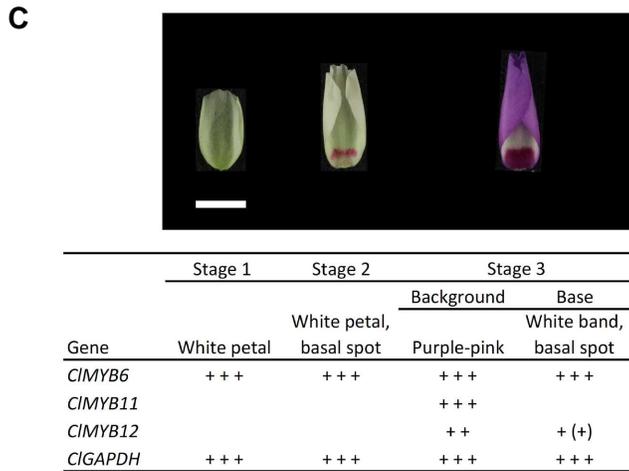


Figure 12: Expression patterns of *MYB6*, *MYB11* and *MYB12* across developmental stages. (A) *Clarkia gracilis* ssp. *albicaulis*; (B) *C. amoena* ssp. *huntiana* and (C) *C. lassenensis* across the flower-bud developmental stages. Based on the PCR-band brightness on the gels, the expression levels were scored as “+”, “(+)”, and blank, respectively representing expressed, weakly expressed and not expressed. A “+” or “(+)” sign shows a data point from one plant. “(NA)” means no data for this plant, because all the small buds collected from this plant already have basal spots. Pictures above the columns designate the bud phenotypes. The scale bar indicates 5 mm.

2.4 Discussion

2.4.1 The petal *R2R3-MYB* genes evolved through ancient duplications prior to polyploidization of *Clarkia gracilis*

In this study, we found that the *R2R3-MYB* genes expressed in the petal of these *Clarkia* (sub)species evolved through ancient duplications occurred before the speciation of *C. gracilis*. This conclusion is supported by the presence of *MYB1*, *MYB6*, *MYB11* and *MYB12* orthologs in the parental species of *C. gracilis*, the conserved spatiotemporal expression patterns of these genes and their phylogenetic relationship. First, *MYB6* has the same spatiotemporal expression pattern in *C. g. ssp. sonomensis*, *C. g. ssp. albicaulis*, *C. amoena* ssp. *huntiana* and *C. lassenensis*, as does does *MYB11*. In these four (sub)species,

MYB6 is expressed throughout the petal and across all the developmental stages, while *MYB11* is expressed in the petal background at the late developmental stage (Figure 12). The expression domain of *MYB1* is either the central or the basal region, forming the central spots in *C. g. ssp. sonomensis* and *C. amoena ssp. huntiana* or the basal spots in *C. g. ssp. albicaulis* (the different positions being determined by different alleles; Martins et al. 2017), and the basal spots in *C. lassenensis*. These patterns imply that the expression domains of these genes were established at least as far back as the common ancestor of *C. amoena ssp. huntiana* and *C. lassenensis*.

2.4.2 Model for the evolution of color patterns in *C. gracilis*

Our results allow us to distinguish between two hypotheses regarding the different *R2R3-MYB* genes in the tetraploid *C. gracilis*: (1) that they represent different orthologous copies from the progenitor species; or (2) that they represent novel duplications that arose within the *C. gracilis* lineage. The phylogenetic analysis clearly indicates that hypothesis (1) is correct. *MYB1*, *MYB11*, and *MYB12* are most similar to the corresponding genes from *C. amoena ssp. huntiana*, suggesting that the copies inherited from *C. lassenensis* have been either lost or downregulated in flower petals and thus do not contribute to pigmentation. By contrast, *MYB6* shows the opposite pattern, with the *C. gracilis* copies being most similar to the *C. lassenensis* sequence, suggesting loss or downregulation of the *C. amoena ssp. huntiana* copy in petals.

The combined results of the expression and phylogenetic analyses allow construction of a model of the evolution of pigment patterns after polyploidization (Figure 13). In this model, *MYB6* is expressed throughout the petal. The petal is divided into two regions, distal (background) and proximal (cup), with the boundary defined by the expression domain of *MYB11*. In addition, a spot, determined by the expression domain of *MYB1*, appears either in the central or basal region of the petal.

In the progenitors, *MYB12* is expressed throughout the petal, whereas *MYB11* was expressed only in the distal region of the petal and *MYB1* was expressed only in spots. This would have produced a petal that had a pigmented background and pigmented spots. Despite the expression of *MYB12^L* in the proximal region of the *C. lassenensis* petal, there is no pigmentation (apart from the spot), which could be explainable by *MYB12^L* being nonfunctional.

Immediately after polyploidization, *C. gracilis* presumably expressed all copies of these genes (Figure 13), which would have produced a petal with a pink background with both central and basal spots and lacking a white patch in the proximal region (because of the expression of the functional *MYB12^A* from *C. amoena* ssp. *huntiana*).

The phylogenetic evidence indicates that prior to subspeciation the copies of *MYB11^L* and *MYB12^L* from *C. lassenensis* were either lost or downregulated in petals. We infer that the *C. lassenensis* copy of *MYB1* was also lost or downregulated because the copies of this gene in the two subspecies of *C. gracilis* are more similar to the *C. amoena*

ssp. huntiana copy than to the *C. lassenensis* copy. Based on this reasoning, the petals of the common ancestor of the two subspecies of *C. gracilis* had a pink background throughout the petal and a single central spot.

After subspeciation, *C. g. ssp. sonomensis* underwent two changes: (1) the expression domain of *MYB12^A* contracted, such that it is expressed only in the proximal (cup) region, and (2) a new, loss-of-function allele, *MYB12^a*, arose to create a polymorphism. In *C. g. ssp. albicaulis*, *MYB12^A* became downregulated or was nonfunctionalized or deleted. Additionally, although the expression domain of *MYB11^A* in *C. g. ssp. albicaulis* remains the same, this domain was divided into two by the difference in expression level (Appendix II, Table A2.3). *MYB11^A* is expressed higher in the most distal region (pink background), and expressed less in the rest of the *MYB11^A* domain, which likely gives the “white band” in the middle of the petal. Finally, based on the orthology of *MYB1* in the two *C. gracilis* subspecies, a novel allele (*MYB1^a*) became fixed in the spot position being basal in *C. g. ssp. albicaulis*, causing a shift in the spot position from central to basal (Martins et al. 2017).

2.4.3 Polyploidization, gene loss, and the evolution of white patches

Overall, the evolution of color patterns in *C. gracilis* involved three regulatory changes (expression domain shift in *MYB1^a*, expression level change of *MYB11^A* in *C. g. ssp. albicaulis*, and expression domain contraction of *MYB12^A* in *C. g. ssp. sonomensis*), five regulatory or loss-of-function mutations (loss or downregulation of *MYB6^A*, *MYB1^L*,

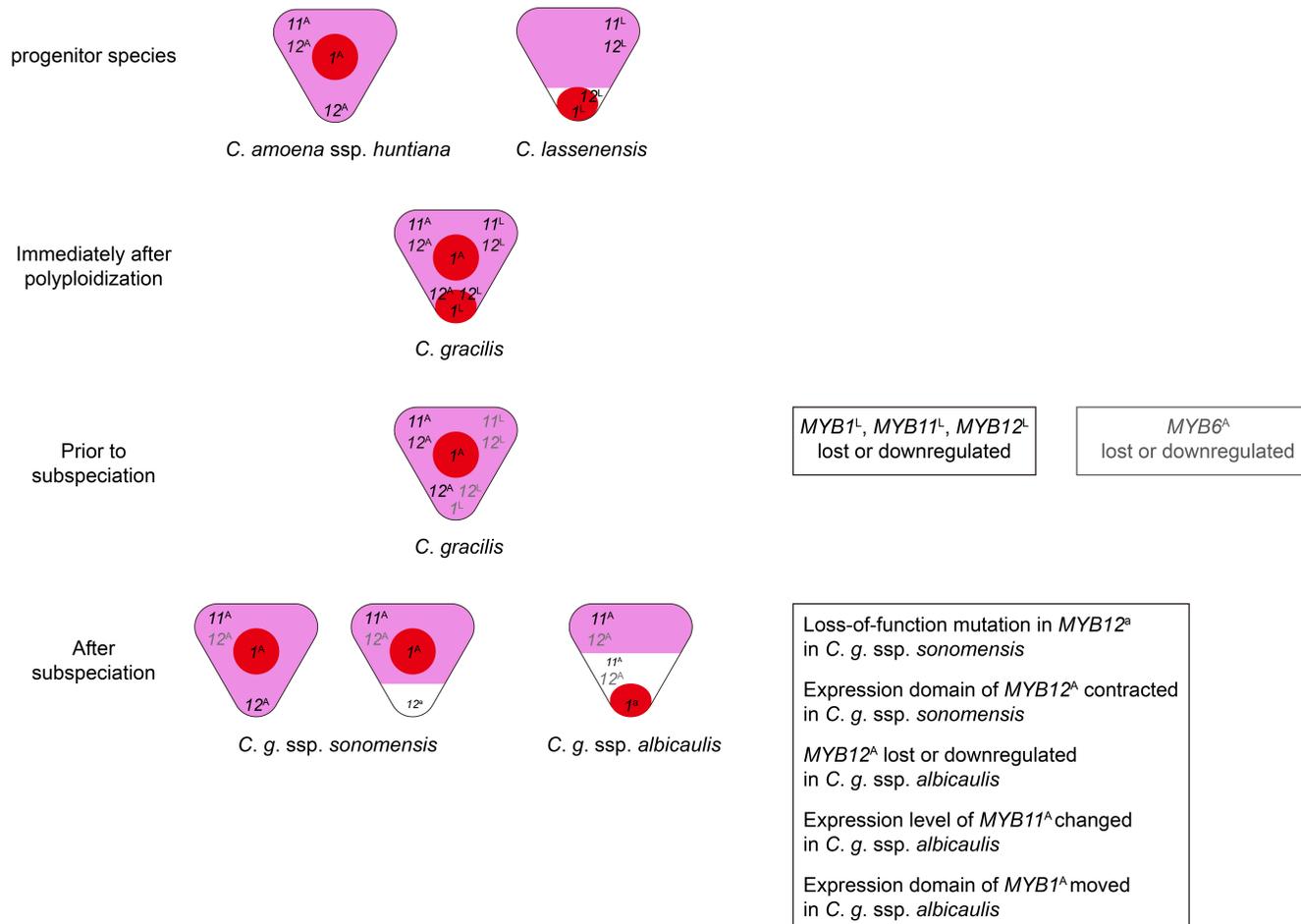


Figure 13: Model of petal pattern element evolution in *Clarkia*. R2R3-MYB genes are designated by number and letter superscript.

For example, *MYB12* in *C. amoena ssp. huntiana* is designated as *12^A*, while *MYB12* in *C. lassenensis* is designated as *12^L*. Genes in *C. gracilis* are those derived from the progenitor species. Genes indicated in gray are either lost, downregulated, or in the case of *C. g. ssp. sonomensis*, not expressed in the distal region. Superscript “a” indicates a different, newly arisen allele of the corresponding gene. For example, *MYB12^A* is the functional copy of *MYB12*, whereas *MYB12^a* is the nonfunctional copy. Different Font sizes indicate the expression level difference. For example, the expression level of *MYB11^A* is higher in the pink background than the white band in *C. g. ssp. albicaulis*. Genetic modifications after polyploidization are shown in the boxes.

MYB11^L, and *MYB12^L* in *C. gracilis* prior to subspeciation and loss or downregulation of *MYB12^A* in *C. g. ssp. albicaulis*), and one definitive loss-of-function mutation (*MYB12^a* in *C. g. ssp. sonomensis*). Determination of whether the five ambiguous changes are regulatory or loss-of-function changes must await further genomic analyses.

Nevertheless, it is clear that both regulatory mutations and loss-of-function mutations in the network of *R2R3-MYB* genes have played important roles in the evolutionary diversification of floral pigment patterns in *Clarkia*.

Polyploidization, which is a form of whole genome duplication, initially creates two copies of orthologous genes from the progenitors. After any transient tetrasomy, tetrasomic inheritance is resolved to diploid inheritance. Our results are consistent with previous studies of gene loss following polyploidization. Specifically, polyploidization seems to be often followed by loss/silencing of one copy in many but not all genes (Kashkush et al. 2002; Ainouche et al. 2004; Kellis et al. 2004; Lai et al. 2004; Paterson et al. 2004; Soltis et al. 2004; Adams & Wendel 2005). In addition, retained duplicates often

undergo functional divergence (subfunctionalization), including divergence in expression domains (Blanc & Wolfe 2004, Throude et al. 2009). This appears to have occurred with *MYB12* in the common ancestor of *C. lassenensis* and *C. amoena* ssp. *huntiana*, with carryover into *C. gracilis*, where expression of the former occurs throughout the petal, while that of the latter is restricted to the distal (cup) region. This pattern produces redundancy in the distal region, which means that downregulation or nonfunctionalization of *MYB12* creates a white patch only in the proximal region, as seen in *C. lassenensis* and *C. g. ssp. sonomensis*.

In summary, the duplication of *R2R3-MYB* genes in the common ancestor of *C. lassenensis* and *C. amoena* ssp. *huntiana* has apparently facilitated petal color pattern diversification in this clade of *Clarkia*. Along with gene loss or downregulation of duplicate copies, it has allowed the emergence of three color pattern elements (background, spots and white patches) that are controlled by different *R2R3-MYB* genes. This independent control has allowed spots to appear and change positions, and white patches to appear and change in size independently. In particular, this independence of control of different pattern elements has allowed four different color patterns to have evolved: central spots with and without a white patch, and basal spots with and without a white patch. It will be interesting to see whether this type of combinatorial evolution of pattern elements occurs in other species.

3. Downregulation of anthocyanin genes contributes to an anther-color polymorphism in two trout lilies (*Erythronium*)

3.1 Introduction

A fundamental goal of evolutionary biology is to understand how phenotypic variation arises and is maintained between and within species. Polymorphisms within a single species or population have attracted particular attention because, as predicted by theory, genetic drift or directional natural selection will eventually eliminate the intra-specific variation. Thus, the evolutionary mechanisms that maintain such variation have long intrigued biologists. For example, the maintenance of floral color polymorphisms has been extensively studied. To date, floral color polymorphisms have been attributed to various forms of balancing selection, including selective forces mediated by pollinator preference (Jones 1996b; Gigord et al. 2001; Jones & Reithel 2001; Eckhart et al. 2006; Austen et al. 2018; Ison et al. 2019), temporal and spatial variation in the magnitude and direction of natural selection (Schemske & Bierzychudek 2001, 2007), opposing selection imposed by herbivores, pathogens and pollinators (Frey 2004), and frequency-dependent pollen transmission bias (Rausher and Fry 1993; Subramaniam & Rausher 2000). However, in all of these studies, the evidence of balancing selection has been provided by field experiments that quantify the pattern of selection on flower-color variants.

There is a potential problem with this approach: because most of these studies have been limited to single populations over just one or a few generations, it is unclear whether the patterns uncovered extend over long periods of time (Rausher & Delph 2015). Therefore, it would be desirable to use other methods of examining balancing selection to corroborate inferences drawn from field experiments. One such alternative method is to detect a signature of long-term selection on the gene(s) underlying the phenotypic polymorphism. Unfortunately, to our knowledge, this approach has not been possible, because the genetic basis of floral-color polymorphisms is usually unknown. In the present study, we attempt to facilitate this approach by identifying the gene(s) responsible for an anther-color polymorphism in the trout lilies (*Erythronium*).

The anther/pollen color polymorphism in *Erythronium* (Liliaceae) has been known for decades (Parks & Hardin 1963; Thomson & Stratton 1985). In general, anther color and pollen color are consistent: purple anthers contain purple pollen and yellow anthers contain yellow pollen. The observed *E. umbilicatum* populations in North Carolina ($N = 16$) contain both purple- and yellow-anthered individuals, and the purple morph is always more common (80-96%; Appendix III, Table A3.1). This polymorphism has been a system for the study of pollination biology such as pollen transport and pollinator effectiveness (Bernhardt 1977; Motten 1983; Thomson & Stratton 1985; Thomson 1986; Thomson et al. 1986; Thomson & Thomson 1989; Holsinger & Thomson

1994; Thomson 2010). Recently, a study of *E. americanum* (Austen et al. 2018) revealed no detectable difference between purple/red and yellow pollen in pollen consumption by pollen-feeding beetles, UV-B tolerance or siring success. However, site-specific pollen-color preferences of pollinators may act as the agent of selection on pollen color. Unfortunately, given the longevity, perenniality and iteroparity (multiple bouts of reproduction) of *Erythronium*, measuring lifetime fitness is difficult. Thus, applying genetic approaches to this system may provide new insight and extend our understanding on the evolution of the anther color polymorphism in *Erythronium*. Here, we present molecular and genetic analyses to determine the pigments and genes involved in anther color differences in two eastern trout lilies, *E. umbilicatum* and *E. americanum*.

In the vast majority of angiosperms, the colors red, pink, purple and blue in flowers are derived from anthocyanin pigments (Grotewold 2006). The production of anthocyanins involves relatively few enzyme-coding genes and associated transcription factors, which are collectively called “Anthocyanin Biosynthetic Pathway (ABP)” (Figure 13). The colors of white, yellow, and green flowers are usually due to lack of anthocyanin accumulation. Loss of pigmentation can result from (i) mutations in the coding region in any of the ABP enzyme-coding genes, leading to the production of dysfunctional enzymes; (ii) mutations in the *cis*-regulatory region in any of the ABP enzyme-coding genes, leading to decreased production of the enzymes; (iii) either

coding or *cis*-regulatory mutations in any of the ABP-associated transcription factors that regulate the expression of the enzyme-coding genes. Despite examples of all these possibilities as shown by spontaneous mutations (Streisfeld & Rausher 2010), evolutionary fixation of unpigmented flowers tends to target regulatory elements, particularly a group of R2R3-MYB transcription factors (Quattrocchio et al. 1999; Schwinn et al. 2006; Streisfeld & Rausher 2010). This is presumably because mutations in R2R3-MYB regulators involve fewer pleiotropic effects than knockouts of enzyme-coding genes (Streisfeld & Rausher 2010). Thus, this pattern, i.e., scenario (iii), is expected for loss of pigmentation in anthers.

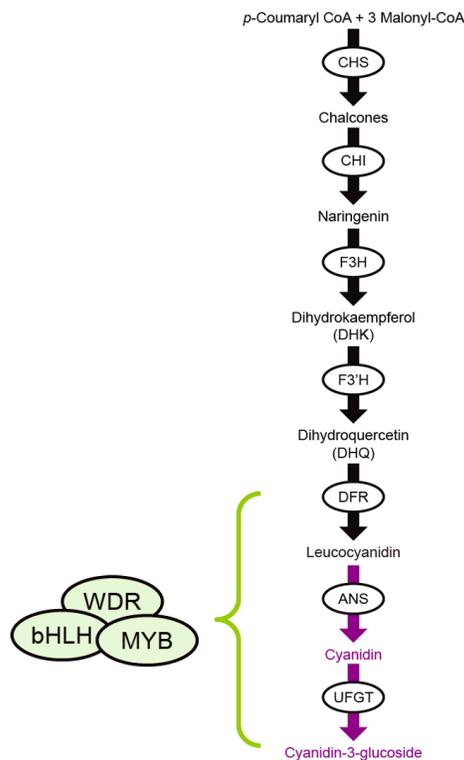


Figure 14: A simplified schematic diagram of the anthocyanin biosynthetic pathway. Enzymes are shown in circles: CHS, chalcone synthase; CHI, chalcone isomerase; F3H,

F3H, flavanone-3-hydroxylase; F3'H, flavonoid 3'-hydroxylase; DFR, dihydroflavonol-4-reductase; ANS, anthocyanidin synthase; UF3GT, UDP-flavonoid-3-O-glucosyl-transferase. The transcription factors are shown in green circles: MYB, myeloblastosis; bHLH, basic Helix-Loop-Helix; WDR, WD (tryptophan-aspartic acid) repeat.

The ABP-associated transcription factors include MYB, bHLH and WDR proteins. These proteins form a highly conserved MYB-bHLH-WDR (MBW) protein activation complex that controls some or all of the enzyme-coding genes in the pathway (Ramsay & Glover 2005; Petroni & Tonelli 2011; Xu et al. 2015). The MYB and bHLH proteins constitute two of the largest transcription factor families in plants. These two groups of transcription factors have been divided into subgroups according to their roles in regulating specific physiological processes (Stracke et al. 2001; Heim et al. 2003; Dubos et al. 2010). Most plant MYB proteins belong to R2R3-MYBs that have two adjacent MYB repeats (R2 and R3). R2R3-MYB proteins of subgroup 6 generally regulate anthocyanin production by physically interacting with the bHLH proteins of subgroup IIIf. The WDR proteins facilitate MYB-bHLH interactions by serving as a platform for the interacting proteins (Baudry et al. 2004).

Such MBW protein complexes are not limited to anthocyanin regulation. They are also involved in other plant developmental processes. For example, TTG1, the WDR protein in *Arabidopsis*, is not only an anthocyanin regulator, but is also central to the development of additional characters, such as trichomes and seed coat mucilage. Similarly, the bHLH anthocyanin regulators in *Arabidopsis* (TT8, GL3, EGL3) are also involved regulation of other developmental processes (Ramsay & Glover 2005).

Therefore, mutations in WDR and bHLH are expected to influence several developmental processes and incur significant deleterious pleiotropic effects. By contrast, the anthocyanin R2R3-MYB regulators usually present in multiple copies in plants, and each copy is expressed in a tissue-specific fashion (Albert et al. 2011; Ramsay & Glover 2005). Mutations in one R2R3-MYB copy that controls anthocyanin pigmentation in a certain part of the flower will eliminate anthocyanins in that part only. Thus, among these three ABP-associated transcription factors, the pleiotropy associated with R2R3-MYB is predicted to be least (Streisfeld & Rausher 2010). Indeed, loss of petal color has been documented to involve changes in anthocyanin regulation by R2R3-MYBs in diverse species (e.g., Quattrocchio et al 1999; Schwinn et al. 2006; Streisfeld et al. 2013). By contrast, anthocyanin regulation in other parts of the flower is less understood. Non-petal pigmentation, including anther color, can be as important as the petal color for pollinator attraction, because when the non-petal color contrasts with the background, it could potentially influence pollinator behavior (Lunau 1991; Nicholls & Hempel de Ibarra 2014). Until now, the genetic investigation of anther pigmentation has been carried out only in two model species: maize (*Zea mays*) and *Petunia* (Cone et al. 1993; Hollick et al. 2000; Albert et al. 2011) . To understand whether natural variation in anther color involves variation in R2R3-MYB genes, we need additional studies from non-model species.

The goals of this study are to identify the pigments and gene(s) that are responsible for the anther colors in *Erythronium*, and to determine which of these three types of mutations -- (i) coding mutations in enzyme-coding genes (ii) *cis*-regulatory mutations in enzyme-coding genes (iii) coding or non-coding mutations in transcription factors -- can explain the anther color difference. We apply our knowledge of the anthocyanin pathway, bioinformatics tools, gene expression analyses and a likelihood-estimation procedure to identify the candidate gene(s) contributing to the anther color polymorphism in *Erythronium*.

3.2 Materials and Methods

3.2.1 Study system

The genus *Erythronium* (Liliaceae) includes 25-30 species that all are perennial spring-ephemerals. Variation in anther/pollen color has been documented in at least three *Erythronium* species (Parks & Hardin 1963; Thomson & Stratton 1985), including *E. umbilicatum* Parks & Hardin subsp. *umbilicatum* (Figure 15A) and *E. americanum* Ker Gawl subsp. *americanum*. The anther/pollen color of these species is either purple/red or yellow. The distributions of *E. umbilicatum* and *E. americanum* are mainly allopatric (Appendix III, Figure A3.1), but with some areas of sympatry: *E. umbilicatum* occurs commonly in the deciduous forest in southeastern United States, while *E. americanum* is widespread in northeastern North America. Both species are largely self-incompatible. Reproductive plants of both species produce two leaves and a single yellow,

hermaphroditic flower. The green leaves are often irregularly mottled with brown-purple splotches that contain anthocyanins.

E. americanum is a tetraploid and also known to reproduce clonally through stolon-like droppers (Robertson 1906). By contrast, *E. umbilicatum* lacks stolons (Parks & Hardin 1963), suggesting that clonal reproduction is relatively uncommon in this species. Possibly because of this difference, populations that are monomorphic for anther color have been reported in *E. americanum* (Austen et al. 2018), but all examined *E. umbilicatum* populations are polymorphic (Appendix III, Table A3.1).

3.2.2 Sample collection

Before anther dehiscence, the nodding floral buds of *E. umbilicatum* were collected in March 2015 and March 2016 in the Oosting Natural Area (35°58'48.5''N, 79°03'54.7''W) of the Duke Forest in Orange County, North Carolina and brought to the lab. After petals and sepals were removed, the anthers were scored as purple or yellow, and then stored at -80°C until use. *E. americanum* samples were collected and provided by Emily Austen (Mount Allison University). *E. americanum* buds were collected in May 2015 and May 2016 in Koffler Scientific Reserve (40°01'47.1''N, 79°32'0.5''W), Ontario, Canada and Gatineau Park (45°36'01.8''N, 76°02'34.08''W), Quebec, Canada. Floral buds at a developmental stage similar to that of the collected *E. umbilicatum* were used in this study. Anthers were scored as either red, orange (the intermediate between red and yellow) or yellow.

3.2.3 Characterization of anthocyanidins in anthers

We characterized the pigments in *Erythronium* anthers using high performance liquid chromatography (HPLC). Anthocyanidins, the aglycone precursors of anthocyanins, were extracted following the method in Harborne (1984) with modifications. We pooled 12 anthers (approximately 60 mg) from each color group in *E. umbilicatum* and six anthers from each color group in *E. americanum*. Anthers were incubated in 1 mL 2N HCl overnight. The 2N HCl supernatants were heated in a boiling water bath for 60 min. The cooled 2N HCl extracts were washed twice with 500 μ L ethyl acetate. The ethyl acetate layer and aqueous layer were separated by centrifugation at 13,000 rpm for 1 min. The ethyl acetate supernatants were removed and evaporated under vacuum, and the remaining aqueous layer was washed once with 150 μ L isoamyl alcohol. After centrifugation, the isoamyl alcohol layer that is enriched with anthocyanidins was collected. The isoamyl alcohol supernatants were dried using a rotary evaporator, resuspended in 100 μ L methanol with 1% (v/v) HCl, and stored in a -20 °C freezer. On the next day, 50 μ L of the elution was injected on a Shimadzu LC-10AT liquid chromatograph with a 4.6 x 150 mm Alltech Prevail reverse phase C18 column (Alltech Associates, Deerfield, IL) at a flow rate of 1 mL/min. Anthocyanidins were separated by gradient elution at 30 °C using solvents A (HPLC-grad water, 0.1% trifluoroacetic acid) and C (1-propanol, 0.1% trifluoroacetic acid) with the following program: 15% C from 0 to 4 min; linear increase to 20% C from 4 to 10 min; 20% C from

10 to 14 min; linear increase to 22.5% C from 14 to 16 min; instantaneous increase to 27.5% C; 27.5% C from 16 to 18 min; instantaneous decrease to 15% C; 15% C from 18 to 21 min. Peaks were detected at 520 and 540 nm. Anthocyanidins were identified by comparison with standard solutions of delphinidin, petunidin and peonidin from Polyphenols Laboratories (Sandnes, Norway), and cyanidin, malvidin, and pelargonidin from Indofine Chemical Company (Hillsborough, NJ, USA).

3.2.4 Transcriptome sequencing

RNA sequencing was performed for different anther morphs in *E. umbilicatum* and *E. americanum*. Total RNA was extracted from one anther of each individual using Spectrum Plant Total RNA Kit (Sigma). Equal amounts of total RNA from the anthers of the same color were then pooled. For *E. umbilicatum*, we pooled RNA from 50 purple- and 50 yellow-anthered individuals. For *E. americanum*, we pooled RNA from 10 red-, 6 orange-, and 3 yellow-anthered individuals. The pooled RNA was run with Bioanalyzer Agilent RNA 6000 Nano Kit (Agilent Technologies) to ensure that all samples had good-quality RNA for library construction. We used KAPA Stranded mRNA-Seq Kit (KAPA Biosystems) with 4 ug of the pooled RNA to construct libraries of each color group. The libraries were barcoded using NEBNext Multiple Oligos for Illumina (New England BioLabs). The quality of the resulting libraries was examined with Bioanalyzer Agilent High Sensitivity DNA Kit (Agilent Technologies). These libraries were sequenced on an

Illumina HiSeq 4000 platform performing 150 bp paired-end reads at the Duke University Sequencing & Genomic Technologies Shared Resource Center.

3.2.5 Bioinformatic analyses

Raw Illumina reads were first trimmed to remove adaptor and barcode sequences using Trimmomatic 0.36 (Bolger et al. 2014) bundled in the Trinity 2.4.0 software package (Grabherr et al. 2011). Only the trimmed, paired-end reads were used in transcriptome assembling. The transcriptome references were *de novo* assembled using Trinity. To search for the anthocyanin gene sequences in the resulting transcriptomes, we ran the tblastx program in NCBI-BLAST-2.6.0+ (<ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/2.6.0/>; Camacho et al. 2009), using the sequences from *Arabidopsis*, *Petunia*, *Lilium*, and *Tulipa* as queries (Appendix III, Table A3.2). Gene expression levels were estimated as FPKM values (the number of RNAseq fragments per kilobase of transcript effective length per million fragments mapped to all transcripts) using RSEM 1.3.0 (Li & Dewey 2011) built into the Trinity package. Scripts used to run these analyses are listed in Appendix III, A3.1. The bitscores of BLAST hits and FPKM values were both considered when selecting the subjects (reference sequences).

3.2.6 Cloning of full-length coding sequences

Based on the gene expression profile revealed in the transcriptome data (see Results), we selected and amplified the full-length coding regions of five anthocyanin

enzyme-coding genes (*EuChs*, *EuF3h*, *EuDfr*, *EuAns* and *EuUf3gt*) and the anthocyanin-regulating transcription factors (*EuMYB3*, *EubHLH2*, *EuWDR1* and *EuWDR2*) from *E. umbilicatum*. *EubHLH1* was not included because, although several attempts have been made, the amplification of *EubHLH1* has never been successful.

The coding regions of *EuChs*, *EuF3h*, *EuMYB3*, *EuWDR1* and *EuWDR2* were amplified with cDNA samples from the anther tissue of one purple- and one yellow-anthered individuals. Because amplification of *EuDfr*, *EuAns*, *EuUf3gt* and *EubHLH2* with yellow anther cDNA was not possible due to absence of expression (see Results), we used genomic DNA (gDNA) from a yellow-anthered individual as templates in PCRs. Additionally, we first examined whether the same copies of these genes are expressed in anthers and leaves, using the anther and leaf cDNA samples from one purple-anthered plant. After the same copies were confirmed, we used cDNA from the leaf tissue of a yellow-anthered individual to obtain the gene sequences from the yellow-anthered individual.

cDNA of each sample was synthesized with total RNA extracted from the anther or leaf tissue. Before making cDNA, gDNA was first removed from RNA using RQ1 RNase-Free DNase (Promega). DNase-treated RNA (0.5 µg) was used to synthesize cDNA in a 20 µL reaction with 200U of M-MuLV Reverse Transcriptase (New England BioLabs) and 1µM of Oligo d(T)₁₈ mRNA Primer (New England BioLabs). gDNA was extracted from the anther tissue using the cetyltrimethylammonium bromide (CTAB)

protocol (Doyle & Doyle 1987). PCR primers (Appendix III, Table A3.3) were designed based on the transcriptome reference sequences directly, except for *EubHLH2*. Because we only retrieved a short fragment of *EubHLH2* from the transcriptome, primers were first designed to anneal to the conserved regions of *bHLH2* from *Lilium* spp. (*LhbHLH2*, GenBank no. AB222076) and *Tulipa fosteriana* (*TfbHLH2*, GenBank no. KF924736). The sequences obtained with these primers were used to design the primers for genome walking using Universal GenomeWalker 2.0 Kit (Clontech) to obtain the 5' region of *EubHLH2*. With the results of genome walking, we then designed primers to amplify the full-length of *EubHLH2*.

PCR reactions were conducted using Q5 High-Fidelity DNA Polymerase (New England BioLabs) with the touchdown PCR program: denaturation at 98 °C for 30 sec, followed by 20 cycles of 98 °C for 10 sec, 68-48 °C for 30 sec (decreasing the annealing temperature by 1 °C per cycle), and 72 °C for 1 min, and 20 cycles of 98 °C for 10 sec, 48 °C for 30 sec, and 72 °C for 1 min, and final extension at 72 °C for 2 min. PCR products were gel-purified using QIAquick Gel Extraction Kit (Qiagen), phosphorylated using T4 Polynucleotide Kinase (New England BioLabs), ligated with the pCR-Blunt vector (Invitrogen), and transformed into 5-alpha *Escherichia coli* cells (New England BioLabs). Multiple colonies per ligation were sequenced by Sanger Sequencing (Eton Bioscience, San Diego, CA). SEQUENCHER 5.0 (Gene Codes, Ann Arbor, MI) was used to correct basecalling errors and align sequence fragments.

To evaluate whether the *R2R3-MYB* and *bHLH* genes we identified from the *E. umblicatum* anthers are homologs to the known anthocyanin regulators, we performed phylogenetic analyses to construct phylogenetic trees of R2R3-MYB and bHLH proteins. The *EuMYB3* copy that is most frequent among the examined purple-anthered individuals (see Results) was chosen. Its coding sequence was translated into amino acid sequence, and aligned with other related R2R3-MYB protein sequences using MUSCLE (Edgar 2004). A neighbor-joining phylogenetic tree was constructed using MEGA 6.06 (Tamura et al. 2013) with the JTT amino acid substitution model. Clade support was estimated by 1000 bootstrap replicates. A phylogenetic tree of the bHLH proteins was constructed following the same procedure with one *EubHLH2* sequence from a purple-anthered individual.

3.2.7 Quantification of gene expression in anthers

Gene expression levels of five enzyme-coding genes (*EuChs*, *EuF3h*, *EuDfr*, *EuAns* and *EuUf3gt*) and two groups of transcription factors (*EubHLH2*, *EuWDR1* and *EuWDR2*) were further analyzed using quantitative real-time PCR (qPCR). The MYB transcription factor (*EuMYB3*) was not included because multiple gene copies were detected (see Results). High sequence similarity among these copies made it impossible to design copy-specific primers. The anther cDNA samples were synthesized as described above and then diluted to 2.5 ng/μL for qPCR.

The qPCR primers (Appendix III, Table A3.3) were designed to amplify 80-150 bp fragments of the selected genes and *EF1 α* (Elongation factor 1-alpha, serving as a reference gene). Primer specificity was examined by visualizing the PCR products on 2% agarose gels and confirmed by sequencing the PCR products. Each 20 μ L qPCR reaction contained 10 μ L of DyNAmo HS SYBR Green qPCR master mix (Thermo Scientific), 0.3 μ M of each primer and 2.5 ng cDNA template. Ten biological replicates for each color group and two technical replicates for each sample were performed. Reactions were run on an Eppendorf Mastercycler RealPlex or a Roche LightCycler 96 with the following conditions: 95°C for 15 min, followed by 55 cycles of 95°C, 15 sec, 60°C for 30 sec, and 72°C for 30 sec. A melting curve analysis was run at the end of the cycle to verify that a single product was amplified. The reactions were repeated if the threshold (Ct) values of technical replicates differed by greater than 10%. A single purple-anthered individual was arbitrarily chosen as a control sample and was included in each run. PCR efficiency of each gene was calculated according to Peirson et al. (2003). The relative expression ratios of target genes were normalized with the expression levels of *EF1 α* , using equation 1 in Pfaffl (2001). Then the relative expression levels were calculated as the logarithm of the ratios.

We used the same qPCR protocol to examine the expression levels of *EaChs*, *EaF3h*, *EaDfr*, *EaAns* and *EaUf3gt* in the red, orange, and yellow anthers of *E. americanum*, except that three biological replicates for each color group were used.

3.2.8 Semi-quantification of gene expression in leaves and anthers

The expression of four genes (*EuDfr*, *EuAns* and *EuUf3gt* and *EubHLH2*) were not detectable in the yellow anthers. However, given the leaves of *E. umblicatum* are often mottled with brown-purple splotches containing anthocyanins, we assumed that these four genes should be expressed in the leaves of yellow-anthered individuals. We tested this assumption by performing PCRs with anther cDNA and leaf cDNA samples from three purple- and three yellow-anthered individuals using the same primers in the qPCR assays. PCR products were visualized on 2% agarose gels. *EF1 α* was included as a positive control.

Because expression of anthocyanin-regulating *R2R3-MYB* genes is often tissue-specific, we expected *EuMYB3* to be expressed in anther tissue but not in leaf tissue. We tested this hypothesis by amplifying the full-length of *EuMYB3* with the primers MYB3-1F and MYB3-Q2R (Appendix III, Table A3.3), using anther and leaf cDNA samples from three purple- and three yellow-anthered individuals. PCR products were visualized on 1.2% agarose.

3.2.9 Likelihood estimation of *cis*-regulatory mutations causing triple downregulation

The qPCR results indicated that three enzyme-coding genes (*EuDfr*, *EuAns* and *EuUf3gt*) are downregulated in yellow anthers of all *E. umblicatum* samples (see Results). To evaluate the hypothesis that independent *cis*-regulatory mutations in all three genes could explain this result, we calculated the maximum probability that a yellow-anthered

individual was homozygous for the downregulated “yellow” allele at each of the three loci. In Appendix III, A3.2, it is shown that this probability is

$$\text{Prob}(ddaaauu) = q^2 s^2 v^2 / \text{Freq}(Y) \quad (1),$$

where, q , s , and v are the unknown frequencies of the “yellow” allele at the three loci (d , a , and u), and $\text{Freq}(Y)$ is the frequency of yellow-anthered individuals in the population. Although this probability depends on the allele frequencies, we obtained a conservative estimate of this probability by finding its maximum, P_{\max} . To find the values of the unknown frequencies that make this probability maximum, we used the FindMaximum function in MATHEMATICA 12 (Wolfram Research, Champaign, IL), subject to the constraint $q^2 (1 - s^2) (1 - v^2) + s^2 + v^2 - s^2 v^2 = \text{Freq}(Y)$, to find the maximum of Equation (1). $\text{Freq}(Y)$ was estimated by performing the anther-color survey in the Oosting Natural Area during the flowering season in 2017 (Appendix III, Table A3.1). The probability that all n individuals of a sample are homozygous “yellow” at all three loci is then $(P_{\max})^n$. A small value of this probability is inconsistent with downregulation at all three loci being caused by independent *cis*-regulatory mutations in the three genes.

3.2.10 Construction of *EuMYB3* gene tree

Initial sequencing of *EuMYB3* indicated the presence of multiple *EuMYB3* copies. We then isolated individual copies through cloning. The full-length of *EuMYB3* was amplified from 10 purple- and 10 yellow-anthered individuals with the primers MYB3-1F and MYB3-Q2R (Appendix III, Table A3.3), using gDNA as templates. PCR products

were cloned as described above, and 10 colonies per construct were sequenced. Sequences were proofread and aligned using SEQUENCHER. The similarity of these sequences was evaluated by the phylogenetic analyses using the maximum likelihood algorithm in MEGA 6.06 (Tamura et al. 2013) with the GTR model. Clade support was assessed by performing 1000 bootstrap replicates

3.3 Results

3.3.1 Cyanidin-derived anthocyanins occur in purple anthers only

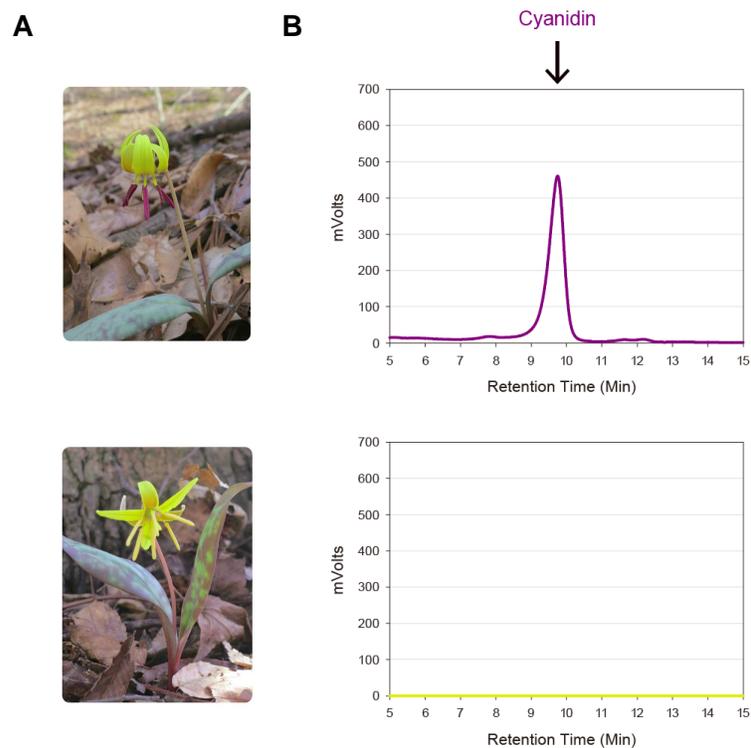


Figure 15: Anthocyanidins in the anthers of *Erythronium umbilicatum*. (A) Purple- and yellow-anthered *E. umbilicatum*. (B) HPLC traces show anthocyanidins in different morphs of anthers.

The purple anthers of *E. umbilicatum* and the red and orange anthers of *E. americanum* have cyanidin-derived anthocyanins (Figure 15B). The amount of

anthocyanins in the *E. americanum* orange anthers is much less than that in the red anthers, which likely explains the difference in color to human eyes. In the yellow anthers of both species, there is no detectable anthocyanin (Figure 16).

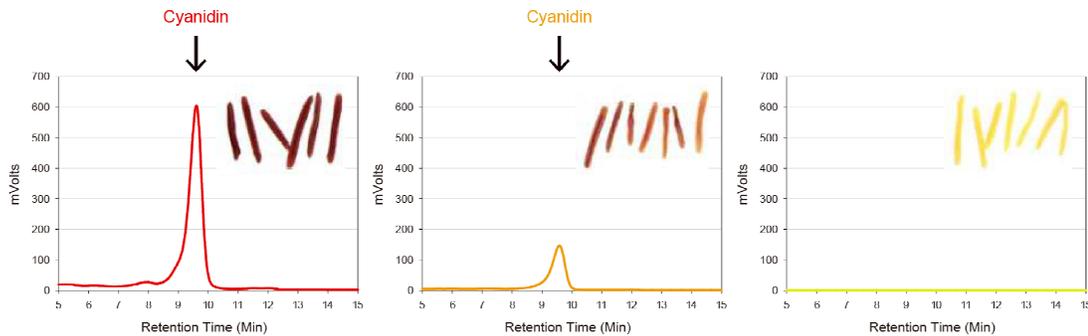


Figure 16: Anthocyanidins in the anthers of *Erythronium americanum*. HPLC traces show anthocyanidins in the red, orange and yellow morph of the anthers.

3.3.2 Identification of anthocyanin genes

BLAST searches against the transcriptomes identified seven ABP enzyme-coding genes (*EuChs*, *EuChi*, *EuF3h*, *EuF3'h*, *EuDfr*, *EuAns* and *EuUf3gt*) and three groups of the ABP-associated transcription factors (*EuMYB3*, *EubHLH1*, *EubHLH2*, *EuWDR1* and *EuWDR2*) (Table 3). Analyses of transcript abundance in the transcriptomes showed that *EuDfr*, *EuAns*, *EuUf3gt* and *EubHLH2* have the FPKM values of zero or very close to zero in the transcriptome of the yellow anthers. Based on these findings, we selected five enzyme-coding genes (*EuChs*, *EuF3h*, *EuDfr*, *EuAns* and *EuUf3gt*) as well as the transcription factors (*EuMYB3*, *EubHLH2*, *EuWDR1* and *EuWDR2*) for further analyses. *EubHLH1* was not included because amplification of *EubHLH1* has never been successful.

Table 3: Gene expression profile in the transcriptomes of purple and yellow anthers of *Erythronium umbilicatum*. Each contig represents a different gene, which has

several isoforms determined by Trinity. *The BLAST bitscore of each contig is shown as the highest bitscore among its isoforms. We found that in most cases, two/three contigs were BLASTed to a single anthocyanin gene, where the sequences of the two/three contigs are complementary (in two different directions). Gene expression levels were estimated as transcript abundance (FPKM values) with two runs that trimmed reads were mapped to the transcriptome references of the purple and yellow anther separately.

<i>E. umbilicatum</i> purple anther transcriptome					
Contig ID	Gene	Direction	BLAST bitscore*	FPKM	
				Purple anther	Yellow anther
EuP_32295_c0_g2	<i>EuChss</i>	3'-5'	893	11.91	25.31
EuP_33596_c0_g1	<i>EuChs</i>	5'-3'	893	11.08	22.98
EuP_37117_c0_g1	<i>EuChi</i>	3'-5'	395	23.44	24.06
EuP_37117_c0_g2	<i>EuChi</i>	5'-3'	398	17.77	17.46
EuP_36256_c0_g1	<i>EuF3h</i>	3'-5'	735	9.97	4.12
EuP_36256_c0_g2	<i>EuF3h</i>	5'-3'	736	9.15	4.35
EuP_38329_c0_g1	<i>EuF3'h</i>	5'-3'	730	11.18	7.55
EuP_38329_c0_g2	<i>EuF3'h</i>	3'-5'	1018	6.23	3.82
EuP_39244_c0_g1	<i>EuDfr</i>	3'-5'	791	20.9	0
EuP_39244_c0_g2	<i>EuDfr</i>	5'-3'	788	23.53	0.05
EuP_35520_c0_g1	<i>EuAns</i>	3'-5'	525	8.38	0
EuP_35520_c1_g1	<i>EuAns</i>	5'-3'	525	7.76	0.05
EuP_39893_c0_g1	<i>EuUf3gt</i>	3'-5'	863	14.02	0
EuP_39893_c0_g2	<i>EuUf3gt</i>	5'-3'	866	7.85	0.04
EuP_21744_c0_g1	<i>EubHLH1</i> (EGL3)	3'-5'	270	1.49	0.32
EuP_52832_c0_g1	<i>EubHLH2</i> (TT8)	3'-5'	219	0.91	0
EuP_30989_c0_g2	<i>EuWDR1</i> (TTG1)	5'-3'	394	2.77	1.86
EuP_33697_c0_g1	<i>EuWDR2</i> (AN11)	3'-5'	289	4.45	4.47

<i>E. umbilicatum</i> yellow anther transcriptom					
Contig ID	Gene	Direction	BLAST bitscore*	FPKM	
				Purple anther	Yellow anther
EuY_40055_c0_g6	<i>EuChs</i>	3'-5'	893	13.73	22.76
EuY_40055_c0_g7	<i>EuChs</i>	5'-3'	893	13.27	22.02
EuY_38665_c0_g1	<i>EuChi</i>	3'-5'	397	32.12	29.88
EuY_38665_c0_g2	<i>EuChi</i>	5'-3'	397	21.86	18.34
EuY_17862_c0_g1	<i>EuF3h</i>	3'-5'	386	8.22	3.04
EuY_25883_c0_g1	<i>EuF3h</i>	5'-3'	736	9.49	3.58

EuY_36308_c1_g8	<i>EuF3'h</i>	3'-5'	281	8.63	4.28
EuY_36308_c1_g9	<i>EuF3'h</i>	5'-3'	603	6.79	4.08
EuY_36308_c0_g2	<i>EuF3'h</i>	3'-5'	269	7.94	4.42
EuY_17157_c0_g2	<i>EuMYB3</i>	5'-3'	280	1.35	1.92
EuY_34064_c1_g2	<i>EuWDR1</i> (TTG1)	3'-5'	282	1.91	1.62
EuY_34064_c1_g1	<i>EuWDR2</i> (AN11)	5'-3'	289	5.85	5.01

With the same procedure, we also identified the same genes from the *E. americanum* transcriptomes (Table 4), which has the similar expression patterns that *EaDfr*, *EaAns*, *EaUf3gt* and *EabHLH2* are downregulated in the yellow anthers.

Table 4: Gene expression profile in the transcriptomes of red, yellow and orange anthers of *Erythronium americanum*. Each contig represents a different gene, which has several isoforms determined by Trinity. *The BLAST bitscore of each contig is shown as the highest bitscore among its isoforms. Gene expression levels were estimated as transcript abundance (FPKM values) with three runs that trimmed reads were mapped to the transcriptome references of the red, yellow and yellow anther separately.

<i>E. americanum</i> red anther transcriptome					
Contig ID	Gene	BLAST bioscore	FPKM		
			Red anther	Yellow anther	Orange anther
RCL1_34292_c0_g1	<i>EaChs</i>	575	2.23	6.98	1.74
RCL1_48421_c0_g2	<i>EaChi</i>	392	16.66	40.67	25.27
RCL1_44998_c0_g1	<i>EaF3h</i>	736	19.67	23.84	22.79
RCL1_38537_c0_g1	<i>EaF3'h</i>	513	3.09	2.3	4.69
RCL1_36156_c0_g1	<i>EaDfr</i>	577	1.89	0.12	1.24
RCL1_41745_c0_g1	<i>EaAns</i>	520	5.18	0.1	6.19
RCL1_29273_c0_g1	<i>EaUf3gt</i>	401	2.06	0	2.6
RCL1_35636_c0_g1	<i>EaUf3gt</i>	261	2.5	0	2.68
RCL1_31925_c0_g1	<i>EaMYB3</i>	214	2.15	3.2	1.63
RCL1_36140_c0_g1	<i>EabHLH1</i>	612	4.17	0.53	0.68
RCL1_5552_c0_g1	<i>EabHLH2</i>	140	0.96	0.52	0
RCL1_87664_c0_g1	<i>EabHLH2</i>	189	0.33	0	0
RCL1_44719_c1_g1	<i>EaWDR1</i>	376	5.48	7.31	7.12
RCL1_44719_c1_g2	<i>EaWDR2</i>	289	22.03	32.04	29.54

E. americanum yellow anther transcriptome

Contig ID	Gene	BLAST bioscore	FPKM		
			Red anther	Yellow anther	Orange anther
RCL2_44576_c0_g1	<i>EaChs</i>	869	1.94	6.03	1.44
RCL2_51181_c0_g3	<i>EaChi</i>	393	28.93	39.85	31.41
RCL2_48601_c0_g1	<i>EaF3h</i>	738	16.99	12.62	14.99
RCL2_43442_c0_g1	<i>EaF3'h</i>	248	2.05	2.01	3.24
RCL2_43543_c0_g1	<i>EaMYB3</i>	208	2.3	3.68	2.64
RCL2_49915_c0_g1	<i>EaWDR1</i>	394	6.78	6.86	7.73
RCL2_49915_c0_g2	<i>EaWDR2</i>	289	13.16	13.12	14.43

E. americanum orange anther transcriptome

Contig ID	Gene	BLAST bioscore	FPKM		
			Red anther	Yellow anther	Orange anther
RCL3_25735_c0_g1	<i>EaChs</i>	620	2.56	6.93	1.31
RCL3_45077_c0_g2	<i>EaChi</i>	243	21.86	32.67	18.41
RCL3_41516_c0_g1	<i>EaF3h</i>	736	18.19	17.21	12.17
RCL3_38915_c0_g1	<i>EaF3'h</i>	509	3.61	2.36	3.94
RCL3_61937_c0_g1	<i>EaDfr</i>	288	1.1	0	1.13
RCL3_38694_c0_g1	<i>EaAns</i>	523	7.17	0.1	5.03
RCL3_39362_c0_g1	<i>EaUf3gt</i>	757	3.05	0	3.52
RCL3_37141_c0_g1	<i>EaMYB3</i>	282	2.11	3.82	2.36
RCL3_29314_c0_g1	<i>EabHLH1</i>	152	1.17	0	0.9
RCL3_40606_c0_g3	<i>EaWDR1</i>	376	6.55	6.97	5.38
RCL3_40606_c0_g2	<i>EaWDR2</i>	289	16.01	19.68	12.42

We compared the coding sequences of the enzyme-coding genes *EuChs*, *EuF3h*, *EuDfr*, *EuAns* and *EuUf3gt* amplified from a purple- and a yellow-anthered individuals, and found no deletions or nonsense mutations. However, there were 1-7 nonsynonymous substitutions in each of these genes (Table 5; Appendix III, Figures S3.2-S3.6).

Table 5: Nucleotide sequence differences in the coding regions of five anthocyanin enzyme-coding genes between purple- and yellow-anthered *E. umblicatum*.

Gene	Full length (bp)	No. of nucleotide differences (bp)	No. of synonymous substitutions	No. of nonsynonymous substitutions
<i>EuChs</i>	1182	1	0	1
<i>EuF3h</i>	1101	8	5	3
<i>EuDfr</i>	1098	6	5	1
<i>EuAns</i>	1083	14	9	5
<i>EuUf3gt</i>	1371	19	12	7

3.3.3 Three enzyme-coding genes are downregulated simultaneously in the yellow anthers

Analyses of transcript abundance in the *E. umblicatum* transcriptomes showed that expression levels are similar in purple and yellow anthers for the upstream enzyme-coding genes *EuChs*, *EuChi*, *EuF3h*, and *EuF3'h*. By contrast, expression levels of the downstream enzyme-coding genes *EuDfr*, *EuAns* *EuUf3gt* are zero or very close to zero in the transcriptome of yellow anthers.

To verify these patterns, we performed qPCR on a subset of these genes. The upstream genes *EuChs* and *EuF3h* exhibit no detectable expression differences between yellow and purple anthers (Student's *t*-test, *EuChs*: $t = -0.006$, $P = 0.995$; *EuF3h*: $t = 0.913$, $P = 0.373$). By contrast, the downstream gene *EuDfr*, *EuAns* and *EuUf3gt* are expressed at greatly reduced levels in yellow anthers compared to purple anthers (Mann-Whitney test, all $P < 0.001$) (Nonparametric Mann-Whitney test was used for downstream genes, instead of Student's *t*-test, because there were non-detectable expression levels in yellow anthers) (Figure 17).

Similar qPCR analyses on the *E. americanum* samples showed the same pattern. The expression levels of *EaChs* and *EaF3h* did not differ significantly between anther color morphs (Student's *t*-test, all $P > 0.4$). However, the expression levels of *EaDfr* and *EaAns* in the yellow anthers were significantly lower than that in the red anthers (*EaDfr*: $t = 5.924$, $P = 0.004$; *EaAns*: $t = 4.412$, $P = 0.012$), while the expression level of *EaUf3gt* was not detectable in the yellow anthers (Figure 18).

3.3.4 *Cis*-regulatory mutations in the enzyme-coding genes are not likely responsible for triple downregulation

Downregulation of the three downstream enzyme-coding genes could be due either to independent *cis*-regulatory mutations in the yellow alleles of each gene or to coordinated downregulation by a change in a transcription factor in yellow-anthered individuals. The former hypothesis seems unlikely because all 10 individuals examined exhibited downregulation of all three genes. If these genes are unlinked or only moderately linked, we would expect to see individuals with different combination of these genes downregulated (e.g., only *EuDfr* downregulated, only *EuAns*, only *EuUf3gt*, *EuDfr* and *EuAns* but not *EuUf3gt*, etc.). We determined how unlikely this hypothesis is by statistical modelling.

Under the hypothesis that all examined yellow-anthered individuals represent independent downregulation of each of the genes *EuDfr*, *EuAns*, and *EuUf3gt* (i.e., the genotype *ddaauu*), we calculated the maximum probability that an individual was triply downregulated given it had yellow anthers by maximizing equation (1) over possible

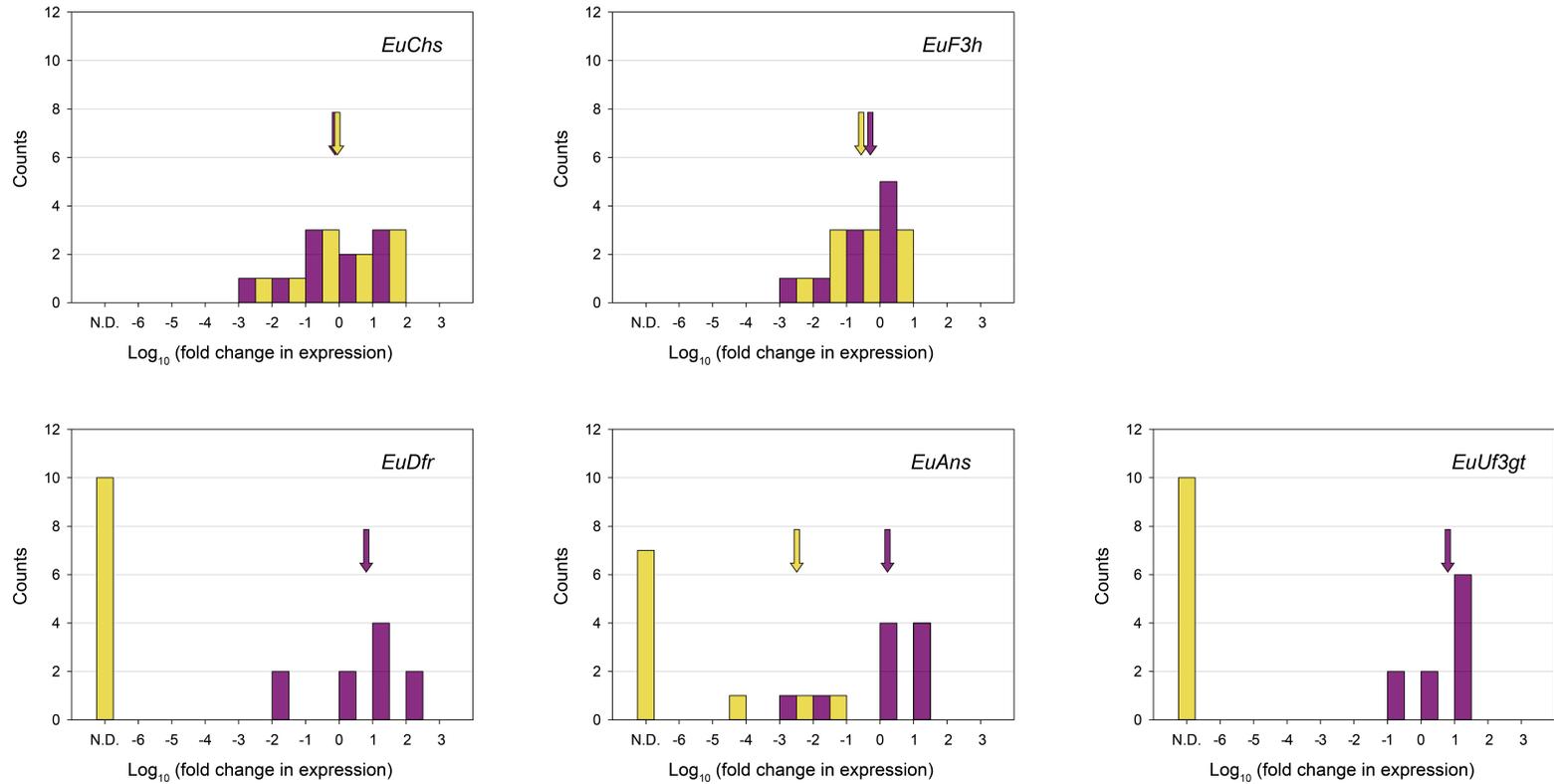


Figure 17: Expression of five enzyme-coding genes between 10 purple- and 10 yellow-anthered *E. umblicatum*. Purple bars represent purple-anthered individuals, and yellow bars represent yellow-anthered individuals. Arrows indicate the mean for each group.

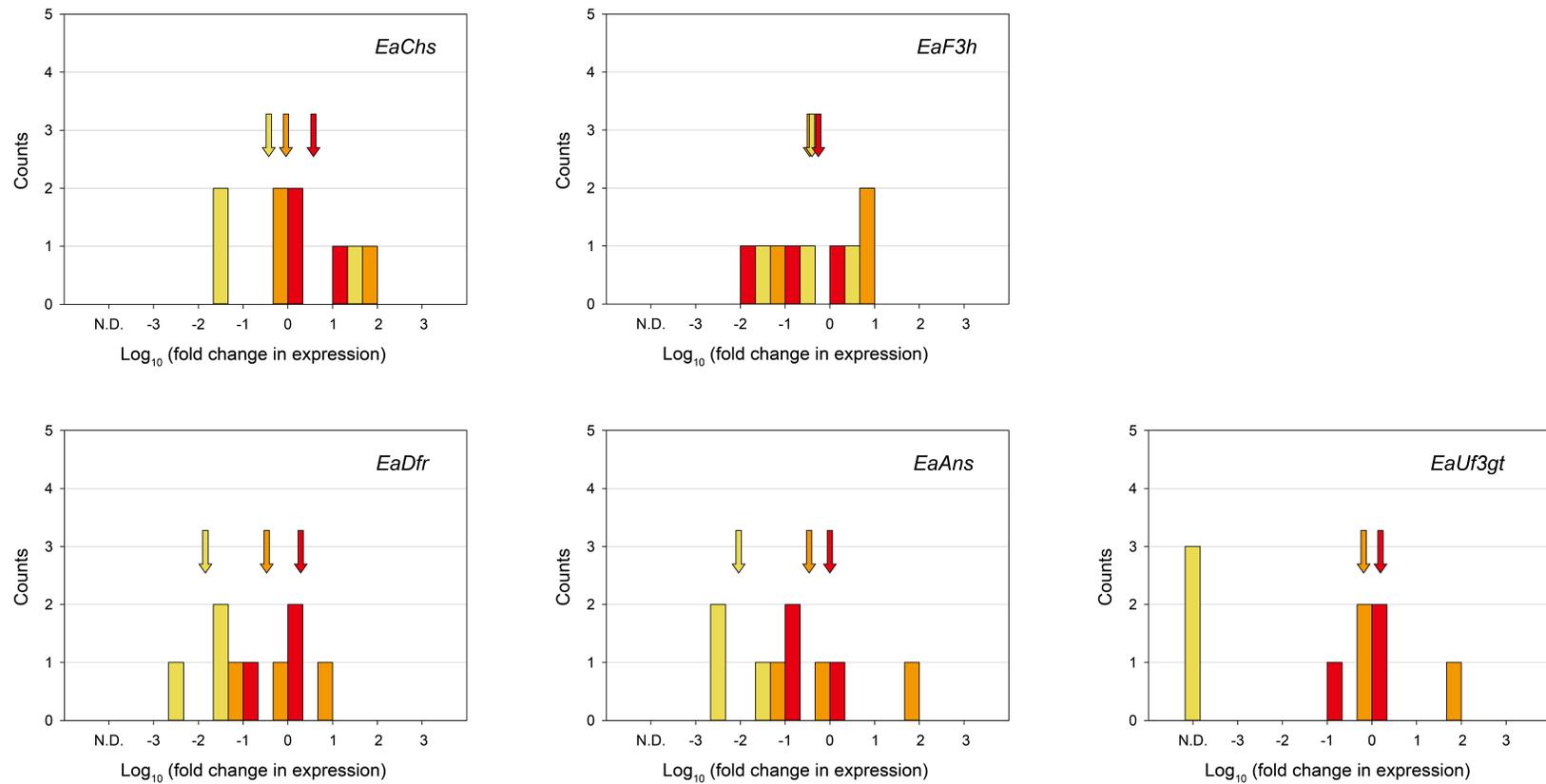


Figure 18: Expression of five enzyme-coding genes between 3 purple-, 3 orange and 3 yellow-anthered *E. americanum*. Red, orange and yellow bars represent red-, orange- and yellow-anthered individuals, respectively. Arrows indicate the mean for each group.

“yellow” allele frequencies. In equation (1), $\text{Freq}(Y)$ was estimated to be 0.183 with the survey over 213 plants in the Oosting Natural Area (Appendix III, Table A3.1). The maximum probability of 0.00151 was obtained when the frequencies of s and u were 0.2553. The probability that all 10 sampled individuals are $ddaaau$ is thus less than or equal to $(0.00151)^{10} = 6.21 \times 10^{-29}$. We thus reject the hypothesis that the joint downregulation is caused by independent *cis*-regulatory mutations in *EuDfr*, *EuAns*, and *EuUf3gt*.

3.3.5 Downregulated genes are not also nonfunctional

While down-regulation of the downstream genes can explain lack of anthocyanin production in yellow anthers, it is also possible that one or more of these enzymes have been functionally inactivated. To examine this possibility, we sequenced these three genes from leaf RNA from purple-anthered plants. The sequences of all three genes were identical to those of the corresponding genes expressed in anthers, indicating that the same gene are expressed in leaves and flowers. In addition, semi-quantitative PCR shows that although these downregulated genes were not expressed in the yellow anthers, they are expressed in the leaves of the yellow-anthered plants (Figure 19). Consequently, because yellow-anthered individuals produce leaf anthocyanins, these genes cannot be non-functional in plants with yellow anthers.

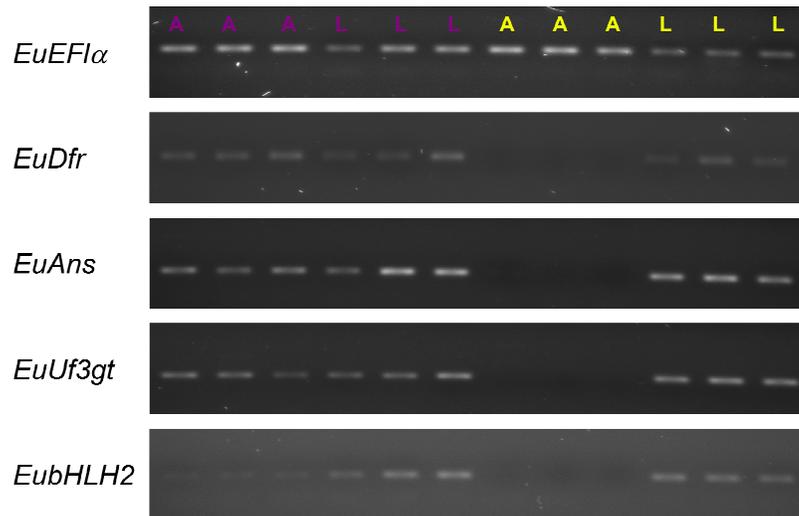


Figure 19: Semi-quantitative real-time PCR of the downregulated genes including three enzyme-coding genes and a *bHLH* gene in the anther (A) and leaf (L) tissues from three purple- and three yellow-anthered individuals. The housekeeping gene *EuEF1α* was included as a positive control.

3.3.6 A single set of transcription factors is likely responsible for triple downregulation

Based on the results described above, we infer that the triple downregulation is likely caused by mutation(s) in a common transcription factor. To evaluate this hypothesis, we examined the nucleotide sequences and expression levels of the ABP-associated transcription factors identified from the transcriptomes. We obtained the partial sequences of *EuWDR1* and the full-length sequences of *EuWDR2* from one purple- and one yellow-anthered individuals. None of the nucleotide differences were nonsynonymous (Appendix III, Figures S3.7, S3.8), indicating that these genes are functional in yellow-anthered individuals. In addition, *EuWDR1* and *EuWDR2* were expressed at the similar levels in purple and yellow anthers (Student's *t*-test, *EuWDR1*: $t = 1.427$, $P = 0.173$; *EuWDR2* : $t = 1.148$, $P = 0.267$; Figure 20).

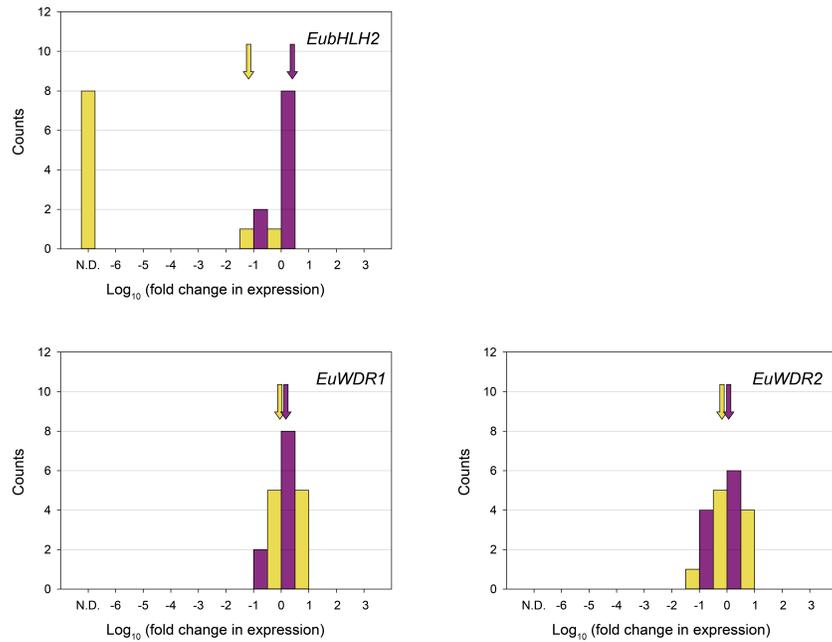


Figure 20: Expression of three anthocyanin-associated transcription factors between 10 purple- and 10 yellow-anthered *E. umblicatum*. Purple bars represent purple-anthered individuals, and yellow bars represent yellow-anthered individuals. Arrows indicate the mean for each group.

We infer that *EubHLH2* is an anthocyanin-regulating transcription factor because it clusters with other anthocyanin-regulating bHLH proteins of the subgroup IIIIf in a gene tree (Figure 21). Like *EuDfr*, *EuAns*, and *EuUf3gt*, the expression of *EubHLH2* was barely detectable in the yellow anthers (Mann-Whitney Test, $P < 0.001$; Figure 20), but was detectable in the leaves of yellow-anthered individuals. The downregulation of this gene likely prevents the formation of the transcription-factor complex that is required for initiating the transcription of the downstream enzyme-coding genes, and thus likely at least in part explains lack of anthocyanins in yellow-anthered individuals.

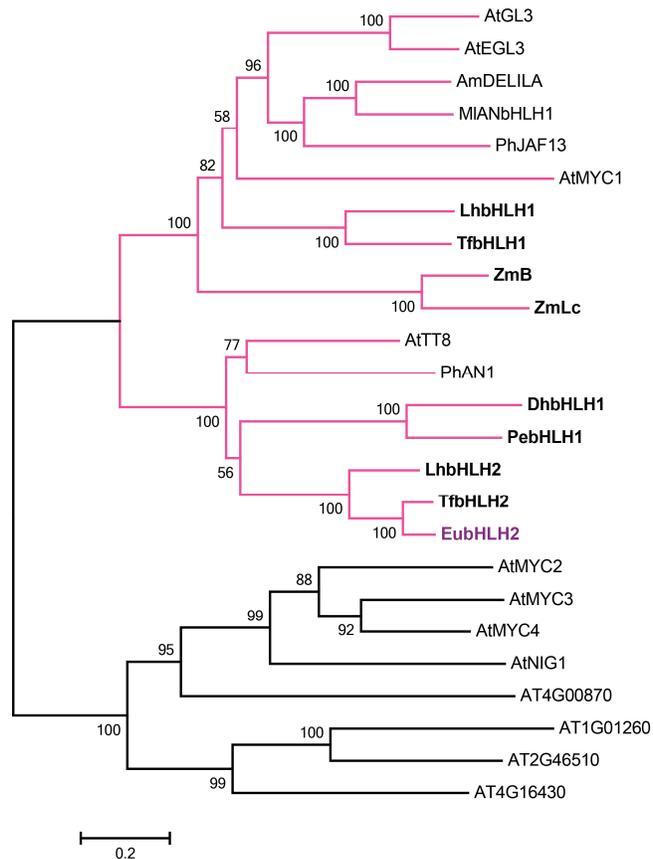


Figure 21: The neighbor-joining phylogenetic tree of bHLH proteins. The clade containing bHLHs of subgroup IIIf is shown in pink. The bHLH from the anther of *Erythronium umbilicatum* is shown in purple. The anthocyanin-regulating bHLHs from monocots are shown in bold. Bootstrap support values > 50% are shown. The *Arabidopsis* sequences were retrieved from TAIR (<https://www.arabidopsis.org/>): AtTT8 (AT4G09820), AtEGL3 (AT1G63650), AtGL3 (AT5G41315), AtMYC1 (AT4G00480), AtMYC2 (AT1G32640), AtMYC3 (AT5G46760), AtMYC4 (AT4G17880), AtNIG1 (AT5G46830). Other sequences were retrieved from GenBank: *Antirrhinum majus* AmDELILA (AAA32663); *Dendrobium* spp. DhbHLH1 (AQS79853); *Lilium* spp. LhbHLH1 (BAE20057), LhbHLH2 (BAE20058); *Mimulus lewisii* MIANbHLH1 (AHJ80985); *Petunia x hybrida* PhAN1 (AAG25928), PhJAF13 (AAC39455); *Phalaenopsis equestris* PebHLH1 (AIS35934); *Tulipa fosteriana* TfbHLH1 (AGL98426), TfbHLH2 (AHY20033); *Zea mays* ZmB (CAA40544), ZmLc (AAA33504).

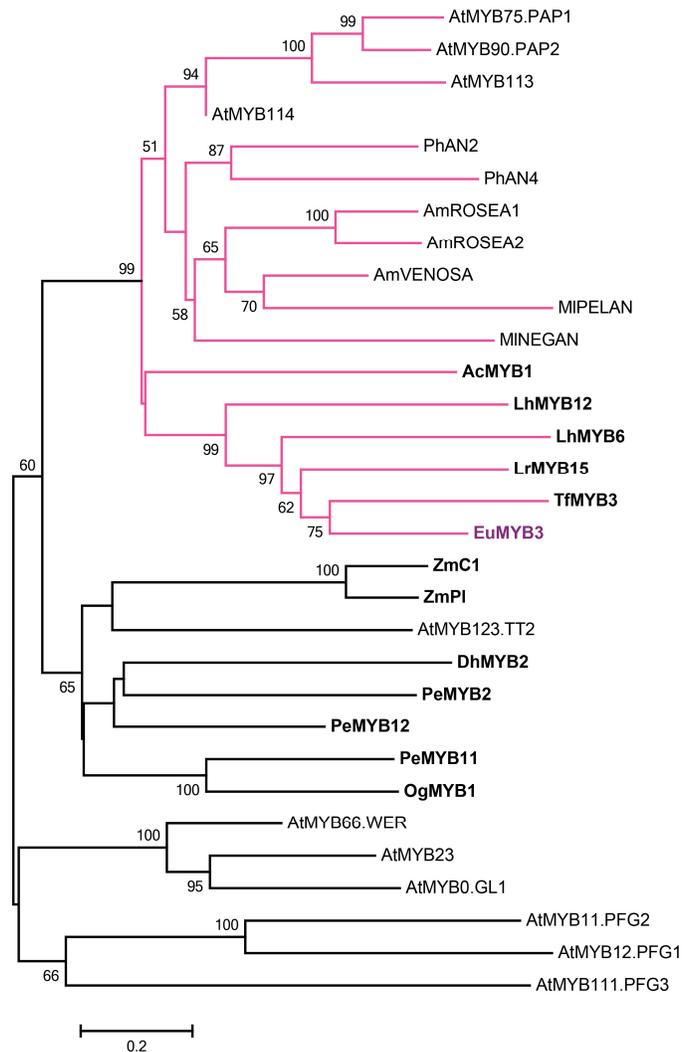


Figure 22: The neighbor-joining phylogenetic tree of R2R3-MYB proteins. The clade containing R2R3-MYBs of subgroup 6 is shown in pink. The R2R3-MYB from the anther of *Erythronium umbilicatum* is shown in purple. The anthocyanin-regulating R2R3-MYBs from monocots are shown in bold. Bootstrap support values > 50% are shown. The *Arabidopsis thaliana* sequences were retrieved from TAIR (<https://www.arabidopsis.org/>): AtMYB0 (AT3G27920), AtMYB11 (AT3G62610), AtMYB12 (AT2G47460), AtMYB23 (AT5G40330), AtMYB66 (AT5G14750), AtMYB75 (AT1G56650), AtMYB90 (AT1G66390), AtMYB111 (AT5G49330), AtMYB113 (AT1G66370), AtMYB114 (AT1G66380), AtMYB123 (AT5G35550). Other sequences were retrieved from GenBank: *Allium cepa* AcMYB1 (AQP25672); *Antirrhinum majus* AmROSEA1 (ABB83826), AmROSEA2 (ABB83827), AmVENOSA (ABB83828); *Dendrobium* spp. DhMYB2 (AQS79852); *Lilium* spp. LhMYB6 (BAJ05399), LhMYB12 (BAJ05398); *Lilium regale* LrMYB15 (BAU29929); *Mimulus lewisii* MIPELAN

(AHJ80987), MINEGAN (AHJ80988); *Oncidium* Gower Ramsey OgMYB1 (ABS58501); *Petunia x hybrida* PhAN2 (AAF66727), PhAN4 (ADQ00392); *Phalaenopsis equestris* PeMYB2 (AIS35919), PeMYB11 (AIS35928), PeMYB12 (AIS35929); *Tulipa fosteriana* TfMYB3 (AHY20034); *Zea mays* ZmC1 (P10290), ZmP1 (AAA19821).

Phylogenetic analysis of the *EuMYB3* sequence suggests that this gene likely also regulates the anthocyanin pathway in *E. umbilicatum*, since *EuMYB3* grouped with other anthocyanin-regulating R2R3-MYB proteins of subgroup 6 (Figure 22) and has a conserved motif “[K/R]P[R/Q]PR” that features all anthocyanin-regulating R2R3-MYB proteins (Appendix III, Figures S3.9). In purple-anthered individuals, *EuMYB3* was expressed in anthers but not in leaves (Figure 23), suggesting that a different R2R3-MYB gene is responsible for regulation of anthocyanins in leaves. In the yellow-anthered individuals, this gene was not expressed in either tissues. It was downregulated in 2 of 3 yellow-anthered individuals, but expressed in a third. The sequence in that individual contained an intron and a premature stop codon (Appendix III, Figure A3.10), suggesting that while expressed in this individual, it is likely nonfunctional.

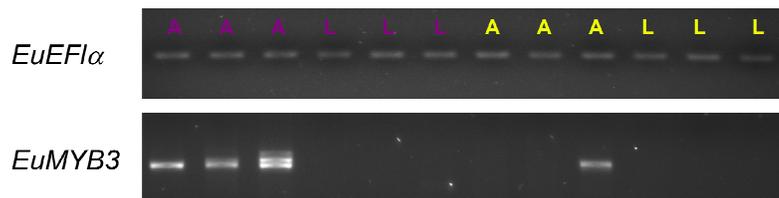


Figure 23: Semi-quantitative real-time PCR of *EuMYB3* in the anther and leaf tissues. cDNA samples from anther (A) and leaf (L) tissues of 3 purple- and 3 yellow-anthered individuals. PCRs were conducted with the primers, MYB3-1F and MYB3-Q2R, to amplify the full-length of *EuMYB3*. The housekeeping gene *EuEF1α* was included as a positive control.

3.3.7 Multiple copies of *EuMYB3*

When cloning *EuMYB3*, we found multiple sequences of this gene in single individuals. Specifically, we cloned the full-length *EuMYB3* gene from 10 purple- and 10 yellow-anthered individuals using gDNA as templates and sequenced 10 colonies for each individual. In total, we obtained 88 distinct *EuMYB3* sequences, 33 from the purple- and 55 from the yellow-anthered individuals (Table 6). None of the sequences were shared by the two morphs.

Table 6: Total number of *EuMYB3* sequences obtained from the purple- and yellow-anthered individuals. The number of sequences having premature stop codon(s) or deletion(s) are also shown.

	Purple	Yellow
Total no. of <i>EuMYB3</i> sequences	33	55
No. of <i>EuMYB3</i> sequences w/ premature stop codon(s)	4	35
No. of <i>EuMYB3</i> sequences w/ deletion(s) in the exons	1	8

The pairwise sequence divergence across all these sequences ranges from 0.1% to 10.1%. The majority (36 out of 55) of the sequences from the yellow-anthered individuals were predicted to be nonfunctional, due to the presence of premature stop codons or deletions in the exons. Such nonfunctional copies were also found in four of the purple-anthered individuals, likely because these individuals were heterozygous. Each

individual has 3-10 different sequences, suggesting that there are at least 2-5 gene copies of *EuMYB3* in *E. umbilicatum* (Table 7).

Table 7: The number of different *EuMYB3* sequences obtained from gDNA of each individual. P: purple-anthered individuals; Y: yellow-anthered individuals.

Sample ID	No. of different <i>EuMYB3</i> sequences	Sample ID	No. of different <i>EuMYB3</i> sequences
P02	5	Y01	8
P05	4	Y03	6
P13	6	Y04	6
P17	3	Y05	10
P19	4	Y06	8
P20	5	Y07	9
P22	6	Y09	6
P31	9	Y10	9
P32	7	Y14	7
P51	3	Y15	7

We performed phylogenetic analyses to elucidate the sequence similarity among these sequences. Most of the sequences from the purple-anthered individuals formed two clades, which differ from each other by 3.3% sequence divergence (Figure 24). The major nucleotide difference between these two clades is the presence of a 12-bp indel in intron 1.

3.4 Discussion

3.4.1 Downregulation of transcription factors explains absence of anthocyanins in yellow anthers

Our results clearly demonstrate that the changes in gene expression are responsible for the anther-color polymorphism in *E. umbilicatum*. The anther color difference is due to the presence/absence of cyanidin-derived anthocyanins: only the purple anthers have these anthocyanins. Absence of anthocyanins is correlated with marked downregulation of three downstream enzyme-coding genes (*EuDfr*, *EuAns* and *EuUf3gt*). Downregulation of multiple enzyme-coding genes suggests that changes in a single set of transcription factors are responsible for lack of anthocyanins in anthers. This conclusion is supported by several lines of evidence that rule out alternative explanations for the elimination of cyanidin production.

One alternative explanation for absence of anthocyanins is that mutations in these enzyme-coding genes have inactivated one or more of the downstream enzyme-coding genes in yellow anthers. We have ruled out this possibility by demonstrating that the functionality of *EuDfr*, *EuAns* and *EuUf3gt* from the yellow-anthered individuals should be intact. Specifically, we show that the same copies of *EuDfr*, *EuAns* and *EuUf3gt* are expressed in the leaf and anther tissues. Although these three genes are not expressed in the yellow anthers, they are expressed in the leaves of the yellow-anthered individuals (Figure 19). Since leaves produce anthocyanins, the proteins produced by these genes must be functional.

A second alternative explanation is that independent *cis*-regulatory mutations in *EuDfr*, *EuAns* and *EuUf3gt* cause downregulation of each of these genes in yellow anthers. Under this hypothesis, only one gene needs to be downregulated to eliminate anthocyanin production. Based on our statistical modeling, this possibility seems very unlikely: the probability of independent downregulation of all three genes in the 10 yellow-anthered individuals examined is vanishingly small.

By contrast, our results are consistent with downregulation of transcription factors causing the downregulation of the downstream enzyme-coding genes. Both *EuMYB3* and *EubHLH2* are markedly downregulated in yellow anthers (Figures 19, 20, 23). We note that this conclusion is valid in spite of the fact that there appear to be several copies of *EubMYB3* expressed in purple anthers. Because we used the same primers to identify the multiple copies and to assess expression levels, our failure to detect the expression of any functional copies in the yellow anthers indicates that all functional copies were downregulated. Although we do not know the mechanism by which this is accomplished, at least two are possible: (1) *cis*-regulatory mutation(s) causing downregulation may have occurred before gene duplication; or (2) there is a mutation in an upstream regulatory gene that activates all of these copies in purple anthers.

Several scenarios can account for the dual downregulation of *EuMYB3* and *EubHLH2*. First, dual downregulation may be due to independent *cis*-regulatory

mutations in each gene. Second, there may be a common upstream transcription factor that activates both of these genes. A third possibility is that one of these genes may activate the other, such that a mutation that downregulates the first will cause downregulation of the second.

Although the role of the MBW complex in regulating the anthocyanin enzyme-coding genes is well studied in the model species and is conserved across the flowering plants, the transcriptional regulation of the MBW components themselves is less clear. In *Arabidopsis thaliana*, the R2R3-MYB regulators, AtTT2 and AtPAP1 participate in the activation of the bHLH protein, AtTT8. This activation might be triggered by a direct binding between the R2R3-MYB and bHLH proteins (Baudry et al. 2006). Other factors regulating the expression of *AtTT8* include the WDR protein (AtTTG1) and auto-activation of *AtTT8* induced by the MBW complex (Baudry et al. 2004, 2006). In *Petunia*, the R2R3-MYB protein, AN4, is required for the expression of the *bHLH* gene, AN1 (Spelt et al. 2000). In Asiatic hybrid lily, *LhbHLH2* interacts with *LhMYB6* or *LhMYB12*. An allele of *LhMYB12* might positively regulate the expression of *LhbHLH2* (Yamagishi et al. 2010, 2014). However, in maize, the *bHLH* genes are not controlled by R2R3-MYBs (Carey et al. 2004). These considerations suggest that the downregulation of *EubHLH2* in the yellow anthers of *E. umbilicatum* may be attributed to *EuMYB3*. The causal *EuMYB3* copy may be not functional or not expressed in the yellow anthers, which could lead to the failure to form binding between *EubHLH2* and *EuMYB3* as well as the MBW

complex in the yellow anthers. As a consequence, the expression of *EubHLH2* would not be upregulated. After the causal copy of *EuMYB3* is determined, further investigations on the interaction of *EubHLH2* and *EuMYB3* will need to examine this hypothesis, as well as the other scenarios described above.

Coordinated expression of multiple enzyme-coding genes is common in the anthocyanin transcriptional regulation. In general, the pathway can be divided into two subsets: early biosynthetic genes (EBGs) and late biosynthetic genes (LBGs), although the genes grouped into EBGs or LBGs vary among species (Martin et al 1991; Xu et al. 2015). The genes within a single subset are regulated coordinately usually by an R2R3-MYB and/or a bHLH transcription factor. In eudicots, EBGs and LBGs are usually regulated separately by different sets of transcription factors, (Martin et al 1991; Quattrocchio et al. 1998; Streisfeld & Rausher 2009b; Yuan et al. 2014), although this does not appear to be true in *Ipomoea purpurea* (Tiffin et al. 1998). By contrast, in monocots, the studies of maize and Asiatic hybrid lily (*Lilium* spp.) revealed that both EBGs and LBGs (i.e., the entire enzyme-coding genes) are regulated by the same set of transcription factors (Dooner 1983; Lai et al. 2012). However, in the white hybrid of the orchid *Dendrobium* spp., only *F3h*, *Dfr*, and *Ans* are found to be downregulated coordinately (Kriangphan et al. 2015). Our findings also show that EBGs and LBGs are controlled separately in some monocots. In *E. americanum* and *E. umbilicatum*, the change in gene expression in *Dfr*, *Ans* and *Uf3gt* occurs simultaneously, but this change is not shared

with *Chs*, *Chi*, *F3h* and *F3'h*, which suggests that *Dfr*, *Ans* and *Uf3gt* belong to LBGs and are regulated coordinately. Interestingly, although this transcriptional regulation pattern is different from the pattern found in another species in Liliaceae and other monocot species, it matches to the pattern in some eudicot species, such as *Petunia* (Quattrocchio et al. 1993). Our results are thus consistent with the suggestion that the specific anthocyanin enzyme-coding genes that are controlled by a given set of activators is evolutionarily very labile (Rausher 2006).

Our finding that variation in anther color is due to modification of transcriptional regulation is consistent with previous findings for loss of pigmentation in floral petals, as well as with theoretical considerations. Many different kinds of mutations have been demonstrated to abolish pigmentation in floral petals, including loss-of-function and downregulating *cis*-regulatory mutations in anthocyanin pathway enzyme-coding genes and similar changes in anthocyanin-associated transcription factors (Rausher and Streifeld 2010). When multiple types of mutations produce the same phenotype, natural selection will tend to preferentially fix those with less deleterious pleiotropy. Loss-of-function mutations in enzyme-coding genes are expected to incur substantial deleterious pleiotropy because they are typically expressed in multiple tissues, and thus are not expected to contribute very often to evolutionary loss of pigmentation. Similarly, *cis*-regulatory mutations in these genes that cause downregulation in all tissues are expected to incur similar deleterious pleiotropy. By

contrast, *cis*-regulatory mutations that cause tissue-specific downregulation would not be expected to incur substantial pleiotropy and thus might contribute to evolutionary change in pigmentation. However, such mutations have seldom been found to become fixed in species that lack pigmentation, possibly because there is a small target size for such mutations (Streisfeld and Rausher 2010).

Among anthocyanin-associated transcript factors, loss-of-function mutations in *WDR* and *bHLH* genes are more likely to incur deleterious pleiotropy than *R2R3-MYB* genes. The former are typically expressed in multiple tissues and regulate other traits, including production of proanthocyanidins, vacuolar acidification, and production of cuticular structures such as trichomes and root hairs (Koes et al. 1994; Mol et al. 1998; Quattrocchio et al. 2006). *R2R3-MYBs*, on the other hand, typically regulate only the anthocyanin pathway and are often highly tissue-specific (e.g., Schwinn et al. 2006; Albert et al. 2011), which appears to be the case in *E. umbilicatum*. It is thus expected that knockout mutations in *R2R3-MYB* genes would preferentially be fixed over knockouts of *bHLH* or *WDR* genes when loss of pigmentation is advantageous. By similar argument, mutations that cause universal (non-tissue-specific) downregulation are expected to incur substantial deleterious pleiotropy than regulatory mutations in *R2R3-MYB* transcription factors.

Finally, as for enzyme-coding genes, *cis*-regulatory mutations that cause tissue-specific downregulation are not likely to cause substantially different levels of

deleterious pleiotropy among the different transcription factors, and thus selection is not expected to differentially favor fixation of any of the transcription factors over others. Overall, these considerations suggest that mutations that cause either loss of function or downregulation of *R2R3-MYB* genes are most likely to contribute to evolutionary loss of pigmentation.

Empirical evidence supports these theoretical arguments. Changes in anthocyanin-associated transcription factors have been documented to cause loss of petal pigmentation in *Petunia* (Quattrocchio et al. 1999), in *Antirrhinum majus* (Schwinn et al. 2006), in *Aquilegia* (Whittall et al. 2006), and in *Mimulus aurantiacus* (Streisfeld & Rausher 2009b; Streisfeld et al. 2013). These changes have involved almost exclusively mutations that either downregulate or abolish function in R2R3-MYB proteins (Streisfeld & Rausher 2011). Our results are consistent with these results, although there are a couple of caveats.

First, we cannot rule out the possibility that downregulation of *EuMYB3* and *EubHLH2* is caused by a mutation in a common upstream regulator. If so, however, this mutation would presumably incur little deleterious pleiotropy since it has no detectable effects on other characters. Such a situation would be consistent with the theoretical expectations above. Second, to the extent that downregulation of *EubHLH2* is independent of downregulation of *EuMYB3*, our results may indicate that change in a *bHLH* gene may contribute to loss of pigmentation. This would not violate theoretical

expectations if the mutation caused anther-specific downregulation since it would then not adversely affect the development of other traits controlled by bHLH proteins.

Alternatively, there may actually be pleiotropy mediated selection against the “yellow” *EubHLH* allele, which may contribute to lack of fixation of yellow anthers in *E.*

umbilicatum. Resolving these caveats must await further investigation.

3.4.2 Similar downregulation observed in *E. americanum* and *E. umbilicatum*

E. americanum and *E. umbilicatum* are the only two *Erythronium* species in the eastern North American that have been documented to exhibit anther-color polymorphisms. We found the same changes in gene expression of the enzyme-coding genes associated with the anther-color polymorphism in both species: the expression levels of *Dfr*, *Ans* and *Uf3gt* are significantly lower or non-detectable in the yellow anthers. We presume the pattern of transcriptional regulation is also similar in these species, though we have not examined it in *E. americanum*. Despite this similarity, the shared downregulation may not evolve independently in these two species. *E.*

americanum is tetraploid and is thought to be of hybrid origin, although its origin has not been examined adequately. Parks and Hardin (1963) proposed that *E. americanum* is a natural hybrid of the two diploid species, *E. rostratum* and *E. umbilicatum*, in the range where the distributions of these two species overlap, because the morphology of *E. americanum* is very similar to that of *E. rostratum* and *E. umbilicatum*. Parks and Hardin (1963) also mentioned that some characters of *E. americanum* are morphologically

intermediate to *E. rostratum* and *E. umbilicatum*. For example, the capsule apex of *E. rostratum* and *E. umbilicatum* are beaked (rostrate) and indented (umbilicate), respectively. In *E. americanum*, the capsule apex ranges in shape from rounded to truncate. Based on sequence similarities between *E. americanum* and other eastern North American *Erythronium* at the ITS region, Allen et al. (2003) agreed on the allotetraploid origin of *E. americanum*, but concluded that a parental species might either be extinct or not have been sampled. Clennett et al. (2012) suggested that *E. americanum* could be seen as autopolyploid of *E. rostratum* or *E. umbilicatum* based on phylogenetic analyses with three markers (two plastid introns and ITS). Despite no consensus on the origin of *E. americanum*, there is no doubt that *E. americanum* and *E. umbilicatum* are closely related. Thus, instead of independent evolution of the anther-color polymorphism in *E. americanum* and *E. umbilicatum*, it seems more likely that the polymorphism in *E. americanum* was derived from *E. umbilicatum* and then maintained separately in these two lineages. This would imply that this polymorphism has existed for a long evolutionary time, in spite of the environmental changes that these species may have experienced. This would be particularly true for *E. americanum* because *E. americanum* probably underwent range expansion after polyploidization, since *E. americanum* and *E. umbilicatum* are mostly allopatric nowadays. Pollinator attraction has been proposed as a selective agent on the anther color in *E. americanum* through field experiments (Austen et al. 2018). Although we have not yet identified the causal locus and therefore could not

perform tests to detect the signature of selection, our findings of the same downregulation underlying this long-standing polymorphism in the two species may correspond to the suggestion by Austen et al. (2018) that this polymorphism is maintained by the evolutionary forces, rather than neutral or random processes.

3.4.3 Conclusion

In this study, we determined that changes in regulatory elements, rather than changes in enzyme-coding genes, are responsible for the anther-color polymorphism in *E. umbilicatum*. These regulatory changes, including the functional or regulatory changes in *EuMYB3* and the change of expression levels in *EubHLH2*, cause the downregulation of three anthocyanin enzyme-coding genes (*EuDfr*, *EuAns* and *EuUf3gt*) in the yellow anthers. *EuMYB3* exhibits an anther-specific expression pattern, although it is present and expressed in multiple copies. Unfortunately, despite several attempts, we have not been able to determine the causal copy of *EuMYB3*. The genetic evaluation of *EuMYB3* is limited by high sequence similarity among the copies, the scarcity of genomic resources mainly due to the huge genome size of *Erythronium* (30-36 Gb, Leitch et al. 2007; Peruzzi et al. 2009), and a long life cycle (> 10 years from sowing to flowering in *E. japonicum*, Yokoi 1976) that makes creating hybrid progeny for genetic mapping infeasible. We hope that in the near future, new techniques or a better knowledge of the *Erythronium* genome will contribute to the identification of the causal *EuMYB3* copy, and encourage further investigations on the evolution of the anther-color polymorphism.

Conclusions

In this dissertation, I investigated genetics and evolution of petal pigmentation patterning in *Clarkia gracilis* ssp. *sonomensis* (Onagraceae) and an anther-color polymorphism in *Erythronium umbilicatum* (Liliaceae). Both polymorphisms involve loss of anthocyanin pigmentation in some individuals in the populations. In *C. g.* ssp. *sonomensis*, pigmentation in the basal region of the petal is polymorphic for pink and white (unpigmented), and in *E. umbilicatum*, anther color is polymorphic for purple and yellow (unpigmented). Although the two polymorphisms exist in two distantly related species, the genetic changes responsible for the polymorphisms located in the same group of transcription factors: the anthocyanin-regulating R2R3-MYBs, which is consistent with the common observation that evolutionary loss of pigmentation typically involves mutations in R2R3-MYB genes (Streisfeld & Rausher 2010). Preferential fixation of mutations in R2R3-MYB genes reflects minimal deleterious pleiotropy associated with inactivation or downregulation of R2R3-MYB genes, given the specificities of the anthocyanin-associated R2R3-MYB genes that regulate anthocyanin production only and also that target peculiar tissues (e.g., floral vs. vegetative) (Ramsay & Glover 2005; Quattrocchio et al. 2006; Streisfeld & Rausher 2010).

In addition, in both cases, the transcriptional regulation patterns of the identified R2R3-MYB genes (in terms of which enzyme-coding genes they regulate) appear to be novel, compared with the close species that have been examined. The anthocyanin

enzyme-coding genes are usually controlled coordinately by either one or two R2R3-MYB transcription factors (Mol et al. 1998; Koes et al. 2005; Xu et al. 2014). In eudicots, the common transcriptional pattern is that one R2R3-MYB protein activates genes that are upstream of the pathway, while a second activates downstream genes (e.g., Martin et al 1991; Quattrocchio et al. 1993). The pattern found in *C. g. ssp. sonomensis* is different from this recognized one: *CgsMYB6* activates both upstream and downstream genes, except for *CgsAns*. The other known transcriptional pattern, which is thought to be common in monocots, but also found in *Ipomoea*, is that the all enzyme-coding genes are controlled by a single R2R3-MYB protein (Lai et al. 2012; Dooner 1983; Tiffin et al. 1998; Morita et al. 2006). However, in *E. umbilicatum*, the downstream genes (*EuDfr*, *EuAns* and *EuUf3gt*) are controlled separately from the upstream genes. These data suggest that the regulatory network of anthocyanin production is evolutionarily labile.

Finally, I found multiple *R2R3-MYB* genes in the petal of *C. g. ssp. sonomensis* and multiple copies of an *R2R3-MYB* gene in *E. umbilicatum*, suggesting that gene duplication is central to phenotypic diversification in floral color. In particular, in *C. g. ssp. sonomensis*, several different genetic changes in the duplicates of the *R2R3-MYB* genes have facilitated the diversification of petal pigmentation patterns

Appendix I

Table A1.1: Voucher information. TRM, seeds provided by Talline R. Martins. LDG, collection of Leslie D. Gottlieb.

Specimen	Voucher	Location	Phenotype	Notes
<i>C. g. ssp. sonomensis</i>	LDG 8513 (TRM 14.6)	Sonoma County, CA	central-spotted, pink-cupped	the same population used in Gottlieb & Ford (1988)
<i>C. g. ssp. sonomensis</i>	LDG 8513 (TRM 14.26)	Sonoma County, CA	central-spotted, white-cupped	the same population used in Gottlieb & Ford (1988)
<i>C. g. ssp. sonomensis</i>	LDG 8920 (TRM L7)		unspotted, pink-cupped	

Table A1.2: Query sequences used in BLAST searches. The sequences of *Arabidopsis thaliana* were retrieved from TAIR (<https://www.arabidopsis.org/>), and the sequences of *Punica granatum* were from GeneBank.

Accession Number	Species	Gene
AT5G13930	<i>Arabidopsis thaliana</i>	<i>Chs</i>
AT3G55120	<i>Arabidopsis thaliana</i>	<i>Chi</i>
AT3G51240	<i>Arabidopsis thaliana</i>	<i>F3h</i>
AT5G07990	<i>Arabidopsis thaliana</i>	<i>F3'h</i>
AT5G42800	<i>Arabidopsis thaliana</i>	<i>Dfr</i>
AT4G22880	<i>Arabidopsis thaliana</i>	<i>Ans</i>
AT5G17050	<i>Arabidopsis thaliana</i>	<i>Uf3gt</i>
AT1G56650	<i>Arabidopsis thaliana</i>	<i>R2R3-MYB; AtMYB075</i>
AT1G66370	<i>Arabidopsis thaliana</i>	<i>R2R3-MYB; AtMYB113</i>
AT1G66380	<i>Arabidopsis thaliana</i>	<i>R2R3-MYB; AtMYB114</i>
AT1G66390	<i>Arabidopsis thaliana</i>	<i>R2R3-MYB; AtMYB090</i>
AT4G09820	<i>Arabidopsis thaliana</i>	<i>bHLH; AtTT8</i>
AT1G63650	<i>Arabidopsis thaliana</i>	<i>bHLH; AtEGL1</i>
AT5G24520	<i>Arabidopsis thaliana</i>	<i>WDR; AtTTG1</i>
KF841615	<i>Punica granatum</i>	<i>Chs</i>
KP726343	<i>Punica granatum</i>	<i>Chs</i>
KU058888	<i>Punica granatum</i>	<i>Chs</i>
KF841616	<i>Punica granatum</i>	<i>Chi</i>
KU058887	<i>Punica granatum</i>	<i>Chi</i>
KF841617	<i>Punica granatum</i>	<i>F3h</i>
KC430328	<i>Punica granatum</i>	<i>F3'h</i>
KP726342	<i>Punica granatum</i>	<i>F3'5'h</i>
KU058892	<i>Punica granatum</i>	<i>F3'5'H</i>
KC430327	<i>Punica granatum</i>	<i>Dfr</i>
KF841618	<i>Punica granatum</i>	<i>Dfr</i>
KP726344	<i>Punica granatum</i>	<i>Dfr</i>
GU376749	<i>Punica granatum</i>	<i>Ans</i>
KF841619	<i>Punica granatum</i>	<i>Ans</i>
KP726345	<i>Punica granatum</i>	<i>Ans</i>
GU371443	<i>Punica granatum</i>	<i>Uf3gt</i>
KF841620	<i>Punica granatum</i>	<i>Uf3gt</i>
KP726346	<i>Punica granatum</i>	<i>Uf3gt</i>
GU371444	<i>Punica granatum</i>	<i>R2R3-MYB</i>
JF747151	<i>Punica granatum</i>	<i>R2R3-MYB</i>
KF841621	<i>Punica granatum</i>	<i>R2R3-MYB</i>
KM014568	<i>Punica granatum</i>	<i>R2R3-MYB</i>
KM881712	<i>Punica granatum</i>	<i>R2R3-MYB</i>
JF747152	<i>Punica granatum</i>	<i>bHLH</i>
KF874658	<i>Punica granatum</i>	<i>bHLH</i>
HQ199314	<i>Punica granatum</i>	<i>WDR</i>

A1.1 The scripts used to run the bioinformatic analyses

Trimmomatic

```
cd /dscrhome/r1124/trinityrnaseq-Trinity-v2.4.0/trinity-plugins/Trimmomatic-0.36/
```

```
./dscrhome/r1124/jdk1.8.0_92/bin/java java -Xmx2g -jar trimmomatic-0.36.jar PE RCL4_S16_L008_R1_001.fastq RCL4_S16_L008_R2_001.fastq -trimlog TrimLogRCL4.log -baseout RCL4trimmed171018.fq ILLUMINACLIP:NebNext.fa:2:40:15 LEADING:20 TRAILING:20 SLIDINGWINDOW:4:20 MINLEN:80
```

Trinity

```
export PATH=/dscrhome/r1124/bowtie2-2.3.0-legacy:/dscrhome/r1124/jdk1.8.0_92/bin:$PATH ./dscrhome/r1124/trinityrnaseq-Trinity-v2.4.0/Trinity --seqType fq --left RCL4trimmed171018_1P.fq --right RCL4trimmed171018_2P.fq --SS_lib_type RF --max_memory 23G --CPU 10 --output /work/r1124/Assembling/TrinityRCL4-20171018
```

BLAST

```
export PATH=/dscrhome/r1124/ncbi-blast-2.6.0+/bin:$PATH makeblastdb -in /work/r1124/Blasting/TrinityRCL4[CgsTop].171018.fasta -input_type fasta -dbtype nucl -logfile dbTOP.log tblastx -query AtPg.txt -db TrinityRCL4[CgsTop].171018.fa -outfmt "7 qseqid sseqid slen length evalue bitscore" -evalue 0.001 -out tblastxRCL4.AtPg.171019.out
```

RSEM

```
/dscrhome/r1124/trinityrnaseq-Trinity-v2.4.0/util/align_and_estimate_abundance.pl --transcripts
```

```
TrinityRCL4[CgsTop].171018.fa --est_method RSEM --aln_method bowtie2 --  
trinity_mode --prep_reference
```

```
/dscrhome/rl124/trinityrnaseq-Trinity-  
v2.4.0/util/align_and_estimate_abundance.pl --transcripts  
TrinityRCL4[CgsTop].171018.fa --left RCL4trimmed171018_1P.fq --right  
RCL4trimmed171018_2P.fq --seqType fq --SS_lib_type RF --est_method RSEM --  
aln_method bowtie2 --trinity_mode --output_dir rsemTOP4
```

```
/dscrhome/rl124/trinityrnaseq-Trinity-  
v2.4.0/util/align_and_estimate_abundance.pl --transcripts  
TrinityRCL4[CgsTop].171018.fa --left RCL5trimmed171018_1P.fq --right  
RCL5trimmed171018_2P.fq --seqType fq --SS_lib_type RF --est_method RSEM --  
aln_method bowtie2 --trinity_mode --output_dir rsemTOP5
```

```
/dscrhome/rl124/trinityrnaseq-Trinity-  
v2.4.0/util/abundance_estimates_to_matrix.pl --est_method RSEM  
rsemTOP4/RSEM.TOP4isoforms.results rsemTOP5/RSEM.TOP5isoforms.results --  
out_prefix TOPisoforms
```

```
/dscrhome/rl124/trinityrnaseq-Trinity-  
v2.4.0/Analysis/DifferentialExpression/run_DE_analysis.pl --matrix  
TOPisoforms.counts.matrix --method edgeR --min_reps_min_cpm 1,1 --dispersion  
0.1 --output edgeR.TOPisoforms
```

```
cut -f 1,3,4 rsemTOP4/RSEM.TOP4isoforms.results > TOPisoforms.transLengths
```

```
/dscrhome/rl124/trinityrnaseq-Trinity-  
v2.4.0/Analysis/DifferentialExpression/run_TMM_normalization_write_FPKM_matrix.  
pl --matrix TOPisoforms.counts.matrix --lengths TOPisoforms.transLengths
```

Table A1.3: Primers used in Chapter 1. *Primers from Martins et al. 2013.
^aPrimers annealing to the vector pENTR2B; ^bPrimers annealing to the vector pBGW418. The underlined sequences are restriction enzyme recognition sites.

Primers for full-length coding region sequencing

Primer Name	Primer Sequence (5'-3')
cCHS-3F	GATCGATCACCGGATTCCAT
CHS _{son} -2R	ATTAAAATCTCCCCAACTTCGA
cCHI-1F	TAACGTAACCCTTCCACCG
cCHI-1R	TAGTAAGTATCCTCTGATTAACGC
F3H-Left	ATTCCACCACCACAATCCCT
cF3H-1R	AATTCATTGACATTATTTCTGCTAA
F3'H-Left	TAAACCACCGCCTTCAAACC
cF3'H-1R	CCATCGTAGTGACCCATCAAAGT
F3'5'H-Left	ACCCATACAAACCCACACCA
cF3'5'H-1R	TTAATCGCCACCCTAACCCA
cDFR-3F	TGGATAGAGTATTGATCTCTAA
cDFR-3R	AGATGCAAATAGATGGAAA
cANS-3F	AAACACCATACATTTTACATTAA
cANS-3R	TAAAGATCAGCGAGTAACAAA
cUF3GT-1F	TCCAAACCGAACAATTCCTTAGT
cUF3GT-1R	AACCCGTGTTTATTATGCCAATT
cMYB6-3F	TGCTACAGAAAGTCTAACGT
cMYB6-1R	ACCGCTGATTTATTTGAAACCCT
cMYB11-5F	AAAAACCAGAAGAAAACCCA
cMYB11-2R	CACAGTTTAATCATTTGATTC
cMYB12-1F	TGCGAAACAAGGACTATCG
cMYB12-4R	CCACGATATATGTAGCACGGAC

Primers for qPCR

Primer Name	Primer Sequence (5'-3')
cCHS-Q1F	AACAAGGTGTCTCAGTCGGA
cCHS-Q1R	CTGTCAAGTGCATGTACCTCG
cCHI-Q1F	TTGAATCCGTCGAGTTCCG
cCHI-Q1R	GGCCGTGAACTTGACAAACT
cF3H-Q2F	GGAGGCCACCAAAGAGTACA
F3H-qR*	GTCAGAGCCTCCTTGTCGAG
cF3'H-Q1F	CCTCCCGAACCTCCCCTA
cF3'H-Q1R	GGATGTGGTAGCCGTTGATC
cF3'5'H-Q1F	CCGGACTTTCTCGACCAGT
F3'5'H1B-qR*	CCATTCGATGATGCTTGATG
cDFR-Q2F	TGTCCGCTCAGTCAAGATGA

cDFR-Q2R	TGGCCCTACAACAAGAGTGG
ANS-qF*	AGCTGTTCTACGAGGGCAAA
ANS-qR*	GACAGCCCAAGAAATCCTGA
cUF3GT-Q1F	CTTGGGTGGCGTTTTGGAC
cUF3GT-Q1R	TGGGATGAAATCTAGTGGGGT
cMYB6-Q1F	GACGAACCCACCGAATCAAG
cMYB6-Q1R	CAGATCGTCCCAGTCCCATT
cMYB11-Q1F	GACAAGGAGGTGATGAATTGGT
cMYB11-Q1R	ATTCTACACATTCATGGAGTCCA
cMYB12-Q1F	GAGCAAGATTCCGGTCAAAGT
cMYB12-1R	TATTCCGTTACAATGAGGCT
GAPDH-Q1F	GAGGCATCAGAGACCCACAT
GAPDH-Q1R	CACGACACGAGCTTCACAAA
AtCHS-Q1F	ACTTCCGCATCACCAACAGT
AtCHS-Q1R	GAATTCCTCCGTCAGATGCA
AtDFR-Q1F	AGGAAGCTACGATGATGCCA
AtDFR-Q1R	ACATTCCATTCACTGTCCGC
AtANS-Q1F	TGCGTATCCTGAAGAGAAGAGA
AtANS-Q1R	CCGACAGAGAGAGCCTTGAA
AtGAPDH-Q1F	CCCATGTTCCGTTGTTGGTGT
AtGAPDH-Q1R	GAGCAAGGCAGTTAGTGGTG

Primers for *CgsMYB12* cosegregation

Primer Name	Primer Sequence (5'-3')
cMYB12-7F	ATACCCGTATTATATTCTTTGAAT
cMYB12-4R	CCACGATATATGTAGCACGGAC

Primers used in construct preparation for *Arabidopsis* transformation

Primer Name	Primer Sequence (5'-3')
pENTRfw ^a	CTACAAACTCTTCCTGTTAGTTAG
SeqL-B ^a	CATCAGAGATTTTGAGACAC
pGWB418-2F ^b	CGACGGTGAACAAAAGTTGATT
pGWB418-2R ^b	GATCGGGGAAATTCGAGCT
BamH1.cMYB6-4F	<u>GGATCCC</u> ATGGGTGGTGTTCCTTGG
Xho1.cMYB6-5R	<u>CTCGAG</u> TACAGAGAGTTATTCCACAGATCG
BamH1.cMYB11-7F	<u>GGATCCA</u> ATGAAGGGAGAATTAAGGAAG
Xho1.cMYB11-4R	<u>CTCGAG</u> TATGGAGTCGACATATACCCAC
BamH1.cMYB12-9F	<u>GGATCCA</u> ATGAAGGAAGGTCTAAGGAA
Xho1.cMYB12-5R	<u>CTCGAG</u> CTACAACTGAAATATGTCGTGATC
BamH1.AtMYB75-1F	<u>GGATCCC</u> ATGGAGGGTTCGTCCAAAG
Xho1.AtMYB75-1R	<u>CTCGAG</u> CTAATCAAATTCACAGTCTCTCCA

A1.2 Reference gene selection for quantitative real-time PCR (qPCR)

In the preliminary qPCR analysis, we found the reference genes previously used in *Clarkia gracilis*, *Actin* and *EFl α* , seemed not expressed stably between the top (pink) and cup (white) regions of the *C. g. ssp. sonomensis* petal. We thus performed reference gene selection to identify a suitable reference gene. We chose 15 genes that were frequently used as reference genes in other studies or whose FPKM values were similar between the transcriptomes of the top and cup regions. These candidate genes are *18S rRNA*, *Actin*, *AP2M*, *DHQS*, *EFl α* , *ESE3*, *F-Box*, *GAPDH*, *RHC*, *RPII*, *SAND*, *TBP*, *TFII*, *TIP41* and *UBC* (Table A1.4). Gene-specific primers were designed using Primer3 (<http://primer3.ut.ee/>), except the primers of *Actin*, which were from Martins et al. (2013). Primer specificity was examined by visualizing the PCR amplicons on 2% agarose gels. In 7 out of the 15 genes, the variation in brightness of the PCR bands amplified with the top and cup cDNA samples had already been observed on the gels. We therefore excluded these seven genes from the qPCR runs. The qPCR reactions were set up for the remaining 8 genes (*Actin*, *EFl α* , *ESE3*, *GAPDH*, *RPII*, *SAND*, *TBP* and *TIP41*) following the conditions described in the text. The cDNA samples of the top and cup regions from 10 white-cupped plants were used as templates, and two technical replicates for each sample were performed. The raw expression values, C_q, were obtained

from Roche LightCycler 96 and were used to assess the expression stability of each gene using three different algorithms: geNorm (Vandesompele et al. 2002), NormFinder (Andersen et al. 2004) and Bestkeeper (Pfaffl et al. 2004). The geNorm algorithm was run using the software qbase+ (Biogazelle) and the R package *NormqPCR* (Perkins et al. 2012). The NormFinder algorithm was run using *NormqPCR*. The Bestkeeper algorithm was run using the Excel-based BestKeeper software (version 1). PCR efficiency of each gene was calculated according to Peirson et al. (2003).

geNorm measures gene stability as the M value, which is defined as the average pairwise variation of the non-normalized expression levels (i.e., Cq) of a particular gene with all other candidate genes. The lower the M values, the higher the gene expression stability. NormFinder calculates the stability value (SV) based on the inter- and intra-group expression variations estimated with an ANOVA-based model. The most stably expressed gene presents the lowest SV value. Bestkeeper ranks the candidate genes based on the calculation of standard deviation (Std Dev) of the Cq values. The lowest Std Dev value indicates the most stable gene.

The results are shown in Table A1.5. The ranks of the candidate genes given by the three algorithms differ from each other. We thus averaged the ranks of each

gene, and selected the one (i.e., *GAPDH*) with the top 1 average rank as the reference gene to be used for normalization in qPCR analyses.

Table A1.4: The candidate genes selected for reference gene selection. Their TAIR (<https://www.arabidopsis.org/>) locus no., gene description, primer sequence, and the length of PCR amplicons are shown. F: forward primer, R: reverse primer. *Primers were designed in this study, except the primers of *Actin*, which are from Martins *et al.* 2013.

Gene	TAIR no.	Gene description	Primer Sequence (5'-3')		Amplicon length (bp)
18S rRNA	AT3G41768	18S ribosomal RNA	F	CAGCAGAACAACCCGAGAAC	82
			R	ACGGAGACAGGATTTGGGAG	
Actin	AT5G09810	Actin 7	F	CATGTATGTTGCGATCCAG	209
			R	TGAACATGTAACCTCTCTCRGT	
AP2M	AT5G46630	Adaptor protein complex 2 subunit MU	F	CTGTTGGTTGGAGGAGGGAG	124
			R	TTGCCAGTCACATCACAACG	
DHQS	AT5G66120	3-dehydroquinate synthase	F	GAAAGAGGGTGCTTGTGGTG	90
			R	CGTTCGGGTTTCCTTTGGTT	
EFl α	AT5G60390	Elongation factor 1-alpha	F	CAGGGTTGAGACTGGTGTGA	133
			R	TCCAACATTGTCACCTGGGA	
ESE3	AT5G25190	Ethylene and salt inducible 3	F	CGCCGATCACTGAAGATGAG	118
			R	TCCACCGAGCCATAATCCAG	
F-Box	AT5G15710	F-box family protein	F	TTCTGTCATGGATCGCAACG	132
			R	GCTAGACGATTCCCCAAATTCA	
GAPDH	AT1G13440	Glyceraldehyde-3- phosphate dehydrogenase	F	GAGGCATCAGAGACCCACAT	149
			R	CACGACACGAGCTTCACAAA	
RHC	AT1G58050	RNA helicase family protein	F	GATCCATCCCAGCCCTTCAA	107
			R	CCAATCCCTGCTCTGTTGC	
RPII	AT2G15430	RNA polymerase II	F	GGCTGATGACTTTGGTGAGC	87

SAND	AT2G28390	SAND family protein	R	AGACAGCTTGAAGATTGTTCGA	
			F	GCTTCGTATTGAGGACCTGC	141
			R	AGTCCAGCTTTCCCACCAAT	
TBP	AT1G55520	TATA binding protein	F	CCAGAAACTTGGCTTCCCTG	100
			R	TGCAAGACCTTCGAGCCTTA	
TFII	AT4G24440	Transcription initiation factor IIA	F	CGCCATTCAAGTTCTCGTCC	112
			R	GTCGCAAAACCTGTACGTGT	
TIP41	AT4G34270	TIP41-like protein	F	ACCCGACTCTACTGTGCATT	108
			R	TCTGGCTGATGATGCTTGGA	
UBC21	AT5G25760	Ubiquitin-conjugating enzyme 21	F	TGGACTGCTCTTATCAAGGGT	86
			R	GGATACTGCTCGGGAATGGA	

Table A1.5: Expression stability analyses of the candidate reference genes for the top (pink) and cup (white) regions of the *Clarkia gracilis* ssp. *sonomensis* petal . The ranks were obtained by running three algorithms: geNorm, NormFinder and BestKeeper. M, average expression stability; SV, stability value; Std Dev, standard deviation; Ave. Rank, the average of the three ranks.

Stability	Rank	geNorm		NormFinder		BestKeeper		Gene	Ave. Rank	PCR efficiency
		Gene	M	Gene	SV	Gene	Std Dev			
most stable	1	TIP41	0.399	GAPDH	0.161	TBP	0.648	GAPDH	1.67	0.941
	2	GAPDH	0.413	EF1 α	0.190	GAPDH	0.730	TIP41	3.00	1.019
	3	SAND	0.423	TBP	0.263	EF1 α	0.731	EF1 α	3.00	0.946
	4	EF1 α	0.453	TIP41	0.273	TIP41	0.856	TBP	3.33	1.068
	5	Actin	0.491	SAND	0.401	ESE3	0.869	SAND	4.67	0.973
	6	TBP	0.543	Actin	0.492	SAND	0.963	Actin	6.33	0.970
	7	ESE3	0.758	RPII	0.733	RPII	1.118	ESE3	6.67	1.053
least stable	8	RPII	1.106	ESE3	0.762	Actin	1.122	RPII	7.33	1.029



Figure A1.1: Five floral phenotypes in the F₂ population. Phenotype I: central-spotted, pink-cupped or central- and basal-spotted, pink-cupped. Phenotype II: central- and basal-spotted, white-cupped. Phenotype III: central-spotted, white-cupped. Phenotype IV: basal-spotted, pink-cupped. Phenotype V: basal-spotted, white-cupped.

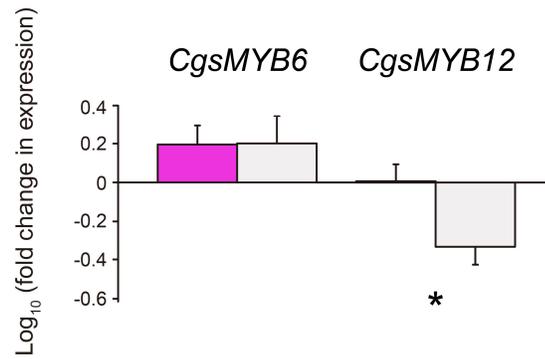


Figure A1.2: Differences in expression of *CgsMYB6* and *CgsMYB12* between the pink cup and white cup. Gene expression was assessed using cDNA samples from flower buds collected 1 day before flowering. Vertical bars represent the means of five replicates and error bars indicate standard error. Pink bars represent pink cup and light gray bars represent white cup. * $P < 0.05$.

CgsMYB12 253.3cup	ATGAAGGAAGGTCTAAGGAAGGGTGTCTTGGAGTGCAGAGAAGATGCTCT	50
CgsMYB12W 216.2cup	ATGAAGGAAGGTCTAAGGAAGGGTGTCTTGGAGTGCAGAGAAGATGCTCT	50
CgsMYB12 253.3cup	CCTCAAGCAATGTATTCAAATTTATGGAGAAGGCAAATGGCATCTTGTTCC	100
CgsMYB12W 216.2cup	CCTCAAGCAATGTATTCAAATTTATGGAGAAGGCAAATGGCATCTTGTTCC	100
CgsMYB12 253.3cup	CCGCCAGAGCAGGGCTAAAATAGGTGTAGAAAAGGTTGCAGATTGAGGTGG	150
CgsMYB12W 216.2cup	CCGCCAGAGCAGGGCTAAAATAGGTGTAGAAAAGGTTGCAGATTGAGGTGG	150
CgsMYB12 253.3cup	CTCAACTATCTGAAGCCAGGCATAAACCTAAAAGAGCTTCAAGATGATGA	200
CgsMYB12W 216.2cup	CTCAACTATCTGAAGCCAGGCATAAACCTAAAAGAGCTTCAAGATGATGA	200
CgsMYB12 253.3cup	AGTTGACTTGATCCTCAAACCTCACAAGCTTCTTGGCAACAATGGTCAC	250
CgsMYB12W 216.2cup	AGTTGACTTGATCCTCAAACCTCACAAGCTTCTTGGCAACAATGGTCAC	250
CgsMYB12 253.3cup	TTATAGCAGGAAGACTTCCGGGAAGAACATGCAATTTATATAAAGAATTAC	300
CgsMYB12W 216.2cup	TTATAGCAGGAAGACTTCCGGGAAGAACATGCAATTTATATAAAGAATTAC	300
CgsMYB12 253.3cup	TGGAACTCCAATATTGCTGCTAAAAAGTGGAAATCAAGAGAAAAGCAGCA	350
CgsMYB12W 216.2cup	TGGAACTCCAATATTGCTGCTAAAAAGTGGAAATCAAGAGAAAAG-AGCA	349
CgsMYB12 253.3cup	AGATTCGGTCAAAGTTAAAGCCATAAGGCCTATTGTACGAAGAGCTCCCA	400
CgsMYB12W 216.2cup	AGATTCGGTCAAAGTTAAAGCCA TAA GGCCTATTGTACGAAGAGCTCCCA	399
CgsMYB12 253.3cup	AAAT A ATCAACTTTGGAATGAATAATAATAATATAGGAACCACCTCTCAG	450
CgsMYB12W 216.2cup	AAAT G ATCAACTTTGGAATGAATAATAATAATATAGGAACCACCTCTCAG	449
CgsMYB12 253.3cup	CCTCATTGTAACGGAATAACT A GGGACGACGAGGTGATGAATTGGTTGGA	500
CgsMYB12W 216.2cup	CCTCATTGTAACGGAATAACT T GGGACGACGAGGTGATGAATTGGTTGGA	499
CgsMYB12 253.3cup	TAGGTTACTAATGGACGATGATGATGATGTTTTATGCTTTTTTGTAAAGGAG	550
CgsMYB12W 216.2cup	TAGGTTACTAATGGACGATGATGATGATGTTTTATGCTTTTTTGTAAAGGAG	549
CgsMYB12 253.3cup	ACGGGGGCTGTACCGCTCACAAGGCCACTGCTCCACCGCAG G TGGTGGC	600
CgsMYB12W 216.2cup	GCGGGGGCTGTACCGCTCACAAGGCCACTGCTCCACCGCAG T TGGTGGC	599
CgsMYB12 253.3cup	GGGTGTATCGACGGCAGTGTCTTAGACGAGCTGTATATTGATCACGACAT	650
CgsMYB12W 216.2cup	GGGTGTATCGACGGCAGTGTCTTAGACGAGCTGTATATTGATCACGACAT	649
CgsMYB12 253.3cup	ATTCAGTTGTAG	663
CgsMYB12W 216.2cup	ATTCAGTTGTAG	662

Figure A1.3: Alignment of the full-length *CgsMYB12* and *CgsMYB12W* sequences from the pink-cupped and white-cupped parental plants (plant IDs. 253 and 216, respectively). Identical sequences are highlighted in a black background. The recognition sequences of the restriction enzyme *BbvI* are highlighted in a blue background. The 1-bp deletion in *CgsMYB12W* is located at 345 bp from the start codon. The premature stop codon in *CgsMYB12W* is highlighted in a red background.

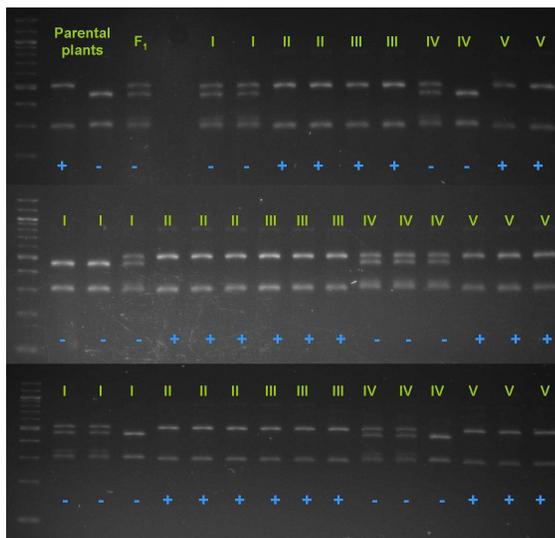


Figure A1.4: Cosegregation of a 1-bp deletion in *CgsMYB12* and the cup phenotypes was examined using the PCR-RFLP method with the parental plants, an F₁ and 40 F₂ plants (eight from each of the flower phenotypes I-V). The white and pink cup colors are scored as “+” and “-”, respectively.

R2 repeat

```

AtMYB75 -----MEGSSKGLRKGAWTTEEDSLLRQCINKEYGEG 31
AtMYB90 -----MEGSSKGLRKGAWTAEEDSLLRLCIDKYGEG 31
AtMYB113 -----MGFSPKGLRKGAWTTFEEDTLLRQCIDKYGEG 31
AtMYB114 -----MEGSSKGLRKGAWTAEEDSLLRQCIGKYGEG 31
PhAN2 -----MSTSNASTSGVRKGAWTEEDDLLRECIKYGEG 34
PhPHZ -----MNTSSTIPKSSGLVRKGAWTEEDVLLRKCIEKYGEG 37
PhAN4 -----MKTSVFTSSGVLKGSWTEEDILLRKCIEKYGEG 35
PhDPL -----MNTSVFTSSGVLKGAWAEEEDILLRKCIEKYGEG 35
AmROSEA1 -----MEKNCRGVRKGTWTKEEDTLLRQCIEEYEGEG 31
AmROSEA2 -----MQKNPRGVRKGTWTKEEDILLMECIDKYGEG 31
AmVENOSA -----MGNNPLGVRKGTWTKEEDILLKQCIEKYGEG 31
MlPELAN -----MEKKKVLGLGVRKGSWTKCEEDSLLRKCIVETYGAG 34
MlNEGAN -----MENTPVGVRKGAWSEEDVLLRKCIEEYEGEG 31
CgsMYB6 -----MGGV---PWTEEDDLLKCKVEQYEGEG 24
VvMYBA1 -----MESLGVKGAWIQEEEDVLLRKCIEKYGEG 29
VvMYBA2 -----MKS LGVRKGAWTQEEEDVLLRKCIEKYGEG 29
FaMYB10 -----MEGFVVRKGAWTKEEDELLKQFIEIHGEG 29
MdMYB10 -----MEGYNENLSVRKGAWTR EEDNLLRQCVEIHGEG 33
CgMYB1C -----MNKVGVKGGWTANEDALLKQCVQTYGEG 29
CgsMYB11 -----MKGE L R K G V W N A E E D A L L K Q C I Q T Y G E G 28
CgsMYB12 -----MKEGLRKGAWSAEEDALLKQCIQTYGEG 28
CgsMYB12W -----MKEGLRKGAWSAEEDALLKQCIQTYGEG 28
LhMYB12 MFQTFIAPANTGTTSP TSAGSGGSPGTRKQWQSK EEDNLLRKCINQYNPV 50
LrMYB15 MRKMPRTMSGKTSDSPTKSQLRTSVSVRKGAWTQAEDELLRSCIEKKGTV 50
LhMYB06 --MSPFRVSATSSSFQMSPPVLRLLVRKGAWTQVEDDLLRSCIERHGVV 48
TfMYB03 ----MSLLTTISSSSSRLPSSVLLRRGAWTQAEEDLLRRCLEKHSGL 45

```

R3 repeat

```

AtMYB75 KWHQVPVVRAGLNRCRKSCRLRWLNLYLKPSTIKRGLKISSDEVDLLIRLHKL 81
AtMYB90 KWHQVPLRAGLNRCRKSCRLRWLNLYLKPSTIKRGLSNDEVDLLIRLHKL 81
AtMYB113 KWHRVPLRTGLNRCRKSCRLRWLNLYLKPSTIKRGLCSDEVDLVRLHKL 81
AtMYB114 KWHQVPLRAGLNRCRKSCRLRWLNLYLKPSTIKRGLFSSDEVDLLIRLHKL 81
PhAN2 KWHLVVVRAGLNRCRKSCRLRWLNLYLRPHIKRGDFSLDEVDLIRLHKL 84
PhPHZ KWHQVPVVRAGLNRCRKSCRLRWLNLYLRPHIKRGDFSEDEVDLIRLHKL 87
PhAN4 KWHQVPVVRAGLNRCRKSCRLRWLNLYLRPHIKRGDFSPDEVDLIRLHKL 85
PhDPL KWHQVPVVRAGLNRCRKSCRLRWLNLYLRPHIKRGDFCPEEVDLIRLHKL 85
AmROSEA1 KWHQVPHRAGLNRCRKSCRLRWLNLYLRPNIKRGRFSRDEVDLIVRLHKL 81
AmROSEA2 KWHQVPLKAGLNRCRKSCRLRWLNLYLRPNIKRGEFSKDEVDLIVRLHKL 81
AmVENOSA KWHQVPIRAGLNRCRKSCRMRLWLNLYLSPNIKRGSFTRDEVDLIVRLHKL 81
MlPELAN KWHLIPLRAGLNRCRKSCRLRWLNLYLRPNIKRGI FDKDEIDLIVRLHKL 84
MlNEGAN KWHLVPLRAGLNRCRKSCRLRWLNLYLRPNIKRGOFNNDEVDLIRLHKL 81
CgsMYB6 KWHRVPLLAGLNRCRKSCRLRWLNLYLRPNIKRGSFTQDEVELIIRLHKL 74
VvMYBA1 KWHLVPLRAGLNRCRKSCRLRWLNLYLKPDIKRGEFALDEVDLIRLHKL 79
VvMYBA2 KWHLVPLRAGLNRCRKSCRLRWLNLYLKPDIKRGEFALDEVDLIRLHKL 79
FaMYB10 KWHHVPLKSGLNRCRKSCRLRWLNLYLKPNIKRGFEADEVDLIRLHKL 79
MdMYB10 KWNQVSYKAGLNRCRKSCRLRWLNLYLKPNIKRGDFKEDEVDLIRLHKL 83
CgMYB1C NWHLVVPDRAGLNRCRKSCRLRWLNLYLKPGLNREEFQDEIDLIRLHKL 79
CgsMYB11 KWHLVPAR TGLNRCRKSCRLRWLNLYLKPGLNREEFQDEVDLIRLHKL 78
CgsMYB12 KWHLVPARAGLNRCRKSCRLRWLNLYLKPGLNLKELQDDEVDLIRLHKL 78
CgsMYB12W KWHLVPARAGLNRCRKSCRLRWLNLYLKPGLNLKELQDDEVDLIRLHKL 78
LhMYB12 KWSHVPKLAGLNRCRKSCRLRWLNLYLDPSTNRGSFSEDEE DLIRLHKL 100
LrMYB15 KWSNVPLLAGLNRCRKSCRLRWLNLYLNPQIDRGTDFDEDENDLIVRLHKL 100
LhMYB06 RWSRVPLAGLNRCRKSCRLRWLNLYLDPRTIRRGFEEDDLIRLHKL 98
TfMYB03 RWCHVARMAGLNRCRKSCRLRWLNLYLDPRLKRGIFEEDEKDLIVRLHKL 95

```

Figure A1.5 (to be continued on the next page)

R3 repeat

```

AtMYB75  GNRWSLIAGRLPGRTANDVKNYWNTHL-SKKHE-PCCKIKMKKRDIPTIP 129
AtMYB90  GNRWSLIAGRLPGRTANDVKNYWNTHL-SKKHESCCCKSKMKKKNIISPP 130
AtMYB113 GNRWSLIAGRLPGRTANDVKNYWNTHL-SKKHDERCCKTKMINKNITSHP 130
AtMYB114 GNRWSLIAGRLPGRTANDVKNYWNTHL-SKKHE-----PCC 116
  PhAN2  GNRWSLIAGRLPGRTANDVKNYWNTHL-RKKLIAPH-----DQKQESKNK 128
  PhPHZ  GNRWSLIAGRLPGRTANDVKNYWNTHL-QRKKLIAP-----PRQEIRKCR 130
  PhAN4  GNRWSLIAGRLPGRTANDVKNYWNTNL-LRRSKFAP-----PQQHDRKCP 129
  PhDPL  GNRWSLIAGRLPGRTANDVKNYWNTHL-LRRSNFAPP-----PQQHERKCT 130
AmROSEA1 GNKWSLIAGRIPGRTANDVKNEFWNTHV-GKNLGEDG-----ERCRKNVMN 125
AmROSEA2 GNKWSLIAGRIPGRTANDVKNEFWNTHV-GKNLGVDG-----ERRKKNVMN 125
AmVENOSA GNRWSLIAGRLPGRTANDVKNEFWNTHF-EKKSG-----ERENTENIN 122
MlPELAN  GNRWSCLIAGRIPGRTANDVKNEFWNTHF-KRKKKPPSSAAATTESSRRSVV 133
MlNEGAN  GNRWSLIAGRLPGRTANDVKNYWNSQI-EKKLLAGGG-----EITPRAAAA 126
CgsMYB6  GNRWSMIAGRLPGRTANDVKNEFWNCHL-SKKLT-----AEQMSIDPE 115
VvMYBA1  GNRWSLIAGRLPGRTANDVKNYWNSHH-FKKEV-----QFQEEGRDK 120
VvMYBA2  GNRWSLIAGRLPGRTANDVKNYWNSHH-LKKKV-----QFQEEGRKK 120
FaMYB10  GNRWSLIAGRLPGRTANDVKNYWNTYQ-RKKDQKTAS----YAKKLVKVP 124
MdMYB10  GNRWSLIAGRLPGRTANDVKNYWNTRL-RID-----SRMKTVKNK 122
CgMYB1C  GNKWSLIAGRLPGRTANDVKNYWYTHI-AKKLPAEPV-----ISQ 118
CgsMYB11 GNRWSLIAGRLPGRTANDVKNYWNAHI-AKKWRSSSKAAPAESKSSSYKG 127
CgsMYB12 GNKWSLIAGRLPGRTANDVKNYWNSNI-AAKKW-----KSREKQQDS 119
CgsMYB12W GNKWSLIAGRLPGRTANDVKNYWNSNI-AAKKW-----KSREKSKIR 119
LhMYB12  GNRWSLIAGRLPGRTANDVKNYWNSHL-SKRKV-----NVEQRTLK 140
LrMYB15  GNRWSLIAGRLPGRTANDVKNHWNSRL-SKKLISG-----IKNDGSRGR 143
LhMYB06  GNRWSLIAGRLPGRTANDVKNYWNSHL-SKKLIP-----QEKKVRACP 140
TfMYB03  GNRWSLIAGRLPGRTANDVKNHWNSRL-NKKLVTEAR----KYGERREVA 140

```

motif 6

```

AtMYB75  T--TPALKNNVYKPRPRSFT----- 147
AtMYB90  T--TPVQKIGVFKPRPRSFS----- 148
AtMYB113 T--SSAQKIDVLKPRPRSFS----- 148
AtMYB114 K--TKIKRINIITP-PNTPA----- 133
  PhAN2  A--VKITENNI IKPRPRTFS----- 146
  PhPHZ  A--LKITENNI IVRPRRTFS----- 148
  PhAN4  KAIKTMAKNA IIRPQPRNLS----- 149
  PhDPL  KEIRTMAKNA IIRPQPRNLS----- 150
AmROSEA1 T--KTIKLTNI VRPRARTFT----- 143
AmROSEA2 T--KNSKETNI IIRPRARTFN----- 143
AmVENOSA P--KLINSSNI IKPQPRTF----- 140
MlPELAN  V--KTITERNI IIRPQPTTFSNRCRSDIQQLMTKTNETNDDDENPKINKNP 181
MlNEGAN  KVQKIITSTNI IVRPRPRAFS----- 146
CgsMYB6  Q--RIDNLVPI IMPQPRNPT----- 133
VvMYBA1  P--QTHSKTKAI KPHPHKFS----- 138
VvMYBA2  P--QTHSKTKAI KPHPHKFS----- 138
FaMYB10  R--ENTIAYT IVRPRPRTFI----- 142
MdMYB10  S--QEMRKTNV IIRPQPKFN----- 140
CgMYB1C  E--NSVKHHA IIRPIAGRPT----- 136
CgsMYB11 K--QOEYSVNI KPIARRAP----- 145
CgsMYB12 V--KVKAIRP IIRRAPKIIN----- 137
CgsMYB12W S--KLKP----- 124
LhMYB12  P----- IIRPQVTLP----- 150
LrMYB15  V----- AAPIRPQRTIP----- 156
LhMYB06  C----- IAAPTIRPQPKCSIKTKTSVDDQQVN----- 167
TfMYB03  P----- PITPQHQCFS----- 151

```

Figure A1.5 (to be continued on the next page)

AtMYB75 -VNNDCNHLNAPPKVDVNP-----CLGLNINNVCDNSII-YNKDKKKDQ 190
 AtMYB90 -VNNGCSHLNGLPEVDLIPS-----CLGLKKNVNCENSIT-CNKDDEKDD 191
 AtMYB113 -DKNSCNDVNILPKVDVPL-----HLGLNANNVYCESSIT-CNKDEQKDK 191
 AtMYB114 -----QK----- 135
 PhAN2 ----RPAMNNFPCWNGKSCN-----KNTIDKNEGDTEIIKFSDEKQKPEE 187
 PhPHZ ---NNAQNISWCSNKSIITS-----TIDKDGSNNECIRINDKKPMAEE 188
 PhAN4 -----KLAKNVSTI-----HKDEHSKQEIIEKPTTAEVV 180
 PhDPL -----KLAKNVSN-----HSTKHKDEYSKQKMFIEKPTTAEVV 184
 AmROSEA1 -----GLHVTWPREVGGKTDEFSNV-RLTTDEIPDCEKQTQFYNDVASPQD 187
 AmROSEA2 -----GLHVTWPREHGKNDAFSNVRITSTTENLDYEKQKPPFHNNAVSTPE 188
 AmVENOSA ----KLRPKETKKQKNIRNV-----CTANDDKQ-----Q-PLSTSGQLE 174
 MlPELAN SSSSSSSSDHSPLLRECGPVNDDRRNRINGDDDDHQNDPKKKDPLIQSSSS 231
 MlNEGAN -----NLSPP-----TTNENPTKNPSSTSSLALAASS 174
 CgsMYB6 ----SVSRKPNKRDQEVGTSIVTLPSVGEDGAMNAVQVID-DGKTNPNNQ 178
 VvMYBA1 -----KALPRFELKTT-----AVDTFDTQVSTSRKPSSTSP 169
 VvMYBA2 -----KALPRFELKTT-----AVDTFDTQVST-----SSKPSSTSP 169
 FaMYB10 -----KRFNFTER-----YANIEHNHSEVSYTS-SLPTEPPQT 174
 MdMYB10 -----RSSYLLSSK-----EPILDHIQSAEDLST-PPQTSSTK 173
 CgMYB1C -----KGM-----QFGMNVTS----- 148
 CgsMYB11 -----KMIDF-----GMTMNNNNICSMSTSQLPPHLDCTG 175
 CgsMYB12 -----FGMNNNNIGTTSQPHCNGITRDE 161
 CgsMYB12W ----- 124
 LhMYB12 -----RNWSWLRMKKQGEA-----EPKMETKVP-DEEHDQWL 182
 LrMYB15 -VRMQKTGHVERKHGEIQPV-----SVVEEDHHTSRMENIIDDDENYNTK 200
 LhMYB06 -----MSELIPQKKKVRACRIIAAPTRPQPRKCSIETKTSVDEQQVNMS 211
 TfMYB03 -----RKQSSPE-----RLQEDLNVTVGSAYQQDNTWVDRLL 184

AtMYB75 LVN-NLIDGDNM**W**--LEK-FLEESQEVD----- 214
 AtMYB90 FVN-NLMNGDNM**W**--LEN-LLGENQEAD----- 215
 AtMYB113 LININLLDGDNM**W**--WES-LLEAD----- 212
 AtMYB114 ----- 135
 PhAN2 SIDDGLQ-----**W**--WAN-LLANNIEIEELVSCNSPTLLHE----- 220
 PhPHZ SRHDGVQ-----**W**--WTS-LLANCNENDETAVENMSYDKLPSLLHEEISP 230
 PhAN4 SRDENVE-----**W**--WTNLLLDNCNGFEKAATESTSAFKNIESLLNEELL 223
 PhDPL SRDNNVE-----**W**--WTNLLLDNCNGFEKAAPESSTFKNIESLLNEELL 227
 AmROSEA1 EVEDCIQ-----**W**--WSK-LLETTEDEGE----- 207
 AmROSEA2 EVDESIR-----**W**--WSNLLLETTEDELE----- 209
 AmVENOSA EVNERIR-----**W**--WSE-LLDFADYVD----- 194
 MlPELAN QDQADEEDEYVR**W**--WRD-LLEMTEKDH----- 256
 MlNEGAN RETDEIVR-----**W**--WRN-LLETTTSTEDGILVAGEEERQTGKLCREN-- 215
 CgsMYB6 EHSVSLGD-----L--PGEFQFDECRLDG----- 200
 VvMYBA1 QPNDDII-----**W**--WES-LLAEHAQMDQETDFSASGEMLIASLRTEETA 211
 VvMYBA2 QPNDDII-----**W**--WES-LLAEHAQMDQETDFSASGEMLIASLWTEETA 211
 FaMYB10 QLLENVTD-----**W**--WKDFSEDSTESID----- 196
 MdMYB10 NGND-----**W**--WET-LLEGEDTFERAAYPSIELE---EELFTSFWF 209
 CgMYB1C ----QPPPLEN**W**--SDWLMDDDDINYD---NGGGCSASEGHCTTAANG 189
 CgsMYB11 IINTDKEVMN---WLRLLDDDDDLIG----- 198
 CgsMYB12 VMN-----**W**--LDRLLMDDDDVYAFCKGDDGGCT----- 188
 CgsMYB12W ----- 124
 LhMYB12 MINDSKHGHENY**Y**--TMNDQMDSNQAD-----FG 210
 LrMYB15 KTERQRE-----ADFSFFDNEGFRE----- 220
 LhMYB06 ESRPSADTANCAV--WQDDLGNVKEMIE-----QLTEATIP 245
 TfMYB03 LYNEEYNKEKTE**W**QHLSDFSLEDVEGFK----- 212

Figure A1.5 (to be continued on the next page)

AtMYB75	----ILVPEATTTEKGDTLAFDV-----DQLW	SLFDGETVKFD--	248
AtMYB90	----AIVPEATTAEHGATLAFDV-----EQLW	SLFDGETVELD--	249
AtMYB113	----VLGPEATEETAKGVTLPLDF-----EQLW	ARFDEETLELN--	246
AtMYB114	-----VDIF-----		139
PhAN2	ETAPSVNAESSLTQGGGSGLSDFSVDI-----DDIW	DLVS-----	255
PhPHZ	T---INGGISNCMQEGQTGWDDFSVDI-----DHLW	NLLN-----	262
PhAN4	SPSINGGTYYPMQETRDMGWSDLSID-----ADLW	ELL-----	256
PhDPL	SASINGGTNYPIQETGDMGWSDFCID-----SDFW	ELLQ-----	262
AmROSEA1	----LGNLFEEAQQIGN-----		220
AmROSEA2	----NLFEDVQQTGK-----MSEW		224
AmVENOSA	-----		194
MlPELAN	GATPLLFSNDPIIDDNNYAIDDGLSSGLCL---DDVW	DLLSSHHDH----	298
MlNEGAN	AADLDEEDGGAAVQEGEADEDEDGLADLLLD---VDIW	ELLSFDDERDDSW	262
CgsMYB6	-----ISSNSRKWDWDDL-----MDMDIDLW	NNSL-----	227
VvMYBA1	TQKKGPMGMIQIQGGEGDFPFD-----VGFWD	TPNTQVNHLI--	250
VvMYBA2	TQKKGTHSKTKAIKPHPHKFSKALPRFELKTTAVDT	FDTQVSTSSKLIHV	261
FaMYB10	-----RTMCSGLGLEDDHDF-----TNFW	VEDMLLSASNDLV	228
MdMYB10	DDRLSRSCANFPEGQSRSEFSFS-----TDLW	NHSKEE-----	243
CgMYB1C	CYDMQIESPWAAIVGGFTEGGNNP-----DELW	FDDIF-----	222
CgsMYB11	----FGGDGGSAASEGHC-----TVGGG	YMSTP-----	223
CgsMYB12	----ASQGHCASTAGGGGCGIDGSVL-----DELW	IDHDIFQL-----	220
CgsMYB12W	-----		124
LhMYB12	FECIYGVGEEETTVDAILQWDGLL-----SD--IKLW	SDSEV-----	246
LrMYB15	-----DEWLMQDGISAWQNLL-----SDLL	TGG-----	243
LhMYB06	SENTEGFAGEGLMQDGVSLWDFI-----FDI	QLSS-----	276
TfMYB03	-----EGMMLEGNLGLDTFLSD-----MQLW	S-----	234
AtMYB75	-----		248
AtMYB90	-----		249
AtMYB113	-----		246
AtMYB114	-----		139
PhAN2	-----		255
PhPHZ	-----		262
PhAN4	-----		256
PhDPL	-----		262
AmROSEA1	-----		220
AmROSEA2	-----		224
AmVENOSA	-----		194
MlPELAN	-----		298
MlNEGAN	GLLGPN		268
CgsMYB6	-----		227
VvMYBA1	-----		250
VvMYBA2	TTTE--		265
FaMYB10	NISYV-		233
MdMYB10	-----		243
CgMYB1C	-----		222
CgsMYB11	-----		223
CgsMYB12	-----		220
CgsMYB12W	-----		124
LhMYB12	-----		246
LrMYB15	-----		243
LhMYB06	-----		276
TfMYB03	-----		234

Figure A1.5: Alignment of anthocyanin-regulating R2R3-MYB proteins (subgroup 6, Strack et al. 2001). If > 70% of amino acids at each column are identical, this common amino acid is highlighted in a black background, and its similar amino acids are highlighted in a gray background. The domains of R2 and R3 repeats and the motif of

subgroup 6 are indicated as bars above the alignment. The motif 6 was defined as "KPRPR[S/T]F" based only on the *Arabidopsis* sequences (Strack et al. 2001). However, our data show "[K/R]P[R/Q]PR" is a more generalized version of this motif. The bHLH interaction domain "[D/E]L_{x2}[R/K]_{x3}L_{x6}L_{x3}R" (Zimmermann et al. 2004) located in the R3 repeat is indicated in yellow letters.

Appendix II

Table A2.1: Voucher information. TRM, seeds provided by Talline R. Martins. NFW, seeds from the collection of Norman F. Weeden. RSABG, seeds provided from Rancho Santa Ana Botanical Garden.

Specimen	Voucher	Location
<i>C. g. ssp. albicaulis</i>	TRM 040/Butte12	Butte County, CA (collected by B. Barringer) The same population used in Martins et al. 2013.
<i>C. lassenensis</i>	NFW 84b	Shasta County, CA
<i>C. lassenensis</i>	NFW 152	Shasta County, CA
<i>C. amoena ssp. huntiana</i>	RSABG 15875	Mendocino County, CA
<i>C. amoena ssp. huntiana</i>	RSABG 15876	Marin County, CA

Table A2.2: Primers used in Chapter 2. *Primers from Martins et al. 2017. ^aPrimers only for amplifying a short fragment of *MYB6* from *Clarkia lassenensis*

Primers for coding region sequencing

Target species: *Clarkia gracilis* ssp. *albicaulis*

Primer Name	Primer Sequence (5'-3')
cMYB6-3F	TGCTACAGAAAGTCTAACGT
cMYB6-1R	ACCGCTGATTTATTTGAAACCCT
cMYB11-3F	GTGAGTCGCATGGCTGT
cMYB11-2R	CACAGTTTAATCATTGATTC

Target species: *Clarkia amoena*

Primer Name	Primer Sequence (5'-3')
cMYB6-3F	TGCTACAGAAAGTCTAACGT
cMYB6-1R	ACCGCTGATTTATTTGAAACCCT
cMYB11-5F	AAAAACCAGAAGAAAACCCA
cMYB11-6R	AATTCACAGTTTAATCATTG
cMYB12-9F	AATGAAGGAAGGTCTAAGGAA
cMYB12-5R	CTACAACCTGAAATATGTCGTGATC
cMYB1-F*	ATGAATAAGGTAGGACTTAGAAAGG
cMYB1-R*	TTAAAATATGTCATCAAAATAAAGCTCATCC

Target species: *Clarkia lassenensis*

Primer Name	Primer Sequence (5'-3')
cMYB6-4F	CATGGGTGGTGTTCCTTGGA
cMYB6-5R	TTACAGAGAGTTATCCACAGATCG
cMYB11-7F	AATGAAGGGAGAATTAAGGAAG
cMYB11-5R	TTACTTGGAATATGATTCTACACA
cMYB12-11F	ATGCACGCTCTACAATAAAACG
cMYB12-6R	TTAGTATAACGGCATAATTCT

Primers for semi-qPCR

Primer Name	Primer Sequence (5'-3')
cMYB6-Q1F	GACGAACCCACCGAATCAAG
cMYB6-Q1R	CAGATCGTCCCAGTCCCATT
CIMYB6-Q1F ^a	CGGTGTCGGTCTTGGAGATA

CIMYB6-Q1R ^a	TCATCCCAATCCCATTTTCTGC
cMYB11-Q1F	GACAAGGAGGTGATGAATTGGT
cMYB11-Q1R	ATTCTACACATTCATGGAGTCGA
cMYB12-Q1F	GAGCAAGATTCGGTCAAAGT
cMYB12-1R	TATTCCGTTACAATGAGGCT
GAPDH-Q1F	GAGGCATCAGAGACCCACAT
GAPDH-Q1R	CACGACACGAGCTTCACAAA

Table A2.3: FPKM estimates of *CgaMYB6* and *CgaMYB11* in the *C. g. ssp. albicaulis* transcriptomes. Each contig represents a different gene, which has several isoforms determined by Trinity. *The BLAST bitscore of each contig is shown as the highest bitscore among its isoforms.

<i>C. g. ssp. albicaulis</i>				
Pink background transcriptome				
SubjectID	Gene	BLAST bitscore*	FPKM	
			Pink background	White band
RCL6_24035_c1_g1	<i>CgaMYB6</i>	211	858.87	595.14
RCL6_21823_c2_g1	<i>CgaMYB11</i>	195	121	46.94
<i>C. g. ssp. albicaulis</i>				
White band transcriptome				
Subject ID	Gene	BLAST bitscore*	FPKM	
			Pink background	White band
RCL7_23760_c0_g1	<i>CgaMYB6</i>	211	664.67	354.33
RCL7_25096_c3_g1	<i>CgaMYB11</i>	211	117.44	36.33

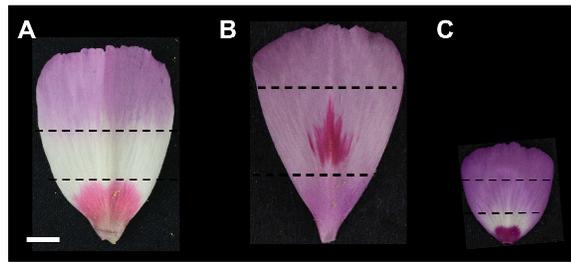


Figure A2.1: Sections dissected from the petal of *Clarkia*. (A) *Clarkia gracilis* ssp. *albicaulis*, top: pink background, middle: white band, base: red spot; (B) *C. amoena* ssp. *huntiana*, top: pink background, base: pink cup; (C) *C. lassenensis*: top: purple-pink background, base: white band and purple-red spot. The scale bar indicates 5 mm.

Appendix III

Table A3.1: The *Erythronium umbilicatum* anther-color polymorphism in 16 sites in North Carolina. These 16 sites are distributed along three main rivers/creeks: Eno River, New Hope Creek/Old Field Creek and Morgan Creek. We conducted the anther-color survey during the flowering season in 2017. In each site, we set up 2 m x 2 m plots that were 6 m apart from each other. Within each plot, four 1 m x 1 m quadrats were laid out. We tallied the anther color for all plants in each quadrat, and counted at least 200 plants in each population. The geographical coordinates of each population, the number of flowers surveyed, the number of flowers having purple or yellow anthers, and the percentage of yellow-anthered flowers in each population are shown.

Population	Location		No. of flowers surveyed	Anther color		
	Latitude	Longitude		Purple	Yellow	% of yellow-anthered flowers
Eno. Old Oxford	36°04'16.8"N	78°51'32.5"W	301	287	14	4.65
Eno. Penny's Bend	36°04'30.6"N	78°52'05.3"W	292	278	14	4.79
NCBG. Stillhouse Bottom	35°53'06.4"N	79°02'49.9"W	210	198	12	5.71
NCBG. Morgan Creek	35°53'27.3"N	79°02'28.5"W	228	199	29	12.72
NCBG. Piedmont Nature Trails	35°53'49.7"N	79°02'01.4"W	266	256	10	3.76
Duke Forest. Oosting	35°58'48.5"N	79°03'54.7"W	213	174	39	18.31
Duke Forest. Gate 23	35°59'32.1"N	79°02'33.6"W	227	209	18	7.93
Duke Forest. Gate 24	35°59'03.8"N	79°02'03.2"W	218	194	24	11.01
Duke Forest. Gate 26	35°59'03.8"N	79°00'56.9"W	268	244	24	8.96
Duke Forest. Gate 27	35°58'50.5"N	79°00'06.0"W	230	201	29	12.61
Eno. Pleasant Green	36°02'30.4"N	79°00'39.7"W	247	217	30	12.15
Eno. Cabe Lands	36°02'36.6"N	78°59'32.6"W	226	209	17	7.52
Eno. Eno Trace Trail	36°04'26.1"N	79°00'27.6"W	256	205	51	19.92
Eno. Pump Station	36°03'40.5"N	78°58'16.7"W	220	196	24	10.91

Eno. Fish Dam	36°04'02.2"N	78°57'19.6"W	235	226	9	3.83
Eno. Cox Mountain Trail	36°04'57.4"N	79°01'17.1"W	224	182	42	18.75

Table A3.2: The query sequences used in the BLAST searches. The sequences of *Arabidopsis thaliana* were retrieved from TAIR (<https://www.arabidopsis.org>). Other sequences were retrieved from GenBank.

Accession Number	Species	Gene
AT5G13930	<i>Arabidopsis thaliana</i>	<i>Chs</i>
AT3G55120	<i>Arabidopsis thaliana</i>	<i>Chi</i>
AT3G51240	<i>Arabidopsis thaliana</i>	<i>F3h</i>
AT5G07990	<i>Arabidopsis thaliana</i>	<i>F3'h</i>
AT5G42800	<i>Arabidopsis thaliana</i>	<i>Dfr</i>
AT4G22880	<i>Arabidopsis thaliana</i>	<i>Ans</i>
AT5G17050	<i>Arabidopsis thaliana</i>	<i>Uf3gt</i>
AT1G56650	<i>Arabidopsis thaliana</i>	<i>R2R3-MYB; AtMYB075</i>
AT1G66370	<i>Arabidopsis thaliana</i>	<i>R2R3-MYB; AtMYB113</i>
AT1G66380	<i>Arabidopsis thaliana</i>	<i>R2R3-MYB; AtMYB114</i>
AT1G66390	<i>Arabidopsis thaliana</i>	<i>R2R3-MYB; AtMYB090</i>
AT4G09820	<i>Arabidopsis thaliana</i>	<i>bHLH; AtTT8</i>
AT1G63650	<i>Arabidopsis thaliana</i>	<i>bHLH; AtEGL1</i>
AT5G24520	<i>Arabidopsis thaliana</i>	<i>WDR; AtTTG1</i>
AF146702	<i>Petunia x hybrida</i>	<i>R2R3-MYB; AN2</i>
HQ428105	<i>Petunia x hybrida</i>	<i>R2R3-MYB; AN4</i>
AF260919	<i>Petunia x hybrida</i>	<i>bHLH; AN1</i>
U94748	<i>Petunia x hybrida</i>	<i>WDR; AN11</i>
AB534587	<i>Lilium</i> hybrid division I	<i>R2R3-MYB</i>
AB222075	<i>Lilium</i> hybrid division I	<i>bHLH</i>
AB222076	<i>Lilium</i> hybrid division I	<i>bHLH</i>
KC261503	<i>Tulipa fosteriana</i>	<i>Chs</i>
KC261502	<i>Tulipa fosteriana</i>	<i>Chi</i>
KC261504	<i>Tulipa fosteriana</i>	<i>F3h</i>
KC256779	<i>Tulipa fosteriana</i>	<i>F3'h</i>
KC261505	<i>Tulipa fosteriana</i>	<i>F3'h</i>
KC261506	<i>Tulipa fosteriana</i>	<i>Dfr</i>
KC261507	<i>Tulipa fosteriana</i>	<i>Ans</i>
KF792732	<i>Tulipa fosteriana</i>	<i>Uf3gt</i>
KF990610	<i>Tulipa fosteriana</i>	<i>R2R3-MYB</i>
KC256778	<i>Tulipa fosteriana</i>	<i>bHLH</i>
KF924736	<i>Tulipa fosteriana</i>	<i>bHLH</i>

A3.1 The scripts used to run the bioinformatic analyses

Trimmomatic

```
cd /dscrhome/r1124/trinityrnaseq-Trinity-v2.4.0/trinity-plugins/Trimmomatic-0.36/

./dscrhome/r1124/jdk1.8.0_92/bin/java java -Xmx2g -jar trimmomatic-0.36.jar PE
RCL1_S13_L008_R1_001.fastq RCL1_S13_L008_R2_001.fastq -trimlog TrimLogEaR.log -
baseout RCL1trimmed171017.fq ILLUMINACLIP:NebNext.fa:2:40:15 LEADING:20 TRAILING:20
SLIDINGWINDOW:4:20 MINLEN:80
```

Trinity

```
export PATH=/dscrhome/r1124/bowtie2-2.3.0-
legacy/:/dscrhome/r1124/jdk1.8.0_92/bin:$PATH

./dscrhome/r1124/trinityrnaseq-Trinity-v2.4.0/Trinity --seqType fq --left
RCL1trimmed171017_1P.fq --right RCL1trimmed171017_2P.fq --SS_lib_type RF --
max_memory 23G --CPU 10 --output /work/r1124/Assembling/TrinityRCL1-20171018
```

BLAST

```
export PATH=/dscrhome/r1124/ncbi-blast-2.6.0+/bin:$PATH

makeblastdb -in /work/r1124/Blasting/PurpleTranscriptome.fasta -input_type fasta -
dbtype nucl -logfile dbPurple.log

tblastx -query AtLily.txt -db PurpleTranscriptome.fasta -outfmt "7 qseqid sseqid
slen length evalue bitscore" -evalue 0.001 -out tblastxEuP.AtLily.170516.out
```

RSEM

```
/dscrhome/r1124/trinityrnaseq-Trinity-v2.4.0/util/align_and_estimate_abundance.pl --  
transcripts PurpleTranscriptome.fasta --est_method RSEM --aln_method bowtie2 --  
trinity_mode --prep_reference
```

```
/dscrhome/r1124/trinityrnaseq-Trinity-v2.4.0/util/align_and_estimate_abundance.pl --  
transcripts PurpleTranscriptome.fasta --left s_7_1_sequence_P.txt --right  
s_7_2_sequence_P.txt --seqType fq --SS_lib_type RF --est_method RSEM --aln_method  
bowtie2 --trinity_mode --output_dir rsemPP
```

```
/dscrhome/r1124/trinityrnaseq-Trinity-v2.4.0/util/align_and_estimate_abundance.pl --  
transcripts PurpleTranscriptome.fasta --left s_7_1_sequence_Y.txt --right  
s_7_2_sequence_Y.txt --seqType fq --SS_lib_type RF --est_method RSEM --aln_method  
bowtie2 --trinity_mode --output_dir rsemPY
```

```
/dscrhome/r1124/trinityrnaseq-Trinity-v2.4.0/util/abundance_estimates_to_matrix.pl -  
-est_method RSEM rsemPP/RSEM.PPisoforms.results rsemPY/RSEM.PYisoforms.results --  
out_prefix PISOforms
```

```
/dscrhome/r1124/trinityrnaseq-Trinity-  
v2.4.0/Analysis/DifferentialExpression/run_DE_analysis.pl --matrix  
PISOforms.counts.matrix --method edgeR --min_reps_min_cpm 1,1 --dispersion 0.1 --  
output edgeR.PISOforms
```

```
cut -f 1,3,4 rsemPP/RSEM.PPisoforms.results > PISOforms.transLengths
```

```
/dscrhome/r1124/trinityrnaseq-Trinity-  
v2.4.0/Analysis/DifferentialExpression/run_TMM_normalization_write_FPKM_matrix.pl --  
matrix PISOforms.counts.matrix --lengths PISOforms.transLengths
```

Table A3.3. Primers used in Chapter 3.

Primers for coding region sequencing	
Primer Name	Primer Sequence (5'-3')
CHS-1F	CAGTTTTTCCCCTTGTCTGTAACA
CHS-1R	AAAAAATAGGATTGTCCCATGC
F3H-3F	CGAGAGAGCTTTCAATCGCA
F3H-2R	CGATAATGTGGCTAAGCATG
DFR-2F	CTCAGAGAAACAGAGAGAGAGGAAA
DFR-2R	TCAATCCTACATCAAACCCGTAT
ANS-1F	CACATACTTCCTCTCAACTCACAACA
ANS-1R	AACGACGAACTGCACTCTTATTT
UF3GT-5F	CACACCACCACCAGCATCG
UF3GT-3R	CATCCATTTAGAATTTGAAACCA
MYB3-1F	ACCTCTCCTCCATGCAATATCA
MYB3-Q2R	TCCTCTTTGTATGAAACCGTGG
bHLH2-9F	TCCGAACAATGGCAAACCTCA
bHLH2-3R	TATAACGTACAGAAATGTATAGCTG
WDR1-1F	TCACCAACTCCAAGTCCTCC
WDR1-1R	TGTCATCGTAGCAACTTCGC
WDR2-2F	CCATTCCCATTCCCCACAA
WDR2-1R	CATATTGCCCCGAATCCACA
Primers used in qPCR with <i>E. umbilicatum</i> samples	
Primer Name	Primer Sequence (5'-3')
EF1a-Q1F	TGGTCAGACTCGTGAGCATG
EF1a-Q1R	TACTTCGGGGTTGTGGCATC
CHS-Q1F	TACTTCCGCATCACCAAGAGC
CHS-Q2R	CAATGAGGGAGCCATGTAGG
F3H-Q2F	CCTCCTTCTCCAAGACCAGG
F3H-Q2R	TCAGCATTCTTGAACCTCCC
DFR-Q1F	GAGGGGACAAAGTGAGCAAGTG
DFR-2R	TCAATCCTACATCAAACCCGTAT
ANS-Q1F	CCTTCAAGCAGCACATCCAGC
ANS-1R	AACGACGAACTGCACTCTTATTT
UF3GT-Q1F	GAAGATGGCGGAGGAGATGA
UF3GT-Q1R	TCCAGAGGTTCACTTAGGCG

bHLH2-Q2F	TCCTCTTCTGCTAACAACACCT
bHLH2-Q2R	ACTTCGACAATGCTGGCTCT
WDR1-Q1F	GAGCACTCCACCATCGTCTA
WDR1-Q1R	GCGCAGGTCCATCTTGTTT
WDR2-2F	CCATTCCCATTCCCCACAA
WDR2-Q4R	AGTTCATGGCGTAGATGTGC

Primers used in qPCR with *E. americanum* samples

Primer Name	Primer Sequence (5'-3')
EF1a-Q1F	TGGTCAGACTCGTGAGCATG
EF1a-Q1R	TACTTCGGGGTTGTGGCATC
EaCHS-Q2F	GGGGTGCTGTTCGGGTTC
EaCHS-1R	AGCAACAGAAGCAGCAATCC
F3H-Q2F	CCTCCTTCTCCAAGACCAGG
F3H-Q2R	TCAGCATTCTTGAACCTCCC
DFR-Q4F	TCTTCACATCCTCTGCTGGG
DFR-Q4R	CGGCTTTCTCTGCTAGGGTT
ANS-Q1F	CCTTCAAGCAGCACATCCAGC
ANS-1R	AACGACGAACTGCACTCTTATTT
UF3GT-Q1F	GAAGATGGCGGAGGAGATGA
UF3GT-Q1R	TCCAGAGGTTCACTTAGGCG

A3.2 Derivation of equation (1)

Here we present the derivation of equation (1). Although we do not assume that the three loci *EuDfr*, *EuAns*, and *EuUf3gt* are unlinked, we do assume that there is no linkage disequilibrium, which allows the expected frequencies of three-locus genotypes to be computed as the product of the individual single-locus genotypes. This assumption seems reasonable for long-standing natural populations.

Let *D* and *d*, *A* and *a*, and *U* and *u* be the alternative alleles at *EuDfr*, *EuAns*, and *EuUf3gt*, respectively, with lower-case alleles corresponding to the downregulated “yellow” alleles (i.e., individuals that are *dd*, *aa*, or *uu* have yellow anthers). Let *q*, *s*, and *v* be the frequencies of *d*, *a*, and *u*, respectively. Then the expected frequencies of the different yellow-anthered genotypes are:

$$\text{Freq}(ddA_U_)=q^2(1-s^2)(1-v^2)$$

$$\text{Freq}(ddaaU_)=q^2s^2(1-v^2)$$

$$\text{Freq}(ddA_uu)=q^2(1-s^2)v^2$$

$$\text{Freq}(ddaaau)=q^2s^2v^2$$

$$\text{Freq}(D_aaU_)=(1-q^2)s^2(1-v^2)$$

$$\text{Freq}(D_aaau)=(1-q^2)s^2v^2$$

$$\text{Freq}(D_A_uu)=(1-q^2)(1-s^2)v^2.$$

The sum of these frequencies is the expected frequency of yellow-anthered individuals, $\text{Freq}(Y)$, is

$$\text{Freq}(Y) = q^2 (1 - s^2) (1 - v^2) + s^2 + v^2 - s^2v^2.$$

The probability that a yellow-anthered individual is homozygous for the “yellow” allele at all three loci (i.e., the genotype *ddaaauu*) is then

$$\text{Prob}(ddaaauu) = q^2 s^2 v^2 / \text{Freq}(Y),$$

which is equation (1).

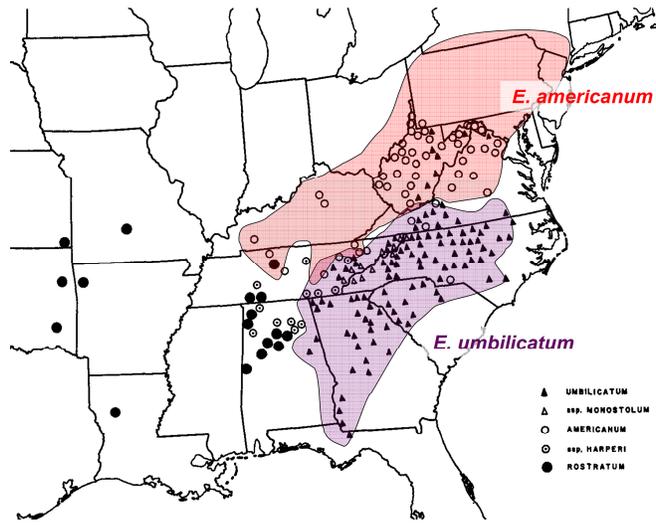


Figure A3.1: Distributions of *Erythronium americanum* and *E. umbilicatum*, modified from Parks & Hardin (1963).

EuCHS.P19cDNA ATGTC TAA GACTGTGCGAGGAGGTGAGGAAGGCTCAGAGAGGCCAGGGGCC 50
 EuCHS.Y05cDNA ATGTC TAA GACTGTGCGAGGAGGTGAGGAAGGCTCAGAGAGGCCAGGGGCC 50
 TfCHS|KC261503 ATGGC --- GAA TGTGCGATGAGATCCGGCAGTCTCAGAGGGCGGAGGGTCC 47

EuCHS.P19cDNA GCGGACTGTGCTGGCCATCGGCACCGCCACCCCTTCCAACGTCATCTACC 100
 EuCHS.Y05cDNA GCGGACTGTGCTGGCCATCGGCACCGCCACCCCTTCCAACGTCATCTACC 100
 TfCHS|KC261503 AGCTACTGTTCTAGCCATCGGCACCGCCACCCAGCCAACGTCATCTACC 97

EuCHS.P19cDNA AGGCCGACTACCCCGACTACTACTTCCGCATCACCAAGAGCGAGCATCTC 150
 EuCHS.Y05cDNA AGGCCGACTACCCCGACTACTACTTCCGCATCACCAAGAGCGAGCATCTC 150
 TfCHS|KC261503 AGTCAGAGTATCCGGACTACTACTTCCAGGATCACCAAGAGCGACCACCTT 147

EuCHS.P19cDNA ACCGACTCTCAAAGAGAAGTTCAAGAGGATGTGCGAGAAATCGATGATCCA 200
 EuCHS.Y05cDNA ACCGACTCTCAAAGAGAAGTTCAAGAGGATGTGCGAGAAATCGATGATCCA 200
 TfCHS|KC261503 ACCGATCTCAAAGAGAAGTTCAAGAGGATGTGTCGACAAATCTATGATCAA 197

EuCHS.P19cDNA AAAGCGTTATATGCAATCTGAACGAGGAGATACTCAAGGAGAACCCCAACA 250
 EuCHS.Y05cDNA AAAGCGTTATATGCAATCTGAACGAGGAGATACTCAAGGAGAACCCCAACA 250
 TfCHS|KC261503 GAAGCGTTACATGCACCTA AACGAGGAGATACTGAAGGAGACCTTAACA 247

EuCHS.P19cDNA TGTGCGCCTACATGGCTCCCTCAT TGGACGCCCGCCAGGACATGGTCGTC 300
 EuCHS.Y05cDNA TGTGCGCCTACATGGCTCCCTCAT TGGACGCCCGCCAGGACATGGTCGTC 300
 TfCHS|KC261503 TGTGTCGCTACATGGCCCTCCCTCC TGGACGCCCGCCAGGACATGGTCGTC 297

EuCHS.P19cDNA GTGGAGGTC CCCAAGCTCGGCAAGGAGGCCGCC TCCAGGGCTATCAAGGA 350
 EuCHS.Y05cDNA GTGGAGGTC CCCAAGCTCGGCAAGGAGGCCGCC TCCAGGGCTATCAAGGA 350
 TfCHS|KC261503 GTGGAGGTC CCCAAGCTCGGCAAGGAGGCCGCC GTCAAGGCCATCAAGGA 347

EuCHS.P19cDNA GTGGGGGCAGCCCAAATCGAAGATCACCCACCTTATCTTCTGTACCACCA 400
 EuCHS.Y05cDNA GTGGGGGCAGCCCAAATCGAAGATCACCCACCTTATCTTCTGTACCACCA 400
 TfCHS|KC261503 GTGGGGGCAGCCCAAATCGAAGATCACCCACCTTATCTTCTGTACTACCA 397

EuCHS.P19cDNA GCGGC GTT GACATGCCCGGC GCAGACTACCAGCTCACCAAGCTCCTCGGC 450
 EuCHS.Y05cDNA GCGGC GTT GACATGCCCGGC GCAGACTACCAGCTCACCAAGCTCCTCGGC 450
 TfCHS|KC261503 GCGGT GTT GACATGCCCGGT GCAGACTACCAGCTCACCAAGCTCCTCGGC 447

EuCHS.P19cDNA CTC CGT CCCTCCGTCAACCGCTTCATGATGTACCAGCAGGGCTGCTTTGC 500
 EuCHS.Y05cDNA CTC TGT CCCTCCGTCAACCGCTTCATGATGTACCAGCAGGGCTGCTTTGC 500
 TfCHS|KC261503 CTC CGC CCCTCCGTCAACCGCTTCATGATGTACCAGCAGGGCTGCTTTGC 497

EuCHS.P19cDNA CGGAGGCACCGTCTTCCGCTTCGCTAAAGACCTGGCTGAGAACAACCGTG 550
 EuCHS.Y05cDNA CGGAGGCACCGTCTTCCGCTTCGCTAAAGACCTGGCTGAGAACAACCGTG 550
 TfCHS|KC261503 CGGAGGCACCGTCTTCCGCTTCGCTAAAGACCTGGCTGAGAACAACCGTG 547

EuCHS.P19cDNA GCGCTCGCGTCCTTGTGCTGCTGCTGAAATCACTGCCGTCACCTTCCGT 600
 EuCHS.Y05cDNA GCGCTCGCGTCCTTGTGCTGCTGCTGAAATCACTGCCGTCACCTTCCGT 600
 TfCHS|KC261503 GCGCTCGCGTCCTTGTGCTGCTGCTGAAATCACTGCCGTCACCTTCCGT 597

EuCHS.P19cDNA GGACCCTCAGAGTCCCACCTCGACAGCCTGGTTGGCCAGGCACTCTTTGG 650
 EuCHS.Y05cDNA GGACCCTCAGAGTCCCACCTCGACAGCCTGGTTGGCCAGGCACTCTTTGG 650
 TfCHS|KC261503 GGACCCTCGAGTCCCACCTCGACAGCCTGGTTGGCCAGGCACTCTTCGG 647

Figure A3.2 (to be continued on the next page)

EuCHS.P19cDNA	CGACGGAGCTGCTGCTGTCATCGTTGGCTCTGATCCCGACACCACCTGTGG	700
EuCHS.Y05cDNA	CGACGGAGCTGCTGCTGTCATCGTTGGCTCTGATCCCGACACCACCTGTGG	700
TfCHS KC261503	CGACGGAGCCGCTGCCGTCACTGTCGGCTCCGATCCCGACACCGCCGTGG	697
EuCHS.P19cDNA	AACGACCTCTGTTCCAGATTGTCTCAGCCAGCCAGACCATCCTCCCAGAT	750
EuCHS.Y05cDNA	AACGACCTCTGTTCCAGATTGTCTCAGCCAGCCAGACCATCCTCCCAGAT	750
TfCHS KC261503	AACGACCTCTGTTCCAGATTGTCTCGGCCAGCCAGACCATCCTCCCAGAT	747
EuCHS.P19cDNA	TCGGAGGGCGCCATCGACGGCCACCTCCGGGAAGTGGGCCTCACTTTCCA	800
EuCHS.Y05cDNA	TCGGAGGGCGCCATCGACGGCCACCTCCGGGAAGTGGGCCTCACTTTCCA	800
TfCHS KC261503	TCGGAGGGCGCCATCGACGGCCACCTCCGGGAAGTGGGCCTCACTTTCCA	797
EuCHS.P19cDNA	TCTGCTCAAGGACGTTCCCGGCCCTCATCTCCAAGAACATTGAGAAGAGCC	850
EuCHS.Y05cDNA	TCTGCTCAAGGACGTTCCCGGCCCTCATCTCCAAGAACATTGAGAAGAGCC	850
TfCHS KC261503	TCTGCTCAAGGACGTTCCCGGCCCTCATCTCCAAGAACATTGAGAAGAGCC	847
EuCHS.P19cDNA	TGGTGCAGGCCCTTCGCGCCACTGGGGATCTCGGACTGGAACCTCCCTCTTC	900
EuCHS.Y05cDNA	TGGTGCAGGCCCTTCGCGCCACTGGGGATCTCGGACTGGAACCTCCCTCTTC	900
TfCHS KC261503	TGGTGCAGGCCCTTCGCGCCACTGGGGATCTCGGACTGGAACCTCCCTCTTC	897
EuCHS.P19cDNA	TGGATTGCTCATCCCGGGGGCCCGGCCATTCTCGACCAGGTGGAGTTGAA	950
EuCHS.Y05cDNA	TGGATTGCTCATCCCGGGGGCCCGGCCATTCTCGACCAGGTGGAGTTGAA	950
TfCHS KC261503	TGGATTGCTCATCCCGGGGGCCCGGCCATTCTCGACCAGGTGGAGTTGAA	947
EuCHS.P19cDNA	GCTGGCTCTGGACAAGGAGAAGATGAAGGCGACAAGGCATGTGCTGAGTG	1000
EuCHS.Y05cDNA	GCTGGCTCTGGACAAGGAGAAGATGAAGGCGACAAGGCATGTGCTGAGTG	1000
TfCHS KC261503	GCTGGCTCTGGACAAGGAGAAGATGAAGGCGACAAGGCATGTGCTGAGTG	997
EuCHS.P19cDNA	AGTACGGCAATATGTCCAGCGCCTGCGTCCTCTTCATCCTTGATGAGATG	1050
EuCHS.Y05cDNA	AGTACGGCAATATGTCCAGCGCCTGCGTCCTCTTCATCCTTGATGAGATG	1050
TfCHS KC261503	AGTACGGCAACATGTCCAGCGCCTGCGTCCTCTTCATCCTTGATGAGATG	1047
EuCHS.P19cDNA	CGCAAGACGTCGGCCGAGCAGGGCAAGGCTACCACGGCGAGGGGGCTAGA	1100
EuCHS.Y05cDNA	CGCAAGACGTCGGCCGAGCAGGGCAAGGCTACCACGGCGAGGGGGCTAGA	1100
TfCHS KC261503	CGCAAGGCCCTCGGCCAGAGCAGGGGAAGGCTACCACCGCGAGGGGGCTAGA	1097
EuCHS.P19cDNA	CTGGGGGGTTCTAATTCGGCTTCGGCCCAGGACTCACCGTCGAGACCGTCCG	1150
EuCHS.Y05cDNA	CTGGGGGGTTCTAATTCGGCTTCGGCCCAGGACTCACCGTCGAGACCGTCCG	1150
TfCHS KC261503	CTGGGGGGTTTGTTCGGCTTCGGCCCAGGACTCACCGTCGAGACCGTCCG	1147
EuCHS.P19cDNA	TCCTGCACAGCCTTCCGATCACGACCAATTGA	1182
EuCHS.Y05cDNA	TCCTGCACAGCCTTCCGATCACGACCAATTGA	1182
TfCHS KC261503	TCCTGCACAGCCTTCCAATCACCTCCAACTGA	1179

Figure A3.2: Sequence alignment of the *EuChs* coding region from one purple- and one yellow-anthered individual and *TfChs* (from *Tulipa fosteriana*; GenBank no. KC261503), which is the BLAST top hit of *EuChs*. Identical sequences are highlighted in a black background. The nonsynonymous substitution between purple- and yellow-anthered individuals is highlighted in a red background. Dashes indicate alignment gaps.

EuF3H.P19cDNA	ATGGCGCCGACCTTCCTCCCCACA	AGTCTCCGACGAGAAGACCCTCCGCGC	50
EuF3H.Y05cDNA	ATGGCGCCGACCTTCCTCCCCACA	AGTCTCCGACGAGAAGACCCTCCGCGC	50
TfF3H KC261504	ATGGCTCCGACCTTCCTCCCCAC	GTCTCCGACGAGAAGACCCTCCGCGC	50
EuF3H.P19cDNA	CAGCTTCGTGCGCGACGAGGACGAGCGCCCCAAGGTCGCCTACAAC	TCT	100
EuF3H.Y05cDNA	CAGCTTCGTGCGCGACGAGGACGAGCGCCCCAAGGTCGCCTACAAC	TCT	100
TfF3H KC261504	CAACTTCGTCCGCGACGAGGACGAGCGCCCCAAGGTCGCCTACAAC	A	100
EuF3H.P19cDNA	TCAGCAACGACATTCCTGTCA	TCTCGCTAGCCGGGCTAGACGATGATGAA	150
EuF3H.Y05cDNA	TCAGCAACGACATTCCTGTAA	TCTCGCTAGCCGGGCTAGACGATGATGAA	150
TfF3H KC261504	TCAGCAACGACATCCCCGTCA	TCTCGCTAGCCGGGCTAGACGATGATGAA	150
EuF3H.P19cDNA	AAGGTTAGAACAGAGATATGCGACAAGATCGTCGCCGCTTGC	GAGGACTG	200
EuF3H.Y05cDNA	AAGGTTAGAACACAGAGATATGCGACAAGATCGTCGCCGCTTGC	GAGGACTG	200
TfF3H KC261504	AGCATCAGATCACAGAGATATGCGACAAGATCGTGCCGCTTGC	GAGGACTG	200
EuF3H.P19cDNA	GGGCATTTTTCAGGTGGTCGATCATGGTGTTCGATGCTGCTCTTGTGGCTG		250
EuF3H.Y05cDNA	GGGCATTTTTCAGGTGGTCGATCATGGTGTTCGATGCTGCTCTTGTGGCTG		250
TfF3H KC261504	GGGCATATTTTTCAGGTGGTTGATCATGGTGTTCGATGCTGCTCTTGTGGATG		250
EuF3H.P19cDNA	ATATGACGAGGCTGGCGAGGGA	CTTCTTTGCGCTGCCGCCGAGAGGAGAAG	300
EuF3H.Y05cDNA	ATATGACGAGGCTGGCGAGGGA	CTTCTTTGCGCTGCCGCCGAGAGGAGAAG	300
TfF3H KC261504	ATATGAACAGGCTGGCGAGGGA	CTTCTTTGCGCTGCCGCCAGAGGAGAAG	300
EuF3H.P19cDNA	TTGAGATTTGATATGTCCGGCGGGAAGAAGGGGGGTTTCATCGTCTCCAG		350
EuF3H.Y05cDNA	CTGAGATTTGACATGTCCGGCGGGAAGAAGGGGGGTTTCATCGTCTCCAG		350
TfF3H KC261504	CTGAGGTTTGATATGTCAAGGTGGGAAGAAGGGGGGTTTCATCGTCTCCAG		350
EuF3H.P19cDNA	CCATCTTCAGGGTGAAGCAGTGCAAGATTGGAGGGAGATAGTGACCTACT		400
EuF3H.Y05cDNA	CCATCTTCAGGGTGAAGCAGTGCAAGATTGGAGGGAGATAGTGACCTACT		400
TfF3H KC261504	CCATCTTCAGGGTGAAGCAGTGCAAGATTGGAGGGAGATAGTGACCTACT		400
EuF3H.P19cDNA	TCTCGTACCCGATCCGTGCCCGTGACTACTCAAGGTGGCCCCGACAAGCCC		450
EuF3H.Y05cDNA	TCTCGTACCCGATCCGTGCCCGTGACTACTCAAGGTGGCCCCGACAAGCCC		450
TfF3H KC261504	TCTCATACCCGATCCGTGCCCGTGACTACTCAAGGTGGCCCCGACAAGCCC		450
EuF3H.P19cDNA	GAGGGCTGGAGGGCCGTTGTTCGAGGCCTACAGCGAGCAATTGATGGGCTT		500
EuF3H.Y05cDNA	GAGGGCTGGAGGGCCGTTGTTCGAGGCCTACAGCGAGCAATTGATGGGCTT		500
TfF3H KC261504	GAGGGCTGGAAAGTCCGTTGTTCGAGGCCTACAGCGAGCAATTGATGGGCTT		500
EuF3H.P19cDNA	GGCCTGCAAGCTACTAGGGGTGTTGTCCGAAGCCATGGGCCTGGACAGGG		550
EuF3H.Y05cDNA	GGCCTGCAAGCTACTAGGGGTGTTGTCCGAAGCCATGGGCCTGGACAGGG		550
TfF3H KC261504	AGCCTGCAAGCTACTAGGGGTGTTATCCGAAGCCATGGGCCTGGACCGTG		550
EuF3H.P19cDNA	AGGCCCTGACCCAGGCTTGC	GTCGACATGGACCAGAAAGTGGTGGTCAAC	600
EuF3H.Y05cDNA	AGGCCCTGACCCAGGCTTGC	GTCGACATGGACCAGAAAGTGGTGGTCAAC	600
TfF3H KC261504	AGGCCCTGACCCAGGCTTGTGTCGACATGGACCAGAAAGTGGTGGTCAAC		600
EuF3H.P19cDNA	TTCTACCCCAAGTGCCCCCAACCTGACCTGACCCCTCGGGCTCAAGCGACA		650
EuF3H.Y05cDNA	TTCTACCCCAAGTGCCCCCAACCTGACCTGACCCCTCGGGCTCAAGCGACA		650
TfF3H KC261504	TTCTACCCCAAGTGCCCCCAACCTGACCTGACCCCTCGGGCTCAAGCGACA		650

Figure A3.3 (to be continued on the next page)

EuF3H.P19cDNA	CACCGACCCGGGCACCATCACCCCTCCTTCTCCAAGACCAGGTCGGC	GGGC	700
EuF3H.Y05cDNA	CACCGACCCGGGCACCATCACCCCTCCTTCTCCAAGACCAGGTCGGC	GGGC	700
TfF3H KC261504	CACCGACCCGGGCACCATCACCCCTCCTTCTCCAAGACCAGGTAGGT	GGGC	700
EuF3H.P19cDNA	TTCAGGCGACCAAGGACGGTGGCGACACTTGGATCACTGTCAAGCCGGTG		750
EuF3H.Y05cDNA	TTCAGGCGACCAAGGACGGTGGCGACACTTGGATCACTGTCAAGCCGGTG		750
TfF3H KC261504	TTCAGGCGACCAAGGACGGTGGCGACACTTGGATCACTGTCAAGCCGGTG		750
EuF3H.P19cDNA	GAAGGTGCTTTCTGTCTGTCAATCTCGGTGACCATGGACATTTCTTAAGCAA		800
EuF3H.Y05cDNA	GAAGGTGCTTTCTGTCTGTCAATCTCGGTGACCATGGACATTTCTTAAGCAA		800
TfF3H KC261504	GAAGGTGCTTTTGTGTGTCAATCTCGGTGACCATGGACATTTCTTGAGCAA		800
EuF3H.P19cDNA	CGGGAGGTTCAAGAATGCTGATCACCAGGCAGTGGTGAACTCGAACTGCA		850
EuF3H.Y05cDNA	CGGGAGGTTCAAGAATGCTGATCACCAGGCAGTGGTGAACTCGAACTGCA		850
TfF3H KC261504	TGGGAGGTTCAAGAATGCTGATCACCAGGCAGTGGTGAACTCGAACAAGCA		850
EuF3H.P19cDNA	GCCGTATATCGATCGCGACATTCCAGAATCCTGCACCAGATGCGATAGTG		900
EuF3H.Y05cDNA	GCCGTATATCGATCGCGACATTCCAGAATCCTGCACCAGATGCGATAGTG		900
TfF3H KC261504	GCCGTATATCGATCGCGACATTCCAGAATCCTGCACCAGATGCGACGGTG		900
EuF3H.P19cDNA	TACCCGCTGGCAATCAGAGAGGGGGAGAAATCAGTGTTGGACGCGCCAAT		950
EuF3H.Y05cDNA	TACCCGCTGGCAATCAGAGAGGGGGAGAAATCAGTGTTGGACGCGCCAAT		950
TfF3H KC261504	TACCCGCTGGCGATCAGAGAGGGGGAGAAATCAGTGTTAGACGCGCCAAT		950
EuF3H.P19cDNA	TACATTCAGTGAAATGTACAGGAAGAAGATGAGTAGAGACATTGAGCTTG		1000
EuF3H.Y05cDNA	TACATTCAGTGAAATGTACAGGAAGAAGATGAGTAGAGACATTGAGCTTG		1000
TfF3H KC261504	GACATTCAGTGAGATGTACAGGAAGAAGATGAGTAGAGACATTGAGCTTG		1000
EuF3H.P19cDNA	CGAAACTCAAGAAGCTGGCTAAGATGGAGGCGCTGGAGATTGCGCATAAG		1050
EuF3H.Y05cDNA	CGAAACTCAAGAAGCTGGCTAAGATGGAGGCGCTGGAGATTGCGCATAAG		1050
TfF3H KC261504	CGAAGCTCAAGAAGTGGCTAAGACTGGAGGCGCTGGATGCTGCGCATATG		1050
EuF3H.P19cDNA	GCTGAGGATGTTGTGCTTCAATCTGCGAAACTTGATCAAATTTCTTGCTTG		1100
EuF3H.Y05cDNA	GCTGAGGATGTTGTGCTTCAATCTGCGAAACTTGATCAAATTTCTTGCTTG		1100
TfF3H KC261504	GCTAAGGATGATGTGCTTCAATCTGAGAAACTTGATCAAATTTCTTGCTTG		1100
EuF3H.P19cDNA	A		1101
EuF3H.Y05cDNA	A		1101
TfF3H KC261504	A		1101

Figure A3.3: Sequence alignment of the *EuF3h* coding region from one purple-anthered individual and one yellow-anthered individual and *TfF3h* (from *Tulipa fosteriana*; GenBank no. KC261504), which is the BLAST top hit of *EuF3h*. Identical sequences are highlighted in a black background. The synonymous and nonsynonymous substitutions between purple- and yellow-anthered individuals are highlighted in green and red backgrounds, respectively.

DFR.P02cDNA	ATGAAGACTGTTAATGGACCTGTGGTGGTGACTGGAGCGAGTGGCTACGT	50
DFR.Y05gDNA_CDS	ATGAAGACTGTTAATGGACCTGTGGTGGTGACTGGAGCGAGTGGCTACGT	50
TfDFR KC261506	ATGAAGGTGTGAAAGGACCTGTCTGGTGACTGGAGCGAGTGGCTACGT	50
DFR.P02cDNA	TGGTTCGTGGCTTGTGATGAAGCTTCTCCAGGATGGCTACACCGTAAGGG	100
DFR.Y05gDNA_CDS	TGGTTCGTGGCTTGTGATGAAGCTTCTCCAGGATGGCTACACCGTAAGGG	100
TfDFR KC261506	TGGTTCGTGGCTCGTAATGAAGCTTCTGGAGAACGGCTATACCGTAAGGG	100
DFR.P02cDNA	CCACCGTTAGAGACCCAAAGGACCTAAGAAAAGACCAAACCTCTGTTGGAT	150
DFR.Y05gDNA_CDS	CCACCGTTAGAGACCCAAAGGACCTAAGAAAAGACCAAACCTCTGTTGGAT	150
TfDFR KC261506	CCACCGTTAGAGATCCAAAGGACCTTAGAAAAGATCAAACCACCTCTAGAT	150
DFR.P02cDNA	CTTCCTGGGGCTGACGAACGACTGACTATTTGGAAATCAGACCTTAACGA	200
DFR.Y05gDNA_CDS	CTTCCTGGGGCTGACGAACGACTGACTATTTGGAAATCAGACCTTAACGA	200
TfDFR KC261506	CTTCGAGGTTCTGATGAACAAGTACTGACTATCTGGAAAGCTGACCTGAATGA	200
DFR.P02cDNA	GGAAGGAAGCTTTGACGACGCGATTACTGGTTGCACCGGGTGTTCATG	250
DFR.Y05gDNA_CDS	GGAAGGAAGCTTTGACGACGCGATTACTGGTTGCACCGGGTGTTCATG	250
TfDFR KC261506	GGAAGGGAGCTTTGATGATGCGTTCAATGGTTGCACTGGAGTGTTCATG	250
DFR.P02cDNA	TCGCAACTCCGATGGATTTTGAGTCCAAAGATCCTGAGAACGAAGTAATA	300
DFR.Y05gDNA_CDS	TCGCAACTCCGATGGATTTTGAGTCCAAAGATCCTGAGAACGAAGTAATA	300
TfDFR KC261506	TCGCAACACCAATGGATTTTGAATCCATAGATCCTGAGAAATGAAGTAATA	300
DFR.P02cDNA	AAGCCAACTATAAATGGAGTGTAAAGTATAATGAAAGTCTTGCAAGAAAGC	350
DFR.Y05gDNA_CDS	AAGCCCACTATAAATGGAGTGTAAAGTATAATGAAAGTCTTGCAAGAAATGC	350
TfDFR KC261506	AAGCCGACATAAATGGAGTGTAGAGTATCTTGAGGTCTTGCAAGAAAGT	350
DFR.P02cDNA	TGGGACTGTCAAGCGTGTTATCTTTCACATCCTCTGCTGGGACGGTGAACG	400
DFR.Y05gDNA_CDS	TGGGACTGTTAAGCGTGTTATCTTTCACATCCTCTGCTGGGACGGTGAACG	400
TfDFR KC261506	TGGGACTGTCAAGCGTGTTATCTTTCACATCCTCTGCTGGGGCAGTGAAAT	400
DFR.P02cDNA	TTCAAGAAGAGCAAATGCCAGAGTATGACGGAAGACTCTTGAGCGGACATT	450
DFR.Y05gDNA_CDS	TTCAAGAAGAGCAAATGCCAGAGTATGAAAGAAGACTCTTGAGCGGACATT	450
TfDFR KC261506	TTCAAGAAGAGCAAATGCTAGAGTATGACGAAAAGTCTTGAGCGGACATT	450
DFR.P02cDNA	GACTTCTGCAGACGTGTGAAGATGACTGGATGGATGTACTTCGTATCCAA	500
DFR.Y05gDNA_CDS	GACTTCTGCAGACGTGTGAAGATGACTGGATGGATGTACTTCGTATCCAA	500
TfDFR KC261506	GACTTCTGCAGACGTGTGAAGATGACTGGATGGATGTACTTTGTATCTAA	500
DFR.P02cDNA	AACCTAGCAGAGAAAGCCGCATGGGAGTTTGCAAAAAGACAATGATATTC	550
DFR.Y05gDNA_CDS	AACCTAGCAGAGAAAGCCGCATGGGAGTTTGCAAAAAGACAATGATATTC	550
TfDFR KC261506	AACCTAGCAGAGAAAGCTGCATGGGAGTTTGCAAAAGGAAAATGATATAC	550
DFR.P02cDNA	AACTCATAAGCATCATTCCAACCTTGGTGGTCTGGTCCTTTTCATCACCCACA	600
DFR.Y05gDNA_CDS	AACTCATAAGCATCATTCCAACCTTGGTGGTCTGGTCCTTTTCATCACCCACA	600
TfDFR KC261506	AACTCATAAGCATCATTCCAACCTTGGTGGTGGTCCTTTTCATCTCCACA	600

Figure A3.4 (to be continued on the next page)

```

DFR.P02cDNA      AGTATGCCCTCCTAGTATGATAACAAGCTTTGTCTTGGATCACAGGAAATGA 650
DFR.Y05gDNA_CDS  AGTATGCCCTCCTAGTATGATAACAAGCTTTGTCTTGGATCACAGGAAATGA 650
TfDFR|KC261506   AGTATGCCCTCCTAGTATGATAACAAGCTTTGTCTTGGATCACAGGAAATGA 650

DFR.P02cDNA      TTCTCACTATTCAATCTTAAAGCAAATCCAACCTTGTTCACTTAGATGACC 700
DFR.Y05gDNA_CDS  TTCTCACTATTCAATCTTAAAGCAAATCCAACCTTGTTCACTTAGATGACC 700
TfDFR|KC261506   TTCTCACTATTCAATCTTAAAGCAAATCCAACCTTGTTCACTTAGATGACC 700

DFR.P02cDNA      TATGTAAAGCACATATTTTCTATTTGAAAATCCTGAAGCGAGTGGGAGA 750
DFR.Y05gDNA_CDS  TATGTAAAGCACATATTTTCTATTTGAAAATCCTGAAGCGAGTGGGAGA 750
TfDFR|KC261506   TATGCGTGGCACATATTTTCTATTTGAAAATCCAAGAACGAGTGGCAGA 750

DFR.P02cDNA      TACATATGTTCTTCTTATGATGCTACCATTTGGGATCTCGCAAGATTGAT 800
DFR.Y05gDNA_CDS  TACATATGTTCTTCTTATGATGCTACCATTTGGGATCTCGCAAGATTGAT 800
TfDFR|KC261506   TACATATGTTCTTCTTATGATGCTACCATTTGGGATCTTGCAAAATTTAT 800

DFR.P02cDNA      GAAAGATAGATACCCCTCAGTATGCCATCCCCAAGAATTTGAAGGTATTG 850
DFR.Y05gDNA_CDS  GAAAGATAGATACCCCTCAGTATGCCATCCCCAAGAATTTGAAGGTATTG 850
TfDFR|KC261506   GAAAGATAGATACCCCTCAGTATGCCATCCCCAAGAATTTGAAGGTATTG 850

DFR.P02cDNA      ATGACCTAATCAAACCAGTGCCTTCTCTTCAAGAAACTGAAGGATCTT 900
DFR.Y05gDNA_CDS  ATGACCTAATCAAACCAGTGCCTTCTCTTCAAGAAACTGAAGGATCTT 900
TfDFR|KC261506   ATGAGCGAATCAAACCAGTGCCTTCTCTTCAAGAAACTCATGGATCTT 900

DFR.P02cDNA      GGCTTCAACTATCAGTATACTGTGAGGAGATGTTTGATGAAGGAATCCG 950
DFR.Y05gDNA_CDS  GGCTTCAACTATCAGTATACTGTGAGGAGATGTTTGATGAAGGAATCCG 950
TfDFR|KC261506   GGCTTCAACTATCAGTATACTATGAGGAGATGTTTGATGAAGGAATCCA 950

DFR.P02cDNA      TTCGTGTACAGAAAAGAAGCTTTTACCCTCCAAACACAAGAAATGTTCT 1000
DFR.Y05gDNA_CDS  TTCGTGTACAGAAAAGAAGCTTTTACCCTCCAAACACAAGAAATGTTCT 1000
TfDFR|KC261506   TTCATGTACTGAAAAGAAGCTTTTACCCTCCAAACACAAGAAATGTTCT 1000

DFR.P02cDNA      ATGTC AATGATAAAAATTGATTTGGGTGG ----- 1028
DFR.Y05gDNA_CDS  ATGTC AATGATAAAAATTGATTTGGGTGG ----- 1028
TfDFR|KC261506   ATATCAATGATAAAAATTGATCTGGGTGGTAGCAAGATGAATTCAATCAAG 1050

DFR.P02cDNA      ----- 1028
DFR.Y05gDNA_CDS  ----- 1028
TfDFR|KC261506   GAGATGATGAGGGGACAAAAGCGAGCAACTGTCTATGTATACTATGGAGGA 1100

DFR.P02cDNA      ----- CAC 1031
DFR.Y05gDNA_CDS  ----- CAC 1031
TfDFR|KC261506   GATGTTTGTATGAAGGAATCCATTCATGTACTGAAAAGAACCTTTTACCAC 1150

DFR.P02cDNA      T----- 1032
DFR.Y05gDNA_CDS  T----- 1032
TfDFR|KC261506   TCCAAACACAAGAATTGTTCTATATCAATGATAAAAATTGATCTGGGTGGT 1200

```

Figure A3.4 (to be continued on the next page)

```

DFR.P02cDNA ---AAGATGAATTCAATCAATGAGATGATGAGGGGACAAAGTGAGCAAGT 1079
DFR.Y05gDNA_CDS ---AAGATGAATTCAATCAATGAGATGATGAGGGGACAAAGTGAGCAAGT 1079
TfDFR|KC261506 AGCAAGATGAATTCAATCAAGGAGATGATGAGGGGACAAAGCGAGCAACT 1250

DFR.P02cDNA GTCCATTGCTTTTCACTAA 1098
DFR.Y05gDNA_CDS GTCCATTGCTTTTCACTAA 1098
TfDFR|KC261506 GTCCACTCTTTTCACTAA 1269

```

Figure A3.4: Sequence alignment of the *EuDfr* coding region from one purple- and one yellow-anthered individual and *TfDfr* (from *Tulipa fosteriana*; GenBank no. KC261506), which is the BLAST top hit of *EuDfr*. Identical sequences are highlighted in a black background. The synonymous and nonsynonymous substitutions between purple- and yellow-anthered individuals are highlighted in green and red backgrounds, respectively. Dashes indicate alignment gaps.

EuANS.P02cDNA ATGCCAACCGTGACTTCACTTCTGCGAGGGTGGAGAGCCTGTCCGACAG 50
 EuANS.Y05gDNA_CDS ATGCCAACCGTGACTGCACTTCTGCGAGGGTGGAGAGCCTGTCCGACAG 50
 TfANS|KC261507 ATGCCAAC-----TCTGCGAGGGTGGAGAGCCTGTCCGACAG 38

EuANS.P02cDNA CGGGCTCGCCAACATCCCAAAGAATACGTCGGCCTGAATCGGAGCGCG 100
 EuANS.Y05gDNA_CDS CGGGCTCGCCAACATCCCAAAGAATACGTCGGCCTGAATCGGAGCGCG 100
 TfANS|KC261507 CGGGCTCGCCAACGATCCCAAAGAATACGTCGGCCTGAATCGGAGCGCG 88

EuANS.P02cDNA ACAACCTCGGGGACGCTTTCGATGAAGCACCAAACTCGACTCTGCCGGT 150
 EuANS.Y05gDNA_CDS ACAACCTCGGGGACGCTTTCGATGAAGCACCAAACTCGACTCTGCCGGT 150
 TfANS|KC261507 ACAACCTCGGGGACGCTTTCGATGAAGCACCAAACTCGACTCTGCCGGT 138

EuANS.P02cDNA CCTCAAAGTGCCATACTCGTCGACCTAGCGGGGTTTGACTCCACGGATGAGAA 200
 EuANS.Y05gDNA_CDS CCTCAAAGTGCCATACTCGTCGACCTAGCGGGGTTTGACTCCACGGATGAGAA 200
 TfANS|KC261507 CCTCAAAGTGCCATGTCGTCGACCTAGCGGGGTTTGACTCCACGGATGAGAA 188

EuANS.P02cDNA GGAGAGGGCGAAGTGCGTGGAGGCGCTGAGGGCGGCGGGCGGAGGACTGGG 250
 EuANS.Y05gDNA_CDS GGAGAGGGCGAAGTGCGTGGAGGCGCTGAGGGCGGCGGGCGGAGGACTGGG 250
 TfANS|KC261507 GGAGAGGGCGAAGTGCGTGGAGGCGCTGAGGAAAGGCGGGCGGAGGACTGGG 238

EuANS.P02cDNA GGGTGATGCACATTGTGAACCATGGAGTGGCACAGGAGCTGATTGAGAAG 300
 EuANS.Y05gDNA_CDS GGGTGATGCACATTGTGAACCATGGAGTGGCACAGGAGCTGATTGAGAAG 300
 TfANS|KC261507 GGGTGATGCACATTGTGAACCATGGAAATGCAAAGGAGGTGATTGAGAAG 288

EuANS.P02cDNA GTGAGGGAGGCGGGGAAGGCCTTTTTTCGGCCTCCCGGTGGGGGAGAAGGA 350
 EuANS.Y05gDNA_CDS GTGAGGGAGGCGGGGAAGGCCTTTTTTCGGCCTCCCGGTGGGGGAGAAGGA 350
 TfANS|KC261507 GTAAGGGAGGTGGGGGAAGGCCTTTTTTCGACCTCCCGGTGGGGGAGAAGGA 338

EuANS.P02cDNA GAAGTATGCGAATGATCAGGAATCCGGGGACATTCAGGGGTACGGGAGTA 400
 EuANS.Y05gDNA_CDS GAAGTATGCGAATGATCAGGAATCCGGGGACATTCAGGGGTACGGGAGTA 400
 TfANS|KC261507 GAAGTATGCGAATGATCAGGAATCCGGGGACATTCAGGGGTACGGGAGTA 388

EuANS.P02cDNA AGCTGGCGAATAACGAGTCTGGGCAGCTTGAGTGGCAGGATTACTTTTTT 450
 EuANS.Y05gDNA_CDS AGCTGGCGAATAACGAGTCTGGGCAGCTTGAGTGGCAGGATTACTTTTTT 450
 TfANS|KC261507 AGCTGGCGAATAACGAGTCTGGGCAGCTTGAGTGGCAGGATTACTTTTTT 438

EuANS.P02cDNA CACCTCATATTTCCGAGGAAAGACCAATTTGGCCTCTGGCCCAAGCA 500
 EuANS.Y05gDNA_CDS CACCTCATATTTCCGAGGAAAGACCAATTTAGCCTCTGGCCCAAGCA 500
 TfANS|KC261507 CACCTCATTTTTCCAGAGGAGAAAGACCAATTTGGCACTCTGGCCCAAGCA 488

EuANS.P02cDNA ACCGGCAGAGTATACAGAGGTAACAAAGGAGTTCGCCAAGCAGCTGAGAG 550
 EuANS.Y05gDNA_CDS ACCGGCAGAGTATACAGAGGTAACAAAGGAGTTCGCCAAGCAGCTGAGAG 550
 TfANS|KC261507 ACCGGCAGAGTATACAGAGGTAACAAAGGAGTTCGCCAAGCAGCTGAGAG 538

EuANS.P02cDNA CGGTGGCGACCAAGATGCTGTCCATGCTCTCTAGGTCTGGGCCTCGAA 600
 EuANS.Y05gDNA_CDS CGGTGGCGACCAAGATGCTGTCCATGCTCTCTCTAGGTCTGGGCCTCGAA 600
 TfANS|KC261507 TGGTGGCGACCAAGATGCTGTCCATGCTCTCTCTGGGTCTGGGCCTCGAA 588

Figure A3.5 (to be continued on the next page)

EuANS.P02cDNA	TCCGGCAAGCTCGAGAAAGAGCTCGGCGGAATGGAGGAGCTTCTGATGCA	650
EuANS.Y05gDNA_CDS	TCCGGCAAGCTTGAGAAAGAGCTCGGCGGAATGGAGGAGCTTCTGATGCA	650
TfANS KC261507	TCCGGCAAGCTCGAGAAAGAGCTCGGCGGAATGGAGGAGCTTCTGATGCA	638
EuANS.P02cDNA	GATGAAAATCAACTACTACCCGAAATGCCGAGCCAGAGCTCGCCCTCG	700
EuANS.Y05gDNA_CDS	GATGAAAATCAACTACTACCCGAAATGTCGCAACCAGATCTCGCCCTCG	700
TfANS KC261507	GATGAAAATCAACTACTACCCCAAATGCCGAGCCAGAGCTAGCCCTCG	688
EuANS.P02cDNA	GCGTGAAGCCACACCCGACGTCAGCTCCCTCACCTTCTCCTCACCAAC	750
EuANS.Y05gDNA_CDS	GCGTGAAGCCACACCCGACGTCAGCTCCCTCACCTTCTCCTCACCAAC	750
TfANS KC261507	GCGTGAAGCCACACCCGACGTCAGCTCCCTCACCTTCTCCTCACCAAC	738
EuANS.P02cDNA	ATGGTCCCCGGCCTTCAGCTCTACTATGACGACAAATGGGTGATCGCGCA	800
EuANS.Y05gDNA_CDS	ATGGTCCCCGGCCTTCAGCTCTACTATGACGACAAATGGGTGATCGCGGA	800
TfANS KC261507	ATGGTCCCCGGCCTTCAGCTCTACTATGACGACAAATGGGTGATCGCGGA	788
EuANS.P02cDNA	GTGCGTCCCTGACTCCCTTCTCGTCCACATTGGCGACACCCTCGAGATCC	850
EuANS.Y05gDNA_CDS	GTGCGTCCCTGACTCCCTTCTCGTCCACATTGGCGACACCCTCGAGATCC	850
TfANS KC261507	GTGCGTCCCTGACTCCCTTCTCGTCCACATTGGCGACACGCTCGAGATCC	838
EuANS.P02cDNA	TCAGCAACGGTAGCTACAGGAGCATTTTACATAGGAGCTTGGTGAACAAG	900
EuANS.Y05gDNA_CDS	TCAGCAACGGTAGCTACAGGAGCATTTTACATAGGAGCTTAGTGAACAAG	900
TfANS KC261507	TCAGCAATGGTAGCTACAGGAGCATTTTACACAGGAGCTTGGTGAACAAG	888
EuANS.P02cDNA	GACAGGGTTCGGATCTCTTGGGCAGTGTTTTGTGAGCCGCCAAAGGAGAC	950
EuANS.Y05gDNA_CDS	GACAGGGTTCGGATCTCTTGGGCAGTGTTTTGTGAGCCGCCAAAGGAGAC	950
TfANS KC261507	GACAGGGTTCGGATACTCTTGGGCAGTGTTTTGTGAGCCGCCAAAGGAGAC	938
EuANS.P02cDNA	GATCGTGCTGCAGCCGCTGCCAGAGCTGGTGAGTGAGGCGGCGCCGGCTA	1000
EuANS.Y05gDNA_CDS	GATCGTGCTGCAGCCGCTGCCAGAGCTGGTGAGTGAGGCGGCGCCGGCTA	1000
TfANS KC261507	GATCGTGCTGCAGCCGCTGCCAGAGCTGGTGAGTGAGGCGGCGCCGGCTA	988
EuANS.P02cDNA	AGTTTCCTCCTCGGACCTTCAAGCAGCACATCCAGCACAAGCTGTTCAAG	1050
EuANS.Y05gDNA_CDS	AGTTTCCTCCTCGGACCTTCAAGCAGCACATCCAGCACAAGCTGTTCAAG	1050
TfANS KC261507	AGTTTCCTCCTCGGACCTTCAAGCAGCACATCCAGCACAAGCTGTTCAAG	1038
EuANS.P02cDNA	AAGACGGAGGAGAACTTCTCTCTCCTTAAATGA	1083
EuANS.Y05gDNA_CDS	AAGACGGAGGAGAACTTCTCTCTCCTTAAATGA	1083
TfANS KC261507	AAGACGGAGGAGAGTGGCTCTCCTTAAATGA	1071

Figure A3.5: Sequence alignment of the *EuAns* coding region from one purple- and one yellow-anthered individual and *TfAns* (from *Tulipa fosteriana*; GenBank no. KC261507), which is the BLAST top hit of *EuAns*. Identical sequences are highlighted in a black background. The synonymous and nonsynonymous substitutions between purple- and yellow-anthered individuals are highlighted in green and red backgrounds, respectively. Dashes indicate alignment gaps.

EuUF3GT.P02cDNA	ATGGGCTCGACCGGAAACCCCCACGTGCGCCCTCATCGCCTTCCCCTTCGG	50
EuUF3GT.Y05gDNA_CDS	ATGGGCTCGACCGGAAACCCCCACGTGCGCCCTCATCGCCTTCCCCTTCGG	50
TfUF3GT KF792732	ATGGGCTCGACCGGAAACCCCCACGTGCGCCCTCATCGCCTTCCCCTTCAG	50
EuUF3GT.P02cDNA	CACCCACGCCGCCCTCTCTTCTCCCTCAC	100
EuUF3GT.Y05gDNA_CDS	CACCCACGCCGCCCTCTCTTCTCCCTCAC	100
TfUF3GT KF792732	CACCCACGCCGCCCTCTCTTCTCCCTCAC	100
EuUF3GT.P02cDNA	CCCCCTCCGCCACCTTCTCCTTCATCAACTCCGCCCGCTCAACGCTTCC	150
EuUF3GT.Y05gDNA_CDS	CCCCCTCCGCCACCTTCTCCTTCATCAACTCCGCCCGCTCAACGCTTCC	150
TfUF3GT KF792732	CCCCCTCCGCCACCTTCTCCTTCATCAACTCCGCCCGCTCAACGCTTCC	150
EuUF3GT.P02cDNA	CTCGCCCGTGACATCTCTGCCGTCCCGTCTGGAGCCAAACATCAGGGTTTA	200
EuUF3GT.Y05gDNA_CDS	CTCGCCCGTGACATCTCTGCCGTCCCGTCTGGAGCCAAACATCAGGGTTTA	200
TfUF3GT KF792732	CTCGCCCGTGACATCTCTGCCGTCCCGTCTGGAGCCAAACATCAGGGTTTA	200
EuUF3GT.P02cDNA	CGACATCTCTGACGGGTGCCCGGACGGGTA	250
EuUF3GT.Y05gDNA_CDS	CGACATCTCTGACGGGTGCCCGGACGGGTA	250
TfUF3GT KF792732	CGACATCTCTGACGGGTGCCCGGACGGGTA	250
EuUF3GT.P02cDNA	CGGAGGAGGAGGTGGTCTGTTTCTT	300
EuUF3GT.Y05gDNA_CDS	CGGAGGAGGAGGTGGTCTGTTTCTT	300
TfUF3GT KF792732	CGGAGGAGGAGGTGGTCTGTTTCTT	300
EuUF3GT.P02cDNA	GAGGCGATGGAGAAGGCAGTGGAAAGGAGCCGGTGGGAGGAGGATTAGCTG	350
EuUF3GT.Y05gDNA_CDS	GAGGCGATGGAGAAGGCAGTGGAAAGGAGCCGGTGGGAGGAGGATTAGCTG	350
TfUF3GT KF792732	GAGGCGATGGAGAAGGCAGTGGAAAGGAGCCGGTGGGAGGAGGATTAGCTG	350
EuUF3GT.P02cDNA	CATTGTCAGCGACGCTTTTATCTGGTTTGCTGGGAAGATGGCGGAGGAGA	400
EuUF3GT.Y05gDNA_CDS	CATTGTCAGCGACGCTTTTATCTGGTTTGCTGGGAAGATGGCGGAGGAGA	400
TfUF3GT KF792732	CATTGTCAGCGACGCTTTTATCTGGTTTGCTGGGAAGATGGCGGAGGAGA	400
EuUF3GT.P02cDNA	TGAGGGTGCCGTGGGTGCCGCTGTGGACTGGCGGACCTTACAGCCTTGCC	450
EuUF3GT.Y05gDNA_CDS	TGAGGGTGCCGTGGGTGCCGCTGTGGACTGGCGGACCTTACAGCCTTGCC	450
TfUF3GT KF792732	TGAGGGTGCCGTGGGTGCCGCTGTGGACTGGCGGACCTTACAGCCTTGCC	450
EuUF3GT.P02cDNA	ACTCATATGTATACTGACTTCCTTCGTCTCAAGTTTGGGGAAACAAGTTAC	500
EuUF3GT.Y05gDNA_CDS	ACTCATATGTATACTGACTTCCTTCGTCTCAAGTTTGGGGAAACAAGTTAC	500
TfUF3GT KF792732	GCTCATATGTATACTGACTTCCTTCGTCTCAAGTTTGGGGAAACAAGTTAC	500
EuUF3GT.P02cDNA	ACCCTCCCGCCTAAGTGAACCTCTGGACTGCATCCCCTACATGTCTGCTC	550
EuUF3GT.Y05gDNA_CDS	ACCCTCCCGCCTAAGTGAACCTCTGGACTGCATCCCCTACATGTCTGCTC	550
TfUF3GT KF792732	ACCCTCCCGCCTAAGTGAACCTCTGGACTGCATCCCCTACATGTCTGCTC	550
EuUF3GT.P02cDNA	TGCAAGTCCGAGAGCTCCCAGAAGGGATCGTTTTTCGGCAACAAGACTCC	600
EuUF3GT.Y05gDNA_CDS	TGCAAGTCCGAGAGCTCCCAGAAGGGATCGTTTTTCGGCAACAAGACTCC	600
TfUF3GT KF792732	TGCAAGTCCGAGAGCTCCCAGAAGGGATCGTTTTTCGGCAACAAGACTCC	600

Figure A3.6 (to be continued on the next page)

EuUF3GT.P02cDNA	GTCTTCGCCCGCCTCGTCCACAGCATGGCCAAGGAGCTCCCGCACGCCAC	650
EuUF3GT.Y05gDNA_CDS	GTCTTCGCCCGCCTCGTCCACAGCATGGCCAAGGAGCTCCCGCACGCCAC	650
TfUF3GT KF792732	GTCTTCGCCCGCCTCGTCCACAGCATGGCCAAGGAGCTCCCGCACGCCAC	650
EuUF3GT.P02cDNA	CACCGTCGCCCTCAACACCTTCCATGGCCTGGACCCCGCTGTAGACCGAG	700
EuUF3GT.Y05gDNA_CDS	CACCGTCGCCCTCAACACCTTCCATGGCCTGGACCCCGCTGTAGACCGGG	700
TfUF3GT KF792732	CACCGTCGCCCTCAACACCTTCCATGGCCTGGACCCCGCTGTAGACCGGG	700
EuUF3GT.P02cDNA	ACTTAGACACCAAGTTCAAGCGCTCCCTCTCCATCGGCCCCCTACACCTC	750
EuUF3GT.Y05gDNA_CDS	ACTTAGACACCAAGTTCAAGCGCTCCCTCTCCATCGGCCCCCTACACCTC	750
TfUF3GT KF792732	ACTTAGACACCAAGTTCAAGCACTCCCTCTCCATCGGCCCCCTACACCTC	750
EuUF3GT.P02cDNA	CTCACCCACAAACCCCGCAGCCAGCCGACTCCTACGGCTGGCTCAAGTG	800
EuUF3GT.Y05gDNA_CDS	CTCACCCACAAACCCCGCAGCCAGCCGACTCCTACGGCTGGCTCAGTG	800
TfUF3GT KF792732	ATCACCCACAAACCCCGCAGCCAGCCGACTCCTACGGCTGGATCAGTG	800
EuUF3GT.P02cDNA	GCTCGACAAACACGGCCCCCGATACAGTGGTCTACATCAGCTTCGGCACCA	850
EuUF3GT.Y05gDNA_CDS	GCTCGACAAACACGGCCCCCGCTACAGTGGTCTACATCAGCTTCGGCACCA	850
TfUF3GT KF792732	GCTCGACAAACAAAGACCCCGCTACAGTGGTCTACATCAGCTTCGGCACCA	850
EuUF3GT.P02cDNA	TCATGTCTCTCCGCCGCCAGAGGCAGCTGCGCTTGGCAGAGGACTCGAG	900
EuUF3GT.Y05gDNA_CDS	TCATGTCTCTCCGCCGCCAGAGGCAGCTGCGCTTGGCAGAGGACTCGAG	900
TfUF3GT KF792732	TCATGTCTCTCCGCCGCCAGAGGCAGCTGCGCTTGGCAGAGGGGCTCGAG	900
EuUF3GT.P02cDNA	GTGAGCGGTGTCCCGTTCATATGGTCCGCTGAAGGCGGGGGCAGGCGTA	950
EuUF3GT.Y05gDNA_CDS	GTGAGCGGTGTCCCGTTCATATGGTCCGCTGAAGGCGGGGGCAGGCGTA	950
TfUF3GT KF792732	GTGAGCGGTGTCCCGTTCATATGGTCCGCTGAAGGCGGGGGCAGGCGTA	950
EuUF3GT.P02cDNA	CCTTCCATCTGGATTCTGGAACGCACCAAGCGGCCGCGGGCTTGTGGTGC	1000
EuUF3GT.Y05gDNA_CDS	CCTTCCATCTGGATTCTGGAACGCACCAAGCGGCCGCGGGCTTGTGGTGC	1000
TfUF3GT KF792732	CCTTCCATCTGGATTCTGGAACGCACCAAGCGGCCGCGGGCTTGTGGTGC	1000
EuUF3GT.P02cDNA	CATGGGTGCCGCAGCTGAAAGTACTCAACCACCTCGGCGGTGGGGGCGTTT	1050
EuUF3GT.Y05gDNA_CDS	CATGGGTGCCGCAGCTGAAAGTACTCAACCACCTCGGCGGTGGGGGCGTTT	1050
TfUF3GT KF792732	CATGGGTGCCGCAGCTGAAAGTACTCAACCACCTCGGCGGTGGGGGCGTTT	1050
EuUF3GT.P02cDNA	GTGACGCATTGTGGAATGGAACCTCGGTGATGGAGAGTGTACCGGCGGTGT	1100
EuUF3GT.Y05gDNA_CDS	GTGACGCATTGTGGAATGGAACCTCGGTGATGGAGAGTGTACCGGCGGTGT	1100
TfUF3GT KF792732	GTGACGCATTGTGGGTGGAACCTCGGTGATGGAGAGTGTACTGGCGGTGT	1100
EuUF3GT.P02cDNA	TCCGATGATTTGCCGTCCCTTCCTGGGGGATCAGAGACTGAATGCTGGGG	1150
EuUF3GT.Y05gDNA_CDS	TCCGATGATTTGCCGTCCCTTCCTGGGGGATCAGAGACTGAATGCTGGGG	1150
TfUF3GT KF792732	TCCGATGATTTGCCGACCCCTTCCTGGGGGATCAGAGACTGAATGCAGGGG	1150
EuUF3GT.P02cDNA	TGGTGTACCGTGTGTGGAAGATCGGGGCGGGATTTGAGGGTGGAAATAGTG	1200
EuUF3GT.Y05gDNA_CDS	TGGTATCGCGTGTGTGGAAGATCGGGGCGGGATTTGAGGGTGGAAATAGTG	1200
TfUF3GT KF792732	TGGTGTCCCGTGTGTGGAAGATCGGGGCGGGATTTGAGGGTGGCATAGTG	1200

Figure A3.6 (to be continued on the next page)

EuUF3GT.P02cDNA	ACTAAGGGAGCGCGGAGAAGGCTTTGC	GCTTGGTTTTGATGGAGGATGG	1250
EuUF3GT.Y05gDNA_CDS	ACTAAGGGAGCGCGGAGAAGGCTTTGC	GCTTGGTTTTGATGGAGGATGG	1250
TfUF3GT KF792732	ACTCAAGGGAGCGCGGAGAAGGCTTTGA	GCTTGGTTTTGATGGAGGACGG	1250
EuUF3GT.P02cDNA	AGGGAAGGCGATGAGGGAGAGGG	TTGGGAAGCTGCGGGAGATGGCGATAG	1300
EuUF3GT.Y05gDNA_CDS	AGGGAAGGCGATGAGGGAGGGG	TTGGGAAGCTGCGGGAGATGGCGATAG	1300
TfUF3GT KF792732	AGGGAAGGCGATGAGGGAGAGGA	TTGGGAAGCTGCGGGAGATGGCATAG	1300
EuUF3GT.P02cDNA	GGGCGACGAAGCCTGGTGA	GAGCTCCACAATAATTTTC	AAGCTTTGGTG 1350
EuUF3GT.Y05gDNA_CDS	GGGCGACAAGCCTGGTGA	GAGCTCCACAATAATTTTC	AAGCTTTGGTG 1350
TfUF3GT KF792732	GGGCAACGAAGCCAAGGTGG	GAGCTCCAGAAATAATCTTG	AAGCTTTGGTG 1350
EuUF3GT.P02cDNA	GAGGTGATTTGTGGATACTGA		1371
EuUF3GT.Y05gDNA_CDS	GAGGTGATTTGTGGATACTGA		1371
TfUF3GT KF792732	GAGGTCATTTGTGGATACTGA		1371

Figure A3.6: Sequence alignment of the *EuUf3gt* coding region from one purple- and one yellow-anthered individual and *TfUf3gt* (from *Tulipa fosteriana*; GenBank no. KF792732), which is the BLAST top hit of *EuUf3gt*. Identical sequences are highlighted in a black background. The synonymous and nonsynonymous substitutions between purple- and yellow-anthered individuals are highlighted in green and red backgrounds, respectively.

```

EuWDR1.P02cDNA ----- 0
EuWDR1.Y05cDNA ----- 0
EgTTG1|XM_010935714 ATGGAGAGGTCGCCCCAAGAATCCCCTCTGACGACGACGACGACGACGCGC 50

EuWDR1.P02cDNA ----- 0
EuWDR1.Y05cDNA ----- 0
EgTTG1|XM_010935714 GACGACGGCGAACCCCTAACTTGGCGAACACGTACACCTTCGACTCCCCTC 100

EuWDR1.P02cDNA ----- 0
EuWDR1.Y05cDNA ----- 0
EgTTG1|XM_010935714 ACCCGGTGTACGCCATGGCCTTCTCCTCCCTCCCGGCCCTCCCCCCG 150

EuWDR1.P02cDNA ----- 0
EuWDR1.Y05cDNA ----- 0
EgTTG1|XM_010935714 CGCCTCGCCCTGGGCTCCTTCATCGAGGACTACGCCAACCGGTTCGACGT 200

EuWDR1.P02cDNA ----- 0
EuWDR1.Y05cDNA ----- 0
EgTTG1|XM_010935714 CGTACCTTCGACGAGGACGCCCGCGCCTTCCGCCCGACCCCTCCCTCT 250

EuWDR1.P02cDNA ----- 0
EuWDR1.Y05cDNA ----- 0
EgTTG1|XM_010935714 CCTTCGACCACCCCTATCCCCCACCAAGCTCATGTTCCACCCCAAGCCC 300

EuWDR1.P02cDNA ----- 0
EuWDR1.Y05cDNA ----- 0
EgTTG1|XM_010935714 CTCCCCAATCCTCCTCCTCTCTCCTCGCCTCCTCCGGCGAGTTCCTTCG 350

EuWDR1.P02cDNA ----- 0
EuWDR1.Y05cDNA ----- 0
EgTTG1|XM_010935714 CCTCTGGCAGCTCGATCACCACGCCGACTCCTCGCCCAAGGCCGAGCTTC 400

EuWDR1.P02cDNA -----GACTTCTCGCCCCCTCACC 21
EuWDR1.Y05cDNA -----GACTTCTCGCCCCCTCACC 21
EgTTG1|XM_010935714 GCGCCATCCTCAACAACAGCAAGTCCAGCGAGTTCTCGCCCCAATCACC 450

EuWDR1.P02cDNA TCCTTCGACTGGAACCCC GCCGAGCCCCGCCGCATCGGCACCTCCTCCAT 71
EuWDR1.Y05cDNA TCCTTCGACTGGAACCCC GCCGAGCCCCGCCGCATCGGCACCTCCTCCAT 71
EgTTG1|XM_010935714 TCCTTCGATGGAACGATGCCGAGCCCCGCCGCATCGGCACCTCCTCGAT 500

EuWDR1.P02cDNA CGACACCACCTGCACCA TCTGGGACCTCGAGCGCTCCGTCTCGAGACCC 121
EuWDR1.Y05cDNA CGACACCACCTGCACCA TCTGGGACCTCGAGCGCTCCGTCTCGAGACCC 121
EgTTG1|XM_010935714 CGACACCACCTGCACCG TCTGGGACATCGAGCGCGGCGCAATCGAGACCC 550

EuWDR1.P02cDNA AGCTCATCGCCCACGACAAGGAGGTCTACGACATCGCCTGGGGCGAGGCC 171
EuWDR1.Y05cDNA AGCTCATCGCCCACGACAAGGAGGTCTACGACATCGCCTGGGGCGAGGCC 171
EgTTG1|XM_010935714 AGCTCATCGCCCACGACAAGGAGGTCTACGACATCGCCTGGGGCGAGGCC 600

EuWDR1.P02cDNA GGCGTCTTCGCCTCCGTCTCCGCCGACGGCTCCGTCCGCATCTTCGACCT 221
EuWDR1.Y05cDNA GGCGTCTTCGCCTCCGTCTCCGCCGACGGCTCCGTCCGCATCTTCGACCT 221
EgTTG1|XM_010935714 GGCGTCTTCGCCTCCGTCTCCGCCGACGGCTCCGTCCGCATCTTCGACCT 650

EuWDR1.P02cDNA CCGCGACAAGGAGCACTCCACCATCGTCTACGAGTCCCCAACCCCGACA 271
EuWDR1.Y05cDNA CCGCGACAAGGAGCACTCCACCATCGTCTACGAGTCCCCAACCCCGACA 271
EgTTG1|XM_010935714 CCGCGACAAGGAGCACTCCACCATCGTCTACGAGAACCCCGCCGGACA 700

```

Figure A3.7 (to be continued on the next page)

```

EuWDR1.P02cDNA C C C C G C T C C T C C G C C T C G C C T G G A A C A A G A T G G A C C T G C G C T A C A T C G C C 321
EuWDR1.Y05cDNA C C C C G C T C C T C C G C C T C G C C T G G A A C A A G A T G G A C C T G C G C T A C A T C G C C 321
EgTTG1|XM_010935714 C C C C G C T C C T C C G T C T C G C C T G G A A C A A G A C T G A C C T C C G C T A C A T G G C C 750

EuWDR1.P02cDNA A C C A T A C T T A T G G A C A G T A A C A A G G T C A T T G T G C T C G A C A T A C G C T C C C C 371
EuWDR1.Y05cDNA A C C A T A C T T A T G G A C A G T A A C A A G G T C A T T G T G C T C G A C A T A C G C T C C C C 371
EgTTG1|XM_010935714 A C C A T C C T C A T G G A C A G C A A C C G C G T C G T C A T C G T C G A C A T C C G C T C C C C 800

EuWDR1.P02cDNA G G C A T G C C C G T G G C C G A G C T G C A G A G G C A C A G G G C C G C G T C A A C G C G G 421
EuWDR1.Y05cDNA G G C A T G C C C G T G G C C G A G C T G C A G A G G C A C A G G G C C G C G T T A A C G C G G 421
EgTTG1|XM_010935714 G G C G G T G C C C G T C G C C G A G C T G C A G C G C C A C C G T G C A A G G T C A A T G C G G 850

EuWDR1.P02cDNA T T G C C T G G G C G C C G C A G G C G C C A G G C A C A T T T G C T C T G C T G G G G A C G A C 471
EuWDR1.Y05cDNA T T G C C T G G G C G C C G C A G G C G C C A G G C A C A T T T G C T C C G C G G G G A C G A C 471
EgTTG1|XM_010935714 T T G C C T G G G C G C C G C A G G C T G C C A G G C A C A T C T G C T C C G C T G G G G A C G A C 900

EuWDR1.P02cDNA G G G A A G G C G C T C A T A T G G G A G C T G C C G G T G G C C G C C G T G G C G C C G C G G T 521
EuWDR1.Y05cDNA G G G A A G G C G C T C A T A T G G G A G C T G C C G G T G G C C G C C G T G G C G C C G C G G T 521
EgTTG1|XM_010935714 G G C C A A G C G C T C A T A T G G G A G T G C C G G C C G C C - - - - G G G C C G - C G G T 944

EuWDR1.P02cDNA G C C G C T G G A T A A T G T C G A C C C G G T G C T G G A G T A T G C G G C G G G G C G G A G A 571
EuWDR1.Y05cDNA G C C G C T G G A T A A T G T C G A C C C G G T G C T G G A G T A T G C G G C G G G G C G G A G A 571
EgTTG1|XM_010935714 G C C G C C G G A C G G G A T C G A T C C G G C G C T G G T T A C T C G G C G G G A G C A G A G A 994

EuWDR1.P02cDNA T T A A T C A G G T T C A G T G G T C G C C G G C G C A G C C G G A T T G G A T C G G G A T T G C G 621
EuWDR1.Y05cDNA T T A A T C A G G T T C A G T G G T C G C C G G C G C A G C C G G A T T G G A T C G G G A T T G C G 621
EgTTG1|XM_010935714 T C A A C C A G C T G C A G T G G T C G C C G C G C A C C C G G A C T G G A T T G G G A T C G C A 1044

EuWDR1.P02cDNA T T C A G T A A C A A G G T G C A G C T G C T G A A A G T T T G A 654
EuWDR1.Y05cDNA T T C A G T A A C A A G G T G C A G C T G C T G A A A G T T T G A 654
EgTTG1|XM_010935714 T T C G C C A A C A A G G T G C A G C T G C T G A G A G T C T G A 1077

```

Figure A3.7: Sequence alignment of the *EuWDR1* coding region from one purple- and one yellow-anthered individual and *EgTTG1* (from *Elaeis guineensis*; GenBank no. XM_010935714), which is the BLAST top hit of *EuWDR1*. Identical sequences are highlighted in a black background. The synonymous substitutions between purple- and yellow-anthered individuals are highlighted in a green background. Dashes indicate alignment gaps.

EuWDR2.P02cDNA ATGGTGGCGAGC--AGCGGCGGCGGCGA CCGAACCTCCCCACCGACC CC 48
 EuWDR2.Y05cDNA ATGGTGGCGAGC--AGCGGCGGCGGCGA CCGAACCTCCCCACCGACC CC 48
 NtWDR|KJ879944 ATGGCCGCCGGGAGCGACACCGGAGA-----CG 30

EuWDR2.P02cDNA TCCGACGAGCAGCAGCGGCGGTCGGAGATCTACACCTACGAGGCGCCATG 98
 EuWDR2.Y05cDNA TCCGACGAGCAGCAGCGGCGGTCGGAGATCTACACCTACGAGGCGCCATG 98
 NtWDR|KJ879944 TCGGAGGAGCAGCAGAAAGCGTTCGGAGATCTACACGTACGAGGCGCCGTG 80

EuWDR2.P02cDNA GCACATCTACGCCATGAAGTGGTCCGTCCGCAAGGACAAGAAGTACCGCC 148
 EuWDR2.Y05cDNA GCACATCTACGCCATGAAGTGGTCCGTCCGCAAGGACAAGAAGTACCGCC 148
 NtWDR|KJ879944 GCACATAACGCCATGAAGTGGTCCGTGAGGCGGACAAGAAGTACCGGC 130

EuWDR2.P02cDNA TCGCCATCTCCTCCCTCCTCGAGCAGTACCCGAACCGCGTCGAGATCGTTC 198
 EuWDR2.Y05cDNA TCGCCATCTCCTCCCTCCTCGAGCAGTACCCGAACCGCGTCGAGATCGTTC 198
 NtWDR|KJ879944 TGGCGATCGCCAGCCTCCTCGAGCAGTACCCGAACCGGGTGGAGATCAATC 180

EuWDR2.P02cDNA CAGCTCGACGACTCCACGGGCGAGATCCGCTCCGACCCCGCCTCGCCTT 248
 EuWDR2.Y05cDNA CAGCTCGACGACTCCACGGGCGAGATCCGCTCCGACCCCGCCTCGCCTT 248
 NtWDR|KJ879944 CAGCTGGACGACTCCACGGGCGAGATCCGCTCCGACCCGGCCTCTCCTT 230

EuWDR2.P02cDNA CGAGCACCCCTACCCCCACCAAGTCCATGTTTCGTCCTCCGACCGCGACT 298
 EuWDR2.Y05cDNA CGAGCACCCCTACCCCCACCAAGTCCATGTTTCGTCCTCCGACCGCGACT 298
 NtWDR|KJ879944 CGAGCACCCCTATCCCCGACCAAGGCCATGTTTCGTCCTCCGACCGTGACT 280

EuWDR2.P02cDNA GCCTCCGCCCCGACCTCCTCGCCACCTCCGCCGACTTCCTCCGCAATCTGG 348
 EuWDR2.Y05cDNA GCCTCCGCCCCGACCTCCTCGCCACCTCCGCCGACTTCCTCCGCAATCTGG 348
 NtWDR|KJ879944 GCCTCCGCCCCGATCTCCTCGCCACCTCCGCCGACTTCCTCCGCAATATGG 330

EuWDR2.P02cDNA CGCATCAACGACGACGGCGTCCAGCTCCAATCCCTCCTCAACGGGAACAA 398
 EuWDR2.Y05cDNA CGCATCAACGACGACGGCGTCCAGCTCCAATCCCTCCTCAACGGGAACAA 398
 NtWDR|KJ879944 CTAGTCGGCCCGACGGGTTGAGCCCGCTCCCTCCTCAATGGCAACAA 380

EuWDR2.P02cDNA GACCTCCGAGTTCTGCGGCCGTTAACCTCCTTCGACTGGAACGAGAACG 448
 EuWDR2.Y05cDNA GACCTCCGAGTTCTGCGGCCGTTAACCTCCTTCGACTGGAACGAGAACG 448
 NtWDR|KJ879944 GAACTCTGAGTTCTGCGGCCCTCTCACTCCTTCGACTGGAACGAGGCCG 430

EuWDR2.P02cDNA AGCCC CGCCGATCGGCACCTCCTCCATCGACACCACCTGCACCGTCTGG 498
 EuWDR2.Y05cDNA AGCCC CGCCGATCGGCACCTCCTCCATCGACACCACCTGCACCGTCTGG 498
 NtWDR|KJ879944 ACCACGCCGGATCGGCACCTCCTCCATCGATACCACCTGCACCAATCTGG 480

EuWDR2.P02cDNA GATGTCGAGGCGGAGACCGTGGAACCCAGCTAATAGCGCACGATAAGGA 548
 EuWDR2.Y05cDNA GATGTCGAGGCGGAGACCGTGGAACCCAGCTAATAGCGCACGATAAGGA 548
 NtWDR|KJ879944 GACATCGAGAAAGGAGGTCGTGACACCCAGCTCATCGCCACGACAAGGA 530

EuWDR2.P02cDNA AGTCTACGACATCGCCTGGGGCGGTGTCGGCGTCTTCGCTCCCGTCTCCG 598
 EuWDR2.Y05cDNA AGTCTACGACATCGCCTGGGGCGGTGTCGGCGTCTTCGCTCCCGTCTCCG 598
 NtWDR|KJ879944 GGTCTACGACATCGCCTGGGGTGGCGTCGGCGTCTTCGCTCCCGTCTCCG 580

EuWDR2.P02cDNA GC GACGGCTCGGTGAGAGTCTTCGACCTGCGCGACAAAAGAGCACTCTACA 648
 EuWDR2.Y05cDNA GC GACGGCTCGGTGAGAGTCTTCGACCTGCGCGACAAAAGAGCACTCTACA 648
 NtWDR|KJ879944 CT GACGGCTCGGTTCGCGTCTTCGACCTTCGTCGACAAGGAGCACTCCACC 630

EuWDR2.P02cDNA ATCATCTATGAATCTAATGACCCGGCGGACACCCCGCTGGTGC GGCTGGG 698
 EuWDR2.Y05cDNA ATCATCTATGAATCTAATGACCCGGCGGACACCCCGCTGGTGC GGCTGGG 698
 NtWDR|KJ879944 ATCATCTACGAGTCTCTCGGACCCGGCTGACACTCCCTAGTTAGATGGG 680

Figure A3.8 (to be continued on the next page)

```

EuWDR2.P02cDNA GTGGAACAAGCAGGACCCGAGGTACATGGCGACTATAATCATGGACAGCG 748
EuWDR2.Y05cDNA GTGGAACAAGCAGGACCCGAGGTACATGGCGACTATAATCATGGACAGCG 748
NtWDR|KJ879944 CTGGAACAAGCAGGATCCGAGAATACATGGCCACCATCATAATGGACAGCG 730

EuWDR2.P02cDNA CCAAGGTGGTGTGTGGACATCCGCTTCCCGACGCTCCCGGTGGTGGAG 798
EuWDR2.Y05cDNA CCAAGGTGGTGGTGTGGACATCCGCTTCCCGACGCTCCCGGTGGTGGAG 798
NtWDR|KJ879944 CCAAGGTCTGTGGTGTGGACATTCGCTTCCCGACGCTGCCCGGTGGTTGAG 780

EuWDR2.P02cDNA CTGCAGCGGCACCAGGCCGGGGTGAATGCCATCGCGTGGGGCCCGCACAG 848
EuWDR2.Y05cDNA CTGCAGCGGCACCAGGCCGGGGTGAATGCCATCGCGTGGGGCCCGCACAG 848
NtWDR|KJ879944 CTGCAGAGGCACCAGGCCAGCGTCAACGCCATCGCGTGGGGCCCGCACAG 830

EuWDR2.P02cDNA CTCCTGCCATATCTGCACATGCAGGGGATGACTCGCAGGCGCTCATATGGG 898
EuWDR2.Y05cDNA CTCCTGCCATATCTGCACATGCAGGGGATGACTCGCAGGCGCTCATATGGG 898
NtWDR|KJ879944 CTCCTGCCACATCTGCACGGCCGGAGACGATTCGCAGGCGCTGATCTGGG 880

EuWDR2.P02cDNA ATCTGTCTTCTATGGGAGGAGGTGGGAATGGGCAGCAGGGGGCGGGGGGCT 948
EuWDR2.Y05cDNA ATCTGTCTTCTATGGGAGGAGGTGGGAATGGGCAGCAGGGGGCGGGGGGCT 948
NtWDR|KJ879944 ACCTTTCCTCCATGGGAGCGGCAGCGGTGGG--GGTGGGAGCGGGGACC- 927

EuWDR2.P02cDNA GCGCGGGCTGCTGCGGCGGATGGCGGGTGGGATCCGATACTGGCGTACAC 998
EuWDR2.Y05cDNA GCGCGGGCTGCTGCGGCGGATGGCGGGTGGGATCCGATACTGGCGTACAC 998
NtWDR|KJ879944 -CAGCA-----GGTTGAGGGGGGTCGGGATCCGATACTGGCGTACAC 968

EuWDR2.P02cDNA GGCTGGGGCGGAGATCGAGCAGCTGCAGTGGTCTCGACCGCAGCCGGACT 1048
EuWDR2.Y05cDNA GGCTGGGGCGGAGATCGAGCAGCTGCAGTGGTCTCGACCGCAGCCGGACT 1048
NtWDR|KJ879944 GGCCGGGGCGGAGATCGAGCAGCTGCAGTGGTCAATCTCGCAGCCAGATT 1018

EuWDR2.P02cDNA GGGTGGCGATTGCTTTCGCGAATAAGCTGCAGATACTCAGGGTTTGA 1095
EuWDR2.Y05cDNA GGGTGGCGATTGCTTTCGCGAATAAGCTGCAGATACTCAGGGTTTGA 1095
NtWDR|KJ879944 GGGTGGCCATTGCTTTCCTCCACCAAGTGCAGATTCTCAGGGTTTGA 1065

```

Figure A3.8: Sequence alignment of the *EuWDR2* coding region from one purple- and one yellow-anthered individual and *NtWDR* (from *Narcissus tazetta*; GenBank no. KJ879944), which is the BLAST top hit of *EuWDR1*. Identical sequences are highlighted in a black background. The synonymous substitutions between purple- and yellow-anthered individuals are highlighted in a green background. Dashes indicate alignment gaps.

R2 repeat

```

AtMYB75  -----MEGSSKG---LRKGAWTTEEDSLLRQCINKY 28
AtMYB90  -----MEGSSKG---LRKGAWTAEEDSLLRLCIDKY 28
AtMYB113 -----MGESPKG---LRKGTWTTEEDI LLRQCIDKY 28
AtMYB114 -----MEGSSKG---LRKGAWTAEEDSLLRQCIGKY 28
LhMYB12  MFQTFIAPANTGTTSPPTSAGSGGSPG---TRKQWSKEEDNLLRKCINQY 47
LhMYB6   -----MSPFRVSATSSSFQMSPPVLRLLVRKGAWTQVEDDLLKRCIERH 45
LrMYB15  ---MRKMPRTMSGKTSDSPTKSQLRTSVSVRKGAWTQAEDDLLRSCIEKH 47
TfMYB3   -----MSLLTTISSSSSSRLPPSSVLLRRGAWTQAEDDLLRRCIEKH 42
EuMYB3   -----MSVLTITASSTSSQPPPSLVILCKGAWTTTEDDLLRRCIEKH 42

```

R3 repeat

```

AtMYB75  GEGKWHQVPVRAGLNRCRKSCRLRWLNLYLKPSIKRGKLSSEDEVDLLLRLH 78
AtMYB90  GEGKWHQVPLRAGLNRCRKSCRLRWLNLYLKPSIKRGRLSNDEVDLLLRLH 78
AtMYB113 GEGKWHRVPLRTGLNRCRKSCRLRWLNLYLKPSIKRGKLCSEDEVDLVLR LH 78
AtMYB114 GEGKWHQVPLRAGLNRCRKSCRLRWLNLYLKPSIKRGKFSSDEVDLLLRLH 78
LhMYB12  NPVKWVSHVPKLAGLNRCRKSCRLRWVNYLDPSINRGSFSEDEEDLII RLH 97
LhMYB6   GVVRSRVPQLAGLNRCRKSCRLRWLNLYLDPRIIRRGQFEDEEDLII RLH 95
LrMYB15  GTVKWSNVPQLAGLNRCRKSCRLRWVNYLNPQIDRGTFEDEEDLII RLH 97
TfMYB3   GSLRWCHVARMAGLNRCRKSCRLRWLNLYLDPRLKRGIFEEDEKDLI VRLH 92
EuMYB3   GAVRWNRVPQLAGLNRCRKSCRLRWLNLYLDPRIKRGRFEEDEEDLII FRLH 92

```

```

AtMYB75  RLLGNRWSLIAGRLPGRTANDVKNYWNTHLSKKHE-PCCKIKMKKRDI TP 127
AtMYB90  KLLGNRWSLIAGRLPGRTANDVKNYWNTHLSKKHESSCCKSKMKKKNI IS 128
AtMYB113 KLLGNRWSLIAGRLPGRTANDVKNYWNTHLSKKHDERCCKTKMINKNITS 128
AtMYB114 KLLGNRWSLIAGRLPGRTANDVKNYWNTHLSKKHE-PCCKTKIKRINI IT 127
LhMYB12  KLLGNRWSLIAGRLPGRTANDIKNYWNSHLSKRV-----NVEQRTL 139
LhMYB6   KLLGNRWSLIAGRLPGRTANDVKNYWNSHLSKKLI--PQEKVVRACPCIA 143
LrMYB15  KLLGNRWSLIAGRLPGRTANDVKNHWNRSRLSKKLI--SGIKNDGSRGRVA 145
TfMYB3   KLLGNRWSLIAGRVPGRTANDVKNHWNRSRLNKKLV--TEARKYGERREVA 140
EuMYB3   KLLGNRWSLIAGRLPGRTANDVKNYWNSHLSKKLE--AGDKKYRGRRRVA 140

```

motif 6

```

AtMYB75  IPTTPALKNNVYKPRPRSFTVNNDCN-----HLN 156
AtMYB90  PPTTPVQKIGVFKPRPRSFSVNNGCS-----HLN 157
AtMYB113 HPTSSAQKIDVLKPRPRSFSDKNSCN-----DVN 157
AtMYB114 PP-----NTPAQKV-----136
LhMYB12  KP-----IRPQPVTLPNWS-----WLR 157
LhMYB6   AP-----TRPQPRKCSIKTTSVDDQQVNMSELIPQKKVRACRII 184
LrMYB15  AP-----IRPQPRTI PVRMQ-----160
TfMYB3   PP-----ITPQHQCFSRKQQ-----155
EuMYB3   TP-----IRPQPRRWSKEEK-----155

```

Figure A3.9 (to be continued on the next page)

AtMYB75	A P P K V D V N P P C L G L N I N N V C D N S I I Y N K D K K K D Q L V N - N L I D G D N M W L E K	205
AtMYB90	G L P E V D L I P S C L G L K K N N V C E N S I T C N K D D E K D D F V N - N L M N G D N M W L E N	206
AtMYB113	I L P K V D V V P L H L G L N N N Y V C E S S I T C N K D E Q K D K L I N I N L L D G D N M W W E S	207
AtMYB114	-----	136
LhMYB12	M K K Q G E A E P K M E T K V P D E E E H D Q W L M I N D S K H G - - - - - H E N Y Y T M N D	199
LhMYB6	A A P T R P Q P R K C S I E T K T S V D E Q Q V N M S E S R P S A D T A N C A V W Q D D L G N V K E	234
LrMYB15	-- K T G H V E R K H G E I Q P V S V V E E - - - - - D H H T S R M E N	189
TfMYB3	---- S S P E R L Q E D L N V T G V S A Y - - - - - Q Q D N T W V D R	182
EuMYB3	-- P I I D I E Q Q Q Q D L T L S G A S A Q Q - - - - - E D D A L W V E S	185
AtMYB75	F L E E S Q E - - - - - V D I L V P E A T T T E K G D T - L A F D V D Q L W S L F D G E T	244
AtMYB90	L L G E N Q E - - - - - A D A I V P E A T T A E H G A T - L A F D V E Q L W S L F D G E T	245
AtMYB113	L L E A - - - - - D V L G P E A T E T A K G V T - L P L D F E Q I W A R F D E E T	242
AtMYB114	----- D I F - - - - -	139
LhMYB12	Q M D S N Q Q - - - - - A D F G F E C I Y G V G E E E - T T V D A I L Q W D G L L S D I	237
LhMYB6	M I E Q L T E - - - - - A T I P S E N T E G F A H E G - L M Q D G V S L W D N F I F D I	272
LrMYB15	I I D D D E N Y N T K K T E R Q R E A D F S F F D N E G F R E D E W L M Q D G I S A W Q N L L S D L	239
TfMYB3	L L L Y N E E Y N K E K T E W Q H L S D F S L E D V E G F K E G - - M M L E G N L G L D T F L S D M	230
EuMYB3	L I R D D E N Y K N E N M N G R G E D N F N L E G M E G F T E - - - - - F W N N L I S D M	225
AtMYB75	V K F D - - - - -	248
AtMYB90	V E L D - - - - -	249
AtMYB113	L E L N - - - - -	246
AtMYB114	-----	139
LhMYB12	K L W S D S E V V	246
LhMYB6	Q L S S - - - - -	276
LrMYB15	L T G G - - - - -	243
TfMYB3	Q L W S - - - - -	234
EuMYB3	P L - - - - -	227

Figure A3.9: Alignment of anthocyanin-regulating R2R3-MYB proteins (subgroup 6, Strack et al. 2001). If > 70% of amino acids at each column are identical, this common amino acid is highlighted in a black background, and its similar amino acids are highlighted in a gray background. The domains of R2 and R3 repeats and the motif of subgroup 6 are indicated as bars above the alignment. The motif 6 was defined as “KPRPR[S/T]F” based only on the *Arabidopsis* sequences (Strack et al. 2001). However, our data show “[K/R]P[R/Q]PR” is a more generalized version of this motif. The bHLH interaction domain “[D/E]Lx₂[R/K]x₃Lx₆Lx₃R” (Zimmermann et al. 2004) located in the R3 repeat is indicated in yellow letters.

EuMYB3.P101A	ATGTCAGT	TCTCACAACA	ACTGCTTCTTCT	ACGTCTTCCC	AGCCTCCGCC	50								
EuMYB3.Y103A	ATGTCAGC	TCTCACAACA	ACTGCTTCTTCT	TTCGTCTTCCC	AGCCTCCGCC	50								
EuMYB3 885.2	ATGTCAGT	TCTCACAACA	ACTGCTTCTTCT	ACGTCTTCCC	AGCCTCCGCC	50								
EuMYB3.P101A	TTCATTGGTAA	TACTCT	GCAAGGGT	GCGTGGACT	TACAACAGAA	AGATGATC	100							
EuMYB3.Y103A	TTCATTGGTAC	TACTCTC	GCAAGGGT	GCGTGGACT	TACAACAGAA	AGATGATC	100							
EuMYB3 885.2	TTCATTGGTAA	TACTCT	GCAAGGGT	GCGTGGACT	TACAACAGAA	AGATGATC	100							
EuMYB3.P101A	TCCTCAGAA	GGTGCA	TAGAGA	AGCATGGT	GCAGTGC	GCGGTGGA	ACCGCGTT	150						
EuMYB3.Y103A	TCCTCAGAAA	AGTGCT	TAGAGC	AGCATGGT	GCAGTGC	GCGGTGGA	ACCGCGTT	150						
EuMYB3 885.2	TCCTCAGAA	GGTGCA	TAGAGA	AGCATGGT	GCAGTGC	GCGGTGGA	ACCGCGTT	150						
				Intron 1										
EuMYB3.P101A	CCTCAGTTAGCAG	-----				163								
EuMYB3.Y103A	CCTCAGTTAGCAG	GGTAATGC	ATCTCT	TTGGTCC	TAAGCTG	ACAAGAAA	ATG	200						
EuMYB3 885.2	CCTCAGTTAGCAG	GGTAATGC	ATCTAT	CGGTC	CTAAGCT	GACAGG	AAACTG	200						
EuMYB3.P101A	-----					163								
EuMYB3.Y103A	CATTTT	TA	-----	CATCCAG	CTGTGAC	ATTTT	TAATCT	TTTCC	238					
EuMYB3 885.2	CATTTCT	TTCTTT	ATATCT	TGCAG	CCAGAT	GAGAC	ATTTT	TACTCT	TTTCT	250				
EuMYB3.P101A	-GG	CTCAAT	AGATG	CCCGAAA	AGCTGT	CGACTG	AGATGG	CTCA	ACTAT	212				
EuMYB3.Y103A	GGC	CTCAAT	AGATG	CCCGAAA	AGCTGT	CGACTG	AGATGG	CTCA	ACTAT	288				
EuMYB3 885.2	GGG	CTCAAT	AGATG	CCCGAAA	AGCTGT	CGACTG	AGATGG	CTCA	ACTAT	300				
EuMYB3.P101A	CGACCC	ACGGAT	CAAG	CGAGGG	AGGTT	CGAAG	AGGAC	GAGG	ATGAT	262				
EuMYB3.Y103A	CGACCC	GCGGAT	TAAG	CGAGGG	AGGTT	CGAAG	AGGAC	GAGG	ATGAT	338				
EuMYB3 885.2	CGACCC	ACGGAT	CAAG	CGAGGG	AGGTT	CGAAG	AGGAC	GAGG	ATGAT	350				
				Intron 2										
EuMYB3.P101A	TCTTCAGG	CTTCATA	AAGCTCT	TTGGG	TAACAG	-----		293						
EuMYB3.Y103A	TCTTCAGG	CTTCATA	AAGCTCT	TTGGG	TAACAG	-----		369						
EuMYB3 885.2	TCTTCAGG	CTTCATA	AAGCTCT	TTGGG	TAACAG	GTGAT	CACCAT	GAATA	AAAT	400				
EuMYB3.P101A	-----					293								
EuMYB3.Y103A	-----					369								
EuMYB3 885.2	TGGAACA	AACCTC	CATTG	AAAAG	AGAGGA	ATCCT	TACCT	C	CAAGTT	TGCTT	450			
EuMYB3.P101A	-----					299								
EuMYB3.Y103A	-----					375								
EuMYB3 885.2	CACACT	TTTT	TAT	TCACT	GATG	TAAAT	CCT	TAC	CTC	TTTT	TTCAG	GTGGTC	500	
EuMYB3.P101A	GCTGAT	AGCGGG	CCGACT	TCCGGG	CAGGAC	AGCGA	ATGAC	GTGA	AGAACT	349				
EuMYB3.Y103A	GCTGAT	AGCGGG	CCGACT	TCCGGG	CAGGAC	AGCGA	ATGAC	GTGA	AGAACT	425				
EuMYB3 885.2	GCTGAT	AGCGGG	CCGACT	TCCGGG	CAGGAC	AGCGA	ATGAC	GTGA	AGAACT	550				
EuMYB3.P101A	ACTGGA	ACTCAC	ACTTA	AAGCA	AAGAACT	CGAAG	CCGG	GAGATA	AAAAAT	399				
EuMYB3.Y103A	ACTGGA	ACTCAC	ACTTA	AAGCA	AAGAACT	CGAAG	CCGG	GAGATA	AAAAAT	475				
EuMYB3 885.2	ACTGGA	ACTCAC	ACTTA	AAGCA	AAGAACT	CGAAG	CCGG	GAGATA	AAAAAT	600				
EuMYB3.P101A	CGAGGG	CGT	CGAAG	AGTT	TGC	GAC	CCGAT	CAG	ACC	ACA	ACCT	CGAAG	ATG	449
EuMYB3.Y103A	CGAGGG	CGT	CGAAG	AGTT	TGC	GAC	CCGAT	CAG	ACC	ACA	ACCT	CGAAG	ATG	525
EuMYB3 885.2	CGAGGG	CGT	CGAAG	AGTT	TGC	GAC	CCGAT	CAG	ACC	ACA	ACCT	CGAAG	ATG	650

Figure A3.10 (to be continued on the next page)

EuMYB3.P101A	GTC	AAAGAGGAAAAACCGAT	CATTGACATTGAGCAGCAACAACAGGACT	499
EuMYB3.Y103A	GTC	AAAGAGGAAAAACCGAG	CATTGACATTGAGCAGCAACAACAGGACT	575
EuMYB3 885.2	GTC	AAAGAGGAAAAACCGAT	CATTGACATTGAGCAGCAACAACAGGACT	700
EuMYB3.P101A	TGACGTTGAGTGGAGCAT	CAGCTCAACAAG	AGGATGATGCATTGTGGGTG	549
EuMYB3.Y103A	TGACGTTGAGTGGAGCA	TAA	GCTCAACAAGGATGATGCATTGTGGGTG	625
EuMYB3 885.2	TGACGTTGAGTGGAGCAT	CAGCTCAACAAG	AGGATGATGCATTGTGGGTG	750
EuMYB3.P101A	GAAAGTTTGATAC	GTGATGATGAAAACTATAAAAAACGAAAAATATGAACGG	599	
EuMYB3.Y103A	GAAAGTTTGATAA	GTGATGATGAAAACTATAAAAAACGAAAAATATGAACGG	675	
EuMYB3 885.2	GAAAGTTTGATAC	GTGATGATGAAAACTATAAAAAACGAAAAATATGAACGG	800	
EuMYB3.P101A	TCGAGGTGAAGACAACCTTCAATTTAGAGGG	TATGGAAGGATTTACAGAA	649	
EuMYB3.Y103A	TCGAGGTGAAGACAACCTTCAATTTAGAGGA	TATGGAAGGATTTACAGAAA	725	
EuMYB3 885.2	TCGAGGTGAAGACAACCTTCAATTTAGAGGG	TATGGAAGGATTTACAGAA	850	
EuMYB3.P101A	TTTGGAAATAATTTGATATCTGATATGCCACTTTAG		684	
EuMYB3.Y103A	TTTGGAAATAATTTGATATCTGATATGCCACTTTAG		760	
EuMYB3 885.2	TTTGGAAATAATTTGATATCTGATATGCCACTTTAG		885	

Figure A3.10: Sequence alignment of the full-length *EuMYB3* amplified with the anther cDNA samples of a purple- (P101) and a yellow-anthered (Y103) individuals and a gDNA sample. P101 is also the first purple-anthered individual shown in Figure 23, and Y103 is the third yellow-anthered individual. Identical sequences are highlighted in a black background. Introns 1 and 2 are indicated as bars above the alignment. The premature stop codon in the yellow copy is highlighted in a red background.

References

- Abdel-Hameed F, Snow R (1968) Cytogenetic studies in *Clarkia*, section *primigenia*. IV. A cytological survey of *Clarkia gracilis*. *American Journal of Botany* **55**: 1047-1054.
- Abdel-Hameed F, Snow R (1972) The origin of the allotetraploid *Clarkia gracilis*. *Evolution* **26**: 74-83.
- Adams KL, Wendel JF (2005) Polyploidy and genome evolution in plants. *Current Opinion in Plant Biology* **8**: 135-141.
- Ainouche M, Baumel A, Salmon A (2004) *Spartina anglica* Hubbard: a natural model system for analysing early evolutionary changes that affect allopolyploid genomes. *Biological Journal of the Linnean Society* **82**: 475-484.
- 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. *The Plant Journal* **65**: 771-784.
- Allen GA, Soltis DE, Soltis PS (2003) Phylogeny and biogeography of *Erythronium* (Liliaceae) inferred from chloroplast *matK* and nuclear rDNA ITS sequences. *Systematic Botany* **28**: 512-523.
- Andersen CL, Jensen JL, Ørntoft TF (2004) Normalization of real-time quantitative reverse transcription-PCR Data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer Research* **64**: 5245-5250.
- Anisimova M, Gascuel O (2006) Approximate likelihood-ratio test for branches: a fast, accurate, and powerful alternative. *Systematic Biology* **55**: 539-552.
- Austen EJ, Lin SY, Forrest JRK (2018) On the ecological significance of pollen color: a case study in American trout lily (*Erythronium americanum*). *Ecology* **99**: 926-937.
- Baudry A, Caboche M, Lepiniec L (2006) TT8 controls its own expression in a feedback regulation involving TTG1 and homologous MYB and bHLH factors, allowing a strong and cell-specific accumulation of flavonoids in *Arabidopsis thaliana*. *The Plant Journal* **46**: 768-779.

- Baudry A, Heim MA, Dubreucq B, Caboche M, Weisshaar B, Lepiniec L (2004) TT2, TT8, and TTG1 synergistically specify the expression of *BANYULS* and proanthocyanidin biosynthesis in *Arabidopsis thaliana*. *The Plant Journal* **39**: 366-380.
- Bernhardt P (1977) The pollination ecology of a population of *Erythronium americanum* Ker (Liliaceae). *Rhodora* **79**: 278-282.
- Blanc G, Wolfe KH (2004) Functional divergence of duplicated genes formed by polyploidy during *Arabidopsis* evolution. *The Plant Cell* **16**: 1679-1691.
- Bolger AM, Lohse M, Usadel B (2014). Trimmomatic: a flexible trimmer for Illumina Sequence Data. *Bioinformatics* **30**: 2114-2120.
- Borevitz JO, Xia Y, Blount J, Dixon RA, Lamb C (2000) Activation Tagging Identifies a Conserved MYB Regulator of Phenylpropanoid Biosynthesis. *The Plant Cell* **12**: 2383-2393.
- Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL (2009) BLAST+: architecture and applications. *BMC Bioinformatics* **10**: 421.
- Carey CC, Strahle JT, Selinger DA, Chandler VL (2004) Mutations in the *pale aleurone color1* regulatory gene of the *Zea mays* anthocyanin pathway have distinct phenotypes relative to the functionally similar *TRANSPARENT TESTA GLABRA1* gene in *Arabidopsis thaliana*. *The Plant Cell* **16**: 450-464.
- Carroll SB (2005) Evolution at two levels: on genes and form. *PLoS Biology* **3**: e245.
- Carroll SB (2008) Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* **134**: 25-36.
- Chang SM, Lu YQ, Rausher MD (2005) Neutral evolution of the nonbinding region of the anthocyanin regulatory gene *Ipmyb1* in *Ipomoea*. *Genetics* **170**: 1967-1978.
- Chau LM, Goodisman MAD (2017) Gene duplication and the evolution of phenotypic diversity in insect societies. *Evolution* **71**: 2871-2884.
- Chiou CY, Yeh KW (2008) Differential expression of MYB gene (*OgMYB1*) determines color patterning in floral tissue of *Oncidium* Gower Ramsey. *Plant Molecular Biology* **66**: 379-388.

- Clennett JCB, Chase MW, Forest F, Maurin O, Wilkin P (2012) Phylogenetic systematics of *Erythronium* (Liliaceae): morphological and molecular analyses. *Botanical Journal of the Linnean Society* **170**: 504-528.
- Clough SJ, Bent AF (1998) Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *The Plant Journal* **16**: 735-743.
- Coberly LC, Rausher MD (2003) Analysis of a chalcone synthase mutant in *Ipomoea purpurea* reveals a novel function for flavonoids: amelioration of heat stress. *Molecular Ecology* **12**: 1113-1124.
- 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.
- Conant GC, Wolfe KH (2008) Turning a hobby into a job: how duplicated genes find new functions. *Nature Reviews Genetics* **9**: 938-950
- Cone KC, Cocciolone SM, Moehlenkamp CA, Weber T, Drummond BJ, Tagliani LA, Bowen BA, Perrot GH (1993) Role of the regulatory gene *pl* in the photocontrol of maize anthocyanin pigmentation. *The Plant Cell* **5**: 1807-1816.
- Cooley AM, Modliszewski JL, Rommel ML, Willis JH (2011) Gene duplication in *Mimulus* underlies parallel floral evolution via independent trans-regulatory changes. *Current Biology* **21**: 700-704.
- Des Marais DL, Rausher MD (2010) Parallel evolution at multiple levels in the origin of hummingbird pollinated flowers in *Ipomoea*. *Evolution* **64**: 2044-2054.
- Dick CA, Buenrostro J, Butler T, Carlson ML, Kliebenstein DJ, Whittall JB (2011) Arctic mustard flower color polymorphism controlled by petal-specific downregulation at the threshold of the anthocyanin biosynthetic pathway. *PLoS ONE* **6**: e18230.
- Dooner HK (1983) Coordinate genetic regulation of flavonoid biosynthetic enzymes in maize. *Molecular and General Genetics* **189**: 136-141.
- Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure from small quantities of fresh leaf tissues. *Phytochemical Bulletin* **19**: 11-15.
- 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.

- Eckhart VM, Rushing NS, Hart GM, Hansen JD (2006) Frequency-dependent pollinator foraging in polymorphic *Clarkia xantiana* ssp. *xantiana* populations: implications for flower-colour evolution and pollinator interactions. *Oikos* **112**: 412-421.
- Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**: 1792-1797.
- Eisen KE, Geber MA (2018) Ecological sorting and character displacement contribute to the structure of communities of *Clarkia* species. *Journal of Evolutionary Biology* **31**: 1440-1458.
- Freeling M, Scanlon MJ, Fowler JE (2015) Fractionation and subfunctionalization following genome duplications: mechanisms that drive gene content and their consequences. *Current Opinion in Genetics & Development* **35**: 110-118.
- Frey FM (2004) Opposing natural selection from herbivores and pathogens may maintain floral-color polymorphism in *Claytonia virginica* (Portulacaceae). *Evolution* **58**: 2426-2437.
- 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.
- Gigord LDB, Macnair MR, Smithson A (2001) Negative frequency-dependent selection maintains a dramatic flower color polymorphism in the rewardless orchid *Dactylorhiza sambucina* (L.) Soo. *Proceedings of the National Academy of Sciences, USA* **98**: 6253-6255.
- Goff SA, Cone KC, Chandler VL (1992) Functional analysis of the transcription activator encoded by the maize B-gene: Evidence for a direct functional interaction between two classes of regulatory proteins. *Genes & Development* **6**: 864-875.
- Gompel N, Prud'homme B, Wittkopp PJ, Kassner VA, Carroll SB (2005) Chance caught on the wing: *cis*-regulatory evolution and the origin of pigment patterns in *Drosophila*. *Nature* **433**: 481-487.
- Gottlieb LD, Ford VS (1988) Genetic studies on the pattern of floral pigmentation in *Clarkia gracilis*. *Heredity* **60**: 237-246.
- Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit A, Adiconis X, Fan L, Raychowdhury R, Zeng Q et al. (2011) Trinity: reconstructing a full-length

- transcriptome without a genome from RNA-Seq data. *Nature Biotechnology* **29**: 644-652.
- Grotewold E (2006) The genetics and biochemistry of floral pigments. *Annual Review of Plant Biology* **57**: 761-780.
- Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology* **59**: 307-21.
- Habu Y, Hisatomi Y, Iida S (1998) Molecular characterization of the mutable *flaked* allele for flower variegation in the common morning glory. *The Plant Journal* **16**: 371-376.
- Harborne JB (1984) *Phytochemical methods: a guide to modern techniques of plant analysis, 2nd edn*. London, UK: Chapman and Hall.
- Hedrick PW (2013) Adaptive introgression in animals: examples and comparison to new mutation and standing variation as sources of adaptive variation. *Molecular Ecology* **22**: 4606-4618.
- Heim MA, Jakoby M, Werber M, Martin C, Weisshaar B, Bailey PC (2003) The basic helix-loop-helix transcription factor family in plants: a genome-wide study of protein structure and functional diversity. *Molecular Biology and Evolution* **20**: 735-747.
- Heuschen B, Gumbert A, Lunau K (2005) A generalised mimicry system involving angiosperm flower colour, pollen and bumblebees' innate colour preferences. *Plant Systematics and Evolution* **252**: 121-137.
- Himmelreich R, Hilbert H, Plagens H, Pirkl E, Li BC, Herrmann R (1996) Complete sequence analysis of the genome of the bacterium *Mycoplasma pneumoniae*. *Nucleic Acids Research* **24**: 4420-4449.
- Hoekstra HE, Coyne JA (2007) The locus of evolution: evo devo and the genetics of adaptation. *Evolution* **61**: 995-1016.
- Hoekstra HE, Hirschmann RJ, Bunday RA, Insel PA, Crossland JP (2006) A single amino acid mutation contributes to adaptive beach mouse color pattern. *Science* **313**: 101-104.

- Hollick JB, Patterson GI, Asmundsson IM, Chandler VL (2000) Paramutation alters regulatory control of the maize *pl* locus. *Genetics* **154**: 1827-1838.
- Holsinger KE, Thomson JD (1994) Pollen discounting in *Erythronium grandiflorum*: mass-action estimates from pollen transfer dynamics. *The American Naturalist* **144**: 799-812.
- Holton TA, Cornish EC (1995) Genetics and biochemistry of anthocyanin biosynthesis. *The Plant Cell* **7**: 1071-1083.
- Hopkins R, Rausher MD (2011) Identification of two genes causing reinforcement in the Texas wildflower *Phlox drummondii*. *Nature* **469**: 411-414.
- Hsu CC, Chen YY, Tsai WC, Chen WH, Chen HH (2015) Three R2R3-MYB transcription factors regulate distinct floral pigmentation patterning in *Phalaenopsis* spp.. *Plant Physiology* **168**: 175-191.
- Innan H, Kondrashov F (2010) The evolution of gene duplications: classifying and distinguishing between models. *Nature Reviews Genetics* **11**: 97-108.
- Irish VF, Litt A (2005) Flower development and evolution: gene duplication, diversification and redeployment. *Current Opinion in Genetics & Development* **15**: 454-460.
- Ison JL, Tuan ESL, Koski MH, Whalen JS, Galloway LF (2019) The role of pollinator preference in the maintenance of pollen colour variation. *Annals of Botany* **123**: 951-960.
- Jones KN (1996a) Fertility selection on a discrete floral polymorphism in *Clarkia* (Onagraceae). *Evolution* **50**: 71-79.
- Jones KN (1996b) Pollinator behavior and postpollination reproductive success in alternative floral phenotypes of *Clarkia gracilis* (Onagraceae). *International Journal of Plant Sciences* **157**: 733-738.
- Jones KN, Reithel JS (2001) Pollinator-mediated selection on a flower color polymorphism in experimental populations of *Antirrhinum* (Scrophulariaceae). *American Journal of Botany* **88**: 447-454.
- Kashkush K, Feldman M, Levy AA (2002) Gene loss, silencing, and activation in a newly synthesized wheat allotetraploid. *Genetics* **160**: 1651-1659.

- Kellis M, Birren BW, Lander ES (2004) Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature* **428**: 617-624.
- Klenk HP, Clayton RA, Tomb JF, White O, Nelson KE, Ketchum KA, Dodson RJ, Gwinn M, Hickey EK, Peterson JD, et al. (1997) The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon *Archaeoglobus fulgidus*. *Nature* **390**: 364-370.
- Koes RE, Verweij W, Quattrocchio F (2005) Flavonoids: a colorful model for the regulation and evolution of biochemical pathways. *Trends in Plant Science* **10**: 236-242.
- Koshikawa S, Giorgianni MW, Vaccaro K, Kassner VA, Yoder JH, Werner T, Carroll SB (2015) Gain of *cis*-regulatory activities underlies novel domains of *wingless* gene expression in *Drosophila*. *Proceedings of the National Academy of Sciences, USA* **112**: 7524-7529.
- Lai J, Ma J, Swigonova Z, Ramakrishna W, Linton E, Llaca V, Tanyolac B, Park YJ, Jeong OY, Bennetzen JL, et al. (2004) Gene loss and movement in the maize genome. *Genome Research* **14**: 1924-1931.
- Lai YS, Shimoyamada Y, Nakayama M, Yamagishi M (2012) Pigment accumulation and transcription of *LhMYB12* and anthocyanin biosynthesis genes during flower development in the Asiatic hybrid lily (*Lilium* spp.). *Plant Science* **193-194**: 136-147.
- Lang D, Weiche B, Timmerhaus G, Richardt S, Riano-Pachon DM, Correa LG, Reski R, Mueller-Roeber B, Rensing SA (2010) Genome-wide phylogenetic comparative analysis of plant transcriptional regulation: a timeline of loss, gain, expansion, and correlation with complexity. *Genome Biology and Evolution* **2**: 488-503.
- Lefort V, Longueville JE, Gascuel O (2017) SMS: Smart Model Selection in PhyML. *Molecular Biology and Evolution* **34**: 2422-2424.
- Leitch IJ, Beaulieu JM, Cheung K, Hanson L, Lysak MA, Fay MF (2007) Punctuated genome size evolution in Liliaceae. *Journal of Evolutionary Biology* **20**: 2296-2308.
- Li B, Dewey CN (2011) RSEM: accurate transcript quantification from RNA-seq data with or without a reference genome. *BMC Bioinformatics* **12**: 323.
- Li WH, Gu Z, Wang H, Nekrutenko A (2001) Evolutionary analyses of the human genome. *Nature* **409**: 847-849.

- Lunau K (1991) Innate flower recognition in bumblebees (*Bombus terrestris*, *B. lucorum*; Apidae): optical signals from stamens as landing reaction releasers. *Ethology* **88**: 203-214.
- Lunau K, Fieselmann G, Heuschen B, van de Loo A (2006) Visual targeting of components of floral colour patterns in flower-naive bumblebees (*Bombus terrestris*; Apidae). *Naturwissenschaften* **93**: 325-328.
- Mackay TFC, Stone EA, Ayroles JF (2009) The genetics of quantitative traits: challenges and prospects. *Nature Reviews Genetics* **10**: 565-577.
- Mallet J, Joron M (1999) Evolution of diversity in warning color and mimicry: polymorphisms, shifting Balance, and speciation. *Annual Review of Ecology and Systematics* **30**: 201-233.
- Martin A, Courtier-Orgogozo V (2017) Morphological evolution repeatedly caused by mutations in signaling ligand genes. In: Sekimura T, Nijhout HF eds. *Diversity and evolution of butterfly wing patterns: an integrative approach*. Singapore: Springer Nature, 59-87
- Nijhout HF (2017) The common developmental origin of eyespots and parafocal elements and a new model mechanism for color pattern formation. In: Sekimura T, Nijhout HF eds. *Diversity and evolution of butterfly wing patterns: an integrative approach*. Singapore: Springer Nature, 3-19.
- Martin A, Papa R, Nadeau NJ, Hill RI, Counterman BA, Halder G, Jiggins CD, Kronforst MR, Long AD, McMillan WO, et al. (2012) Diversification of complex butterfly wing patterns by repeated regulatory evolution of a *Wnt* ligand. *Proceedings of the National Academy of Sciences, USA* **109**: 12632-12637.
- Martin A, Reed R (2010) *wingless* and *aristaless2* define a developmental ground plan for moth and butterfly wing pattern evolution. *Molecular Biology and Evolution* **27**: 2864-2878.
- Martin A, Reed RD (2014) *Wnt* signaling underlies evolution and development of the butterfly wing pattern symmetry systems. *Developmental Biology* **395**: 367-378.
- Martins TR, Berg JJ, Blinka S, Rausher MD, Baum DA (2013) Precise spatio-temporal regulation of the anthocyanin biosynthetic pathway leads to petal spot formation in *Clarkia gracilis* (Onagraceae). *New Phytologist* **197**: 958-969.

- Martins TR, Jiang P, Rausher MD (2017) How petals change their spots: *cis*-regulatory re-wiring in *Clarkia* (Onagraceae). *New Phytologist* **216**: 510-518.
- Medel R, Botto-Mahan C, Kalin-Arroyo M (2003) Pollinator-mediated selection on the nectar guide phenotype in the Andean monkey flower, *Mimulus luteus*. *Ecology* **84**: 1721-1732.
- Mol J, Grotewold E, Koes R (1998) How genes paint flowers and seeds. *Trends in Plant Science* **3**: 212-217.
- Morita Y, Saitoh M, Hoshino A, Nitasaka E, Iida S (2006) Isolation of cDNAs for R2R3-MYB, bHLH, and WDR transcriptional regulators and identification of *c* and *ca* mutations conferring white flowers in the Japanese morning glory. *Plant and Cell Physiology* **47**: 457-470.
- Motten AF (1983) Reproduction of *Erythronium umbilicatum* (Liliaceae): pollination success and pollinator effectiveness. *Oecologia* **59**: 351-359.
- Myburg AA, Grattapaglia D, Tuskan GA, Hellsten U, Hayes RD, Grimwood J, Jenkins J, Lindquist E, Tice H, Bauer D, et al. (2014) The genome of *Eucalyptus grandis*. *Nature* **510**: 356-362.
- Nachman MW, Hoekstra HE, D'Agostino SL (2003) The genetic basis of adaptive melanism in pocket mice. *Proceedings of the National Academy of Sciences, USA* **100**: 5268-5273.
- Nakagawa T, Suzuki T, Murata S, Nakamura S, Hino T, Maeo K, Tabata R, Kawai T, Tanaka K, Niwa Y, et al. (2007) Improved Gateway binary vectors: high-performance vectors for creation of fusion constructs in transgenic analysis of plants. *Bioscience, Biotechnology, and Biochemistry* **71**: 2095-2100.
- Nicholls E, Hempel de Ibarra N (2014) Bees associate colour cues with differences in pollen rewards. *Journal of Experimental Biology* **217**: 2783-2788.
- Nijhout HF (2017) The common developmental origin of eyespots and parafocal elements and a new model mechanism for color pattern formation. In: Sekimura T, Nijhout HF eds. *Diversity and evolution of butterfly wing patterns: an integrative approach*. Singapore: Springer Nature, 3-19.
- Ohno S (1970) *Evolution by gene duplication*. New York, USA: Springer-Verlag.

- Ota M, Fukushima H, Kulski JK, Inoko H (2007) Single nucleotide polymorphism detection by polymerase chain reaction-restriction fragment length polymorphism. *Nature Protocols* **2**: 2857-2864.
- Parks CR, Hardin JW (1963) Yellow Erythroniums of the eastern United States. *Brittonia* **15**: 245-259.
- Paterson A, Bowers JE, Chapman BA (2004) Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. *Proceedings of the National Academy of Sciences, USA* **101**: 9903-9908.
- Peirson SN, Butler JN, Foster RG (2003) Experimental validation of novel and conventional approaches to quantitative real-time PCR data analysis. *Nucleic Acids Research* **31**: e73.
- Perkins JR, Dawes JM, McMahon SB, Bennett DLH, Orengo C, Kohl M (2012) ReadqPCR and NormqPCR: R packages for the reading, quality checking and normalisation of RT-qPCR quantification cycle (Cq) data. *BMC Genomics* **13**: 296.
- Peruzzi L, Leitch IJ, Caparelli KF (2009) Chromosome diversity and evolution in Liliaceae. *Annals of Botany* **103**: 459-475.
- Petroni K, Tonelli C (2011) Recent advances on the regulation of anthocyanin synthesis in reproductive organs. *Plant Science* **181**: 219-229.
- Pfaffl MW (2001) A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Research* **29**: e45.
- Pfaffl MW, Tichopad A, Prgomet C, Neuvians TP (2004) Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper – Excel-based tool using pair-wise correlations. *Biotechnology Letters* **26**: 509-515.
- Quattrocchio F, Baudry A, Lepiniec L, Grotewold E (2006) The regulation of flavonoid biosynthesis. In: Grotewold E, ed. *The science of flavonoids*. New York, USA: Springer, 97-122.
- Quattrocchio F, Wing JF, Leppen HTC, Mol JNM, Koes RE (1993) Regulatory genes controlling anthocyanin pigmentation are functionally conserved among plant species and have distinct sets of target genes. *The Plant Cell* **5**: 1497-1512.

- 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. *The 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 (2006) The evolution of flavonoids and their genes. In: Grotewold E, ed. *The science of flavonoids*. New York, USA: Springer, 175-211.
- Rausher MD, Delph LF (2015) Commentary: When does understanding phenotypic evolution require identification of the underlying genes? *Evolution* **69**: 1655-1664.
- Rausher MD, Fry JD (1993) Effects of a locus affecting floral pigmentation in *Ipomoea purpurea* on female fitness components. *Genetics* **134**: 1237-1247.
- Reed RD, Papa R, Martin A, Hines HM, Counterman BA, Pardo-Diaz C, Jiggins CD, Chamberlain NL, Kronforst MR, Chen R, et al. (2011) *optix* drives the repeated convergent evolution of butterfly wing pattern mimicry. *Science* **333**: 1137-1141.
- Rensing SA (2014) Gene duplication as a driver of plant morphogenetic evolution. *Current Opinion in Plant Biology* **17**: 43-48.
- Robertson A (1906) The 'droppers' of Tulipa and Erythronium. *Annals of Botany* **20**: 429-440.
- Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**: 139-140.
- Robinson MD, Oshlack A (2010) A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biology* **11**: R25.
- Rubin GM, Yandell MD, Wortman JR, Gabor Miklos GL, Nelson CR, Hariharan IK, Fortini ME, Li PW, Apweiler R, Fleischmann W, et al. (2000) Comparative genomics of the eukaryotes. *Science* **287**: 2204-2215.
- Sainz M, Grotewold E, Chandler VL (1997) Evidence for direct activation of an anthocyanin promoter by the maize C1 protein and comparison of DNA binding by related Myb domain proteins. *The Plant Cell* **9**: 611-625.

- Shang Y, Venail J, Mackay S, Bailey PC, Schwinn KE, Jameson PE, Martin CR, Davies KM (2011) The molecular basis for venation patterning of pigmentation and its effect on pollinator attraction in flowers of *Antirrhinum*. *New Phytologist* **189**: 602-615.
- Shapiro MD, Marks ME, Peichel CL, Blackman BK, Nereng KS, Jonsson B, Schluter D, Kingsley DM (2004) Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. *Nature* **428**: 717-723.
- Schemske DW, Bierzychudek (2001) Perspective: evolution of flower color in the desert annual *Linanthus parryae*: Wright revisited. *Evolution* **55**: 1269-1282.
- Schemske DW, Bierzychudek (2007) Spatial differentiation for flower color in the desert annual *Linanthus parryae*: was Wright right? *Evolution* **61**: 2528-2543.
- Schmitz JF, Zimmer F, Bornberg-Bauer E (2016) Mechanisms of transcription factor evolution in Metazoa. *Nucleic Acids Research* **44**: 6287-6297.
- 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*. *The Plant Cell* **18**: 831-851.
- Smith SD, Rausher MD (2011) Gene loss and parallel evolution contribute to species difference in flower color. *Molecular Biology and Evolution* **28**: 2799-2810.
- Soltis DE, Soltis PS, Pires JC, Kovarik A, Tate JA, Mavrodiev E (2004) Recent and recurrent polyploidy in *Tragopogon* (Asteraceae): cytogenetic, genomic and genetic comparisons. *Biological Journal of the Linnean Society* **82**: 485-501.
- Soltis P. 1986. Anthocyanidin variation in *Clarkia*. *Biochemical Systematics and Ecology* **14**: 487-489.
- Spelt C, Quattrocchio F, Mol JNM, Koes R (2000) *anthocyanin1* of *Petunia* encodes a basic helix-loop-helix protein that directly activates transcription of structural anthocyanin genes. *The Plant Cell* **12**: 1619-1631.
- Stern DL, Frankel N (2013) The structure and evolution of *cis*-regulatory regions: the *shavenbaby* story. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **368**: 20130028.

- Stern DL, Orgogozo V (2008) The loci of evolution: how predictable is genetic evolution? *Evolution* **62**: 2155-2177.
- Stoddard MC, Hauber ME (2017) Color, vision and coevolution in avian brood parasitism. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **372**: 20160339.
- 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 (2009a) Genetic changes contributing to the parallel evolution of red floral pigmentation among *Ipomoea* species. *New Phytologist* **183**: 751-763.
- Streisfeld MA, Rausher MD (2009b) Altered *trans*-regulatory control of gene expression in multiple anthocyanin genes contributes to adaptive flower color evolution in *Mimulus aurantiacus*. *Molecular Biology and Evolution* **26**: 433-444.
- Streisfeld MA, Rausher MD (2010) Population genetics, pleiotropy, and the preferential fixation of mutations during adaptive evolution. *Evolution* **65**: 629-642.
- 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 Genetics* **9**: e1003385.
- Subramaniam B, Rausher MD (2000) Balancing selection on a floral polymorphism. *Evolution* **54**: 691-695.
- Tamura K, Stecher G, Peterson D, Filipinski A, Kumar S (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution* **30**: 2725-2729.
- Tiffin P, Miller RE, Rausher MD (1998) Control of expression patterns of anthocyanin structural genes by two loci in the common morning glory. *Genes and Genetic Systems* **73**: 105-110.
- The *Arabidopsis* Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**: 796-815.
- Thomson JD (1986) Pollen transport and deposition by bumble bees in *Erythronium*: influences of floral nectar and bee grooming. *Journal of Ecology* **74**: 329-341.

- Thomson JD (2010) Flowering phenology, fruiting success and progressive deterioration of pollination in an early-flowering geophyte. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **365**: 3187-3199.
- Thomson JD, Price MV, Waser NM, Stratton DA (1986) Comparative studies of pollen and fluorescent dye transport by bumble bees visiting *Erythronium grandiflorum*. *Oecologia* **69**: 561-566.
- Thomson JD, Stratton DA (1985) Floral morphology and cross-pollination in *Erythronium grandiflorum* (Liliaceae). *American Journal of Botany* **72**: 433-437.
- Thomson JD, Thomson BA (1989) Dispersal of *Erythronium grandiflorum* pollen by bumblebees: implications for gene flow and reproductive success. *Evolution* **43**: 657-661.
- Throude M, Bolot S, Bosio M, Pont C, Sarda X, Quraishi UM, Bourgis F, Lessard P, Rogowsky P, Ghesquiere A, et al. (2009) Structure and expression analysis of rice paleo duplications. *Nucleic Acids Research* **37**: 1248-1259.
- Tomb JF, White O, Kerlavage AR, Clayton RA, Sutton GG, Fleischmann RD, Ketchum KA, Klenk HP, Gill S, Dougherty BA, et al. (1997) The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature* **388**: 539-547.
- Van de Peer Y, Mizrahi E, Marchal K (2017) The evolutionary significance of polyploidy. *Nature Reviews Genetics* **18**: 411-424.
- Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, Speleman F (2002) Accurate normalisation of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biology* **3**: research0034.1-0034.11.
- Velasco R, Zharkikh A, Affourtit J, Dhingra A, Cestaro A, Kalyanaraman A, Fontana P, Bhatnagar SK, Troglio M, Pruss D, et al. (2010) The genome of the domesticated apple (*Malus x domestica* Borkh.). *Nature Genetics* **42**: 833-839.
- Wallbank RWR, Baxter SW, Pardo-Diaz C, Hanly JJ, Martin SH, Mallet J, Dasmahapatra KK, Salazar C, Joron M, Nadeau N, et al. (2016) Evolutionary Novelty in a Butterfly Wing Pattern through Enhancer Shuffling. *PLoS Biology* **14**: e1002353.
- Waser NM, Price MV (1983) Pollinator behaviour and natural selection for flower colour in *Delphinium nelsonii*. *Nature* **302**: 422-424.

- 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 (2014a) Predictability and irreversibility of genetic changes associated with flower color evolution in *Penstemon barbatus*. *Evolution* **68**: 1058-1070.
- Wessinger CA, Rausher MD (2014b) Ecological transition predictably associated with gene degeneration. *Molecular Biology and Evolution* **32**: 347-354.
- Wittkopp PJ, Vaccaro K, Carroll SB (2002) Evolution of *yellow* gene regulation and pigmentation in *Drosophila*. *Current Biology* **12**: 1547-1556.
- Whibley AC, Langlade NB, Andalo C, Hanna AI, Bangham A, Thébaud C, Coen E (2006) Evolutionary paths underlying flower color variation in *Antirrhinum*. *Science* **313**: 963-966.
- Whittall JB, Voelckel C, Hodges SA (2006) Convergence, constraint and the role of gene expression during adaptive radiation: floral anthocyanins in *Aquilegia*. *Molecular Ecology* **15**: 4645-4657.
- Wray GA (2007) The evolutionary significance of *cis*-regulatory mutations. *Nature Reviews Genetics* **8**: 206-216.
- Wu CA, Streisfeld MA, Nutter LI, Cross KA (2013) The genetic basis of a rare flower color polymorphism in *Mimulus lewisii* provides insight into the repeatability of evolution. *PLoS ONE* **8**: e81173.
- Xu W, Dubos C, Lepiniec L (2015) Transcriptional control of flavonoid biosynthesis by MYB-bHLH-WDR complexes. *Trends in Plant Science* **20**: 176-185.
- Yamagishi M (2018) Involvement of a LhMYB18 transcription factor in large anthocyanin spot formation on the flower tepals of the Asiatic hybrid lily (*Lilium* spp.) cultivar "Grand Cru". *Molecular Breeding* **38**:60.
- Yamagishi M, Shimoyamada Y, Nakatsuka T, Masuda K (2010) Two R2R3-MYB genes, homologs of petunia *AN2*, regulate anthocyanin biosyntheses in flower tepals, tepal spots and leaves in Asiatic hybrid lily. *Plant and Cell Physiology* **51**: 463-474.
- Yamagishi M, Toda S, Tasaki K (2014) The novel allele of the *LhMYB12* gene is involved in splatter-type spot formation on the flower tepals of Asiatic hybrid lilies (*Lilium* spp.). *New Phytologist* **201**: 1009-1020.

- Yamagishi M, Uchiyama H, Handa T (2018) Floral pigmentation pattern in Oriental hybrid lily (*Lilium* spp.) cultivar 'Dizzy' is caused by transcriptional regulation of anthocyanin biosynthesis genes. *Journal of Plant Physiology* **228**: 85-91.
- Yokoi Y (1976) Growth and reproduction in higher plants II. Analytical study of growth and reproduction of *Erythronium japonicum*. *The botanical magazine = Shokubutsu-gaku-zasshi* **89**: 15-31.
- Yuan Y, 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.
- Yuan YW, Sagawa JM, Frost L, Vela JP, Bradshaw HD (2014) Transcriptional control of floral anthocyanin pigmentation in monkeyflowers (*Mimulus*). *New Phytologist* **204**: 1013-1027.
- Zhang J (2003) Evolution by gene duplication: an update. *Trends in Ecology & Evolution* **18**: 292-298.
- Zimmermann IM, Heim MA, Weisshaar B, Uhrig JF (2004) Comprehensive identification of *Arabidopsis thaliana* MYB transcription factors interacting with R/B-like BHLH proteins. *The Plant Journal* **40**: 22-34.

Biography

Rong-Chien Lin attended Tunghai University (Taichung, Taiwan) and received a Bachelor of Science degree in Biology in June 2000. She earned a Master of Science degree from Department of Life Science, National Taiwan Normal University (Taipei, Taiwan) in January 2007. The title of her master thesis is “Genetic evidences of post last glacial expansion and low population differentiation for a fig-pollinating wasp of *Ficus septica* in Taiwan”. She first-authored three papers based on her master thesis. From 2007 – 2012, she worked at National Taiwan Normal University in the lab of Dr. Shou-Hsien Li, working on several projects on evolutionary genetics of birds in Taiwan. She started her PhD research at Duke University in 2012, and expects to complete her PhD in May 2020.

While pursuing her doctorate, Rong-Chien received the following fellowships: Sigma Xi Grant-in-Aid of Research, Duke Biology Grant-in-Aid of Research, Duke Biology One-semester Fellowship, Ray J. Tysor Graduate Summer Research Fellowship and Duke Graduate School Conference Travel Fellowships. She has also been the recipient of the Taiwanese Government Scholarship for studying abroad.

First-authored publications

Lin RC, Yeung CKL, Fong JJ, Tzeng HY, Li SH (2011) The lack of pollinator specificity in a dioecious fig tree: sympatric fig-pollinating wasps of *Ficus septica* in southern Taiwan. *Biotropica* **43**: 200-207.

- Lin RC, Huang CC, Li SH, Yao CT (2009)** Isolation and characterization of 15 tetranucleotide microsatellite loci from the gray-cheeked fulvetta (*Alcippe morrisonia morrisonia*). Available from Molecular Ecology Resources Database (<http://tomato.biol.trinity.edu/manuscripts/9-6/mer-09-0109.pdf>). Summarized in *Molecular Ecology Resources* (2009) **9**: 1460-1559.
- Lin RC, Yeung CKL, Li SH (2008)** Drastic post-LGM expansion and lack of historical genetic structure of a subtropical fig-pollinating wasp (*Ceratosolen* sp. 1) of *Ficus septica* in Taiwan. *Molecular Ecology* **17**: 5008-5022.
- Lin RC, Huang CC, Li SH (2007)** Development and characterization of microsatellite markers for a fig-pollinating wasp (*Ceratosolen* sp. 1) of *Ficus septica*. *Molecular Ecology Notes* **7**: 1360-1362.
- Lin RC, Yao CT, Lo WS, Li SH (2007)** Characterization and the broad cross-species applicability of 20 anonymous nuclear loci isolated from the Taiwan Hwamei (*Garrulax taewanus*). *Molecular Ecology Notes* **7**: 156-159.