

## CHAPTER 7

# THE EVOLUTION OF FLAVONOIDS AND THEIR GENES

MARK D. RAUSHER

*Department of Biology, Box 90338, Duke University, Durham, NC 27708 USA*

### 1. INTRODUCTION

Flavonoids constitute a diverse array of plant secondary compounds that perform a wide variety of physiological and ecological functions (Figure 7.1). The role of anthocyanin pigments as visual signals in angiosperms for attracting pollinators and fruit dispersal agents is well known, but these functions were acquired late in the evolutionary diversification of flavonoids. Less well-known and probably more ancient functions of flavonoids include protection against the detrimental effects of UV radiation; mediation of interactions between pollen and stigma; defense against bacteria, pathogenic fungi, and herbivores; mediation of interactions between plants and mutualistic mycorrhizal fungi; and regulators of hormonal activity. Because recent, thorough reviews of these functions are available elsewhere (Koes et al., 1994; Shirley, 1996), they will not be discussed in detail here. Instead, I focus on the evolutionary processes by which these functions arose and continue to evolve.

As other chapters in this volume make clear, the flavonoid pathway has served as a model system for understanding gene regulation in plants. In a similar way, flavonoid pathway genes, both structural and regulatory, have served as a model system for understanding a variety of evolutionary processes, including the role of gene duplication in facilitating the evolution of novel characters, the causes of evolutionary rate variation among genes, and the relative importance of structural and regulatory genes in evolution of ecologically important characters. In this chapter, I review how examination of patterns of change in flavonoid genes has contributed to our understanding of these and other evolutionary issues, beginning with an examination of our understanding of the historical evolution of the flavonoid biosynthetic pathway.

## 2. HISTORICAL EVOLUTION OF THE FLAVONOID PATHWAY

The flavonoid pathway is a classic example of a pathway that has evolved piecemeal, gradually lengthening as new products and new functions were added. Although many details of its evolutionary construction are still hazy, it is possible to reconstruct the main evolutionary events with confidence. The following account is an elaboration of the reconstruction originally suggested by Stafford (1991). As will be seen, a general theme emerges from this reconstruction: gene duplication has provided the raw material for building the pathway. There is a satisfying symmetry to this in that, as we shall see below, once the flavonoid pathway was essentially complete, duplication and divergence of genes from the pathway has led to the emergence of new classes of compounds. Through gene duplication, pathways are not only born themselves, but they spawn additional pathways.

Equally importantly, the evolution of the flavonoid pathway provides an excellent example of how a biochemically complex trait may be built up in stages, with each addition being adaptive. It thus provides an illuminating counterexample to the claim of the “Intelligent Design” movement that complex biochemical adaptations that are irreducibly complex cannot be assembled gradually by natural selection. In this case, the complex biochemical adaptation is flower color produced by anthocyanin pigments. Its complexity lies in the fact that anthocyanin production requires at least six sequential biochemical reactions enabled by six different enzymes. Looking at this character by itself, there is no question that it is irreducibly complex. As numerous knockout mutations in the pathway demonstrate, removal of one enzyme eliminates the production of anthocyanins. It is only because the intermediate products that were invented along the way—the diverse collection of flavonoids produced by land plants—retain their important ecological and physiological functions, and so have not been superseded in these functions by other secondary metabolites, that we are able to recognize that the irreducible complexity of floral pigment production actually evolved gradually.

When considering the evolutionary construction of the flavonoid biochemical pathway, it should be borne in mind that early flavonoid production may have been quite “leaky.” Initially, enzymes of primary metabolism already may have produced small quantities of many different simple flavonoids, simply due to lack of complete substrate specificity in those enzymes. In this context, if an environmental change caused the production of one or a subset of these compounds to become advantageous, natural selection would act to enhance its production. At first, the mechanism of enhancement may have been quite crude (e.g., up-regulating a key enzyme of primary metabolism) and probably would have had deleterious pleiotropic effects. If, however, the advantages of increasing the production of flavonoids were large enough, this would evolve despite the pleiotropic effects. Moreover, it would automatically establish selection pressures to lessen the pleiotropic effects. Two particularly effective types of genetic change that could respond to this selection would be specialization of a duplicate gene for carrying out early flavonoid synthesis and the evolution of regulatory control over such a gene.

### 2.1. Evolution of the flavonoid enzymes

A major clue to the stepwise development of the flavonoid pathway is provided by the distribution of both different types of flavonoids and different flavonoid enzymes among land plant taxa (Figure 7.2). This distribution allows us to recognize a series of successive stages in the evolution of the pathway.

#### 2.1.1. Stage 1: Algae

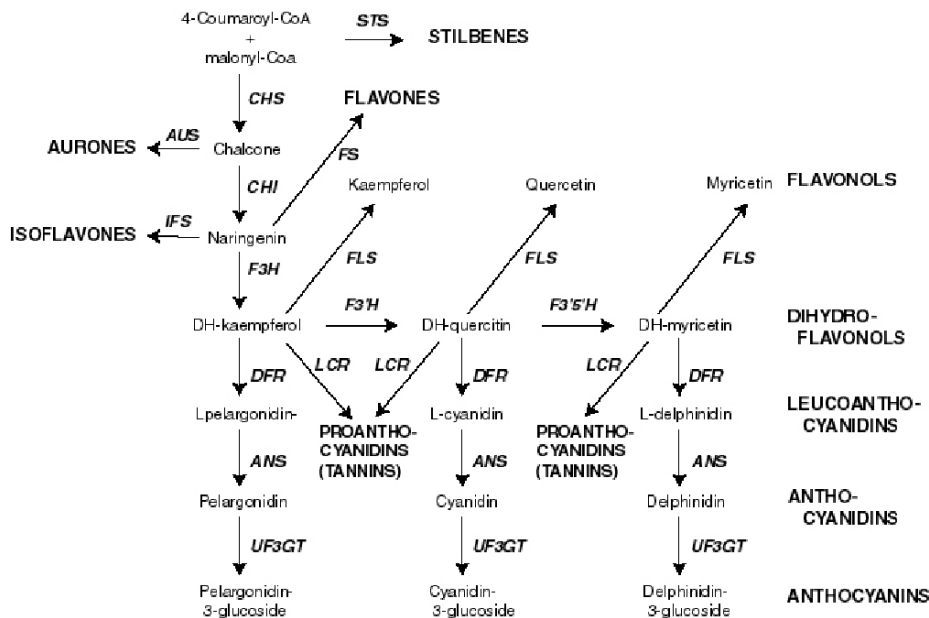
Land plants are believed to be derived from organisms like green algae (Charales). Although extensive surveys have been conducted (Markham, 1988), flavonoids consistently fail to be found among the algae. Moreover, although several algal genomes have been sequenced, none contain open reading frames that show homology to the coding sequences of any known flavonoid enzymes. The initial evolution of the flavonoid pathway thus probably took place after the colonization of land.

#### 2.1.2. Stage 2: Bryophytes (mosses), liverworts, and hornworts

This paraphyletic group represents the earliest plants to colonize land. It also is the oldest plant group to produce flavonoids. Among the types of flavonoids produced by these plants are chalcones, flavonols, and flavones (Markham, 1988), which are derived from the first three enzymes of the flavonoid pathway (Figure 7.1). In addition, EST (expressed sequence tags) databases report sequences that appear to represent these three enzymes: chalcone synthase, chalcone–flavanone isomerase, and flavanone 3-hydroxylase. All three enzymes appear to have been derived via gene duplication from genes coding for enzymes of primary metabolism. CHS exhibits strong sequence similarity to bacterial genes coding polyketide synthases, particularly those involved in fatty acid synthesis (Verwoert et al., 1992). Moreover, the condensation reactions of these enzymes also are similar, as is the set of substrates used. Based on similarity in sequence and enzymology, F3H belongs to the oxoglutarate-dependent dioxygenase family of enzymes and is presumably derived from a duplication of one of its members (Winkel-Shirley, 2001). The origin of CHI is less clear, since it seems to be unrelated in sequence and tertiary structure to any other plant enzyme (Jez et al., 2000). However, enzymes with similar sequence and secondary structures have been reported from some bacteria and fungi (Gensheimer and Mushegian, 2004), and an enzyme of unknown sequence but exhibiting CHI activity has been isolated from the bacterium *Eubacterium ramulus* (Herles et al., 2004). Interestingly, although this enzyme is capable of converting naringenin chalcone into naringenin, it is believed that its function *in vivo* is the catabolism of naringenin.

Two hypotheses have been put forward regarding the initial advantages associated with producing flavonoids. One suggests that flavonoids evolved as an effective sunscreen protecting against UV radiation as plants began colonizing land

(Markham, 1988; Koes et al. 1994; Shirley 1996). Supporting this suggestion are the observations that even simple flavonoids such as chalcones, aurones, and flavanones absorb UV wavelengths strongly, and that flavonoid knockout mutants often are extremely susceptible to UV damage (Li et al., 1993; Lois and Buchanan, 1994). By contrast, Stafford (1991) has argued that the first function of flavonoids was to regulate or chaperone plant hormones. Stafford argues that this hypothesis is more likely than the UV protection hypothesis because presumably early flavonoid enzymes were not as efficient as current enzymes, and therefore large quantities of flavonoids likely did not accumulate early. Moreover, it has been shown in angiosperms that flavonoids contribute to regulation of auxin transport (Brown et al., 2001; Peer et al., 2004).



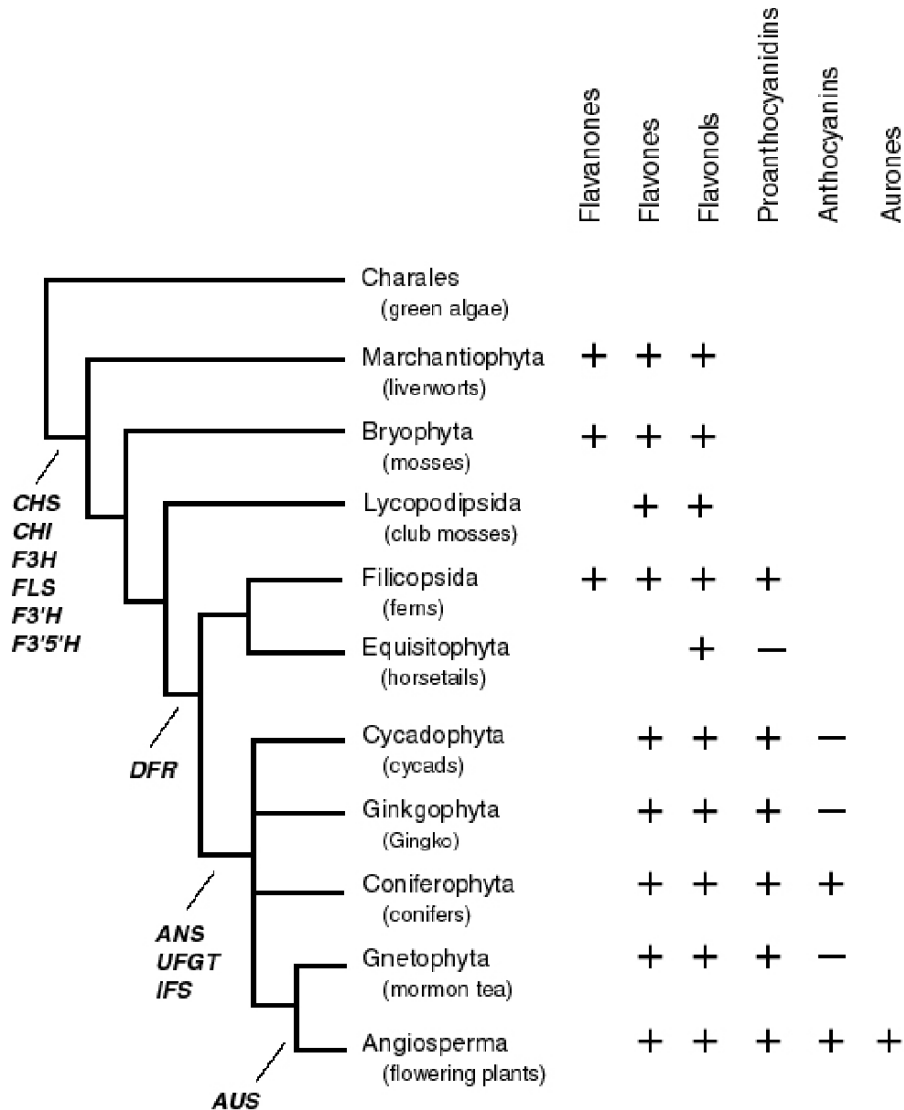
**Figure 7.1** Schematic portrayal of the flavonoid pathway. Specific compounds are listed in lower case. Classes of compounds are listed in bold upper case. Enzymes are listed in bold italics. STS: stilbene synthase. CHS: chalcone synthase. CHI: chalcone-flavanone isomerase. F3H: flavanone-3-hydroxylase. DFR: dihydroflavonols 4-reductase. ANS: anthocyanidin synthase. UF3GT: UDP flavonoid glucosyltransferase. F3'5'H: flavonoid 3'5'hydroxylase. LCR: leucoanthocyanidins reductase. FS: flavone synthase. FLS: flavonol synthase. IFS: isoflavone synthase. AUS: aureusidin synthase. DH="dihydro." L="leuco."

These three enzymes and the flavonoids produced by them are unlikely to have arisen and evolved their new functions simultaneously. A likely, though currently

unproven, sequence of origin is that CHS evolved first, followed by F3H, and then CHI. CHS seems a likely first enzyme because it is the initial enzyme in the pathway. Perhaps equally important, though, is the fact that naringenin chalcone, the product of CHS, spontaneously isomerizes to the flavanone naringenin even in the absence of CHI (Holton and Cornish, 1995). This “leakiness” means that it is likely that early plants with only CHS function could produce a variety of flavonoids that could collectively provide UV protection. The evolutionary incorporation of CHI into the nascent pathway could then have been favored, not because it provided a new function, but because it improved the efficiency of an existing function. The incorporation of F3H into the pathway could have either proceeded or followed CHI, allowing the production of dihydroflavonols and perhaps flavonols. While we can probably never be certain what advantage this conferred, one likely possibility is that these compounds evolved as defenses against pathogenic fungi and bacteria, a function that has been demonstrated in contemporary plants (Kemp and Burden, 1986; Curir et al., 2005).

In addition to the evolution of the CHS-CHI-F3H backbone of the rudimentary flavonoid pathways, this stage also saw the evolutionary development of the three major branches of that pathway (Figure 7.1). These branches correspond to different degrees of hydroxylation of the flavonoid B ring. The first branch produces compounds lacking hydroxylation at either the 3' or 5' position, such as dihydrokaempferol and kaempferol. The second branch produces compounds hydroxylated at the 3' position (dihydroquercetin and quercetin), while compounds produced by the third branch (dihydromyricetin and myricetin) are hydroxylated at both the 3' and 5' positions. Evidence for this claim comes from extensive surveys of flavonoid compounds produced by mosses and liverworts (e.g., Markham, 1988), which reveal the species that produce derivatives of all three branches. Moreover, EST databases from the moss *Physcomitrella patens* contain sequences that are very F3'H like, which in higher plants hydroxylates the 3' position. Sequences similar to F3'5'H, which in higher plants hydroxylates the 5' position, have not yet been reported from mosses or liverworts. Both of these enzymes have apparently been recruited to the pathway through duplication of genes in the cytochrome P450 hydroxylase family (Dixon and Steele, 1999). Although one can only speculate as to the selective advantages associated with addition of branches to the pathway, differential effects of hydroxylated vs. nonhydroxylated flavonoids are known. For example, quercetin seems to be a more effective photoprotectant than kaempferol because it is a more effective antioxidant (Ryan et al., 2001, 2002).

Another enzymatic innovation that appears to have occurred at this stage is the evolution of flavonol synthase. All orders of land plants, including bryophytes, produce flavonols, which are produced from dihydroflavonols by one or more flavonol synthase (FLS) enzymes (Figures 7.1 and 7.2). *FLS* is derived from the 2-oxoglutarate-dependent dioxygenase gene family (Holton et al., 1993). While the presence of flavonols in mosses, liverworts, or ferns indicates that these groups must possess *FLS* genes, they have not yet been detected and characterized, almost certainly because of the little attention that has been paid to the genomes of these groups.



**Figure 7.2** Land plant phylogeny showing the time of origin of flavonoid enzymes. +, Documented presence of flavonoid; —, possible evolutionary loss of flavonoid. Phylogeny from Savolainen and Chase (2003).

In contrast to flavonols, isoflavonoids occur only sporadically throughout the land plants (Dewick, 1988). Isoflavonoids have been reported in the moss *Bryum*

*capillare* (Anhut et al., 1984), suggesting that the key enzyme for producing these compounds, isoflavone synthase (IFS) may have originated at this stage of plant evolution. However, the isoflavonoids have not been reported from any other plant outside of the gymnosperms and angiosperms (Dewick, 1988). This pattern suggests that the production of isoflavonoids may have evolved independently in *Bryum* and in the seed plants. Moreover, until isoflavonoids are detected in other bryophytes, the acquisition of the ability of *Bryum* to synthesize these compounds may represent a relatively recent evolutionary event. Since *IFS* has not yet been cloned from *Bryum*, it is not currently possible to evaluate this hypothesis.

#### 2.1.3. Stage 3: *Filicophyta* (ferns)

The ferns and allies (lycopsids and equisetopsids) are believed to be derived from bryophytelike ancestors. They are the oldest group of plants known to produce proanthocyanidins (also called leucoanthocyanidins). Proanthocyanidins are commonly found in polymerized form, molecules known as tannins. In higher plants, proanthocyanidins are produced by the enzyme dihydroflavonol-4-reductase (DFR), which uses dihydroflavonols as substrates. Although this enzyme has not been reported in ferns, the genomic characterization of ferns is in its infancy. It would be very surprising if DFR were not eventually found in these plants.

One of the primary functions of tannins in plants is believed to be defense against bacterial and fungal pathogens, as well as herbivores (Feeny, 1970). This is likely to be one of the primary advantages that drove the evolution of the capability to produce proanthocyanidins in the lineage leading to ferns. As with other enzymes, DFR seems to have been derived from a duplicated gene associated with primary metabolism, in this case NADPH-dependent reductases associated with steroid metabolism (Baker and Blasco, 1992). Additionally, the three branches of the flavonoid pathway established in the bryophytes are maintained in the ferns, with procyanidin and prodelphinidin, as well as the flavonols kaempferol, quercetin, and myricetin being produced by species in many different fern families (Markham, 1988).

#### 2.1.4. Stage 4: *Gymnosperms and angiosperms*

It is within these two groups of land plants that anthocyanins finally made their appearance (Timberlake and Bridle, 1980; Niemann, 1988). The key addition to the flavonoid pathway involved in this step was the recruitment of the enzyme anthocyanidin synthase (ANS), presumably as a result of gene duplication from the family of 2-oxo-glutarate-dependent oxygenases. This enzyme catalyzes the production of colored anthocyanins from colorless leucoanthocyanidins. It is not clear, however, whether it was the production of color per se that constituted the original function of anthocyanins. Although color signaling, especially to pollinators and fruit dispersal agents, is clearly a primary function of anthocyanins in angiosperms, such signaling is rare if not absent in gymnosperms from which the angiosperms arose.

In gymnosperms and angiosperms, anthocyanins are typically sequestered in vacuoles (see Chapter 5 for a detailed description of the vacuolar transport of flavonoids). This requires two additional enzymes, one for the addition of sugar moieties to the anthocyanidin skeleton to make the compound more soluble and one to transfer the resulting anthocyanin across the vacuolar membrane. In angiosperms, these two functions are accomplished, respectively, by enzymes such as UF3GT, which glycosylates anthocyanidins, and glutathione-*S*-transferase (GST). The former are derived from the large family of sugar transferases (e.g., UDP glucosyltransferases) (Mackenzie et al., 1997). The latter has apparently been derived independently at least twice in angiosperms from the large glutathione transferase family (Alfenito et al., 1998).

As in the case of the first three flavonoid enzymes, it is not likely that ANS, UFGT, and GST were recruited simultaneously. Once anthocyanidins began to be produced, it is likely that nonspecific glycosyltransferases provided sufficient activity to convert them to anthocyanins. Similarly, nonspecific GSTs may have initially allowed some degree of vacuolar sequestration. Indeed, experimental transformation of anthocyanin-GST deficient maize kernels with GSTs not associated with anthocyanin conjugation produced vacuolar accumulation of anthocyanins (Alfenito et al., 1998). The efficiency of these functions then could have been improved through gene duplication and specialization on anthocyanin substrates.

Two other classes of flavonoids may have made their appearance in the seed plants. Except for the moss *Bryum*, isoflavonoids appear to be restricted to gymnosperms and angiosperms (Dewick, 1988), suggesting two separate origins of these compounds and the enzymes that produce them (see above). While isoflavonoids are widely distributed throughout the seed plants, their occurrence is sporadic (Figure 7.3). This taxonomic distribution may represent either multiple origins of isoflavone synthase and related enzymes or a single origin with multiple secondary losses of the ability to produce these compounds. These hypotheses can be distinguished in principle by constructing a phylogeny of *IFS* and related cytochrome P450 hydroxylases, as has been done for the origin of stilbene synthases from chalcone synthases (see below). To date, however, only one *IFS* gene from a plant outside the Fabaceae (legumes) has been characterized definitively, preventing at present a phylogenetic reconstruction.

Aurones are the second class of compounds that appear to have originated at this stage. They have never been found in ferns, fern allies (horsetails, lycopods, whisk ferns), or in gymnosperms (Bohm, 1975; Markham, 1988). By contrast, they have been found in at least seven orders of eudicots, where they typically occur as yellow pigments in flowers and in sedges (Bohm, 1975). They also have been reported from one species of bryophyte and one genus of liverworts (Markham, 1988), though it is unclear whether these represent mistaken chemical identification or independent origins. The gene coding the enzyme aureusidin synthase (*AUS*), which catalyzes the conversion of naringenin chalcone into the aurone aureusidin, recently has been cloned from *Antirrhinum* (Nakayama et al., 2000, 2001) and is derived by duplication from a plant polyphenol oxidase gene. As with *IFS*, the hypothesis of independent origins in angiosperms and bryophytes cannot be tested at this point because *AUS* has not been cloned from any other species.

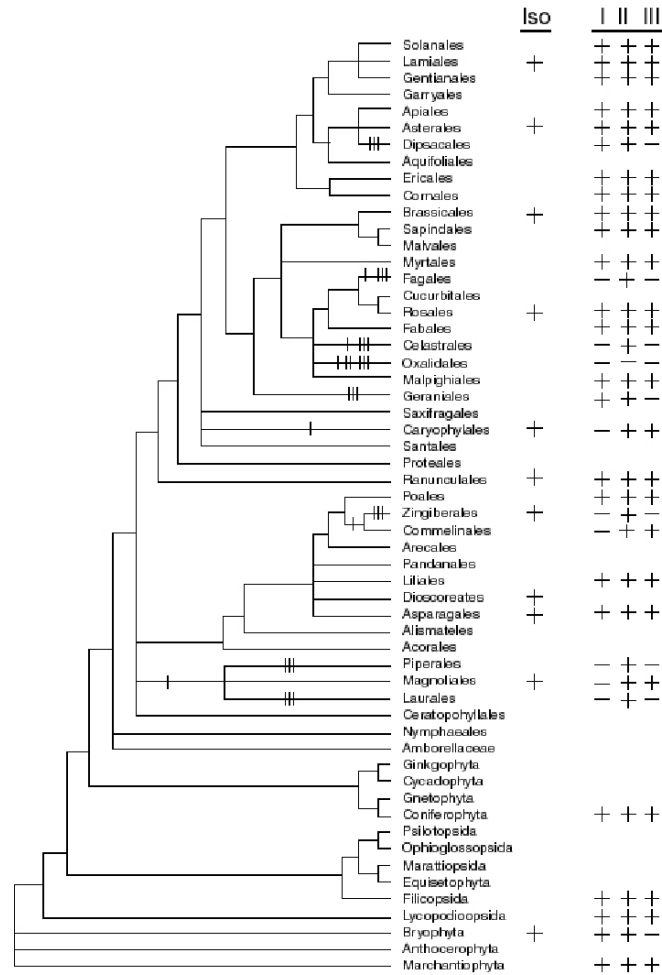
## 2.2. Historic evolution of flavonoid gene regulation

Evolution of regulatory control of the flavonoid structural genes is poorly understood, but there are some intriguing patterns. In most dicots that have been examined, the core structural genes are coordinately regulated in two sets: a group of “early” genes and a group of “late” genes. In *Arabidopsis*, for example, light activation of flavonoid production first results in the production of transcripts of *CHS*, *CHI*, *F3H*, and *FLS*. Somewhat later, *DFR* and *ANS* are activated (Kubasek et al., 1992; Pelletier and Shirley, 1996; Pelletier et al., 1997). Moreover, activation of the late genes requires the product of the *TtG1* gene, whereas activation of the early genes does not, indicating that these two sets of genes regulated in different fashion. The patterns in *Antirrhinum* and *Petunia* are similar, although in *Antirrhinum*, *F3H* belongs to the set of “late” genes (Martin et al., 1991)

It is intriguing that the “early” genes in *Arabidopsis* correspond to those that evolved in early land plants (bryophytes and liverworts), while the “late” genes evolved in ferns and seed plants. This correspondence may indicate that a system of regulatory control over flavonoid production was established early in the assembly of the flavonoid pathway and has remained intact for more than 100 million years. Activation of this control system is induced by UV radiation (van Tunen et al., 1988; Kubasek et al., 1992; Pelletier et al., 1997), exactly what would be expected if flavonoids evolved originally to provide photoprotection as plants moved onto land. Under this scenario, a second regulatory control system would have been added as additional enzymes were incorporated into the flavonoid pathway, giving rise to the coordinated control over the “late” genes. Since the products of these “late” genes provide functions other than photoprotection, and since these products were needed in other tissues, it is not surprising that a separate regulatory system evolved to control them.

Although the regulatory genes that activate the “early” genes have not yet been identified, circumstantial evidence points to the involvement of genes from the MYB family of transcription factors. In particular, Moyano et al. (1996) demonstrated that two MYB proteins from *Antirrhinum* flowers activate *CHS* and *CHI* when expressed in yeast. Homologous MYB proteins also control expression of “late” genes in angiosperms, although these require the coparticipation of bHLH transcription factors (Mol et al., 1998) (see Chapter 4 for a description of the factors involved in the transcriptional regulation of flavonoid genes).

The observation that biosynthetic genes that presumably evolved at different times are regulated coordinately raises the question of how coordinate regulation arose. Although we currently do not have the information needed to answer this question, it is useful to speculate on how coordinate regulation *may* have evolved, if for no other reason than to generate concrete hypotheses that can be tested empirically.



**Figure 7.3** Distributions of isoflavonoids and flavonoids from the three main branches of the flavonoid pathway in land plants. Pluses in the “Iso” column indicate orders for which isoflavonoids have been reported. Pluses in the “I, II, and III” columns indicate presence of flavonoids in the pelargonidin, cyanidin, and delphinidin branches, respectively, of the pathway. Minuses indicate putative absence. Hash marks on phylogeny indicate most parsimonious locations of putative losses of flavonoids from the three branches. |, ||, and |||, respectively, represent loss of pelargonidin, cyanidin, and delphinidin branches. Phylogeny from Savolainen and Chase (2003). Data for taxonomic distribution of flavonoids are from Markham (1988), Giannasi (1988), Niemann (1988), and Williams and Harborne (1988). A more detailed analysis of distributions in plant families would reveal many more cases of probable loss of function in one or more of the pathway branches.

In this spirit, I offer the following model for the evolution of coordinate regulation (Figure 7.4). Imagine a time very early in the evolution of land plants in which chalcone synthase was the only flavonoid enzyme that had evolved. Presumably it was regulated by a MYB transcription factor (*MYB1* in Figure 7.4). When it became advantageous to incorporate another step into the pathway (e.g., flavonone-3-hydroxylase; this may have preceded the incorporation of CHI since the reaction catalyzed by CHI can proceed without the enzyme, albeit more slowly), this enzyme (*F3H* in Figure 7.4B) was derived via duplication from an oxoglutarate-dependent oxygenase (OGDO), which presumably was already regulated by its own transcription factor(s) (*TF2* in Figure 7.4A). Because timing and spatial pattern of expression controlled by *MYB1* and *TF2* probably were dissimilar, natural selection would probably have favored a more coordinated expression of *CHS* and *F3H* to increase the efficiency at which dihydroflavonols were produced. This could be accomplished easily by adding MYB1-protein binding sites to the promoter region of *F3H* (Figure 7.4C). At this point, *F3H* would be activated by either *MYB1* or *TF2*. Subsequently, selection also might favor the elimination of *tf2*-protein binding sites, if activation of *F3H* by *TF2* were deleterious (Figure 7.4D). This sequence of events thus would establish coordinated regulation of *CHS* and *F3H* by *MYB1*.

In a similar manner, additional enzymes could be brought under the control of *MYB1* as they evolved to produce novel flavonoids. In addition, the independent coordinated regulation of “late” biosynthetic genes could be established in similar fashion. In this case, the first enzyme, DFR, was derived by duplication from an NADH-dependent reductase, which presumably was already regulated by its own transcription factor(s), possibly a MYB protein (*MYB2* in Figure 7.4E). Duplication of the reductase, followed by divergence, created the DFR enzyme, which presumably still was regulated by *MYB2* (Figure 7.4F).

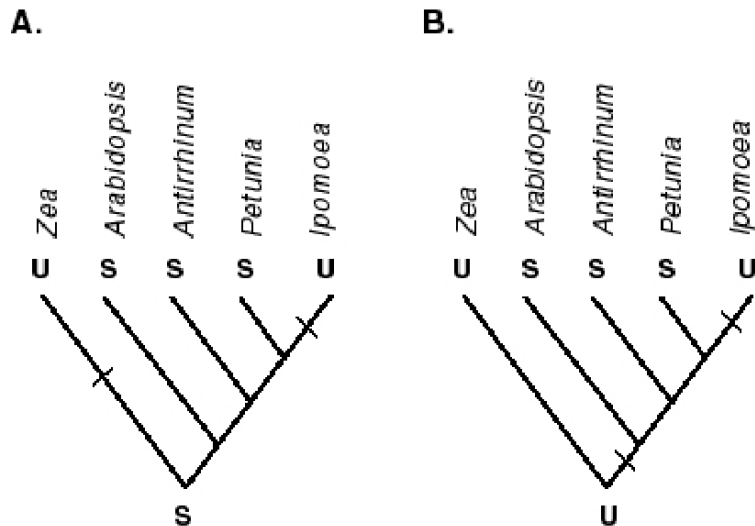
Efficient production of leucoanthocyanidins requires the coordinated expression of *DFR* with the “early” genes. Because at this stage *DFR* is regulated independently of the “early” genes, coordination is likely to be poor. Consequently, selection would probably favor the evolution of a mechanism that would produce increased coordination. One way of accomplishing this would involve duplication of *MYB2* to produce *MYB2a*, which could then evolve to activate *DFR* in a way that was better coordinated with the “early” genes (Figure 7.4G). This would establish a regulatory control system of *DFR* expression, to which then could be added subsequent “late” enzymes (*ANS*, *UF3GT*) as they evolved, by the same mechanism that established the coordinated regulation of the early enzymes.



While this scenario is hypothetical, it does make two very strong predictions: (1) Coordination of expression within a set of biosynthetic genes should have been brought about by evolutionary change in the promoter modules of the biosynthetic genes themselves, rather than by change in transcription factors to “capture” control over additional enzymes; and (2) the transcription factor(s) controlling “late” genes should be more closely related to those controlling the expression of the NADH-dependent reductase gene that gave rise to *DFR* than to the transcription factor controlling the expression of the “early” genes (i.e., *MYB2a* should be more closely related to *MYB2* than to *MYB1*) (Figure 7.4).

Although evaluating the second prediction must await further characterization of the actual transcription factors associated with the flavonoid pathway, some evidence already supports prediction (1). For example, during the production of anthocyanins in maize, both the “early” and “late” biosynthetic genes are coordinately activated by a single MYB transcription factor, acting with a single bHLH factor. Transient expression experiments indicate that the “late” genes were brought under the control of the “early” transcription factor by changes in the *cis*-regulatory regions of the “late” genes. In these experiments, transcription factors from *Petunia* and maize were reciprocally bombarded into tissues of each of these two species. The *Petunia* transcription factors used were those that activate the “late” genes in *Petunia*. Both sets of factors activated the “early” genes in maize, while neither set of factors activated the “early” genes in *Petunia*, demonstrating that the difference between maize and *Petunia* in control over “early” genes is due to changes in the promoter regions of those genes (Quattrocchio et al., 1993, 1998).

The hypothetical scenario described above assumes that control over “early” genes by the same transcription factor that controls “late” genes is a derived character in maize. This assumption is also testable. Currently, we have information about control of “early” and “late” genes during anthocyanin synthesis in five taxa. In three of the taxa (*Arabidopsis*, *Antirrhinum*, and *Petunia*), the two sets of genes are under separate control, while in maize and *Ipomoea* there is unified control of both sets. The phylogenetic relationships among these taxa are shown in Figure 7.5, along with the present character states. With this information, inferring that the ancestral state was either separate or unified, control is equally parsimonious, since in either case only two transitions are required to produce the observed character states. It should be possible, however, to distinguish between these possibilities when regulatory control is deciphered for additional taxa.



**Figure 7.5** Evolution of coordinate regulation of flavonoid biosynthetic genes. Character states indicated by U or S. U: Anthocyanin regulator has unified control over “early” and “late” genes. S: Control of “early” genes separate from control of “late” genes. A and B show that a minimum of two character state changes (hash marks) whether the ancestral state was S or U. Phylogeny from Savolainen and Chase (2003).

### 3. IS PHENOTYPIC EVOLUTION MEDIATED BY STRUCTURAL OR REGULATORY GENES?

A central issue in evolutionary biology regards the nature of genes involved in adaptive evolutionary change. Substantial evidence is accruing to indicate that regulatory sequences, including both *cis*-regulatory regions as well as transcription factors, play a major role in morphological evolution in plants and animals (King and Wilson, 1975; Britten and Davidson, 1969, 1971; Dickinson, 1991; Doebley, 1993; Doebley and Lukens, 1998; Wray et al., 2003). By contrast, much physiological adaptation clearly involves changes in structural genes (Watt, 1977, 1983; Hochachka and Somero, 1984; Crawford and Powers, 1989; Gillespie, 1991; Yokoyama et al., 1993; Newcomb et al., 1997; Purrington and Bergelson, 1997). It still is not clear, however, whether this apparent contrast represents a fundamental difference in the genetic underpinnings of morphological and physiological evolution or instead perhaps reflects the recent simultaneous burgeoning of interest in gene regulation and in evolution of development. More generally, evolutionary biology currently lacks a coherent framework for predicting whether any particular type of trait is likely to evolve primarily by changes in structural genes or changes in regulatory genes.

Investigations of the evolution of flower color are beginning to provide some insight into these evolutionary issues. Flower color is a particularly appropriate trait

for addressing these issues for several reasons: (1) the biochemical pathway that produces anthocyanins, probably the most important floral pigments, is relatively simple, and thus the connection between genotype and phenotype is straightforward and easily interpreted; (2) both structural genes and regulatory genes of the pathway have been extensively characterized in many different plants; (3) there is a tremendous amount of naturally occurring variation in flower color, both within as well as among species, that can be used to dissect the genetic factors responsible for evolutionary change; and (4) much is understood about the ecological significance of differences in flower color, which provides a context for interpreting evolutionary change in this character.

Evolutionary transitions in flower color frequently accompany or are accompanied by, changes in floral morphology that are believed to enhance the efficiency of interactions with new pollinators. Indeed, this is such a widespread phenomenon that “pollinator syndromes” have been recognized by plant evolutionary biologists for decades (Faegri and van der Pijl, 1966). For example, bee-pollinated flowers typically are blue-purple, have relatively short, broad tubes, broad limbs that serve as landing platforms, small amounts of concentrated nectar, and inserted anthers and stigmas. By contrast, hummingbird-pollinated flowers typically have reddish flowers, long narrow tubes, small limbs, copious dilute nectar, and exerted anthers and stigmas. Moth- and bat-pollinated flowers tend to be white, fragrant, and open at night. Many evolutionary changes in flower color thus seem to be adaptations associated with pollinator attraction.

Interestingly, many floral color transitions appear to be unidirectional. In the genus *Ipomoea*, for example, ancestral flowers were blue-violet and adapted to bee pollination. From this state, there have been numerous independent transitions to red, white and yellow flowers, but no documented cases of a transition from these colors back to blue-violet (Figure 7.6). Similarly, in the genus *Penstemon* there have been many independent transitions from blue, bee-pollinated flowers to red, hummingbird-pollinated flowers, without any evidence of the reverse transition (Wilson et al., 2004).

A simple hypothesis for this pattern is that these evolutionary transitions result from loss-of-function (LOF) mutations, for which reversions are unlikely. Several cases of naturally occurring LOF mutations of this type are known. For example, in the normally blue-flowered common morning glory, *Ipomoea purpurea*, red-flowered variants result from inactivation of the enzyme F3'H caused by a transpositional insertion into the *F3'H* gene that produces a premature stop codon (Zufall and Rausher, 2003; Hoshino et al., 2003). In the same species, natural white-flowered variants result from a similar transpositional inactivation of the enzyme CHS (Habu et al., 1998; Coberly, 2003) and from a deletion and frameshift in the coding region of a MYB transcription factor *IpMYB1* (Chang et al., 2005). In *Petunia axillaris*, white flowers have evolved from purple ancestors. The primary cause seems to have been a transposon-mediated deletion/frameshift in *AN2*, encoding a MYB transcription factor homologous to *IpMYB1*, although changes in several anthocyanin structural genes also contribute to lack of pigmentation (Quattrocchio et al. 1999). In the red-flowered, hummingbird-pollinated morning

glory *Ipomoea quamoclit*, which has evolved from blue-flowered, bee-pollinated ancestors, the transition from blue to red pigmentation is caused by production of pelargonidin-based rather than cyanidin-based anthocyanins. This shift has resulted from some combination of three different LOF mutations that block the cyanidin branch of the pathway: an almost complete down-regulation of F3'H, a knockout of F3'H function *in planta*, and a loss in the enzyme DFR of the ability to metabolize dihydroquercetin (Zufall and Rausher, 2004). Finally, in the white-flowered *I. alba* and *I. aquatica*, expression of the structural genes *CHS* and *DFR* is markedly reduced, though it is not known whether changes in other anthocyanin genes also may contribute to lack of floral pigmentation in these species (Durbin et al., 2003).

The recognition that LOF mutations often may contribute to flower color evolution suggests that there may be a relatively straightforward conceptual framework for predicting and explaining whether flower color evolution is more likely to involve structural or regulatory genes. Two extreme possibilities can be envisioned, depending on the magnitude of deleterious pleiotropy associated with LOF mutations. One possibility is that mutations in regulatory sequences incur substantially lower levels of deleterious pleiotropy than mutations in structural genes. Under this hypothesis, whenever selection favors a novel flower color, regulatory mutations will have a higher net selection coefficient than structural gene mutations, because they experience lower deleterious pleiotropy. If the magnitude of pleiotropy is large enough, it is possible that only the regulatory mutations will have a positive net selection coefficient, and thus contribute to flower color evolution. Even if the magnitude of deleterious pleiotropy is not sufficient to completely offset the flower color advantage in structural gene mutations, the net selection coefficient will be smaller for those mutations than for mutations in regulatory sequences. Because the probability that an advantageous mutation actually will become fixed is roughly proportional to its selection coefficient (Hartl and Clark, 1989), this means that it will still be more likely for flower color evolution to be caused by changes in regulatory sequences.

The second extreme possibility is that when a new flower color becomes advantageous, the associated fitness benefit dwarfs the fitness cost of pleiotropy in any anthocyanin gene, structural or regulatory. Under this circumstance, selective coefficients of LOF mutations will not differ substantially among genes and all will have a roughly equal probability of fixation. Chance will predominate and determine whether regulatory or structural gene mutations ultimately give rise to a novel flower color.

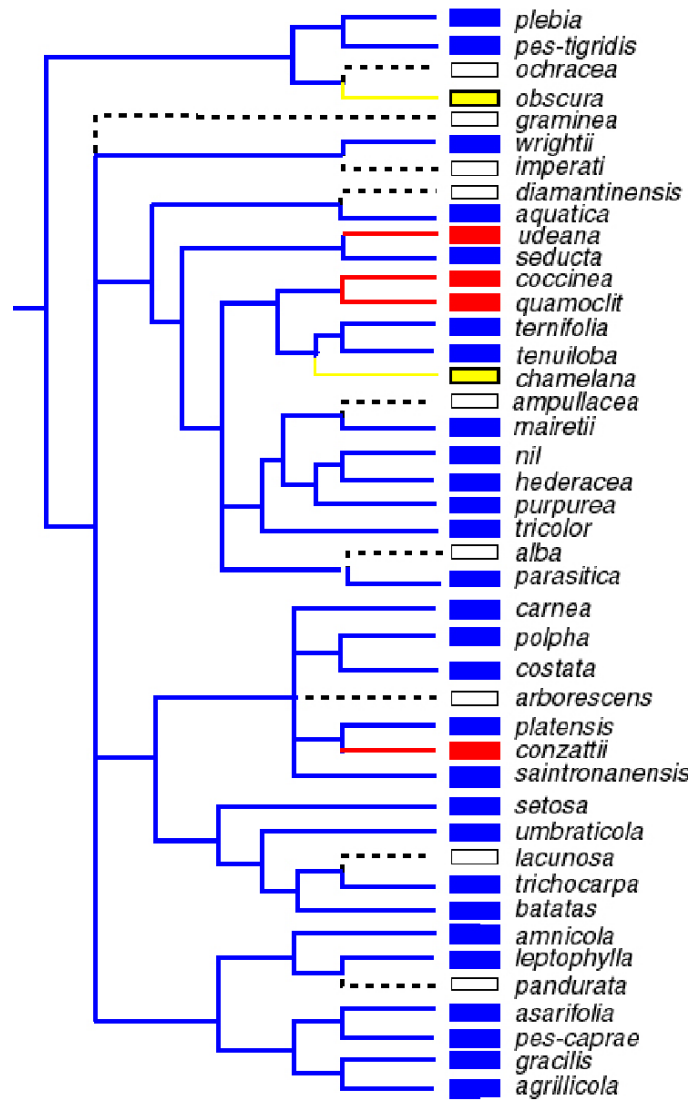
### 3.1. How common is differential pleiotropy?

Substantial differences in the magnitude of deleterious pleiotropy between structural and regulatory LOF mutations may be expected for several reasons. First, anthocyanin structural genes often have broad expression domains (e.g., Durbin et al., 2003). For example, all structural genes in *Arabidopsis thaliana*, except flavonol synthase, are single-copy genes, and thus are expressed in all tissues that accumulate anthocyanins (Winkel-Shirley, 2001). Moreover, the upstream genes are expressed

in all tissues that accumulate any type of flavonoid. In other species, some structural genes exist as small gene families (Koes et al., 1989; Durbin et al., 1995; Inagaki et al., 1999). Nevertheless, in many cases, one member of the gene family appears to be the primary copy expressed in most tissues (Koes et al., 1989; Inagaki et al., 1999; Durbin et al., 2003). Because of this broad expression domain, a knockout of a structural gene typically would be expected not only to produce altered flower color but also reduce or eliminate production of some anthocyanins and other flavonoids in many other tissues (Quattrocchio et al., 1993; Huits et al., 1994; van Houwelingen et al., 1998). Because these compounds serve a variety of often-crucial physiological and ecological functions in plants (Koes et al., 1994; Shirley, 1996), such a reduction is likely to substantially counteract any fitness advantages associated with the production of altered flower color.

By contrast, LOF in anthocyanin regulatory sequences are expected *a priori* to have fewer deleterious consequences for at least two reasons. First, anthocyanin transcription factors often exist as multigene families, with each member of the family having a restricted domain of expression (Ludwig and Wessler, 1990; Cone et al., 1993; Quattrocchio et al., 1993; Huits et al., 1994). As a consequence, knockouts of these genes can eliminate pigment production in flowers (or even in parts of flowers), without affecting anthocyanin or general flavonoid production in other tissues. Second, as is expected because of the general modularity of *cis*-regulatory sequences, expression of anthocyanin genes is commonly regulated independently in different tissues (Ludwig et al., 1990; Radicella et al., 1992; Patterson et al., 1995). In particular, mutations in *cis*-regulatory regions can differentially alter anthocyanin production in different tissues (e.g., Hansen et al., 1996). These observations suggest that mutations in *cis*-regulatory regions can greatly down-regulate anthocyanin production in flowers without necessarily reducing flavonoid production in other tissues. Such tissue-specific structural gene down-regulation is correlated with absence of floral pigmentation in several *Ipomoea* species (Durbin et al. 2003). Although in this case it has not been directly demonstrated that the changes in expression level are due to changes in *cis*-regulation rather than alteration of a transcription factor, the former seems likely because coordinate down-regulation of blocks of structural genes is not observed. In the white-flowered *I. alba*, for instance, *CHS* is down-regulated but *CHI* is not. Our own unpublished data indicate that only *CHS* and *DFR* among the seven core structural genes are down-regulated in *I. alba* flowers. Similarly, in the red-flowered *I. quamoclit*, only *F3'H* is down-regulated; absence of F3'H results in the absence of blue cyanidin-based pigments (Zufall and Rausher, 2003, 2004).

Experimental examination of the fitness consequences of anthocyanin structural and regulatory gene LOF in *Ipomoea purpurea* supports the hypothesis that structural gene mutants that produce white flowers experience substantially greater negative pleiotropy than regulatory mutants. In natural populations of this species, mutations at two different loci (*A* and *W*) produce very similar white-flower phenotypes (Ennos and Clegg, 1983; Epperson and Clegg, 1987).



**Figure 7.6** Unidirectional evolution of flower color in *Ipomoea*. Phylogeny and ancestral state reconstruction based on Miller *et al.* (1999, 2004) and Zufall (2003). Boxes indicate flower color. All species are in the genus *Ipomoea*. See Color Section for figure in colors.

At the *A* locus, white flowers result from a mutation in the structural gene *CHS-D* (Habu *et al.*, 1998; Coberly, 2003), the primary copy of chalcone synthase that is

expressed in most plant tissues (Fukada-Tanaka et al., 1997; Durbin et al., 2003). This mutation is caused by a transposon insertion into the sole intron that results in a truncated transcript. By contrast, white flowers at the *W* locus are a consequence of an out-of-frame deletion that introduces a premature stop codon into *IpMYB1* (Chang et al., 2005), which coordinately activates many of the anthocyanin structural genes in *I. purpurea* (Tiffin et al., 1998).

The direct effect of the white flowers produced by these two mutations on the mating system, and thus the direct fitness effects of the white phenotype, are very similar (Brown and Clegg, 1984; Rausher et al., 1983; Fehr and Rausher, 2004). Pleiotropic effects differ substantially, however. Compared to “wild-type” plants with blue flowers (*AAWW* genotype), white-flowered *aa* plants exhibit reduced germination and early survival, increased susceptibility to insect herbivores and fungal pathogens, and reduced seed set and pollination success at high (though natural) temperatures (Coberly, 2003; Coberly and Rausher, 2003). By contrast, white-flowered *ww* plants exhibit similar germination rates, viability, and seed set compared to blue-flowered plants (Rausher and Fry, 1993; Mojonner and Rausher, 1997) and show no differences in susceptibility to herbivores or pathogens (Fineblum and Rausher, 1997). This difference in the magnitude of the deleterious pleiotropy is not only consistent with the expectation of greater negative effects of LOF mutations in structural genes, but may explain why the *w* allele is much more common in natural populations than the *a* allele (Coberly and Rausher, 2003; Fehr and Rausher, 2004).

Whether this pattern of differential pleiotropy holds in general remains to be explored. Numerous examples of pleiotropic effects of both anthocyanin structural and regulatory gene mutations have been documented (e.g., Coe et al., 1981; Mo et al., 1992; Van der Meer et al. 1992; Li et al., 1993), but the relative magnitude of pleiotropy is seldom assessed in the same species in fitness units under natural conditions, so it is unclear whether the fitness effects of negative pleiotropy tend to be higher in structural genes. Nor is it clear whether floral color adaptation is accomplished more frequently through modification of regulatory genes than of structural genes. Only a handful of attempts have been made to discern the genetic changes responsible for evolutionary change in flower color (Quattrocchio et al., 1999; Zufall and Rausher, 2004; Durbin et al., 2003), and even in these cases it is unclear whether redundant LOF changes in genes not examined might contribute substantially to the observed change in floral hue.

#### 4. EVOLUTIONARY RATE VARIATION AMONG ANTHOCYANIN GENES

Variation in evolutionary rates among proteins is ubiquitous (Li, 1997). Rates of amino acid substitution vary over several orders of magnitude. Some proteins, such as histones, have undergone essentially no amino acid substitutions over hundreds of millions of years. By contrast, the *Drosophila* gene *OdsH* has undergone 10 amino acid replacements in approximately 1 million years (Ting et al., 1998). Two explanations have commonly been proposed to account for this variation: variation in selective constraint and variation in the frequency of advantageous substitutions.

At the extremes, these explanations are probably correct. Histones interact directly with DNA and other histones in the formation of the nucleosome. Because most amino acids in histone molecules participate in these interactions, it is likely that the optimal amino acid configuration was achieved early in the history of life and that there have been few subsequent replacements because virtually any amino acid substitution would disrupt function. In other words, selective constraint is almost complete in these proteins. In the case of *OdsH*, repeated episodes of advantageous substitutions are indicated by the very high Ka/Ks ratio.

For the large majority of proteins that do not exhibit such extremely high or low substitution rates, however, the causes of rate variation are less clear. Although population genetic tools that would permit a determination of the relative importance of variation in constraint and variation in positive selection have been available for nearly a decade, they seldom have been applied to addressing the causes of rate variation. Recently, however, application of these techniques to substitution patterns in anthocyanin pathway genes has begun to illuminate the causes of rate variation among these genes. In addition, they have begun to suggest that the rate at which particular genes evolve may in part be determined by the position and role that their protein products in the anthocyanin biochemical pathway.

#### 4.1. Rate variation among structural genes

Rausher et al. (1999) quantified rates of amino acid substitution in six core anthocyanin structural genes (*CHS*, *CHI*, *F3H*, *DFR*, *ANS*, and *UF3GT*) by comparing maize sequences to *Antirrhinum* and *Ipomoea* sequences. They found more than a fivefold difference in the nonsynonymous substitution rate between the most rapidly evolving gene (*UF3GT*) and the most slowly evolving gene (*CHS*). Because of the broad taxonomic comparison in this study, substitutions at synonymous sites were saturated, and it thus was not possible to distinguish between alternative explanations for rate variation.

In a subsequent taxonomically more restricted study, Lu and Rausher (2003) quantified both synonymous and nonsynonymous substitution rates in *CHS*, *ANS*, and *UF3GT* by examining six species within the genus *Ipomoea*. As in the previous study, *CHS* exhibited the lowest amino acid substitution rate, *UF3GT* the highest, and *ANS* an intermediate rate. Differences in the underlying mutation rates were ruled out as a cause of this variation because the gene with the lowest nonsynonymous substitution rate had the highest synonymous substitution rate. In addition, codon-based analyses of Ka/Ks ratios (Yang et al., 2000) revealed very little evidence that positive selection had caused extensive amino acid substitution in any of the three genes: no positively selected sites were detected for either *CHS* or *UF3GT*, and only two positively selected sites (out of 349) were detected for *ANS*. Together, these results suggest that variation in evolutionary rates among anthocyanin structural genes is most likely due to a variation in selective constraint. This inference must be viewed as tentative, however, because the analyses used to detect positive selection can be very conservative and can fail to detect selection if

advantageous substitutions are scattered throughout the gene rather than being concentrated in certain sites.

A striking pattern that emerged from the Rausher et al. (1999) and Lu and Rausher (2003) analyses of rate variation is that downstream enzymes evolve faster than upstream enzymes. This pattern implies that genes corresponding to the downstream enzymes are under lower selective constraint than upstream genes. Although this pattern could be due simply to chance, there is another intriguing explanation: the position of an enzyme in the pathway determines in part the degree of selective constraint it experiences.

Upstream anthocyanin pathway enzymes also are required for the production of other flavonoid compounds, whereas downstream enzymes are not (Figure 7.1). Slightly deleterious mutations in upstream genes that reduce catalytic efficiency are thus likely to reduce flavonoid production in addition to anthocyanin production. By contrast, a similar mutation in a downstream gene potentially will affect only anthocyanin production. In other words, equivalent mutations are likely to have greater deleterious pleiotropy in upstream enzymes. This greater pleiotropy is in turn likely to increase the magnitude of the selection coefficient associated with the mutation, making it less likely that the mutation will be fixed by genetic drift. This scenario thus envisions that at evolutionary equilibrium most nonsynonymous substitutions represent chance fixation of slightly deleterious mutations and that the rate of such fixation is higher in downstream enzymes. While this hypothesis is consistent with observed patterns of rate variation, there is currently very little direct evidence supporting it.

#### *4.2. Rapid evolution of regulatory genes*

Because of their role in controlling developmental processes, it has been hypothesized that plant transcription factors may play a central role in the evolution of plant morphology (Doebley, 1993; Doebley and Lukens, 1998; Purugganan, 1998). An intriguing observation that is in accordance with this hypothesis is that plant transcription factors often evolve at elevated rates compared to the structural genes they regulate (Purugganan et al., 1995; Purugganan, 1998; Barrier et al., 2001; Remington and Purugganan, 2002; Dias et al., 2003), a pattern that also appears to hold for anthocyanin gene transcription factors (Purugganan and Wessler, 1994; Rausher et al. 1999). Usually, this elevated rate of nonsynonymous substitutions is concentrated in certain protein domains, while other domains (e.g., DNA-binding domains) are highly constrained. In the rapidly evolving domains,  $K_a/K_s$  ratios often approach 1. Moreover, high rates of insertion/deletion often are found in these rapidly evolving domains (Dias et al., 2003; Chang et al., 2005; Shavorskaya and Lagercrantz, unpublished manuscript). As is the case for the rapid evolution of structural genes, this rapid rate of evolution can be explained in two ways: by greatly relaxed selective constraint or by substantially enhanced, repeated positive selection. Confirmation of the latter explanation would constitute support for the importance of transcription factors in morphological evolution.

Analysis of substitution patterns in the *IpMYB1* anthocyanin transcription factor in *Ipomoea* does not provide such support (Chang et al., 2005). This gene, which is a member of the R2R3-MYB family of transcription factors, can be divided into two domains: an approximately 350 bp encoding the DNA-binding domain and an approximately 550 bp encoding the non-DNA-binding domain region. While the binding domain is believed to function in both DNA sequence recognition and binding to partner bHLH transcription factors (Goff et al., 1992; Williams and Grotewold, 1997; Sainz et al., 1997; Grotewold et al., 2000), little is known about the function of the variable region other than that it contains a transcription activation domain (Goff et al., 1991).

As expected, the DNA binding domain of this gene is highly conserved, exhibiting a relatively low Ka/Ks ratio of 0.235. The nonbinding domain, on the other hand, exhibits a Ka/Ks ratio of approximately 0.75. A series of independent analyses indicate that approximately 25% of the amino acid sites in the nonbinding domain are very highly conserved, while the remaining sites are evolving effectively neutrally. At these sites, Ka/Ks = 1, the proportions of radical and conservative amino acid substitutions do not differ from neutral expectations, and neither interspecific nor intraspecific patterns of variation reveal any evidence that positive selection has contributed to the elevated replacement rate. It thus appears that the rapid evolution of this transcription factor is caused by greatly reduced selective constraints. Interestingly, the highly conserved sites are scattered randomly through the non-DNA-binding domain region, suggesting that the domain as a whole serves one or more crucial functions. Moreover, these conserved sites contain a disproportionate number of negatively charged, acidic amino acids. These observations suggest that the known function of transcriptional activation may be the primary function of nonbinding domain. The important functional feature of transcription activation domains of eukaryotic transcription factors appear to consist of negatively charged amino acids aligned along the same edge or face of alpha helices or other secondary structures, while the composition and three-dimensional structure of the remainder of the domain appears to be largely irrelevant (Ptashne, 1988). One thus would expect to see strong conservation at sites where acidic amino acids are present, as well as at sites that maintain the proper orientation of those sites but little constraint on other sites, a pattern that is consistent with what is observed in *IpMYB1*.

The absence of detectable positive selection in *IpMYB1* is consistent with results of other investigations of selection on rapidly evolving plant transcription factors. Remington and Purugganan (2002) failed to detect positive selection on growth-regulating transcription factors in Hawaiian Silverswords, and Shavorskaya and Lagercrantz (unpublished results) similarly detected no positive selection on *Col1* in the *Brassicaceae*. Although it is dangerous to generalize from a small number of investigations, current evidence suggests that rapidly evolving domains of plant transcription factors do not contribute to adaptive phenotypic evolution.

The results of two recent investigations of DNA-binding domain evolutionary diversification in members of the R2R3-MYB transcription factor family (Jia et al., 2003, 2004), which includes several anthocyanin regulatory genes, may at first seem difficult to reconcile with this conclusion (but see Dias et al., 2003). In both cases,

positive selection was demonstrated to be responsible for DNA-binding domain substitutions. However, these investigations examined the divergence of paralogs within the same species. This divergence presumably occurred early in the evolutionary history of the gene family, when different paralogs were acquiring different regulatory functions. By contrast, the three studies referred to in the previous paragraph all examined evolutionary divergence of orthologous copies. During the divergence of these orthologs, a common function for the binding domains (e.g., activation of anthocyanin structural genes, in the case of *IpMYB1*) has been maintained. One thus would not expect to see major functional divergence and thus substantial positive selection in the binding domains of these genes, and one does not.

One of the most remarkable aspects of rapidly evolving anthocyanin gene transcription factors is that function is highly conserved. Ectopic expression of *C1* (a *MYB* gene) and *R* (a *bHLH* gene) from maize elicits anthocyanin production in *Petunia*, *Arabidopsis*, and *Nicotiana* (all dicots) (Lloyd et al., 1992; Quattrocchio et al., 1993), despite the common ancestors of maize and these species having lived approximately 100–125 million years ago (Savolainen and Chase, 2003). This common function has been maintained despite, in some cases, the absence of recognizable homology in the rapidly evolving non-DNA-binding domains.

Explaining how an essentially neutrally evolving region at the same time can maintain its function over long periods of evolutionary time is a challenging evolutionary mystery. One possibility, alluded above, is that for some regions the only functional requirement is the presence of certain types of amino acid (e.g., acidic). Another possibility is that only the three-dimensional configuration, rather than the specific amino acid composition, of the region is of primary importance to function (e.g., in mediating contact between transcription factors). If this hypothesis is correct, then function may be quite tolerant to virtually any individual amino acid substitution at most sites, although at any one time there may be a minority of sites that are crucial for maintaining configuration, and thus at least over a relatively short evolutionary period, are under severe constraint, as seen in *IpMYB1*. Over longer time periods, however, turnover may be possible at constrained sites also if, for example, substitutions at other previously unconstrained sites relieve constraint by introducing new foci for stabilization of the three-dimensional configuration.

Although, as suggested above, much flower color evolution appears to involve loss-of-function mutations in the anthocyanin pathway, it also is clear that this pathway has contributed frequently to the evolution of novel characters. Arguably, the most important process yielding new function is gene duplication followed by the evolution of novel function in one of the duplicated copies (neofunctionalization). Evolutionary analyses of flavonoid gene families suggests that this process has repeatedly given rise to novel classes of secondary compounds in plants.

## 5. GENE DUPLICATION AND THE EVOLUTION OF NOVEL FUNCTION

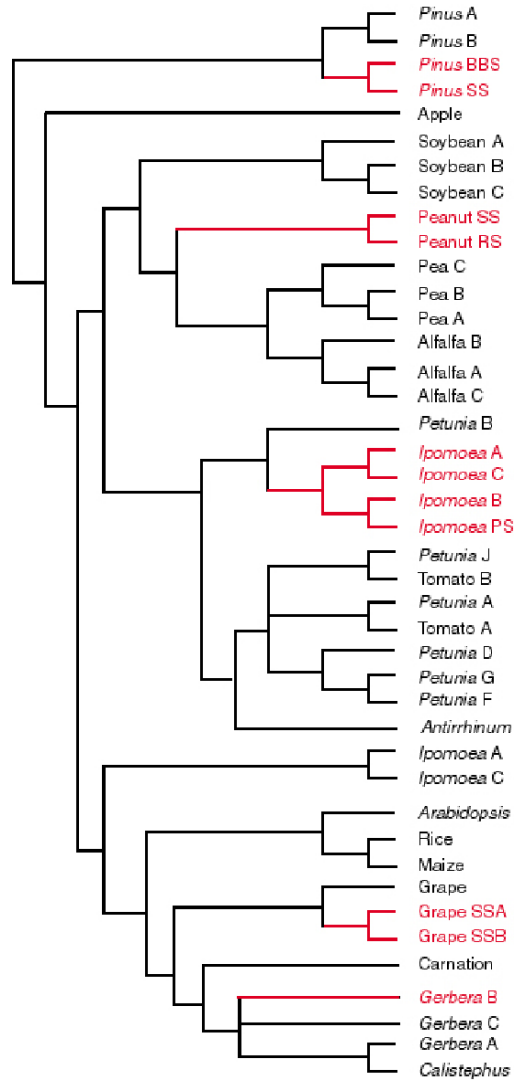
### 5.1. Duplication and divergence in chalcone synthase

A striking example of the evolution of new function is provided by stilbene synthases (Tropf et al., 1994). Stilbene phytoalexins are produced sporadically throughout the plant kingdom as defenses against fungal pathogens (Figure 7.7). The backbone of these compounds is synthesized by the enzyme stilbene synthase.

Stilbene synthases exhibit moderate to high amino acid sequence similarity to chalcone synthases and catalyze a similar condensation reaction. Using a phylogenetic analysis of chalcone and stilbene synthases, Tropf et al. (1994) showed that stilbene synthases have been derived independently several times from chalcone synthase, presumably following *CHS* duplication. They also demonstrated, using site-directed mutagenesis and *in vitro* enzyme assays, that modification of only three or four amino acids in chalcone synthase is needed to produce stilbene synthase activity.

Duplication and divergence of chalcone synthases apparently has produced a variety of other novel enzymes. For example, the common morning glory, *Ipomoea purpurea*, contains a six-member gene family exhibiting sequence similarity to chalcone synthases from other plants. Five of these copies appear to produce functional enzymes, while the sixth copy is apparently a pseudogene (Clegg and Durbin, 2003). Of the five functional copies, *CHS-D* is the primary transcript in most tissues, including flowers (Fukada-Tanaka et al., 1997; Durbin et al., 2003). This copy, along with *CHS-E*, has been shown to be capable of catalyzing the canonical chalcone synthase reaction: condensation of 4-coumaroyl-CoA and malonyl-CoA to form naringenin chalcone, the first step of the flavonoid pathway (Clegg and Durbin, 2003). By contrast, three of the copies, *CHS-A*, *-B*, and *-C*, are not able to catalyze this reaction, suggesting that they have evolved alternate as yet unknown functions. This suggestion is supported by the observation of an accelerated nonsynonymous substitution rate in the lineage leading to these copies (Durbin et al., 1995; Rausher et al., 1999), as well as the demonstration that at least some of the substitutions on this branch were positively selected (Yang et al., 2004).

Helariutta et al. (1996) described a similar situation in the Asterids. They isolated a gene with deduced amino acid similarity to asterid chalcone synthases and stilbene synthases. But the expression domain of this gene differs from the typical expression domains of chalcone and stilbene synthases, as does its substrate specificity *in vitro*, indicating that this gene has evolved some novel unknown function. A survey of several Asterid species revealed that this gene itself had duplicated and formed a small gene family, the members of which have diverged in both expression domain and substrate specificity. Nevertheless, the gene tree they obtained indicated that this novel family was derived from chalcone synthase. In addition to these novel enzymes, duplicates of chalcone synthase have evolved into resveratrol synthases and bibenzyl synthases (Figure 7.7) (Clegg and Durbin, 2003).



**Figure 7.7** Evolution of stilbene synthases (SS), bibenzyl synthases (BBS), and resveratrol synthases (RS) from chalcone synthases. Black: chalcone synthases. Red: modified chalcone synthases. In cases in which the modified chalcone synthase is not listed as another enzyme (*Ipomoea A, B, C, Gerbera B*), these enzymes have been shown to preferentially utilize substrates other than 4-coumaroyl-CoA and/or malonyl-CoA, the normal chalcone synthase substrates. Modified from Tropp et al. (1994) and Clegg and Durbin (2003). See Color Section for figure in colors.

### 5.1. Duplication and subfunctionalization in *CHS*

While some duplicate copies of *CHS* have developed new enzyme functions, it is also common for duplicate copies to retain the ancestral canonical function. In *Ipomoea purpurea*, for example, while *CHS-A*, *-B*, and *-C* have acquired novel catalytic properties, *CHS-D* and *-E* both retain standard chalcone synthase activity (Clegg and Durbin, 2003). This situation raises the evolutionary issue of why both of these copies are retained to perform the same enzymatic function. Unless a double dose of this enzyme is needed to produce an optimal level of flux, an unlikely scenario, one duplicate is expected in most cases (except in presumably rare cases of neofunctionalization) to accumulate inactivating mutations and become a pseudogene.

One possible explanation for the maintenance of catalytically similar copies is simply that there has not been enough time for pseudogenization. This explanation seems unlikely to account for preservation of *CHS-D* and *-E* in *Ipomoea* because selective constraint, as measured by Ka/Ks ratios, remains strong on both copies (Yang et al., 2004). If one of these copies were undergoing pseudogenization, the Ka/Ks ratio would tend approach 1, rather than the observed low value of 0.055.

An alternative explanation is that the two copies are maintained because they have undergone subfunctionalization (Hughes, 1994; Force et al., 1999). Under this process, the expression domains may diverge through the accumulation of neutral mutations in *cis*-regulatory regions that reduce or eliminate expression in certain tissues or under certain conditions (e.g., drought stress, pathogen attack). Such mutations are expected to be neutral as long as they affect only one of the duplicate copies. Once one or more such mutations have accumulated in each copy, both copies are necessary to realize the same expression domain as the single ancestral copy. Consequently, both copies will be retained indefinitely by purifying selection.

If this explanation is correct, it would be expected that duplicate copies would exhibit at least partially distinct expression domains. This appears to be the case for *CHS-D* and *-E* (Durbin et al., 2003), as well as for *DFR-A*, *-B*, and *-C* in *Ipomoea* (Inagaki et al., 1999), for four distinct *CHS* genes in *Petunia* (Koes et al., 1989), and for the duplicate *CHS* genes *C2* and *WHP* in maize (Holton and Cornish, 1995).

Although the idea that subfunctionalization can explain maintenance of duplicate gene copies has become very popular, there are very few convincingly documented cases, including anthocyanin genes. Most claims of subfunctionalization rest on simple demonstration of noncongruent expression domains, as described above. However, noncongruent domains also can result from acquiring novel expression domains in different copies (e.g., Hansen et al., 1996). The only way to distinguish between these two causes of noncongruence is to carefully reconstruct phylogenetically the ancestral, single-copy expression domain and determine if the domains of the current duplicate copies are subsets of that ancestral domain. Flavonoid structural genes are an excellent candidate for this approach.

### 5.2. Duplication and divergence in flavonoid regulatory genes

Gene duplication and divergence also has played a significant role in the evolution of flavonoid regulation. Four classes of transcription factors have been implicated across angiosperms as contributing to regulation (Mol et al., 1998). Two of these, belonging to the *MYB* and *bHLH* gene families, themselves are found in multiple copies in plants. In maize, for example, MYB transcription factors include C1 and PL1. The two exhibit greater than 90% amino acid identity in their amino- and carboxyl-terminal domains, which participate in regulatory function. These two duplicates have diverged greatly, however, in their expression domains: *Pl1* activates anthocyanin production in vegetative tissues and flowers, while *C1* activates anthocyanin biosynthesis in the aleurone layer of kernels (Cone et al., 1993).

bHLH-Type flavonoid transcription factors are represented in maize by a half-dozen paralogs (Ludwig and Wessler, 1990) grouped in two sets. One set of apparent tandem duplicates is found on chromosome 10, while a second set is found on chromosome 2. These copies have largely nonoverlapping expression domains, and thus activate anthocyanin production in different tissues (Ludwig and Wessler, 1990). Nevertheless, particle bombardment experiments demonstrate that any of these copies, when driven by a constitutive promoter, will induce pigment formation in virtually any tissue to which it is introduced (Ludwig et al., 1990; Goff et al., 1990). Multiple copies of these transcription factors also are found in dicots. For example, in *Petunia*, at least three distinct MYB factors control pigmentation in different parts of the flower (Kroon, 2004), while *Arabidopsis* also has at least three such copies (Nesi et al., 2001; Borevitz et al., 2000).

Transient and stable transformation of anthocyanin regulatory genes from one species into corresponding LOF mutants in distantly related species repeatedly has demonstrated that copies from different species are functionally exchangeable (Quattrocchio et al., 1993, 1998; Lloyd et al., 1992). This observation, coupled with differentiation of expression domains among duplicated copies suggests that maintenance of multiple copies of these transcription factors may be the result of subfunctionalization, though the caveats described above apply here also. Regardless of the evolutionary processes that maintain the duplicate copies, however, their noncongruent expression domains presumably allow pigmentation patterns to evolve somewhat independently in different tissues, including different floral parts. Divergence among species in complex floral color patterning thus presumably has been rendered possible by prior duplication of anthocyanin regulatory genes.

## 6. PATHWAY DEGENERATION AND EVOLUTIONARY POTENTIAL

Anthocyanins are a highly diverse group of flavonoids. Several hundred different anthocyanins have been described from the gymnosperms and angiosperms. Nevertheless, this great diversity is built upon structural elaboration of a small

number of anthocyanidins, which form the backbone of every anthocyanin molecule. There are six commonly occurring anthocyanidins, plus another ten or so that occur sporadically in these plant groups (Timberlake and Bridle, 1980). The diversity of anthocyanins results from the myriad of ways these few anthocyanidins can be decorated by the addition of sugars.

The six common anthocyanidins are the product of three different branches of the anthocyanin pathway (Figure 7.1). One branch gives rise to a single unmethylated anthocyanidin, pelargonidin, which tends to produce red or orange anthocyanins. A second branch gives rise to two anthocyanidins, the unmethylated cyanidin and the singly methylated peonidin. Anthocyanins derived from these compounds tend to be blue or magenta. The branching enzyme leading to this branch is F3'H, which adds a hydroxyl group to the 3' carbon of the anthocyanidin skeleton. Finally, a third branch gives rise to three anthocyanidins, the unmethylated delphinidin, the singly methylated petunidin, and the doubly methylated malvidin, which tend to produce blue-purple pigments. The branching enzyme associated with this part of the pathway is F3'5'H, which adds hydroxyl groups to the 3' and 5' carbons of the skeleton.

Flower color is in large part determined by which of these branches is most active in a particular species. In *Ipomoea*, for example, blue- and purple-flowered species tend to produce almost exclusively cyanidin-based anthocyanins. Pathway flux in these species is almost entirely down the second branch. However, mutations that knock out the enzyme F3'H redirect flux down the pelargonidin branch and result in red flowers (Zufall and Rausher, 2003; Hoshino et al., 2003). Red-flowered *Ipomoea* species also almost always produce pelargonidin-based rather than cyanidin-based, anthocyanins (Zufall and Rausher, 2004; Zufall, 2003). A similar pattern is seen in *Penstemon* (Scrophulariaceae), in which blue/purple, bee-pollinated flowers tend to produce delphinidin-derived anthocyanins, while red, hummingbird-pollinated flowers tend to produce pelargonidins (Scogin and Freeman, 1987). Adjusting the relative amounts of flux down the different pathway branches thus seems to be a common way of altering flower color.

Several lines of evidence indicate that these three branches evolved very early in the diversification of plants. First, the ability to produce all three classes of anthocyanidins is exhibited by most angiosperm orders as well as many gymnosperms (Figure 7.3). Second, the hydroxylating enzymes F3'H and F3'5'H, exhibit high-sequence homology throughout these two groups of plants, indicating that they were each recruited into flavonoid metabolism once very early in the diversification of the land plants. Flavonols derived from all three of the branches have been reported in ferns (Wollenweber and Schneider, 2000) and from the first two branches in bryophytes (Webby et al., 1996).

Despite the near taxonomic ubiquity at the level of plant orders of these three branches of the flavonoid pathway there are numerous known instances in which one or more of the branches has been inactivated (Table 7.1). At least two types of genetic changes have been associated with this inactivation. In some genera (*Arabidopsis*, *Petunia*, *Cymbidium*), enzymes below the branch point, such as DFR, have evolved to be substrate specialists. For example, in *Petunia* no pelargonidin-

derived anthocyanins are produced because the *Petunia* DFR does not metabolize dihydrokaempferol, a precursor of pelargonidin. In other genera (*Ipomoea*, *Chrysanthemum*, *Rosa*, *Dianthus*), the branching enzymes themselves have been inactivated. Thus, morning glories (*Ipomoea*) do not produce derivatives of delphinidin because they lack the enzyme F3'5'H. Extensive surveys of land plants for the types of flavonoids present (Markham, 1988; Giannasi, 1988; Niemann, 1988; Williams and Harborne, 1988) also suggest that knockouts of one or more of these three branches may be common, since there are numerous plant orders in which compounds associated with particular branches are apparently absent (Figure 7.3).

**Table 7.1** Taxa for which a branch of the anthocyanidins pathway has been inactivated

<i>Taxon</i>	<i>Family</i>	<i>Branch Inactivated</i>	<i>Implicated Enzyme</i>
<i>Arabidopsis</i>	Brassicaceae	pelargonidin	DFR substrate specialization
<i>Chrysanthemum</i>	Asteridae	delphinidin	F3'5'H activity absent
<i>Cymbidium</i>	Orchidaceae	pelargonidin	DFR substrate specialization
<i>Dianthus</i>	Caryophyllaceae	delphinidin	F3'5'H activity absent
<i>Ipomoea</i>	Convolvulaceae	delphinidin	F3'5'H activity absent
<i>Petunia</i>	Solanaceae	pelargonidin	DFR substrate specialization
<i>Rosa</i>	Rosaceae	delphinidin	F3'5'H activity absent

On theoretical grounds, the evolution of redundant knockouts in a branch of the pathway may be expected, because once a branch has been inactivated by a knockout in one enzyme it is unlikely that purifying selection will act on other enzyme functions associated with that pathway. Thus, in the absence of a DFR that metabolizes dihydrokaempferol, selection on ANS and UF3GT to maintain the ability to utilize leucopelargonidin and pelargonidin will be removed. The genes coding for these enzymes therefore will tend to accumulate mutations that inhibit these activities.

Although it is clear that in the species listed in Table 7.1, at least one enzyme has been inactivated, for most of these species it is not known whether other enzymes associated with the same pathway branch also may have been redundantly inactivated. In *Petunia*, however, transformation of a strain lacking DFR, F3'H, and F3'5'H with a *Gerbera* DFR, which efficiently metabolizes both DHK and DHQ, produces copious pelargonidin-based anthocyanins (Johnson et al., 1999), indicating that enzymes downstream of DFR have not degenerated in *Petunia*. By contrast, transformation of maize DFR into a DFR- and F3'H-deficient *Arabidopsis* strain produces only minimal amounts of pelargonidin-derived anthocyanins (Dong et al., 2001), suggesting that one or more downstream genes in *Arabidopsis* may be at least partially degenerate with respect to metabolizing pelargonidin precursors or derivatives.

This type of pathway “degeneration” is likely to have important implications for future evolutionary potential. When a branch of the pathway has been inactivated because of a change in only one enzyme, it is possible for back-mutation to restore that branch if it becomes advantageous to do so (e.g., if a different flower color becomes advantageous). However, once two or more enzymes acquire inactivating changes, the chances of restoring the lost branch become much more remote for the simple reason that the chances of two or more back mutations occurring simultaneously is exceedingly small. Branch inactivation thus is likely to be permanent, thus restricting the evolutionary possibilities for ecologically important characters such as flower color. This phenomenon is likely to reinforce the apparent tendency for unidirectionality of flower color change (Figure 7.6) (see above).

A recent investigation of changes in anthocyanin pathway enzymes in the morning glory *Ipomoea quamoclit* provides an apparent example of this type of pathway degeneration (Zufall and Rausher, 2004). *I. quamoclit* is a member of a small clade of species, all of which have red or orange flowers pollinated by hummingbirds. In addition to color, flowers of these species exhibit several morphological features characteristic of the bird pollination syndrome (Faegeri and van der Pijl, 1996). Their closest known relatives, *I. ternifolia* and *I. tenuiloba*, exhibit the ancestral bee pollination syndrome, which includes blue/purple flowers. It thus appears that the evolutionary transition from blue to red flowers in the ancestral lineage leading to the *I. quamoclit* group constituted an adaptation to exploit a novel pollinator.

As described above, three different genetic changes in the anthocyanin pathway have been identified in *I. quamoclit*. In flowers, the branching enzyme F3'H has been significantly down-regulated. When transformed using a constitutive promoter into an *Arabidopsis* strain lacking F3'H function, the *F3'H* gene from *I. quamoclit* fails to restore anthocyanin production, whereas the same gene from the blue-flowered *I. purpurea* successfully complements the *Arabidopsis* knockout. This result suggests that the functioning of *I. quamoclit* F3'H *in planta* also is impaired. Finally, as demonstrated both by enzyme assays and by complementation tests in *Arabidopsis*, DFR from *I. quamoclit* has lost the ability to metabolize dihydroquercetin, the precursor of cyanidin.

Although it has not yet been definitively proven, it is likely that any one of these changes is sufficient to inactivate the cyanidin branch of the anthocyanin pathway in *I. quamoclit*, thus forcing flux down the pelargonidin branch to produce red pigments. This conclusion, in turn, suggests that the evolutionary transition from blue to red flowers was caused by one of these genetic changes, while the other two represent subsequent degeneration of that branch of the pathway. Moreover, it seems unlikely that species in the *I. quamoclit* group would be able to reinstate a functional cyanidin branch because it would require multiple simultaneous mutations. If in the future ecological conditions make it advantageous for one of these species to utilize bees rather than hummingbirds as pollinators, prior adaptation and accompanying pathway degeneration may make this nearly impossible. It will be of interest to determine whether the other species listed in Table 7.1 have similarly undergone pathway degeneration, and thus have a reduced evolutionary potential.

## 7. CONCLUSIONS

Several important themes regarding the evolution of flavonoids and their genes emerge from the above considerations. First, the flavonoid metabolic network has been and continues to be evolutionarily very fluid, both in terms of its structure and of its regulation. Gene duplications have been co-opted repeatedly for the production of novel enzymes that produce novel flavonoids. In many cases (e.g., stilbene synthases, glutathione *S*-transferases, and probably isoflavone synthase), enzymes with the same function have been recruited independently two or more times. While this creative assembly of the network has been occurring constantly since plants colonized land, loss of components of the network apparently also has occurred frequently (e.g., parallel inactivation of pelargonidin, cyanidin, and/or delphinidin branches in different taxa, perhaps loss of ability to produce isoflavonoids). These two processes, working in tandem, are in large part responsible for the enormous variation in types of flavonoids produced by different plants.

A second important theme is that pleiotropy is likely to play a major role in determining which flavonoid genes participate in adaptive evolutionary change and how rapidly those genes evolve. While evolutionary biologists have long recognized that pleiotropic costs theoretically can constrain adaptive evolutionary change, it has been difficult to study this phenomenon empirically. The flavonoid pathway holds great promise as a model system for examining the evolutionary consequences of pleiotropy because in several plant groups there have been repeated parallel instances of shifts in flower color. Nature thus has provided us with a natural experiment from which we should be able to determine whether evolution has repeatedly used the same genes to create similar phenotypes, and if so whether it has done so because changes in those genes incur relatively small pleiotropic fitness decrements.

A final theme emerging from this review is the importance of loss-of-function mutations in adaptive evolutionary change. Often biologists tend to think of adaptation as a constructive process that adds novel characters to a preexisting phenotype. As should be clear, the assembly of the flavonoid pathway reflects just this type of process and demonstrates the importance of gene duplication and neofunctionalization in creating new characters. However, the numerous examples described above of inactivation of branches of the flavonoid pathway suggest that evolutionary change often involve destructive processes. In a few cases, such as the evolution of red flowers in hummingbird-pollinated *Ipomoea*, it is clear that pathway inactivation was the mechanism of natural selection used to produce an adaptive phenotype. In most other cases, however, it remains to be determined whether pathway inactivation arises from fixation of neutral inactivating mutations by genetic drift when certain flavonoids are no longer needed or whether selection actually turns off the pathway. In either case, however, loss of function is likely to constrain the possible directions in which future evolutionary change can occur.

## 8. REFERENCES

- Alfenito, M. R., Souer, E., Goodman, C. D., Buell, R., Mol, J., Koes, R., and Walbot, V., 1998, Functional complementation of anthocyanin sequestration in the vacuole by widely divergent glutathione S-transferases, *Plant Cell* **10**: 1135-1149.
- Anhut, S., Zinsmeister, H. D., Mues, R., Barz, W., Mackenbrock, K., Koster, J., and Markham, K. R., 1984, The first identification of isoflavones from a bryophyte, *Phytochem* **23**: 1073-1075.
- Baker, M. E., and Blasco, R., 1992, Expansion of the mammalian 3 $\beta$ -hydroxysteroid dehydrogenase/plant dihydroflavonols reductase superfamily to include bacterial cholesterol dehydrogenase, a bacterial UDP-galactose-4-epimerase and open reading frames in vaccinia virus and fish lymphocystis disease virus, *FEBS Lett* **301**: 89-93.
- Barrier, M., Robichaux, R. H., and Purugganan, M. D., 2001, Accelerated regulatory gene evolution in an adaptive radiation, *Proc Natl Acad Sci USA* **98**: 10208-10213.
- Bohm, B. A., 1975, Chalcones, aurones and dihydrochalcones, in *The Flavonoids*, Harborne, J. B., Mabry, T. J., and Mabry, H., eds. Academic Press, New York, pp. 442-504.
- Britten, R., and Davidson, E., 1969, Gene regulation for higher cells: A theory, *Science* **165**: 349-357.
- Britten, R., and Davidson, E., 1971, Repetitive and non-repetitive DNA sequences and a speculation on the origins of evolutionary novelty, *Quart Rev Biol* **46**: 111-133.
- Brown, B. A., and Clegg, M. T., 1984, Influence of flower color polymorphism on genetic transmission in a natural population of the common morning glory, *Ipomoea purpurea*, *Evolution* **38**: 796-803.
- Brown, D. E., Rashotte, A. M., Murphy, A. S., Normanly, J., Tague, B. W., Peer, W. A., Taiz, L., and Muday, G. K., 2001, Flavonoids act as negative regulators of auxin transport *in vivo* in *Arabidopsis*, *Plant Physiol* **126**: 524-535.
- Borevitz, J. O., Xia, Y., Blount, J., Dixon, R. A., and Lamb, C., 2000, Activation tagging identifies a conserved MYB regulator of phenylpropanoid biosynthesis, *Plant Cell* **12**: 2383-2394.
- Chang, S.-M., Lu, Y., and Rausher, M. D., 2005, Neutral evolution of the non-binding region of the anthocyanin regulatory gene *Ipmyb1* in *Ipomoea*, *Genetics* **170**: 1967-1978.
- Clegg, M. T., and Durbin, M. L., 2003, Tracing floral adaptations from ecology to molecules, *Nat Rev Gen* **4**: 206-215.
- Coberly, L. S., 2003, The cost of white flowers: pleiotropy and the evolution of flower color, Ph.D. dissertation, Duke University, Durham, NC.
- Coberly, L. C., and Rausher, M. D., 2003, Analysis of a chalcone synthase mutant in *Ipomoea purpurea* reveals a novel function for flavonoids: amelioration of heat stress, *Mol Ecol* **12**: 1113-1124.
- Coe, E. H., McCormick, S., and Modena, S. A., 1981, White pollen in maize, *J Heredity* **72**: 318-320.
- Cone, K. C., Cocciolone, S. M., Burr, F. A., and Burr, B., 1993, Maize anthocyanin regulatory gene *pl* is a duplicate of *cl* that functions in the plant, *Plant Cell* **5**: 1795-1805.
- Crawford, D. L., and Powers, D. A., 1989, Molecular basis of evolutionary adaptation at the lactate dehydrogenase-B locus in the fish *Fundulus heteroclitus*, *Proc Natl Acad Sci USA* **86**: 9365-9369.
- Curir, P., Dolci, M., and Galeotti, F., 2005, A phytoalexin-like flavonol involved in the carnation (*Dianthus caryophyllus*)-*Fusarium oxysporum* f. Sp *dianthi* pathosystem, *J Phytopath* **153**: 65-67.
- Dewick, P. M., 1988, Isoflavonoids, in *The Flavonoids. Advances in Research Since 1980*, Harborne, J. B., ed., Chapman and Hall, London, pp. 125-209.
- Dias, A. P., Braun, E. L., McMullen, M. D., and Grotewold, E., 2003, Recently duplicated maize *R2R3 Myb* genes provide evidence for distinct mechanisms of evolutionary divergence after duplication, *Plant Physiol* **131**: 610-620.
- Dickinson, W., 1991, The evolution of regulatory genes and patterns in *Drosophila*, *Evol Biol* **25**: 127-174.
- Dixon, R. A., and Steele, C. L., 1999, Flavonoids and isoflavonoids—a gold mine for metabolic engineering, *Trends Plant Sci* **10**: 394-400.
- Doebley, J., 1993, Genetics, development and plant evolution, *Curr Opin Genet Dev* **3**: 865-872.
- Doebley, J., and Lukens, L., 1998, Transcriptional regulators and the evolution of plant form, *Plant Cell* **10**: 1075-1082.
- Dong, X., Braun, E.L., and Grotewold, E. (2001). Functional conservation of plant secondary metabolic enzymes revealed by complementation of *Arabidopsis* flavonoid mutants with maize genes. *Plant Physiol*. **127**, 46-57.

- Durbin, M. L., Learn, G. H., Huttley, G. A., and Clegg, M. T., 1995, Evolution of the chalcone synthase gene family in the genus *Ipomoea*, *Proc Natl Acad Sci USA* **92**: 3338-3342.
- Durbin, M. L., Lundy, K. E., Morrell, P. L., Torres-Martinez, C. L., and Clegg, M. T., 2003, Genes that determine flower color: the role of regulatory changes in the evolution of phenotypic adaptations, *Mol Phyl Evol* **29**: 507-518.
- Ennos, R. A., and Clegg, M. T., 1983, Flower color variation in the morning glory, *Ipomoea purpurea*, *J Heredity* **74**: 247-250.
- Epperson, B. K., and Clegg, M. T., 1987, Instability at a flower color locus in the morning glory, *J Heredity* **78**: 346-352.
- Faegri, D.T., and van der Pijl, L., 1966, *The Principles of Pollination Ecology*, Pergamon, Oxford.
- Feeny, P., 1970, Seasonal changes in oak leaf tannins and nutrients as a cause of spring feeding by winter moth caterpillars, *Ecology* **51**: 565-581.
- Fehr, C., and Rausher, M. D., 2004, Effects of variation at the flower-color *A* locus on mating system parameters in *Ipomoea purpurea*, *Mol Ecol* **13**: 1839-1847.
- Fineblum, W. L., and Rausher, M. D., 1997, Do genes influencing floral pigmentation also influence resistance to herbivores and pathogens? The *W* locus in *Ipomoea purpurea*, *Ecology* **78**: 1646-1654.
- Force, A., Lynch, M., Pickett, F. B., Amores, A., Yan, Y., and Postlethwait, J., 1999, Preservation of duplicate genes by complementary, degenerative mutations, *Genetics* **151**: 1531-1545.
- Fukada-Tanaka, S., Hoshino, A., Hisatomi, Y., Habu, Y., Hasebe, M., and Iida, S., 1997, Identification of new chalcone synthase genes for flower pigmentation in the Japanese and common morning glories, *Plant Cell Physiol* **38**: 88-96.
- Gensheimer, M., and Mushegian, A., 2004, Chalcone isomerase family and fold: no longer unique to plants, *Protein Sci* **13**: 540-544.
- Giannasi, D. E., 1988, Flavonoids and evolution in the dicotyledons, in *The Flavonoids. Advances in Research Since 1980*, Harborne, J. B., ed., Chapman and Hall, London, pp. 479-504.
- Gillespie, J. H., 1991, *The Causes of Molecular Evolution*, Oxford Univ. Press, Oxford, UK.
- Goff, S. A., Cone, K. C., and Chandler, V. L., 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 Dev* **6**: 864-875.
- Goff, S. A., Cone, K. C., and Fromm, M. E., 1991, Identification of functional domains in the maize transcriptional activator C1: comparison of wild-type and dominant inhibitor proteins, *Genes Dev* **5**: 298-309.
- Goff, S. A., Klein, T. M., Roth, B. A., Fromm, M. E., Cone, K. C., Radicella, J. P., and Chandler, V. L., 1990, Transactivation of anthocyanin biosynthetic genes following transfer of *B*-regulatory genes into maize tissues, *EMBO J* **9**: 2517-2522.
- Grotewold, E., Sainz, M.B., Tagliani, L., Hernandez, J.M., Bowen, B., and Chandler, V.L., 2000, Identification of the residues in the Myb domain of maize C1 that specify the interaction with the bHLH cofactor *R*. *Proc Natl Acad Sci USA* **97**, 13579-13584.
- Habu, Y., Hisatomi, Y., and Iida, S., 1998, Molecular characterization of the mutable *flaked* allele for flower variegation in the common morning glory, *Plant J* **16**: 371-376.
- Hansen, M. A., Gaut, B. S., Stec, A. O., Fuerstenberg, S. I., Goodman, M. M., Coe, E. H., and Doebley, J. F., 1996, Evolution of anthocyanin biosynthesis in maize kernels: The role of regulatory and enzymatic loci, *Genetics* **143**: 1395-1407.
- Hartl, D. L., and Clark, A. G., 1989, *Principles of Population Genetics, 2<sup>nd</sup> Edition*, Sinauer, Sunderland, MA.
- Helariutta, Y., Kotilainen, M., Elomaa P., Kalkkinen, N., Bremer, K., Teeri, T. H. and Albert, V. A., 1996, Duplication and functional divergence in the chalcone synthase gene family of Asteraceae: Evolution with substrate change and catalytic simplification, *Proc Natl Acad Sci USA* **93**: 9033-9038.
- Herles, C., Braune, A., and Blaut, M., 2004, First bacterial chalcone isomerase isolated from *Eubacterium ramulus*, *Arch Microbiol* **181**: 428-434.
- Hochachka, P. W., and Somero, G. N., 1984, *Biochemical Adaptation*, Princeton Univ. Press, Princeton.
- Holton, T. A., and Cornish, E. C., 1995, Genetics and biochemistry of anthocyanin biosynthesis, *Plant Cell* **7**: 1071-1083.
- Holton, T. A., Brugliera, F., and Tanaka, Y., 1993, Cloning and expression of flavonol synthase from *Petunia hybrid*, *Plant J* **4**: 1003-1010.

- Hoshino, A., Moriga, Y., Choi, J.-D., Saito, N., Toki, K., Tanaka, Y., and Iida, S., 2003, Spontaneous mutations of the flavonoid 3'-hydroxylase gene conferring reddish flowers in the three morning glory species, *Plant Cell Physiol* **44**: 990-1001.
- Hughes, A. L. 1994, The evolution of functionally novel proteins after gene duplication, *Proc Roy Soc Lond B* **256**: 119-124.
- Huits, H.S.M., Gerats, A.G.M., Kreike, M.M., Mol, J.N.M., and Koes, R.E., 1994, Genetic control of dihydroflavonol 4-reductase gene expression in *Petunia hybrida*, *Plant J* **6**: 295-310.
- Inagaki, Y., Johzuka-Hisatomi, Y., Mori, T., Takahashi, S., Hayakawa, Y., Peyachoknagul, S., Ozeki, Y., and Iida, S., 1999, Genomic organization of the genes encoding dihydroflavonols 4-reductase for flower pigmentation in the Japanese and common morning glories, *Gene* **226**: 181-188.
- Jez, J. M., Bowman, M. E., Dixon, R. A., and Noel, J. P., 2000, Structure and mechanism of chalcone isomerase: an evolutionarily unique enzyme in plants, *Nat Struct Biol* **7**: 786-791.
- Jia, L., Clegg, M. T., and Jiang, T., 2003, Excess nonsynonymous substitutions suggest that positive selection episodes operated in the DNA-binding domain evolution of *Arabidopsis* R2R3-MYB genes, *Plant Mol Biol* **52**: 627-642.
- Jia, L., Clegg, M. T., and Jiang, T., 2004, Evolutionary dynamics of the DNA-binding domains in putative R2R3-myb genes identified from rice subspecies *indica* and *japonica* genomes, *Plant Physiol* **134**: 575-585.
- Johnson, E. T., Hankuil, Y., Shin, B., Oh, B.-J., Cheong, H., and Choi, G., 1999, *Cymbidium hybrida* dihydroflavonols 4-reductase does not efficiently reduce dihydrokaempferol to produce orange pelargonidin-type anthocyanins, *Plant J* **19**: 81-85.
- Kemp, M. S., and Burden, R. S., 1986, Phytoalexins and stress metabolites in the sapwood of trees, *Phytochemistry* **25**: 1261-1269.
- King, M.-C., and Wilson, A. C., 1975, Evolution at two levels in humans and chimpanzees, *Science* **188**: 107-116.
- Koes, R.E., Spelt, C.E., and Mol, J.N., 1989, The chalcone synthase multigene family of *Petunia hybrida* (V30): Differential, light-regulated expression during flower development and UV light induction, *Plant Mol Biol* **12**: 213-225.
- Koes, R. E., Quattrocchio, R., and Mol, J. N. M., 1994, The flavonoid biosynthetic pathway in plants: function and evolution, *BioEssays* **16**: 123-132.
- Kroon, A. R., 2004, *Transcription Regulation of the Anthocyanin Pathway in Petunia hybrida*, Ph.D. dissertation, Vrije Universiteit, Amsterdam, The Netherlands.
- Kubasek, W. L., Shirley, B. W., McKillop, A., Goodman, H. M., Briggs, W., and Ausubel, F. M., 1992, Regulation of flavonoid biosynthetic genes in germinating *Arabidopsis* seedlings, *Plant Cell* **4**: 1229-1236.
- Li, W.-H., 1997, *Molecular Evolution*, Sinauer, Sunderland, MA.
- Li, J., Ou-Lee, T. M., Raba, R., Amundson, R. G., and Last, R. L., 1993, *Arabidopsis* flavonoid mutants are hypersensitive to UV-B radiation, *Plant Cell* **5**: 171-179.
- Lloyd, A. M., Walbot, V., and Davis, R. W., 1992, *Arabidopsis* and *Nicotiana* anthocyanin production activated by maize regulators *R* and *Cl*, *Science* **258**: 1773-1775.
- Lois, R., and Buchanan, B. B., 1994, Severe sensitivity to ultraviolet radiation in an *Arabidopsis* mutant deficient in flavonoid accumulation, *Planta* **194**: 504-509.
- Lu, Y., and Rausher, M. D., 2003, Evolutionary rate variation in anthocyanin pathway genes, *Mol Biol Evol* **20**: 1844-1853.
- Ludwig, S., Bowen, B., Beach, L. and Wessler, S., 1990, A regulatory gene as a novel visible marker for maize transformation, *Science* **247**: 449-450.
- Ludwig, S.R., and Wessler, S.R., 1990, Maize *R* gene family: Tissue specific helix-loop-helix proteins, *Cell* **62**: 849-851.
- Mackenzie, P.I., Owens, I. S., Burchell, B., Bock, K. W., Bairoch, A., Belanger, A., FournelGigleux, S., Green, M., Hum, D. W., Iyanagi, T., Lancet, D., Louisot, P., Magdalou, J., Chowdhury, J. R., Ritter, J. K., Schachter, H., Tephly, T. R., Tipton, K. F., and Nebert, D. W., 1997, The UDP glycosyltransferase gene superfamily: Recommended nomenclature update based on evolutionary divergence, *Pharmacogenetics* **7**: 255-269.
- Markham, K. R., 1988, Distribution of flavonoids in the lower plants and its evolutionary significance, in *The Flavonoids. Advances in Research Since 1980*, Harborne, J. B., ed., Chapman and Hall, London, pp. 427-468.

- Martin, C., Prescott, A., Mackay, S., Bartlett, J., and Vrijlandt, E., 1991, Control of anthocyanin biosynthesis in flowers of *Antirrhinum majus*, *Plant J* **1**: 37-49
- Miller, R. E., McDonald, J. A., and Manos, P. S., 2004, Systematics of *Ipomoea* subgenus *Quamoclit* (Convolvulaceae) based on ITS sequence data and a Bayesian phylogenetic analysis, *Am J Bot* **91**: 1208-1218.
- Miller, R. E., Rausher, M. D., and Manos, P. S., 1999, Phylogenetic systematics of *Ipomoea* (Convolvulaceae) based on ITS and *waxy* sequences, *Syst Bot* **24**:209-227.
- Mo, Y., Nagel, C., and Taylor, L. P., 1992, Biochemical complementation of chalcone synthase mutants defines a role for flavonols in functional pollen, *Proc Nat Acad Sci USA* **89**:7213-7217.
- Mojonnier, L., and Rausher, M. D., 1997, Selection on a floral color polymorphism in the common morning glory (*Ipomoea purpurea*): The effects of overdominance in seed size, *Evolution* **51**: 608-614.
- Mol, J., Grotewold, E., and Koes, R., 1998, How genes paint flowers and seeds. *Trends Plant Sci.* **3**, 212-217.
- Moyano, E., Mart'nez-Garcia, J.F. and Martin, C., 1996, Apparent redundancy in *myb* gene function provides gearing for the control of flavonoid biosynthesis in *Antirrhinum* flowers, *Plant Cell* **8**: 1519-1532.
- Nakayama, T., Yonekura-Sakakibara, K., Sato, T., Kikuchi, S., Fukui, Y., Fukuchi-Mizutani, M., Ueda, T., Nakao, M., Tanaka, Y., Kusumi, T., and Nishino, T., 2000, Aureusidin synthase: a polyphenol oxidase homolog responsible for flower coloration, *Science* **290**: 1163-1166.
- Nakayama, T., Sato, T., Fukui, Y., Yonekura-Sakakibara, K., Hayashi, H., Tanaka, Y., Kusumi, T., and Nishino, T., 2001, Specificity analysis and mechanism of aurone synthesis catalyzed by aureusidin synthase, a polyphenol oxidase homolog responsible for flower coloration, *FEBS Lett.* **499**: 107-111.
- Nesi, N., Jond, C., Debeaujon, I., Caboche, M., and Lepiniec, L., 2000, The *TT8* gene encodes a basic helix-loop-helix domain protein required for expression of *DFR* and *BAN* genes in *Arabidopsis* siliques, *Plant Cell* **13**: 2099-2114.
- Nesi, N., Jond, C., Debeaujon, I., Caboche, M., and Lepiniec, L., 2001, The Arabidopsis *TT2* gene encodes an R2R3 MYB domain protein that acts as a key determinant for proanthocyanidin accumulation in developing seed. *Plant Cell* **13**, 2099-2114.
- Newcomb, R. D., Campbell, P. M., Ollis, D. L., Cheah, E., Russell, R. J., and Oakeshott, J. G., 1997, A single amino acid substitution converts a carboxylesterase to an organophosphorus hydrolase and confers insecticide resistance on a blowfly, *Proc Nat Acad Sci USA* **94**: 7464-7468.
- Niemann, G. J., 1988, Distribution and evolution of the flavonoids in gymnosperms, in *The Flavonoids. Advances in Research Since 1980*, Harborne, J. B., ed., Chapman and Hall, London, pp. 469-478.
- Patterson, G. I., Kubo, K. M., Shrover, T. and Chandler, V. L., 1995, Sequences required for paramutation of the maize *b* gene map to a region containing the promoter and upstream sequences, *Genetics* **140**: 1389-1406.
- Pelletier, M. K., Murrell, J. R., and Shirley, B. W., 1997, Characterization of flavonol synthase and leucoanthocyanidins dioxygenase genes in Arabidopsis, *Plant Physiol* **113**: 1437-1445.
- Pelletier, M.K., and Shirley, B.W., 1996, Analysis of flavanone 3-hydroxylase in *Arabidopsis* seedlings. Coordinate regulation with chalcone synthase and chalcone isomerase, *Plant Physiol* **111**: 339-345.
- Peer, W. A., Bandyopadhyay, A., Blakeslee, J. JU., Makam, S. N., Chen, R. J., Masson, P. H., and Murphy, A. S., 2004, Variation in expression and protein localization of the PIN family of auxin efflux facilitator proteins in flavonoid mutants with altered auxin transport in *Arabidopsis thaliana*, *Plant Cell* **16**: 1898-1911.
- Ptashne, M., 1988 How eukaryotic transcriptional activators work, *Nature* **335**: 683-689.
- Purrginton, C. B., and Bergelson, J., 1997, Fitness consequences of genetically engineered herbicide and antibiotic resistance in *Arabidopsis thaliana*, *Genetics* **145**: 807-814.
- Purugganan, M. D., 1998, The molecular evolution of development, *BioEssays* **20**: 700-711.
- Purugganan, M. D., Rounsley, S. D., Schmidt, R. J., and Yanofsky, M. F., 1995, Molecular evolution of flower development: diversification of the plant MADS-box regulatory gene family, *Genetics* **140**: 345-356.
- Purugganan, M. D., and Wessler, S. R., 1994, Molecular evolution of the plant *R* regulatory gene family, *Genetics* **138**: 849-854.
- Quattrocchio, F., Wing, J.F., Leppen, H.T.C., Mol, J.N.M., and Koes, R.E., 1993, Regulatory genes controlling anthocyanin pigmentation are functionally conserved among plant species and have distinct sets of target genes, *Plant Cell* **5**: 1497-1512.

- Quattrocchio, F., Wing, J.F., van der Woude, K., Mol, J.N.M., and Koes, R., 1998, Analysis of bHLH and MYB-domain proteins: Species-specific regulatory differences are caused by divergent evolution of target anthocyanin genes, *Plant J* **13**: 475–488.
- Quattrocchio, F., Wing, J., van der Woude, K., Souer, E., de Vetten, N., Mol, J., and Koes, R., 1999, Molecular analysis of the *anthocyanin2* gene of *Petunia* and its role in the evolution of flower color, *Plant Cell* **11**: 1433-1444.
- Radicella, J. P., Brown D., Tolar, L. A. and Chandler, V. L., 1992, Allelic diversity of the maize *B* regulatory gene: different leader and promoter sequences of two *B* alleles determine distinct tissue specificities of anthocyanin production, *Genes Dev* **6**: 2152-2164.
- Rausher, M.D., Augustine, D., and Vanderkoi, A., 1983, Absence of pollen discounting in a genotype of *Ipomoea purpurea* exhibiting increased selfing, *Evolution* **47**: 1688–1695.
- Rausher, M. D., and Fry, J. D., 1993, Effects of a locus affecting floral pigmentation in *Ipomoea purpurea* on female fitness components, *Genetics* **134**: 1237-1247.
- Rausher, M. D., Miller, R. E., and Tiffin, P., 1999, Patterns of evolutionary rate variation among genes of the anthocyanin biosynthetic pathway, *Mol Biol Evol* **16**: 266-274.
- Remington, D. L., and Purugganan, M. D., 2002, *Gal* homologues in the Hawaiian Silversword alliance (Asteraceae-Madiinae): molecular evolution of growth regulators in a rapidly diversifying plant lineage, *Mol Biol Evol* **19**: 1563-1574.
- Ryan, K. G., Swinny, E. E., Winefield, C., and Markham, K. R., 2001, Flavonoids and UV photoprotection in *Arabidopsis* mutants, *Zeitschrift für Naturforschung (Section C)* **56**: 745-754.
- Ryan, K. G., Swinny, E. E., Markham, K. R., and Winefield, C., 2002, Flavonoid gene expression and UV photoprotection in transgenic and mutant *Petunia* leaves, *Phytochem* **2002**: 59:23-32.
- Sainz, M. B., Goff, S. A., and Chandler, V. L., 1997, Extensive mutagenesis of a transcriptional activation domain identifies single hydrophobic and acidic amino acids important for activation *in vivo*, *Mol Cell Biol* **17**: 115-122.
- Savolainen, V., and Chase, M. W., 2003, A decade of progress in plant molecular phylogenetics, *Trends Genet* **19**: 717- 724.
- Scogin, R., and Freeman, C. E., 1987, Floral anthocyanins of the genus *Penstemon*: Correlations with taxonomy and pollination, *Bioch Syst Ecol* **15**: 355-360.
- Shirley, B. W., 1996, Flavonoid biosynthesis: “new” functions for an “old” pathway, *Trends Plant Sci* **1**: 377-382.
- Stafford, H. A., 1991, Flavonoid evolution: an enzymatic approach, *Plant Physiol* **96**: 680-685.
- Tiffin, P., Miller, R. E., and Rausher, M. D., 1998, Control of expression patterns of anthocyanin structural genes by two loci in the common morning glory, *Genes Gen Syst* **73**: 105-110.
- Timberlake, C. F., and Bridle, P., 1980, The anthocyanins, in *The Flavonoids*, Harborne, J. B., Mabry, T. J., and Mabry, H., eds., Academic Press, New York, pp. 214-266.
- Ting, C. T., Tsaur, S. C., Wu, M. L., and Wu, C.-I., 1998, A rapidly evolving homeobox at the site of a hybrid sterility gene, *Science* **282**: 1501-1504.
- Tropf, S., Lanz, T., Rensing, S. A., Schroeder, J., and Schroeder, G., 1994, Evidence that stilbene synthases have developed from chalcone synthases several times in the course of evolution, *J Mol Evol* **38**: 610-618.
- Van der Meer, I. M., Stam, M., van Tunen, A. J., Mol, J. N. M., and Stuitje, A. R., 1992, Antisense inhibition of flavonoid biosynthesis in *Petunia* anthers results in male sterility, *Plant Cell* **4**: 253-262.
- van Houwelingen, A., Souer, E., Spelt, C., Kloos, D., Mol, J., and Koes, R., 1998, Analysis of flower pigmentation mutants generated by random transposon mutagenesis in *Petunia hybrida*, *Plant J* **13**: 39–50.
- Van Tunen, A. J., Koes, R. E., Spelt, C. E., van der Krol, A. R., Stuitje, A. R., and Mol, J. N. M., 1988, Cloning of the two chalcone flavanone isomerase genes from *Petunia hybrida*: coordinate, light-regulated and differential expression of flavonoid genes, *EMBO J* **7**: 1257-1263.
- Verwoert, I. I. G. S., Verbree, E. C., Van der Linden, K. H., Nijkamp, H. J. J., and Stuitje, A. R., 1992, Cloning, nucleotide sequence and expression of the *Escherichia coli fabD* gene, encoding malonyl coenzyme A-acyl carrier protein transacylase, *J Bact* **174**: 2851-2857.
- Watt, W. B., 1977, Adaptation at specific loci I. Natural selection on phosphoglucose isomerase of *Colias* butterflies—biochemical and population aspects, *Genetics* **87**: 177–194.

- Watt, W.B., 1983, Adaptation at specific loci. 2. Demographic and biochemical-elements in the maintenance of the *Colias* PGI polymorphism, *Genetics* **103**: 691-724.
- Webby, R. F., Markham, K. R., and Smith, R. I. L., 1996, Chemotypes of the Antarctic moss *Bryum algens* delineated by their flavonoid constituents, *Bioch Syst Ecol* **24**: 469-475.
- Williams, C. E., and Grotewold, E., 1997, Differences between plant and animal Myb domains are fundamental for DNA binding activity and chimeric Myb domains have novel DNA-binding specificities, *J Biol Chem* **272**: 563-571.
- Williams, C. A., and Harborne, J. B., 1988, Distribution and evolution of flavonoids in the monocotyledons, in *The Flavonoids. Advances in Research Since 1980*, Harborne, J. B., ed., Chapman and Hall, London, pp. 505-524.
- Wilson, P., Castellanos, M. C., Hogue, J. N., Thomson, J. D. and Armbruster, W. S., 2004, A multivariate search for pollination syndromes among penstemons, *Oikos* **104**: 345-361.
- Winkel-Shirley, B., 2001, Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology and biotechnology, *Plant Physiol* **126**: 485-492.
- Wollenweber, E., and Schneider, H., 2000, Lipophilic exudates of Pteridaceae—chemistry and chemotaxonomy, *Bioch Syst Ecol* **28**: 751-777.
- Wray, G.A., Hahn, M. W., Abouheif, E., Balhoff, J. P., Pizer, M., Rockman, M. V., and Romano, L. A., 2003, The evolution of transcriptional regulation in eukaryotes, *Mol Biol Evol* **20**: 1377- 1419.
- Yang, J., Gu, H., and Yang, Z., 2004, Likelihood analysis of the chalcone synthase genes suggests the role of positive selection in morning glories (*Ipomoea*), *J Mol Evol* **58**: 54-63.
- Yang, Z., Nielsen, R., Goldman, N., and Pedersen, A. K., 2000, Codon-substitution models for heterogeneous selection pressure at amino acid sites, *Genetics* **155**: 431-449.
- Yokoyama, S., Starmer, W. T., and Yokoyama, R., 1993, Paralogous origin of the red- and green-sensitive visual pigment genes in vertebrates, *Mol Biol Evol* **10**: 527-538.
- Zufall, R. A., 2003, *Evolution of Red Flowers in Ipomoea*, Ph.D. dissertation, Duke University, Durham, NC.
- Zufall, R. A., and Rausher, M. D., 2003, The genetic basis of a flower-color polymorphism in the common morning glory, *Ipomoea purpurea*, *J Heredity* **94**: 442-448.
- Zufall, R. A., and Rausher, M. D., 2004, Genetic changes associated with floral adaptation restrict future evolutionary potential, *Nature* **428**: 847-850.