New Phytologist

Research

Predictable patterns of constraint among anthocyanin-regulating transcription factors in *Ipomoea*

Matthew A. Streisfeld[1], Danya Liu[2] and Mark D. Rausher[2]

[1]Center for Ecology and Evolutionary Biology, 5289 University of Oregon, Eugene, OR 9703-5289, USA; [2]Department of Biology, Duke University, Box 90338, Durham, NC 27708, USA

Author for correspondence
Matthew A. Streisfeld
Tel.: 541 346 4816
Email: mstreis@uoregon.edu

Received: *27 October 2010*
Accepted: *17 January 2011*

New Phytologist (2011)
doi: 10.1111/j.1469-8137.2011.03671.x

Key words: anthocyanin, constraint, flower colour evolution, molecular evolution, pleiotropy, transcription factor.

Summary

- Transcription factors (TFs) may play a central role in plant morphological evolution. Variation in the nonsynonymous to synonymous nucleotide substitution rate (d*N*/d*S*) ratio among TFs can be attributed to either differences in constraint or the frequency of adaptive substitution. However, the relative contribution of these forces to the variation in d*N*/d*S* ratios is unknown.
- We synthesize current and previous results comparing the variation in d*N*/d*S* ratios among members of the MYB–bHLH–WDR complex of TFs that regulates floral anthocyanin pigmentation in *Ipomoea*.
- Low values of d*N*/d*S* in a *WDR* gene are the result of exceptionally strong purifying selection, with no evidence of positive selection. *bHLH* and *MYB* genes also fail to show evidence for positive selection, but have higher d*N*/d*S* ratios, indicating reduced selective constraint.
- Differences in constraint are consistent with expectations based on the intrinsic features and regulatory network properties among these proteins. Significantly elevated d*N*/d*S* ratios in the *MYB* gene suggest that mutations experience reduced magnitudes of deleterious pleiotropy compared with the rest of the complex. Although reduced constraint may account for the observation that *Myb* mutations disproportionately contribute to differences in floral pigmentation, the lack of detectable positive selection in any of these TF proteins suggests that amino acid substitutions contribute little to flower colour evolution.

Introduction

The combinatorial interaction of transcription factor (TF) proteins with their DNA-binding targets specifically determines the spatial and temporal patterning of eukaryotic gene expression. Moreover, because of their role in controlling developmental processes, it has been suggested that plant TFs play a central role in the evolution of plant morphology (Doebley, 1993; Doebley & Lukens, 1998; Shepard & Purugganan, 2002; Martin *et al.*, 2010). One observation consistent with this hypothesis is that many plant TF genes exhibit significantly higher nonsynonymous to synonymous nucleotide substitution rate (d*N*/d*S*) ratios than the genes that they regulate (Purugganan & Wessler, 1994; Purugganan *et al.*, 1995; Rausher *et al.*, 1999; Langercrantz & Axelsson, 2000; Barrier *et al.*, 2001; Lukens & Doebley, 2001; Remington & Purugganan, 2002), which could reflect a greater frequency of adaptive substitutions (Rausher *et al.*, 2008). Another observation is that some amino acid substitutions in TFs have been shown to alter morphology (Flowers *et al.*, 2009). However, the fraction of substitutions in TFs that contribute to morphological evolution is currently unknown, and it may be that most substitutions in TFs are a result of genetic drift rather than positive selection for developmental alteration. If true, the higher d*N*/d*S* ratios of TF evolution would reflect reduced constraint rather than increased adaptive evolution.

Previous work on the R2R3-MYB (MYB) and basic helix-loop-helix (bHLH) TFs that regulate the anthocyanin biosynthetic pathway has shown that high d*N*/d*S* ratios result exclusively from relaxed constraint in the genus *Ipomoea* (Chang *et al.*, 2005; Streisfeld & Rausher, 2007). Positive selection was not detected among species for either the MYB or bHLH regulators, indicating that

nonsynonymous substitutions in these TFs do not reflect adaptive evolution of flower colour or any other trait. One objective of this study was to examine whether substitutions in a third anthocyanin-regulating TF may have been caused by adaptive evolution.

A second objective was to examine whether variation in d*N*/d*S* among the anthocyanin TFs can be ascribed to differences in constraint that are predicted from a knowledge of their functional properties. It is generally believed that differences in the magnitude of mutational pleiotropy among genes contribute to variation in constraint, and thus to differences in d*N*/d*S* (Kimura, 1977; Li, 1997). However, only in a very few cases has a link between pleiotropy and constraint been established. As described below, the three classes of anthocyanin TF that combine to form the transcription activation complex differ in the breadth of their functional activities. This difference suggests that mutations in TFs with broader activity are likely to experience greater deleterious pleiotropy and thus greater evolutionary constraint. In this study, we test whether this expectation is upheld.

Among angiosperms, evolutionary transitions in flower colour are common, and variation in the intensity and type of anthocyanin pigments produced is usually responsible for this diversity (Rausher, 2008). The biochemical and regulatory control of this pigmentation pathway is extremely well characterized (Holton & Cornish, 1995; Koes *et al.*, 2005; Grotewold, 2006). Detailed investigation from *Arabidopsis*, maize and *Petunia* has provided a nearly complete description of an interacting complex of TF proteins that regulates not only anthocyanin pigmentation, but also many other developmental processes involved in epidermal cell differentiation (Mol *et al.*, 1998; Broun, 2005; Koes *et al.*, 2005; Quattrocchio *et al.*, 2006). This complex is conserved across all species that have been examined, and is made up of proteins encoded by members from three large gene families: the R2R3-MYB (MYB), basic helix-loop-helix (bHLH) and WD40-repeat (WDR) families (Broun, 2005; Koes *et al.*, 2005; Ramsay & Glover, 2005). Hereafter, we refer to this complex as the MBW complex. Expression of the anthocyanin pathway genes is controlled coordinately at the level of transcription via the interaction of the MBW complex with the *cis*-regulatory elements of multiple enzyme-encoding pathway genes. At present, the MBW complex represents one of the best studied and nearly complete regulatory networks known in plants.

In addition to the conserved role in regulating the expression of anthocyanin pigmentation, the MBW complex also controls alternative developmental pathways that affect various aspects of epidermal cell fate, including the initiation of trichomes, alteration of vacuolar pH, production of root hairs, seed coat mucilage and proanthocyanidin synthesis in seeds (Broun, 2005; Koes *et al.*, 2005; Ramsay & Glover, 2005). From mutant studies in *Arabidopsis*, it has been shown that a single WDR protein (TTG-1) is central for the differentiation of all of these cell types. Studies of protein interactions using the yeast two-hybrid system have shown that this same WDR protein has the capacity to interact with multiple bHLH (e.g. GL3, EGL3, TT8) and MYB (e.g. TT2, PAP1, PAP2, WER, GL1; Broun, 2005) proteins. Cell-type specificity appears to be the result of the particular MYB protein that joins the complex (Ramsay & Glover, 2005). For example, different copies of MYB proteins are known to confer tissue- and cell-specific anthocyanin regulation in several species, whereas other copies are responsible for trichome initiation and proanthocyanidin synthesis in seeds (Broun, 2005; Koes *et al.*, 2005; Quattrocchio *et al.*, 2006; Schwinn *et al.*, 2006; Morita *et al.*, 2006; Gonzalez *et al.*, 2008). This pattern suggests that duplication of MYB proteins has allowed different paralogues to become specialized in their expression and developmental control of specific traits. By contrast, this specialization does not appear to have occurred for WDR proteins, whose expression domain is substantially broader than that of the MYB proteins (Ramsay & Glover, 2005; Koes *et al.*, 2005; Morita *et al.*, 2006). Moreover, although bHLH proteins appear to have partially restricted expression domains and only control a subset of WDR's functions, they also have a propensity to form homo- and hetero-dimers, which may place additional constraints on their evolutionary potential.

Based on a knowledge from model organisms of the properties of the three classes of protein that comprise this regulatory complex, we generated a series of evolutionary predictions about the relative differences in constraint among the members of the MBW complex. For example, because of its widespread expression, greater number of protein–protein interactions and regulatory control of multiple aspects of epidermal cell differentiation, WDR proteins would be expected to experience greater constraint than bHLH or MYB proteins (Ramsay & Glover, 2005; Koes *et al.*, 2005). However, as a consequence of the high degree of tissue specificity associated with multiple MYB copies, MYB proteins that regulate floral anthocyanins are expected to experience substantially lower constraint than WDR proteins. Finally, intermediate levels of control and partially restricted expression suggest that constraints on bHLH proteins lie somewhere between those of WDR and MYB proteins. These expectations would be congruent with the observation that altered function or expression of MYB proteins has been responsible for all of the documented evolutionary transitions in floral colour that involve TFs (Streisfeld & Rausher, 2010). Presumably, because of the fewer constraints on MYB proteins, mutations – whether regulatory or in coding sequences – in the genes that encode them are expected to experience less deleterious pleiotropy and thus have a greater chance of being fixed than

mutations in other TFs (Streisfeld & Rausher, 2010). Interestingly, however, no evidence for positive selection in the MYB protein has been detected previously among species of *Ipomoea* (Chang *et al.*, 2005).

Because the function of the MBW complex is conserved among all species investigated so far (Ramsay & Glover, 2005; Koes *et al.*, 2005), these are general predictions that should be applicable to all species. However, to date, most research into this complex has focused on the molecular functions and interactions among its members. Thus, little is known about the molecular evolution of the MBW complex. The current study uses the morning glory genus (*Ipomoea*), which varies widely in the type and intensity of floral anthocyanin pigmentation, as a test case to address these predictions. Our previous studies have examined variation in constraint among different copies of MYB and bHLH proteins known to regulate the expression of floral anthocyanins in *Ipomoea* species with diverse flower colours (Chang *et al.*, 2005; Streisfeld & Rausher, 2007). Both projects demonstrated elevated d*N*/d*S* ratios in these proteins compared with the genes they regulate, but failed to detect evidence for adaptive substitution. Both analyses reported higher d*N*/d*S* ratios for MYB compared with bHLH proteins, in accordance with our predictions of differential constraint, but neither study evaluated whether these ratios differed statistically. Here, we extend our analysis to document d*N*/d*S* ratios for the WDR gene and determine whether evidence exists for positive selection in this gene, as would be expected if substitutions contributed to flower colour variation. We then compare statistically the d*N*/d*S* ratios for the three TFs to determine whether they differ in constraint as predicted.

Materials and Methods

Assessment of constraint

Variation in d*N*/d*S* ratios among genes has often been interpreted as reflecting differences in constraint (Li, 1997; Yang & Nielsen, 2000; Yang & Bielawski, 2000). This interpretation is natural under the Neutralist view of evolution, in which adaptive substitutions are rare and virtually all nondeleterious, nonsynonymous mutations are neutral (Kimura, 1968, 1983). Under this view, differences among genes in d*N*/d*S* ratios reflect differences in the probability that a nonsynonymous mutation is deleterious, which is a reasonable definition of constraint. Thus, in a comparison of two genes that vary in their d*N*/d*S* ratios, the gene with the smaller ratio would be expected to demonstrate tighter constraint because a larger proportion of the nonsynonymous mutations experienced in that gene are deleterious. However, differences in d*N*/d*S* among genes may also reflect differences in the frequency of adaptive substitution, even for genes with d*N*/d*S* < 1. The emerging consensus that a substantial fraction of amino acid substitutions are adaptive (Smith & Eyre-Walker, 2002; Charlesworth & Eyre-Walker, 2006; Begun *et al.*, 2007; Halligan *et al.*, 2010) indicates that this possibility must be taken seriously, and that equating differences in d*N*/d*S* ratios to differences in constraint is only valid if it is shown first that the frequencies of adaptive substitutions do not differ among genes. Here, we adopt this strict definition of constraint. In particular, we only infer differences in constraint based on differences in d*N*/d*S* ratios if the frequency of positive selection does not appear to differ among genes.

WDR sequencing

The WDR proteins represent a highly diverse superfamily of regulatory proteins that are involved in multiple cellular functions and biochemical processes. The diagnostic feature of members of the WDR gene family is a series of two to seven imperfect tandem repeats of *c.* 40 amino acids each that together are believed to be important for interaction with other proteins (van Nocker & Ludwig, 2003). In *Arabidopsis*, there are an estimated 239 proteins with at least four recognizable WD motifs (van Nocker & Ludwig, 2003). Screening of a floral cDNA library in *Ipomoea nil* and *I. purpurea* revealed two *WDR* copies with sequence similarity to previously identified floral anthocyanin regulators (Morita *et al.*, 2006). However, a frame shift mutation in one of these copies (*WDR1*) has been shown to eliminate the expression of multiple anthocyanin pathway enzymes and has resulted in the absence of anthocyanin pigmentation in flowers (Morita *et al.*, 2006). In addition, putative orthologues of *WDR1* from *Arabidopsis* (TTG-1) and *Petunia* (AN11) have been shown to be involved in the regulatory network controlling anthocyanin synthesis (de Vetten *et al.*, 1997; Broun, 2005). Therefore, we focused our efforts on *WDR1* and sequenced the full-length coding region from 12 additional morning glory species across the tribe Ipomoeeae (Convolvulaceae) (Table 1). The species that we selected broadly encompass the range of genetic and flower colour diversity sampled in our previous studies of the molecular evolution of anthocyanin-regulating TFs (Table 1; Fig. 1), and were obtained from our glasshouse collections (Chang *et al.*, 2005; Streisfeld & Rausher, 2007). Total RNA was extracted from floral buds using the Qiagen RNeasy kit (Valencia, CA, USA). First-strand cDNA synthesis was performed using 500 ng RNA and the Invitrogen M-MLV reverse transcriptase (Carlsbad, CA, USA), according to the manufacturer's specifications. Sets of overlapping, degenerate PCR primers were designed from conserved regions of the *WDR1* sequence from closely related species (Supporting Information Table S1). Following PCR, we directly sequenced amplicons on ABI 3730 sequencers using Big Dye technology.

Table 1 A list of the species sequenced at the *WDR1* gene and their flower colour characteristics

Species name	Flower colour	Flower intensity
Ipomoea alba	White	na
I. coccinea	Red	Intense
I. hochstetteri	Blue	Intermediate
I. horsfalliae	Red	Intense
I. muricata	White and purple	Pale
*I. nil**	Blue	Intermediate
I. obscura	Yellow and white	na
*I purpurea**	Purple	Intermediate
I. quamoclit	Red	Intense
I. tricolor	Blue, white and yellow	Intermediate
I. trifida	White and purple	Pale
I. udeana	Red	Intense
I. violacea	White	na
Stichtocardia tilifolia	Purple	Intermediate

Sequences have been deposited in Genbank under accession numbers: HQ875561–HQ875572.
*Sequences were publicly available under accession numbers AB232777 (*I. purpurea*) and AB232779 (*I. nil*).
na, not applicable.

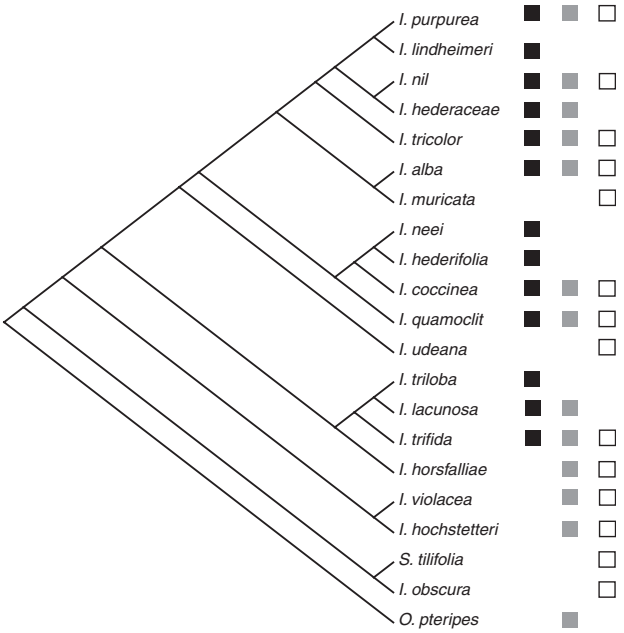

Fig. 1 Representation of the evolutionary relationships among all species used to compare the nonsynonymous to synonymous nucleotide substitution rate (d*N*/d*S*) ratios in the four anthocyanin-regulating members of the MYB–bHLH–WDR (MBW) complex. Symbols next to each species denote which gene(s) were sequenced. Black boxes, *MYB* (Chang *et al.*, 2005); grey boxes, *bHLH1* and *bHLH2* (Streisfeld & Rausher, 2007); white boxes, *WDR1* (this study). Data on relationships taken from Miller *et al.* (1999), Manos *et al.* (2001) or were kindly provided by R. Miller (pers. comm.).

Data analysis

Overlapping sequence reads from each species were combined to generate full-length sequences of the *WDR1* gene. Sequences were aligned manually using BioEdit and codon-based maximum likelihood models of d*N*/d*S* variation were conducted using the codeml program of PAML version 4.2 (Yang, 1997; Nielsen & Yang, 1998). Specifically, we first tested whether the d*N*/d*S* ratio (referred to as ω in PAML) varied across lineages. The lineage-based models assume a single ω for all codon sites. A previously established phylogeny of these species was used to generate the topology for these analyses (Fig. 1; Miller *et al.*, 1999; Manos *et al.*, 2001). We compared estimates from a model in which a single ω was constrained on all branches of the phylogeny with a free-ratio model, in which ω was allowed to vary independently on each branch. We used the standard likelihood ratio framework to evaluate whether the free-ratio model fit the data significantly better than the single-ω model. Specifically, two times the difference in log-likelihood scores from each model was compared with a χ^2 distribution with degrees of freedom (df) equal to the difference in estimated parameters between the models.

We next evaluated whether particular amino acid residues were subject to positive selection using codon site models, where the mean value of ω was held constant on each branch, but ω was estimated independently across sites. For these analyses, a series of nested models was run, and likelihood ratio tests were conducted to test for the presence of positive selection on classes of amino acids. Model M0 represents the null model that all sites have a single ω value. Model M1a estimates two different site classes that are either constrained ($0 < \omega < 1$) or evolving neutrally ($\omega = 1$). Model M2a is identical to M1a, except that it adds a third site class of amino acids that may be positively selected ($\omega > 1$). Models M7 and M8 both model ω according to a beta distribution, with the parameters of the distribution estimated from the data. Model M8 differs from M7 in that it contains an extra site category with $\omega \geq 1$. Likelihood ratio tests were constructed between Models M1a and M2a and between M7 and M8. If Models M2a and/or M8 fit the data significantly better than Models M1a or M7, respectively, and some of the ω values for individual amino acids were estimated according to an empirical Bayes procedure to be significantly greater than unity, positive selection was indicated.

Comparison of d*N*/d*S* variation among transcriptional regulators

To determine whether d*N*/d*S* differed as predicted among the floral anthocyanin-regulating TFs in *Ipomoea*, we compared ω values from Model M0 of PAML among four of these genes: *WDR1*, *bHLH1*, *bHLH2* and *MYB*. In addition to the *WDR1* sequences reported here, we used information from Chang *et al.* (2005) and Streisfeld & Rausher (2007) to obtain sequences for the *MYB*, *bHLH1*

and *bHLH2* genes (Fig. 1). Moreover, because previous investigations of TF evolution in plants have revealed that they often consist of different functional domains that evolve at different rates (Purugganan & Wessler, 1994; Purugganan *et al.*, 1995; Langercrantz & Axelsson, 2000), we also compared d*N*/d*S* ratios among specific domains in these genes. For example, in the anthocyanin bHLH and MYB factors, conserved and variable domains exist (Purugganan & Wessler, 1994; Chang *et al.*, 2005; Streisfeld & Rausher, 2007). The conserved domains are responsible for DNA-binding and protein–protein interactions, and the variable domains are thought to contain transcriptional activation regions (Goff *et al.*, 1991, 1992; Gong *et al.*, 1999; Grotewold *et al.*, 2000; Kroon, 2004), However, there is no information about whether such domains exist in the WDR1 protein.

The study of Chang *et al.* (2005) only reported sequence from the *MYB* gene's conserved domain for six species, but reported variable domain sequence for 13 species. To complete this dataset, we sequenced the *MYB* conserved domain from the remaining seven species using the methods and primers described in Chang *et al.* (2005).

Because a slightly different set of species was analysed in each dataset, we used a bootstrapping approach to determine whether ω values differed significantly among the genes and domains. Specifically, we obtained 1000 bootstrapped datasets for each gene using PAML, where resampling was performed across codons. For each dataset, we then ran Model M0 of PAML to obtain a distribution of ω values. To determine whether the estimated ω values from any two genes (or domains) differed from each other, we then calculated the probability of overlap between the bootstrapped distributions of ω values from the two genes. More specifically, let the null hypothesis be that gene 1 has an ω value equal to gene 2, whereas the alternative hypothesis, based on our *a priori* expectations, is that the value for gene 1 is less than for gene 2 (a one-tailed test). For each of the 1000 bootstrapped ω values estimated for gene 1, we determined the proportion of ω values from gene 2 that were less than or equal to the value in gene 1 (this proportion represents cases consistent with our null hypothesis, but inconsistent with the one-tailed alternative hypothesis). This proportion was then averaged over all gene 1 values to obtain an estimate of the probability that a randomly chosen value from the gene 1 distribution was greater than or equal to a randomly chosen value from the gene 2 distribution. If this probability was less than 0.05, we rejected the null hypothesis and concluded that the ω value for gene 1 was significantly less than for gene 2.

In order to determine whether particular domains within each of the genes could account for the observed differences in ω across genes, we partitioned the *MYB*, *bHLH1* and *bHLH2* gene sequences into their known conserved and variable domains according to Chang *et al.* (2005) and Streisfeld & Rausher (2007). The observed amino acid sequence of the WDR1 protein is highly conserved across the entire protein (see the Results section), so we did not partition *WDR1* sequences. We then repeated the bootstrapping procedure described above to obtain distributions for each domain and determined whether individual domains demonstrated significant rate variation both within and among genes.

Results

The *WDR1* gene from these *Ipomoea* species consists of a single exon of 1032 bp that is highly conserved across all of the species sequenced (ω = 0.036). In addition, there is only a single amino acid insertion/deletion event across all of the sequences. By contrast, indels are common in the *MYB* and *bHLH* sequences (seven in *bHLH1*, *c.* 25 each in *bHLH2* and *MYB*; Chang *et al.*, 2005; Streisfeld & Rausher, 2007), suggesting that *WDR1* is under greater constraint. Finally, there were no indications in any of the sequences of premature stop codons or frame shift mutations that would appear to cause nonfunctional proteins.

Patterns of evolution in WDR1

To determine the selective forces operating on the WDR1 protein, we first examined whether there was heterogeneity in d*N*/d*S* ratios among lineages by performing branch tests. Likelihood ratio tests comparing single-ratio and free-ratio models do not differ significantly (Λ = 21.81; df = 24; $P > 0.5$), providing no reason to reject the null hypothesis that the gene has evolved with similar constraint along different lineages.

We then compared models that evaluate the evolutionary forces acting on individual codon sites. Models M1a, M2a, M7 and M8 of PAML all result in nearly identical

Table 2 Results of codon-based models of molecular evolution for the *WDR1* gene in *Ipomoea*

Model	l[a]	ω[b]	Parameters[c]
M0 (single ω)	2992.90	0.036	ω = 0.03562
M1a (nearly neutral)	2963.82	0.043	ω_0 = 0.012 (p_0 = 0.969)
			ω_1 = 1.000 (p_1 = 0.031)
M2a (selection)	2963.82	0.043	ω_0 = 0.012 (p_0 = 0.969)
			ω_1 = 1.000 (p_1 = 0.019)
			ω_2 = 1.000 (p_2 = 0.012)
M7 (β)	2965.52	0.041	α = 0.032; β = 0.627
M8 (β + ω)	2964.18	0.040	α = 0.008; β = 206
			p_0 = 0.976
			p_1 = 0.024; ω_1 = 1.000

[a]Negative log-likelihood of the data.
[b]Mean d*N*/d*S* ratio for the entire gene.
[c]p_i, the proportion of codons that fall into each category.

log-likelihood scores (Table 2), providing no reason to reject Model M1a that classifies sites as either highly constrained or evolving neutrally. Moreover, Model M1a indicates that 97% of sites are located in the highly constrained site class, with ω estimated as 0.012, suggesting that all but a few amino acid sites experience exceptionally strong purifying selection. Furthermore, for both of the selection models (M2a and M8), estimates of ω are not greater than unity for any site class, again providing no evidence for positive selection in the *WDR1* gene. Empirical Bayes estimation of individual amino acid sites belonging to the 'selected' site class in model M8 indicates that only four amino acid sites throughout the entire protein have an estimated $\omega > 1$. For these sites, the maximum ω is 1.33 ± 0.47, and none of these sites is estimated to have a posterior probability of > 0.79 of belonging to the rapidly evolving site class. These results provide consistently strong evidence that the WDR1 protein has experienced substantial selective constraint and little evidence for positive selection at any amino acid site.

Variation in constraint among transcriptional regulators

Estimates of the d*N*/d*S* ratio for each of the genes are presented in Table 3. Because none of the TFs exhibit evidence of positive selection, it is appropriate to use d*N*/d*S* ratios to compare levels of constraint among these proteins (see the Methods section). Comparison of ω values among the four genes strongly supports our predictions of differential constraint of this regulatory complex. Among 1000 bootstrapped replicates, all estimates of ω in *WDR1* are less than observations of ω from the other three genes (Fig 2; Table 4). By contrast, ω values for the *MYB* gene are significantly higher than the other three genes, which supports our prediction that MYB proteins are functionally less constrained than the other anthocyanin-regulating TFs. Finally, as predicted, *bHLH1* and *bHLH2* have ω values that lie in between *WDR1* and *MYB*, with *bHLH1* showing a lower d*N*/d*S* ratio compared with *bHLH2*, as documented by Streisfeld & Rausher (2007).

Table 3 Estimates of the d*N*/d*S* ratio for each of the genes

	d*N*/d*S*		
	Full gene	Conserved domain	Variable domain
WDR1	0.036[a]	na	na
bHLH1	0.192[b]	0.136[b]	0.207[b]
bHLH2	0.289[b]	0.130[b]	0.405[b]
MYB	0.496[a]	0.151[a]	0.803[c]

[a]Data obtained from this study.
[b]Data obtained from Streisfeld & Rausher (2007).
[c]Data obtained from Chang *et al.* (2005).
na, not applicable.

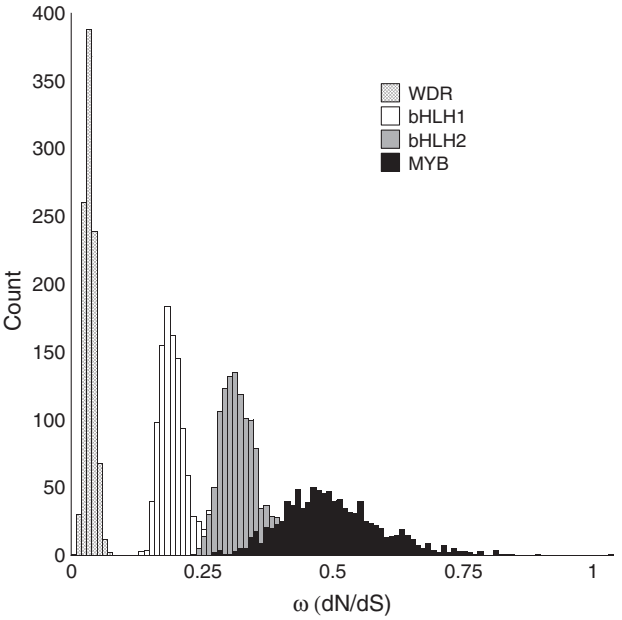

Fig. 2 Distributions of ω from Model M0 of PAML from 1000 bootstrapped datasets for each of the anthocyanin-regulating transcription factors analysed in this study.

Table 4 Pairwise comparisons of the probability that ω differs among the genes

	WDR1	bHLH1	bHLH2	MYB
WDR1	–			
bHLH1	0	–		
bHLH2	0	< 0.001	–	
MYB	0	0	0.016	–

The value reported represents the probability that ω for the gene in each column is greater than or equal to ω for the gene in each row. A probability of less than 0.05 indicates that the gene in each column is more constrained than the gene in each row. Based on 1000 bootstrapped replicates.

We next assessed the extent to which differences in conserved and variable domains contribute to the observed heterogeneity in constraint among these proteins. For *bHLH1*, *bHLH2* and *MYB*, the conserved domain has a significantly lower ω value than its corresponding variable domain (Table 5). Although ω does not differ among the conserved domains of these three genes, significant differences in d*N*/d*S* ratios are found among the variable domains. For example, ω for the variable domain of *MYB* is significantly elevated compared with the variable domains of either *bHLH* gene (Table 5; Fig 3). Moreover, the variable domain of *bHLH1* appears to be constrained significantly more than that of *bHLH2*. Differences in

Table 5 Pairwise comparisons of the probability that ω differs among functional domains of each gene

	Full coding sequence	Conserved domain			Variable domain		
	WDR1	bHLH1	bHLH2	MYB	bHLH1	bHLH2	MYB
WDR1 – full coding	–						
bHLH1 – conserved	< 0.001	–					
bHLH2 – conserved	< 0.001	0.374	–				
MYB – conserved	0.006	0.427	0.605	–			
bHLH1 – variable	0	0.039	0.055	0.209	–		
bHLH2 – variable	0	< 0.0001	0	0.013	< 0.0001	–	
Myb – variable	0	0	0	< 0.001	0	0.007	–

The value reported represents the probability that ω for the domains in each column is greater than or equal to ω of the domains in each row. A probability of less than 0.05 indicates that the gene in each column is more constrained than the gene in each row. Based on 1000 bootstrapped replicates.

overall constraint among these proteins thus appear to result primarily from differences within the variable domain.

As mentioned above, the diagnostic feature of members of the WDR gene family is a series of two to seven imperfect repeats of *c.* 40 amino acids (van Nocker & Ludwig, 2003). There are five such repeats in the WDR1 protein of *Ipomoea* as well as its closely related orthologue from *Petunia hybrida* (AN11; de Vetten *et al.*, 1997; Morita *et al.*, 2006). These repeats are highly conserved among species of *Ipomoea* and comprise *c.* two-thirds of the total amino acid sequence (not shown). As there are no apparent variable domains in this sequence, we compared the entire *WDR1* coding sequence with the conserved and variable regions of the other genes. In each case, *WDR1* had a significantly lower ω (Table 5; Fig 2). *WDR1* thus appears to have a lower overall ω because it is constrained more than even the conserved domains of the other genes.

Discussion

A common observation among both plant and animal TFs is that they often exhibit high d*N*/d*S* ratios compared with the structural genes they regulate (Purugganan & Wessler, 1994; Doebley & Lukens, 1998; Barrier *et al.*, 2001; Remington & Purugganan, 2002), in large part as a result of high d*N*/d*S* ratios in certain domains. Despite the observation of this pattern previously in the *MYB* and *bHLH* anthocyanin regulators in *Ipomoea* (Chang *et al.*, 2005; Streisfeld & Rausher, 2007), we were unable to document evidence for high d*N*/d*S* in *WDR1*. Instead, the *WDR1* gene as a whole (ω = 0.036) exhibits a d*N*/d*S* ratio that is equal to or lower than those of the structural genes that it regulates (ω = 0.034–0.277; Lu & Rausher, 2003). Moreover, d*N*/d*S* is substantially lower for *WDR1* than for the *MYB* and *bHLH* genes (Fig. 2).

These patterns do not appear to be caused by differences in rates of adaptive substitutions. As found previously for the other anthocyanin-regulating TFs (Chang *et al.*, 2005; Streisfeld & Rausher, 2007) and the anthocyanin pathway enzymes (Lu & Rausher, 2003; Rausher *et al.*, 2008), we did not detect evidence for any adaptive substitutions in *WDR1*. Although we note that our method has low power to detect small numbers of adaptive substitutions (Kosakovsky Pond & Frost, 2005; McClellan *et al.*, 2005), such rare changes would seem to be unlikely to account for the large differences in d*N*/d*S* that we have documented. It thus appears that differential constraint best explains the differences in d*N*/d*S* ratios observed here. This conclusion is also supported by the much lower rate of insertion–deletion events in *WDR1*: although indels are common in *MYB* and *bHLH* genes, especially in variable regions (Chang *et al.*, 2005; Streisfeld & Rausher, 2007), only one indel was identified across the entire *WDR1* sequence.

The relative magnitude of differences in constraint in *Ipomoea* is consistent with our *a priori* expectations that are based on the well-established intrinsic and network properties of the proteins that form the MBW complex (Ramsay & Glover, 2005). In Arabidopsis and *Petunia*, the putative orthologues of the WDR1 protein are known to be expressed more widely, to be more centrally located within the network and to be involved in more protein–protein interactions than the MYB proteins, which are more specific and presumably experience fewer deleterious pleiotropic effects of substitutions (Broun, 2005; Koes *et al.*, 2005). From what is known in these model organisms, individual bHLH proteins that join the complex appear to control a subset of WDR's functions (Koes *et al.*, 2005) and, as predicted, these proteins have ω values that are in between those of the WDR1 and MYB proteins. Thus, our predictions of relative differences in evolutionary constraint that were based initially on our knowledge of the functional interactions among members of this complex from model organisms also appear to be supported by estimates of their d*N*/d*S* ratios among species of *Ipomoea*.

The overall differences in constraint detected here arise from a combination of three different components. First, MYB and bHLH proteins have a modular structure with conserved and variable domains, whereas no evidence

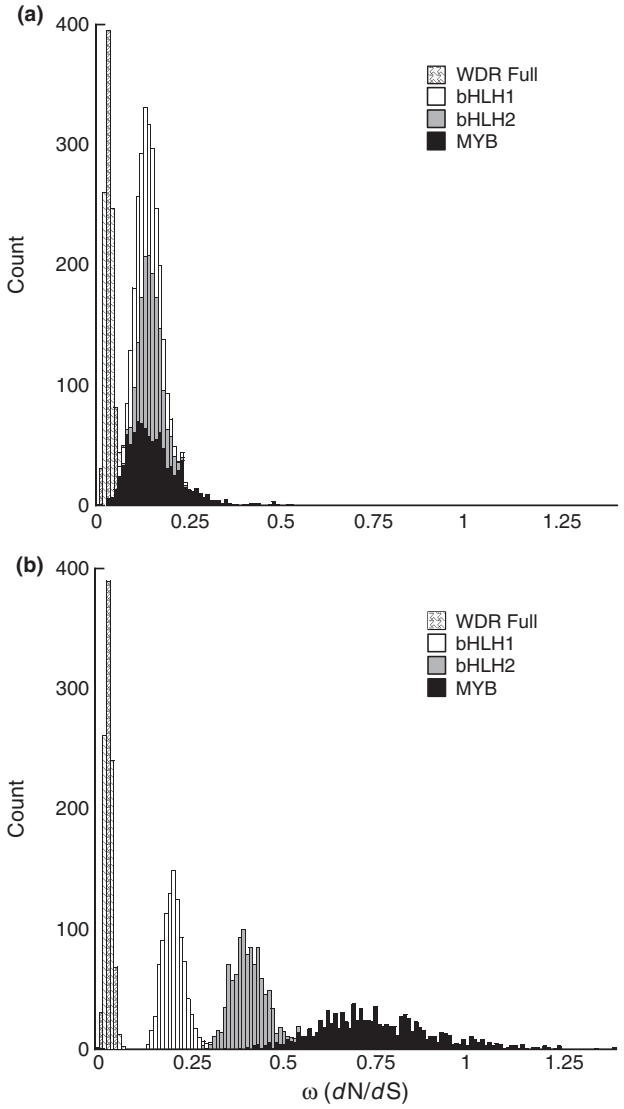

Fig. 3 Distributions of ω from Model M0 of PAML from 1000 bootstrapped datasets for the conserved domains (a) and variable domains (b) for the bHLH and MYB proteins. WDR1 Full represents the frequency distribution of ω values from the complete coding sequence of the WDR protein.

currently exists for different domains in WDR1. Second, the entire WDR1 protein evolves under greater constraint than the conserved domains of the other regulators. Finally, the variable domain of MYB is less constrained than the variable domain of the bHLH proteins.

The conserved domains of the MYB and bHLH proteins play a role in DNA-binding and protein–protein interactions (Goff *et al.*, 1991, 1992; Grotewold *et al.*, 2000; Kroon, 2004), but the main role for the WDR protein in this complex appears to be as a scaffold that holds the complex together (Broun, 2005). Thus, compared with the conserved domains of the other regulators, a larger fraction of amino acid sites across WDR may be required to accommodate this large number of protein interactions. In particular, the five WD repeat motifs found in WDR1 comprise more than two-thirds of the length of the gene sequence and are highly conserved among species of *Ipomoea*. If these motifs are each involved in protein–protein interactions, it suggests the possibility of greater constraint across the entire *WDR1* gene compared with the conserved regions of the other regulators, and thus it appears that the entire WDR1 protein evolves as one conserved unit.

The less-constrained evolution of the MYB variable domain, compared with the variable domains of the bHLH proteins, can also be understood from known functional properties determined from model organisms. The variable regions of each protein are believed to contain transcriptional activation domains (Goff *et al.*, 1991; Gong *et al.*, 1999). In addition, the variable domain of the bHLH proteins appears to have the potential to affect binding to other proteins in the complex. For example, yeast two-hybrid assays in *Petunia hybrida* have shown that the C-terminal regions of the variable domains of JAF13 and AN1 (orthologues of *Ipomoea* bHLH1 and bHLH2, respectively) are required to form homo- and hetero-dimers (Kroon, 2004). By contrast, for MYB, all of the known sites required for binding to target gene promoters and bHLH proteins reside within the defined conserved domain and not the variable domain (Grotewold *et al.*, 2000). Thus, the MYB variable domain appears to perform fewer functions, and therefore is expected to be less constrained and more free to evolve than the variable domains of the bHLH proteins.

Additional evidence for variation in constraint among the anthocyanin transcriptional regulators is provided by analyses of which of these regulators contribute to floral colour differences between species. In particular, current evidence indicates that evolutionary transitions to white flowers and, more generally, changes in the intensity of pigmentation differentially involve mutations affecting MYB TFs rather than bHLH or WDR proteins (Streisfeld & Rausher, 2010). This preferential use of MYBs does not appear to be caused by differences in mutation rates among TF families; rather, it appears to be a result of a higher probability of fixation for mutations that arise in MYBs, which can be caused only by reduced constraint (Streisfeld & Rausher, 2010). Therefore, the differences in d*N*/d*S* ratios documented here and the preferential use of MYBs during flower colour evolution consistently point to lower constraint on MYBs than on the other anthocyanin TFs. We suggest (as have others) that this reduced constraint derives from fewer pleiotropic effects associated with mutations in *MYB* genes because the MYB proteins are more tissue and function specific than the other TFs (Martin *et al.*, 2010).

The *Ipomoea* species included in our analysis differ substantially in floral pigment hue and intensity (Table 1). Yet, despite the central role of this protein regulatory complex in controlling floral pigment production, and despite

the fact that many evolutionary transitions in flower colour are a result of genetic changes in MYB proteins, the lack of evidence for positive selection in any of these TFs suggests that floral colour diversification among these *Ipomoea* species is not caused by repeated episodes of adaptive amino acid substitutions in the coding regions of these proteins. Of course, it is possible that floral colour transitions involve only a small number of substitutions, and hence would not be detectable by the statistical tests used here (Kosakovsky Pond & Frost, 2005; McClellan *et al.*, 2005). However, experimental results demonstrating that TFs from one plant species often can activate anthocyanin production in taxonomically distant species add support to our inference that amino acid substitutions are not involved in adaptive differentiation in these species (Lloyd *et al.*, 1992; Quattrocchio *et al.*, 1993). Moreover, in all cases to date in which the genetic changes in *Myb*s associated with the evolution of floral pigment intensity have been characterized sufficiently, these changes are either in *cis*-regulatory regions (Schwinn *et al.*, 2006; Hopkins & Rausher, 2011) or are a result of frame shift-causing indels (Quattrocchio *et al.*, 1999; Chang *et al.*, 2005; Hoballah *et al.*, 2007); none are caused by amino acid substitutions in the coding regions of TFs. Amino acid substitutions in the anthocyanin-regulating TFs thus do not appear to be involved readily in flower colour evolution. Previous studies have led to the same conclusion regarding amino acid substitutions in genes coding for enzymes in the anthocyanin pathway (Toleno *et al.*, 2010).

Conclusions

Among plants, perhaps the two most widely studied regulatory networks are the MBW complex discussed here and the network that controls flowering time. Consistent with the results reported here on the MBW complex, few signatures of strong positive selection were identified among proteins of the flowering time network in *Arabidopsis* (Flowers *et al.*, 2009). Instead, deleterious polymorphisms segregating at a low frequency among *Arabidopsis* accessions were common. Despite detailed investigation into the flowering time network, little is still known about the role of pleiotropic constraints affecting the molecular evolution of these genes. Thus, *a priori* predictions about expected differences in constraint among the members of the flowering time regulatory network are difficult to formulate.

By contrast, we have shown that the individual proteins that form the MBW complex of anthocyanin-regulating TFs in *Ipomoea* evolve in a predictable manner based on differences in their expected levels of constraint. In particular, the WDR1 protein, which is the most highly integrated into this network, exhibits greater evolutionary constraint than the bHLH and MYB proteins. By contrast, the MYB protein, which exhibits the least constraint, is expressed more specifically and controls fewer functions than the other members of the complex. The patterns of evolution exhibited by these TFs are thus consistent with the view that network properties of proteins have a substantial impact on the magnitude of evolutionary constraint (Pal *et al.*, 2006). Moreover, it is notable that our predictions of differential constraint among the members of the MBW complex were upheld by sequence data from *Ipomoea* species, even though our predictions were generated on the basis of molecular genetic data compiled from model organisms. This suggests the possibility that the differences in constraint identified here may reflect similar levels of constraint among other species as well.

Finally, we were unable to find evidence either for a pattern of elevated d*N*/d*S* ratios in any region of *WDR1* or for positive selection at any site in this gene that might account for flower colour diversification among species of *Ipomoea*. This lack of adaptive substitutions is consistent with previous studies of MYB and bHLH proteins (Chang *et al.*, 2005; Streisfeld & Rausher, 2007) and suggests that sequence divergence among *Ipomoea* species at each of the anthocyanin TFs does not contribute in a major way to flower colour evolution. Instead, flower colour evolution seems more likely to be accomplished by changes in regulatory sequences, as has been demonstrated in several cases (Durbin *et al.*, 2003; Whittall *et al.*, 2006; Streisfeld & Rausher, 2009, 2010; Des Marais & Rausher, 2010; Hopkins & Rausher, 2011), or by loss-of-function mutations (Quattrocchio *et al.*, 1999; Hoballah *et al.*, 2007).

Acknowledgements

We would like to thank R. Miller and members of the Rausher laboratory for discussions that greatly improved the manuscript. This research was supported by the National Science Foundation.

References

Barrier M, Robichaux RH, Purugganan MD. 2001. Accelerated regulatory gene evolution in an adaptive radiation. *Proceedings of the National Academy of Sciences, USA* **98**: 10208–10213.

Begun DJ, Holloway AK, Stevens K, Hillier LW, Poh YP, Hahn MW, Nista PM, Jones CD, Kern AD, Dewey CN *et al.* 2007. Population genomics: whole-genome analysis of polymorphism and divergence in *Drosophila simulans*. *Plos Biology* **5**: 2534–2559.

Broun P. 2005. Transcriptional control of flavonoid biosynthesis: a complex network of conserved regulators involved in multiple aspects of differentiation in Arabidopsis. *Current Opinion in Plant Biology* **8**: 272–279.

Chang SM, Lu Y, Rausher MD. 2005. Neutral evolution of the nonbinding region of the anthocyanin regulatory gene *Ipmyb1* in *Ipomoea*. *Genetics* **170**: 1967–1978.

Charlesworth J, Eyre-Walker A. 2006. The rate of adaptive evolution in enteric bacteria. *Molecular Biology and Evolution* **23**: 1348–1356.

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

Doebley J. 1993. Genetics, development and plant evolution. *Current Opinion in Genetics and Development* **3**: 865–872.

Doebley J, Lukens L. 1998. Transcriptional regulators and the evolution of plant form. *Plant Cell* **10**: 1075–1082.

Durbin ML, Lundy KE, Morrell PL, Torres-Martinez CL, Clegg MT. 2003. Genes that determine flower color: the role of regulatory changes in the evolution of phenotypic adaptations. *Molecular Phylogenetics and Evolution* **29**: 507–518.

Flowers JM, Hanzawa Y, Hall MC, Moore RC, Purugganan MD. 2009. Population genomics of the *Arabidopsis thaliana* flowering time gene network. *Molecular Biology and Evolution* **26**: 2475–2486.

Goff SA, Cone KC, Chandler VL. 1992. Functional analysis of the transcription activator encoded by the maize B-gene: evidence for a direct functional interaction between two classes of regulatory proteins. *Genes and Development* **6**: 864–875.

Goff SA, Cone KC, Fromm ME. 1991. Identification of functional domains in the maize transcriptional activator C1: comparison of wild-type and dominant inhibitor proteins. *Genes and Development* **5**: 298–309.

Gong ZZ, Yamagishi E, Yamazaki M, Saito K. 1999. A constitutively expressed *Myc*-like gene involved in anthocyanin biosynthesis from *Perilla frutescens*: molecular characterization, heterologous expression in transgenic plants and transactivation in yeast cells. *Plant Molecular Biology* **20**: 33–44.

Gonzalez A, Zhao M, Leavitt JM, Lloyd AM. 2008. Regulation of the anthocyanin biosynthetic pathway by the TTG1/bHLH/Myb transcriptional complex in Arabidopsis seedlings. *Plant Journal* **53**: 814–827.

Grotewold E. 2006. *The science of flavonoids*. New York, NY, USA: Springer.

Grotewold E, Sainz MB, Tagliani L, Hernandez JM, Bowen B, Chandler VL. 2000. Identification of the residues in the myb domain of maize C1 that specify the interaction with the bHLH cofactor R. *Proceedings of the National Academy of Sciences, USA* **97**: 13579–13584.

Halligan DL, Oliver F, Eyre-Walker A, Harr B, Keightley PD. 2010. Evidence for pervasive adaptive protein evolution in wild mice. *PLoS Genetics* **6**: e1000825.

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

Holton TA, Cornish EC. 1995. Genetics and biochemistry of anthocyanin biosynthesis. *Plant Cell* **7**: 1071–1083.

Hopkins RH, Rausher MD. 2011. Identification of two genes causing reinforcement in the Texas wildflower *Phlox drumondii*. *Nature* **469**: 411–414.

Kimura M. 1968. Evolutionary rate at the molecular level. *Nature* **217**: 624–626.

Kimura M. 1977. Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. *Nature* **267**: 275–276.

Kimura M. 1983. *The neutral theory of molecular evolution*. London, UK: Cambridge University Press.

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

Kosakovsky Pond SL, Frost SDW. 2005. Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Molecular Biology and Evolution* **22**: 1208–1222.

Kroon AR. 2004. *Transcription regulation of the anthocyanin pathway in* Petunia hybrida. PhD thesis, Vrije Universiteit, Amsterdam, the Netherlands.

Langercrantz U, Axelsson T. 2000. Rapid evolution of the family of *CONSTANS LIKE* genes in plants. *Molecular Biology and Evolution* **17**: 1499–1507.

Li WH. 1997. *Molecular evolution*. Sunderland, MA, USA: Sinauer Associates.

Lloyd AM, Walbot V, Davis RW. 1992. *Arabidopsis* and *Nicotiana* anthocyanin production activated by maize regulators *R* and *C*. *Science* **258**: 1773–1775.

Lu Y, Rausher MD. 2003. Evolutionary rate variation in anthocyanin pathway genes. *Molecular Biology and Evolution* **20**: 1844–1853.

Lukens L, Doebley J. 2001. Molecular evolution of the teosinte branched gene among maize and related grasses. *Molecular Biology and Evolution* **18**: 627–638.

Manos PS, Miller RE, Wilkin P. 2001. Phylogenetic analysis of *Ipomoea*, *Argyreia*, *Stictocardia*, and *Turbina* suggests a generalized model of morphological evolution in morning glories. *Systematic Botany* **26**: 585–602.

Martin C, Ellis N, Rook F. 2010. Do transcription factors play special roles in adaptive variation? *Plant Physiology* **154**: 506–511.

McClellan DA, Palfreyman EJ, Smith MJ, Moss JL, Christensen RG, Sailsberry JK. 2005. Physicochemical evolution and molecular adaptation of the Cetacean and Artiodactyl cytochrome *b* proteins. *Molecular Biology and Evolution* **22**: 437–455.

Miller RE, Rausher MD, Manos PS. 1999. Phylogenetic systematics of *Ipomoea* (Convolvulaceae) based on ITS and *waxy* sequences. *Systematic Botany* **24**: 209–227.

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

Morita Y, Saitoh M, Hoshino A, Nitasaka E, Iida S. 2006. Isolation of cDNAs for R2R3-MYB, bHLH, and WDR transcriptional regulators and identification of *c* and *ca* mutations conferring white flowers in the Japanese morning glory. *Plant Cell Physiology* **47**: 457–470.

Nielsen R, Yang Z. 1998. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* **148**: 929–936.

van Nocker S, Ludwig P. 2003. The WD-repeat protein superfamily in Arabidopsis: conservation and divergence in structure and function. *BMC Genomics* **4**: 50.

Pal C, Papp B, Lercher MJ. 2006. An integrated view of protein evolution. *Nature Reviews Genetics* **7**: 337–348.

Purugganan MD, Rounsley SD, Schmidt RJ, Yanofsky MF. 1995. Molecular evolution of flower development: diversification of the plant MADS-Box regulatory gene family. *Genetics* **140**: 345–356.

Purugganan MD, Wessler SR. 1994. Molecular evolution of the plant *R* regulatory gene family. *Genetics* **138**: 849–854.

Quattrocchio F, Baudry A, Lepiniec L, Grotewold E. 2006. The regulation of flavonoid biosynthesis. In: Grotewold E., ed. *The science of flavonoids*. New York, NY, USA: Springer, 97–122.

Quattrocchio F, Wing JF, Leppen HTC, Mol JNM, Koes RE. 1993. Regulatory genes controlling anthocyanin pigmentation are functionally conserved among plant species and have distinct sets of target genes. *Plant Cell* **5**: 1497–1512.

Quattrocchio F, Wing J, van der Woude K, Souer E, de Vetten N, Mol JNM, Koes R. 1999. Molecular analysis of the *anthocyanin2* gene of *Petunia* and its role in the evolution of flower color. *Plant Cell* **11**: 1433–1444.

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

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

Rausher MD, Lu Y, Meyer K. 2008. Variation in constraint versus positive selection as an explanation for evolutionary rate variation among anthocyanin genes. *Journal of Molecular Evolution* **67**: 137–144.

Rausher MD, Miller RE, Tiffin P. 1999. Patterns of evolutionary rate variation among genes of the anthocyanin biosynthetic pathway. *Molecular Biology and Evolution* **16**: 266–274.

Remington DL, Purugganan MD. 2002. *GAI* homologues in the Hawaiian Silversword alliance (Asteraceae-Madiinae): molecular evolution of growth regulators in a rapidly diversifying plant lineage. *Molecular Biology and Evolution* **19**: 1563–1574.

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

Shepard KA, Purugganan MD. 2002. The genetics of plant morphological evolution. *Current Opinion in Plant Biology* **5**: 49–55.

Smith NG, Eyre-Walker A. 2002. Adaptive protein evolution in *Drosophila*. *Nature* **415**: 1022–1024.

Streisfeld MA, Rausher MD. 2007. Relaxed constraint and evolutionary rate variation between the basic helix-loop-helix (bHLH) floral anthocyanin regulators in *Ipomoea*. *Molecular Biology and Evolution* **24**: 2816–2826.

Streisfeld MA, Rausher MD. 2009. Altered *trans*-regulatory control of gene expression in multiple anthocyanin genes contributes to adaptive flower color evolution in *Mimulus aurantiacus*. *Molecular Biology and Evolution* **26**: 433–444.

Streisfeld MA, Rausher MD. 2010. Population genetics, pleiotropy, and the preferential fixation of mutations during adaptive evolution. *Evolution*. doi: 10.1111/j.1558-5646.2010.01165.x.

Toleno DM, Durbin ML, Lundy KE, Clegg MT. 2010. Extensive evolutionary rate variation in floral color determining genes in the genus *Ipomoea*. *Plant Species Biology* **25**: 30–42.

de Vetten N, Quattrocchio F, Mol J, Koes R. 1997. The *an11* locus controlling flower pigmentation in petunia encodes a novel WD-repeat protein conserved in yeast, plants, and animals. *Genes and Development* **11**: 1422–1434.

Whittall JB, Voelckel C, Kliebenstein DJ, Hodges SA. 2006. Convergence, constraint, and the role of gene expression during adaptive radiation: floral anthocyanins in *Aquilegia*. *Molecular Ecology* **15**: 4645–4657.

Yang Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *CABIOS* **13**: 555–556.

Yang Z, Bielawski JP. 2000. Statistical methods for detecting molecular adaptation. *Trends in Ecology and Evolution* **15**: 496–503.

Yang Z, Nielsen R. 2000. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Molecular Biology and Evolution* **17**: 32–43.

Supporting Information

Additional supporting information may be found in the online version of this article.

Table S1 A list of the PCR primers used to amplify and sequence *WDR1* from *Ipomoea* species

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.