Relaxed Constraint and Evolutionary Rate Variation between Basic Helix-Loop-Helix Floral Anthocyanin Regulators in *Ipomoea*

Matthew A. Streisfeld and Mark D. Rausher

Department of Biology, Duke University, Durham, NC

Regulatory genes are believed to play a large role in morphological diversification and are often characterized by elevated rates of evolution. Whether this rapid evolution is primarily due to adaptive differentiation or relaxed selective constraint remains an open question. We attempted to distinguish between these alternative outcomes in 2 transcription factors known to regulate the expression of anthocyanin pigmentation genes in flowers. We cloned the full-length coding region from 2 basic helix-loop-helix transcription factors from several species of *Ipomoea* with diverse flower colors and determined the selective forces operating on them. In both genes, rapidly evolving sites and indel mutations are clustered in nonbinding domains, but the extent of rate acceleration in these domains is reduced relative to most previously characterized plant transcription factors. Moreover, codon models of substitution rates and models evaluating the magnitude of change to physical amino acid properties demonstrate little evidence for adaptive evolution and suggest that elevated nonsynonymous substitution rates in these domains represent relaxed selective constraint. Although both genes show qualitatively similar patterns, their rates of evolution differ significantly due to an increased rate of nonsynonymous substitutions in the nonbinding domains in one copy, suggesting substantial differences in functional constraint on each gene. In general, these results provide additional evidence demonstrating that decreased constraint as opposed to positive selection is largely responsible for the frequently observed pattern of rapid evolution in particular domains of plant transcription factors. More specifically, they suggest that most of the amino acid substitutions are neutral and do not implicate a role for natural selection on these regulatory genes in the diversification of flower color in *Ipomoea*.

Introduction

Rates of morphological evolution are often greatly decoupled from rates of protein sequence evolution. A popular explanation to account for this pattern is that differences in gene regulation rather than diversification of structural genes contribute most to morphological evolution (King and Wilson 1975; Doebley 1993; Doebley and Lukens 1998; Carroll 2005). Developmental genetic analyses routinely have shown that altering the timing and/or location of gene expression can have profound effects on morphological diversity (e.g., Weigel and Meyerowitz 1993). However, it is still largely unclear whether mutations in structural or regulatory genes contribute most to phenotypic evolution.

One observation supporting the role of regulatory genes in plant diversification is that plant transcription factors often display elevated nonsynonymous to synonymous substitution rate ratios (ω or d_N/d_S) compared with structural genes (Purugganan and Wessler 1994; Purugganan et al. 1995; Rausher et al. 1999; Langercrantz and Axelsson 2000; Barrier et al. 2001; Lukens and Doebley 2001; Remington and Purugganan 2002). These higher overall rates across the genes are primarily due to particular domains with ω ratios that often approach 1, whereas other regions are highly conserved (Purugganan and Wessler 1994; Purugganan et al. 1995; Purugganan 1998; Rausher et al. 1999; Langercrantz and Axelsson 2000; Chang et al. 2005). If repeated positive selection accounts for these domain-specific elevated ω's, then those substitutions may have contributed to phenotypic diversification among species. Conversely, if these substitutions are largely nonadaptive, it is unlikely that rapid evolution of these regulatory genes has played a role in morphological divergence.

Despite this evidence for rapid evolution in plant transcription factors, there have been few attempts to determine whether these elevated rates of nonsynonymous substitution are due to the effects of positive selection or whether they represent a relaxation of selective constraint. In the few cases where this has been explicitly tested, analyses repeatedly have found that particular domains in these plant transcription factors evolve at elevated rates, but rarely because repeated positive selection has driven diversification (Remington and Purugganan 2002; Fan et al. 2004; Chang et al. 2005; but see Hernández-Hernández et al. 2007). Our primary objective in this study was to extend these analyses by attempting to distinguish between relaxed selective constraint and repeated positive selection as explanations for high rates of nonsynonymous substitutions in 2 transcription factors that control the expression of anthocyanin pathway genes in the genus *Ipomoea* (morning glories).

The anthocyanin biosynthetic pathway has been a model for the study of gene regulation for some time, and the structural genes as well as their regulatory elements are well characterized and conserved across the angiosperms (Koes et al. 1994; Mol et al. 1998; Rausher 2006). According to the most recent models of transcriptional regulation of floral anthocyanin production, at least 3 gene families are believed to be involved: 1) members of the R2R3-MYB family, 2) two members of the *R*-like, basic helix-loop-helix (bHLH) family that are known to interact directly with each other and with their partner MYB protein, and 3) members of the WD-40 family that are believed to be important in posttranslational control of the MYB and bHLH proteins (Mol et al. 1998; Kroon 2004). The complex formed by the MYB and bHLH proteins has been shown to bind directly to the promoter regions of target structural anthocyanin genes to coordinately control transcription (Mol et al. 1998). Previous estimates of ω ratios in one bHLH regulator suggest that particular functional

Key words: Ipomoea, positive selection, rate variation, anthocyanin, basic helix-loop-helix, transcription factor.

E-mail: mstreisf@duke.edu.

Mol. Biol. Evol. 24(12):2816–2826. 2007
doi:10.1093/molbev/msm216
Advance Access publication October 5, 2007

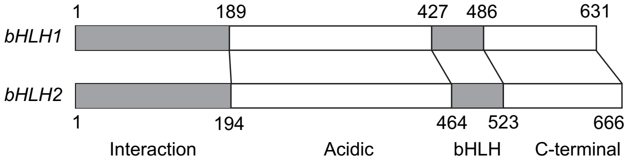

FIG.1.—A schematic representation of the 2 *bHLH* genes sequenced in this study. The numbers represent the amino acid position of each functional domain in the *Ipomoea purpurea* sequence. The domains shaded in gray denote the binding domains, whereas those in white represent the nonbinding domains.

domains within these genes have greatly elevated rates of amino acid substitution (Purugganan and Wessler 1994; Rausher et al. 1999; Fan et al. 2004), again showing the familiar pattern that plant transcription factors often appear to evolve rapidly relative to the genes they regulate. However, these analyses did not rigorously test for the effects of positive selection and so cannot explicitly establish whether accelerated substitution rates were due to adaptive evolution or relaxed constraint.

The *bHLH* gene family is characterized by a highly conserved helix-loop-helix motif that is preceded by a series of basic amino acids. This motif is present in over 200 copies among vertebrates (Massari and Murre 2000), and there have been 133 bHLH factors identified in the *Arabidopsis thaliana* genome (Heim et al. 2003). Members of the *R*-like subfamily have been described as transcriptional regulators of anthocyanin synthesis in several model species. In particular, at least 2 gene copies that function in anthocyanin regulation have been identified in *Zea mays*, *Petunia hybrida*, *Perilla frutescens*, and *Antirrhinum majus* (Kroon 2004). In *Ipomoea nil* and *Ipomoea purpurea*, orthologs of the above 2 genes have been characterized and named *bHLH1* and *bHLH2* (Morita et al. 2006; Park et al. 2007). A third copy (*bHLH3*) in *I. nil* also has been described and appears to be derived from a recent duplication event in the ancestor of *bHLH1* (Morita et al. 2006).

Molecular genetic studies have identified 4 functional domains within the proteins: interaction, acidic, bHLH, and C-terminal (fig. 1). The highly conserved interaction domain consists of approximately 190 amino acids at the N-terminus of the protein and has been shown to bind directly with partner R2R3-MYB factors (Goff et al. 1992). The acidic domain is composed of a region of approximately 230 amino acids that contains poorly conserved stretches of negatively charged acidic amino acids required to activate transcription of the target structural genes (Gong et al. 1999). The bHLH domain contains 59 highly conserved amino acids presumed to be involved in DNA binding, whereas the final approximately 150 amino acids on the C-terminus of the protein are required for subunit dimerization with other bHLH proteins (Kroon 2004). Rates of nonsynonymous substitution have been shown to be greatest in the acidic and C-terminal domains (Purugganan and Wessler 1994). This offers additional support for the observation that plant transcription factors often show domain-specific elevated d_N/d_S ratios, and it implies that differences in functional constraint exist among the domains. Therefore, a second goal of the current study is to determine whether the bHLH factors involved in regulation of anthocyanin genes in the genus *Ipomoea* also display domain-specific elevated d_N/d_S ratios.

A final objective of this investigation is to examine whether differences in the extent of pleiotropy manifested by the 2 *bHLH* paralogs has resulted in variable levels of evolutionary constraint. Mutant analyses of *bHLH2* in *Ipomoea* species have demonstrated that the gene controls multiple additional functions besides the regulation of floral anthocyanins. Mutants deficient for the *bHLH2* gene in *I. purpurea* and *Ipomoea tricolor* have reductions in flower and seed coat pigmentation, yielding pale flowers and ivory-colored seeds (Park et al. 2004, 2007). In addition, other seed epidermal traits that are known to be important in plant defense are negatively affected in the *I. purpurea* mutants. These include smaller and fewer unbranched trichomes and reduced accumulation of the proanthocyanidin and phytomelanin secondary metabolites in the ivory seeds (Park et al. 2007). In contrast, regulation of anthocyanin synthesis is the only function currently assigned to the orthologs of the *bHLH1* gene in the flowering plant species that have been characterized (Goff et al. 1992; Quattrocchio et al. 1998; Spelt et al. 2000). These results suggest that there may be substantial differences in the extent of constraint on the 2 loci, and this has implications for their relative rates of evolution. Because of the multiple functions assigned to *bHLH2*, it is reasonable to predict that *bHLH2* evolves more slowly than *bHLH1*, whose only known function is in the regulation of anthocyanin synthesis.

In summary, we address here 3 questions concerning the rates and patterns of molecular evolution of 2 bHLH transcription factors in *Ipomoea*. First, we ask how well these bHLH transcription factors conform to the patterns of regulatory gene evolution seen in other plant transcription factors. Specifically, is there evidence for rapid evolution of the coding region as a whole, and if so, can this be attributed to particular domains with elevated ω ratios? Second, as mentioned above, orthologs of these *bHLH* genes appear to be essential for the transcriptional regulation of floral anthocyanins in several model species. Because floral pigmentation is an ecologically important character that is often linked to pollinator attraction, rapid evolution of these *bHLH* regulatory genes due to repeated positive selection may be indicative of adaptive evolution. Therefore, we select a set of species in the genus *Ipomoea* that shows substantial flower color diversification and ask whether there is evidence for positive selection in either of the bHLH transcriptional regulators. Among several of the species included in this investigation, flower color has been shown to contribute to adaptive differentiation (Machado and Sazima 1987; Rausher and Fry 1993). Finally, we test whether the gene copies evolve at different rates, as would be predicted by what is known about relative differences in functional constraint on the individual genes.

Materials and Methods

Species

The genus *Ipomoea* (Convolvulaceae) contains approximately 1,000 species that are distributed throughout the tropics and subtropics. In this study, we selected 12

Table 1
A List of the Species Sequenced in This Study, with Their Associated Flower Colors and GenBank Accession Numbers

		GenBank Accession Numbers	
Species	Flower Color	*bHLH1*	*bHLH2*
Ipomoea purpurea	Purple	AB252664*	AB252663*
Ipomoea nil	Blue	AB232774*	AB232775*
Ipomoea hederacea	Blue	EU192087	EU192096
Ipomoea tricolor	Blue	EU192088	AB154370*
Ipomoea alba	White	EU192089	EU192097
Ipomoea quamoclit	Red	EU192091	EU192098
Ipomoea coccinea	Red	EU192090	EU192099
Ipomoea horsfalliae	Red	EU192094	EU192103
Ipomoea lacunosa	White	EU192092	EU192101
Ipomoea trifida	Lavender	EU192093	EU192100
Ipomoea hochstetteri	Blue	EU192086	EU192104
Ipomoea violacea	White	EU192095	EU192102
Operculina pteripes	Salmon	EU192085	EU192105

* These GenBank accession sequence numbers were previously available (Park et al. 2004, 2007; Morita et al. 2006).

species from within *Ipomoea* as ingroup species as well as the more distantly related *Operculina pteripes* as an outgroup. The phylogenetic relationships among these species are well established (Miller et al. 1999; Manos et al. 2001), and they display substantial variation of floral colors (table 1). Plant material used for isolation of nucleic acids was obtained from our greenhouse collections. Voucher information for several of these species is available in Miller et al. (1999) or can be obtained directly from the authors.

Sequencing of *bHLH* Genes

We sequenced *Ipomoea bHLH* genes using polymerase chain reaction (PCR) amplification of first-strand complementary DNA (cDNA), derived from total RNA extracted from the distal half of floral bud tips, approximately 1 day before the flower opened. RNA was extracted using the Spectrum Plant Total RNA extraction kit (Sigma-Aldrich, St. Louis, MO) and first-strand cDNA was generated using M-MLV reverse transcriptase (Invitrogen, Carlsbad, CA), both according to the manufacturers' specifications. Orthologs of 2 previously identified *bHLH* genes in *I. nil* (*InbHLH1* and *InbHLH2*; Morita et al. 2006) were PCR amplified in this study from all 13 species, with the exception of *I. purpurea*, *I. nil*, and *I. tricolor* (*bHLH2* only), whose sequences were available on public databases. A third copy of this gene in *I. nil*, *InbHLH3*, was not sequenced in this study as it did not appear to be expressed in corolla tissues (Morita et al. 2006). PCR primers were designed based on copy-specific conserved regions throughout the coding sequence by manually aligning the publicly available sequences. In most cases, 2 to 3 overlapping PCR products were obtained that covered the entire coding sequence. After PCR amplification, products were gel extracted (Qiagen, Valencia, CA) and cloned into the pCR 2.1 TOPO vector (Invitrogen). A minimum of 3 clones were sequenced for each product using an ABI 3730 capillary sequencer and the Big-Dye protocol (Applied Biosystems, Foster City, CA). Both strands of plasmid DNA were sequenced using the M13 forward and reverse primers. All PCR primers used are listed in Supplementary Material online.

Phylogenetic Reconstructions

Overlapping sequencing products were combined to form single contiguous reads for each species. Nucleotide sequences were translated to the predicted amino acid sequence and aligned separately for each copy using ClustalW, followed by manual adjustment of both the amino acid and nucleotide sequences. Only short stretches of amino acid residues were conserved well enough between the paralogs to perform unambiguous alignments. As a result, all subsequent analyses evaluated each gene copy separately. Phylogenetic reconstructions were first performed for each *bHLH* gene to determine whether the recovered gene trees reflected the established species relationships (Miller et al. 1999). In addition, internal transcribed spacer (ITS) sequences for these species were provided by R. Miller (Southeastern Louisiana University), aligned, and used for separate phylogenetic reconstructions. Phylogenetic trees were constructed using MrBayes v. 3.1 (Huelsenbeck and Ronquist 2001; Ronquist and Huelsenbeck 2003). We analyzed 4 independent runs, each using the general time reversible (GTR) model plus gamma-distributed rates (GTR + γ), as determined by Modeltest v3.7 (Posada and Crandall 1998). The data were partitioned by codon position (for the *bHLH* genes), and we ran 5 million generation Markov Chain Monte Carlo simulations with 4 separate chains, with the first 250,000 generations discarded as burn-in. Trees were summarized for each independent run and compared to check for concordant topologies. The consensus tree of all compatible groupings among all runs was used in all analyses.

Estimating Rates of Amino Acid Substitution

We evaluated codon- and lineage-based maximum likelihood models to characterize the selective processes that may have previously acted on each *bHLH* gene (Nielsen and Yang 1998; Yang and Nielsen 2000; Yang et al. 2000). Although these models may have low power to detect positive selection if it occurred infrequently at only a limited number of sites, it has been shown that repeated episodes of positive selection are readily detectable using these methods. All analyses were implemented using the codeml program of PAML version 3.15 (Yang 1997). The ratio of nonsynonymous to synonymous substitution rates (d_N/d_S or ω) was compared in these models. An estimate significantly less than 1 implies purifying selection, whereas an estimate of 1 suggests the presence of neutral evolution. Positive selection is indicated if the estimated ω is significantly greater than 1.

We first determined whether there was evidence for heterogeneous rates of evolution along any of the lineages leading to the extant species we studied. These lineage-based models assume a single ω for all codon sites. We compared estimates from a model where a single ω was constrained on all branches of the phylogeny with a free-ratio model, where ω was allowed to vary independently on each branch. We used the standard likelihood ratio framework (Hocking 1985) to evaluate whether the free-ratio model fit the data significantly better than the single-ω model. Specifically, 2 times the difference in log-likelihood scores from each model were 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 by using codon site models, where the mean value of ω was held constant on each branch, but ω was estimated independently for each codon. For these analyses, a series of nested models was run, and likelihood ratio tests (LRTs) were conducted to test for the presence of positive selection on individual amino acids (Yang et al. 2000). Model M0 represents the null model that all sites have a single-ω value. Model M1a estimates 2 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 ($\omega > 1$). In Model M3 (discrete), each codon is assigned to 1 of 3 site classes that have been estimated without any constraint. 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 $\omega \geq 1$. LRTs 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 are estimated according to an empirical Bayes procedure to be greater than 1, then positive selection is indicated.

Analysis of Amino Acid Substitutions

The ability to accurately detect positive selection using methods that rely only on the d_N/d_S ratio may be quite poor, particularly if adaptive evolution targeted only a limited subset of amino acid changes and/or occurred infrequently (Kosakovsky Pond and Frost 2005; McClellan et al. 2005). Additional methods for examining the effects of nonsynonymous substitutions on protein evolution have focused on changes to physicochemical properties of particular amino acid residues. Because these properties are important in determining the 3-dimensional structure and conformation of the protein, and hence its biochemical function, it may be more appropriate to evaluate selection in terms of the magnitude of change in physicochemical property.

For each *bHLH* gene, we examined whether nonsynonymous substitutions caused significantly more or less change in 31 different physicochemical amino acid properties than expected under neutrality, as implemented in the program TreeSAAP version 3.2 (Wooley et al. 2003). For all possible pairwise amino acid changes, the range of effect size for each of the 31 properties was determined and equally divided into 8 magnitude categories. Lower number categories represented more conservative changes, and higher category numbers indicated more radical changes. Ancestral nucleotide sequences at each node of the phylogeny were then reconstructed using the baseml program of PAML v.3.15, with the GTR model of nucleotide substitution. For each inferred amino acid substitution in the data, the magnitude of effect size for each physicochemical property was determined and assigned to 1 of the 8 categories. An expected distribution that modeled the effects of neutral evolution was calculated by characterizing the 9 possible nucleotide changes from each codon in the data set, evaluating each amino acid substitution for the magnitude of change in each property, and assigning each change to 1 of the 8 categories defined above. These changes were summed across the data to construct a set of relative frequencies of change for all of the magnitude classes. If the distribution of observed changes fit the data significantly better than the expected distribution according to χ^2 goodness-of-fit tests, we rejected the null hypothesis that physicochemical property changes in each *bHLH* gene were neutral. For all properties that differed significantly from neutrality, *Z*-scores were then calculated in each magnitude category to determine which classes contributed to this deviation. Significantly positive *Z*-scores indicated an overrepresentation of substitutions of that magnitude relative to neutrality, and significantly negative *Z*-scores indicated an underrepresentation of those substitutions relative to neutral expectations. We used a Bonferroni correction to account for the effects of multiple comparisons by adjusting the critical significance level for both goodness-of-fit and *Z*-score tests to $P < 0.001$ (Sokal and Rohlf 1995). Substitutions of magnitude categories 1 and 2 were considered to be the most conservative, whereas substitutions of categories 7 and 8 were the most radical. Radical substitutions affecting a particular property that occurred more frequently than expected by chance constituted the signature of adaptive evolution (McClellan et al. 2005).

Domain-Specific Differences in Rates of Amino Acid Substitution

In order to determine if rapidly evolving amino acid sites were clustered into particular functional domains of the *bHLH* genes, we first partitioned the sequence of each *bHLH* gene into binding domains (including the interaction and bHLH domains) and nonbinding domains (acidic and C-terminal), according to figure 1. We used a method described previously by Lu and Rausher (2003) to evaluate whether the d_N/d_S ratios were significantly different between the domains for each gene. Briefly, the d_N/d_S ratio was estimated separately for each domain using Model M0 from the codeml program of PAML. We next attempted to find a value of ω that satisfied each of 2 criteria: 1) its value lied in between the ω value for each gene and 2) it was significantly different than the ω for each gene. If such a value could be found, then it would suggest that the confidence intervals around the estimates of ω did not overlap between the 2 genes and the ω values were significantly different from each other. In order to assess significance, we performed LRTs by comparing the log likelihood of the model where ω was estimated to the log likelihood of the same model where ω was constrained to particular values.

We also compared the locations of the most rapidly evolving sites, as identified by Model M8 in PAML. Because our analysis of the *bHLH2* gene indicated that approximately 15% of sites occurred in the rapidly evolving site class with $\omega = 1.2$ (see below), we used empirical Bayes estimation to designate the 15% of sites with the greatest estimated ω ratio as the most rapidly evolving sites. We then noted the functional domain where each identified site occurred (binding or nonbinding) and used a *G*-test (correcting for differences in domain size) to examine the hypothesis that rapidly evolving sites occurred more frequently in the nonbinding domain than expected by chance. We performed the same analysis with the 15% of sites to be evolving most rapidly in the *bHLH1* gene as well.

Comparing d_N/d_S between Genes

In order to determine whether rates of evolution differed between the 2 *bHLH* genes, we compared d_N/d_S ratios using 2 separate methods. The first procedure was described above and attempted to find a value of ω that was significantly different than the estimated ω for each gene (Lu and Rausher 2003). We used Model M0 from PAML to estimate the ω ratio for the full sequence of each *bHLH* gene.

The second approach followed the methods of Yang and Swanson (2002), where variation in site heterogeneity was estimated between previously defined partitions in a combined data set consisting of both genes. We partitioned the data by gene and ran Model C, which estimated different substitution patterns and equilibrium codon frequencies between the partitions, but the same transition/transversion rate ratio (κ) and nonsynonymous/synonymous rate ratio (ω). We compared that model with the nested-Model E, which was identical to Model C, except that it estimated κ and ω separately between the partitions. We used a likelihood ratio test with 2 df to determine if Model E fit the data significantly better than Model C, which would indicate that κ and/or ω evolved at different rates between the 2 genes. In these analyses, a significant likelihood ratio statistic cannot unambiguously determine that ω differs between the genes because in both models, ω is simultaneously estimated with κ. However, we compared the point estimates of each parameter between the partitions in Model E to evaluate whether 1 or both parameters changed substantially between the partitions. Finally, to determine if differences in rates between the genes could be attributed to particular functional domains, we ran 2 additional analyses that partitioned either the binding or the nonbinding domains of each gene. We then compared which domains contributed most to variation in evolutionary rates.

Results

In order to characterize the selective forces that have operated on the bHLH anthocyanin transcriptional regulators, we cloned and sequenced the entire coding region from 2 *bHLH* genes from 13 species of *Ipomoea*. Sequence lengths varied among species from 1,878 to 1,896 bp in *bHLH1*, whereas *bHLH2* was slightly larger and ranged from 1,980 to 2,055 bp. The amino acid alignments of both genes are provided in Supplementary Material online. The effective number of codons used and GC nucleotide content (calculated with DNAsp v. 4.0; Rozas et al. 2003) were highly similar among species and genes. There were no indications in any of the sequences of premature stop codons or frameshift mutations that would appear to cause a nonfunctional enzyme, even though insertion–deletion (indels) mutations were common.

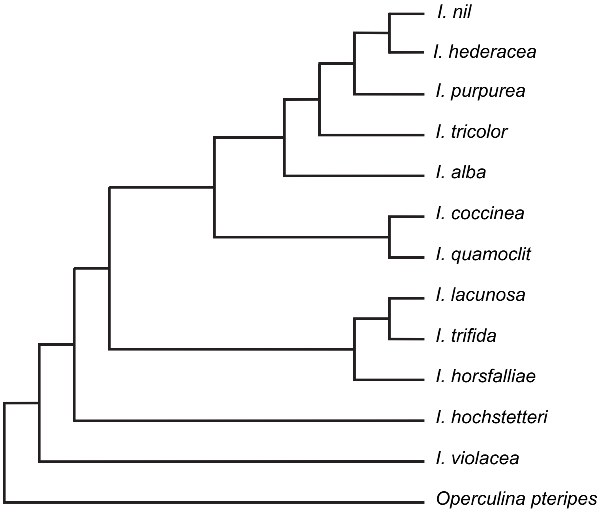

FIG. 2.—A Bayesian phylogeny of the 13 species used in this study, derived from ITS sequence data. This topology was used in all analyses. Branch lengths connote no information.

Bayesian phylogenetic reconstructions of the ITS sequences for these 13 taxa recovered the established species relationships (fig. 2; Miller et al. 1999). The individual *bHLH* gene trees also consistently recovered the relationships among species. This indicates that the gene sequences we analyzed were likely to be orthologous, but it should be noted that the molecular genetic analyses to test protein function have not been performed for most species. However, conserved function of these bHLH transcription factors has been maintained since the split of the monocots and dicots (Lloyd et al. 1992; Quattrocchio et al. 1993), suggesting similar function among species of *Ipomoea*. The only major incongruity among the gene trees was in the placement of *Ipomoea hochstetteri*. The *bHLH1* gene tree reconstructed it as subtending the clade containing *I. nil* and *Ipomoea hederaceae*, whereas the ITS and *bHLH2* consensus trees placed it basal to all *Ipomoea* species except *Ipomoea violaceae*. For all further analyses, we used the topology derived from the ITS sequences (fig. 2), although analyses using each *bHLH* gene tree provided qualitatively similar results (data not shown).

Domain-Specific Differences in Rates of Amino Acid Substitution

The nonbinding domain shows elevated ω ratios relative to the binding domain in *bHLH2* but not in *bHLH1* (table 2). In *bHLH2*, the ω ratio in the nonbinding domain is more than 2 times greater than the binding domain, whereas the difference between the domains in *bHLH1*

Table 2
Estimates of ω from Codon-Based Model M0 of PAML, Partitioned for the Binding and Nonbinding Domains for the *bHLH1* and *bHLH2* Genes

	Binding Domain		Nonbinding Domain	
Model	ω	*l*	ω	*l*
bHLH1				
Model M0	0.136	1,761.08	0.207	3,513.46
Freely estimated	(0.150)	1,761.20	(0.150)	3,513.51
Constrained	(0.171)	1,761.79	(0.171)	3,514.93
	(0.200)	1,762.12	(0.200)	3,513.50
LRTs (df = 1)[a]		$\Lambda < 2.08$ $P > 0.05$		$\Lambda < 2.94$ $P > 0.05$
bHLH2				
Model M0				
Freely estimated	0.130	2,139.68	0.405	4,045.09
Constrained	(0.276)	2,151.53	(0.276)	4,045.09
LRTs (df = 1)[a]		$\Lambda = 23.7$ $P < 0.0001$		$\Lambda = 18.74$ $P < 0.0001$

[a] The log likelihood (*l*) of the data for the model where ω was estimated was compared with the log likelihood of the same model where ω was constrained to intermediate values, and significance of the models was evaluated using LRTs with 1 df. Statistically significant differences between the domains of each gene are evaluated by attempting to find an intermediate value of ω that is significantly different than the estimated ω for each domain (Lu and Rausher 2003).

is more modest and does not vary significantly (table 2). To further characterize differences in domain-specific rates of amino acid substitution, we determined whether the 15% of sites with the highest ω ratio in each gene, as described by Bayesian estimation, are located more frequently in the nonbinding domains. *G*-tests indicate that these rapidly evolving sites do occur more frequently in the nonbinding domains for both genes (*bHLH1*: $G = 11.26$, df = 1, $P < 0.001$; *bHLH2*: $G = 26.87$, df = 1, $P < 0.0001$), suggesting that rapidly evolving sites in both genes are significantly clustered in the nonbinding domains.

In addition, indel mutations occur more frequently in the nonbinding domains. For *bHLH1*, no indels are present in either the interaction or the bHLH domains, but 7 indels occur in the nonbinding domains. The *bHLH2* gene contains considerably more indels than *bHLH1*, and they again appear to avoid the binding domains (binding = 3; nonbinding = 26). Finally, in regions outside of indels, invariant amino acid sites are found more frequently in the binding domains as compared with the nonbinding domains for both *bHLH1* ($G = 6.09$, df = 1, $P < 0.05$) and *bHLH2* ($G = 14.64$, df = 1, $P < 0.001$), indicating that rates of evolution differ between the domains. Although these data are consistent with previous analyses of plant transcription factors showing that certain domains have elevated rates of evolution relative to other regions, the overall low ω ratios in the nonbinding domains of each bHLH regulator indicate strong constraint on these regions as well (table 2).

Rates and Patterns of Amino Acid Substitution

In order to determine the selective forces operating on these *bHLH* genes, we first examined whether there was heterogeneity in d_N/d_S ratios among lineages for each gene by performing branch tests. LRTs comparing single-ratio and free-ratio models do not differ significantly for either gene (*bHLH1*: $\Lambda = 30.46$, df = 22, $P > 0.1$; *bHLH2*: $\Lambda = 20.7$, df = 22, $P > 0.5$), providing no reason to reject the null hypothesis that the genes have evolved at constant rates along lineages.

We then compared models that evaluate the evolutionary forces acting on individual codon sites within each gene. For *bHLH1*, Models M1a, M2a, M3, M7, and M8 of PAML all result in nearly identical log-likelihood scores (table 3), providing no reason to reject Model M1a that classifies sites as either highly constrained or evolving neutrally. Model M1a indicates that 85% of sites are located in the highly constrained site class, with ω estimated as 0.087. Furthermore, for both of the selection models (M2a and M8) and the discrete Model (M3), estimates of ω are not greater than 1 for any site class, again providing no evidence for positive selection in the *bHLH1* gene. Empirical Bayes estimation of individual amino acid sites belonging to the "selected" site class in Model M8 indicates that only 9 amino acids sites throughout the entire gene have an estimated ω greater than 1. For these sites, the maximum ω is 1.29 ± 0.369, and none of these sites is estimated to have a posterior probability greater than 0.73 of belonging to the rapidly evolving site class. These results provide consistently strong evidence that the *bHLH1* gene has experienced substantial selective constraint and little evidence for positive selection at any amino acid site.

A similar analysis of the *bHLH2* gene provides nearly identical results to *bHLH1* and is unable to identify putatively, positively selected sites in this gene either. Models M1a–M3 result in log-likelihood scores that are indistinguishable from each other (table 3). Model M1a estimates that approximately 82% of sites are highly constrained with $\omega = 0.136$. Even though additional parameters are estimated in the M2a and M3 models, they do not fit the data significantly better than a model that assumes all sites to be either highly constrained or evolving neutrally. In fact, the discrete model (M3), which freely estimates site classes based on the data, defines only 2 different classes, one with $\omega = 0.15$ and the other with $\omega = 1.19$, a value not substantially greater than 1. Approximately 85% of the amino acid sites are assigned to the more constrained class of sites in this model. Model M8 fits the data significantly better than Model M7 ($\Lambda = 11.0$, df = 2, $P < 0.01$), and because 15% of sites are located in the positively selected site class with $\omega = 1.20$, weak positive selection may be indicated with these comparisons. However, it has been found that a poor fit of the data to a beta distribution may result in a high frequency of significant tests when comparing Models M7 and M8 even in the absence of positive selection. As a result, Swanson et al. (2003) devised a likelihood ratio test to account for these apparently elevated type I error rates. Briefly, the original Model M8 that contains an extra site class constrained to have $\omega \geq 1$ is compared with a more restricted null model (M8a), where the extra site class is constrained to have $\omega = 1$. This test has the added benefit of directly testing whether the estimated ω in the extra site class is significantly greater than 1. The likelihood ratio score is compared with a χ^2 distribution with 1 df. When performing

Table 3
Results of Codon-Based Models of Molecular Evolution for the *bHLH1* and *bHLH2* Genes in *Ipomoea*, Using the Tree Topology Derived from ITS Sequence Data

Model	l^a	ω^b	Parameters^c
bHLH1			
M0 (single ω)	5,338.11	0.192	$\omega = 0.192$
M1a (nearly neutral)	5,311.71	0.222	$\omega_0 = 0.087$ ($p_0 = 0.857$) $\omega_1 = 1.000$ ($p_1 = 0.143$)
M2a (selection)	5,311.71	0.218	$\omega_0 = 0.087$ ($p_0 = 0.857$) $\omega_1 = 1.000$ ($p_1 = 0.092$) $\omega_2 = 1.000$ ($P_2 = 0.051$)
M3 (discrete)	5,310.95	0.206	$\omega_0 = 0.062$ ($p_0 = 0.000$) $\omega_1 = 0.062$ ($p_1 = 0.785$) $\omega_2 = 0.728$ ($p_2 = 0.215$)
M7 (β)	5,311.22	0.204	$\alpha = 0.258$; $\beta = 0.997$
M8 (β + ω)	5,311.20	0.206	$\alpha = 0.309$; $\beta = 1.501$ $p_0 = 0.955$ $P_1 = 0.046$; $\omega_1 = 1.000$
bHLH2			
M0 (single ω)	6,203.35	0.289	$\omega = 0.289$
M1a (nearly neutral)	6,162.72	0.296	$\omega_0 = 0.136$ ($p_0 = 0.815$) $\omega_1 = 1.000$ ($p_1 = 0.185$)
M2a (selection)	6,162.27	0.308	$\omega_0 = 0.151$ ($p_0 = 0.849$) $\omega_1 = 1.000$ ($p_1 = 0.000$) $\omega_2 = 1.190$ ($p_2 = 0.151$)
M3 (discrete)	6,162.27	0.308	$\omega_0 = 0.151$ ($p_0 = 0.218$) $\omega_1 = 0.151$ ($p_1 = 0.631$) $\omega_2 = 1.190$ ($p_2 = 0.151$)
M7 (β)	6,167.82	0.293	$\alpha = 0.404$; $\beta = 0.972$
M8 (β + ω)	6,162.32	0.308	$\alpha = 17.98$; $\beta = 99$ $p_0 = 0.852$ $P_1 = 0.148$; $\omega_1 = 1.201$

[a] Log likelihood of the data.

[b] Mean d_N/d_S ratio for the entire domain.

[c] p_i, the proportion of codons that fall in each category.

this analysis with the sequence data from the *bHLH2* gene, Model M8 does not significantly differ from Model M8a ($\Lambda = 0.94$, df = 1, $P > 0.3$), indicating that the estimated $\omega = 1.20$ is not significantly different than 1 and that there is little indication of positive selection in this gene either.

Analysis of Amino Acid Substitutions

Methods relying on d_N/d_S ratios typically have low power to detect positive selection, especially if selection targeted only a limited number of amino acid sites or occurred infrequently. As a result, a new class of models that evaluates the effects of amino acid substitutions on the magnitude of change in physicochemical amino acid properties has recently been developed. We used the program TreeSAAP (Wooley et al. 2003) to model how 31 different physicochemical properties were affected by amino acid substitutions in both *bHLH* genes. Consistent overrepresentation of radical amino acid changes (i.e., categories 7 and 8) would suggest repeated adaptive substitution (McClellan et al. 2005).

According to both goodness-of-fit tests and *Z*-scores, amino acid substitutions in the *bHLH1* gene affect the magnitude of change of 4 physicochemical properties more frequently than would be expected if all substitutions are equally likely to occur (table 4). Conservative substitutions of magnitude category 3 occur more frequently than expected by chance for 2 of the properties (β-structure tendencies and hydropathy), whereas slightly more radical substitutions (i.e., those found in category 6) affecting alpha-helical tendencies and the power to be at the C-terminal also are overrepresented in the data set (table 4). However, for none of the properties do the most radical amino acid substitutions (i.e., those found in categories 7 and 8) occur more frequently than expected under neutrality, indicating that these substitutions do not appear to have large effects on the physical structure of the protein. These results provide additional evidence that the entire *bHLH1* gene is highly constrained and has not experienced repeated episodes of adaptive evolution.

For *bHLH2*, 13 different properties show significant deviations from neutral expectations (table 4). Interestingly, for 12 of these 13 properties, amino acid substitutions of particular magnitude categories occur less frequently than expected by chance, suggesting that purifying selection eliminates these classes of substitutions. Of these 12 properties, 3 of them also show an overrepresentation of particular types of amino acid substitutions, but these occur only in the most highly conservative classes (categories 1–3). For only one property (turn tendencies) are marginally radical amino acid substitutions (category 6) overrepresented. For 2 properties related to polarity, radical amino acid substitutions from categories 7 and 8 occur significantly less often than expected due to random substitution, suggesting that drastically altering the polarity of the protein is strongly selected against. Consistent with our analyses comparing d_N/d_S ratios, these results indicate that amino acid substitutions have not radically altered the physicochemical makeup of the protein, providing little evidence that repeated positive selection in either *bHLH* gene contributed to adaptive differentiation of flower color in *Ipomoea*.

Comparing d_N/d_S between Genes

In order to determine whether *bHLH1* and *bHLH2* evolved at different rates, we followed previously described methods from Lu and Rausher (2003). Model M0 from PAML estimates the ω ratio for *bHLH1* and *bHLH2* to be 0.192 and 0.289 for the 2 genes, respectively. The log-likelihood scores for these models each fit the data significantly better than a model, where ω is constrained to the intermediate value of $\omega = 0.241$ (*bHLH1*: $\Lambda = 5.84$, df = 1, $P < 0.02$; *bHLH2*: $\Lambda = 5.3$, df = 1, $P < 0.03$). Therefore, we have 1) found a value of ω that lies in between the estimated values for each gene and 2) that value of ω is significantly different than the values individually estimated by PAML for each gene, satisfying both criteria and allowing us to conclude that the d_N/d_S ratios for *bHLH1* and *bHLH2* are significantly different from each other.

A second method for comparing the d_N/d_S ratio between the genes provides similar results, consistent with the idea that the *bHLH2* gene is evolving faster than *bHLH1*. By evaluating a combined data set partitioned according to each gene, we observe significant differences between Models C and E ($\Lambda = 18.9$, df = 2, $P < 0.0001$), indicating that both the transition/transversion rate ratio (κ) and ω vary significantly between the genes. The

Table 4
A List of Statistically Significant Physicochemical Amino Acid Properties for A) *bHLH1* and B) *bHLH2*, As Implemented in the TreeSAAP v. 3.2 Software Program (Wooley et al. 2003)

Property	Goodness of Fit[a]	*Z*-score (Category)[b]
bHLH1		
α-Helical tendencies	59.88	6.70 (6)
β-Structure tendencies	41.65	−3.40 (2); 3.12 (3)
Hydropathy	29.44	3.62 (3)
Power to be at the C-terminal	32.88	4.17 (6)
bHLH2		
Buriedness	28.28	−3.44 (6)
Chromatographic index	56.21	3.60 (1); −3.61 (6); −3.70 (7)
Helical contact area	46.84	3.64 (2); −4.06 (4)
Isoelectric point	28.59	−3.52 (6)
Long-range nonbonded energy	45.86	−3.19 (4)
Molecular volume	37.43	3.28 (2); −3.34 (4)
Polar requirement	48.14	−4.08 (5); −3.13 (8)
Polarity	45.32	−3.10 (4); −3.28 (7)
Refractive index	30.98	−3.18 (4)
Solvent accessible reduction ratio	30.75	−3.94 (6)
Surrounding hydrophobicity	31.16	−3.19 (6)
Thermodynamic transfer hydrophobicity	47.22	4.18 (1); −3.24 (4)
Turn tendencies	39.99	4.27 (6)

[a] χ^2 goodness-of-fit statistic to determine whether observed and expected distributions of amino acid substitutions are significantly different from each other. The critical value for rejecting the null hypothesis of neutrality with $P = 0.001$ and 7 df is 24.32.

[b] *Z*-score statistics for significant magnitude categories (in parentheses). A category with a low number indicates conservative changes to the protein, whereas higher categories denote more radical substitutions. The critical values for $P = 0.001$ are 3.09, indicating positive selection on that magnitude category, and −3.09, which indicates negative selection.

estimated ω is higher for *bHLH2* than for *bHLH1* (0.289 vs. 0.193) and κ is higher for *bHLH1* than for *bHLH2* (2.47 vs. 1.94).

The higher estimated d_N/d_S ratio for *bHLH2* can be attributed to an increase in the nonsynonymous substitution rate as opposed to a decrease in the synonymous substitution rate. When Model M0 is run for each gene separately in the site-based analyses, the tree lengths for synonymous and nonsynonymous substitutions differ. The ratio of *bHLH2* to *bHLH1* tree lengths for synonymous substitutions (synonymous substitutions per codon) is 1.12, indicating that synonymous substitutions occur at roughly similar rates between the 2 loci. However, the ratio of nonsynonymous tree lengths between *bHLH2* and *bHLH1* is 1.68, which is 50% greater than the rate of synonymous substitution. These different rates between the genes as a whole can be attributed to a higher d_N/d_S ratio in the nonbinding domains. When we compare only the binding domains of each gene, we find no significant difference between the genes for Models C and E ($\Lambda = 0.98$, df = 2, $P > 0.6$). The freely estimated ω ratios for each gene from Model E are similar (*bHLH1*: 0.136; *bHLH2*: 0.130). However, when we run the analysis only comparing the nonbinding domains between the genes, we obtain a significant likelihood ratio statistic ($\Lambda = 36.68$, df = 2, $P < 0.0001$). The estimated ω ratio for this region of the 2 genes is nearly twice as high for *bHLH2* than for *bHLH1* (0.408 vs. 0.210). These differences are due to greater rates of nonsynonymous substitution in *bHLH2* as the ratio of the nonsynonymous tree lengths is nearly twice the ratio of the synonymous tree lengths (1.751 vs. 0.893). These results indicate that *bHLH2* is experiencing significantly reduced selective constraint as compared with *bHLH1* and this difference is restricted to the nonbinding domains. Furthermore, measures of nucleotide composition were similar among species and among genes, suggesting that selection for codon usage does not appear to account for differences in subsitution rates between the genes.

Discussion

Regulatory Gene Evolution

Recent investigations into the patterns of molecular evolution of plant transcription factors have revealed 2 common patterns (Purugganan and Wessler 1994; Purugganan et al. 1995; Rausher et al. 1999; Langercrantz and Axelsson 2000; Barrier et al. 2001; Remington and Purugganan 2002): 1) transcription factors frequently exhibit elevated rates of nonsynonymous substitutions compared with the structural genes they regulate and 2) elevated rates are often restricted to certain domains, whereas other regions remain highly conserved. Purugganan (1998) has termed these patterns "rapid mosaic" evolution, but the functional significance of these regions remains an open question. In some cases, it is clear that the rapidly evolving regions play essential roles in protein function (Gong et al. 1999; Kroon 2004), whereas in others it appears that most substitutions are essentially neutral (e.g., Chang et al. 2005). Results from our study largely conform to the pattern of rapid mosaic evolution, although the rate acceleration in the nonbinding domains is not as great as is commonly observed. In *bHLH2*, the rate in the nonbinding domain (0.41) is approximately 2-fold lower than that exhibited by variable regions in some transcription factors (e.g., 0.8 in a *bHLH* gene in the *Poaceae*; Purugganan and Wessler 1994), but comparable to that exhibited by the *tb1* gene in grasses (0.39; Lukens and Doebley 2001). Moreover, the d_N/d_S ratio in the nonbinding domain of *bHLH2* in *Ipomoea* is substantially higher than that in the structural genes of the anthocyanin pathway (0.034–0.277; Lu and Rausher 2003). By contrast, although rapidly evolving sites in *bHLH1* are clustered in the nonbinding domains, the d_N/d_S ratio is not significantly different than that of the binding domains (table 2). Furthermore, the d_N/d_S ratios for both the binding and nonbinding domains of *bHLH1* fall within the range of values for the structural genes. Therefore, *bHLH1* does not appear to exhibit as rapid evolution in nonbinding domains as many plant transcription factors, including *bHLH2*.

Our results for *bHLH1* also differ somewhat from those reported for its orthologous copies in the *Poaceae* (Purugganan and Wessler 1994) and in *Cornus* (Fan et al. 2004). Although the *Poaceae* study reports very similar d_N/d_S ratios for the binding domains (0.14 for *Ipomoea* vs.

0.12 for *Poaceae*), the ratio in the nonbinding domains is 4.5 times greater (0.89 vs. 0.21) for the *Poaceae* than for *Ipomoea*. In *Cornus*, the ratios in all 4 of the individual domains (interaction = 0.20; acidic = 0.48; bHLH = 0.38; C-terminal = 0.42; Fan et al. 2004) show somewhat higher values than either the binding or nonbinding domains in *Ipomoea*. However, despite sites in *Cornus* with estimated ω ratios that approach 1.0, no class of sites consistently demonstrated positive selection. Because attempts to detect selection on the grass *bHLH* genes have not been undertaken, it is currently unknown whether the more rapid rate of evolution in the *Poaceae* is due to relaxed constraint or an increase in the frequency of adaptive substitution.

Adaptive Evolution

The primary objective of this study was to determine whether accelerated rates of evolution in *bHLH1* and *bHLH2* nonbinding domains are due primarily to increased positive selection or decreased constraint. Our results provide very little support for the positive selection hypothesis. Although within the nonbinding domains approximately 20% of sites exhibit high ω ratios, in neither paralogs is there a class of sites for which ω is significantly greater than 1.0. Moreover, in neither copy do radical amino acid substitutions occur in substantially greater frequencies than would be expected by chance. Additionally, the absence of indels in the binding domains implies that these regions are structurally more constrained than the nonbinding domains. In fact, it has been shown that indel mutations in the binding domains can reduce or completely abolish transcription of target anthocyanin genes in *bHLH* orthologs of maize and *P. hybrida* (Liu et al. 1998; Spelt et al. 2002).

Further evidence for relaxed constraint is provided by the observation that ectopic expression of the maize *R*-like bHLH anthocyanin regulator (which appears orthologous to *bHLH1* in *Ipomoea*) elicits expression of anthocyanin pigmentation in *Petunia*, *Nicotiana*, and *Arabidopsis* (Lloyd et al. 1992; Quattrocchio et al. 1993). Despite the maintenance of function between these widely diverged taxa for over 100 Myr, there is little sequence homology among species in the nonbinding domains, regions known to be essential for transcriptional activation and dimerization with other bHLH proteins. This pattern suggests that proper function of these domains appears to depend not on particular amino acids being present at particular locations but more loosely on conservation of general properties of these regions such as size, shape, and charge. This is evident from the significant underrepresentation of particular classes of substitutions affecting 12 physicochemical amino acid properties in *Ipomoea bHLH2* (table 4), suggesting that these properties remain constrained even though amino acid substitutions are common.

These patterns are very similar to those found for another anthocyanin pathway transcription factor in *Ipomoea*. Chang et al. (2005) report that despite a very high d_N/d_S ratio (0.80) in the nonbinding domain of *Ipmyb1*, adaptive substitution could not be detected either by analysis of d_N/d_S ratios or by analysis of properties of substituted amino acids. Based on this evidence, it seems doubtful that the elevated rates of substitution in the nonbinding regions of these genes contribute substantially to adaptive floral color differentiation in *Ipomoea*. More generally, the failure to detect adaptive substitution in other plant transcription factors (e.g., Lukens and Doebley 2001; Remington and Purugganan 2002; Fan et al. 2004), along with the common observation that indels are restricted to nonbinding regions (Lukens and Doebley 2001; Chang et al. 2005), suggests that constraint is relaxed and most substitutions are neutral in variable domains of these genes. By contrast, positive selection has been implicated in the evolution of substitutions in the binding domains of several plant transcription factors (Jia et al. 2003, 2004). These patterns suggest that, to the extent that phenotypic change is caused by transcription factor evolution, it will involve primarily substitutions within conserved binding domains.

Functional Constraint

A secondary objective of this investigation was to test the hypothesis that *bHLH2* would evolve more slowly than *bHLH1* because of its greater pleiotropy. In both *Ipomoea* and *Petunia*, multiple functions are ascribed to *bHLH2* and its orthologs, including production of seed coat color and trichomes, acidification of the vacuole, and production of secondary metabolites involved in plant defense (Spelt et al. 2002; Park et al. 2004, 2007). In contrast, regulation of anthocyanin synthesis is the only function currently assigned to the orthologs of *bHLH1* in the flowering plant species that have been characterized (Goff et al. 1992; Quattrocchio et al. 1998; Spelt et al. 2000).

Surprisingly, we observed the opposite pattern: in *Ipomoea*, the *bHLH2* gene evolves at a significantly higher rate than *bHLH1*, due primarily to an increase in the rate of nonsynonymous substitutions in its nonbinding domains. These apparently conflicting results potentially can be resolved when we consider that the orthologs of *bHLH1* are involved in more protein–protein interactions than orthologs of *bHLH2* (Kroon 2004). This suggests that despite more pleiotropic effects of *bHLH2* relative to *bHLH1*, the nonbinding domains of the *Ipomoea bHLH2* gene may be released from some of the constraint associated with ensuring that all protein-binding sites are maintained. Therefore, this may account for the observed significant increase in the ω ratio in *bHLH2* relative to *bHLH1*.

Supplementary Material

Supplementary materials are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

We thank L. Xu for assistance in the laboratory; R. Miller for providing ITS sequence data; S. Smith and D. Des Marais for helpful discussions with analyses; and D. McClellan for assistance with the TreeSAAP software.

This work was supported by an award from the National Science Foundation.

Literature Cited

Barrier M, Robichaux RH, Purugganan MD. 2001. Accelerated regulatory gene evolution in an adaptive radiation. Proc Natl Acad Sci USA. 98:10208–10213.

Carroll SB. 2005. Evolution at two levels: on genes and form. PLoS Biol. 3:1159–1166.

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.

Doebley J. 1993. Genetics, development and plant evolution. Curr Opin Genet Dev. 3:865–872.

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

Fan C, Purugganan MD, Thomas DT, Wiegmann BM, Xiang QY. 2004. Heterogeneous evolution of the *Myc-like* anthocyanin regulatory gene and its phylogenetic utility in *Cornus* L. (Cornaceae). Mol Phylogenet Evol. 33:580–594.

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. Gene Dev. 6:864–875.

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 Mol Biol. 41:33–44.

Heim MA, Jakoby M, Werber M, Martin C, Weisshaar B, Bailey PC. 2003. The basic helix-loop-helix transcription factor family in plants: a genome-wide study of protein structure and functional diversity. Mol Biol Evol. 20:735–747.

Hernández-Hernández T, Martínez-Castilla LP, Alvarez-Buylla ER. 2007. Functional diversification of B MADS-box homeotic regulators of flower development: adaptive evolution in protein-protein interaction domains after major gene duplication events. Mol Biol Evol. 24:465–481.

Hocking DR. 1985. The analysis of linear models. Monterey (CA): Brooks/Cole Pub. Co.

Huelsenbeck JP, Ronquist F. 2001. MRBAYES: Bayesian inference of phylogeny. Bioinformatics. 17:754–755.

Jia L, Clegg MT, 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 MT, 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.

King MC, Wilson AC. 1975. Evolution at two levels in humans and chimpanzees. Science. 188:107–116.

Koes RE, Quattrocchio F, Mol JNM. 1994. The flavonoid biosynthetic pathway in plants: function and evolution. Bioessays. 16:123–132.

Kosakovsky Pond SL, Frost SDW. 2005. Not so different after all: a comparison of methods for detecting amino acid sites under selection. Mol Biol Evol. 22:1208–1222.

Kroon AR. 2004. Transcription regulation of the anthocyanin pathway in *Petunia hybrida* [Dissertation]. [Amsterdam (Netherlands)]: Vrije Universiteit.

Langercrantz U, Axelsson T. 2000. Rapid evolution of the family of *CONSTANS LIKE* genes in plants. Mol Biol Evol. 17:1499–1507.

Liu Y, Wang L, Kermicle J, Wessler S. 1998. Molecular consequences of *Ds* insertion into and excision from the helix-loop-helix domain of the maize *R* gene. Genetics. 150:1639–1648.

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. Mol Biol Evol. 20:1844–1853.

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

Machado ICS, Sazima M. 1987. A comparative study in floral biology of two weed species *Ipomoea hederifolia* and *I. quamoclit* Convolvulaceae. Rev Bras Biol. 47:425–436.

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. Syst Bot. 26:585–602.

Massari ME, Murre C. 2000. Helix-loop-helix proteins: regulators of transcription in eucaryotic organisms. Mol Cell Biol. 20:429–440.

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. Mol Biol Evol. 22:437–455.

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

Mol J, Grotewold E, Koes R. 1998. How genes paint flowers and seeds. Trends Plant Sci. 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 Physiol. 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.

Park KI, Choi J, Hoshino A, Morita Y, Iida S. 2004. An intragenic tandem duplication in a transcriptional regulatory gene for anthocyanin biosynthesis confers pale-colored flowers and seeds with fine spots in *Ipomoea tricolor*. Plant J. 38:840–849.

Park KI, Ishikawa N, Morita Y, Choi JD, Hoshino A, Iida S. 2007. A *bHLH* regulatory gene in the common morning glory, *Ipomoea purpurea*, controls anthocyanin biosynthesis in flowers, proanthocyanidin and phytomelanin pigmentation in seeds, and seed trichome formation. Plant J. 49:641–654.

Posada D, Crandall KA. 1998. Modeltest: testing the model of DNA substitution. Bioinformatics. 14:817–818.

Purugganan MD. 1998. The molecular evolution of development. Bioessays. 20:700–711.

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, 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 JF, van der Woude K, Mol JNM, 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.

Rausher MD. 2006. The evolution of flavonoids and their genes. In: Grotewold E, editor. The science of flavonoids. New York: Springer. p. 175–212.

Rausher MD, Fry JD. 1993. Effects of a locus affecting floral pigmentation in *Ipomoea purpurea* on female fitness components. Genetics. 134:1237–1247.

Rausher MD, Miller RE, Tiffin P. 1999. Patterns of evolutionary rate variation among genes of the anthocyanin biosynthetic pathway. Mol Biol Evol. 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. Mol Biol Evol. 19:1563–1574.

Ronquist F, Huelsenbeck JP. 2003. MYBAYES 3: Bayesian phylogenetic inference under mixed models. Bioinformatics. 19:1572–1574.

Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics. 19:2496–2497.

Sokal RR, Rohlf FJ. 1995. Biometry, 3rd ed. New York: WH Freeman and Co.

Spelt C, Quattrocchio F, Mol J, Koes R. 2002. ANTHOCYANIN1 of Petunia controls pigment synthesis, vacuolar pH, and seed coat development by genetically distinct mechanisms. Plant Cell. 14:2121–2135.

Spelt C, Quattrocchio F, Mol JNM, Koes R. 2000. *anthocyanin1* of Petunia encodes a basic helix-loop-helix protein that directly activates transcription of structural anthocyanin genes. Plant Cell. 12:1619–1631.

Swanson WJ, Nielsen R, Yang Q. 2003. Pervasive adaptive evolution in mammalian fertilization proteins. Mol Biol Evol. 20:18–20.

Weigel D, Meyerowitz EM. 1993. Activation of floral homeotic genes in *Arabidopsis*. Science. 261:1723–1726.

Wooley S, Johnson J, Smith MJ, Crandall KA, McClellan DA. 2003. TreeSAAP: selection on amino acid properties using phylogenetic trees. Bioinformatics. 19:671–672.

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

Yang Z, Nielsen R. 2000. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. Mol Biol Evol. 17:32–43.

Yang Z, Nielsen R, Goldman N, Pedersen AM. 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics. 155:431–449.

Yang Z, Swanson WJ. 2002. Codon-substitution models to detect adaptive evolution that account for heterogeneous selective pressures among site classes. Mol Biol Evol. 19:49–57.

Patricia Beldade, Associate Editor

Accepted September 27, 2007