J Mol Evol (2008) 67:137–144
DOI 10.1007/s00239-008-9105-5

Variation in Constraint Versus Positive Selection as an Explanation for Evolutionary Rate Variation Among Anthocyanin Genes

Mark D. Rausher · Yingqing Lu · Kyle Meyer

Received: 9 October 2007 / Accepted: 18 March 2008 / Published online: 25 July 2008
© Springer Science+Business Media, LLC 2008

Abstract It has been argued that downstream enzymes in metabolic pathways are expected to be subject to reduced selective constraint, while upstream enzymes, particularly those at pathway branch points, are expected to exhibit more frequent adaptive substitution than downstream enzymes. We examined whether these expectations are met for enzymes in the anthocyanin biosynthetic pathway in *Ipomoea.* Previous investigations have demonstrated that downstream enzymes in this pathway have substantially higher rates of nonsynonymous substitution than upstream enzymes. We demonstrate here that the difference in rates between the most upstream enzyme (CHS) and the two most downstream enzymes (ANS and UFGT) is explained almost entirely by differences in levels of selective constraint. Adaptive substitutions were not detected in any of these genes. Our results are consistent with suggestions that constraint is greater on enzymes with greater connectivity.

Keywords Anthocyanin pathway · Evolutionary rate variation · Evolutionary constraint · Positive selection · *Ipomoea*

M. D. Rausher (✉) · K. Meyer
Department of Biology, Duke University, Box 90338, Durham, NC 27708-0338, USA
e-mail: mrausher@duke.edu

Y. Lu
State Key Laboratory of Systematics and Evolution, Institute of Botany, Chinese Academy of Sciences, 20 Nanxincun, Xiangshan, Beijing 100093, China

Present Address:
K. Meyer
Department of Biological Sciences, University of Wisconsin—Milwaukee, Lapham Hall, 3209 North Maryland Avenue, Milwaukee, WI 53211, USA

Introduction

Proteins vary over several orders of magnitude in the rate of amino acid substitution (Li et al. 1985), and accounting for this variation has been a major objective of the study of molecular evolution. At one extreme, very rapid, repeated substitution often occurs in reproductive proteins, which typically exhibit dN/dS ratios of $\gg 1$ (Whitefield et al. 1993; Wyckoff et al. 2000; Swanson et al. 2001; Lu 2002). At the other extreme, histones and other proteins believed to be under extreme constraint exhibit low substitution rates and dN/dS ratios close to 0. The vast majority of proteins, however, exhibit both intermediate substitution rates and dN/dS ratios between 0 and 1 (Li 1997). While evolutionary rate variation among this class of proteins is typically ascribed to differences in degree of selective constraint [i.e., differences in the strength of purifying selection (Kimura 1977; Li 1997; Hirsh and Fraser 2001; Hahn and Kern 2005; Lemos et al. 2005; Lin et al. 2007)], it may also be due to variation in the frequency of adaptive substitution. Because very few studies have attempted to distinguish explicitly between these two possibilities, however, especially for genes exhibiting dN/dS ratios <1, very little is know about the relative contribution of differential constraint and differential adaptation to variation in evolutionary rates (Flowers et al. 2007)

A number of recent investigations have suggested that both the degree of constraint and the rate of adaptive substitution in proteins, and hence the overall rate of substitution, are likely influenced not just by intrinsic properties of particular proteins, but also by their network properties. Such properties include position in a metabolic network (Eanes 1999; Rausher et al. 1999; Cork and Purugganan 2004; Flowers et al. 2007), level of expression (Pal et al. 2001), number of interacting protein partners (Fraser

et al. 2002; Hahn et al. 2004), and, for enzymes, number of other enzymes sharing the same metabolic reactants or products (Vitkup et al. 2006). While each of these properties has been shown to correlate with rate of protein evolution, it is seldom demonstrated directly that differences in evolutionary rates are due to differences in constraint rather than differences in rates of adaptive substitution.

Metabolic pathways constitute a convenient type of network for examining how position of an enzyme in a network influences both constraint and rate of adaptive substitution (Flowers et al. 2007). In metabolic pathways, the magnitude of the selective advantage of a mutation is expected to be proportional to its effect on flux through the pathway because the phenotype is expected to be proportional to flux. And because the probability of fixation of an adaptive mutation is approximately proportional to its selective advantage (Hedrick 2000), it has been argued that that natural selection primarily targets enzymes with the greatest control on flux (Hartl et al. 1985; Eanes 1999; Watt and Dean 2000). Although we are unaware of any formal theoretical analysis of the evolution of relative flux control among pathway enzymes, it has been argued that in branching pathways, flux allocation is controlled primarily by enzymes at pathway branch points (LaPorte et al. 1984; Stephanopoulos and Vallino 1991; Eanes 1999). If this is true, then adaptive substitutions are expected to be more common in enzymes at major pathway branch points than in downstream enzymes. This expectation has recently received some empirical support (Flowers et al. 2007).

A corollary of this argument is that slightly deleterious mutations should be more strongly selected against in enzymes at branch points than in downstream enzymes. As a consequence, slightly deleterious nonsynonymous mutations are expected to accumulate at a higher rate in downstream genes. In effect, differential flux control among pathway enzymes is expected to cause downstream enzymes to be under reduced selective constraint.

In this report, we test these expectations using three enzymes in the anthocyanin biosynthetic pathway. Anthocyanins constitute the largest class of floral pigments. The anthocyanin pathway consists of seven core enzymes (Van der Meer et al. 1993; Shirley 1996). Many of these enzymes, especially the upstream enzymes, are also responsible for the production of other flavonoid compounds, which are important both physiologically and ecologically for plants (Shirley 1996). By contrast, the two most downstream enzymes are responsible for only the production of anthocyanin pigments. Thus, alteration of upstream pathway enzymes is likely to be associated with a greater magnitude of deleterious pleiotropy than alteration of the downstream genes (Rausher et al. 1999; Lu and Rausher 2003). This phenomenon, along with the more general argument presented above, suggests that upstream enzymes are likely to be subject to greater selective constraint than downstream enzymes in this pathway.

Here we examine three anthocyanin-pathway enzymes: chalcone synthase (CHS), anthocyanidin synthase (ANS), and UDP flavonoid:3-*O*-glucosyltransferase (UFGT). CHS is the first committed enzyme in the anthocyanin pathway, which branches off the general phenylpropanoid pathway (Van der Meer et al. 1993). It is thus a branching enzyme in the sense of Flowers et al. (2007). By contrast ANS and UFGT are the two most downstream enzymes in the anthocyanin pathway.

Previously, we have reported that nonsynonymous substitution rates, as well as dN/dS ratios, are substantially higher in the two downstream enzymes than in CHS (Rausher et al. 1999; Lu and Rausher 2003). This pattern is consistent with the expectation that selection is relaxed in the downstream enzymes, but could also be explained by a higher rate of adaptive substitution in the downstream enzymes than in CHS. In a previous analysis (Lu and Rausher 2003), we attempted to distinguish between these two possibilities using codon-based tests of selection (Yang et al. 2000). We were unable to detect differences in positive selection between the downstream enzymes and CHS, suggesting that rate differences are caused by differences in the magnitude of selective constraint, although we provided no direct evidence of differential constraint. However, the ability to accurately detect positive selection using these methods may often be quite poor (Kosakovsky Pond and Frost 2005; McClellan et al. 2005). In particular, if adaptive substitutions are distributed across many amino acid sites such that dN/dS is not elevated above 1 for any site, these tests will not reveal positive selection. In this situation, alternative tests may have greater power to detect positive selection.

Here we extend our examination of the causes of evolutionary rate variation in anthocyanin enzymes by applying an alternative approach for distinguishing between differential constraint and differential positive selection. Specifically, we conduct two types of tests. The first examines within-species variation and asks whether there is direct evidence for differential constraint between CHS and the two downstream enzymes. The second asks whether these enzymes exhibit evidence of adaptive substitution and whether the rate of adaptive substitution differs among the enzymes.

Materials and Methods

Species Examined

Ipomoea trifida (Convolvulaceae) is a perennial vine that is widely distributed throughout central and northern South

America. Because this species is closely related to cultivated sweet potato, *I. batatas*, numerous accessions are available from these regions. We sampled two greenhouse-grown plants from each of 15 accessions. These accessions constituted 15 of the 16 whose collection locations are given by Chang et al. (2005) (we did not use accession PI 543819 in this study). Four accessions were from Colombia, two from Venezuela, three from Costa Rica, two from Guatemala, and four from Mexico. One copy of the gene for each enzyme examined was obtained from each of two plants from each accession. In addition, one copy of each gene was obtained from *I. purpurea* for determining between-species divergence (Chang et al. 2005). This species has more extensive and more intense floral pigmentation than *I. trifida*, indicating that some evolutionary divergence in anthocyanin production in flowers has occurred.

Enzymes Examined

Coding-region sequences were obtained for three genes coding for enzymes in the anthocyanin pathway: chalcone synthase (CHS), the most upstream enzyme in the pathway, which exhibits a low dN/dS ratio of 0.034 in *Ipomoea* (Lu and Rausher 2003); anthocyanidin synthase (ANS), the penultimate enzyme in the pathway (dN/dS = 0.197); and the most downstream enzyme, UDP-glucose flavonoid:3-oxy-glucosyltransferase (UFGT; dN/dS = 0.277). *ANS* and *UFGT* are single-copy genes in *Ipomoea*, whereas the copy of *CHS* examined, *CHS-D*, is one member of a five-member family of chalcone-synthase-like genes in *Ipomoea*. *CHS-D* is readily distinguishable in sequence from the other four members of this family and is the copy that is most highly expressed in most tissues and is necessary for floral pigmentation in *I. purpurea* (Fukada-Tanaka et al. 1997; Durbin et al. 2003). *CHS-D* sequences were those reported previously (Lu and Rausher 2003). *ANS* and *UFGT* sequences were newly obtained for this report.

Cloning and Sequencing

Based on known sequences of *ANS* and *UFGT* from *I. trifida* (Lu and Rausher 2003), we designed a PCR primer pair for each gene. The pair for *ANS* was P32 (CAACT GTTCCCAGCAGGGTG) and P42R (TTCATCAGGTT TGGGCGTGTCAGC), while the pair for *UFGT* was 3F (ACAGCCGAAGAAACCATTTC) and 42R (GCGTTTG CGCCGGGTAACTTT). These primers amplified an ~1073-bp segment of the 1134-bp coding region of *ANS* and an ~1036-bp segment of the 1370-bp coding region of *UFGT*. Genomic DNA was isolated for PCR using the DNeasy Plant Mini Kit (Qiagen) as per the manufacturer's instructions. PCR reactions were performed using the high-fidelity Plantinum *Pfx* DNA polymerase (Invitrogen). PCR fragments were cloned into the pcr-Blunt II-TOPO vector (Invitrogen) and sequenced off the vector using the Big Dye protocol (Applied Biosystems). Sequence data were collected by an ABI 3700 automated sequencer (Applied Biosystems). Sequences were deposited in GenBank (accession numbers EU852739–EU852798).

Analysis of Differential Constraint

If upstream enzymes are subject to greater constraint than downstream enzymes, two patterns ought to be observed. First, the amount of nonsynonymous variation within a species should be greater for the downstream enzymes, while the amount of synonymous variation should be similar for all enzymes. Second, the allele-frequency spectrum for nonsynonymous variation should exhibit fewer sites with intermediate frequencies in the upstream enzymes, compared to downstream enzymes, while synonymous polymorphisms should exhibit a neutral frequency spectrum.

To examine the first pattern, we calculated the average number of pairwise differences per site (π), and from this the dN/dS ratio, separately for synonymous and nonsynonymous variation, using DnaSP (Rozas et al. 1999). To examine the second pattern, we calculated Tajima's *D* statistic separately for both types of variation. The number of sites, the number of segregating sites, and π for each type of variation were calculated for each enzyme using DnaSP; these numbers were then used to calculate *D* using the formula given by Tajima (1989). Coalescent simulations using 1000 runs, as implemented by DnaSP, were used to determine whether *D* was significantly negative, as would be expected under purifying selection.

Analysis of Positive Selection

To determine whether any of the three enzymes exhibited evidence of positive selection, we performed a standard McDonald–Kreitman (1991) test on each gene, as implemented by DnaSP, using published sequences from *I. purpurea* for the estimation of divergence. Significance of the test results was determined by a two-way *G*-test and, in cases involving reduced samples, a Fisher exact test (Sokal and Rohlf 1969). To determine whether enzymes differed in the extent of positive selection, we used a three-way *G*-test (Bishop et al. 1975), in which the factors were (1) synonymous versus nonsynonymous substitutions, (2) polymorphism versus divergence, and (3) enzyme, to ascertain whether the magnitude of adaptive substitution differed for the three enzymes. In this analysis, a significant three-way interaction indicates that the excess number of nonsynonymous differences between species, relative to that expected under neutrality, differs for the three

enzymes. Analyses were performed using (1) all polymorphisms and (2) only polymorphisms for which the frequency of the mutant allele was >0.068. This frequency threshold was chosen based on preliminary analyses of the dN/dS ratios for different frequency classes as representing the most likely classes with elevated ratios. The latter analysis was undertaken to correct for effects of deleterious nonsynonymous alleles segregating at low frequencies (Fay et al. 2001; Shapiro et al. 2007) (see Results). Mutations were polarized based on *I. purpurea* sequences.

We also performed an HKA test (Hudson et al. 1987) for nonsynonymous substitutions using the program HKA by J. Hey (http://lifesci.rutgers.edu/~heylab/HeylabSoftware.htm). This test asks whether nonsynonymous divergence differs from neutrality for any of the genes examined, as would be expected if one or two of the genes experienced more frequent adaptive substitution.

Analysis of Geographic Substructure

Because the *I. trifida* accessions were distributed widely across central and northern South America, and because among-population differentiation can influence the results of the tests described above, we also used AMOVA, as implemented by Arlequin (Excoffier et al. 2005), to determine the degree of geographic structuring of variation in the genes for the three enzymes. Two comparisons were performed. One comparison divided the accessions into two geographic regions (Central America, including accessions from Mexico, Guatemala, and Costa Rica; and South America, including accessions from Colombia and Venezuela). The second comparison divided the accessions into three regions (Mexico, the rest of Central America, and South America). In this analysis, accessions were nested within region, and individuals were nested within accessions. AMOVA was used to calculate the proportion of variation associated with differences between regions (F_{CT}). In addition, we used DnaSP to determine the numbers of synonymous and nonsynonymous fixed differences between regions for each gene.

Results

General Characteristics of Sequence Variation

Of 30 sequences sampled from *I. trifida*, there were 25 unique haplotypes for *CHS*, 27 for *ANS*, and 26 for *UFGT*. The average pairwise number of differences (synonymous + nonsynonymous) was 15.6, 14.9, and 19.1 for the same genes, respectively. The total number of segregating sites (and total number of sites) was 70 (867), 66 (1032), and 112 (991), respectively. The dN/dS ratios for divergence between *I. trifida* and *I. purpurea* are 0.062, 0.261 and 0.277 for *CHS*, *ANS*, and *UFGT* respectively.

Analysis of Differential Constraint

Two patterns of nucleotide variation within *I. trifida* indicate that the downstream enzymes ANS and UFGT are subject to less constraint than the upstream enzyme CHS. First, the *dN/dS* ratios for polymorphisms in ANS and UFGT are 5.5 and 8 times higher, respectively, than the corresponding ratio for CHS (Table 1). This pattern arises because while the average pairwise difference per site for synonymous sites (π_s) is similar for all three enzymes, the corresponding values for nonsynonymous sites (π_n) are substantially higher for the downstream enzymes than for CHS (Table 1).

Second, the allele frequency spectra of these enzymes, as reflected in Tajima's *D* statistic, also exhibit a pattern consistent with relaxed constraint on the downstream enzymes. Purifying selection results in a negative value for *D* because it prevents slightly deleterious mutations from

Table 1 Synonymous and nonsynonymous variation in three anthocyanin-pathway enzymes for 15 accessions of *Ipomoea trifida*

Gene	Site type	No. polymorphic sites	No. sites[a]	π[b]	*dN/dS*[c]	Tajima's *D*	Conf int[d]	*p*
CHS	Syn	59	211.94	0.068	0.027	−0.119	(−1.60, 1.85)	0.52
	Nonsyn	11	655.06	0.002		−1.821	(−1.69, 1.88)	**0.014**
ANS	Syn	36	238.00	0.043	0.15	0.0424	(−1.64, 1.89)	0.73
	Nonsyn	30	793.92	0.006		−1.193	(−1.73, 1.74)	0.11
UFGT	Syn	59	235.65	0.047	0.220	−0.948	(−1.68, 164)	0.180
	Nonsyn	43	751.35	0.01		−1.028	(−1.65, 1.75)	0.15

Note: Syn, synonymous; nonsyn, nonsynonymous. Significant values of Tajima's *D* statistic indicated in boldface

[a] Average number of synonymous or nonsynonymous sites in enzyme

[b] Average pairwise differences per site

[c] Ratio π_{nonsyn}/π_{syn}

[d] Confidence interval for Tajima's *D* statistic based on coalescent simulations

rising to intermediate values. Moreover, as the strength of purifying selection increases, *D* becomes more negative. Because in both CHS and ANS synonymous and nonsynonymous polymorphisms exhibited significantly different *D* values, we analyzed these two classes of substitution separately. For synonymous polymorphisms in *I. trifida*, *D* is generally low and does not deviate significantly from neutrality for any of the enzymes (Table 1). For nonsynonymous polymorphisms, although *D* is negative for all three enzymes, it is most negative for CHS, intermediate for ANS, and least negative for UFGT. Additionally, only for CHS is *D* significantly different from zero. This pattern suggests that purifying selection is strongest on CHS.

Analysis of Positive Selection

A McDonald–Kreitman analysis of each enzyme individually provides no evidence of positive selection on any of the enzymes (Table 2). None of the tests were statistically significant (by either *G*-test or Fisher's exact test), suggesting that most amino acid substitutions between species were neutral. This conclusion is also supported by a three-way M–K test that analyzes all three proteins simultaneously (Table 3). The three-way interaction is not significant, indicating no difference among enzymes in the proportion of nonsynonymous fixed differences that are due to positive selection. Finally, the site-type (synonymous versus nonsynonymous) × variation-type (polymorphism versus fixed difference) effect is also very far from significant, indicating that even when the data from all three proteins are pooled, there is no detectable positive selection.

The significant Tajima's *D* statistic for nonsynonymous variation in CHS, as well as its near significance for ANS and UFGT, suggests that slightly deleterious nonsynonymous alleles may be segregating at a low frequency at these loci. Because this type of segregation may reduce the power to detect significance in an M–K test, we performed a modified M–K analysis, in which both synonymous and nonsynonymous sites at which alleles are segregating at a low frequency are eliminated from the data before performing the test (Fay et al. 2001; Shapiro et al. 2007). In CHS and ANS, the ratio of number of nonsynonymous sites to number of synonymous sites was more than four times higher for segregating sites with mutants at a frequency of <0.068 than for segregating sites with mutant frequencies >0.068 (Table 4). This difference is highly significant for both enzymes, indicating that sites with mutant alleles segregating at a frequency of ≤0.068 should not be included in the M–K analysis. For consistency, we also eliminated this class of sites from the data for UFGT, even though it did not exhibit a significantly elevated ratio in this protein (Table 4).

The modified M–K test remained nonsignificant for all three enzymes (Table 2). Moreover, the three-way interaction in the three-way test also remained nonsignificant by either *G*-test or Fisher exact test (Table 3). Adjusting for segregating, slightly deleterious mutations thus does not alter our original conclusion that there is no detectable history of positive selection acting on any of the enzymes.

Finally, an HKA test using only nonsynonymous polymorphisms and divergent sites was also not significant ($\chi^2 = 0.333$, df = 2, $p > 0.84$). This result is consistent with an absence of variation among the genes in the frequency of adaptive substitutions.

Table 3 Three-way *G*-test for differences among three anthocyanin-pathway enzymes in frequency of adaptive substitution

Source	df[c]	With all polymorphic sites[a]		Without low-frequency polymorphic sites[b]	
		G[d]	p	G	p
E × S × V	2	0.11	ns	1.53	ns
E × S	2	34.56	<0.0001	31.41	<0.0001
S × V	1	0.03	ns	1.67	ns
V × E	2	6.24	<0.05	3.71	ns

Note: E, enzyme; S, synonymous versus nonsynonymous; V, variable within species versus fixed difference between species

[a] Includes all polymorphic sites in analysis

[b] Includes only polymorphic sites with mutant frequency >0.068

[c] Degrees of freedom

[d] *G*-statistic

Table 2 McDonald–Kreitman analysis for three anthocyanin pathway enzymes in *I. trifida*

	CHS			ANS			UFGT		
	Polym	Adj	Fixed	Polym	Adj	Fixed	Polym	Adj	Fixed
Nonsynonymous	11	2	10	30	10	20	43	21	23
Synonymous	59	36	52	36	25	23	59	31	26
G-statistic	0.004	2.93		0.012	2.66		0.307	0.44	
	(ns)	(ns)		(ns)	(ns)		(ns)	(ns)	

Note: ns, nonsignificant. The outgroup was *I. purpurea*. Polym, number variable sites within *I. trifida*; Adj, number of variable sites with mutant frequency >0.068; Fixed, number of sites with fixed differences between species

Table 4 Number of segregating sites and ratios of number of nonsynonymous segregating sites to number of synonymous segregating sites for mutations in different frequency classes

Enzyme	Site type[a]	Frequency of mutant allele					p[b]
		<0.07	0.07–0.33	0.34–0.67	0.68–0.99	>0.07	
CHS	n	9	1	0	1	2	0.009
	s	22	17	12	7	36	
	n/s	0.41	0.06	0	0.14	0.06	
ANS	n	20	3	4	3	10	0.003
	s	11	11	7	7	25	
	n/s	1.82	0.27	0.57	0.43	0.4	
UFGT	n	21	11	0	10	21	0.48
	s	28	18	0	13	31	
	n/s	0.75	0.61	–	0.77	0.68	

[a] n, nonsynonymous; s, synonymous

[b] Probability that n/s is equal for sites with frequency <0.07 and sites with frequency >0.07

Table 5 Among-region proportions of variation (F_{ct}) for three anthocyanin-pathway genes in *I. trifida*, as calculated by AMOVA

	2 regions	3 regions
CHS	0.096	0.113
ANS	0.244	0.234
UFGT	0.060	0.145

Analysis of Geographic Substructure

Because unaccounted-for population structure can bias estimates of both within and between species diversity, we assessed the degree to which the three genes exhibited geographic structure. Whether the accessions were divided into two or three regions, there was significant geographic differentiation for all three genes (Table 5). However, the proportion of variation ascribable to between-region differences was low and similar for *CHS* and *UFGT*, and somewhat higher for *ANS*. Despite this differentiation, there were no fixed differences between any of the regions at any of the genes.

Discussion

Enzymes of the anthocyanin pigment pathway in angiosperms exhibit an interesting evolutionary pattern: downstream enzymes evolve more rapidly than upstream enzymes (Rausher et al. 1999; Lu and Rausher 2003). The pattern of divergence between *I. trifida* and *I. purpurea* reported here is consistent with this pattern: dN/dS for divergence is more than four times higher for ANS and UFGT than for CHS.

One explanation for this pattern is that constraint is higher for upstream enzymes than for downstream enzymes. One reason for expecting this pattern of relative constraint is that flux control in branching pathways is believed to normally reside primarily in the most upstream enzymes, with downstream enzymes exerting relatively little control over flux (LaPorte et al. 1984; Stehanopoulos and Vallino 1991; Eanes 1999). Consequently, deleterious mutations with similar effects on enzyme kinetic properties are expected to have greater deleterious fitness effects for the most upstream enzyme and, thus, are less likely to be effectively neutral and fixed by genetic drift. A second reason for expecting greater constraint in upstream enzymes is that in the anthocyanin pathway, downstream enzymes are responsible only for the production of anthocyanins, whereas upstream enzymes are involved in the production of several other types of flavonoids that serve a variety of physiological and ecological functions (Shirley 1996). An alternative explanation for more rapid evolution of downstream enzymes, however, is that they have evolved faster because they have undergone more frequent adaptive change, perhaps as a mechanism for fine-tuning the amounts and types of anthocyanins produced by a particular plant.

Our results implicate differential constraint and argue against differential adaptive change as explanations for more rapid evolution of downstream enzymes. In particular, two of our results provide direct evidence for differential constraint. First, within the species *I. trifida*, segregating nonsynonymous variation is much lower for the upstream enzyme CHS than for the downstream enzymes ANS and UFGT. This is precisely the pattern expected if upstream genes experience greater selective constraint, which means that purifying selection is stronger on upstream genes. Second, purifying selection, reflected in Tajima's D statistic, appears to be stronger on CHS than on the other two genes.

Alternative explanations for these results can be ruled out. For example, both the reduced variation and the value of Tajima's D for CHS might conceivably be due to a recent selective sweep in that gene. If that were true, however, we would expect also to see low variation and a low D at synonymous sites, which we do not. Similarly, a recent population expansion could have produced the generally negative D values for nonsynonymous variation. However, such a demographic event is expected to influence all variation similarly (Hahn et al. 2002) and should have produced equally negative D values for synonymous variation, which was not observed. Moreover, population expansion cannot account for the large differences among the genes in the amount of nonsynonymous variation.

Unaccounted-for geographic divergence could also conceivably explain some of the patterns we observed.

Divergence artificially increases *D*. Therefore, if downstream enzymes experience greater amounts of adaptive divergence, their *D* values could appear higher than for the upstream enzyme. Moreover, a greater amount of divergent fixation of alternate alleles between areas could make the apparent amount of nonsynonymous polymorphism in the species as a whole greater for downstream enzymes. We believe, however, that these effects are probably of minor importance in our study. We did detect significant divergence between regions for all three genes, but the degree of divergence was not large. Because the magnitude of divergence did not differ for CHS and UFGT, the effect of divergence on calculated nonsynonymous variation and on *D* would likely be very similar and, thus, could not account for the observed differences in these two parameters for the two enzymes.

While our results provide substantial evidence for differential constraint among upstream and downstream enzymes, they provide no evidence that differences in the rates of adaptive substitution contribute to the observed differences in evolutionary rates. When tested individually, all three enzymes exhibited essentially equal dN/dS ratios for segregating variation and for fixed differences between species, the expectation under neutrality. And when tested simultaneously in a single analysis, no evidence for positive selection, or for differences among the genes in the frequency of positive selection, was obtained. Finally, a comparison of divergence versus polymorphism across genes (HKA test) using only nonsynonymous substitutions revealed no heterogeneity among genes of the type expected if they had experienced different levels of adaptive substitution. These results do not mean that there have been no adaptive substitutions in any of these enzymes, only that they account for, at most, only a small fraction of the nonsynonymous changes. It is thus difficult to envision how differential adaptive substitution could account for the more rapid evolution of downstream enzymes.

One potential limitation to our analysis is that geographic differentiation can reduce the power of the M–K test to detect positive selection when data from across the range of a species are pooled. In particular, fixed differences between regions due to positive selection will be counted in the analysis as within-species polymorphisms, artificially raising the dN/dS ratio for within-species variation, to which the dN/dS ratio for divergence between species is compared. We believe, however, that our analysis does not suffer from this limitation. Although we did detect geographic substructuring in our data, there were no fixed differences between regions that would inflate the dN/dS ratio for polymorphisms. Moreover, because between-region divergence was similar for all three genes, any effect of substructuring on the M–K test is likely to have been similar for the three genes and, thus, should not contribute to our failure to find *differences* in rates of positive substitution among the genes.

Even though the overall rate of nonsynonymous substitution appears to be higher for ANS and UFGT than for CHS, it is nevertheless possible that the rate of adaptive substitution is higher for the upstream enzyme. Such a pattern is again expected if enzymes at pathway branch points have the greatest control over flux (Flowers et al. 2007). Our analysis of adaptive substitutions fails to support this pattern. However, we detected little evidence of positive selection in any of the three enzymes examined, which may simply mean that there has been little historical opportunity for more frequent adaptive substitutions in CHS.

Conclusion

Our results are consistent with the expectation that downstream enzymes in metabolic pathways are less constrained than upstream enzymes (Rausher et al. 1999; Vitkup et al. 2006; Flowers et al. 2007) and, thus, with the more general proposition that system properties influence the degree to which particular enzymes are evolutionarily constrained. It is of course possible that the enzymes differ in constraint because of differences in intrinsic properties, such as structure and composition, and that constraint is just by chance higher in the upstream genes. Differentiating between these hypotheses will constitute a significant challenge.

References

Bishop YMM, Feinberg SE, Holland PW (1975) Discrete multivariate analysis. MIT Press, Cambridge, MA

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

Cork JM, Purugganan MD (2004) The evolution of molecular genetic pathways and networks. Bioessays 5:479–484

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. Mol Phylogenet Evol 29:507–518

Eanes WF (1999) Analysis of selection on enzyme polymorphisms. Annu Rev Ecol Syst 30:301–326

Excoffier L, Laval G, Schneider S (2005) Arlequin ver. 3.0: an integrated software package for population genetics data analysis. Evol Bioinform Online 1:47–50

Fay JC, Wyckoff GJ, Wu C-I (2001) Positive and negative selection on the human genome. Genetics 158:1227–1234

Flowers JM, Sezgin E, Kumagi S, Duvernell DD, Matzkin LM, Schmidt PS, Eanes WF (2007) Adaptive evolution of metabolic pathways in *Drosophila*. Mol Biol Evol 24:1347–1354

Fraser HB, Hirsh AE, Steinmetz LM, Scharfe C, Feldman MW (2002) Evolutionary rate in the protein interaction network. Science 296:750–752

Fukada-Tanaka S, Hoshino A, Hisatomi Y, Habu Y, Hasebe M, Iida S (1997) Identification of new chalcone synthase genes for flower

pigmentation in the Japanese and common morning glories. Plant Cell Physiol 38:88–96
Hahn MW, Kern AD (2005) Comparative genomics of centrality and essentiality in three eukaryotic protein-interaction networks. Mol Biol Evol 22:803–806
Hahn MW, Rausher MD, Cunningham CW (2002) Distinguishing between selection and population expansion in an experimental lineage of Bacteriophage T7. Genetics 161:11–20
Hahn MW, Conant GC, Wagner A (2004) Molecular evolution in large genetic networks: does connectivity equal constraint? J Mol Evol 58:302
Hartl DL, Dykhuizen DE, Dean AM (1985) Limits of adaptation: the evolution of selective neutrality. Genetics 111:655–674
Hedrick PW (2000) Genetics of populations, 2nd edn. Jones and Bartlett, Sudbury, MA
Hirsh AE, Fraser HB (2001) Protein dispensability and rate of evolution. Nature 411:1046–1049
Hudson RR, Kreitman M, Aguadé M (1987) A test of neutral molecular evolution based on nucleotide data. Genetics 116:153–159
Kimura M (1977) Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. Nature 267: 275–276
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
LaPorte DC, Walsh K, Koshland DE (1984) The branch point effect: ultrasensitivity and subsensitivity to metabolic control. J Biol Chem 259:4068–4075
Lemos B, Bettencourt BR, Meiklejohn CD, Hartl DL (2005) Evolution of proteins and gene expression levels are coupled in Drosophila and are independently associated with mRNA abundance, protein length, and number of protein-protein interactions. Mol Biol Evol 22:1345–1354
Li WH (1997) Molecular evolution. Sinauer Associates, Sunderland, MA, p 185
Li WH, Wu CI, Luo CC (1985) A new method of estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. Mol Biol Evol 2:150–174
Lin YS, Hsu WL, Hwang JK, Li WH (2007) Proportion of solvent-exposed amino acids in a protein and rate of protein evolution. Mol Biol Evol 24:1005–1011
Lu Y (2002) Molecular evolution at the self-incomatpbility locus of *Physalis longifolia* (Solanaceae). J Mol Evol 54:784–793
Lu Y, Rausher MD (2003) Evolutionary rate variation in anthocyanin pathway genes. Mol Biol Evol 20:1844–1853
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
McDonald JH, Kreitman M (1991) Adaptive protein evolution at the *ADH* locus in *Drosophila*. Nature 351:652–654
Pal C, Papp B, Hurst LD (2001) Highly expressed genes in yeast evolve slowly. Genetics 158:927–931
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
Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R (1999) DnaSP, DNA polymorphism analysis by the coalescent and other methods. Bioinformatics 19:2496–2497
Shapiro JA, Huang W, Zhang C, Hubisz MJ, Lu J, Turissini DA, Fang S, Wang H-Y, Hudson RR, Nielsen R, Chen Z, Wu C-I (2007) Adaptive genic evolution in the Drosophila genomes. Proc Natl Acad Sci USA 104:2271–2276
Shirley BW (1996) Flavonoid biosynthesis: 'new' functions for an 'old' pathway. Trends Plant Sci 1:377–382
Sokal RR, Rohlf JJ (1969) Biometry. The principles and practice of statistics in biological research. W. H. Freeman, San Francisco
Stephanopoulos G, Vallino JJ (1991) Network rigidity and metabolic engineering in metabolite overproduction. Science 252: 1675–1681
Swanson WJ, Yang Z, Wolfner MF, Aquadro CF (2001) Positive Darwinian selection derives the evolution of several female reproductive proteins in mammals. Proc Natl Acad Sci USA 98:2509–2514
Tajima F (1989) Statistical method for testing the neutral mutation hypothesis. Genetics 123:585–595
Van der Meer IM, Stuitje AR, Mol HNM (1993) Regulation of general phenylpropanoid and flavonoid gene expression. In: Verma DPS (ed) Control of plant gene expression. CRC Press, Boca Raton, FL
Vitkup D, Karchenko P, Wagner A (2006) Influence of metabolic network structure and function on enzyme evolution. Genome Biol 7:R39
Watt WB, Dean AM (2000) Molecular functional studies of adaptive genetic variation in prokaryotes and eukaryotes. Annu Rev Genet 34:593–622
Whitefield L, Lovell-Badge R, Goodfellow P (1993) Rapid sequence evolution of the sex-determining gene SRY. Nature 364:713–715
Wyckoff GJ, Wang W, Wu CI (2000) Rapid evolution of male reproductive genes in the descent of man. Nature 403:304–309
Yang Z, Nielsen R, Goldman N, Pedersen AK (2000) Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics 155:431–449