Patterns of Variation in the Evolution of Carotenoid Biosynthetic Pathway Enzymes of Higher Plants

KEVIN LIVINGSTONE AND STEPHANIE ANDERSON

From the Department of Biology, Trinity University, One Trinity Place, San Antonio, TX 78212.

Address correspondence to Kevin Livingstone at the address above, or e-mail: klivings@trinity.edu.

Abstract

Cellular metabolic networks depend on the products of many loci for proper functioning. These interrelationships between loci at the phenotypic level raise the question of whether they evolve independently. Previous research has demonstrated that in the anthocyanin pathway, which produces important secondary metabolites in plants, the genes encoding downstream enzymes show an increased rate of change at nonsynonymous sites when compared with upstream loci due to relaxed constraint. To test whether this pattern exists more broadly, we compared a set of 4 genes encoding enzymes of the carotenoid biosynthetic pathway, which produces a set of distinct colored secondary metabolites in plants. Comparisons between copies of phytoene desaturase, \( \varepsilon \)-carotene desaturase, lycopene \( \beta \)-cyclase, and zeaxanthin epoxidase from 6 taxa indicate that the 3 upstream enzymes (phytoene desaturase, \( \varepsilon \)-carotene desaturase, and lycopene \( \beta \)-cyclase) have similar proportions of codons under selective constraint, whereas the most downstream enzyme (zeaxanthin epoxidase) has more codons evolving under relaxed constraint. Overall, nonsynonymous substitution rates appear to be highest for zeaxanthin epoxidase, whereas synonymous substitution rates were highest for the intermediate enzyme lycopene \( \beta \)-cyclase. Analysis of codon bias shows that only lycopene \( \beta \)-cyclase may be under slight selection pressure for codon usage. Taken together, these results show that the enzymes of the carotenoid biosynthetic pathway are under strong selective constraint but that the most downstream enzyme is under the least constraint.

Key words: carotenoid, evolution, metabolic pathway

The spectacular recent improvements in high-throughput DNA sequencing technology are now providing molecular evolutionists with increasingly large data sets. Importantly, these data sets are both deeper within species and more widespread across taxa, allowing for broad comparisons encompassing whole genomes (Clark et al. 2007). These data sets, together with the concomitant advances in computational methods, now offer an unprecedented scale and sophistication for testing global hypotheses of major causative factors explaining the different rates of protein evolution seen across genes and taxa.

Our knowledge of the chromosomal location, structure, and patterns of expression for many genes has allowed these characteristics to be tested as explanatory variables in studies of protein evolution. To date, the single largest factor influencing a protein’s evolutionary rate appears to be a composite measure of expression including codon bias, level of transcription, and amount of protein in the cell (Drummond et al. 2006). Beyond expression, other factors shown to have significant but smaller effects include intron number, genomic position, recombination rate, and essentiality (Pál et al. 2006; Larracuente et al. 2008).

These tests have explicitly considered what we may call structural factors leading to expression, but a gene’s biochemical function in the greater cellular context after expression is of equal importance (Cork and Purugganan 2004). Gene products act and interact within a cell in multiple ways. Many proteins function together as enzymes in biochemical pathways, causing interrelationships between loci in the formation of phenotypes. These interactions can also be physical in that the gene products from different loci may participate in macromolecular complexes necessary to carry out a reaction. These groups of loci, however, may be participating in disparate areas of the metabolome, raising the question of whether any of these factors systematically affect a gene’s evolutionary rate.

At a gross level, differences can be seen between the mean levels and variation of \( \omega \), the ratio of nonsynonymous (\( d_n \)) to synonymous (\( d_s \)) mutation rates, across different functional categories (Clark et al. 2007). When taking into
account the interactions between gene products, the number of interactions a protein has and its position in the interaction network appear to lead to a decreased \( d_N \) (Fraser et al. 2002; Hahn and Kern 2005; but see Jordan et al. 2003). Although these studies give broad overviews, they have not looked specifically at how the structure of a metabolic pathway may have influenced protein evolution rates.

The most comprehensive examples to date of molecular evolution studies that explicitly take into account pathway structure have used *Drosophila* and the anthocyanin biosynthetic pathway in angiosperms. In *Drosophila*, Flowers et al. (2007) have found that genes at branch points of glucose catabolism show evidence of divergence between species driven by positive selection, which they interpret as selection acting on the parts of the pathway that control metabolic flux rates. More recently, an overall analysis by Greenberg et al. (2008) has confirmed that pathway architecture does appear to affect protein evolutionary rates, although the main finding of this study was that core metabolic enzymes and those that share metabolites evolve very slowly.

An alternative system for these studies is the plant anthocyanin biosynthetic pathway developed by Rausher and colleagues. In the first of 3 studies, broad comparisons between *Zea*, *Antirrhinum*, and *Ipomoea* showed that upstream enzymes in this pathway have reduced rates of nonsynonymous substitution when compared with the downstream enzymes (Rausher et al. 1999). A second study examining multiple species within *Ipomoea* found the same pattern of increased \( d_N \) for downstream enzymes and suggested relaxed constraint on downstream enzymes to be the cause, as changes to these enzymes would be less pleiotropic and/or have lesser effects on flux through the pathway (Lu and Rausher 2003). This conclusion was confirmed in the most recent work by Rausher et al. (2008), which showed that within *Ipomoea trifida* the downstream enzymes in this pathway are under relaxed constraint. Taken together, the *Drosophila* and anthocyanin studies suggest that pathway structure is another global factor influencing protein evolution rates.

Our goal in this study was to test for the pattern seen in the anthocyanin pathway in another plant system, the carotenoid biosynthetic pathway. The carotenoid biosynthetic pathway provides many parallels to the anthocyanin pathway: the products of both pathways are used for both conserved, housekeeping functions in green tissues (Cuttriss and Pogson 2004; Koes et al. 2005), as well as derived functions important for coloring reproductive structures (Bradshaw and Schemske 2003; Clegg and Durbin 2003; Zufall and Rausher 2003; Howitt and Pogson 2006). In addition, the nutritional importance of carotenoids has led to extensive characterization of the biosynthetic genes from several species, allowing broad comparisons to be made. Here we use a sequence data set of 4 genes from 6 species to ask whether downstream genes in the carotenoid pathway show increased rates of evolution and examine factors that may cause these differences.

![Figure 1](https://example.com/carotenoid_pathway.png)  

*Figure 1.* Partial schematic of the carotenoid biosynthetic pathway in higher plants. This pathway shows the formation of the first C40 carotenoid, phytoene, and subsequent conversions to neoxanthin (not shown are some intermediates and the branch at lycopene which makes lutein). The gene symbols used for the enzymes are as follows: phytoene synthase, *Psy*; phytoene desaturase, *Pds*; \( \zeta \)-carotene desaturase, *Zds*; carotenoid isomerase, *Crtiso*; lycopene \( \beta \)-cyclase, *Lcy-B*; \( \beta \)-carotene hydroxylase, *\( \beta \)-OH*; zeaxanthin epoxidase, *Zep*; neoxanthin synthase, *Nxs*. (modified from Hirschberg 2001 and Howitt and Pogson 2006).

### Materials and Methods

#### Sequences and Phylogenetic Tree Construction

In higher plants, there are 10 enzymes in the central carotenoid biosynthetic pathway that convert geranylgeranyl pyrophosphate to the end products neoxanthin or lutein. In this study, we focused on the neoxanthin branch (Figure 1), selecting the enzymes phytoene desaturase (*Pds*; note that enzyme names and gene symbols vary among organisms), \( \zeta \)-carotene desaturase (*Zds*), lycopene \( \beta \)-cyclase (*Lcy-B*), and zeaxanthin epoxidase (*Zep*) for analysis. These genes were chosen because there is evidence for ancestral and derived duplication of the other loci in the pathway (Cunningham 2002; DellaPenna and Pogson 2006; Galpaz et al. 2006; Just et al. 2007). Although *Lcy-B* also appears to have undergone an ancestral duplication, inclusion of this gene was justified based on prior work that allowed for recognition of each ortholog (Livingstone K, unpublished data).
Solanum lycopersicum (tomato) was chosen to be the representative species based on extensive previous research characterizing each enzyme (Bramley 2002; Galpaz et al. 2006). We were able to obtain representative complementary DNA sequences for each of the tomato enzymes listed above from GenBank. We then identified similar genes through BLAST searches (TBLASTX) against all plant gene indices from the Institute for Genomic Research (Lee et al. 2004), PlantGDB (Dong et al. 2005), and GenBank databases. Homology of the tomato and database sequences was determined through percent identities and the length of the match between sequences. In addition, we used sequences that were published in studies of the carotenoid enzymes from Chrysanthemum (Kishimoto and Ohmiya 2006) and Daucus carota (Just et al. 2007).

The selected sequences were initially trimmed down to the coding sequences and then translated using BIOEDIT 7.0.9.0 (Hall 1999). We next created alignments of the edited peptide sequences (Hall 2004) using CLUSTALX 2.0.5 (Larkin et al. 2007). We obtained predicted chloroplast leader sequences from the CHLOROP 1.1 server ( Emanuelsson et al. 1999), and these results were combined with subjective analysis to demarcate the portion of the alignment that appeared to represent the leader sequences. The DNA sequences corresponding to the predicted leader sequences were then removed, and manual adjustment of aligned DNA sequences was performed as necessary. We also created a concatenated sequence for each species by combining DNA sequence alignments. We constructed a maximum likelihood phylogeny of the 6 species analyzed using the concatenated DNA sequences and the program DNAML, a part of the PHYLIP package (Felsenstein 2008). Sequences from Oryza sativa were used as an out-group, and confidence in the phylogeny was assessed by analyzing 1000 bootstrap replicates.

Comparing Codon Evolution Models to Determine Selective Constraint

We fit 3 models of molecular evolution to the data for each gene using the CODEML program of the PAML suite according to published parameters (Yang 2007). The first model, M0, is a null model that assumes all positions across a given set of sequences have the same levels of \( d_{NS} \) and \( d_{S} \) and these values and their ratio (\( \omega = d_{NS}/d_{S} \)) are estimated. The parameter \( \omega \) is a good indicator of protein level selection, with 0 < \( \omega \) < 1 indicating stabilizing/purifying selection, \( \omega = 1 \) indicating neutral evolution, and \( \omega > 1 \) indicating directional/positive selection. The M1a model uses 2 site classes: Class 0 is a proportion of codons with 0 < \( \omega_0 \) < 1 and the remainder of codons form the other class with fixed \( \omega_1 = 1 \). Model M1a estimates both the proportion of Class 0 sites and their \( \omega_0 \). The fit of model M0 versus model M1a can be evaluated by the likelihood ratio test, comparing twice the difference in log likelihoods with \( \chi^2 \) distribution whose degrees of freedom equals the difference in the number of parameters from each model (Yang 2007). A significantly better fit for model M1a indicates that there are codons evolving neutrally, as well as codons under selective constraint. Model M2a further divides the codons into 3 site classes with 0 < \( \omega_0 \) < 1, \( \omega_1 = 1 \), and \( \omega_2 \geq 1 \) and estimates both the proportions and variable \( \omega \) values. Again, the relative fits of model M1a and model M2a can be compared by a likelihood ratio test and accepting model M2a over M1a provides evidence that there is a subset of codons that have experienced positive selection.

To test whether differences in the estimated parameter values between genes were significant, the methods of Lu and Rausher (2003) and Streisfeld and Rausher (2008) were implemented. Briefly, this method uses model M0, but instead of using the program’s estimates of \( \omega \), fixed values of \( \omega \) are supplied, and the resulting likelihoods are compared to the model with the original estimate of \( \omega \). Values of \( \omega \) between the values estimated for each gene are used to determine whether an intermediate value of \( \omega \) exists that is not significantly better than the original estimates for each gene. If such a value exists, then the original estimated values of \( \omega \) for each gene are said to be not significantly different. Alternatively, if no such intermediate value exists, then the values of \( \omega \) for the 2 genes are said to be significantly different. Finally, pairwise \( d_{NS} \) and \( d_{S} \) values were also calculated for each gene using the CODEML program from the PAML 4.0 suite (Yang 2007).

Codon Bias Estimation

The effective number of codons (ENC) was calculated using the methods of Wright (1990) as implemented in DNASP 4.50.1 (Rozas et al. 2003). Because codon bias is often associated with GC content at the third positions across codons (GC3), a graphical comparison of ENC versus GC3 was used to control for this possible bias. The observed values of ENC versus GC3, along with a curve describing the expected values, were graphed using the Ne-plot technique (Wright 1990).

Results

Sequences and Alignments

We were able to find a large number of sequences for each enzyme from a diverse set of taxa. Many of these sequences were discarded, however, because they did not contain the majority of the coding region. In addition, this analysis required that each alignment includes the same set of species, so we pruned the set of putative orthologs to a final set including S. lycopersicum, Capsicum annuum, Gentiana lutea, Chrysanthemum × morifolium, Citrus spp., and D. carota (Table 1). We were also able to obtain a complete set of sequences from O. sativa, and these were used as out-group sequences during phylogenetic tree construction. Most of these sequences were generated as part of published studies; however, several were direct depositions into the database without peer review. Although these sequences were not verified, it was assumed that they contained no more error than the sequences associated with publications and that this
error rate was low relative to interspecific differences. In no case were putative paralogs of *Psy*, *Pds*, or *Zep* discovered in the databases. It is still possible, however, that species-specific duplications of these loci exist and that they have not been discovered despite directed searches in the case of extensively studied taxa such as *S. lycopersicum*.

Alignments of the translated coding sequence for each gene showed an amino terminal region with low conservation, followed by a region of high similarity. These genes are encoded by the nucleus, but the enzymes function in the chloroplast, so this pattern could be explained by the presence of chloroplast targeting signal sequences. The CHLOROP server was used to provide evidence for such leaders and approximate the length of the leader sequence for each gene. The strength of predictions and lengths of the leader sequences were extremely variable: CHLOROP predicted each copy of *Pds* had a leader, but the lengths varied from 7 to 70 amino acids. In other cases, no leader was predicted for 1 to several of the species in an alignment. Thus, CHLOROP predictions were used as a guide, but subjectivity was also employed to determine where sequence alignments were trimmed. In addition, because the alignments must contain the same number of base pairs for each sequence, all sequences were trimmed at the same point in the alignment on the 3′ ends. A small amount of internal sequence, corresponding to 9–10 codons for *Lcy-B* and 6–11 codons for *Zep*, was removed because of indel polymorphisms. The amount of the *S. lycopersicum* coding sequence used was 73% for *Pds*, 85% for *Zds*, 89% for *Lcy-B*, and 84% for *Zep*. All alignments are available as Supplementary Materials.

The phylogeny constructed from the concatenated sequence alignment (Figure 2) shows evolutionary relationships between the selected species that are in accordance with known phylogenetic data (Soltis P and Soltis D 2004). Bootstrap analysis revealed strong support for this tree, suggesting that true orthologs had been obtained from each species.

### Sequence Evolution

In order to explore patterns of sequence variation, we fit different models of molecular evolution to each set of sequences. Comparisons of these models showed that M1a provided a better fit for each gene when compared with M0 (0.01) but that model M2a was not a better fit than M1a for any of the genes (Table 2). Rejecting model M2a in favor of model M1a indicates that there is no support for positive selection on any of the codons across all 4 genes. Instead, model M1a includes only 2 classes of sites, a class under strong purifying selection (0 < α0 < 1) and another class evolving neutrally (α0 = 1), and both the proportion of codons under purifying selection (p0) and the estimates of α0 from model M1a are plotted for each gene in Figure 3. This analysis shows that the estimated proportions of codons under purifying selection are high overall, but this proportion appears lower for *Zep*. The estimates of the strength of purifying selection for these codons, however, do not seem to vary systematically with position in the pathway.

Although model M1a is a better fit to the data than model M0, model M0 has still been shown to be a way to

---

### Table 1. Database accession number (GI) and subsequences (nt) of genes selected for analysis

<table>
<thead>
<tr>
<th>Gene</th>
<th>Pds</th>
<th>Zds</th>
<th>Lcy-B</th>
<th>Zep</th>
</tr>
</thead>
<tbody>
<tr>
<td>Species</td>
<td>GI nt</td>
<td>GI nt</td>
<td>GI nt</td>
<td>GI nt</td>
</tr>
<tr>
<td><em>Citrus spp.</em></td>
<td>18073983 385–1668</td>
<td>18073985 266–1768</td>
<td>14550424 577–1917</td>
<td>17402596 226–1941</td>
</tr>
<tr>
<td><em>Capsicum annuum</em></td>
<td>17950 325–1608</td>
<td>6006400 208–1710</td>
<td>999440 419–1759</td>
<td>1673405 238–1947</td>
</tr>
<tr>
<td><em>Gentiana lutea</em></td>
<td>18146804 448–1731</td>
<td>6681691 421–1923</td>
<td>117935906 361–1701</td>
<td>1673405 238–1947</td>
</tr>
</tbody>
</table>

a Some internal nucleotides removed due to indels.
Table 2. Parameter estimates and tests of selection using PAML.

<table>
<thead>
<tr>
<th>Gene</th>
<th>M0</th>
<th>M1a</th>
<th>M2a</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>lnL ( ^a )</td>
<td>( \omega_0 )</td>
<td>( \omega_0 )</td>
</tr>
<tr>
<td>Pds</td>
<td>-4387.5</td>
<td>0.051</td>
<td>-4382.6</td>
</tr>
<tr>
<td>Zds</td>
<td>-5293.1</td>
<td>0.052</td>
<td>-5227.0</td>
</tr>
<tr>
<td>Lcy-B</td>
<td>-4866.8</td>
<td>0.041</td>
<td>-4826.5</td>
</tr>
<tr>
<td>Zep</td>
<td>-7143.8</td>
<td>0.107</td>
<td>-6990.7</td>
</tr>
</tbody>
</table>

\( ^a \) Log likelihood of model.
\( ^b \) Parameter estimate assuming a single \( d_{ls}/d_{kt} \) ratio per gene.
\( ^c \) Estimated \( d_{ls}/d_{kt} \) for proportion of codons (\( \omega_0 \)) under purifying selection; other codons assumed to be evolving neutrally.
\( ^d \) Estimated proportion of codons under purifying selection.
\( ^e \) Test of M1a versus M0.
\( ^f \) Estimated \( d_{ls}/d_{kt} \) for proportion of codons (\( \omega_0 \)) under purifying selection.
\( ^g \) Estimated proportion of codons under purifying selection.
\( ^h \) Estimated proportion of codons under positive selection.
\( ^i \) Test of M2a versus M1a.

To further investigate this pattern, we also calculated all 15 pairwise estimates of \( d_{N} \). These data show that the highest level of nonsynonymous substitutions for each comparison is seen with Zep (Figure 4a). In addition, for all but the closest comparison, S. lycopersicum and C. annuum (Figure 4a, bottom line), the values of \( d_{N} \) are more tightly grouped for Pds and Zds and more variable for the downstream genes Lcy-B and Zep. Although the set of all comparisons is not independent, the overall trend across all comparisons is that the highest substitution rate observed is always Zep; therefore, each of the sets of independent contrasts would be expected to provide the same results.

To determine whether this pattern for \( d_{N} \) could be explained by increased mutation rates for the downstream genes, we also examined the values for pairwise \( d_t \) comparisons (Figure 4b). Interestingly, these comparisons reveal a distinct pattern, with generally higher and more variable substitution rates at Lcy-B than Zep. If overall mutation rates were higher for downstream genes, then these 2 graphs would be similar. The differences between the observed patterns of \( d_t \) and \( d_{N} \) suggest that there are not higher overall mutation rates at Zep; therefore, this cannot be an explanatory factor for the higher rate of \( d_{N} \) for this gene. Taken together, these data suggest that the majority of codons in these 4

Test whether there are significant differences in evolutionary rates between genes in similar analyses (Lu and Rausher 2003; Streisfeld and Rausher 2008). In this study, we tested whether pairwise estimates of an overall \( \omega_0 \) were different by fixing \( \omega_0 \) values for each gene in model M0 and then testing whether these values still fit the data well (\( P > 0.05 \)). We found that a value of \( \omega_0 = 0.047 \) could be substituted for the estimated value of \( \omega_0 \) from model M0 for Pds, Zds, and Lcy-B without a significant difference in the model’s fit, indicating that these 3 genes were not evolving at significantly different rates. When these 3 genes were compared with Zep, however, it was found that a value of \( \omega_0 = 0.07 \) could not be substituted for the estimated value for any of the 4 genes (\( P < 0.05 \)), indicating that Pds, Zds, and Lcy-B showed an evolutionary pattern significantly different from Zep.

The difference between Zep and the set of the other 3 genes seen in the M0 analyses with fixed \( \omega_0 \) could be due to 1 of 2 differences. As shown in the model M1a analyses, these genes are fit best by a model with 2 classes of codons (under purifying and neutral evolution). Thus, overall differences seen when using a single \( \omega_0 \) for a gene must be due to differences in either the relative proportions of these 2 classes or varying levels of purifying selection. As shown in Figure 3, the most noticeable difference between Zep and the other 3 genes is in the proportion of codons under purifying selection, making this the most likely explanation for the overall differences seen in the M0 analyses.

Figure 3. Percentage of codons with \( 0 < \omega < 1 \) (open circles) and the corresponding \( \omega \) (closed squares) as a function of enzyme position in the carotenoid biosynthetic pathway. Values shown are estimates from model M1a calculated by CODEML (Yang 2007).
genes are under similar levels of strong purifying selection but that the fraction of codons evolving neutrally is higher in the most downstream enzyme (Zep).

**Codon Bias**

ENC is a measure of the extent of codon usage, ranging from 20, where codon bias is at a maximum and only 1 codon is used for each amino acid, to 61, where there is no codon bias and all codons are used. However, codon usage can also be affected by factors other than usage bias, such as mutation bias. To distinguish between these alternative explanations, we graphed ENC as a function of GC3, along with the expectation of ENC given GC3 (Figure 5). Overall, the enzymes are tightly clustered in a narrow range of GC3 content (Figure 5 inset). On closer inspection, the only visible trend is that the points for Lcy-B are more tightly clustered than any other enzyme, and in general, they lie furthest from the expected values.

**Discussion**

Our work parallels analyses by Rausher and colleagues of the enzymes that produce anthocyanins, an important class of secondary metabolites in plants (Rausher et al. 1999, 2008; Lu and Rausher 2003). These previous studies revealed that genes encoding downstream enzymes have an accelerated rate of nonsynonymous substitution when compared with upstream loci, in both broad (monocot–eudicot) and more restricted (within *Ipomoea*) comparisons, because of relaxed constraint on downstream enzymes. Although this pattern was not explicitly addressed in the study of the *Drosophila* glycolytic pathway (Flowers et al. 2007), data included in that paper could also be interpreted as showing an increase in $d_N$ for enzymes in the linear portion of the pathway from glucose to pyruvate. Together, these studies suggest a general pattern of protein evolution influenced by position in a metabolic pathway, along with the causative factor of relaxed selective constraint on the downstream loci. We approached these questions using another pathway that produces colored secondary metabolites in plants and a taxonomic range intermediate to the 3 anthocyanin studies.

Our analysis confirms the pattern that the most downstream locus tends to evolve more quickly. In general,
all 4 enzymes are under similarly strong levels of purifying selection over most of their sequence, a finding that is commonplace in genomic comparisons (Clark et al. 2007; Foxe et al. 2008). Each gene does, however, have codons that appear to be evolving neutrally. This fraction of codons appears to be higher for Zep than the other 3 loci, indicating that relaxed selective constraint over a greater portion of Zep is the most likely explanation for its increased rate of protein evolution, as opposed to positive selection for novel enzymatic activities (Copley 2003) and/or physical associations (Winkel 2004).

When the substitutions are divided into nonsynonymous and synonymous substitutions, it is apparent that there are differences between these 2 classes. For nonsynonymous substitutions, there is a trend toward increases in the level and variation of \( d_N \) for at least the 2 most terminal enzymes (\( Lcy-B \) and \( Zep \)), although there appears to be a higher level of \( d_s \) seen for the penultimate enzyme analyzed (\( Lcy-B \)). The differences between \( d_N \) and \( d_s \) argue against a simple explanation of higher mutation rates for downstream loci.

The results of Drummond et al. (2006) suggest that the observed variation in evolutionary rates between these loci may be due to factors influencing expression. We tested these loci for evidence of selection on codon usage, but our analysis indicates that of the 4 genes, only \( Lcy-B \) shows evidence for at most mild selective constraint for codon usage. \( Lcy-B \) is also the gene with the highest rate of synonymous substitutions, which is counterintuitive: it is expected that greater selection for codon bias would cause a decrease in synonymous substitution rates, but this same apparent anomaly was seen in the anthocyanin study (Lu and Rausher 2003). It is intriguing to note that in both cases, the genes were the only duplicated genes included in the analysis. Overall, however, the general lack of variation amongst species and between genes implies that codon bias has remained low and relatively stable for genes of the carotenoid biosynthetic pathway, both within and across species.

The expression profiles of these genes do vary, however, because of the different roles carotenoids play in plants. On the one hand, carotenoids are important ancillary members of the light harvesting complex, and studies in developing tobacco chloroplasts, where carotenoids are first needed, have demonstrated similar levels of coordinated expression (Woitsech and Römer 2003). On the other hand, the regulation of carotenoid biosynthetic genes is also coordinated in organs where carotenoid hyperaccumulation occurs, such as petals and fruit. Although coordinately regulated in these tissues, the specific constellation of upregulated genes varies, leading to differences in carotenoid accumulation profiles (Bramley 2002; Kato et al. 2004; Kishimoto and Ohmiya 2006). Zeaxanthin is an important component of the photosynthetic apparatus of green tissues but is not present in all species’ flowers and fruits (Goodwin 1980), leaving open the possibility that increases in \( d_N \) for Zep could be due to reduced expression in the organs of some species.

### Supplementary Material

Supplementary material can be found at http://www.jhered.oxfordjournals.org/.

### Funding

Trinity University Summer Undergraduate Research Fellowship to S.A.

### Acknowledgments

We thank Brian Malpede for critical readings and 2 anonymous reviewers for suggestions that strengthened the manuscript.

### References


Received June 20, 2008; Revised April 8, 2009; Accepted April 16, 2009

Corresponding Editor: John Stommel