Plant Molecular Biology **42:** 79–92, 2000.

Molecular evolution of the chalcone synthase multigene family in the morning glory genome

Mary L. Durbin*, Bonnie McCaig and Michael T. Clegg
*Botany and Plant Sciences Department, University of California, Riverside, CA 92521, USA (*author for correspondence)*

Key words: anthocyanin, chalcone synthase, gene duplication, *Ipomoea*, rate of nucleotide substitution, stilbene synthase

Abstract

Plant genomes appear to exploit the process of gene duplication as a primary means of acquiring biochemical and developmental flexibility. Thus, for example, most of the enzymatic components of plant secondary metabolism are encoded by small families of genes that originated through duplication over evolutionary time. The dynamics of gene family evolution are well illustrated by the genes that encode chalcone synthase (CHS), the first committed step in flavonoid biosynthesis. We review pertinent facts about CHS evolution in flowering plants with special reference to the morning glory genus, *Ipomoea*. Our review shows that new CHS genes are recruited recurrently in flowering plant evolution. Rates of nucleotide substitution are frequently accelerated in new duplicate genes, and there is clear evidence for repeated shifts in enzymatic function among duplicate copies of CHS genes. In addition, we present new data on expression patterns of CHS genes as a function of tissue and developmental stage in the common morning glory (*I. purpurea*). These data show extensive differentiation in gene expression among duplicate copies of CHS genes. We also show that a single mutation which blocks anthocyanin biosynthesis in the floral limb is correlated with a loss of expression of one of the six duplicate CHS genes present in the morning glory genome. This suggests that different duplicate copies of CHS have acquired specialized functional roles over the course of evolution. We conclude that recurrent gene duplication and subsequent differentiation is a major adaptive strategy in plant genome evolution.

Introduction

Genetic redundancy is a prime feature of plant genomes. Virtually all plant genes so far examined are represented within most plant genomes as small gene families that originate through duplication. To persist over long periods of evolutionary time, new gene duplications must be positively selected, otherwise the relentless accumulation of mutations would ultimately convert the duplicate copy to a pseudogene. There are three ways that newly duplicated genes may be of adaptive value to the plant: (1) genetic redundancy may be favored because it increases the production of a limiting product; (2) the protein may take on a new or modified enzymatic function; and (3) there may be a specialization in developmental expression that also permits subtle shifts in catalytic properties so as to optimize pathway throughput in different tissues and at different points in the life cycle of the plant.

Our objective in this article is to explore these latter two possibilities with respect to the genes that encode the enzyme chalcone synthase (CHS), the first committed step of flavonoid biosynthesis. We will consider previously published evidence that there have been repeated shifts in CHS enzymatic function over evolutionary time and we will present new data on CHS gene expression patterns in the common morning glory (*Ipomoea purpurea*) that suggest divergence in developmental expression among duplicated genes. Finally, we will present data which suggest that a single CHS gene mutation is responsible for an albino phenotype in the common morning glory, despite con-

siderable redundancy in CHS gene copies within the morning glory genome establishing a one-to-one correspondence between a single CHS gene and a major phenotype.

CHS catalyzes the initial condensation reaction that results in the 15-carbon three-ring structure that is the backbone of flavonoid biosynthesis [31]. Specifically, CHS catalyzes the condensation of three molecules of malonyl-CoA with one molecule of 4-coumaroyl-CoA to form the naringenin chalcone molecule. Chalcone synthase functions as a homodimer with a total size of ca. 82 kDa [31]. Whereas 4-coumaroyl-CoA is the preferred substrate for CHS in most plants, there are reports that CHS in some plants can utilize caffeoyl-CoA or feruloyl-CoA as substrates *in vitro* [19]. The pH optima of the enzyme reactions are 7.5–8.5 for 4-coumaroyl-CoA and 6.5–6.8 for caffeoyl-CoA. There are no co-factor requirements for the reaction. A cysteine at amino acid 169 is thought to be part of the site that binds 4-coumaroyl-CoA and is required for enzyme activity [31]. The product of the condensing reaction of CHS is naringenin-chalcone which spontaneously isomerizes to form naringenin. Isomerization is also accomplished *in vivo* by the enzyme chalcone isomerase (CHI), the second enzyme in the flavonoid pathway. The product of CHS may be further modified in a number of subsequent biochemical steps to yield many different end products (Figure 1). One of the end products is anthocyanin, which is a pigment responsible for flower color in plants. CHS was thought to be a cytosolic enzyme [2] although there is now evidence that it is associated with the cytoplasmic side of the endoplasmic reticulum [16]. This association with the endoplasmic reticulum has implications for the targeting of the enzyme into metabolic complexes, which we discuss later.

The biochemical pathway that commences with CHS is important in plant disease defense, pigment biosynthesis, UV protection, and pollen viability. Each of these processes is clearly important to plant adaptation and we may assume that selection has played, and continues to play, a determining role in the evolution of the genes that comprise the pathway. Thus, for example, the anthocyanin pigments that determine floral color are clearly important to plant reproductive success by virtue of their role in the attraction of insect pollinators. Similarly, the disease defense pathway is of obvious adaptive importance.

The flavonoid biosynthetic pathway is believed to have evolved directly from other primary pathways

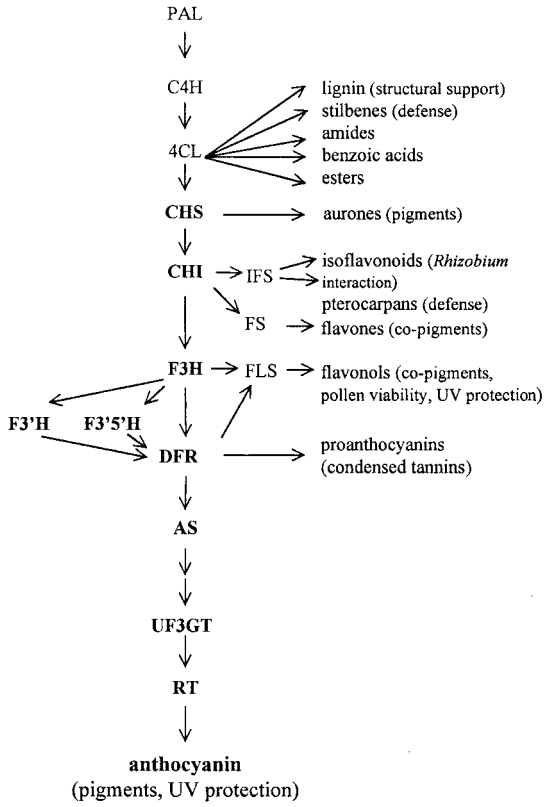

Figure 1. Anthocyanin biosynthetic pathway. Enzymes involved in anthocyanin biosynthesis and side branches leading to related flavonoid pathways are shown. Symbols in bold type indicate anthocyanin enzymes. PAL, phenylalanine ammonia-lyase; C4H, cinnamate 4-hydroxylase; 4CL, coumarate:CoA ligase; CHS, chalcone synthase; CHI, chalcone isomerase; F3H, flavanone 3-hydroxylase; DFR, dihydroflavonol 4-reductase; AS, anthocyaninidin synthase (same as leucoanthocyanidin dioxygenase); UF3GT, UDP-glucose flavonoid 3-oxy-glucosyltransferase.

[12]. Thus, chalcone synthase is thought to have arisen from an enzyme involved in fatty acid synthesis [41]. The other enzymatic steps in the pathway are thought to have then been recruited sequentially over evolutionary time [41]. The full pathway appears to have been assembled by the early phases of seed plant evolution based on the presence of anthocyanins in gymnosperms but not in the more primitive ferns [26].

The structure of the CHS gene is conserved in flowering plants. All CHS genes studied so far contain one intron at a conserved site, with the exception of *Antirrhinum majus*, which has a second intron in exon 2 [38]. CHS mutations resulting in a phenotypic change have been described in many species. Such mutations have been reported in *Antirrhinum majus*, *Hordeum vulgare*, *Lycopersicon esculentum*, *Pisum*

sativum, and *Zea mays.* Generally loss of CHS function results in a lack of anthocyanin and an albino flower color phenotype. This is not unexpected considering that CHS is a key enzyme in the anthocyanin pathway. However, it is now known that CHS is encoded by a small multigene family in many species [31] including those species containing mutations that result in loss of CHS activity. As many as eight CHS genes are found in bean [34] and eight or more are found in petunia [25]. Why so many CHS genes? Is there redundancy of function or is each CHS gene serving a different purpose? If there is redundancy in function, then presumably another CHS gene family member would assume the function of a lost CHS member and there would be no loss of pigmentation observed in mutant phenotypes.

Genetic redundancy must be of some adaptive value to the plant to persist over evolutionary time. This follows from the observation that mutational pressure will eventually convert newly duplicated genes to pseudogenes owing to insertion/deletion events, nonsense mutations and frameshifts [46]. Once converted to a pseudogene the likely fate is to continue to accumulate mutations at all sites at the maximum rate, ultimately diverging beyond recognition. The persistence of a duplicate gene is therefore *prime facie* evidence that the duplicate is positively selected.

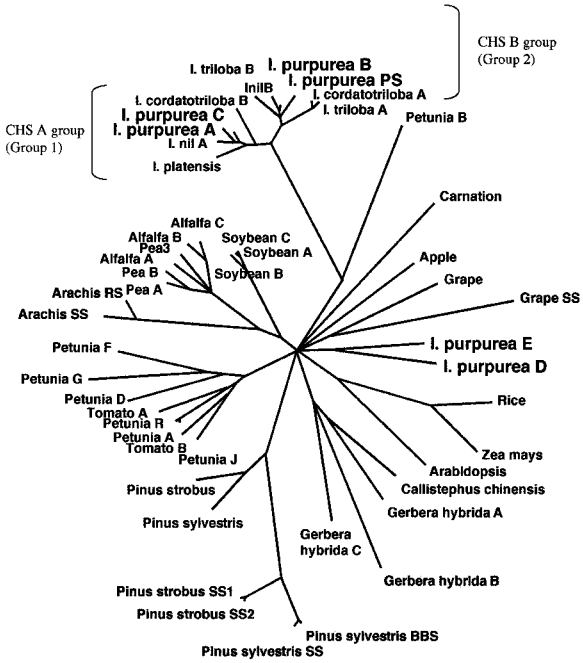

Figure 2. Neighbor joining tree of Kimura [22] 2-parameter distances. Multiple sequence alignment performed by ClustalW [42]. Phylogeny includes many of the CHS gene sequences currently in GenBank as well as the related stilbene synthases (SS), bibenzyl synthases (BBS), and resveratrol synthases (RS). Genes for which a shift in enzyme function has been determined are shown in shaded boxes. PS: pseudogene. Numbers or letters designate multiple genes.

Evidence for divergence in enzymatic function among duplicate CHS genes

There is growing evidence of repeated divergence in function of CHS gene family members. One functional shift that appears to have occurred repeatedly in plant evolution is the shift to stilbene synthase (SS) activity. Stilbene synthases are similar to CHS in that they are polyketide synthases, but they are involved in the production of stilbene phytoalexins rather than flavonoids [45]. Stilbene phytoalexins have anti-fungal properties that are thought to be important in defense against plant pathogens [44]. SS uses the same substrate as CHS to produce identical products up to the tetraketide intermediate. Ring closure is then accomplished by a conformational change in the intermediate with closure occurring at different atoms to produce either chalcone or stilbene. The SS reaction also includes a decarboxylation step that is not found in the CHS reaction [44]. Tropf *et al.* [44] analyzed the evolution of CHS and SS genes and showed that SS has evolved from CHS independently several times over the course of evolutionary history. Tropf *et al.* [44] used site-directed mutagenesis to demonstrate that only three amino acid changes in a CHS/SS hybrid construct were necessary to obtain SS enzymatic activity. This suggests that in nature very few amino acid changes are required to change the enzymatic function of a CHS gene. Figure 2 is a neighbor-joining tree showing the relationship of the SS genes with CHS genes from several species. CHS-like genes that have undergone a known functional shift, including stilbene synthase (SS), bibenzyl synthase (BBS), and resveratrol synthase (RS), are highlighted. SS, BBS, and RS may occur in more species, but have not yet been reported.

A second example of functional divergence is provided by *Gerbera hybrida* where gene duplication of CHS was followed by functional divergence. In this case, Helariutta *et al.* [15] analyzed the enzymatic properties of CHS-like genes and showed that the properties are distinct from both chalcone synthase and the stilbene synthases. Substrate testing showed that the novel CHS gene was unable to use 4-coumaroyl-CoA as a substrate but was able to use

benzoyl-CoA. A novel product is produced by the enzymatic reaction, although the role of this product in the plant is presently unknown. In addition to this change in substrate specificity, the novel CHS reaction has been truncated, which indicates a reversal to a more simplified catalytic reaction [15].

Another example of functional divergence is found in barley where Christensen *et al.* [3] have recently isolated a novel CHS-like gene that uses feruloyl-CoA and cafferoyl-CoA as the preferred substrate over cinnamoyl-CoA or 4-coumaroyl-CoA. This new CHS gene is pathogen- and UV light-induced and may be responsible for the production of a methylated phytoalexin [3]. Southern blot analysis indicates the possibility that there are seven genes in the barley genome that have sequence similarity with this novel CHS gene. This work indicates that genomes in which CHS genes have been previously reported might well contain as yet uncharacterized CHS and CHS-like genes.

Yet another example is found in *Pinus* where Schröder *et al.* [37] have characterized two new CHS-related genes and have demonstrated that whereas the CHS-related gene PStrCHS1 is capable of performing the typical CHS reaction, PStrCHS2 was inactive with malonyl-CoA and coumaroyl-CoA as substrates. Schröder [36] points out that very few of the reported CHS genes have been characterized in terms of their enzymatic properties and shown to have CHS activity. As more CHS genes are being characterized it is becoming apparent that CHS is a large, functionally divergent gene family.

CHS gene redundancy in the morning glory genus *(Ipomoea)*

The morning glory genus is characterized by a rich diversity of flower color polymorphisms. Figure 3 shows just a few of the many flower color variations found in *I. purpurea.* These particular morning glory phenotypes are the subject of our study in this paper. The Japanese morning glory (*I. nil*) has been a long-term subject of genetic study in Japan [20]. In addition, our laboratory has studied the genetics of flower color variation in the common morning glory (*I. purpurea*) over the past 15 years [7–9]. In recent years this activity has turned to the molecular characterization of genes involved in flavonoid biosynthesis. To put this work in an evolutionary context, we first consider salient features of research on petunia CHS gene family organization.

Petunia and tomato are members of the Solanaceae and are in the same order, Solanales, as is the family Convolvulaceae, which includes *Ipomoea.* As a consequence, petunia and tomato serve as a useful comparison to *Ipomoea.* The time of divergence between Solanaceae and Convolvulaceae is about 70 million years ago (assuming molecular clock arguments [6]), so a substantial level of molecular divergence has accumulated between these plant genomes. Eight complete CHS genes and four partial genes were cloned from *Petunia hybrida* [23]. On the basis of gene sequences, some of the *Petunia* CHS genes can be divided into two subfamilies. Intron sizes vary from about 700 to almost 4000 bp in the 8 fully sequenced genes but could range up to 15 kb in the four partially characterized genes based on restriction mapping [24]. *Petunia* CHS A is the main CHS transcript in the corolla, with CHS J also expressed but at a much lower level. CHS A and J transcript is found in floral tissue but not in leaf, root, or stems of plants grown under normal greenhouse conditions. CHS A, J, B, and G are expressed in *P. hybrida* seedlings in response to UV illumination; however, levels of CHS B and G are very low. No transcripts were detected for the remaining genes or gene fragments under these conditions [24]. *Petunia* CHS B is very different from the rest of the *Petunia* CHS sequences. As seen in Figure 2, the *Petunia* CHS B gene clusters with the *Ipomoea* CHS, B, C and PS genes. The long branch length for *Petunia* CHS B indicates an accelerated rate of evolution as is the case for the *Gerbera* B CHS, *Ipomoea* CHS A, B, C, PS and the SS, BBS, and RS genes. *Petunia* CHS B is located on chromosome V closely linked with CHS genes A, D, F, J and I. *Petunia* CHS C, G, L and E are located on chromosome II, also closely linked to one another [25].

To date six CHS genes have been described in *Ipomoea* [6, 10]. These data reveal two subfamilies that are highly diverged. One subfamily characterized by Durbin *et al.* [6] is composed of the CHS genes designated as A, B, C and PS (a putative pseudogene) and the second subfamily, characterized by Fukada-Tanaka *et al.* [10], is composed of CHS D and E. It is also important to note that CHS A and B are adjacent on the chromosome within 6 kb of one another and are probably subject to occasional interlocus exchange events [6, 18]. The linkage relations among the remaining genes are unknown. The CHS A, B, C and PS cluster can be further classified into two subgroups

Figure 3. The genotypes of the *I. purpurea* flowers pictured are as described [7, 8, 9]. A. Purple limb = AAPPIIWW. B. Intense pink limb = AAppiiWW. C. Light pink limb = AAppIIWW. D. Albino limb = aaPPIIWW. E. Albino limb with purple sectors = a*a*PPIIWW. F. Albino limb with purple rays = AAPPiiWW. Locus A is CHS D (this paper and Habu *et al.* [14]). Loci P, I, and W are as yet uncharacterized.

(group 1 defined by CHS A and group 2 defined by CHS B), which are distinguished by nine amino acid changes [6]. As already noted (Figure 2), CHS A, B, C, and PS have long branch lengths within the phylogeny, suggesting accelerated rates of nucleotide substitution. The long branch lengths trace back to a common lineage with the rapidly evolving *Petunia* CHS B gene. In contrast, CHS D and E have shorter branch lengths and slower rates of nucleotide substitution. Assuming average amino acid substitution rates of 8.8×10^{-10} (calculated from the individual rate data in Table 1), we estimate the divergence between the CHS A, B cluster and the CHS D, E cluster to have occurred more than 100 million years ago. Molecular clock-based calculations are extremely uncertain at such deep levels of evolution, but using an average of the CHS A, B and CHS D, E nucleotide substitution rates yields an estimate of 146 million years. An estimate slightly in excess of 100 million years is not inconsistent with the deep placement of the primary CHS duplication events within angiosperm evolution.

Evolutionary rates are heterogeneous among classes of *Ipomoea* CHS genes

Accelerated rates of amino acid replacement are often observed in conjunction with shifts in enzyme function [1, 29]. To examine this possibility for the *Ipomoea* CHS gene family, absolute rates of nonsynonymous substitution between the two CHS subfamilies were calculated (CHS A, B, C versus CHS D, E) (Table 1). These calculations reveal that the CHS A, B subfamily is evolving 2.7 times faster than the CHS D, E subfamily. The average synonymous to nonsynonymous ratios also varies between the two subfamilies, although the CHS D, E contrast may be biased by multiple substitutions at synonymous sites (Table 2).

Table 1. Rates of nonsynonymous substitution ($\times 10^{-9}$ per site, per year). Based on Jukes-Cantor [21] corrected distances for nonsynonymous substitutions between *I. purpurea* and the presumed *P. hybrida* orthologues (*Petunia* A for CHS D and E; *Petunia* B for CHS A, B, C and PS) and the estimated date of divergence between *Ipomoea* and *Petunia* of 70 million years.

CHS D, E subfamily	
CHS D	0.469 ± 0.066
CHS E	0.471 ± 0.066
CHS A, B subfamily	
CHS A	1.30 ± 0.117
CHS B	1.20 ± 0.111
CHS C	1.35 ± 0.119
CHS PS	1.24 ± 0.113

Table 2. Synonymous to nonsynonymous ratio of Jukes-Cantor [21] corrected distances. Distances calculated by MEGA (Molecular Evolutionary Genetics Analysis) [27].

CHS D, E subfamily	
CHS D to CHS E	12.70
CHS A, B subfamily	
CHS A to CHS C	3.43
CHS A to CHS B	5.54
CHS A to CHS PS	5.85
CHS B to CHS C	5.00
CHS B to CHS PS	9.43
CHS C to CHS PS	4.90
Average	*5.70*

Despite this potential bias, which tends to reduce the number of synonymous substitutions relative to nonsynonymous substitutions, the ratio calculated within the CHS D, E subfamily is more than two times higher than the average ratio within the CHS A, B groups, which strongly suggests greater selective constraint on replacement changes relative to silent changes in the CHS D, E subfamily.

Li and Bousquet [28] and Muse and Gaut [11, 32] relative rate tests were performed to test for heterogeneous rates of CHS evolution. Rate tests show that none of the lineages differ significantly in rates of synonymous substitution, although at distant levels of comparison this result may be confounded by saturation of synonymous positions. The rate of nonsynonymous substitution is, however, significantly faster in the CHS A, B groups than in the more conserved CHS D, E genes. The nonsynonymous rates do not differ significantly between CHS D and CHS E or between the Group 1 and Group 2 CHS genes in the CHS A, B subfamily.

Evidence of divergence in expression among CHS genes in morning glory

Figure 4 shows the expression of different CHS genes in various tissues and stages of development in common morning glory as determined by hybridization of radioactive DNA probes from the respective genes to total RNA on northern blots (methods described in figure legends). Tissue was dissected from flower bud stages 2 through 7 of *I. purpurea* genotype AAppii-WW (Figure 3). Each stage represents a time point taken at successive 24 h stages of floral development as described in Figure 5 culminating with stage 7 represented by an open flower. Figure 5 shows the dissected bud stages representing these time points; note that floral pigmentation can be seen as early as stage 4. CHS A, B, and C gene data are grouped because of their high degree of sequence similarity to one another and in some cases their low degree of expression, which makes it difficult to distinguish message from each gene separately based on hybridization. For some of the tissues RT-PCR was employed (as described in the figure legends) to determine the specific CHS gene present in the population of mRNA obtained from a particular tissue and developmental stage.

CHS D is by far the most abundantly expressed CHS gene in the limb. The peak level of CHS D mRNA expression corresponds to the appearance of pigment at stage 4. CHS D message falls off dramatically by stage 7. CHS A type message appears early in floral limb development and is expressed at a low level compared to CHS D. RT-PCR identified CHS A as the predominant if not the sole transcript among the CHS A, B and C group in this tissue. CHS E is weakly expressed in the floral limb stages 4 through 6 at a fairly constant level. In the tube of the flower, CHS D is expressed at a lower level than either CHS A type transcript or CHS E. RT-PCR identified the CHS C transcript in the tube. The peak of expression of CHS C and E genes in the tube occurs earlier than in the floral limb. Finally, there is a shorter transcript of unknown origin also hybridizing to the CHS D probe in the floral tissues.

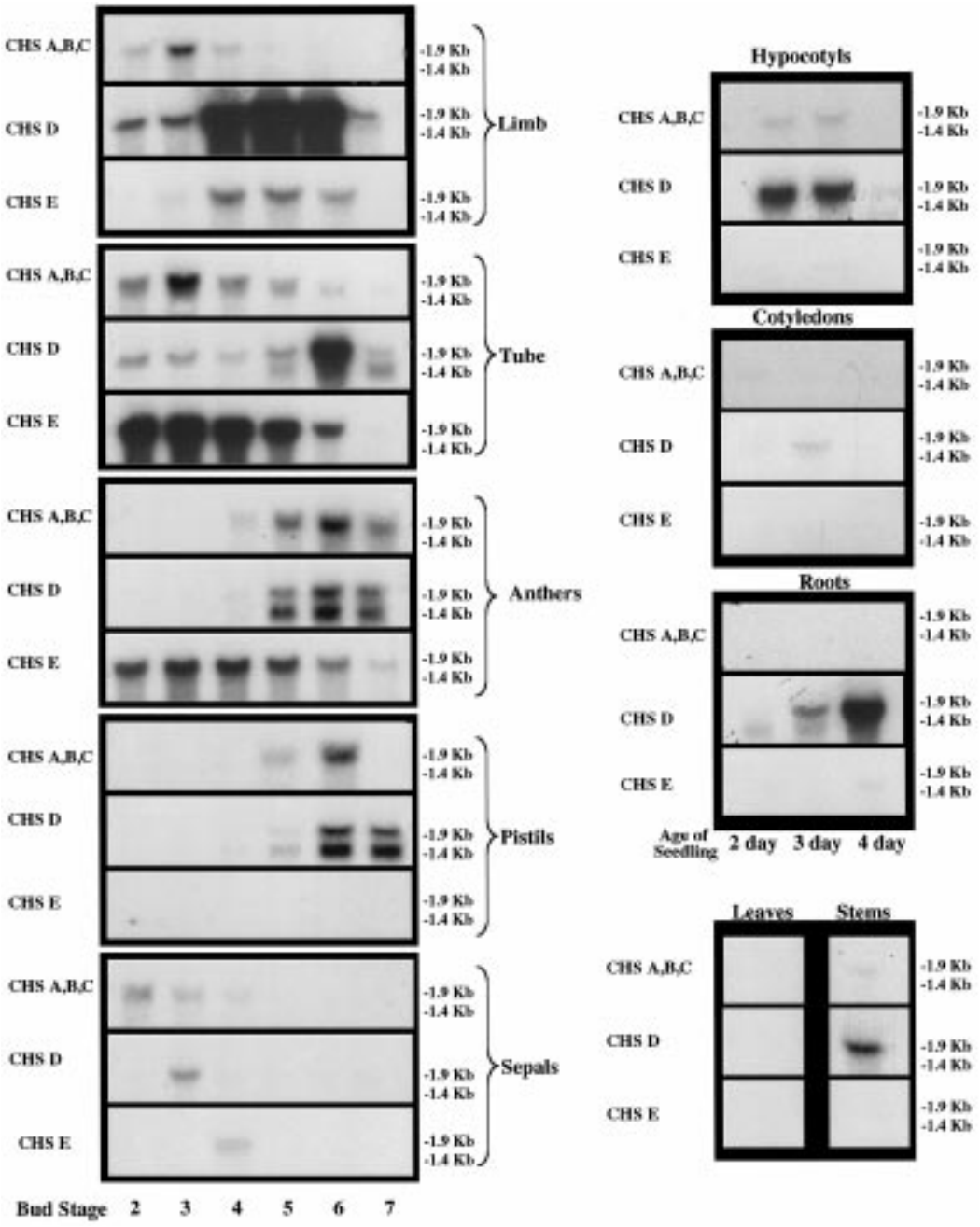

Figure 4. Expression of the different CHS genes in various tissues and stages of development in *I. purpurea* genotype AAppWWii (intense pink limb in Figure 3). A 15 μg portion of total RNA (as described [33]) was transferred from formaldehyde gels [30] to GeneScreen as per the manufacturer's instructions (New England Nuclear, Boston, MA) and UV-crosslinked at 120 mJ in a Ultra Lum Ultraviolet Crosslinker (Ultra Lum, Carson, CA). Photographing the ethidium bromide-stained gels and observing the intensity of ribosomal bands assessed equal loading of RNA. The blots were probed with a ^{32}P-labeled random-primed probe (Boehringer Mannheim, Indianapolis, IN). Hybridization was at 42 °C in the GeneScreen hybridization solution containing 50% formamide. Washes were at 50 °C, in 1× SSC, 2% SDS. Under these conditions, there was no cross-hybridization of the CHS A probe to either CHS D or CHS E or of CHS D to CHS E. The CHS A probe was a PCR product from a genomic clone and contained sequence from 123 from the ATG start site to 1198. CHS D probe was a PCR product from a cDNA clone and contained sequence from −121 of the ATG start site to 1001. CHS E probe was a PCR product form a cDNA clone and contained sequence from −39 of the ATG start site to 978. RT-PCR was used to determine which of the CHS A, B, and C genes were being expressed. cDNA was made from DNAse-treated total RNA using SuperScript II RNase H$^-$ Reverse Transcriptase according to the manufacturer's instructions (Gibco-BRL, Grand Island, NY). The cDNA was purified using phenol/chloroform and then PCR-amplified with primers designed to amplify CHS A, B, and C respectively. The products were sequenced to verify the identity of the product and to verify that it was indeed a cDNA and not a genomic contaminant. X-ray exposure time was 24 h.

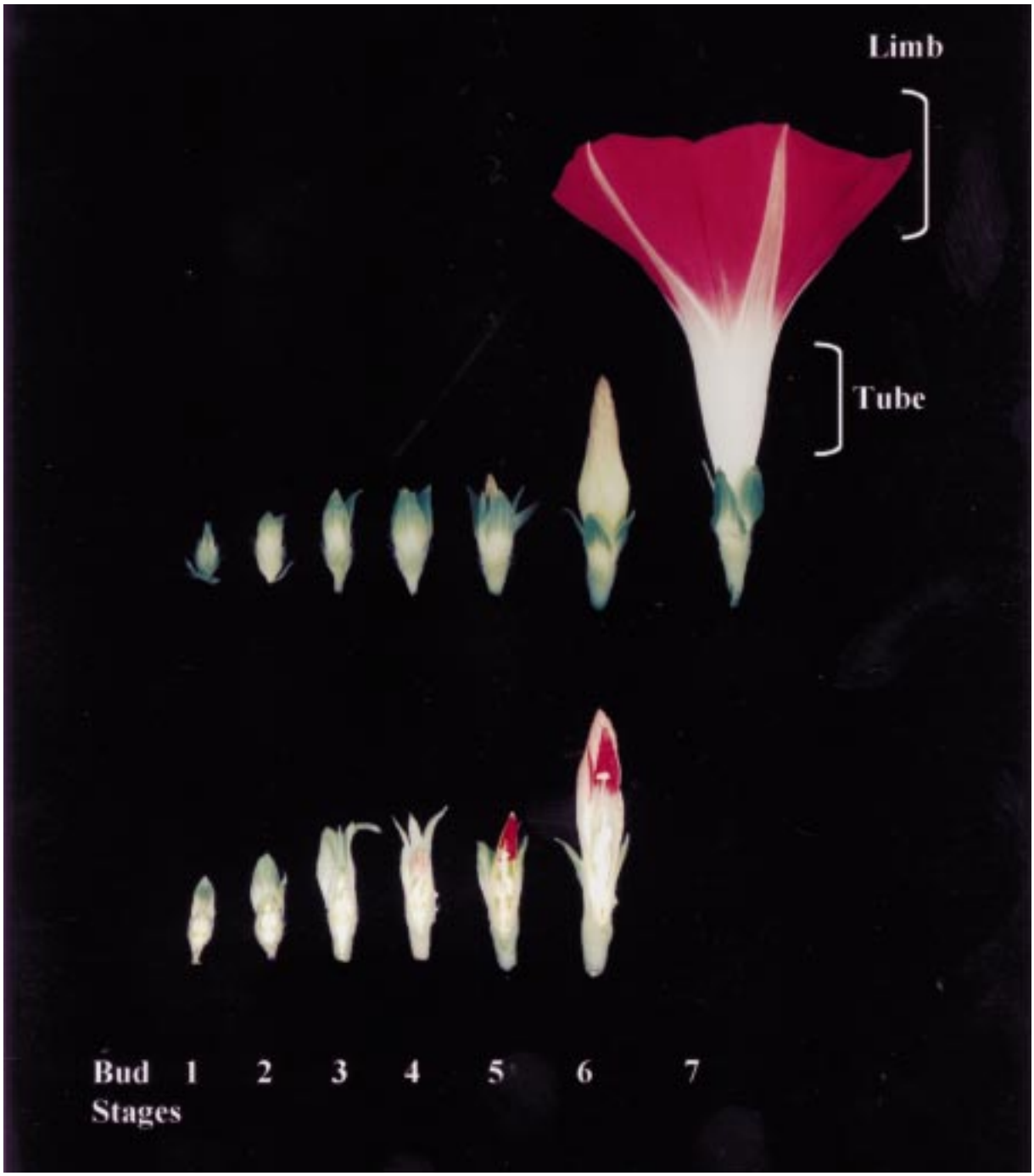

Figure 5. Developmental stages of the flower in *Ipomoea purpurea*. Stage 1 may persist for several days. The subsequent stages represent ca. 24 h periods. The flower is ephemeral, opening in the morning and senescing within a few hours. Stage 1 is ca. 8–10 mm. All organs are differentiated, pollen is close to maturity and the stigma is mature. Stage 2 is 10–15 mm with nearly mature ovary and anther length. Stage 3 is ca. 15 mm. Sepals have reached full length and the style is elongated above the anthers. In stage 4, which is also about 15 mm, pigment is present in the limb. Filaments of the stamens are elongating. The stigma, anthers and pollen are mature. In stage 5, the bud is 15–20 mm with fully pigmented limb tissue expanded beyond the sepals. There is massive elongation of the style and filaments. In stage 6 the bud is over 20 mm. There is further expansion and elongation of the tube, limb, filaments and style. In stage 7 the flower is fully open. The limb and tube are fully expanded and pollen is dehiscing.

CHS E is expressed earlier in the anthers than either CHS A type or CHS D transcripts. RT-PCR identified both CHS A and C transcript in the anther RNA population. Because the RNA was derived from anthers with pollen adhering to them, it is impossible to say from which tissue these two transcripts originated. CHS D is weakly expressed in anther tissue (compared to its abundant expression in the pigmented limb) and there is again the appearance of a shorter unknown transcript hybridizing with the CHS D probe. The shorter transcript is at about the same level of expression as the full-length CHS D transcript. In the pistils, the CHS A type is weakly expressed in stages 4 through 6 whereas E is not detected. CHS D and the shorter transcript are present in stages 5 through 7.

In sepals, CHS A type transcript is found to be expressed at a low level in stages 2 through 4 whereas CHS D and E are only weakly expressed in stages 3 and 4, respectively. In the hypocotyls of seedlings, pigment appears within 2 days of germination in the dark. Thus the production of pigment in the hypocotyl is developmentally regulated rather than light-regulated. CHS A transcript was detected by RT-PCR in the hypocotyl RNA. CHS D is the most abundant CHS transcript in the hypocotyl. CHS E was not detected. In the cotyledon, CHS B was determined by RT-PCR to be expressed at day 2 but the level of transcript is so low that it is not detected on the 24 h film exposure in Figure 4. This is the only tissue in which we have confirmed that CHS B is expressed. CHS D is weakly expressed at day 3 and CHS E was not expressed at a detectable level in cotyledons.

In dark-grown roots of seedlings the CHS A type gene transcripts are not detected. CHS E is weakly expressed at day 4. CHS D is weakly expressed at days 2 and 3, with expression levels increasing at day 4. There was no CHS gene expression detected in mature leaves at the stage selected for assay. In the pigmented mature stem, the CHS A type transcript was present at a low level. CHS D was expressed at a higher level, but there was no transcript detected for CHS E.

It is apparent from these results that the timing and tissue specificity of expression of the various CHS genes in *I. purpurea* is quite variable, suggesting that there has been considerable divergence in developmental regulation of individual members of the gene family. Both CHS A, B, C and E are most highly expressed in the unpigmented tube. Low levels of these genes are also found in most of the other tissues though they differ in expression patterns from each other and from CHS D. The levels of expression of CHS A, B, and C are significantly lower than that observed for CHS D and it takes a much longer exposure time (24 h versus 4 h or less for CHS D) to detect bands on the film.

An albino phenotype is associated with loss of CHS D expression

To determine if any of the CHS genes in *Ipomoea purpurea* are responsible for any of the flower color variations that occur in southeastern US populations of the common morning glory, we probed northern blots of RNA from the limb with clones for each of the duplicate CHS genes. The entire limb was used including the rays because of the difficulty in dissecting out the ray in the smaller bud stages. Results obtained from limb tissue excluding the ray for stages 5 through 7 gave the same results as shown in Figure 6.

Panel A in Figure 6 shows the expression pattern of CHS genes in morning glory with purple pigmented limbs (genotype AAPPIIWW shown in Figure 3). CHS D is again the most abundant CHS mRNA with the highest expression at stages 4, 5, and 6. CHS A was used as a probe, which will detect the presence of CHS B and C although RT-PCR indicates that it is CHS A that is expressed in the floral limb.

Panel B in Figure 6 shows the pattern of expression in the albino floral limb of *I. purpurea* (genotype aaIIPPWW shown in Figure 3). There is no detectable message for CHS D at the appropriate size for the full-length transcript. Instead there is the appearance of the shorter transcript that is also seen in other tissues such as the flower throat, anthers, and pistils (described above). The shorter transcript is not nearly as abundant as the full-length CHS D transcript normally observed in pigmented flower limbs. To determine if the shorter transcript is missing a portion of the CHS D gene, the blots were probed sequentially with labeled DNA made from exon 1 and a probe from exon 2 respectively. When this same blot was probed with labeled DNA that included only the 5′ portion of exon 1, the short transcript is not detected even after 4 days exposure (Figure 7, panel B). In contrast to CHS D, CHS A and CHS E show the same pattern of expression in the albino phenotype as in the pigmented flower limb (Figure 6).

Panel C in Figure 6 shows the pattern of expression in the limb of the morning glory with an albino flower limb with purple sectors (genotype a*a*PPIIWW shown in Figure 3). Again there is no

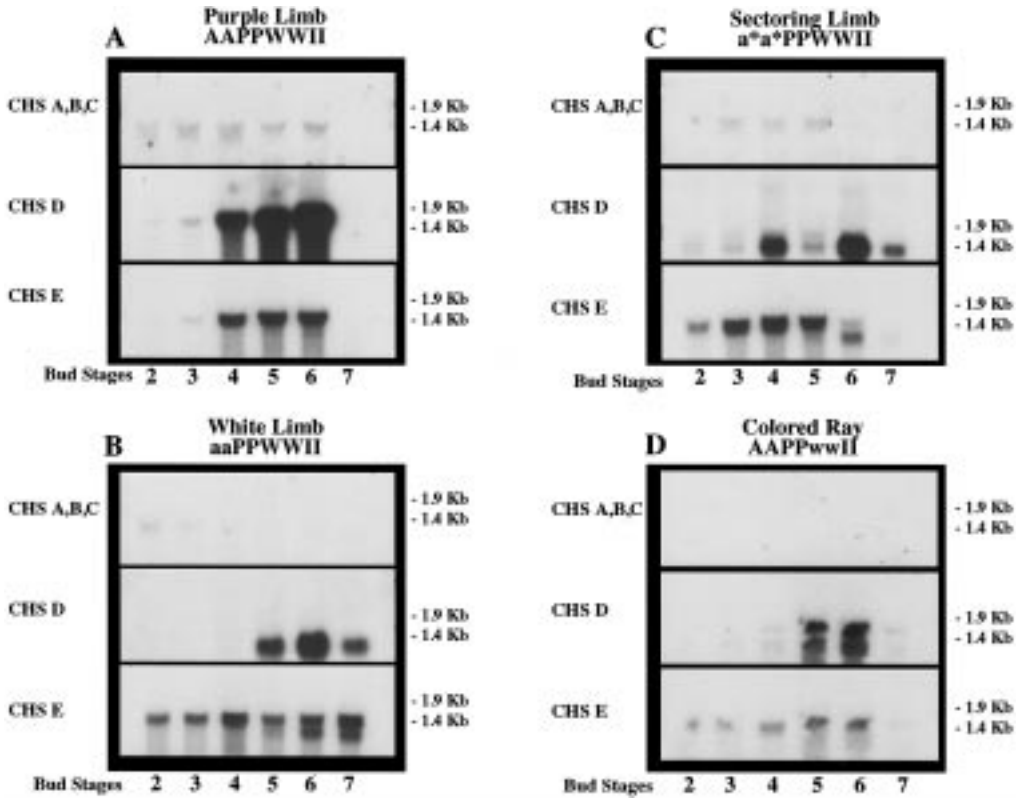

Figure 6. Expression of CHS genes in the limb of various *I. purpurea* phenotypes. Methods and probes are as described in Figure 4. The genotypes are as described in Figure 3.

full-length transcript for CHS D; however there is the shorter transcript, which does not hybridize with the probe from exon 1 (Figure 7, panel C). CHS A and CHS E again show the same pattern of expression as in the pigmented limb.

Panel D in Figure 6 shows the expression of CHS genes in the phenotype with the white corolla with pigment in the floral rays (genotype AAIIPPww shown in Figure 3). CHS D is expressed, as is the shorter transcript, although at a much lower level than in the fully pigmented limb. Again, when probed with exon 1 alone, only the full-length transcript is detected (Figure 7, panel D). CHS E is expressed but at a much lower level than in the fully pigmented limb. CHS A transcript was not detected.

The assays of albino and pigmented floral phenotypes point towards the conclusion that CHS D is solely responsible for anthocyanin production. This conclusion is based on the fact that there is a total absence of normal CHS D transcript in the floral limb of the albino phenotype whereas there is an abundance of CHS D message in the pigmented floral limb. The other structural genes in the pathway show the same expression pattern in the albino flowers as in the pigmented flowers suggesting that the albino phenotype is not the result of a downstream enzyme (unpublished data). It is not apparent why CHS D would be expressed in unpigmented tissues such as roots, anthers, and pistils, however, other tissues such as the floral throat, stem, hypocotyls, sepals, leaves and cotyledons are observed to have varying amounts of pigment associated with them depending on genotype.

The lower-molecular-weight transcript that is detected by the full-length CHS D probe is not detected in the pigmented floral limb where the full-length CHS D transcript is abundant. This shorter transcript may be due to defective RNA splicing or it may be another CHS locus, closely related to CHS D that lacks the first exon. The expression of the shorter transcript is tissue-specific, being more abundant in the anthers, pistils, throat and albino tissues.

The data presented here strongly implicate CHS D as the A/a locus. In the aa and a*a* genotypes the full-length CHS D transcript is undetected whereas the

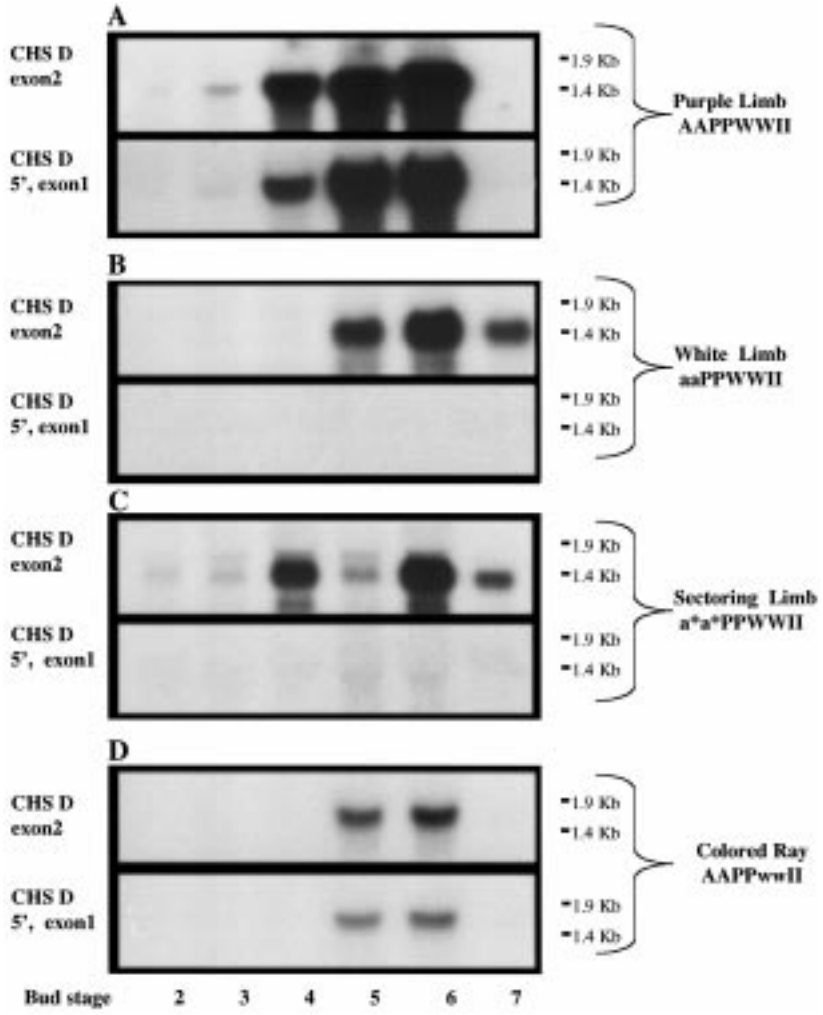

Figure 7. Expression pattern of CHS D in the various phenotypes when probed with exon 1 versus exon 2. The probe for exon 1 was a PCR product generated from a cDNA clone and includes sequences from −121 of the ATG start site to 158 from the start site. For exon 2 the probe was a PCR product generated from a cDNA clone and includes sequences 500 from the start site to 1001 from the start site in exon 2. We used a lengthy exposure time (4 days) in order to detect any weak signal. Since the probes are short (about 280 bp for exon 1) the signal is not as intense as when using the full-length probe (see Figure 6).

rest of the CHS genes are expressed normally as are the rest of the structural genes in the pathway (unpublished data). In addition, Habu *et al.* [13] report a reduction of CHS D transcript in one of their lines of common morning glory that bears white flowers. The expression data also implicates CHS D as the locus with unstable alleles (a*) and the insertion site for a transposable element [9]. This has recently been confirmed by Habu *et al.* [14] who have characterized a new transposable element (Tip100) and shown that the mutable a* allele in common morning glory is due to the insertion of the element in the intron of CHS D. Excision of the element from CHS D restores production of full-length transcript and restores pigmentation in the limb. These results illustrate the importance of using gene-specific probes to distinguish members of a gene family. Tiffin *et al.* [43] examined a single member of each gene family in the anthocyanin pathway in *I. purpurea* and having found transcripts for each of those particular genes concluded that one of the transcripts must therefore be defective in the a and a* mutants. Gene-specific probing of northerns with all the CHS genes revealed the lack of full-length transcript for CHS D in the a and a* mutants. We now have conclusive evidence that the A/a locus in *I. purpurea* is CHS D and that the lack of pigment in the a and

a* mutants is a result of the deficiency of full-length transcription of CHS D.

In the ww genotype CHS D and E are expressed although at reduced levels. CHS A was undetected after 24 h of exposure of the film. A longer exposure time might be necessary to detect CHS A, as its transcript in the fully pigmented limb is very low. Previous genetic analysis in our laboratory indicate that the W/w locus is a regulatory gene because normal patterns of gene expression are observed in the corolla rays but not in the mid petal regions [7]. The stem and hypocotyl of this genotype are also pigmented.

Conclusions

Molecular evolution occurs at several hierarchical levels. The level of nucleotide substitution modifies the protein or nucleotide sequence, which has a primary function, based on sequence specificity. A second level is the informational unit such as the gene or regulatory motif where the duplication of strings of nucleotides creates redundancy in information. The second level of adaptive change must be important in plant evolution because new genes are evidently recruited recurrently in plant lineages over time [4]. New duplicate gene copies are liable to mutation or to ectopic exchange with other related gene copies. Without positive selection for the maintenance of sequence fidelity, redundant gene copies will erode over time owing to mutational accumulation [46]. In previous work we estimated gene recruitment rates for the CHS, *rbcS* and ADH gene families and found a high rate of recruitment for CHS genes [4]. Based on this work, and the arguments above, we would expect functional shifts to accumulate relatively rapidly in CHS gene evolution. Our study of CHS gene evolution appears to confirm these expectations because biochemical analyses and phylogenetic evidence suggest repeated functional shifts in enzyme specificity during the course of seed plant evolution. In addition, we find considerable evidence for shifts in developmental expression among gene family members in *I. purpurea.*

An interesting aspect of CHS gene family in the lineage leading to the morning glory genus is the very early duplication that defines two major groups of CHS genes. This split is also mirrored in the petunia lineage (Figure 2). One class of *Ipomoea* CHS genes (D and E) is slowly evolving and appears to be under much greater selective constraint. We have shown in this article that CHS D is evidently the sole CHS gene responsible for the accumulation of pigment in the outer limb of the flower since lack of CHS D transcript results in an unpigmented phenotype. In contrast, CHS A, B and C are rapidly evolving with more than twice the rate of accumulation of amino acid change. It is not yet established whether this rapid rate of change is involved in a shift in enzymatic function, but there is clear precedence for such shifts in CHS function in other plant lineages. In addition, the CHS A, B, C and PS subgroup underwent a rapid expansion through duplication within the past 20 million years [6]. Subsequent to this expansion, there has been divergence between CHS A and B at several key amino acid sites that are thought to be important in enzyme specificity.

Stafford [41] postulated that the flavonoid pathway might be organized into aggregates or complexes such as exist in other pathways [39, 40, 17]. Such complexes or 'metabolons' may be very transient or long-lived. Metabolons offer many advantages in terms of kinetics, channeling of intermediates, and protection of labile intermediates [5]. Saslowsky *et al.* [35] have used immunolocalization to show that CHS and CHI are localized at overlapping positions around the cell wall and vacuoles. They further report that a variety of *in vitro* studies have shown that the two proteins interact. Srere [39] postulates that if a protein were to function as part of a metabolon then it must have conserved binding sites for maintenance of the complex and that it may be these binding sites that distinguish different isozymes. It is therefore possible that the amino acid differences observed in the *Ipomoea* CHS genes could be involved in directing the protein into the correct complex.

We are entering an era where comparative genomics will provide important guidance to research. Comparative genomics seeks to identify important adaptive signals from the analysis of evolutionary change among different lineages for genes deemed to be of interest. This task will be complicated in plant genomes by genetic redundancy. An exhaustive description of gene families is likely to be a prerequisite to informed analysis. The study of CHS evolution in the morning glory genome may help provide some guide to the strengths and limitations of the comparative genomics approach.

Acknowledgements

The work described in this article was supported in part by the Alfred P. Sloan Foundation. We would like to thank Dr Elizabeth Lord (University of California, Riverside) for her anatomical descriptions of the dissected bud stages in Figure 5. We would also like to thank Dr Shigeru Iida (National Institute for Basic Biology, Okazaki, Japan) for reading the manuscript and for his helpful suggestions.

References

1. Axelsen KB, Palmgren MG: Evolution of substrate specificities in the P-type ATPase superfamily. J Mol Evol 46: 84–101 (1998).
2. Beerhues L, Robeneck H, Wiermann R: Chalcone synthases from spinach (*Spinacia oleracea* L.). Planta 173: 544–553 (1988).
3. Christensen AB, Gregersen PL, Schröder J, Collinge DB: A chalcone synthase with an unusual substrate preference is expressed in barley leaves in response to UV light and pathogen attack. Plant Mol Biol 37: 849–857 (1998).
4. Clegg MT, Cummings ML, Durbin ML: The evolution of plant nuclear genes. Proc Natl Acad Sci USA 94: 7791–7798 (1997).
5. Debnam PM, Shearer G, Blackwood L, Kohl DH: Evidence for channeling of intermediates in the oxidative pentose phosphate pathway by soybean and pea nodule extracts, yeast extracts, and purified yeast enzymes. Eur J Biochem 246: 283–290 (1997).
6. Durbin ML, Learn GH, Huttley GA, Clegg MT: Evolution of the chalcone synthase gene family in the genus *Ipomoea.* Proc Natl Acad Sci USA 92: 3338–3342 (1995).
7. Ennos RA, Clegg MT: Flower color variation in the morning glory, *Ipomoea purpurea.* J Hered 74: 247–250 (1983).
8. Epperson BK, Clegg MT: Genetics of flower color polymorphism in the common morning glory (*Ipomoea purpurea*). J Hered 79: 64–69 (1988).
9. Epperson BK, Clegg MT: Unstable white flower color genes and their derivatives in the morning glory. J Hered 83: 405–409 (1992).
10. Fukada-Tanaka S, Hoshino A, Hisatomi Y, Habu Y, Hasebe M, Iida S: Identification of new chalcone synthase genes for flower pigmentation in the Japanese and common morning glories. Plant Cell Physiol 38: 754–758 (1997).
11. Gaut BS, Morton BR, McCaig BC, Clegg MT: Substitution rate comparisons between grasses and palms: synonymous rate differences at the nuclear gene *Adh* parallel rate differences at the plastid gene *rbcL*. Proc Natl Acad Sci USA 93: 10274–10279 (1996).
12. Gottlieb O: Phytochemicals: differentiation and function. Phytochemistry 29: 1715–1724 (1990).
13. Habu, Y, Fukada-Tanaka, S, Hisatomi Y, Iida, S: Amplified restriction fragment length polymorphism-based mRNA fingerprinting using a single restriction enzyme that recognizes a 4-bp sequence. Biochem Biophys Res Comm 234: 516–521 (1997).
14. Habu, Y, Hisatomi Y, Iida S: Molecular characterization of the mutable *flaked* allele for flower variegation in the common morning glory. Plant J 16: 371–376 (1998).
15. Helariutta Y, Kotilainen M, Elomaa P, Kalkkinen N, Bremer K, Teeri TH, Albert VA: Duplication and functional divergence in the chalcone synthase gene family of Asteraceae: evolution with substrate change and catalytic simplification. Proc Natl Acad Sci USA 93: 9033–9038 (1996).
16. Hrazdina G, Zobel AM, Hoch HC: Biochemical, immunological and immunocytochemical evidence for the association of chalcone synthase with endoplasmic reticulum membranes. Proc Natl Acad Sci USA 84: 8966–8970 (1987).
17. Hrazdina G, Jensen RA: Spatial organization of enzymes in plant metabolic pathways. Annu Rev Plant Physiol Plant Mol Biol 43: 241–267 (1992).
18. Huttley GA, Durbin ML, Glover DE, Clegg MT: Nucleotide polymorphism in the chalcone synthase: A locus and evolution of the chalcone synthase multigene family of common morning glory *Ipomoea purpurea.* Mol Ecol 6: 549–558 (1997).
19. Ibrahim RK, Varin L: Flavonoid enzymology. Meth Plant Biochem 9: 99–131 (1993).
20. Imai Y: Analysis of flower colour in *Pharbitis nil.* J Genet 24: 203–224 (1931).
21. Jukes TH, Cantor CR: Evolution of protein molecules. In: Mums HN (ed), Mammalian Protein Metabolism, pp. 21–132. Academic Press, New York (1969).
22. Kimura M: A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16: 111–120 (1980).
23. Koes RE, Spelt CE, Mol JNM, Gerats AG: The chalcone synthase multigene family of *Petunia hybrida* (V30): sequence homology, chromosomal localization and evolutionary aspects. Plant Mol Biol 10: 159–169 (1987).
24. Koes RE, Spelt CE, Mol JNM: The chalcone synthase multigene family of *Petunia hybrida* (V30): differential, light-regulated expression during flower development and UV light induction. Plant Mol Biol 12: 213–225 (1989).
25. Koes RE, Spelt CE, van den Elzen PJM, Mol JNM: Cloning and molecular characterization of the chalcone synthase multigene family of *Petunia hybrida.* Gene 81: 245–257 (1989).
26. Koes RE, Quattrocchio F, Mol JNM: The flavonoid biosynthetic pathway in plants: function and evolution. BioEssays 16: 123–132 (1994).
27. Kumar S, Tamura K, Nei M: MEGA: Molecular Evolutionary Genetics Analysis, version 1.0. Pennsylvania State University, University Park, PA 16802 (1993).
28. Li P, Bousquet J: Relative-rate test for nucleotide substitutions between two lineages. Mol Biol Evol 9: 1185–1189 (1992).
29. Li W-H: Molecular Evolution. Sinauer Associates, Sunderland, MA (1997).
30. Maniatis T, Fritsch EF, Sambrook J: Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1982).
31. Martin CR: Structure, function, and regulation of the chalcone synthase. Int Rev Cytol 147: 233–284 (1993).
32. Muse SV, Gaut BS: A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to the chloroplast genome. Mol Biol Evol 11: 715–724 (1994).
33. O'Neill SD, Tong Y, Sprolein B, Forkmann G, Yoder J: Molecular genetic analysis of chalcone synthase in *Lycopersicon esculentum* and an anthocyanin-deficient mutant. Mol Gen Genet 224: 279–288 (1990).

34. Ryder TB, Hedrick SA, Bell JN, Liang X, Clouse SD, Lamb CJ: Organization and differential activation of a gene family encoding the plant defense enzyme CHS in *Phaseolus vulgaris*. Mol Gen Genet 210: 219–233 (1987).
35. Saslowsky DE, Shirley BW: Immunolocalization of two flavonoid enzymes. In: Arabidopsis. 9th International Conference on Arabidopsis Research, p. 108. University of Wisconsin-Madison (1998).
36. Schröder J: A family of plant-specific polyketide synthases: facts and predictions. Trends Plant Sci 2: 373–378 (1997).
37. Schröder J, Raiber S, Berger T, Schmidt A, Schmidt J, Soares-Sello A, Bardshiri E, Strack D, Simpson TJ, Veit M, Schröder G: Plant polyketide synthases: a chalcone synthase-type enzyme which performs a condensation reaction with methylmalonyl-CoA in the biosynthesis of C-methylated chalcones. Biochemistry 37: 8417–8425 (1998).
38. Sommer H, Saedler H: Structure of the chalcone synthase gene of *Antirrhinum majus*. Mol Gen Genet 202: 429–434 (1986).
39. Srere PA, Sümegi B, Sherry AD: Organizational aspects of the citric acid cycle. Biochem Soc Symp 54: 173–182 (1987).
40. Srivastava DK, Bernhard SA: Metabolite transfer via enzyme-enzyme complexes. Science 234: 1081–1084 (1986).
41. Stafford HA: Flavonoid evolution: an enzymic approach. Plant Physiol 96: 680–685 (1991).
42. Thompson JD, Higgins DG, Gibson TJ: Clustal W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight-matrix choice. Nucl Acids Res 22: 4673–4680 (1994).
43. Tiffin P, Miller R, Rausher M: Control of expression patterns of anthocyanin structural genes by two loci in the common morning glory. Genes Genet Syst 73: 105–110 (1998).
44. Tropf S, Lanz T, Rensing SA, Schröder J, Schröder G: Evidence that stilbene synthases have developed from chalcone synthases several times in the course of evolution. J Mol Evol 38: 610–618 (1994).
45. Tropf S, Kärcher B, Schröder G, Schröder J: Reaction mechanisms of homodimeric plant polyketide synthases (stilbene and chalcone synthase). A single active site for the condensing reaction is sufficient for synthesis of stilbenes, chalcones, and 6′-deoxychalcones. J Biol Chem 270: 7922–7928 (1995).
46. Walsh JB: How often do duplicated genes evolve new functions? Genetics 139: 421–428 (1995).