Characterization of Flavonol Synthase and Leucoanthocyanidin Dioxygenase Genes in Arabidopsis[1]

Further Evidence for Differential Regulation of "Early" and "Late" Genes

Matthew K. Pelletier, Julie R. Murrell, and Brenda W. Shirley*

Department of Biology, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061–0406

As part of an ongoing investigation into the organization and regulation of the flavonoid biosynthetic pathway, two *Arabidopsis thaliana* expressed sequence tag (EST) clones (153O10T7 and YAY780) with high homology to leucoanthocyanidin dioxygenase (LDOX) or flavonol synthase (FLS) were identified. EST YAY780 was sequenced and found to encode a protein 49 to 78% identical to all LDOX sequences in the database. EST 153O10T7 was used to isolate a genomic clone encoding a protein with 59 to 61% sequence identity to petunia (*Petunia hybrida*) and potato (*Solanum tuberosum*) FLS. DNA blot analysis was used to screen the Arabidopsis genome for sequences related to FLS and LDOX and to determine the positions of the two clones on the RI map. The expression patterns of FLS and LDOX in etiolated seedlings moved to white light and in two putative regulatory mutants (*ttg* and *tt8*) were determined by RNA blot analysis. These studies indicate that FLS is an "early" flavonoid gene in Arabidopsis seedlings, whereas LDOX is a "late" gene. Furthermore, FLS is the first flavonoid enzyme identified in Arabidopsis that may be encoded by a gene family.

Flavonoids are important secondary metabolites derived from malonyl-CoA and the aromatic amino acid Phe. These compounds are found throughout the plant kingdom and to date more than 3000 different flavonoids have been identified. Individual plant species can synthesize a variety of flavonoid compounds, which function in providing pigmentation to attract pollinators, in defending plants against pathogens, in acting as signal molecules in plant-microbe interactions, and in protecting plants from UV radiation (reviewed in Dooner et al., 1991; Koes et al., 1993; Shirley, 1996). However, these roles are not always conserved across plant species; for example, flavonoids are required for male fertility in maize (*Zea mays*) and petunia (*Petunia hybrida*), but not in Arabidopsis (Mo et al., 1992; van der Meer et al., 1992; Burbulis et al., 1996). The steps involved in flavonoid biosynthesis are well-characterized biochemically, and genes encoding both structural enzymes and regulatory proteins have been cloned from a variety of plant species (Fig. 1).

Numerous studies indicate that the expression of individual flavonoid genes is controlled by distinct mechanisms in a tissue- or species-specific manner (Martin et al., 1991; Jackson et al., 1992; Quattrocchio et al., 1993; Deboo et al., 1995; Shirley et al., 1995; Pelletier and Shirley, 1996). RNA blot analyses of etiolated seedlings moved to light and of plants carrying mutations at regulatory loci have revealed that genes encoding flavonoid enzymes can be divided into two groups designated as "early" and "late" (Martin et al., 1991; Kubasek et al., 1992; Koes et al., 1993; Shirley et al., 1995; Pelletier and Shirley, 1996). The "early" genes appear to be coordinately regulated and encode enzymes functioning near the beginning of the pathway, whereas the "late" genes exhibit a distinct pattern of expression. It is interesting that the individual genes that make up the "early" and "late" groups differ, possibly due to the unique functions of flavonoids in a given plant species or tissue. For example, F3H is coordinately regulated with the "early" genes in Arabidopsis seedlings, but with "late" genes in snapdragon (*Antirrhinum majus*) flowers (Martin et al., 1991; Kubasek et al., 1992; Pelletier and Shirley, 1996).

Significantly less is known about the expression patterns of FLS and LDOX, as genes encoding these enzymes have been cloned from only a few plant species. Soluble enzyme preparations from petunia and *Matthiola incana* flower buds and illuminated parsley cells were shown to catalyze the conversion of the flavanones kaempferol and quercetin to the flavonols dihydrokaempferol and dihydroquercetin (Britsch et al., 1981; Spribille and Forkmann, 1984; Forkmann et al., 1986). Full activity in these extracts required the presence of 2-oxoglutarate, ascorbate, and Fe^{2+}. Based on these requirements, the enzyme that catalyzes this reaction, FLS, belongs to a family of 2-oxoglutarate-dependent dioxygenases that also includes the flavonoid enzymes F3H, flavone synthase I, and LDOX (Britsch, 1990b; Britsch et al., 1993). The first gene encoding FLS was later isolated from petunia by differential screening of cDNA from *FLFL* and *flfl* lines (Holton et al., 1993). By

Abbreviations: CHI, chalcone isomerase; CHS, chalcone synthase; DFR, dihydroflavonol reductase; DIG, digoxygenin; EST, expressed sequence tag; F3H, flavanone 3-hydroxylase; FLS, flavonol synthase; LDOX, leucoanthocyanidin dioxygenase; *tt*, transparent testa.

Figure 1. Schematic of the Arabidopsis flavonoid biosynthetic pathway. Enzymes for which genes have been cloned in Arabidopsis are shown in bold, whereas those described in this study are also boxed. The names of specific compounds are given below the chemical structures and flavonoid classes are indicated in boldface. F3'H, flavonoid 3'-hydroxylase; UFGT, UDP-flavonoid 3-O-glucosyl transferase.
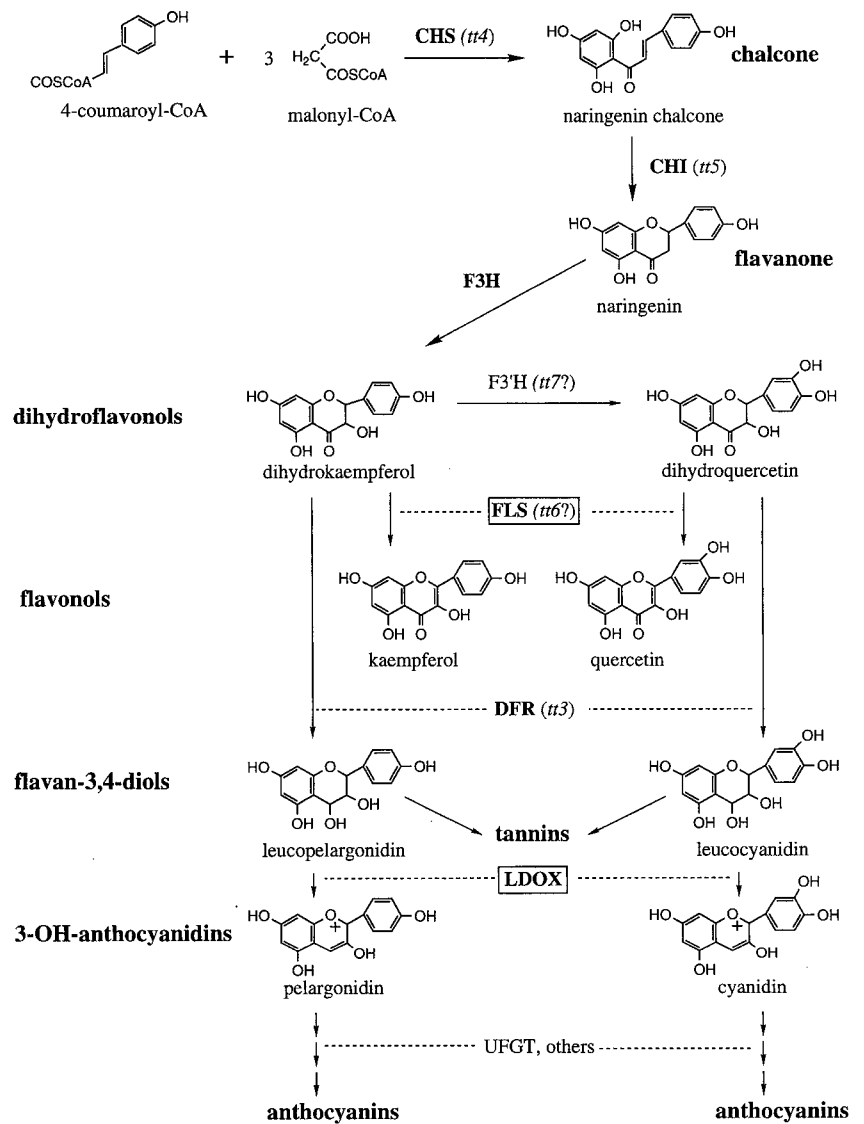
overexpressing the cloned gene in yeast and assaying extracts for activity, the identity of the putative FLS was confirmed. Although low-stringency DNA blot analysis indicated that a single gene encodes FLS in petunia, transgenic lines expressing an FLS antisense construct produced appreciable levels of flavonols, suggesting that multiple FLS genes may exist (Holton et al., 1993). The only other FLS clone that has been isolated to date is a cDNA from potato (GenBank accession no. X92178).

Genetic loci involved in the conversion of flavan-3,4-diols to anthocyanidins (Fig. 1) have been identified in maize (*A2*) (Reddy and Coe, 1962) and snapdragon (*Candi*) (Martin et al., 1991). The maize gene was cloned by transposon tagging, and particle bombardment into the maize aleurone confirmed that the cloned gene complemented the *A2* mutation (Menssen et al., 1990). Two genes encoding "late" flavonoid enzymes were identified and cloned from snapdragon by differential screening using a regulatory mutant (*del*) that had reduced activities of several enzymes

acting late in flavonoid biosynthesis (Martin et al., 1991). Linkage analysis showed that DNA homologous to one of the clones had been deleted from the *candi* mutant, and feeding experiments revealed that this mutant was blocked at an enzymatic step after DFR but before UDP-flavonoid 3-O-glucosyl transferase. The *Candi* clone was subsequently used to isolate homologs from apple and grape (Davies, 1993; Sparvoli et al., 1994), and a petunia gene with homology to *A2* and *Candi* was isolated by differential screening (Weiss et al., 1993). The proteins encoded by these genes have been referred to as LDOX and anthocyanidin hydroxylase. Although the precise reaction mechanism for the conversion of flavan-3,4-diols to anthocyanidins remains unknown, sequence similarity indicates that A2 and its homologs belong to the 2-oxoglutarate-dependent family of dioxygenases. We will therefore refer to the protein as LDOX throughout this manuscript.

To further elucidate the regulation of flavonoid genes in Arabidopsis seedlings, we have obtained clones encoding

FLS and LDOX and have characterized these at the molecular level. Low-stringency DNA blot analysis indicated that LDOX is encoded by a single gene, whereas several sequences related to FLS are present in the Arabidopsis genome. The map positions of the cloned genes were determined using recombinant inbred lines in an effort to correlate these genes with previously identified *tt* mutants. RNA blot analysis of the expression of these genes in response to white light and in two regulatory mutants indicated that the cloned FLS gene is an "early" gene, whereas LDOX is a "late" gene in Arabidopsis seedlings.

MATERIALS AND METHODS

Identification, Isolation, and Sequence Analysis of Arabidopsis LDOX and FLS Clones

Arabidopsis thaliana EST clones YAY780 and 153O10T7 were the gift of Jerome Giraudat (Institut des Sciences Vegetales, Gif, France) and Tom Newman (Michigan State University, East Lansing, MI), respectively. YAY780 plasmid DNA was isolated according to the method of Ish-Horowicz and Burke (1981). Sequencing was performed on both strands of the cDNA insert using T3 and T7 primers and primers derived from this primary sequence data with a DNA sequencer (377 Prism, Applied Biosystems) at the University of Virginia Biomolecular Research Facility (Charlottesville). Sequence analysis was carried out using the software package from DNASTAR (Madison, WI). The *SalI/NotI* insert of clone 153O10T7 was subcloned into pBluescript KS+ (Stratagene) to create pFLS.cDNA, and additional sequence data were obtained as described above. A genomic clone corresponding to this FLS gene was isolated from a λ FIX library of total Landsberg genomic DNA (Voytas et al., 1990). This library was plated on *Escherichia coli* strain ER1458, and approximately 30,000 plaques were screened for the presence of sequences with FLS homology. Plaque lifts were made using 0.2-μm nylon filters (Biotrans, ICN), which were then autoclaved for 10 min and irradiated using a UV cross-linker (Fisher Scientific). The nylon filters were processed as described for DNA and RNA gel-blot analysis using a pFLS.cDNA-derived probe and a hybridization temperature of 62°C. Two clones containing sequences homologous to FLS were purified to homogeneity and DNA was prepared from liquid lysates according to the method of Grossberger (1987). Restriction mapping and DNA blot analysis were performed on one of the clones. A 3-kb *EcoRI/SalI* fragment that appeared to contain the entire coding region of FLS and approximately 1 kb of promoter sequence was subcloned into pBluescript KS+ (Stratagene) to create plasmid pFLS3.0. Sequence analysis was performed as described above.

Sequencing of FLS from *tt6*

Genomic DNA was prepared from *tt6* plants as described previously (Pelletier and Shirley, 1996). The coding region of FLS was amplified by PCR using 100 ng of this DNA as a template and the following primers with dena-turing at 94°C for 45 s, annealing at 51°C for 45 s, and polymerizing for 4 min at 72°C: FLSs3, 5' CACTCTA-GAAACCAAGTC 3'; and FLSa5, 5' CAACTCAATAAC-GAACC 3'. The PCR product was purified (QiaQuick column, Qiagen, Chatsworth, CA) and the sequence was analyzed as described above.

Probe Synthesis

An Arabidopsis cDNA encoding amino acids 6 to 336 of the Arabidopsis LDOX protein was synthesized by reverse-transcription-PCR using 10 μg of total RNA from 3-d-old Landsberg seedlings grown in continuous white light as described previously (Shirley and Hwang, 1995). The following primers were used in this reaction: LDOXs1, 5' CGCGGATCCAGAGTTGAGAGTCTAGC 3'; and LDOXa1, 5' CGCGGATCCTTAAGCAAAAGTCCGTGGA-G 3'. Denaturation was carried out at 94°C for 30 s, annealing at 51°C for 30 s, and polymerization at 72°C for 90 s, and the product was purified from an agarose gel (Qia-Quick column). This fragment was subsequently cloned into the *Bam*HI site of pBluescript KS+ (Stratagene) to create pLDOX.cDNA. A LDOX probe was synthesized from this clone by PCR using the LDOXs1 and LDOXa1 primers. In this reaction, DIG-labeled dUTP (Boehringer Mannheim) was included in the deoxynucleotide mixture at a ratio of 1:9 (DIG-dUTP:dTTP), and the probe was purified from unincorporated nucleotides using a Qia-Quick column. A FLS probe for library screening was synthesized and purified in the same manner, except that pFLS.cDNA was used as template with T3 and T7 primers. The FLS probe for RNA blot analysis was synthesized by PCR using the T7 and either the T3 or FLSs1 (5'CATGC-CATGGAGTTCATCAGATCAGAG 3') primers and annealing temperatures of 46°C (T3) or 48°C (FLSs1).

DNA and RNA Gel-Blot Analysis

Genomic DNA was isolated from Landsberg and Columbia plants grown in soil as described previously (Pelletier and Shirley, 1996). The DNA (5 μg) was digested with restriction enzyme, fractionated in a 0.8% agarose gel with Tris-borate-EDTA buffer, and transferred to a 0.2 μM nylon membrane (Biotrans, ICN) in 25 mM sodium phosphate (pH 6.5) by capillary action. Hybridization was carried out using the FLS or LDOX probes at 55 or 65°C in buffer containing 0.5 M sodium phosphate (Church and Gilbert, 1984) as described previously. Probe detection was carried out as described in Pelletier and Shirley (1996), except that CDP Star (Boehringer Mannheim) was used as the chemi-luminescent substrate. The membrane that was hybridized at 65°C was stripped according to the manufacturer's protocol and rehybridized with fresh FLS probe at 55°C.

For RNA blot analysis, total RNA was isolated from 3-d-old seedlings according to the method of Simon et al. (1992). Ten micrograms of total RNA was fractionated in formaldehyde gels (Gerard and Miller, 1986), and the gels were then stained with ethidium bromide to ensure that all lanes had been loaded equally. The RNA was subsequently transferred to a nylon membrane and hybridized with

DIG-labeled probes as described for DNA blot analysis, except that hybridization was at 65°C.

Mapping of FLS and LDOX

A collection of recombinant inbred lines constructed by Lister and Dean (1993) was obtained from the Arabidopsis Biological Resource Center (Columbus, OH). The following 28 lines were used to determine the map positions of FLS and LDOX: CL4, CL5, CL13, CL33, CL37, CL115, CL190, CL191, CL194, CL217, CL231, CL238, CL242, CL245, CL263, CL267, CL283, CL284, CL288, CL295, CL302, CL303, CL332, CL356, CL358, CL367, CL370, and CL377. Among 17 enzymes tested (BamHI, BclI, BglII, BstEII, CfoI, DdeI, DraI, EcoRI, EcoRV, HindIII, NcoI, SacI, SalI, Sau3A, XbaI, XhoI, and XmaI), only CfoI revealed a polymorphism between Landsberg and Columbia ecotypes when either a pFLS.cDNA- or pLDOX.cDNA-derived probe was used for DNA blot analysis. Therefore, this enzyme was used to digest 1 μg of genomic DNA from each of the recombinant inbred lines, and blot analysis was carried out as described above at a hybridization and washing temperature of 65°C. Segregation data were analyzed using Mapmaker (Lander et al., 1987) to establish the locations of the two genes on the recombinant inbred map (Lister and Dean, 1993).

RESULTS

Cloning and Sequencing of FLS and LDOX

To further elucidate the regulation of flavonoid genes in Arabidopsis seedlings, clones encoding FLS and LDOX were identified and characterized at the molecular level. An Arabidopsis EST clone (153O10T7) with high homology to petunia and potato FLS was obtained from Michigan State University. Sequence analysis of this clone indicated that it did not encode a full-length FLS protein, based on an alignment of its deduced amino acid sequence with petunia and potato FLS. The clone was therefore used to screen an Arabidopsis (Landsberg) genomic library (see "Materials and Methods") and two phage clones were purified to homogeneity (Voytas et al., 1990). Restriction mapping was performed on one of these clones, and a 3-kb EcoRI/SalI fragment that contained the entire coding region and approximately 1 kb of promoter region was subcloned into pBluescript, creating pFLS3.0.

The sequence of the FLS gene, including 451 bp of promoter, the entire coding region, and 216 bp of trailer sequence, was determined. FLS was found to contain two introns, based on comparison with the sequences of the genomic and EST clones. These introns occur at positions identical to those found in the Arabidopsis gibberellin 20-oxidase and tomato E8 genes and correspond to two of four positions that are conserved among all plant dioxygenases characterized to date (Prescott and John, 1996). The open reading frame encodes a protein 336 amino acids in length, with an apparent molecular mass of 38.3 kD. The Arabidopsis FLS protein is very similar to the two published sequences from petunia and potato, showing 61 and 59% amino acid identity, respectively (Holton et al., 1993;

GenBank accession no. X92178) (Table I). The amino acid sequence also has 38% identity to Arabidopsis LDOX (described below) and 26% identity with Arabidopsis F3H (Pelletier and Shirley, 1995).

An Arabidopsis EST clone (YAY780) with high homology to grape, petunia, maize, snapdragon, and apple LDOX (Menssen et al., 1990; Martin et al., 1991; Davies, 1993; Weiss et al., 1993; Sparvoli et al., 1994) was obtained from the Institut des Sciences Vegetales. Sequence analysis of the entire insert of this clone indicated that it encoded a full-length protein that is 356 amino acids in length with an apparent molecular mass of 40.4 kD. The Arabidopsis protein is highly homologous to all previously characterized LDOX sequences, with the maize A2 gene product being the least similar (49% amino acid identity) and the apple anthocyanidin synthase (LDOX; 78% identical) the most similar (Menssen et al., 1990; Davies, 1993; Weiss et al., 1993; Sparvoli et al., 1994; C. Martin, personal communication) (Table I). The clone also exhibited significant homology to an ethylene-forming enzyme from white spruce (GenBank accession no. L42466) and Arabidopsis SRG1 protein (GenBank accession no. X79052).
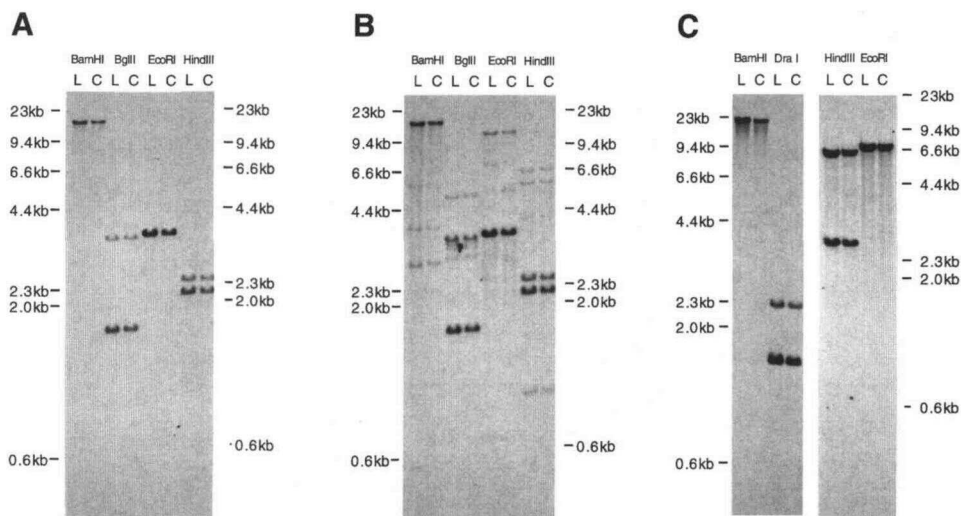
DNA Blot Analysis and Mapping of FLS and LDOX

To determine the number of sequences related to FLS and LDOX in Arabidopsis, genomic DNA from ecotypes Landsberg and Columbia was digested with several different restriction enzymes and blot analysis was carried out at high (65°C) and low stringency (55°C). For FLS, a single band was observed when an enzyme that did not cut within the probe region was used and hybridization was carried out under stringent conditions (Fig. 2A). However, multiple DNA fragments were detected when the same filter was rehybridized at lower stringency, indicating that more than one sequence with homology to FLS exists in the Arabidopsis genome (Fig. 2B). In contrast, only a single band was detected when an enzyme that did not cut within the probe region was used and filters were hybridized with the LDOX probe at low stringency (Fig. 2C). In addition, polymorphisms were detectable between Landsberg and Columbia ecotypes when CfoI was used to digest the genomic DNA and filters were hybridized with either probe (data not shown).

To ascertain whether FLS or LDOX corresponded to any of the previously characterized tt loci, the map positions of

Table I. *Comparison of FLS and LDOX protein sequences to the deduced amino acid sequences of the Arabidopsis genes*

Gene	Percent Identity
FLS	
Potato	61
Petunia	59
LDOX	
Apple	78
Snapdragon	74
Petunia	73
Grape	61
Maize	49

A

BamHI BglII EcoRI HindIII
L C L C L C L C

23kb —
9.4kb—
6.6kb—
4.4kb—
2.3kb—
2.0kb—
0.6kb—

— 23kb
— 9.4kb
— 6.6kb
— 4.4kb
— 2.3kb
— 2.0kb
·0.6kb

B

BamHI BglII EcoRI HindIII
L C L C L C L C

23kb —
9.4kb—
6.6kb—
4.4kb—
2.3kb—
2.0kb—
0.6kb—

— 23kb
— 9.4kb
— 6.6kb
— 4.4kb
— 2.3kb
— 2.0kb
— 0.6kb

C

BamHI Dra I   HindIII EcoRI
L C L C   L C L C

23kb—
9.4kb—
6.6kb—
4.4kb—
2.3kb—
2.0kb—
0.6kb—

— 23kb
— 9.4kb
— 6.6kb
— 4.4kb
— 2.3kb
— 2.0kb
— 0.6kb

Figure 2. Gene copy analysis of Arabidopsis FLS and LDOX. Landsberg (L) and Columbia (C) genomic DNA (5 µg) was digested with the indicated enzymes and hybridized with probes for either FLS (A and B) or LDOX (C). A, The filter was hybridized and washed at a temperature of 65°C. B, The filter shown in A was stripped and then rehybridized and washed at a temperature of 55°C. C, The filter was hybridized and washed at a temperature of 55°C.

the cloned genes were determined using the RI lines. DNA recombinant inbred from 28 RI lines was digested with *Cfo*I and blot analysis recombinant inbred was performed (data not shown). Based on analysis of segregation data, the Arabidopsis LDOX gene was determined to be on chromosome 4, less than 1 cM from *AG* and 10.7 cM from M600 (C. Lister, personal communication). Of the 14 *tt* loci for which a map position is known, none are found in this region of the genome (M. Koornneef, personal communication; Koornneef et al., 1982; Shirley et al., 1995), indicating that a mutation in this gene has not yet been identified in Arabidopsis. The FLS gene was determined to be on chromosome 5, 1.6 cM from marker g3837 and 2.4 cM from *tt4* (C. Lister, personal communication). Whereas three *tt* loci are located at the top of chromosome 5 (*tt4* and *tt7*, Shirley et al., 1995; *tt6*, J. Campanella and C. Town, personal communication), *tt4* has been previously shown to be the CHS locus and *tt7* is believed to be defective in flavonoid 3'-hydroxylase activity (Koornneef et al., 1982; Shirley et al., 1995). *tt6* plants have decreased levels of flavonols, but contain wild-type levels of other flavonoids, and could therefore have reduced FLS activity (Shirley et al., 1995).

As a first step toward determining whether *tt6* was defective in the cloned FLS gene, we performed blot analysis on total RNA isolated from 3-d-old wild-type and *tt6* seedlings grown in continuous white light. Based on the results of the DNA blot analysis, it is expected that under the conditions utilized in this analysis the FLS probe was gene-specific. However, neither the size nor abundance of FLS mRNA was affected in *tt6* relative to wild type, indicating that either *tt6* is not defective in the cloned FLS gene or that the mutation only affects the enzymatic activity of the protein (Fig. 3). To discriminate between these possibilities, the coding region of FLS was amplified from *tt6* by PCR. However, the sequence of the FLS coding region from *tt6* was identical to the sequence in the Landsberg wild type,

indicating that *tt6* is not defective in FLS activity encoded by this clone.

Expression Patterns of FLS and LDOX

Light induction of flavonoid mRNA in seedlings has been used to distinguish between "early" and "late" genes in Arabidopsis (Kubasek et al., 1992; Pelletier and Shirley, 1996). To determine the expression of the cloned FLS and LDOX genes relative to other flavonoid genes, blot analysis was performed on total RNA extracted from 3-d-old etiolated seedlings that were subsequently shifted to white light for 0, 1, 2.5, 5, or 8 h. Low-molecular-weight mRNA species were reproducibly observed in seedlings harvested 2.5 to 8 h after the shift to light, suggesting that FLS and LDOX transcripts synthesized in response to light induction are rapidly degraded. The highest level of FLS mRNA accumulated 2.5 h after the seedlings were moved to white light, whereas mRNA for LDOX accumulated to its highest level slightly later, at approximately 5 h (Fig. 4A). This experiment indicates that FLS expres-

La   *tt6*

FLS
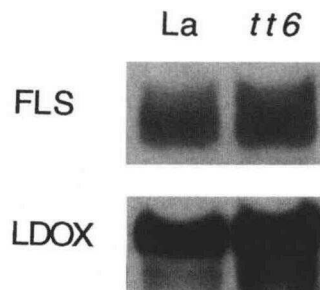
LDOX

Figure 3. Expression of FLS and LDOX in *tt6*. Total RNA (10 µg) from 3-d-old light-grown wild-type (Landsberg) or *tt6* seedlings was fractionated in formaldehyde gels and used for blot analysis. Filters were hybridized with the indicated probes and washed at 65°C.

A

FLS

LDOX

B

FLS

LDOX
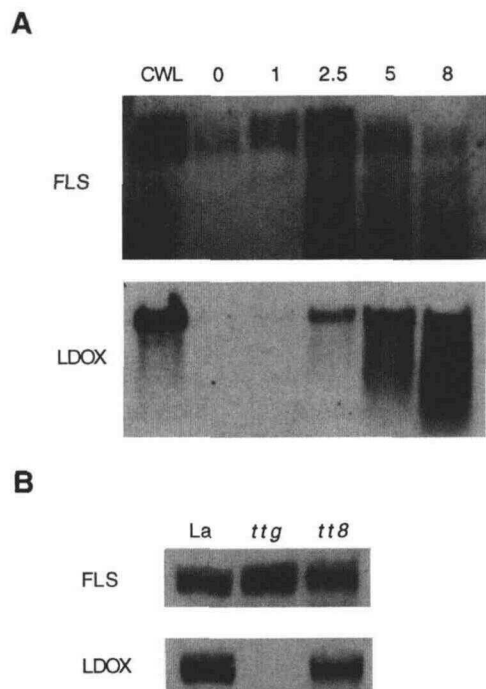
Figure 4. Expression analysis of "early" and "late" genes. A, Arabidopsis wild-type (Landsberg) seedlings were grown in continuous white light (CWL) or darkness for 3 d and then shifted to white light. Seedlings were harvested after 0, 1, 2.5, 5, or 8 h of light exposure. Total RNA (10 μg) was used for blot analysis and filters were hybridized with the indicated probe and washed at 65°C. B, Arabidopsis wild-type (Landsberg), *ttg*, and *tt8* seedlings were grown in continuous white light for 3 d. Total RNA (10 μg) was used for blot analysis and filters were hybridized with the indicated probes and washed at 65°C.

sion in Arabidopsis seedlings is coordinated with the "early" genes (CHS, CHI, and F3H), whereas the expression of LDOX is similar to that observed for the "late" gene (DFR).

Regulatory mutants that affect the expression of some but not all flavonoid genes have been isolated from barley, maize, petunia, snapdragon, and Arabidopsis, and provide further evidence for the differential regulation of the pathway (Dooner, 1983; Cone et al., 1986; Ludwig et al., 1989; Taylor and Briggs, 1990; Martin et al., 1991; Meldgaard, 1992; Quattrocchio et al., 1993; Deboo et al., 1995; Shirley et al., 1995; Pelletier and Shirley, 1996). As an independent means of determining the control of flavonoid gene expression in Arabidopsis seedlings, we performed blot analysis on RNA isolated from 3-d-old wild-type seedlings and from two putative regulatory mutants (*ttg* and *tt8*) grown in continuous white light. We have shown previously that CHS, CHI, and F3H gene expression is unaffected in *ttg* and *tt8*, whereas DFR mRNA is undetectable or reduced in these mutants (Shirley et al., 1995; Pelletier and Shirley, 1996). The results of this experiment were consistent with the light induction experiment, because FLS mRNA levels were unaffected in the *ttg* and *tt8* mutants while LDOX mRNA was undetectable in *ttg* (Fig. 4B), providing further evidence that this FLS is an "early" gene and LDOX is a "late" gene. The level of
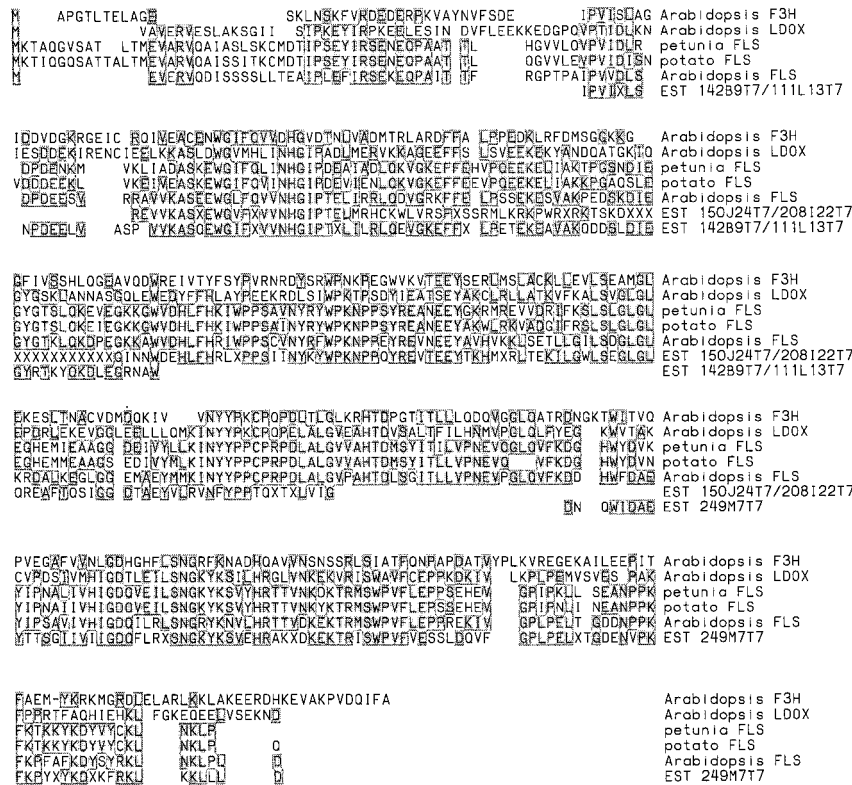
LDOX mRNA was not significantly reduced in *tt8*, in contrast to previous results for DFR (Fig. 4B) (Shirley et al., 1995; Pelletier and Shirley, 1996).

DISCUSSION

Isolation and Characterization of Arabidopsis FLS and LDOX Genes

By using a probe derived from an Arabidopsis EST clone (153O10T7) with high homology to petunia and potato FLS (Holton et al., 1993; GenBank accession no. X92178), we have isolated and characterized a corresponding genomic clone from Arabidopsis. To our knowledge, this is the first genomic clone encoding FLS that has been isolated to date. Sequence analysis revealed the presence of two introns and a deduced amino acid sequence with a high level of homology to both the petunia and potato deduced sequences (Table I). Furthermore, the expression pattern of this gene in seedlings, both in response to white light and in the *ttg* and *tt8* regulatory mutants, was identical to that observed for three other flavonoid genes (CHS, CHI, and F3H) in Arabidopsis (Pelletier and Shirley, 1996). We also identified and characterized an EST clone (YAY780) that appeared to contain a full-length coding region with high homology to LDOX genes from five other plant species (Menssen et al., 1990; Martin et al., 1991; Davies, 1993; Weiss et al., 1993; Sparvoli et al., 1994). The full sequence of the EST clone was determined, revealing a deduced amino acid sequence that was very similar to all other LDOX sequences in the database, with the maize A2 protein being the least similar (49% amino acid identity) and apple anthocyanidin hydroxylase the most similar (78% identity) (Table I).

The gene copy number of FLS and LDOX in Arabidopsis was investigated by DNA blot analysis. Low-stringency hybridization revealed that more than one sequence with homology to FLS is present in the Arabidopsis genome, whereas LDOX appears to be encoded by a single gene. It is possible that two or more genes encode different isoforms of FLS in Arabidopsis, particularly because five additional EST sequences with high sequence homology to FLS have appeared in the database since we began this project. Alignment of the DNA sequences with the previously sequenced FLS genes indicated that the five clones represent two to three genes with homology to FLS that have not yet been characterized. The amino acid sequences deduced from the partial DNA sequences reported in the EST database for 150J24T7/208I22T7, M249M7T7, and 142B9T7/111L13T7 are 53, 64, and 71% identical (based on unambiguous residues only), respectively, to the deduced amino acid sequence of the Arabidopsis FLS gene reported here (Fig. 5). These sequences are also very similar to the petunia and potato FLS proteins. Furthermore, amino acids that do not match the Arabidopsis sequence are often identical to those present in the petunia or potato FLS proteins (Fig. 5). In contrast, the partial amino acid sequences deduced from clones 150J24T7/208I22T7, 142B9T7/111L13T7, and M249M7T7 are only 28, 27, and 27% identical, respectively, to Arabidopsis F3H, and 34, 40, and 41% identical,

```
M   APGTLTELAGE           SKLNSKFVRDEDERPKVAYNVFSDE        IPVISLAG  Arabidopsis F3H
M       VAVERVESLAKSGII  STPKEYIRPKEELESIN DVFLEEKKEDGPOVPTIDLKN  Arabidopsis LDOX
MKTAQGVSAT  LTMEVARVQAIASLSKCMDTIIPSEYIRSENEQPAAT IIL  HGVVLOVPVIDLR  petunia FLS
MKTIQGOSATTALTMEVARVQAISSITKCMDTIIPSEYIRSENEQPAAT IIL  QGVVLEVPVIDISN  potato FLS
M           EVERVQDISSSSLLTEAIPLEFIRSEKEQPAIT IIF  RGPTPAIPVVDLS  Arabidopsis FLS
                                            IPVIXLS  EST 142B9T7/111L13T7

IDDVDGKRGEIC  ROIMEACBNWGIFCVVDHGVDTNUVADMTRLARDFFA LPPEDKLRFDMSGGKKG  Arabidopsis F3H
IESDDEKIRENCIEELKKASLDWGVMHLINHGIPADLMERVKKAGEEFFS LSVEEKEKYANDOATGKIIO  Arabidopsis LDOX
DPDENKM     VKLIADASKEWGIFQLINHGIPDEATADLOKVGKEFFEHVPGEEKELIAKTPFGSNDIE  petunia FLS
VDDDEEKL    VKEIMEASKEWGIFQVINHGIPDEVIIENLQKVGKEFFEEVPQEEKELIAKKPGAOSLE  potato FLS
DPDEESV     RRAVVKASEEWGLFQVVNHGIPTELIRRLQDVGRKFFE LRSSEKESVAKPEDSKDIE  Arabidopsis FLS
            REVVKASXEWGVFIXVVNHGIPTELMRHCKWLVRSFIXSSRMLKRKPWRXRKTSKDXXX  EST 150J24T7/208I22T7
NPDEELM     ASP MVYKASOEWGIFIXVVNHGIPTIXLILRLQEVGKEFFIX LBETEKBAVAKODDSLDIE  EST 142B9T7/111L13T7

QFIVESHLOGEAVQDWREIVTYFSYPVRNRDYSRWPNKPEGWVKVTEEVSERLMSLACKLLEVLSEAMGL  Arabidopsis F3H
GYGSKLANNASGOLEWEDYFFRLAYPEEKRDLSIWPKTPSDYIEATSEYAKCLRLLATKVFKALSVGLGL  Arabidopsis LDOX
GYGTSLOKEVEGKKGWVDHLFHKIWPPSAVNYRYWPKNPPSYREANEEYGKRMREVVDRIFKSLSLGLGL  petunia FLS
GYGTSLOKEIEGKKGWVDHLFHKIWPPSAINYRYWPKNPPSYREANEEYAKWLRKVADGIFRSLSLGLGL  potato FLS
GYGTKLQKDPEGKKAWVDHLFHRIWPPSCVNYRFWPKNPREYREVNEEYAVHVKKLSETLLGILSDGLGL  Arabidopsis FLS
XXXXXXXXXXQINNWDEHLFHRLXPPSIINYKVWPKNPROYREVTEEVTKHMXRLTEKILGWLSEGLGL  EST 150J24T7/208I22T7
GYRTKYOKDLEGRNAW                                                        EST 142B9T7/111L13T7

EKESLTNACVDMOOKIV   VNYYRKCROPDDTLGLKRHTDPGTITLLLODQVGGLOATRDNGKTWITVQ  Arabidopsis F3H
EPDRLEKEVGGLEBLLLOMKINYYPKCROPELALGVEAHTDVSALTFILHNMVPGLQLRYEG KMVTIAK  Arabidopsis LDOX
EQHEMIEAAGG DEIVYLLKINYYPPCPRPDLALGVVAHTDMSYITILVPNEVOGLQVFKDG HWYQVK  petunia FLS
EQHEMMEAAGS EDIVYMLKINYYPPCPRPDLALGVVAHTDMSYITLLVPNEVQ  VFKDG HWYDVN  potato FLS
KRDALKEGLGG EMAEYMMKINYYPPCPRPDLALGVPAHTDLSGITLLVPNEVPGLOVFKDD HWFDAE  Arabidopsis FLS
OREAFTOSIGG DTAEYVLRVNFYPPTQXTXLVIG                                    EST 150J24T7/208I22T7
                                         DN QWIDAE  EST 249M7T7

PVEGAFVVNLGDHGHFLSNGRFKNADHQAVWNSNSSRLSIATPONPAPDATMYPLKVREGEKAILEEPIT  Arabidopsis F3H
CVPDSDVMHTGDTLETLSNGKYKSILHRGLVNKEKVRISWAVFICEPPKDKIV  LKPLPEMVSVES PAK  Arabidopsis LDOX
YIPNALIVHIGDQVEILSNGKYKSVYHRTTVNKDKTRMSWPVFLEPPSEHEM  GRIPKLL SEANPPK  petunia FLS
YIPNAIIVHIGDQVEILSNGKYKSVYHRTTVNKYKTRMSWPVFLEPSSEHEM  GRIPNLI NEANPPK  potato FLS
YIPSAVVIVHIGDQILRLSNGRYKNVLHRTTVDKEKTRMSWPVFLEPPREKIV  GPLPELT QDDNPPK  Arabidopsis FLS
MTTSGIIVIIGDQFLRXSNGRYKSVEHRAKXDKEKTRISWPVFVESSLDQVF  GPLPELXTGDENVPK  EST 249M7T7

PAEM-YKRKMGRDDELARLKKLAKEERDHKEVAKPVDQIFA                          Arabidopsis F3H
PPBRTFAQHIEHKL FGKEQEEDVSEKND                                      Arabidopsis LDOX
FKTKKYKDYYVCKL    NKLP                                             petunia FLS
FKTKKYKDYYVCKL    NKLP    Q                                        potato FLS
FKPFAFKDYSYRKL    NKLPQ   D                                        Arabidopsis FLS
FKPYXYKQXKFRKL    KKLLL   D                                        EST 249M7T7
```

Figure 5. Evidence for a gene family encoding FLS in Arabidopsis. The deduced amino acid sequences of FLS from petunia, potato, Arabidopsis, and three Arabidopsis EST clones with homology to FLS were aligned (MegAlign, DNASTAR, Madison, WI). Residues that are identical in the majority of the sequences are shaded.

respectively, to Arabidopsis LDOX. This strongly suggests that at least two other genes encode FLS in Arabidopsis. We will pursue the characterization of these EST clones in future studies.

Although the physiological significance of multiple FLS genes is unknown, this result is surprising, because it provides the first evidence for a multigene family encoding a flavonoid enzyme in Arabidopsis (Feinbaum and Ausubel, 1988; Shirley et al., 1992; Pelletier and Shirley, 1996). It is worth noting that in Arabidopsis, different flavonols are predominant in different tissues, with quercetin being the primary seed flavonol and kaempferol the main flavonol found in whole flowers (Shirley et al., 1995), whereas similar levels of these two flavonols are found in stamens (Burbulis et al., 1996). Spribille and Forkmann (1984) have shown that dihydrokaempferol is the preferred substrate for FLS in *Matthiola incana* flowers. It is therefore possible that FLS isozymes with different substrate specificities control the amount and types of flavonols present in a given tissue.

The positions of both FLS and LDOX on the recombinant inbred map were determined, and it was found that FLS was present in the same region of chromosome 5 as the *TT6* locus, whereas LDOX was located on chromosome 4. None of the previously characterized *tt* loci are located in this region of chromosome 4, and therefore an Arabidopsis LDOX mutant remains to be identified (Shirley et al., 1995). *TT6* has previously been identified as a possible FLS locus, because mutations at *TT6* cause a reduction in flavonols but not other flavonoids (Shirley et al., 1995). However, RNA blot analysis of *tt6* revealed no alterations in the abundance or size of FLS mRNA in this mutant compared

with wild type. Furthermore, the sequence of the FLS coding region in *tt6* was identical to the wild-type sequence. Therefore, *tt6* is not defective in the FLS activity encoded by the cloned gene. It remains possible that *tt6* is the structural or regulatory locus for a different FLS gene.

Expression of FLS and LDOX in Arabidopsis Seedlings

We have previously shown that CHS, CHI, and F3H mRNA expression is coordinate and precedes DFR mRNA expression in Arabidopsis seedlings in response to white light (Pelletier and Shirley, 1996). In addition, it was demonstrated that TTG is required for the accumulation of DFR mRNA, but not for CHS, CHI, or F3H mRNA in seedlings. This type of differential expression of "early" and "late" flavonoid genes has been demonstrated in a number of plant species, although the individual genes that compose these groups differ (Martin et al., 1991; Quattrocchio et al., 1993; Shirley et al., 1995; Pelletier and Shirley, 1996). In the current study, we have characterized the expression patterns of Arabidopsis FLS and LDOX genes in response to white light and in the regulatory mutants *ttg* and *tt8*. Both of these experiments indicate that LDOX is a "late" gene and FLS an "early" gene in Arabidopsis, and that TTG is required for expression of DFR and LDOX. Whereas DFR mRNA levels are significantly reduced in *tt8* mutants, the level of LDOX mRNA was equivalent to wild type. This result may indicate that TTG is required for the expression of both DFR and LDOX, whereas TT8 is a specific regulator of DFR. The classification of LDOX as a "late" gene was expected, based on its position within the biosynthetic

pathway (Fig. 1) and the previous characterization of DFR. In contrast, to our knowledge, no studies have been performed on the expression pattern of FLS relative to other flavonoid genes, and its classification as an "early" gene is a significant finding. However, it remains possible that different FLS genes may be regulated independently of one another in a developmental- or tissue-specific manner, because it appears that more than one gene may encode FLS in Arabidopsis. Because flavonols are the active agents in a wide range of physiological functions attributed to flavonoids (Shirley, 1996), an understanding of the regulation of this enzyme's activity will be of particular importance.

Much is now known about the regulation of flavonoid genes in Arabidopsis seedlings at the level of mRNA expression, because genes encoding six of the biosynthetic enzymes have been cloned and characterized. Despite this, much remains to be learned about the regulation of the flavonoid pathway at the level of protein expression, as antibodies have been developed only against buckwheat, parsley, rye, and maize CHS, petunia and bean CHI, and petunia F3H (Schroder and Schafer, 1980; Robbins and Dixon, 1984; Hrazdina et al., 1986; Van Tunen and Mol, 1987; Peters et al., 1988; Britsch, 1990a; Pollak et al., 1993). Our laboratory has developed antibodies against Arabidopsis CHS, CHI, F3H, FLS, and LDOX that will allow us to explore the regulation of the pathway at the protein level (M.K. Pelletier, C.C. Cain, and B.W. Shirley, unpublished data). Furthermore, whereas 14 tt mutants with a yellow or pale brown seed coat, some of which disrupt flavonoid biosynthesis in vegetative tissues, have been identified, the molecular defect is unknown for 8 of these loci (Shirley et al., 1995). What causes the lack of seed-coat pigmentation in these 8 tt mutants? Does the differential regulation of the pathway observed at the level of mRNA accumulation reflect the regulation of the pathway at the protein level? It is possible that some of these loci encode regulatory proteins. Alternatively, the tt phenotype may be caused by metabolic "crosstalk" resulting from mutations in other pathways. For example, trp2-1 plants, which carry a mutation in a tryptophan synthase β-subunit gene, also produce pale brown seeds (Last et al., 1991). It is clear that much work remains to be done. However, the large collection of cloned genes and mutants now available for the study of the phenylpropanoid pathway in Arabidopsis, together with the antibodies that are being developed, will provide a unique combination of tools for investigating the diverse mechanisms that regulate plant metabolism.

ACKNOWLEDGMENTS

LITERATURE CITED

Britsch L (1990a) Purification of flavanone 3 beta-hydroxylase from *Petunia hybrida*: antibody preparation and characterization of a chemogenetically defined mutant. Arch Biochem Biophys **276**: 348–354

Britsch L (1990b) Purification and characterization of flavone synthase I, a 2-oxoglutarate-dependent desaturase. Arch Biochem Biophys **282**: 152–160

Britsch L, Dedio J, Saedler H, Forkmann G (1993) Molecular characterization of flavanone 3 beta-hydroxylases: consensus sequence, comparison with related enzymes and the role of conserved histidine residues. Eur J Biochem **217**: 745–754

Britsch L, Heller W, Grisebach H (1981) Conversion of flavanone to flavone, dihydroflavonol, and flavonol with an enzyme system from cell cultures of parsley. Z Naturforsch **36**: 742–750

Burbulis IE, Iacobucci M, Shirley BW (1996) A null mutation in the first enzyme of flavonoid biosynthesis does not affect male fertility in Arabidopsis. Plant Cell **8**: 1013–1025

Church GM, Gilbert W (1984) Genomic sequencing. Proc Natl Acad Sci USA **81**: 1991–1995

Cone KC, Burr FA, Burr B (1986) Molecular analysis of the maize anthocyanin regulatory locus *C1*. Proc Natl Acad Sci USA **83**: 9631–9635

Davies KM (1993) A *Malus* cDNA with homology to the *Antirrhinum candica* and *Zea A2* genes. Plant Physiol **103**: 1015

Deboo G, Albertsen M, Taylor L (1995) Flavanone 3-hydroxylase transcripts and flavonol accumulation are temporally coordinate in maize anthers. Plant J **7**: 703–713

Dooner HK (1983) Coordinate genetic regulation of flavonoid biosynthetic enzymes in maize. Mol Gen Genet **189**: 136–141

Dooner HK, Robbins TP, Jorgensen RA (1991) Genetic and developmental control of anthocyanin biosynthesis. Annu Rev Genet **25**: 173–199

Feinbaum RL, Ausubel FM (1988) Transcriptional regulation of the *Arabidopsis thaliana* chalcone synthase gene. Mol Cell Biol **8**: 1985–1992

Forkmann G, de Vlaming P, Spribille R, Wiering H, Schram AW (1986) Genetic and biochemical studies on the conversion of dihydroflavonols to flavonols in flowers of *Petunia hybrida*. Z Naturforsch **41**: 179–186

Gerard GF, Miller K (1986) Comparison of glyoxal and formaldehyde gels for sizing rRNAs. Focus **8**: 5–6

Grossberger D (1987) Minipreps of DNA from bacteriophage lambda. Nucleic Acids Res **15**: 6737

Holton TA, Brugliera F, Tanaka Y (1993) Cloning and expression of flavonol synthase from *Petunia hybrida*. Plant J **4**: 1003–1010

Hrazdina G, Lifson E, Weeden NF (1986) Isolation and characterization of buckwheat (*Fagopyrum esculentum* M.) chalcone synthase and its polyclonal antibodies. Arch Biochem Biophys **247**: 414–419

Ish-Horowicz D, Burke JF (1981) Rapid and efficient cosmid cloning. Nucleic Acids Res **9**: 2989–2998

Jackson D, Roberts K, Martin C (1992) Temporal and spatial control of expression of anthocyanin biosynthetic genes in developing flowers of *Antirrhinum majus*. Plant J **2**: 425–434

Koes RE, Quattrocchio F, Mol JNM (1993) The flavonoid biosynthetic pathway in plants: function and evolution. BioEssays **16**: 123–132

Koornneef M, Luiten W, de Vlaming P, Schram AW (1982) A gene controlling flavonoid-3'-hydroxylation in Arabidopsis. Arabidopsis Inf Serv **19**: 113–115

Kubasek WL, Shirley BW, McKillop A, Goodman HM, Briggs W, Ausubel FM (1992) Regulation of flavonoid biosynthetic genes in germinating *Arabidopsis* seedlings. Plant Cell **4**: 1229–1236

Lander ES, Green P, Abrahamson J, Barlow A, Day MJ, Lincoln SE, Newberg L (1987) Mapmaker: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics **1**: 174–181

Last RL, Bissinger PH, Mahoney DJ, Radwanski ER, Fink GR (1991) Tryptophan mutants in *Arabidopsis*: the consequences of duplicated tryptophan synthase β genes. Plant Cell **3**: 345–358

Lister C, Dean C (1993) Recombinant inbred lines for mapping RFLP and phenotypic markers in *Arabidopsis thaliana*. Plant J 4: 745–750

Ludwig SR, Habera LF, Dellaporta SL, Wessler SR (1989) *Lc*, a member of the maize *R* gene family responsible for tissue-specific anthocyanin production, encodes a protein similar to transcriptional activators and contains the *myc*-homology region. Proc Natl Acad Sci USA 86: 7092–7096

Martin C, Prescott A, Mackay S, Bartlett J, Vrijlandt E (1991) Control of anthocyanin biosynthesis in flowers of *Antirrhinum majus*. Plant J 1: 37–49

Meldgaard M (1992) Expression of chalcone synthase, dihydroflavonol reductase, and flavanone-3-hydroxylase in mutants of barley deficient in anthocyanin and proanthocyanidin biosynthesis. Theor Appl Genet 83: 695–706

Menssen A, Hohmann S, Martin W, Schnable SS, Peterson PA, Saedler H, Gierl A (1990) The En/Spm transposable element of *Zea mays* contains splice sites at the termini generating a novel intron from a dSpm element in the *A2* gene. EMBO J 9: 3051–3057

Mo Y, Nagel C, Taylor LP (1992) Biochemical complementation of chalcone synthase mutants defines a role for flavonols in functional pollen. Proc Natl Acad Sci USA 89: 7213–7217

Pelletier MK, Shirley BW (1995) A genomic clone encoding flavanone 3-hydroxylase (accession No. U33932) from *Arabidopsis thaliana*. Plant Physiol 109: 1125

Pelletier MK, Shirley BW (1996) Analysis of flavanone 3-hydroxylase in Arabidopsis seedlings. Coordinate regulation with chalcone synthase and chalcone isomerase. Plant Physiol 111: 339–345

Peters A, Schneider-Poetsch HAW, Schwarz H, Weissenböck G (1988) Biochemical and immunological characterization of chalcone synthase from rye leaves. J Plant Physiol 133: 178–182

Pollak PE, Vogt T, Mo Y, Taylor LP (1993) Chalcone synthase and flavonol accumulation in stigmas and anthers of *Petunia hybrida*. Plant Physiol 102: 925–932

Prescott AG, John P (1996) Dioxygenases: molecular structure and role in plant metabolism. Annu Rev Plant Physiol Plant Mol Biol 47: 245–271

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

Reddy GM, Coe EH (1962) Intertissue complementation: a simple technique for direct analysis of gene-action sequence. Science 138: 149

Robbins MP, Dixon RA (1984) Induction of chalcone isomerase in elicitor-treated bean cells. Comparison of rates of synthesis and appearance of immunodetectable enzyme. Eur J Biochem 145: 195–202

Schröder J, Schafer E (1980) Radioiodinated antibodies, a tool in studies on the presence and role of inactive enzyme forms: regulation of chalcone synthase in parsley cell suspension cultures. Arch Biochem Biophys 203: 800–808

Shirley BW (1996) Flavonoid biosynthesis: "new" functions for an "old" pathway. Trends Plant Sci 1: 377–382

Shirley BW, Hanley S, Goodman H (1992) Effects of ionizing radiation on a plant genome: analysis of two Arabidopsis *transparent testa* mutations. Plant Cell 4: 333–347

Shirley BW, Hwang I (1995) The interaction trap: in vivo analysis of protein-protein associations. Methods Cell Biol 49: 401–416

Shirley BW, Kubasek WL, Storz G, Bruggemann E, Koornneef M, Ausubel FM, Goodman HM (1995) Analysis of *Arabidopsis* mutants deficient in flavonoid biosynthesis. Plant J 8: 1229–1236

Simon AE, Li XH, Lew JE, Stange R, Zhang C, Polacco M, Carpenter CD (1992) Susceptibility and resistance of *Arabidopsis thaliana* to turnip crinkle virus. Mol Plant-Microbe Interact 5: 496–503

Sparvoli F, Martin C, Scienza A, Gavazzi G, Tonelli C (1994) Cloning and molecular analysis of structural genes involved in flavonoid and stilbene biosynthesis in grape (*Vitis vinifera* L.). Plant Mol Biol 24: 743–755

Spribille R, Forkmann G (1984) Conversion of dihydroflavonols to flavonols with enzyme extracts from flower buds of *Matthiola incana* R. Br. Z. Naturforsch 39c: 714–719

Taylor L, Briggs WR (1990) Genetic regulation and photocontrol of anthocyanin accumulation in maize seedlings. Plant Cell 2: 115–127

van der Meer IM, Stam ME, van Tunen AJ, Mol JNM, Stuitje AR (1992) Antisense inhibition of flavonoid biosynthesis in petunia anthers results in male sterility. Plant Cell 4: 253–262

Van Tunen AJ, Mol JN (1987) A novel purification procedure for chalcone flavanone isomerase from *Petunia hybrida* and the use of its antibodies to characterize the Po mutation. Arch Biochem Biophys 257: 85–91

Voytas DF, Konieczny A, Cummings MP, Ausubel FA (1990) The structure, distribution and evolution of the Tal retrotransposable element family of *Arabidopsis thaliana*. Genetics 126: 713–721

Weiss D, van der Luit AH, Kroon JTM, Mol JN, Kooter JM (1993) The petunia homologue of the *Antirrhinum majus candi* and *Zea mays A2* flavonoid genes; homology to flavanone 3-hydroxylase and ethylene-forming enzyme. Plant Mol Biol 22: 893–897