Molecular characterization of myb-homologous transcriptional factors of the flavonoid pathway in Zea mays

Peifen Zhang

Iowa State University
INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.
Molecular characterization of myb-homologous transcriptional factors of the flavonoid pathway in *Zea mays* 

by 

Peifen Zhang 

A dissertation submitted to the graduate faculty 
in partial fulfillment of the requirements for the degree of 

DOCTOR OF PHILOSOPHY 

Major: Genetics 

Major Professor: Thomas A. Peterson 

Iowa State University 

Ames, Iowa 

1999
Graduate College

Iowa State University

This is to certify that the doctoral dissertation of

Peifen Zhang

has met the dissertation requirements of Iowa State University

Signature was redacted for privacy.

\textbackslash Major Professor

Signature was redacted for privacy.

For the Major Program

Signature was redacted for privacy.

For the Graduate College
# TABLE OF CONTENTS

## CHAPTER 1. GENERAL INTRODUCTION

- Biological functions of flavonoids: 1
- The flavonoid biosynthetic pathway: 5
- Regulation of flavonoid biosynthesis: 8
- Tissue-specific anthocyanin and phlobaphene pigmentation patterns and the molecular mechanisms of tissue-specific pigmentation: 13
- Duplicated genes: 19
- Evolution of the flavonoid pathway: 21
- Dissertation organization: 23
- References: 24

## CHAPTER 2. GENE DUPLICATION AND RETROELEMENT INSERTION GENERATED LINKED MYB-HOMOLOGOUS GENES WITH DISTINCT EXPRESSION PATTERNS IN MAIZE

- Abstract: 32
- Introduction: 33
- Materials and Methods: 35
- Results: 38
- Discussion: 45
- Acknowledgments: 54
- References: 55

## CHAPTER 3. FUNCTIONAL ANALYSIS OF THE MAIZE P2 GENE: EVIDENCE FROM GENETIC MUTANTS AND TRANSFORMATION STUDIES

- Abstract: 68
- Introduction: 69
- Results: 70
- Discussion: 75
- Materials and Methods: 77
- Acknowledgments: 80
- References: 80

## CHAPTER 4. CELLULAR DISTRIBUTION OF PHLOBAPHENE PIGMENT AND TRANSCRIPTS OF THE REGULATORY GENE P1 IN MAIZE KERNEL

- Introduction: 89
- Results and Discussion: 92
- Materials and Methods: 98
- Acknowledgments: 103
- References: 103

## CHAPTER 5. GENERAL SUMMARY

Acknowledgments: 111

References: 115
CHAPTER 1. GENERAL INTRODUCTION

Biological functions of flavonoids

Flavonoids are a class of phenolic compounds produced in secondary metabolism: they include flavonols, flavanones, proanthocyanidins, anthocyanins, and isoflavonoids. These flavonoids can be further modified by hydroxylation, methylation, acylation, or glucosylation, and thereby produce the enormous diversity of flavonoids observed in nature (for a review, see Heller and Forkmann, 1994). Flavonoids are widely distributed in the plant kingdom, from mosses and ferns, to flowering plants. Man’s original attention to flavonoids was perhaps attracted by the bright color of flowers (for a review, see Mol et al. 1998). The majority of flower pigments are red and purple anthocyanins, or yellow chalcones and aurones. However, new colors can occur in mutants, for example, mutation of the maize *al* gene (which encodes dihydroflavonol reductase in the anthocyanin pathway) results in brown pigment (O’Reilly et al. 1985). Many other factors can also affect flower color (Mol et al. 1998; Brouillard and Dangle 1994). In petunia, mutations that increase the anthocyanin vacuole pH produce blue flower color. In snapdragon, the shape of cells that accumulate anthocyanin influences flower color by influencing light reflection. In wild type, petals are bright red, and the epidermal cells are a conical shape, whereas in the *mixta* mutant, petals are pink, and the epidermal cells are a flattened shape. Other flavonoids, such as flavonols and flavanones are themselves colorless, but they can alter the flower color to blue through the formation of complexes with anthocyanin or metal ions. Interestingly, the yellow color of pollen in almond is caused by glycosylated flavonol. Phlobaphene is another flavonoid pigment that is synthesized in a
parallel branch of the anthocyanin pathway (Styles and Ceska, 1977). The red phlobaphene pigment accumulates mainly in maize floral organs, such as kernel pericarp and cob glume. Red phlobaphene pigment has also been observed in sorghum (Zanta et al. 1994). Flower colors are thought to contribute to pollination by attracting insects or birds. In field experiments, removal of flower petals results in dramatic decrease of pollination (Waser et al. 1983 a and b). In plants that are wind pollinated, such as maize, the function of pigmentation is unclear. In addition to the bright pigments contributed by flavonoids, flavonoids also play a number of important biological functions to plants (for a review, see Koes et al. 1993):

Flavonoids protect plants from UV damage

Sunlight is the energy resource for plant growth. But the UV component of sunlight is deleterious because it can damage DNA and other macromolecules involved in essential physiological processes (Stapleton 1992). Flavonoids are strong-UV-absorbing compounds, and plants may protect themselves from UV damage by accumulating flavonoids in epidermal cells (Schmelzer et al. 1988; Tevini et al. 1991). It has been reported that expression of genes in flavonoid biosynthesis is induced by UV light (Kubasek et al. 1992). In arabidopsis, mutations in the CHS and CHI genes essential for flavonol accumulation result in mutant plants which are growth-retarded when placed under UV light (Li et al. 1993). This additionally demonstrates the role of flavonoids in the protection of plants against UV light.
Flavonoids are also important to plant reproduction

It was shown in maize and petunia that flavonol is essential to pollen development and pollen function. In petunia, plants with mutations in the gene encoding CHS (chalcone synthase) are self-sterile. Pollen of the mutant fails in pollen tube growth as assayed on pollen germination medium. The deficient pollen is, however, functional on wild type plant stigmas. This suggests that a component supplied from the wild type stigma can rescue the mutant pollen function. By comparison of compounds extracted from wild type and mutant stigmas, flavonol was identified as the compound required for pollen tube growth and pollen function (Mo et al. 1992). In maize, plants which carry mutations in both of the CHS (chalcone synthase) genes, c2 and whp, are self-sterile (Coe et al. 1981).

Flavonoids are involved in plant pathogen and insects resistance

Plants are susceptible to bacterial and fungal infections. In most cases of the infection, plants undergo a hypersensitive response, in that cells near the infection sites quickly die to inhibit the spread of the infection to the surrounding non-infected cells. This suicide defense can be observed on leaves as small lesions. In legumes, activation of flavonoid biosynthesis genes and accumulation of isoflavonoids at the infection site are closely correlated to the hypersensitive response (Dixon 1986; Lamb et al. 1989).

Many insects are plant-feeding pests. Flavonoids are generally not very toxic to most insects. But some flavonoids can act as feeding deterrents to insects. Among them, tannins were extensively shown to be active in deterring insect larvae and aphids from feeding on leaves and petioles. This deterrence by tannins may involve various mechanisms. In soluble
form, tannins may deter insects because of their taste, whereas in insoluble form, tannins contribute to lignification that causes hardening of plant tissues. In addition, tannins can also hinder the ability of insects to digest their food by binding to proteins. Finally, tannins may inhibit insect growth and development by interfering with some steps of metabolism (Harborne and Grayer, 1994). In maize a glycosyl-flavone has been shown to be active in deterring feeding by corn earworm (*helicoverpa zea*). Corn earworm is a major silk- and kernel-feeding insect pest of maize, cotton, soybean, tomato, and other plants. In maize, eggs of corn earworm are laid on fresh silks. After hatching, the larvae will feed on silks, and may migrate to the ear where they feed on developing kernels. The host maize plant resistance to corn earworm relies on both antibiosis and morphological features of ear. Glycosol flavone (maysin isolated from silk is a C-glycosyl flavone) can inhibit corn earworm early development. A significant correlation was found between maysin concentration in silk and reduced corn earworm larvae growth (Byrne et al. 1996a; Byrne et al. 1996b).

Flavonoids contribute to the structure of plant

Proanthocyanidins (tannins) derived by polymerization of leuco-anthocyanidin are important in plant tissue reinforcement. The proanthocyanidins are found in large amounts in the seed coat of many plant species, and in the heartwood and bark of trees. It was shown that mutant snapbean seed lacking proanthocyanidins are more sensitive to mechanical and water stress than wild type seeds (Moore 1972).
Potential health benefits of flavonoids

Flavonoids are abundant in fruits, vegetables and medicinal plants, and they most likely contribute to the antioxidant activity of these foods and medicinal plants. Certain kinds of fruit juice were demonstrated to be able to lower blood pressure or inhibit cancer cell proliferation. The flavonoids contained in the fruit juices were suggested to be the bioactive compounds (Carroll et al. 1998; Samman et al. 1998).

The flavonoid biosynthetic pathway

The easily scored bright pigments produced in the flavonoid biosynthetic pathway are not essential to the plant, thus making the flavonoid pathway an ideal system for genetic and biochemical study. Tremendous knowledge about the pathway, the genes involved, and their regulation has been accumulated in a variety of plant species, most notably maize, snapdragon, and petunia.

The flavonoid pathway is summarized in Figure 1 (for a comprehensive review, see Heller and Forkmann, 1994; Styles and Ceska, 1989). Flavonoid synthesis starts with the simple organic compound phenylalanine. Phenylalanine is converted to 4-coumaroyl-CoA. The condensation of 4-coumaroyl-CoA with malonyl-CoA by chalcone synthase (CHS) produces a yellow chalcone. The isomerization of chalcone to a colorless flavanone can occur either spontaneously at a low rate, or catalyzed by the enzyme chalcone-flavanone isomerase (CHI). Flavanone is hydroxylated by flavanone 3-hydroxylase (F3H) to give a colorless dihydroflavonol. Dihydroflavonol is then reduced to leucoanthocyanidin by the enzyme dihydroflavonol 4-reductase (DFR). Leucoanthocyanidin is converted to a colored
Figure 1. The flavonoid biosynthesis pathway. In maize, regulatory genes R/B and C1/P1 regulate biosynthesis steps in box A and B, whereas P1 regulates steps in box C (after Byrne et al. 1996 b).
anthocyanidin by anthocyanidin synthase (AS). The glycosylation of anthocyanidin to anthocyanin is catalyzed by the enzyme UDPglucose flavonoid 3-oxy-glucosyltransferase (UF3GT). The last step to transfer anthocyanin to the vacuole is through glutathione S-transferase. Phlobaphene synthesis has two initial steps in common with anthocyanin synthesis. The reduction of flavanone by DFR to flavan-4-ol is the first divergent step of the two pathways. Phlobaphene is produced by a non-enzymatic polymerization of flavan-4-ol. Flavonol, another important flavonoid compound, is produced from dihydroflavonol by flavonol synthase (FLS).

Isolation of the genes encoding enzymes for flavonoid synthesis was achieved by means of genetic and molecular strategies. Anthocyanins are produced differentially in different plant lines. Thus, comparison of mRNA from tissues where anthocyanins are accumulated to mRNA from the same tissue of a different genotype where anthocyanins are not accumulated can reveal genes involved in anthocyanin synthesis. For example, a cDNA library prepared from the anthocyanin-accumulating genotype can be screened using probes made from RNA of a non-anthocyanin-accumulating genotype. Those non-hybridizing clones represent genes putatively involved in anthocyanin synthesis. An alternative strategy is using transposon tagging, in which tagged genes can be isolated using the transposon as a probe. Once a gene is isolated from a species, homologous genes of another species can be isolated based on their conserved sequence homology. This approach will also enable the isolation of homologous genes from the same species. By these approaches, flavonoid biosynthetic structural genes were isolated from snapdragon, petunia, maize, and other plant species. The
structural genes isolated from maize are listed in Figure 1 (for review, see Dooner et al. 1991: Holton and Cornish 1995).

**Regulation of flavonoid biosynthesis**

Structural genes encoding enzymes that catalyze reactions in the flavonoid pathway are regulated at the transcriptional level (for review, see Holton and Cornish 1995). Transcriptional regulatory genes can be recognized by their effects on structural gene activity using either enzyme or mRNA assays. On Northern blots, for example, the presence and absence of functional alleles of a regulatory gene will correlate with the presence and absence of mRNA from one or more structural genes, respectively (Grotewold et al. 1991). In general, more than one structural gene is regulated by a regulatory gene. Transcriptional regulation may also be characterized by the binding of the protein encoded by the regulatory gene to the promoter of a structural gene. The DNA-binding activity of a regulatory gene product can be detected by electromobility shift assay. In this approach, a preparation of protein encoded by a regulatory gene is incubated with the promoter fragment of a structural gene, and the complexes are analyzed by gel electrophoresis. The binding of a regulatory protein on a promoter DNA fragment is indicated as a retarded band on the gel (Ausubel et al. 1989). The DNA-binding characteristics and transcriptional activation activity of a regulatory gene product can be further evaluated by a transgenic study. In this approach, a reporter gene such as GUS is fused to the promoter of a structural gene, and co-transformed with a regulatory gene. The expression and function of the reporter gene indicates the role of the co-transformed regulatory gene (McCarty et al. 1991).
Regulation of anthocyanin biosynthesis

Regulation of anthocyanin biosynthesis is best understood in maize, where two groups of regulatory genes have been well characterized. The R/B gene family encodes myc-homologous regulatory proteins (Ludwig and Wessler 1990). The protein N-terminal region has the acidic characteristics of a transcriptional activator. The C-terminal region contains a helix-loop-helix DNA-binding domain of the myc family of proto-oncogenes. The Cl/Pl gene family encodes myb-like regulatory proteins (Cone et al. 1993). The protein N-terminal domain resembles the myb class of proto-oncogenes that have DNA binding capacity, whereas the C-terminal region contains an acidic domain characteristic of transcriptional activation domains. Transcriptional activity of anthocyanin synthetic structural genes requires the combined function of one member of each of the R/B and Cl/Pl families (Goff et al. 1990; Ludwig et al. 1990; Roth et al. 1991). In a transient transgenic assay, transformation of the B or the Cl gene alone could not provide DNA binding or transcriptional activation of the structural genes, whereas co-transformation of the B and Cl genes together results in transcriptional activation. The R/B and Cl/Pl have been shown to regulate all the known steps in the anthocyanin pathway, from CHS to the last step of transport of anthocyanin into vacuoles.

Regulatory genes have also been isolated from snapdragon and petunia. In snapdragon, the Delila gene was isolated by transposon tagging, and was found to be homologous to the maize R gene family (Goodrich, J. et al. 1992; Martin, C. et al. 1991). Delila was shown to regulate only the later steps of the anthocyanin pathway, including genes encoding F3H, DFR, and UF3GT. As far as is known, the Delila gene seems to act alone, and
does not require a counterpart, such as $CI/Pl$ in maize. In petunia, $an1$, $an2$, $an10$, and $an11$ regulate a subset of structural genes, encoding enzymes for DFR and subsequent steps. The $an2$ mutant can be complemented by the maize $Lc$ gene, a member of $R$ family, suggesting that $an2$ is a $R$ homolog. The $an11$ mutant can be complemented by the maize $Cl$ gene, suggesting that $an11$ is a $Cl$ homolog. Co-transformation of $Lc$ and $Cl$ gave a stronger complementation of $an2$ or $an11$ than transformation of either gene alone (Quattrocchio et al. 1993). This suggests that, like in maize, anthocyanin synthesis in petunia may be enhanced by the combined function of members of two regulatory gene families. The complementation of petunia $an$ mutants by maize anthocyanin regulatory genes, or vice versa (Quattrocchio et al. 1998), suggest that the functions of anthocyanin regulatory genes are broadly conserved in plant species. Interestingly, the petunia $an1$, $an2$, and $an11$ genes have a differential regulatory role on the $chs$ genes: they regulate $chsJ$, but not $chsA$. The maize $Lc$ and $Cl$ genes expressed in petunia can also act on the $chsJ$ gene, but not $chsA$ (Quattrocchio et al. 1993). Moreover, the petunia $an2$ gene expressed in maize can activate the $c2$ gene (Quattrocchio et al. 1998). These results suggest that, in addition to the conserved function of these regulatory genes, the differential regulation of structural genes in different plant species is also conserved.

**Regulation of phlobaphene pathway in maize**

Regulation of maize phlobaphene synthesis does not appear to be as complex as that of anthocyanin synthesis. The three steps required, CHS, CHI, and DFR, are shown to be under $Pl$ regulation (Grotewold et al. 1991, and 1994). The protein encoded by the maize $Pl$ gene
has a myb-homologous DNA-binding domain at the N-terminus. The P1 myb-DNA binding
domain shares 80% amino acid sequence homology with the C1 myb domain. The P1 protein
also contains an acidic region close to the C-terminus that is presumably involved in
transcriptional activation. The P1 gene alone seems to be sufficient to activate the
phlobaphene pathway. No other regulatory genes of the phlobaphene pathway have been
isolated so far. The regulatory role of the P1 gene is restricted to the CHS, CHI, and DFR
steps. P1 does not regulate biosynthetic steps that are necessary for anthocyanin synthesis,
including F3H, AS, UF3GT, and the bz2- encoded glutathione S-transferase. In an in-vitro
DNA-binding assay, the P1 protein did not bind to the bz1 (encodes UF3GT in maize)
promoter fragment even though this fragment contains a myb-binding site where the C1
protein binds (Grotewold et al. 1994). This indicates that the sequence differences between P1
and C1 in their myb domains distinguish their functions in the regulation of two different
branches of flavonoid synthesis.

Regulation of maysin synthesis in maize silk

Maysin is a C-glycosyl flavone derived from flavanones. The accumulation of maysin
in silk renders the corn plant resistant to earworm feeding. Structural genes such as chs and
chi are required for maysin synthesis. However, a mutation in the a1 gene (encodes DFR)
enhances production of C-glycosyl flavonones, an indication of competition between maysin
and phlobaphene, and/or anthocyanin synthesis (Styles and Ceska, 1989). The concentration
of silk maysin is a quantitative trait. In a genetic study, it was shown that the P1 locus is
coincident with a major QTL trait affecting maysin biosynthesis, and accounting for 58% of
silk maysin levels (Byrne et al. 1996). The effect of $PL$ on maysin concentration is additive. Plants with homozygous $PL$ produce twice as much maysin as plants with heterozygous $PL$ (Byrne et al. 1996). In maize BMS cell transformation, the synthesis of C-glycosyl flavone was shown to be regulated by the transformed $PL$ gene. The level of C-glycosyl flavone is correlated with the level of $PL$ expression, in that higher expression of $PL$ produces higher amounts of C-glycosyl flavone (Grotewold et al. 1998). These results also support the conclusion that the effect of the $PL$ gene on maysin concentration is additive.

**Regulation of maize pollen flavonol synthesis**

Pollen flavonol is important to pollen development and function. Studies in petunia and maize showed that pollen-flavonol-deficient plants are self-sterile. Flavonol synthesis requires the structural genes $chs$, $chi$, $f3h$, and $fls$. (the latter one has not been cloned in maize). Expression of the $f3h$ gene in developing anther was closely correlated to flavonol accumulation (Deboo et al. 1995). The $f3h$ gene is also required for anthocyanin synthesis. In maize, the regulatory gene $R$ regulates anthocyanin production in anther. The regulatory role of $R$ on $f3h$ gene was clearly illuminated by mRNA analysis using seedling material. In a dominant $R$ background, light- induced anthocyanin accumulation in seedlings is correlated with the increased levels of $f3h$ mRNA. Whereas, in a recessive $R$ background, there is no anthocyanin accumulation, and no $f3h$ expression (Deboo et al. 1995). However, in maize anther with a recessive $R$ background and without anthocyanin pigment, flavonol still accumulates. It indicates that pollen flavonol synthesis is independent of $R$ gene regulation (Deboo et al. 1995). It is possible that $R$ gene is only functional in anther cells where
anthocyanin accumulates, for example, in the epidermal cell layer, but not in the tapetal cell layer where flavonol is synthesized. Other regulatory genes of the flavonoid pathway that specifically express in tapetal cells would be candidates for flavonol regulation.

Tissue-specific anthocyanin and phlobaphene pigmentation patterns and the molecular mechanisms of tissue-specific pigmentation

In maize the \( R/B \) gene family and the \( C1/Pl \) gene family work together to regulate anthocyanin pigment accumulation. Many patterns of anthocyanin pigmentation can be observed in maize plants (for an illustration of maize plant, see Figure 2), and the distinct patterns were found to be determined by the different forms or combinations of the two regulatory gene families (for a review, see Dooner et al. 1991). The \( R \) gene typically conditions pigmentation of seed aleurone. In fact, there are two separable components of \( R \): the \( S \) (seed) component, conditions kernel aleurone color, whereas the \( P \) (plant) component, conditions pigmentation of anther, coleoptile, and seedling leaf tip (Stadler and Neuffer 1953). The \( Lc \) gene is a linked \( R \) homolog which typically conditions pigmentation in leaf blade (Ludwig et al. 1989). The \( Sn \) gene is another \( R \) homolog which mainly conditions pigmentation in the scutellar node and root mesocotyl (Tonelli et al. 1991). Molecular characterization of \( R(S), R(P), Lc, \) and \( Sn \) showed that they all encode similar \( R \) proteins. These results suggest that the different tissue-specific expression patterns reflect differential regulation of the \( R \) genes. The \( B \) gene conditions pigmentation in mature plant tissues, including leaf sheath and husk leaf (Chandler et al. 1989). The \( B-I \) and the \( B-peru \) alleles
Figure 2. Diagram of the maize plant (from Styles et al. 1973).
specify different pigmentation patterns, and molecular analysis provided evidence that the
different tissue specificity is determined by the \( B \) gene promoter sequences (Radicella et al.
1992). The \( B-I \) allele conditions dark purple leaf sheath, husk, and tassel, whereas the \( B-peru \)
allele conditions moderately pigmented tassel, and intensely and uniformly pigmented seed
aleurone. The \( B-I \) and \( B-peru \) alleles are 98% identical in coding sequence and 90% identical
in 3' flanking sequence. However, there is no sequence homology in their 5' flanking region
and non-translated 5' leader region. In transient maize transformation, the full genomic
sequence of \( B-peru \) produced aleurone pigmentation, whereas the full genomic sequence of
\( B-I \) did not produce aleurone pigment. These results mimic maize plant pigmentation by
endogenous \( B \) alleles. Transformation with a chimeric construct which contains the 5' flanking
and 5' UTR sequences from the \( B-peru \) allele, and the coding and 3' flanking sequences from
the \( B-I \) allele, induced pigment accumulation in aleurone. A counterpart construct containing
the coding and 3' flanking sequences of \( B-peru \) driven by the 5' flanking and 5' UTR
sequences of \( B-I \) did not produce pigment in aleurone. These transformation experiments
further support the hypothesis that the aleurone- specific expression pattern of the \( B-peru \)
allele is determined by the promoter or 5' UTR sequence of \( B-peru \).

The regulatory function of the \( R/B \) gene family requires the presence of a co-activator
encoded by a member of the \( CI/Pl \) regulatory gene family. The \( CI \) gene is required for
anthocyanin accumulation in seed aleurone (Paz-Ares et al. 1987), whereas the \( Pl \) gene is
required for pigmentation in plant tissues (Cone et al. 1993). A recessive \( pl \) allele, sun-red.
can confer light-dependent pigmentation in plant tissues. With the sun-red \( pl \) allele, only
tissues exposed to light become pigmented. Sequence comparison between the sun-red and
the wild type $Pl$ allele revealed that they encode very similar proteins; however, the two alleles
have very different promoters (Cone et al. 1993b). This is another example showing that
promoter sequences of regulatory genes can be responsible for the tissue-specific
pigmentation patterns conferred by the regulatory genes. Other than promoter differences,
methylation also plays a role in tissue-specific pigmentation patterns. A $Pl$ allele, $Pl-Bh$.

pigments plant tissues in blotched pattern. Additionally, the $Pl-Bh$ allele can also pigment seed
aleurone where the wild type $Pl$ allele can not produce anthocyanin. Sequence comparison
between the $Pl-Bh$ and wild type $Pl$ alleles for 2 kb of upstream sequence, full length coding
sequence, and 1 kb downstream sequence revealed no difference. However, DNA methylation
level is very different in the two alleles. The $Pl-Bh$ allele is hyper-methylated, and this higher
methylation level correlates with lower mRNA amount. It has been proposed that novel seed
pigmentation pattern specified by $Pl-Bh$ possibly results from DNA methylation of upstream
regulatory sequences to prevent the binding of a repressor that usually represses $Pl$ expression
in kernel (Cocciolone et al. 1993).

**Tissue specific accumulation of phlobaphene**

In maize, phlobaphene pigment synthesis is regulated by the $Pl$ (stands for pigmented
pericarp) gene. The red phlobaphene pigments accumulate in maize floral organs, most
notably in the ear. The maize kernel pericarp and the cob glume can be both pigmented: this
pattern is specified by $P-rr$, for red pericarp and red cob. Alternatively, an ear can be $P-wr$
(white pericarp and red cob) phenotype, or $P-rw$ (red pericarp and white cob) phenotype
(Fig. 3).
Figure 3. Phlobaphene pigmentation patterns of maize ears carrying different $P$ alleles.
Unlike the $R-r$ locus where two independent components control pigment accumulation in seed and plant parts, the pericarp pigment and the cob pigment seems to be regulated by a single component or gene. Genetic mutations that involve transposable element insertions in the $PI$ locus (Athma et al. 1992) or chromosome deletion (Athma and Peterson 1991) of the $PI$ locus affected pigmentation in both pericarp and cob, simultaneously. The $PI$ gene ($Prr$ allele) was isolated from a transposon tagged maize line (Lechelt et al. 1989). Later, the $Pwr$ allele was isolated using $Prr$ homologous probes (Chopra et al. 1996). The $Pwr$ allele shares very high homology (96% average) to $Prr$ throughout the gene coding region and the 5' flanking region. The major difference between the two alleles was derived from a small chromosome fragment rearrangement near the 3' end of the genes. This rearrangement results in different C terminal regions of the proteins encoded by the two alleles. The difference might be responsible for the different tissue-specific patterns of the two alleles as previously proposed (Chopra et al. 1996). Alternatively, $Pwr$ is observed to have a lower expression level in pericarp, and this may be the reason for the absence of pericarp pigmentation.

Interestingly, the $Pwr$ allele contains amplified $P$ gene sequences. The intact $Pwr$ gene unit (including the 5' flanking sequence and the coding sequence) is repeated 6-fold in a head- to-tail tandem array. Recently, it was proposed that the amplified gene structure may cause some level of gene silencing at $Pwr$. Indeed. DNA methylation level at the $Pwr$ allele is higher than that at the $Prr$ allele (Chopra et al. 1998), and this methylation difference may be related to the observed expression differences.
Duplicated genes

The maize genome contains many duplicated genes, and in many cases the duplicated genes are located on different chromosomes (some examples are listed in table 1. Gaut and Doebley 1998). In fact, the maize genome contains duplicated chromosome segments which share colinear gene arrangements. Maize is thought to have originated as a segmental allotetraploid (Gaut and Doebley 1998). In addition, some duplicated maize genes are found very close to each other on the same chromosome. Some good examples of adjacent duplicated genes include the *R-r* complex locus (Walker et al. 1995) and the zein genes (Heidecker et al. 1991).

Gene duplication produces an additional copy of a gene, and selection pressure on the extra gene copy is then reduced. Rapid accumulation of mutations in one gene copy may inactivate it. Alternatively, the rapid accumulation of amino acid replacements in one copy may lead to a new gene function which could be fixed by positive selection. In other occasions, both copies of the duplicated genes are subjected to purifying selection, in that null mutations are repressed. In this way, both copies retain the original function and the amount of gene product is increased. For example, maize zein proteins are major endosperm storage proteins encoded by the zein multigene family. The standard allele for the 27-kDa-zein contains tandemly duplicated genes, and each of them is a functional zein gene (Das et al. 1991). The 19-kDa-zein and the 22-kDa-zein proteins account for 75%-80% of the total zein proteins. Their genes are highly duplicated and clustered. Many of the copies perform the same expression and function (Heidecker et al. 1991). Thus, zein gene duplication may enable higher level zein expression required for endosperm development.
Another important outcome of duplicated genes retaining the same original function is that one copy may acquire new regulatory sequences that specify a new expression pattern. This type of gene duplication can be well illustrated by the regulatory genes controlling anthocyanin synthesis. The R/B myc-like regulatory gene family of the flavonoid pathway contains duplicated genes that perform the same function (regulation of anthocyanin accumulation) but in distinct tissue-specific patterns (see above). The R and B genes are duplicated genes as evidenced by the high homology of their encoded proteins and the similar function of the proteins. The R gene contains additional complex duplicated genes. The Lc and Sn genes are duplicated R genes located only 1 map unit away from the R locus. Each of them retained the same R protein function, but they are expressed in distinct tissue-specific manner. Moreover, the R locus contains two adjacent duplicated components (S and P) that regulate anthocyanin synthesis in different plant organs. In the myb-like regulatory gene family, the C1 gene and the P1 gene are also duplicated genes that maintained the same regulatory function but act in different tissues. It is interesting to note from the above examples that alteration in the patterns of gene regulation can produce more rapid and significant evolutionary changes than changes in protein functions. In the duplicated genes, new tissue-specific regulatory sequences may be added to one gene copy. From several known cases, the new regulatory sequences were likely generated from transposable elements. The promoter of the R gene S component contains sequences derived from the CACTA family of transposable elements. It was proposed that the aleurone expression of the R (S) component could have been conferred by the transposable element sequences (May et al. 1998). Comparison of the B-peru and B-I alleles indicates that the two sequences diverged
completely at position +43. It has been demonstrated that the 2.5 kb promoter sequence of the
*B*-peru allele, and not the coding sequence, is responsible for generating the novel expression
pattern of B-peru. The 2.5 kb divergent upstream sequence in B-peru is flanked by a 8 bp
target site sequence duplication. This suggests that the 2.5 kb upstream sequence in B-peru is
an insertion that may be a relic of a transposable element, or have resulted from a transposon-
mediated event (Radicella et al. 1992).

In contrast to the gene duplication events that occurred at the regulatory genes of the
anthocyanin pathway, the regulatory gene of the phlobaphene pathway seems quite simple.
The only known gene is PI, and no duplicated PI gene was isolated or reported previously.

Evolution of the flavonoid pathway

Different classes of flavonoids appeared in the plant kingdom sequentially during
evolution (for a review, see Koes et al. 1993). There are no flavonoids in algae. Flavonones.
flavonols and chalcones appeared first in moss, and then spread widely in land plants.
Proanthocyanins first appeared in ferns. Anthocyanins appeared only when flowering plants
were evolved. The sequential appearance of different flavonoids at different evolutionary time
points seems consistent with their biological functions. Flavonones and flavonols are involved
in UV protection, an important function to all plants and thus might evolve first. Whereas,
anthocyanins are pigments that attract insects to flowers for pollination, and their occurrence
may be correlated with the origin of flowering plants.

Enzymes encoded by the structural genes *chs, chi,* and *f3h* catalyze the synthesis of
chalcone, flavonone, and flavonol. Thus the *chs, chi,* and *f3h* genes should be generated early
during evolution. These three genes are considered as early genes. The expression of the early
genes in moss and ferns are perhaps regulated by an as yet unknown set of regulatory genes.
Enzymes encoded by \textit{dfr}, \textit{as}, and \textit{uf3gt} genes along with the enzymes encoded by the early
genes can together produce anthocyanins in flowering plants. The \textit{dfr}, \textit{as}, and \textit{uf3gt} genes are
thus considered as later genes. In petunia, snapdragon, and maize, the synthesis of
anthocyanins is regulated at the transcriptional level in a tissue-specific manner. Because
anthocyanin synthesis requires the coordinate expression of all the structural genes involved,
the expression of the later genes would need to be coupled with the expression of the early
genes. This coordinate regulation could have been achieved by addition to the early genes of
new cis-elements, similar to the cis-elements present in the later genes. This would then allow
similar transcriptional control for both early and later genes by specific regulators, such as the
\textit{R} gene in maize, resulting in anthocyanin synthesis in specific tissues. Alternatively, the plants
could have a duplicated set of early genes that harbor the same cis-elements as that of the later
genes. In fact, duplicated early genes are found in petunia. There are two \textit{chs} genes, \textit{chsA} and
\textit{chsJ} in petunia. The \textit{chsJ} is under the same regulatory control as of \textit{dfr}, \textit{as}, and \textit{uf3gt} genes.
whereas expression of the \textit{chsA} gene is under an unknown regulator. Similarly, the expression
of the petunia \textit{chi} gene (\textit{chiA}), and \textit{f3h} gene (\textit{an3}) seems to be under an unknown regulator.
In snapdragon, the isolated \textit{chs} gene (\textit{nivea}) and \textit{chi} gene (\textit{chi}) are not regulated by \textit{Delila} as
are the \textit{dfr}, \textit{as}, and \textit{uf3gt} genes. In maize, there are also two genes that encode CHS: \textit{c2} and
\textit{whp}. The \textit{c2} gene is regulated by the \textit{R/B} and \textit{C1/Pl} gene families for anthocyanin synthesis.
whereas the \textit{whp} gene is not (Franken et al. 1991). All the evidence listed above suggests that
in flowering plants there are two sets of early genes: one set is coupled with later genes for
anthocyanin synthesis, and the other set of early genes for chalcone, flavonone, and flavonol
synthesis may retain an original, as yet unknown, regulatory gene control. Further
identification of new regulatory genes from flowering plants, mosses and ferns should provide
new insights into the evolution of the flavonoid pathway.

Dissertation organization

Following this general introduction are three chapters of research reports and a general
summary. Chapter 2 and chapter 3 were written in a manuscript format that will be submitted
for publication. Co-author Surinder Chopra (in chapter 2) provided sequence of the sorghum
Yl gene. The contributions of co-authors in chapter 3 are: Jianbo Zhang for maize material
that has been used for P2 deletion mutants screening; Jianbo Zhang also provided helpful
advice in characterization of the mutants; Sheila Maddock for valuable advice and assistance
in maize BMS cell transformation; Yibin Wang for technical assistance in characterization of
BMS transformants.
References


CHAPTER 2. GENE DUPLICATION AND RETROELEMENT INSERTION
GENERATED LINKED MYB-HOMOLOGOUS GENES WITH DISTINCT EXPRESSION PATTERNS IN MAIZE

A paper to be submitted to Proceedings National Academy of Science USA
Peifen Zhang, Surinder Chopra, and Thomas Peterson

Abstract

The maize *P1* gene encodes a myb-homologous protein that regulates flavonoid pigment biosynthesis in floral organs including the kernel pericarp, cob glume and tassel glume. The *P1-rr* allele coding sequence is flanked by long (5.2 kbp) direct repeats. To clarify the origin of the unusual *P-rr* gene structure, we isolated and characterized a second, tightly linked, maize gene (*P2*), and a related gene (*P2-t*) from teosinte parviglumis, the closest wild relative of maize. The maize *P2* gene and the teosinte *P2-t* gene are homologous to the maize *P1* gene in the coding region and 3' flanking region, but not in the 5' flanking region. The 5' flanking regions of maize *P2* and teosinte *P2-t* are similar to each other, but different from that of the maize *P1* gene. We propose that the maize *P1* and *P2* genes were generated by a tandem (head-to-tail) gene duplication in which the 3' flanking sequence of the progenitor *P* gene was duplicated at the 5' side of the incipient *P1* gene, thereby generating the long direct repeats that flank the *P1-rr* allele. This duplication event occurred approximately 2.75 MYA, and subsequent insertions of retroelements separated the *P1* and *P2* genes. Pigmentation phenotypes and transcript analysis indicate that the two duplicated maize *P* genes and the teosinte *P2-t* gene have distinct tissue specific expression patterns. The
different 5' flanking sequences of the \( P1 \) and \( P2 \) genes are likely responsible for their distinct tissue specific expression properties.

**Introduction**

Anthocyanin and phlobaphene pigments are end products of parallel branches in the flavonoid biosynthetic pathway. In maize, anthocyanin biosynthesis is controlled by the combined function of two groups of regulatory genes: the \( Cl/PII \) gene family encodes myb-homologous transcriptional activators, and the \( R/B \) gene family encodes helix-loop-helix co-activators. Expression of these regulatory genes in different tissues produces distinct patterns of anthocyanin pigmentation in maize (reviewed in Dooner et al., 1991). In contrast to the anthocyanin pathway, the \( P1 \) gene is so far the only known regulatory gene required for phlobaphene biosynthesis. The \( P1 \) gene encodes a myb-like transcriptional activator (Grotewold et al. 1991; Grotewold et al. 1994) of the structural genes \( C2 \), \( Chi \), and \( Ai \) which encode chalcone synthase, chalcone isomerase and dihydroflavonol reductase, respectively. The sequential function of these enzymes converts simple organic compounds to polyphenolic flavonoids including phlobaphene pigment (Styles and Ceska, 1977). The most notable red pigmentation pattern conferred by the \( P1 \) gene is in the kernel pericarp and cob glume of the maize ear. The \( P-rr \) allele specifies red pericarp and red cob. \( P-wr \) specifies white (colorless) pericarp and red cob, and \( P-www \) specifies white (colorless) pericarp and cob. Unlike the \( R-r \) gene which has two independently mutable and separable components for seed color and plant color (Stadler and Neuffer, 1953), the pericarp color and cob color are controlled by a single
coding sequence of $Pl$-$rr$. Deletion of the $Pl$-$rr$ coding region results in colorless pericarp and colorless cob simultaneously (Athma and Peterson, 1991). Molecular isolation and characterization of the $Pl$-$rr$ and $Pl$-$wr$ alleles indicates that they are highly homologous (Chopra et al. 1996). Interestingly, the $Pl$-$rr$ allele has an unusual gene structure, with a single coding region flanked by two 5.2 kbp direct repeats (Lechelt et al. 1989). The $Pl$-$wr$ allele has an even more complex structure, with 5–6 copies of 12.6 kbp $P$ gene sequences, including regulatory and coding sequences, amplified in a head-to-tail tandem array (Chopra et al. 1998). Each of the $Pl$-$wr$ gene copies is highly homologous (99%) to the $Pl$-$rr$ gene, except for a short rearrangement (1269 bp) near the 3' end of $Pl$-$wr$.

It is unclear how the complex $Pl$-$rr$ and $Pl$-$wr$ gene structures originated and evolved, but it seems likely that gene duplication events were involved in both cases. Gene duplication is considered to have played an important role in the evolution of plant genomes. It has been proposed that the $Cl/PiI$ and $R/B$ genes represent the descendants of ancient gene duplications (Cone et al. 1993; Chandler et al. 1989). This view is supported by their conserved coding sequences and the exchangeable functions of their encoded proteins (Goff et al. 1990; Ludwig et al. 1990). Furthermore, the $R$ gene family includes other linked loci, such as the $Sn$ and $Lc$ genes, that confer additional distinct pigmentation patterns (Dooner et al. 1976; Ludwig et al. 1989; Tonelli et al. 1991). These diverse expression patterns have been proposed to reflect differences in expression elicited by different promoter sequences. This idea has been confirmed in the case of two alleles of the $B$ gene by transformation analysis (Radicella, 1992). Gene duplication and subsequent divergence in their expression pattern by acquisition of different gene regulatory sequences may have occurred multiple times in the
aforementioned regulatory genes of anthocyanin biosynthesis pathway. Despite the prevalence of gene duplication and its significance in shaping both plant and animal genomes, little is known about the mechanism(s) involved.

We report here the isolation of a second *Pl* gene homolog (*P2*) from maize that is tightly linked to the *Pl-rr* gene, and a single copy *P* allele (*P2-t*) from teosinte that resembles the maize *P2* gene. Molecular analysis of their gene sequences and expression patterns suggests a model of maize *P* gene duplication and subsequent divergence to generate distinctly regulated homologous genes. Our results illustrate how, through gene duplication, the regulatory sequences of one of the duplicated gene copies can be altered, leading to the acquisition of new patterns of expression.

**Materials and Methods**

**Stocks**

The *P-ww-1112* allele carries a deletion of *Pl-rr* (Athma and Peterson, 1991). Teosinte parviglumis *Itis81* stock was obtained from Dr. John Doebley at the University of Minnesota. One plant (*Itis 81-5*) was identified as carrying a single copy *P* homologue (*P2-t*). Analysis of progeny derived from outcrossing and self-pollination indicate that the *P2-t* is allelic to the maize *P* gene, and is homozygous in the *Itis 81-5* line.
Genomic library construction and screening

A genomic library used to clone the maize $P2$ gene was made from young leaf DNA of maize $P$-ww-1112. Genomic DNA was prepared using CTAB reagent, partially digested with $Sau3A1$, and ligated to partially filled-in $XhoI$ site of lambda FixII vector (Stratagene). About 500,000 independent phage clones were screened using $Pl$-$rr$ gene coding region probes.

PCR1 and genomic fragment 12. PCR1 is a PCR fragment amplified with primers EP5-8/EP3-13 (primers were previously described in Grotewold et al. 1991) located at the 5' end of $Pl$-$rr$ gene transcribed region, whereas fragment 12 (Lechelt et al. 1989) is at the 3' end of $Pl$-$rr$ gene transcribed region. Clones that hybridized to both probes were selected. Clone 3 contains the largest insert (20 kbp), and its $Not$I fragments were subcloned to pBluescript for sequencing.

To clone the teosinte $P2$-$t$ gene, a genomic library was made from DNA of parviglumis Itis81-5. The DNA was partially digested with $Sau3A1$, ligated to Lambda fixII vector (Stratagene), and approximately 800,000 independent clones were screened using probe PCR1, and $Pl$-$rr$ genomic fragments 13 and 15. Probe PCR1 encompasses the 5' region encoding the Myb-homologous domain, whereas fragments 13 and 15 (Lechelt et al. 1989) are located at the 3' end of $Pl$-$rr$ gene transcribed region, and 3' flanking region, respectively. Clone 1, that contains the largest insert (19 kbp) and hybridizes to all three probes, was selected for further study. The $Sal$I fragments of Clone 1 were subcloned to pBluescript and sequenced at the Iowa State University Nucleic Acids Facility.
RNA isolation and RT-PCR

Pericarps were dissected from developing maize kernels at 20 DAP (Day-After-Pollination). Tassel glumes of *Pl-rr* and teosinte Itlis 81-5 were dissected from tassels at anthesis. Immature anthers of *Pl-rr, P-ww-1112*, and teosinte Itlis 81-5 were collected at early pollen development stage (between uninucleate and binucleate stages). Samples were frozen in liquid N2 and ground to a fine powder with a mortar and pestle. Total RNA was extracted using Trizol reagent (GIBCO BRL), and treated with Dnase at 37 °C for 45 min to remove residual genomic DNA. Aliquots of RNA (1 - 2 µg) were reverse transcribed using either Oligo dT primer or EP3-12 (5'-AAGCTTGAATTCGAGTTC CAGTAGTTCTTGATC-3'). an oligonucleotide homologous to *P* gene exon 3. SuperScript reverse transcriptase was from GIBCO BRL, and reverse transcription was carried at 42 °C to alleviate possible secondary structure influence. The reaction was stopped by heating at 95°C for 5 min to destroy the enzyme. The first strand cDNA pool was then diluted 5 fold in dH2O, and one tenth was used in PCR amplification. Primers EP5-8 (5'ACGCGCGACCAGCTGCT AACCGTG-3'; homologous to the 5' untranslated region of the maize *Pl* gene) and EP3-13 (5'-AGGAATTCCGCCC GAAGGTAGTTGATCC-3'; homologous to *Pl* gene exon 2) were used to amplify either the maize *Pl* or *P2* transcripts. Primers P2-5 (5'-CTCGATTGGCGGGAC CAGC-3'; homologous to the maize *P2* untranslated region) and EP3-13 were used to specifically amplify the maize *P2* and the teosinte *P2-t* transcript. PCR reactions contain 1.5 mM MgCl2, and the PCR condition was: 94 °C for 4 min, followed by 30 cycles of 94 °C for 45 sec, 60 °C for 1 min, and 72 °C for 1 min, and a final extension for 10 min at 72 °C. RT-PCR reactions were repeated at least once to verify the results.
Results

Isolation of a second P gene from maize

Previously-reported Southern blot hybridizations indicated that a second P-homologous sequence is present and tightly linked to the Pl-rr locus (Athma and Peterson, 1991; Das and Messing, 1994). We cloned this second P-homologous sequence (P2) from a maize line (P-ww-l112) in which the Pl coding sequence was deleted as a result of recombination between the 5.2 kbp repeats flanking the Pl-rr gene (Materials and Methods). Additional evidence for the chromosomal location of P2 came from a deletion mutant, P-del-2. The deletion spanned the region from exon 3 of P2 to intron 2 of the Pl-rr gene, and thereby joined the non-deleted 5' region of P2 to the 3' portion of the Pl-rr gene (J. Zhang, P. Zhang, and T. Peterson, unpublished). The molecular structure of the P-del-2 deletion confirms the tight linkage of P2 and Pl-rr, and additionally indicates that the two genes are in the same transcriptional orientation with the P2 gene 5' of the Pl-rr gene.

Isolation of a P gene homologue from teosinte

To clarify the origin of the maize P gene duplications, we investigated the structure of P gene homologs in teosinte, a wild relative of maize. Due to the expected diversity of P alleles in teosinte, we screened a number of teosinte lines for the presence of a functional P gene as indicated by the presence of pigmented tassel glume margins. In maize, the Pl gene confers red/brown phlobaphene pigmentation to the tassel glume. A similar reddish brown tassel glume margin pigmentation pattern was observed in several accessions of teosinte.
including representatives of teosinte parviglumis and teosinte mexicana. One line with pigmented tassel glumes, teosinte parviglumis Ilitis 81-5, was selected for further analysis because of its relatively simple $P$ gene structure. Genomic DNA of teosinte parviglumis Ilitis 81-5 was digested with different enzymes, and hybridized to maize $P$ gene specific probes. Hybridization with either a 5' end probe of the $P1$ coding region (fragment PCR1) encompassing the Myb domain or an intron 2 probe (fragment 8B) detected a single strongly-hybridizing band with each of five individual enzyme digestions (data not shown). Additionally, 22 progeny plants from a single self-pollinated plant gave the same band pattern as the parental plant (data not shown). Together, these results indicate that the Ilitis 81-5 stock is homozygous for a single copy $P$ gene homologue. We cloned this teosinte $P$ gene ($P2-t$) from Ilitis 81-5 (Materials and Methods). To determine the allelic relationship of the teosinte $P$ gene, the Ilitis 81-5 stock was crossed to a maize $P2$-ww line containing the $P2$ gene ($P-ww-11/2$) and then crossed to a second, structurally distinct, $P-ww$ line ($P-ww-4Co63$). Among 46 progeny, 25 contained only the teosinte $P2-t$ gene band, while 21 contained only the maize $P2$ gene band (data not shown). The approximate 1:1 ratio, and the fact that no plants contained both the maize $P2$ and the teosinte $P2-t$ gene bands, indicates that the teosinte $P2-t$ gene and the maize $P2$ gene segregate as alleles in repulsion.

**Structural comparison of $P$ gene homologs**

We isolated and sequenced 12 kbp from both the maize $P2$ gene and the teosinte $P2-t$ gene. The structures of the maize $P2$ gene and the teosinte $P2-t$ gene were compared to two alleles of the maize $P1$ gene, $P1-rr$ and $P1-wr$ (Figure 1). Both the maize $P2$ gene and the
teosinte P2-t gene have the same exon/intron structure as the maize PI gene. Moreover, the sequences of the 5' UTR and the two introns of the four genes are greater than 91% similar. Interestingly, the PI-rr and PI-wr alleles contain two insertions that are absent in both the maize P2 gene and the teosinte P2-t gene: one is a 80 bp insertion in the 5' UTR, and a second is a 734 bp insertion in intron 2.

The sequence similarity of the ORF regions is 98% or higher among the four genes, except for a region of divergence near the 3' end of the gene. As mentioned above, the PI-wr allele differs from the PI-rr allele at the 3' end, due to a short chromosome rearrangement (Chopra et al. 1996). In the PI-wr allele, a 210 bp fragment (wr61) replaces a 660 bp fragment (14) that is present in the PI-rr allele. Following the wr61 fragment, PI-wr contains a 1 kbp fragment (fragment C in figure 1) that is absent from the PI-rr 3' end, but present at the PI-rr 5' end approximately 8 kbp from the PI-rr transcription start site (Figure 1; Chopra et al., 1996). Further 3', the PI-rr and PI-wr alleles are 99% identical over a 3.5 kbp region, comprising fragments 15, 6 and 7A (Figure 1). It was previously proposed (Chopra et al. 1996) that the 3' end structure of the PI-wr allele is similar to that of the ancestral P gene, and that the wr61 fragment was replaced by fragment 14 during the generation of the PI-rr allele. Interestingly, the 3' flanking regions of both the teosinte P2-t gene and the maize P2 gene resemble the PI-wr allele, in that they too contain the wr61 fragment in place of fragment 14. Moreover, the similarity of the teosinte P2-t gene to the maize PI-wr allele extends approximately 3.5 kbp further 3', including the fragment C, and fragments 15 and 6. These results support the hypothesis that the 3' region of the ancestral P gene was more similar to that of the maize PI-wr and teosinte P2-t genes.
In the 5' flanking regions, the maize $P1-rr$, $P1-wr$, $P2$ genes and the teosinte $P2-t$ gene all contain a homologous, 98% identity, 90 bp region upstream from the transcription start site, including the TATA box. Further 5' of this 90 bp region, the maize $P2$ and teosinte $P2-t$ genes are similar to each other, but distinct from the $P1-rr$ and $P1-wr$ genes (Figure 1). The $P1-rr$ gene contains a 5.2 kbp 5' flanking sequence that is also present as a direct repeat 3' of the coding sequence. The $P1-rr$ and $P1-wr$ 5' flanking sequences are nearly identical (99.3% similarity over 5.2 kbp), whereas, the maize $P2$ and teosinte $P2-t$ genes share a 170 bp region of similarity (93%), and a 1 kbp block of similarity (96%) located further 5'. The 170 bp and 1 kbp regions of similarity are separated by non-homologous sequence blocks of 2.6 kbp and 120 bp in the maize $P2$ and teosinte $P2-t$ genes, respectively. The maize $P2$ gene contains a region similar to retrotransposon $Prem-2$ (Turcich et al. 1996) located 3.9 kbp 5' of the transcription start site. In contrast, the teosinte $P2-t$ gene contains a 3' LTR of retrotransposon $Prem-1$ (Turcich et al. 1994), located about 2 kbp 5' of the transcription start site.

**Conserved protein functional domains and diverged C terminus**

The deduced amino acid sequences of the maize $P2$ gene and the teosinte $P2-t$ gene products were aligned with those of the $P1-rr$ and $P1-wr$ alleles of the maize $P1$ gene (Figure 2). Important functional domains such as the myb-DNA binding domain and the putative transcriptional activation domain (Grotewold et al. 1991; Grotewold et al. 1994) are highly conserved in all four genes. Outside of the two functional domains, there are several amino acid substitutions and in/del mutations. The most striking difference is found at the protein C
terminus. The maize P2 and teosinte P2-t gene translation products diverge from that of the maize Pl gene shortly after the presumptive activation domain. However, their nucleotide sequences are 98–99% identical in the same region. The divergence in amino acid sequences reflects a reading frame difference caused by a single base in/del between the Pl and P2 nucleotide sequences. The deduced P2 and teosinte P2-t proteins are shorter than both the Pl-rr and Pl-wr encoded proteins. Further C-terminal, beyond the region of comparison with P2, the Pl-rr- and Pl-wr-encoded proteins differ from each other as reported previously (Chopra et al. 1996).

Another related P gene homologue (teosinte clone-4) was isolated from the same genomic library of teosinte Itis 81-5. The nucleotide sequence of teosinte Clone-4 has greatest homology to P gene sequences in the exons (81–94%), and lower homology in the adjacent non-coding sequences (39–60% identity over sequenced regions). The clone-4 gene encodes a protein which appears to be a more distant member of the P gene family (76% overall similarity to maize P2 gene product). Most importantly, the clone-4 gene product C-terminal sequence is homologous to that of the P2 and teosinte P2-t genes (Figure 2). Because the clone-4 gene diverged prior to the divergence of the Pl and P2 genes, we conclude that the C terminal sequences shared by clone 4, P2-t, and maize P2 protein more closely resemble that of the progenitor P gene. This supposition is further supported by analysis of the YI gene, a P gene homologue that controls phlobaphene pigmentation in sorghum seed coat (Surinder Chopra and Tom Peterson, unpublished): the deduced YI protein C terminal region is unlike Pl. but it contains blocks of amino acid sequences with high homology to the maize P2 gene product (Figure 2).
P2 gene and teosinte P gene have distinct expression pattern

Both the P1-rr and P1-wr alleles of the maize P1 gene control phlobaphene pigment accumulation in kernel pericarp, cob and tassel glumes. Moreover, the presence of P1-rr and P1-wr transcripts in pericarp and cob has been confirmed by Northern blot and/or RT-PCR analysis (Grotewold et al. 1991; Chopra et al. 1996). In contrast, the P2 gene does not induce detectable phlobaphene pigmentation, based on the absence of phlobaphene pigments in plants in which P1-rr function is destroyed, but which still retain an intact P2 gene (Athma and Peterson, 1991; Athma et al. 1992). As noted above, the maize P2 gene and the teosinte P2-t gene 5' flanking sequences are similar to each other, yet different from P1. To determine the P2 transcript expression pattern, we employed a sensitive RT-PCR method to detect RNA transcripts in kernel pericarp (20DAP), tassel glume, young silk, young anther, and young leaf. RNA samples from these tissues were reverse transcribed, and the resulting cDNA preparations were PCR amplified using oligonucleotide primers EP5-8 and EP3-13, located in the P gene 5' UTR and exon 2, respectively. This primer pair can amplify both P1 transcripts and P2 transcripts, and produces a different size for each: the P1 gene product is 380 bp, whereas the P2 gene product is 300 bp due to a 80 bp deletion in the P2 5' UTR. To further verify expression from the P2 gene, we used oligonucleotide primer P2-5 specifically designed from the 5' UTR sequence. Figure 3 clearly shows that the maize P1 and P2 genes have distinct patterns of transcript accumulation. Using RNAs isolated from P1-rr pericarp and tassel glume, primer pair EP 5-8 and EP 3-13 amplified only a 380 bp band, derived from P1-rr transcripts. In contrast, using RNA isolated from silk, the same primer pair amplified both the 380bp and 300bp bands, which corresponds to P1 and P2 transcripts respectively. In
developing anther, the primer pair amplified a stronger 300bp band of P2, and a much weaker 380 bp band of P1. In vegetative tissue (young leaf), neither P1 nor P2 were detectably expressed. Expression of the P2 gene was confirmed using RNA isolated from a maize line carrying the P-ww-1112 mutation, in which P1 is deleted but P2 is present. RNA from silk and young anther was amplified with P2-specific primer pair P2-5 and EP 3-13 to produce a 240 bp band of P2 (Figure 3). Like the maize P2 gene, the teosinte P2-t gene is also expressed in silk and young anther, as indicated by RT-PCR assay using primers EP5-8/EP3-13 that yield a 300 bp product, and P2-5/EP3-13 that yield a 240 bp product (Figure 3). Additionally, RT-PCR analysis of RNA from teosinte tassel glume using primers EP 5-8/EP 3-13 produces a 300 bp band. In teosinte, transcripts of the homologous gene clone-4 could also be amplified by primers EP5-8/EP3-13 yielding a product of approximately 300 bp which would co-migrate with the P2-t specific product. That the P2-t gene is expressed in silk and young anther was confirmed using the primers P2-5/EP3-13, which can amplify transcripts from P2-t, but not clone-4. The 300 bp band amplified in tassel glume, using primers EP5-8/EP3-13, could be generated from either the teosinte P2-t or clone-4 genes. However, we consider it most likely to be derived from the P2-t gene, based on our observations that, in 46 backcross progeny segregating for the P2-t and clone-4 genes, tassel glume pigmentation is correlated with presence of the P2-t gene, but not the clone-4 gene.
Discussion

P gene duplication and genesis

We have isolated a second P gene (P2) from maize that is tightly linked to the P1 gene. The P2 coding sequence is very similar to the P1 gene, and the Myb-like DNA binding domain and putative transcriptional activation domain are highly conserved, indicating that functionality of the P2 gene has been maintained by selection. The P2 gene 3' flanking region is similar to that of the P1 gene (P1-wr allele), whereas the P2 and P1 5' flanking regions are non-homologous. We have also isolated a single copy P gene from teosinte (P2-t) that segregates as an allele of the maize P1/P2 complex. Interestingly, the teosinte P2-t gene is more closely related to the maize P2 gene throughout both 5' and 3' flanking regions and coding sequences. Moreover, the maize P2 and teosinte P2-t genes share similar tissue specific expression patterns that are different from P1.

Based on the tight linkage and high homology of the maize P1 and P2 genes, we propose that these were generated by a gene duplication event. The duplication occurred approximately 2.75 million years ago based on the formula $R = K / 2T$ (Li et al. 1991), using synonymous sites for K (Nei and Gojobori, 1986), and using $6 \times 10^{-9}$ as the rate of substitution per synonymous site per year for grass nuclear genes (Gaut 1998). It is thought that maize was domesticated from teosinte parviglumis ~7500 years ago (Iltis 1983; Doebley et al. 1984). Thus, P gene duplication occurred long before maize domestication. The single copy P gene we isolated from teosinte parviglumis Iltis 81-5 may have been retained as the original non-duplicated gene and evolved independently. The divergence time between the
teosinte *P2-t* gene and either of the maize *P1* or *P2* genes was estimated to be roughly the same (2.75 MYA).

A complete gene duplication produces two identical genes. Both copies may retain their original function, enabling the organism to produce a higher quantity of RNA or proteins, as in the case of rRNA, tRNA, and histone genes. Alternatively, one of the copies may become a functionless pseudogene by accumulation of deleterious mutations, or one gene copy may mutate to gain a new function (Li et al. 1991). Finally, one copy may retain the same coding sequence function, but it may acquire new regulatory elements that specify a different pattern of expression. The latter process has occurred in some cases of tissue specific divergence of duplicated genes or alleles, including some tissue-specific regulatory genes in the anthocyanin biosynthetic pathway (see Introduction). Variation in the regulation of duplicated genes could be generated in a single step by a genome rearrangement that places new regulatory sequences next to a gene. Transposable elements have been proposed as being involved in inducing rearrangements of gene flanking regions (Habu et al. 1997), but little is known about the actual mechanisms involved. Here, we present evidence that the maize *P1* and *P2* genes were generated by a gene duplication event, and we propose a mechanism to explain the generation of new regulatory sequences for one of the duplicated *P* genes.

Previous studies of the structure of the *Pl-rr* allele indicated that the *P* coding sequence is flanked on both the 5' and 3' sides by 5.2 kbp direct repeat sequences. Transgenic analysis of *Pl-rr* 5' flanking sequences indicates that the 5' flanking region contains regulatory elements sufficient to confer *P1* gene tissue specificity (Sidorenko et al. 1999). However, it was unknown how the unusual *Pl-rr* gene structure was generated. The
repeats do not resemble transposons in structure, because their terminal sequences do not have the inverted repeats of transposons, nor are the repeats flanked by target site duplications that are normally generated upon transposon insertion (Athma and Peterson 1991). Analysis of the maize P2 and teosinte P2-t gene structure shows that their 5' flanking sequences are similar to each other, but different from that of Pl. In contrast, the 3' flanking sequences of the maize P2 and teosinte P2-t are similar to the Pl gene flanking sequence: this similarity is more clearly seen in the teosinte P2-t gene. These results suggest that the progenitor allele of the P gene had a single copy of the direct repeat sequence, and that it would have been located at the 3' end of gene. To generate the Pl-rr type of structure, the 3' flanking sequence must have been duplicated and placed on the 5' side of the gene. Thus, the gene could have come under the regulatory influence of the new 5' sequences. We propose that the initial gene duplication event included not only the 3' repeat sequences, but also the P coding sequence to generate the P2 and Pl genes.

Figure 4 illustrates our model of P gene duplication and subsequent generation of the P2 and Pl genes. The structure of the progenitor P gene (P') is predicted from the teosinte P2-t gene. We propose that the P' gene had a 5' flanking region containing the two 5' flanking sequence fragments (1 kbp and 170 bp) common to the teosinte P2-t gene and the maize P2 gene (patterned box). The P' gene would also have the conserved 90 bp promoter sequence (gray box) common among the teosinte P2-t gene, maize P2 gene, and Pl-rr and Pl-wr alleles of the maize Pl gene. The P' 3' flanking region would resemble the 3' region of the Pl-wr and P2-t genes, and would contain fragment wr61, followed by fragment C and fragments 15, 6, 7A, and possibly 7B (see discussion below). We propose that the progenitor
Gene sequence, from the 5' end of the conserved 90 bp promoter sequence to the 3' end of fragment 7A where the 5.2 kbp direct repeat sequence ends, approximately 10 kbp, was duplicated in a tandem head-to-tail arrangement. This duplication placed the formerly 3' flanking sequences on the 5' side of the downstream gene. Following this duplication, sequential retrotransposon insertions truncated the P2 3' flanking region and separated the two duplicated genes. The P2 gene and Pl gene backbones were thus formed. The Pl gene backbone is retained in the Pl-wr allele, except that Pl-wr was tandemly amplified 5–6 fold, possibly by un-equal crossing over. The Pl-rr allele has a short chromosome rearrangement in its flanking regions as mentioned above. From our model and the estimated maize P gene duplication date of approximately 2.75 MYA, we can also conclude that multiple retrotransposon insertions in the Pil/P2 complex region occurred within the last 2.75 million years. This time range is consistent with a recent report that multiple retrotransposon insertions occurred in maize genome within the last three million years (SanMiguel et al. 1998).

Sequence information at the duplicated P2/Pl-rr genes border supports the above model. The P2 gene 3' flanking sequence is similar to the P2-t and Pl-wr sequences for 250 bp 3' of the translational stop: the homology ends about 30 bp prior to the Psrl site in fragment wr61 (Figure 5). The homology is interrupted by a 540 bp sequence of unknown origin, followed by a Prem-2 element, a highly repetitive retrotransposon found in the ends of many maize genes. The Prem-2 insertion continues for 2.2 kbp to the end of the P2 genomic clone. Interestingly, the 30 bp sequence similarity begins again adjacent to the Psrl site in Pl-rr, and the following fragment C sequence that are absent from the P2 3' end are found.
upstream of the 5.2 kbp 5' flanking sequence of \textit{Pl-rr} (Figure 5 and Figure 1). Thus, it appears that the break point of \textit{P2} and \textit{Pl-rr} were separated by insertions. In \textit{Pl-rr}, further 5' of this 30 bp sequence is a 620 bp sequence of unknown origin (Figure 5). Sequences immediately upstream of the 620 bp segment resemble internal sequences of the repetitive retroelement \textit{opie-2}. The \textit{Pl-rr} locus genomic clone terminates at an \textit{EcoRI} site within the region of homology to \textit{opie-2}. The physical distance between the 3' end of \textit{P2} and the 5' end of \textit{Pl-rr} is not known. However, this interval does not contain any essential gene based on the fact that plants homozygous for an interstitial deletion that removes this region are viable and show no obvious abnormalities. Possibly, the chromosome region between \textit{P2} and \textit{Pl} may be occupied only by retroelements. In maize, the intervals between genes are often found to be composed of retroelement sequences, and it has been proposed that retroelement sequences comprise at least 50% of the maize genome (SanMiguel et al. 1996). Nevertheless, the 540 bp sequence 3' of the \textit{P2} gene and the 620 bp sequence near the \textit{Pl-rr} gene 5' end (Figure 5) are not direct repeat sequences as expected for retrotransposon LTRs, nor are they homologous to known sequences in database. One possibility is that they are truncated LTRs of an undefined retroelement. Retroelement LTRs commonly have short (4 to 12 bp) inverted repeats at their ends that are presumably required for integration. For example, the yeast \textit{Ty} retrotransposon family (Boeke 1989), \textit{Drosophila copia} retrotransposon (Bingham et al. 1989), and maize retrotransposons (SanMiguel et al. 1996) all contain the sequences TGTTG and CAACA at the LTR ends. The 620 bp presumptive sequence next to the \textit{Pl-rr} gene 5' end terminates in CAACA, suggesting that this 620 bp sequence is derived from a retrotransposon LTR. More importantly, a 5 bp direct duplication (AAGAC) is found
adjacent to the 540 bp and 620 bp segments at the 3' end of P2 gene and 5' end of Pl-rr gene, respectively (Figure 5). Retrotransposon insertions in maize are generally flanked by 5 bp target site sequence duplications (SanMiguel et al. 1996). The 5 bp target site duplication found at the P2 gene 3' border and the Pl-rr gene 5' border strongly supports the hypothesis that retrotransposon(s) inserted and separated the two copies of the original tandemly (head-to-tail) duplicated P genes. It has been reported that retroelements commonly insert into another retroelement, and occasionally insert within LTRs (SanMiguel et al. 1996). Similarly, a retrotransposon could have inserted between the duplicated P2 and Pl genes, followed by multiple hits of other retrotransposons including Prem-2 and Opie-2.

It is unclear how P gene duplication occurred at the first step in the Figure 4 model. Tandem duplication can occur via unequal crossing over in repeated sequences flanking a single copy gene. For example, unequal crossing over between two copies of a transposable element inserted on either side of the Drosophila Bar gene was shown to generate a tandem duplication associated with the Bar-B allele (Tsubota et al. 1989). To determine whether the same mechanism was involved in duplication of the P locus, we examined the sequence at the putative duplication breakpoint located in the interval 1 kbp (fragment 7B) 5' of the Pl-rr transcription start site. The duplication breakpoint region is delimited by the 3' end of the 5.2 kbp direct repeat at position -1034 upstream of Pl-rr and the beginning of the 90 bp conserved promoter sequence at position -92 (Figure 4, fragment 7B). This 923 bp sequence contains a number of inverted and direct sequence repeats, but does not have the terminal inverted repeats that are a characteristic feature of many transposons, nor does it share
sequence homology with any known transposable elements. Thus, it is not clear if the 923 bp sequence mediated an un-equal crossing over in the P locus.

**Possible functions of maize P2 gene and the homologous teosinte P gene**

The maize Pl gene controls the synthesis of a red flavonoid pigment in maize floral organs including kernel pericarp, cob glume and tassel glume. The deduced amino acid sequences of the proteins encoded by the maize P2 and teosinte P2-t genes show that they are nearly identical to the Pl protein at the Myb DNA binding domain and the putative acidic transcriptional activation domain. Thus the teosinte P2-t gene and the maize P2 gene could potentially have the same function as the Pl gene. However, plants that carry deletions or other mutations that destroy Pl-rr function, yet leave P2 intact, have colorless pericarp, cob, and tassel glumes. The absence of pigmentation could be due to a lack of transcripts: indeed, RT-PCR experiments indicate that P2 transcripts are not present in pericarp or tassel glume. In contrast, transcripts of the Pl-rr gene are readily detected in these tissues from plants carrying the Pl-rr allele (Figure 3).

Like the maize P2 gene, the teosinte P2-t gene does not pigment pericarp or cob glume. In the test crosses of teosinte Iltis 81-5 to P-ww maize lines, the progeny plants have colorless kernel pericarp and cob glume regardless of the presence of the teosinte P2-t gene. However, the tassel glume margin of the teosinte Iltis 81-5 stock is slightly reddish-brown pigmented, and this tassel glume pigmentation co-segregates with the teosinte P2-t gene when testcrossed into maize. Transcripts of the teosinte P2-t gene were detected by RT-PCR analysis using tassel glume RNA. This suggests that the protein encoded by P2-t is capable of
conferring pigmentation as long as the gene is expressed in a tissue competent for phlobaphene accumulation. The lack of pigmentation in pericarp and cob glume is likely due to the different tissue specific expression pattern of the \textit{P2-t} gene. Taken together, these results suggest that the \textit{P2-t} protein, the \textit{P2} protein, and by extension the protein encoded by the progenitor \textit{P} gene, all have the potential to activate phlobaphene biosynthesis.

In contrast to the \textit{P1} gene, the gene expression profile indicates that the \textit{P2} gene and the teosinte \textit{P2-t} gene are expressed in silk and young anther. It is very likely that the different 5' flanking sequence of the \textit{P2} gene and the teosinte \textit{P2-t} gene are responsible for their distinct tissue specific expression pattern. Silk and anther are important reproductive organs of maize, and the myb-homologous regulatory \textit{P} protein could have an important biological function in these organs. In fact, the maize \textit{P1} gene is involved in silk browning, a phenotype observed in wounded silks as a result of dihydroxyl flavone oxidation (Levings III et al. 1971; Coe et al. 1986; Han et al. 1987; Widstrom, 1987; Coe et al. 1988). In addition, the \textit{P1-wr} allele has been shown to regulate levels of silk maysin synthesis, a C-glycosyl flavone compound that provides resistance to feeding by corn earworm (Byrne et al. 1996a). The \textit{P1-rr} and \textit{P1-wr} alleles have been previously implicated in silk browning, but the following observations suggest that, in addition, the maize \textit{P2} gene is additive for the silk browning reaction. Maize plants of genotype \textit{P-ww-1112}, in which the \textit{P1} gene is deleted but the \textit{P2} gene is intact, still retain a certain degree of silk browning. In contrast, the mutant (\textit{P-del-2}) carrying the interstitial deletion of \textit{P1} and \textit{P2} has no silk browning indicating that loss of \textit{P2} function is coincident with loss of residual silk browning. The teosinte \textit{P2-t} gene may also control silk browning, based on the observation that mashed teosinte silk turned
brownish. Silk browning phenotype was reported to correlate with maysin synthesis (Byrne et al. 1996b), hence the maize \( P2 \) gene and the teosinte \( P2-t \) gene may also be involved in maysin biosynthesis.

As mentioned above, the maize and teosinte \( P \) genes are highly conserved in the Myb-like DNA binding domain, and the putative transcriptional activation domains. Immediately after the putative activation domain, the \( P \) gene homologs have varying C-terminal regions. The \( P1-rr \) and \( P1-wr \) proteins have nearly identical C-terminal sequences until a point of chromosomal rearrangement that produces completely divergent C-terminal ends. In contrast, the maize \( P2 \), teosinte \( P2-t \), clone-4, and the sorghum \( Y \) proteins share a region of considerable homology over approximately 100 AA residues at the C-terminus (Fig. 2). The fact that this distinct C-terminal domain is found in sorghum further supports the idea that the maize \( P2 \) and teosinte \( P2-t \) genes, and not the \( P1-rr \) allele, more closely resemble the ancestral \( P \) gene. The C-terminal region of \( P2 \) may provide additional functions to the maize \( P2 \) protein that are not present in the \( P1 \)-encoded proteins that lack this domain. For example, this C-terminal domain could facilitate interactions with other co-factors and thereby allow the protein to regulate other steps in flavonoid biosynthesis than those regulated by \( P1 \) alleles.

One speculation is that \( P2 \) may regulate the biosynthesis of flavonol, a compound required for pollen development and pollen function in petunia and maize (Mo et al. 1992). Maize plants that produce flavonol-deficient pollen are male sterile. It has previously been suggested that the regulator of pollen flavonol biosynthesis is a \( P \)-like protein (Deboo et al. 1995). However, pollen of \( P1 \) deletion plants is viable, thus excluding \( P1 \) as the sole regulator of pollen flavonol synthesis. With the isolation of the Myb-homologous \( P2 \) gene and the presence of \( P2 \)
transcripts in anther, it raised the possibility that $P2$ is the pollen flavonol regulator. The hypothesized function of $P2$ can be tested by expression of $P2$ cDNA in transgenic maize BMS cultured cells, a similar experiment as in the report by Grotewold (Grotewold et al. 1998). Transformed cells expressing $P2$ can be assayed for transcription of the structural genes such as $C2$, $Chi$, $A1$, or $F3H$, and the accumulation of corresponding flavonoid compounds. The results should prove informative not only in establishing the function of the $P2$ gene, but also from this we can infer the function of the ancestral $P$ gene.

In conclusion, we have isolated examples of $P$-homologous genes from maize and teosinte. The maize $P2$ gene structure significantly differs from the $P1$ gene in 5’ flanking region, and its tissue specific expression pattern is different from $P1$. The maize $P1$ and $P2$ genes arose from a tandem gene duplication, followed by retrotransposon insertions to generate the extant maize $P2/P1$ gene complex. The conserved amino acid sequence of $P2$ gene and the observation of mutant phenotype suggest that it function in flavonoid biosynthesis, yet the function remains to be further determined.

**Acknowledgments**

We thank Jianbo Zhang for providing the original maize material carrying $P$-del-2, and very helpful discussion in characterization of the deletion allele: Dr. Suzy M. Cocciolone for helpful advice on RT-PCR; Drs. John Imsande, Johnathan Wendel, Phil Becraft for critical reading of the manuscript. This research was supported in part by grants to T. Peterson from USDA.
References


**Figure Legends**

Figure 1. Structure comparison among $P$-homologous genes in maize and teosinte. Vertical arrow indicates transcription start site. Black boxes are three exons, and gray boxes are two introns and a 5' untranslated region. The 90 bp conserved promoter sequence is also shown in gray box. The $P_{I-rr}$ gene coding region is flanked by two long direct repeats (5.2 kbp) as indicated by horizontal arrows. Numerical boxes were previously used for the location of $P$ gene fragments (8), and used here for homologous sequence comparison. Patterned boxes also indicate homologous sequences (for homology higher than 90%). The draw is not to scale in order to show details.

Figure 2. Amino acid sequence alignment using pileup program of GCG software. Amino acids of the Myb DNA binding domain and putative activation domain are shown in bold. Boxed region contains conserved blocks of amino acids among maize $P2$, teosinte $P2-t$, teosinte clone 4, and sorghum $Yl$.

Figure 3. RT-PCR detection of $P$ gene transcripts. Panel A shows the primers location on $P$ gene. The forward primer (F) is EP5-8, or P2-5, and the reverse primer (R) is EP3-13, see material and method. In panel B, 1: $P_{I-rr}$, 2: $P_{-ww-ll12}$, 3: teosinte lltis81-5, 4: genomic DNA from $P_{I-rr}$ as control. Primers used are EP5-8 and EP3-13 that amplifies both $P1$ and $P2$ transcripts. The abbreviate pc stands for pericarp, and tg for tassel glume. In panel C, primers P2-5 and EP3-13 were used to specifically amplify $P2$ transcript. The RT-PCR agarose gel was also blotted and hybridized to probe PCR1, which covers the PCR amplified
region. The RT-PCR specifically amplified bands mentioned above (see Results) were confirmed by the hybridization results. Additionally a slightly larger band that is not seen on EtBr gel hybridized to the probe, and the hybridization band is always associated with P2 or teosinte P2-t gene expression (data not shown). Thus it could be non-spliced or alternatively spliced P2 or teosinte P2-t gene transcripts.

Figure 4. Mechanism of maize P gene duplication and the generation of P2 and Pl genes. The P gene coding region is indicated by a horizontal arrow. The 90 bp conserved promoter sequence is indicated by a gray box. The duplicated region is shown between two vertical dash lines.

Figure 5. Sequences at 3' end of the P2 gene and 5' end of the Pl-rr gene were interrupted by retroelements. A 540 bp sequence at the P2 gene 3' end, indicated by a arrow head, is presumably a partial 5' LTR of a undefined retroelement. The black arrowhead is 5' LTR sequence of retroelement Prem-2 with the typical TGTTG terminal sequence. A 620 bp sequence at the Pl-rr gene 5' end indicated by a arrow head is presumably a partial 3' LTR of the undefined retroelement with the typical CAACA terminal sequence. The 5 bp target sequence (AAGAC) duplication are shown in bold. At the top of the figure is 3' end sequence of teosinte P2-t gene. At the bottom of the figure is 3' end sequence of Pl-wr gene.
Figure 2.
Figure 3.
Figure 4.
Figure 5.
CHAPTER 3. FUNCTIONAL ANALYSIS OF THE MAIZE P2 GENE: EVIDENCE FROM GENETIC MUTANTS AND TRANSFORMATION STUDIES

A paper to be submitted to Plant Cell

Peifen Zhang, Jianbo Zhang, Sheila Maddock, Yibin Wang, and Thomas Peterson

Abstract

The maize P1 gene encodes a myb-like transcriptional factor that regulates phlobaphene pigment biosynthesis in kernel pericarp and cob glume. In addition, P1 was indicated to be a major QTL for silk maysin levels which is important in deterring silk and kernel feeding by corn earworm. A recent study indicates that there are tandem duplicated P genes. The P2 gene encodes a protein with a highly conserved myb-like DNA binding domain and putative activation domain. The P2 gene is specifically expressed in silk and anther, but not in kernel pericarp, because its upstream regulatory sequences are different from those of P1. To elucidate the functional role of the P2 gene, we analyzed its regulatory action in a maize cell culture system by a transformation approach and in the maize plant by deletion mutants analysis. Our results indicate that P2 can activate a similar subset of structural genes, as does P1. Maize Deletion mutants that remove both P1 and P2 genes produce non-detectable silk maysin level, and lost completely a correlated silk browning reaction, and therefore support a regulatory role for the P2 gene in silk maysin biosynthesis. The two duplicated P genes may have similar functions but in response to different maize developmental or environmental signals in maysin production, which may provide the maize plant a more effective protection from corn earworm damage.
Introduction

Flavonoids are plant secondary metabolic products that are widely distributed in plants. In addition to providing bright pigmentation, flavonoids help to protect plants from the effects of ultra violet light irradiation, pathogen infection, and insect feeding. Flavonoids also play important roles in plant reproduction and plant tissue reinforcement (reviewed in Koes et al. 1993). Moreover, the antioxidation properties of flavonoids may provide significant beneficial impacts in human health and medical applications (reviewed in Carroll et al. 1998; Samman et al. 1998). Long before the diverse functions of flavonoids were characterized, the pigmentation of flavonoids has provided a powerful tool for genetic analysis of gene regulation in plants. A number of structural and regulatory genes required for flavonoid biosynthesis have been isolated and characterized in several plant species, including maize (reviewed in Dooner et al. 1991). The maize PI gene has been shown to encode a transcriptional regulator for synthesis of a flavonoid pigment, phlobaphene (Grotewold et al. 1991). The red phlobaphene pigment accumulates in specific floral organs, most notably in kernel pericarp and cob glume. The PI gene also regulates synthesis of another flavonoid compound, maysin (a C-glycosyl-flavone), in maize silk (Byrne et al. 1996). Maysin plays an important role in deterring feeding by corn earworm (Helicoverpa zea) on silk and kernel. Recently, a PI gene homologue, P2, has been isolated (P. Zhang, S. Chopra, and T. Peterson, in preparation). The P2 gene is tightly linked to the PI-rr allele, and the local duplication of the two genes was estimated to have occurred approximately 2.75 million years ago. The P2 gene encodes a protein with a highly conserved myb-like DNA binding domain and acidic
putative transcriptional activation domain. Although the \( P2 \) gene appears to be functionally conserved, the \( P2 \) gene does not regulate phlobaphene pigment synthesis in kernel pericarp and cob glume. We showed before that the \( P2 \) gene is not expressed in those tissues, apparently because its upstream regulatory sequences are different from those of the \( P1 \) gene. The \( P2 \) gene is, however, expressed in silk, as is \( P1 \). In addition, \( P2 \) transcripts were also detected in developing anther. To further elucidate the functional role of the \( P2 \) gene, we characterized its regulatory action in a maize cell culture system and in the maize plant. The results from this study indicate that the \( P2 \) gene is a functional duplicate copy of \( P1 \). Both genes encode transcriptional regulators of the flavonoid pathway, and thereby contribute to silk maysin production.

**Results**

**Transformation of \( P2 \) in maize BMS cells**

Previous sequence alignment data showed that the \( P2 \) and the \( P1-rr \) genes are highly homologous (99% identity in amino acid sequence) in the myb-like DNA binding domain and acidic transcriptional activation domain (P. Zhang, S. Chopra, and T. Peterson, in preparation). This high level of identity suggests that the \( P \) gene function is selectively conserved in the \( P2 \) gene. In order to characterize the \( P2 \) gene function, we transformed a \( P2 \) cDNA plasmid construct into maize BMS cells by particle gun bombardment. The \( P2 \) cDNA was fused to a maize ubiquitin gene promoter (ubi::\( P2 \)), and co-transformed with rice actin::PAT for herbicide resistance selection. From Basta resistant callus, \( P2 \) -expressing lines
were screened by Northern blot analysis using P2 cDNA as a probe. The non-transformed BMS cells showed no P gene expression, as expected for a negative control (Figure 1, lane 3). Whereas, the P2-transformed lines showed varying levels of P2 expression (Figure 1, lanes 6-10). Additionally, a construct of Ubi::P1 was also co-transformed with actin::PAT, and individual P1-expressing lines (Figure 1, lanes 4 and 5) were selected for parallel comparisons.

A subset of structural genes for flavonoid synthesis, C2, Chi, and A1, were previously shown to be regulated by the P1 gene (Grotewold et al. 1991). Whereas the structural genes, F3H, A2, Bzl, and Bz2 encoding enzymes required for anthocyanin and flavonol synthesis, are not regulated by P1 (Grotewold et al. 1998). Figure 1 shows that in P1-expressing cell lines, II 2-5 and II 3-9 (lane 4 and 5), the C2 and A1 transcripts accumulate in proportion to the transcription level of P1. This result is consistent with previous reports, and confirms the regulatory role of P1 in vivo. Likewise, in P2-expressing lines I 9-6, I 12-6, II 15-5, and I 11-1 (lane 6 through 9), the accumulation of transcripts of C2 and A1 corresponds with levels of P2 expression. In contrast, a Ubi:: P2 transformant, line I 12-13 (lane 10), shows no P2 transcript, and also no C2, A1 transcripts as well. To test the potential role of P2 in activating other biosynthetic genes, the Northern blots were probed with a cDNA from the F3H (flavonol 3-hydroxylase) gene, which is required for anthocyanin and flavonol biosynthesis. However, no F3H transcript can be detected in either P1- or P2-expressing lines. A RNA sample from anthocyanin pigmented husk was included as a control for hybridization with F3H (lane 11). Our results indicate that, like the P1 gene, P2 can activate a similar subset of genes in the flavonoid pathway. Our data also indicate that P2 is not sufficient to activate structural genes necessary for both flavonol and anthocyanin synthesis, such as the F3H gene.
RNA from $P$-rr pericarp was also loaded in the same blot (lane 1) for positive control of hybridization. It is interesting to note that even though the amount of $P^I$ transcripts in pericarp is much lower than $P^I$ transcripts in BMS line II 2-5 (lane 4), the transcripts level of $C2$ and $A1$ are however much higher in pericarp than in BMS cell. This suggests that the efficiency of target gene activation by $P^I$ might be very different in BMS cultured cells compared to pericarp.

**Deletion mutants suggest that $P^2$ controls maysin synthesis**

Transcript analysis by RT-PCR (P. Zhang, S. Chopra, and T. Peterson, in preparation). showed that the $P^2$ gene is expressed in emerging silk and developing anther, but not in pericarp. To determine the functional role of the $P^2$ gene in the maize plant, we characterized genetic stocks containing deletions of the $P^I$ and $P^2$ genes. Deletions were obtained from the $P$-vv-$9D9A$ allele, which contains one intact and one truncated $Ac$ transposable element in the $P^I$-$rr$ gene. This configuration of $Ac$ transposable element sequences can generate chromosome deletions which start at the $Ac$ insertion site and continue upstream of the $P^I$-$rr$ gene. The deletions are caused by a non-linear sister chromatid translocation mechanism of the two $Ac$ transposons (J. Zhang and T. Peterson, in preparation). Because the deletions remove the 5' end sequences of the $P^I$-$rr$ gene, they can be visualized as colorless kernel pericarp sectors on a background of red-striped kernels, due to $Ac$ excision from the $P$-vv-$9D9A$ allele. Additionally, whole colorless ears can be obtained, depending upon the timing of development at which deletions occur. We screened more than 30 white-sectorred ears or whole white ears on genomic Southern blots, and identified ten
independent deletions of the 5' portion of \(PI-rr\) gene. One deletion, \(P_{-}\text{del}-774\), deleted the \(PI-rr\) gene, but left the \(P2\) gene intact. A second deletion, \(P_{-}\text{del}-2\), deleted the \(PI-rr\) gene and also the 3' portion of the \(P2\) gene. The remaining eight deletions extend beyond the 5' portion of \(P2\).

A restriction map of the \(PI-rr\) locus carrying \(Ac\) insertions (maize stock \(P-vv-9D9A\)) and a restriction map of the \(P2\) locus are presented in Figure 2A. A representative Southern blot analysis of deletions is shown in figure 2B. Genomic DNA was digested with \(SalI\) and hybridized with \(P\) locus fragment 15. In the \(PI-rr\) gene (lane 1), the fragment 15 probe detects a 3.0 kb band which is about 10 kb upstream of the \(PI-rr\) gene transcription start site, a 3.4 kb band derived from the 3' part of the coding region, and a 1.2 kb band containing doublet. The two 1.2 kb fragments are derived from direct repeated sequences flanking the \(PI-rr\) gene at both 5' and 3' ends. The probe fragment 15 hybridizes to all these four bands because it is repeated at four sites in the \(PI-rr\) region (Figure 2A). When the \(Ac\) transposons inserted in the intron 2 of the \(PI-rr\) locus, the 3.4 kb band will be lifted to a bigger size. Upon \(Ac\) sister-chromatid transposition, deletion events can be indicated as the loss of the 3.0 kb band that is located at the 5' flanking region of the \(PI-rr\) locus. Ten deletions were selected based on the loss of the 3.0 kb band: lanes 8 through 12 in figure 2B are examples.

Because the \(P2\) locus does not contain sequences homologous to fragment 15 (Figure 2A), hybridization with fragment 15 can not be used to detect \(P2\) deletions. Fragment 13, however, is derived from the exon 3 region of the \(PI-rr\) and \(P2\) genes (Figure 2A), and it hybridizes to a 2.0 kb band from the \(P2\) gene (Figure 2C, lanes 4-8). Deletions which extend to the \(P2\) gene can be detected by loss of the 2.0 kb \(SalI\) band. Rehybridizing the same blot
with fragment 13, several \textit{P2} deletion mutants were identified (Figure 2C, lanes 9-12 are examples; \textit{P-del-2} is in lane 9). A total of nine \textit{P2} deletion mutants were obtained in this way. One deletion (\textit{P-del-774}, in lane 8) was indicated as a non-\textit{P2} deletion.

In order to further characterize how far the nine \textit{P2} deletion mutants extend, we used PCR to amplify the genomic DNA of mutants. Primer P2-5 is located in the 5' untranslated region of \textit{P2}; primer EP3-13 is located in the exon 2 of \textit{P2}. If deletion mutations extend beyond the 5' coding region of \textit{P2}, the PCR will not produce a \textit{P2} band, at the size of 340 bp. We found eight deletion mutants are beyond the \textit{P2} gene 5' coding region, whereas one deletion (\textit{P-del-2}) mutant still produces a 340 bp amplified band (lane 4 in Figure 3) indicating that the deletion ends within the \textit{P2} gene. We then PCR amplified the deletion break point in \textit{P-del-2} using one primer (ZPF2) from exon 3 of \textit{P2}, and a second primer (Ac3) from the \textit{fAc} transposon 3' end. The sequence of the PCR product shows that the deletion ends where the 3' end of the \textit{fAc} transposon is fused to the \textit{P2} gene, at 63 bp 5' of the translation stop of \textit{P2} (Figure 4). Thus, the \textit{P-del-2} deletion removed the 5' portion of the \textit{Pl-rr} gene (upstream from the \textit{fAc} insertion site), the chromosome region between the \textit{Pl-rr} and \textit{P2} genes, and the 3' region of the \textit{P2} gene.

The deletion mutations thus obtained were characterized for their effects on expression of flavonoid biosynthesis in silk. In particular, we examined the effects of these mutations on production of maysin (a C-glycosyl flavone), by detecting the correlated silk browning phenotype. The silk browning phenotype can be scored by cutting the tip of ear shoot when silks begin to emerge from the husks. After several minutes, the cut silk end will turn from green to brown, indicating the oxidation of dihydroxyl flavone (Levings and Stuber, 1971).
QTL analysis has implicated the *Pl* gene as a major regulator of silk maysin synthesis (Byrne et al. 1996). As shown in figure 5, the *P-rr* stock which contains both *Pl-rr* and *P2* genes have strongly browning silks. In contrast, the *P-del-2* stock has completely lost silk browning. A similar non-browning phenotype was observed in independent mutants that have larger deletions than the *P-del-2* allele (data not shown). However the *P-del-774* mutant that deleted the 5' coding sequence of *Pl-rr* but left an intact *P2* coding sequence retains partial silk browning. Silks of *P-del-2* homozygous plants showed no detectable maysin accumulation (M. Snook, data not shown). The complete loss of silk browning and maysin production in the deletions which remove both the *Pl-rr* and *P2* genes indicate that *P2*, in addition to *Pl*, can induce silk browning and maysin synthesis.

Discussion

The *Pl* gene was previously suggested to regulate maysin synthesis. And the regulatory role of *Pl* on the structural genes *C2* and *A1* was clearly demonstrated in previous studies (Grotewold et al. 1994). A QTL analysis of maysin synthesis indicated that a major QTL for silk maysin levels that accounted for 58% of the variation in silk maysin levels mapped to the *Pl* locus (Byrne et al. 1996). Additionally, the expression of *Pl* in maize BMS cell cultures directly showed that *Pl* is necessary and sufficient for accumulation of C-glycosyl flavone, a class of flavonoid that includes maysin (Grotewold et al. 1998). Taking a similar BMS transformation approach in this study, we showed here that the *P2* gene is capable to activate a similar subset of structural genes as does *Pl*. Finally, the deletion mutants studied
above also show that the \textit{P2} gene is functional in the flavonoid pathway, particularly in regulating silk maysin synthesis.

Flavonol has been identified as an important flavonoid compound required for pollen viability in both maize and snapdragon (Mo et al. 1992). The biosynthetic pathway of flavonol shares some of the early steps of anthocyanin synthesis. However, the genes responsible for regulation of flavonol synthesis in maize anther have not been identified to date. The \textit{R} gene that regulates anthocyanin synthesis in anther is not required for pollen production and pollen function (Deboo et al. 1995). Likewise, deletion of the \textit{Pl} gene in the \textit{P-ww-112} line does not abolish pollen development. The recently isolated \textit{P2} gene is expressed in developing anther; \textit{P2} is homologous to \textit{Pl}, yet its deduced protein C terminal region is divergent from \textit{Pl} (P. Zhang, S. Chopra, and T. Peterson, in preparation). We intended to see whether \textit{P2} regulates flavonol synthesis and therefore pollen viability. However, the homozygous mutants that have deletions of the \textit{P2} locus showed no notable defects in anther, pollen, or seed set. The results indicate that either \textit{P2} does not regulate pollen flavonol synthesis, or that there is a redundant regulatory function in the genome. The latter seems unlikely to be the case, since \textit{P2} is unable to active \textit{F3H} transcription in BMS cells.

The expression of the duplicated \textit{Pl} and \textit{P2} genes in the same silk tissue, and their similar roles in transcriptional control of structural genes suggest that the two myb-like \textit{P} genes are functionally redundant. Similar functional redundancy of myb-homologous genes has been reported recently in the snapdragon Myb305 and Myb340 genes (Moyano et al. 1996). These genes encode proteins with highly homologous DNA binding domains and in fact they bind to and activate common structural genes in the flavonoid pathway for flavonol
synthesis in snapdragon flowers. Interestingly, the Myb340 gene is expressed early in flower development when there is a corresponding rapid increase of flavonoid gene expression. Whereas, the Myb305 gene is expressed later in flower development, when flavonoid gene expression is maintained at a moderate level. Interestingly, the Myb340 protein has a higher transcriptional activation potential, but a weaker DNA-binding activity, compared to the Myb305 protein. At later developmental stage, the Myb305 protein successfully competes with Myb340 protein for binding to the same target structural genes, and thereby switches the structural gene expression to a moderate level. Their study presented a good example of how functionally redundant genes may interact to adjust the rate of gene expression in response to different stages of plant development. A similar example came from a recent study of chalcone synthase genes in Gerbera hybrida (Helariutta et al. 1995). The gchs1 and gchs3 genes encode structurally and functionally similar chalcone synthase proteins. The gchs3 expression correlates with early synthesis of flavonols, whereas the gchs1 expression correlates with late synthesis of both flavonols and anthocyanins. Similarly, the PI and P2 genes may have similar functions in regulating flavonoid synthesis but may response to different maize developmental or environmental signals.

Materials and Methods

Maize materials

The P1-rr-4B2 allele was derived from P-vv by excision of Ac. The P-ww-1112 allele was derived from a deletion of the P1-rr coding sequence as a result of homologous recombination
of direct repeated sequences flanking the \textit{Pl-rr} coding region (Athraa and Peterson, 1991).

The \textit{P-vv-9D9A} allele has an intact \textit{Ac} transposon and a fractured \textit{Ac} (fAc) inserted in the \textit{Pl-rr} locus, as described in J. Zhang and T. Peterson, in preparation. The deletion mutants studied here were derived from \textit{P-vv-9D9A}. The stocks containing the \textit{Pl-rr-4B2}, \textit{P-ww-1112} and \textit{P-vv-9D9A} alleles also contain the \textit{P2} gene tightly linked to the \textit{Pl} locus.

**Genomic DNA isolation, Southern blot and PCR analysis of mutants**

Young leaf of individual plants was ground in liquid N2, and genomic DNA was extracted using CTAB reagent. Genomic DNA was digested with \textit{SalI} for at least 6 hours, and electrophoresed through 0.8\% agarose gel, and transferred to nylon membrane according to the method of Southern (Sambrook et al. 1989). Membranes were prehybridized for at least 1 hour at 65 °C in 25mM NaH2PO4 (pH 7.2), 7\% SDS, and hybridized in the same solution containing probe overnight at 65 °C. Membranes were then washed in 25 mM NaH2PO4, 1\% SDS, at 65 °C for 30 min twice, and exposed to Kodak film. Primers used in Genomic DNA PCR amplifications were EP5-8 (5' ACGCGCGACCAGCTGCTAACCGTG 3') and EP3-13 (5' AGGAATTCCGCCGAAGGTAGTTGATGC 3'), or ZPF2 (5' TGACGCCCTGATGCTTAT 3') and Ac3 (5' GGATTCGTTTTCTGTTACCGGTATATC 3').
**P gene expressing plasmid construction**

Plasmid pF2cDNA containing a full length *P1-rr* cDNA in a pBluescript vector was from E. Grotewold (Grotewold et al. 1998). The *PstI/XhoI* fragment of pF2cDNA containing the 3' part of *P1-rr* cDNA was replaced by a *PstI/SalI* fragment from a *P2* genomic clone to generate plasmid pP2cDNA, which encodes a protein identical to that encoded by the native genomic *P2* gene. The Ubi::P1 plasmid was constructed by ligating a *BamHI/KpnI* fragment of pF2cDNA to *BamHI* linearized pAHC17 (Christensen et al. 1992) using a *BamHI/KpnI* linker (5'-GATCGTAC-3'). The pAHC17 contains the maize ubiquitin promoter, a NOS (nopaline synthase) gene 3' untranslated sequence and polyadenylation signal. To construct Ubi::P2, the *P2* cDNA insert from pP2cDNA was PCR amplified using primer T7 and T3, and cloned to a A/T vector pT7Blue T-vector (Novagen). The resulting clone was digested with *BamHI*, and the *BamHI* fragment containing the *P2* cDNA insert was then ligated into pAHC17. The cloning junctions of all constructs were confirmed by sequencing, and the *P2* cDNA insert in Ubi::P2 was sequenced to rule out any PCR-introduced errors. Plasmid constructs were sequenced with gene or vector specific primers using the Applied Biosystems fluorescent sequencing system at the Iowa State University Nucleic Acid Facility.

**Transformation**

Cultured maize BMS (Black Mexican Sweetcorn) cells were transformed by microprojectile bombardment using a Bio-Rad particle delivery gun PDS-1000/He (Bio-Rad). Approximately 0.5 ug of Ubi::P1 or Ubi::P2 with 0.5 ug of Actin::PAT were used per bombardment. PAT-resistant callus were selected approximately 5 weeks later on selection media, and maintained
on selection media. Cell culture conditions were as described previously (Grotewold et al. 1998).

RNA isolation and Northern blot analysis

Kernel pericarps were dissected from 20 DAP (day-after-pollination) kernels, frozen in liquid N2, and ground with a mortar and pestle. Total RNA was extracted in Trizol reagent (GibcoBRL). BMS callus was ground in Trizol reagent for RNA extraction. Approximately 20-30 μg total RNA was loaded per lane on 1% agarose gel containing 18% formaldehyde. The gel was blotted to a nylon membrane, and hybridized with probes of: P2 cDNA (see materials and methods at above), C2 cDNA (Wienand et al. 1986), A1 cDNA (Schwarz-Sommer et al. 1987), and F3H cDNA (Deboe et al. 1995).

Acknowledgments

P. Zhang was supported by a predoctoral fellowship from Pioneer Hi-bred International, Inc.

References


Figure Legends

Figure 1. Northern blot of transformed BMS cell lines. Lane 1 and lane 2 are RNAs from pericarp of P-rr as a positive control and of P-ww as a negative control, respectively. Lane 3 is non-transformed BMS cell. Lanes 4 and 5 are PI transformants, and lanes 6 through 10 are
P2 transformants. Lane 11 is RNA from anthocyanin pigmented husk, as a positive control for probe F3H. Probes used are indicated to the right of each panel.

Figure 2. A: Chromosome organization and the gene structure of P-vv-9D9A and P2. Arrow heads indicate transcription start for P2 and P1-rr genes. Black boxes indicate the three exons of P2, and gray boxes indicate the three exons of P1-rr. SaI sites are indicated as S. SaI fragments which hybridize with probes 13 and 15 are indicated. Panel B: Southern blot analysis of genomic DNAs digested with SaI and hybridized with probe 15. Deletion mutants are selected based on the absence of the 3.0 kb band. Panel C: the same blot was rehybridized with probe 13. Deletions that extended to the P2 locus are indicated by the absence of the 2.0 kb band. Lane 1 through 3 are controls: 1. P-rr-4B2; 2. P-wr-w23; 3. P-ww-1112.

Figure 3. Genomic PCR of all ten deletion mutants selected. Lane 2 is P-vv-9D9A as a control showing that the P2 gene 5' end is specifically amplified to a 340 bp band. P-del-774 in lane 3 is a deletion that only deleted P1-rr, but does not extend to the P2 locus. P-del-2 in lane 4 is a P2 deletion that indicated on Southern blot (lane 9 on figure 2C). However PCR indicates that the deletion does not extend to the 5' end of P2.

Figure 4. Gene structure of the two particular deletion mutants, P-del-774 and P-del-2. Special labels are the same as in figure 2A.

Figure 5. Silk browning phenotype of the two deletion mutants P-del-774 and P-del-2 compared to wild type P-rr that carries both intact P2 and P1-rr genes.
Figure 1.
Figure 2.
Figure 4.
Figure 5.
CHAPTER 4. CELLULAR DISTRIBUTION OF PHLOBAPHENE PIGMENT AND TRANSCRIPTS OF THE REGULATORY GENE PI IN MAIZE KERNEL

Introduction

The pericarp is the outer most layer of the maize kernel. It functions as a physical barrier to penetration by pathogens, and also protects the kernel from other stress and damage from the outside world. The thickness of the pericarp also affects the drying down rate of mature kernels, and is an important factor in determining the tenderness and thus the quality of sweet corn (Bailey and Bailey, 1938; Gaessler et al., 1940).

The pericarp is derived from the ovary wall after pollination. It usually contains about 10-20 layers of cells (Figure 1D); however, the number of cell layers varies among different maize lines (Tracy and Galinat, 1987). The outer most cell layer is epidermis, and the internal layers are termed sub-epidermal layers. Terms of outer- and inner- part of pericarp are also used in describing pericarp morphology (Haddad, 1931). Cells in the outer half of pericarp are much smaller and more elongated. Their cell walls are thicker, and the cells are tightly arranged. Cells in the inner half of pericarp are larger and round shaped. Their cell walls are thinner, and the cells are more loosely arranged. Starting at 10 days after pollination, there is a clear zone dividing the pericarp between the outer half and the inner half layers. Cells in this central zone region begin to disorganize and collapse, and the degeneration of cells continues toward the inner half of pericarp. From 20 days after pollination until maturity, cells in the outer half of pericarp remain intact, and the cell walls increase in thickness. However, cells in
the inner half of pericarp continue to degenerate and are further compressed. At maturity, the pericarp consists of mostly the external half of its cell layers. The much thickened and extensively lignified cell walls are closely appressed and form a firm protective covering for the seed (Haddad, 1931; Randolph, 1936).

The kernel pericarp can be pigmented with a red phlobaphene compound produced in the flavonoid pathway (Styles and Ceska, 1977). The biosynthesis of phlobaphene is tissue-specifically controlled by a regulatory gene PL (Pigmented pericarp). The PL gene encodes a myb-homologous transcription factor that transactivates the expression of a set of structural genes encoding biosynthetic enzymes in the phlobaphene biosynthesis pathway (Grotewold et al. 1991). The P-rr gene produces red kernel pericarp, whereas the P-vv allele, which has an Ac transposon associated with the PL-rr gene, produces red stripes on the pericarp. In earlier studies of pericarp pigmentation by Emerson, two derivatives from striped pericarp were investigated (Emerson, 1917). The near-self type of kernel is wholly red except at the kernel crown, the top part of the kernel around the silk attachment region. In contrast, the dark-crown type of kernel is very lightly pigmented except at the crown region, which is deep red pigmented. Emerson found that the near-self pattern is heritable, but the dark-crown pattern is not. Pigment distribution at the cellular level of the pericarp revealed that sub-epidermal layers were pigmented in the near-self type of kernel (Figure 3A), whereas only the epidermal layer was pigmented in the dark-crown kernel (Figure 3B). To explain the relationship between pigment distribution and the pattern of heritability, Emerson proposed the two-cell lineage hypothesis, in which epidermal and sub-epidermal cells are derived from two distinct cell lineages that were separated early in development. The changes from striped pericarp to near-
self are associated with mutations in sub-epidermal layers that are in the same lineage as the egg cell, and thus are heritable. The changes from striped pericarp to dark-crown are associated with mutations only in epidermal cell layer that do not affect the reproductive cells, and therefore would not be transmitted to the progeny.

Following Emerson's study, Randolph further investigated kernel morphology and phlobaphene pigment distribution (Randolph, 1926). Randolph's results were not in complete agreement with Emerson's hypothesis. She found that, in the near-self type, pigment was not restricted to sub-epidermal layers, as some regions of the epidermis were also pigmented (Figure 3C). Whereas, in the dark-crown type, not all the regions of epidermis are pigmented (Figure 3D). Randolph stated that an assumption must be added to the previous hypothesis, in that the effect of a mutation may extend to surrounding cells so that a pigmentation pattern may be formed by a group of genetically different cells. In this assumption, for example, the pigment found in some epidermal cells of the near-self type might be transported from sub-epidermal cells. In other words, the pigment may not be cell autonomous. However, this assumption could not be tested at that time without an appropriate molecular approach. Now that the $P$ gene has been isolated, we applied in-situ hybridization techniques to $P-rr$ pericarp and asked whether the phlobaphene pigmentation is cell autonomous.

We also compared the cell-type specific expression of two alleles of the $P$ gene which give differential pericarp pigmentation. $P-rr$ gives both red pericarp and cob, while $P-wr$ gives red cob, and colorless pericarp. Sequence comparison of the two alleles revealed that they are nearly identical through out the coding region and more than 5 kb of the 5' flanking region. The major difference between the two alleles is a short chromosome rearrangement near the
genes' 3' end (Chopra et al. 1996). It is not clear what causes the different pigmentation patterns of the two $P$ alleles, in other words, why the $P$-wr allele does not pigment pericarp. It was shown by Northern blot analysis that colorless $P$-wr pericarps do contain $P$ transcripts, although at lower levels than in $P$-rr pericarp (Chopra et al. 1996). Here, we address the question of whether expression of the $P$-wr allele is in a subset of pericarp cells that may not support phlobaphene synthesis. We first investigated the phlobaphene pigment distribution pattern in kernel pericarp of plants carrying a $P$-rr allele. We then detected the location of $P$-rr and $P$-wr transcripts in pericarp by in-situ hybridization.

Finally, an improved practical protocol for maize kernel paraffin embedding and sectioning was presented in this study.

Results and Discussion

Cellular morphology and phlobaphene pigment distribution in pericarp

Maize kernels with $P$-rr genotype have fully pigmented red pericarp. The red pigment starts to develop at 10 – 12 days after pollination, and is first seen at the silk attachment region. The pigment then spreads over the crown (top of kernel) and continues to develop towards the gown (sides of kernel), and gradually fades at the base of kernel. At 20 – 24 days after pollination, most of the pericarp is red pigmented. The red phlobaphene pigment darkens when the kernel becomes mature and dry. On longitudinal sections of kernels at 20 days after pollination, the red phlobaphene pigment was observed in all the cell layers of the pericarp at both crown and gown regions. In the crown region (Figure 1E), the outer part of cell layers
are composed of smaller, tightly arranged cells. Intensive red pigment was found evenly distributed within the cells (indicated as an arrowhead in Figure 1G), or occasionally appeared as small particles (indicated as an arrowhead in Figure 1G). Less intensive pigment (pink color) was located within cell walls and intercellular spaces. Cells of the inner part layers are larger, round shaped and loosely arranged comparing to those of the outer part layers. Here, red pigment was less intense and found mostly in the cell walls and intercellular spaces. In the gown region (Figure 1F), intense pigment was observed in epidermal cells. For the remaining cell layers, the pigment was less intense, and mostly along the cell walls and intercellular spaces. Cells in the gown region appeared larger and more elongated than cells in the crown region of pericarp. However, the elongated cell shape also appears in the crown region, as more clearly viewed in scanning electronic microscopy (Figure 1C).

The distribution of natural phlobaphene pigment, produced under P1 gene regulation in maize kernel pericarp cells, indicates that subcellular localization of phlobaphene differs from that of anthocyanin pigments. The red phlobaphene pigment was variously located in the cell lumen, small particles, which could be vesicles as reported in another study (Grotewold et al. 1998), cell wall, and intercellular spaces. In contrast, anthocyanin is solely deposited in central vacuole (Marrs et al. 1995). Our results agree with a recent study by Grotewold and colleagues, in which subcellular distribution of flavonoids was investigated by ectopic expression of P1 gene in cultured BMS (Black-Mexican-Sweet corn) cells (Grotewold et al. 1998). Two types of autofluorescent bodies were found in P1 expressing cells at different subcellular locations: large olive-green bodies were found filling most of the cell lumen, whereas small greenish orange bodies were localized to spherical vesicles. The latter were
proposed to fuse to the cell membrane and empty their vesicle contents to cell wall. Thus the
$PI$ gene regulated compounds appeared to be transported via different trafficking pathways,
in sharp contrast to the vacuolar localization of anthocyanins. In addition, our observation
shows that the cell lumen location of the pigment is mostly restricted to the outer layer of
pericarp cells. Whereas, pigmentation in the inner cell layers is mostly located in the cell wall
and intercellular spaces, possibly as a result of small vesicle deposition.

The $P$-$vv$ allele has an $Ac$ transposon inserted in the $PI$ locus that disrupts the $P$ gene
function (Lechelt et al. 1989). Upon $Ac$ excision, the $P$ gene function can be restored.
Transposition of the $Ac$ transposon during kernel development will give rise to red stripes on a
white kernel background (Figure 1A). A single red stripe is derived from a single event of $Ac$
excision, thus cells in the stripe are from the same cell lineage. Therefore, the red pigment
pattern distributed in cell clones can be informative about pericarp development. However, the
nature of the cell-lineage relationships during pericarp development has been controversial.
Greenblatt (1985) proposed that the pericarp is composed of both tunica and corpus lineages.
The tunica is thought to be formed in early embryogenesis, and gives rise to the epidermal
layers in all organs. Within the tunica is the corpus, which gives rise to the sub-epidermal
tissues including the megaspore-mother-cell. The two-cell lineage idea is supported by
cytological study: it was once observed that epidermis cell divisions are anticlinal, so that
epidermis cells give rise to only epidermis cells, but not to underlying cells (Randolph. 1926).
If the so-called two-cell lineage hypothesis of pericarp development is true, in a single red
stripe on $P$-$vv$ kernel, one should not expect to see pigment in both epidermal and sub-
epidermal layers in the same sector. Cross sections of $P$-$vv$ kernels (Figure 4B) revealed that
the red pigment distribution within pericarp layers is different for individual stripes. For some red stripes, pigment was restricted only to the outer cell layers, while for some other red stripes, pigment was restricted to inner layers. Most importantly, in some sectors the pigment was found in both epidermis and sub-epidermis cells within a single stripe. These latter observations of pigmentation patterns in $P$-$vv$ kernels are not in accord with the two-cell lineage hypothesis. An alternative explanation could be that the pigment distribution may not reflect cell lineage boundaries (Randolph, 1926), i.e., that the pigment may diffuse between epidermal and subepidermal cells. This possibility was tested using in situ hybridization analysis of $P$-$vv$ pericarp (see below).

**In-situ detection of $P$ gene expression**

The maize $P$-$rr$ allele pigments kernel pericarp and cob glumes, while a highly homologous $P$-$wr$ allele pigments only cob glumes, but not pericarp. Northern blot analysis of pericarp RNA from $P$-$wr$ kernels 22 days after pollination showed that $P$ gene transcripts were present in pericarp, though at about 30% of the amount in $P$-$rr$ pericarp. One possible explanation is that the level of $P$-$wr$ transcripts is below a threshold level required for pigmentation. However, this explanation seems unlikely, based on the observation that $P$-$wr$ cob glumes are intensely pigmented, yet they contain only 15% the level of transcripts of $P$-$rr$ cob glumes (Chopra, 1996).

A possible explanation for why $P$-$wr$ does not pigment pericarp even though $P$ transcripts are present is that $P$-$wr$ is expressed in a restricted region of pericarp cells which are
not competent to produce or accumulate pigment. To address this question, we used RNA in-situ hybridization to examine P gene expression at the cellular level.

In P-rr kernel sections cut longitudinally and hybridized to P gene antisense probe, signals of P gene transcripts were detected specifically in pericarp, but not in endosperm or embryo (Figure 2A). This is consistent with previous Northern results (Grotewold, 1991), and confirms that the P gene is tissue-specifically expressed in pericarp. The P gene expression signals were detected in the crown and gown regions of kernel, and in both epidermis and sub-epidermis layers. No distinct region for P gene expression in pericarp were observed. For cells in outer layers of the pericarp that are smaller and tightly arranged, signals are more intense. Cells in the inner layers that are larger and loosely arranged had weaker signals (Figure 2E). This corresponds with the observation in this study that pigment is more intense in cells of outer layers than in inner layers. In the larger and round shaped and loosely connected inner layer cells, red pigment was only associated with cell walls, and the cell lumen appeared empty. Possibly, the inner layer cells may already be in an inactive physiological stage, and thus give a weaker P gene expression signal. The overall P-rr signal distribution in pericarp is consistent with the phlobaphene pigmentation pattern, and the results show that the P transcripts are not restricted to a specific cell layer.

In a parallel in-situ hybridization experiment using sections of P-wr kernels, no P-wr transcript signals could be detected above background (Figure 2B). Although we could not identify in which cell types P-wr is expressed, the fact that all pericarp layers are competent to produce pigment as shown in the P-rr in situ hybridization argues against the aforementioned model.
As noted above, the threshold explanation is difficult to reconcile with the observed cob glume pigmentation induced by \( P\text{-wr} \). Recently, evidence for transcriptional and post-transcriptional control mechanisms possibly involving tissue-specific silencing have been presented (Chopra et al., 1996; Chopra et al., 1998).

In-situ hybridization was also applied to cross-sections of \( P\text{-vv} \) kernels. In one section, two red stripes appeared close to each other (Figure 4B). The red phlobaphene pigment was distributed in outer pericarp layers including epidermis in one stripe, and the inner layers of the other stripe. Interestingly, \( P \) gene expression signals were detected specifically in the red striped regions of pericarp. In the single red stripe having pigment in the outer pericarp (4D), including epidermis, signals were observed in the same pattern that also included epidermis (Figure 4F). These observation strongly suggest that pigmentation is cell autonomous, at least in the regions of the pericarp examined here. Thus the pigmentation pattern is a real reflection of the expression state of the \( P \) locus.

The results of our study can be summarized as follows: first, the localization of phlobaphene pigment in cell lumen, small particles, and cell walls indicates at least two different trafficking pathways for deposition of phlobaphene pigment. Second, all cells in the pericarp are competent to produce pigment, and hence the colorless phenotype of \( P\text{-wr} \) pericarp is not explained by expression in non-competent cell types. Thirdly, the co-localization of phlobaphene pigment and \( P \) transcripts in maize kernel pericarp suggest that the pigmentation phenotype is cell autonomous. And fourth, observations of pigmented sectors indicate that the epidermal and sub-epidermal cells of the pericarp are not derived
exclusively from separate cell lineages as predicted by the two-cell lineage model. Or, alternatively, there might be multiple epidermal cell layers.

Materials and Methods

Maize lines

P-rr-4B2 carries P-rr allele: P-wr (w23) carries P-wr allele: P-ww-1112 is a deletion mutant of P-rr allele: P-vv has a Ac transposon inserted in the P-rr allele.

Sample preparation

Kernels at 20 Day-After-Pollination (DAP) were collected. Two slices of longitudinal cuts were made at the sides to expose the inside of kernel. Fixation and embedding were according to an improved method described below. Kernels were finally embedded in paraplast (56 °C melting temperature, from Fisher). Sections were cut at 8 μm using a microtome, and collected in a drop of adhesive solution (freshly diluted from a 10x stock containing: 2% polyvinyl alcohol, 0.2% vinyltriethoxysilane, 0.05% methyl-4-hydroxybenzoate, ref). The sections were either stained or viewed directly for the natural red phlobaphene pigment. Sequential sections from the same cutting were used for in-situ hybridization.

An improved method of maize kernel paraffin sectioning

Paraffin sections of a variety of tissues are often used for cellular morphological studies or in-situ RNA hybridizations. Preparation of paraffin embedded samples includes fixation,
dehydration, paraffin (wax) infiltration and embedding. For many tissues, such as leaf, stem, petiole, bud, embryo, and anther, a standard protocol generally produces satisfactory sections. When the same protocol is applied to maize kernels, the section quality is very low in our hands, resulting in dissociation of the pericarp and endosperm, and distortion of the pericarp cells, peeling of the pericarp. In this study we developed an improved protocol that reliably produces high quality maize kernel sections for in-situ hybridization and morphological study.

**Choice of fixative**

Fixation of tissue is critical in order to preserve cell components including macromolecules, and retain the natural morphology of the cell/tissue. There are two most commonly used fixatives for in-situ hybridization of plant tissues. One has an ethanol-acetic acid formula that nonspecifically precipitates macromolecules such as RNA. The other fixative uses aldehyde to cross-link RNA to protein. In general, the aldehyde fixative results in better preservation of cellular morphology, and superior RNA retention and hybridization (Raikhel et al. 1989). Paraformaldehyde was shown to be a good fixative in my protocol.

**Choice of solvent in dehydration**

Ethanol is the most commonly used dehydration solvent, followed with xylene as a transition solvent to wax infiltration. However, this produces excessive brittleness in maize kernel sectioning. As an alternative, dioxane has been shown to be ideal as both a dehydrant and wax solvent (Sass, 1945). But due to the cost and the toxicity of dioxane, butyl alcohol was suggested as an alternative. Tertiary butyl alcohol was successfully used in this protocol.
Extended time for each step in sample preparation

Dehydration and wax infiltration require gradual changes of solutions to avoid tissue damage due to excessive shrinkage or swelling. For most tissues, thirty minutes time intervals in each change of dehydration solution is considered sufficient. For tough tissues like maize kernels, extending the period of time for each step of sample preparation was found to be extremely important.

Protocol

1. Harvest an ear, remove husks, and dissect kernels from the cob. Make a longitudinal cut on each side of the kernel to expose the inner part of kernel, making it more accessible to solutions. Remove cob glumes, which can cause the kernel to float on the surface of solutions. Handling of the kernel must be done with care. The slightest pressure on kernel will result in peeling of the pericarp, disintegration of the endosperm, and loss of the embryo during the subsequent processes (Sass, 1945).

2. Immediately drop the kernel into freshly made fixative. 4% paraformaldehyde in phosphate buffered saline. Approximately 5 kernels can be placed in a 10 ml fixative vial. A vacuum can be applied to remove air from the tissue and help the kernels sink to the bottom of the vial. Vacuum should be applied and released gradually. The fixative is then changed, and fixation is continued at 4 °C overnight.

3. Pour off the fixative, and wash twice with phosphate buffered saline for 10 min.

4. Start dehydration in 10%, 25%, 35%, 50% ethanol. 4 hours for each step at room temperature.

6. Change to pure TBA, 4 hours, twice. Keep sample on a warmer plate, because TBA will crystallize at room temperature.

7. Pour off 1/3 of the TBA, fill with melted paraffin, and keep in a 60 °C oven for 1 day. Gently swirl the vial in the middle of time interval to help homogenize the solution.

8. Pour off ½ of the mixture, fill with paraffin, and incubate at 60 °C for 24 hours. Then repeat with pure paraffin.

9. Change paraffin for 2 more times. Then cast the paraffin infiltrated sample into a mold.

This protocol was used for paraffin embedding kernels ranging from 7 to 24 days after pollination. Sections were cut at 8 or 10 μm. Pre-chilling the wax block and soaking the exposed tissue surface in water (DEPC H2O for in-situ hybridization) is very helpful in sectioning. Good sections were routinely obtained using this protocol (Figure 2B and 2D).

**RNA probe preparation**

To prepare antisense and sense probes, a 300 bp PCR amplified P gene (P-wr allele) cDNA fragment was cloned into pT7 vector (Novagen) in both directions (Provided by Dr. Surinder Chopra). The primers used for PCR amplification were EP5-15: 5’-TGCCCTCGAAGGGCTT-3’, and SC002: 5’-CTAGCGGCGTCGACCATA-3’. The PCR product contains the 3’ end of P-wr cDNA, and can hybridize to both P-rr allele and P-wr allele because of the high homology between the two alleles. The plasmid clones pWRP59 and
pWRC70 were linearized by BamHI digestion, and in-vitro transcribed using T7 polymerase (Promega) and labeled with 35S-UTP for antisense and sense probes respectively. The condition for in-vitro transcription was from Raikhel et al. 1989. The probes were then purified by passing through a P-60 gel (Bio-Rad) column prepared in a 3ml syringe barrel.

In-situ hybridization

ISH was done according to a protocol modified from Raikhel (Raikhel et al. 1989). The slides were pretreated with 0.2 M HCl for 20 min at room temperature, followed by proteinase K (1 ug/ml) treatment for 30 min at 37 °C, and finally an acetic anhydride (0.5% in 0.1 M triethanolamine) treatment for 10 min at room temperature. About 5-10 x 10^5 cpm denatured probe in 30 ul hybridization solution (0.3 M NaCl, 10 mM Tris pH 6.8, 10 mM Na2HPO4 pH 6.8, 5 mM EDTA, 50% formamide, 10% dextran sulfate, 10 mM DTT, 1x Denhart's, 500 ug/ml tRNA, 500 ug/ml polyA RNA) was applied to each slide. The slide was covered by a piece of parafilm, and hybridization was carried out at 65 °C overnight in a humid chamber. Slides were washed as follows: in 50% formamide, 2 x SSC, washing for 3 hours at 65 °C, followed by RNase (20 ug/ml) treatment at 37 °C for 30 min, and a final wash for 1.5 hours at 65 °C. After drying, the slides were coated with Kodak NTB-2 emulsion, and exposed at 4 °C for 5-12 days.
Acknowledgments

I thank Ms. Marianne Smith for very helpful assistance in in-situ hybridization; Dr. Bruce Wagner for advice and assistance in Scanning Electronic Microscopy; Dr. Harry Horner for very helpful advice in paraffin embedding and sectioning; Dr. Surinder Chopra for providing the plasmid clone used in RNA probe preparation; and Dr. Suzy Coccio lone for helpful discussion.

References


Haddad, E.S. 1930. Morphological development of sweet corn pericarp in two inbred lines and their F1 hybrid. A Ph.D. thesis of Haddad, Purdue University.


Randolph, F.R. 1926. A cytological study of two types of variegated pericarp in maize. Cornell University Agricultural experiment station.


Figure Legends

Figure 1. Anatomy of maize kernel and phlobaphene pigment distribution in kernel pericarp.

1A: whole kernel view of $P$-$vv$ showing red stripes on pericarp. 1B: a longitudinal section of paraffin-embedded maize kernel (20 DAP, day-after-pollination) prepared using the improved protocol presented in this study, stained with Hemalum. En, endosperm; Em, embryo; Pc, pericarp. 1C: scanning electron microscopy of a maize kernel (20 DAP) showing surface of kernel at silk attachment region. Note the elongated cell shape. 1D: higher magnification of the Hemalum stained paraffin section showing cellular structure of the pericarp. 1E: red phlobaphene pigment distribution in pericarp cells at the crown region. 1F: phlobaphene pigment distribution in gown region. 1G: enlarged view of 1E showing details of pigment location. Arrow heads indicate pigment location in cell lumen, or in small particles.

Figure 2. In-situ hybridization of $P$ transcripts in kernel pericarp (20 DAP). 2A: antisense probe detection of $P$ transcripts in $P$-$rr$ kernel. 2B: $P$-$wr$ kernel hybridized with antisense probe. 2C: $P$-$ww$ kernel hybridized with antisense probe as a negative control. 2D: sense
probe hybridized to $P$-$rr$ kernel. 2E: higher magnification of 2A. 2F: higher magnification of 2D.

Figure 3. Previous investigation of phlobaphene pigment distribution in pericarp. 3A and 3C are near-self type kernels. 3B and 3D are dark-crown type kernels (Figure is from Randolph 1926).

Figure 4. Pigment localization in $P$-$vv$ (20DAP) red stripes, and in-situ hybridization of $P$-$vv$:
4A. $P$-$vv$ kernel; 4B. cross section of red striped region; 4C. antisense probe hybridization of 4B; 4D. enlarged view of 4B, showing one stripe that pigments the outer layers of pericarp;
4E. DIC (interface contrast device) view of 4D; 4F. enlarged view of antisense probe hybridization of the area shown in 4D; 4G. a non-pigmented area adjacent to 4F.
Figure 1.
Figure 4.
CHAPTER 5. GENERAL SUMMARY

The bright pigments produced in the flavonoid biosynthetic pathway make it an ideal system for genetic analysis of gene structure, function, and the regulation of gene expression in plants. Moreover, flavonoid compounds have a number of important biological functions in plants: hence, knowledge regarding flavonoid biosynthesis and accumulation may have important practical applications in plant improvement.

The structural genes, which encode enzymes that catalyze steps of flavonoid synthesis, are temporally and spatially regulated, and the specific mechanism for their regulation has served as a model for other plant genes. Expression of the structural genes is controlled by two major families of regulatory genes: the myb-homologous transcriptional factor $C_l/P_l$ genes. function together with $R/B$ genes that encode co-transcriptional factors with a helix-loop-helix domain. It was shown that the generation of tissue-specific anthocyanin pigmentation patterns results from differential expression of duplicated regulatory genes. For example, the $C_l$ and $P_l$ genes, as well as the $R$ and $B$ genes, are duplicated genes. The $C_l$ gene typically pigments the aleurone of the maize kernel, whereas the $P_l$ gene pigments most parts of mature plant tissues. The $R$ gene pigments aleurone, embryo scutellum, anther, coleoptile, and seedling leaf tip, whereas the $B$ gene pigments mature plant tissues including leaf sheath and husk. In addition, the duplicated $R$ genes, known as $Lc$ and $Sn$ confer distinct patterns of anthocyanin pigmentation to scutellar node, root mesocotyl, and leaf blade. Novel pigmentation patterns can also result from differential expression by different alleles of regulatory genes. For example, plants with the $B-I$ allele have dark purple husks, leaf sheaths.
and tassel glumes, while plants with the B-peru allele have moderate pigmentation of tassel glume, and strong pigmentation of kernel aleurone. The wild type PI gene evenly pigments plant tissues, but not aleurone. In contrast, Pl-Bh allele gives blotched plant pigmentation pattern, and additionally extends pigmentation to the kernel aleurone. The mechanism for different tissue-specificity of the regulatory genes listed above involves different regulatory sequences in the case of B alleles, and altered methylation conditions of the genes, in the case of PI alleles.

In contrast to the anthocyanin pathway, regulation of the phlobaphene pathway seems simple. The myb-homologous PI gene is the only known regulatory gene that is necessary and sufficient for phlobaphene pigment accumulation in maize floral organs, most notably the kernel pericarp and cob glume. The report of this thesis work extends our understanding of the molecular evolution of the PI gene and its paralogous copy, the P2 gene. First, the isolation and characterization of the P2 gene (in Chapter 2) provide precise molecular evidence that the P region contains duplicated genes, and that the duplicated PI (Pl-rr allele in here) and P2 genes have distinct tissue-specific expression. The PI gene is expressed in pericarp, cob glume, tassel glume, and silk, whereas the P2 gene is expressed in silk and developing anther, but not in pericarp, cob, and tassel glume. The two genes have highly conserved coding regions but very diverged 5' regulatory sequences. The diverged 5' regulatory sequences are proposed to confer their different tissue-specificity. Second, the isolation and characterization of a single copy P-homologous gene (P2-t) from teosinte (in Chapter 2), and gene structure and sequence comparisons among the PI, P2, and P2-t genes support a model for generation of the PI/P2 complex by gene duplication followed by
retroelement insertions (Figure 4 of Chapter 2). The $P2$ and $P1$ genes were originally tandemly (head-to-tail) duplicated from a single copy ancestral $P$ gene (resembling the teosinte $P2-t$ gene). The duplication included the $P$ gene coding region and 3' flanking region. This tandem duplication thus placed a $P$ gene coding sequence immediately following $P$ gene 3' flanking sequence. In this single event, a new regulatory pattern was generated for the $P1$ gene. Subsequent to the gene duplication event, retroelement insertions (including $Prem-2$ and $Opie-2$ retrotransposons) separated the two genes. The retroelement insertions truncated the $P2$ gene 3' flanking region, whereas the $P1$ gene retained the flanking sequence duplication at the 5' and 3' ends. This additionally explains the origin of the unusual $Pl-rr$ gene structure, in which the $P$ coding sequence is flanked by direct repeated sequences. Third, this thesis presents evidence regarding the biological function of the $P2$ gene (in Chapter 3). By screening maize materials for deletion mutants generated by transposon-mediated chromosome deletions. deletions of $P1$, or $P1$ and $P2$ were obtained and characterized. The mutants were examined for the occurrence of phenotypic changes, especially in silk and anther where the $P2$ gene is expressed. It was noted that complete loss of silk browning is coincident with the deletion of both $P1$ and $P2$ genes. Silk browning was previously reported to be closely correlated to maysin synthesis, a flavonoid compound that inhibits corn earworm feeding. A previous study showed that 58% of the maysin QTL trait maps to the $P1$ locus, suggesting that the $P1$ gene regulates maysin synthesis in silk. The results presented here indicate that the $P2$ gene is another contributor to maysin synthesis. The regulatory function of the $P2$ gene was further characterized using a transgenic approach. A construct containing the $P2$ cDNA driven by the constitutive promoter of a maize ubiquitin gene was transformed
into cultured maize BMS cells. The transcriptional activation of the structural genes C2 and A1 by the ectopically expressed P2 gene was demonstrated by Northern blot analysis. The results support a regulatory role for P2 in flavonoid (maysin) biosynthesis. The redundant functions of the duplicated P genes in silk maysin synthesis may be temporally or developmentally regulated. For example, the P1 and P2 genes may be differentially activated in response to developmental or environmental signals, and thus both genes may be important for effective earworm antibiosis. Further investigation of the similarities and differences in expression of the P1 and P2 genes would be informative. Finally, the cellular expression pattern of two P1 alleles, P1-rr and P1-wr, in maize kernel pericarp was compared using in-situ hybridization (in Chapter 4). The results indicate that the pigmented pericarp vs. non-pigmented pericarp in P1-rr and P1-wr alleles, respectively, is not due to expression of the two alleles in different cell layers of pericarp. Rather, the lack of pigmentation in P1-wr allele may be due to the overall low expression of P1-wr, or to an unknown mechanism of tissue-specific silencing. Additionally, in-situ hybridization of P gene transcripts and microscopy of phlobaphene pigment distribution in pericarp of P-vv suggest that the phlobaphene pigmentation conferred by the P gene is cell autonomous. Microscopic examination of cell clones marked by the presence of phlobaphene pigment has proven informative with regard to the development of the pericarp. The observation that some cell clones include both epidermal and sub-epidermal cells is not in accord with a previous hypothesis involving separate and distinct cell lineages for epidermal and sub-epidermal cell types.
ACKNOWLEDGMENTS

First of all, I would like to thank my major professor Dr. Thomas Peterson for his excellent guidance and support in my research. As it is said that the direction of basic science research is hard to predict, the belief, encourage, and patience of him are especially appreciated. I wish to thank the members of my dissertation committee, Dr. Phil Becraft, Dr. John Imsande, Dr. Pat Schnable, and Dr. Jonathan Wendel for their time, comments, and advice. I wish to extend my appreciation to Dr. John Mayfield for providing me a good start of research in Ames as a foreign student.

I would like to thank people in Dr. Peterson’s lab, Suzy, Luda, Surinder, and Yongli, and Jianbo. I am greatly nurtured by them in every day’s lab life.

I am very grateful to my parents, my sister, and my brother. They always provide immense love and support in everything I do, and they make my life full of joy. Finally, I can not say enough thanks to my husband. Without his emotional support, understanding, and encourage I could never accomplish this study.