Transcriptional regulation of the maize Myb-homologous genes *p1* and *p2*, and molecular analysis of a novel variegated pericarp allele

by

Yibin Wang

A thesis submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Major: Genetics

Program of Study Committee:
Thomas A. Peterson, Major Professor
Drena L. Dobbs
Kan Wang

Iowa State University

Ames, Iowa

2003

Graduate College
Iowa State University

This is to certify that the master's thesis of

Yibin Wang

has met the thesis requirements of Iowa State University

Signatures have been redacted for privacy

TABLE OF CONTENTS

CHAPTER 1. GENERAL INTRODUCTION

Plant transposable elements and their effects on gene expression

Transposable elements (TEs) were first described by McClintock in maize in the 1940s and were termed "controlling elements" for their unique ability to move, control gene expression and induce heritable mutations (McClintock 1948, 1956ab). Since then transposable elements have been discovered in a wide variety of organisms including both prokaryotes and eukaryotes. Transposable elements are widespread in plants and animals and are a major component of some genomes. It is estimated that sequences derived from transposable elements make up at least 45% of the human genome, and an even greater proportion of some plant genomes including maize (Feschotte et al. 2002). Transposable elements can be divided into two groups according to their transposition mechanism: Class 1 elements (retrotransposons) transpose via an RNA intermediate, and Class 2 elements transpose via a DNA intermediate (Craig et al. 2002). The Class 2 transposable elements in plants can be grouped into several families based on their conserved sequences at the element termini and other genetic and molecular characteristics, such as target site duplication upon insertion and the similarity of the element-encoded proteins. The majority of the transposons in plants are grouped into four families: (1) *hAT* family, including the maize *Ac/Ds* elements that were first described by McClintock; (2) CACTA superfamily, which contains the maize *En/Spm* element and has so far only been found in plants; (3) *Mutator* family, which is now widely used as a genetic tool for cloning new maize genes, and (4) MITEs family, which contains high copy number miniature inverted repeat elements and may be still active (Kunze and Weil, 2002; Jiang et al. 2003). Transposition of transposable elements can affect gene expression and induce chromosomal rearrangements such as duplication, deletion and inversion. It is believed that transposable elements play a major role in plant genome evolution (Fedoroff 2000).

Plant CACTA transposable element superfamily

McClintock's elegant studies on maize transposable elements focused on two transposon systems: *Ac/Ds* (*Activator/Dissociation*) and *Spm* (*Suppressor-mutator*) (McClintock 1948, 1954). Subsequently it was shown that another transposable element *En* (*Enhancer*) that was independently discovered by Peterson (Peterson 1953) was genetically and molecularly nearly identical to *Spm* (O'Reilly et al. 1985; Pereira et al. 1986; Masson et al. 1987). Like most other transposable element systems, members of the *En/Spm* element family also exist in several forms, including autonomous elements (*En/Spm*) which encode trans-factors required for transposition, and numerous non-autonomous elements (*Inhibitor*(*I*)/defective *Spm*(*dSpm*)) that can transpose only when the corresponding active autonomous elements are present in the genome. In some cases, non-autonomous elements are immobile due to the absence of cis-elements required for transposition. After the initial discovery of the maize *En/Spm* element, a large number of transposable elements with similar element termini and other structural features were identified in several other plant species, including *Antirrhinum*, sorghum, soybean and petunia (Kunze and Weil, 2002). Due to the conserved sequence 5'-CACTA-3' at the element termini, these transposable elements are grouped into the CACTA superfamily. Some members belonging to plant CACTA transposable element family are listed in Table 1. The maize *En/Spm* element is the best studied and provided much information on the novel properties of CACTA transposons.

The maize *En/Spm* element is 8.3 kb in length and exhibits several structural features that are characteristic of the CACTA superfamily: it generates a 3-bp target site duplication upon insertion; it contains 13-bp terminal inverted repeats (TIRs) with 5'-CACTA-3' at termini; and it contains structured subterminal repetitive regions (SRRs) with 9 copies of a 12-bp conserved sequence at the 5'-end and 15 copies at the 3'-end (Pereira et al. 1985, 1986; Table 1). TIRs and SRRs are cis-elements required for *En/Spm* transposition (Grant et al.

Table 1. Plant CACTA transposable elements superfamily

Name	Species	TIR length (bp) and sequence	Length	SRR repeat
En/Spm	*Zea mays*	13 CACTACAAGAAAA	8287 bp	TAAGAGTGTCGG
Cs1	*Sorghum bicolor*	20 CACTATGTG--(20bp)	23018 bp	CGTCTGTAATA
Tam1	*Antirrhinum majus*	13 CACTACAACAAAA	15164 bp	GTGTCCAA
Tgm1	*Glycine max*	13 CACTATTAGAAAA	3.5 kb	ACATCGG
Pis1	*Pisum sativum*	12 CACTACGCCAAA	2.5 kb	
Tpn1	Japanese morning glory	28 CACTACAAG--(28bp)	6412 bp	GACAACGGTT
Tnr3	*Oryza sativa*	13 CACTAGAAGGGAT	1536 bp	TGTGATG

Autonomous elements are in boldface. After Kunze and Weil (2002) in Mobile DNA II:565-610 (eds Craig et al., American Society for Microbiology Press, Washington D.C., 2002).

1993). Deletion of the outermost two nucleotides of the element's TIR reduced transposition frequency (Kunze and Weil, 2002).

En/Spm encodes several transcripts by alternative splicing of a single precursor transcript (Masson et al. 1989). Tobacco transformation analysis revealed that two proteins, TNPA and TNPD which are encoded by two major transcripts, are required for the two functional components of *En/Spm* (suppressor and mutator) (Frey et al 1990; Grant et al. 1990; Masson et al. 1991). TNPA protein contains 621-amino acids and is the most abundant *En/Spm* encoded product. TNPA binds specifically to the 12-bp motif in the SRRs. This binding ability is critical for both its suppression function and element transposition. It is proposed that when TNPA proteins bind to the ends of elements that insert in the coding region of the genes, they hinder RNA polymerase read-through during transcription elongation and result in premature transcription termination and thus suppress the host gene expression (Schwarz-Sommer et al. 1985; Gierl 1996). However, at present no molecular evidence for the potential sites of the premature termination of transcription was reported. In addition to the suppressor function, TNPA is also required for transposition. It is thought that TNPA can promote and stabilize synapsis of the transposon ends, enable the proper alignment of transposon TIRs, and even cause bending of *En/Spm* ends (Kunze and Weil 2002). TNPD is a 132-kDa protein and is present at much lower levels in maize. TNPD itself does not bind specific DNA sequences, but is thought to bind TNPA protein at the element ends and stabilize the transposome complex (Raina et al. 1998; Figure 1). TNPD is also required for transposition and may function as an endonuclease and cut the very ends of the elements during transposition (Gierl et al. 1989; Kunze and Weil 2002). The suppressor and mutator functional components of *En/Spm* elements were also observed in other CACTA elements including *Tpn1* in Japanese morning glory (Hoshino et al. 1997). Sequence comparison of some of the cloned CACTA elements revealed that TNPD-homologous proteins exhibit greater sequence conservation than TNPA. TNPD-homologous proteins were predicted from

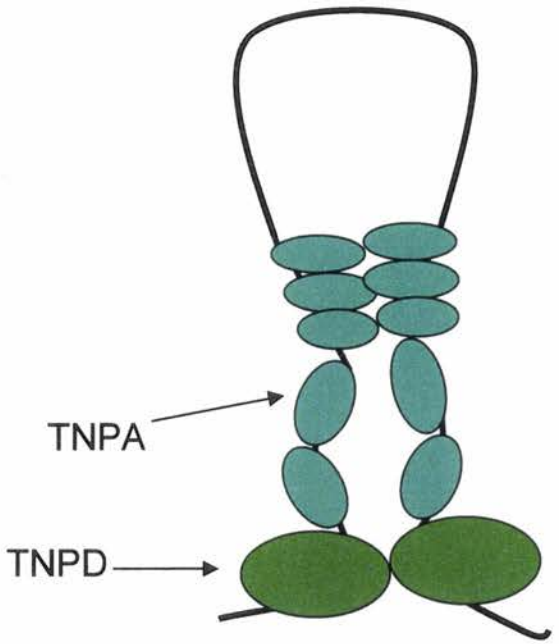


Figure 1. The maize *En/Spm* transposome complex model.
(Gierl A. Curr. Top. Microbiol. Immunol. 1996)

several transposable element sequences, including Tnp2 by *Tam1* in snapdragon (Nacken et al. 1991), DOPD by *Doppia* in maize (Bercury et al. 2001) and a protein encoded by *Tgm1* in soybean (Rhodes and Vodkin 1998). The DOPA protein encoded by *Doppia* in maize was shown to have 47% homology with TNPA (Bercury et al. 2001). The predicted proteins in other CACTA elements didn't reveal significant similarity to the TNPA protein. Accordingly, the repetitive motif present in the subterminal regions of *En/Spm* which has been demonstrated to be a binding site for TNPA is not conserved in other subterminal motifs that have been identified in several CACTA elements (Kunze and Weil 2002; Table 1).

Effects of transposable elements on plant gene expression

Mutant phenotypes caused by transposable element insertion had been observed long before the scientific explanation by McClintock. Insertion of transposable element can cause changes in gene expression both qualitatively and quantitatively, leading to striking phenotypic variation in plants. For example, in 1917 Emerson described a maize *p1* mutant allele with red stripes on a colorless kernel pericarp (Emerson 1917). Later it was found that this mutation is caused by an *Ac* transposable element insertion into intron 2 of the *p1* gene (Lechelt et al. 1989). Insertion of the transposable element disrupts the normal function of the *p1* gene that controlls red pigments biosynthesis in kernel pericarp and cob, and leads to the colorless pericarp. Excision of the *Ac* element in some cells during kernel development restores *p1* gene function and generats red pigments in the pericarp. Analysis of maize *En/Spm* and other transposon-induced mutant alleles also provided good examples for the striking effects of transposable elements on gene expression. The maize *bz1* gene is a structural gene involved in anthocyanin pigments synthesis. In the allele *bz-m13,* insertion of a *dSpm* element into exon 2 of *bz1* reduced *bz1* expression, but the mutant allele still retained a relative high level of *bz1* encoded enzyme due to the ability of *dSpm* to be spliced out of *bz1* transcripts (Kim et al. 1987). Thus, the *bz-m13* allele still showed almost full purple color in

aleurone layer. However this pigmentation was only produced when the autonomous element was absent in the genome. When an active *En/Spm* was present, this residual gene expression was completely suppressed (the suppressor function component of *En/Spm*), resulting in a bronze coloration in the endosperm aleurone layer. Excision of *dSpm* during kernel development from the insertion site was able to restore the normal functional gene form of *bz1* and produced kernels with full color spots on a bronze background in aleurone (the mutator function component of *En/Spm*). Additionally, the sizes of colored sectors varied according to the time of transposon excision during kernel development: earlier excision resulted in larger color sectors.

Transposons not only can affect gene expression quantitatively, but also can alter gene expression qualitatively. Transposon insertions can interfere with the action of transcription regulators of adjacent genes. In *Antirrhinum*, insertion of transposable element *Tam3* into the intron of the *plena* gene induced ectopic expression of the gene which caused the sterile floral organs be replaced by sex organs. It was proposed that the *plena* intron contained a negative regulator which repressed its expression in certain organs. The insertion of *Tam3* disrupted the negative regulator and generated the gain-of-function phenotype (Bradley et al. 1993; Girard and Freeling 1999). Transposition of transposons can induce genome rearrangement and affect gene organization and expression significantly. Even when such rearrangement is confined to a very small stretch of chromosome it can still induce altered organ-specific gene expression pattern. This is demonstrated in maize *AdhI* mutants induced by maize *Mutator* transposon. The excision of a *Mu* element from the TATA box of the *AdhI* promoter generated alleles with deletion, inversion or inverted duplication over a 430 bp region around the TATA box and resulted in allelic variation of *AdhI* gene expression in an organ-specific pattern (Kloeckner-Gruissen and Freeling 1995; Girard and Freeling 1999). The distinct tissue-specific expression observed in two maize *b* gene alleles are also thought to be the result of transposon-induced genome rearrangements in the 5'-regulatory regions (Radicella et al.

1992; Selinger et al. 1998).

Additionally, transposable elements themselves can function as regulatory elements for host genes and thereby control gene expression. Transposable elements have been found to function as promoters and initiate transcription in or near the ends of transposon insertions. The maize *hcf106* allele contains a *Mu1* element inserted at the 5'-UTR of *high chlorophyll fluorescence* gene. When the *Mutator* element system was active, the expression of the *hcf106* gene was blocked due to the element-encoded trans-factor binding to the *Mu1* insertion element and interfering with transcription initiation. However when the *Mutator* system was inactive the wild type phenotype reappeared. It was found that transcription of the *hcf106* gene was initiated at new sites that were in or just downstream of the end of the *Mu1* element. The end of the *Mu1* element now functioned as a promoter of the *hcf106* gene (Martienssen et al. 1990; Barkan and Martienssen 1991). Transposable element encoded proteins can also function as transcription regulators for adjacent gene expression. McClintock isolated a maize mutant allele *a1-m2* with anthocyanin pigmentation variegation. This allele contains a *dSpm* element inserted in the 5'- promoter region of the *a1* gene. The expression of *a1-m2* requires the presence of an active *En/Spm* element. When *En/Spm* is present this allele shows a leaky background expression, with some revertant sectors due to excision of the insertion element. However when *En/Spm* is absent, *a1* expression is completely eliminated. Here the *En/Spm* encoded protein TNPA is thought to function as a transcriptional activator instead of suppressor of the *a1* gene (Masson et al. 1987). It is believed that some transposable elements have been assimilated by adjacent genes during evolution to become a component of the cis-regulatory elements of the genes (Girard and Freeling 1999).

The *Myb*-gene family in plants

The *Myb*-gene was first identified in *avian myeloblatosis* virus as the form of *v-myb* oncogene that encoded a DNA-binding motif called Myb-domain. Since then, large numbers

of *Myb*-related genes were discovered in a wide variety of eukaryotes including animals, plants, fungi and cellular slime molds (Lipsick 1996). Myb-proteins in animals and slime molds have three copies of the Myb-domain. For example, there are three adjacent imperfect Myb-repeats classified as R1, R2 and R3 in C-Myb protein in vertebrates (Lipsick 1996). In contrast, most of the Myb-proteins in plants have two repeats corresponding to R2 and R3 repeats in C-Myb (Jin and Martin 1999).

Many *Myb*-gene encoded proteins are believed to act as transcription regulators by binding to specific DNA sequences and influence gene expression. Only a limited number of *Myb*-genes are expressed in vertebrates, fungi and cellular slime mold (Lipsick 1996). Functional analyses of the three major Myb-proteins (A-, B-, C-Myb) in vertebrates have shown that those Myb-proteins act as transcriptional regulators for cell cycle-progression, cell division and differentiation (Oh and Reddy 1999). The *Myb*-related genes have also been identified in a variety of plant species including mosses, monocots and dicots. In contrast to the number of *Myb*-genes in animals, plants possess a large number of *Myb*-genes. Approximately 40 *Myb*-genes have been identified to date in *Petunia hybrida,* while more than 100 *Myb*-genes are found in the Arabidopsis genome (Romero *et al.* 1998; Stracke et al. 2001). Studies of monocot *Zea mays* revealed that maize expressed more than 82 *Myb*-genes with distinct organ-specific expression patterns (Rabinowicz *et al.* 1999). The functions of most plant *Myb*-genes are still unknown. However analysis of some of the identified plant *Myb*-genes has revealed that they have diverse functions, including cell cycle regulation, control of cellular morphogenesis, regulation of secondary metabolism and plant responses to hormone molecules, viral infections, and environmental changes (Jin and Martin 1999; Martin and Paz-Ares 1997). It is suggested that *Myb*-genes in plants have expanded and diversified their functions in conjunction with the development of novel functions of higher plants. Thus the large number of plant *Myb*-genes and their distinct functions are important

for the evolution of the metabolic and morphological diversity of plants (Jin and Martin 1999).

Regulation of flavonoid biosynthesis by *Myb*-homologous regulatory genes

Among numerous *Myb*-genes identified in plants, several regulatory genes involved in flavonoid biosynthesis have been intensively studied. Flavonoids are a group of secondary metabolites with diverse structures and are widely distributed in numerous plant species. Flavonoids are involved in a variety of processes in plants such as response to pathogens and insect feeding, protection from ultraviolet light irradiation, and plant tissue reinforcement (Koes et al. 1994). The most apparent effect of flavonoids in plants is the bright pigments they produced in different plant tissues. The biochemical steps of the synthesis of flavonoid and derived pigments are studied intensively and are relatively well understood now (Styles and Ceska 1977; Byrne et al. 1996; Figure 2). Structural genes encoding catalytic enzymes for the flavonoids biosynthetic pathway have been identified in several plants, including maize, petunia and snapdragon, by genetic and molecular approaches. Previous studies have showed that expressions of these structural genes are regulated by several transcriptional regulators. The regulatory genes encoding Myb- and other transcription factors that control maize anthocyanin and phlobaphene pigments biosynthesis are the best studied. Because the presence or absence of pigments is not essential for plant viability, genes involved in anthocyanin and phlobaphene pigmentation have been proved to be ideal genetic markers for the study of gene regulation and transposable element transposition in plants (Wessler 1988).

Two groups of regulatory genes are involved in maize anthocyanin pigmentation in several plant tissues. The *c1* and *pl* genes encode Myb-like transcription factors that activate transcription of a subset of structural genes such as *a1*, *c2*, *chi*, *bz1* and *bz2* of the anthocyanin synthetic pathway. However this transcriptional activation requires the presence of members of the *r/b* regulatory gene family which encode Myc-homologous regulators with an acidic

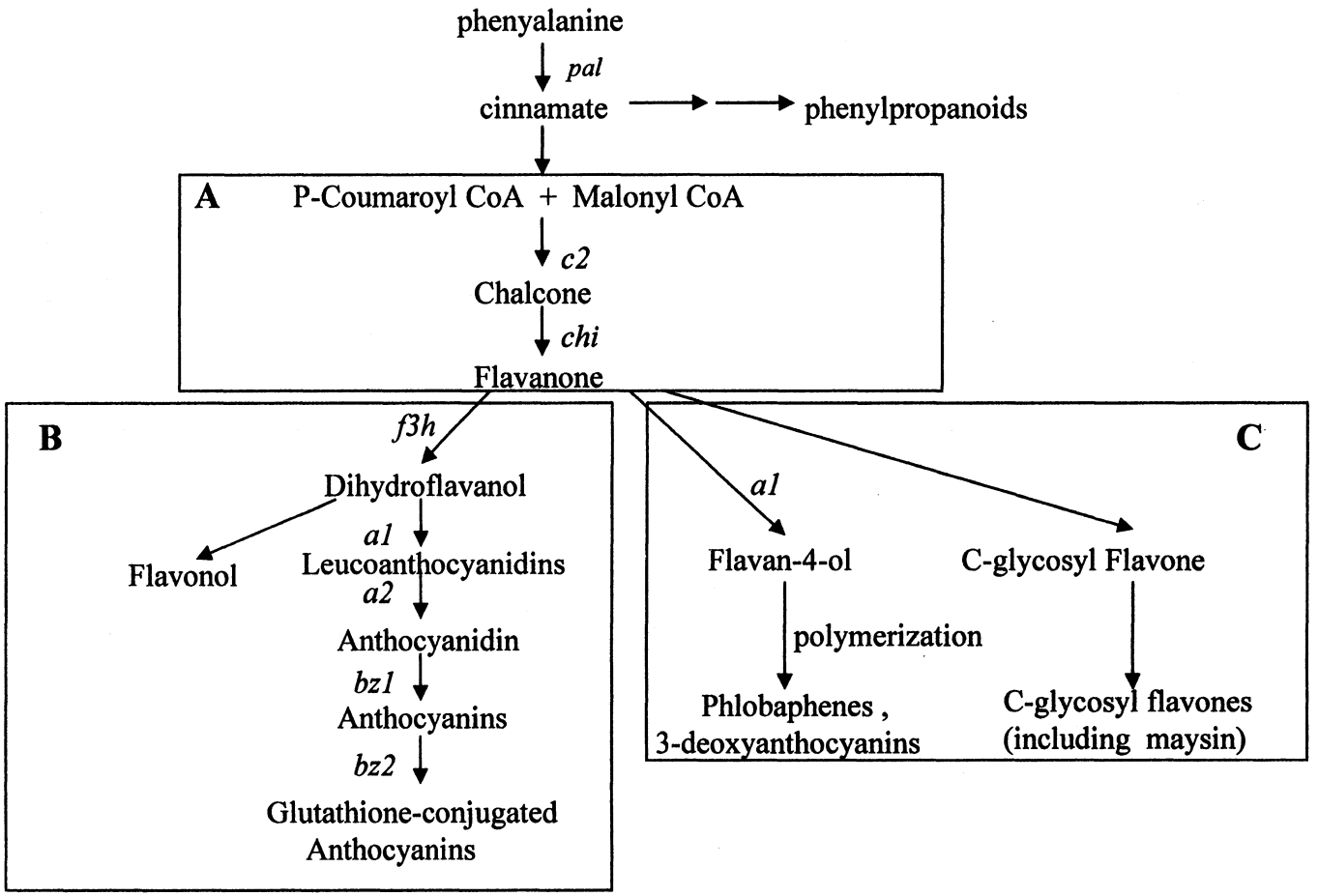


Figure 2. The flavonoid biosynthesis pathway in maize. Box A represents steps regulated by both *p1* and *c1/pl* plus *r/b* genes. Box B represents steps regulated by *c1/pl* plus r/b. Box C represents steps regulated by *p1* gene. (Byrne et al. 1996)

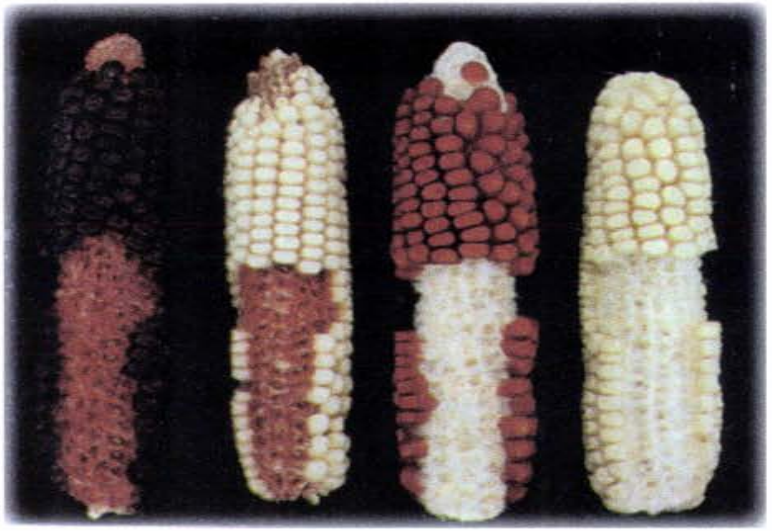
transcription activation domain and a helix-loop-helix DNA-binding domain (Ludwig and Wessler 1990). Though the genes in each family encode homologous proteins with similar functional domains, the expression pattern of each gene is quite different. Different combinations of *c1/pl* and *r/b* gene family can generate different patterns of anthocyanin pigmentation. The *c1* gene controls anthocyanin pigments accumulation in the kernel endosperm aleurone layer (Paz-Ares et al. 1987), while *pl* regulates pigmentation in the plant body (Cone et al. 1993). The *r* locus is complex and contains several separable components: the S component controls seed aleurone pigmentation and the P component conditions pigmentation in coleoptile and anther (Stadler and Neuffer 1953). The *b* gene family regulates anthocyanin pigments accumulation in husk, anther and leaf sheath. One of the *b* gene alleles, *B-peru*, also conditions pigmentation in seed aleurone (Chandler et al. 1989). Homologous regulatory genes controlling anthocyanin biosynthesis in other plant species have also been identified, including the *r* gene family homolog *Delila* in snapdragon and *an2* in petunia, and a *c1* homologous gene *an11* in petunia. These homologous genes also activate transcription of a subset of anthocyanin structural genes. Mutations in *an2* and *an11* in petunia can be complemented by *Lc*, a member of the maize *r* gene family, and the maize *c1* gene, respectively, indicating the conservation of the regulatory functions among these transcriptional regulatory genes (Goodrich et al. 1992; Quattrocchio et al. 1993)

In contrast to the multiple genes involved in anthocyanin biosynthesis, the accumulation of another major flavonoid pigment, phlobaphene, in maize floral organs requires a single *Myb*-homologous regulatory gene *p1* (Athma et al. 1992). The synthetic pathway of red pholobaphene pigments shares its early steps with the anthocyanin pathway. The maize *p1* gene encodes a transcription factor with a Myb DNA-binding domain and an acidic transcription activation domain (Grotewold et al. 1991). The *p1* gene regulates transcription activation of structural genes such as *a1*, *c2* and *chi* on the steps shared by anthocyanin pathway and the branching steps leading to red phlobaphene pigments in maize floral organs,

including kernel pericarp and cob glumes. However, the *p1* gene can not activate the transcription of structural genes such as *bz1* and *bz2* on the anthocyanin branching pathway (Grotewold et al. 1991; Figure 2). Further analysis of P1 protein binding ability to target DNA sites indicated that P1 transcription factor can bind specific sites in the maize *a1* gene promoter but can not bind the *bz1* promoter. This is contrary to the binding ability of C1 Myb-homologous transcription factor which can bind specific sequences in both *a1* and *bz1* (Grotewold et al. 1994). In addition to the regulation of visible phlobaphene accumulation in pericarp and cob, the *p1* gene is also expressed in other maize tissues including silk, husk and tassel glumes. Low-level expression of *p1* gene is also detected in vegetative tissues (Cocciolone et al. 2001). A major QTL determining silk maysin level in maize was found to be correlated with the *p1* locus (Byrne et al. 1996). Maysin is a C-glycosyl-flavone compound that retards corn earworm larvae growth. Recently, a *p1* homologous gene termed *p2* was isolated from maize. The *p2* gene is closely linked to *p1* in several *p1*-locus alleles. It shares high homology in the coding region with *p1* and encodes a transcription factor with a conserved Myb DNA-binding domain and a transcription activation domain, but with a divergent C-terminal domain compared to P1 factor (Zhang et al. 2000). The *p2* gene-encoded protein can also activate the transcription of structural genes on the phlobaphene synthetic pathway in transformed maize BMS cells (Zhang et al. 2003). Analysis of maize alleles with interstitial deletions of *p1/p2* complex revealed that *p2* was also involved in the regulation of flavone biosynthesis. Deletion of both *p1* and *p2* eliminates the accumulation of maysin and its related compounds apimaysin and neomaysin in fresh maize silks (Zhang et al. 2003). Two *p1* homologous genes, the *y1* gene in sorghum and *p2-t* in teosinte that regulate phlobaphene pigmentation and flavone compound synthesis, were isolated and characterized. Both genes encode transcription factors with conserved Myb DNA-binding domains (Zhang et al. 2000).

Tissue-specific expression and transcriptional regulation of maize *p1* and homologous genes

Except for a small number of genes that are expressed constitutively, most genes express variably among different cell types. Proper spatial and temporal expression of genes is essential for development processes and the responses to external stimuli in various organisms including plants. Changes in the spatial-temporal gene expression pattern in plants, especially of those regulatory genes, can produce novel phenotypes and physiological traits (Meisel and Lam 1997). Like the diverse tissue-specific expression pattern observed in maize *c1/pl* and *r/b* gene families, maize *p1* and its homologous genes also showed different expression patterns. Several maize *p1*-locus alleles with striking pigmentation patterns in pericarp and cob have been characterized. As shown in Figure 3A, *p1*-locus alleles can either accumulate red phlobaphene pigments in both kernel pericarp and cob (specified as *P1-rr* allele for red pericarp and red cob) or produce pigments in only one of the tissues, such as the phenotype observed in *P1-wr* (white pericarp and red cob) and *P1-rw* allele (red pericarp and white cob). In contrast to the multiple components of maize *r*-locus that controls different tissue pigmentation, the accumulation of red phlobaphene pigments in pericarp and cob is controlled by a single gene. Mutation of the *p1* gene affects the pigmentation in both organs (Athma et al. 1992). The pigmentation level in pericarp and cob in each allele is correlated with *p1* transcripts level, as in *P1-wr* allele the amount of *p1* transcripts in pericarp is only about 30% of that of *P1-rr* allele (Chopra et al. 1996). The *p1* gene has a complex gene structure with a ~ 7 kb coding region flanked by two 5.2 kb direct repeat sequences at 5'- and 3'- ends (Lechelt et al. 1989; Figure 3B). The maize *P1-wr* allele contains 6 copies of a *p1*-homologous unit arranged as tandem repeats. The 5'-flanking region and the coding region of each unit share high homology with the *P1-rr* gene, while the 3'-flanking region is similar to a *p1*- homologous gene *p2-t* in teosinte (Chopra et al. 1998; Zhang et al. 2000). Despite the redundant function of the *p1* and *p2* genes in transcriptional regulation of



P1-rr *P1-wr* *P1-rw* *p1-ww*

Figure 3A. Maize *p1*-locus alleles with distinct pigmentation pattern in pericarp and cob.

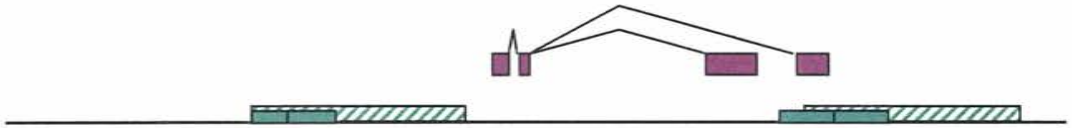


Figure 3B. The maize *p1* gene structure. The hatched boxes indicate 5.2 kb large repeats.The green boxes indicate small repeat sequences in the large repeats. The exons and introns of the *p1* gene are indicated by boxes and thin lines, respectively, in the upper map.

flavonoid compounds biosynthesis, *p2* has a different expression pattern from *p1*. The *p2* gene is expressed in husk, silk and anther, but not in pericarp and cob. Sequence comparison revealed that the *p2* gene is homologous to *p1* in coding regions but divergent in both 5'- and 3'- flanking regions. Another *p1*-homologous gene isolated from teosinte, *p2-t*, was shown to be more similar to the maize *p2* gene in both sequence and expression pattern. The *p2-t* gene expresses in silk, anther and tassel glumes and not in pericarp and cob in maize. It was proposed that *p1/p2* gene complex in maize are generated by a segmental gene duplication of a precursor gene similar to the *p2-t* gene and separated subsequently by retroelement insertion (Zhang et al. 2000).

Regulation of gene expression can be exerted at several levels including both transcriptional and translational controls. A common level of plant gene regulation occurs at transcription, especially the initiation of transcription. Interaction of cis-acting regulatory elements such as enhancer and promoters with RNA Polymerase II transcription-initiation complex and transcription factors are required for the active transcription of genes. Regulatory sequence changes often can alter gene expression. Analysis of the expression of two maize *b* gene alleles, *B-I* and *B-peru*, which have high homology in their coding and 3'-flanking regions but with different expression patterns, provided good evidence for the importance of promoter sequences in allelic expression diversity. Transient assays of the 5'-leader and promoter sequences of *B-I* and *B-peru* demonstrated that their different expression patterns were determined by their divergent 5'-regulatory sequences (Radicella et al. 1992). It is proposed that the segmental gene duplication of the precursor *p* gene placed new regulatory sequences upstream of *p1* gene and resulted in more intense and broader expression of *p1*. The finding of cis-regulatory elements in the 5'- promoter region of the *p1* gene and its broader and more intense expression in maize organs strongly supported this hypothesis. Functional analysis of the *p1* gene promoter in both transient and stable transformation identified several cis-regulating elements in the 5'-flanking region of the *p1* gene, including a

distal and a proximal enhancer element and a basal promoter fragment, that were important for its tissue-specific expression (Sidorenko et al. 1999, 2000). Transposon mutagenesis analysis indicated that insertion of transposable elements into these cis-regulatory regions disrupted normal *p1* gene transcription and induced mutant phenotypes with less pigmented pericarp and cob. Transgenic plants containing transgenes with the *p1* promoter fused with either *GUS* reporter gene or *p1* cDNA regenerated *p1* tissue-specific expression pattern (Sidorenko et al 2000; Cocciolone et al. 2000, 2001). These results suggest that the *p1* gene promoter contains cis-regulatory elements that are important for its tissue-specific expression. The divergent flanking regions among *p1* and its homologous genes might be responsible for their distinct expression patterns. However, at present the exact mechanisms for the tissue-specific expression pattern of *p1* and its homologous genes are still unknown. Several cis-elements that are important for maize regulatory gene tissue-specific expression and their response to external signals have been identified in the promoter regions (Selinger et al. 1998; Kao et al. 1996). However the cis-elements required for gene expression in certain plant tissues may be gene-specific and not conserved in different genes that are expressed in the same tissue. For example, several maize pollen-specific elements have been identified, but they are not universally present in all the pollen-specific or pollen-expressed genes (Hamilton et al. 1998).

Duplicated genes are frequently observed in plant genomes. Large regulatory gene families generated by gene expansion during evolution are observed in various plant species. The new copies of duplicated genes can either lose function or retain their original functions. Frequently the duplicated genes may acquire new regulatory elements and display new expression patterns. In particular, changes in the cis-regulatory elements of transcriptional factors are thought to be the predominant mechanism for the generation of novel phenotypes (Doebley and Lukens, 1998). The expression of maize *p1* and its homologous genes are tightly regulated by the maize developmental process. The transcription of *p1* gene is only

active late during plant development. No external stimuli are reported to be required for its transcription activation. Also, mutations in *p1* gene are readily detected by the change of pericarp color and the mutations do not affect plant viability. These make the *p1* gene an extremely good genetic marker for the study of both plant gene regulation and the effects of transposable elements. The analyses of maize *Ac* transposable element-induced *p1* gene mutant alleles have provided important information on the *Ac* transposition mechanism and the effect of transposons on plant genome evolution (Athma and Peterson 1991; Zhang and Peterson 1999; Xiao and Peterson 2002). Analysis of transposon-induced *p1* mutant alleles will also help to elucidate the mechanisms of the diverse effects of transposble element on plant gene expression. Similarly, the elucidation of transcriptional regulation of *p1* and its homologous genes can lead to a better understanding of plant gene regulation.

Thesis organization

The thesis contains four chapters: chapter 1 is a general introduction; chapters 2 and 3 are research reports; chapter 4 is a general summary. Chapter 2 presents molecular analysis of a maize *p1*-locus mutant allele *p1-vv85*. Genetic and molecular analysis suggested that the *p1* gene transcription in the *p1-vv85* mutant allele was affected by a CACTA transposable element. We present evidence that epigenetic regulation of the insertion element is correlated with the phenotype change of *p1-vv85* and its related alleles. Chapter 3 presents the functional analysis of the maize *Myb*-homologous *p2* gene promoter and leader regions in both transient and transgenic plants assays. Chapter 2 and chapter 3 are written in journal paper format that will be submitted for publication. Dr. Peter Peterson (co-author in chapter 2) isolated the mutant *p1-vv85* allele.

References

Athma, P., and Peterson, T. (1991) *Ac* induces homologous recombination at the maize *P*

locus. Genetics 128: 163-173.

Athma, P., Grotewold, E., and Peterson, T. (1992) Insertional mutagenesis of the maize *P* gene by intragenic transposition of *Ac*. Genetics 131:199-209.

Barkan, A., and Martienssen, R.A. (1991) Inactivation of maize transposon *Mu* suppresses a mutant phenotype by activating an outward-reading promoter near the end of *Mu1*. Proc. Natl. Acad. Sci. USA 88(8): 3502-3506.

Bercury, S.D., Panavas, T., Irenze, K., and Walker, E.L. (2001) Molecular analysis of the *Doppia* transposable element of maize. Plant Mol. Biol. 47(3): 341-351.

Bradley, D., Carpenter, R., Sommer, H., Hartley, N., and Coen, E. (1993) Complementary floral homeotic phenotypes result from opposite orientations of a transposon at the *plena* locus of *Antirrhinum*. Cell 72(1): 85-95.

Byrne, P., McMullen, M., Snook, M., Musket, T., Theuri, J., Widstrom, N., Wiseman, B., and Coe, E. (1996) Quantitative trait loci and metabolic pathways: genetic control of the concentration of maysin, a corn earworm resistance factor, in maize silks. Proc. Natl. Acad. Sci. USA 93: 8820-8825.

Chandler, V.L., Radicella, J.P., Robbins, T.P., Chen, J. and Turks, D. (1989) Two regulatory genes of the maize anthocyanin pathway are homologous: isolation of *B* utilizing *R* genomic sequences. Plant Cell 1(12): 1175-1183.

Chopra, S., Athma, P., and Peterson, T. (1996) Alleles of the maize *P* gene with distinct tissue specificities encode Myb-homologous proteins with C-terminal replacements. Plant Cell 8: 1149-1158.

Chopra, S., Athma, P., and Peterson, T. (1998) A maize Myb-homologue is encoded by a stable multicopy gene complex. Mol. Gen. Genet. 260: 372-378.

Cocciolone, S.M., Sidorenko, L.V., Chopra, S., Dixon, P.M., and Peterson, T. (2000) Hierarchical patterns of transgene expression indicate involvement of developmental mechanisms in the regulation of the maize *P1-rr* promoter. Genetics 156: 839 - 846.

Cocciolone, S.M., Chopra, S., Flint-Garcia, S.A., McMullen, M.D., and Peterson, T. (2001) Tissue-specific patterns of a maize Myb transcription factor are epigenetically regulated. Plant J. 27(5): 467-478.

Cone, K.C., Cocciolone, S.M., Burr, F.A., and Burr, B. (1993) Maize anthocyanin regulatory gene *pl* is a duplicate of *c1* that functions in the plant. Plant Cell 5(12): 1795-1805.

Craig, N.L., Craigie, R., Gellert, M., and Lambowitz, A.M. (2002) Mobile DNA II. American Society fro Microbiology Press, Washington D.C., 2002.

Doebley, J., and Lukens, L. (1998) Transcriptional regulators and the evolution of plant form.

Plant Cell 10(7): 1075-1082.

Emerson, R.A. (1917) Genetical studies of variegated pericarp in maize. Genetics 141: 347-360.

Feschotte, C., Jiang, N., Wessler, S.R. (2002) Plant transposable elements: where genetics meets genomics. Nat. Rev. Genet. 3(5): 329-341.

Frey, M., Reinecke, J., Grant, S., Saedler, H., and Gierl, A. (1990) Excision of the *En/Spm* transposable element of *Zea mays* requires two element-encoded proteins. EMBO J. 9: 4037-4044.

Gierl, A., Cuypers, H., Lutticke, S., Pereira, A., Schwarz-Sommer, Z., Dash, S., Peterson, P.A., and Saedler, S. (1989) Structure and function of the *En/Spm* transposable element system of *Zea mays*: identification of the suppressor component of *En*. In Plant Transposable Elements: 115-120 (eds O.E. Nelson, Jr., Plenum Press Corp., New York, N.Y.).

Gierl, A. (1996) The *En/Spm* transposable element of maize. Cuur. Top. Microbiol. Immunol. 204: 145-159.

Girard, L. and Freeling, M. (1999) Regulatory changes as a consequence of transposon insertion. Dev. Genet. 25: 291-296.

Goodrich, J., Carpenter, R., and Coen, E.S. (1992) A common gene regulates pigmentation pattern in diverse plant species. Cell 1992 68(5): 955-964.

Grant, S.R., Gierl, A., and Saedler, H. (1990) *En/Spm* encoded TnpA protein requires a specific target sequence for suppression. EMBO J. 9: 2029-2035.

Grant, S.R., Hardenack, S., Trentmann, S., and Saedler, H. (1993) Functional cis-element sequence requirements for suppression of gene expression by the TNPA protein of the *Zea mays* transposon *En/Spm*. Mol Gen Genet. 241(1-2): 153-160.

Grotewold, E., Athma, P., and Peterson, T. (1991) Alternatively spliced products of the maize *P* gene encode proteins with homology to the DNA-binding domain of Myb-like transcription factors. Proc. Natl. Acad. Sci. USA 88: 4587-4591.

Grotewold, E., Drummond, B. J., Bowen, B., and Peterson, T. (1994) The *myb* homologous *P* gene controls phlobaphene pigmentation in maize floral organs by directly activating a flavonoid biosynthetic gene subset. Cell 76: 543-553.

Hoshino, A., Abe, Y., Saito, N., Inagaki, Y., and Iida, S. (1997) The gene encoding flavanone 3-hydroxylase is expressed normally in the pale yellow flowers of the Japanese morning glory carrying the speckled mutation which produce neither flavonol nor anthocyanin but accumulate chalcone, aurone and flavanone. Plant Cell Physiol. 38(8): 970-974.

Jiang, N., Bao, Z., Zhang, X., Hirochika, H., Eddy, S.R., McCouch, S.R., and Wessler, S.R.

(2003) An active DNA transposon family in rice. Nature 421(6919):163-167.

Jin, H. and Martin, C. (1999) Multifunctionality and diversity within the plant MYB-gene family. Plant Mol Biol. 41(5): 577-585.

Kao, C.Y, Cocciolone, S.M, Vasil, I.K, McCarty, D.R. (1996) Localization and interactionof the cis-acting elements for abscisic acid, VIVIPAROUS1, and light activation of the *C1* gene of maize. Plant Cell 8(7): 1171-1179.

Kim, H.Y., Schiefelbein, J.W., Raboy, V., Furtek, D.B., and Nelson, O.E. Jr. (1987) RNA splicing permits expression of a maize gene with a defective *Suppressor-mutator* transposable element insertion in an exon. Proc. Natl. Acad. Sci. USA 84(16): 5863-5867.

Kloeckner-Gruissem, B., and Freeling, M. (1995) Transposition-induced promoter scrambling: A mechanism for the evolution of new alleles. Proc. Natl. Acad. Sci. U S A 92:1836-1840.

Koes, R.E., Quattrocchio, F., and Mol, J.N. (1993) The flavonoid biosynthetic pathway in plants: function and evolution. BioEssays 16(2): 123-132.

Kunze, R., and Weil, C.F. (2002) The *hAT* and CACTA superfamilies of plant transposons. In Mobile DNA II: 565-610 (eds Craig, N.L., Craigie, R., Gellert, M. and Lambowitz, A.M., American Society for Microbiology Press, Washington D.C., 2002).

Lipsick, J.S. (1996) One billion years of Myb. Oncogene 13: 223-235.

Lechelt, C., Peterson, T., Laird, A., Chen, J., Dellaporta, S.L., Dennis, E., Peacock, W.J., and Starlinger, P. (1989) Isolation and molecular analysis of the maize *P* locus. Mol Gen Genet. 219(1-2): 225-234.

Ludwig, S.R., and Wessler, S.R. (1990) Maize *R* gene family: tissue-specific helix-loop-helix proteins. Cell 62(5): 849-851.

Martienssen, R., Barkan, A., Taylor, W.C., and Freeling, M. (1990) Somatically heritable switches in the DNA modification of *Mu* transposable elements monitored with a suppressible mutant in maize. Genes & Dev. 4(3): 331-343.

Martin, C., Paz-Ares, J. (1997) MYB transcription factors in plants. Trends Genet. 13(2): 67-73.

Masson, P., Surosky, R., Kingsbury, J.A, and Fedoroff, N.V. (1987) Genetic and molecular analysis of the *Spm*-dependent *a-m2* alleles of the maize *a* locus. Genetics 117(1): 117-137.

Masson, P., Rutherford, G., Banks, J.A., and Fedoroff, N. (1989) Essential large transcripts of the maize *Spm* transposable element are generated by alternative splicing. Cell 58(4): 755-765.

Masson, P., Strem, P., and Fedoroff, N. (1991) The *tnpA* and *tnpD* gene products of the *Spm*

element are required for transposition in tobacco. Plant Cell 3: 73-85.

Meisel, L., and Lam, E. (1997) Switching on gene expression: analysis of the factors that Spatially and temporally regulate plant gene expression. Genetic Engineering 19:183-199.

McClintock, B. (1948) Mutable loci in maize. Carnegie Inst .Wash. Year Book 47:155- 169.

McClintock, B. (1954) Mutations in maize and chromosomal aberrations in Neurospora. Carnegie Inst .Wash. Year Book 53: 254-260.

McClintock, B. (1956a) Intranuclear systems controlling gene action and mutation. Brookhaven Symp. In Biol. 8: 58-74.

McClintock, B. (1956b) Controlling elements and the gene. Cold Spring Harbor Symp. Quant. Biol. 21: 197-216.

Nacken, W.K., Piotrowiak, R., Saedler, H., and Sommer, H. (1991) The transposable element *Tam1* from *Antirrhinum majus* shows structural homology to the maize transposon *En/Spm* and has no sequence specificity of insertion. Mol Gen Genet. 228(1-2):201-208.

Oh, I.H., and Reddy, E.P. (1999) The myb gene family in cell growth, differentiation and apoptosis. Oncogene 18(19): 3017-3033.

O'Reilly, C., Shepherd, N.S., Pereira, A., Schwarz-Sommer, Z., Bertram, I., Robertson, D.S., Peterson, P.A., and Saedler, H. (1985) Molecular cloning of the *c1* locus of *Zea mays* using the transposable elements *En* and *Mu1*. EMBO J. 4: 877-882.

Paz-Ares, J., Ghosal, D., Wienand, U., Peterson, P.A. and Saedler, H. (1987) The regulatory *c1* locus of *Zea mays* encodes a protein with homology to myb proto-oncogene products and with structural similarities to transcriptional activators. EMBO J. 6(12): 3553-3558.

Pereira,A., Schwarz-Sommer, Z., Gierl, A., Bertram, I., Peterson, P.A., Saedler, H.(1985) Genetic and molecular analysis of the *Enhancer(En)* transposable element system of *Zea mays*. EMBO J. 4(1): 17-23.

Pereira, A., Cuypers, H., Gierl, A., Schwarz-Sommer, Z., and Saedler, H. (1986) Molecular analysis of the *En/Spm* transposable element system of *Zea mays*. EMBO J. 5: 835-841.

Peterson, P.A. (1953) A mutable *pale green* locus in maize. Genetics 45:115-133.

Quattrocchio, F., Wing, J.F., Leppen, H., Mol, J. and Koes, R.E. (1993) Regulatory genes controlling anthocyanin pigmentation are functionally conserved among plant species and have distinct sets of target genes. Plant Cell 5(11): 1497-1512.

Rabinowicz, P.D., Braun, E.L., Wolfe, A.D, Bowen, B., and Grotewold, E. (1999) Maize R2R3 Myb genes: Sequence analysis reveals amplification in the higher plants. Genetics 153(1): 427-444.

Radicella, J.P., Brown,D., Tolar, L.A., and Chandler, V.L.(1992) Allelic diversity of the maize

B regulatory gene: different leader and promoter sequences of two B alleles determine distinct tissue specificities of anthocyanin production. Genes & Development 6: 2152-2164.

Rhodes, P.R., and Vodkin, L.O. (1998) Organization of the *Tgm* family of transposable elements in soybean. Genetics 120(2): 597-604.

Romero, I., Fuertes, A., Benito, M.J., Malpica, J.M., Leyva, A. and Paz-Ares, J. (1998) More than 80R2R3-MYB regulatory genes in the genome of *Arabidopsis thaliana*. Plant J. 14(3): 273-284.

Selinger, D.A., Lisch, D., and Chandler, V.L. (1998) The maize regulatory gene *B-Peru* contains a DNA rearrangement that specifies tissue-specific expression through both positive and negative promoter elements. Genetics 149(2): 1125-1138.

Sidorenko, L.V., Li, X., Tagliani, L., Bowen, B., Peterson, T. (1999) Characterization of the regulatory elements of the maize *P-rr* gene by transient expression assays. Plant Mole. Bio. 39: 11-19.

Sidorenko, L.V., Li, X., Cocciolone, S.M., Tagliani, L., Chopra, S., Bowen, B., Daniels, M., and Peterson, T. (2000) Complex structure of a maize Myb gene promoter: functional analysis in transgenic plants. Plant J. 22: 471-482.

Stadler, L.J. and Neuffer, M.G. (1953) Problems of gene structure. II. Separation of R-r elements (S) and (P) by unequal crossing over. Science 117: 471-472.

Stracke, R., Werber, M., Weisshaar, B. (2001) The R2R3-MYB gene family in Arabidopsis thaliana. Curr Opin Plant Biol. 4(5): 447-456.

Styles, E.D. and Ceska, O. (1977) The genetic control of flavonoid synthesis in maize. Can J. Genet. Cytol. 19: 289-302.

Wessler, S.R. (1988) Phenotypic diversity mediated by the maize transposable elements *Ac* and *Spm*. Science 242(4877): 399-405.

Xiao, Y., and Peterson, T. (2002) *Ac* transposition is impaired by a small terminal deletion. Mol. Genet. Genomics 266: 720-731.

Zhang, J., and Peterson, T. (1999) Genome rearrangements by nonlinear transposons in maize. Genetics 153: 1403-1410.

Zhang, P., Chopra, S., and Peterson, T. (2000) A segmental gene duplication generated differentially expressed Myb-homologous genes in maize. The Plant Cell 12: 2311-2322.

Zhang, P., Wang, Y., Zhang, J., Maddock, S., Snook, M., and Peterson, T. (2003) A maize QTL for silk maysin levels contains duplicated Myb-homologous genes which jointly regulate flavone biosynthesis. Plant Mol Biol. 52(1): 1-15.

CHAPTER 2. INSERTION OF A CACTA TRANSPOSABLE ELEMENT LEADS TO SOMATIC VARIEGATION AND GERMINAL INSTABILITY OF *p1* GENE EXPRESSION IN THE ABSENCE OF ELEMENT EXCISION

A paper to be submitted to *Genetics*

Yibin Wang, Peter A. Peterson, and Thomas Peterson

ABSTRACT

The maize *p1* gene encodes a Myb-like factor that activates transcription of structural genes such as *a1* and *c2* required for red phlobaphene pigmentation in certain floral organs including kernel pericarp and cob. Here we report the molecular analysis of a new *p1* mutant allele *p1-vv85* that exhibits both somatic and germinal instability. A novel CACTA transposon-like element was isolated from the second intron of the *p1* gene in both *p1-vv85* and its stable pigmented "revertant" allele *P1-rr'85*. Although *p1* gene transcription level in *p1-vv85* pericarp was greatly reduced relative to *P1-rr'85*, no restriction fragment length polymorphism was detected in *p1*-locus between these two alleles. Transcriptional analysis indicated that transcription of the *p1* gene in *p1-vv85* terminated prematurely in the vicinity of the CACTA insertion site. Southern blot analysis demonstrated that the *P1-rr'85* revertant allele exhibited more cytosine methylation at the ends of the insertion element than *p1-vv85*. These results are consistent with a model in which binding of a suppressor-like factor to the CACTA insertion element interferes with transcription elongation. We propose that the reversion of *p1-vv85* to a functional P1-rr state does not result from transposable element excision, but rather from relief of suppression which is correlated with the methylation of the CACTA insertion element.

INTRODUCTION

Since McClintock's elegant genetic studies of transposable elements in maize, numerous transposons have been identified and analyzed in a variety of organisms including both prokaryotes and eukaryotes (Craig et al. 2002). Extensive genetic and molecular analyses revealed that the widespread DNA transposable elements in plants could be categorized into several superfamilies, including *hAT*, CACTA, *Mutator* and MITEs, based on similarities in terminal sequences and other characteristics (Kunze and Weil 2002). Transposition of plant transposable elements not only can disrupt normal gene function and induce genome rearrangement but also can bring the genes in which they reside under the control of transposons and lead to novel spatial and temporal expression of the host genes (Girard and Freeling 1999). On the other hand, the genetic activity of transposable elements themselves is also under developmental control. Transposable elements in the plant CACTA family, especially maize *En/Spm* elements, provided good examples for such regulations.

Plant CACTA transposable elements are identified by the conserved terminal sequence 5'-CACTA-3'. They also share some other characteristics including the creation of 3-bp target site duplications and the presence of highly structured subterminal repetitive regions. CACTA elements have been isolated and characterized in several plant species, including maize (*En/Spm*, Peterson 1953; McClintock 1954), sorghum (*Cs1*, Chopra et al. 1999), *Arabidopsis* (*CACI*, Miura et al. 2001), snapdragon (*Tam1*, Bonas et al. 1984) and other species (Kunze and Weil 2002). Among the various CACTA elements, maize *En/Spm* is the most extensively studied. Autonomous *En* (*Enhancer*) and *Spm* (*Suppressor-Mutator*) elements were discovered independently by Peterson (Peterson 1953) and McClintock (McClintock 1954) and they have been shown to be molecularly and genetically nearly identical (Pereira et al. 1986; Masson et al. 1987). There are also numerous non-autonomous elements *I/dSpm* (*Inhibitor*/defective *Spm*) present in the maize genome. The functions of the

genetic components of *En/Spm* element, "suppressor" and "mutator", are carried out by two trans-factors encoded by two major transcripts. TNPD, which is necessary for element transposition, is required for the "mutator" function. The "suppressor" function is due to the presence of the other factor TNPA, which is also necessary for transposition (Frey et al. 1990; Masson et al. 1991). In several maize mutant alleles, such as *a1-m1* and *bz-m13*, *dSpm* insertion into structural genes does not completely disrupt gene expression. When *Spm* is absent, these mutants can have some level of expression. However when an active *Spm* is present such residual expression is completely suppressed (Schwarz-Sommer et al 1985; Kim et al. 1987). It is proposed that binding of TNPA to sets of a 12-bp consensus motif in subterminal repetitive regions of *dSpm* elements blocks transcription read-through and results in the suppression of residual expression of the *dSpm*-containing mutant alleles (Schwarz-Sommer et al 1985; Gierl 1996). These mutant alleles are also called "*Spm*-suppressible" alleles. TNPA is also involved in the element's transcriptional activation of adjacent genes in "*Spm*-dependent" alleles in which the expression of *dSpm*-containing genes requires the presence of an active *Spm* (Masson et al. 1987).

Plant genes involved in pigment biosynthesis have been used extensively in transposon isolation and characterization. The maize *p1* gene encodes a Myb-like transcription activator and controls the biosynthesis of a kind of flavonoid pigments, red phlobaphene, in certain floral organs including kernel pericarp and cob. Mutations in *p1* can lead to variegated or null pigmentation in both pericarp and cob, indicating that a single gene controls the synthesis of red pigments in both organs (Athma et al 1992). Emerson described a mutant allele with variegated pigmentation phenotype in kernel pericarp and cob and termed it *p-vv* (Emerson 1917). Subsequent molecular analysis revealed that the variegation in *p-vv* was caused by a transposable element *Ac* that inserted in intron2 of the *p1* gene (Lechelt et al 1989). Among numerous transposon-induced *p1*-locus mutant alleles that were analyzed so far, all are

caused by the insertion of maize transposable element *Ac/Ds* into coding or regulatory non-coding regions (Athma et al 1991, 1992; Zhang and Peterson 1999; Xiao et al. 2002).

Here we report molecular and genetic analysis of a new maize *p1*-locus mutant allele *p1-vv85*. This allele displays somatic instability with red stripes on colorless pericarp as well as variegated pigmentation in cob (Figure 1), similar to the phenotypes of some *Ac*-induced *p1-vv* alleles. One striking feature of this new allele is its highly germinal instability. It frequently reverted to fully pigmented P1-rr form, which specifies red pericarp and red cob (Figure 1). In some lines the frequency of germinal reversion of *p1-vv85* to normal functional form can be as high as 80%. Genetic analysis suggested that the variegated phenotype of *p1-vv85* is not associated with either maize *Ac* or *En/Spm* transposable element (data not shown). We present evidence from molecular analysis that the instability of *p1*-locus in *p1-vv85* is due to the insertion of an 1168-bp CACTA transposon-like element into intron 2 of *p1* gene. This element has terminal inverted repeats (TIRs) similar to a maize CACTA transposable element *MpiI* (Weydemann et al. 1988) and shares characteristics of CACTA transposon family, including 3 bp target site duplication and highly structured subterminal repetitive regions with multiple copies of a 12 bp consensus motif. Insertion of the CACTA element caused transcription of *p1* gene in *p1-vv85* to terminate prematurely in the vicinity of the insertion site. Molecular and genetic analyses conducted in this study, however, indicates that the germinal reversion frequently observed in *p1-vv85* is not due to excision of the insertion element. DNA methylation analysis of *p1-vv85* and its revertant alleles using methylation sensitive enzymes indicated that the TIRs in the revertant *P1-rr'85* allele are hypermethylated relative to that of *p1-vv85* allele. We propose that in *p1-vv85*, the CACTA insertion in *p1* binds a "suppressor" factor, possibly encoded by a corresponding autonomous element. Whereas, the hypermethylation in the *P1-rr'85* revertant allele greatly reduces its ability to bind the suppressor, leading to relief from suppression. This can explain the apparent reversion of *p1-vv85* in the absence of element excision.

MATERIALS AND METHODS

Maize genetic stocks

Maize *p1* gene alleles are identified by a two-letter suffix, which indicates their expression in kernel pericarp and cob. *P1-rr* specifies red pericarp and red cob, *p1-vv* specifies variegated pericarp and cob and *p1-ww* specifies white pericarp and cob. The dominant alleles are represented in italics with the first letter capitalized; the recessive alleles are represented with lower case italics. Maize stocks used in this study including: *P1-rr4B2*, derived from *p1-vv* by excision of *Ac* (Grotewold et al. 1991); *p1-ww1112*, derived from *p1-vv* by *Ac* induced deletion of *p1* coding region (Athma and Peterson 1991). The maize line *P1-rr107B* and two inbred lines A619 with genotype *p1-www* (white pericarp and cob and non-silk browning, both *p1* gene and a *p1* homologous gene *p2* are deleted, Zhang et al. 2000) and W23 with genotype *P1-wr* were obtained from Maize Genetics Cooperation Stock Center, Urbana, Illinois. The new *p1* gene mutant allele *p1-vv85* was obtained from the *P1-rr107B* as follows: plants of genotype *P1-rr107B/P1-rr107B* were crossed by maize lines containing maize *En/Spm* transposable element, which are maintained by Dr. Peter Peterson, Iowa State University. The original *p1-vv85* was crossed with inbred line W23. The *p1-vv85* stocks used in this study were self-crossed for two generations with relatively stable light variegated pericarp and cob. *P1-rr'85* were revertants from *p1-vv85* allele and the *P1-rr'85* stocks used in this study were self crossed for two generations with stable fully pigmented red pericarp and cob.

DNA isolation and Southern blot hybridization

Tissues including either seedling leaves or fresh pericarp of plants homozygous for alleles *P1-rr4B2*, *P1-rr107B*, *p1-vv85* and *P1-rr'85* or pericarp from plants of *p1-vv85/ p1-www* (A619) and *P1-rr'85/p1-www* (A619) were ground in liquid nitrogen and genomic

DNAs were extracted using cetyltrimethylammonium bromide (CTAB, Sigma) as described previously (Porebski et al. 1997). Approximately 5 μg genomic DNAs were digested with indicated enzymes according to manufacturer's instructions and then fractionated in 0.7% agarose gel and transferred to nylon membranes. Southern blot hybridization was performed as described previously (Zhang et al. 2003). Filters were stripped by washing for 10 minutes twice in boiling water with 0.5% SDS before re-hybridization.

PCR Amplification and Sequencing

Genomic DNA used for Southern blot analysis was polymerase chain reaction (PCR)-amplified using Expand Long Template PCR System (Roche Molecular Biochemicals). PCR conditions were performed according to manufacturer's instructions. Primers used in PCR were *p1* gene specific primers in intron 2. Their sequences are:

(1) PA-A6 (5'-CTGTTTAGCTACTGTAGTG-3')
and PA-B3 (5'-TGGATGATTTTTACCAAATCG-3'),

(2) Yb-1803f (5'-TCATCAGCGAAAGACAAAAGAGAGA-3')
and Yb-2616r (5'-TGCGAAGTCATAAGAAGTGG-3').

The PCR products from fresh pericarp were sequenced directly. The PCR products from seedling leaves were cloned into pGEM-T easy vector (Promega) and 3 independent clones from each allele were sequenced. The sequences covering the whole insertion element were determined by using Applied Biosystems fluorescent sequencing system at Iowa State University DNA Sequencing and Synthesis Facility.

RNA isolation, Northern blot hybridization and Reverse Transcription-PCR

Total RNAs were extracted using TRIzol reagent (Gibco-BRL) from pericarp 18 days after pollination (DAP) from plants homozygous for alleles in this study. About 20 μg of total RNA from each allele was fractionated in 1.2% agarose gel containing 18% formaldehyde

and transferred to nylon membrane. The filter was hybridized with probes in the sequence: *p1* cDNA (Grotewold et al. 1991), *c2* cDNA (Wienand et al. 1986), *a1* cDNA (Schwarz-Sommer et al. 1987) and *tubulin* (*tub*) cDNA (Villemur et al. 1992). Filters were stripped as above before re-hybridization.

For Reverse Transcription-PCR (RT-PCR), total RNAs prepared for RNA-gel analysis were treated with Amplification grade DNase I (Invitrogen) to remove any residual genomic DNA. A fraction of total RNA sample from each allele was further treated with PolyA polymerase (Invitrogen). The first strand cDNA synthesis was performed with 3'-RACE System (Invitrogen). Primers used for cDNA amplification were:

(1) Yb-231f (5'-CGCGACCAGCTGCTAACCGT-3') in 5'-UTR of *p1* gene;

(2) Yb-737r (5'-CGTCCGCCCGAAGGTAGTTGAT-3') in *p1* gene exon2;

(3) Yb-1803f (5'-TCATCAGCGAAAGACAAAAGAGAGA-3')
and Yb-2135f (5'-GTCACTGGCTCAGCATGTGTCT-3') in *p1* gene intron2;

(4) Yb-6245f (5'-GTGCGTGGGTCTTGGTTCAG-3') in *p1* gene exon3.

(5) tub-1 (5'-GTGCATCTCCGTCCACATCG-3')
and tub-2 (5'-GCTCAAGAGGAGAGAACCAAG-3'), which were used for *α-tubulin* cDNA amplification as internal control.

Target cDNA amplifications were performed with HotStar polymerase (QIAGEN) and the cycle conditions were set according to manufacturer's instruction. PCR products were cloned into pGEM-T easy vector (Promega). Several independent clones were sequenced at Iowa State University DNA Sequencing and Synthesis Facility.

Nuclei Isolation and Nuclear Run-on Transcription Analysis

Pericarp from plants with genotype *p1-vv85/p1-www* (A619) and *P1-rr'85/p1-www* (A619) were peeled at stage 17 days after pollination (DAP) and were frozen in liquid nitrogen before nuclei isolation. Nuclei were prepared and run-on assays were performed as

described previously (Dorweiler et al. 2000) with minor modification. Approximately 2g pericarps from each sample were used for nuclei isolation. Purified DNA fragments including: (1) 5'-UTR+exon1+exon2 of *p1* gene; (2) exon3+3'-UTR of *p1* cDNA; (3) *c2* cDNA; (4) *a1* cDNA; (5) *ubiquitin* (*ubi*) cDNA (Christensen et al. 1992) and (6) linear pUC18 plasmid vector DNA as negative control. The DNA fragments were denatured and for 120ng~150ng of each DNA fragment was added per slot and transferred to nitrocellulose membrane (Hybond-ECL, Amersham) using a slot blotter. The transcripts were treated with RNase-free DNase I and Proteinase K (Boehringer), extracted with Phenol: Chloroform (1:1) and Chloroform: Isoamyl alcohol (100:1) and precipitated with 0.2 volume of 10M NH_4Ac and 2.5 volume of 95% ethanol. The pellet was dissolved and treated with RNase-free DNaseI again. Extraction and precipitation were repeated as described above and the labeled RNAs were dissolved in 150 µl RNase-free TE buffer. Filters were exposed and quantified using phophorimager and ImageQuant software at the Protein Facility of Iowa Sate University Office of Biotechnology. The hybridization values of each gene were normalized to the *ubiquitin* gene (*ubi*) hybridization value.

RESULTS

No restriction fragment length polymorphism between *p1-vv85* and *P1-rr'85* allele

Several maize *p1*-locus alleles exhibiting tissue-specific expression patterns have been isolated and analyzed at molecular level (Lechelt et al. 1989; Athma et al. 1992; Chopra et al. 1998). These analyses revealed that the *p1* gene has a complex structure. For example in a standard *P1-rr* allele *P1-rr4B2*, the *p1* gene contains a ~7 kb coding region flanked by two 5.2 kb direct repeats (Figure 2A). The *p1-vv* mutant allele conditions variegated pigmentation in pericarp and cob and the variegations in several *p1-vv* alleles are induced by insertion of maize transposable element *Ac* into coding or upstream regulatory non-coding regions of *p1*

locus (Emerson 1917; Athma et al. 1992). Previous genetic analysis suggested that the mutant phenotype observed in *p1-vv85* is not associated with maize *Ac* or *En/Spm* transposable element systems (data not shown). To determine whether the variegated phenotype observed in *p1-vv85* allele was caused by any other transposable element in *p1* locus, Southern blot hybridization was carried out using *p1* specific probes (Figure 2A). Representative enzyme digestion and hybridization results are depicted in Figure 2B. Genomic DNA from seedling leaves of *p1-vv85*, progenitor allele *P1-rr 107B*, revertant allele *P1-rr'85* and a reference allele *P1-rr4B2* were digested with *BamH*I and hybridized with probe 15. As shown in Figure 2B, *P1-rr4B2* gives two bands at 10 kb and 7.1 kb (lane 1). The 10 kb band is also present in all other 3 alleles. In parental *P1-rr107B*, as well as in *p1-vv85* and *P1-rr'85*, the 7.1 kb band is shifted to spproximately 6 kb, suggesting a deletion in the 3'-end of *p1* gene compared to the *P1-rr4B2* allele. Digestion with *Sal*I and hybridization with probe #15 gives three bands in *P1-rr4B2*, including a 3.4 kb band, a 3.0 kb band and a 1.2 kb band with two copies of a 1.2 kb SalI fragment from 5'- and 3'- flanking regions of the *p1* gene. While in *P1-rr107B*, *p1-vv85* and *P1-rr'85* the 3.4 kb and 3.0 kb bands are unchanged, the signal strength of the 1.2 kb band is reduced compared to that of *P1-rr4B2*, indicating deletion of one copy of the 1.2 kb fragment. Genomic DNA digested by *Xho*I further confirms that this deletion happened in the 3'-flanking region. The 1.2 kb fragment in the 5' region of *P1-rr* is thought to function as a transcriptional enhancer and is important for *P1-rr* paramutation (Sidorenko et al. 1999, 2001). Since both *P1-rr107B* and *p1-vv85* lack the 3' 1.2 kb fragment, the loss of this fragment is unlikely to be associated with the observed mutant phenotype in *p1-vv85*. The *BamH*I blot was then re-hybridized with probe 8B, which detect a 6.4 kb band in *P1-rr4B2*, containing 2.6 kb of *p1* gene promoter region and 3.8 kb of part of *p1* coding region (Figure 2A). This 6.4 kb band is also present in the progenitor *P1-rr107B* allele. But in *p1-vv85* and *P1-rr'85* the 6.4 kb band shifts up to approximately 7.5 kb. Hybridization of DNA digested with *Sal*I and *Kpn*I with probe #8B detects two bands, 2 kb and 2.5 kb in size, in both

P1-rr107B and *P1-rr4B2*. The 2 kb band comes from a *p1* homologous gene *p2* which is not expressed in pericarp and cob (Zhang et al. 2000). This 2 kb band is also present in *P1-rr'85* and *p1-vv85*. The 2.5 kb band in *P1-rr4B2* and *P1-rr107B* comes from intron 2 of the *p1* gene and it shifts up to approximately 3.6 kb in both *p1-vv85* and *P1-rr'85*, indicating that a ~1.1 kb fragment inserted into the second intron of *p1* gene (Figure 2A). However, no polymorphism between *p1-vv85* and *P1-rr'85* was detected within a ~23 kb region of *p1* locus by Southern analysis using several restriction enzymes (data not shown). Southern blot hybridization and PCR analysis with DNA extracted from either leaf or pericarp of several independent mutant *p1-vv85* and revertant *P1-rr'85* plants also failed to detect any polymorphism within *p1* locus.

_p1-vv85_ contains a CACTA element insertion

The inserted fragment detected in Southern analysis was amplified from pericarp DNA by PCR using *p1* gene specific primers from intron 2 (Materials and Methods). The amplified products were sequenced directly. Sequence analysis revealed that both *p1-vv85* and its revertant *P1-rr'85* allele contain an identical transposon-like element inserted into the second intron of *p1* gene, 1539 bp downstream of exon2 (Figure 2A). This element is 1168 bp in length and shares characteristics of CACTA transposable element superfamily in plants: it contains the conserved 5'-CACTA-3' sequence at the termini of the 11 bp perfect terminal inverted repeats (TIRs, 5'-CACTACCGGAA-3') and it is bounded by 3 bp target site duplication (Figure 3). The 13 bp of the reversed 3'-end is identical to 5' TIR (5'-CACTACCGGAATT-3') of a maize transposable element *Mpi*I which is also a member of CACTA transposable element superfamily (Weydemann et al. 1988). This insertion element also has structured subterminal repetitive regions with multiple copies of direct or inverted repeats. These repeats share a 12 bp consensus sequence (5'-CTTTGCCGAGTG-3') with 12 copies in the 220 bp region at the 5'-end and 6 copies in the 170 bp region at the 3'-end

(Figure 3). Maize *En/Spm* transposable element also has a highly structured subterminal repetitive region with multiple copies of a 12 bp consensus motif which has been shown to be binding sites for transposon encoded regulatory protein TNPA (Grant et al 1993). This binding is important for the excision and genetic activity regulation of the transposon. A similar organization of multiple copies of a sequence motif is also observed in several other CACTA transposable elements (Kunze and Weil 2002). Comparing the sequence of the 12 bp motif in the CACTA insertion fragment in *p1-vv85* with the consensus sequences in other CACTA elements revealed no significant similarity. Analysis of the insertion sequence in *p1-vv85* didn't reveal any large ORF. Considering the size of the element it is suggested that this element is a non-autonomous transposable element. Southern blot hybridization using the CACTA insertion fragment as probe detected multiple bands in both *P1-rr4B2* and *p1-vv85* (data not shown), indicating that this insertion element has multiple copies of homologous sequences in the maize genome. According to the result of public database searching, about 550 bp internal region of the insertion element in *p1-vv85* and *P1-rr'85* shares high sequence similarity (>92%) with a region in several maize *shrunken 2* locus alleles. An internal region of a maize transposable element *ILS-1*, which was isolated from the *shrunken 2* locus, also shares 95% homology with this 550bp sequence (Alrefai et al. 1994; Figure 3). The *ILS-1* element is thought to be a non-autonomous element and shares no homology to other known plant transposable elements.

Transcription analysis of *p1-vv85* and *P1-rr'85*

Because no *p1* locus polymophism was detected between *p1-vv85* and its fully pigmented red revertant *P1-rr'85*, we asked whether the variegated pigmentation phenotype observed in *p1-vv85* is indeed due to instability of *p1* locus. To answer this question we used several techniques to examine the expression of *p1* gene in *p1-vv85* and its revertant, including RNA gel blot analysis, nuclear run-on transcription analysis and RT-PCR. First we compared the steady-state *p1* transcripts level in *p1-vv85* and the revertant allele *P1-rr'85*.

Previous research has shown that *p1* gene produces two major transcripts, 1.8 kb and 1.0 kb in size, due to alternative splicing of a 7 kb precursor transcript (Grotewold et al.1991). The 1.8 kb transcript-encoded protein itself can activate the transcription of a subset of structural genes such as *a1* and *c2* required for synthesis of red phlobaphene pigments. As shown in Figure 4, Northern blot of total RNA from pericarp of *p1-vv85*, *P1-rr'85* and progenitor *P1-rr107B* alleles at stage 18 days after pollination were hybridized sequentially with *p1* cDNA, *a1* cDNA, *c2* cDNA and *tubulin* cDNA. Total RNA from *p1-ww1112* in which *p1* gene coding region has been deleted (Athma et al. 1991) and *P1-rr4B2* were also included in Northern blot analysis as negative and positive controls, respectively. As in previous reports, no *p1* transcripts were detected in *p1-ww1112* pericarp (Figure 4, lane 1 for *p1* cDNA). While in *P1-rr4B2* the 1.8 kb and 1.0 kb mRNA as well as the 7 kb unspliced transcripts of *p1* gene were observed (lane2). Two major bands with size similar to 7 kb and 1.8 kb were observed in *P1-rr107B*; some RNA bands with size smaller than 1.8 kb were also present in *P1-rr107B,* but the major 1.0 kb transcript in *P1-rr4B2* is reduced or absent (lane 3, *p1* cDNA). This result is consistent with the Southern blot data indicating a 1.2 kb deletion at the 3'-end of the *p1* gene in *P1-rr107B*. This deletion deletes the third exon of the alternatively-spliced1.0 kb transcript, but has no effect on the 1.8 kb transcript (Figure 2A). A similar level of 1.8 kb transcripts was also detected in *P1-rr'85* but the 7 kb unspliced transcripts now shifted up to approximately 9 kb (lane 4). This suggests that the 1168 bp insertion element in *P1-rr'85* is transcribed along with the *p1* gene and is spliced out together with intron 2. Significant levels of *a1* and *c2* gene transcripts were observed in both *P1-rr107B* and *P1-rr'85*, indicating that *p1* transcripts encoded proteins in this two alleles functioned normally as transcription activator. This result further confirmed that the small 1.0 kb transcript of *p1* gene is not necessary for P regulatory function (Grotewold et al. 1994). Comparing to the *p1* transcripts in *P1-rr'85*, two transcripts with different size and much reduced levels were detected in pericarp from *p1-vv85* (lane5). The sizes of the two transcripts are approximately 2.2 kb and 1.5 kb. These *p1* transcripts do not encode functional regulator because neither *a1* nor *c2* transcripts can be detected in *p1-vv85* (Figure 4). Taken together, these results indicate that the primary lesion in *p1-vv85* is the disruption of *p1* transcription.

The different level of steady-state transcripts between *p1-vv85* and *P1-rr'85* observed in Northern blot analysis may reflect the result of differences in transcription rate or transcript stability. To distinguish between these possibilities, we used nuclear run-on transcription analysis. Nuclei from pericarp of plants with genotype *p1-vv85/p1-www* (A619) and *P1-rr'85/p1-www* (A619) were extracted. Radioactively-labeled nascent transcripts were hybridized with slot blots containing fragments of 5'-*p1* cDNA, 3'- *p1* cDNA, *c2* cDNA, *a1* cDNA, *ubiquitin* cDNA and plasmid pUC18 (Figure 5A, Materials and Methods). The relative transcription rates of these genes in *p1-vv85* and *P1-rr'85* are compared in Figure 5B. When hybridized with 5'-*p1* cDNA, similar levels of transcripts were detected in *p1-vv85* and *P1-rr'85* (Figure 5); this indicates that transcription in *p1-vv85* initiated at a similar rate as in *P1-rr'85*. However, when hybridized with the 3'- *p1* cDNA, transcripts are detected with *P1-rr'85*, but not *p1-vv85* (Figure 5). This indicates that transcription elongation was blocked in *p1-vv85*. The difference in the production of full-length *p1* transcripts was also reflected in the different transcription rates of *c2* and *a1* genes, both of which require a functional P regulator for their transcription activation. The transcription rates of *a1* and *c2* were both increased in *P1-rr'85* comparing to those in *p1-vv85*, though the transcription rate of *c2* was increased much more significantly than that of *a1* (Figure 5B). A similar difference was also observed in other report (Chopra et al. 2003) and this may reflect a differential response of *c2* and *a1* genes in transcriptional activation by *p1*.

Transcription in *p1-vv85* terminates prematurely in the vicinity of the CACTA element insertion site

Previous studies of maize *En/Spm* and *dSpm* transposable elements revealed that TNPA, one of element encoded protein, can bind to sequences in the subterminal repetitive regions of *dSpm* elements and function as a trans-acting suppressor to repress gene expression (Gierl 1996). A model had been proposed that TNPA binding hindered RNA polymerase elongation and thus terminated transcription prematurely (Schwarz-Sommer et al. 1985; Gierl 1996). The nuclear run-on transcription analysis in this study indicates that transcription in *p1-vv85* initiates normally but is apparently blocked between exon 2 and 3. This observation suggests

that the CACTA insertion element in *p1-vv85*, like *dSpm*, may bind to a suppressor factor(s) that may be encoded by an active autonomous transposable element present in the genome. Binding of the suppressor might block *p1* transcription in the vicinity of the CACTA element insertion site in *p1-vv85*. To determine the transcription termination site in *p1-vv85* we then performed 3' RACE (Frohman et al.1988). Total RNAs for RNA gel blot analysis were reverse transcribed using oligo-dT adapter primer (Material and Methods). The converted first strand cDNAs were then amplified using *p1* specific primers (Materials and Methods, Figure 6A). The PCR products were then fractionated by agarose gel electrophoresis. As shown in Figure 6B panel I, primers yb-231f and yb-737r amplified a 383 bp product in progenitor allele *P1-rr107B* (lane 2), while in *p1-ww1112* which has a *p1* deletion, no bands were detected (lane 1). A weak band at size 506 bp was also observed in *P1-rr107B*. The maize *p1* gene contains a 108 bp intron I and primers yb-231f and yb-737r which are flanking intron 1 will amplify a 506 bp product from genomic DNA (Figure 6A). However, since no PCR-products could be detected on ethidium bromide gel when first strand reactions were performed without reverse transcriptase (data not shown), such 506 bp should not come from genomic DNA contamination. We conclude that the 506 bp product is derived from the unspliced 7 kb precursor *p1* transcripts which were detected by RNA-gel analysis (Figure 4). Amplification in *P1-rr'85* yielded a 383 bp band (lane 3) as well as the 506 bp band at a similar level compared to *P1-rr107B*. In *p1-vv85,* the 383 bp PCR product was also observed though at a much reduced level relative to that of *P1-rr'85* (lane 4). In addition to the weak 383 bp band, a 506 bp band in *p1-vv85* could be detected. These results confirm that *p1-vv85* has a reduced level of *p1* transcripts including unspliced precursor mRNA. To examine the 3'-end of *p1* transcripts in these alleles, the first strand cDNAs were then amplified using primers in exon 3 of the *p1* gene. From the result shown in Figure 6B panel II, products of 400 bp size are observed in both *P1-rr107B* (lane 2) and *P1-rr'85* (lane 3). In *p1-vv85* a very weak band of similar size can be detected on ethidium bromide gel (lane 4) and was

confirmed to be *p1* sepecific by hybridization of the PCR products with a *p1* cDNA probe (data not shown). These resuls suggest that *p1-vv85* may produce a very low level of full-length *p1* mRNA which could not be detected by Northern blot or in nuclear run-on transcription analysis. Such low level of *p1* transcripts apparently cannot support normal P1 transcriptional activation function.

To further test whether transcription elongation in *p1-vv85* is blocked in the vicinity of CACTA element insertion site we then performed 3' RACE with total RNAs treated with polyA polymerase before reverse transcription. The first-strand cDNAs were then PCR-amplified using a primer specific for *p1* gene intron 2, together with a 3'-RACE adapter primer (Figure 6A). As shown in figure 6B panel III, a weak band with size about 600 bp was observed in both *P1-rr'85* and *p1-vv85* but not detected in *P1-rr107B*. In contrast to the low level of PCR-amplified products of *p1-vv85* in panel I and panel II relative to those of *P1-rr'85*, the level of the 600 bp product in *p1-vv85* in panel III was higher than that in *P1-rr'85*. PCR amplification using another *p1* specific primer in intron 2 produced similar results (panel IV), indicating that more premature or truncated transcripts were present in *p1-vv85*. The ~600bp band was then purified, cloned and sequenced. Among 9 sequenced independent clones, 6 have *p1* transcripts terminating around CACTAC $\pm$ 3 bp at the 5'-end of the insertion element, 1 ends at 110 bp inside the element end, and the other 2 clones end at 24 bp and 53 bp, respectively, 5'- upstream of the element insertion site (Figure 6C). Since any premature transcripts are expected to be unstable and degraded soon, these RT-PCR clones may represent some of the degraded *p1* transcripts in *p1-vv85*. However, the relative abundance of transcripts terminated around the CACTA element insertion site in *p1-vv85* but not in *P1-rr'85* suggested that in *p1-vv85* the transcription elongation of *p1* gene is blocked in the vicinity of the 5'end of the CACTA element.

Genetic Analysis of *p1-vv85* and its revertant allele *P1-rr'85*

Previous studies have shown that the *En/Spm* encoded proteins are important for the element's "suppression and mutation" functions. For example, the presence of an active *En/Spm* in the genome can completely suppress the expression of some genes with *I/dSpm* elements inserted. The absence of *En/Spm*, either due to segregation or inactivation, will allow some level of expression of these *I/dSpm*-containing genes (Gierl 1996, Kunze and Weil 2002). The results described above suggest that the variegated phenotype in *p1-vv85* might due to the insertion of a CACTA element in the second intron of *p1* gene. Because this insertion element is still present in the fully pigmented revertant allele *P1-rr'85*, *p1* expression observed *in P1-rr'85* cannot be attributed due to excision of the insertion element. One possibility for this non-excision reversion mechanism is that *p1* expression in *P1-rr'85* is also due to segregation or inactivation of a suppressor-encoding element, while the null mutant phenotype observed in *p1-vv85* is due to the continuing presence of the active suppressor element. To test this hypothesis we performed a genetic analysis of *p1-vv85* and its related alleles. If the hypothesis is true, then re-introducing the trans-acting element by a genetic cross into *P1-rr'85* should restore the variegated mutant phenotype.

Plants homozygous for *p1-vv85* were outcrossed as female parents with plants homozygous for *P1-rr'85*. Since pericarp is maternal tissue, the phenotype of the progeny kernels is consistent with the genotype of female parent. All the ears from such crosses showed light variegated pericarp just as that of the parent *p1-vv85* allele. Kernels from the resulting ears were planted and more than 200 progeny ears were scored. The results are summarized in Figure 7. All the progeny heterozygous ears showed red pericarp and red cob. Some kernels on the progeny ears had very small light pigmented stripes; approximately ~ 60 out of ~ 30,000 kernels showed larger sectors (less than half of the whole kernel) of colorless or red/white variegated pericarp. Representative progeny ears are shown in Figure 7. Similar results were also observed in ears resulting from reciprocal crosses in which homozygous *p1-vv85* plants were crossed as males with *P1-rr'85*. This result that all progeny ears are of

P-rr phenotype is contrary to the prediction of the model postulating the presence of an active suppressor in *p1-vv85*. However, the *p1* expression observed in those *p1-vv85/P1-rr'85* heterozygous ears could possibly result from inactivation of the putative suppressor element during kernel development. To test this, plants grown from red *p1-vv85/P1-rr'85* heterozygous kernels were either self-pollinated or outcrossed with inbred line A619 (Figure 7). Among the 41 progeny ears derived from self-pollinated ears, 38 of them showed red pericarp and red cob, and 3 ears had very light variegated pericarp and cob similar to the mutant phenotype of their grand parent *p1-vv85*. Among 59 ears from outcrossed plants, 35 showed red pericarp and cob and 24 ears had *p1-vv85* phenotype (Figure 7). According to the inactivation model, if *p1* expression in heterozygous *p1-vv85/P1-rr'85* plants is due to inactivation of a trans-acting element during plant development, then the restoration of the mutant phenotype should be the result of spontaneous reactivation of such element. The frequency of such reactivation must be low, as *p1* expression in revertant plants is highly stable. However, the low frequency of the *p1-vv85* mutant phenotype in a whole ear and the stable pigmentation in the *P1-rr* like ears suggested that the pigmentation observed in heterozygous kernels was not likely due to inactivation of a trans-acting element. On the contrary, the genetic analysis results that we got so far suggested that *p1-vv85* and *P1-rr'85* are segregating rather like independent alleles.

Correlation between mutant phenotype and demethylation at TIRs of insertion element

The genetic activity of transposable elements is correlated with DNA modifications. Plant transposable elements silencing, which has been observed in several maize transposons like *Spm*, *Ac* and *Mu*, has been shown to be associated with changes in methylation in certain regions in the elements (Hiroyuki and Hirochika 2001). The genetic and molecular analysis of *p1-vv85* and its revertant *P1-rr'85* allele in this study suggested that the *p1* gene expression in these two alleles might be regulated by an epigenetic mechanism. To examine whether the

mutant and revertant phenotype in *p1-vv85* and *P1-rr'85* are associated with any DNA modification, especially methylation change during kernel development, DNA samples from kernel pericarp of homozygous *p1-vv85* and *P1-rr'85* plants were digested with methylation-sensitive enzymes *Sal*I, *Msp*I and *Hpa*II. *Msp*I and *Hpa*II are isoenzymes with differential sensitivity to deoxycytosine methylation in CpG. The CACTA insertion element contains a *Msp*I site at each TIR end. Southern blot analysis revealed no difference between *p1-vv85* and *P1-rr'85* in *Sal*I digestion (data not shown). Southern blot analysis of DNA digested with *Msp*I and *Hpa*II showed that no methylation differences could be detected between *p1-vv85* and *P1-rr'85* in regions flanking the element insertion site (Figure 8). However when the blot hybridized with probe #8p, a 5.2 kb band was observed in *P1-rr'85* but not in *p1-vv85* (Figure 8A) or at a much reduced level (data not shown). Methylation of cytosine residue in CpG and CpNpG nucleotides in the two *Msp*I sites at the 5'- and 3'-TIR of insertion element in either *p1-vv85* or *P1-rr'85* will generate the 5.2kb fragment. Hybridization of blot with probe #9p (Figure 8A) confirmed that the *Msp*I sites in *P1-rr'85* were methylated, since a 5.2 kb band was again observed in *P1-rr'85* but absent in *p1-vv85*. These results indicate that the cytosine residues in both CpG and CpNpG sites at the 5'- and 3'-end TIRs in *P1-rr'85* are methylated in *P1-rr'85*, but not in *p1-vv85*. Similar results were obtained using pericarp DNA from other plants homozygous or heterozygous for *p1-vv85* and *P1-rr'85* (data not shown). However, the methylation analysis of leaf DNA from *p1-vv85 and P1-rr'85* didn't always give the same results as obtained using pericarp DNA. For example, some cytosine residues at the end of *p1-vv85* in leaf DNAs were also methylated, giving the same methylation pattern as in *P1-rr'85* in *Hpa*II digestion and a less methylated pattern in *Ms*pI digestion (data not shown). This suggested that the modification of CACTA element in *p1-vv85* might be the result of a demethylation process during kernel development.

DISCUSSION

CACTA insertion element in *p1-vv85* affects *p1* gene transcription

Here we describe the isolation and molecular analysis of a new maize mutant allele *p1-vv85* that showed both somatic and germinal instability in *p1* expression. We isolated a CACTA element in the second intron of the *p1* gene in both *p1-vv85* and its revertant allele *P1-rr'85*. Effects of transposable element (TE) insertion on gene expression are of great diversity and can cause both quantitative and qualitative phenotypic changes. Transposon insertion into regulatory or coding regions of a gene usually result in loss-of-function phenotypes. Whereas, transposon insertions into coding regions may still permit some level of gene expression due to the ability of insertion elements to splicing out as introns (Kim et el. 1987). Orientation of transposable element insertion relative to host gene transcription direction can also have different effects on gene expression (Bradley et al. 1993). Transposable elements can also bring transcription of target genes under the transposon's control. In some cases transposable elements can function as either enhancers or promoters for host genes, such as the maize *Mu1* element insertion in *hcf106* allele (Barkan Martienssen 1991). However, in a number of transposon-induced mutant alleles, the effects of transposable element insertion are dependent upon trans-acting factors. Extensive studies on maize *En/Spm* transposable element induced mutant alleles provided great information on the interaction of trans-acting factors and cis-elements in the corresponding transposons (Kunze and Weil 2002). Comparing the CACTA element in *p1-vv85* with other known members of plant CACTA transposable element family revealed that several features were conserved among these transposable elements. One striking structural feature is the presence of multiple copies of consensus sequences arranged as direct and inverted repeats in the subterminal regions. These arrangements create potential complex secondary structure. Though the consensus sequences in those identified transposons are not conserved, they may play similar

roles in transposition and regulation of transposon activity. In maize *En/Spm* elements, the transposon-encoded TNPA protein specifically binds to a subset of a 12 bp motif in subterminal regions (Grant 1990), and promotes and stabilizes the synapsis of the two ends of the transposon. The binding affinity of TNPA to DNA is not markedly affected by number of binding sites and is not cooperative. However DNA fragments with three or more binding sites formed intermolecular TNPA-DNA complexes cooperatively (Rina et al. 1998; Kunze and Weil 2002). TNPA contains both DNA-binding and dimerization domains (Trentmann et al. 1993). Dimerization is greatly stimulated by the presence of two 12 bp motifs in tail-to-tail arrangement and these two motifs are shown to be minimal cis-elements required for the *En/Spm* suppression function in transgenic tobacco protoplasts (Grant et al. 1990). Sequence analysis of another maize CACTA transposable element *Doppia* indicated that it encodes two proteins, DOPD and DOPA, with 68% and 47% similarity to TNPD and TNPA of *En/Spm*, respectively. DOPA was also shown to bind to a 14 bp subterminal repeat motif in the *Doppia* element ends and to promote the formation of intermolecular complexes (Bercury et al. 2001). Though no other CACTA elements have been reported to encode proteins with strong similarity to TNPA, an *Antirrhinum majus Tam1* encoded protein indeed was shown to bind to subterminal repeats (Nacken et al.1991). Binding of TNPA protein to the subterminal repetitive region is important for the *En/Spm* "suppressor and mutator" functions. Another CACTA element *Tpn1* in Japanese morning glory also has been found to have "suppressor" and "mutator" functions. In an *f3h* mutant allele induced by a defective *Tpn1* element insertion, the presence of an active element in the genome suppressed the pale yellow color in the flower to colorless (suppression) with excision-reversion sectors (mutation) (Hoshino et al. 1997).

Nuclear run-on and RT-PCR analyses in this study suggested that *p1* transcription elongation in *p1-vv85* was blocked in the vicinity of a CACTA element insertion site in the second intron of the *p1* gene, probably at the 5'end of insertion element. This result is

consistent with the hypothetical model for maize *En/Spm* "suppressor" function. Thus *p1-vv85* may be regarded as a "suppressible" allele. The TIRs and the subterminal consensus motifs that are arranged in direct and inverted orientation provide potential binding sites for trans-factors. Since the *p1-vv85* insertion element apparently does not encode any large transcripts, the suppressor function is probably encoded by a corresponding autonomous element in the genome. The similarity between the TIRs of *p1-vv85* insertion element with that of maize *MpiI* transposable element implies that the CACTA element in *p1-vv85* is related to *MpiI*. The *MpiI* element is about 9 kb long. Part of it has been cloned though the sequence has not been reported. Previous Northern blot analysis of *MpiI* element from a maize *c2-m3* mutant allele indicated that it encoded a 2 kb transcript (Weydemann et al. 1988). The *MpiI* element may be a good candidate for the suppressor-encoding element in *p1-vv85*. Analysis of *En/Spm* encoded TNPA in transgenic tobacco indicated that suppression of target gene transcription ranged from 85% to 99.5% (Bunkers et al. 1993). Though we detected some *p1* transcripts in *p1-vv85* pericarp in RT-PCR analysis, no detection of *p1* transcripts in Northern blot analysis indicated that only trace levels of full length *p1* transcripts are present in *p1-vv85*, and such low transcripts levels are not sufficient to produce any visible pigmentation in pericarp and cob.

In addition to the requirement of element-encoded trans-factors for *En/Spm*'s "suppressor and mutator" functions, the structure and size of *Spm* or *dSpm* elements can also affect their genetic activity. Integrity of TIRs and presence of subterminal repeat motifs are important for element's transposition. Deletion of nucleotides in either TIR or subterminal motifs will reduce excision rate (Kunze and Weil 2002). In our study we didn't detect any excision in *p1-vv85*. One possibility is that the *p1-vv85* allele used in this study is relatively stable and gives small red stripes on only about 4~5 kernels out of an ear. Some of these late somatic reversions may be the result of excision-reversion, but due to the low percentage of such revertant sectors it may not be detectable in our Southern or PCR analysis. Another

possibility concerns the integrity of cis-elements required for excision. Deletion and mutation of these cis-elements can completely abolish transposition (Xiao and Peterson 2002; Kunze and Weil 2002). Due to the lack of any knowledge about the corresponding autonomous element of the CACTA element in *p1-vv85*, it's not clear whether the TIRs and subterminal repetitive regions in this insertion element are intact or not. Thus this insertion element in *p1-vv85* may represent a non-mobile deletion-derivative of a previously mobilizable element.

DNA modification at the ends of insertion element in revertant *P1-rr'85* allele is correlated with reversion

An interesting characteristic of *p1-vv85* is that reversion from mutant to normal functional phenotype is not due to excision of the insertion element or genetic segregation of a possible trans-acting autonomous element. Methylation analysis of *p1-vv85* and several independent revertant *P1-rr'85* alleles using methylation sensitive enzymes suggested that change of the *p1* gene from mutant to normal results by relief of suppression of through an epigenetic mechanism.

Epigenetic regulation of gene expression and transposon activity has been observed in prokaryotes and eukaryotes including mammals and plants. DNA methylation is thought to be important in mammals for X-chromosome inactivation and genomic imprinting. DNA modification especially cytosine methylation in CpG and CpNpG in plants is also associated with heritable changes in gene expression and transposon silencing (Kakutani 2002). In her extensive studies of maize *Ac* and *Spm* transposable elements, McClintock noticed that sometime transposable elements could cycle between active and inactive form, which she referred to as "changes of phase". This inactivation is both reversible and heritable. An inactive element can also be reactivated transiently by an active element (McClintock 1958, 1971). Later, as transposable elements (TEs) were described in other organisms, cycling between active and inactive phases was also observed in other transposable elements,

including bacterial elements and transposons in other plant species. Subsequent molecular analysis indicates that changes in TEs are actually associated with DNA modification at certain regions in the elements, and that these changes affected both element transposition and target gene expression (Okamotoand Hirochika 2001; Craig et al. 2002). Recent evidence has provided additional linkes of DNA methylation to chromatin remodeling. In Arabidopsis, the *DDM1* gene encodes a protein similar to chromatin remodeling factor SWI2/SNF2. In a *ddm1* mutant induced hypomethylation background, several silent repeated sequences are reactivated. *CAC1*, which was similar to maize *En/Spm* and not mobile in wildtype background, became transcriptionally and transpositionally active in a *ddm1* mutant line (Miura et al. 2001).

Our analysis of *p1-vv85* and its functional revertant alleles *P1-rr'85* detected no structural difference between them in the *p1*-locus. Transcription in *p1-vv85* was blocked in the vicinity of a CACTA insertion element in the second intron of *p1* gene. But in *P1-rr'85*, which contained an identical insertion element at the same site and same orientation, *p*1 transcription was not affected and normal functional *p1* transcripts were detected both in Northern and RT-PCR analyses. We propose that a trans-acting factor encoded by an element at an independent locus in the genome binds to the ends of the CACTA insertion element in *p1-vv85* and blocks transcription. No such blocking occurs in *P1-rr'85*, and our genetic analysis indicated that this was not due to genetic segregation of a trans-factor. These results suggest that *p1* transcription in *p1-vv85* and *P1-rr'85* is regulated by an epigenetic mechanism. One possibility is the inactivation of the trans-acting encoding element. However when we crossed *p1-vv85* with *P1-rr'85* the F1 heterozygous ears gave P1-rr phenotype, not a *p1-vv* phenotype as predicted by the inactivation model. Additionally the active element is still present in the *P1-rr'85/ p1-vv85* heterozygous kernels, since p1-vv phenotype reappeared in the next generation when F1 heterozygotes were either selfed or outcrossed to other *p1*-gene alleles. But unlike the even pigmented P1-rr phenotype that we observed in

homozygous *P1-rr'85* kernels, the pigmentation in kernels with genotype *p1-vv85/P1-rr'85* did show some level of variation in which small light pigmented stripes appeared on the red pericarp. This indicated that *p1-vv85* allele contained some factors that could affect *p1* transcription in *P1-rr'85* late in the kernel development. Based on these results we propose a model that "phase" change occurred not in the trans-factor coding element but in the CACTA insertion element in *p1-vv85* and *P1-rr'85*. This "phase" change is associated with DNA methylation at the ends of CACTA insertion element that are the putative binding sites for trans-acting factors. As depicted in the model in Figure 9, DNA methylation at the ends of the CACTA insertion element in *P1-rr'85*, either at TIRs or subterminal repeat motifs or both, greatly reduces binding of the trans-acting factor. We propose that in *P1-rr'85*, no stable transposome complex are formed to interfere with transcription elongation of RNA polymerase through intron 2 of *p1* gene. The insertion element is transcribed along with the *p1* gene and spliced out with intron 2, generating mature normal *p1* transcripts. However, in *p1-vv85* the ends of the insertion element are unmodified and can bind to a trans-factor which suppresses *p1* gene transcription, generating truncated *p1* transcripts. This model is supported by methylation analysis of CACTA insertion element using methylation sensitive enzyme *Msp*I and *Hpa*II. As in Figure 8, cytosine residues of the CCGG sequence at the 5'- and 3'-TIRs of the CACTA insertion element are methylated in *P1-rr'85* but not in *p1-vv85*. Differential methylation may also occur at other cytosine residues in 12 bp repeat motif, CTTTGCCGAGTG, and in the subterminal repetitive region at both ends in the CACTA insertion element, but these are not disclosed in our analysis. It has been observed in both *Mu* and *Ac* maize transposable elements that the binding of transposon-encoded proteins to element sequences can be inhibited by cytosine methylation (Barkan and Martienssen 1991; Settles et al. 2001; Ros and Kunze 2001). Previous analyses of maize *En/Spm*-encoded TNPA factor showed that hemimethylation at the CCG sequence in its 12 bp motif reduced TNPA binding ability by 5 to 10 fold, and full methylation reduced the binding even further (Kunze

and Weil 2002). The data presented here are best explained by a model in which a trans-acting factor present in *p1-vv85* genome could be functionally analogous protein to *En/Spm*'s TNPA factor. Modification at the end of the TIRs of the insertion element, and not a change in activity of an autonomous element, is responsible for the suppression of p1-vv phenotype.

REFERENCES

Alrefai, R., Orozco, B., and Rocheford, T. (1994) Detection and sequencing of the transposable element ILS-1 in the Illinois long-term selection maize strains. Plant Physiol. 106: 803-804.

Athma, P., and Peterson, T. (1991) *Ac* induces homologous recombination at the maize *P* locus. Genetics 128: 163-173.

Athma, P., Grotewold, E., and Peterson, T. (1992) Insertional mutagenesis of the maize *P* gene by intragenic transposition of *Ac*. Genetics 131:199-209.

Barkan, A., and Martienssen, R.A. (1991) Inactivation of maize transposon *Mu* suppresses a mutant phenotype by activating an outward-reading promoter near the end of Mu1. Proc. Natl. Acad. Sci. USA. 88(8): 3502-3506.

Bercury, S.D., Panavas, T., Irenze, K., and Walker, E.L. (2001) Molecular analysis of the *Doppia* transposable element of maize. *Plant Mol Biol.* 47(3): 341-351.

Bonas, U., Sommer, H., and Saedler, H. (1984) The 17-kb *Tam1* element of *Antirrhinum majus* induces 3-bp duplication upon integration into the chalcone synthease gene. EMBO J. 5: 1015-1020.

Bradley, D., Carpenter, R., Sommer, H., Hartley, N., and Coen E. (1993) Complementary floral homeotic phenotypes result from opposite orientations of a transposon at the *plena*

locus of *Antirrhinum*. Cell 72(1): 85-95.

Bunkers, G., Nelson, O.E. Jr., and Raboy, V. (1993) Maize *Bronze 1: dSpm* insertion mutations that are not fully suppressed by an active *Spm*. Genetics 134: 1211-1220.

Chandler, V.L, Walbot, V. (1986) DNA modification of a maize transposable element correlates with loss of activity. Proc Natl Acad Sci U S A 83(6): 1767-1771.

Chopra, S., Athma, P., and Peterson, T. (1996) Alleles of the maize *P* gene with distinct tissue specificities encode myb-homologous proteins with C-terminal replacements. Plant Cell 8: 1149-1158.

Chopra, S., Athma, P., and Peterson, T. (1998) A maize Myb-homologue is encoded by a stable multicopy gene complex. Mol. Gen. Genet. 260:372-380.

Chopra, S., Brendel, V., Zhang, J., Axtell, J.D., and Peterson, T. (1999) Molecular characterization of a mutable pigmentation phenotype and isolation of the first active transposable element from *Sorghum bicolor*. Proc Natl Acad Sci U S A 96(26): 15330-15335.

Chopra, S., Cocciolone, S.M., Bushman, S., Sangar V., McMullen, M.D., Peterson, T. (2003) The maize *Unstable factor for orange1* is a dominant epigenetic modifier of a tissue specifically silent allele of *pericarp color1*. Genetics 163: 1135-1146.

Christensen, A.H., Sharrock, R.A., and Quail, P.H. (1992) Maize *polyubiquitin* genes: structure, thermal perturbation of expression and transcript splicing, and promoter activity following transfer to protoplasts by electroporation. Plant Mol Biol. 18(4): 675-689.

Cui, H., and Fedoroff, N.V. (2002) Inducible DNA demethylation mediated by the maize *Suppressor-mutator* transposon-encoded TnpA protein. Plant Cell 14(11): 2883-2899.

Dooner, H.K., Robbins, T.P., Jorgensen, R.A. (1991) Genetic and developmental control of anthocyanin biosynthesis. Annu Rev Genet. 25:173-199.

Dorweiler, J.E., Carey, C.C., Kubo, K.M., Hollick, J.B., Kermicle, J.L., Chandler, V.L. (2000)

Mediator of paramutation1 is required for establishment and maintenance of paramutation at multiple maize loci. Plant Cell. 12(11): 2101-2118.

Emerson, R.A. (1917) Genetical studies of variegated pericarp in maize. Genetics 141: 347-360.

Fedoroff, N.V. (1995) DNA methylation and activity of the maize *Spm* transposable element.Curr Top Microbiol Immunol. 197:143-164.

Fedoroff, N.V. (1999) The *Suppressor-mutator* element and the evolutionary riddle of transposons. Genes to Cells 4:11-19.

Feschotte, C., Jiang, N., Wessler, S.R. (2002) Plant transposable elements: where genetics meets genomics. Nat Rev Genet. 3(5): 329-341.

Frohman, M.A., Dush, M.K., and Martin, G.R. (1988) Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. Proc Natl Acad Sci U S A. 85(23): 8998-9002.

Frey, M., Reinecke, J., Grant, S., Saedler, H., and Gierl, A. (1990) Excision of the *En/Spm* transposable element of *Zea mays* requires two element-encoded proteins. EMBO J. 9: 4037-4044.

Gierl, A., Schwarz-Sommer, Z., and Saedler, H. (1985) Molecular interactions between the components of the *En-I* transposable element system of *Zea mays.* EMBO J. 4: 579-583.

Gierl, A., Lutticke, S., and Saedler, H. (1988) TnpA product encoded by the transposable element *En-I* of *Zea mays* is a DNA binding protein. EMBO J. 7: 4045-4053.

Gierl, A. (1996) The En/Spm transposable element of maize. Cuur. Top. Microbiol. Immunol. 204: 145-159.

Girard, L. and Freeling, M. (1999) Regulatory changes as a consequence of transposon insertion. Dev. Genet. 25: 291-296.

Girard, L., Freeling, M. (2000) *Mutator*-Suppressible alleles of *rough sheath1* and *liguleless 3* in maize reveal mutiple mechanisms for suppression. Genetics 154: 437-446.

Grant, S.R., Gierl, A., and Saedler, H. (1990) *En/Spm* encoded TnpA protein requires a specific target sequence for suppression. EMBO J. 9: 2029-2035.

Grant, S.R., Hardenack, S., Trentmann, S., and Saedler, H. (1993) Functional cis-element sequence requirements for suppression of gene expression by the TNPA protein of the *Zea mays* transposon *En/Spm*. Mol Gen Genet. 241(1-2): 153-160.

Greene, B., Walko, R., Hake, S. (1994) *Mutator* insertions in an intron of the maize *knotted1* gene result in dominant suppressible mutations. *Genetics* 138: 1275-1285.

Grotewold, E., Athma, P., and Peterson, T. (1991) Alternatively spliced products of the maize P gene encode proteins with homology to the DNA-binding domain of myb-like transcription factors. *Proc. Natl. Acad. Sci. USA* 88: 4587-4591.

Grotewold, E., Drummond, B. J., Bowen, B., and Peterson, T. (1994) The *myb*-homologous *P* gene controls phlobaphene pigmentation in maize floral organs by directly activating a flavonoid biosynthetic gene subset. Cell 76: 543-553.

Okamoto, H., and Hirochika, H. (2001) Silencing of transposable elements in plants. Trends Plant Sci. 2001 6(11):527-534.

Hoshino, A., Abe, Y., Saito, N., Inagaki, Y., and Iida, S.(1997) The gene encoding flavanone 3-hydroxylase is expressed normally in the pale yellow flowers of the Japanese morning glory carrying the speckled mutation which produce neither flavonol nor anthocyanin but accumulate chalcone, aurone and flavanone. Plant Cell Physiol. 1997 38(8): 970-974.

Inagaki, Y., Hisatomi, Y., Suzuki, T., Kasahara, K., and Iida S. (1994) Isolation of a *Suppressor-mutator/Enhancer*-like transposable element, *Tpn1*, from Japanese morning glory bearing variegated flowers. Plant Cell 6(3): 375-383.

Kakutani, T.(2002) Epi-alleles in plants: inheritance of epigenetic information over generations. Plant Cell Physiol. 43(10):1106-1111.

Kim, H.Y., Schiefelbein, J.W., Raboy, V., Furtek, D.B., and Nelson, O.E. Jr. (1987) RNA splicing permits expression of a maize gene with a defective *Suppressor-mutator* transposable element insertion in an exon. Proc Natl Acad Sci U S A. 84(16): 5863-5867.

Kloeckner-Gruissem, B. and Freeling, M. 1995) Transposition-induced promoter scrambling: A mechanism for the evolution of new alleles. Proc Natl Acad Sci U S A. 92:1836-1840.

Kunze, R., and Weil, C.F. (2002) The *hAT* and CACTA superfamilies of plant transposons. In Mobile DNA II: 565-610. (eds Craig, N.L., Craigie, R., Gellert, M. and Lambowitz, A.M., American Society for Microbiology Press, Washington D.C., 2002)

Lechelt, C., Peterson, T., Laird, A., Chen, J., Dellaporta, S.L., Dennis, E., Peacock, W.J., and Starlinger, P. (1989) Isolation and molecular analysis of the maize *P* locus. Mol Gen Genet. 219(1-2): 225-234.

Ludwig, S.R., and Wessler, S.R. (1990) Maize *R* gene family: tissue-specific helix-loop-helix proteins. Cell. 62(5): 849-851.

Masson P, Surosky R, Kingsbury JA, Fedoroff NV. (1987) Genetic and molecular analysis of the *Spm*-dependent *a-m2* alleles of the maize *a* locus. Genetics 117(1): 117-137.

Masson, P., Fedoroff, N.V. (1989) Mobility of the maize *Suppressor-mutator* element in transgenic tobacco cells. Proc. Natl. Acad. Sci. USA 86: 2219-2223.

Masson, P., Rutherford, G., Banks, J.A., Fedoroff, N.V. (1989) Essential large transcripts of the maize *Spm* element are generated by alternative splicing. Cell 58:755-765

Masson, P., Strem, P. and Fedoroff, N. (1991) The *tnpA* abd *tnpD* gene products of the *Spm* element are required for transposition in tobacco. Plant Cell 3: 73-85.

Martienssen, R., Barkan, A., Taylor, W.C., and Freeling, M. (1990) Somatically heritable

switches in the DNA modification of *Mu* transposable elements monitored with a suppressible mutant in maize. Genes Dev. 4(3): 331-343.

McClintock, B. (1948) Mutable loci in maize. Carnegie Inst .Wash. Year Book 47:155-169.

McClintock, B. (1954) Mutations in maize and chromosomal aberrations in *Neurospora*. Carnegie Inst .Wash. Year Book 53: 254-260.

McClintock, B. (1958) The *suppressor-mutator* system of control of gene action in maize. Carnegie Inst .Wash. Year Book 57: 415-429.

McClintock, B. (1971) The contribution of one component of a control system to versatility of gene expression. Carnegie Inst .Wash. Year Book 70: 5-17.

Miura ,A., Yonebayashi, S., Watanabe, K., Toyama, T., Shimada, H., and Kakutani, T. (2001) Mobilization of transposons by a mutation abolishing full DNA methylation in *Arabidopsis*. Nature 411(6834): 212-214.

Motohashi, R., Ohtsubo, E., and Ohtsubo, H. (1996) Identification of *Tnr3*, a *suppressor-mutator/enhancer*-like transposable element from rice. Mol Gen Genet. 250(2): 148-152.

Nacken, W.K., Piotrowiak, R., Saedler, H., and Sommer, H. (1991) The transposable element *Tam1* from *Antirrhinum majus* shows structural homology to the maize transposon *En/Spm* and has no sequence specificity of insertion. Mol Gen Genet. 228(1-2): 201-208.

Pereira,A., Schwarz-Sommer, Z., Gierl, A., Bertram, I., Peterson,P.A., and Saedler ,H. (1985) Genetic and molecular analysis of the *Enhancer*(*En*) transposable element system of *Zea mays*. EMBO J. 4(1):7-23.

Pereira, A., Cuypers, H., Gierl, A., Schwarz-Sommer, Z., and Saedler, H. (1986) Molecular analysis of the En/Spm transposable element system of *Zea mays*. EMBO J. 5:835-841.

Peterson, P.A. (1953) A mutable *pale green* locus in maize. Genetics 45:115-133.

Porebski, S., Bailey, L.G., and Baum, B.R. (1997) Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. Plant Mol. Biol. Reporter 15: 8-15.

Rhodes, P.R., and Vodkin, L.O. (1998) Organization of the *Tgm* family of transposable elements in soybean. Genetics 120(2): 597-604.

Ros, F., and Kunze, R. (2001) Regulation of activator/dissociation transposition by replication and DNA methylation. Genetics. 157(4):1723-1733.

Settles, A.M., Baron, A., Barkan, A., Martienssen, R.A. (2001) Duplication and suppression of chloroplast protein translocation genes in maize. Genetics. 157(1):349-360.

Schwarz-Sommer, Z., Shepherd, N., Tacke, E., Gierl, A., Rohde, W., Leclercq, L., Mattes, M., Berndtgen, R., Peterson, P.A., and Saedler, H. (1987) Influence of transposable elements on the structure and function of the *A1* gene of *Zea mays*. EMBO J. 6: 287-294.

Sidorenko, L.V., and Peterson, T. (2001) Transgene-Induced Silencing Identifies Sequences Involved in the Establishment of Paramutation of the Maize *p1* Gene. Plant Cell 13(2): 319-335.

Trentmann, S.M., Saedler, H., and Gierl, A. (1993) The transposable element *En/Spm*-encoded TNPA protein contains a DNA binding and a dimerization domain. Mol Gen Genet. 238(1-2): 201-208.

Villemur, R., Joyce, C.M., Haas, N.A., Goddard, R.H., Kopczak, S.D., Hussey, P.J., Snustad, D.P., and Silflow, C.D. (1992) Alpha-tubulin gene family of maize (Zea mays L.). Evidence for two ancient alpha-tubulin genes in plants. J Mol Biol. 227(1): 81-96.

Weydemann, U., Wienand, U., Niesbach-Klosgen, U., Peterson, P.A., and Saedler, H. (1988) cloning of the transposable element *Mpi1* from *c2-m3*. MNL 62: 48

Wienand, U., Weydemann, U., Niesbach-Klosgen, U., Peterson, P., and Saedler, H. (1986)

Molecular cloning of the *c2* locus of *Zea mays*, the gene encoding chalcone synthase. Mol. Gen. Genet. 203: 202-207.

Xiao, Y., and Peterson, T. (2002) *Ac* transposition is impaired by a small terminal deletion. Mol. Genet. Genomics 266: 720-731.

Zhang, J., and Peterson, T. (1999) Genome rearrangements by nonlinear transposons in maize. Genetics 153: 1403-1410.

Zhang, P., Wang, Y., Zhang, J., Maddock, S., Snook, M., and Peterson, T. (2003) A maize QTL for silk maysin levels contains duplicated Myb-homologous genes which jointly regulate flavone biosynthesis. Plant Mol Biol. 52(1): 1-15

FIGURE LEGENDS

Figure 1. Representative kernel phenotypes of *p1-vv85* and its revertant *P1-rr'85* allele.

Figure 2. Southern analysis of *p1* gene structure in *p1-vv85*, revertatant *P1-rr'85* and progenitor *P1-rr107B*.

A. Structures of *P1-rr4B2*, *P1-rr107B*, *p1-vv85* and *P1-rr'85* alleles. The hatched boxes indicate the 5.2 kb direct repeats flanking *p1* gene. Two alternatively spliced *p1* transcripts are indicated at top with exons (filled boxes) connected by introns (thin lines). The black boxes and the patterned box indicated at bottom represent fragments #15 and #8B, respectively, used as probes in Southern analysis. The triangle in *p1-vv85* and *P1-rr'85* indicates the insertion of CACTA element. The three thin lines in the middle show the size of fragments obtained by digestion with *BamH*I (B), *Sal*I (S) and *Xho*I (X).

B. Southern blots of leaf genomic DNA digested with restriction enzyme *BamH*I, *Sal*I, *Xho*I and *Sal*I plus *Kpn*I. The *p1* locus fragment 15 and 8B used as probes are indicated at the bottom of each blot. The lanes are designed as follows: 1= *P1-rr4B2*; 2= *P1-rr107B*; 3= *P1-rr'85*; 4= *p1-vv85*.

Figure 3. Structure of the CACTA insertion element in *p1-vv85* and *P1-rr'85*.

The sequences flanking the insertion site in *p1-vv85* and progenitor *P1-rr107B* are shown under the enlarged insertion element. The 3 bp target site duplication is indicated underlined. Hatched arrows represent the 12 bp repeat motifs.

Figure 4. Northern blot analysis of steady-state *p1* transcripts.

Northern blot of total RNA from 18 DAP pericarp from plants homozygous for *p1-ww1112*, *P1-rr4B2*, *P1-rr107B*, *P1-rr'85* and *p1-vv85* are hybridized with probes *p1* cDNA, *c2* cDNA and *a1* cDNA sequentially. The quantity and quality of total RNA from each sample are monitored by hybridization with *tubulin* (*tub*) cDNA.

Figure 5. Nuclear run-on analysis of *p1* transcription.

A. Slot blot with target gene fragments are arranged as: *p1*-5'= 5'-UTR plus exon 1 plus exon 2 plus ~100 bp of exon 3 of *p1* cDNA; *p1*-3'= exon 3 of *p1* cDNA; *a1*= *a1* cDNA; *c2*= *c2* cDNA; *ubi*= *ubiquitin* cDNA; pUC18= linealized pUC 18 plasmid. Slot blot is hybridized with nascent transcripts from 17 DAP pericarp of *p1-vv85/p1-www* (A619) and *P1-rr'85/p1-www* (A619).

B. Hybridization signals were quantified by phophorimager and normalized to *ubiqutin* level (set as 1). Relative transcripts level of each target genes are calculated and represented at y-axis.

Figure 6. RT-PCR analysis of *p1* transcription.

A. Locations of primers in *p1-vv85*. Hatched box represents 5'-UTR, black boxes indicate exons. Triangle indicates insertion site of the CACTA element in *p1-vv85*. F indicates forward primer and R indicates reverse primer. Four pairs of *p1* locus primers used in RT-PCR are designed as follows:

I. F=yb-231f, R=yb-737r;

II. F= yb-6245f, R=3'RACE adapter;

III. F=yb-1803f, R= R=3'RACE adapter;

IV. F=yb-2135f, R=3'RACE adapter;

and V. primers tub-1 and tub-2 used for amplification of *tubulin* cDNA as normalization control.

B. RT-PCR products from total RNA from pericarp of indicated *p1* alleles are fractioned on ethidium bromide gel. Lane1=*p1-ww1112*, lane 2=*P1-rr107B*, lane 3=*P1-rr'85*, lane *4=p1-vv85*, M=100 bp DNA ladder.

C. Sequences of the independent clones of the ~600 bp PCR products in Panel III lane 4. The terminating site of the *p1* transcripts in each clone is indicated by an asterisk.

Figure 7. Diagram of crosses for genetic analysis of *p1-vv85* and *P1-rr'85*.

Representative ears show pigmentation phenotypes. RR indicates ears with red pericarp and red cob, VV represents variegated pericarp and cob. The *p1-www* (A619) is the inbred line A619 which showed white pericarp and white cob. Variegation in *p1-vv85* is very light with small red stripes on only about 4-5 kernels per ear.

Figure 8. Methylation analysis of *p1-vv85* and *P1-rr'85*.

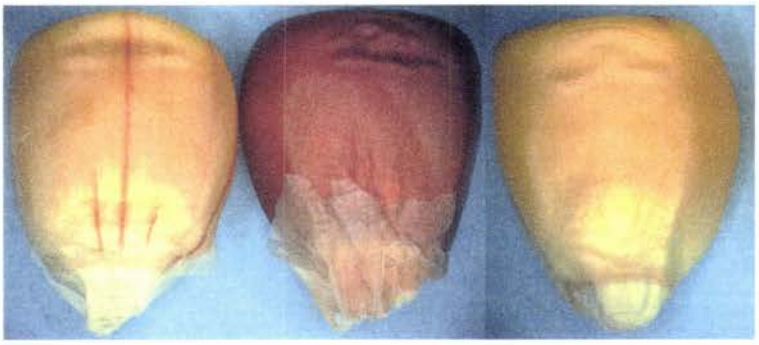
A. Methylation map of *p1-vv85* and *P1-rr'85*. Vertical lines represents *Msp*I/*Hpa*II sites. Open circles indicate unmethylated sites, black circles indicates methylated sites. Not all sites are shown in regions with clustered sites. Open boxes with numbers represent *p1* locus fragments used as probes. Triangle indicates insertion of CACTA element in *p1-vv85* and *P1-rr'85*.

B. Southern blot analysis of genomic DNA from pericarp of homozygous plants of *p1-vv85* and *P1-rr'85*. DNAs were digested with *Msp*I (M) and *Hpa*II (H) and hybridized with *p1* locus fragments as indicated on the bottom of each blot. The bands resulting from methylation of *Msp*I/*Hpa*II sites in *P1-rr'85* are indicated by arrows.

Figure 9. Model for epigenetic regultion of CACTA insertion element and its effect on *p1* gene transcription in *p1-vv85* and *P1-rr'85*.

Hatched boxes represent the 5.2 kb direct repeats flanking the *p1* gene. Solid boxes indicate

exons. Bent arrows indicate transcription start site in *p1* gene. Wavy lines represent *p1* transcripts. Triangle represents CACTA insertion element in *p1-vv85* and *P1-rr'85*. Solid line at the right corner represents a potential trans-factor encoding element. The trans-factors are represented by open circles. It is proposed that in *p1-vv85*, trans-factors bind to the end of the CACTA insertion element in intron 2 of the *p1* gene and block *p1* transcription elongation. However in *P1-rr'85*, methylation of the ends of the insertion element (asterisks) inhibits trans-factor binding, allowing *p1* gene transcription and the production of functional *p1*-encoded products (solid circles).



p1-vv85 *P1-rr'85* *p1-vv85*

Figure 1.

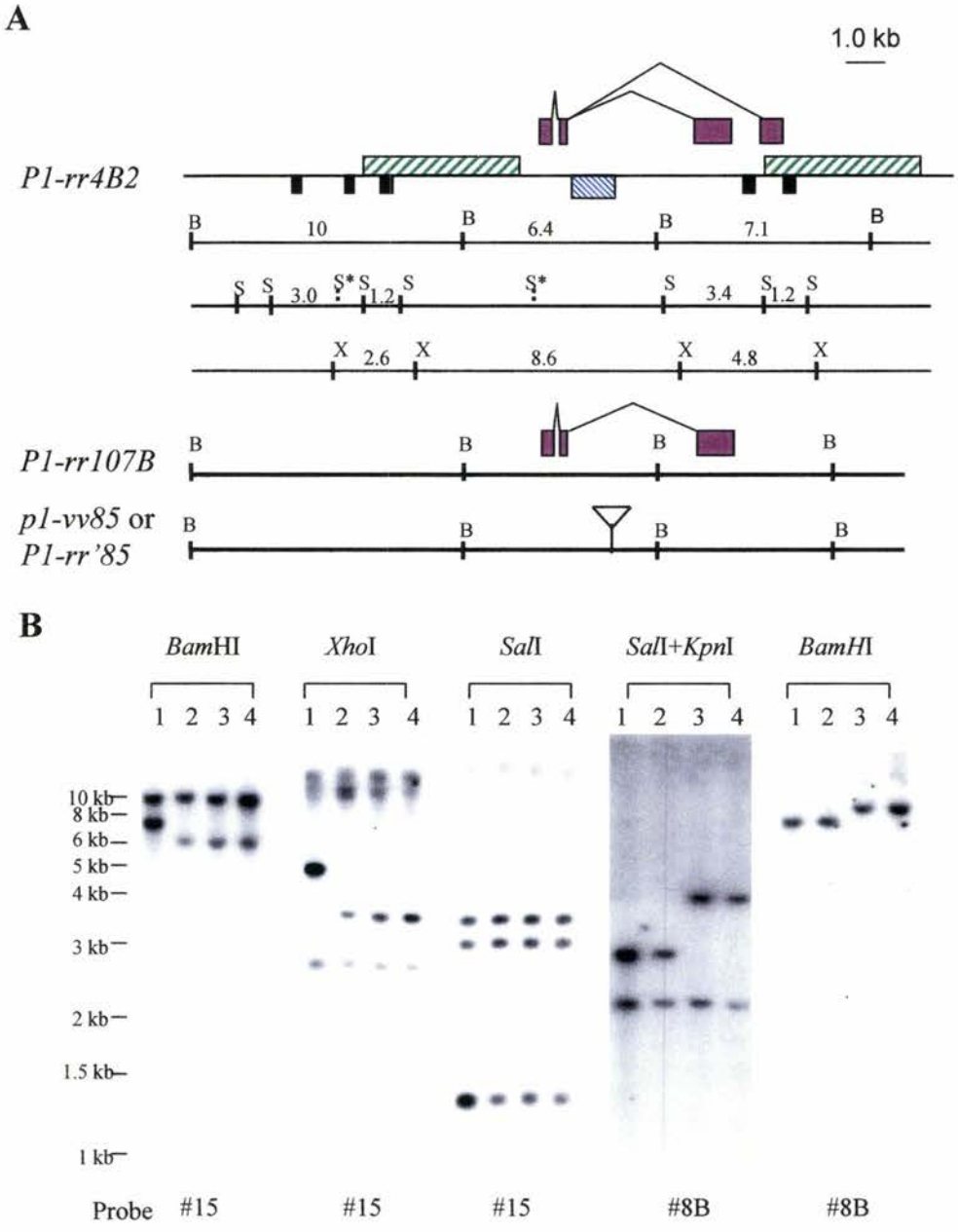

A
1.0 kb
P1-rr4B2
B
10
6.4
7.1
S
3.0
1.2
3.4
X
2.6
8.6
4.8
P1-rr107B
p1-vv85 or
P1-rr'85
B
BamHI
XhoI
SalI
SalI+KpnI
BamHI
1 2 3 4
10 kb
8 kb
6 kb
5 kb
4 kb
3 kb
2 kb
1.5 kb
1 kb
Probe
#15
#15
#15
#8B
#8B

Figure 2.

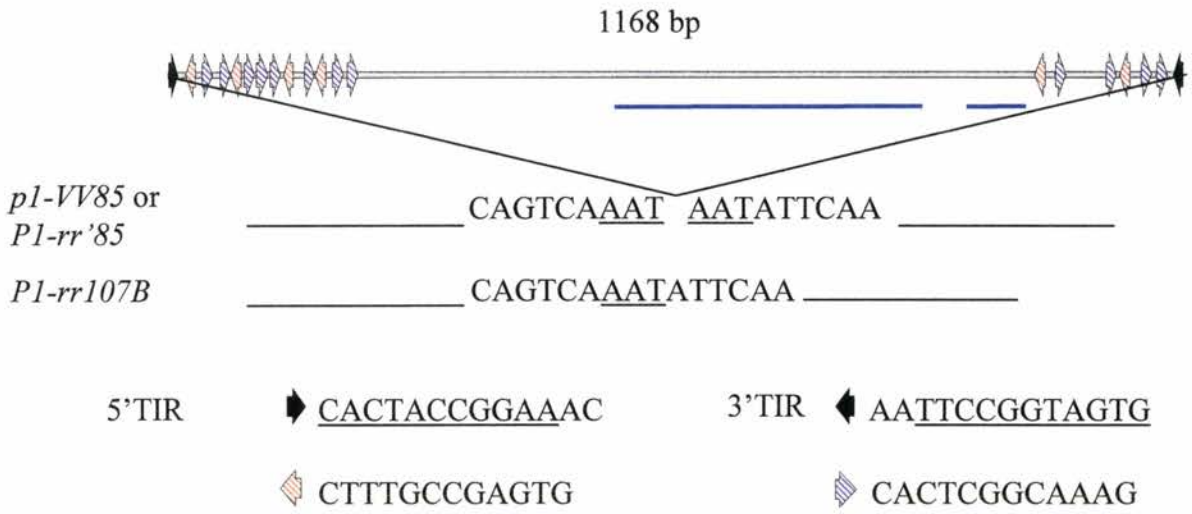

1168 bp
p1-VV85 or
P1-rr'85
CAGTCAAAT AATATTCAA
P1-rr107B
CAGTCAAATATTCAA
5'TIR
CACTACCGGAAAC
3'TIR
AATTCCGGTAGTG
CTTTGCCGAGTG
CACTCGGCAAAG
Regions of homology to putative maize transposable element ILS-1.

Figure 3.

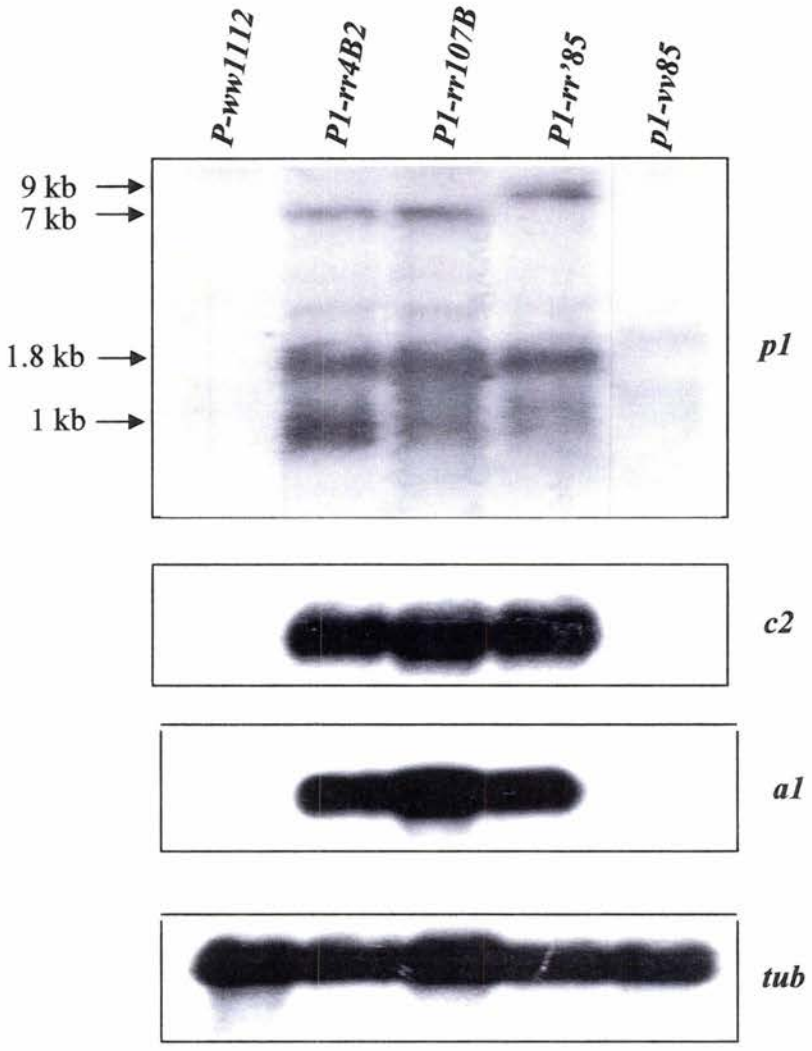

P-ww1112
P1-rr4B2
P1-rr107B
P1-rr'85
p1-vv85
9 kb
7 kb
1.8 kb
1 kb
p1
c2
a1
tub

Figure 4.

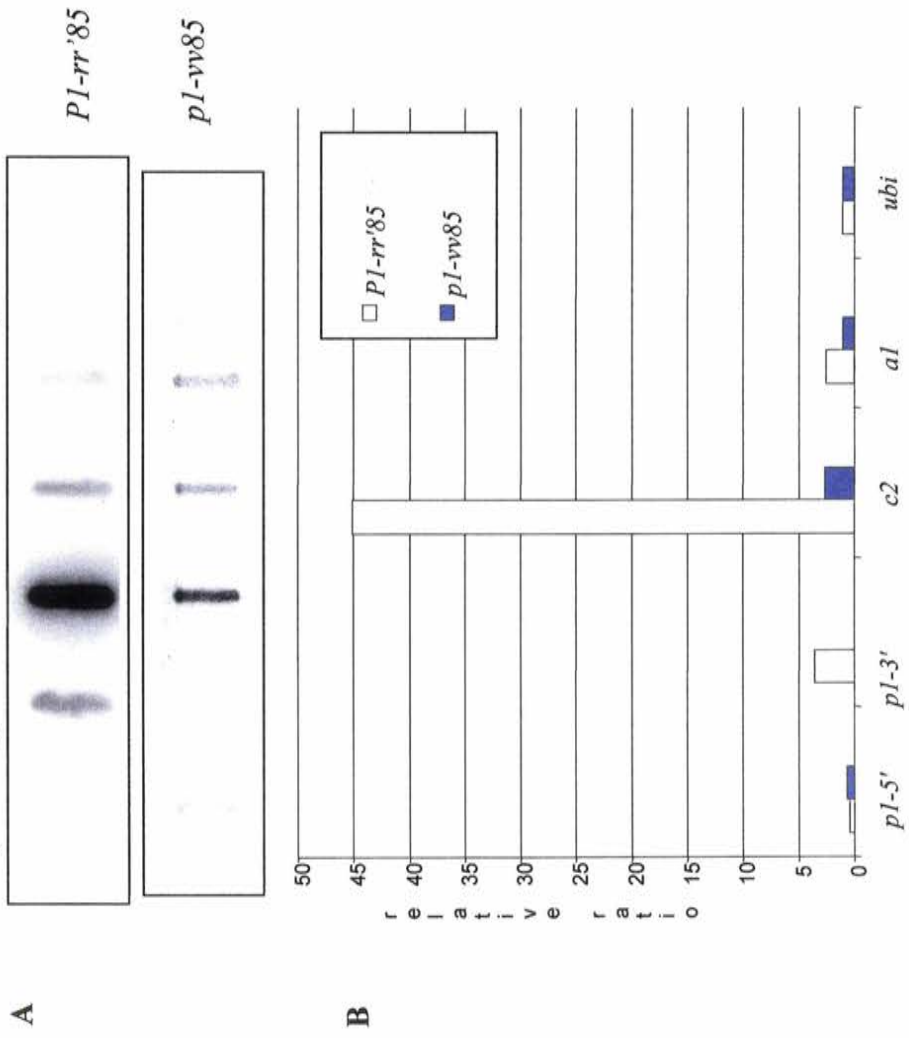

A
p1-5'
p1-3'
c2
a1
ubi
pUC18
P1-rr'85
p1-vv85
B
P1-rr'85
p1-vv85
relative ratio
50
45
40
35
30
25
20
15
10
5
0
p1-5'
p1-3'
c2
a1
ubi

Figure 5.

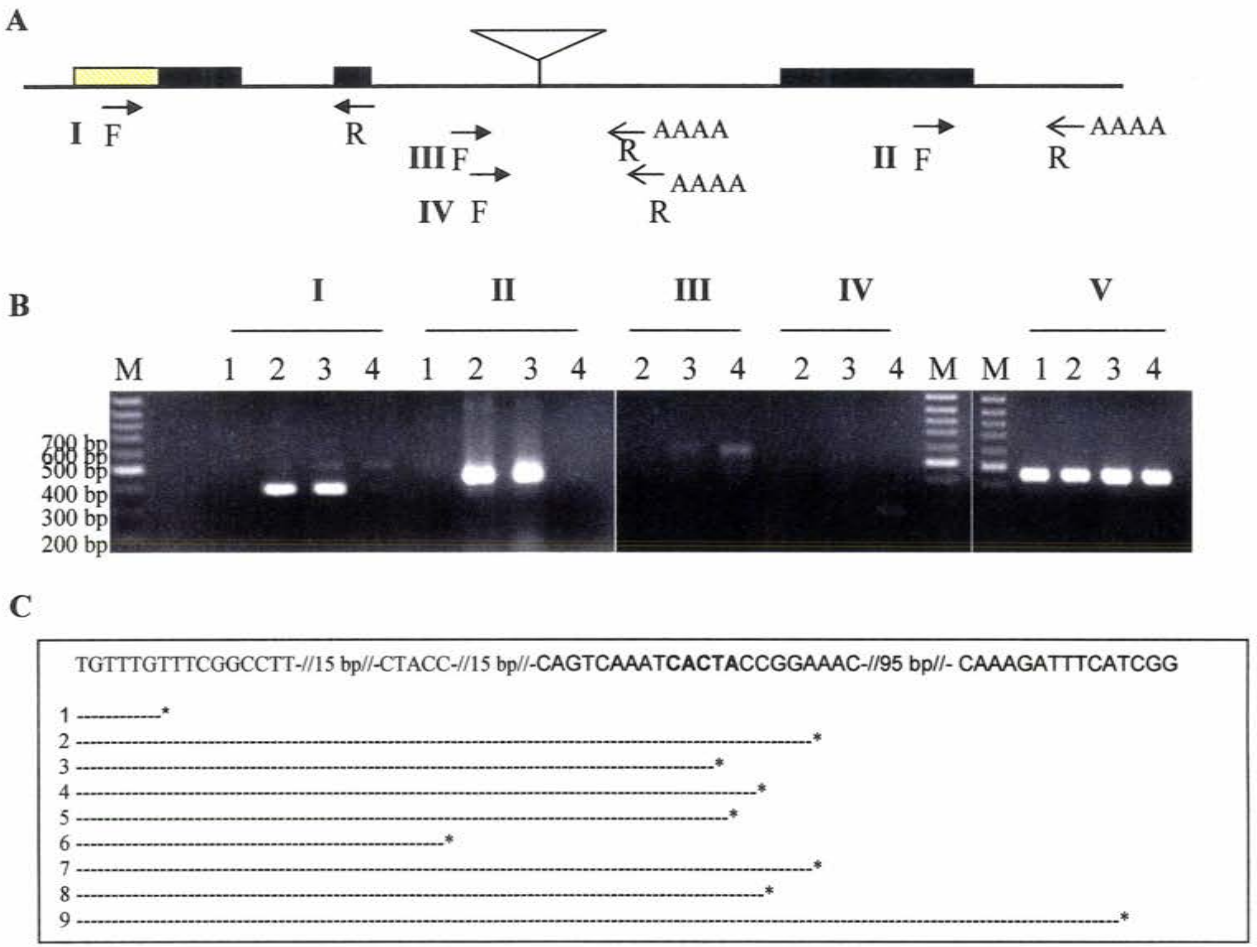

A
I F
R
III F
IV F
AAAA
R
AAAA
R
II F
AAAA
R
B
I
II
III
IV
V
M 1 2 3 4 1 2 3 4 2 3 4 2 3 4 M M 1 2 3 4
700 bp
600 bp
500 bp
400 bp
300 bp
200 bp
C
TGTTTGTTTCGGCCTT-//15 bp//-CTACC-//15 bp//-CAGTCAAATCACTACCGGAAAC-//95 bp//- CAAAGATTTCATCGG
1 ----------*
2 ----------*
3 ----------*
4 ----------*
5 ----------*
6 ----------*
7 ----------*
8 ----------*
9 ----------*

Figure 6

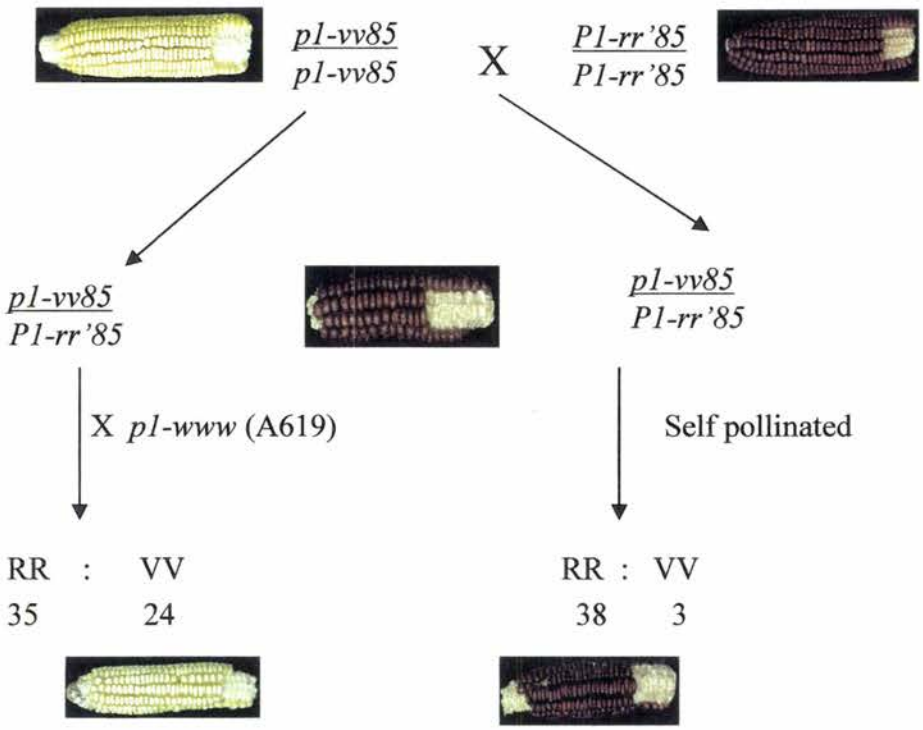

p1-vv85
p1-vv85
X
P1-rr'85
P1-rr'85
p1-vv85
P1-rr'85
p1-vv85
P1-rr'85
X p1-www (A619)
Self pollinated
RR : VV
35 24
RR : VV
38 3

Figure 7.

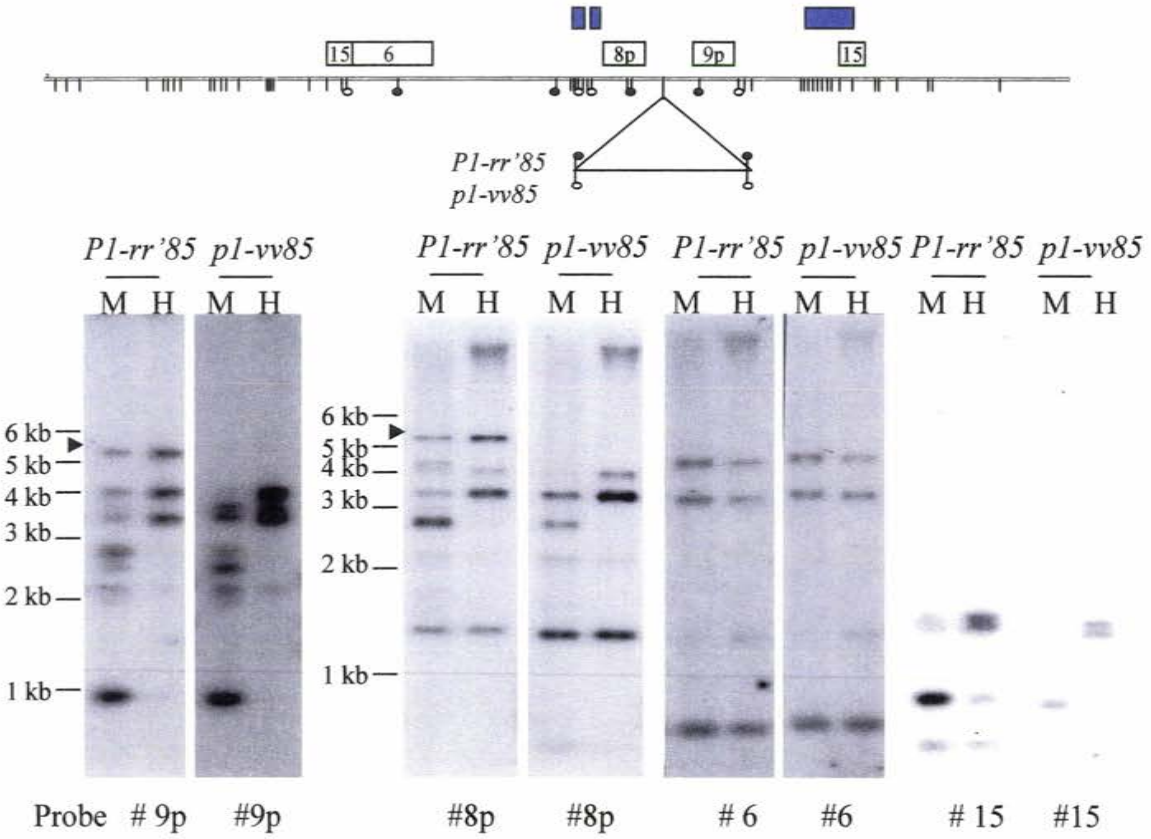

15
6
8p
9p
15
P1-rr'85
p1-vv85
P1-rr'85 p1-vv85
M H M H
6 kb
5 kb
4 kb
3 kb
2 kb
1 kb
P1-rr'85 p1-vv85 P1-rr'85 p1-vv85 P1-rr'85 p1-vv85
M H M H M H M H M H M H
6 kb
5 kb
4 kb
3 kb
2 kb
1 kb
Probe # 9p #9p #8p #8p # 6 #6 # 15 #15

Figure 8.

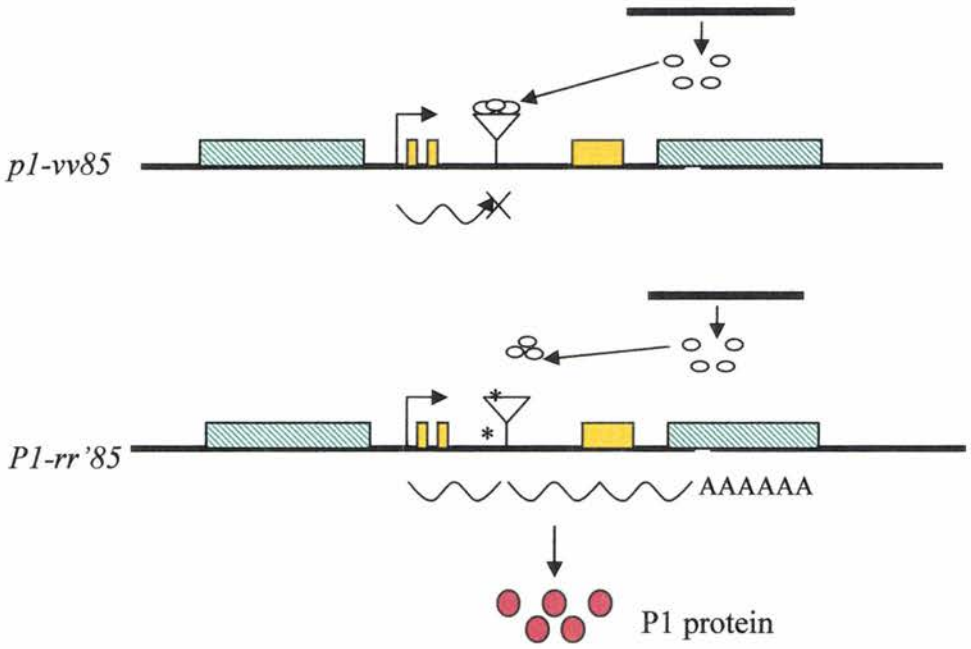


Figure 9.

CHAPTER 3. FUNCTIONAL ANALYSIS OF THE PROMOTER OF THE MAIZE *Myb*-HOMOLOGOUS *p2* GENE

A paper to be submitted to *Plant Physiology*

Yibin Wang and Thomas Peterson

ABSTRACT

The maize *p2* gene encodes a Myb-like transcription factor which activates transcription of a subset of structural genes of the flavonoid biosynthetic pathway. The *p2* gene is expressed in certain floral organs including husk, anther and silk, where it co-regulates the synthesis of a maize host resistant factor, maysin. Here we report the functional analysis of the *p2* gene promoter in both transient and stable transformation assays. In transient assays a series of plasmids with *p2* promoter segments fused to a GUS reporter gene (p2-promoter::GUS) were transformed into maize culture suspension cells and fresh silks by particle bombardment. The TATA box region in the *p2* gene promoter was found to be required for GUS expression in both cultured cells and silks. A small fragment -36 to +219 relative to the transcription start site in *p2* promoter still showed significant level of GUS expression in silks. Whereas, a segment between -265 to -81 was found to contain potential cis-elements that suppress GUS expression in maize BMS cells. Analyses of transgenic maize containing p2-promoter::GUS constructs revealed that GUS expression was mainly restricted to the floral tissues. However, GUS expression was also observed in the immature kernel pericarp and cob. Possible mechanisms for the *p2* expression patterns observed in both transient and stable assays are discussed.

INTRODUCTION

Flavonoids are a group of plant secondary metabolites that are involved in a variety of processes such as response to pathogens and insect feeding, protection from ultraviolet light irradiation and pollen tube germination (Koes et al. 1994). The biochemical steps of synthesis of flavonoids and derived pigments have been studied intensively (Styles and Ceska 1977; Byrne et al. 1996). In maize there are two major flavonoid pigments that are both derived from flavanones: anthocyanins and phlobaphenes. Since the presence of pigments is not essential for plant viability, flavonoid pigmentation also provides a good indicator for analysis of gene regulation in plants. Genetic analyses have identified several genes required for the synthesis of anthocyanin and phlobaphene pigments, and these two groups of pigments share some early steps in their biosynthesis (Dooner et al. 1991). Among the numerous genes identified so far several regulatory genes have been the subjects of extensive studies and have provided great deal of information on plant gene regulation. Previous studies found that four regulatory genes (*c*1, *pl, b1, r1*) controlled the purple and red anthocynin pigmentation in various maize plant tissues. But *c*1 and *pl* encode Myb-homologous transcription factors that activate the transcription of structural genes encoding enzymes on the anthocyanin biosynthetic pathway. The *c*1 and *pl* genes function together with another class of genes *r*1 and *b*1, which encode proteins with basic helix-loop-helix (bHLH) domains (Paz-Ares et al. 1987; Ludwig and Wessler 1990). Many maize lines also accumulate red flavonoid pigment termed phlobaphene in certain floral organs including kernel pericarp and cob glumes. Phlobaphene biosynthesis in maize is regulated by the *p*1 gene, which also encodes a transcription factor with a Myb-like DNA binding domain and a transcription activation domain (Grotewold et al. 1991). The P1 protein can activate the transcription of several structural genes, including *c2*, *chi* and *a1,* for early steps of phlobaphene biosynthesis that are shared by the anthocyanin synthetic pathway. Though both *p*1 and *c1/pl* together with *r1/b1*

can activate transcription of *c2, chi* and *a1*, the *p1* gene can't activate transcription of genes on the branching pathway leading to anthocyanin pigments (Grotewold et al. 1994). Also unlike *c*1 and *pl*, P1 transcription factor itself can directly activate the transcription of *a1* and *c2* genes (Grotewold et al. 1994). In addition to regulating phlobaphene pigments synthesis, the *p1* gene is also involved in other flavonoid compounds synthesis. QTL analysis suggests that *p1* gene regulates the synthesis of a flavonoid compound, maysin, in maize silks. Maysin is a host-resistant factor that retards the growth of corn earworm larvae that feed on fresh silk and developing kernels (Byrne et al. 1996ab). Recently a *p*1 homologous gene, *p2*, was isolated in maize (Zhang et al. 2000). The *p2* gene was generated by a segmental gene duplication and is tightly linked to the *p1* gene in several *p1* haplotypes. The *p2* gene shares high homology in nucleotide sequence to *p1* in coding regions and encodes a protein with a Myb-like DNA binding domain and a putative acidic transcription activation domain that are highly homologous to those in the P1 transcription factor. However, the C-terminal region of P2 is different from P1, but is conserved in the other two transcription factors encoded by two *p1*-homologous genes, *p2-t* in teosinte and *y1* in sorghum. Despite the divergent C-termini, the *p2*- encoded protein can still activate the transcription of structural genes *c2*, *chi* and *a1* on the pathway of phlobaphene biosynthesis in transformed maize BMS cells (Zhang et al. 2003). Also like P1 factor, the *p2* encoded protein does not activate transcription of a subset of structural genes on the branching pathway to anthocyanin pigments. Though *p1* and *p2* have similar function on transcription activation, the expression pattern of *p1* and *p2* are different. No transcripts of *p2* can be detected in pericarp and cob and thus *p2* does not induce red pigmentation in those floral organs (Athma and Peterson 1991; Grotewold et al.1991). However transcriptional analysis indicated that *p2* is expressed in fresh silk and emerging anthers (Zhang et al. 2000). HPLC analysis of silk and BMS cells showed that *p1* and *p2* co-regulate synthesis of maysin and several other flavonoid compounds in maize plants and BMS cultured cells (Zhang et al. 2000, 2003).

Temporal and spatial regulations of transcription factor expression are important for plant development as well as accumulation of secondary metabolites (Meisel and Lam 1997). Changes of regulatory sequences in duplicate genes are often associated with their distinct expression patterns (Doebley and Lukens, 1998). Although much research has been directed to the regulatory function of those transcription factors on structural gene transcription, relatively little information is available about the regulation of the transcription factors themselves. Previously transposon mutagenesis and transient and stable expression assay in maize BMS cells and transgenic maize plants have identified several cis-acting regulatory regions in *p1* promoter, including a proximal and a distal enhancer element and a basal promoter region (Sidorenko et al. 1999, 2000). The basal promoter region of *p1* gene contains a 235 bp promoter segment and 326 bp 5'-UTR. This fragment can drive a significant level of GUS reporter gene expression in both transient assay in cultured cells and in certain floral tissues in transgenic plants, including pericarp, cob, husk, tassel glumes and silk. It was suggested that the cis-elements required for tissue-specific expression of the *p1* gene are retained in this small fragment. The maize *p2* gene carries out a similar function of transcriptional activation of flavonoid biosynthetic genes. However little is known about how its expression is regulated during plant development.

Several *p1* homologous genes and *p1*-locus alleles with distinct tissue-specific expression patterns have been cloned and characterized (Chopra et al. 1998; Zhang et al 2000). Comparing the sequences of those *p1*-locus alleles with sequences of *p2* gene in maize and *p2-t* from teosinte revealed that in addition to the coding region for the functional domains, these genes also share a ~ 400 bp region of high sequence similarity, including 5'–UTR (untranslated region) and a 90 bp fragment just upstream of transcription start site (Zhang et al. 2000). The 90 bp fragment is present in the basal promoter fragment identified in the *p1* gene and contains some structural features that are potential cis-regulating elements for the *p1* gene transcription regulation, including a 16 bp palindrome with four GTAC repeats and a putative

TATA box region (Figure 1B). Palindrome sequences have been shown to be important for gene regulation (Spiro and McMurray 1997). The three copies of ACGT motif in the 16 bp palindrome may be potential binding sites for plant bZIP DNA-binding regulatory proteins (Izawa et al. 1993).

In this study we sought to test whether any cis-acting regulatory elements present in the *p2* gene promoter and untranslated leader region are required for its tissue-specific expression, especially expression in silk. Since both *p1* and *p2* gene are expressed in silk and the only common sequences present in both genes are the 90 bp fragment and the 5'-UTR region, it implies that some cis-elements important for the silk expression of the *p1* and *p2* gene may reside in this small region. We used two approaches for the functional analysis of *p2* gene promoter. The first approach was transient expression assay of the *GUS* reporter gene in maize suspension cultured cells and fresh silks. The second approach was analysis of stable transgenic plants containing *p2* promoter regions fused with *GUS* reporter gene. Our preliminary results from transient assay suggested that cis-elements in the *p2* promoter region might have different responses to transcription factors in different testing systems. Some elements may function as repressor in the cultured maize BMS cells but not in silks. Interestingly, transgenic plants of *p2* promoter constructs showed ectopic GUS expression in tissues that *p2* gene not express normally. We discussed possible mechanisms for these unexpected observations.

MATERIALS AND METHODS

Plasmid construction

The plasmid Pb::GUS and two reference plasmids (PHP687 and PHP1528) which were used as internal controls for normalization in transient expression assays were obtained from Dr. Sidorenko and were described previously (Sidorenko et al 1999). For p2-265::GUS

construction, a 484 bp fragment extending from 265 bp 5'of *p2* transcription start site to 219 bp 3' of *p2* transcription start site was PCR-amplified and cloned into pGEM T-easy vector. Two enzyme sites, *Sal*I at 5'- end and *BamH*I at 3'-end were introduced by PCR. The *Sal*I-*BamH*I fragment was cut out and the 562 bp *Sal*I-*BamH*I fragment of *p1* gene promoter in Pb::GUS was replaced with this 484 bp fragment. The 484 bp *p2* gene promoter fragment in p2-265::GUS was replaced with a series of DNA fragments from maize *p2* gene promoter to generate derivative plasmids for both transient and stable expression analyses (Figure 2B). Except p2::GUS construct, the promoter fragments in all constructs that contain the *GUS* reporter gene were PCR-amplified and sequenced to eliminate any PCR introduced mutation. The 4.4 kb *p2* promoter region in p2::GUS plasmid was derived from a *Kpn*I–*Sac*I fragment from plasmid containing ~23 kb region of the maize *p2* gene (Zhang et al. 2000) and cloned into p2-265::GUS plasmid digested with *Kpn*I plus *Sac*I. The junctions of ligated fragments in p2::GUS were sequenced. Negative control p0::GUS was obtained from plasmid LP1.2-1 (Sidorenko, unpublished data) by cutting out the *p1* promoter fragment. Thus p0::GUS contains only AdhI intron, GUS gene and PinII terminator on a pBluescript backbone.

Construct p1::P2 contains a ~6 kb *p1* gene promoter region followed by maize AdhI intron I, fused with *p2* cDNA and NOS terminator. Construction of *p2* cDNA fragment in p1::P2 was as follows: first an *Nco*I-*Pst*I segment was PCR-amplified from pF2cDNA plasmid containing *p1* cDNA (Grotewold et al. 1994). The sequence encoding an Alanine on the 4th amino acid in *p1* was mutated by PCR to encode a Threonine as in *p2*. Secondly a *Pst* I-*Sal*I fragment in *p2* cDNA was obtained from plasmid containing ~23 kb region of *p2* gene (Zhang et al. 2000). Then the 3'-UTR part of *p2* cDNA was PCR-amplified from this plasmid and two enzyme sites (*Sal*I-*Xho*I) were introduced. Finally the *Xho*I-*Eco*RI fragment containing NOS terminator sequence was obtained from plasmid pF2cDNA. The resulting *Nco*I-*Eco*RI *p2* cDNA fragment was then ligated with a *Bgl*II-*Nco*I fragment from plasimid LP1.0-4. The *Bgl*II-*Nco*I fragment contains a 560 bp *p1* promoter region and 5'-UTR fused

with AdhI intron. The resulting *Bgl*II-*EcoR*I fragment was ligated with an *EcoR*I–*Bgl*II fragment from plasmid pRRARR (Cocciolone et al. 2001) which contains the ~ 6 kb *p1* promoter segment. Finally, the *EcoR*I-*EcoR*I fragment was inserted into an *EcoR*I site in pPTF101.1 (Plant Transformation Facility, Iowa State University) which lies between T-DNA right and left border.

Construct p2::P1 contains a 4.4 kb *p2* promoter fused with *p1*cDNA and NOS terminator. The *p2* promoter fragment was obtained from p2::GUS plasmid described above. The BamHI-BglII fragment containing *p1* cDNA and NOS terminator was obtained from pRRARR described previously (Cocciolone et al. 2001). The EcoRI-BglII fragment was then inserted into pTF101.1 digested with EcoRI-BamHI. The junctions in p1::P2 and p2::P1 plasmids were all sequenced.

Microprojectile bombardment and Transient assays

For each bombardment about 0.2 pmol DNA of test constructs and 0.05 pmol internal control plasmid were mixed with gold particle (0.6μm, Bio-Rad) and precipitated as described previously (Sidorenko et al 1999). Bombardments were performed with Biolistic particle gun (DuPont). Each bombardment comprised at least three duplicates and for each construct the bombardments were performed at least three times.

Suspension maize cultured BMS cells were obtained from Pioneer Hi-Bred Inc., Johnston, Iowa. Cells were treated with 0.25 M mannitol in MS medium for 12 hours before bombardment. After bombardment cells were kept on MS medium in darkness at 28 °C for 2 days before collection for analysis. The GUS light assay and Luciferase activity measure were performed according to manufacturer's instruction (Applied Biosystems). The value of GUS relative light unit (RLU) was divided by Luciferase RLU value and the ratio of GUS/Luc for construct p2::GUS was set to equal 100. The ratios of GUS/Luc value of p2-promoter::GUS test plasmids to p2::GUS value were calculated.

For maize silk transient assay, plants from inbred line A619 which has genotype *p1-www* (both *p1* and *p2* are deleted) were grown until silk emergence. About 6 cm ear tips from ears 1 day after silk emergence were cut and sprayed with 70% ethanol. The husks were peeled off after the ear tips were further divided into 3 parts (bottom, middle and top). For each part the silks were about 2 cm long. The silks from each part were mixed in liquid MS medium and arranged in a 2cm x 2cm square (about 10 silks) on filter paper soaked with MS medium. Bombardments were duplicated 3 times using silks from each part once. Bombardment of silks for each construct was performed at least 3 times. After bombardment the silks were kept on MS medium in darkness for 2 days. The red anthocyanin spots were counted under low magnification microscope. Then the silks were stained with x-Gluc for 2 days at 37 °C and number of blue spots were counted. The ratio of blue spots to red spots was calculated.

Transgenic plants analysis

Transformation of maize plants with test plasmids was performed by the Plant Transformation Facility at Iowa State University. The plasmids containing *GUS* reporter gene were co-bombarded by particle gun with a plasmid conferring resistance to the herbicide bialaphos. Plasmids p1::P2 and p2::P1 were transformed using Agrobaterium mediated transformation. The *GUS* reporter gene constructs generated transgenic plants (T0) were outcrossed to inbred line B73 and the transgenic plants in T1 segregation population were identified by resistance to herbicide treatment. Mature leaf, leaf auricle, tassel, husk, silk and kernels at stage 16-18 days after pollination were collected from T0 plants and stained with x-gluc at 37 °C for 2 days. Then the samples were kept in 80% ethanol for several days before checking for GUS expression. The results of GUS expression levels of each sample were recoded as none, weak, moderate and strong. The data was analyzed by using the SAS macro 'glimmix' which handles logistic regression with random variables. The probability of GUS

expression was calculated from the least squired means lines in SAS (Heilmann, personal communication).

RESULTS

Transient expression assay of *p2* promoter in maize suspension BMS cells

Previous studies have demonstrated that transient assay techniques are useful in functional analysis of promoter segments of plant regulatory genes and have been applied in identification of cis-regulatory element responsible for tissue-specific expression of maize genes (Radicella et al. 1992; Sidorenko et al. 1999). The *p2* gene does not confer visible pigmentation in maize, and even the *p1* gene whch induces red pigmentation in pericarp and cob does not induce visible pigmentation in transient assays in pericarp (Sidorenko et al. 1999). Therefore, we chose *GUS* as reporter gene for the analysis of the promoter activity of *p2*. As shown in Figure 2A, a series of constructs with segments from the *p2* promoter and 5'-unstranslated region were fused with the maize AdhI intron and GUS coding region. A negative plasmid p0::GUS which contains only AdhI intron and GUS gene was also included in all transient assays as negative control. Inclusion of the AdhI intron is thought to increase reporter gene expression (Callis et al. 1987). Constructs lacking the AdhI intron failed to produce GUS expression levels above background (data not shown). The same result was obtained in previous transient assay of *p1*gene promoter, indicating that both *p1* and *p2* promoters have relatively low activity (Sidorenko et al. 1999). The results of GUS expression of these test plasmids were normalized to Luciferase activity and the relative ratio of GUS/Luciferase activity of these p2-promoter::GUS plasmids to that of plasmid p2::GUS is summarized in Figure 2B. The p2::GUS contains a 4.4 kb fragment from -4409 to +219 (transcription start site=+1, Figure 2A) of *p2*. The GUS expression level of p2::GUS is significantly higher than that of the promoter-lacking plasmid p0::GUS. Deletion of *p2* promoter fragment from 4.4 kb in p2::GUS to 265 bp in p2-265::GUS did not change the GUS

expression level very much. Interestingly, further deletion of the *p2* promoter fragment from 265 bp 5'-of transcription start site to only 81 bp in p2-81::GUS, resulted a 2-fold increase in GUS expression leve. This result may imply the existence of a repressor-like element in the region between –265 and –81 relative to transcription start site. When the promoter fragment was deleted further to 66 bp 5'-of transcription start site in p2-66::GUS, the expression level of GUS dropped to about the same level of p2-265::GUS. When the + 10 to +219 (transcription start site =+1) of 5'-UTR region of *p2* gene in p2-265::GUS was replaced with a segment from +1 to +23 of another maize gene, *c1*, which controls anthocyanin pigmentation in maize plant, the GUS expression level of p2-265-c1::GUS dropped to approximately 40% of the GUS expression level of p2-265::GUS, though it was still significantly higher than p0::GUS. This result indicates that some sequences in 5'-UTR of the *p2* gene might be important for its transcriptional activity in suspension cultured BMS cells Plasmid p2-265-m7::GUS in which the putative TATA box TATATTAT in the *p2* gene was mutated to TAGATCTT by PCR, was also included in the transient assay. This mutation significantly decreased the GUS expression level comparing to that of p2-265::GUS and is similar to the expression level of p0::GUS. This observation confirms the improtance of the TATA box in the *p2* promoter.

Transient expression assay of *p2* promoter in maize silks

To test whether the *p2* promoter contains cis-elements important for expression in silk, we used fresh maize silk for transient assays. The p2-promoter::GUS constructs were co-bombarded with internal control plasmid PHP687 which contains double cauliflower mosaic virus 35S promoter fused with maize *c1* and *r1* gene coding region that induce red anthocyanin pigments in maize cells (Sidorenko et al. 1999). Number of blue spots which were visible in both silk lobes and stigmatic hairs after staining with x-Gluc was used to represent the GUS expression level of test plasmids in silk and the level was normalized to the

numbers of red foci produced by PHP687, which were also visible in both silk stigmatic hairs and lobes. As shown in Figure 2B, plasmid p2-265::GUS produced significant high GUS expression level relative to p0::GUS but is only approximately 70% of the expression level of p2::GUS. In contrast to the observation in suspension BMS cells that deletion from –265 to –81 of promoter fragment in p2-81::GUS increased GUS expression level, the p2-81::GUS expression level in silk was similar to that of p2-265::GUS. Also the replacement of 5'-UTR of *p2* with 5'-UTR from *c1* did not affect much the GUS expression level, as p2-265-c1::GUS or p2-81-c1::GUS showed a similar level of GUS expression as p2-265::GUS or p2-81::GUS, respectively. However, mutation of TATA box region of the *p2* gene again reduced GUS expression significantly as the same observation in BMS cell transient assay. One unexpected observation is that further deletion of *p2* promoter fragment did not affect the promoter activity of the fragment in driving *GUS* reporter gene expression. Plasmid p2-36::GUS which contains only 36 bp 5'- of transcription start site and 5'-UTR still showed significant GUS expression level compared to the promoter-lacking p0::GUS or the TATA box mutant p2-265-m7::GUS. It is suggested that though the 36 bp sequence is a very small fragment it still has promoter activity. To our knowledge, no such small promoter fragment has been reported previously.

Analysis of *p2* promoter activity in transgenic plants

Transient assays are useful for rapid identification of promoter regulatory sequences. However, the normal *in planta* tissue specific expression of *p1* gene can not be reproduced in transient assays of p1-promoter::GUS plasmids (Sidorenko et al. 1999). This may be due to the inability of transforming DNA to assume chromatin structure that is important for proper temporal and spatial expression of genes. Based on the transient assays of GUS expression in several p2-promoter::GUS plasmids, we selected four constructs (p2::GUS, p2-81::GUS, p2-81-c1::GUS, p2-81-m7::GUS) for further analysis in stable transformed maize plants. The

p2-promoter::GUS constructs were introduced into Type II callus induced from maize Hi-II (genotype p1-ww) embryos. The tissues from the generated transgenic plants were stained and GUS expression levels were recorded as none, weak, moderate and strong as previously described (Sidorenko et al. 2000). The results of GUS expression from several tissues, including mature leaves, auricle, tassel, husk, silk, pericarp and cob, were summarized in Table 1. The GUS activity in the transgenic plants varied in both intensity and distribution in the tested plants. Representative staining patterns of plant tissues are shown in Figure 3. One striking feature of the expression of *GUS* reporter gene in these stable transformants is that several independent transgenic plant lines from each construct showed significant GUS expression in pericarp and cob, tissues in which no endogenous *p2* transcripts could be detected by RT-PCR (Zhang et al. 2000). This might result from an increased sensitivity of GUS detection, a higher expression level of GUS gene due to the effect of AdhI intron, or the complexity of transgene organization in the plant genome. For example the fragment of p2-promoter::GUS may be inserted downstream of a stronger promoter or enhancer that drives ectopic GUS expression in pericarp and cob. Southern blot analysis of genomic DNA from several plants with GUS expression in pericarp and cob in T1 generation indicated the presence of complex transgene structures, including multiple copies of p2-promoter::GUS fragments (data not shown). Comparing the GUS expression patterns of *p2* promoter to that of *p1* promoter-generated transgenic plants revealed several similarities. First both kinds of constructs produced plants with GUS expression in leaf auricle in which no *p1* regulated pigmentation was reported previously. Recently, Cocciolone and colleagues reported that a *p1* transgene in which a full length *p1* promoter fused with *p1* cDNA could induce pigmentation in transgenic plant auricle. RT-PCR and RNA blot analysis of non-transgenic plants also detected low level of *p1* transcripts in several vegetative tissues that were not reported previously for *p1* expression (Cocciolone et al. 2001). The observation of auricle expression in p2-promoter::GUS transgenic plants indicates that similar to *p1*, *p2* may also express in

some vegetative tissues at low levels. Second, three classes of tissue-specific expression pattern of GUS in pericarp and cob were also observed in p2-promoter::GUS as in p1-promoter transgenic plants (Sidorenko et al. 2000; Cocciolone et al. 2001). The three classes are designed as RR, WR and RW based on the tissue specific pigmentation of *p1*-locus alleles that displayed distinct pigmentation in pericarp and cob: RR represents red pericarp and red cob as in *P1-rr* allele, WR represents white pericarp and red cob as observed in *P1-wr* allele and RW represents red pericarp and white cob as in *P1-rw* allele (Athma and Peterson 1991; Chopra et al. 1996; Cocciolone et al.2001). Pigmentations in the two tissues are correlated with the level of *p1* transcripts in them (Grotewold et al. 1991; Chopra et al. 1996). GUS expression of immature kernels from p2-promoter::GUS generated transgenic plants also showed distinct expression pattern in pericarp and cob among and within independent transformation events. As shown in Figure 3, some kernels showed GUS expression in both pericarp and cob (RR), while some only have GUS expression in cob but not pericarp (WR). Also several lines produced kernels with GUS expressed in pericarp but not in cob (RW, data not shown). The mechanism of tissue specific expression of *p1* locus allele is not known yet but it was proposed that it might be regulated epigenetically (Cocciolone et al. 2000, 2001). Analysis of GUS expression in mature leaves from initial transgenic plants (T0 generation) didn't detect GUS expression in most plants in which GUS expression was detectable in other floral organs. Also, no GUS activity was detected in seedling leaves from plants in T1 generations, which indicated that the absence of expression of *p2* gene in leaf in endogenous alleles was retained in those transgenic plants. Statistical analysis of GUS reporter gene expression in the transgenic plants generated from the four p2-promoter constructs indicated that the expression level of GUS in silk is relatively weak compared to the GUS expression levels in other tissues like auricle, husk and even pericarp and cob (Table 1). Plasmid p2::GUS with full length *p2* promoter fragment has the highest probability to obtain GUS expression in husk, silk, pericarp and cob. Though p2-81::GUS has relatively low expression

level in all the tissues tested, it still has higher frequency to obtain plants with GUS activity in all tested tissues than those of p2-81-c1::GUS and p2-81-m7::GUS. However statistical tests of the GUS activities in the tested tissues did not detect significant difference in the GUS gene expression level among the four tested constructs. The relative small number of observations, especially small samples of pericarp and cob, may affect the statistical tests for significant differences.

DISCUSSION

The *p2* gene from maize and *p2-t* from teosinte share high homology with the maize *p1* gene in the coding region and a ~90 bp region just upstream of transcription start site (Zhang et al. 2000; Figure 1). Further sequence analysis of *p2* and *p2-t* revealed that the 5'-flanking regions of these two genes also have other conserved sequences. As shown in Figure 1, several segments with length ranging from 30 bp to 1 kb in the 5'- flanking region of *p2* gene have high sequence similarity (93% to 100%) to several segments in *p2-t.* Moreover, these segments are arranged in the same order as in *p2-t.* The expression pattern of *p2* is also similar to that of *p2-t.* Both genes are expressed in maize silk and anther, but not in pericarp and cob. However *p2-t* also expresses and induces brown pigmentation in tassel glumes, just as *p1*, while neither transcripts of *p2* nor brown pigmentation were detected in tassel glumes from plants in which *p1* is deleted but *p2* is intact. It is proposed that the *p1* and *p2* gene complex was generated by a tandem duplication of an ancestral *p* gene (p^{pre}) which was similar to *p2-t* (Zhang et al. 2000). Redundant function of duplicated genes is thought to be evolutionarily unstable; eventually, one of the duplicated genes may either lose its function or gain a novel function. Acquisition of new regulatory sequences is often responsible for the new expression pattern of duplicated genes (Doebley and Lukens, 1998). Previous analysis of *p1* gene promoter suggested that the higher expression level and broadened transcription pattern of maize *p1* gene expression in floral organs including pericarp and cob might be the result of

acquiring new regulatory sequences at 5'-flanking regions of *p1* following the segmental gene duplication that generated the *p1-p2* complex (Sidorenko et al. 1999; Zhang et al. 2000). However, comparison of the expression pattern of *p1* and its homologous genes such as *p2* and *p2-t* suggested that these genes still retained some redundant functions. One major function of the *p2* gene is the co-regulation with *p1* of flavonoid compound synthesis in maize silks. Sequence analysis of *p1*, *p2* and *p2-t* promoter region and 5'- untranslated region revealed highly conserved sequences with some potential regulatory factor binding sites and small direct and inverted repeats (Figure 1B). Previous results of *p1* promoter analyses implied the existence of cis-regulatory elements important for *p1* tissue-specific expression in a relatively small region, including the conserved promoter regions between *p1* and *p2* (Sidorenko et al. 1999, 2000). Thus it is possible that there are some cis-acting elements important for the *p2* expression in silk. We carried out both transient and stable transformation analysis of the *p2* promoter. However preliminary results did not reveal any cis-acting elements that are required for tissue-specific expression of the *p2* gene, except for the TATA box that is crucial for GUS reporter gene expression in both suspension cultured BMS cells and fresh maize silks.

Both *p1* and *p2* are regulatory genes with low levels of endogenous transcription. The low expression level of *p1* and *p2* are reflected in the observation that GUS expression is not detectable in transient assays when the AdhI intron is not included in the test constructs (Sidorenko et al. 1999 and this study). Adding AdhI intron greatly increased reporter gene expression level, possibly due to the effects on precursor RNA splicing (Callis et al. 1987). Inclusion of the AdhI intron may make small promoter activity changes indistinguishable. However, deletion of sequences crucial for *p2* promoter activity should be detectable. Transient assays of p2-promoter construct with 5'-UTR and only 36 bp fragment upstream of transcription start site still gives significantly high GUS expression level in fresh silk, suggesting that such small fragment still retained the basal promoter activity. No such small

promoter fragment was reported previously in plant genes. One possibility is that the relatively long 5'-UTR of both *p1* and *p2* may contain regulation elements. For example, in some constructs like p2-36::GUS, though the promoter region is quite small, some elements retained within the 5'- untranslated region may complement the functions of elements deleted from the upstream promoter region. Replacement of the 5'-UTR in p2-265-c1::GUS and p2-81-c1::GUS decreased the GUS expression level in BMS cells, but didn't affect the reporter gene expression in silks. This may be the result of the difference between testing systems (See below) or may be due to the presence of some elements still retained in the 81 bp small fragment that can still drive basal promoter activity. Unlike the effect of the TATA box on promoter activity, deletion of some of the elements may decrease but not totally eliminate the promoter activity. Further analyses of constructs with different *p2* promoter segments and 5'-UTR are needed before any clear conclusion can be made.

Comparing the promoter activity of the tested constructs in suspension BMS cells and silks revealed that the results are not consistent between these two different testing systems. Deletion of *p2* promoter fragment from –4409 in p2::GUS to –81 in p2-81::GUS resulted in increased GUS expression in suspension BMS cells, but little change in GUS expression in silk. In contrast, replacing the *p2* 5'-UTR with 5'-UTR from maize *c1* gene in p2-265-c1::GUS decreased the GUS expression level to about half of that of p2-265::GUS in BMS cells. A corresponding decrease was not observed in silk transient assay. In fact the expression levels of GUS reporter gene of p2-265-c1::GUS and p2-81-c1::GUS are even higher than that of p2-265::GUS and p2-81::GUS, respectively. This inconsistency of promoter activity of the constructs in different testing systems may reflect the presence of different regulatory factors in suspension maize culture cells and silks. Some factors in cultured maize BMS cells may recognize some sequences upstream of –81 in *p2* promoter and function as repressors of expression, but in silk such factors may not be present. Alternatively the inconsistency may due to the different sensitivity of techniques used in measuring GUS

and internal control gene expression levels in the two systems. The GUS and anthocyanin control level in transient assays in silk are represented by counting blue and red spots in bombarded silks. The relative intensity of the pigmentation can not been taken into account when the relative GUS expression level are calculated. Thus some may have GUS levels much higher than that of other cells, but the high level of expression was not considered in the analysis since only numbers of GUS foci were scored. This could generate some measure errors and result in less sensitivity of measuring promoter activity than the assay used in suspension BMS cells.

Previously the *p2* gene expression in husk is not tested. RT-PCR analysis of RNA from maize *p* gene alleles *P1-rr4B2* which contains both *p1* and *p2* gene showed that the *p2* gene is also expressed in husk (data not shown). Transgenic analysis of *p2*-promoter constructs also suggested that the GUS expression level in husk is relatively higher than that in silk. The observation of GUS expression in pericarp and cob of transgenic plants containing p2-promoter constructs is somewhat surprising. However recent RT-PCR analysis of RNA from pericarp of a maize *p1* allele *p1-ww1112* in which *p1* gene coding region is deleted but *p2* is intact detected a very low level of *p2* transcripts when the PCR products were blotted onto nylon membrane and hybridized with *p1* cDNA probe (data not shown). It suggested that the *p2* gene might have a very low level of transcription in pericarp. Such low level expression is not detectable in Northern blot analysis and even hard to detect in RT-PCR. The maize *p1* gene may also have a relative low expression level in all the tissues that have been tested previously, including both floral and vegetative tissues. However due to the presence of enhancer elements in the flanking region of the *p1* gene, probably in a tissue-specific manner, *p1* expression level is greatly intensified (Sidorenko et al. 1999, 2000). The absence of such enhancer elements in *p2* gene flanking regions retains *p2* gene expression weak. The visible expression of GUS gene in transgenic kernels may result from the more sensitive GUS staining assay or from other factors that are not related with the tissue-specific activity of *p2*

promoter.

Recently a new p-homologous gene derived from the promoter and the 5'-end coding regions until the middle of intron 2 from *p2* gene and the 3'-end of intron 2 and exon 3 plus 3'-flanking region of *p1* gene was identified (Zhang et al. in preparation). This gene induced orange pigmentation in both pericarp and cob, indicating the expression of the new gene in these two tissues. This observation implies that the *p2* gene promoter indeed contains cis-elements that are required for expression in pericarp and cob. Previously it has been shown that *p2* encoded protein can function normally as transcription activator in structural gene transcription activation in transformed cultured maize BMS cells (Zhang et al. 2003). Whether the encoded protein can function normally in transgenic plants to induce any pigmentation in floral tissues is unknown. Also it is not clear whether the *p2* promoter is strong enough to induce visible phlobaphene pigmentation in maize floral tissues when fused with *p1* cDNA. Two constructs with *p1* promoter driving *p2* cDNA (p1::P2) and *p2* promoter driving *p1* cDNA (p2::P1) have been transformed into maize (Material and Methods). If red pigments are observed in transgenic plants generated from both constructs it will provide more convincing support that the non-detectable *p2* induced pigmentation in pericarp and cob is neither due to the function ability of *p2* encoded protein nor due to the lacking of cis-acting elements required for tissue-specific expression of *p2* gene in floral tissues, but may due to other mechanisms that need to be clarified.

ACKNOWLEDGEMENTS

We would like to thank Dr. Kan Wang, Lise Marcell, Bronwyn Frame and Tina Fonger at the Plant Transformation Facility, Iowa State University for their help in generating transgenic plants. I would also like to thank Cory Heilmann in the Department of Statistics, Iowa State University, for his help in the statistical analysis of the transgenic research data.

REFERENCES

Athma, P., and Peterson, T. (1991) *Ac* induces homologous recombination at the maize *P* locus. Genetics 128: 163-173.

Athma, P., Grotewold, E., and Peterson, T. (1992) Insertional mutagenesis of the maize *P* gene by intragenic transposition of *Ac*. Genetics 131:199-209.

Byrne, P. F., Darrah, L. L., Snook, M. E., Wiseman, B. R., Widstrom, N. W., Moellenbeck, D. J., and Barray, B. D. (1996a) Maize silk-browning, maysin content, and antibiosis to the corn earworm, *Helicoverpa Zea* (Boddie). Maydica 41: 13-18.

Byrne, P., McMullen, M., Snook, M., Musket, T., Theuri, J., Widstrom, N., Wiseman, B., and Coe, E. (1996b) Quantitative trait loci and metabolic pathways:genetic control of the concentration of maysin, a corn earworm resistance factor, in maize silks. Proc. Natl. Acad. Sci. USA 93: 8820-25.

Callis, J., Fromm, M., and Walbot, V. (1987) Introns increase gene expression in cultured maize cells. Genes & Development 1: 1183-1200.

Chopra, S., Athma, P., and Peterson, T. (1996) Alleles of the maize *P* gene with distinct tissue specificities encode myb-homologous proteins with C-terminal replacements. Plant Cell 8: 1149-1158.

Chopra, S., Athma, P., and Peterson, T. (1998) A maize Myb-homologue is encoded by a stable multicopy gene complex. Mol. Gen. Genet. 260: 372-378.

Cocciolone, S.M., Sidorenko, L.V., Chopra, S., Dixon, P.M., and T. Peterson. (2000) Hierarchical patterns of transgene expression indicate involvement of developmental mechanisms in the regulation of the maize P1-rr promoter. Genetics 156: 839 – 846.

Cocciolone, S.M., Chopra, S., Flint-Garcia, S.A., McMullen, M.D., and Peterson, T. (2001) Tissue-specific patterns of a maize Myb transcription factor are epigenetically regulated. Plant J. 27(5): 467-478.

Cone, K.C., Cocciolone, S.M., Burr, F.A. and Burr B. (1993) Maize anthocyanin regulatory gene *pl* is a duplicate of *c1* that functions in the plant. Plant Cell. 5(12):1795-1805.

Doebley, J., and Lukens, L. (1998) Transcriptional regulators and the evolution of plant form. Plant Cell. 10(7):1075-1082.

Dooner, H.K., Robbins, T.P., and Jorgensen, R.A. (1991) Genetic and developmental control of anthocyanin biosynthesis. Annu. Rev. Genet. 25:173-199.

Grotewold, E., Athma, P., and Peterson, T. (1991) Alternatively spliced products of the maize *P* gene encode proteins with homology to the DNA-binding domain of myb-like transcription factors. Proc. Natl. Acad. Sci. USA 88: 4587-4591.

Grotewold, E., Drummond, B. J., Bowen, B., and Peterson, T. (1994) The myb homologous *P* gene controls phlobaphene pigmentation in maize floral organs by directly activating a flavonoid biosynthetic gene subset. Cell 76: 543-553.

Izawa, T., Foster, R., and Chua, N.H. (1993) Plant bZIP protein DNA binding specificity. J Mol Biol. 230(4):1131-1144.

Koes, R.E., Quattrocchio, F., and Mol, J.N. (1993) The flavonoid biosynthetic pathway in plants: function and evolution. BioEssays 16(2): 123-132.

Ludwig, S.R., and Wessler, S.R. (1990) Maize *R* gene family: tissue-specific helix-loop-helix proteins. Cell. 62(5):849-851.

Paz-Ares, J., Ghosal, D., Wienand, U., Peterson, P.A. and Saedler, H. (1987) The regulatory *c1* locus of *Zea mays* encodes a protein with homology to myb proto-oncogene products and with structural similarities to transcriptional activators. EMBO J. 6(12):3553-3558

Radicella, J.P., Brown,D., Tolar, L.A., Chandler V.L.(1992) Allelic diversity of the maize *B* regulatory gene: different leader and promoter sequences of two *B* alleles determine distinct tissue specificities of anthocyanin production. Genes& Development 6: 2152-2164.

Sidorenko, L.V., Li, X., Tagliani, L., Bowen, B., Peterson, T. (1999) Characterization of the regulatory elements of the maize *P-rr* gene by transient expression assays. Plant Mole. Bio.

39: 11-19.

Sidorenko, L.V., Li, X., Cocciolone, S.M., Tagliani, L., Chopra, S., Bowen, B., Daniels, M., and T. Peterson. (2000) Complex structure of a maize Myb gene promoter: functional analysis in transgenic plants. The Plant Journal 22: 471-482.

Spiro, C., and McMurray, C.T. (1997) Switching of DNA secondary structure in proenkephalin transcriptional regulation. J. Biol. Chem. 272: 33145-33152.

Styles, E.D., and Ceska, O. (1977) The genetic control of flavonoid synthesis in maize. Can. J. Genet. Cytol. 19: 289-302.

Zhang, P., Chopra, S., T. Peterson. (2000) A segmental gene duplication generated differentially expressed Myb-homologous genes in maize. The Plant Cell 12: 2311-2322.

Zhang, P., Wang, Y., Zhang, J., Maddock, S., Snook, M., and Peterson, T. (2003) A maize QTL for silk maysin levels contains duplicated Myb-homologous genes which jointly regulate flavone biosynthesis. Plant Mol Biol. 52(1):1-15.

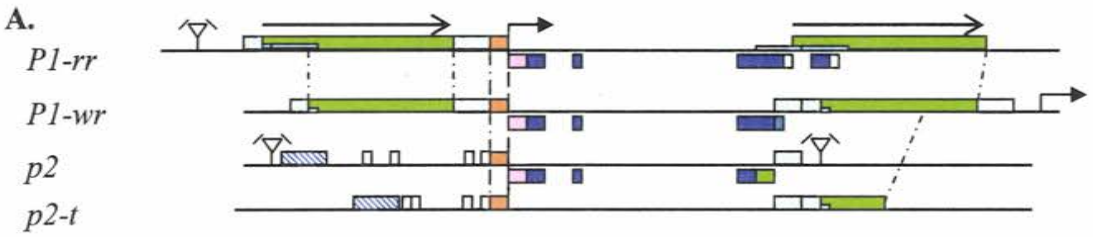
FIGURE LEGENDS

Figure 1. A. Structural comparison of *p1* homologous genes. The *p1* gene structure in two *p1*-alleles *P1-rr* and *P1-wr* are shown on the top. *P1-wr* has 6 copies of *P1-rr* gene arranged as tandem repeats. Patterned boxes represent homologous sequences at flanking regions at maize *p1*, maize *p2* and teosinte *p2-t* genes. Boxes below each line represent exons. Patterned triangle at the ends of *p1-rr* and *p2* represent retroelements insertions. Large arrows indicate two 5.2 kb direct repeats in the flanking regions of *p1* in *P1-rr*. Bent arrows represent transcription start site. **B.** Nucleotide sequence alignment of ~150 bp region upstream of transcription start site in *p1* (*P1-rr* and *P1-wr*), *p2* and *p2-t* genes. The TATA box sequences are highlighted. Arrows represent repeat sequences. Bent arrow represents transcription start site.

Figure 2. A. Maps of maize *p2* and teosinte *p2-t* promoter and 5'-UTR. Bent arrows indicate

transcription start site. Patterned boxes represent homologous sequences. **B.** Schematic maps of p2-promoter::GUS constructs and the relative GUS expression level of each construct in transient expression assays in suspension maize BMS cells and in fresh maize silks. The *p2* promoter and 5'-UTR are indicated by boxes whose position and length are corresponding to the map above. The dark colored boxes in p2-81-c1::GUS indicate 5'-UTR from maize *c1* gene. The AdhI introns are represented by angled lines. GUS reporter gene and PinII terminitor are represented by black and grey boxes, respectively. GUS levels are normalized to either Luciferase activity or number of anthocyanin pigment foci. The promoter activity of the 4.4 kb *p2* fragment in p2::GUS is set to 100 and the relative activities of promoter fragments in other test constructs are listed.

Figure 3. GUS activity of p2-promoter::GUS in tissues from transgenic plants. Representative samples of stained tissues are shown. A. kernels with GUS expression in both pericarp and cob (RR phenotype); B. kernels with GUS expression in cob but not in pericarp (WR phenotype); C. mature leaf and leaf auricle; D. husk; E. silk; F. tassel; G. anther.

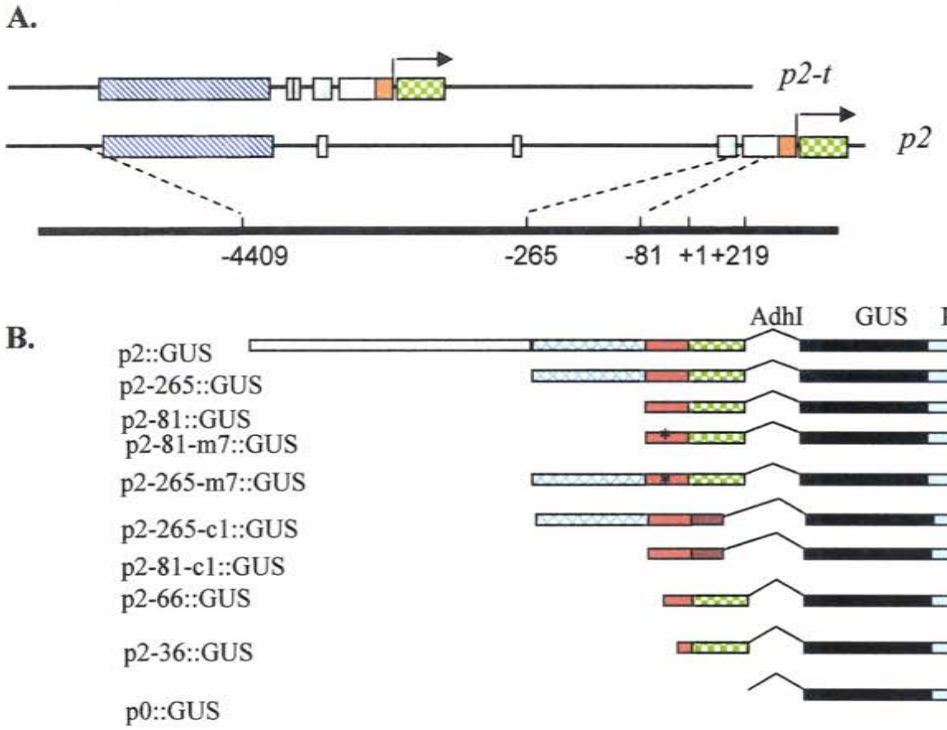


B.

```
P1-rr  CGTGATTTTACTTCTGTTTAAGGGCTAAAATACA.................AACGTGCACTCTGCACTC
       ||||||||||||||||||||||||||||||||||                 ||||||||||||||||||
P-wr   CGTGATTTTACTTCTGTTTAAGGGCTAAAATACAAAAAATTCGGTCGGTCCGTAACGTGCACTCTGCACTC
                                                 ||||||||||||||||   |||||||
P2     TGCTCGCTCGGTTCCCTTCCCCCCAGTCCCACGTAACGGACAGGTCGGTCCGTAACG.......TGCACTC
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||   |||||||
P2-t   TGCTCGCTCGGTTCCCTTCCCCCCAGTCCCACGTAACGGACAGGTCGGTCCGTAACG.......TGCACTC

P1-rr  TACTAAGCGCTAGTGTACGTACGTACGTACTCCGTCCGCTGCTATATTATGGCCGGCCGTGGCGTGCCCTCTC
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
P-wr   TACTAAGCGCTAGTGTACGTACGTACGTACTCCGTCCGCTGCTATATTATGGCCGGCCGTGGCGTGCCCTCTC
       |||||| |||||||    | ||||||||||||||||||||||||||||||||||||||||||||||||||
P2     TACTAAACGCTAGT....GCACGTACGTACTCCGTCCGCTGCTATATTATGGCCGGCCGTGGCGTGCCCTCTC
       |||||| |||||||    ||||||||||||||||||||||||||||||||||||||||||||||||||||
P2-t   TACTAAGCGCTAGT....GTACGTACGTACTCCGTCCGCTGCTATATTATGGCCGGCCGTGGCGTGCCCTCTC
```

Figure 1.



	Relative GUS Expression level	
	BMS cells	Silk
p2::GUS	100	100
p2-265::GUS	91	79
p2-81::GUS	216	80
p2-81-m7::GUS	/	17
p2-265-m7::GUS	4	7
p2-265-c1::GUS	46	113
p2-81-c1::GUS	/	114
p2-66::GUS	107	67
p2-36::GUS	/	74
p0::GUS	3	24

Figure 2.

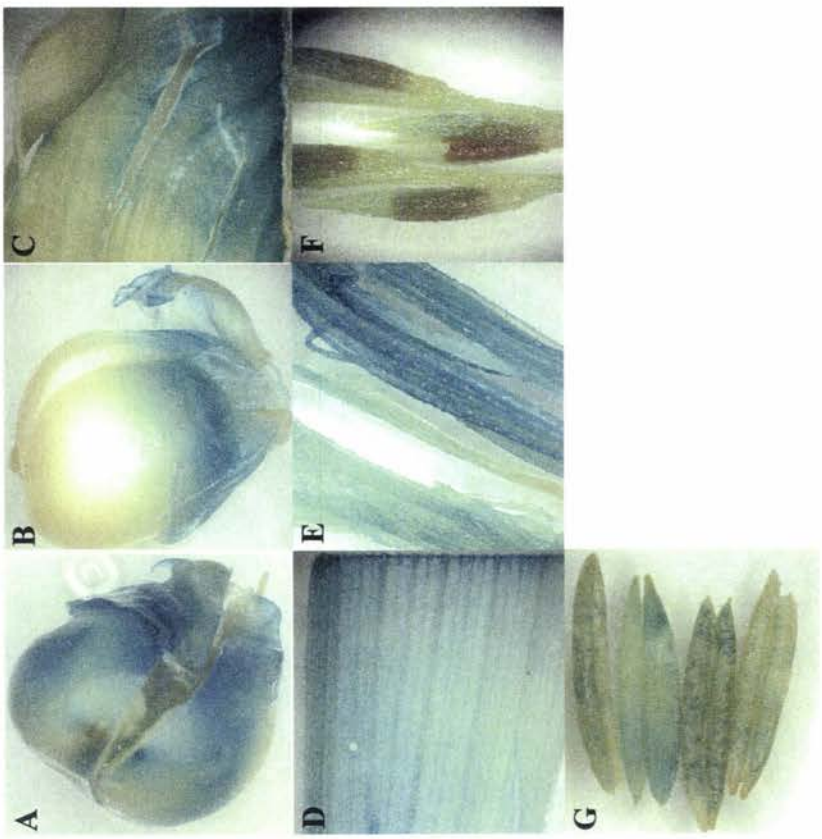


Figure 3.

Table 1. GUS activity of p2-promoter::GUS transgenic plant tissues

Plasmid	GUS activity	Tassle	MatureLeaf	Leaf auri	Husk	Silk	Pericarp
	Total	12(44)	12(44)	12(44)	10(30)	10(30)	5(15)
	None	11(32)	12(40)	9(27)	7(12)	10(23)	1(1)
p2::GUS	Weak	4(6)	2(2)	4(10)	5(13)	3(6)	4(7)
	Moderate&stron	3(6)	1(2)	3(7)	5(9)	2(5)	3(7)
	P(GUS)	0.17	/	0.32	0.69	0.22	0.73
	Total	11(44)	11(44)	11(44)	9(32)	9(32)	10(23)
	Non	10(34)	9(33)	7(25)	8(20)	9(23)	7(14)
p2-81::GUS	Weak	4(8)	3(11)	4(14)	6(10)	3(6)	3(3)
	Moderate&stron	1(2)	0(0)	2(5)	3(5)	3(6)	3(6)
	P(GUS)	0.19	/	0.36	0.34	0.21	0.28
	Total	10(35)	10(35)	10(35)	10(32)	10(32)	7(18)
	Non	9(29)	9(30)	8(21)	8(24)	8(24)	5(12)
p2-81-M7::GUS	Weak	1(2)	1(1)	4(6)	1(2)	1(1)	2(3)
	Moderate&stron	2(4)	2(4)	3(8)	3(6)	3(7)	2(3)
	P(GUS)	0.08	/	0.33	0.12	0.14	0.21
	Total	12(45)	12(47)	12(47)	12(41)	12(41)	9(22)
	Non	12(47)	12(47)	12(38)	10(28)	12(35)	7(17)
p2-81-c1::GUS	Weak	0(0)	0(0)	4(6)	4(10)	4(4)	2(2)
	Moderate&stron	0(0)	0(0)	2(3)	3(4)	3(3)	3(3)
	P(GUS)	0.08	/	0.11	0.25	0.11	0.16

Number before the parenthesis indicates total number of independent events giving at least one plant with GUS activity at corresponding level. The number in parenthesis indicates total number of plants with GUS activity at corresponding level. P(GUS) means probability of obtaining plants with GUS expression.

CHAPTER 4. GENERAL CONCLUSION

Proper spatial and temporal regulation of gene expression has been shown to be essential for plant development and the response to external stimuli. Changes in the spatial-temporal gene expression pattern of certain regulatory genes produced novel phenotype and physiological traits in plants. Changes of cis-regulatory sequences in the flanking region of duplicated regulatory genes are believed to be an important mechanism for the generation of distinct tissue-specific gene expression patterns. Alleleic diversity of several plant genes with distinct tissue-specific expression patterns is found to be the result of the divergent regulatory sequences in the promoter regions. Several cis-regulatory sequences required for tissue-specific gene expression or response to external stimuli have been identified both in regulatory and structural genes. The functions of the maize *p1* gene and its homologous genes as regulators of flavonoid biosynthesis have been extensively studied and are relatively well-understood. However, the transcriptional regulation of *p1* and its homologous genes, especially the regulation of their tissue-specific expression is still at an early stage. Previous transgenic analysis of *p1* gene promoter suggested that the cis-elements required for the tissue-specific expression of *p1* gene are retained in a 561 bp fragment containing 235 bp promoter region just upstream of the transcription start site and 326 bp 5'-UTR. The maize *p2* gene shares high homology in the 5'-UTR and a 90bp fragment in the promoter region with *p1*.

In this thesis we used both transient and transgenic transformation approaches for functional analysis of the *p2* gene promoter (Chapter 3). Specifically, this study sought to determine whether any cis-regulatory elements important for the tissue-specific expression of *p2* gene are present in the promoter and leader regions. In transient assays a series of transgenes with segments from the *p2* promoter and the 5'-UTR fused with GUS reporter gene were introduced into suspension maize BMS culture cells and fresh silks by particle

bombardment. Due to the relative low expression level of *p1* and *p2* gene in plants, an AdhI intron was included in all the constructs to increase the transgene expression level. The expression levels of the *GUS* gene were normalized to internal control Luciferase or anthocyanin expression level and the relative expression level of the testing constructs were compared. The results indicated that the TATA box region in *p2* promoter is important for its promoter activity in both maize BMS cells and silk. A small region between –265 and –81 upstream of transcription start site in *p2* promoter responds differently in the BMS and silk transient assay testing system. Deletion of this small region increased the *GUS* reporter gene expression level in suspension BMS cells but had no effects on GUS expression in silk. Replacement of the 5'-UTR of *p2* gene with 5'-UTR from maize *c1* gene also affect *gus* gene expression differently in the two systems. Replacement of the *p2* gene 5'-UTR decreased GUS expression level in BMS cells but slightly increased the GUS expression level in silk. This difference may reflect the different transcription factors present in these two cell types. Further analysis of these constructs in other maize tissues and cells may help to explain the observation in this study. A construct with only 36 bp fragment of promoter region just upstream of the transcription start site and 219 bp 5'-UTR of *p2* gene still gives significant GUS expression in silk. Further analysis of this small fragment in suspension BMS cells are needed to more precisely define the promoter sequences. Also a series of mutation constructs with multiple site mutation in the small *p2* promoter region may be required to detect any potentail cis-elements that are important for the promoter activity of *p2* and its homologous genes. Analysis of transgenic plants generated from four constructs containing either the *p2* promoter and 5'-UTR segments or the *p2* promoter with 5'-UTR from maize *c1* gene showed that the expression of *gus* gene was mainly restricted to the floral tissues such as husk and silk. But the ectopic expression of *gus* gene in pericarp and cob was also observed in several independent transgenic lines. This expression may due to the complex organization of transgene in trangenic plant genomes. The expression pattern of *p2* promoter constructs in

transgenic plants is similar to that of the *p1* promoter constructs. The tissue-specific expressions of *GUS* gene in pericarp and cob that are similar to several maize *p1*-locus alleles were also observed in p2-promoter::GUS generated transgenic plants. Further analysis of these transgenic plants may provide a clearer understanding of the mechanism of tissue-specific expression of *p1*-locus alleles. A new plasmid with *p2* promoter and *p1* cDNA was constructed to test whether the *p2* promoter is sufficient to induce any visible red phlobaphene pigmentation in pericarp and cob in transgenic plants. In order to further analyze the function of the *p2*-encoded protein in flavonoid compound biosynthesis in plants a construct with *p1* promoter and *p2* cDNA was also constructed for stable transformation to test whether P2 protein can activate the transcription of structural genes for phlobaphene biosynthesis and induce any visible pigmentation in pericarp and cob in trangenic plants.

Mutant phenotype observed in plants as well as in other organisms can be result not only for spontaneous mutations in regulatory and coding regions of the gene, but also can be induced by transposable element insertion and excision. Transposable elements have been found to be widespread in plants and can be classified into several groups, including the CACTA superfamily which is observed only in plants. Transposon insertion into coding or regulatory sequences of plant genes has been shown to induce both qualitative and quantitative alteration of gene expression and even can place the expression of adjacent genes under the control of the transposable elements. The transposition and transcription of tranposons are also under the developmental regulation of the host plants. The transposon activity can be regulated epigenetically. DNA methylation, especially cytosine methylation, has been found to be correlated with transposition and transcription activity in several maize transposable elements, including *Ac*, *Mutator* and *En/Spm*.

Insertion of maize *Ac/Ds* transposable element into coding and non-coding regulatory regions of *p1* gene induced several mutant alleles with variegated phlobaphene pigmentation in both pericarp and cob. These mutant alleles are specified as *p1-vv* alleles. In chapter 2 we

report the analysis of a new maize mutant allele *p1-vv85*. This allele showed somatic and germinal instability in the *p1*-locus. Southern analysis of *p1-vv85* and its stable pigmented revertant allele *P1-rr'85* revealed that an identical 1168 bp CACTA transposable element-like fragment was inserted in both *p1-vv85* and *P1-rr'85* at the same site and orientation. No excision of this element was observed in several independent revertant plants. To understand this observation we analyzed transcription of *p1-vv85* and *P1-rr'85* using Northern blot, nuclear run-on and RT-PCR techniques. The results revealed that the rate of transcription initiation of *p1* gene in *p1-vv85* was normal. But transcription elongation was blocked in *p1-vv85*,and not in *P1-rr'85*. RT-PCR analysis suggested that *p1-vv85* contains truncated *p1* transcripts that are terminated in the vicinity of the 5'-end of the CACTA insertion element. These results suggested that the transcription of *p1* gene in *p1-vv85* was blocked in the vicinity of the 5'-end of the CACTA element insertion site. This observation provided strong evidence for a previously proposed model for the "suppressor" function of maize *En/Spm* transposable elements and suggested that the mechanism of the suppression function is conserved among CACTA elements. It also is the first direct molecular analysis of the site of the premature transcription termination caused by CACTA elements insertion. The observation of DNA methylation difference at the ends of insertion element between *p1-vv85* and P1-rr'85 suggested a model in which the revertant phenotype of *P1-rr'85* is the result of relief of suppression by trans-acting factors associated with reduced factor binding to methylated insertion element ends. Further analysis of the methylation status of p1-vv85 allele and its revertant allele in different plant tissues will help explain whether the difference in methylation of the CACTA element is due to an active demethylation process or to the failure of maintenance of the methylation in *p1-vv85*. Also the isolation of the potential trans-factor may further demonstrate the model proposed in this study. Overall, these results provide further evidence for the diverse effects of transposable elements on gene expression.

ACKNOWLEDGEMENTS

I would like to thank my major professor Dr. Thomas Peterson for his effective guidance, inspiring discussions, great patience and generous support for my research work. It is a great pleasure for me to have the opportunity to work in his lab and learn both scientific knowledge and experimental techniques which will certainly benefit my rest life.

I would also like to thank Dr. Drena Dobbs and Dr. Kan Wang for serving as my POS committee members and giving me great advice on my thesis.

I would like to thank Dr. Kan Wang, Lise Marcell, Bronwyn Frame and Tina Fonger at the Plant Transformation Facility, Iowa State University for their help in generating transgenic plants. I would also like to thank Cory Heilmann in Department of Statistics, Iowa State University, for his help in the statistical analysis of my research data.

I would like to express my thanks to Dr. Lyuda Sidorenko, Dr. Surinder Chopra, Dr. Suzy Cocciolone, Dr. Yongli Xiao, Dr. Jianbo Zhang, Dr. Peifen Zhang, Cizhong Jiang, Lakshminarasimhan Krishnaswamy, Zhuying Li, Terry Olson, Diane Sickau and Feng Zhang for their advice and help on my research work and friendship during my stay in the lab.

Finally my great thanks to my husband, my parents and my sister and my friends for their endless love, encouragements and supports for me to finish my study.