Organ-specific regulation and molecular evolution of the maize *pericarp color1* gene

by

Feng Zhang

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Genetics

Program of Study Committee:
Thomas A. Peterson, Major Professor
Diane Bassham
Phil Becraft
Steven Whitham
Dan Voytas

Iowa State University

Ames, Iowa

2005

Graduate College
Iowa State University

This is to certify that the doctoral dissertation of

Feng Zhang

has met the dissertation requirements of Iowa State University

Signature was redacted for privacy.

Major Professor

Signature was redacted for privacy.

For the Major Program

TABLE OF CONTENTS

ABSTRACT

The maize *pericarp color1* (*p1*) gene encodes an R2R3 Myb-like transcription factor that regulates the flavonoid biosynthetic pathway in floral organs, most notably kernel pericarp and cob. Alleles of the *p1* gene condition distinct tissue-specific pigmentation patterns; to elucidate the molecular basis of these allele-specific expression patterns, we characterized two novel *P1-rw* (red pericarp/white cob) alleles, *P1-rw1077* and *P1-rw751::Ac*. Structural analysis of *P1-rw1077* indicated that this allele was generated by recombination between *p1* and the tightly linked paralogous gene, *p2*. In the resulting gene, the *p1* coding sequence was replaced by the *p2* coding sequence, while the flanking *p1* regulatory sequences remained largely preserved. The red pericarp color specified by *P1-rw1077* suggests that the *p1*- and *p2*-encoded proteins are functionally equivalent as regulatory factors in the flavonoid biosynthesis pathway. Sequence analysis shows that the *P1-rw1077* allele lacks a 386 bp sequence in a distal enhancer region 5 kb upstream of the transcription start site. An independently-derived *P1-rw* allele contains an *Ac* insertion into the same sequence, indicating that this site likely contains cob glume-specific regulatory elements.

The *p1* alleles exhibit a high degree of genetic and phenotypic diversities. To understand the driving force(s) for generation of those diversities as well as the evolutionary history of the *p1* locus, 6 single copy natural *p1* alleles, including 4 new *p1* alleles and the previously cloned *P1-rr4B2* and *P1-rw1077* alleles, were characterized and compared. Structural and sequence analyses revealed 3 distinctive types of the distal enhancer in these alleles. The structure of the distal enhancer region correlates with the expression of each allele in cob glumes. Further investigation suggested that homologous recombination

between the 5' and 3' non-coding direct repeats promotes diversification of the distal enhancer. As a result, novel distal enhancer types were generated during the *p1* evolution, giving rise to a novel phenotypic variation. Moreover, based on the structures in the non-coding regions, we proposed a stepwise model to account for evolutionary pathways of the distinct *p1* alleles. In this model, recombination between duplicated non-coding repeats is the major driving force in diversifying the *p1* alleles, and creating novel cis-regulatory regions. Because tandem and segmental duplication are common in both animal and plant genomes, homologous recombination between non-coding duplicated sequences could have been an important means to generate genetic and phenotypic variations.

CHAPTER 1. GENERAL INTRODUCTION

The colorful genetic system to study gene regulation

In higher eukaryotes, the expression of most genes is finely regulated in response to a wide range of tissue, developmental and environmental signals. Thus, understanding of mechanisms of gene regulation is one of the most important and challenging fields in basic biology research. In plants, genes involved in flavonoid pigments biosynthetic pathway provide an excellent system for understanding diverse mechanisms of gene regulation (Winkel-Shirley, 2001).

Flavonoids constitute a class of aromatic compounds that are derived from secondary metabolism of phenylalanine and malonyl-coenzyme A (Winkel-Shirley, 2001). These compounds include flavones, flavonols, anthocynins, proanthocynins (condensed tannins), aurones, and isoflavonoids. Further modifications, such as hydroxylation, methylation, acylation, or glucosylation, make flavonoids even more diverse in nature. All subgroups of flavonoids have been found in most higher plants, and provide the majority of resources for red, blue and purple pigments in plant vegetative and floral organs. The well known functions of flavonoids have been shown as: pollinator attraction, UV damage protection, plant pathogen and insects resistance, germination of pollen tubes. (for a comprehensive review, see (Winkel-Shirley, 2001; Winkel-Shirley, 2002).

Because of easily scored phenotypes, the genes in the flavonoid biosynthetic pathway have served as genetic marker systems for various studies, such as elucidation of basic genetic principles, mechanisms of Ac/Ds transposition (Athma et al, 1992; Moreno et al 1992), paramutation (for recent review, (Chandler and Stam, 2004), and organ-specific gene

regulation (Chopra et al 1996; Chopra 1998;Sidorenko et al 1999). In recent years, the flavonoid biosynthetic pathway has been elucidated in maize, snapdragon (*Antirrhinum majus*), petunia, and Arabidopsis (for review, see Holton and Cornish, 1995; Mol et al., 1998). Two best studied flavonoid synthetic pathways in maize are shown in Figure 1. Both pathways start with phenylalanine, which is in turn converted to 4-coumaroyl-CoA. 4-coumaroyl-CoA is joined with malonyl-CoA to produce chalcone, catalyzed by chalcone synthase (CHS). The step from chalcone to flavanone can be achieved either catalytically by chalcone isomerase (CHI), or spontaneously at very low rate. As a major branch point, the flavanone can lead to different pathways through catalysis by different sets of enzymes. Catalyzed stepwise by flavanone 3-hydroxylase (F3H), dihydroflavonol 4-reductase (DFR, encoded by *a1*), anthocyanidin synthase (AS, encoded by *a2*), and UDPglucose flavonoid 3-oxy-glucosyltransferase (UF3GT, encoded by *bz1*), flavonone is converted to anthocyanin. In contrast, catalyzed by dihydroflavonol 4-reductase (DFR) followed by polymerization, flavanone can go to phlobaphene synthetic pathway.

It has been shown that two sets of regulatory genes control the anthocyanin and phlobaphene pathways respectively. The R2R3 Myb-like transcription factors, *c1* (*cololess1*) and *pl1* (*purple plant1*) gene, interact with bHLH transcription factors, *r1* (*red1*) and *b1* (*booster1*), to regulate synthesis of purple pigments, anthocyanin (Paz-Ares et al., 1987; Ludwig et al., 1989; Radicella et al., 1991; Cone et al., 1993), while the other R2R3 Myb transcription factor, *p1* (*pericarp color1*), acts alone to control accumulation of red pigments, phlobaphene (Grotewold et al., 1994). In each regulatory locus, natural variations confer distinct expression patterns at maize vegetative or floral organs. Studies of these naturally

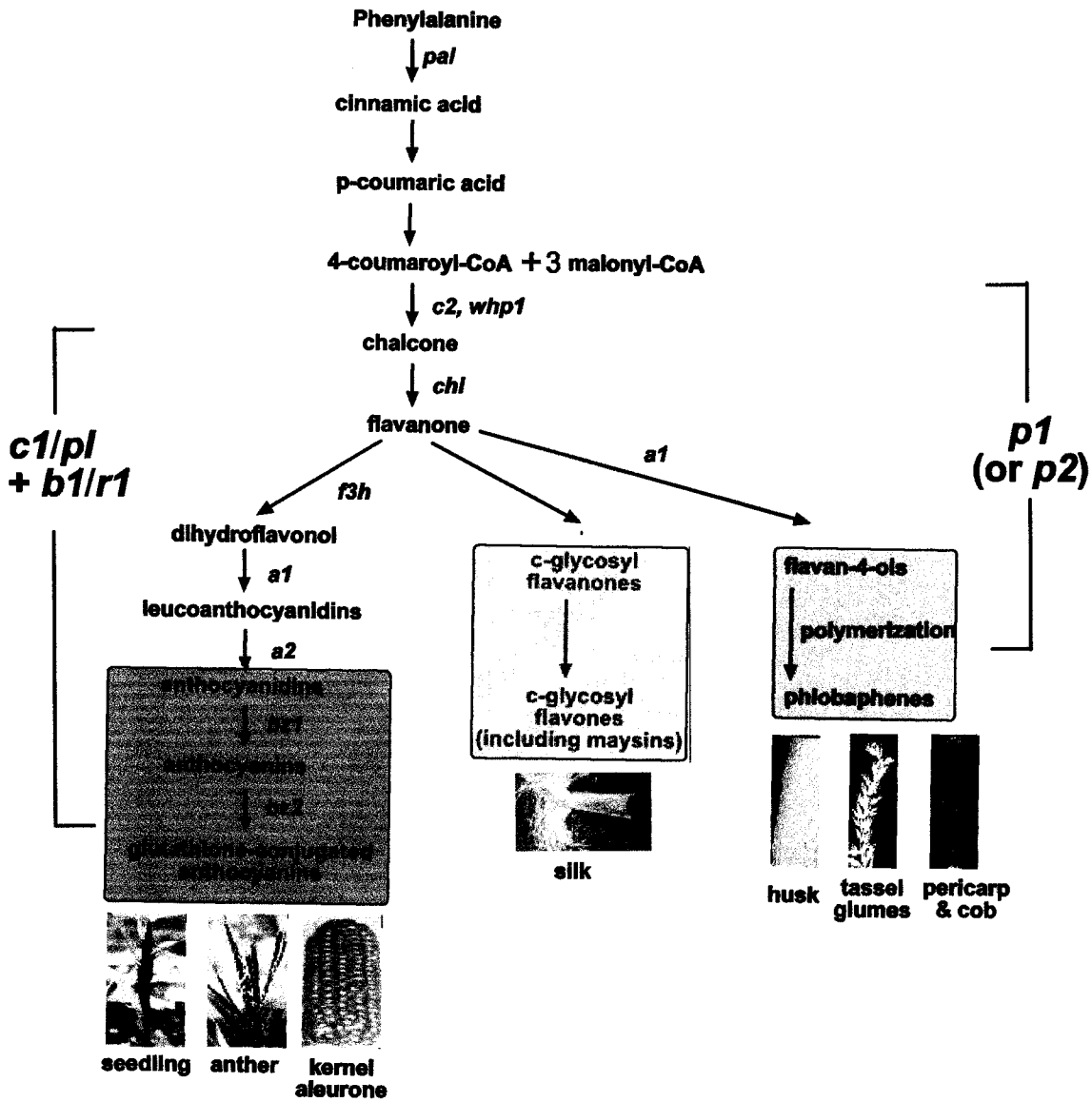


Figure 1. Flavonoid biosynthetic pathways in maize. The *c1/pl* interacting with *b1/r1* genes regulate production of anthocyanins, while the *p1* gene regulates phlobaphene biosysthesis. Photos, showing accumulation of anthocynins in diverse maize organs, are courtesy of maizeGDB (www.maizegdb.com; Lawrence et al., 2004).

occurring alleles have provided insights in gene regulation in higher organisms. For instance, investigation of the *b1* alleles suggested that the distinct expression patterns result from differences in cis-regulatory regions, which could be generated by recent DNA rearrangement (Selinger et al., 1998; Selinger and Chandler, 1999). Functional analysis of upstream non-coding regions identified an aleurone-specific regulatory region (Selinger et al., 1998). In addition, studies of the *r1* genes revealed three genetic components: S, for pigmentation in seed aleurone; P, for anthers and coleoptile; Lc, for midrib, ligule, auricle, glume, lemma, palea, and pericarp tissues (Ludwig et al., 1989; Walker et al., 1995). Results from Ludwig et al (1989) suggested that the diverse pigment patterns conferred by the distinct components are also due to different promoter regions.

The regulatory loci in the flavonoid pathways are also valuable to study a widespread epigenetic phenomenon, paramutation. Paramutation was coined by Brink in (1968) to describe the observation that combining two specific *r1* alleles results in heritable reduction in gene expression of one of the alleles. The allele with the altered expression is termed as the paramutable allele, while the allele that induces the change is termed as the paramutagenic allele. The expression states of the paramutable alleles are unstable, and can be reversed to the original states with various frequencies. The paramutation-like phenomena were later observed in other regulatory loci for pigment production, such as *b1*, *pl1* and *p1* (Coe, 1959; Hollick et al., 1995; Hollick et al., 2000; Sidorenko and Peterson, 2001), as well as many other genes from diverse species (Chandler and Stam, 2004). Because of the easily scored phenotype, the regulatory loci for pigment synthesis offer an attractive system to study cis and trans factors, which are required for establishment and maintenance of paramutation. Recent studies from Chandler's group have successfully

identified cis-regulatory region participating in paramutaion (Stam et al., 2002a; Stam et al., 2002b), and also characterized a serious of mutants that affect paramutation, i.e. *mediator of paramutation* (*mop*) (Dorweiler et al., 2000; Hollick and Chandler, 2001).

The *pericarp color1* gene: the genetic marker for a century

As one of the major regulators of flavonoid pigment biosynthesis pathways in maize, the *p1* (*pericarp color1*) gene confers red pigmentation primarily in floral organs, such as kernel pericarp, cob glumes, tassel glumes, and husk (Figure 1). Due to the remarkable pigmented phenotype in kernel pericarp and cob glumes, the *p1* gene has been used as a genetic marker since 1911 (Emerson, 1911). Like other regulatory loci in the pathway, a large number of the *p1* alleles has been collected in terms of pigmentation patterns in kernel pericarp and cob glumes (Brink and Styles, 1966; Figure 2). According to the presence or absence of pigments in pericarp and cob glumes, the *p1* alleles are designated as *P1-rr* (red pericarp/reb cob), *P1-wr* (white pericarp/red cob), *P1-rw* (red pericarp/white cob) and *p1-ww* (white pericarp/white cob).

The first-cloned *p1* allele was the *P1-rr* allele, using the transposon tagging technique (Lechelt et al., 1989). The *p1* gene encodes R2R3 Myb-like transcription factor, which contains a conserved N-terminal DNA-binding domain (Myb-domain). The *P1-rr* allele contains a complex gene structure with an 8 kb coding region (three exons and two introns) flanked by 5.2kb long direct repeats. In 1998, Chopra et al cloned another *p1* allele, *P1-wr*, which consists of more than six copies of *p1* sequences arranged in a head to tail tandem array (Chopra et al., 1998). The gene unit of the *P1-wr* complex shares high sequence

similarity with *P1-rr*: 99% similarity in the 5.2 kb upstream promoter region and 99.9% similarity in the coding region.

Regulation of the *p1* expression

Cis-regulatory regions in the *P1-rr* allele

The cis-regulatory regions of *P1-rr* have been characterized by transposon mutagenesis, transient expression assay, and transgenic analysis (Athma et al., 1992; Moreno et al., 1992; Sidorenko et al., 1999; Sidorenko et al., 2000). The results revealed three distinct regulatory regions: the basal promoter region (from -235 to +326; the transcription start site as +1); the proximal enhancer, located within a ~1kb fragment immediate upstream of 5' of the basal promoter; and the 1.2 kb distal enhancer region, located ~5kb upstream of transcription start site. Because most plant gene promoters analyzed to date appear to be relatively compact (for review, see (Singh, 1998) and references therein), characterization of the 1.2 kb distal enhancer at the *p1* locus is of most interest. A number of studies have been conducted on this enhancer region, which are summarized as following:

1. Insertions of Ac transposable elements in the 1.2 kb distal enhancer region dramatically reduce red pigmentation in both kernel pericarp and cob glumes (Athma et al., 1992; Moreno et al., 1992)
2. Functional analysis of the 1.2 kb distal enhancer region by transient and transgenic assays indicated that this region is capable of significantly increasing the expression of a GUS (β-glucuronidase) reporter driven by the *p1* basal promoter (Sidorenko et al., 1999; Sidorenko et al., 2000).

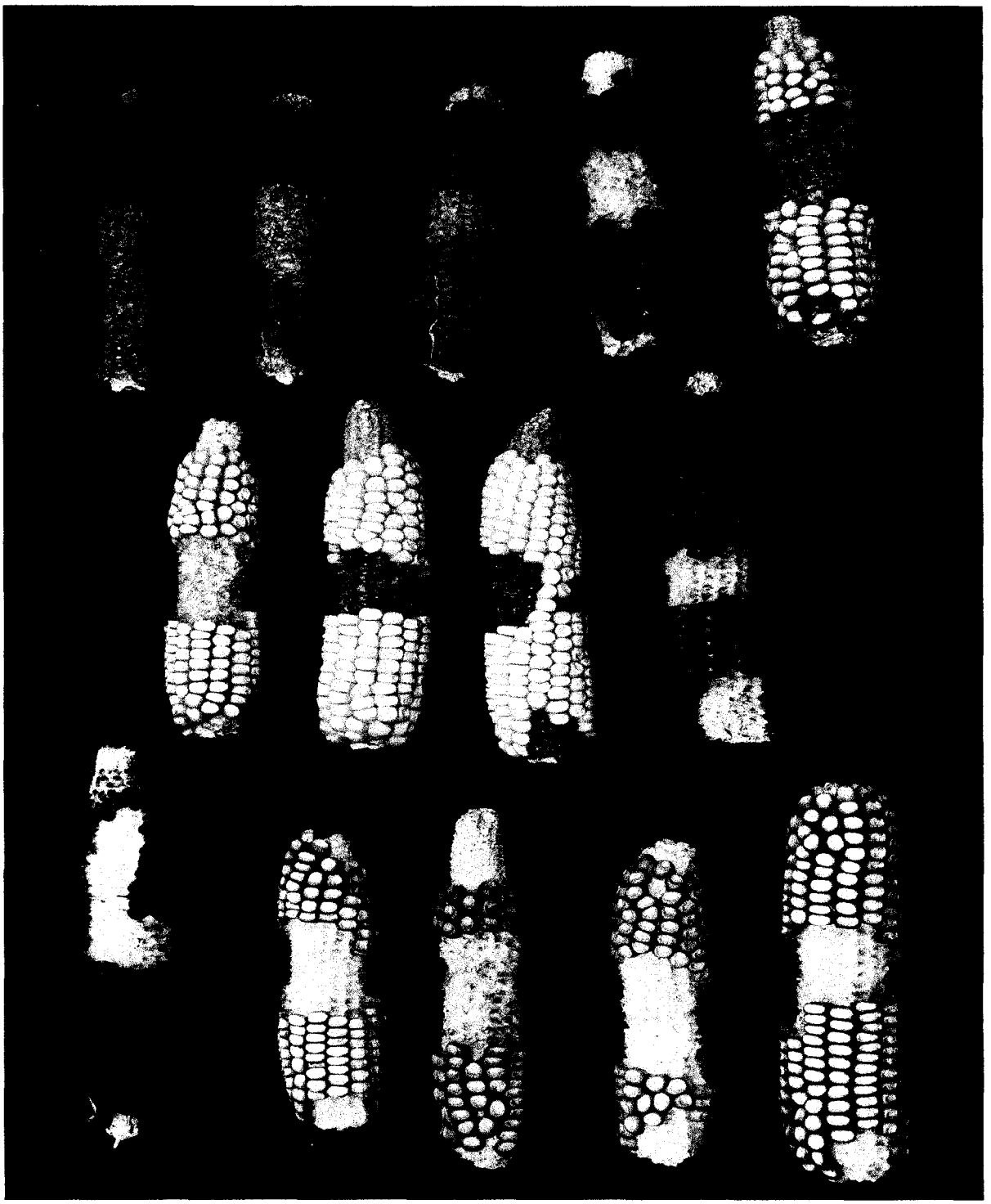


Figure 2. Phenotypic diversity in the natural *p1* alleles. All *p1* alleles are introgressed into 4co63 background (Brink and Styles, 1966).

3. A spontaneous epi-mutant of *P1-rr*, *P1-pr* (patterned pericarp and red cob glumes), was identified as exhibiting decreased pigmentation in kernel pericarp. Characterization of the *P1-pr* allele showed that pigment reduction in kernel pericarp is correlated with hyper-methylation and condensed chromatin structure in the 1.2 kb distal enhancer region (Das and Messing, 1994; Lund G et al., 1995).
4. The 1.2 kb distal enhancer region in *p1*-expressing organs, such as kernel pericarp, has more open chromatin structure than the region in *p1*-nonexpressing organs, such as leaves (Das and Messing, 1994; Lund G et al., 1995).

Taken together, these data indicted that the 1.2 kb distal enhancer region plays a critical role in regulating expression of the *p1* gene. The function of this enhancer could be controlled by chromatin structure alteration and DNA methylation pattern changes. It is well known that chromatin structure alteration and DNA methylation changes can affect accessibility of regulatory elements to transcription factors (Emerson, 2002). Further dissection of the 1.2 kb distal enhancer region is necessary to isolate cis-regulatory motifs, and identify trans-factors that regulate *p1* expression.

Organ-specific regulation of the *p1* expression

Lessons from the *p1* natural alleles: Perhaps the most intriguing observation about the *p1* gene is organ-specific expression of the *P1-wr* and *P1-rw* alleles, which only show red pigments in either cob glumes or kernel pericarp. As mentioned above, the *P1-wr* allele has been characterized in great detail. Sequence comparisons have shown high sequence similarity in both coding region and non-coding regions between *P1-wr* and *P1-rr*, suggesting that DNA polymorphisms could not explain organ-specific expression observed

on *P1-wr*. In addition, recent studies by Chopra et al (2003) indicated that the phenotype of *P1-wr* can be converted to that of *P1-rr* in the presence of an unlinked genetic factor, *Ufo1* (*Unstable factor for orange1*). The gain of pigmentation in kernel pericarp of *P1-wr* accompanies hypomethylation in the multiple-copy *p1* gene complex, which is otherwise hypermethylated compared to *P1-rr*. Furthermore, a recent survey of gene structure of natural *P1-rr* alleles revealed that many *P1-rr* alleles contain multiple copy gene structure resembling *P1-wr* (Cocciolone et al., 2001). Taken data together, these results suggested that epigenetic regulation may account for organ-specific regulation of the *P1-wr* allele. The possible mechanism is that the loss of pigmentation in kernel pericarp results from organ-specific gene silencing induced by repetitive gene structure at the *P1-wr* allele (Chopra et al., 1998; Cocciolone et al., 2001). Repeats-induced gene silencing (RIGS) was reported from studies of trangenes in many species (Dorer and Henikoff, 1994; Meyer, 1996; Ye and Signer, 1996; Jensen et al., 1999). It has been suggested that repetitive sequences could induce formation of heterochromatin, i.e. transcriptional gene silencing, or post-transcriptional gene silencing (see recent review by (Fagard and Vaucheret, 2000). As an endogenous locus, the *P1-wr* allele provides a unique system to study mechanisms of repetitive-structure related gene silencing. Further isolation of trans-factors that affect *P1-wr* gene silencing will undoubtedly help us better understand epigenetic pathways.

Another naturally occurring allele that has been useful in understanding organ-specific *p1* regulation is the *P1-pr* (pattern pericarp and red cob glumes) allele. As mentioned previously, *P1-pr* is a spontaneous allele from the *P1-rr4B2* allele, which shows dramatic pigment reduction in kernel pericarp but not in cob glumes (more precisely, the pigment reduction in kernel pericarp is more severe than in cob glumes). No structural

differences were detected between *P1-pr* and *P1-rr4B2*. *P1-pr*, however, is hyper-methylated and has condensed chromatin structure in several regions, including the 1.2 kb distal enhancer region (Das and Messing, 1994; Lund G et al., 1995). Thus, the phenotypic alteration between *P1-rr* and *P1-pr* are also likely due to epigenetic regulation as observed in *P1-wr*. Because the pigment reduction seems to be uncoupled between kernel pericarp and cob glumes, it has been suggested that two discrete regulatory regions exist at the *p1* non-coding region that are responsible for *p1* expression in pericarp and cob glumes respectively (Das and Messing, 1994; Lund G et al., 1995).

Lessons from transgenic studies: In addition to studies of the *p1* natural alleles, transgenic analyses also suggested epigenetic regulation of *p1* expression. As described above, the capability and specificity of the *P1-rr* promoter have been tested using a transgenic construct that carries the 6.2 kb *P1-rr* full promoter driving a GUS reporter gene (Cocciolone et al., 2000; Sidorenko et al., 2000). Interestingly, expression patterns of GUS gene resemble not only RR pattern, but also WR and RW patterns. The plants with WR pattern frequently showed hypermethylation in the transgene locus, compared to the plants with RW pattern. This observation is consistent with the data from the endogenous *p1* gene. Moreover, a hierarchical expression pattern was observed in the P1-rr::GUS transgenes, that is, if the GUS gene construct in a transgenic plant expresses in early-initiated organs, it frequently expresses in the late initiated organs. Such non-random expression patterns suggested developmental regulation of the *p1* promoter, and was explained by a chromatin structure transition model (Cocciolone et al., 2000). The model proposes that the accessibility of the *p1* promoter to transcription factors is restricted by closed chromatin structure, and that transition of chromatin conformation from close to open state during

development will increase the accessibility and potentiate transcription. After this transition occurs, the chromatin structure will remain open in successive developmental stages. Although this model can nicely account for most observations in this transgenic assay, the *P1-wr* like expression pattern would be an exception of the model because the GUS gene is expressed in the early initiated organ, cob glumes, but not in the later developed organ, kernel pericarp.

Although the *P1-rr* and *P1-wr* alleles share high sequence similarity in both coding and non-coding regions, there are still many small DNA polymorphisms between two alleles. In order to test if sequence polymorphisms determine tissue-specificity of the *P1-wr* allele, the transgene constructs were made containing the *P1-rr* promoter driving the *P1-wr* coding sequence and *P1-wr* promoter driving *P1-rr* coding sequence (Cocciolone et al., 2001). The result showed that both constructs can confer *P1-rr* phenotype, which further confirmed that DNA variations are not responsible for WR phenotype. Again, inverse correlations between DNA methylation and pericarp pigmentation were also observed in this transgenic experiment. Taken together, these results convincingly suggested that expression of the *p1* gene could be epigenetically controlled, and produces WR pigmentation patterns. Epigenetic regulation, however, is not the whole story of the *p1* alleles. In chapter 2 of this dissertation, we will present evidence based on investigation of the *P1-rw* alleles for the existence of a cob glumes-specific enhancer region.

Paramutation of the *p1* gene

An unexpected result from the *p1* trangene studies is that the 1.2 kb distal enhancer region, but not other regulatory regions, can induce gene silencing of the endogenous *P1-rr*

allele (Sidorenko and Peterson, 2001). The silenced *P1-rr* allele, showing reduced pigmentation in both pericarp and cob glumes, was converted into a paramutagenic allele, which can induce silencing of a naïve *P1-rr* (the paramutable allele). Silencing of the *P1-rr* allele is associated with the hypermethylation of the distal enhancer regions, suggesting again involvement of epigenetic mechanisms in the *p1* regulation. It is interesting to note that the 1.2 kb region has capability to both enhance gene expression and induce gene silencing. A similar observation was also reported in the maize *b1* locus. In *b1*, the cis-elements participating in paramutation were located in the same region that exhibits enhancer activity. An intriguing question is whether the region(s) required for enhancer activity are separable from the region(s) required for paramutation. Further dissection of the 1.2 kb region will help address this question.

Origin and evolution of the *p1* locus

In the maize genome, a homologous gene of *p1*, *p2*, was identified recently (Zhang et al., 2000). The *p2* gene is located upstream of *p1*, in the same transcription orientation. Sequence analysis indicated that *p2* shares high sequence similarity with *p1* in the coding sequence. In the promoter region, however, the two genes are completely divergent from a point immediately upstream of TATA box (-90 from the transcription start site). Like many duplicated genes, the *p1* and *p2* genes have distinct expression patterns, possibly due to diverged non-coding regulatory regions. The *p1* gene expresses in kernel pericarp, cob glumes, tassel glumes, and silks, while *p2* transcripts are found in developing anthers and silks. The sequence data suggested that the *p1* and *p2* gene might be generated by a recent duplication event. To understand the origin and evolution of the *p1* and *p2* complex, a *p*-

homologous gene was isolated from *Z. mays* subbsp. *Parviglumis*, the ancestor of modern maize. Based on sequence analysis of the ancestral *p* gene, a segmental gene duplication model was proposed to account for the origin and structural organization of *p1* and *p2* genes (Figure 3). This model postulated that the *p1*/*p2* gene complex was generated by a recent tandem duplication event from a single copy *p* progenitor gene. The duplicated region was ~10 kb long, extending from the -90 promoter region to ~6 kb downstream of the last exon. As the result, the previous 3' flanking region was inserted into the 5' side of the downstream gene, thus generating the *p2* and *p1* backbone. Subsequent insertions of retro-transposons separated *p2* from the *p1* gene, and left *p1* with its current structure containing the long direct non-coding repeats. One consequence of this duplication event is the creation of novel regulatory regions for the newly formed *p1* gene, which could cause differential expression patterns observed in *p1* and *p2*. After formation of the *p1* gene backbone, unequal crossover between the long direct repeats could generate the multiple-copy structure of the *P1-wr* allele, and other further modification will give rise to other single copy *p1* alleles, which will be discussed in chapter 3.

Dissertation organization

This dissertation consists of four chapters. Chapter 1 is the general introduction that provides the background of my Ph.D. thesis project. Chapter 2 and Chapter 3 are research manuscripts. In Chapter 2, by investigating the *P1-rw* alleles, we identified paralogous gene recombination between the *p1* and *p2* genes, and a cob glumes-specific enhancer region. In Chapter 3, we focused on evolutionary dynamic of the *p1* non-coding regulatory regions, and proposed a model of *p* gene evolution to account for genetic and phenotypic diversity

observed in the distinct *p1* alleles. Chapter 4 contains general summary of the research data present in chapters 2 and 3.

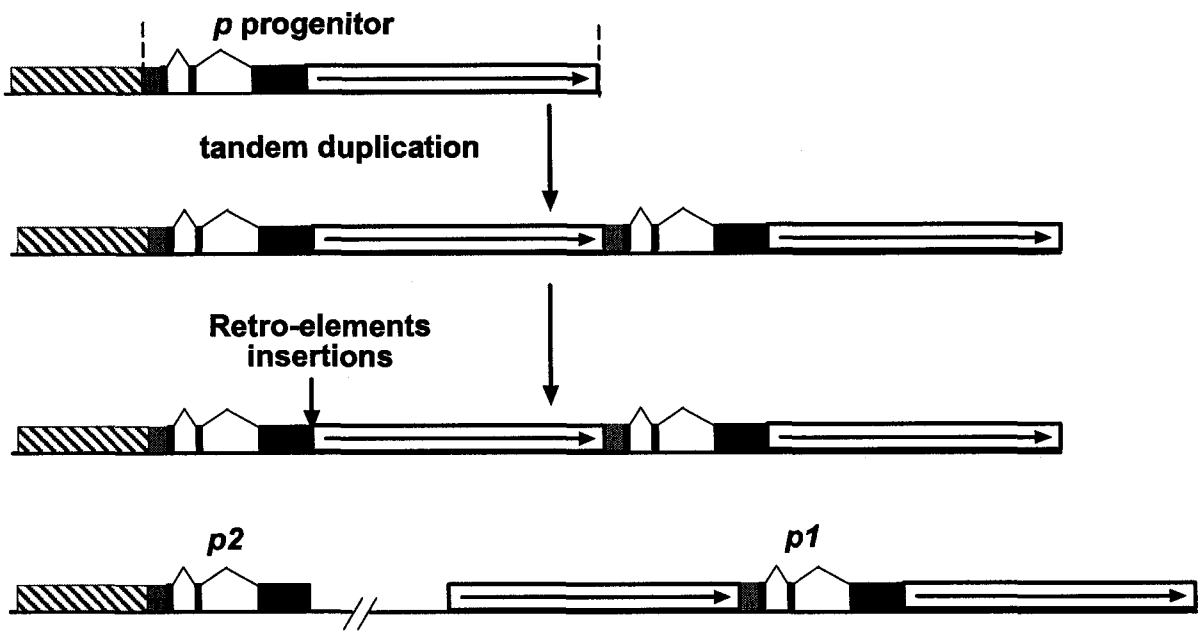


Figure 3. Duplication model for origin of the *p1* and *p2* genes (reproduced from Zhang et al (2000)). The black boxes connected with thin lines represent exon and intron structures. The gray boxes indicate the basal promoter and 5' UTR region that are shared by *p1* and *p2*. The hatched boxes represent the upstream promoter region in the *p* progenitor gene and the *p2* gene. The region between dashed lines is the ~10 kb sequence that was duplicated. The open boxes with black arrows indicate the duplicated non-coding region. The drawing is not to scale.

References

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

Brink, R., and Styles, D. (1966). A collection of pericarp factors. Maize Genet. Cooperation Newsletters 40, 149-160.

Brink, R.A., Styles, E.D., and Axtell, J.D. (1968). Paramutation: directed genetic change. Paramutation occurs in somatic cells and heritably alters the functional state of a locus. Science 159, 161-170.

Chandler, V.L., and Stam, M. (2004). Chromatin conversations: mechanisms and implications of paramutation. Nat Rev Genet 5, 532-544.

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

Chopra, S., Cocciolone, S.M., Bushman, S., Sangar, V., McMullen, M.D., and 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.

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, 467-478.

Coe, E.H. (1959). A regular and continuing conversion-type phenomenon at *b* locus in maize. Maydica 24, 49-58.

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, 1795-1805.

Das, O.P., and Messing, J. (1994). Variegated phenotype and developmental methylation changes of a maize allele originating from epimutation. Genetics 136, 1121-1141.

Dorer, D.R., and Henikoff, S. (1994). Expansions of transgene repeats cause heterochromatin formation and gene silencing in Drosophila. Cell 77, 993-1002.

Dorweiler, J.E., Carey, C.C., Kubo, K.M., Hollick, J.B., Kermicle, J.L., and Chandler, V.L. (2000). mediator of paramutation1 Is Required for Establishment and Maintenance of Paramutation at Multiple Maize Loci. Plant Cell 12, 2101-2118.

Emerson, B.M. (2002). Specificity of gene regulation. Cell 109, 267-270.

Emerson, R.A. (1911). Genetic correlation and spurious allelomorphism in maize. Ann. Rept. Nebraska Agric. Exp. Sta. 24, 59-90.

Fagard, M., and Vaucheret, H. (2000). (TRANS)GENE SILENCING IN PLANTS: How Many Mechanisms? Annual Review of Plant Physiology and Plant Molecular Biology 51, 167-194.

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.

Hollick, J.B., and Chandler, V.L. (2001). Genetic Factors Required to Maintain Repression of a Paramutagenic Maize pl1 Allele. Genetics 157, 369-378.

Hollick, J.B., Patterson, G.I., Asmundsson, I.M., and Chandler, V.L. (2000). Paramutation Alters Regulatory Control of the Maize pl Locus. Genetics 154, 1827-1838.

Hollick, J.B., Patterson, G.I., Coe-Jr., E.H., Cone, K.C., and Chandler, V.L. (1995). Allelic Interactions Heritably Alter the Activity of a Metastable Maize pl Allele. Genetics 141, 709-719.

Jensen, S., Gassama, M.P., and Heidmann, T. (1999). Taming of transposable elements by homology-dependent gene silencing. Nat Genet 21, 209-212.

Lawrence, C.J., Dong, Q., Polacco, M.L., Seigfried, T.E., and Brendel, V. (2004). MaizeGDB, the community database for maize genetics and genomics. Nucleic Acids Res 32, D393-397.

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, 225-234.

Ludwig, S.R., Habera, L.F., Dellaporta, S.L., and Wessler, S.R. (1989). Lc, a member of the maize R gene family responsible for tissue-specific anthocyanin production, encodes a protein similar to transcriptional activators and contains the myc-homology region. Proc Natl Acad Sci U S A 86, 7092-7096.

Lund G, Prem D, and J., M. (1995). Tissue-specific DNase I-sensitive sites of the maize P gene and their changes upon epimutation. Plant J 7, 797-807.

Meyer, P. (1996). Repeat-induced gene silencing: common mechanisms in plants and fungi. Biol Chem Hoppe Seyler 377, 87-95.

Moreno, M.A., Chen, J., Greenblatt, I., and Dellaporta, S.L. (1992). Reconstitutional mutagenesis of the maize P gene by short-range Ac transpositions. Genetics 131, 939-956.

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, 3553-3558.

Radicella, J.P., Turks, D., and Chandler, V.L. (1991). Cloning and nucleotide sequence of a cDNA encoding B-Peru, a regulatory protein of the anthocyanin pathway in maize. Plant Mol Biol 17, 127-130.

Selinger, D.A., and Chandler, V.L. (1999). Major recent and independent changes in levels and patterns of expression have occurred at the b gene, a regulatory locus in maize. PNAS 96, 15007-15012.

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, 1125-1138.

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

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, 319-335.

Sidorenko, L.V., Li, X., Cocciolone, S.M., Chopra, S., Tagliani, L., 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.

Singh, K.B. (1998). Transcriptional Regulation in Plants: The Importance of Combinatorial Control. Plant Physiol. 118, 1111-1120.

Stam, M., Belele, C., Dorweiler, J.E., and Chandler, V.L. (2002a). Differential chromatin structure within a tandem array 100 kb upstream of the maize b1 locus is associated with paramutation. Genes Dev 16, 1906-1918.

Stam, M., Belele, C., Ramakrishna, W., Dorweiler, J.E., Bennetzen, J.L., and Chandler, V.L. (2002b). The regulatory regions required for B' paramutation and expression are located far upstream of the maize b1 transcribed sequences. Genetics 162, 917-930.

Walker, E., Robbins, T., Bureau, T., Kermicle, J., and Dellaporta, S. (1995). Transposon-mediated chromosomal rearrangements and gene duplications in the formation of the maize R-r complex. EMBO J. 14, 2350-2363.

Winkel-Shirley, B. (2001). Flavonoid Biosynthesis. A Colorful Model for Genetics, Biochemistry, Cell Biology, and Biotechnology. Plant Physiol. 126, 485-493.

Winkel-Shirley, B. (2002). Biosynthesis of flavonoids and effects of stress. Curr Opin Plant Biol 5, 218-223.

Ye, F., and Signer, E.R. (1996). RIGS (repeat-induced gene silencing) in Arabidopsis is transcriptional and alters chromatin configuration. PNAS 93, 10881-10886.

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

CHAPTER 2. COMPARISONS OF MAIZE *PERICARP COLOR 1* ALLELES REVEAL PARALOGOUS GENE RECOMBINATION AND AN ORGAN-SPECIFIC ENHANCER REGION

A paper published in the Plant Cell[1]

Feng Zhang and Thomas Peterson

Abstract

The maize *p1* gene encodes an R2R3 Myb-like transcription factor that regulates the flavonoid biosynthetic pathway in floral organs, most notably kernel pericarp and cob. Alleles of the *p1* gene condition distinct tissue-specific pigmentation patterns; to elucidate the molecular basis of these allele-specific expression patterns, we characterized two novel *P1-rw* (red pericarp/white cob) alleles, *P1-rw1077* and *P1-rw751::Ac*. Structural analysis of *P1-rw1077* indicated that this allele was generated by recombination between *p1* and the tightly linked paralogous gene, *p2*. In the resulting gene, the *p1* coding sequence was replaced by the *p2* coding sequence, while the flanking *p1* regulatory sequences remained largely preserved. The red pericarp color specified by *P1-rw1077* suggests that the *p1*- and *p2*-encoded proteins are functionally equivalent as regulatory factors in the flavonoid biosynthesis pathway. Sequence analysis shows that the *P1-rw1077* allele lacks a 386 bp sequence in a distal enhancer region 5 kb upstream of the transcription start site. An independently-derived *P1-rw* allele contains an *Ac* insertion into the same sequence, indicating that this site likely contains cob glume-specific regulatory elements.

---

[1]Reprinted with permission of the Plant Cell, 2005, 17, 903-914

Introduction

In multicellular organisms, most genes exhibit tissue-specific and/or developmental stage-specific expression. The precise regulation of gene expression requires the coordinated interplay of many components at distinct levels (Singh, 1998; Emerson, 2002). Moreover, alleles of individual genes can exhibit strikingly different expression patterns. Examples of such allele-specific gene expression have been demonstrated in a number of plant genes that regulate flavonoid biosynthetic pathways (Radicella et al., 1992; Consonni et al., 1993; Scheffler et al., 1994; Chopra et al., 1998; Edwards et al., 2001; Pilu et al., 2003). Elucidation of the mechanism(s) by which different alleles of a single gene can confer distinct expression patterns will provide valuable insights into the regulation of gene expression in higher organisms.

In maize, the *p1* (*pericarp color 1*) gene conditions red flavonoid pigment, phlobaphene, in floral organs, such as kernel pericarp (seed coat), cob glumes (floral bracts subtending the kernel), tassel glumes and silk (Coe et al., 1988). The *p1* gene encodes an R2R3 Myb-like transcription factor that regulates structural genes for flavonoid biosynthesis including *c2* (chalcone synthase), *chi* (chalcone flavonone isomerase), and *a1* (dihydroflavonol reductase) (Grotewold et al., 1994). Over 100 *p1* alleles have been described, each of which specify distinct pigmentation patterns (Brinks and Styles, 1966; Cocciolone et al., 2001). Prototype alleles conferring distinct pigmentation patterns of pericarp and cob glumes include *P1-rr* (red pericarp/red cob), *P1-wr* (white pericarp/red cob), *P1-rw* (red pericarp/white cob) and *p1-ww* (white pericarp/white cob) (Figure 1A).

Two alleles, *P1-rr4B2* and *P1-wr*, have been cloned and compared at the molecular level (Lechelt et al., 1989; Chopra et al., 1996; Chopra et al., 1998). The *P1-rr4B2* allele

contains a single coding sequence flanked by 5.2 kb direct repeats, whereas *P1-wr* has six or more *p1* gene copies organized in a head-to-tail tandem array. The gene unit of the *P1-wr* complex shares high sequence similarity with *P1-rr4B2*: 99% similarity in the 5.2 kb upstream promoter region and 99.9% similarity in most of the coding region (Chopra et al., 1998). Analyses of natural *p1* alleles and transgenic plant experiments suggest that the distinct expression pattern of the *P1-wr* allele is governed by epigenetic controls, and not by sequence polymorphisms (Cocciolone et al., 2001). A tissue-specific repeat-induced gene-silencing model was proposed to account for the lack of pigmentation in the pericarp of *P1-wr* (Chopra et al., 1998). The silenced state of *P1-wr* can be relieved by the presence of a *trans*-acting factor, *Ufo1*, which results in red color pericarp with *P1-wr*. Moreover, *Ufo1*-induced reactivation of *P1-wr* expression in pericarps is associated with demethylation in the *P1-wr* gene complex (Chopra et al., 2003).

To better understand mechanism(s) of allele-specific gene expression patterns, in this study we characterized two novel *P1-rw* alleles, *P1-rw1077* and *P1-rw751::Ac*, which exhibit little or no red pigmentation in cob glumes. Structural analysis of *P1-rw1077* showed that this natural allele was generated by recombination between *p1* and the tightly-linked paralogous gene, *p2* (Zhang et al., 2000). Further comparison of sequence and expression properties of the *P1-rw* and *P1-rr* alleles points to the existence of a cob glume-specific regulatory sequence in the distal enhancer region. The absence of this sequence in the *P1-rw1077* allele, or insertion of an *Ac* transposable element in this sequence in the *P1-rw751::Ac* alleles, results in dramatic reduction or complete loss of pigmentation in cob glumes. These results show how alterations in regulatory sequences can generate distinct

patterns of expression of paralogous coding sequences, leading to a high degree of phenotypic diversity.

Results

Pigment accumulation pattern of *P1-rw1077* differs from that of *P1-rr4B2* both spatially and temporally

The standard *P1-rr4B2* allele specifies red pigmented kernel pericarp and cob glumes, whereas the *P1-rw1077* allele specifies red pigments in pericarp and colorless cob glumes. Both alleles are dominant to a *p1-ww* (white pericarp/white cob) allele (Anderson, 1924). However, in contrast to the relatively even pigmentation of *P1-rr4B2* pericarp, the pigmentation of *P1-rw1077* pericarp is concentrated on the gown region (sides of the kernel), while the crown region (top of kernel) is nearly colorless (Figure 1A). The differential spatial distribution of red color specified by *P1-rw1077* is associated with a delay in the onset of pigmentation compared with that observed for *P1-rr4B2*. On *P1-rr4B2* kernels, red pigmentation begins at approximately 10 ~ 12 DAP (days after pollination), and is first seen at the silk attachment point. Subsequently, the pigment spreads over the crown region and progresses through the gown region towards the base of the kernel. At 20 ~ 24 DAP, most of the pericarp is red. During maturation and dessication, the red color becomes darker (Zhang, 1999). In contrast, on *P1-rw1077* kernels, red color does not appear until 16 DAP, and the pigmentation begins in the gown region and does not spread to the crown (Figure 1B). Similar to *P1-rr4B2*, the pericarp color of *P1-rw1077* becomes darker during kernel maturation, although the color intensity is much less than that of *P1-rr4B2* (Figure 1A, B).

Expression profile of *P1-rw1077* is correlated with pigment accumulation patterns

Previous studies showed that the steady state levels of *P1-rr4B2* transcripts are developmentally regulated, and the timing of their accumulation is correlated with the formation of visible pericarp pigmentation (Sidorenko et al., 2000). As described above, *P1-rw1077* exhibits pronounced differences in the onset and intensity of phlobaphene pigmentation relative to *P1-rr4B2*. To examine whether or not the accumulation of *p1* transcripts is correlated with pigment accumulation in *P1-rw1077*, RNA gel blot analysis was performed on kernel pericarp total RNAs from both alleles at several developmental stages following pollination. The blot was sequentially hybridized with probes from the maize *p1*, *a1* (which is regulated by *p1*) and *actin* cDNAs (Figure 2A). The hybridization signals were normalized to *actin* transcript levels. In *P1-rr4B2*, the *p1* cDNA detected 1.8 kb and 1.0 kb transcripts, which arise from alternative splicing (Grotewold et al., 1991); whereas, in *P1-rw1077*, only a 1.4 kb band was detected (Figure 2A). Consistent with previous results (Sidorenko et al., 2000), accumulation of *P1-rr4B2* transcripts starts at a very early stage of ear development as traces of *p1* transcripts can be detected in the 0 DAP sample (Figure 2A, B). The transcript levels of *P1-rr4B2* peak at 12 DAP and 16 DAP, and start to decrease at 21 DAP. In contrast, *P1-rw1077* transcript levels are barely detectable at early developmental stages (0 DAP to 6 DAP), weakly present at 8 DAP, and peak at 16 DAP. In both lines, the expression pattern of the *a1* gene, which is directly activated by *p1* (Grotewold et al., 1994), temporally follows the onset of *p1* transcript accumulation (Figure 2A, B). Thus, the *P1-rw1077* allele exhibits a delay in *p1* transcript accumulation and a lower peak level of *p1* transcripts. This RNA expression pattern correlates with the pattern of red pigment accumulation observed in *P1-rw1077* kernel pericarp.

The *P1-rw1077* allele has completely colorless cob glumes. To examine whether the loss of cob glume pigmentation results from the absence of *p1* transcripts, we used RT-PCR to amplify transcripts in kernel pericarp and cob glume of both *P1-rw1077* and *P1-rr4B2* at 16 DAP (Figure 3). RNA samples from these tissues were reverse transcribed, and the resulting first strand cDNAs were subject to nested PCR amplification (see Material and Methods). The expected 570 bp PCR products from *p1* transcripts were detected from pericarp of *P1-rr4B2* and *P1-rw1077* as well as cob glumes of *P1-rr4B2*, whereas no *p1* transcript was detected from *P1-rw1077* cob glumes (Figure 3). This result was further confirmed by blotting the agarose gel of RT-PCR products and hybridizing to a *p*-specific probe (Figure 3).

P1-rw1077 transcripts contain both *p1* and *p2* sequences

As the RNA gel bolt analysis showed, the size of *P1-rw1077* transcripts (~1.4 kb) is smaller than that of the major *P1-rr4B2* transcripts (~1.8 kb). To identify the differences between *P1-rw1077* and *P1-rr4B2* transcripts, RT-PCR was performed to isolate *p1* transcripts from *P1-rw1077* pericarp. Because a GC-rich region in *p1* exon 3 prevents the amplification of full length *p1* transcripts, we isolated 5' and 3' regions of *P1-rw1077* transcripts separately (see Material and Methods). Sequence analyses showed that the 5' region of *P1-rw1077* transcripts (380 bp in length; extending from the 5' UTR into exon 3; see Materials and Methods) is nearly identical to that of *P1-rr4B2* except for two single nucleotide changes in exon 1 and 3 nucleotide changes in exon 3. Whereas the 3' end (320 bp in length) is identical to that of the *p2* gene, a tightly linked paralog of *p1* (Zhang et al, 2000). Similar to the *p2* gene in the *P1-rr4B2* stock, the *p2* gene in *P1-rw1077* contains an

80 bp deletion in the 5' UTR region (see below). The RT-PCR results shown in Figure 3 indicate that, like in *P1-rr4B2*, the *p2* gene in the *P1-rw1077* stock does not express in pericarp and cob glumes (otherwise a *p2* band, 80 bp-smaller than the *p1* band, should be present in the RT-PCR analysis. Zhang et al, 2000). Thus, the 3' RT-PCR product cannot be derived from the *p2* gene. These results indicate that *P1-rw1077* transcripts resemble *p1* in the 5' UTR and *p2* in exon 3. The chimeric structure of the *P1-rw1077* transcripts was subsequently corroborated by genomic DNA sequence analysis (see below).

***P1-rw1077* is a single copy gene generated by paralogous gene recombination**

Genomic DNA gel blot analyses were used to compare the structures of *P1-rw1077* and *P1-rr4B2*, *P1-wr*, and *p1-ww1112;* this latter haplotype has a deletion of *p1,* but retains the linked *p2* gene (Athma and Peterson, 1991). Genomic DNA was digested with *Xba*I and hybridized with genomic fragment 8B; this probe is derived from the second intron of *P1-rr4B2*, and it also cross-hybridizes with the *p2* gene. The *P1-rw1077* and *P1-rr4B2* alleles have similar hybridization patterns, i.e. both have a 3.7-kb band corresponding to the *p1* gene and a 6-kb band derived from *p2* (Figure 4A). The *p1-ww1112* allele lacks the 3.7 kb *p1* band, but has the 6 kb band corresponding to the *p2* gene. In contrast, the *P1-wr* allele has a 6-kb band (*p2*) as well as an intense 3.7-kb band derived from the tandem-repeat structure of *P1-wr* (Chopra et al., 1998). These results show that *P1-rw1077* has a single *p1* sequence similar to *P1-rr4B2*. However, additional DNA gel blot analyses revealed a number of polymorphisms between *P1-rw1077* and *P1-rr4B2*. For example, genomic DNAs from *P1-rw1077* and *P1-rr4B2* were digested by *Eco*RI and *Sal*I, and hybridized with a *p1*-specific genomic fragment (Fragment 15) as a probe, which is repeated 4 times at sites upstream and

downstream of the *P1-rr4B2* coding sequence (Figure 5). The different hybridization patterns indicated that several distinct polymorphisms exist between *P1-rw1077* and *P1-rr4B2* (Figure 4B).

We isolated genomic sequences of *P1-rw1077* by screening a genomic *P1-rw1077* λ library with *p1* genomic probes 8B and 15 (Figure 5). Two classes of λ clones were isolated. The first class hybridizes with probe 8B but not probe 15, and also lacks an 80 bp sequence present in the 5' UTR of *P1-rr4B2* and *P1-wr* (Zhang et al., 2000); partial sequence analysis of these clones indicated that they were derived from the *p2* gene (Zhang et al., 2003). The second class hybridizes with both probes 15 and 8B, has the 80bp sequence in the 5' UTR sequence, and thus appears to contain the *p1* gene. Two overlapping *p1*-carrying λ clones (covering 22270 bp sequences) were sequenced (Figure 5). The general structure of the *P1-rw1077* allele is similar to that of *P1-rr4B2*, e.g. it has a single coding sequence flanked by two long (6.3 kbp) direct repeats. However, the *P1-rw1077* coding sequence is chimeric, consisting of a *p1*-like 5' UTR, followed by *p2*-like exons and introns. Further downstream, the *P1-rw1077* allele has a region similar to the *Ji* retroelement fused with a truncated *p1*-like exon 3 (Figure 5). A 6915 bp region of *P1-rw1077* extending from the 5' UTR to the end of the *Ji*-homologous sequence is 99.5% similar to that of the *p2* gene, with only 11 1-bp mismatches and 8 small gaps. Most of these nucleotide changes also differ from the corresponding *P1-rr4B2* sequences. Thus, the *P1-rw1077* allele appears to have originated by recombination between *p1* and *p2* genes.

By comparing the sequences of the *P1-rw1077*, *P1-rr4B2*, *P1-wr* alleles and the *p2* gene, the postions of the recombination break points can be inferred. The 5' break point is located in the first exon, in the interval between position -8 and position +60 of *P1-rw1077*

(relative to the first base of the start codon as position +1; Figure 6A). The 3' break point is located at position +6916 of *P1-rw1077*, joining the *Ji*-1 retrotransposon sequence from *p2* to the 3' region of *p1* exon 3 at a 4-bp micro-homology sequence (CGCC) (Figure 6B). In summary, the *P1-rw1077* allele contains a 6.9 kb segment of the *p2* coding sequence that replaces most of the *p1* coding sequence. The resulting gene is expected to generate transcripts with *p1*-like sequences in the 5' UTR, and *p2*-like sequences in the remaining coding region. The data from RNA gel blot analysis, RT-PCR, and sequence analysis of *P1-rw1077* cDNAs indicate that the *P1-rw1077* produces a single mature transcript that includes exons 1, 2 and 3 as shown in Figure 5. There was no evidence for alternative splicing products as described for *P1-rr4B2* (Figure 2; Grotewold et al., 1991), nor for the inclusion of any portion of the downstream *Ji* sequences or the truncated exon 3 within the *P1-rw1077* transcript. The *P1-rw1077* cDNA sequence (1274 nt) was deduced and translated. Amino acid sequence comparison between *P1-rw1077* and *p2* showed that the *P1-rw1077* protein is nearly identical to the *p2* protein except for a single amino acid change at position eleven (Zhang et al, 2000).

A 386 bp sequence is absent from the distal enhancer region of *P1-rw1077*

Both *P1-rr4B2* and *P1-rw1077* have long direct repeats flanking the coding sequences. The upstream long direct repeat in *P1-rr4B2* has previously been shown to contain sequences important for *p1* expression, including a basal promoter region, a 1.0 kb proximal enhancer region and a 1.2 kb distal enhancer region located 5 kb upstream of the transcription start site (Sidorenko et al 1999; 2000). To examine the polymorphisms between the promoter regions of *P1-rr4B2* and *P1-rw1077*, we compared the nucleotide sequences

from ~10-kb upstream of the start codons of both alleles. The results showed 20 single-nucleotide and 16 small indel (insertion/deletion) polymorphisms between *P1-rr4B2* and *P1-rw1077*; however, the most striking difference in this region is located in the 1.2 kb distal enhancer region. In *P1-rr4B2*, there are two copies of Fragment 15 sequences flanking Fragment 14 in direct orientation (Figure 5). The upstream Fragment 15 in *P1-rr4B2* is interrupted near its midpoint by a 1.6 kb transposon-like sequence, whereas the *P1-rw1077* allele lacks this element (Figure 5). In addition, the *P1-rw1077* allele lacks a 386 bp sequence in this region, and instead contains a 54 bp sequence with some weak similarity to the corresponding sequence of *P1-rr4B2* (Figure 7). Because this region of *p1* has been shown to be important for *P1-rr4B2* expression (Sidorenko et al., 1999; Sidorenko et al., 2000; Sidorenko and Peterson, 2001), we were interested to determine its possible role in *P1-rw* expression.

_Ac_ insertion in the 386 bp sequence converts RR to RW-like phenotype

In previous studies, *Ac* transposable elements have been used to characterize regulatory elements important for *P1-rr4B2* expression (Sidorenko et al., 2000). A number of alleles with *Ac* transposable element insertions in the *P1-rr4B2* sequence were isolated on the basis of the distinct pigmentation patterns they specified. One of those alleles, designated *P1-ovov1114*, contains an *Ac* transposon in the second intron in the opposite orientation as *p1* transcripts, resulting in variegated orange phenotype in both pericarp and cob glumes (Peterson, 1990). Transposition of the *Ac* element in *P1-ovov1114* generated a series of new *p1* alleles (Athma et al., 1992), one of which showed a novel RW-like phenotype with deep orange pericarp and colorless cob phenotype (Figure 8A). Southern analysis of DNA from

P1-rr4B2 and the novel *P1-rw* plants showed that there are two *Ac* elements associated with this *P1-rw* allele. As can be seen in Figure 8B (Lane 4 and 5), DNA from *P1-rr4B2* plants digested with *Sal*I and hybridized with Fragment 15 produces four bands at 3.4 kb, 3 kb and 1.2 kb (doublet), whereas, in *P1-rw* plants, the 3.4 kb and 3 kb bands shift to 7.9 kb and 7.5 kb, respectively. The size of *Ac* elements is ~4.5 kb; and there is no *Sal*I site in *Ac* elements. Thus, this result indicated that, in *P1-rw*, two *Ac* transposons are inserted in the *P1-rr4B2* sequence: one in the 3.4 kb *Sal*I fragment, and a second in the 3 kb *Sal*I fragment. Additional DNA gel blots and PCR analyses showed that the *Ac* element in the 3.4 kb *Sal*I fragment is at the same location and orientation as the donor *Ac* element in the *P1-ovov1114* allele (see Materials and Methods for details)

The approximate position of *Ac* in the 3.0 kb *Sal*I fragment was estimated from the *Eco*RI digestion pattern (Figure 8B). The precise insertion site and orientation of the *Ac* element in the 3 kb *Sal*I fragment of the new *P1-rw* allele (*P1-rw512A::2Ac*) was determined by genomic PCR using primers from the deduced insertion region and from 5' and 3' *Ac* sequences (see Material and Methods for details). As shown in Figure 9, *Ac* is inserted in the same orientation as the direction of *p1* transcripts. The insertion is located 78 bp from the 3' end of the 3.0 kb *Sal*I fragment. Notably, the *Ac* element is inserted within the 386 bp sequence that is absent from *P1-rw1077* (Figure 7).

Does the RW-like phenotype of *P1-rw512A::2Ac* result from the *Ac* insertion in the 386 bp region, or alternatively from effects imposed by insertion of 2 *Ac* elements? An example of this latter effect was described by English et al (1998), who reported that two *Ac* elements in the direct repeat configuration could silence the expression of a streptomycin resistance gene located between them. To distinguish these two possibilities, we used a

genomic PCR screen to identify derivative alleles in which the *Ac* element in the second intron excised (see Material and Methods). Among 100 plants screened, we identified two plants, *P1-rw751::Ac4* and *P1-rw751::Ac5*, that lack *Ac* insertions in the second intron (data not shown). This result was further confirmed by Southern analysis: in a *Sal*I digestion hybridized with fragment 15, the 7.9 kb band in *P1-rw512A::2Ac* is replaced by 3.4 kb bands in *P1-rw751::Ac4* and *P1-rw751::Ac5*, while the 7.5 kb bands remain unchanged. In addition, the *P1-rw751::Ac4* and *P1-rw751::Ac5* alleles exhibit a ~25 kb *Eco*RI band rather than the 14.7 kb doublet bands in *P1-rw512A::2Ac*, which arise by cutting at the *Eco*RI sites within *Ac* elements (Figure 8B; Figure 9).

In kernel pericarp, the *P1-rw751::Ac4* and *P1-rw751::Ac5* alleles condition uniformly darker red pigmentation relative to *P1-ovov1114* and *P1-rw512A::2Ac* (Figure 8A). In contrast, both alleles exhibit dramatic reduction in cob glume pigmentation: the cob of *P1-rw751::Ac4* is nearly colorless, while the cob of *P1-rw751::Ac5* has a light orange color (Figure 8A). Germinal excision of the *Ac* element from the 386 bp region restores red pigmentation in cob glumes (Zhang, F. unpublished data). Therefore, insertion of a single *Ac* element in the 386 bp sequence of the distal enhancer region reduces cob glume pigmentation, and excision of this element restores pigmentation.

It is unclear why *P1-rw751::Ac4* and *P1-rw751::Ac5* have different cob color intensity, although they contain identical *p1* gene structures. It has been reported that transposon insertions can affect the expression of nearby genes by epigenetic mechanisms (Barkan and Martienssen, 1991; Girard and Freeling, 1999; Lippman et al., 2004). Possibly, in the *P1-rw751::Ac* alleles, *Ac* insertion in the 386 bp region can impose variable epigenetic effects on the adjacent regulatory elements, leading to different levels of *p1* expression in cob

glumes. Further analysis will be required to determine the stability of these different expression states.

Discussion

P1-rw1077 was generated by recombination between the *p1* and *p2* genes

Alleles of the maize *p1* gene exhibit a number of diverse, tissue-specific pigmentation patterns. In this study, two independent *P1-rw* alleles, specifying red pericarp and white cob glumes, were characterized and compared to the previously described *P1-rr* and *P1-wr* alleles (Lechelt et al., 1989; Chopra et al., 1996; Chopra et al., 1998; Sidorenko et al., 2000). Analyses of the *P1-rw1077* allele indicated that this natural allele was apparently generated by recombination between *p1* and its paralog *p2*. The *p1* and *p2* genes were formed by recent gene duplication, and arranged as a tandem gene cluster (Zhang et al., 2000.) Like many duplicate genes, *p1* and *p2* exhibit differential expression patterns: the *p2* gene is expressed primarily in silk and anther wall, while the *p1* gene is expressed mainly in kernel pericarp, cob glumes, husks and silk (Zhang et al., 2000). The *p1* and *p2* coding sequences are very similar, but their flanking regulatory sequences differ. This has led to the proposal that *p1* and *p2* encode functionally equivalent proteins whose expression patterns differ due to the distinct regulatory sequences flanking each coding sequence (Zhang et al., 2000, 2003).

Closely-linked paralogous loci have been reported to be good substrates for DNA recombination including deletion, duplication and gene conversion (for a review of paralogous recombination in plants, see Lichtenstein, 1994). Some of the best examples of recombination between paralogous genes have been reported in studies of the maize *r1* gene,

which regulates kernel aleurone pigmentation. The *R1-stippled* allele consists of several duplicated *r1* genes; unequal crossover events among *r1* components can change the number of coding segments, as well as recombine distinct coding and regulatory regions (Eggleston et al., 1995). Here, we show that the *P1-rw1077* allele was formed by a paralogous gene conversion event that generated the chimeric structure (*p1-p2-p1*): i.e., the *p1* coding sequence was replaced by the *p2* coding sequence, while the flanking *p1* regulatory sequences were preserved. The net effect is that the recombinant *P1-rw1077* gene resembles a *p2* coding sequence controlled by *p1* regulatory sequences. The resulting red color in the *P1-rw1077* pericarp supports the hypothesis that the *p1* and *p2* genes encode proteins that are functionally equivalent as regulatory factors in the phlobaphene biosynthesis pathway (Zhang et al., 2003). It is not clear whether the *p1/p2* gene conversion event was interchromosomal or intrachromosomal. By either mechanism, recombination between the paralogous *p1* and *p2* genes may have contributed to the high degree of genetic diversity observed among *p1* alleles, including phenotypic variation (Brink and Styles, 1966) and structural polymorphism (Zhang et al., 2000; Cocciolone et al., 2001).

Most reported examples of paralogous gene recombination contain homologous sequences at both junctions (Liao, 2000; Matzkin and Eanes, 2003; Jelesko et al., 2004); in contrast, the *P1-rw1077* allele possesses a homologous junction at the 5' end and a nonhomologous junction at the 3' end, with a 4 base pair micro-homology region. This interesting structure can be interpreted as arising from the one-sided invasion (OSI) mechanism for the repair of double-strand breaks (DSBs) in both animals and plants (Belmaaza and Chartrand, 1994; Puchta et al., 1996). According to the OSI model, after the occurrence of a DSB in the recipient sequences and generation of free 3' ends by

exonucleolytic degradation, the 3' end from one side of the DSB invades the homologous region in the donor sequence, and serves as a primer for DNA synthesis. The newly synthesized strand could be extended beyond the homologous region, released and then ligated to the non-invaded end by the Non-Homologous End-Joining (NHEJ) mechanism. This would lead to a recombinant product with a homologous junction at one end and a non-homologous junction at the other. Evidence for this mechanism has been observed following the experimental induction of DSBs (Puchta et al., 1996). To our knowledge, the *P1-rw1077* allele described here is the first reported case of an OSI-type of structure found as a natural allele of a functional plant gene.

The 386bp region in the distal enhancer may contain cob glume-specific regulatory elements

Most plant gene promoters characterized to date appear to be relatively compact (for review, see Singh, 1998 and references therein); notable exceptions include the promoters of the maize genes *r1* (Li et al. 2001); *b1* (Stam et al., 2002b); and *p1* (Sidorenko et al., 1999), all of which encode transcriptional regulators of flavonoid pigment biosynthesis. The maize *p1* gene promoter has been shown to contain an enhancer within a 1.2 kb sequence that is located approximately 5 kbp upstream of the transcription start site. This 1.2 kb region was initially identified by transposon mutagenesis of *P1-rr4B2*. Insertions of *Ac* transposable elements in this region lead to reduced *p1* gene expression in both pericarp and cob glumes of *P1-rr4B2* (Athma et al., 1992; Moreno et al., 1992). This region was later functionally demonstrated to have enhancer activity in transient and transgenic expression assays (Sidorenko et al., 1999; Sidorenko et al., 2000).

In this study, analysis of two independent *P1-rw* alleles provides strong evidence that a 386 bp sequence in a distal enhancer region contains elements that are specifically required for *p1* expression in cob glumes. The 386 bp sequence is contained within the repeat sequences flanking the *p1* gene, and thus is present 4 times in *P1-rr4B2* (2 copies upstream and 2 copies downstream of the coding sequence), and 2 times in *P1-rw1077* (1 copy upstream and 1 downstream) (Figure 5). Lack of one of the upstream copies of the 386 bp sequence in the *P1-rw1077* allele is correlated with the absence, in the cob glumes, of *P1-rw* transcripts and pigmentation. Moreover, in *P1-rr4B2*, an *Ac* element insertion in one of the upstream 386 bp sequences converts RR into a RW like phenotype, as shown in the *P1-rw512A::2Ac*, *P1-rw751::Ac4* and *P1-rw751::Ac5* alleles. In contrast, insertion of *Ac* elements into two flanking sites (2374 bp upstream or 914 bp downstream of the 386 bp sequence) did not affect cob glume pigmentation specifically (Sidorenko et al., 2000). So far as is known, alleles carrying *Ac* insertions in the vicinity of the 386 bp do not exhibit properties of suppressible alleles as have been reported for certain other maize transposon insertions (Masson et al., 1987; Barkan and Martienssen, 1991; reviewed in Girard and Freeling, 1999).

Comparisons of *P1-rw1077* and *P1-rr4B2* indicate that these alleles have certain differences in the structures of the repeated sequences downstream of their respective coding sequences. These changes can provide clues to the molecular evolution of *p1* alleles and will be described in detail elsewhere (Zhang et al., in preparation). We cannot exclude the possibility that these structural differences may affect the expression of *P1-rw1077*. However, our results showing that alterations of the 5' 386 bp sequence are sufficient to

confer a P1-rw phenotype strongly support the hypothesis that this sequence contains elements required for *p1* expression in cob glumes.

Homology searches using BLAST indicate that the 386 bp sequence does not have extensive similarity with sequences from other maize loci, nor from other plants (not shown). Searches of the PlantCare promoter database (PlantCare, Lescot et al., 2002; web site: http://intra.psb.ugent.be:8080/PlantCARE/index.html), identified three ACGT motifs and one putative RY element within the 386 bp sequence (Figure 8). Similar ACGT and RY motifs have been shown to be important for expression of seed-specific genes, such as maize *c1*, wheat Em-1, bean β-phaseolin, and Arabidopsis napin (Vasil et al., 1995; Kao et al., 1996; Ezcurra et al., 1999; Chandrasekharan et al., 2003). The maize VP1 protein and its homologs FUS3 and ABI3 in Arabidopsis specifically bind RY elements, while many bZIP proteins bind ACGT motifs (Mikami et al., 1994; Suzuki M, 1997; Reidt et al., 2000; Siberil et al., 2001). VP1 is reported to interact and enhance the DNA binding activity of certain bZIP proteins including TRAB1, EmBP1 and o2 (Hill et al., 1996; Hobo et al., 1999), and thereby activate target gene expression. Interestingly, the *Ac* transposon in the *P1-rw751::Ac* alleles is inserted 4 bp downstream of the ACGT motif I (Figure 7) where it may interfere with protein-DNA interactions. Further analysis will be required to determine whether VP1 homologs and/or bZIP proteins are involved in the regulation of *p1* expression.

Previous studies have indicated that certain *p1* expression patterns could be attributed to epigenetic regulation rather than DNA sequence polymorphism (Chopra et al., 1999; Cocciolone et al., 2001; Chopra et al., 2003). The results we present here show that diversity in *p1* expression could also arise from DNA sequence changes in a flanking enhancer region. Taken together, these studies indicate that the combined effects of variation in tissue-specific

regulatory elements and epigenetic controls can give rise to the wide range of spatial and temporal phenotypic diversity observed in *p1* alleles. Interestingly, the distal enhancer region of *P1-rr* described here has been shown to induce *p1* paramutation, an allelic interaction leading to epigenetic silencing ((Sidorenko and Peterson, 2001). The colocalization of distal enhancer sequences with sequences required for paramutation has been recently reported at the maize *b1* (*booster 1*) gene, which is also a regulatory gene in flavonoid biosynthetic pathway (Stam et al., 2002a; Stam et al., 2002b). The regulatory sequences that are required for both enhancer activity and paramutablity are located ~100 kb upstream of *b1* transcription start site. An intriguing question regarding these observations is whether sequences for enhancer function are separable from sequences for paramutation. The dissection of the distal enhancer region in *p1* by using *P1-rw* alleles will allow us to address this question and shed further light on the mechanisms of long-distance cis and trans gene regulation.

Material and Methods

Maize Genetic stocks

The *P1-rr4B2* allele (Grotewold et al., 1991) was in inbred line 4Co63 background, and the *P1-wr* allele was from inbred line W23 (Chopra et al., 1996; Chopra et al., 1998). The *P1-rw1077* allele in 4Co63 background was obtained from the Maize Genetics Cooperation Stock center, Urbana, Illinois. The inbred line 4Co63 (*p1-ww*) was obtained from the National Seed Storage Laboratory, Fort Collins, CO, USA. The allelism of *P1-rw1077* and *P1-wr* was tested by crossing *P1-wr* with *P1-rw1077*. The F1 plants (*P1-rr* like phenotype) were subsequently crossed to *p1-ww*, 4Co63. Among 74 progeny plants, 40

plants were P1-wr phenotype, and 34 were P1-rw phenotype. This ~1:1 ratio, and the absence of P1-rr and P1-ww phenotypes, indicate that *P1-wr* and *P1-rw1077* segregate as alleles in repulsion.

Genomic library construction and screening

Genomic DNA was extracted from silk of *P1-rw1077* homozygous plants, partially digested with *Sau*3AI, and ligated to the λ FIX II/*Xho* I partial fill-in vector (Stratagene, La Jolla, CA). Approximately 1.5 x 10^6 independent plaques were screened by using two probes of *P1-rr4B2*, Fragment 8B and Fragment 15 (Figure 5). The Fragment 8B probe detects a homologous sequence in the *p2* gene, whereas the Fragment 15 probe is unique to the *p1* gene. Seven clones hybridized only with Fragment 8B but not Fragment 15; while four clones hybridized with both probes, indicating that they contain *p1* sequences. PCR using primers EP5-8 (5'-ACGCGCGACCAGCTGCTAACCGTG-3'; homologous to the 5' untranslated region of *p1*) and EP 3-13 (5'-AGGAATTCCGCCCGAAGG TAGTTGATCC-3'; homologous to *p1* exon 2) showed that the 7 clones which did not hybridize with Fragment 15 resemble the *p2* gene, which does not contain an 80 bp insertion in the 5' UTR region (Figure 6). The 4 *p1*-like λ clones were subject to further analyses, including restriction digest and sequencing of ends. Two overlapping clones, which together cover 22.2 kb region of *p1* locus, were selected, subcloned into pBluescript plasmid vectors, and sequenced with the EZ:TN <KAN-2> Insertion Kit (Epicenter Technologies, Madison, WI) following the manufacturer's instructions at the Iowa State University Nucleic Acid Facility.

DNA and RNA gel blots

Genomic DNA preparation and DNA gel blot analyses were conducted as described by Sidorenko et al (2000). The kernel pericarp and cob glumes were dissected at different developmental stages as described by Sidorenko et al (2000). Total RNAs were extracted from frozen samples using the RNeasy plant mini kit (Qiagen, Valencia, CA), and treated with DNase (Qiagen) to remove residual genomic DNA. RNA gel blot analysis was done as described previously (Chopra et al., 1996). The *p1* probe was amplified from the plasmid containing *P1-rr4B2* cDNA with primer EP5-17 (5'-GGGAGGACGCCGTGCTGC-3'; homologous to *p1* exon 1) and EP3-10 (5'-CTGTCGGCCTCCCCCCAGACTAGG-3', homologous to *p1* exon 3). The probes of maize *a1* cDNA and *actin* cDNA were obtained as described previously (Grotewold et al., 1991). RNA gel blot hybridization signals were quantified by ImageQuant software (Molecular Dynamics, Sunnyvale, CA). Transcript levels of *p1* and *a1* were normalized by dividing by the level of *actin* transcripts.

Reverse Transcription-Polymerase Chain Reaction

To compare levels of steady-state *p1* transcripts in *P1-rr4B2* and *P1-rw1077*, 1 μg total RNA from 16 DAP pericarp and cob glumes of *P1-rr4B2* and *P1-rw1077* were reverse transcribed using StrataScript Reverse Transcriptase (Stratagene, La Jolla, CA) with oligo(dT) at 42°C. Because *p1* expression in cob glumes is very low, nested PCR was used to amplify first strand cDNA. In the 1st round PCR amplification, 5 μl cDNA was used with primers EP5-8 and P1-23568 (5'-GCAGCTTGCTCATGTCGATGGC-3'); while, in 2nd round PCR, the 0.1 μl PCR product from 1st amplification was used with primers EP5-8-1 (5'-GCTGCTAACCGTGCGCAA

GTAG -3') and ZFRT-8 (5'- CAGGGACCACCTGTTG CCGAG-3'; spanning the junction of exon2 and exon3 of *p1* to avoid amplification from genomic DNA). PCR amplification was performed in the same conditions as above, except that the annealing temperature was 50°C. The PCR products were subject to 1.5% agarose gel electrophoresis. The same amount of cDNA used in detection of *p1* transcripts was amplified with *tubulin* primers as positive control. The sequences of *tubulin* primers are: tub-1, 5'-AGCCCGATGGCACCATGCCCAGTGATACCT-3'; and tub-2, 5'-AACACCA AGAATCCCT GCAGCCCAGTGC-3' (Danilevskaya et al., 2003).

To isolate 5' regions of *P1-rw1077* transcripts, 1 μg of total RNA isolated from 16 DAP *P1-rw1077* pericarp tissues was reverse transcribed using StrataScript Reverse Transcriptase (Stratagene, La Jolla, CA) with oligo(dT) at 42°C. 5μl from the first strand cDNAs was used in a PCR reaction with primers EP5-8 and EP3-13. To obtain 3' ends of *P1-rw1077* transcripts, the 3' adapter primer (5'-GGCCACGCGTCGACTAGTAC$(T)_{17}$-3') was used to perform first strand cDNA synthesis. 5μl of reverse transcribed product was used in PCR reaction with primers EP5-16 (5'-GACGATCGCGAGCTGG-3'; homologous to *p1* and *p2* exon 3) and AP (5'- GGCCACGCGTCGACTAGTAC-3'). PCR amplifications were performed using HotStart Taq DNA polymerase (Qiagen, Valencia, CA) with the following cycle conditions: 95°C for 15 min, followed by 35 cycles of 94°C for 45 sec, 58°C for 1 min, and 72°C for 1 min, and a final extension for 8 min at 72°C. The PCR products from both 5' end and 3' end of *P1-rw1077* were cloned into pGEM-T vector (Promega, Madison, WI). 3 independent clones from 5' PCR products and 2 independent clones from 3' PCR products were sequenced with the same primer sets used in the RT-PCR at Iowa State University Nuclei Acid Facility.

Mapping of *Ac* transposable elements in the *P1-rw512A::2Ac* allele and sequencing of insertion site

The positions of *Ac* transposable elements in the *P1-rw512A::2Ac* allele were determined by DNA gel blot analysis of genomic leaf DNA as described previously (Athma et al., 1992). Genomic DNA was digested with *Sal*I and *Eco*RI respectively, and hybridized with Fragment 15. Comparing the hybridization patterns of the *P1-rw512A::2Ac*, *P1-rr4B2* and *P1-ovov1114* alleles allowed the approximate locations of *Ac* elements to be mapped (Figure 8 and 9).

The orientations and insertion sites of *Ac* elements in *P1-rw512A::2Ac* were determined by PCR using primers homologous to *Ac* and flanking *p1* genomic sequences. The *p1* primers used were ZFPrr-4 (5'-ATGTGTCATTGCCTCGTTGG-3'), EP3-7 (5'-CACGCACCTAAAG CAGAAGCGAAC-3'), PA-A13 (5'-TTGTGGATCCGGCCCCTG-3') and PP1' (5'-GACCG TGACCTGTCCGCTC-3'). Primers homologous to *Ac* sequences were iAc3-2 (5'-TTATCCC GTTCGTTTTCGTTACC-3') and Ac123 (5'-ATCCCGTTTCCGTTCCGTTTTC-3'). Locations of these primers are shown in Figure 9. PCR amplifications were performed using HotStart Taq DNA polymerase (Qiagen, Valencia, CA) with the following cycle conditions: 95°C for 15 min, followed by 35 cycles of 94°C for 45 sec, 60°C for 1 min, and 72°C for 1 min, and a final extension for 8 min at 72°C. Sequences were determined at the Iowa State University Nuclei Acid Facility.

PCR screen for plants with *Ac* excised from the second intron

Homozygous *P1-rw512A::2Ac* plants were crossed with *p1-ww,* and 100 F1 seedlings were grown in groups of 10. Genomic DNA was prepared from pooled leaves (10 seedlings per pool) and used in PCR with primers PA-A13 and PP1'. Excision of *Ac* from the second intron of the *p1* gene will yield a 690 bp PCR product. Following identification of pooled DNA samples which show strong 690 bp PCR products, DNA of individual plants in those pools were extracted and used for PCR amplification with the Primer set: PA-A13 and PP1'. PCR using the primer set PA-A13 and iAc3-2, which amplifies across the junction of *Ac* and *p1* sequences, was also used to confirm *Ac* excision.

Acknowledgements

We thank Ruizhong Shen and Surinder Chopra for isolating 5' and 3' ends of *P1-rw1077* transcripts. We thank Erica Unger-Wallace, Lyudmilla Sidorenko, Erik Vollbrecht, Diane Bassham, Steve Whitham and Dan Voytas for advice and comments on the manuscript. This material is based upon work supported by the National Science Foundation under Grant No. 9601285. This journal paper of the Iowa Agriculture and Home Economics Experiment Station, Ames, Iowa, was supported by Hatch Act and State of Iowa funds.

References

Anderson, E.G. (1924). pericarp studies in maize II. the allelomorphism of a series of factors for pericarp color. Genetics 9, 442-453.

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. (1991). Inactivation of Maize Transposon Mu Suppresses a Mutant Phenotype by Activating an Outward-Reading Promoter Near the End of Mu1. PNAS 88, 3502-3506.

Belmaaza, A., and Chartrand, P. (1994). One-sided invasion events in homologous recombination at double-strand breaks. Mutat Res 314, 199-208.

Brinks, R., and Styles, D. (1966). A collection of pericarp factors. Maize Genet. Cooperation Newsletters 40, 149-160.

Chandrasekharan, M.B., Bishop, K.J., and Hall, T.C. (2003). Module-specific regulation of the beta-phaseolin promoter during embryogenesis. Plant J 33, 853-866.

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., Li, X.G., and Peterson, T. (1998). A maize Myb homolog is encoded by a 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, 15330-15335.

Chopra, S., Cocciolone, S.M., Bushman, S., Sangar, V., McMullen, M.D., and 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.

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, 467-478.

Coe, E.H., Neuffer, M.G., and Hoisingto, D.A. (1988). The genetics of corn. In: Sprague GF, Dubley J(eds) Corn and Corn improvement (American Society of Agronomy, Madison, WI), pp. 81-258.

Consonni, G., Geuna, F., Gavazzi, G., and Tonelli, C. (1993). Molecular homology among members of the R gene family in maize. Plant J 3, 335-346.

Danilevskaya, O.N., Hermon, P., Hantke, S., Muszynski, M.G., Kollipara, K., and Ananiev, E.V. (2003). Duplicated fie Genes in Maize: Expression Pattern and Imprinting Suggest Distinct Functions. Plant Cell 15, 425-438.

Edwards, J., Stoltzfus, D., and Peterson, P.A. (2001). The C1 locus in maize (Zea mays L.): effect on gene expression. Theoretical and Applied Genetics 103, 718 - 724.

Eggleston, W.B., Alleman, M., and Kermicle, J.L. (1995). Molecular Organization and Germinal Instability of R-stippled Maize. Genetics 141, 347-360.

Emerson, B.M. (2002). Specificity of gene regulation. Cell 109, 267-270.

English, J.J., and Jones, J.D. (1998). Epigenetic instability and trans-silencing interactions associated with an SPT::Ac T-DNA locus in tobacco. Genetics 148, 457-469.

Ezcurra, I., Ellerstrom, M., Wycliffe, P., Stalberg, K., and Rask, L. (1999). Interaction between composite elements in the napA promoter: both the B-box ABA-responsive

complex and the RY/G complex are necessary for seed-specific expression. Plant Mol Biol 40, 699-709.

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

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 U S A 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.

Jelesko, J.G., Carter, K., Thompson, W., Kinoshita, Y., and Gruissem, W. (2004). Meiotic recombination between paralogous RBCSB genes on sister chromatids of Arabidopsis thaliana. Genetics 166, 947-957.

Kao, C.Y., Cocciolone, S.M., Vasil, I.K., and McCarty, D.R. (1996). Localization and Interaction of the cis-Acting Elements for Abscisic Acid, VIVIPAROUS1, and Light Activation of the C1 Gene of Maize. Plant Cell 8, 1171-1179.

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, 225-234.

Lescot, M., Dehais, P., Thijs, G., Marchal, K., Moreau, Y., Van de Peer, Y., Rouze, P., and Rombauts, S. (2002). PlantCARE, a database of plant cis-acting regulatory elements and a portal tools for in silico analysis of promoter sequences. Nucleic Acids Res. 30, 325-327.

Liao, D. (2000). Gene Conversion Drives Within Genic Sequences: Concerted Evolution of Ribosomal RNA Genes in Bacteria and Archaea. Journal of Molecular Evolution 51, 305-317.

Lichtenstein, C.P., Paszkowski, J., Hohn, B. (1994). Intrachromosal recombination between genomic repeats. In Homologous recombination and gene silencing, J. Paszkowki, ed (Dordecht, the Netherlands: Kluwer Academic), pp. 95-122.

Lippman, Z., Gendrel, A.-V., Black, M., Vaughn, M.W., Dedhia, N., Richard McCombie, W., Lavine, K., Mittal, V., May, B., Kasschau, K.D., Carrington, J.C., Doerge, R.W., Colot, V., and Martienssen, R. (2004). Role of transposable elements in heterochromatin and epigenetic control. Nature 430, 471-476.

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, 117-137.

Matzkin, L., and Eanes, W. (2003). Sequence Variation of Alcohol Dehydrogenase (Adh) Paralogs in Cactophilic Drosophila. Genetics 163, 181-194.

Mikami, K., Sakamoto, A., and Iwabuchi, M. (1994). The HBP-1 family of wheat basic/leucine zipper proteins interacts with overlapping cis-acting hexamer motifs of plant histone genes. J Biol Chem 269, 9974-9985.

Moreno, M.A., Chen, J., Greenblatt, I., and Dellaporta, S.L. (1992). Reconstitutional mutagenesis of the maize P gene by short-range Ac transpositions. Genetics 131, 939-956.

Peterson, T. (1990). Intragenic transposition of Ac generates a new allele of the maize P gene. Genetics 126, 469-476.

Pilu, R., Piazza, P., Petroni, K., Ronchi, A., Martin, C., and Tonelli, C. (2003). pl-bol3, a complex allele of the anthocyanin regulatory pl1 locus that arose in a naturally occurring maize population. Plant J 36, 510-521.

Puchta, H., Dujon, B., and Hohn, B. (1996). Two different but related mechanisms are used in plants for the repair of genomic double-strand breaks by homologous recombination. Proc Natl Acad Sci U S A 93, 5055-5060.

Radicella, J., Brown, D., Tolar, L., and Chandler, V. (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 Dev. 6, 2152-2164.

Reidt, W., Wohlfarth, T., Ellerstrom, M., Czihal, A., Tewes, A., Ezcurra, I., Rask, L., and Baumlein, H. (2000). Gene regulation during late embryogenesis: the RY motif of maturation-specific gene promoters is a direct target of the FUS3 gene product. plant J 21, 401-408.

SanMiguel, P., Tikhonov, A., Jin, Y.-K., Motchoulskaia, N., Zakharov, D., Melake-Berhan, A., Springer, P.S., Edwards, K.J., Lee, M., Avramova, Z., and Bennetzen, J.L. (1996). Nested Retrotransposons in the Intergenic Regions of the Maize Genome. Science 274, 765-768.

Scheffler, B., Franken, P., Schutt, E., Schrell, A., Saedler, H., and Wienand, U. (1994). Molecular analysis of C1 alleles in Zea mays defines regions involved in the expression of this regulatory gene. Mol Gen Genet 242, 40-48.

Siberil, Y., Doireau, P., and Gantet, P. (2001). Plant bZIP G-box binding factors. . Modular structure and activation mechanisms. Eur J Biochem 268, 5655-5666.

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

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, 319-335.

Sidorenko, L.V., Li, X., Cocciolone, S.M., Chopra, S., Tagliani, L., 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.

Singh, K.B. (1998). Transcriptional Regulation in Plants: The Importance of Combinatorial Control. Plant Physiol. 118, 1111-1120.

Stam, M., Belele, C., Dorweiler, J.E., and Chandler, V.L. (2002a). Differential chromatin structure within a tandem array 100 kb upstream of the maize b1 locus is associated with paramutation. Genes Dev 16, 1906-1918.

Stam, M., Belele, C., Ramakrishna, W., Dorweiler, J.E., Bennetzen, J.L., and Chandler, V.L. (2002b). The regulatory regions required for B' paramutation and expression are located far upstream of the maize b1 transcribed sequences. Genetics 162, 917-930.

Suzuki M, K.C., McCarty DR. (1997). The conserved B3 domain of VIVIPAROUS1 has a cooperative DNA binding activity. Plant Cell 9, 799-807.

Vasil, V., Marcotte, W.R., Jr., Rosenkrans, L., Cocciolone, S.M., Vasil, I.K., Quatrano, R.S., and McCarty, D.R. (1995). Overlap of Viviparous1 (VP1) and abscisic acid response elements in the Em promoter: G-box elements are sufficient but not necessary for VP1 transactivation. Plant Cell 7, 1511-1518.

Zhang, P. (1999). Molecular characterization of myb-homologous transcriptional factors of the flavonoid pathway in Zea mays (Dissertation, Iowa State University).

Zhang, P., Chopra, S., and Peterson, T. (2000). A segmental gene duplication generated differentially expressed myb-homologous genes in maize. 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, 1-15.

Figure Legends

Figure 1. Distinct phlobaphene pigmentation patterns of different *p1* alleles

A. Mature ear pigmentation patterns specified by the *p1* alleles: *P1-rr4B2*, *P1-wr*, *P1-rw1077* and *p1-ww1112* (from left to right). All alleles are homozygous in 4Co63 background.

B. Ear pigmentation patterns of *P1-rr4B2* and *P1-rw1077* allele at 16 and 28 DAP (DAP, Days After Pollination).

Figure 2. Accumulation of *p1* transcripts in *P1-rw1077* and *P1-rr4B2* during development.

A. RNA gel blot analysis of *p1* transcripts of *P1-rw1077*, *P1-rr4B2* and *p1-ww1112* at different developmental stages Samples were taken from whole ears (0 DAP), developing kernels and cob (2-6 DAP), whole kernel (8 DAP) and kernel pericarp (12-21 DAP). 10 μg

total RNA was loaded in each lane. Hybridization probes are indicated at right. The *actin* cDNA probe was used as a loading control.

B. Quantified RNA levels of *p1* and *a1* from RNA gel blot in A are normalized to *actin* to calculate relative transcript levels (y-axis).

Figure 3. RT-PCR analyses of *p1* transcripts in kernel pericarp and cob glumes.

Nested RT-PCR was performed on total RNA extracted from pericarp and cob glumes of *P1-rr4B2* and *P1-rw1077* at 16 DAP. Upper panel shows ethidium bromide-stained PCR products (570 bp) amplified with *p1* specific primers (EP5-8-1 and ZFRT-8). Middle panel shows the results following blotting of the agarose gel and hybridization with *p*-specific probe. The lower panel shows the same samples amplified with α-*tubulin*-specific primers as positive controls.

Figure 4. Genomic Southern analyses of different *p1* alleles.

A. Genomic DNA of individual *p1* alleles were digested with *Xba*I and hybridized with *p1* Fragment 8B (Figure 5). The *p1-ww1112* (Lane 1) allele contains only the 6-kb *p2* band, while *P1-rw1077* (Lane 2), *P1-rr4B2* (Lane 3) and *P1-wr* (Lane 4)contain both the 3.7-kb *p1* bands and the 6-kb *p2* bands. In *P1-wr*, the high intensity *p1* band results from the tandem repeat of *p1* sequence (Chopra et al., 1998).

B. Genomic DNAs of *P1-rw1077* (lane 1) and *P1-rr4B2* (lane 2) were digested with *Sal*I and *Eco*RI, and probed with *p1* Fragment 15 (Figure 5). A number of RFLPs are apparent.

Figure 5. Gene structure comparisons of *P1-rr4B2*, *P-rw1077* and *p2*.

The open boxes with black arrows indicate long direct repeats in the flanking regions of *P1-rr4B2* and *P1-rw1077*. The black boxes connected by thin lines represent exon regions transcribed in *P1-rr4B2*, *P1-rw1077* and *p2* (in *P1-rr4B2*, only the major spliced product is shown). Bent arrow indicates transcription start site in *P1-rr4B2* (Grotewold et al., 1991). The hatched boxes represent 5' UTR and 90 bp promoter regions conserved among *p1* alleles and the *p2* gene (Zhang et al., 2000). Positions of 5' and 3' breakpoints in *P1-rw1077* chimeric structure are indicated by offset lines. Sequences sharing homology to Fragment 15 and 8B are indicated by the numbered open boxes. The boxes labeled *Ji* in both the *P1-rw1077* allele and the *p2* gene indicate the *Ji-1* retrotransposon sequence (SanMiguel et al., 1996), previously reported as *Prem-2* retrotransposon (Zhang et al., 2000). In *P1-rw1077*, the black box following the *Ji* sequence represents a truncated *p1* exon 3 region. The white box upstream of the *p2* promoter indicates the 500 bp sequence homologous with retroelement *Prem-2*. Triangles indicate the 80 bp, 734 bp and 1.6 kb insertions in *P1-rr4B2;* the 80 bp insertion is also present in *P1-rw1077* (Sidorenko et al., 2000; Zhang et al., 2000). The polymorphic distal enhancer regions in *P1-rr4B2* and *P1-rw1077* are located between the vertical dashed lines (from *Sac*I site to *Sal*I site). The 1.2 kb distal enhancer region in *P1-rr4B2* is delineated by the two *Sal*I sites. The regions marked 15* indicate the partial Fragment 15 sequences in *P1-rw1077*, which contains only the first 200 bp Fragment 15 sequences. The gray bar between the *Pst*I and *Sac*I sites in *P1-rr4B2* represents the 386 bp sequence, which is absent from the distal enhancer of *P1-rw1077*. The thin lines below *P1-rw1077* indicate two overlapping *P1-rw1077* genomic λ clones. Restriction sites shown are: E, *Eco*RI; S, *Sac*I; Sl, *Sal*I. Not all restriction sites are shown. The black arrow heads

indicate primers that were used to isolate 5' and 3' end of *P1-rw1077* transcripts: 1, EP5-8; 2, P1-23568; 3, EP5-16; 4, AP. Genbank accession number of *P1-rw1077* is AY702552.

Figure 6. Localization of *p1* and *p2* recombination break points in *P1-rw1077*.

A. The 5' region of *P1-rw1077* was aligned with *P1-rr4B2, P1-wr* and *p2*. The sequence in boldface indicates the exon 1 ORF, and the first base of the start codon is indicated as position +1. The triangle at position -8 in the 5' UTR represents an 80 bp sequence that is present in *p1* alleles but absent from *p2* (Zhang et al., 2000). The arrows at +60 and +138 indicate nucleotide polymorphisms in *P1-rw1077* which match the *p2* sequence. Thus, the 5' breakpoint of *p2-p1* crossover lies between the 80 bp insertion at -8 bp and the polymorphism at +60 bp.

B. The 3' region of *P1-rw1077* was aligned with *p1* and *p2* sequences. The exon 3 regions and *Ji*-homologous sequences in *p1* and *p2* are shown as in Figure 5. The 4-bp microhomology sequence CGCC at the breakpoint is underlined.

Figure 7. Alignment of the distal enhancer region in *P1-rr4B2* and *P1-rw1077*.

DNA sequences of *P1-rr4B2* and *P1-rw1077* were aligned at the distal enhancer region (between the two *Sac*I sites shown in Figure 5). Restriction sites are shaded; the sequence between the first *Sac*I site and the *SalI* site corresponds to Fragment 15, while the sequence after the second *Sac*I site corresponds to Fragment 14. The dashed line represents gaps in the alignment. The black bar above the *P1-rr4B2* sequence indicates the 386 bp region that is replaced by a 54 bp sequence in *P1-rw1077*. The gray arrowhead represents the 1.6 kb transposon-like sequence present in *P1-rr4B2*, and the black arrowhead represents the *Ac*

element in the *P1-rw512A::2Ac* and *P1-rw751::Ac* alleles. The 8-bp underlined sequence indicates the target site duplicated upon transposon insertion. The boxed sequences are three putative ACGT motifs, designated as I, II and III, and the putative RY element.

Figure 8. Phenotypes and DNA gel blot analysis of *P1-rw512A::2Ac* and *P1-rw751::Ac*.

A. Phenotypic comparisons of *P1-ovov1114*, *P1-rw512A::2Ac*, *P1-rw751::Ac4*, *P1-rw751::Ac5* and *P1-rr4B2* alleles (from left to right). All alleles are homozygous except for *P1-rw512A::2Ac* that is heterozygous with *p1-ww* [4Co63].

B. Genomic Southern blot analysis of *P1-rw751::Ac4* (lane 1), *p1-ww* [4Co63](lane 2), *P1-rw751::Ac5* (lane 3), *P1-rw512A::2Ac* (lane4), and *P1-rr4B2* (lane 5). DNA samples were digested with *Eco*RI and *Sal*I, respectively, and probed with *p1* Fragment 15. The *P1-rw751::Ac4*, *P1-rw751::Ac5* alleles are heterozygous with *p1-ww* [4Co63], while the *P1-rw512A::2Ac* allele is homozygous.

Figure 9. Gene structures of *P1-rr4B2*, *P1-rw512A::2Ac* and *P1-rw751::Ac* alleles.

The bent arrow indicates the *p1* transcription start site. The black boxes with connected lines represent the exon/intron structure of the major *P1-rr4B2* splicing product (Grotewold *et al.*, 1991). The open box indicates the 1.6 kb transposon-like sequence inserted into the furthest 5' copy of Fragment 15 in the *P1-rr4B2* allele (Figure 5). The triangles with arrows represent *Ac* transposable elements, and the black arrows indicate the orientations of *Ac* (from 5' to 3'). The gray boxes represent sequence homology to Fragment 15. The black arrow heads with numbers indicate primers that were used to determine *Ac* insertion sites: 1,

iAc3-2; 2, Ac-123; 3, ZFPrr-4; 4, EP3-7; 5, PA-A13; 6, PP1'. Restriction sites shown are: S, *Sal*I site; S*, methylated *Sal*I site; E, *Eco*RI site. The drawing is not to scale.

A

B

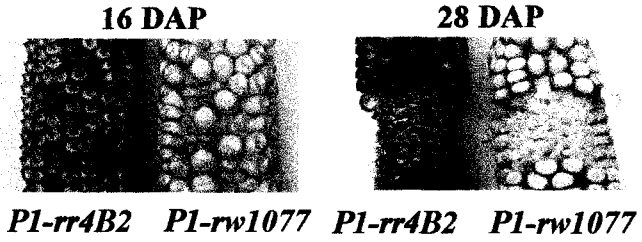
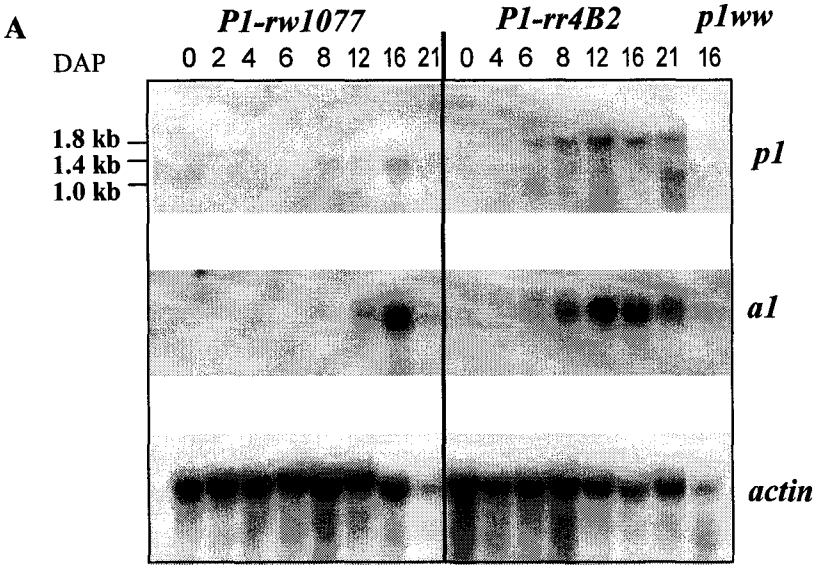


Figure 1


A
P1-rw1077
P1-rr4B2
p1ww
DAP
0 2 4 6 8 12 16 21
0 4 6 8 12 16 21
16
1.8 kb
1.4 kb
1.0 kb
p1
a1
actin

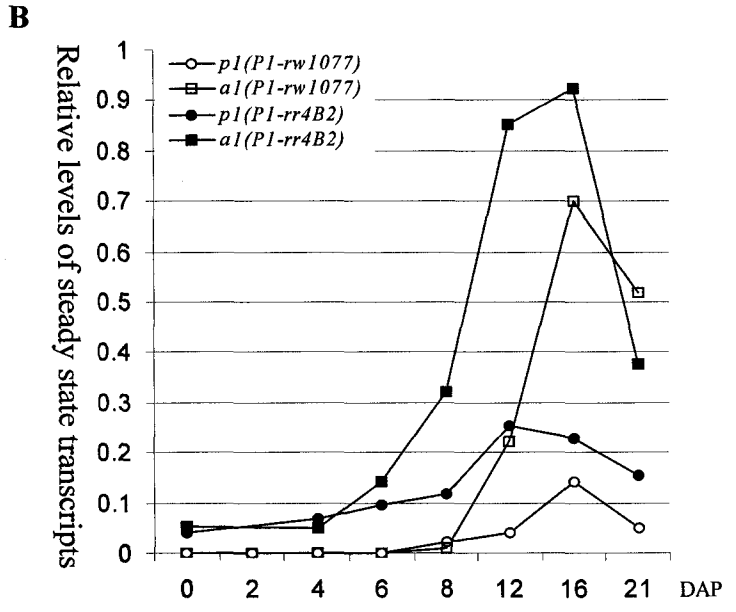

B
Relative levels of steady state transcripts
p1(P1-rw1077)
a1(P1-rw1077)
p1(P1-rr4B2)
a1(P1-rr4B2)
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0 2 4 6 8 12 16 21 DAP

Figure 2

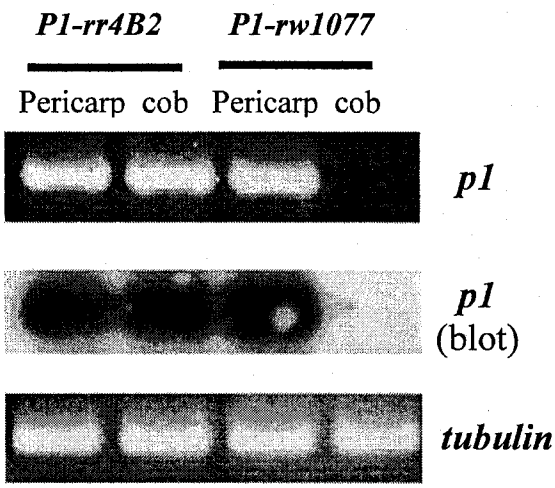

P1-rr4B2
P1-rw1077
Pericarp cob Pericarp cob
p1
p1
(blot)
tubulin

Figure 3

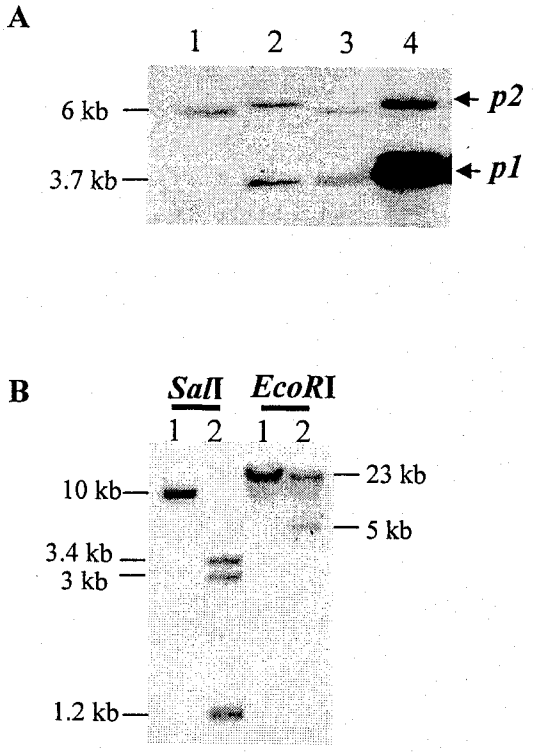

A
1
2
3
4
6 kb
3.7 kb
p2
p1
B
SalI
EcoRI
1
2
1
2
10 kb
3.4 kb
3 kb
1.2 kb
23 kb
5 kb

Figure 4

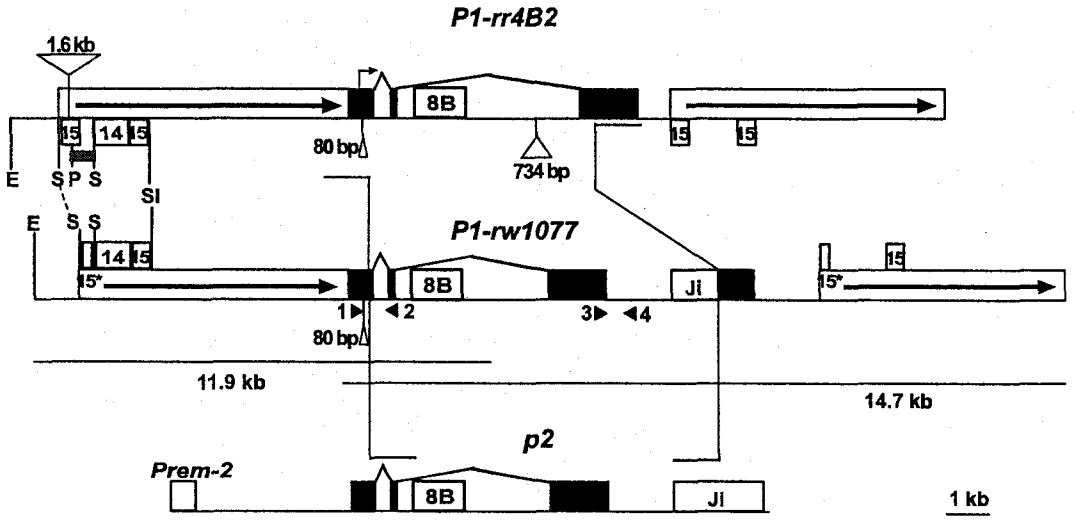
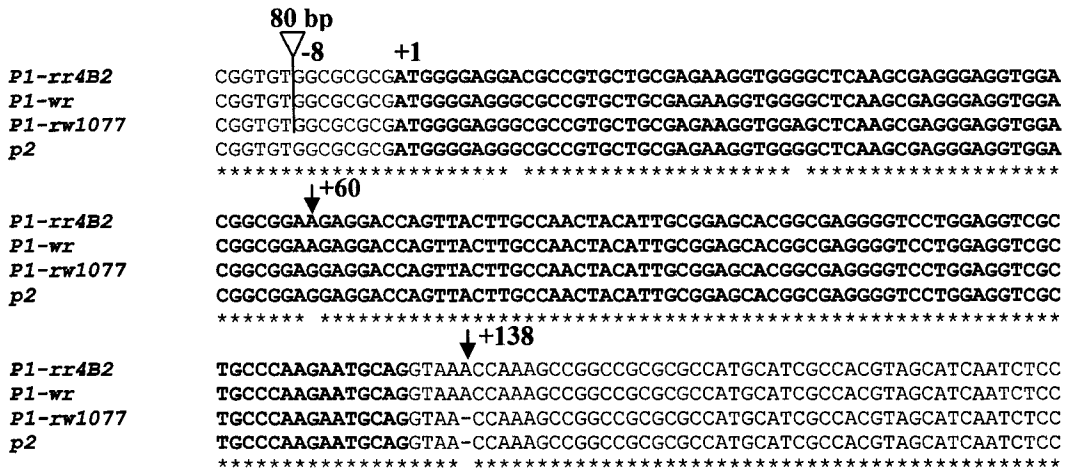

P1-rr4B2
1.6kb
15
14
15
E
S
P
S
SI
8B
80 bp
734 bp
15
15
E
S
S
14
15
P1-rw1077
15*
8B
Ji
15
15*
1
2
3
4
80 bp
11.9 kb
14.7 kb
p2
Prem-2
8B
Ji
1 kb

Figure 5

A


80 bp
-8
+1
P1-rr4B2 CGGTGTGGCGCGCGATGGGGAGGACGCCGTGCTGCGAGAAGGTGGGGCTCAAGCGAGGGAGGTGGA
P1-wr CGGTGTGGCGCGCGATGGGGAGGGCGCCGTGCTGCGAGAAGGTGGGGCTCAAGCGAGGGAGGTGGA
P1-rw1077 CGGTGTGGCGCGCGATGGGGAGGGCGCCGTGCTGCGAGAAGGTGGAGCTCAAGCGAGGGAGGTGGA
p2 CGGTGTGGCGCGCGATGGGGAGGGCGCCGTGCTGCGAGAAGGTGGGGCTCAAGCGAGGGAGGTGGA
+60
P1-rr4B2 CGGCGGAAGAGGACCAGTTACTTGCCAACTACATTGCGGAGCACGGCGAGGGGTCCTGGAGGTCGC
P1-wr CGGCGGAAGAGGACCAGTTACTTGCCAACTACATTGCGGAGCACGGCGAGGGGTCCTGGAGGTCGC
P1-rw1077 CGGCGGAGGAGGACCAGTTACTTGCCAACTACATTGCGGAGCACGGCGAGGGGTCCTGGAGGTCGC
p2 CGGCGGAGGAGGACCAGTTACTTGCCAACTACATTGCGGAGCACGGCGAGGGGTCCTGGAGGTCGC
+138
P1-rr4B2 TGCCCAAGAATGCAGGTAAACCAAAGCCGGCCGCGCGCCATGCATCGCCACGTAGCATCAATCTCC
P1-wr TGCCCAAGAATGCAGGTAAACCAAAGCCGGCCGCGCGCCATGCATCGCCACGTAGCATCAATCTCC
P1-rw1077 TGCCCAAGAATGCAGGTAA-CCAAAGCCGGCCGCGCGCCATGCATCGCCACGTAGCATCAATCTCC
p2 TGCCCAAGAATGCAGGTAA-CCAAAGCCGGCCGCGCGCCATGCATCGCCACGTAGCATCAATCTCC

B

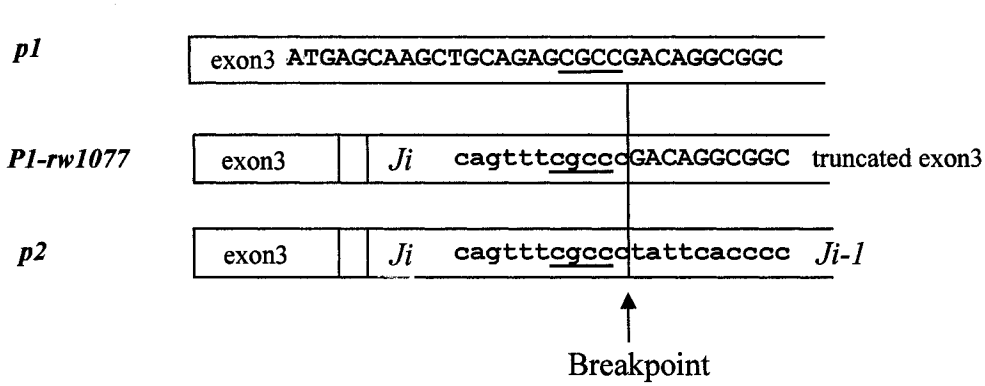

p1
exon3 ATGAGCAAGCTGCAGAGCGCCGACAGGCGGC
P1-rw1077
exon3
Ji
cagtttcgcccGACAGGCGGC
truncated exon3
p2
exon3
Ji
cagtttcgccctattcacccc
Ji-1
Breakpoint

Figure 6

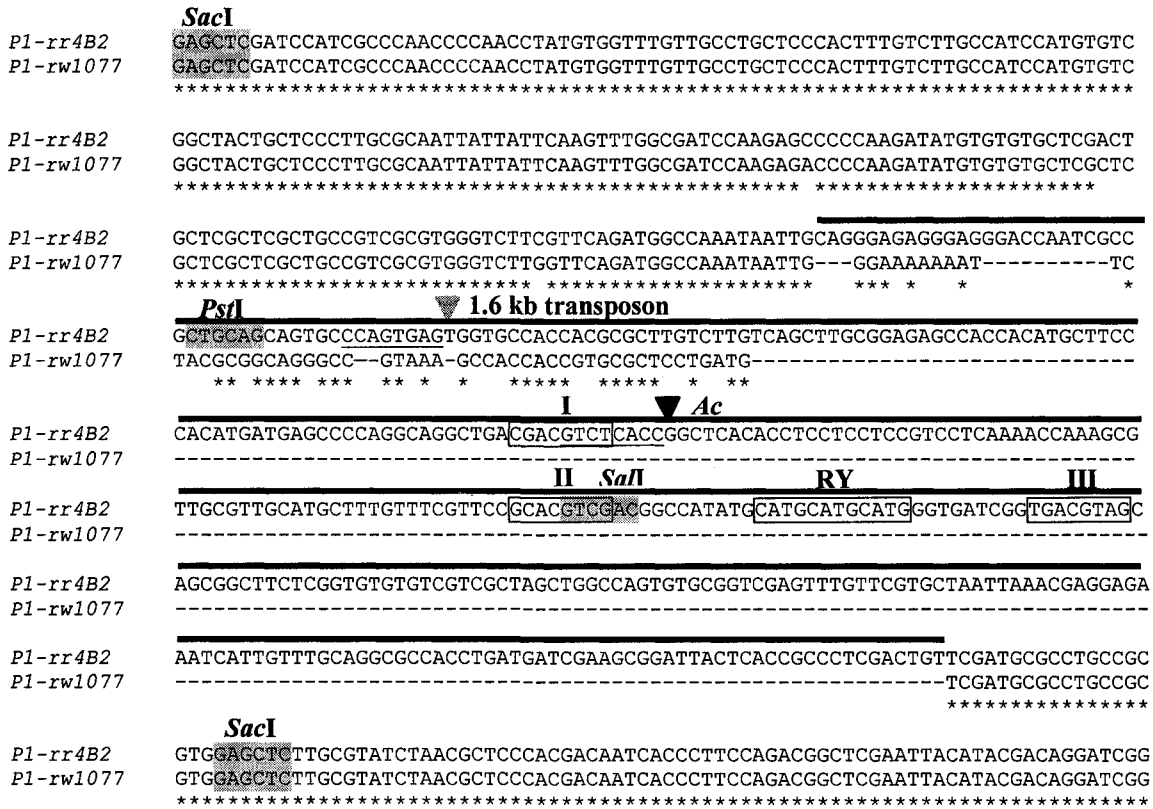
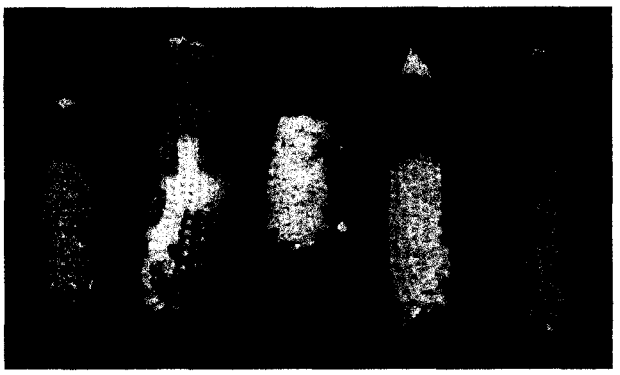

SacI
P1-rr4B2 GAGCTCGATCCATCGCCCAACCCCAACCTATGTGGTTTGTTGCCTGCTCCCACTTTGTCTTGCCATCCATGTGTC
P1-rw1077 GAGCTCGATCCATCGCCCAACCCCAACCTATGTGGTTTGTTGCCTGCTCCCACTTTGTCTTGCCATCCATGTGTC
P1-rr4B2 GGCTACTGCTCCCTTGCGCAATTATTATTCAAGTTTGGCGATCCAAGAGCCCCCAAGATATGTGTGTGCTCGACT
P1-rw1077 GGCTACTGCTCCCTTGCGCAATTATTATTCAAGTTTGGCGATCCAAGAGACCCCAAGATATGTGTGTGCTCGCTC
P1-rr4B2 GCTCGCTCGCTGCCGTCGCGTGGGTCTTCGTTCAGATGGCCAAATAATTGCAGGGAGAGGGAGGGACCAATCGCC
P1-rw1077 GCTCGCTCGCTGCCGTCGCGTGGGTCTTGGTTCAGATGGCCAAATAATTG---GGAAAAAAAT----------TC
PstI
1.6 kb transposon
P1-rr4B2 GCTGCAGCAGTGCCCAGTGAGTGGTGCCACCACGCGCTTGTCTTGTCAGCTTGCGGAGAGCCACCACATGCTTCC
P1-rw1077 TACGCGGCAGGGCC--GTAAA-GCCACCACCGTGCGCTCCTGATG----------
I
Ac
P1-rr4B2 CACATGATGAGCCCCAGGCAGGCTGACGACGTCTCACCGGCTCACACCTCCTCCTCCGTCCTCAAAACCAAAGCG
P1-rw1077 ----------
II
SalI
RY
III
P1-rr4B2 TTGCGTTGCATGCTTTGTTTCGTTCCGCACGTCGACGGCCATATGCATGCATGCATGGGTGATCGGTGACGTAGC
P1-rw1077 ----------
P1-rr4B2 AGCGGCTTCTCGGTGTGTGTCGTCGCTAGCTGGCCAGTGTGCGGTCGAGTTTGTTCGTGCTAATTAAACGAGGAGA
P1-rw1077 ----------
P1-rr4B2 AATCATTGTTTGCAGGCGCCACCTGATGATCGAAGCGGATTACTCACCGCCCTCGACTGTTCGATGCGCCTGCCGC
P1-rw1077 ----------TCGATGCGCCTGCCGC
SacI
P1-rr4B2 GTGGAGCTCTTGCGTATCTAACGCTCCCACGACAATCACCCTTCCAGACGGCTCGAATTACATACGACAGGATCGG
P1-rw1077 GTGGAGCTCTTGCGTATCTAACGCTCCCACGACAATCACCCTTCCAGACGGCTCGAATTACATACGACAGGATCGG

Figure 7

A



B

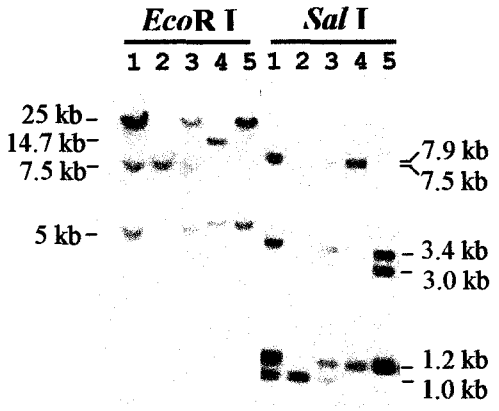

*Eco*R I
Sal I
1 2 3 4 5 1 2 3 4 5
25 kb
14.7 kb
7.5 kb
5 kb
7.9 kb
7.5 kb
3.4 kb
3.0 kb
1.2 kb
1.0 kb

Figure 8

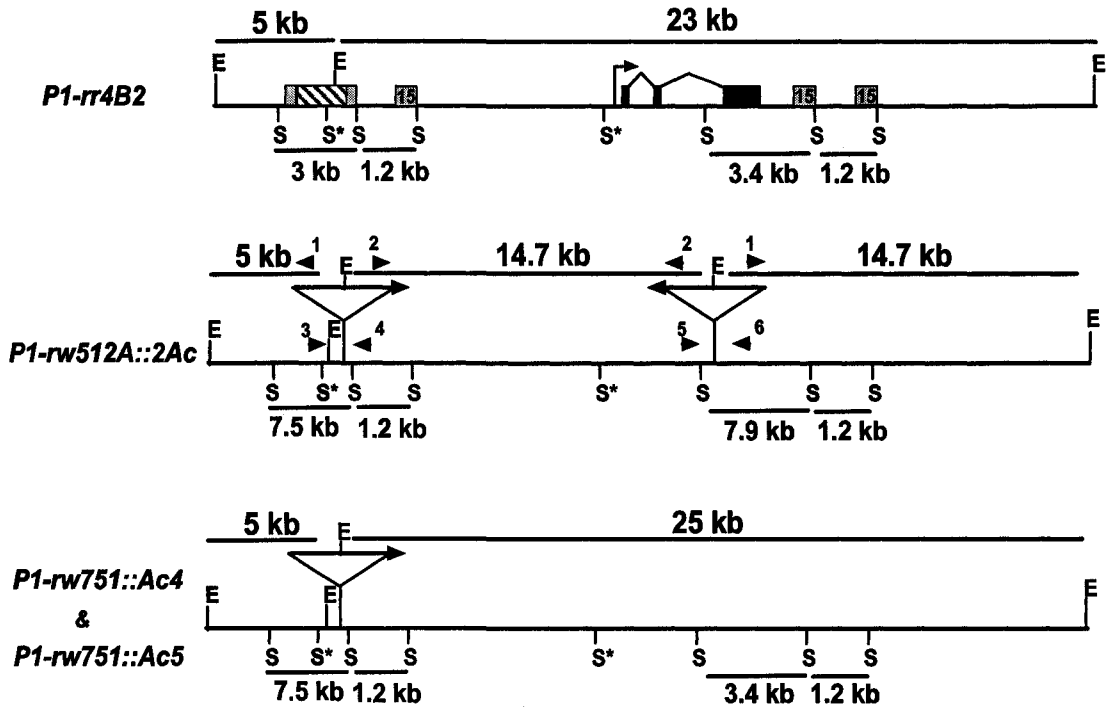

5 kb
23 kb
P1-rr4B2
E
E
E
15
15
15
S
S*
S
S
S*
S
S
S
3 kb
1.2 kb
3.4 kb
1.2 kb
5 kb
1
2
14.7 kb
2
1
14.7 kb
P1-rw512A::2Ac
E
3
E
4
5
6
E
S
S*
S
S
S*
S
S
S
7.5 kb
1.2 kb
7.9 kb
1.2 kb
5 kb
E
25 kb
P1-rw751::Ac4
&
P1-rw751::Ac5
E
E
E
S
S*
S
S
S*
S
S
S
7.5 kb
1.2 kb
3.4 kb
1.2 kb

Figure 9

CHAPTER 3. GENE CONVERSION BETWEEN NON-CODING DIRECT REPEATS PROMOTES GENETIC AND PHENOTYPIC DIVERSITY AT THE MAIZE *PERICARP COLOR1* LOCUS

A paper to be submitted to Genetics

Feng Zhang and Thomas Peterson

Abstract

While evolution of coding sequences has been intensively studied, diversification of non-coding regulatory regions remains poorly understood. In this study, molecular evolution of a distal enhancer region, 5 kb upstream of transcription start site, was investigated at the maize pericarp color1 (*p1*) locus. The *p1* gene encodes an R2R3 Myb-like transcription factor that regulates flavonoid biosynthetic pathway in maize floral organs. Characterization of 6 single-copy *p1* alleles revealed 3 distinct types of the distal enhancer. Our study suggests that homologous recombination between the 5' and 3' *p1* non-coding regions promotes rapid evolution of the distal enhancer. As a result, novel distal enhancer types were generated during the *p1* evolution, giving rise to novel phenotypic variation. A stepwise model was proposed to account for the evolutionary pathway of the distinct *p1* alleles. The importance of recombination between duplicated non-coding regions in the generation of genetic and phenotypic diversity is discussed.

Introduction

Evolution of cis-regulatory elements has been indicated as a major contributor to phenotypic variation, because changes in regulatory regions can induce temporal and spatial expression pattern changes (CARROLL *et al.* 2004; DOEBLEY and LUKENS 1998; LUDWIG 2002; WRAY *et al.* 2003). Despite the importance of cis-elements in regulation of gene expression, the molecular basis of cis-regulatory region evolution remains poorly understood. In recent studies, comparisons of cis-regulatory regions between natural variants has been indicated as a powerful approach to investigate the evolution of regulatory sequences (HANSON *et al.* 1996; LUDWIG *et al.* 2000; PURUGGANAN 2000; WANG *et al.* 1999). In this study, we used natural variations of the gene specifying flavonoid pigment patterns in maize to gain insight into the evolutionary dynamic of cis-regulatory sequences. The genetic loci, including the *c1*, *p1*, *r1*, *b1* and *pl1* (not to be confused with the *p1* gene) genes, that regulate flavonoid biosynthetic pathways of maize have been well characterized at the molecular level. All of these regulatory genes display diverse patterns of gene expression in both floral and vegetative organs of maize (CONE *et al.* 1993; LUDWIG and WESSLER 1990; PROCISSI *et al.* 1997; SELINGER *et al.* 1998). Allelic variation has been investigated at the *c1*, *b1* and *r1* loci in great detail. Variations in regulatory regions have been linked to distinct allelic expression patterns and phenotypic diversity (LI *et al.* 2001; SELINGER *et al.* 1998). Moreover, studies on the *c1* gene have suggested that the cis-regulatory region could be subject to selection, resulting in increased frequency of one *c1* haplotype and giving rise to the pigmented kernel phenotype (HANSON *et al.* 1996).

Recent analyses of the *p1* gene provide further evidence that variations in cis-regulatory regions are involved in phenotypic diversification. The *p1* gene encodes a R2R3

Myb-like transcription factor. Expression of the *p1* gene is primarily restricted to the maize floral organs, most notably the kernel pericarp and cob glumes. According to the pigmentation patterns in these two organs, *p1* alleles are commonly classified into four types: *P1-rr* (red pericarp/red cob), *P1-wr* (white pericarp/red cob), *P1-rw* (red pericarp/white cob) and *p1-ww* (white pericarp/white cob). Three distinct *p1* alleles, *P1-rr4B2*, *P1-rw1077* and *P1-wr[w22]* have been characterized and compared at molecular level. All of these *p1* alleles share highly similar coding regions, but differ in the distal enhancer regions (CHOPRA *et al.* 1998; SIDORENKO *et al.* 2000). Our recent data have demonstrated that changes in the distal enhancer regions result in phenotypic variations between *P1-rr4B2* and *P1-rw1077* (ZHANG and PETERSON 2005).

More than 100 natural variants of *p1* have been reported, each exhibiting distinctive patterns and intensity of pigmentation in kernel pericarp and cob glumes (BRINK and STYLES 1966). This rich collection of natural occurring alleles provides an excellent system to study the evolution of cis-regulatory regions at the *p1* locus. In this study, 6 distinct single-copy *p1* alleles were characterized and compared. The goals of the study were: (i) to investigate DNA variations in non-coding regions of the *p1* locus, particularly the distal enhancer regions; (ii) to examine correlation between variations in the distal enhancer and phenotypic changes; (iii) to identify the forces affecting evolution of the distal enhancer regions, and the generation of genetic and phenotypic diversity observed at the *p1* locus.

Materials and methods

Collection of the *p1* alleles: The *p1* alleles used in this study are homozygous in the 4Co63 genetic background. As shown in Table 1, the *p1* alleles with an assigned CFS

number were from Brink's *p1* collection (BRINK and STYLES 1966). The *P1-rr4B2* and *P1-rw1077* alleles are the same as used in previous studies (LECHELT *et al.* 1989; ZHANG and PETERSON 2005).

Survey of the *p1* genotypes: DNA gel blot analysis was used to survey the genotypes of the *p1* alleles. Genomic DNA was made from fresh leave tissues by the CTAB method (SAGHAI-MAROOF *et al.* 1984) and digested with restriction enzymes according to manufacturer's instructions. The electrophoresis and hybridization procedures were conducted as described in previous studies (SIDORENKO *et al.* 2000). The blot was probed with the *p1* specific probe, genomic Fragment 15 (Figure 1). The *p1* alleles with identical hybridization patterns were grouped as one allelic type.

Amplification and sequencing of the *p1* non-coding regions: Nested genomic PCR were performed to amplify the 5' non-coding regions. The primer pair, P1rr-18 and PA-B4, was used in the first round reaction. A 1 μl aliquot of the first-round PCR product was subject to a second round PCR with the primer pair, P1rr-14 and PA-B6. The 3' non-coding regions were amplified in two overlapping pieces separately by using the primers EP5-16 and P1rr-16, as well as nested primers, P1rr-13r and P1rr-30; P1rr-16r and P1rr-29. The 723 bp sequences in the second intron were amplified with primers, 723-5 and 723-3. Locations and sequences of all primers are shown in Figure 1, Figure 4 and Table 4. The PCR reactions were performed using Enhanced DNA polymerase, Elongase® (Invitrogen, Carlsbad, CA) with ~1 min of extension time for every 1 kb fragment size. Optimal PCR parameters were followed as suggested by the manufacturer. The PCR products were purified using gel extraction kit (Qiagen, Valencia, CA), and sequenced from both directions at Iowa State

University DNA Sequencing Facility. The final sequences were inspected and assembled using the software package, Vector NTI Advance™ 9.0 (InforMax™, Frederick, MD).

Sequence analysis: Sequences from the *P1-rr1088*, *P1-rrCFS36*, *P1-rwCFS302* and *P1-rwCFS342* were aligned with those from *P1-rr4B2* and *P1-rw1077* using the software package, Vector NTI Advance™ 9.0 (InforMax™, Frederick, MD). An unrooted phylogenetic tree was constructed using MEGA, version 2.1 (KUMAR *et al.* 2001). The Neighbor-Joining method was conducted with Kimura's (1980) two parameter distance. The robustness of the tree topologies were tested by performing 1000 bootstrap replicates. DNA diversity estimation and sliding-window analysis on the *p1* alleles were conducted using program package, DnaSP 4.0 (ROZAS and ROZAS 1999).

Results

Characterization of *p1* alleles

More than 100 natural *p1* variations have been collected and classified according to their pigmentation patterns in kernel pericarp and cob glumes (BRINK and STYLES 1966). Previous studies on three distinct *p1* alleles indicated that the *p1* gene could be either a single copy locus, as in *P1-rr4B2* and *P1-rw1077* (LECHELT *et al.* 1989; ZHANG and PETERSON 2005), or a multiple copy complex, as in *P1-wr[w22]*(CHOPRA *et al.* 1998). A DNA gel blot survey with *p1* specific probes has been conducted on 24 *P1-rr*, 6 *P1-rw* and 9 *P1-wr* alleles (COCCIOLONE *et al.* 2001), McMullen personal communication; Zhang, F unpublished data). The results showed that all examined *P1-wr* alleles contain multiple copies of *p1* sequences, while all *P1-rw* alleles are single copy genes; the *P1-rr* alleles could be either single copy or multiple copy. Because of the complicated nature of multiple copy *p1* alleles (CHOPRA *et al.*

1998), we focused on the characterization of single copy *p1* alleles in this study. According to the DNA gel blot analyses, the single copy *p1* alleles, including 6 *P1-rw* and 9 *P1-rr* alleles, can be classified as 6 allelic types: three *P1-rr* types and three *P1-rw* types (Table 1). The *P1-rr4B2* and *P1-rw1077* allelic types have been characterized and compared previously (LECHELT *et al.* 1989; SIDORENKO *et al.* 2000; ZHANG and PETERSON 2005). Representative alleles (*P1-rrCFS36*, *P1-rr1088*, *P1-rwCFS304* and *P1-rwCFS342*) from the other four allelic types were chosen for further analysis (Table 1).

Structural comparisons of *p1* alleles

Previous studies have shown that the coding regions of different *p1* alleles share high sequence similarity; and DNA polymorphisms in non-coding regulatory regions rather than in coding regions could be responsible for phenotypic variations of the *p1* alleles (LECHELT *et al.* 1989; SIDORENKO *et al.* 2000; ZHANG and PETERSON 2005). To examine sequence polymorphisms in non-coding regions of the single copy *p1* alleles studied here, we used genomic PCR with *p1*-specific primers to amplify both 5' and 3' non-coding regions of the *P1-rr* and *P1-rw* alleles. In each allele, approximately 6 kb PCR products from both regions were sequenced and assembled (see Materials and Methods). The structures of the non-coding regions of all six single-copy *p1* alleles were compared as shown in Figure 1. Like those in *P1-rr4B2* and *P1-rw1077*, the 5' and 3' non-coding regions in the 4 new *p1* alleles are also featured as long direct repeats flanking the *p1* coding sequences. The repeat unit could extend as long as 6.3 kb (as in *P1-rw1077*) from a 1.1 kb region, termed as Fragment C, to a 15 bp small repeat $(5'\text{-GCGGGAGTGCGGCCT-}3')_2$.

Structural and sequence comparisons between the simplex *p1* alleles revealed a number of DNA polymorphisms in both 5' and 3' non-coding direct repeats. In the 5' repeats, the most notable polymorphic regions are located in the previously identified enhancer regions (Figure 1). The distal enhancer region is located 5 kb upstream of the *p1* transcription start site, and includes the 405 bp Fragment 15 and a 666 bp sequence termed Fragment 14 (SIDORENKO *et al.* 2000; ZHANG and PETERSON 2005; Figure 1). According to the copy number of Fragment 15 and Fragment 14 sequences in the distal enhancer regions, 3 distinct enhancer types can be defined in 6 simplex *p1* alleles: (i) 1 copy of Fragment 15 and 0 copy of Fragment 14, i.e. the enhancer region contains only Fragment 15, as shown in *P1-rwCFS302* and *P1-rwCFS342*; (ii) 1½ copies of Fragment 15 and 1 copy of Fragment 14, i.e. the distal enhancer region contains a partial copy of Fragment 15 followed by Fragment 14 and a full copy of Fragment 15, as seen in *P1-rw1077*; (iii) 2 copies of Fragment 15 and 1 copy of Fragment 14, i.e. the distal enhancer region contains a full copy of Fragment 15 followed by Fragment 14 and additional full copy of Fragment 15, as shown in *P1-rr1088*, *P1-rrCFS36* and *P1-rr4B2*. The type iii enhancer can be further classified as two subtypes: (i) *P1-rrCFS36 and P1-rr4B2*, in which a 1.6 kb transposon-like sequence (SIDORENKO *et al.* 2000) is present in the midpoint of the upstream Fragment 15; (ii) *P1-rr1088*, in which the 1.6 kb sequence is absent in the distal enhancer (Figure 1). Further sequence analysis indicated that a 8 bp (CCAGTGAG) putative target site duplication (TSD) accompanied insertion of the 1.6 kb transposon-like sequence. Because no footprints, presumably left by excision of the 1.6 kb transposon, were found in *P1-rr1088*, the sequence polymorphism between the two subtypes appears to be consequence of transposon insertion rather than deletion.

In the 3' non-coding regions, the sequences comprising Fragment 15 and Fragment 14 are also highly variable. Similar to the 5' non-coding sequences, the 3' regions of the simplex *p1* alleles can also be classified as 3 distinct types: (i) 1 copy of Fragment 14 and 1 copy of Fragment 15 as shown in *P1-rwCFS342*, *P1-rrCFS36* and *P1-rr1088*; (ii) 1½ copies of Fragment 15 and 1 copy of Fragment 14 as seen in *P1-rw1077*; (iii) 2 copies of Fragment 14 and 2 copies of Fragment 15 as shown in *P1-rr4B2* (equivalent to two copies of type (i) sequences arranged in direct orientation).

Sequence alignment also revealed two additional large polymorphic regions in the 5' and 3' non-coding regions of the *p1* simplex alleles: a 549 bp sequence (738 bp downstream of Fragment 15 in the 3' coding region) that is absent from both 5' and 3' non-coding direct repeats of *P1-rwCFS342* (Figure 1); and a 148 bp sequence (~3500 bp downstream of Fragment 15 in the 3' non-coding region) that is absent from the 3' non-coding regions of *P1-rwCFS342*. Sequence analysis of a *p1* homologous gene from teosinte (*Zea mays* subsp *parviglumis*; (ZHANG *et al.* 2000)), the immediate progenitor of modern maize, indicated that both 549 bp and 148 bp sequences are present in teosinte (data not shown). This result suggested that these two polymorphic regions likely result from deletions in *P1-rwCFS342* rather than insertions in other *p1* alleles.

Variations in the distal enhancers correlate with changes of pigmentation patterns in cob glumes

Comparisons of the simplex *p1* alleles revealed three distinct types of distal enhancers. Interestingly, all *P1-rr* alleles share the same enhancer type, which differs from that of the *P1-rw* alleles. Previous studies on the *P1-rr4B2* and *P1-rw1077* alleles have identified a 386

bp cob glume-specific regulatory region in the distal enhancer (ZHANG and PETERSON 2005). In *P1-rr4B2*, this 386 bp region (Figure 1) overlaps with the upstream copy of Fragment 15 in the distal enhancer. In contrast, this particular region is absent from the distal enhancer in *P1-rw1077* (Figure 1), and thereby accounts for the loss of *p1* expression as well as pigmentation in cob glumes (ZHANG and PETERSON 2005). The correlation between the presence of the cob glume-specific regulatory region in the distal enhancer and red pigmentation in cob glumes was examined in the additional 4 *p1* simplex alleles. In *P1-rrCFS36* and *P1-rr1088*, the distal enhancer regions of both alleles contain the 386 bp regulatory region as in *P1-rr4B2* (Figure 1). In contrast, in *P1-rwCFS302* and *P1-rwCFS342*, the distal enhancers, which contain only one copy of Fragment 15, lack the cob glume-specific region (Figure 1). Therefore, in the single copy *p1* alleles examined in this study, the presence/absence of the cob glumes-specific regulatory region is correlated with gain/loss of pigmentation in cob glumes.

Insertions and deletions in the *p1* non-coding regions

As one component in the distal enhancer region, Fragment 14 is not present in the distal enhancer of all *p1* alleles. Based on the evidence from the *p* homologous gene in teosinte, it has been proposed that the *p1* progenitor gene did not contain Fragment 14 sequence, and contained only one copy of Fragment 15 sequence in both 5' and 3' non-coding regions (ZHANG *et al.* 2000). Thus, the appearance of Fragment 14 in the *p1* locus must have been the result of an insertion event. Previous study has indicated that Fragment 14 sequence is repetitive in maize genome (~4-5 copies; (CHOPRA *et al.* 1998). Maize sequence database search was performed to identify homologous sequences of Fragment 14

(PlantGDB, http://www.plantgdb.org/; (DONG *et al.* 2004). Six maize genomic survey sequences, ranging from 703 bp to 945 bp, were identified with the highest match scores (E value = 0). Two of those sequences belong to an assembly contig (contig number ZmGSStuc11-12-04.23582.1). Sequence analysis indicated that this 3582-bp contig shares 99% sequence identity with Fragment 14 over a 467 bp region, and 68% similarity with a Mutator-like transposase (Genbank accession number: AC146468) over a 2570 bp region. These observations suggest a link between Fragment 14 sequences and Mutator-like transposable elements. Possibly, the insertion of Fragment 14 into the *p1* locus was mediated by insertion of a transposable element. Notably, some *p1* alleles, i.e. *P1-rw1077*, *P1-rr4B2*, *P1-rrCFS36* and *P1-rr1088*, contain Fragment 14 in both 5' and 3' non-coding regions. Our data suggested that those structures likely resulted from gene conversion rather than independent insertion events (see below).

In addition to insertion, sequence comparisons in the 3' non-coding regions between *P1-rwCFS302* and *P1-rwCFS342* indicated that deletion can also contribute to the generation of molecular diversity in the *p1* locus. Deletion breakpoints were identified by aligning sequences from the 3' non-coding region of *P1-rwCFS302* and *P1-rwCFS342*. As shown in Figure 2, a 1407 bp sequence, including Fragment C and the partially-duplicated Fragment 15, was deleted from the *P1-rwCFS342* alleles. A 13 bp insertion occurred at the site of the deletion; 9 nucleotides of this insertion are similar to a nearby sequence located 51 bp downstream of the 3' deletion end points, suggesting a possible origin as filler DNA. This observation of deletion associated with small filler DNA is similar to that reported for the maize *waxy* and *bronze* loci (RALSTON *et al.* 1988; WESSLER *et al.* 1990). It has been suggested that such deletions could result from slip mispairing during replication. The

deleted type of the 3' non-coding region is also shared in all *P1-rr* alleles, which may have been derived from the *P1-rwCFS342* allele (see discussion for detail).

Gene conversion between non-coding regions in the *p1* locus

The 5' and 3' non-coding regions of the *p1* locus are arranged as direct repeats, separated by the ~6 kb transcribed region. The observation that a 549-bp deletion is located at identical positions in both 5' and 3' non-coding regions of *P1-rwCFS342* (Figure 1) suggests that gene conversion events occurred between the non-coding direct repeats in the *p1* locus. To assess the role of gene conversion in the creation of genetic diversity in the *p1* non-coding regions, particularly in the distal enhancer regions, phylogenetic analysis was performed on a 602-bp sequence, which includes Fragment 15 and a 180 bp sequence immediate downstream of Fragment 15. This particular region is the key component in the distal enhancer, and present in both 5' and 3' non-coding regions of all single copy *p1* alleles. A total of 18 sequences were used to generate a neighbor-joining phylogenetic tree (Figure 3): 16 sequences from 5' and 3' repeats of the simplex *p1* alleles; one sequence from the multiple copy *p1* allele, *P1-wr* (CHOPRA *et al.* 1998); and one sequence from the 3' non-coding region of the teosinte *p* gene (*p2t*) (ZHANG *et al.* 2000). In the different *p1* alleles, if the 5' repeats evolved independently from the 3' repeats, the 5' sequences and the 3' sequences would form distinct groups in the phylogenetic tree. Gene conversion events between the 5' and 3' repeats, however, could change the phylogenetic patterns by clustering the 5' sequence with the 3' sequence from the same *p1* allele. Based on this logic, from the resulting phylogenetic tree, gene conversion events were detected between the 5' and 3' sequences from 4 different *p1* alleles, i.e. *P1-rr4B2*, *P1-rw1077*, *P1-rrCFS36* and *P1-rr1088*.

Gene conversion between non-coding repeats could increase nucleotide diversity

Nucleotide diversity was estimated in both the 5' and 3' non-coding repeats of the simplex *p1* alleles. As indicated in Figure 4, the regions that were sampled in this analysis are shared in all simplex *p1* alleles. In the 5' non-coding regions, a total of 19 indels and 83 polymorphic sites were detected, while, in the 3' non-coding regions, a total of 42 indels and 197 polymorphic sites were detected. The estimated values (π and θ) for DNA diversity also indicated that the 3' regions are more diversified than the 5' regions (e.g. π value: 0.02929 vs. 0.00906; Table 2). The possible explanation for this observation is that the 5' non-coding regions, but not the 3' non-coding regions, contain critical regulatory regions for *p1* expression and are therefore under functional constraints.

To further investigate the distribution of polymorphic sites, sliding-window analysis was performed on both 5' and 3' non-coding regions. This analysis revealed that the polymorphic sites are distributed throughout the 3' non-coding regions, while, in the 5' non-coding regions, polymorphic sites were frequently observed in the first 2 kbp region but are much less frequent in the remaining regions (Figure 4). Unequal functional constraints on the two regions could be one possible explanation for heterogeneity of nucleotide diversity in the 5' non-coding regions. By *Ac* transposon mutagenesis assay, however, no functional elements were identified in the low diversity region yet, while the distal enhancer was found in the first 2 kb region (MORENO *et al.* 1992; SIDORENKO *et al.* 2000). Further investigation is required to examine whether functional constraints differ on these two regions.

Alternatively, the uneven distribution of polymorphic sites in the 5' non-coding regions could result from unequal frequency of recombination events. As mentioned above,

gene conversion events could have occurred between the 5' and 3' non-coding direct repeats. It has been suggested that gene conversion between duplicated sequences could increase divergence of orthologous sequences more rapidly than any other mutational processes (ANGERS *et al.* 2002; OHTA 1995). Possibly, gene conversion events may have occurred more frequently in the first 2 kbp region between the 5' and 3' direct repeats, than in the further downstream regions. As a result, DNA diversity in the first 2 kbp region in the 5' non-coding region was increased by converting nucleotide variations present in the 3' non-coding regions, which probably accumulated more freely there due to less functional constraints. Although it is not clear why recombination frequency would differ between these adjacent regions, uneven distribution of recombination events have been reported in a1, b1 and r1 loci (EGGLESTON *et al.* 1995; PATTERSON *et al.* 1995; YAO *et al.* 2002). Because of the small sample size (6 *p1* alleles) used in this study, however, investigation of addition *p1* variations is necessary to test either hypothesis.

If gene conversion can increase the DNA diversity of duplicated sequences at the *p1* locus, then one might predict that non-duplicated *p1* sequences should have lower DNA diversity than the repeated sequences. The ideal sequence to test this prediction is the 723 bp sequence located in the second intron of *p1* (Figure 4), because (i) this sequence is present only in the *p1* locus but not in the *p2* gene, the tightly-linked paralogous gene of *p1*, which has potential to undergo gene conversion with *p1* (ZHANG and PETERSON 2005; ZHANG *et al.* 2000); (ii) this sequence seems to be subject to no functional constraints as it is dispensable for expression of *P1-rw1077* (ZHANG and PETERSON 2005). Consistent with the hypothesis, the values for DNA diversity of the 723 bp sequence are 8-26 times lower than those of the 5' and 3' non-coding regions (Table 2).

Estimated divergence time of the *p1* alleles

As stated above, the 723 bp sequence in the *p1* second intron likely evolved neutrally; and nucleotide diversity of this region could not have been affected by recombination. Thus, this sequence can be used to estimate the divergence time of the *p1* alleles. The pair wise nucleotide substitution rates were estimated between the distinct *p1* alleles, including the multiple-copy *P1-wr* allele and 5 simplex *p1* alleles (Table 3). Because *P1-rw1077* lacks the 723 bp sequence, the substitution rates between *P1-rw1077* and other *p1* alleles cannot be determined. Given the estimate of substitution rates in grass nuclear gene is ~6 x 10^{-9} substitute per synonymous site per year (GAUT 1998), the divergence time between the multiple copy *P1-wr* allele and the common ancestor of the simplex *p1* allels is ~1.3-1.4 millions years ago. In contrast, within the simplex *p1* alleles, the divergence time of two *P1-rw* alleles, *P1-rwCFS342* and *P1-rwCFS302* is estimated as 230,000 years, while the time between *P1-rw*, either *P1-rwCFS342* or *P1-rwCFS302*, and *P1-rr* is ~120,000 years (Figure 5). The time of divergence between the *P1-rr* alleles can not be estimated, because no substitutions were detected between these alleles.

Discussion

Origin and diversification of the *p1* gene: Alleles of the *p1* gene exhibit a great degree of genetic and phenotypic diversity (COCCIOLONE *et al.* 2001; ZHANG and PETERSON 2005). In a previous study, Zhang et al (2000) suggested that the *p1* gene was generated by a tandem duplication event ~2.8 millions years ago. How was the observed high degree of diversity created at the *p1* locus in this short period of evolutionary time? In this study, investigation

of 6 distinct *p1* alleles suggested that the non-coding repeats flanking the *p1* coding sequence may have played an important role in diversification of the *p1* locus during evolution.

Sequence comparisons of 6 distinct single-copy *p1* alleles as well as the multiple-copy *P1-wr* allele have revealed that all the *p1* alleles contain the long direct non-coding repeats that flank the *p1* coding sequence. It has been proposed that the long direct repeats resulted from the recent tandem duplication event (ZHANG *et al.* 2000). The gene structure of a common ancestor of all *p1* alleles has been deduced based on analysis of a *p* homologous gene in teosinte ((ZHANG *et al.* 2000). As indicated in Figure 5, in the *p1* progenitor gene, the long direct repeat extends from the sequence, termed as Fragment C, to a small repeat $(5'\text{-GCGGGAGTGCGGCCT-}3')_2$. Subsequent sequence rearrangements in both 5' and 3' repeats diversified the *p1* alleles.

Structural comparisons of the distinct *p1* alleles as well as the evidence from phylogenetic analysis and nucleotide diversity analysis allowed us to propose an evolutionary model for diversification of the *p1* alleles (Figure 5). In this model, the multiple-copy *P1-wr* allele diverged early from the progenitor gene at ~1.4 mya (million years ago), possibly due to unequal crossover between the non-coding direct repeats (CHOPRA *et al.* 1998), while the single copy *p1* allele split from the latest common ancestor in more recent times (starting at ~0.23 mya). During the evolution of the simplex *p1* alleles, we propose a stepwise pathway to most parsimoniously account for generation of the distinct single copy *p1* alleles. In this pathway, as indicated in Figure 5, the DNA rearrangements started at the 3' non-coding repeats with insertion of Fragment 14 and partial duplication of Fragment 15, i.e. the structure shown in *P1-rwCFS302*. Such rearrangements could be converted to the 5' non-coding region by gene conversion as shown in *P1-rw1077*; or they can be further modified

by deletion, giving rise to the structures of the 3' non-coding regions observed in the *P1-rwCFS342* and *P1-rr* alleles (Figure 2). In our model, all *P1-rr* alleles share a common ancestor with the *P1-rwCFS342* allele, which is consistent with the observation that the *P1-rr* alleles form a monophyletic group in the phylogenetic analysis (Figure 3). Then how could the *p1* locus change from *P1-rwCFS342*-like structure, which contains only one copy of Fragment 15 in the distal enhancer, to a *P1-rr*-like structure that has 2 copy of Fragment 15 and 1 copy of Fragment 14 in the distal enhancer (Figure 1 and Figure 5)? It is unlikely that such structural changes resulted from an independent insertion of Fragment 14 and duplication of Fragment 15 in the 5' coding region of the *P1-rr* common ancestor, because the chance of independent insertions of Fragment 14 in the same region in both 5' and 3' non-coding repeats is small. More importantly, the phylogenetic analysis based on the sequences from the distal enhancer indicated that the second copy of the duplicated Fragment 15 sequences at the 5' non-coding region is more closely related to the Fragment 15 in the 3' non-coding region rather that to its 5' duplicated copy. Taken together, we proposed that gene conversion between the 5' and 3' non-coding repeats could result in the structural changes in the distal enhancer from the *p1* ancestral state to the *P1-rr* state. It is not clear how such gene conversion events occurred. Possibly, the 8 bp microhomology sequence (TCGATGCC), which is present at both ends of Fragment 15 and the beginning of Fragment 14 (Figure 5), could be involved in the gene conversion event.

Although the proposed evolution model is not in agreement with the phylogenetic tree in every detail, both scenarios state that the *P1-rw* alleles were present early in the evolutionary pathway, while the *P1-rr* alleles likely diverged in more recent times. The divergence time of the *p1* alleles, estimated based on the 723 bp sequence from the *p1* second

intron (Figure 5), postdates the estimated birth date of the *p1* progenitor gene (~2.7 myrs, ZHANG *et al.* 2000). This is consistent with the hypothesis that all *p1* alleles are derived from a single *p1* ancestral gene, generated by a tandem duplication. On the other hand, the diversification of the *p1* gene appears to predate domestication of maize from teosinte (about 7500 year ago, (ILTIS 1983). Most likely, introgression of various *p1* alleles from teosinte into the domesticated maize gene pool was subject to positive human selection due to the remarkable pigmentation phenotypes. However, our estimation does not preclude the possibility that divergence of the *p1* alleles occurred during or after maize domestication, as the DNA diversity between the *p1* alleles are very low. Further investigation of *p1*-homologous genes in representative teosinte stocks is needed to address the question regarding divergence time.

Evolution of the distal enhancer and phenotypic variations at the *p1* locus

As described above, three distinct types of the distal enhancer have been characterized in the simplex *p1* alleles (Figure 1). Our model suggested that the distal enhancer evolved from simple to complex structures, i.e. from a single copy of Fragment 15 to duplicated Fragment 15 sequences with insertion of Fragment 14. In the recent study by Zhang and Peterson (2005), a cob glume-specific regulatory region has been indicated in the distal enhancer type that contains two copies of Fragment 15 (Figure 1). This particular region is associated with all *P1-rr* alleles, but absent from the *P1-rw* alleles. Thus, the evolution of the distal enhancer region appears to accompany the evolution of phenotypic variations observed at the *p1* gene. Based on the evolution model, we speculated that the *p1* alleles conferred RW (red kernel pericarp and white cob glumes) phenotype in the early stage of the *p1* evolution. A later event may have converted the distal enhancer into a novel type

that contains the cob glume-specific regulatory region. Such change resulted in novel expression of the *p1* gene in cob glumes, and gave rise to RR (red kernel pericarp and red cob glumes). As discussed above, gene conversion between the 5' and 3' non-coding repeats could drive evolution of the distal enhancer region.

Taken together, the results of this investigation of the different *p1* alleles indicate that homologous recombination between the non-coding repeat sequences could be a major driving force for generation of genetic and phenotypic diversity at the *p1* locus. The present results suggested that genetic exchanges between non-coding repeats could generate regulatory variations. Such regulatory changes could subsequently produce novel expression and phenotypic variations, which are subject to natural and human selections. Because tandem and segmental duplication are common in both animal and plant genomes (BLANC *et al.* 2000; MCLYSAGHT *et al.* 2002; YU *et al.* 2005), recombination between non-coding duplicated sequences could have a major impact on molecular evolution and diversity in gene expression.

Literature cited

ANGERS, B., K. GHARBI and A. ESTOUP, 2002 Evidence of gene conversion events between paralogous sequences produced by tetraploidization in Salmoninae fish. J Mol Evol **54:** 501-510.

BLANC, G., A. BARAKAT, R. GUYOT, R. COOKE and M. DELSENY, 2000 Extensive Duplication and Reshuffling in the Arabidopsis Genome. Plant Cell **12:** 1093-1102.

BRINK, R., and D. STYLES, 1966 A collection of pericarp factors. Maize Genet. Cooperation Newsletters **40:** 149-160.

CARROLL, S. B., J. K. GRENIER and S. D. WEATHERBEE, 2004 *From DNA to Diversity: Molecular Genetics and the Evolution of Animal Design*. Blackwell Science, Malden, Massachusetts.

CHOPRA, S., P. ATHMA, X. G. LI and T. PETERSON, 1998 A maize Myb homolog is encoded by a multicopy gene complex. Mol Gen Genet **260:** 372-380.

COCCIOLONE, S. M., S. CHOPRA, S. A. FLINT-GARCIA, M. D. MCMULLEN and T. PETERSON, 2001 Tissue-specific patterns of a maize Myb transcription factor are epigenetically regulated. Plant J **27:** 467-478.

CONE, K. C., S. M. COCCIOLONE, F. A. BURR and B. BURR, 1993 Maize anthocyanin regulatory gene pl is a duplicate of c1 that functions in the plant. Plant Cell **5:** 1795-1805.

DOEBLEY, J., and L. LUKENS, 1998 Transcriptional Regulators and the Evolution of Plant Form. Plant Cell **10:** 1075-1082.

DONG, Q., S. D. SCHLUETER and V. BRENDEL, 2004 PlantGDB, plant genome database and analysis tools. Nucleic Acids Res **32 Database issue:** D354-359.

EGGLESTON, W. B., M. ALLEMAN and J. L. KERMICLE, 1995 Molecular Organization and Germinal Instability of R-stippled Maize. Genetics **141:** 347-360.

GAUT, B. S., 1998 Molecular clocks and nucleotide substitution rates in higher plants, pp. 93?20 in E*volutionary Biology,* edited by W. C. S. A. B. W. M. K. HECHT. Plenum, New York.

HANSON, M. A., B. S. GAUT, A. O. STEC, S. I. FUERSTENBERG, M. M. GOODMAN *et al.,* 1996 Evolution of Anthocyanin Biosynthesis in Maize Kernels: The Role of Regulatory and Enzymatic Loci. Genetics 1**43:** 1395-1407.

ILTIS, H. H., 1983 From teosinte to maize: the catastrophic sexual transmutation. Science **222:** 886-894.

KIMURA, M., 1980 A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol **16:** 111-120.

KUMAR, S., K. TAMURA, I. B. JAKOBSEN and M. NEI, 2001 MEGA2: molecular evolutionary genetics analysis software. Bioinformatics **17:** 1244-1245.

LECHELT, C., T. PETERSON, A. LAIRD, J. CHEN, S. L. DELLAPORTA *et al.,* 1989 Isolation and molecular analysis of the maize P locus. Mol Gen Genet **219:** 225-234.

LI, Y., J. P. BERNOT, C. ILLINGWORTH, W. LISON, K. M. BERNOT *et al.,* 2001 Gene Conversion Within Regulatory Sequences Generates Maize r Alleles With Altered Gene Expression. Genetics **159:** 1727-1740.

LUDWIG, M. Z., 2002 Functional evolution of noncoding DNA. Curr Opin Genet Dev **12:** 634-639.

LUDWIG, M. Z., C. BERGMAN, N. H. PATEL and M. KREITMAN, 2000 Evidence for stabilizing selection in a eukaryotic enhancer element. Nature **403:** 564-567.

LUDWIG, S. R., and S. R. WESSLER, 1990 Maize R gene family: tissue-specific helix-loop-helix proteins. Cell **62:** 849-851.

MCLYSAGHT, A., K. HOKAMP and K. H. WOLFE, 2002 Extensive genomic duplication during early chordate evolution. nature genetics **31:** 200-204.

MORENO, M. A., J. CHEN, I. GREENBLATT and S. L. DELLAPORTA, 1992 Reconstitutional mutagenesis of the maize P gene by short-range Ac transpositions. Genetics **131:** 939-956.

OHTA, T., 1995 Gene conversion vs point mutation in generating variability at the antigen recognition site of major histocompatibility complex loci. J Mol Evol **41:** 115-119.

PATTERSON, G. I., K. M. KUBO, T. SHROYER and V. L. CHANDLER, 1995 Sequences Required for Paramutation of the Maize b Gene Map to a Region Containing the Promoter and Upstream Sequences. Genetics **140:** 1389-1406.

PROCISSI, A., S. DOLFINI, A. RONCHI and C. TONELLI, 1997 Light-Dependent Spatial and Temporal Expression of Pigment Regulatory Genes in Developing Maize Seeds. Plant Cell **9:** 1547-1557.

PURUGGANAN, M. D., 2000 The molecular population genetics of regulatory genes. Mol Ecol **9:** 1451-1461.

RALSTON, E. J., J. J. ENGLISH and H. K. DOONER, 1988 Sequence of Three bronze Alleles of Maize and Correlation With the Genetic Fine Structure. Genetics **119:** 185-197.

ROZAS, J., and R. ROZAS, 1999 DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. Bioinformatics **15:** 174-175.

SAGHAI-MAROOF, M. A., K. M. SOLIMAN, R. A. JORGENSEN and R. W. ALLARD, 1984 Ribosomal DNA spacer-length polymorphisms in barley: mendelian inheritance, chromosomal location, and population dynamics. Proc Natl Acad Sci U S A **81:** 8014-8018.

SELINGER, D. A., D. LISCH and V. L. CHANDLER, 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:** 1125-1138.

SIDORENKO, L. V., X. LI, S. M. COCCIOLONE, S. CHOPRA, L. TAGLIANI *et al.,* 2000 Complex structure of a maize Myb gene promoter: functional analysis in transgenic plants. Plant J **22:** 471-482.

WANG, R. L., A. STEC, J. HEY, L. LUKENS and J. DOEBLEY, 1999 The limits of selection during maize domestication. Nature **398:** 236-239.

WESSLER, S., A. TARPLEY, M. PURUGGANAN, M. SPELL and R. OKAGAKI, 1990 Filler DNA is Associated with Spontaneous Deletions in Maize. PNAS **87:** 8731-8735.

WRAY, G. A., M. W. HAHN, E. ABOUHEIF, J. P. BALHOFF, M. PIZER *et al.,* 2003 The Evolution of Transcriptional Regulation in Eukaryotes. Mol Biol Evol **20:** 1377-1419.

YAO, H., Q. ZHOU, J. LI, H. SMITH, M. YANDEAU *et al.,* 2002 From the Cover: Molecular characterization of meiotic recombination across the 140-kb multigenic a1-sh2 interval of maize. PNAS **99:** 6157-6162.

YU, J., J. WANG, W. LIN, S. LI, H. LI *et al.,* 2005 The Genomes of Oryza sativa: A History of Duplications. PLoS Biology **3:** e38.

ZHANG, F., and T. PETERSON, 2005 Comparisons of Maize pericarp color1 Alleles Reveal Paralogous Gene Recombination and an Organ-Specific Enhancer Region. Plant Cell **17:** 903-914.

ZHANG, P., S. CHOPRA and T. PETERSON, 2000 A segmental gene duplication generated differentially expressed myb-homologous genes in maize. Plant Cell **12:** 2311-2322.

Figure legends

Figure 1. — Structural comparisons of the simplex *p1* alleles. The black boxes connected by thin lines represent exon regions. The open boxes with black arrows indicate long direct repeats in 5' and 3' non-coding regions. The bent arrow represents transcription start site. The hatched boxes represent 5' UTR and 90 bp promoter regions conserved among the *p1* alleles. The open boxes labeled with C indicate a 1.1 kb region, termed as Fragment C, present at the 5' direct repeats of all *p1* haplotypes, but deleted in 3' direct repeats in several *p1* alleles. The numbered open boxes represent the sequences, termed as Fragment 15 and Fragment 14. The distal enhancer regions in each *p1* alleles are highlighted by black bars, which include Fragment 15 as well as Fragment 14 in many cases. The cob glume-specific regulatory region is indicated as gray boxes in the distal enhancers of *P1-rr* haplotypes. The regions marked 15* indicate the partial Fragment 15 sequence, which contains only the first 200 bp Fragment 15 sequences. The triangles located in the first copy of Fragment 15 in the 5' non-coding repeats of *P1-rr4B2* and *P1-rrCFS36* represent insertions of 1.6 kb transposon-like sequence. The vertical wavy lines indicate 549 bp and 148 bp deletions in the non-coding repeats of *P1-rwCFS342*. The black arrows represent primers used to amplify 5' and 3' non-coding regions of *p1*: 1, P1rr-18; 2, PA-B4; 3, P1rr-14; 4, PA-B6; 5, EP5-16; 6, P1rr-16; 7, P1rr-13r; 8, P1rr-30; 9, P1rr-16r; 10 P1rr-29.

Figure 2. — Deletion at the 3' non-coding region of *p1*. Polymorphic regions at the 3' non-coding regions between *P1-rwCFS342* (as well as all *P1-rr* alleles) and *P1-rwCFS302* (as well as *P1-rw1077*) were aligned. The numbered open boxes are as shown in Figure 1. The 1.4 kb region, including Fragment C and 15*, in *P1-rwCFS302* was deleted and replaced

with 13 bp filler DNA in *P1-rwCFS342*. The homologous region of the filler DNA, 51 bp downstream of the deletion breakpoint, was indicated in a triangle in *P1-rwCFS302*. The 4-bp microhomology sequence CCAA in the deletion breakpoint and the filler DNA region is underlined.

Figure 3. — Phylogenetic analysis of the *p1* distal enhancer regions. An unrooted neighbor-join tree was generated by using the 602 bp sequence, which includes Fragment 15 plus 180 bp downstream sequence. The numbers at nodes represent bootstrap values derived from 1000 replicates. Sequences from the 5' and 3' non-coding repeats are distinguished by labelling with upstream or downstream. Sequences from the same repeat are indicated in numeric order from 5' to 3'.

Figure 4. — Sliding-window analysis of non-coding regions at the *p1* locus. The open boxes with black arrows represent the 5' and 3' non-coding direct repeats. The position of Fragment 15 is indicated with the numbered box. The black boxes connected with thin lines represent the exon and intron structure of the *p1* locus. The triangle in the second intron indicates the 723 bp sequence that is also subject to nucleotide diversity analysis. The black arrows represent the primers for amplification of the 723 bp region: 1, 723-5; 2, 723-3.

Figure 5. — Sequential model for *p1* evolution. The stepwise progression from *p* progenitor gene to the distinct *p1* alleles is shown. The open boxes with black arrows indicated the non-coding direct repeats flanking the *p1* coding sequence. The numbered open boxes are the same as those shown in Figure 1. The black boxes indicate exons of the *p1* and *p2* genes.

The vertical arrows between the *p1* and *p2* regions represent retro-element insertions that separated the *p1* gene from *p2* (ZHANG *et al.* 2000). In each step, the rearranged regions are highlighted by black bars. The black arrowheads in the progenitor gene of *P1-rwCFS342* indicate the 8 bp microhomology sequence (TCGATGCC), which could be involved in the gene conversion in generating *P1-rr* alleles. The divergence time between the *p1* alleles is indicated at the branch points (mya for millions year ago).

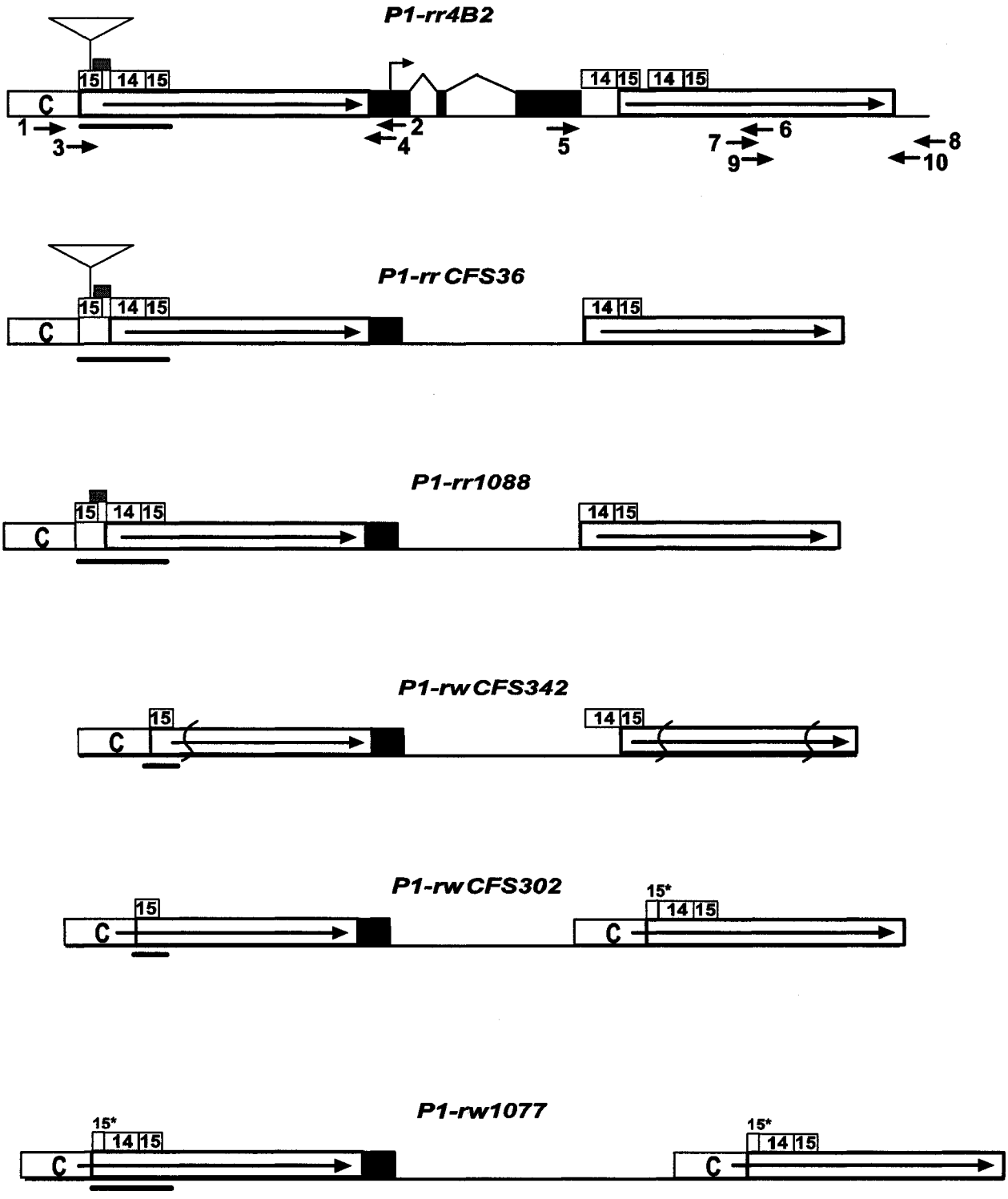

P1-rr4B2
15
14 15
C
14 15
14 15
1
3
2
4
5
7
9
6
8
10
P1-rr CFS36
15
14 15
C
14 15
P1-rr1088
15
14 15
C
14 15
P1-rwCFS342
15
C
14 15
P1-rwCFS302
15
C
15*
14 15
C
P1-rw1077
15*
14 15
C
15*
14 15
C
1 kb

Figure 1

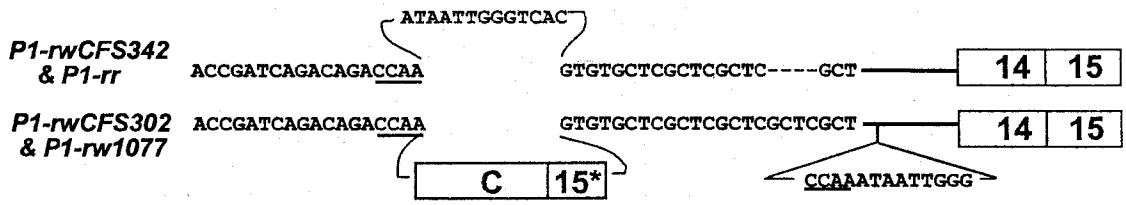

ATAATTGGGTCAC
P1-rwCFS342
& P1-rr
ACCGATCAGACAGACCAA
GTGTGCTCGCTCGCTC----GCT
14
15
P1-rwCFS302
& P1-rw1077
ACCGATCAGACAGACCAA
GTGTGCTCGCTCGCTCGCTCGCT
14
15
C
15*
CCAAATAATTGGG

Figure 2

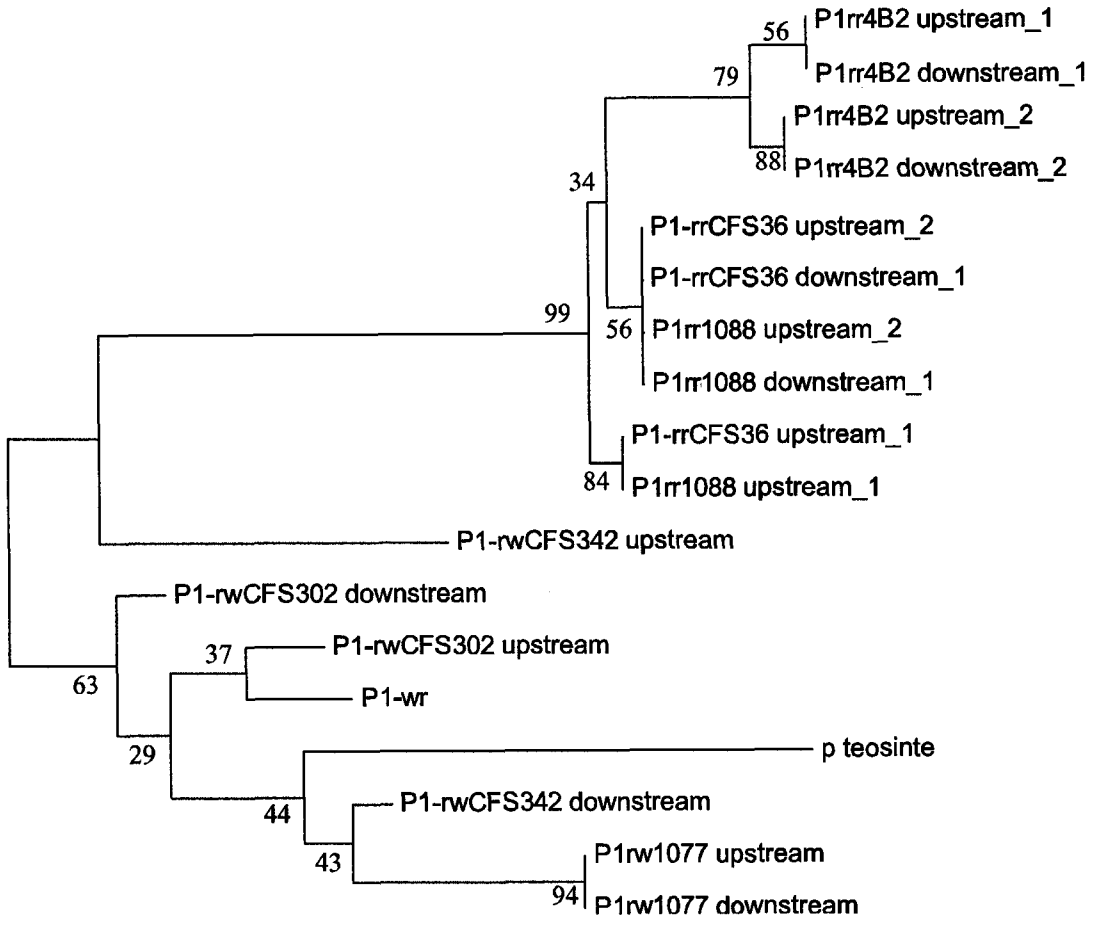

P1rr4B2 upstream_1
P1rr4B2 downstream_1
P1rr4B2 upstream_2
P1rr4B2 downstream_2
P1-rrCFS36 upstream_2
P1-rrCFS36 downstream_1
P1rr1088 upstream_2
P1rr1088 downstream_1
P1-rrCFS36 upstream_1
P1rr1088 upstream_1
P1-rwCFS342 upstream
P1-rwCFS302 downstream
P1-rwCFS302 upstream
P1-wr
p teosinte
P1-rwCFS342 downstream
P1rw1077 upstream
P1rw1077 downstream
56
79
88
34
99
56
84
37
63
29
44
43
94

Figure 3

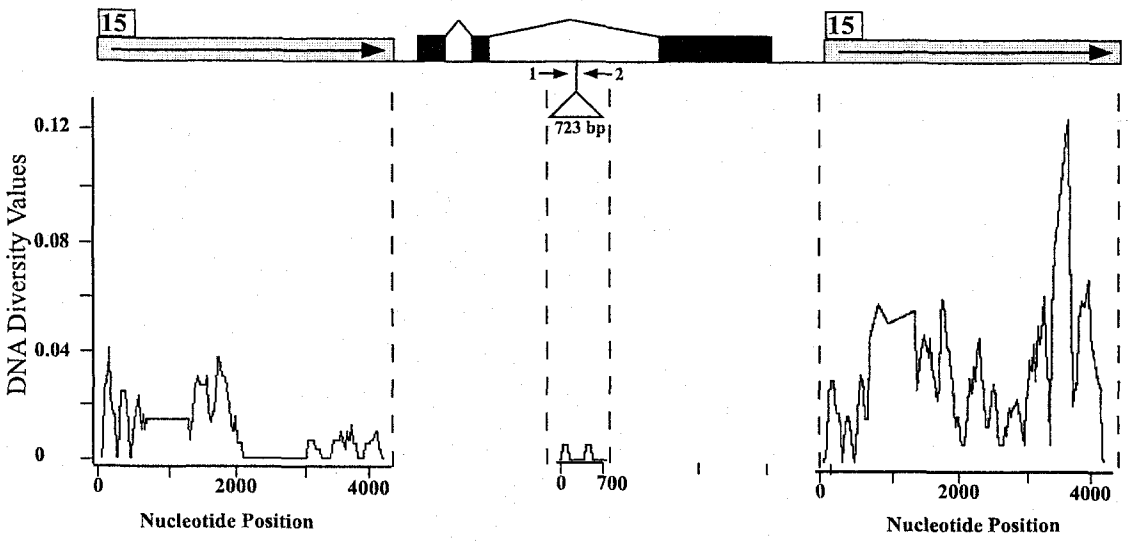

15
15
1
2
723 bp
0.12
0.08
0.04
0
DNA Diversity Values
0
2000
4000
Nucleotide Position
0
700
0
2000
4000
Nucleotide Position

Figure 4

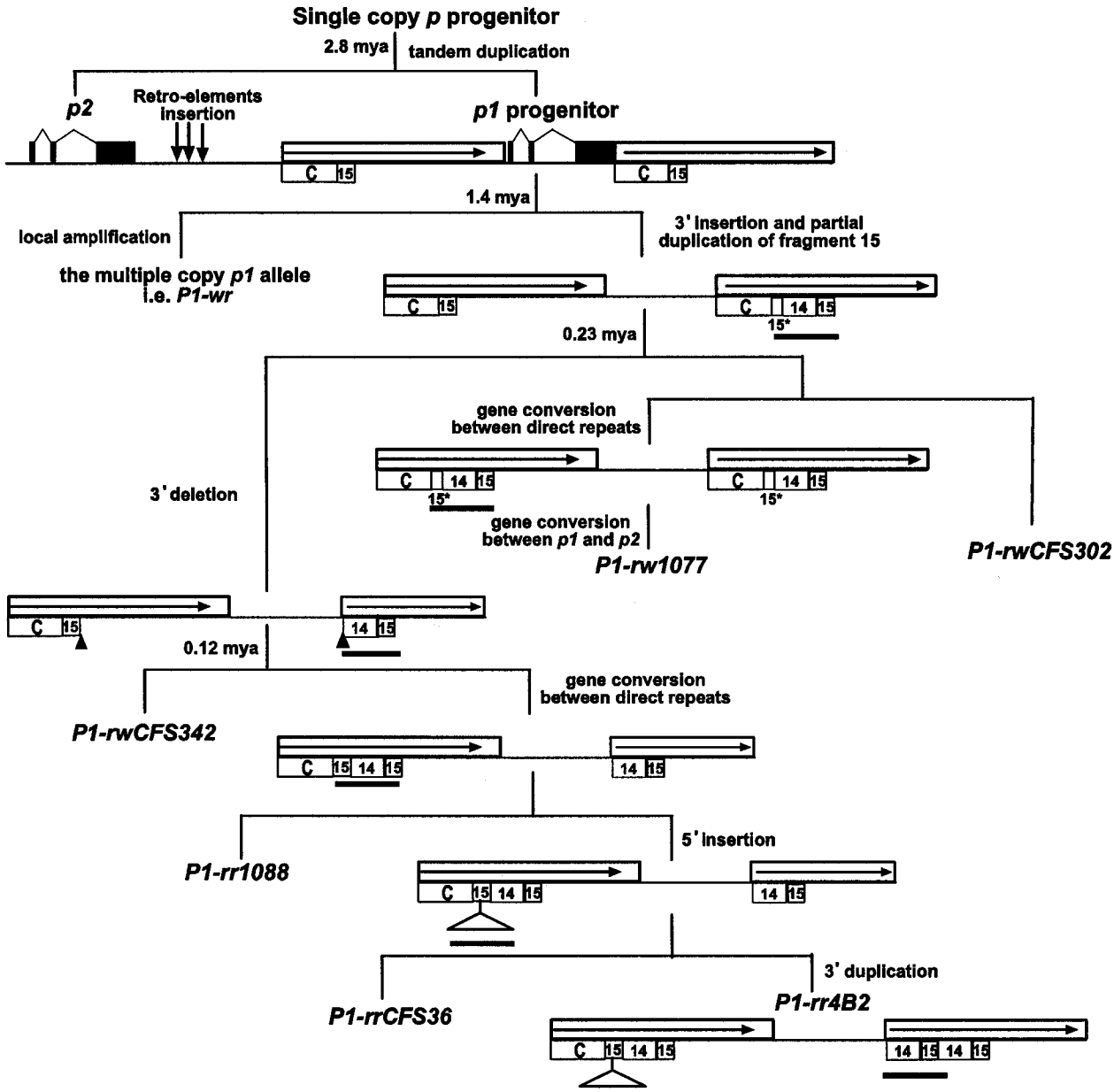

Single copy p progenitor
2.8 mya
tandem duplication
p2
Retro-elements insertion
p1 progenitor
C 15
C 15
1.4 mya
local amplification
3' insertion and partial duplication of fragment 15
the multiple copy p1 allele i.e. P1-wr
C 15
C 14 15
15*
0.23 mya
gene conversion between direct repeats
3' deletion
C 14 15
15*
C 14 15
15*
gene conversion between p1 and p2
P1-rw1077
P1-rwCFS302
C 15
14 15
0.12 mya
gene conversion between direct repeats
P1-rwCFS342
C 15 14 15
14 15
5' insertion
P1-rr1088
C 15 14 15
14 15
3' duplication
P1-rrCFS36
P1-rr4B2
C 15 14 15
14 15 14 15

Figure 5

Table 1

Collections of single-copy *p1* alleles

Allele name	Haplotype	Phenotype	Source
P1-rr4B2	*P1-rr4B2*	red pericarp; red cob	Derivative from *P1-vv* [a]
P1-rrCFS181	*P1-rr4B2*	red pericarp; red cob	Brink's collection [b]
P1-rr13:255 A10	*P1-rr4B2*	red pericarp; red cob	From *P1-vv*
P1-rr1088	*P1-rr1088*	red pericarp; red cob	To be determined
P1-rrCFS36	*P1-rrCFS36*	red pericarp; red cob	Brink's collection
P1-rrCFS33	*P1-rrCFS36*	red pericarp; red cob	Brink's collection
P1-rrCFS305	*P1-rrCFS36*	red pericarp; red cob	Brink's collection
P1-rrCFS548	*P1-rrCFS36*	red pericarp; red cob	Brink's collection
P1-rrCFS272	*P1-rrCFS36*	red pericarp; red cob	Brink's collection
P1-rw1077	*P1-rw1077*	red pericarp; white cob	Maize Genetic Coop
P1-rwCFS325	*P1-rw1077*	red pericarp; white cob	Brink's collection
P1-rwCFS302	*P1-rwCFS302*	red pericarp; white cob	Brink's collection
P1-rwCFS332	*P1-rwCFS302*	red pericarp; white cob	Brink's collection
P1-rwCFS342	*P1-rwCFS342*	red pericarp; white cob	Brink's collection
P1-rwCFS334	*P1-rwCFS342*	red pericarp; white cob	Brink's collection

a. (LECHELT *et al.* 1989);
b. (BRINK and STYLES 1966).

Table 2

Nucleotide diversity in the simplex *p1* alleles *

Regions	n	No. of Silent sites	No. of polymorphic sites	Θ	π	Tajima's D
Upstream repeat	6	3526	83	0.01031	0.00906	-0.7840^{NS}
Downstream repeat	6	3278	197	0.02483	0.02929	-1.1599^{NS}
723 bp sequence	5	723	2	0.00133	0.00111	-0.97256^{NS}

Table 3.

Nucleotide substitution rates between the *p1* alleles

	P1-wr	*P1-rw CFS302*	*P1-rw CFS342*	*P1-rw 1077*	*P1-rr CFS36*	*P1-rr 1088*	*P1-rr 4B2*
P1-wr	—						
P1-rw CFS302	0.0166 (12)	—					
P1-rw CFS342	0.0166 (12)	0.0028 (2)	—				
P1-rw 1077	N.A.	N.A.	N.A.	—			
P1-rr CFS36	0.0152 (11)	0.0014 (1)	0.0152 (11)	N.A.	—		
P1-rr 1088	0.0152 (11)	0.0014 (1)	0.0014 (1)	N.A.	0 (0)	—	
P1-rr 4B2	0.0152 (11)	0.0014 (1)	0.0014 (1)	N.A.	0 (0)	0 (0)	—

Numbers in parentheses are the numbers of nucleotide substitutions over 723 silent sites.

Table 4
Primers used to amply the *p1* sequences

Primer name	Primer sequence
P1rr-18	TGAGTCCTGACCGACAGTCT
PA-B4	TGCCTTCCATACTTGCACTGC
P1rr-14	GAAGGCAGACGATGAGGAGA
PA-B6	CACAACCTTTCACATACAGAG
EP5-16	CGAGACTTGGCTCCTGT
P1rr-16	TCTCAGAGTATAGCAACAC
P1rr-13r	CTCATCAACGTGCTGTTCC
P1rr-30	CGTCGTCAAGAACTCAAGAT
P1rr-16r	GTGTTGCTATACTCTGAGA
P1rr-29	GGCTTGGTGCGTTGCTGA
723-5	TCTAGGCACTTTCTCGTG
723-3	GTAGAAATAAAGTCTGAGCA

The orientation and approximate position of these primers were shown in Figure 1 and Figure 4

CHAPTER 4. GENERAL CONCLUSIONS

As one of the well-defined regulatory loci that control flavonoid pigments biosynthetic pathways in maize, the *pericarp color1* (*p1*) gene regulates phlobaphene synthesis in the floral organs. The striking phenotypes conferred by the *p1* gene in maize kernel pericarp and cob glumes have made it an excellent genetic system to study transposable elements, developmental biology, population and evolution genetics, and gene regulation in plants. More than 100 distinct *p1* alleles have been collected and classified according to the pigmentation patterns in kernel pericarp and cob glumes. Investigation of such a rich collection of natural *p1* variations will provide valuable insights in mechanisms of gene regulation and evolution of genetic and phenotypic diversity.

Multiple levels of regulation for diversity in *p1* gene expression patterns

How is the high degree of phenotypic variability in *p1* expression achieved? A number of previous studies have indicated that a wide variety of *p1* expression patterns could be attributed to epigenetic regulation rather than DNA sequence polymorphism. Das et al (1994) reported that the *P1-rr4B2* allele can spontaneously change to an epiallele termed *P1-pr* (patterned pericarp and red cob), that confers variegated or nearly colorless kernel pericarp, and reduced cob pigmentation. Compared with the progenitor allele, *P1-pr* exhibits a hypermethylated state and condensed chromatin structure in the 1.2 kb distal enhancer region (Das and Messing, 1994; Lund G et al., 1995). *P1-rr* can also be suppressed by paramutation induced by transgene constructs carrying the 1.2 kb distal enhancer (Sidorenko and Peterson, 2001), and this suppressed state of *P1-rr* is also correlated with increased DNA methylation. Additionally, *p1* transgenic plants containing identical integrated transgene structures exhibit

a variety of tissue-specific expression patterns and pigmentation intensities in pericarp, cob, and other floral organs, irrespective of whether the promoter and coding sequences were derived from *P1-rr* or *P1-wr* (Cocciolone et al., 2001). The *p1* expression levels of the transgenic plants were also inversely correlated with DNA methylation of *p1*. Moreover, the *P1-wr* allele can be ectopically activated in the presence of *Ufo1* (*Unstable factor for orange1*), leading to red pigmentation in kernel pericarp and vegetative tissues. Increased pigmentation in *P1-wr Ufo1* plants is associated with increased *p1* transcript levels and demethylation in the *P1-wr* complex (Chopra et al., 2003).

The examples above indicate that epigenetic regulation can modulate the pattern and intensity of *p1* expression. Other studies indicate that the timing of expression during development may also be a determinant of organ-specific pigmentation. Specifically, distinct *p1* pigmentation patterns may result from the activation of *p1* expression at different times during the development of floral organs (Cocciolone et al., 2000). It was proposed that *p1* gene expression is induced during the transition from vegetative to floral growth, and tends to remain active during subsequent developmental stages. Changes in the timing of activation could then result in different patterns of expression due to the timing of organ development. In addition, the progressive accumulation of *p1* transcripts during early stage of ear development in both *P1-rr4B2* and *P1-rw1077* could also reflect a temporal transition in chromatin structure during development (Chapter 2).

Our analysis of *P1-rw* demonstrates that cis-element regulated transcriptional control can also contribute to diversity in *p1* gene expression (Chapter 2). It is possible that the *p1* promoter has a modular composition, i.e. there could be discrete regulatory modules for *p1* expression in one or several floral organs, i.e. kernel pericarp, cob glumes, tassel glumes,

husk and silk; mutations in each of these modules could produce distinct tissue-specific expression patterns. In addition, the accessibility of these regulatory elements could be controlled by epigenetic regulation, such as alterations in chromatin structure and DNA methylation (Li et al., 2002). For the *P1-rr4B2* allele, it has been reported that the chromatin structure within the 1.2 kb distal enhancer region is less condensed in *p1*-expressing than in non-expressing tissues (Lund et al., 1995). If the condensed chromatin structure and/or hypermethylated state covers certain tissue-specific regulatory elements but not others, this could result in tissue-specific *p1* expression patterns. For example, in *P1-pr* and *P1-wr*, it is possible that the pericarp-specific regulatory elements could be kept from the binding of transcription factors by a condensed chromatin structure and DNA methylation, while cob glumes-specific regulatory elements remain accessible.

Thus, the diversity in *p1* expression could arise from the interactions of various control levels, including tissue-specific regulatory elements as described here, developmental transitions in chromatin structure, and overlying epigenetic controls. The multiple-level regulation of *p1* expression results in a wide range of spatial and temporal phenotypic diversity observed in *p1* alleles. Recent studies have shown that allelic expression variation occurs commonly in both mammals and plants (Cowles et al., 2002; Yan et al., 2002; Guo et al., 2004). It has been suggested that regulatory genetic variations play important roles in modulating allele-specific gene expression (Knight, 2004). Our study demonstrates that such regulatory variations could result from a combination of mechanisms acting at distinct levels of gene expression, and giving rise to a great number of allele-specific expression variants.

Paralogous recombination promotes genetic diversity at the *p1* locus

In addition to a high degree of phenotypes variations, the *p1* locus also exhibits extensive genetic diversity with distinct gene structures. The *p1* alleles can be roughly classified into two distinct groups: multiple copy *p1* alleles and single copy *p1* alleles. Investigation of the simplex alleles revealed numerous DNA rearrangements in the *p1* locus, especially in the non-coding regions. Our results also suggested a correlation between sequence polymorphisms in the distal enhancer region and *p1* expression in cob glumes. Based on structural and sequence analyses, we proposed a model to account for generation of genetic diversity observed at the *p1* locus. In this model, recombination between duplicated repeats plays an important role in diversifying the *p1* alleles. Such recombination events could have occurred between the *p1* and *p2* paralogous genes as shown in the *P1-rw1077* allele (Chapter 2), and also occurred between the duplicated direct non-coding repeats. The direct non-coding repeats resulted from the tandem duplication that generated *p1* and *p2* paralogs. The direct repeats with the close proximity could serve as excellent substrates for homologous recombination. Unequal crossover between the direct repeats could have given rise to the multiple-copy *p1* alleles, while gene conversion between them could have increased variability at the cis-regulatory regions. Both of these events have provided materials for creating phenotypic variations. Thus, our studies suggested that the homologous recombination could be the driving force for generation of both genetic and phenotypic diversity at the *p1* locus.

In the past decade, complete genomes of numerous organisms have been sequenced and analyzed in detail. One common observation in various genomes, from single cell bacteria to mammals, is the extensive amount of gene duplication (Sankoff, 2001). Many duplicated genes are shown to cluster together forming a tandem array, which resembles the

p1 and *p2* duplication or the *P1-wr* complex. Homologous recombination between the duplicated genes has been documented in various species (see discussion in Chapter 3). Many studies have suggested that homologous recombination in the coding regions of paralogous genes contributes to functional divergence of duplicated genes (Leister, 2004). Our investigation, however, provides an example that recombination between duplicated non-coding sequences could generate novel cis-regulatory regions and result in changes in gene expression patterns. Due to the prevalence of gene duplication, this mechanism might be an additional important means to create evolutionary novelty.

References

Chopra, S., Cocciolone, S.M., Bushman, S., Sangar, V., McMullen, M.D., and 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.

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, 467-478.

Cowles, C.R., Hirschhorn, J.N., Altshuler, D., and Lander, E.S. (2002). Detection of regulatory variation in mouse genes. Nat Genet 32, 432-437.

Das, O.P., and Messing, J. (1994). Variegated phenotype and developmental methylation changes of a maize allele originating from epimutation. Genetics 136, 1121-1141.

Guo, M., Rupe, M.A., Zinselmeier, C., Habben, J., Bowen, B.A., and Smith, O.S. (2004). Allelic Variation of Gene Expression in Maize Hybrids. Plant Cell 16, 1707-1716.

Knight, J.C. (2004). Allele-specific gene expression uncovered. Trends Genet 20, 113-116.

Leister, D. (2004). Tandem and segmental gene duplication and recombination in the evolution of plant disease resistance gene. Trends Genet 20, 116-122.

Li, G., Hall, T.C., and Holmes-Davis, R. (2002). Plant chromatin: development and gene control. Bioessays 24, 234-243.

Lund G, Prem D, and J., M. (1995). Tissue-specific DNase I-sensitive sites of the maize P gene and their changes upon epimutation. Plant J 7, 797-807.

Sankoff, D. (2001). Gene and genome duplication. Curr Opin Genet Dev 11, 681-684.

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, 319-335.

Yan, H., Yuan, W., Velculescu, V.E., Vogelstein, B., and Kinzler, K.W. (2002). Allelic Variation in Human Gene Expression. Science 297, 1143.

ACKNOWLEDGEMENTS

First of all, I would like to thank my major professor Dr. Thomas Peterson for his excellent guidance, generous support in my research work. Throughout my doctoral work he encouraged me to develop independent thinking and research skills. He continually stimulated my analytical thinking and greatly assisted me with scientific writing.

I am also very grateful for having an excellent doctoral committee and wish to thank Dr. Dan Voytas, Dr. Diane Bassham, Dr. Phil Becraft and Dr. Steven Whitham for their generous time, and continuos support.

I wish to extend my thanks to former and current member in the P team: Dr. Erica Unger-Wallace, Dr. Jianbo Zhang, Dr. Yongli Xiao, Dr. Surinda Chopra, Dr. Lyuda Sidorenko, Dr. Suzy Cocciolone, Cizhong Jiang, Zhuying Li, Terry Olson, Diane Sickau, Yibin Wang, Lakshminarasimhan Krishnaswamy, and Tanya Rogers for their advice and help on my research. They have brought fun and friendship into my everyday life.

Finally, I wish to thank my family. I am grateful to my mother, father and younger brother for their love, support and encouragement. I am especially thankful to my wife, Jing Zhang, for her love, understanding and selfless support. She has been helping me keep my life in proper perspective and balance. I could not accomplish my study without her.