Comparisons of Maize pericarp color1 Alleles Reveal Paralogous Gene Recombination and an Organ-Specific Enhancer Region

Feng Zhang*a,b and Thomas Petersona,b,1

*Department of Genetics, Development, and Cell Biology, Iowa State University, Ames, Iowa 50010
b Department of Agronomy, Iowa State University, Ames, Iowa 50010

The maize (Zea mays) p1 (for pericarp color1) 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-rw1077 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, whereas 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.

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 several 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 (Zea mays), the p1 (for pericarp color1) 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 (for chalcone synthase), chi (for chalcone flavonone isomerase), and a1 (for dihydroflavonol reductase) (Grotewold et al., 1994). More than 100 p1 alleles have been described, each of which specifies a distinct pigmentation pattern (Brinks and Styles, 1966; Cocciolone et al., 2001). Prototype alleles conferring distinct pigmentation patterns of pericarp and cob glumes include P1-rr (for red pericarp/red cob), P1-wr (for white pericarp/red cob), P1-rw (for red pericarp/white cob), and P1-ww (for 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, 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, Unstable factor for orange1 (Ufo1), which results in red pericarp color 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 the 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...
alleles are dominant to a p1-ww allele (Anderson, 1924). However, in contrast with the relatively even pigmentation of P1-rr4B2 pericarp, the pigmentation of P1-rw1077 pericarp is concentrated on the gown region (the sides of the kernel), whereas the crown region (the top of the 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 ~10 to 12 d after pollination (DAP) and is first seen at the silk attachment point. Subsequently, the pigment spreads over the crown region and progresses through the gown region toward the base of the kernel. At 20 to 24 DAP, most of the pericarp is red. During maturation and desiccation, the red color becomes darker (Zhang, 1999). By 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 (Figures 1A and 1B).

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 that 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 after pollination. The blot was sequentially hybridized with probes from the maize p1 allele, or insertion of an Ac transposable element in this 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-rw1077 allele, 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
were detected from pericarp of P1-rw1077 and P1-rr4B2 at different developmental stages. Samples were taken from p1-ww1112 and P1-rr4B2 during Development.

The expected 450-bp PCR products from cDNAs were subjected to nested PCR amplification (see Methods). The total RNA was loaded in each lane. Hybridization probes are indicated at kernel (8 DAP), and kernel pericarp (12 to 21 DAP). Ten micrograms of whole ears (0 DAP), developing kernels and cob (2 to 6 DAP), whole kernel (8 DAP), and kernel pericarp (12 to 21 DAP). Ten micrograms of 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 the RNA gel blot in (A) are normalized to actin to calculate relative transcript levels (y axis).

P1-rw1077 is a Single-Copy Gene Generated by Paralogous Gene Recombination

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 subjected to nested PCR amplification (see Methods). The expected 450-bp PCR products from p1 transcripts were detected from pericarp of P1-rr4B2 and P1-rw1077 as well as from cob glumes of P1-rw1077, 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 blot 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 Methods). Sequence analyses showed that the 5’ region of P1-rw1077 transcripts (606 bp in length, extending from the 5’ untranslated region [UTR] into exon 3; see Methods) is nearly identical to that of P1-r4B2 except for two single-nucleotide changes in exon 1 and three nucleotide changes in exon 3. The 3’ end (320 bp in length) is identical to that of p2, a tightly linked paralog of p1 (Zhang et al., 2000). Similar to the p2 gene in the P1-rr4B2 stock, p2 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, as in P1-r4B2, p2 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 results (Zhang et al., 2000)]. Thus, the 3’ RT-PCR product cannot be derived from p2. 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).

**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 to 6 DAP), whole kernel (8 DAP), and kernel pericarp (12 to 21 DAP). Ten micrograms of 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 the 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.**

Nestled RT-PCR was performed on total RNA extracted from pericarp and cob glumes of P1-rr4B2 and P1-rw1077 at 16 DAP. The top panel shows ethidium bromide–stained PCR products (570 bp) amplified with p1–specific primers (EP5-8-1 and ZFRT-8). The middle panel shows the results after blotting of the agarose gel and hybridization with the p1–specific probe. The bottom panel shows the same samples amplified with α-tubulin–specific primers as positive controls.
Genomic DNAs of tandem repeat of the 6-kb band. Several restriction fragment length polymorphisms are apparent. (lane 3), and digested with contains only the 6-kb band, whereas P1-rw1077 (lane 2), P1-rr4B2 (lane 3), and P1-wr (lane 4) alleles contain both the 3.7-kb band and the 6-kb p2 band. In P1-wr, the high-intensity p1 band results from the tandem repeat of p1 sequence (Chopra et al., 1998).

Figure 4. Genomic DNA Gel Blot Analyses of Different Alleles.

(A) Genomic DNAs of individual p1 alleles were digested with XbaI and hybridized with p1 fragment 8B (Figure 5). The p1-ww1112 (lane 1) allele contains only the 6-kb p2 band, whereas P1-rw1077 (lane 2), P1-rr4B2 (lane 3), and P1-wr (lane 4) alleles contain both the 3.7-kb p1 band and the 6-kb p2 band. 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 SalI and EcoRI and probed with p1 fragment 15 (Figure 5). Several restriction fragment length polymorphisms are apparent.

p1-ww1112 allele lacks the 3.7-kb p1 band but has the 6-kb band corresponding to p2. By 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 sequence similar to P1-rr4B2. However, additional DNA gel blot analyses revealed several polymorphisms between P1-rw1077 and P1-rr4B2. For example, genomic DNAs from P1-rw1077 and P1-rr4B2 were digested by EcoRI and SalI and hybridized with a p1-specific genomic fragment (fragment 15) as a probe, which is repeated four 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 with 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 p2 (Zhang et al., 2003). The second class hybridizes with both probes 15 and 8B and has the 80-bp sequence in the 5’ UTR sequence; thus, it appears to contain the p1 gene. Two overlapping p1-carrying λ clones (covering 22,270-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 kb] direct repeats). However, the P1-rw1077 coding sequence is chimeric, consisting of a p1-like 5’ UTR followed by p2-like exons and introns. Farther 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 p2, 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.

By comparing the sequences of the P1-rw1077, P1-rr4B2, and P1-wr alleles and the p2 gene, the positions of the recombination break points can be inferred. The 5’ break point is located in the first exon, in the interval between positions −8 and +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 microhomology 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) or 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 nucleotides) was de-duced 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 11 (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
Figure 5. Gene Structure Comparisons of P1-rr4B2, P1-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). The bent arrow indicates the transcription start site in P1-rr4B2 (Grotewold et al., 1991). The hatched boxes represent 5’ UTRs and 90-bp promoter regions conserved among p1 alleles and the p2 gene (Zhang et al., 2000). The positions of 5’ and 3’ break points in the P1-rw1077 chimeric structure are indicated by offset lines. Sequences sharing homology with fragments 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 after 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 upstream SacI sites (dashed line) and downstream SalI sites (solid lines). The regions marked 15” indicate the partial fragment 15 sequences in P1-rw1077, which contain only the first 200 bp of the fragment 15 sequences. The gray bar between the PsI and SacI 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 clones. Restriction sites are as follows: E, EcoRI; S, SacI; Sl, SalI. Not all restriction sites are indicated. The black arrowheads indicate primers that were used to isolate the 5’ and 3’ ends of P1-rw1077 transcripts: 1, EP5-8; 2, ZFRT-8; 3, EP5-16; 4, AP. The GenBank accession number of P1-rw1077 is AY702552.

Single-nucleotide and 16 small 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, 2000; Sidorenko and Peterson, 2001), we wanted 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). Several 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 a 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 (Figure 8A). DNA gel blot 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 (lanes 4 and 5), DNA from P1-rr4B2 plants digested with SalI and hybridized with fragment 15 produces four bands at 3.4, 3, and 1.2 kb (doublet), whereas in P1-rw plants, the 3.4- and 3-kb bands shift to 7.9 and 7.5 kb, respectively. The size of Ac elements is ~4.5 kb, and there is no SalI 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 SalI fragment and a second in the 3-kb SalI fragment. Additional DNA gel blot and PCR analyses showed that the Ac element in the 3.4-kb SalI fragment is at the same location and orientation as the donor Ac element in the P1-ovov1114 allele (see Methods for details).

The approximate position of Ac in the 3.0-kb SalI fragment was estimated from the EcoRI digestion pattern (Figure 8B). The precise insertion site and orientation of the Ac element in the 3-kb
A 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 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 SalI 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 the insertion of two Ac elements? An example of the latter effect was described by English and Jones (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 was excised (see 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

Figure 6. Localization of p1 and p2 Recombination Break Points in P1-rw1077.

(A) The 5' region of P1-rw1077 was aligned with P1-rw2B, P1-wr, and p2. The sequence in boldface indicates the exon 1 open reading frame, and the first base of the start codon is indicated as position +1. The triangle at position +60 and +138 indicate nucleotide polymorphisms in P1-rw1077 that match the p2 sequence. Thus, the 5' break point of p2-p1 crossover lies between the 80-bp insertion at +60 bp and the polymorphism at +138 bp.

(B) The 3' region of P1-rw1077 was aligned with p1 and p2 sequences. The exon 3 regions and Jh-homologous sequences in p1 and p2 are shown as in Figure 5. The 4-bp microhomology sequence GCCG at the break point is underlined.

Figure 7. Alignment of the Distal Enhancer Region in P1-rw2B and P1-rw1077.

DNA sequences of P1-rw4B2 and P1-rw1077 were aligned at the distal enhancer region (between the two SacI sites shown in Figure 5). Restriction sites are shaded; the sequence between the first SacI site and the SalI site corresponds to fragment 15, whereas the sequence after the second SacI site corresponds to fragment 14. The dashed lines represent gaps in the alignment. The black bar above the P1-rw4B2 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-rw4B2, 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 I, II, and III, and the putative RY element.
result was further confirmed by DNA gel blot analysis: in a SalI 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, whereas the 7.5-kb bands remain unchanged. In addition, the P1-rw751::Ac4 and P1-rw751::Ac5 alleles exhibit an ∼25-kb EcoRI band rather than the 14.7-kb doublet bands in P1-rw512A::2Ac, which arise by cutting at the EcoRI sites within Ac elements (Figures 8B and 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). By contrast, both alleles exhibit dramatic reduction in cob glume pigmentation: the cob of P1-rw751::Ac4 is nearly colorless, whereas 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 (F. Zhang, 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 even though they contain identical p1 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 p1 and p2**

Alleles of maize p1 exhibit several diverse, tissue-specific pigmentation patterns. In this study, two independent P1-rw alleles, specifying red pericarp and white cob glumes, were characterized and compared with the previously described P1-rr and P1-wr alleles (Lechelt et al., 1989; Chopra et al., 1996, 1998; 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.
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: p2 is expressed primarily in silk and anther wall, whereas p1 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 as a result of 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 et al., 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: that is, the p1 coding sequence was replaced by the p2 coding sequence, whereas 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 p1 and p2 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 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); by contrast, the P1-rw1077 allele possesses a homologous junction at the 5′ end and a nonhomologous junction at the 3′ end, with a 4-bp microhomology region. This interesting structure can be interpreted as arising from the one-sided invasion 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 one-sided invasion 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 noninvaded end by the nonhomologous end-joining mechanism. This would lead to a recombinant product with a homologous junction at one end and a nonhomologous junction at the other end. Evidence for this mechanism has been observed after the experimental induction of DSBs (Puchta et al., 1996). The P1-rw1077 allele described here is an OSI-type of structure found as a natural allele of a functional plant gene.

The 386-bp 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 promoter has been shown to contain an enhancer within a 1.2-kb sequence that is located ~5 kb upstream of the transcription start site. This 1.2-kb region was initially identified by transposon mutagenesis of P1-r4B2. Insertions of Ac transposable elements in this region lead to reduced p1 expression in both pericarp and cob glumes of P1-r4B2 (Athaia 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, 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 four times in P1-r4B2 (two copies upstream and two copies downstream of the coding sequence) and two times in P1-rw1077 (one copy upstream and one copy 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-r4B2, an Ac element insertion in one of the upstream 386-bp sequences converts RR to a RW-like phenotype, as shown in the P1-rw512A::2Ac, P1-rw751:A4c, and P1-rw751:A5c alleles. By 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 sequence 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-r4B2 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 (F. Zhang and T. Peterson, unpublished data). 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 or from other plants (data not shown). Searches of the PlantCare promoter database (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 the expression of seed-specific genes, such as maize c1, wheat (Triticum aestivum) Em-1, bean (Phaseolus vulgaris) β-phaseolin, and Arabidopsis thaliana 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, whereas many basic domain/leucine zipper (bZIP) proteins bind ACGT motifs (Mikami et al., 1994; Suzuki and McCarty, 1997; Reidt et al., 2000; Siberil et al., 2001). VP1 is reported to interact with 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-rr751:Ac alleles is inserted 4 bp downstream of the ACGT motif (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, 2003; Coccioniolo et al., 2001). The results we present here show that diversity in p1 expression could also arise from DNA sequence changes in a flanking enhancer region. 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 parmutation, an allelic interaction leading to epigenetic silencing (Sidorenko and Peterson, 2001). The colocalization of distal enhancer sequences with sequences required for parmutation has been reported at the maize b1 (for booster1) gene, which is also a regulatory gene in the flavonoid biosynthetic pathway (Stam et al., 2002a, 2002b). The regulatory sequences that are required for both enhancer activity and paramutability are located ~100 kb upstream of the b1 transcription start site. An intriguing question regarding these observations is whether sequences for enhancer function are separable from sequences for parmutation. The dissection of the distal enhancer region in p1 using P1-rr alleles will allow us to address this question and elucidate the mechanisms of long-distance cis and trans gene regulation.

METHODS

Maize Genetic Stocks

The maize (Zea mays) P1-rr4B2 allele (Grotewold et al., 1991) was inbred line 4Co63 background, and the P1-rr allele was from inbred line W23 (Chopra et al., 1996, 1998). The P1-rr77 allele in the 4Co63 background was obtained from the Maize Genetics Cooperation Stock Center (Urbana, IL). The inbred line 4Co63 (p1-ww) was obtained from the National Seed Storage Laboratory (Fort Collins, CO). The alleleism of P1-rr77 and P1-rr was tested by crossing P1-rr with P1-rr77. The F1 plants (P1-rr-like phenotype) were subsequently crossed to p1-ww, 4Co63. Among 74 progeny plants, 40 plants had the P1-rr phenotype and 34 plants had the P1-rr phenotype. This ~1:1 ratio, and the absence of P1-rr and P1-ww phenotypes, indicate that P1-rr and P1-rr77 segregate as alleles in repulsion.

Genomic Library Construction and Screening

Genomic DNA was extracted from silk of P1-rr77 homozygous plants, partially digested with Sau3AI, and ligated to the λ FIX II/XhoI partial fill-in vector (Stratagene, La Jolla, CA). Approximately 1.5 × 10⁵ independent plaques were screened using two probes of P1-rr4B2, fragment 8B and fragment 15 (Figure 5). The fragment 8B probe detects a homologous sequence in p2, whereas the fragment 15 probe is unique to p1. Seven clones hybridized only with fragment 8B but not fragment 15; four clones hybridized with both probes, indicating that they contain p1 sequences.

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 RNasy plant minikit (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′-GGAGAGGACGCCGCGC-3′; homologous with the 5′ UTR of p1) and EP3-13 (5′-AGGAATTCCGCCC-3′; homologous with p1 exon 1) and showed that the seven clones that did not hybridize with fragment 15 resemble p2, which does not contain an 80-bp insertion in the 5′ UTR region (Figure 6). The four p1-like λ clones were subjected to further analyses, including restriction digest and sequencing of ends. Two overlapping clones, which together cover a 22.2-kb region of the p1 locus, were selected, subcloned into pBluescript KS- plasmid vectors, and sequenced with the EZ:TN <KAN-2> Insertion Kit (Epicenter Technologies, Madison, WI) according to the manufacturer’s instructions at the Iowa State University Nucleic Acid Facility (Ames, IA).

RT-PCR

To compare the levels of steady state p1 transcripts in P1-rr4B2 and P1-rr77, 1 μg of total RNA from 16-DAP pericarp and cob glumes of P1-rr4B2 and P1-rr77 were reverse-transcribed using StrataScript Reverse Transcriptase (Stratagene) 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 first round of PCR amplification, 5 μL of cDNA was used with primers EP5-8 and P1-23568 (5′-GGAGAGGACGCCGCGC-3′); in the second round of PCR, the 0.1-μL PCR product from the first amplification was used with primers EP5-8-1 (5′-GGAGAGGACGCCGCGC-3′); spanning the junction of exon 2 and exon 3 of p1 to avoid amplification from genomic DNA. PCR amplification was performed in the same conditions described above, except that the annealing temperature
was 50°C. The PCR products were subjected to 1.5% agarose gel electrophoresis. The same amount of cDNA used in the detection of p1 transcripts was amplified with tubulin primers as a positive control. The sequences of tubulin primers are as follows: tub-1, 5'-AGCCCGATGG-CACCAGCCAGTTGACTTCT-3'; tub-2, 5'-AACACAAAGATCCTGCAAGCGGCTGAC-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) with oligo(dT) at 42°C. Five microliters from the first-strand cDNAs was used in a PCR with primers EP5-8 and P1-23568. To obtain 3' oligo(dT) at 42°C isolated from 16-DAP exon 3) and AP (5'-AACACCGCAGCTTGAGG-3'); homologous with p1 and p2 (exon 3) and AP (5'-GGCCAGCGCTGACTGAGTAC-3'). PCR amplifications were performed using HotStar Taq DNA polymerase (Qiagen) with the following cycle conditions: 95°C for 15 min, followed by 35 cycles of 94°C for 45 s, 58°C for 1 min, and 72°C for 1 min, with a final extension for 8 min at 72°C. The PCR products from both the 5' and 3' ends of P1-rw1077 were cloned into pGEM-T vector (Promega, Madison, WI). Three independent clones from the 5' PCR products and two independent clones from the 3' PCR products were sequenced with the same primer sets used in the RT-PCR at the Iowa State University Nucleic Acid Facility.

**Mapping of Ac Transposable Elements in the P1-rw512A::2Ac Allele and Sequencing of the 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 SalI and EcoRI and hybridized with fragment 15. Comparing the hybridization patterns of the P1-rw512A::2Ac, P1-rwMB2, and P1-ovov1114 alleles allowed the approximate locations of Ac elements to be mapped (Figures 8 and 9).

The orientations and insertion sites of Ac elements in P1-rw512A::2Ac were determined by PCR using primers homologous with Ac and flanking p1 genomic sequences. The p1 primers used were ZFP1n-4 (5'-ATGTCATCGCTGTTGG-3'), EP3-7 (5'-CCAGCGCATCCTAAGCAGAAC-GCGAAC-3'), PA-A13 (5'-TTGTTGATCCGGCGCCCCGT-3'), and PP1' (5'-GACCGTACGGTCCGTCCG-3'). Primers homologous with Ac sequences were iAc3-2 (5'-TTATCCTGGCTTGTGTTACC-3') and Ac123 (5'-ATCCGGTTCTCGTCCGTTC-3'). The locations of these primers are shown in Figure 9. PCR amplifications were performed using HotStar Taq DNA polymerase (Qiagen) with the following cycle conditions: 95°C for 15 min, followed by 35 cycles of 94°C for 45 s, 60°C for 1 min, and 72°C for 1 min, with a final extension for 8 min at 72°C. Sequences were determined at the Iowa State University Nucleic 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 p1 yields a 690-bp PCR product. After identification of pooled DNA samples that show strong 690-bp PCR products, DNA of individual plants in those pools was 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.

Sequence data from this article have been deposited with the EMBL/GenBank data libraries under accession number AY702552.

**ACKNOWLEDGMENTS**

We thank Ruizhong Shen and Surinder Chopra for isolating the 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 on work supported by National Science Foundation Grant 9601285. This journal paper of the Iowa Agriculture and Home Economics Experiment Station (Ames, IA) was supported by Hatch Act and State of Iowa funds.

Received November 24, 2004; accepted December 21, 2004.

**REFERENCES**


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.


