The Maize Regulatory Gene *B-Peru* Contains a DNA Rearrangement That Specifies Tissue-Specific Expression Through Both Positive and Negative Promoter Elements

David A. Selinger, Damon Lisch and Vicki L. Chandler

Department of Plant Sciences, University of Arizona, Tucson, Arizona, 85721

Manuscript received December 15, 1997
Accepted for publication March 17, 1998

ABSTRACT

The *B-Peru* allele of the maize *b* regulatory gene is unusual relative to most *b* alleles in that it is expressed in the aleurone layer of the seed. It is also expressed in a subset of plant vegetative tissues. Transgenic maize plants containing the *B-Peru* gene with the first 710 bases of upstream sequence conferred the same levels of aleurone expression as nontransgenic *B-Peru* plants, but no pigment was made in vegetative tissues. Transient transformation assays in aleurone tissue localized the aleurone-specific promoter to the first 176 bases of the *B-Peru* upstream region and identified two critically important regions within this fragment. Mutation of either region alone reduced expression greater than fivefold. Surprisingly, the double mutation actually increased expression to twice the native promoter level. Our results suggest that these two critical sequences, which lie close together in the promoter, may form a negative regulatory element. Several lines of evidence suggest that the *B-Peru* promoter arose through the translocation of an existing aleurone-specific promoter to the *b* locus. Immediately upstream of the aleurone-specific promoter elements and in the opposite orientation to the *b* coding sequence is a pseudogene sequence with strong similarity to a known class of proteins. Our findings that novel aleurone-specific promoter sequences of the *B-Peru* transcription factor are found adjacent to part of another gene in a small insertion are quite unexpected and have interesting evolutionary implications.

TISSUE-specific gene expression is essential for the proper development of multicellular organisms and the maintenance of their component tissues and organs. One important level of control is the tissue-specific regulation of transcription factors, which then activate pathways appropriate for particular tissues. The maize *b* gene and its homologue *r* encode a basic helix-loop-helix transcription factor that, in combination with the product of either the *c1* or *pl* gene, induces the transcription of the biosynthetic genes of the maize anthocyanin pathway (Cone *et al.* 1986; Chandler *et al.* 1989; Ludwig *et al.* 1989; Cone *et al.* 1993). The enzymes encoded by the biosynthetic genes produce the red and purple anthocyanin pigments found in maize and many other flowering plants. The existence of many alleles at *b* and *r*, which show a diverse array of tissue-specific pigmentation patterns, provides an excellent system for studying both the mechanisms and elements of tissue-specific gene regulation, as well as the evolution of novel patterns of regulation. Most alleles of *b* are expressed solely in the vegetative parts of the plant. In contrast, two alleles of *b*, *B-Peru*, and *B-Bolivia*, along with many alleles of *r*, are expressed in the seed (Styles *et al.* 1973; Coe 1979).

Corresponding author: Vicki L. Chandler, Department of Plant Sciences, 303 Forbes Hall, University of Arizona, Tucson, AZ 85721. E-mail: chandler@ag.arizona.edu

Two alleles of *b*, *B-Peru*, and *B-I*, which have markedly different expression phenotypes in the seed and the plant, have been cloned and extensively restriction mapped. One dramatic difference is that *B-Peru* is highly expressed in the aleurone, while *B-I* is not expressed in this tissue (Radicella *et al.* 1992). Introduction of chimeric constructs, consisting of the upstream region of *B-Peru* and the coding and downstream regions of *B-I*, into isolated maize aleurones induced the expression of a reporter gene to levels identical to that produced by the intact *B-Peru* construct. Conversely, constructs with the upstream region of *B-I* and the *B-Peru* coding and downstream sequences failed to induce expression in the aleurone (Radicella *et al.* 1992). Intragenic recombinants between *B-Peru* and *B-I*, which have breakpoints in the 5′ end of the transcribed region and thus contain the *B-Peru* upstream region and the *B-I* coding region, have been isolated. These recombinants have aleurone-specific and plant-specific expression patterns indistinguishable from those produced by the native *B-Peru* allele (Patterson *et al.* 1995). The plant expression of *B-Peru* is readily distinguishable from that of *B-I* (Radicella *et al.* 1992). These studies indicate that the tissue-specific regulatory elements controlling *B-Peru* seed and plant expression lie upstream of the coding region. Restriction mapping of the upstream regions of the *B-I* and *B-Peru* alleles revealed a 2.5-kbp block of divergent sequence found immediately up-

stream of the start of transcription in *B-Peru* (Radicella *et al.* 1992; Harris *et al.* 1994; Patterson *et al.* 1995).

In this study, we have used two complementary approaches to test whether the divergent block of sequence is responsible for the aleurone and plant expression of *B-Peru*. The first approach involved the construction of stable transgenic lines containing portions of the *B-Peru* upstream region along with the native *B-Peru* coding and downstream regions. The second was a quantitative assay based on the transient transformation of reporter gene constructs into isolated maize aleurones. Our results from both of these assays indicate that a small part of the 2.5-kbp sequence is both necessary and sufficient for aleurone expression. This sequence is not able to confer expression of *B-Peru* in plant tissues. We have characterized the specific parts of that sequence that are important for aleurone expression. Our results suggest that multiple sequence elements within a small region of 140 bp are involved in producing the aleurone expression of *B-Peru*. We discuss the significance of these elements and a possible mechanism for the evolution of the *B-Peru* allele.

MATERIALS AND METHODS

DNA gel blot analysis: DNA blot analysis and hybridizations were performed as described in Patterson *et al.* (1995). Hybridization probes were the BIu4 and Spe/G700 probes used by Patterson *et al.* (1995) and the 0.5-kb *Pst*I *B-P* probe of Harris *et al.* (1994).

Transient transformation assay: Kernels homozygous for *b r-g* alleles were sterilized in 20% bleach, 0.1% Tween 20 for 20 min, washed, and soaked overnight in sterile water. The pericarp, embryo, and scutellum were removed. After splitting the kernel in half, four halves were placed in the middle of an MS medium plate with the aleurones facing up and incubated overnight. The aleurones were subjected to microprojectile bombardment as described (Klein *et al.* 1989). Each DNA precipitation onto microprojectiles consisted of 5 μg of pBP35ScDNA (*B-Peru* cDNA driven by the CaMV 35S promoter, Goff *et al.* 1990), 5 μg of pCIB4436 (C1cDNA driven by the CaMV 35S promoter, gift of Ciba-Geigy), 5 μg of pJB4 [*uidA* (GUS) driven by the CaMV 35S promoter; Bodeau and Walbot 1992], and 2.5 μg of a chimeric *B-Peru* promoter, firefly luciferase reporter plasmid. All of the constructs contain the maize *adh1* intron1 between the promoter and the appropriate cDNA, as well as the Agrobacterium *nopaline synthase* 3′ end. The *B-Peru* and *C1* cDNA plasmids indicate bombardment efficiency by inducing the expression of the anthocyanin pathway, which results in purple cells. The GUS construct allows quantitation of transformation efficiency, and the chimeric *B-Peru*/luciferase plasmid assays the *B-Peru* promoter activity of the segment cloned in front of the luciferase cDNA. A total of 17.5 μg of either CsCl or Qiagen purified DNA (Qiagen, Chatsworth, CA) was precipitated onto 125 μl of 50 mg/ml 1.8-μm tungsten microprojectiles (Bio Rad, Richmond, CA) as described (Radicella *et al.* 1992). Bombardments into L6 suspension cells were done as described (Sainz *et al.* 1997). After bombardment, the tissues were incubated at 28° for 48 hr in the dark. Tissue extracts were prepared (Radicella *et al.* 1992), and luciferase activity (Callis *et al.* 1987) and GUS activity were assayed (Sainz *et al.* 1997).

Use of *Renilla* luciferase as a control for transformation: In some experiments, we introduced a plasmid, p35S-Rluc, consisting of the CaMV 35S promoter and ADH1-Intron1 fused to the *Renilla* luciferase gene (Lorenz *et al.* 1991) as an additional transformation control. The plasmid was constructed by inserting a blunt-ended, *Nhe*I/*Xba*I fragment from pRL-null (Promega, Madison, WI) into the blunted *Eco*RI site of pMF6 (Callis *et al.* 1987). We included either 1.0 or 0.1 μg (the latter amount gave more uniform results) of this construct in the DNA precipitations and assayed the activity of this enzyme by Promega's Dual Luciferase Assay.

Analysis of luciferase data: After correcting the luciferase and GUS activity for background levels, the ratio of luciferase activity over GUS activity was used to normalize for transformation. This ratio was then divided by the normalized value of the no promoter control (pABR4) that was included in all experiments. This step allows the values of different experiments to be compared. Standard error levels were calculated, and Student's *t*-test was used to determine statistical significance levels.

Construction of the 5′ deletion series: The plasmid pABPlucdK consists of 2.5 kbp of the upstream sequence of *B-Peru* from the *Spe*I site at ~−2500 to the *Sac*I site at +29, including all of exon1, fused to the maize *adh1* intron1, the firefly luciferase cDNA, and the NOS 3′ end, and cloned into pTZ18U with the *Kpn*I site in the polylinker deleted. The −1044, −710, and −176 deletions were obtained by partial digestion of pABPlucdK with *Xba*I; the −559 and −144 deletions were obtained by complete digestion with *Pst*I or *Kpn*I, respectively. The −127 and −80 deletions were produced by PCR using a primer encoding a *Bam*HI site at the 5′ end and complementary to the *B-Peru* sequence starting at the 5′ end of the deletion. A subclone (pBP X/S) of *B-Peru* from the *Xba*I site at −176 to the *Sac*I site at +29 in pTZ19U was used as the template, and the M13 universal primer was the 3′ end primer. The product was cloned into pABPlucdK using the *Bam*HI site in the polylinker and the *Sna*BI site at +2.

PCR based linker-scanner mutagenesis: Single and double linker-scan (L-S) mutations were generated by designing a pair of PCR primers containing a complementary 12-bp linker sequence at their 5′ ends. Primer A has a 12-bp tail (GAATTC CTCGAG) followed by 18 bp of sequence complementary to the *B-Peru* sequence extending from the 5′ end of the insertion site upstream. Primer B has the complementary tail (CTCGAG GAATTC) followed by 18 bp complementary to *B-Peru* sequence from the 3′ end of the insertion site downstream. Universal and reverse primers were used with the appropriate A and B primers to generate two products from the pBP X/S template. After gel purification, the A and B products were mixed, and PCR was performed with universal and reverse primers. The reaction was set up so that the first three rounds allowed the 12-bp complementary overlap of the A and B primers to anneal; extension from the 3′ ends of these annealed products created a full-length template for amplification by the universal and reverse primers. The second product was cloned into the luciferase expression plasmid pABPluc using the *Bam*HI and *Sna*BI sites. The mutated *B-Peru* segment of the resulting plasmid was sequenced to check all mutations.

Construction of −710 L-S mutations: Additional sequence from *B-Peru* corresponding to the 534 bp between the −710 and −176 *Xba*I sites was subcloned into the *Xba*I site upstream of two L-S mutations (LS120 and LS96) that were cloned into the luciferase expression plasmid. This subcloning created the L-S mutations in the context of the −710 promoter deletion.

Construction of insertion mutations within L-S mutations: Mutations containing a 4-bp insertion in LS156, LS132, LS120, and LS60 were created by cutting the *Xho*I site in the linker sequence, filling in with Klenow, and religating. Mutations with 21-bp insertions in LS132, LS120, and LS60 were pro-

duced by ligating a 21-bp segment of DNA with *Xho*I-compatible overhangs (produced from two complementary oligonucleotides encoding the sequence of the pBR322 *tetR* gene from +1 to +15, with *Xho*I compatible overhangs) into the *Xho*I site of the linker sequence.

Heterologous promoter constructs with the CaMV −35S basal promoter: Complementary oligonucleotides containing *B-Peru* sequence and designed to anneal to leave a four-nucleotide overhang compatible with either *Hin*dIII or *Xho*I were ordered, phosphorylated, and annealed. The annealed oligos were ligated at either 1:1 or 4:1 molar ratios with a −73 CaMV 35S promoter construct (Roth *et al.* 1991). A larger piece of the *B-Peru* promoter, from −176 to −36, was fused to the −73 CaMV 35S promoter by excising the desired region from the LS36 luciferase plasmid using the *Bam*HI site in the polylinker and the *Xho*I site of the L-S mutation. Pieces from −176 to −60 and −176 to −120 were fused to a −56 CaMV 35S promoter construct (Grotewold *et al.* 1994) using the same technique.

Production of transgenic maize: We followed the methods of Koziel *et al.* (1993). Thirty to 40 immature embryos (0.5–1.5 mm in length, ~14 days after pollination) were dissected from the line CG00526 (gift of Ciba-Geigy) and plated on each target plate with the embryonic shoot facing the media. Two days after isolation, the embryos were placed on 12% sucrose plates and bombarded 4–6 hr later. Two bombardments were done at ~1500 psi per target plate. After 16 hr, the bombarded embryos were transferred to low (2%) sucrose media containing a sublethal amount (10 mg/liter) of the herbicide Basta. Embryos were bombarded with 1.0- or 0.3-μm gold microprojectiles, onto which we had precipitated 2 μg of pCIB3064 (CaMV 35S promoter driving the bar gene; Ciba-Geigy) and 2 μg of either pBP2 (−2500 5′ end point of *B-Peru*), pBP710 (−710 5′ endpoint), or pBP176 (−176 5′ endpoint). After 3 wk of sublethal selection, we subjected the resulting embryonic callus to lethal (20 mg/liter Basta) selection for 10–12 wk and then regenerated the stably transformed callus. Regeneration was done either with 10 mg/liter Basta selection or later with no selection. Regenerated plants were grown in a greenhouse, and the T_0 plants outcrossed to a *b Pl C1 r-g* line. DNA blots and visual observation of the kernel phenotype on the ears of T_0 plants were used to determine which plants contained functional transgenes (neither CG00526 *B* nor *r* alleles are expressed in the aleurone). After two generations of backcrosses to separate the transgene from the endogenous CG00526 *B* and *r-r* alleles, plant phenotypes of the transgenes were scored.

RESULTS

The *B-Peru* allele contains two regions of divergent sequence relative to the *B-I* allele: Our current understanding of the structure of the *B-Peru* allele is summarized in Figure 1. The previously cloned portion of the *B-Peru* allele showed high sequence identity with the *B-I* allele through the coding and 3′ flanking regions (Radicella *et al.* 1992). As previously described, the two sequences diverged completely at position +43 in exon1 of *B-I*. In *B-Peru*, this diverged sequence continued for the full 2.5 kbp of cloned upstream sequence (Figure 1).

Sequence analysis showed that the cloned 2.5-kbp block of divergent sequence ends serendipitously at the *Spe*I site that was previously used for cloning (Figure 1). The 5′ end of this divergent sequence ends with the same TATATATA sequence that is found at the 3′ end of the divergent sequence. Based on Southern blot analysis, the sequence upstream of the *Spe*I site consists of several hundred bases that is homologous to the *B-I* sequences immediately upstream of the *B-I* start of transcription. A second region of divergent sequence that is at least 14 kbp in size separates the block of sequence that hybridizes to the BIu4 and H/S 400 probes from that which hybridizes to the S/G 700 probe (Figure 1). In *B-I*, the H/S 400 and S/G 700 probes overlap by ~50 bases. In *B-Peru*, these two probes are estimated to be separated by a fragment with a minimum size of 14 kbp. This minimum size is based on the assumption that several restriction sites mapped upstream and downstream of two probes are the same sites; these probes are contiguous in *B-I* (see materials and methods for details). The addition of the sizes of these pairs of fragments generated the minimum size. Sequences in *B-Peru* between the S/G 700 and 1.6 B′V probes were spaced similarly to the region in *B-I* and shared some restriction fragments with *B-I* (Patterson *et al.* 1995). The blocks of distinct sequence coupled with the great difference in tissue specificity of the two alleles suggested that the unique sequences in *B-Peru* might be conferring its tissue-specific expression. We thus set out to explicitly test the ability of the proximal region of divergence to produce the aleurone-specific phenotype of *B-Peru*.

Transgenic maize plants show that the −710 to +1 region of *B-Peru* is both necessary and sufficient for aleurone expression: Transgenic maize plants were produced carrying one of three *B-Peru* constructs. Each construct, consisting of various amounts of 5′ nontranscribed sequence, the native *B-Peru* coding region, and ~4 kbp of 3′ flanking sequence, was cotransformed along with a selectable, herbicide-resistance marker into immature embryos via particle gun bombardment. Twelve days after bombardment, initial herbicide-resistant callus responses were scored and dissected. The frequency of response varied in the different experiments. Between 40 and 68% of the embryos gave initial responses; the average frequency of response is 54% (see Table 1). After 12 wk of herbicide selection, the remaining lines were considered to be, and behaved as if, they were stably transformed. The percentage of initial responses that gave rise to stable lines varied between 0.5 and 5.9%, with an average of 3%. These numbers reflect callus lost to selection and to contamination. Stable lines were then regenerated into plantlets. Most of the lines were eventually regenerated, although a significant number remained refractory to regeneration despite many attempts (Table 1). The resulting T_0 plants were crossed to an appropriate tester line containing recessive alleles of *b* and *r* and functional alleles of all other anthocyanin genes. We identified six BP2500 and three BP710 plants that expressed anthocyanins in the aleurone (see Figure 2 for diagrams and phenotypes of

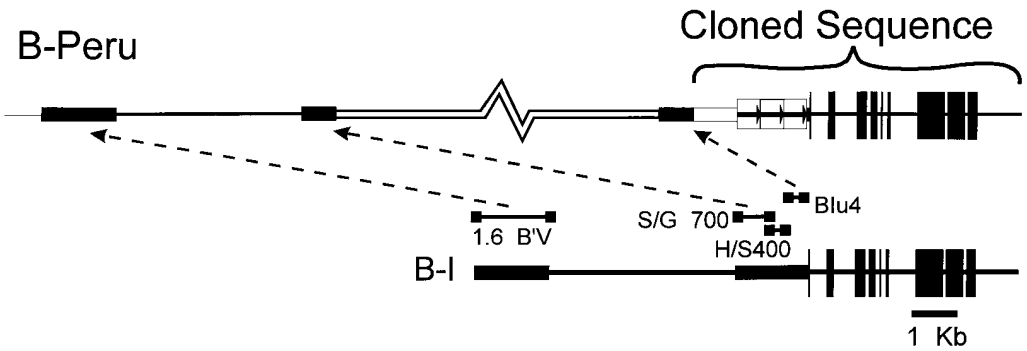

Figure 1.—Structure of *B-Peru*. Approximately 12 kbp of *B-Peru* sequence was previously cloned (Chandler *et al.* 1989; Radicella *et al.* 1992) and is denoted by the bracket. Sequences unique to *B-Peru* are indicated by open boxes, and sequences shared with *B-I* are indicated by black boxes. The sequence immediately upstream of the transcription start site is completely divergent from that of *B-I* (shown below) and contains a 534 bp sequence, denoted by the arrows, that is repeated three times. Use of probes complementary to the upstream region of *B-I* (1.6 B′V, S/G 700, H/S 400, and Blu4) revealed that these sequences are present in *B-Peru* along with a second region of divergence (Patterson *et al.* 1995). This second region of divergence is quite large, spanning at least 14 kbp.

transgenes). These represented ~40% of the herbicide-resistant lines. All nine transgenic lines with anthocyanin expression in the aleurone had phenotypes that mimic the aleurone phenotype of the native *B-Peru* allele. In all cases, the transgenic maize lines segregated one functional transgenic locus, indicated by the segregation of 50% colored and 50% colorless kernels upon outcrosses to recessive testers. Southern blots showed that all the expressing lines carry at least one intact copy of the *B-Peru* transgene construct (Table 1). The transgene copy number varied among the individual lines from just a few copies to more than 10. Eight of the nine lines that initially expressed anthocyanin pigment in the aleurone continued to show expression in subsequent generations. One initially expressing line failed to produce progeny that expressed the transgene (based on the absence of pigment), and one line that did not initially express pigment appears to contain an intact transgene, based on Southern blotting. These latter two lines were probably subject to transgene inactivation (reviewed in Matzke and Matzke 1995) although this has not been investigated further. The ability of the proximal 710 bp of upstream sequence to reproduce the aleurone phenotype of *B-Peru*, together with the inability of transgene constructs with *B-I* upstream regions to express in the aleurone (K. Kubo and V. Chandler, personal communication), demonstrate that our cloned portion of the *B-Peru* upstream region is both necessary and sufficient for the aleurone expression of *B-Peru*.

TABLE 1

Transgenic plant data

Experiment	Transgene	Embryos bombed	Percentage showing an initial response	Stably transformed lines	Percentage of initial lines that survive selection	Regenerated lines	Regenerated lines with intact transgenes	Lines with aleurone expression
JEG99	BP2500	188	46	4	4.65	4	2	2
JEG101	BP2500	144	59	5	5.88	5	2	2
VLC22	BP2500	384	68	5	1.92	3	0	0
VLC23	BP2500	387	62	6	2.52	5	2	2
Summary of BP2500		1103	59	20	3.74	17	6	6
VLC10	BP710	375	62	2	0.86	1	0	0
VLC11	BP710	369	57	11	5.24	6	3	2
VLC12	BP710	427	40	1	0.58	1	1	1
Summary of BP710		1171	53	14	2.23	8	4	3
VLC17	BP176	708	43	4	1.30	2	0	0
VLC19	BP176	865	51	10	2.27	5	0	0
VLC29	BP176	755	60	24	5.32	13	1	0
VLC30	BP176	654	49	11	3.43	4	0	0
Summary of BP176		2982	51	49	3.08	24	1	0

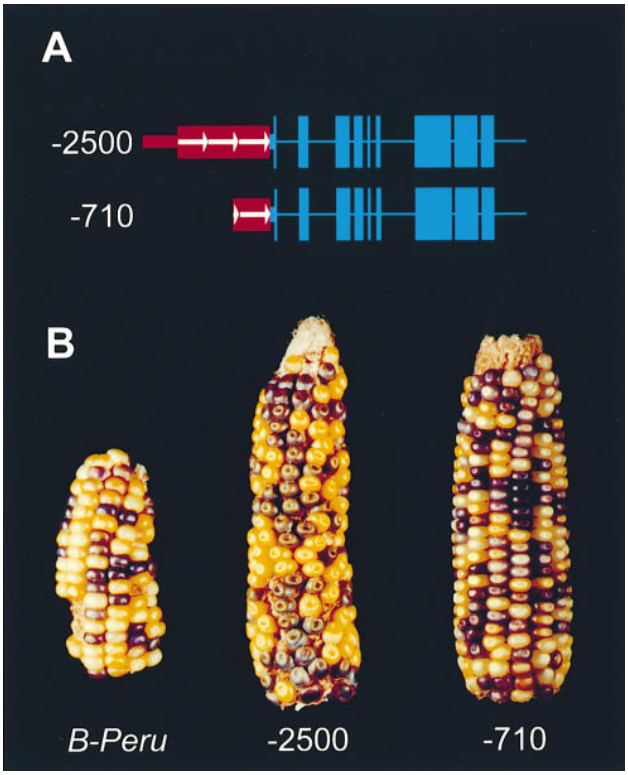

Figure 2.—Aleurone phenotype of transgenic constructs. (A) The BP2500 (−2500 to +1 of upstream sequence) and BP710 (−710 to +1 of upstream sequence) constructs are shown schematically. The transcribed region and 3′ flanking sequences are shown in blue. the portion of the promoter sequence of *B-Peru* that differs from *B-I* is shown in red. (B) Ears from plants heterozygous for a *b* allele and either the native *B-Peru* allele, a BP2500 transgene, or a BP710 transgene are shown. The transgenic ears are from a cross between the T_0 plant and a tester line homozygous for functional copies of all anthocyanin pathway genes, except *b* and *r*. The *B-Peru* ear is from a cross of the same tester with a plant heterozygous for *B-Peru*, and the inbred line is used as the transformation recipient.

Although the *B-Peru* transgenics reproduce the aleurone phenotype, they do not express the typical plant pigment phenotype of *B-Peru*. *B-Peru* plants have darkly pigmented tassels and variable pigmentation of the sheath, culm, and husk (see Radicella *et al.* 1992 for a detailed description). At least eight transgenic plants from the second backcross generation, derived from each of the nine independent transgenic lines, were compared to sibling plants without the transgene. We observed no pigmentation in either the transgenic plants or their siblings. The inability of the full 2.5 kbp of upstream sequence to produce any plant-specific expression suggests that the promoter elements contained within it are aleurone specific and not general elements, and that the plant-specific expression of *B-Peru* is generated from elements located elsewhere. Intragenic recombination experiments with *B-I*, which produces a distinct pattern of plant pigmentation, suggest that these sequences lie farther upstream (Patterson *et al.* 1995).

In addition to the −2500 and −710 constructs, we introduced a construct containing the proximal 176 bp of upstream sequence. This construct has promoter activity equivalent to the −2500 construct in the transient expression assay (see below). We were unable to recover any expressing lines from 24 regenerates. Upon examination by Southern blotting, only one line appeared to have an intact transgene. This was strikingly different from the situation for transgenic lines obtained with the other two constructs (Table 1). One potential explanation for the lack of expressing clones is that inappropriate expression of the *b* protein and/or the anthocyanin pathway is toxic, and the −176 construct, but not the −710 or −2500 construct, is inappropriately expressed in the embryonic callus or regenerating plants. We did not see a significant difference in expression between BP2500 and BP176 in callus and embryo-transient transformation assays (data not shown). However, in both cases, expression was very low compared to aleurone levels and, thus, it may be that there is a subtle but still important difference either in callus initiation, maintenance, or during regeneration.

The exon 1 sequence of *B-Peru* is important for expression but does not confer tissue specificity: The aleurone phenotype of *B-Peru* can be reproduced in a transient transformation assay using the *B-Peru* genomic clone (Radicella *et al.* 1992). In these assays, the expression of *B-Peru* was measured indirectly by looking at luciferase expression from a target promoter, the *bronze1* promoter, fused to luciferase. This assay is very sensitive, and in the case of *B-Peru* aleurone expression, takes very little of the *B-Peru* plasmid to saturate the system. To directly measure *B-Peru* expression, we fused the *B-Peru* upstream region, including the nontranslated exon1 to the firefly luciferase cDNA (Figure 3A). The first intron of the maize *adh1* gene was included between the *B-Peru* upstream sequence and the firefly luciferase cDNA to improve the expression of the heterologous cDNA (Callis *et al.* 1987). This construct, when introduced into maize aleurones via particle gun bombardment, gave a robust 364-fold increase in luciferase expression over that produced by the no promoter control pABR4. To determine whether exon1 sequences, which are homologous to the 3′ end of the *B-I* exon1, are responsible for the aleurone-specific differences between the two alleles, we tested *B-Peru* and *B-I* constructs that lacked exon1 (Figure 3A). The *B-Peru* construct that lacks exon1 produced a luciferase level that was 30% of the level produced by the construct containing exon1. Both *B-I* constructs expressed very little luciferase. The *B-I* construct with exon 1 produced 3% of the luciferase expressed by the *B-Peru* construct. Deletion of exon 1 reduced the expression to 1%, which is ~3% of the level obtained with the *B-Peru* construct deleted for exon 1. Thus, inclusion of exon1 increases the ex-

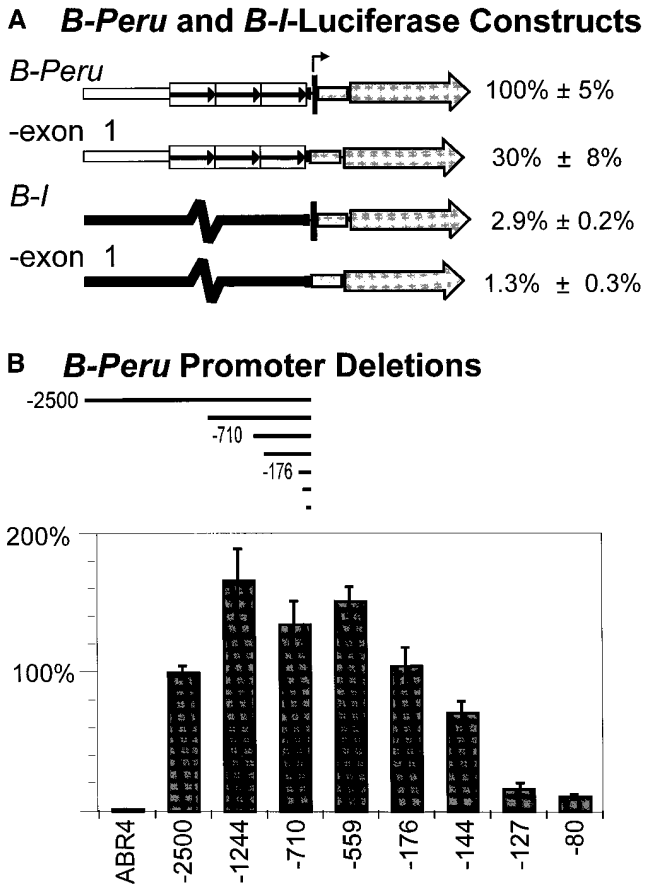

Figure 3.—Transient assay measuring *B-Peru*, *B-I*, and *B-Peru* 5′ deletion construct expression. (A) The structure of the chimeric *B-Peru*:luciferase and *B-I*:luciferase constructs, with or without exon 1 from *B-Peru*, are shown with their relative expression levels normalized to *B-Peru* + exon1 (set at 100%). (B) The endpoints of the 5′ deletion series are indicated by lines that correspond to the *B-Peru* structures shown in A. Apart from the different fragments of the *B-Peru* upstream region, all the deletion and substitution mutation constructs in this paper have the same structure. Each construct contains a part of the *B-Peru* upstream region, *B-Peru* exon1, the intron 1 from the maize ADH1 gene, and firefly luciferase cDNA sequence followed by the nopaline synthase 3′ sequence. The histogram shows the results of testing a 5′ deletion series for activity in maize aleurones. Each bar represents data from at lease five individual samples, and the error bars show standard error.

pression of the chimeric constructs approximately threefold but does not affect aleurone specificity.

Deletions of the *B-Peru* upstream sequence reveal that the proximal 176 bases of upstream sequence are required for full expression in the aleurone: To define the minimal sequence required for aleurone expression, we produced a series of 5′ deletions of the *B-Peru* upstream sequence (Figure 3B). Deletions with upstream endpoints at −176 or further upstream showed no decrease in expression relative to the full 2.5 kbp of upstream sequence. Deletion to −144 resulted in a modest decrease in expression, while deletions to −127 and −80 produced more severe decreases (Figure 3B and Table 2). These results localize the important promoter elements to the proximal 176 bases of upstream sequence and show that the cloned *B-Peru* sequences are sufficient for aleurone expression. The decrease in expression from 70% produced by the −144 construct to 15% produced by the −127 construct suggests that critical positive elements are at least partially contained within this 17-bp region.

Substitution mutagenesis of the proximal 176 bp of the *B-Peru* promoter reveals at least two critically important elements for aleurone expression: To identify the *cis*-acting elements responsible for the aleurone expression of *B-Peru*, we used a substitution mutagenesis strategy referred to as L-S mutagenesis. We used a PCR-based strategy to introduce 10- or 12-bp substitution mutations without affecting the spacing of potentially important elements. This "linker" sequence was then "scanned" over the region from −176 to +1 by producing a series of adjoining mutations. Two of the mutations produced large decreases in expression (Figure 4). The first, between −120 and −108, produced an approximately fivefold drop in expression and is flanked by mutations with little or no effect. The second, between −96 and −84, reduced expression ~15-fold and is flanked on the 5′ side by a region that has no effect when mutated. On the 3′ side of the −96 mutation lies a 48-bp region stretching to the TATA box in which all L-S mutations resulted in modest but significant losses of expression. This region may consist of several individual elements or may be part of the −96 element. Alternatively, the sequence context of this region may affect the function of the −96 region. Based on the decreased expression of the −127 deletion construct to 15% of the level seen with the −176 construct, we predicted the region between −176 and −120 to contain one or more positive elements. Three of the L-Ss in this region, the −176, −144, and −132 L-Ss showed modest decreases in expression. Although none of these mutations resulted in a large decrease in expression, by simply multiplying them together, they cumulatively produce a decrease in expression to 20%, which is close to the 15% level produced by the −127 deletion. Interestingly, the two L-Ss, that did not decrease expression, −166 and −156, showed increases in expression of 283 and 142%, respectively. The nearly threefold increase in expression produced by the −166 mutation suggests the presence of a negative regulatory element.

Both of the critically important elements, defined by the mutations at −120 and −96 (which will subsequently be referred to as E1 and E2, respectively) are found in the portion of the *B-Peru* sequence that is repeated three times in the upstream region. We wondered if the upstream copies of these elements could compensate for the loss of the proximal copy. To test this idea, we took advantage of the naturally occurring redundancy in the *B-Peru* promoter by introducing either of the two L-S

TABLE 2

A comparison of deletion, linker-scan, and insertion mutations

	Deletions			Linker-scans			Insertions		
Position	Expression	SE (%)	*t*-test[a]	Expression	SE (%)	*t*-test[a]	Expression	SE (%)	*t*-test[b]
−2500	100	5							
−176	106	14	65	75	12	95			
−166				283	43	>99			
−156				142	10	>99	57	4	>99
−144	71	8	>99	43	3	>99			
−132				62	6	>99	116	7	>99
−120	15	6	>99	18	2	>99			
−108				109	17	68	29	3	>99
−96				7	1	>99			
−84	10	2	>99	30	2	>99			
−72				37	6	>99			
−60				56	8	>99	59	5	62

[a,b] Student's *t*-test for significance was used to compare the data from the deletion or linker-scan mutation with that of the −2500 construct[a] or, for the insertion mutations, the corresponding linker-scan mutation[b]. Values >95% are considered to be statistically significant.

mutations into the proximal site of the first 710 bp of upstream sequence. This sequence contains all of one copy of the repeated sequence plus a partial second copy that duplicates the sequences mutated by the L-S (Figure 5). We found that the −710 LS120 construct, with a mutated E1 module, produced an expression level that is 94% of the intact −176 construct (Figure 5). In contrast, the LS96 mutation, with a mutated E2 module, produced no increase in expression in the context of the −710 construct over that produced by the −176 construct. These data demonstrate that although sequences upstream of −176 can compensate for the loss of E1 at position −120, these sequences, which include an upstream copy of E2, cannot compensate for the loss of E2 at position −96.

Insertion mutations suggest that factors binding to the E1 and E2 elements interact: To test for position dependence of the elements identified by the L-S experiment, we created a series of mutations with insertions within previously described L-Ss that had either no effect or a minor effect (Figure 4). Filling in the *Xho*I site in the linker sequence produced 4-bp insertions. Because the 4-bp insertions change both the spacing and helical orientation between elements, we also tested 21-bp insertions, which have two complete helical repeats, to determine whether an alteration of helical orientation or spacing produced the effects seen in the 4-bp insertions. Insertions of 4 and 21 bp into the LS60 L-S resulted in no decrease in activity over that seen with the L-S mutation (59 and 76% *vs.* 56% for the L-S, Table 2). Both 4- and 21-bp insertions into the LS132 linker-scan produced modest increases in expression from 62 to 110% and 102%, respectively (Table 2). In contrast, a 4-bp insertion into the LS156 L-S resulted in a significant loss of activity, from 142 to 57% (Table 2). This reduction may result from changing the distance between the positive regulatory element identified by the −176 L-S and −144 deletion mutations and downstream elements, or from increasing the effect of the possible negative element identified by LS166. Interestingly, insertion of 4 bp between E1 and E2 (into LS108) reduced expression from 109 to 29%. This decrease is partially ameliorated by the 21-bp insertion, which had an expression of 45% (Table 2). The effect of the insertions between E1 and E2 at LS108, which decreased expression, *vs.* insertions downstream of both at LS60, which did not decrease expression, suggests that the spacing between E1 and E2 is important. This result is consistent with a direct interaction between factors binding to E1 and E2.

E1, but not E2, is capable of promoting a modest increase in the aleurone expression of a heterologous minimal promoter: The L-S results suggest that E1 and E2 are two critically important elements in the *B-Peru* promoter. To test if these elements either alone or together can confer aleurone expression on a heterologous basal promoter, we cloned oligonucleotides containing each element in front of a −73 CaMV 35S promoter that was fused to the *adh1* intron1 and the firefly luciferase cDNA (Roth *et al.* 1991). The −176 *B-Peru* promoter construct had an expression level in aleurone tissue that was ~11-fold higher than the −73CaMV construct alone. On the basis of the L-S results, we chose the sequence between −126 and −102 to test as E1 and the sequence between −96 to −72 to test as E2. A doublet of E1 induced an approximately twofold increase in expression (not shown), and the triplet of E1 (E1T:35S) induced an almost threefold increase over the basal expression (Figure 6). Although these inductions are smaller than expected based on

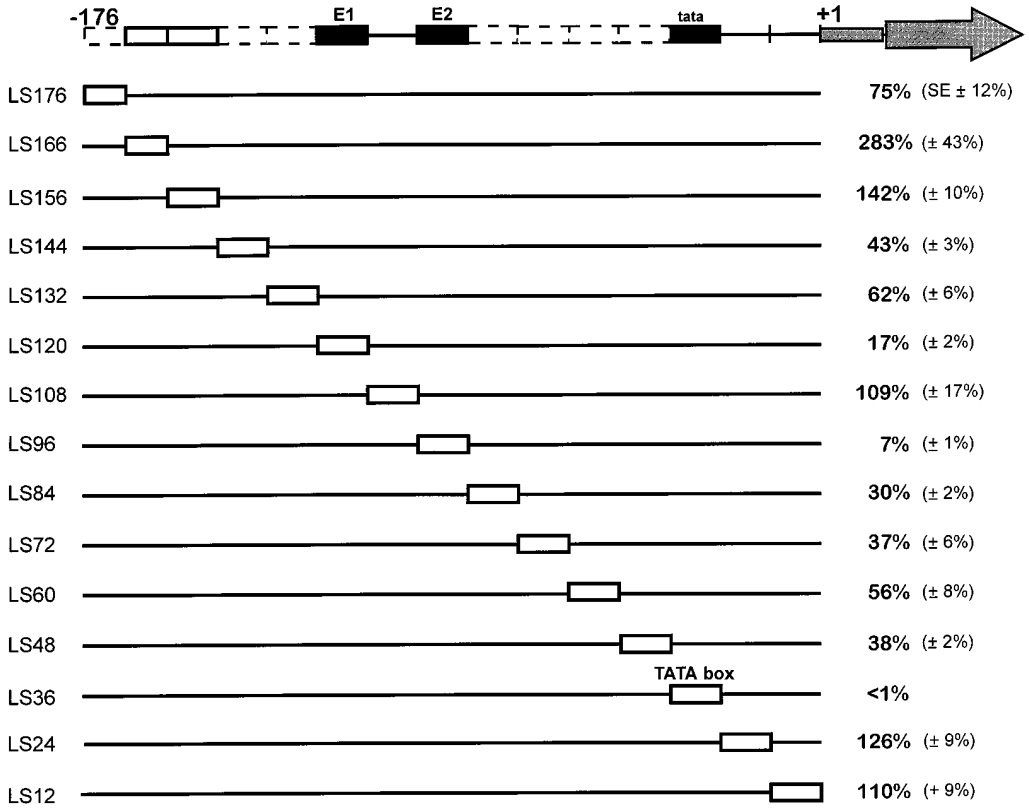

Figure 4.—Results of L-S mutagenesis of the *B-Peru* upstream sequence. A 12-bp linker sequence containing an *Eco*RI and an *Xho*I site was used to replace 12-bp blocks of *B-Peru* sequence in the −176 deletion construct (Figure 3). In the case of the LS176 and LS166 mutations, a 10-bp linker that is missing the last two bases of the 12-bp linker was used. Across the top is a diagram of the promoter with the effects of the mutations shown by a bar for no effect, an empty box for an increase in expression, a dashed box for a moderate reduction, or a solid box to denote severe reductions in expression. Below that are boxes showing the positions of the mutations and their expression normalized to the −2500 construct. The standard error of measurement is shown in parentheses.

the L-S results, these increases are statistically significant and reproducible. Although the L-S in the E2 region had the most severe effects, addition of E2 to the basal promoter (E2D:35S) had no effect on its expression (Figure 6).

Because introduction of E2 did not affect the CaMV 35S minimal promoter expression, we tested the two elements in concert to see if E2 could amplify the effect of E1. A construct containing a triplet of E1 upstream of a doublet of E2 in the −73 CaMV 35S promoter construct (E1T,E2D:35S) produced an induction of 2.3 over the basal CaMV promoter alone. This induction was not statistically different from the 2.8-fold induction of the E1 triplet alone (Figure 6). Because the combination of the E1 and E2 elements did not produce a greater level of expression, we tested the *B-Peru* promoter sequence between −176 and the TATA box (−36) fused to the −73 CaMV 35S promoter. The 2.1-fold increase in expression that this construct (BP176-36:35S) produced was not statistically different from the introduction of the E1 trimer alone. To test whether the change in distance between E2 and the TATA box of 33 bp in the previous construct is responsible for the lack of expression, we tested a similar construct containing the −176 to −60 region of *B-Peru* fused to a −59 CaMV 35S promoter (Grotewold *et al.* 1994). This construct results in a fusion where the distance between E2 and the TATA box is almost the same as in the native promoter. The −59 construct (BP176-60:35S) gave a 3.1-fold increase in expression over the −59 CaMV 35S promoter, indicating that the distance between E2 and the TATA is not the critical factor contributing to the low induction. These results suggest that although mutation of either E1 or E2 produced a large decrease in expression, neither of these elements is a strong activator of transcription on their own. In addition, the −176 to −60 region cannot confer *B-Peru* levels of expression when fused to a heterologous TATA and initiator element that does not contain exon 1.

Because the results from the deletion mutations suggest the presence of important elements contributing to activation in the region between −176 and −127, we also tested the activity of a construct consisting of the region between −176 and −120 fused to the −59 CaMV 35S promoter. This construct (BP176-120:35S) did not produce an increase in activity over the −59 CaMV 35S promoter alone. This result suggests that there are no independently acting positive elements between −176 and −120, and that the putative activator elements in this region require elements downstream of −120 to act.

Deletion or mutation of the region between −120 and −84 increases expression: If factors binding to E1 and E2 directly interact, then deleting or mutating both E1 and E2 may produce no greater decrease in expression than that produced by either mutation alone. Three mutations were made to test this idea. The first was a substitution of the 36-bp region from −120 to −84, with the 12-bp sequence used in the L-S experiments. Surprisingly, this deletion mutation produced a higher expression level (38%) than those of the LS120 (17%)

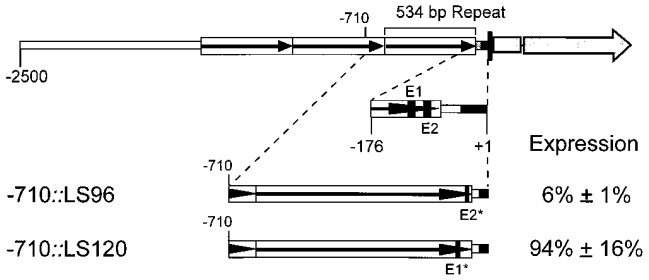

Figure 5.—Two L-S mutations behave differently in the context of the −710 sequence. To test if sequences farther upstream could complement for the loss of important promoter elements, the LS120 and LS96 mutations, which showed the largest decreases in expression in the substitution mutagenesis, were placed in constructs containing 710 bp of upstream sequence. This construct contains a second copy of the region found between −176 to −66, which contains the LS120 and LS96 L-S mutations in addition to the full proximal 534-bp repeat sequence.

and LS96 (7%) single mutations (Figure 7). To more cleanly test the effect of eliminating E1 and E2 and to see if the sequence between −108 and −96, which was mutated without effect in LS108, was contributing to the deletion phenotype, we created two double L-S mutations. One of these double mutations combined the LS120 and LS96 mutations, and the other combined the LS108 and LS96 mutations. Both double mutations gave higher levels of expression than the LS96 mutation alone. The LS120:LS96 mutation dramatically increased expression to nearly twice that produced by the full-length promoter (196%). The LS108:LS96 mutation produced an expression level of 26%, which was nearly fourfold higher than the LS96 mutation alone, but much less than the LS120:LS96 mutation. Because these two mutations share the same linker sequence, the large increase in expression of the LS120:LS96 mutation does not result from a fortuitous activator sequence in the linker. The large increase in expression produced by the double mutation, in contrast to the large decreases in expression produced by either of the single mutations, suggests that E1 and E2 are components of a negative regulator. At the very least, these results indicate that something more complicated than a simple model of activation sites is required to explain the *B-Peru* promoter.

The LS120, the LS96, and the LS120:LS96 double mutations do not show increased expression in a tissue in which *B-Peru* is not normally expressed: If the E1 and E2 elements are part of a negative regulatory system, the purpose of such an element may be to ensure tissue-specific expression. Mutation of these elements may then increase the very low level of expression produced by the *B-Peru* promoter in nonaleurone tissues. To test this, we transiently transfected L6 suspension cells with the −176 *B-Peru* luciferase, the LS120, LS96, and LS120:LS96 double-mutation constructs. The −176 *B-Peru* construct produced a luciferase level that was 3.32 ± 0.33 that of the no promoter control. The LS120 construct produced a level that was 2.20 ± 0.10, the LS96 mutation produced a level of 1.21 ± 0.44, and the double mutation produced a level of 3.03 ± 0.94. None of these mutations allow an increase over the very low level of expression produced by the intact promoter in maize callus cells.

The upstream region of *B-Peru* appears to contain a pseudogene sequence that does not overlap the promoter sequences: A database search using the entire *B-Peru* upstream sequence (−2489 to +1) identified homology to a group of proteins containing a conserved ATPase domain, the AAA module. The AAA module is highly conserved and found in divergent proteins (reviewed in Confalonieri and Duguet 1995). We found that the upstream sequence of *B-Peru* from −2239 to −1677 contained three regions of homology to this group of proteins. The location of these regions and the alignment between the *B-Peru* upstream region and three representative AAA domain proteins is shown in Figure 8. Interestingly, the homologous region in *B-Peru* is broken into three blocks, suggesting exon/intron structure. The translation reading frame of the pseudogene is running in the opposite direction relative to *B-Peru* transcription. A search using the 534-bp repeated sequence that contains the aleurone-specific promoter elements did not find any homologous sequences in GenBank.

DISCUSSION

The upstream region of *B-Peru* contains two large-scale rearrangement events relative to *B-I*. We have shown that the sequence found in the rearrangement that is proximal to the start of transcription contains the promoter activity that is necessary and sufficient for *B-Peru* expression in the aleurone layer of the seed. Expression appears limited to the aleurone as this sequence cannot induce expression in undifferentiated callus cells in transient assays, and transgenic plants carrying this sequence did not produce anthocyanin in any other tissues. In the nine transgenic lines that expressed aleurone pigmentation, there was absolutely no plant expression. The fact that these lines expressed pigment in the seed, and the production of plant and seed pigmentation by a transgenic plant carrying the *B-Peru* cDNA under the control of the CaMV 35S promoter (data not shown), enable the interpretation that the −2500 or −710 fragments do not contain sequences sufficient for plant expression. Our observations, together with the results of Patterson *et al.* (1995), who localized the tissue-specific regulatory sequences of *B-I* and *B-Peru* to the upstream regions of these alleles, suggests two hypotheses. First, *B-Peru* plant color may be the result of the upstream sequences that are homologous to those in *B-I*, an allele with strong plant expres-

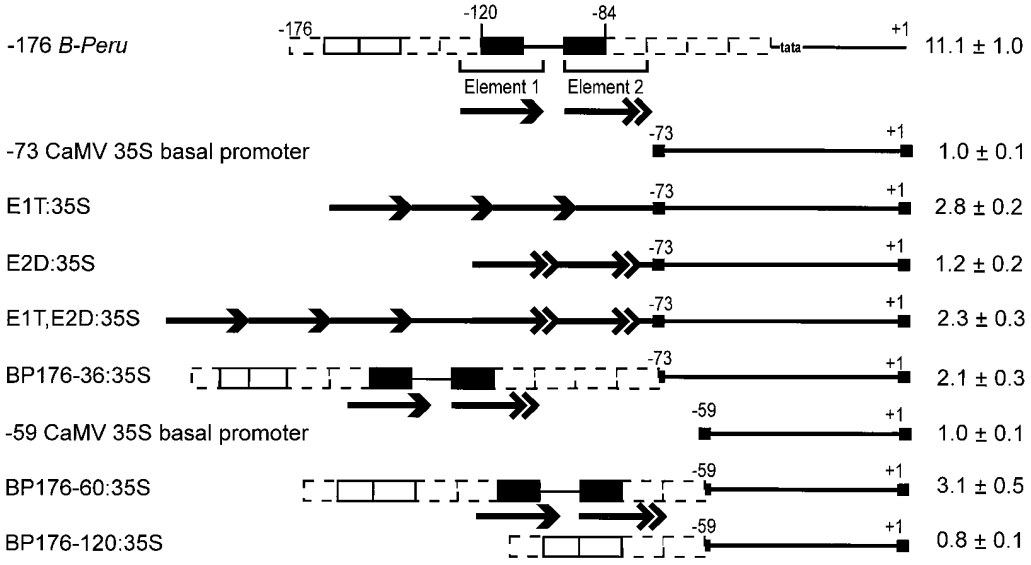

Figure 6.—The −176 to −60 promoter fragment of *B-Peru* and E1 alone, but not E2, modestly increase the aleurone expression of a heterologous minimal promoter. Oligonucleotides containing 24 bases of *B-Peru* sequence from −126 to −102, including the −120 to −108 E1 element (indicated by a single arrowhead), and from −96 to −72, including the −96 to −84 E2 element (indicated by a double arrowhead), were placed upstream of the −73 CaMV 35S promoter driving luciferase (E1T:35S; E2D:35S; E1T,E2D:35S). The region from −176 to −36 was also fused to the −73 CaMV 35S promoter (BP176-36:35S). Expression of each construct was normalized to the −73 CaMV 35S promoter alone, set at 1.0. Two regions, consisting of −176 to −60 (BP176-60:35S) or −176 to −120 (BP176-120:35S), of *B-Peru* were fused to the −56 CaMV promoter. Expression of these constructs was normalized to the −56 promoter that has a very similar, low level of expression as that of the −73 promoter.

sion. Second, the upstream block of divergent sequence found in *B-Peru* may be responsible for the plant expression.

The segmentation of the *B-Peru* promoter into a proximal region that is aleurone specific and more distal elements specific to plant expression is reminiscent of Drosophila promoters found in genes expressed in multiple tissues (Duncan 1987; Geyer and Corces 1987; Shimell *et al.* 1994). In contrast to the *B-Peru* allele, an *R-r* allele of the maize *r* gene, which is expressed in multiple tissues, has multiple promoters coupled to multiple functional coding regions (Walker *et al.* 1995). Of the other alleles of the *r* gene that have been characterized, the *R-stippled* allele also has multiple coding regions (Eggleston *et al.* 1995). However, the *R-navajo* allele, which produces a spot of color in the aleurone and produces some plant color, appears to have a single coding region (Dellaporta *et al.* 1988). In another maize regulatory gene with multiple alleles, the *p* gene, the tissue-specific differences between two alleles, *Prr* and *Pwr*, appear to be mediated by changes in the C terminus of the proteins produced by these alleles (Chopra *et al.* 1996).

Although the differences between the *B-Peru* and *B-I* alleles are caused by rearrangements in the promoter sequences, we do not know which allele is ancestral. There are, however, several pieces of evidence suggesting that *B-Peru* may be the derived allele. First, the 5′ end of the *B-I* coding region and sequences immediately upstream are homologous to the same region of the *r* allele *Lc*, which, like *B-I*, colors vegetative parts of the plant. Since the *r* and *b* genes diverged before the genus Zea was established, this suggests that *B-I* is closer to the ancestral form in its promoter structure. Second, in *B-Peru*, identical TATATATA sequences are present at both ends of the proximal insertion, whereas in *B-I*, a single copy of this sequence is present at the point of divergence (Radicella *et al.* 1992). This apparent target site duplication suggests that the block of divergent sequence in *B-Peru* is an insertion that may be a relic of a transposon or the result of a transposon-mediated event. Database searches did not turn up any significant homology between the entire 2.5-kbp *B-Peru* upstream sequence and any known transposon, but they did pick up significant identity between regions upstream of the triplicate repeat region and members of a class of proteins containing a conserved ATPase domain known as

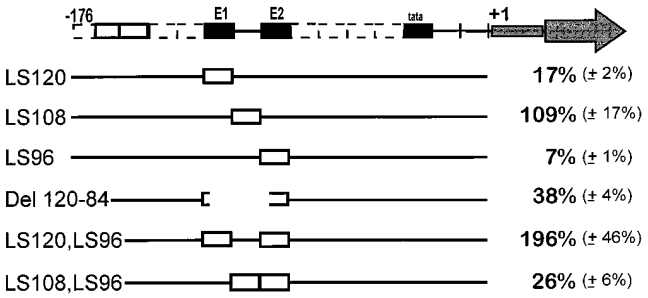

Figure 7.—Mutation of E1 and E2 together results in a higher level of expression than either mutation alone. A deletion mutation, indicated by a broken box, was produced using the *Xho*I sites of the LS120 and LS96 substitution. The double-substitution mutations have somewhat different linker sequences than the original mutations to avoid producing a repeat. These three mutations are in the same −176 promoter context as the single L-S mutations, and their expression is normalized to the −2500 construct.

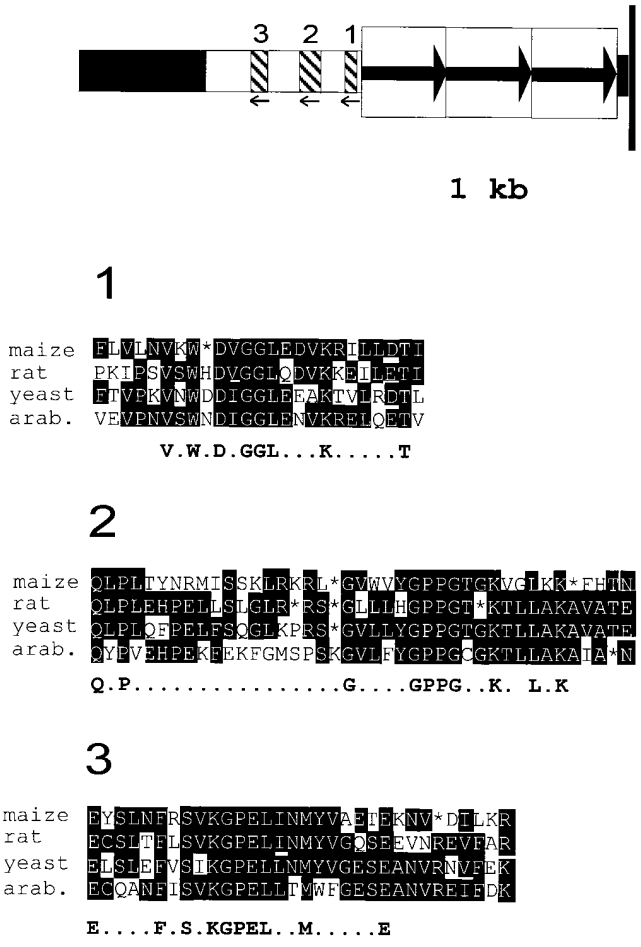

Figure 8.—The upstream region of *B-Peru* contains an apparent pseudogene. A homology search of GenBank using the BlastX utility revealed strong identity between the nonrepeated part of the upstream region of *B-Peru* and a family of proteins containing the AAA module. The diagram at the top shows the upstream region of *B-Peru*. The portion of the upstream sequence homologous to *B-I* is shown in black, the three 534-bp repeats are shown as white boxes with black arrows, and the first exon is represented by a tall, narrow black box. The three blocks of sequence with strong homology are numbered and cross-hatched. The small arrows underneath the cross-hatched boxes indicate the orientation of the reading frames, which are opposite of *b* transcription. The first block starts 9 bp upstream of the distal repeat (−1678) and runs upstream to −1852. The second block is frameshifted relative to the first and runs from −1886 to −2003. The third block runs from −2143 to −2239 in the same reading frame as the first block. Shown below are alignments of the potential amino acid sequences of these three blocks with the rat PAF-2 gene (GenBank accession number 1709557), a putative protein from *Schizosaccaromyces pombe* (GenBank accession number 2370473), and a CDC48 homolog from *Arabidopsis thaliana* (GenBank accession number 1705677). This alignment shows the strong sequence similarity between the four amino acid sequences.

the AAA module. The presence of part of a coding region from another gene adjacent to the promoter elements and running in the opposite direction as the *B-Peru* transcription unit is strong evidence that the *B-Peru* promoter region evolved elsewhere in the genome and was translocated to the *b* locus. Given the strong similarity of a part of the *B-Peru* upstream region with members of a conserved class of proteins, we are very interested in identifying the source gene for the *B-Peru* promoter. Southern blots revealed relatively few copies (6 to 15) of the *B-Peru* triplicate repeat region, which contains the promoter (data not shown, see Harris *et al.* 1994). This observation is consistent with either a low copy number transposon or with a promoter sequence shared by a multigene family. Regarding a transposon-based translocation as the origin of these sequences, two recent studies have shown that the *Mutator* transposable element in maize can induce complex rearrangements in promoter regions (Harris *et al.* 1994; Kloeckner-Gruissem and Freeling 1995). In addition, several recent studies have shown the importance of retrotransposons in the evolution in maize of genomic structure and possibly in the evolution of gene expression patterns (White *et al.* 1994; SanMiguel *et al.* 1996).

Our finding that the promoter region, rather than the coding region, of *B-Peru* is responsible for generating a novel expression pattern and, hence, a novel function fits a growing paradigm in which evolution acts on the regulatory sequences of a gene rather than on the protein-coding sequence. Expression of anthocyanins in the maize aleurone is likely to be a recent development given that the progenitor of maize, teosinte, has a cupulate seed case produced by the glume that completely encases the aleurone layer of the seed (Dorweiler *et al.* 1993). This case renders pigmentation in the aleurone invisible, as it cannot be seen unless the outer case is removed, which is not easily done, and, in teosinte, removal is not necessary for germination. To produce anthocyanin pigments in the aleurone, a functional allele of *c1* is required in addition to a *b* or *r* allele capable of seed color. A recent study concluded that the *C1* allele capable of seed color is a relatively recent derivative found in domesticated maize but not in most teosintes. The authors argued that selection for a visible kernel phenotype may have changed the frequencies of *c1* alleles and possibly enriched for a recombination event generating the *C1* allele capable of seed color (Hanson *et al.* 1996). It is intriguing that *B-Peru*, which appears to be derived from a progenitor that does not show aleurone expression, and the S component of the *R* standard allele (Walker *et al.* 1995), which is responsible for the aleurone color of this allele, both have promoter regions whose structures are reminiscent of transposon-mediated events.

The structure of the aleurone-specific promoter of *B-Peru* is similar to other maize genes in that the promoter region is quite small and close to the start of transcription. Other maize promoters have been characterized, and it is not unusual to find multiple regions required for activation. In the case of *c1*, hormonal and

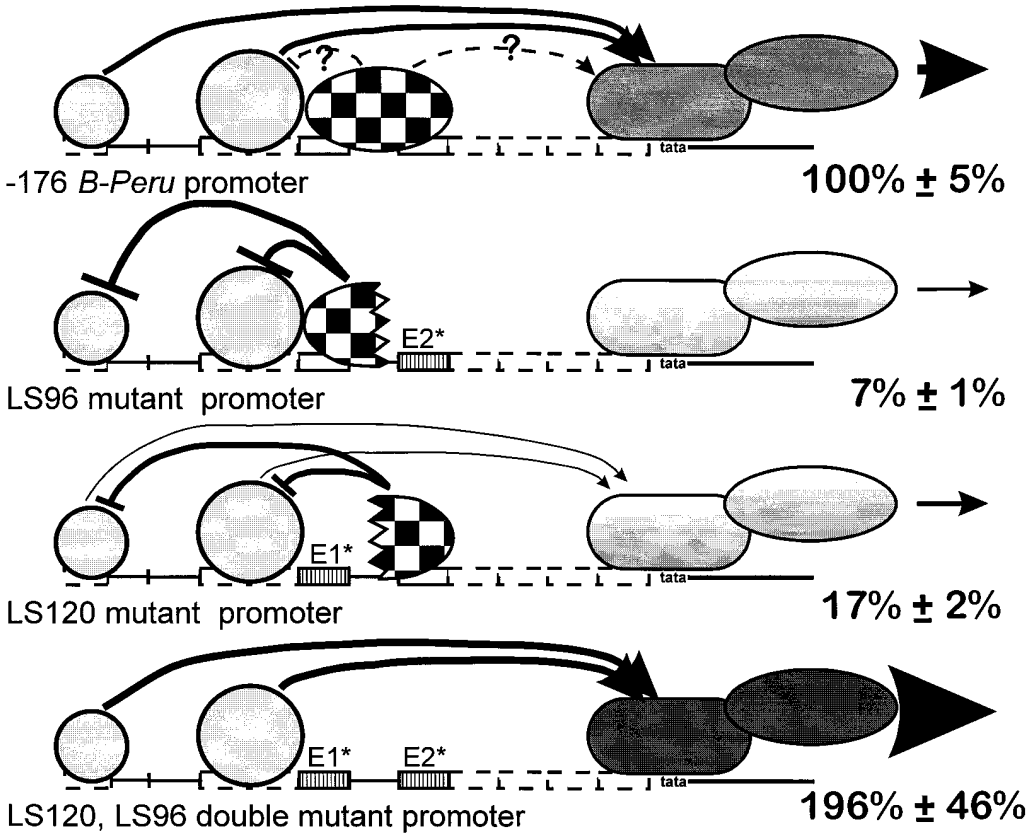

Figure 9.—A hypothetical model for the action of the E1 and E2 sequence elements. We propose that E1 and E2 are two parts of a bipartite regulatory element that is bound by a multicomponent factor, which is shown as the checkered oval. When complete, this factor allows the activation of the *B-Peru* promoter. If either one of the two parts is mutated, this prevents the interaction of some components of the hypothetical factor and results in the negative regulation of the promoter by the remaining components. When both sites are mutated, none of the components of the factor is able to bind, and activation, driven by other positive elements, is even stronger than in the native promoter. The placement of positive elements in this model upstream of E1 is based on the results of the deletion mutations; however, other locations for positive-acting factors are entirely consistent with our data. Sufficiency experiments indicate that positive-acting sequences in this region must be interacting with other factors bound downstream. These downstream factors were not diagramed for simplicity.

light regulation are mediated by nearby separate small elements (Hattori *et al.* 1992; Kao *et al.* 1996). Members of the anthocyanin biosynthetic genes (Roth *et al.* 1991; Tuerck and Fromm 1994; Bodeau and Walbot 1996; Lesnick and Chandler 1998) also have multiple nearby elements required for activation. The *B-Peru* promoter appears to consist of several elements whose interactions are complex. The results from the deletion mutations suggest that deletion of the nonrepeat sequences (from −2500 to −1244) increased expression. This increased expression was then lost as the extra repeat sequences are deleted. Although the simple interpretation of the results from the deletion mutations is that the most important region for activation is between −176 and −127, none of the L-S mutations in the −176 to −120 region produced a large loss of activity. In fact, two L-S mutations in this region produced increases in activity. Additionally, the inability of this region to increase expression of a heterologous promoter requires that the putative activator elements in this region act through sequences downstream of −120.

The L-S mutations do identify two critically important regions, E1 and E2, in the −120 to −84 region. Each of the single mutations dramatically reduced promoter activity. However, the increase in activity observed with a deletion spanning these two elements and with the LS120:LS96 double mutation suggests that E1 and E2 may function as negative elements rather than as positive ones. One model that fits the disparate pieces of data on E1 and E2 is shown in Figure 9. We hypothesize that E1 and E2 compose a bipartite negative regulator that requires the presence of two different factors (one binding to E1 and one to E2) to allow activators to properly function (Figure 9). In the absence of either site or factor, activation is repressed. In the plant, we envision that one of the two elements is bound in most or all tissues, and the second factor is aleurone specific. If both elements are removed, then expression is increased because of the loss of both parts of the regulatory element. In the model, we show the activators bound to upstream elements. This arrangement is based on the positive elements identified by the −127 and

−80 deletion mutations. It is entirely possible that other regions of the promoter bind to activators as well. Based on the L-S results and the inability of the −176 to −120 region to induce aleurone expression of a heterologous promoter, the putative elements shown in the model require elements downstream of −120 in the *B-Peru* promoter. These required elements may include the elements identified by the L-S mutations between −84 and −36, the specific TATA box sequence of *B-Peru*, or even the specific initiator and exon 1 sequences present in the native promoter that are not in the heterologous promoter.

An interaction between E1 and E2 is suggested by several experiments. The insertion scan mutations within LS108 provide strong evidence that E1 and E2 interact. The 4-bp insertion between the two elements dramatically reduced expression, potentially by altering an interaction either between E1 and E2 or between E1 and the TATA box. Because insertions at LS60, which is between both elements and the TATA box, produced no decrease in expression, the simplest explanation is that the interaction between E1 and E2 is perturbed by the altered distance. The small recovery of activity by inserting 21 bases instead of 4, thereby restoring the helical orientation of the two elements, also argues for an interaction. This interaction may not be sensitive to longer distances, as shown by the −710 L-S mutations. These mutations showed that the upstream region, including copies of E1 and E2, can compensate for the loss of the proximal copy of E1, but not E2. These observations may be explained by looping the DNA to bring the upstream E1 or other similar acting elements and E2 in contact when they are 534 bp apart. Insertions of 4 and 21 bases would disrupt the interaction between factors bound at these sites because looping is not possible over such short distances.

In addition to its possible role as a negative regulator, the E1 region can induce a small increase in the expression of a heterologous promoter in the aleurone. The 5′ deletion experiments suggest the existence of positive elements upstream of E1 and the L-Ss in this region identify several elements with small positive or small negative effects. The possible dual role of these various elements suggests that they may form composite elements with activities that vary depending on the composition of the protein complexes bound to them (reviewed in Kel *et al.* 1995). Composite elements were first proposed to explain the behavior of glucocorticoid response elements. In the case of the mouse proliferin gene promoter, it is the subunit composition of a DNA-binding factor that binds to a site within the same DNA sequence bound by the glucocorticoid receptor that determines whether it is a positive or negative regulator (Diamond *et al.* 1990).

Together, our experiments suggest that the positive activation of the *B-Peru* promoter requires many elements, most of which are tissue-specific in action. These elements are apparently further regulated by negative elements and induce high levels of expression with the aid of sequences in the nontranslated first exon of *B-Peru*. The complex regulation of this nonessential gene, together with the presence of an ATPase pseudogene on the upstream side of the promoter, lends credence to the idea that the promoter originated from another gene that has an important and tightly regulated function in the aleurone. It is also possible that the complex regulation evolved after the rearrangement that produced *B-Peru* because there is no apparent selection against the expression of *B-Peru* in the aleurone. Future work, including genetic and biochemical screens to isolate the factors binding to the aleurone-specific promoter of *B-Peru*, should allow us to clarify the roles of the different elements in this promoter. Additionally, this effort should allow us to characterize many of the genes required for the normal development and maintenance of the aleurone layer.

We thank Marc Lesnick and Wim Van Heeckeren for a critical reading of this manuscript. We are grateful to scientists at Ciba-Geigy (now Novartis) for assisting us in establishing the transgenic maize technology in our laboratory. We want to especially thank Stephen Evola and Gary Pace for enabling the technology transfer, as well as Jan Suttie, Erik Dunder, and John Dawson for sharing their time and expertise with us. We thank Devon Brown for construction and testing of the −2500 to −144 deletion mutations, and Susan Belcher for help in producing and maintaining the transgenic lines. D.A.S. was supported by a postdoctoral fellowship from the Jane Coffin-Childs Memorial Fund for Medical Research, and the work was supported by a grant 96-35301-3179 from the U.S. Department of Agriculture.

LITERATURE CITED

Bodeau, J. P., and V. Walbot, 1992 Regulated transcription of the maize *Bronze-2* promoter in electroporated protoplasts requires the *C1* and *R* gene products. Mol. Gen. Genet. **233:** 379–387.

Bodeau, J. P., and V. Walbot, 1996 Structure and regulation of the maize *Bronze2* promoter. Plant Mol. Biol. **32:** 599–609.

Callis, J., M. Fromm and V. Walbot, 1987 Introns increase gene expression in cultured maize cells. Genes Dev. **1:** 1183–1200.

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

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

Coe, E. H., Jr., 1979 Specification of the anthocyanin biosynthetic function by *B* and *R* in maize. Maydica **XXIV:** 49–58.

Cone, K. C., F. A. Burr and B. Burr, 1986 Molecular analysis of the maize anthocyanin regulatory locus C1. Proc. Natl. Acad. Sci. USA **83:** 9631–9635.

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.

Confalonieri, F., and M. Duguet, 1995 A 200-amino acid ATPase module in search of a basic function. BioEssays **17:** 639–650.

Diamond, M. I., J. N. Miner, S. K. Yoshinaga and K. R. Yamamoto, 1990 Transcription factor interactions: selector of positive or negative regulation from a single DNA element. Science **249:** 1266–1272.

Dellaporta, S. L., I. Greenblatt, J. L. Kermicle, J. B. Hicks and S. R. Wessler, 1988 Molecular cloning of the maize *R-nj* allele by transposon tagging with *Ac*, pp. 263–282 in *Chromosome Structure*

and Function: Impact of New Concepts, edited by J. P. Gustafson and R. Appels. Plenum Press, New York.

Dorweiler, J., A. Stec, J. Kermicle and J. Doebley, 1993 *Teosinte glume architecture 1*: a genetic locus controlling a key step in maize evolution. Science **262:** 233–235.

Duncan, I., 1987 The bithorax complex. Annu. Rev. Genet. **21:** 285–319.

Eggleston, W. B., M. Alleman and J. L. Kermicle, 1995 Molecular organization and germinal instability of *R-stippled* maize. Genetics **141:** 347–360.

Geyer, P. K., and V. G. Corces, 1987 Separate regulatory elements are responsible for the complex pattern of tissue-specific and developmental transcription of the *yellow* locus in *Drosophila melanogaster*. Genes Dev. **1:** 996–1004.

Goff, S. A., T. M. Klein, B. A. Roth, M. E. Fromm, K. C. Cone *et al.*, 1990 Transactivation of anthocyanin biosynthetic genes following transfer of *b* regulatory genes into maize tissues. EMBO J. **9:** 2517–2522.

Grotewold, E., B. J. Drummond, B. Bowen and T. Peterson, 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.

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 **143:** 1395–1407.

Harris, L. J., K. Currie and V. L. Chandler, 1994 Large tandem duplication associated with a *Mu2* insertion in *Zea mays B-Peru* gene. Plant Mol. Biol. **25:** 817–828.

Hattori, T., V. Vasil, L. Rosenkrans, L. C. Hannah, D. R. McCarty *et al.*, 1992 The *viviparous-1* gene and abscisic acid activate the *C1* regulatory gene for anthocyanin biosynthesis during seed maturation in maize. Genes Dev. **6:** 609–618.

Kao, C.-Y., S. M. Cocciolone, I. K. Vasil and D. R. McCarty, 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.

Kel, O. V., A. G. Romaschenko, A. E. Kel, E. Wingender and N. A. Kolchanov, 1995 A compilation of composite regulatory elements affecting gene transcription in vertebrates. Nucleic Acids Res. **23:** 4097–4103.

Klein, T. M., B. A. Roth and M. E. Fromm, 1989 Regulation of anthocyanin biosynthetic genes introduced into intact maize tissues by microprojectiles. Proc. Natl. Acad. Sci. USA **86:** 6681–6685.

Kloeckner-Gruissem, B., and M. Freeling, 1995 Transposon-induced promoter scrambling: a mechanism for the evolution of new alleles. Proc. Natl. Acad. Sci. USA **92:** 1836–1840.

Koziel, M. G., G. L. Beland, C. Bowman, N. B. Carozzi, R. Crenshaw *et al.*, 1993 Field performance of elite transgenic maize plants expressing an insecticidal protein derived from *Bacillus thuringiensis*. Bio/Technology **11:** 194–200.

Lesnick, M. L., and V. L. Chandler, 1998 Activation of the maize anthocyanin gene *a2* is mediated by an element conserved in many anthocyanin promoters. Plant Phys. (in press).

Lorenz, W. W., R. O. McCann, M. Longiaru and M. J. Cormier, 1991 Isolation and expression of a cDNA encoding *Renilla reniformis* luciferase. Proc. Natl. Acad. Sci. USA **88:** 4438–4442.

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

Matzke, M. A., and A. J. M. Matzke, 1995 How and why do plants inactivate homologous (trans)genes? Plant Physiol. **107:** 679–685.

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.

Radicella, J. P., D. Brown, L. A. Tolar and V. L. Chandler, 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.

Roth, B. A., S. A. Goff, T. M. Klein and M. E. Fromm, 1991 C1- and R-dependent expression of the maize bz1 gene requires sequences with homology to mammalian myb and myc binding sites. Plant Cell **3:** 317–325.

Sainz, M. B., S. A. Goff and V. L. Chandler, 1997 Extensive mutagenesis of a transcriptional activation domain identifies single hydrophobic and acidic amino acids important for activation *in vivo*. Mol. Cell. Biol. **17:** 115–122.

SanMiguel, P., A. Tikhonov, Y-K. Jin, N. Motchoulskaia, D. Zakharov *et al.*, 1996 Nested retrotransposons in the intergenic regions of the maize genome. Science **274:** 765–768.

Shimell, M. J., J. Simon, W. Bender and M. B. O'Conner, 1994 Enhancer point mutation results in a homeotic transformation in *Drosophila*. Science **264:** 968–971.

Styles, E. D., O. Ceska and K. Seah, 1973 Developmental differences in action of *R* and *B* alleles in maize. Can. J. Genet. Cytol. **15:** 59–72.

Tuerck, J. A., and M. E. Fromm, 1994 Elements of the maize *A1* promoter required for transactivation by the anthocyanin *B*/*C1* or phlobaphene *P* regulatory genes. Plant Cell **6:** 1655–1663.

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

White, S. E., L. F. Habera and S. R. Wessler, 1994 Retrotransposons in the flanking regions of normal plant genes: a role for *copia*-like elements in the evolution of gene structure and expression. Genetics **91:** 11792–11796.

Communicating editor: J. Chory