Major recent and independent changes in levels and patterns of expression have occurred at the *b* gene, a regulatory locus in maize

David A. Selinger and Vicki L. Chandler*

Department of Plant Sciences, University of Arizona, 303 Forbes Hall, Tucson, AZ 85721

Communicated by Susan R. Wessler, University of Georgia, Athens, GA, October 25, 1999 (received for review August 3, 1999)

The *b* locus encodes a transcription factor that regulates the expression of genes that produce purple anthocyanin pigment. Different *b* alleles are expressed in distinct tissues, causing tissue-specific anthocyanin production. Understanding how phenotypic diversity is produced and maintained at the *b* locus should provide models for how other regulatory genes, including those that influence morphological traits and development, evolve. We have investigated how different levels and patterns of pigmentation have evolved by determining the phenotypic and evolutionary relationships between 18 alleles that represent the diversity of *b* alleles in *Zea mays*. Although most of these alleles have few phenotypic differences, five alleles have very distinct tissue-specific patterns of pigmentation. Superimposing the phenotypes on the molecular phylogeny reveals that the alleles with strong and distinctive patterns of expression are closely related to alleles with weak expression, implying that the distinctive patterns have arisen recently. We have identified apparent insertions in three of the five phenotypically distinct alleles, and the fourth has unique upstream restriction fragment length polymorphisms relative to closely related alleles. The insertion in *B-Peru* has been shown to be responsible for its unique expression and, in the other two alleles, the presence of the insertion correlates with the phenotype. These results suggest that major changes in gene expression are probably the result of large-scale changes in DNA sequence and/or structure most likely mediated by transposable elements.

The *b* gene encodes a transcriptional activator of the anthocyanin pigment pathway in maize. Anthocyanins are one of the main classes of pigments involved in flower color and thus in attracting pollinators. In maize and other grasses that are wind pollinated, anthocyanin expression is frequently found in many of the vegetative tissues, and its function in grasses is unknown. The anthocyanin pigments are related to other flavonoid compounds that are important for fertility and for defense against insects and pathogens (reviewed in refs. 1, 2). Thus, it is possible that anthocyanins may have some role in defense. Maize plants that lack the ability to produce anthocyanins exhibit no growth defects relative to those that do. Several roles for anthocyanin pigmentation in maize have been proposed. Anthocyanin pigments absorb UV light as well as visible light and have been suggested to serve as UV protectants (3). It has also been proposed that anthocyanin expression in the leaf sheaths and stems of maize and teosinte plants growing at higher altitudes is an adaptation to cooler temperatures and may help the plants attain higher plant temperatures during the day (4). Because maize is a cultivated plant, human selection over the past ≈7,000 years of maize cultivation has probably led to the selection of plants with dramatic pigment patterns purely for aesthetic reasons.

The basic helix-loop-helix protein encoded by the *b* gene (5) activates the transcription of the biosynthetic genes (6) and is ≈70% identical at the amino acid sequence level to, and is functionally interchangeable with, the product of the maize *r* gene (7, 8). The synthesis of anthocyanin pigment in the plant and seed is very sensitive to the expression levels of the *b* gene over at least a 10-fold range of expression (6). A large number of different alleles conferring distinct tissue-specific anthocyanin synthesis have been described for both the *b* and *r* genes (9, 10). We have investigated the relationships between the phenotypically diverse *b* alleles and the underlying causes of the phenotypic differences between these alleles. Because *b* is a regulatory gene, understanding how phenotypic diversity is produced and maintained at the *b* locus should provide models for how other regulatory genes, including those that influence morphological traits and development, evolve.

Previous studies have focused on two very distinct *b* alleles: *B-I*, which confers intense pigmentation of most vegetative tissues but does not induce pigmentation of the seed tissues, and *B-Peru*, which confers strong expression in the seed but induces only weak pigmentation in some vegetative parts of the plant. *In vitro* assays and intragenic recombination experiments with these two alleles found that the region upstream of the transcription start site determines the pattern of pigmentation (11, 12). The upstream sequences of these two alleles are completely different for the first 2.5 kb upstream, but the transcribed regions are greater than 95% identical (12). Characterization of a transposon-derived allele of *B-Peru* and careful restriction mapping revealed that the sequences in *B-Peru,* upstream of this 2.5-kb block, are homologous to those immediately upstream of the start of transcription in *B-I* (11, 13). A thorough characterization of this 2.5-kb block of sequence revealed that it contains all of the elements required for the seed-specific expression seen in *B-Peru*, but that it conferred no plant expression (14). The finding that part of the 2.5-kb upstream sequence has potential amino acid homology to a conserved class of proteins suggested that the *B-Peru* allele resulted from the translocation of the aleurone-specific promoter from a different maize gene into the upstream region of *B-Peru* (14).

On the basis of previous findings with *B-I* and *B-Peru*, we hypothesized that *b* alleles with major differences in expression would have large-scale structural differences in the upstream region. To test this idea, and to determine the relationships between some of the many known *b* alleles and possible progenitor alleles from wild relatives of cultivated maize, we cloned a 600- to 700-bp region of the *b* gene from 18 different alleles and determined the relationships between these alleles. To correlate phylogeny with phenotype, we crossed all of the alleles into a standard background and determined that the pigment phenotype segregated with the *b* locus. Superimposing the phenotypes on the molecular phylogeny reveals that the relatively few alleles with strong and distinctive patterns of expression are closely

EVOLUTION

Abbreviations: MITE, miniature inverted-repeat transposable element; RFLP, restriction fragment length polymorphism.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. AF205792–AF205807).

*To whom reprint requests should be addressed. E-mail: chandler@ag.arizona.edu.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

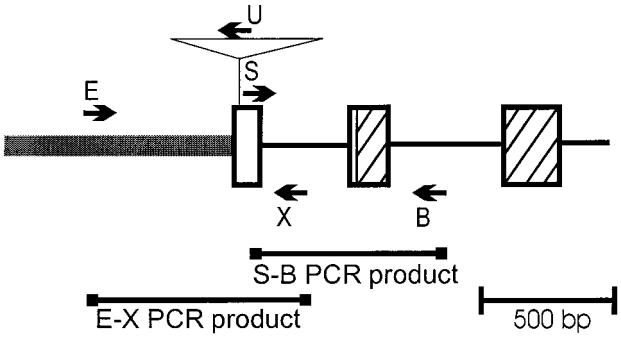

Fig. 1. Diagram of PCR primers and products. The map shows the upstream region (thick line), exon 1 (open box), which is not translated, exons 2 and 3 (the cross-hatched area indicates translated sequence), and introns 1, 2, and 3 (thin lines) of the *b* gene. Exons 4–9 and introns 4–8 are not shown. The insertion (depicted by the triangle) represents sequences found in *B-Peru* and *B-Bolivia*, and U is a primer specific to this region that was used to amplify *B-Bolivia* upstream sequences. The locations of the primers used in this paper and the PCR products generated are shown. See *Materials and Methods* for the primer sequences and exact locations.

related to alleles with weak expression, implying that the distinctive patterns have arisen recently. In addition, three of the five alleles with strong phenotypes, including *B-Peru*, have insertions in the promoter proximal region that are unique to those alleles.

Materials and Methods

Maize Nomenclature. Consistent with maize genetics nomenclature (details at http://www.agron.missouri.edu), dominant alleles of genes are capitalized (e.g., *B-Peru*), and recessive alleles are in lower case (e.g., *b-W23*). Gene names are in lower case but are not hyphenated (e.g., *b*).

PCR and Cloning. The E, S, and B primers were designed on the basis of previously cloned sequences of the *B-I* (GenBank accession no. X70790), *B-Peru* (X57276) and *b-W23* alleles, and primer X was designed on the basis of an alignment of 20 different S-B PCR product sequences. A diagram of the locations and orientations of the primers is found in Fig. 1. Primer E (5′-ATAGTAGCCATATACAACTTTGC-3′) is at −648, primer S (5′-GTGGAGGTGAGCTCCTCCT-3′) is at +98, primer X (5′-TACAACAACTCGATTCTAGATCGG-3′) is at +242, and primer B (5′-CTACAATCGCATCACGGGATCC-3′) is at +735; these coordinates are relative to the start of transcription in *B-I*. Primer U (5′-TGAGCCGTTAGATTTCATCTCC-3′) is 623 bp upstream of the start of transcription in *B-Peru* (Fig. 1). We supplemented the buffer supplied with the *Taq* polymerase (Qiagen, Valencia, CA) with $MgCl_2$ to a final concentration of 2.5 mM and used the following touchdown PCR program to amplify *b* allele sequences: 2 min at 94°C, [30 sec at 94°C, 30 sec at 53C° (−1°/cycle), 2 min at 72°C] for 10 cycles, then 25 cycles with a constant 43° annealing temperature. PCR reactions were subjected to gel electrophoresis; those with a single band were purified via a Qiagen PCR purification spin column, and for those that produced two bands, each band was gel purified. We ligated the purified DNA into pGEM-T-Easy (Promega) according to the manufacturer's directions and electroporated into XL1-blue MRF′ cells.

Sequencing, Finishing, Alignment, and Phylogenetic Analysis. At least two clones of each *b* allele were sequenced on both strands by using Applied Biosystems Sequencers (Arizona Research Laboratories, University of Arizona, Tucson, AZ). The machine-called sequence was checked against the electropherograms by using the CHROMAS program (available from http://www.microbiology.adelaide.edu.au/www/ftp/molbiol/chromas), edited, and the sequence exported to the SEQUED program (Wisconsin package, ver. 9.1, Genetics Computer Group, Madison, WI). Finished sequences were aligned by using the CLUSTAL W program (http://dot.imgen.bcm.tmc.edu:9331/multi-align) and adjusted by hand to correct misaligned insertion/deletion polymorphisms. The multiple sequence alignment was used for phylogenetic analysis with the DNAPARS, DNADIST, DNAML, DNAMLK, and NEIGHBOR programs from the PHYLIP package (15).

Results

Phenotypic Characterization of *b* Alleles from Diverse Sources. Seeds for a variety of exotic maize land races (cultivars of maize grown by indigenous peoples) and from several collections representing all three of the recognized wild subspecies of *Zea mays* (*Zm*), teosintes, were obtained from a wide geographic area in Central and South America (Table 1). We crossed the plants grown from these accessions with our standard *b* tester stock that has homozygous dominant functional alleles of all of the anthocyanin genes except for those of *b* and *r*, for which it carries recessive nonfunctional alleles. The resulting plants were scored for color phenotype, and those displaying pigmentation were repeatedly backcrossed and selected by phenotype for three to five generations. During this process of introgression, we tested whether each *b* allele cosegregated with plant pigmentation using two linked morphological markers: *glossy2* (*gl2*), 19 map units distal to *b,* and *white tip* (*wt*), 11 map units centromere proximal to *b*. The recessive alleles of *gl2* and *wt* confer easily scored phenotypes on homozygous seedlings. For each allele, we created a population segregating the new *b* allele and the marked *gl2 b wt* allele. All plants were scored for pigment phenotype when they began to flower, and linkage between the new *b* allele and the pigment phenotype was detected by cosegregation of the anthocyanin pigment phenotype with wild-type phenotypes for both the *gl2* and *wt* markers. In all cases, the new *b* allele showed tight linkage to the pigment phenotype (Table 1).

Most of the alleles exhibited a pigmentation pattern consisting of a spray-like pattern of very small spots on the sheath and streaks of color on the culm of the plant (Fig. 2*A*) or of distinct typically circular nonclonal spots (Fig. 2*B;* see Table 1 for phenotype classes). These patterns were usually combined to some degree. Almost all the alleles pigmented the glume bar, which is the node where the tassel glumes originate (Fig. 2*C*). Although some alleles produced more color than others, these differences were more quantitative than qualitative (supplemental Fig. 5; see www.pnas.org., which shows the phenotypes of all the alleles). Because quantitative differences in pigmentation could be caused by spontaneous reversion of the *Pl-Rh* allele of the *pl* locus, which is required for strong pigmentation, to the weaker expressing *Pl′-mah* allele (16), we crossed the plants carrying exotic alleles to tester plants carrying the *r-r* allele, which produces dark anther color with the *Pl-Rh* allele but not with *Pl′-mah*. By scoring anther color, we were able to find *Pl-Rh* plants to phenotype and photograph for each allele. We were unable to perform crosses between the two *Zm huehuetenangensis* accessions and the tester. However, the plants of this subspecies exhibited pigmentation that was similar to the *mexicana* and *parviglumis* teosinte plants.

Although most of the *b* alleles from cultivars of maize were phenotypically similar to the teosinte alleles, five differed significantly in pattern and intensity. Two alleles from the exotic lines *B-Gua31* and *B-Mex7b* had distinct patterns of pigmentation. The *B-Gua31* allele induced dark pigmentation of the culm and most of the sheath (Fig. 2*D*) that was very similar to that of the previously characterized *B-Bolivia* allele (Fig. 2*E*). Unlike *B-Bolivia*, *B-Gua31* confers no seed pigmentation. The *B-Mex7b*

Table 1. Source and characterization of *b* alleles used in this study

Allele	Source	Subspecies	Geographic location	Phenotype class*	Segregation analysis†	Clade‡
B-bar	E.H. Coe, Jr.§	Cultivated	North America	A	–	II
B-Bolivia	E.H. Coe, Jr.	Cultivated	Bolivia	B,C	–	I
B-I	E.H. Coe, Jr.	Cultivated	North America	D	–	III
B-Marker	E.H. Coe, Jr.	Cultivated	North America	A	–	I
B-Peru	E.H. Coe, Jr.	Cultivated	Peru	C,E	–	I
B-615	Novartis CG00526	Cultivated	North America	A	36/38	III
b-W23	W23 inbred	Cultivated	North America	–	–	II
B-Gua31	M. Goodman¶	Cultivated	Guatemala	B	10/10	I
B-Mag466	M. Goodman	Cultivated	Venezuela	A	8/8	I
B-Mex7a	M. Goodman	Cultivated	Mexico	A	19/19	III
B-Mex7b	M. Goodman	Cultivated	Mexico	F	12/13	III
B-M031	J. Doebley‖	*huehuetenangenis*	Guatemala	A	N.D.	
B-M033	J. Doebley	*huehuetenangensis*	Guatemala	A	N.D.	
B-M075	J. Doebley	*mexicana*	Mexico	A	16/18	I
B-M092	J. Doebley	*mexicana*	Mexico	A	34/35	I
B-M046	J. Doebley	*parviglumis*	Mexico	A	51/51	I
B-M063	J. Doebley	*parviglumis*	Mexico	A	18/20	I, II**
B-M106	J. Doebley	*parviglumis*	Mexico	A	23/23	III

§U.S. Department of Agriculture—Agricultural Research Service, Columbia, MO; ¶North Carolina State University, Raleigh, NC; ‖University of Wisconsin, Madison, WI; *Class A has weak pigmentation (see Fig. 2 *A–C*), B has strong pigment in some plant parts (Fig. 2 *D*, and *E*), C indicates seed pigmentation, D is strong in most plant parts (Fig. 2*G*), E is like class A but with much more tassel pigment (Fig. 2*I*), F is strong along margins of the sheath and around nodes (Fig. 2*F*).

†The numbers shown indicate the number of plants with a pigment phenotype/the total number of plants that were phenotypically wild type for both recessive markers that are linked to the *b*-tester allele. ND, not determined because of an inability to introgress these alleles. A 1/N-linkage had been previously determined by others. We observed a slightly higher level, 3.4% (8/235), of apparent double crossover events than expected (~2%), probably because of occasional misscoring of markers.

‡The roman numerals indicate which clade each allele belongs to based on phylogenetic analysis.

**This accession segregated two *b* alleles; we reported the sequence of the allele in clade II. The clade I allele is very similar to the *B-M046* allele.

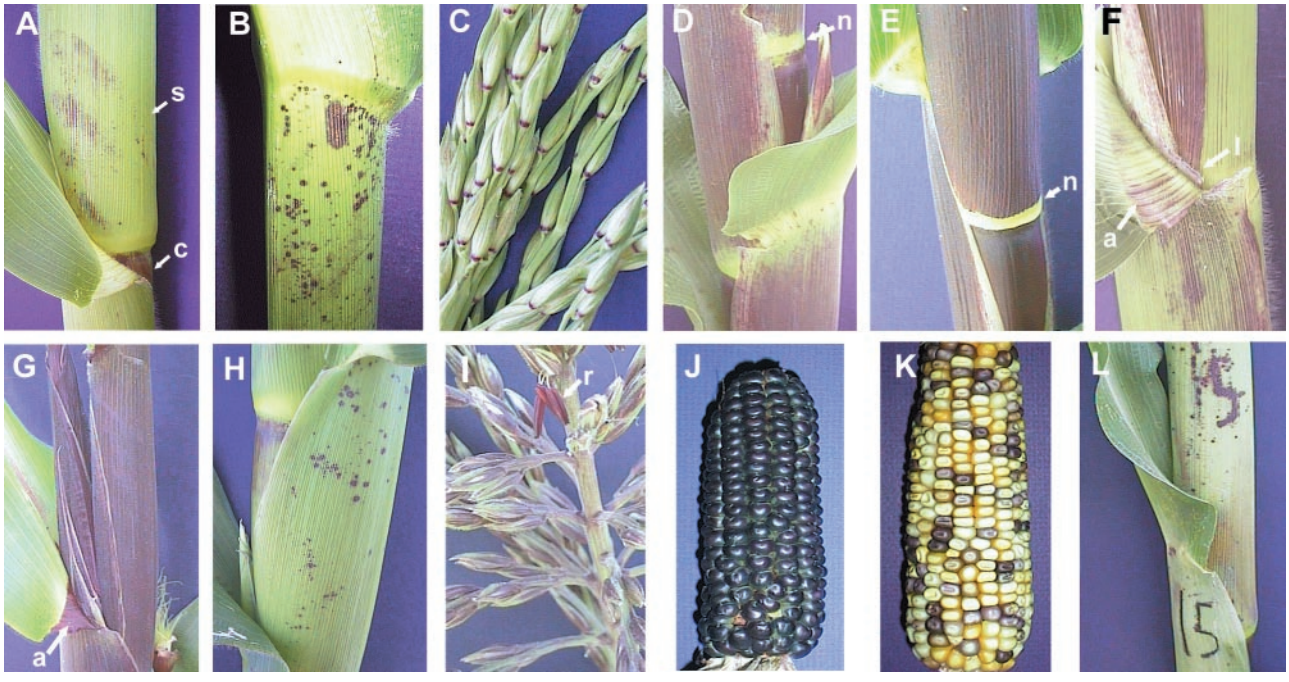

Fig. 2. Phenotypes of *b* alleles. *A–D* and *F* are photographs of plants carrying exotic or teosinte-derived *b* alleles, which are the result of at least three backcross generations into our *b* tester stock. (*A*) culm (c) and sheath (s) of *B-Mag466* showing the ''spray'' pattern on the sheath. (*B*) Sheath of *B-Mex7a* showing nonclonal spots. (*C*) Tassel branches showing the ''glume bar'' phenotype from *B-M063*. (*D*) *B-Gua31* plant. Note the unpigmented node (n). (*E*) *B-Bolivia* plant. Note the similarity to *B-Gua31*, including the unpigmented node (n). (*F*) auricle (a), ligule (l), and sheath of *B-Mex7b*. (*G*) Auricle (a) and sheath of *B-I*. (*H*) Sheath and culm of *B-Peru*. (*I*) Tassel branches of *B-Peru* [anther (r) color is caused by an allele of the *r* gene]. (*J*) Ear from a *B-Peru* homozygote. (*K*) Ear from a *B-Bolivia* homozygote. (*L*) Consecutive sheaths of an F_1 plant from the cross between Gua31 and the *b* tester stock showing the ''carbon-copy'' response. The plant was numbered when the upper sheath was mostly covered by the lower one, several weeks before this picture was taken. This plant also illustrates the effect of genetic background on expression. Compare F_1 in *L* to a plant with the same allele after three backcross generations into our *b* tester stock (*D*).

EVOLUTION

allele was isolated from the Cacahuacintle (Mex7) stock and produces dark pigmentation in a wide band starting at the margins of the sheath, in the auricle, which separates the sheath from the leaf, in the ligule (Fig. 2*F*), and in the nodes (not shown). This pattern contrasts with *B-Gua31* and *B-Bolivia*, which are pigmented in the central part of the sheath but not in the auricle, nodes, or margins. The *B-I* and *B-Peru* alleles have distinct patterns that have been described previously (9, 12). *B-I* intensely colors most plant tissues (Fig. 2*G*), including all parts of the tassel (not shown). *B-Peru* weakly colors the sheath and culm (Fig. 2*H*), but produces significant pigment in the tassel (Fig. 2*I*). The only alleles that conferred seed color were the *B-Peru* and *B-Bolivia* alleles (Fig. 2 *J* and *K*).

The basic pattern of spots on the sheath may be related to defense against pathogens. We observed that anthocyanin expression often occurred in the cells under and surrounding the numbers that we write on the sheaths of the plants to keep track of them. In one extreme case, several members of a family exhibited a "carbon-copy" response to the writing of a number on the overlying sheath. As the plants grew, the nodes elongated, and the sheath that was below the one written on became exposed and displayed a perfect copy of the number in anthocyanin "ink" (Fig. 2*L*).

Structural Characterization of the Upstream Region. As a first step toward determining the relationships between the alleles, we used DNA blots to characterize restriction fragment length polymorphisms (RFLPs) in several of the alleles (data not shown). Almost all of the alleles had identical RFLP maps in the transcribed region, but most differed in the upstream region. To get a better idea of the relationships between the alleles, we used PCR to clone the 5′ end of the transcribed region between the S and B primers (Fig. 1). Comparison of the sequence of this clone from various alleles led to the discovery of a large number of recombination events in this region, and thus we were not able to determine a reliable phylogeny from these data (D.S. and V.C., unpublished results). We next used PCR to amplify the region immediately upstream of the start of transcription. Using the E and X primers, we successfully amplified a fragment from all alleles tested except for *B-Bolivia*. Because of a large insertion in *B-Bolivia*, we used a new downstream primer (U) that is based on the sequence of the related *B-Peru* allele with the E primer to produce a 2-kbp E-U product, part of which is homologous to the E-X region.

Cloning and Sequencing of the Upstream Region of *b* Alleles. We cloned and sequenced PCR products of the E-X region from the 18 different alleles. Comparison of the aligned sequences revealed the presence of several small transposon or transposon-like insertions, which are diagrammed in Fig. 3. Some of these insertions represent previously characterized transposons, and several, to our knowledge, are novel, as described below.

The *B-bar* and *b-W23* alleles contain a 406-bp insertion 28 bp downstream of the exon1/intron 1 junction that is flanked by GCCTAA direct repeats and has an 11-bp inverted repeat. This insertion, which we have named *Snowbird*, has homology to two other sequences in maize: 90% identity over 299 bp to an insertion in a *Ds* element (GenBank accession no. L33921) and 85% identity over 269 bp to an insertion in the 3′ flanking region of the *histone H2B* gene [European Molecular Biology Organization no. X69960; see supplemental Fig. 7 (www.pnas.org)]. One allele from M063 is closely related to *B-bar* and *b-W23* but lacks this insertion.

The PCR products from five alleles share a 127-bp insertion previously identified from the *B-I* sequence as *Tourist-Zm23* (17). In addition to this insertion, the *B-Mex7b* allele has a unique 862-bp insertion located 220 bp upstream of the start of transcription relative to the *B-I* allele (Fig. 3). This element,

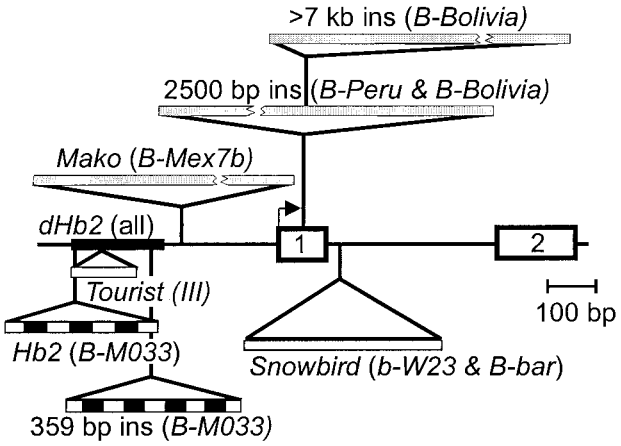

Fig. 3. Insertions in the various *b* alleles. The upstream region and the 5′ end of the transcribed sequence of *b* are shown. The numbered boxes denote exons, and the arrow shows the start of transcription. The deleted *Hb2* (*dHb2*) MITE sequence, which is conserved in all *b* alleles studied, is shown as a black box in the upstream region. The positions are shown to scale, as are the sizes of the small (<500-bp) insertions. The >7-kb insertion in *B-Bolivia*, 2.5-kb insertion in *B-Peru* and *B-Bolivia* and the *Mako* element in *B-Mex7b,* are shown in gray to indicate that they cause or correlate with the pigment phenotype. The *Tourist* element and the *Snowbird* MITEs are unshaded to indicate that they are conserved between multiple alleles and do not affect phenotype. The *Hb2* and 359-bp insertions, which are striped, are unique to *B-M033,* and their affect on expression is not known.

which we have named *Mako,* is flanked by 8-bp direct repeats and has 33- to 35-bp inverted repeats. The *Mako* sequence has 72% identity to a 423-bp region in the 5′ nontranscribed sequence of the maize *Cyp71c1* gene (European Molecular Biology Organization no. X81828) and 63% identity to a 95 bp region of the first intron of the *dapA* gene (GenBank accession no.U61730) of *Coix lacryma-jobi* [see supplemental Fig. 8 (www.pnas.org)].

The *B-M033* allele contains two insertions of 316 bp and 359 bp that, when removed, yield a sequence similar in size to that of the other *Zm huehuetenangensis* allele, *B-M031*. The 316-bp insertion in *B-M033*, flanked by 5-bp CTTAG direct repeats, has 90% identity to the 316-bp *Hb2* miniature inverted repeat transposable element (MITE) first identified in the maize *waxy* gene (ref. 18; European Molecular Biology Organization no. X06934). The second insertion in *B-M033* is 359 bp in size and has 84% identity to 189 bp of the 3′ nontranslated region of the maize *oleosin* KD18 sequence (GenBank accession no. J05212).

Phylogenetic Analysis of the *b* Sequences. The 18 aligned sequences averaged 594 bp and contained 116 single-nucleotide polymorphisms. In addition to the large insertions already described, we found 30 small insertions from 1 and 15 bp in size; 20 of these were full or partial direct repeats of the adjacent sequence, suggesting that they are footprints of now excised transposons [details of the polymorphisms found in the aligned sequence are in supplemental Fig. 6 (www.pnas.org.)]. To facilitate alignment, we recoded the large insertions as small insertions by removing the insertion sequence, leaving just the direct repeats. We used parsimony analysis to construct phylogenetic trees using both insertion/deletion and single-nucleotide polymorphisms. This approach generated a phylogenetic tree in which most branches were well supported when tested by bootstrap methods (ref. 19; Fig. 4*A*). To provide an additional check on the validity of the tree, we analyzed the data using a neighbor-joining algorithm that uses the distances between the sequences. Because the distance program ignores insertions and deletions, the neighbor-joining tree, which is shown in Fig. 4*B*, is derived from

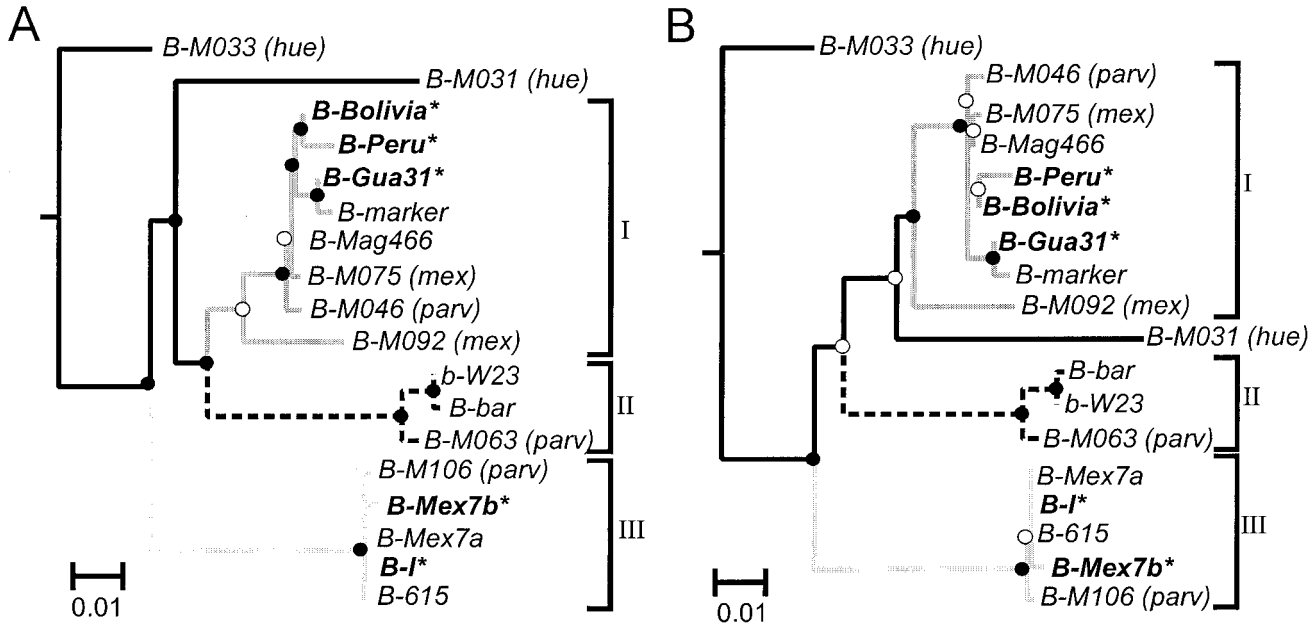

Fig. 4. Phylogenetic relationships of the *b* alleles. (*A*) The tree resulting from parsimony analysis. (*B*) The tree resulting from neighbor joining analysis. In both trees, confidence levels for branches were determined by bootstrap analysis. A black dot at the node indicates an 80–100% confidence level, an open dot indicates 50–80% confidence, and branches with less than 50% support were collapsed down one level. The bar indicates a distance of 0.01 substitutions/bp (1% sequence difference). Branch lengths for both trees were determined by maximum likelihood analysis. The members of clades I, II, and III are indicated by brackets and with gray lines for I and III and dashed lines for II. Alleles with distinct phenotypes are starred and in bold type.

the single-nucleotide polymorphisms only. The two trees have very similar topologies and, by using maximum likelihood analysis, the statistical likelihood of the trees is not significantly different. Therefore, the two trees are equally likely to be correct. Using maximum likelihood, we also tested whether a molecular clock was valid for comparing all the lineages. The log likelihood value for the consensus parsimony tree computed without assuming a clock hypothesis is significantly lower than the value computed for the same tree with the molecular clock hypothesis. This result rejects the molecular clock hypothesis, because at least some of the lineages have evolved at different rates.

The results of the phylogenetic analysis showed three well-supported clades in both trees, which we have labeled I, II, and III. Although each of these clades contains distinct insertion/deletion polymorphisms, the strong support for these clades in the neighbor-joining tree, which is based solely on nucleotide polymorphisms, indicates that the clades are defined independently of the insertion/deletion polymorphisms. Interestingly, 12 of the 20 small insertion polymorphisms that appear to be transposon footprints occur in one member of a closely related pair of alleles. In fact, six of these footprints are polymorphic in allelic pairs that differ by zero or one nucleotide substitution. These results indicate that most of the insertions at *b* are recent events.

By averaging the distances between the *b* alleles computed from nucleotide substitutions, we found that the alleles within each clade are much more closely related than are alleles in different clades. Depending on which pair was used, the average distances between alleles of clades I, II, and III were 5- to 50-fold greater than the average for alleles within each clade. These findings indicate that the alleles within each clade have arisen quite recently compared with the divergence of the three different clades.

Discussion

Although most of the *b* alleles have similar patterns of pigmentation, five have distinct patterns of expression. Interestingly, two closely related alleles in clade I, *B-Peru* and *B-Gua31*, have completely different phenotypes. Yet both of these closely related alleles share strong phenotypic similarities with distinct phenotypic aspects of a third closely related allele, *B-Bolivia.* *B-Peru* and *B-Bolivia* are the only alleles with seed pigment and are closely related on the basis of results of this study and of shared insertional sequences previously investigated. The 2.5-kb insertion in *B-Peru* confers seed expression but not plant expression (14). *B-Bolivia* has at least 2 kb of this 2.5-kb insertion including the promoter elements necessary for seed expression. Within this *B-Peru*-like sequence, *B-Bolivia* has an insertion of at least 7 kb in size that contains a sequence not found in other *b* alleles (D.S. and V.C., unpublished results). It is possible that this insertion reduces the amount of seed expression and produces the variable seed expression associated with this allele. The *B-Gua31* allele has a plant pigment phenotype that is very similar to *B-Bolivia* but lacks seed pigment. *B-Gua31* is missing the sequences found in *B-Peru* and *B-Bolivia* that confer seed expression. However, *B-Gua31* has a direct repeat that is consistent with the footprint expected if the insertion of the seed-specific sequences found in *B-Peru* and *B-Bolivia* were removed. The strong similarities in the plant phenotypes of *B-Gua31* and *B-Bolivia* suggest that they share regulatory elements. Because the two alleles do not share the large insertion in *B-Bolivia*, the plant-specific regulatory elements are likely to be located elsewhere.

All of the alleles in clade III are very closely related, and most have a weak pigment phenotype. However, two alleles in this clade have distinctive strong pigment phenotypes: *B-Mex7b*, which pigments the margins of the sheath, and *B-I*, which strongly pigments most of the plant tissues, are the only alleles that pigment the auricle and ligule tissues. Although *B-I* has no allele-specific insertions in the region studied, it appears to have RFLPs that define a region required for its phenotype much farther upstream (M. Stam and V.C., unpublished results). Further studies will be necessary to show that the RFLPs are responsible for the *B-I* phenotype. Like *B-Peru* and *B-Bolivia*, the

B-Mex7b allele has an insertion upstream of the start of transcription. The 862-bp *Mako* element has homology to part of the upstream region of another maize gene and to an intron sequence in a gene of the grass Coix. This distribution is similar to that found for other small maize transposons such as *Tourist* and *Stowaway* (17, 20). Some insertions of these small transposons appear to contain regulatory regions for the adjacent genes (20).

Although *Mako* may produce the distinct expression of *B-Mex7b*, not all insertions upstream in the *b* locus alter expression. The 406-bp *Snowbird* transposon in intron 1 of *B-bar* and *b-W23* and the *Tourist* element found upstream of the alleles in clade III do not seem to affect expression, because they are common to alleles with distinct expression patterns. We do not know whether the 316-bp *Hb2* or the 359-bp insertion in the *B-M033* allele have any effect on expression, because we could not introgress this allele into our maize tester stocks.

The overall arrangement of the phylogenetic trees is consistent with results from other loci; however, the relative distances between the alleles are quite different from that found in other maize phylogenetic studies. The positioning of the two alleles from *Zm huehuetenangensis* in basal positions and the location of *Zm parviglumis* alleles interspersed with the cultivated maize alleles are consistent with other findings (21, 22). Our finding that maize and *Zm parviglumis* alleles fall into three very distinct clades composed of closely related alleles that are distantly related to alleles in other clades is quite different from that seen in the other maize genes that have been studied (*adh1*, ref. 23; *adh2*, ref. 24; *c1*, ref. 22; *glb1*, ref. 25; and rDNA, ref. 21). In fact, the distances between alleles in different clades are much larger than expected, implying either that they diverged a long time ago, potentially before the genus *Zea* diverged, or that the upstream region of *b* may have accumulated nucleotide substitutions at a much higher rate than other loci studied. We favor the latter model, because the relationships between *b* alleles that we determined match the accepted phylogeny of the *Zea mays* subspecies based on two other loci (21, 22). A recent study demonstrates that small regions within genes can evolve at extremely high rates (26). We are currently cloning and sequencing additional alleles of *b* to discriminate between these models.

The phenotypic similarity displayed by *b* alleles that have, between them, many single-nucleotide differences suggests that point mutations play a very small role, if any, in determining the phenotypic diversity of this regulatory gene. Our finding that haplotype diversity in the upstream region does not correlate with phenotypic diversity is similar to results from several other studies in maize (22, 27). In contrast, we have determined that *b* alleles with strong and distinct patterns of expression have arisen recently in evolutionary time, and that the distinct phenotypes are either caused by (*B-Peru* and *B-Bolivia*) or correlate with insertions (*B-Mex7b*) or RFLPs (*B-I*, *B-Gua31*) in the upstream region. Although other studies have found evidence for phenotypic selection being focused on nontranslated regions (28, 29), our findings actually identify polymorphisms that change the phenotype (*B-Peru* and *B-Bolivia*) and identify transposon-mediated insertions as the mechanism behind the evolution of phenotypic variants at *b*. We suspect that the evolution of novel expression patterns at other loci has occurred through similar means.

We thank Dave Hawthorne (University of Maryland, College Park, MD) for acquiring the exotic and teosinte stocks and performing the initial crosses; Susan Belcher and Ken Kubo for helping with the introgressions; Jose Lopez-Valenzuela for assisting with DNA blots and PCR; William Birky, John Doebley, Damon Lisch, Lucinda McDade, Margaret Kidwell, and members of the Kidwell lab for helpful discussions; and two anonymous reviewers for insightful criticisms. This work was supported by a grant to V.L.C. from the National Research Initiative Competitive Grants Program/U. S. Department of Agriculture, no. 96-35301-3179.

1. Dooner, H. K., Robbins, T. P. & Jorgensen, R. A. (1991) *Annu. Rev. Genet.* **25,** 173–199.
2. Grotewold, E., Chamberlin, M., Snook, M., Siame, B., Butler, L., Swenson, J., Maddock, S., St. Clair, G. & Bowen, B. (1998) *Plant Cell* **10,** 721–740.
3. Stapleton, A. E. & Walbot, V. (1994) *Plant Physiol.* **105,** 881–889.
4. Doebley, J. F. (1984) *Ann. Mo. Bot. Gard.* **71,** 1100–1113.
5. Chandler, V. L., Radicella, J. P., Robbins, T. P., Chen, J. & Turks, D. (1989) *Plant Cell* **1,** 1175–1183.
6. Patterson, G. I., Thorpe, C. & Chandler, V. L. (1993) *Genetics* **135,** 881–894.
7. Goff, S. A., Klein, T. M., Roth, B. A., Fromm, M. E., Cone, K. C., Radicella, J. P. & Chandler, V. L. (1990) *EMBO J.* **9,** 2517–2522.
8. Ludwig, S. R., Bowen, B., Beach, L. & Wessler, S. R. (1990) *Science* **247,** 449–450.
9. Coe, E. H., Jr. (1979) *Maydica* **XXIV,** 49–58.
10. Styles, E. D., Ceska, O. & Seah, K. T. (1973) *Can. J. Genet. Cytol.* **15,** 59–72.
11. Patterson, G. I., Kubo, K. M., Shroyer, T. & Chandler, V. L. (1995) *Genetics* **140,** 1389–1406.
12. Radicella, J. P., Brown, D., Tolar, L. A. & Chandler, V. L. (1992) *Genes Dev.* **6,** 2152–2164.
13. Harris, L. J., Currie, K. & Chandler, V. L. (1994) *Plant Mol. Biol.* **25,** 817–828.
14. Selinger, D., Lisch, D. & Chandler, V. L. (1998) *Genetics* **149,** 1125–1138.
15. Felsenstein, J. (1993) PHYLIP *(Phylogeny Inference Package)* ver. 3.5c. Department of Genetics, University of Washington, Seattle, WA.
16. Hollick, J. B., Patterson, G. I., Coe, E. H., Jr., Cone, K. C. & Chandler, V. L. (1995) *Genetics* **141,** 709–719.
17. Bureau, T. E. & Wessler, S. R. (1994a) *Proc. Natl. Acad. Sci. USA* **91,** 1411–1415.
18. Spell, M. L, Baran, G. & Wessler, S. R. (1988) *Mol. Gen. Genet.* **211,** 364–366.
19. Felsenstein, J. (1985) *Evolution* **39,** 783–791.
20. Bureau, T. E. & Wessler, S. R. (1994b) *Plant Cell* **6,** 907–916.
21. Buckler, E. S., IV, & Holtsford, T. P. (1996) *Mol. Biol. Evol.* **13,** 612–622.
22. Hanson, M. A., Gaut, B. S., Stec, A. O., Fuerstenberg, S. I., Goodman, M. M., Coe, E. H., Jr. & Doebley, J. F. (1996) *Genetics* **143,** 1395–1407.
23. Eyre-Walker, A., Gaut, R. L., Hilton, H., Feldman, D. L. & Gaut, B. S. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 4441–4446.
24. Goloubinoff, P., Paabo, S. & Wilson, A. C. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 1997–2001.
25. Hilton, H. & Gaut, B. S. (1998) *Genetics* **150,** 863–872.
26. Ting, C.-T., Tsaur, S.-C., Wu, M.-L. & Wu, C.-I (1998) *Science* **282,** 1501–1504.
27. Polidoros, A. N. & Scandalios, J. G. (1998) *Genetics* **149,** 405–415.
28. Schulte, P. M., Gomez-Chiarri, M. & Powers, D. A. (1997) *Genetics* **145,** 759–769.
29. Wang, R.-L., Stec, A., Hey, J., Lukens, L. & Doebley, J. (1999) *Nature (London)* **398,** 236–239.