The genomic sequences of three bronze alleles from Zea mays, Bz-McC, Bz-W22 and bz-R, are presented together with their flanking sequences. The bronze locus encodes UDPglucose flavonoid glucosyl-transferase (UFGT), an anthocyanin biosynthetic enzyme. The wild-type alleles Bz-McC and Bz-W22 condition purple phenotypes in the seed and plant, while bz-R conditions a bronze color. A full length cDNA corresponding to the Bz-McC allele was cloned and sequenced. Primer extension and RNase protection experiments were used to verify the 5' end of the bronze transcript. The Bz-McC allele has a 1416-bp coding region, a 100-bp intron and an approximately 83-bp 5' leader. Upstream of the message initiation site the sequences CTAACT and TATAA occupy the positions where the eukaryotic consensus CCAAT and TATA boxes are normally found. The alleles Bz-McC and bz-R each have different large insertions with characteristics of transposable elements in their 5' flanking regions. The bz-R allele is distinguished by a 340-bp deletion starting within the intron and including 285 bp of the second exon. The Bz-McC and Bz-W22 isoleles are known to differ in two genetically defined locations. The uts and uqv sites from the Bz-McC allele condition, respectively, lowered thermostability for the UFGT enzyme and increased amount of UFGT activity when compared with the corresponding sites in the Bz-W22 allele. The uts site maps to a region of the gene encoding two adjacent amino acid differences, either or both of which might alter the thermostability of the UFGT enzyme. The difference in UFGT levels conditioned by the uqv site is shown here to be correlated with variation in the bronze mRNA level. A likely cause of this decreased bronze mRNA level in Bz-W22 is a 6-bp duplication near the sequence CTAACCT located 74 bp upstream of the bronze message initiation site. This region is therefore tentatively identified as the uqv site.
cally for material that would cross-react with UFGT-specific antibodies. The mutation, *bz-R*, the standard reference allele for the *bronze* locus (RHOADES 1952) produces no enzymatic activity and no cross-reacting material (CRM·) (DOONER and NELSON 1977b). We have found that it is associated with a 340-bp deletion of the transcribed region (RALSTON, ENGLISH and DOONER 1987). The UFGT enzymes in *Bz-W22*, the allele from the inbred line W22, and in *Bz-McC*, a wild-type isoallele from B. McClinotck's stocks and the progenitor of the transposable element mutation *bz-m2*(Ac), differ in their thermal stability. In addition, the total amount of UFGT activity in *Bz-McC* and *Bz-W22* differs by two- to threefold. In a large scale fine structure mapping experiment, DOONER (1986) mapped genetically the position of the protein thermostability difference in these alleles to the proximal side of the insertion site of *Ds* in *bz-m2*(DI), a deletion derivative of *bz-m2* (Ac) (DOONER et al. 1986), and the site responsible for the level of UFGT activity to the centromere-proximal end of the *bz* locus.

Since UFGT is not an abundant enzyme, traditional methods for cloning the gene involving cDNA or immunological methods are not readily applicable. However, the well characterized transposable element mutations of *bronze* made it the obvious first plant gene of choice for cloning by transposable element tagging (FEDEROFF, FURTEK and NELSON 1984; DOONER et al. 1985).

To determine the exact 5' and 3' boundaries of the *bronze* transcript and its intron-exon structure we have cloned and sequenced the genomic DNA corresponding to *Bz-McC* and a full length cDNA from this allele. In addition we have cloned and sequenced two other *bronze* alleles, *Bz-W22* and *bz-R*. As found for other genes in maize (Adh: SACHS et al. 1986; Sh: WERR et al. 1985), there is a large amount of sequence diversity between alleles especially in the region 5' of the transcription initiation site. The natural diversity can be used to help define regions of DNA important for *bronze* transcript regulation and for the stability of the UFGT protein.

Recently, several other genes in the anthocyanin pathway have been cloned using transposon tagging (*a*: O'REILLY et al. 1985; *c*: PAZ-ARES et al. 1986; CONE, BURR and BURR et al. 1986) *c2*: WIENAND et al. 1986; *b22*: THERES, SCHEELE and STARLINGER 1986; McLAUGHLIN and WALBOT 1987). Since *c2*, and probably *a* and *b22*, are expressed coordinately with UFGT (DOONER 1983), a comparison of their promoter

**MATERIALS AND METHODS**

**Maize stocks:** All of the alleles used in this study were derived from or incorporated into the common genetic background of the inbred W22. The *Bz-McC* allele was extracted from a *C-I sh Bz* chromosome obtained from a maize stock of Barbara McClintock by selecting *C-I Sh Bz* crossovers from *C-I sh Bz/C sh bz-R* heterozygotes. The *Bz* allele resulting from this selection was thereafter designated *Bz-McC* (DOONER and NELSON 1977b). Other wild-type isoalleles derived from McClintock stocks may differ from *Bz-McC*. The *bz-R* allele is derived from stocks obtained from RHOADES and was designated *bz-R* by DOONER and NELSON (1977b). This is presumably the mutation that originally defined the *bronze* locus (RHOADES 1952; RALSTON, ENGLISH and DOONER 1987). The *Bz-W22* allele is the normal allele carried in the W22 inbred.

**Molecular analysis:** DNA extractions, Southern analysis, and genomic library constructions were performed as described previously (DOONER et al. 1985). DNA from a maize stock carrying the *Bz-W22* allele was digested with BamHI, size-fractionated by centrifugation through a 10% to 40% glycerol gradient, purified, ligated to EMBL3 arms and packaged. DNA from a maize stock carrying the *bronze* allele *bz-R* was partially digested with Sau3A1 treated as above and ligated to EMBL4 arms. Phage libraries were screened with pAGS528, the KpnI fragment internal to the transcribed region of *Bz-McC* (Figure 1). Restriction fragments from phage hybridizing to the *bronze* probe were subcloned into pUC19 (YANISCH-PERRON, VIEIRA and MESSING 1985).

**DNA sequencing:** Ordered deletions obtained by Bal 31 digestions (DEAN et al. 1985) for *Bz-McC*, *Bz-W22*, and *bz-R* were subcloned into *mpl8* (MESSING, CREA and SEEBURG 1981). Single strand DNA was sequenced using the dyeoxy method (SANGER, NICKLEN and COULSON 1977). Regions of high GC content, where compression of the bands on the sequencing gel made the sequence difficult to read, were resequenced using inosine nucleoside triphosphate in place of guanosine (MILLS and KRAMER 1979).

**RNA analysis:** RNA isolation and Northern analysis

---

**Figure 1.** —Restriction map of the *Bz-McC* allele showing location of the subclones used as Southern and Northern probes and of the transcribed region. The sequenced region is demarcated by asterisks.

**Figure 2.** —Comparison of three *bronze* alleles. The sequence of *Bz-McC* allele (the center line) is compared with the sequences of the *Bz-W22* (top) and *bz-R* (bottom) alleles. Labeled bases are indicated by stars. Single base differences are shown where they occur and insertions or deletions are shown as gaps. Dashes indicate regions that were not sequenced. Large insertions and deletions are also indicated on the right margin. Direct repeats (DR) and inverted repeats (IR) for the insertions in *Bz-McC* and *bz-R* are indicated above the sequence as are certain restriction sites, a mRNA initiation site, the stop codon and poly(A) addition sites. Possible regulatory sequences are indicated by boxes. Asterisks above the *Bz-W22* sequence mark DNA differences that cause changes in the corresponding protein sequence.
were performed as previously described (DOONER et al. 1985). A cDNA library was constructed in lambda gt10 from 4 μg of poly(A) RNA isolated from husks of Bz-McC; B; Pl plants by the procedure of HUVNH, YOUNG and DAVIS (1984). RNase protection experiments were performed as described by the Promega Biochemicals's publication "Riboprobe Gene Analysis System" using either 20 μg of total RNA or 1 μg of poly(A) RNA. Primer extension analysis was performed as described by McKnight et al. (1981). The oligonucleotide GGCGGGCAGGACTCG was synthesized with a Pharmacia DNA synthesizer and purified on a 20% polyacrylamide gel.

RESULTS

Genomic sequences of bronze alleles

The sequence of the bronze allele Bz-McC and its flanking regions was obtained by sequencing 60 overlapping subclones for one strand and 40 for the opposite strand. The alleles bz-R and Bz-W22 were sequenced in one direction, but any ambiguous regions were redone and any differences with the Bz-McC allele were rechecked in all three alleles. In addition, the Bz-McC allele was independently sequenced by D. FURTEK, J. SCHIEFELBEIN and O. E. NELSON (personal communication) who generously exchanged sequences with us before publication. A comparison of these two sequences showed no differences. The sequences of Bz-McC, bz-R and Bz-W22 are presented in Figure 2. The Bz-McC sequence is given as the center line, M, and the sequences for Bz-W22 and bz-R are above, W, and below, R. In regions where one of the alleles has an insertion relative to the other alleles, the sequence of the insertion is given, and a gap is present in the other alleles. The numbers at the left reflect the number of nucleotides from the initial ATG of the bronze message for each allele. The restriction maps in this paper and the nucleotide sequence presented in Figure 1 follow the molecular convention of placing the 5' end of the gene on the left. Since the direction of transcription of the bronze gene is away from the centromere, the sequence is inverted relative to our earlier restriction map of the gene (DOONER et al. 1985) which was oriented according to the genetic convention of placing the centromere of 9S to the right.

Some of the major features of the Bz locus and its flanking sequence are indicated in Figure 2, and will be discussed subsequently. The A in the initial ATG of the message is assigned a position of 1, sequences 5' to this are assigned negative numbers. A message initiation site at -83, the TGA stop codon at 1514 bp, and three poly(A) addition sites are marked. The single intron is located from 524 to 624 in the Bz-McC allele. The Bz-W22 allele has a slightly longer intron (5 bp). The 5' flanking regions of the three bronze alleles have segments of high homology. Bz-McC and bz-R each have different large insertions at different positions in their 5' flanking sequences which are marked along the (right) side of Figure 2. The region between these insertions shows a greater amount of mismatch than the rest of the sequence. Bz-McC and bz-R also have identical 62-bp insertions relative to Bz-W22 in their transcribed but untranslated 3' tails and further downstream, Bz-McC has 163 bp not found in Bz-W22. The bz-R sequence does not extend to the end of this insertion, but its sequence appears to be similar to that of Bz-McC. Experimental verification of these placements and their significance are discussed below.

The bronze transcript

cDNA analysis: The genomic region homologous to the bronze transcript and the direction of transcription have been determined by Northern analysis (DOONER et al. 1985). To determine the exact start and stop positions and the positions of any introns, further RNA and cDNA analysis was undertaken.

Husk material from plants with the anthocyanin regulatory factors B and Pl is deeply pigmented and is the best source of bronze RNA. When these studies were initiated the only B; Pl corn lines available with a characterized Bz allele carried the isoallele from the inbred line W22 (Bz-W22). A cDNA library was made from this material and several bronze hybridizing cDNAs were recovered. From these clones it was discovered that the bronze mRNA had several polyadenalation sites (DEAN et al. 1985). The sequence of these clones also revealed substantial polymorphism between the Bz-McC genomic DNA and the Bz-W22 cDNA. These differences include a 62-bp insertion in Bz-McC relative to Bz-W22 and approximately 2% base changes.

Experiments to determine the 5' end of the Bz-McC transcript using Bz-W22 RNA would be inconclusive if the same level of polymorphism existed at the 5' end of the message as is present at the 3' end. Therefore the Bz-McC allele was introduced into a B; Pl background. RNA from these plants was used to synthesize a cDNA library and to determine the 5' end of the transcript by primer extension and by RNase protection experiments.

Approximately 140,000 phage from a lambda gt10 cDNA library made with poly(A) RNA from B; Pl; Bz-McC husk material were screened with probes for the middle and the 5' end of the bronze message (pAGS528 and pAGS526, respectively, Figure 1). Approximately 100 phage (0.07%) were detected with the 3' probe and 8 of these also hybridized with the 5' probe. These 8 potentially full length cDNA clones were purified, subcloned, and characterized. The two longest ones were sequenced and the 5' ends of these were found to be within 20 bases of each other. Further analysis, described below, confirmed that one of these was a full length message.
A comparison of the Bz-McC cDNA sequence with the genomic sequence revealed a single intron of 100 bp located at +523 bp (Figure 1). The sequence around the splice sites, AGG(T)GCGT and TCGTTGCCAGC, have 7/9 and 8/11 bases (underlined) in common with the 5' and 3' consensus splice junction sequences, respectively (Breathnach and Chambon 1981). Similar agreements are obtained by comparison with the more recently derived consensus sequences from animal (Mount 1982) and plant (Brown, Feix and Frendeway 1986) introns.

**Primer extension and riboprobe protection:** A 15-base oligonucleotide was synthesized which would hybridize to the RNA sequence 17 bases downstream of the translational start site. The oligonucleotide position is shown in Figure 2. To find the start of the bronze message, this primer was hybridized to RNA from B; P1; Bz-McC husks and extended with reverse transcriptase. In addition, the oligonucleotide was used as a primer for a sequencing reaction that was used as a calibration ladder. The length of the extended fragment was 112 ± 3 bp (Figure 3). This corresponds to the 5' end of the longest cDNA clone. The family of bands in Figure 3 could be due to multiple transcriptional start sites (Lebowitz and Ghosh 1982). A faint 460-bp fragment was also present (not shown), possibly representing a second start site for the bronze message. However, when the oligonucleotide primer hybridization was carried out at different temperatures, the 112-bp and the 460-bp products had different optimal hybridizing temperatures, suggesting that the high molecular weight band was an artifact.

To determine whether the transcription start sites are the same in aleurone (where the bronze locus is under R and C regulation) and husk (under B and P1 regulation), RNA was isolated from these two different tissues. A RNase protection experiment was performed using this RNA and a riboprobe (corresponding to pAGS526 in Figure 1) complementary to the bronze message. Riboprobe can be labeled to a high specific activity and thus the low amounts of bronze message isolated from aleurone can be detected. Figure 4 shows that the protected fragment is the same for aleurone (lane A) and husk (lanes B and C). The size of the protected fragment agrees with the expected fragment size (329 bp) predicted from the primer extension experiment above. No high molecular weight band was detected in this experiment. The results from both the primer extension and the riboprobe protection confirm that the longest cDNA clone is essentially full length.

The first ATG in the message, located 83 bp 3' to the transcription start identified above, is the likely translation initiation site. The sequence around this site, AAG ATGG, shows high homology with the consensus eukaryotic translation initiation site described by Kozak (1981). Since the UFGT enzyme is present in low amounts, verification of this assignment by amino acid sequencing has not been attempted.

**Differences in the transcripts and predicted proteins encoded by the bronze alleles:** The nucleotide sequences of the two functional bronze alleles in this
study, Bz-McC and Bz-W22, differ from each other and from the mutant allele bz-R. The bz-R allele has a 340-bp deletion which begins in the intron 54 bp past the 5' splice site, removes the 3' splice site and extends 285 bp into the second exon. There is an 8-bp insertion relative to Bz-McC in the portion of the intron left in bz-R. This “insertion” of TACTACAG seems to be closer to each other than to Bz-McC.

Within the 1413 bp that constitute the translated region of the bronze transcript there are 21 bp differences between the functional alleles Bz-McC and Bz-W22. Of these, 9 produce changes in the amino acid composition of UFGT. These changes are listed in Table 1 and are indicated with stars in Figure 2. Excluding the region deleted in bz-R, there are 8 bp changes between bz-R and Bz-McC and 3 bp changes between bz-R and Bz-W22 that would result in amino acid differences. Thus Bz-W22 and bz-R seem to be closer to each other than to Bz-McC.

A proteina of 471 amino acids with the composition shown in Table 2 is predicted from the Bz-McC cDNA sequence. The molecular weight of this protein, 48.8 kD, and the pl calculated from the average pIs of the amino acids, 9.0, agree roughly with the measured values of 50 kD and 5.0 (Dooner 1981). The previously estimated acidic pl of the UFGT protein allows us to be confident of the correct reading frame of our sequence. The GC content of the three coding positions are 74% (47.6% are G), 56%, and 94%, respectively. Due to this distribution of nucleotides, any artificial frame shift would result in a protein with a drastically higher pl.

Expression of different bronze alleles

Expression of the bz-R allele. Analysis of the enzymatic and antigenic properties of endosperm proteins isolated from plants homozygous for bz-R has demonstrated that no UFGT activity or cross-reacting material to UFGT-specific antibodies is detectable, suggesting that no protein is made (Dooner and Nelson 1977b; Dooner 1986). The sequence of bz-R reveals a 340-bp deletion starting 46 bp inside the intron and extending 285 bp beyond the 3' splice site of the intron (Figure 2). For these reasons, the sequence of Bz-McC in the 5' upstream regions raise the possibility that the promoter region of bz-R might also be altered in function. To determine bz-R transcript levels, crosses were made to derive bz-R/bz-R homozygous and bz-R/Bz-McC heterozygous plants with the regulatory factors B and Pl. RNA from these plants was

---

**TABLE 1**

<table>
<thead>
<tr>
<th>Position</th>
<th>Allele</th>
<th>Codon</th>
<th>Amino acid</th>
</tr>
</thead>
<tbody>
<tr>
<td>265</td>
<td>M</td>
<td>ACC</td>
<td>THR</td>
</tr>
<tr>
<td></td>
<td>W</td>
<td>TCC</td>
<td>SER</td>
</tr>
<tr>
<td>642</td>
<td>M</td>
<td>GGG</td>
<td>GLY</td>
</tr>
<tr>
<td></td>
<td>W</td>
<td>GAG</td>
<td>GLU</td>
</tr>
<tr>
<td>645</td>
<td>M</td>
<td>CTA</td>
<td>LEU</td>
</tr>
<tr>
<td></td>
<td>W</td>
<td>CCA</td>
<td>PRO</td>
</tr>
<tr>
<td>1045</td>
<td>M</td>
<td>GAC</td>
<td>ASP</td>
</tr>
<tr>
<td></td>
<td>W</td>
<td>GCC</td>
<td>ALA</td>
</tr>
<tr>
<td>1089</td>
<td>M</td>
<td>CCG</td>
<td>PRO</td>
</tr>
<tr>
<td></td>
<td>W</td>
<td>ACG</td>
<td>THR</td>
</tr>
<tr>
<td>1094</td>
<td>M</td>
<td>CAC</td>
<td>HIS</td>
</tr>
<tr>
<td></td>
<td>W</td>
<td>CTC</td>
<td>LEU</td>
</tr>
<tr>
<td>1236</td>
<td>M</td>
<td>TTG</td>
<td>LEU</td>
</tr>
<tr>
<td></td>
<td>W</td>
<td>GTG</td>
<td>VAL</td>
</tr>
<tr>
<td>1364</td>
<td>M</td>
<td>ACG</td>
<td>THR</td>
</tr>
<tr>
<td></td>
<td>W</td>
<td>GCG</td>
<td>ALA</td>
</tr>
<tr>
<td>1426</td>
<td>M</td>
<td>GAG</td>
<td>GLU</td>
</tr>
<tr>
<td></td>
<td>W</td>
<td>GTG</td>
<td>VAL</td>
</tr>
</tbody>
</table>

The nucleotide position identifies the base in the Bz-McC sequence that corresponds to the first position of the codons specifying amino acid differences between the two “wild-type” alleles.

---

**TABLE 2**

<table>
<thead>
<tr>
<th>Amino acid</th>
<th>No.</th>
<th>Amino acid</th>
<th>No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ala</td>
<td>90</td>
<td>Met</td>
<td>8</td>
</tr>
<tr>
<td>Cys</td>
<td>9</td>
<td>Asn</td>
<td>8</td>
</tr>
<tr>
<td>Asp</td>
<td>23</td>
<td>Pro</td>
<td>38</td>
</tr>
<tr>
<td>Glu</td>
<td>24</td>
<td>Gln</td>
<td>9</td>
</tr>
<tr>
<td>Phe</td>
<td>20</td>
<td>Arg</td>
<td>30</td>
</tr>
<tr>
<td>Gly</td>
<td>44</td>
<td>Ser</td>
<td>30</td>
</tr>
<tr>
<td>His</td>
<td>12</td>
<td>Thr</td>
<td>14</td>
</tr>
<tr>
<td>Ile</td>
<td>6</td>
<td>Val</td>
<td>43</td>
</tr>
<tr>
<td>Lys</td>
<td>4</td>
<td>Trp</td>
<td>10</td>
</tr>
<tr>
<td>Leu</td>
<td>45</td>
<td>Try</td>
<td>4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Base</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>14.4</td>
<td>25.9</td>
<td>1.9</td>
</tr>
<tr>
<td>C</td>
<td>26.8</td>
<td>35.0</td>
<td>50.1</td>
</tr>
<tr>
<td>A</td>
<td>11.3</td>
<td>17.8</td>
<td>3.2</td>
</tr>
<tr>
<td>G</td>
<td>47.6</td>
<td>21.2</td>
<td>44.8</td>
</tr>
</tbody>
</table>

G + C  74.4  56.2  94.9

Forty-seven percent of the first bases in the codons for UFGT are G. Thus 47% of the amino acids in UFGT are restricted to the 5 amino acids specified by this codon class and are indicated by asterisks.
separable from the transcript of the functional alleles by gel electrophoresis. Thus the level of bz-R RNA can be used as an internal standard for the levels of Bz-McC and Bz-W22 RNA as measured by Northern analysis.

Northern blots of RNA from Bz-McC/bz-R and Bz-W22/bz-R seedlings were probed with the insert from pAGS551 (Figure 1) cloned behind the SP6 promoter. The similarity in the intensity of the bz-R and Bz-McC bands in the Bz-McC/bz-R heterozygote (Figure 5, lane c), suggests that these two alleles have functionally similar promoters. On the other hand, the pronounced differences in intensity between the bz-R and Bz-W22 band in the respective heterozygote (Figure 5, lane b) suggest that the Bz-W22 allele has a weaker promoter than bz-R. The difference in intensity between the Bz-McC and Bz-W22 bands relative to the bz-R band is at least as large as the three fold difference in the enzymatic activity levels found in aleurone tissue (Dooner 1986). These observations suggest that the promoters for Bz-McC and Bz-W22 are functionally different and that the difference between the promoters most likely correspond to the previously mapped uqv site.

Insertions in the sequences flanking the bronze locus: Comparison of the 5' region of the 3 alleles showed that Bz-McC and bz-R each have independent insertions with the characteristics of transposable elements. The Bz-W22 allele has no insertion. The insertion in Bz-McC at ~834 to ~563 has a 4-base direct repeat at its ends, CTAA, which is found in only one copy in bz-R and Bz-W22. Possibly these bp were duplicated when the insertion occurred, a feature of transposable element insertions. As shown in Figure 2 and Figure 6, this element also has terminal inverted repeats which would allow a “stem and loop” structure postulated to play a role in excision (Saedler and Nevers 1985). We have called this insertion Ins1 (see Discussion).

Sequences structurally related to Ins1 are found elsewhere in the genome. When the Bz-McC lambda gt10 cDNA library was screened for full length bronze messages with the pAGS526 probe (Figure 1) which contains Ins1 in addition to the 5' end of the bronze message, two clones were isolated which did not hybridize to the bronze gene. The sequence of these clones revealed homology to Ins1 (Figure 6). One of these clones (Figure 6, #1) has the entire element flanked by a new 4-bp direct repeat and ends with a poly(A) tail beyond the right end of the element. The other (Figure 6, #2) terminates inside of the insertion sequence with a poly(A) tail close to the left end of the element (shown as a poly(T) in Figure 6). These three sequences are highly homologous and may have been derived from a common ancestor. Some of the differences in Figure 6 are in the inverted repeats which may have been more extensive in the

---

**FIGURE 5.**—Expression of Bz-McC, Bz-W22 and bz-R analyzed by Northern blots. Lane a contains 1 μg of poly(A) RNA from a Bz-McC homozygous plant. Lanes b, c, and d each contain 20 μg of total RNA extracted from Bz-W22/bz-R (lane b), Bz-McC/bz-R (lane c) heterozygous plants, or a bz-R/bz-R (lane d) homozygous plant. The Bz and bz-R transcripts were separated by gel electrophoresis. The Northern blots were hybridized to the bronze specific probe pAGS528. The arrows indicate RNA m. w. markers of 3.0, 2.25, and 0.9 kb.

compared by Northern analysis (Figure 5). The bz-R message is smaller than the Bz-McC message and the two messages can be separated in RNA from the heterozygous plant. The similarity in intensities of the bz-R and the Bz-McC bands in RNA from a heterozygous plant RNA suggests that similar amounts of RNA are being made from the promoters of these two alleles. In addition, riboprobe protection experiments with RNA from bz-R/Bz-McC heterozygotes (Figure 4, lane B) shows only one band, suggesting that the mRNA initiation sites are identical. Thus, the promoter region of bz-R is functional, a smaller message is made, yet no cross-reacting material to the UFGT enzyme can be detected. Apparently, any protein made from this message is unstable and quickly degraded.

Expression of the Bz-W22 allele: Dooner (1986) has shown that Bz-McC and Bz-W22 stocks differ in their amounts of enzymatic activity and has mapped genetically the site responsible for this difference. Northern analysis was used to determine if the difference in enzyme levels reflects a difference in message levels. It would be difficult to quantify accurately the difference in expression between two plants homozygous for Bz-McC and Bz-W22 because the bz allele is only expressed in a small fraction of the cell types and its expression is influenced by light levels which differ from leaf to leaf even on the same plant. Instead each of these alleles was made heterozygous with bz-R which makes a smaller transcript

---

E. J. Ralston, J. J. English and H. K. Dooner
ancestral form. For instance, the 12-bp "insert" in cDNA #1 makes a fourth inverted repeat (IR4), with all three sequences. Genomic Southern analysis revealed approximately 50 bands or more that hybridize to *Ins1*. Thus, the three elements that we have cloned and sequenced are members of a large family. At least two members of this family are inserted in regions of the genome that are transcribed.

The insertional element in the 5' flanking region of *bz-R* is located 299 bp upstream of the first ATG in the *bronze* message (Figure 2). The 10-bp direct repeats at its ends (shown underlined in Figure 2) are found in one copy in the other two alleles, suggesting again that the duplication occurred as a consequence of the insertion. The inverted terminal repeats characteristic of transposable elements are also indicated in Figure 2. This element, designated *Ins2* (see discussion), is found to hybridize with many bands when used as a probe for genomic Southern analysis. We have not isolated any other members of this family to see what level of divergence is present.

Both *bz-R* and *Bz-McC* are found to have a 62-bp insertion in their 3' transcribed, but untranslated, ends when their sequences are compared with the sequence of *Bz-W22*. While this sequence is flanked by a four base pair direct repeat ACAT it does not have the more extensive level of internal inverted and direct repeats usually found in transposable elements, and also found in the 5' insertions. Further downstream, at 2097 bp in *Bz-McC*, a 163-bp insert in *Bz-McC* or a deletion in *Bz-W22* is found, again with no internal structures similar to those of a transposable element.

**DISCUSSION**

The *bronze* locus has been studied extensively both genetically and biochemically. Three *bronze* alleles have been sequenced in this investigation. In the following discussion we will attempt to integrate the sequence data with the existing genetic and biochemical data.

**The transcribed region:** The *bronze* transcribed region is simple. The *Bz-McC* allele has only one intron of 100 bp located 525 bases past the translation start codon. The intron in *Bz-W22* and the portion of the intron left in *bz-R* are identical to the *Bz-McC* intron except for additional 5 and 8 bp, respectively. The intron length variation may be due to DNA replication errors or to other deleterious events in *Bz-McC* and *Bz-W22*, since the region deleted is flanked by the direct repeat, TACTAC. We previously reported that multiple polyadenylation sites are used in the *bronze* message (Dean et al. 1986). Multiple poly(A) addition sites have also been found in the *Adh* (Sachs et al. 1986) and *Wx* (Klosgen et al. 1986) loci in maize and in the petunia CAB and small subunit of RUBISCO gene families (Dean et al. 1986). Perhaps the most unusual structural feature of the *bronze* sequence is its high GC content. Excluding the intron which has 50% GC, the coding sequence has an overall GC content of 75% and has long stretches of almost 100% GC.

The allele *bz-R*, the first or reference mutation described in the *bronze* locus (Rhoades 1952), has been shown to be CRM\(^-\), i.e., to accumulate no cross-reacting material to UFGT antibodies (Dooner and Nelson 1977a; Dooner 1986). The sequence of the *bz-R* allele shows a 340-bp deletion starting in the intron and extending 285 bp past the normal splice site for *Bz-McC*. Potential secondary 3' splice sites exist 9 and 117 bp beyond the deletion end point. If the *bz-R* allele was otherwise undamaged, a shorter spliced mRNA might be made. However, the deletion in *bz-R* may also include the internal intron splicing signal described for animal systems by Keller and Noon (1984). An analysis of 36 introns from maize *Adh* (Sachs et al. 1986), *Sh* (Werr et al. 1985) and *Wx* (Klosgen et al. 1986) shows a consensus sequence of C(21), T(34), G(27), A(36), C or T(29) to 45 bases upstream of the 3' splice junction. The corresponding sequences in the *bronze* intron, ATGAT or TTGAT, are deleted in *bz-R*. An alternate internal splice signal,
CCGAC, is found 10 bp past the deletion end point, however no 3' splice sites, i.e., AG, occur within 45 bp downstream. Further analysis of the bz-R message is needed to determine if an alternate splice site is used.

A comparison of the coding regions common to the bz-R, Bz-McC and Bz-W22 alleles shows a high degree of divergence, similar to that found in comparisons of other maize alleles studied so far (Sh, Adh). Considering only nucleotide differences that do not result in amino acid changes, Bz-McC and Bz-W22 differ at 12 positions, Bz-McC and bz-R differ at 9 positions, and Bz-W22 and bz-R differ at 8 positions. The percent difference in silent bases as calculated by the method of Perler et al. (1980), and the rate of divergence for silent changes in plants estimated by Wilson, Ochman and Prager (1987) predict that three bronze alleles separated from each other 2 to 3 x 10^6 yr ago. Werr et al. (1985) have found a similar degree of divergence in two alleles of sh.

Unlike sh and adh, the bronze alleles sequenced here also have many nucleotide changes that would affect the amino acid composition of the UFGT enzyme. Considering the nucleotide differences that would result in amino acid changes, Bz-McC differs from Bz-W22 and bz-R at 9 and 8 positions, respectively, while Bz-W22 and bz-R differ at only 3 positions. The fact that almost half of the nucleotide differences between Bz-McC and Bz-W22 result in amino acid changes suggests that the UFGT enzyme can tolerate more change than sucrose synthase or ADH. The observation that the divergence in silent substitutions is approximately the same for all three alleles, yet the divergence in amino acid replacement substitutions is not, suggests that some of the amino acid changes are the result of selection pressure not experienced by all three alleles.

The UFGT enzymes found in Bz-McC and Bz-W22 have been studied biochemically, and found to differ in their thermal stability. The us (UFGT thermal stability) site responsible for this difference has been mapped genetically to be 5' of the Ds insertion site in bz-m2(DI) (Dooner 1986). By cloning and sequencing the appropriate alleles, the position of this Ds insertion has been identified and found to be identical to the Ac insertion in bz-m2 which is located at +749 bp in Figure 2 (Dooner et al. 1986; Dooner, Ralston and English 1988). The us site has also been placed by Southern analysis of selected Bz intragenic crossovers to be 3' to the AatII site which differs in Bz-McC and Bz-W22 at +514 bp in Bz-McC (H. K. Dooner, unpublished data). Between +514 and +749 there are two amino acid differences in these alleles. These changes, a leucine to a proline and a glycine to a glutamine in Bz-McC and Bz-W22 respectively, are located 3 bases apart at +642 and +645 and thus would behave essentially as one genetic site. Since either or both of these changes could cause a change in the protein structure and hence in thermal stability, we conclude that this location probably represents the us site. The insertion site of Ds in bz-m2(DI) is also known to influence the thermal stability of UFGT since reversions that arise from excision of Ds result in proteins with large variation in their thermal stability (Dooner and Nelson 1979b).

The insertions outside the coding region: Outside of the coding regions, at both the 3' and the 5' ends, the bronze alleles differ from each other by the presence of large insertions or deletions. The insertions upstream of the 5' ends of the bz-R and Bz-McC alleles are flanked by 10- and 4-bp direct repeats, respectively, and have terminal inverted repeats characteristic of transposable elements. They are also members of separate large families as judged by genomic Southern hybridization. Thus, it is likely that these elements are derived from transposable element insertions. However, since these elements have not been observed to move, we cannot determine whether they are still mobile or remnants of an old transposable element system.

We have compared the sequences of the inverted terminal repeats in the insertions upstream of Bz-McC and bz-R with those of other maize insertions (Doring and Starling 1984; Shepherd et al. 1984; Zack, Ferl and Hannah 1986). The insertion upstream of the Bz-McC allele has a 6-bp inverted repeat near its termini and is flanked by a 4-bp target site direct repeat. The inverted repeat shares 5 of 6 bases with the inverted terminal repeat of cin1, an element that also causes a 4-bp target site duplication, but otherwise, the two elements have little internal homology. We have designated the family of elements homologous to the insertion upstream of Bz-McC as Ins1 (Insertion 1). The terminal inverted repeat in the insertion upstream of bz-R is not homologous with the terminal repeat of any sequenced maize insertion. We have designated this family of insertions as Ins2 (Insertion 2).

The 3' transcribed but untranslated ends of bz-R and Bz-McC have a 62-bp insertion not present in the Bz-W22 sequence. This insertion does not have extensive characteristics of a transposable element. Further downstream, Bz-McC and Bz-W22 differ by a 163-bp insertion in Bz-McC which also seems to be present in the bz-R allele (our bz-R sequence terminates within this insertion). This insertion does not have terminal repeats or extensive internal repeats.

It is puzzling that these 3' insertions are present in Bz-McC and bz-R, but not in Bz-W22, since when the coding regions are compared, bz-R is closer to Bz-W22. While bz-R and Bz-McC differ by 1-2% through most of the coding region, the sequences
flanking the 62-bp insertion at the 3' end, expected to be more divergent since they are outside of the coding region, are 100% homologous for over 300 bases. Possibly, a crossover event between the 3' ends of bz-R and Bz-McC has occurred recently, leading to homogenization of the 3' flanking sequences. In fact, a crossover event between sh and bz (i.e., on the 3' side of bz) was selected initially when a Sh Bz-McC chromosome was extracted from a sh Bz-McC/Sh bz-R heterozygote (see MATERIALS AND METHODS). Other unselected crossover events could have occurred subsequently during the incorporation of Bz-McC into a W22 genetic background by repeated backcrossing to a W22 bz-R stock.

The bronze 5' end: The 5' regions of the three alleles studied here can be divided into 3 parts separated by Ins1 and Ins2 (Figure 7). Region I, from −1 to around −300 is highly conserved. The first 193 bp 5' to the first ATG are 100% conserved in bz-R and Bz-McC. This homology is broken by the deletion of 9 bp (GCGTGCGCG) in bz-R relative to Bz-McC. Further upstream lies a 100-bp region of 95% homology delimited by the insertion in bz-R. No consensus “TATA” box is found within 80 bases of the transcription start site. Instead, the sequence AATAAAAA is found 32 bp before the start of transcription, a normal location for the “TATA” box (Messing et al. 1983). The sequences CCAACC and CTAACT are found 68 and 34 bp upstream of this AATAAAA, respectively. Both of these are within the normal range for the consensus CCAAT box (Nussinov, Owens and Maizel 1986). While the Bz-W22 promoter region has a few single base changes in this region compared to Bz-McC and bz-R, the only major difference is the duplication of the sequence GTCTAA two bases upstream of the CTAACT sequence, creating a near duplication of this potential sequence, creating a near duplication of this potential

![Figure 7](https://example.com/f7.png)

**Figure 7.**—Composite structure of the Bz-McC, Bz-W22, and bz-R alleles. Insertions relative to the other sequences are shown as triangles: A, insertion in Bz-McC only (Ins1); B, insertion in bz-R only (Ins2); C and D, insertions in both Bz-McC and bz-R and not in Bz-W22. The protein coding region is shown as a dark bar, the direction and length of the RNA transcript is indicated by the arrow, and the deletion in the bz-R allele is marked. The KpnI site within the coding region is found only in the Bz-McC allele. The sequenced region presented in Figure 2 is indicated by **stars**. Except for the region between inserts A and B the sequences of the three alleles are 98 to 99% homologous.

functionally equivalent for two reasons. Their mRNA initiation site is the same (i.e., only one riboprobe protected band is observed with RNA from bz-R/Bz-McC heterozygous plants, Figure 4, lane B), and they produce similar amounts of message as judged by the intensity of the bz-R and Bz-McC bands in Northern blots of bz-R/Bz-McC heterozygous plants (Figure 5). The Bz-W22 allele is shown above to have a functionally less efficient promoter. The difference in transcript levels between Bz-McC and Bz-W22 correlates with the difference in UFGT activities of these alleles. The genetic site responsible for this enzyme activity difference, uqv, has been mapped genetically to be 5' of uts (Dooner 1986) which is located between +642 and +645 bp in Bz-McC. From genetic data it is also known that the uqv site lies close or 5' to the Ds insertion site in bz-ml. This site has been localized by Southern analysis to the region from +120 to +234 flanked by the KpnI and BglII sites (Dooner et al. 1985).

In an attempt to correlate the uqv site with a specific sequence, we can scan the 5' upstream regions of the different bronze alleles for any significant differences starting at +642 bp. Of the 6-bp differences within the coding region between Bz-W22 and Bz-McC upstream of uts, only one difference results in an amino acid change, a minor change of a threonine to a serine. Most of these nucleotide differences are also found in bz-R which seems to make the same amount of RNA as does Bz-McC. Therefore the most likely location for the uqv site is 5' of the message. The sequence CTAACT, at −157, has an upstream 6 bp duplication in Bz-W22, as discussed above. If this sequence is involved in protein binding, it is possible that the related sequence CTAAGT, generated in the duplication, would serve as a competing binding site. Alternatively, the approximately one-half turn of the DNA helix resulting from a 6 base insertion could disrupt any protein binding sites that normally span the insertion site. A comparison of the sequences of the three bz alleles in regions II and III shows that
many of the differences between Bz-Mc and Bz-W22 are also found in bz-R and are thus unlikely to be responsible for the low Bz-W22 message levels. Therefore it is likely that the uvg site corresponds to the presence versus absence of the 6-bp insertion at -157. Similar structure-function relationships have been observed in the promoter region of the Pal gene of Antirrhinum majus. Changes directly upstream of the sequence CCAACCT due to excision of the transposable element Tam3 apparently affect the level of expression of that gene (Coen, Carpenter and Martin 1986).

To date every allele of B: that has been sequenced, has had insertions in different positions outside of the coding region, (this paper and D. Furtek, J. Schieffelin and O. Nelson, personal communication). These insertions or deletions can be identified only when the sequences of several alleles are compared. A large insertion 5' to the message of a wild-type allele helps define the boundary of the region involved in direct transcriptional control and, of course, allows one to discount the sequence of the insertion itself as playing a role in regulation. If the natural insertional diversity surrounding genes is a common feature in maize or plants in general, the only when the sequences of several alleles are compared. A large insertion

We would like to thank D. Furtek, J. Schieffelin and O. Nelson for preprocessing exchanges of bronze sequences, R. Jorgensen, W. Tucker and M. Ralston for reviewing the manuscript, and C. Stephens for the artwork. This is Paper No. 5-7 from Advanced Genetic Sciences.

LITERATURE CITED


Dooner, H. K., E. Ralston, and J. English 1988 Deletions and insertions in different positions outside of the coding region, (this paper and D. Furtek, J. Schieffelin and O. Nelson, personal communication). These insertions or deletions can be identified only when the sequences of several alleles are compared. A large insertion 5' to the message of a wild-type allele helps define the boundary of the region involved in direct transcriptional control and, of course, allows one to discount the sequence of the insertion itself as playing a role in regulation. If the natural insertional diversity surrounding genes is a common feature in maize or plants in general, the utility of sequencing several different alleles to define the conserved region surrounding the message is apparent.

We would like to thank D. Furtek, J. Schieffelin and O. Nelson for preprocessing exchanges of bronze sequences, R. Jorgensen, W. Tucker and M. Ralston for reviewing the manuscript, and C. Stephens for the artwork. This is Paper No. 5-7 from Advanced Genetic Sciences.

LITERATURE CITED


Dooner, H. K., E. Ralston, and J. English 1988 Deletions and insertions in different positions outside of the coding region, (this paper and D. Furtek, J. Schieffelin and O. Nelson, personal communication). These insertions or deletions can be identified only when the sequences of several alleles are compared. A large insertion 5' to the message of a wild-type allele helps define the boundary of the region involved in direct transcriptional control and, of course, allows one to discount the sequence of the insertion itself as playing a role in regulation. If the natural insertional diversity surrounding genes is a common feature in maize or plants in general, the utility of sequencing several different alleles to define the conserved region surrounding the message is apparent.

We would like to thank D. Furtek, J. Schieffelin and O. Nelson for preprocessing exchanges of bronze sequences, R. Jorgensen, W. Tucker and M. Ralston for reviewing the manuscript, and C. Stephens for the artwork. This is Paper No. 5-7 from Advanced Genetic Sciences.
Sequence of Maize bronze Alleles


Sachs, M. M., E. S. Dennis, W. L. Gerlach and W. J. Peacock, 1986 Two alleles of maize alcohol dehydrogenase 1 have 3' structural and poly(a) addition polymorphisms. Genetics 113: 449–467.


Communicating editor: W. F. Sheridan