A new allele of the duplicated 27kD zein locus of maize generated by homologous recombination

O.Prem Das, Eduard Poliak, Kathleen Ward and Joachim Messing*
Waksman Institute, Rutgers, the State University of New Jersey, Piscataway, NJ 08855-0759, USA

Received March 8, 1991; Revised and Accepted May 6, 1991 EMBL accession no. X58197

ABSTRACT

The 27kD zein storage protein locus in many inbred lines of maize consists of a tandem duplication of 12kb, with an expressed gene in each repeat, termed A and B. A single-copy allele with only the A gene can be generated from this duplication in particular stocks of the maize inbred line A188 by a mitotic process that includes a crossover at the 3′ regions of the two genes (1). Here, we characterize a second single-copy allele with only the B gene, found in different stocks of A188. This allele arises from a homologous recombination at the highly conserved 5′ regions of the two repeats, and cloning and sequencing of this allele define the crossover region. The A and B genes in the duplicated allele were previously shown to be expressed at different levels; this difference remains unchanged in either recombinant allele. Therefore, the crossover points of these two recombinant alleles define the borders of cis-acting sequences responsible for the differential expression of the two genes.

INTRODUCTION

The genes for the zein storage proteins of maize are among the best characterized plant genes and extensive sequence and mapping data have been accumulated for these genes (2). Since they are the major protein components of the maize endosperm and are thus of economic importance, considerable information has also been obtained at the protein level. Significant variation in the composition profiles of zeins has been observed among commonly used inbred lines (3). Since most of these inbred lines are of related lineage and represent a relatively small genetic pool (4), this variation suggests mechanisms effecting quantitative and qualitative changes in the expression of these genes.

Zeins can be divided into two classes based on their genetic organization. The more abundant zeins, z1, are encoded by a large multigene family, which may include up to a hundred genes in total (2). As for most multigene families, unequal crossovers and gene conversion (5) have been suggested to be the major influences on the evolution of these genes (6). The less abundant zeins, z2, are encoded by only one or two genes for each molecular weight class (7, 8, 9). One of these, the 27kD zein, has been extensively studied in our laboratory.

Some inbred lines of maize carry the *S* or standard allele of the 27kD zein locus, which has two genes within tandem repeats of 12kb (9). Figure 1 shows a restriction map of the duplicated *S* allele in A188, with the 12kb repeats, designated A and B, drawn on separate lines to illustrate their homology (9, 13). Each repeat contains an expressed copy of the gene, shown by an open box, and the B repeat contains a unique 1.8kb insertion 3′ of the coding region. The *S* alleles in the maize inbred lines W22 and W23 are very similar to the one in A188 (9, 10). Sequence comparisons of the *S* alleles of A188 and W22 indicate copy correction mechanisms that probably operate by gene conversion, which have preserved sequences during the evolution of this allele (10). It is interesting that these copy correction processes introduce different polarities of sequence conservation that originate from the same crossover region as described below for the *Ra* recombinant allele (10).

There are many stocks of A188 which are homozygous for the duplicated *S* allele (Das and Messing, unpublished data). However, particular stocks were found to contain a single-copy allele generated by a two step process consisting of a homologous crossover 3′ of the coding sequences and a second rearrangement, which may be a deletion, insertion or inversion, at the 5′ flanking sequence (1). Since this leads to the deletion of the B gene, but leaves the A gene intact, we called this recombinant allele *Ra*. Surprisingly, this process can occur during the development of the plant rather than during meiosis (1). Since plants do not have a germline, the product of such a mitotic event can be inherited. Since the same rearranged allele has been found in other maize inbreds, mitotic recombination may provide an alternate mechanism for genetic diversity.

Although the A and B coding sequences are very conserved between W22 and A188 (10), two differences, being only 9bp apart, allowed for the construction of mRNA specific oligonucleotide probes. Using these probes we were able to show that B mRNA accumulated to 2.5-fold higher levels than A mRNA for the *S* allele (9). Accumulation of mRNA from the A gene of the *Ra* allele appeared to be the same as from the A gene of the duplicated allele (9). Since endosperm tissue is triploid, with one copy of the paternal genes and two of the maternal genes, we could demonstrate gene dosage for B mRNA in reciprocal crosses of *Ra* and *S* alleles. In this report, we describe the expression and structure of a second allele of this

* To whom correspondence should be addressed

locus (termed *Rb* for recombinant with the B gene) that has been found in different A188 stocks.

MATERIALS AND METHODS

Maize stocks

A188 stocks used for the experiments described were all obtained from Dr.Ronald Phillips of the University of Minnesota, St Paul, MN. Some of these were homozygous for the *S* allele and others were segregating for *S* and *Rb*, but none carried *Ra*. Two other stocks, obtained from Dr. Burle Gengenbach, also of the University of Minnesota and from Molecular Genetics Inc., Minnetonka, MN, carried only the *Ra* and *S* alleles, and not the *Rb* allele (1).

Isolation and analysis of DNA and RNA

Self-pollinated A188 plants were harvested at days 8, 10, 12, 14, 16, 18, 20, 22, 24 and 28 after pollination, with one ear harvested for each time point. Immature endosperms were manually separated from embryos and frozen away at −70°C. DNA and RNA were isolated from frozen endosperms as described (11). For Fig. 2, 10 μg per well of RNA from pooled endosperms from individual ears was loaded on slot blots, and for Fig. 3, a tenth of the RNA isolated from individual endosperms was loaded as described (11). These blots were probed with the gene-specific oligonucleotide probes for the A and B genes (9).

DNA was isolated from the same endosperm samples as the RNA and from leaves as described (11). Genomic Southern blots of these samples were prepared after digestion with the appropriate restriction enzymes, and hybridized to a labelled mixture of 3.7kb *Sal* I fragments of the A and B genes isolated from a cosmid clone of the *S* allele from the inbred line W22 (See Fig 1). All DNA hybridizations were performed in 50% formamide at 42°C, and stringent washes were for 60 min. at 65°C in 0.1×SSC.

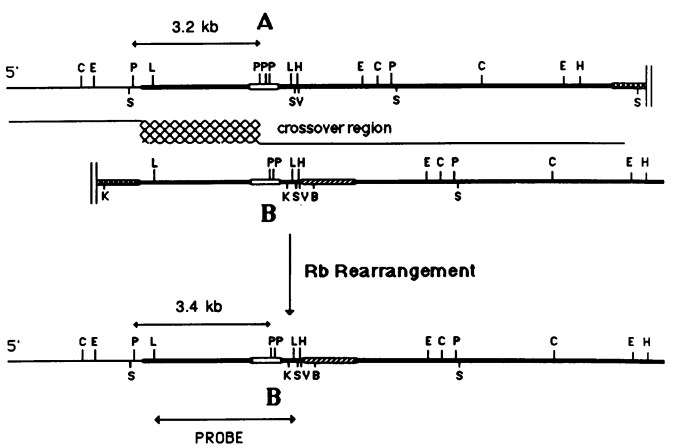

Figure 1. Restriction maps of the *S* and *Rb* alleles: The upper two restriction maps represent the *S* allele, with the A and B copies of the duplications drawn on separate lines to highlight homology. The map is split with parallel vertical lines in the inter-repeat region that is indicated by a patterned box. The lower map represents the *Rb* allele. The coding sequences are shown by open boxes and the 1.8kb insertion unique to the B repeat is represented by a striped box. Transcription direction is from left to right. The crossover leading to *Rb* is indicated by the line between the repeats of the upper panel, and the crossover region deduced from restriction analysis is crosshatched. Restriction site key: B = *Bgl* II, C = *Sca* I, E = *Eco* RI, H = *Hind* III, K = *Cla* I, L = *Sal* I, P = *Pst* I, S = *Sac* I.

Cloning and sequencing of *Rb*

Genomic DNA from pooled endosperms of two selfed A188 plants segregating for the *Rb* and *S* alleles was isolated as above and further purified on a CsCl gradient. The DNA was digested with *Bam* HI and fractionated on a 10−40% (w/v) linear sucrose gradient to isolate a fraction enriched for the *Rb* band of 12kb. DNA from this fraction (500 ng) was ligated to 1 μg *Bam* HI-digested λ EMBL3 arms, and packaged *in vitro*. This library was plated on the host strain K802, screened with the same probe as above, and 13 positive plaques were purified. DNA from several positive clones was digested with *Bam* HI, and subcloned into pUC119. The 3.5kb *Pst* I fragment from one of these (see Fig 1) was further subcloned, and oriented deletions were made from either end with the exonucleaseIII-mung bean nuclease method (12). Double stranded DNA from these clones was sequenced using the dideoxy chain termination method. Sequence data (X58197) were compiled and comparisons were made using the sequence analysis software package from the University of Wisconsin Genetics Computer Group.

RESULTS

Detection and expression of the *Rb* Allele

The *Rb* allele was discovered through experiments designed to compare the expression of the A and B genes through endosperm development in A188. RNA was isolated from immature endosperms of selfed A188 plants on various days after pollination (DAP); one sibling of these plants had been shown to be

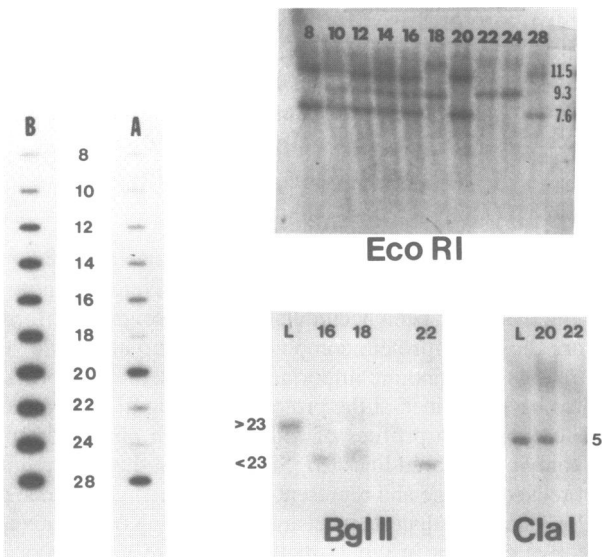

Figure 2. RNA and DNA analysis of endosperms from selfed A188 plants: RNA and DNA were isolated from approximately 2 g of frozen immature endosperms from ten self-pollinated A188 plants harvested 8, 10, 12, 14, 16, 18, 20, 22, 24 and 28 days after pollination. For all panels, these days after pollination (DAP) numbers are used to indicate particular samples. The two panels labelled A and B on the left are autoradiograms of slot blots of total RNA probed sequentially with oligonucleotides specific for the A and B genes respectively. The panel labelled *Eco* RI is an autoradigram of genomic Southern blots of DNA from the same samples. The panels labelled *Bgl* II and *Cla* I are similar autoradiograms, using the indicated enzyme and sample. Lanes marked L are from leaf DNA of an *S/S* A188 plant. All Southern blots were hybridized to a mixture of 3.7kb *Sal* I fragments derived from the A and B genes of W22. Molecular sizes in kb are indicated beside each panel.

homozygous for the *S* allele. When these were probed with the B gene-specific oligonucleotide probe (Fig. 2, panel B), mRNA accumulation through development appeared to be the same as in W23 (9). Note that in Fig. 2, each DAP corresponds to a different sibling plant. However, when the same blot was hybridized to the A gene-specific oligonucleotide probe (Fig. 2, panel A), some samples appeared to have either no detectable level (DAPs 18, 24) or a reduced level (DAPs 14 and 16) of A gene transcripts, whereas others had normal levels (DAP 20). This was different from W23, which had a constant ratio of A to B transcripts at all times after pollination (9). The reduction in A transcript levels in A188 can be due either to a lack of expression of the A gene of the *S* allele in endosperms of particular plants or to the presence of a new segregating allele which did not contain the A gene. Differential methylation, which affects the activity of the maize transposon *Ac* (14), provides a mechanism for the first possibility, and formation of the *Ra* allele in other A188 stocks (1) is a precedent for the second.

To distinguish between these, endosperm DNA from these samples was compared to DNA from W23 endosperms by Southern blot hybridization using various restriction enzymes (Fig. 2). With *Eco* RI, two bands of 11.5 and 7.5kb represent the duplicated allele. Comparison of the *Eco* RI pattern in Fig. 2 with the RNA data shows clear correspondence between the loss of A gene transcripts and the appearance of a new band of 9.3kb. In a plant with normal levels of A transcripts (DAP 20), this band is absent whereas in those with no A transcripts (DAPs 22, 24) only the new band and neither of the two progenitor bands are present. In cases of intermediate A transcript levels (DAPs 12, 14, 16), this band is present in intermediate levels relative to the 11.5 and 7.5kb bands of the *S* allele (the faint bands of higher MW in some lanes stem from incomplete digestion of genomic DNA). Similar results were obtained with all tested enzymes, which included *Apa* I, *Bam* HI, *Bgl* II, *Cla* I, *Dra* I, *Hind* III, *Kpn* I, *Pst* I, *Pvu* I, *Sac* I, *Sal* I, *Sca* I and *Xba* I. These observations are difficult to explain by methylation differences, since it is unlikely that methylation can affect sites for all enzymes tested.

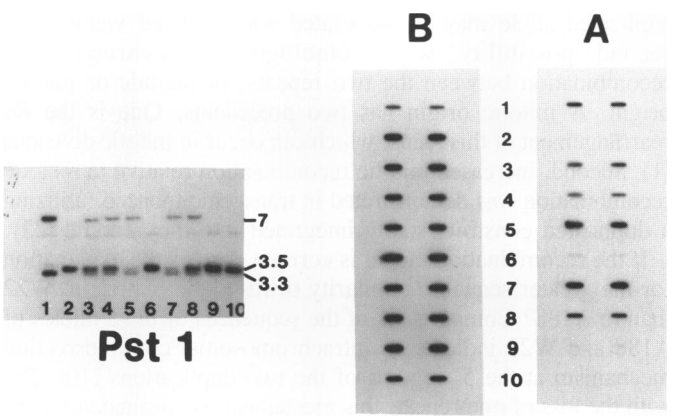

Figure 3. RNA and DNA analysis of individual endosperms: RNA and DNA were isolated from ten individual endosperms of the plant corresponding to DAP 16. Slot blots of duplicate samples of RNA were prepared, probed sequentially with the A and B oligonucleotide probes, and autoradiograms are shown in the panels on the right, with the samples labelled 1 through 10. DNA from the same endosperms was digested with *Pst* I, and a genomic Southern blot of these probed as above is shown in the left panel. The *S* allele gives the two bands of 7 and 3.3kb, and the *Rb* allele generates the 3.5kb band.

To investigate this on the genetic level, DNA and RNA were prepared from single endosperms of a plant corresponding to DAP16 with an intermediate level of A gene transcripts. RNA was loaded on slot blots and probed sequentially with the gene-specific probes, and DNA was digested with *Pst* I, which gives bands of 7 and 3.3kb from the *S* allele (Fig. 3). Note that the oligonucleotide probe for B was diluted three-fold to give comparable band intensities for A and B transcripts. Here also, complete correspondence was observed between the appearance of a new band of 3.5kb and the reduction of A transcripts. In addition, the relative ratios of B to A transcripts (quantified by densitometry, data not shown) and ratios of the new 3.5 and the old 3.3kb bands are consistent with the 1:2 and 2:1 ratios expected for triploid endosperm from a selfed heterozygous plant. These results suggest a segregating allele of the 27kD locus containing only the B gene. The new 9kb *Eco* RI and 3.5kb *Pst* I bands and the loss of the binding site for the A probe in this allele (termed *Rb* for recombinant with the B gene) are best explained by a homologous recombination event between the highly similar 5′ flanking sequences of the A and B genes.

This should give rise to a deletion of genomic DNA. For instance, the *S* allele has one *Bgl* II site 3′ of the B gene while the next 5′ site is >25kb away. In Fig. 2, *S*/*S* A188 leaf DNA has a band of >23kb, whereas DNA from DAP 22 and 18, which have only the new allele has only a lower band of <23kb, as expected for a deletion. Similarly, *Cla* I has one site 3′ of the B gene, one in the region between the repeats, and the next 5′ site is >20kb away. Recombination causes the deletion of the site between the repeats, and the 5kb band is shifted to a large one (Fig. 2). Since *Cla* I cuts maize DNA infrequently (6), the large fragment is not visualized clearly.

We do not know whether this allele is formed during meiosis

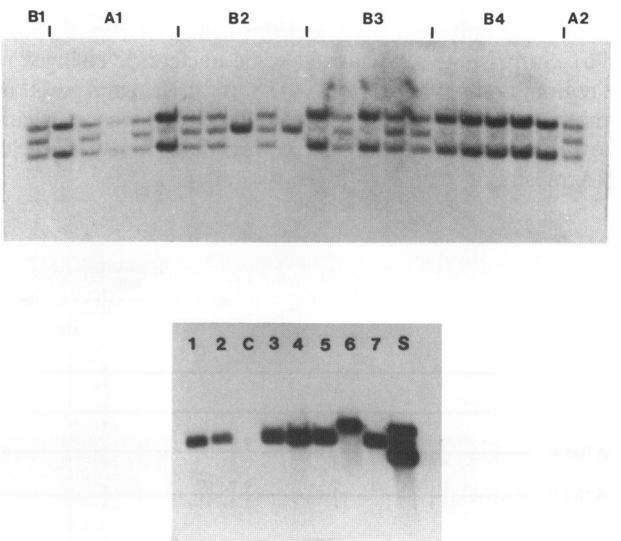

Figure 4. Southern blots of segregating A188 progeny and *Rb* clones: The upper panel shows autoradiograms of genomic Southern blots of leaf DNA from individual A188 plants after *Bam* HI digestion probed for the 27kD genes. Lanes A1 correspond to five progeny of a plant whose kernels, when harvested 30 DAP and analyzed, showed the presence of segregating *S* and *Rb* alleles. Lane A2 represents one progeny of a selfed sibling of this plant. Lanes B1 through B4 correspond to progeny of selfed siblings of the plant in A2. The lower panel is an autoradiogram of DNA from seven l clones (lanes 1−7) and a cosmid clone of the *S* allele (lane S). Lane C is a DNA control from a negative λ clone. All DNAs were digested with *Bam* HI and probed as before.

or early in development during mitosis, a pathway similar to the formation of *Ra*. In either case, the newly formed *Rb* allele should become heritable. Indeed, in Fig. 4, upper panel, leaf DNA of several progeny obtained from one plant were observed to segregate for the *S* and *Rb* alleles (samples A1). One progeny of siblings of this plant was also heterozygous (A2). Second generation progeny of siblings of this plant (samples B1–B4) were also found to have inherited the *Rb* allele. Most A188 stocks from Dr. Phillips were, however, homozygous for the *S* allele. No isolates that carried the *Rb* allele had the *Ra* allele (1) and vice versa. Limited surveys of other maize inbred lines have identified several carrying the *Ra* allele, though none have been found so far with *Rb*.

Cloning and sequence analysis of the *Rb* allele

Segregation of a heritable *Rb* allele simplified the cloning and molecular characterization of the recombination product from a single event. Size fractionated endosperm DNA, enriched for the *Bam*HI fragment of the *Rb* allele from plants corresponding to samples d20 and d22, was used to make a partial genomic library in λ EMBL-3, and 13 positive plaques were purified. Seven of these were compared by Southern blot hybridization to a cosmid clone of the *S* allele from W22 (Fig. 4, lower panel). Six of the positive clones had the single *Bam*HI band diagnostic for the *Rb* allele (see upper panel); this was also the case for the other six (not shown). However, one clone contained the larger of the two *Bam*HI fragments of the *S* allele, and probably arose from incomplete separation in the sucrose gradient.

Sequence data from the *Sal*I site at the 5′ flanking region to the *Pst*I site in the coding sequence of *Rb* (Fig. 1) was compared to the corresponding regions of the A and B genes from A188 (10), and to the W22 A and B genes (13). Fig. 5 is a schematic representation of the comparisons, which span 3100 bases. Note that both A genes contained three *Pst* I sites whereas *Rb* and the B genes had only two (first and third lines of Fig. 5, position 2896). Surprisingly, *Rb* carried some nucleotide changes in its 5′ region compared to the A188 *S* allele, and was more homologus to the *S* allele of W22. This was particularly noteworthy since A188 sequences were determined from cloned DNA of sibling plants with the two alleles.

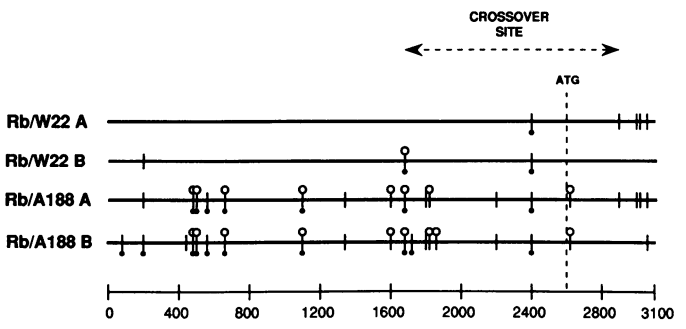

Figure 5. Sequence comparisons of *Rb* to *S*: Each line represents a pairwise sequence comparison of *Rb* to the particular gene indicated. Vertical lines indicate positions which differ, an open circle above the vertical line indicates a difference involving more than one base and a filled circle below each line indicates an insertion/deletion difference. The comparisons span 3.1kb as indicated on the linear scale below, and extend from the conserved *Sal* I site in the 5′ flanking sequence of all genes to the next *Pst* I site in the 3′ direction that is present in all genes (note that the A genes have a site not found for B). The start codon and the crossover region are indicated.

The *Rb* crossover site could not be defined as narrowly as the *Ra* site due to the extensive homology in the crossover region. The *Pst* I site mentioned above, which is absent in *Rb* and the B genes, but present in the A genes, defined the 3′ end of this crossover zone. The 5′ end was defined by the mutation at position 1728, where the B gene had eleven repeats of the dinucleotide TA, whereas *Rb* and the A gene had only ten. At the only other difference 5′ of this point between the A and B genes, *Rb* was similar to the A gene. There was no obvious insertion or deletion similar to those created at excision sites of transposons in maize (15), although the deletion of an A residue at position 2402 within the crossover region may be a candidate.

DISCUSSION

Both the *Rb* and the *S* alleles of A188 were cloned from the same stock of A188. Three possibilities can explain the heterozygosity at the 27kD locus in this stock, namely pollen contamination, residual heterozygosity from the original parents and recombination of the *S* allele in A188. Pollen contamination is always a possibility, particularly since *Rb* is more similar to W22 than A188. Still, this is unlikely since we have not yet identified another maize line with *Rb* and since A188 has a distinct dwarf phenotype whereas its hybrid progeny are tall due to heterosis. Also, A188 kernels are white, due probably to the *y* allele, while most inbred lines, including W22, are yellow. Pollen contamination is easily detected by color in kernels and by unmistakable yellow-white segregation in the next generation. Residual heterozygosity from the original parents is also unlikely, since the stock used in these studies dates from approximately 40 years after the release of A188 (23), and should, therefore, be homozygous. The third possibility is recombination of the *S* allele during the propagation of A188, which has a precedent in the generation of the *Ra* allele. The *Rb* allele could have been formed from *S* by unequal crossover in meiosis, as for the duplicated genes of the *R* locus in maize (21). This, however, should also yield a triplication, with two A genes and one B gene. Though we have not observed a triplicated allele in A188 stocks, our surveys may not have been extensive enough, or the triplicated allele may be associated with reduced viability. A second possibility is a homologous *intra*chromosomal recombination between the two repeats, of meiotic or mitotic origin. A mitotic origin has two precedents. One is the *Ra* rearrangement of this locus, which can occur in mitotic divisions (1). Second, increased mitotic recombination relative to meiotic recombination was demonstrated in transgenic tobacco, utilizing a duplicated construct stably integrated into the genome (22).

If the recombination model is correct, what is the explanation for the greater sequence similarity of *Rb* to the *S* allele in W22 than to A188? Comparisons of the sequences of the *S* alleles of A188 and W22 indicate an intrachromosomal copy correction mechanism at the 5′ regions of the two duplications (10). But with the loss of one repeat, this mechanism is eliminated for *Rb* and copy correction in A188 would cause sequence changes only in the duplicated allele. This would explain why the W22 allele is virtually identical to *Rb*, whereas otherwise, one may have expected copy correction in W22 to generate differences from *Rb* that are conserved between the repeats in W22. This model should be testable by PCR sequencing of the relevant regions.

In any case, we have now derived from the *S* allele two recombinant germinal alleles, whose crossover sites flank the coding region of the 27kD zein gene. Gene dosage results

obtained previously (9) and here (Fig. 3, and densitometric scans of this data) demonstrate good correlation between mRNA accumulation and dosage of the corresponding A and B genes. Dosage changes of either gene were rendered possible by the expression of the *S* allele in heterozygous states with either the *Ra* or the *Rb* allele. This correlation suggests *cis*-acting sequences for gene expression that are conserved, despite the elimination of parts of the *S* allele by recombination. On the other hand, recombination yields hybrid A and B genes with the fusions of the two sequences defined by the respective crossover sites. Given that hybrid genes can be formed by exchanging either the 5′ or the 3′ flanking sequences of the two genes without apparent change in the accumulation of either mRNA, one would expect that any regulatory signals in these flanking regions are interchangable. However, the differential accumulation of A and B messages should correlate to sequence differences within the regions flanked by the 5′ and 3′ crossover sites. Sequence differences within this region can in turn be utilized to localize a potential *cis*-acting signal for a mechanism that lead to the differential accumulation of mRNA.

The sequences of the 5′ flanking regions of the two genes are identical for 2.8kb in W22, except for two differences. *Rb* is identical to the A gene at these positions and further upstream; therefore, differences here and further upstream cannot cause the expression difference. *Ra* is identical to the B gene at positions 3′ of the *Ra* crossover site, between 188 and 220bp 3′ of the stop codon (between positions 857 and 889 in Fig. 6). Therefore, sequences 3′ of this point cannot be responsible, including the 'GT box' of sequence TGTGTTT (position 897, Fig. 6), that had been previously considered as a possible target for the differential expression of these genes (9, 16). This also argues against distant enhancer or silencer elements. Models in which the presence of a transcribed upstream gene increases transcription of a downstream gene (17), are ruled out since the single *Rb* gene maintains high expression. DNA modifications, such as differential methylation, should also act on only the sequences between these endpoints, since the recombinant alleles are likely to have inherited parental methylation in the flanking regions. Opposite strand transcription has been shown to be absent for these genes (Wu, L., Ueda, T. and Messing, J. unpublished).

Therefore, the coding sequences and 188bp of 3′ flanking region are likely to be the sites of *cis*-acting sequence differences (Fig. 6). Of the 188bp 3′, the first 147 bases are identical between all A and B genes. The remaining 37 bases contain eight differences in W22 (between positions 821 and 858, Fig. 6). However, seven of these are not found in A188, and are unlikely to cause differential expression, since these inbred lines show similar expression of each gene (data not shown). One difference, a G−A change, is conserved between the three A genes relative to the two B genes (position 826, Fig. 6). This single base change at position 826 is immediately adjacent to the second of the two poly-A signals (AATGAA, Fig. 6), and precedes the 3′ end of the A transcript at position 842 (18). The location of this sequence difference points to potential difference in transcript stability or processing (19). Within the coding sequences, only three differences are conserved (positions 339, 438 and 448). The location of these differences also point to differential transcript stability or processing (20). In summary, these four base changes help us define potential signal sequences for factors involved in the differential accumulation of mRNA. Nuclear run-off analysis of transcription in developing endosperm tissue (11) from homozygous *Ra* and *Rb* plants, and gene fusion experiments in transgenic or transient expression systems (24) should clarify whether the differential regulation occurs at the DNA or RNA level, and which of the above sequence differences acts as the cis-acting signal.

```
           1                                                          50
    W22-A  ATGAGGGTGT TGCTCGTTGC CCTCGCTCTC CTGGCTCTCG CTGCGAGCGC
    A188-A ATGAGGGTGT TGCTCGTTGC CCTCGCTCTC CTGGCTCTCG CTGCGAGCGC
    A188-B ATGAGGGTGT TGCTCGTTGC CCTCGCTCTC CTGGCTCTCG CTGCGAGCGC
     W22-B ATGAGGGTGT TGCTCGTTGC CCTCGCTCTC CTGGCTCTCG CTGCGAGCGC
 CONSENSUS ATGAGGGTGT TGCTCGTTGC CCTCGCTCTC CTGGCTCTCG CTGCGAGCGC
           ***
           51                                                         100
    W22-A  CACCTCCACG CATACAAGCG GCGGCTGCGG CTGCCAGCCA CCGCCGCCGG
    A188-A CACCTCCACG CATAACAGCG GCGGCTGCGG CTGCCAGCCA CCGCCGCCGG
    A188-B CACCTCCACG CATAACAGCG GCGGCTGCGG CTGCCAGCCA CCGCCGCCGG
     W22-B CACCTCCACG CATACAAGCG GCGGCTGCGG CTGCCAGCCA CCGCCGCCGG
 CONSENSUS CACCTCCACG CATA..AGCG GCGGCTGCGG CTGCCAGCCA CCGCCGCCGG
                      **
           101                                                        150
    W22-A  TTCATCTACC GCCGCCGGTG CATCTGCCAC CTCCGGTTCA CCTGCCACCT
    A188-A TTCATCTACC GCCGCCGGTG CATCTGCCAC CTCCGGTTCA CCTGCCACCT
    A188-B TTCATCTACC GCCGCCGGTG CATCTGCCAC CTCCGGTTCA CCTGCCACCT
     W22-B TTCATCTACC GCCGCCGGTG CATCTGCCAC CTCCGGTTCA CCTGCCACCT
 CONSENSUS TTCATCTACC GCCGCCGGTG CATCTGCCAC CTCCGGTTCA CCTGCCACCT
           151                                                        200
    W22-A  CCGGTGCATC TCCCACCGCC GGTCCACCTG CCGCCGCCGG TCCACCTGCC
    A188-A CCGGTGCATC TCCCACCGCC GGTCCACCTG CCGCCGCCGG TCCACCTGCC
    A188-B CCGGTGCATC TCCCACCGCC GGTCCACCTG CCGCCGCCGG TCCACCTGCC
     W22-B CCGGTGCATC TCCCACCGCC GGTCCACCTG CCGCCGCCGG TCCACCTGCC
 CONSENSUS CCGGTGCATC TCCCACCGCC GGTCCACCTG CCGCCGCCGG TCCACCTGCC
           201                                                        250
    W22-A  ACCGCCGGTC CATGTGCCGC CGCCGGTTCA TCTGCCGCCG CCACCATGCC
    A188-A ACCGCCGGTC CATGTGCCGC CGCCGGTTCA TCTGCCGCCG CCACCATGCC
    A188-B ACCGCCGGTC CATGTGCCGC CGCCGGTTCA TCTGCCGCCG CCACCATGCC
     W22-B ACCGCCGGTC CATGTGCCGC CGCCGGTTCA TCTGCCGCCG CCACCATGCC
 CONSENSUS ACCGCCGGTC CATGTGCCGC CGCCGGTTCA TCTGCCGCCG CCACCATGCC
           251                                                        300
    W22-A  ACTACCCTAC TCAACCGCCC CGGCCTCAGC CTCATCCCCA GCCACACCCA
    A188-A ACTACCCTAC TCAACCGCCC CGGCCTCAGC CTCATCCCCA GCCACACCCA
    A188-B ACTACCCTAC TCAACCGCCC CGGCCTCAGC CTCATCCCCA GCCACACCCA
     W22-B ACTACCCTAC TCAACCGCCC CGGCCTCAGC CTCATCCCCA GCCACACCCA
 CONSENSUS ACTACCCTAC TCAACCGCCC CGGCCTCAGC CTCATCCCCA GCCACACCCA
           301                                                        350
    W22-A  TGCCCGTGCC AACAGCCGCA TCCAAGCCCG TGCCAGCTGC AGGGAACCTG
    A188-A TGCCCGTGCC AACAGCCGCA TCCAAGCCCG TGCCAGCTGC AGGGAACCTG
    A188-B TGCCCGTGCC AACAGCCGCA TCCAAGCCCG TGCCAGCTAC AGGGAACCTG
     W22-B TGCCCGTGCC AACAGCCGCA TCCAAGCCCG TGCCAGCTAC AGGGAACCTG
 CONSENSUS TGCCCGTGCC AACAGCCGCA TCCAAGCCCG TGCCAGCT.C AGGGAACCTG
                                                  *******
           351                                                        400
    W22-A  CGGCGTTGGC AGCACCCCGA TCCTGGGCCA GTGCGTCGAG TTCCTGAGGC
    A188-A CGGCGTTGGC AGCACCCCGA TCCTGGGCCA GTGCGTCGAG TTCCTGAGGC
    A188-B CGGCGTTGGC AGCACCCCGA TCCTGGGCCA GTGCGTCGAG TTCCTGAGGC
     W22-B CGGCGTTGGC AGCACCCCGA TCCTGGGCCA GTGCGTCGAG TTCCTGAGGC
 CONSENSUS CGGCGTTGGC AGCACCCCGA TCCTGGGCCA GTGCGTCGAG TTCCTGAGGC
           401                                                        450
    W22-A  ATCAGTGCAG CCCGACGGCG ACGCCCTACT GCTCGCCTCA GTGCCAGTCG
    A188-A ATCAGTGCAG CCCGACGGCG ACGCCCTACT GCTCGCCTCA GTGCCAGTCG
    A188-B ATCAGTGCAG CCCGACGGCG ACGCCCTACT GCTCGCCCCA GTGCCAGGCG
     W22-B ATCAGTGCAG CCCGACGGCG ACGCCCTACT GCTCGCCCCA GTGCCAGGCG
 CONSENSUS ATCAGTGCAG CCCGACGGCG ACGCCCTACT GCTCGCC.CA GTGCCAG.CG
                                                  **********
           451                                                        500
    W22-A  TTGCGGCAGC AGTGTTGCCA GCAGCTCAGG CAGGTGGAGC CGCAGCACCG
    A188-A TTGCGGCAGC AGTGTTGCCA GCAGCTCAGG CAGGTGGAGC CACAGCACCG
    A188-B TTGCGGCAGC AGTGTTGCCA GCAGCTCAGG CAGGTGGAGC CACAGCACCG
     W22-B TTGCGGCAGC AGTGTTGCCA GCAGCTCAGG CAGGTGGAGC CACAGCACCG
 CONSENSUS TTGCGGCAGC AGTGTTGCCA GCAGCTCAGG CAGGTGGAGC CaCAGCACCG
           **
           501                                                        550
    W22-A  GTACCAGGCG ATCTTCGGCT TGGTCCTCCA GTCCATCCTG CAGCAGCAGC
    A188-A GTACCAGGCG ATCTTCGGCT TGGTCCTCCA GTCCATCCTG CAGCAGCAGC
    A188-B GTACCAGGCG ATCTTCGGCT TGGTCCTCCA GTCCATCCTG CAGCAGCAGC
     W22-B GTACCAGGCG ATCTTCGGCT TGGTCCTCCA GTCCATCCTG CAGCAGCAGC
 CONSENSUS GTACCAGGCG ATCTTCGGCT TGGTCCTCCA GTCCATCCTG CAGCAGCAGC
           551                                                        600
    W22-A  CGCAAAGCGG CCAGGTCGCG GGGCTGTTGG CGGCGCAGAT AGCGCAGCAA
    A188-A CGCAAAGTGG CCAGGTCGCG GGGCTGTTGG CGGCGCAGAT AGCGCAGCAA
    A188-B CGCAAAGTGG CCAGGTCGCG GGGCTGTTGG CGGCGCAGAT AGCGCAGCAA
     W22-B CGCAAAGTGG CCAGGTCGCG GGGCTGTTGG CGGCGCAGAT AGCGCAGCAA
 CONSENSUS CGCAAAGtGG CCAGGTCGCG GGGCTGTTGG CGGCGCAGAT AGCGCAGCAA
           601                                                        650
    W22-A  CTGACGGCGA TGTGCGGCCT GCAGCAGCCG ACTCCATGCC CCTACGCTGC
    A188-A CTGACGGCGA TGTGCGGTCT GCAGCAGCCG ACTCCATGCC CCTACGCTGC
    A188-B CTGACGGCGA TGTGCGGTCT GCAGCAGCCG ACTCCATGCC CCTACGCTGC
     W22-B CTGACGGCGA TGTGCGGTCT GCAGCAGCCG ACTCCATGCC CCTACGCTGC
 CONSENSUS CTGACGGCGA TGTGCGGtCT GCAGCAGCCG ACTCCATGCC CCTACGCTGC
           651                                                        700
    W22-A  TGCCGGCGGT GTCCCCCACT GAAGAAACTA TGTGCTGTAG TATAGCCGCT
    A188-A TGCCGGCGGT GTCCCCCACT GAAGAAACTA TGTGCTGTAG TATAGCCGCT
    A188-B TGCCGGCGGT GTCCCCCACT GAAGAAACTA TGTGCTGTAG TATAGCCGCT
     W22-B TGCCGGCGGT GTCCCCCACT GAAGAAACTA TGTGCTGTAG TATAGCCGCT
 CONSENSUS TGCCGGCGGT GTCCCCCACT GAAGAAACTA TGTGCTGTAG TATAGCCGCT
                           * **
```

```
          701                                                   750
    W22-A GCCCGCTGGC TAGCTAGCTA GTTGAGTCAT TTAGCGGCGA TGATTGAGTA
   A188-A GCCCGCTGGC TAGCTAGCTA GTTGAGTCAT TTAGCGGCGA TGATTGAGTA
   A188-B GCCCGCTGGC TAGCTAGCTA GTTGAGTCAT TTAGCGGCGA TGATTGAGTA
    W22-B GCCCGCTGGC TAGCTAGCTA GTTGAGTCAT TTAGCGGCGA TGATTGAGTA
CONSENSUS GCCCGCTGGC TAGCTAGCTA GTTGAGTCAT TTAGCGGCGA TGATTGAGTA

          751                                                   800
    W22-A ATAATGTGTC ACGCATCACC ATGGGTGGCA GTGTCAGTGT GAGCAATGAC
   A188-A ATAATGTGTC ACGCATCACC ATGGGTGGCA GTGTCAGTGT GAGCAATGAC
   A188-B ATAATGTGTC ACGCATCACC ATGGGTGGCA GTGTCAGTGT GAGCAATGAC
    W22-B ATAATGTGTC ACGCATCACC ATGGGTGGCA GTGTCAGTGT GAGCAATGAC
CONSENSUS ATAATGTGTC ACGCATCACC ATGGGTGGCA GTGTCAGTGT GAGCAATGAC

          801                                                   850
    W22-A CTGAATGAAC AATTGAAATG *AAAAGAAAA AAGT**AT*T GTTCCAAATT
   A188-A CTGAATGAAC AATTGAAATG *AAAAGAAAA TACTCCATCT GTTCCAAATT
   A188-B CTGAATGAAC AATTGAAATG **AAAAAAAA TACTCCATCT GTTCCAAATT
    W22-B CTGAATGAAC AATTGAAATG AAAAAAAAAA TACTCCATCT GTTCCAAATT
CONSENSUS CTGAATGAAC AATTGAAATG *aAAA.AAAA tAcTccATcT GTTCCAAATT
            ******      **** ***           *
          851                                                   900
    W22-A AAA***CGTT TTAACCTTTT AATAGGTTTA TACAATAATT GATATA**TG
   A188-A AAAATTCATT TTAACCTTTT AATAGGTTTA TACAATAATT GATATA**TG
   A188-B AAAATTCATT TTAACCTTTT AATAGGTTTA TACAATAATT TATATATGTG
    W22-B AAAATTCATT TTAACCTTTT AATAGGTTTA TACAATAATT TATATATGTG
CONSENSUS AAAattCaTT TTAACCTTTT AATAGGTTTA TACAATAATT .ATATA..TG
                  ************************************       ****

          901              919
    W22-A TTTTCTGTAT ATGTCTA*A
   A188-A TTTTCTGTAT ATGTCTA*A
   A188-B TTTTCTATAT ATATCTAGA
    W22-B TTTTCTATAT ATATCTAGA
CONSENSUS TTTTCT.TAT AT.TCTA.A
          ***
```

Figure 6. Nucleotide sequence comparisons of the coding and 3′ regions: The sequences of, respectively, the A gene from W22, the A gene from A188, the B gene from A188, the B gene from W22 and the consensus sequence are shown. Position 1 corresponds to the start codon. Blanks in the consensus sequence indicates positions which differ, and asterisks in the sequence text represent spaces introduced for alignment. Asterisks below each block indicate, respectively, the start codon (position 1), a copy correction site (65–66), the *Pst* I site unique to the A genes (337–342), the binding site for the gene-specific oligonucleotides (435–452), the stop codon (670), the poly-A signals (804–810 and 817–823), the 3′ end of the transcript (842), the Ra crossover site (859–890) and the 'GT' box (897–903). Entries in boldface show nucleotides conserved between A188 and W22, and different between the A and B genes.

ACKNOWLEDGEMENTS

We thank Drs. Carl Price and Thomas Peterson for reviewing the manuscript. This work was supported by a grant from the National Institutes of Health (GM43261).

REFERENCES

1. Das, O. P., Levi-Minzi, S., Koury, M., Benner, M. and Messing, J. (1990b) *Proc. Natl. Acad. Sci. USA* **87**, 7809–7813.
2. Messing, J. (1987) in *Genetic Engineering 6.* Rigby, P. (ed.), Academic Press, New York, pp 1–47.
3. Wilson, C. M., (1986) *Plant Physiol.* **82**,196–202.
4. Goodman, M. M. and Brown, W. L. (1988) In Corn and Corn Improvement, Third Edition, Sprague G. F. and Dudley, J. W. (eds.) American Society of Agronomy, Madison, WI.
5. Maeda, N. and Smithies, O. (1986) *Ann. Rev. Genet.* **20**, 81–108.
6. Heidecker, G., Chaudhuri, S. and Messing, J. (1991) *Genomics,* In Press.
7. Kirihara, J., Petri, J. B., and Messing, J. (1988) *Gene* **71**, 359–370.
8. Pederson, K., Argos, P., Narayana, S. V. L. and Larkins, B. A. (1986) *J. Biol. Chem.* 261, 6279–6284.
9. Das, O. P. and Messing, J. (1987) *Mol. Cell. Biol.* **7**, 4490–4497.
10. Das, O. P., Ward, K., Ray, S. and Messing, J., in preparation, (X56117, X56118).
11. Das, O. P., Cruz-Alvarez, M., Chaudhuri, S. and J. Messing (1990a) *Methods Mol. Cell. Biol.* **1**, 213–222.
12. Messing, J. and Bankier, A. T. (1989) in *Nucleic Acids Sequencing: A Practical Approach* Ward, E. S. and Howe, C. J. (eds.), IRL Press Ltd., Oxford, pp 1–36.
13. Geraghty, D. E., 1985 Ph. D. Thesis, University of Minnesota, St. Paul, MN.
14. Chomet, P. S., Wessler, S. and Dellaporta, S. L. (1987) *EMBO J.* **6**, 295–302.
15. Peterson, T. P. (1990) *Genetics* **126**, 469–476.
16. Elliston, K. E. and Messing, J. (1987) in *Architecture of Eukaryotic Genes,* Kahl, G. (ed.), VCH Verlagsgesellschaft mbH, Weinheim, FRG pp 22–56.
17. McStay, B. and Reeder, R. H. (1986) *Cell* **47**, 913–920.
18. Wang, S.-Z. and Esen, A. (1986) *Plant Physiol.* **81**, 70–74.
19. Chodchoy, N., Levine, B. J., Sprecher, C., Skoultchi, A. I. and Marzluff, W. F. (1987) *Mol. Cell. Biol.* **7**, 1039–1047.
20. Brawerman, G. (1986) *Cell* **48**, 5–6.
21. Dooner, H. K. and Kermicle, J. L. (1971) *Genetics* **67**, 427–436.
22. Peterhans, A., Schlupmann, H., Basse, C. and Paszkowski, J. (1990) *EMBO J.* **11**, 3437–3445.
23. Maize Breeder's Handbook, MBS Inc., Ames, Iowa.
24. Ueda, T., and Messing, J. (1991) *Theoret. Appl. Genet.*, In Press.