The Plant Cell, Vol. 4, 811–820, July 1992

Alternative Splicing Induced by Insertion of Retrotransposons into the Maize *waxy* Gene

Marguerite J. Varagona,[1] Michael Purugganan, and Susan R. Wessler[2]

Department of Botany, University of Georgia, Athens, Georgia 30602

The molecular basis for the low level expression of three alleles of the maize *waxy* (*Wx*) gene has been described. Each allele contains a retrotransposon in intron sequences. These insertions represent previously undescribed elements, and their association with three *wx* alleles indicates that retrotransposon elements are important agents of spontaneous mutation in maize. For each allele, element sequences are spliced from pre-mRNA with the surrounding intron even though the insertions increase intron length by approximately 40- to 60-fold. In addition, despite differences in element sequences, insertion sites, and relative orientations, each element disrupts long-range splice site recognition leading to novel *Wx* transcripts where exons both upstream and downstream of the insertion site are skipped. The expression of *wx* alleles with large insertions in introns provides support for studies that indicate that the primary *cis* requirement for maize introns is the splice donor and acceptor sites.

INTRODUCTION

Insertion of a transposable element into a gene can lead to changes in spatial and temporal patterns of gene expression or to novel responses to developmental and environmental stimuli (Weil and Wessler, 1990; Smith and Corces, 1991). These new patterns of expression usually result from complex interactions between regulatory and processing signals residing in both the element and the gene. For example, insertion into 5′ flanking sequences can place element promoters, which are usually terminally located, near the TATA box of adjacent genes. This juxtaposition can dramatically alter gene expression because protein binding sites involved in element transcription and transposition can now become new *cis*-acting modules that influence transcription initiation of the adjacent gene. Insertion of transposable elements into exons or introns can in turn result in new patterns of pre-mRNA processing.

The *waxy* (*Wx*) gene of maize has proven to be an excellent system for studying the effects of transposable element insertion on gene expression in a higher plant. The gene encodes a starch granule–bound ADP-glucose-glucosyl transferase responsible for amylose biosynthesis in the triploid endosperm of the developing kernel and in haploid pollen and embryo sac (Nelson and Rines, 1962). The locus derives its name from the waxy appearance of mutant kernels, reflecting the lack of amylose in the endosperm. The *waxy* phenotype can also be detected by staining endosperms and pollen grains with I/KI, which turns blue-black in the presence of amylose. This easily scorable, nonlethal mutant phenotype has facilitated the isolation of more than 40 mutant *wx* alleles since the turn of the century (Nelson, 1968).

Molecular characterization of the *wx* mutations has provided the first comprehensive picture of the lesions that underlie mutant phenotypes in plants. Half of the *wx* mutants have been shown to result from the insertion of DNA into the gene (Wessler and Varagona, 1985). Cloning of two of these mutations, *wx-m9* and *wx-844*, led to the isolation of the *Activator* (*Ac*) and *Suppressor-mutator* (*Spm*) transposable elements (Fedoroff et al., 1983; Pereira et al., 1985). Recent studies of three *wx* alleles—*wx-m9*, *wx-B4*, and *wx-m1*—have demonstrated that the transposable *Dissociation* (*Ds*) element can function as an intron (Wessler, 1989, 1991a). For these alleles, it was found that three structurally distinct *Ds* elements inserted into three different *Wx* exons are spliced in a similar manner. Splicing of the 4.3-kb *Ds* element from exon 10 of the *wx-m9* mutation is responsible for this allele's intermediate phenotype (Wessler, 1991b). The intermediate expression of two transposable element alleles of the *bronze* and *a2* loci—*bz-m13* and *a2-m1*—results from the splicing of a second maize transposable element, *defective Spm* (*dSpm*), from coding sequences of these genes (Kim et al., 1987; Raboy et al., 1989; Menssen et al., 1990).

The involvement of RNA splicing in the residual expression of several maize mutations led us to examine *Wx* transcripts from several phenotypically stable alleles that condition intermediate *waxy* phenotypes. In a previous survey of stable *wx* mutations, we found that six of 17 contained large insertions of DNA, ranging in size from 4.5 to 6.1 kb (Wessler and Varagona, 1985). Four of these—*wx-Stonor*, *wx-G*, *wx-B5*, and *wx-M*—were of particular interest because they displayed in-

[1] Current address: Michigan State University, MSU-DOE Plant Research Laboratory, East Lansing, MI 48824.

[2] To whom correspondence should be addressed.

termediate levels of *Wx* expression in the endosperm. In addition, *wx-Stonor* and *wx-M* exhibited a tissue-specific phenotype: both are null for *Wx* expression in the pollen. In contrast, *wx-G* and *wx-B5* are expressed in both endosperm and pollen.

In this report, we have determined the lesions responsible for three of these mutations—*wx-Stonor*, *wx-G*, and *wx-B5*—and the molecular basis for their residual *Wx* expression. We found that each mutation contains an insertion with the structural features of retrotransposons. Two of the elements are inserted in introns, whereas the third is at an intron/exon junction. For each allele, *Wx* expression results from the splicing of element sequences along with the surrounding intron and the consequent generation of wild-type *Wx* transcripts. In addition, despite differences in element sequence, insertion sites, and relative orientation, each element disrupts long-range splice site recognition, creating *Wx* transcripts with novel alternative splicing patterns.

RESULTS

Expression Despite Large Insertions in the *Wx* Gene

The *wx-Stonor*, *wx-G*, and *wx-B5* alleles were previously shown to contain 5.2-, 5.0-, and 6.1-kb insertions, respectively, in a 2.1-kb SalI fragment that spans exons 1 to 9 of the 14 *Wx* exons (Wessler and Varagona, 1985). These alleles condition low levels of amylose in endosperm tissues, indicating that there is some *Wx*-encoded ADP glucose-glucosyl transferase activity despite large insertions in the gene. Residual expression is apparently not due to element excision as the three alleles display uniformly low levels of amylose in all endosperm cells (as revealed by I/KI staining). Element excision would be expected to produce sectors with a revertant or intermediate *Wx* phenotype. In addition, gel blots of genomic DNA from *wx-Stonor*, *wx-B5*, and *wx-G* failed to detect significant levels of somatic excision (data not shown).

Wx enzymatic activity was quantified in these strains by assaying endosperm tissue for starch granule–bound glucosyl transferase. These activities are compared in Table 1 and reported as a percent of the wild-type control: *wx-G*, *wx-B5*, and *wx-Stonor* encode 5, 7.8, and 9.5% of nonmutant activity, respectively.

Table 1. *Waxy* Expression

Allele	Phenotype		Endosperm *Wx* Activity	
	Endosperm	Pollen[a]	^{14}C Incorporation	% Wild Type
Wx	Wild type	Wild type	2663	100.0
wx-Stonor	Intermediate	Null	254	9.5
wx-G	Intermediate	Wild type	135	5.0
wx-B5	Intermediate	Wild type	207	7.8

[a] Intermediate levels of *Wx* activity appear as wild type in pollen.

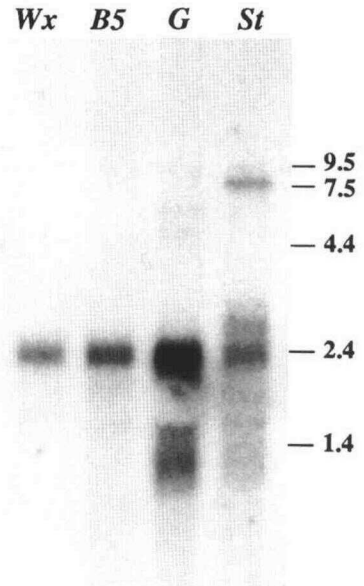

Figure 1. RNA Gel Blot Analysis of *Wx* Transcripts.

Poly(A)$^+$ RNA isolated from immature endosperms (20 to 22 days after pollination) was resolved following electrophoresis in a denaturing agarose gel, transferred to a membrane, and probed with *Wx* DNA. The amount of RNA loaded was as follows: *Wx*, 0.5 μg; *wx-B5*, 5 μg; *wx-G*, 5 μg; *wx-Stonor*, 5 μg. The molecular size standards are indicated at the right in kilobases.

Due to the positions of insertion within the *Wx* transcription unit, we suspected that gene expression resulted from the splicing of element sequences from pre-mRNA. To investigate this possibility, the size and relative abundance of steady state mutant transcripts were compared to wild-type transcripts on RNA gel blots, as shown in Figure 1. Poly(A)$^+$ RNA isolated from immature endosperms of strains homozygous for *wx-Stonor*, *wx-B5*, or *wx-G* contains wild-type size (~2.4 kb) *Wx* transcripts. Additional *Wx* transcripts are evident among RNA from *wx-Stonor* and *wx-G* endosperms. Whereas both of these alleles encode transcripts smaller than 2.4 kb, only *wx-Stonor* encodes a large transcript (7.6 kb) that corresponds in size to wild-type mRNA plus the 5.2-kb insertion. Although all alleles encode wild-type sized transcripts, the steady state level of *Wx* transcripts is reduced approximately 10-fold in the mutants.

Retrotransposon Insertions in *wx-Stonor*, *wx-G*, and *wx-B5*

The insertions and flanking sequences from *wx-Stonor*, *wx-B5*, and *wx-G* were cloned and partially sequenced (see Methods). All three insertions have several features that are characteristic of retrotransposons, including (1) long terminal repeats

A

```
  1  TGTTAGGATG TGTCCTCTAC ACTAGCCGGT
 31  TAACAGTAAT CGGGGTAGGG TTAATCCTAA
 61  CCGTCTAACC GCGCCCCTGA TCGGGGGGCT
 91  ATCCTAACCG TTGGTTGCAG CCCCCGGTCA
121  CACGACAGCC ATATAAATAG ATTAAGGGGG
151  CCGACGGGTT TAGCCAACGC TAATCGCTCG
181  ACCACAAACC CTAATCGACA CTTTGAGAGG
211  AACCGATCCC CTAGGCGCTC GGAGTGACTG
241  GAATCGCTGT GACTATGTCT CAGATCCCCG
271  TTATCCATCG GGATCTGAGA TTCCCCCTGC
301  AGGCAACATT GGAGGAGCTG ATCCACCGCC
331  GATCAGCGAC GGGAACCAAT CTCCTCGACA
361  CCAGCAGGGA GATCGTCAAG AAGGAGGAGG
391  CGGGTCTTCG CTGCAAGACA CCGCCGCAAT
421  GATGGCTAGG AACAACAACG ACGGTATGTT
451  AACCTATACT CCTCAGTAGA ATAGATCTAC
481  TCATCATCGA AGTATTTTAG ACACGGATTG
511  GTATTCTAGG GACTTAGGAA ACGATCCTAA
541  GCACCAGTAA TTTCTAACA
```

B

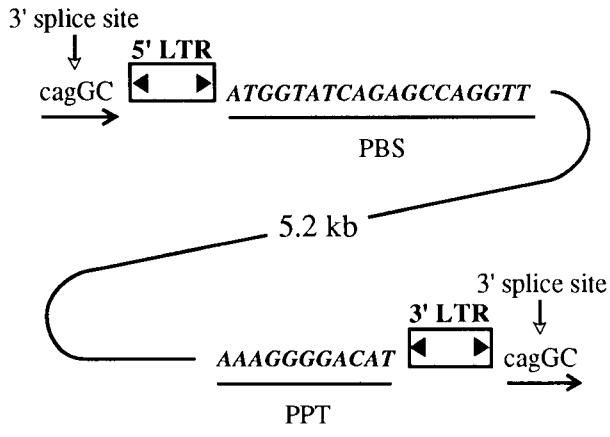

Figure 2. Structural Features of the Stonor Element.

(A) Nucleotide sequences of the Stonor LTR with the 5-bp inverted repeat termini underlined.

(B) Structural features of the Stonor element and flanking *Wx* sequences. The presumptive primer binding site (PBS) and polypurine tract (PPT) are indicated. The PBS is identical to an 18-nucleotide sequence near the 3′ end of the initiator methionyl tRNA from wheat (Ghosh et al., 1982). The designation of 5′ and 3′ LTRs is based on the fact that tRNA-mediated priming of reverse transcription from retroviral or retrotransposon RNA usually occurs a few base pairs downstream of the 5′ LTR. A horizontal arrow indicates the position of the 5-bp direct repeat of the *Wx* sequence generated upon insertion. This sequence includes the end of intron 5 (lowercase) and the beginning of exon 6 (uppercase). The 3′ splice site downstream of the 3′ LTR is utilized in *wx-Stonor*.

(LTRs), (2) short inverted repeats at the termini of the LTRs with TG . . . CA as their terminal dinucleotides, (3) a $tRNA^{Met}$ primer binding site (PBS) adjacent to one copy of the LTR, (4) a polypurine tract (PPT) adjacent to the other LTR, and (5) a 5-bp direct repeat of *Wx* sequence flanking the insertion. These features are summarized in Figures 2 and 3.

Based on their LTR sequences, the three elements (now referred to as Stonor, B5, and G) are not related to any previously described plant retrotransposon (nor to any other eukaryotic element described to date). However, B5 and G appear to be members of a family of retrotransposons distinct from Stonor. Although B5 and G differ in size by about 1 kb, their LTRs differ only by the presence of four extra bases in the LTR of B5 (Figure 3A).

A

```
  1  TGTCAGCTAC CAGCACGTGC GCCCGGCGCT
 31  GAGGAAGGCG AGGAAGAGAA CGAGAGAGGA
 61  ACAGAGAGAT GTTAGGCGAA TATTCTTTCT
 91  TGCATAAGGG TTCACAGCCT CCAGTGTTTC
121  CCTTAATAAC TGCTTACAAG TTTACGGCCC
151  AGCTAATTGA CCCATACTCC ACCCATTACA
181  CGGCCATTCG GCTGCGGACA CGACGGCCCA
211  TAGTTCCTTG CTCTTGGCTG CTTGCACGTG
241  ATAGCTCATC CGACGCGACG GAGTCACTGG
271  GCACTTCCTT GCTGGCTTGA GGGCTGACA
```

B

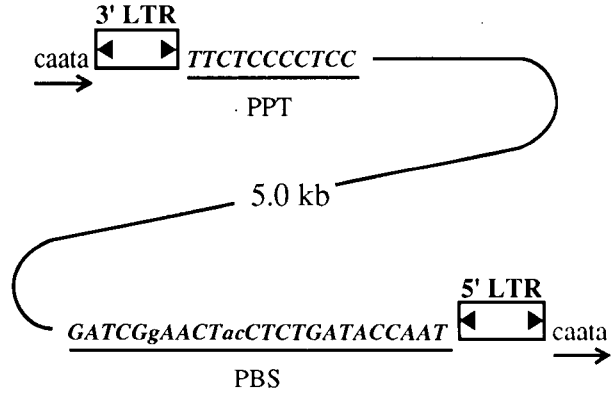

Figure 3. Structural Features of the G and B5 Elements.

(A) Nucleotide sequence of the B5 LTR with the 6-bp inverted repeat termini underlined. Dots are over the four nucleotides that are not found in the LTR of G.

(B) Structural features of the G element and flanking *Wx* sequences. The positions of the PBS and PPT are indicated. The PBS shares 18 of 23 nucleotides found near the 3′ end of the initiator methionyl tRNA from wheat (Ghosh et al., 1982). A horizontal arrow indicates the position of the 5-bp direct repeat generated upon insertion of the G element.

Insertion Sites and Relative Orientations of Stonor, B5, and G

The location of each element within the *Wx* gene is shown in Figure 4. Stonor is inserted at the intron 5–exon 6 junction; the 5-bp host duplication of host sequence resulting from insertion is composed of 3 bp at the 3′ end of intron 5 and 2 bp at the 5′ end of exon 6. As a result of insertion, the splice acceptor site of intron 5 is duplicated and now flanks the Stonor element (Figure 2B). In contrast, both the G and B5 elements are inserted into introns: B5 in intron 2 and G in intron 8. Neither element is near the splice sites: B5 is 54 bp from the 3′ splice site of the 132-bp intron 2, whereas G is 28 bp from the 3′ splice site of the 81-bp intron 8.

The tRNA binding site is found at the 5′ end of retroelement transcription units. Thus, the position of the putative primer binding sites serves to orient the Stonor, B5, and G elements with respect to the direction of *Wx* transcription. Based on these criteria, the putative transcription unit of the Stonor element is in the same (parallel) orientation as *Wx* transcription, whereas B5 and G are inserted in the opposite (antiparallel) orientation (Figure 4).

Alternative Splicing of *wx-Stonor*

To determine the molecular basis for the intermediate *Wx* phenotype conditioned by *wx-Stonor*, *wx-B5*, and *wx-G*, and to investigate the possibility of alternative splicing suggested by RNA gel blot analysis (Figure 1), RNA processing events were characterized. For *wx-Stonor*, splicing events were analyzed by first isolating recombinant phage containing *Wx* sequences from a cDNA library generated from endosperm poly(A)$^+$ RNA. Seventeen clones were identified and their inserts characterized by restriction analysis prior to determining the DNA sequence of the insertion site and flanking DNA. The results are summarized in Figure 5. Seven of the 17 clones recovered from the primary cDNA library were indistinguishable from wild type, indicating that splicing proceeded from the intron 5 splice donor to the duplicated splice acceptor site 5.3 kb downstream and adjacent to the 3′ LTR (Figure 2B). Although the presence of the Stonor LTR alters the context of the downstream splice acceptor site, this sequence is still recognized by the splicing machinery. These data indicate that, despite the presence of a 5.2-kb insertion at an intron–exon junction, element sequences and one copy of the direct repeat are spliced from pre-mRNA, generating a wild-type transcript. Presumably, the low level of enzymatic activity encoded by this allele reflects the reduced steady state level of wild-type transcript found in *wx-Stonor* endosperm tissue (Figure 1).

The DNA sequences of the remaining cDNA clones reveal perturbation of normal splicing downstream of the Stonor element. Five of the 17 cDNA clones define an RNA processing event that involves the splicing of exon 5, the exon immediately upstream of the Stonor insertion site, to exon 12. Other alternative splicing events, represented by a single cDNA clone each, show splicing of exon 5 to exon 8 and exon 5 to exon 9. In all cases, alternative splicing involved exon skipping: transcripts lacked one or more *Wx* exons in addition to all element sequences. Furthermore, the natural *Wx* splice sites were always used; that is, cryptic sites within the *Wx* gene were not observed.

The remaining three cDNAs identify transcripts containing element sequences (data not shown). In one case, exon 5 was spliced to a 125-nucleotide sequence within the Stonor LTR. This exon was then spliced to *Wx* exon 6, resulting in a chimeric transcript with element sequences between exons 5 and 6. The final two cDNA clones represent a truncated transcript containing *Wx* exon 5 spliced to Stonor sequences prior to a poly(A) tract.

Alternative Splicing of *wx-G* and *wx-B5*

To determine the molecular basis for *wx-B5* and *wx-G* expression, first-strand cDNAs were synthesized from endosperm

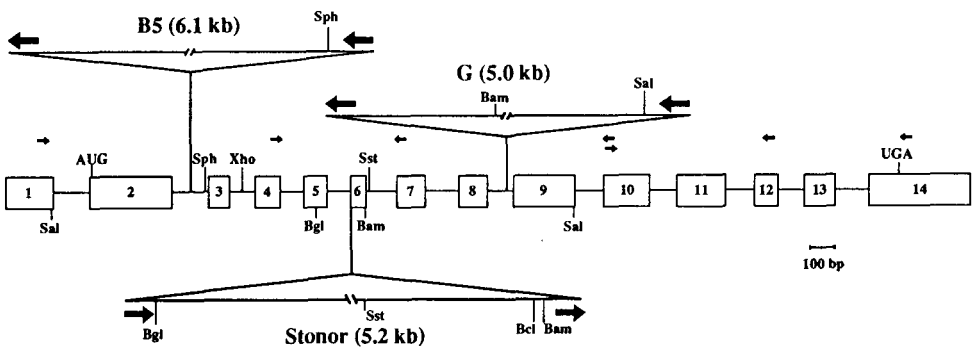

Figure 4. The Position and Orientation of the Retrotransposon with Respect to the Exons and Introns of the *Wx* Gene.

The limits of the Wx protein are indicated by the translation start and stop codons. The positions of relevant restriction sites and oligonucleotide primers used for PCR amplification (small horizontal arrows) are also noted. Exons, open boxes; introns, connecting lines.

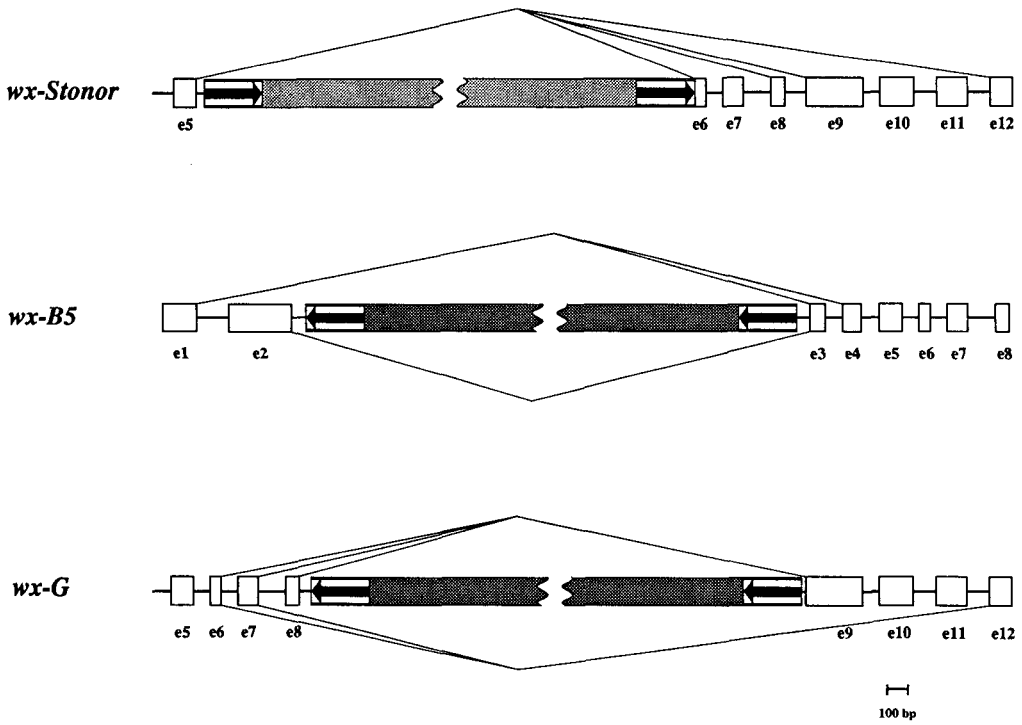

Figure 5. Alternatively Spliced Transcripts Encoded by *wx-Stonor*, *wx-G*, and *wx-B5*.

Elements are shaded and terminate with horizontal arrows representing the LTRs and indicating the orientation of each element's putative transcription unit. Diagonal lines connect the 5′ and 3′ splice sites. In each case, only the splicing events that remove element sequences from *Wx* pre-mRNA are shown. Splicing upstream and downstream of the events shown appears normal (Figure 6).

poly(A)$^+$ RNA and appropriate regions amplified by the polymerase chain reaction (PCR) prior to DNA sequencing. For *wx-B5*, the sequences between exons 1 and 7 were amplified. For *wx-G*, the sequences between exon 4 and exon 10 or 12 were amplified in independent reactions (see Figure 4 for the position of upstream and downstream primers). Amplified DNA was subcloned and restriction mapped prior to sequencing. For both *wx-B5* and *wx-G*, wild-type sequences resulted from normal splicing of exon 2 to exon 3 (for *wx-B5*) and exon 8 to exon 9 (for *wx-G*) (Figure 5). These transcripts account for the wild-type sized bands seen in the RNA gel blot and are presumably responsible for the low levels of activity encoded by these alleles.

In addition to wild-type splicing, alternative splicing events were also detected (Figure 5). For *wx-B5*, two exon-skipping events involved the joining of exon 1 to exon 3 or 4, downstream of the element. Four exon-skipping events, involving exons 6 and 7 spliced to exons 9 and 12, were detected among the DNA amplified from *wx-G* cDNA. For both *wx-G* and *wx-B5*, all splicing events utilized the normal splice sites in the *Wx* gene; in no case were cryptic sites within *Wx* exons or introns activated. Finally, although all three alleles encode alternatively spliced *Wx* transcripts, the absence of smaller transcripts among *wx-B5* RNAs (Figure 1) suggests that, for this allele, the splicing of exon 2 to exon 3 occurs much more frequently than the other splicing events.

Alternative Splicing Is Associated with the DNA Insertions

Detection of alternative splicing among three retrotransposon-induced mutations of *waxy* led us to question whether alternative splicing was in fact aberrant and element induced or whether exon skipping is a normal feature of *Wx* expression. Exon skipping along the entire length of *Wx* transcripts was assessed by amplifying overlapping regions of *Wx* cDNA. In addition, identical primers were used to amplify *wx-B5* and *wx-G* cDNAs to see if alternative splicing is widespread or restricted to the RNA surrounding element sequences. DNA gel blot analysis of the mutant and wild-type PCR products is shown in Figure 6.

Examination of the amplified wild-type products revealed only a single band for each reaction, indicating that alternative splicing is not a feature of nonmutant gene expression. For *wx-B5* cDNAs, the multiple fragments associated with exon skipping were only observed when the sequences between exons 1 and 7 were amplified. Detection of wild-type sized bands among

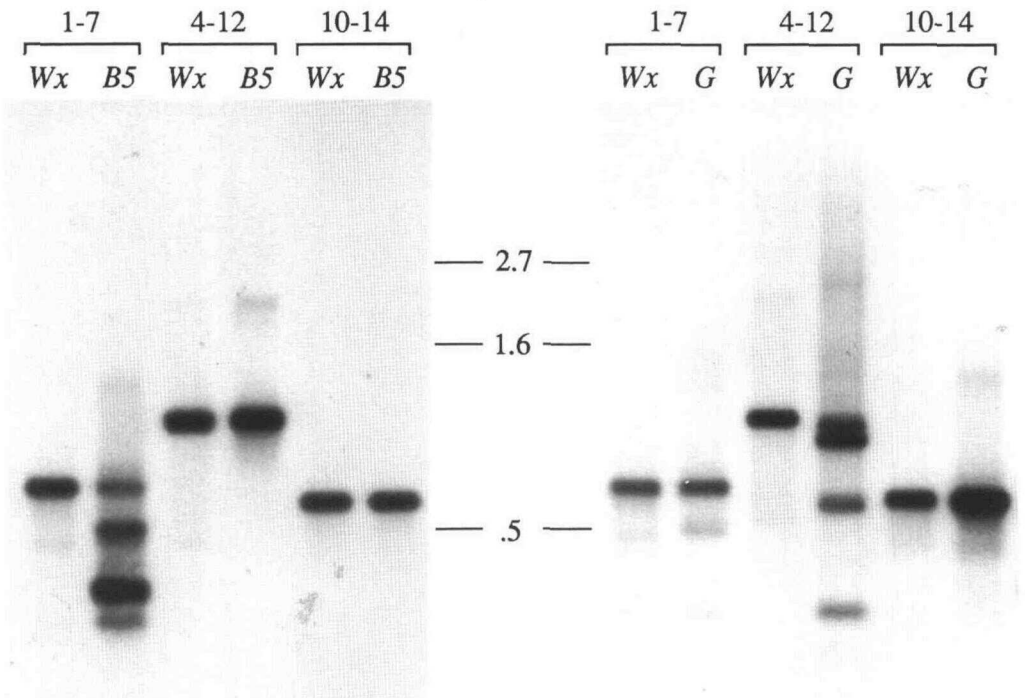

Figure 6. DNA Gel Blot Analysis of PCR Products of *Wx*, *wx-B5*, and *wx-G* cDNAs.

Poly(A)$^+$ RNA from immature endosperms of the three strains was used as substrate for single-stranded cDNA synthesis primed by oligo(dT). Overlapping regions of cDNA were amplified following incubation of approximately 10 ng of cDNA with the primer pair indicated (*Wx* exons 1 and 7, 4 and 12, and 10 and 14). Reaction products were resolved on 1.2% agarose gels, transferred to a membrane, and probed with full-length *Wx* cDNA. The molecular size standards are in kilobase pairs.

the *wx-B5* products of exons 4 to 12 and 10 to 14 amplifications served to further delimit the region of alternative splicing to exons 1 to 4. For *wx-G*, multiple fragments were only associated with sequences between exons 4 to 12. Furthermore, the presence of only a wild-type sized band for the exons 10 to 14 reaction indicated that this region is normally spliced in those transcripts that still contain exon 10. Based on these data, we concluded that alternative splicing of *wx-B5* and *wx-G* transcripts is restricted to the events depicted in Figure 6. Restriction analysis of the 17 *wx-Stonor* cDNA clones also demonstrated that for this allele alternative splicing is confined to the events shown in Figure 5.

DISCUSSION

Retrotransposons Are Important Agents of Spontaneous Mutation in Maize

Retrotransposons are the most prevalent class of mobile elements in Drosophila and yeast (Finnegan and Fawcett, 1986; Boeke and Corces, 1989). In Drosophila, retrotransposons, including the gypsy, copia, and 412 families, comprise at least 2% of the Drosophila genome (Bingham and Zachar, 1989) and may account for as much as 80% of spontaneous mutations (Green, 1988). In yeast, the only known mobile elements are those that belong to the Ty retrotransposon family (Boeke and Corces, 1989). In contrast, relatively little is known about retrotransposons in plants. Of seven retrotransposons that have been characterized to date in higher plants, only two, the maize Bs1 and the tobacco Tnt1 elements, have been associated with mutational events (Johns et al., 1985; Grandbastien et al., 1989). The remaining five elements were isolated as RFLPs or as high copy number repeats; there is no evidence that any of these elements is still capable of transposition (Haberd et al., 1987; Voytas and Ausubel, 1988; Sentry and Smyth, 1989; Smyth et al., 1989; Camirand et al., 1990; Voytas et al., 1990; Moore et al., 1991).

The *waxy* mutations represent a unique opportunity to understand the molecular basis of spontaneous mutation in a higher plant. Seven of 17 spontaneous mutations result from DNA insertions, and six of these seven have large inserts ranging in size from 4.5 kb to 6.1 kb (Wessler and Varagona, 1985). In this study, three of these alleles—*wx-Stonor*, *wx-B5*, and *wx-G*—were found to contain retrotransposons that, based on a comparison of their LTR sequences, are not members of any known family in plants or animals. The remaining insertion

alleles—*wx-M*, *wx-I*, and *wx-K*—have recently been characterized and also contain retrotransposons in *Wx* coding sequences (S. White, L. Hebara, M. Purugganan, and S. R. Wessler, unpublished data). The predominance of retrotransposons among the stable *wx* insertions illustrates the importance of this class of transposable elements in maize and suggests that their relative importance in plant genomes has been seriously underestimated.

Element Insertion Sites

Although this study has highlighted certain similarities in the mutational processes of maize, Drosophila, and yeast, comparison of insertion sites may indicate underlying differences. The Stonor, G, and B5 insertions are all within the *Wx* coding region; *wx-M*, *wx-I*, and *wx-K* also contain coding region insertions (S. White, M. J. Varagona, and S. R. Wessler, unpublished data). In contrast, the majority of retrotransposon insertions in yeast and Drosophila are in 5′ flanking sequences (Bingham and Zachar, 1989; Boeke, 1989).

To explain an apparent insertion site preference in yeast and Drosophila, it has been suggested that preferential insertion into DNAse I hypersensitive sites may be responsible, presumably because of the open chromatin structure of 5′ flanking regions (Voelker et al., 1990). The possibility of bias in mutant selection procedures has also been suggested (Zachar and Bingham, 1982). It seems unlikely that mutant selection bias can explain the preference for coding region insertions into the *Wx* gene. Selection for residual expression might favor insertions into intron sequences if *Wx* were an essential gene. However, the viability of plants homozygous for a deletion of the entire gene argues against this interpretation (Wessler and Varagona, 1985). Rather, the distribution of retrotransposons is the same as that of other *waxy* mutations; the vast majority are located in the coding region. The association of insertion sites with coding sequences probably reflects the relatively large target size rather than sequence preference. These data, taken together with the recent finding that three null alleles of the tobacco *nia* gene contain the Tnt1 element in exon sequences (Pouteau et al., 1991), suggest that there are no preferred insertion sites for retrotransposons in plant genes.

Gene Expression Despite the Presence of Large Insertions

The maize genome contains a large fraction of repetitive DNA, much of which is, or at one time was, transposable (Freeling, 1984). The presence of a large amount of transposable DNA in a plant genome poses unique problems. Plant germ cells differentiate from apical meristems after prolonged vegetative growth, unlike animal germ cells which are set aside early in development and protected. For this reason, somatic mutations in plants can be heritable. Plants may confront these mutation pressures by ensuring that genes that have sustained insertions can still function.

Although *Wx* is not an essential gene, examination of its mutant alleles has revealed several mechanisms used to ensure gene expression despite element insertion. Previous studies of three *wx* alleles containing *Ds* elements in exons (*wx-m9*, *wx-B4*, *wx-m1*) demonstrated that *Ds* can function as an intron (Wessler, 1989, 1991a). Each of these alleles encodes wild-type sized *Wx* mRNA reflecting the splicing of most *Ds* sequences from pre-mRNA. This study has focused on three alleles containing large insertions in *Wx* introns. Like the *Ds* alleles, splicing of element sequences from these genes is also responsible for their wild-type sized *Wx* transcripts.

In Drosophila, residual gene expression has been associated with transposable element insertions into 5′ flanking sequences (Corces and Geyer, 1991), untranslated exons (Levis et al., 1984), and large introns (Scott et al., 1983). Among this latter class are alleles of the *white* locus that condition low levels of gene expression despite the presence of a variety of elements in the 3-kb intron 1 or the 1.5-kb intron 2 (Levis et al., 1984). In contrast, the B5, G, and Stonor elements are inserted into introns of 131, 81, and 97 bp, respectively. Despite the enormous increase in intron size and a dramatic change in intron sequence, normal 5′ and 3′ splice sites are still recognized.

Recent findings regarding the *cis* requirements for splicing in maize provide a possible explanation for why maize genes can still function with large insertions. Goodall and Filipowicz (1989) found that plant introns do not require a conserved branchpoint sequence like yeast introns do, nor do they require a polypyrimidine tract upstream of the 3′ splice site as do mammalian introns. Rather than these features, A + T richness was found to be an important component of plant introns. A more recent analysis of intron requirements in maize revealed that A + T content is not as important in monocots as in dicotyledonous plants as long as consensus 5′ and 3′ splice sites are available (Goodall and Filipowicz, 1991). The apparent minimal *cis* requirements for maize introns imply that insertion of DNA is not likely to disrupt *cis* elements needed for intron definition. For these reasons, the maize splicing machinery may be unusually tolerant of sequence changes within introns as long as splice sites are unaltered.

Unlike *wx-B5* and *wx-G*, the context of the 3′ splice site has been altered in *wx-Stonor* (Figure 2B). However, examination of steady state transcripts encoded by the three alleles (Figure 1) reveals that only *wx-Stonor* contains a larger transcript. The 7.6-kb *wx-Stonor* transcript hybridizes with *Wx* and Stonor sequences (data not shown) and presumably represents a chimeric transcript where the altered intron 5 is unspliced due to the sequence changes in the vicinity of its 3′ splice site. Disruption of this 3′ splice site may also be responsible for the tissue-specific expression of *wx-Stonor*. Unlike *wx-B5* and *wx-G*, which are expressed in both endosperm and pollen, *wx-Stonor* is null for pollen expression, as shown in Table 1. This suggests that recognition of the splice site required to generate a wild-type transcript may occur more efficiently in endosperm

than in pollen. Such tissue-specific expression could result from a deficiency in a splicing factor in pollen relative to its concentration in endosperm cells.

Alternative Splicing of *Wx* Exons

In addition to producing wild-type transcripts, each insertion allele encodes novel alternative splicing events where nonconsecutive exons are ligated in the mature transcript. That these events are element induced is indicated by the fact that exon skipping is restricted to exons around the insertion site and that a nonmutant *Wx* gene is not alternatively spliced (Figure 6). Although element induced, alternative splicing occurs irrespective of the orientation of the element, its sequence, or its position within the *Wx* gene. These data suggest that the identity of the insertions as retrotransposons may not be relevant when considering their effect on RNA processing. Rather, an important factor to focus on in attempting to understand alternative splicing may be the impact of large insertions in maize introns.

Mutations that lead to exon skipping have been reported previously in animal systems and usually involve alterations in intron sequences recognized by the splicing machinery. Splicing defect mutations responsible for thalassemias (Felber et al., 1982; Treisman et al., 1982), albuminerism (Ruffner and Dugaiczy, 1988), and Tay-Sachs disease (Arpaia et al., 1988) all contain changes in the GT/AG sequences at the splice junctions, and each involves skipping of only a single exon. The lesion responsible for neurofibromatosis type 1 in one patient was recently shown to be a 320-bp Alu element in intron 5 of the NF1 gene (Wallace et al., 1991). The authors of this study suggest that mutant transcripts lack exon 6 because Alu insertion interferes with branchpoint recognition.

Skipping of *Wx* exons differs from previous examples in two ways: (1) mutations do not destroy 5′ and 3′ splice sites, and (2) exons that are both upstream and downstream of the insertion site are skipped. The association of exon skipping and DNA insertions leads us to suggest that the presence of element sequences in pre-mRNA delays the processing of nearby introns. Such a delay would result in the simultaneous availability of multiple 5′ and 3′ splice sites and may lead to the abnormal pairing of natural splice sites. It is only after removal of element sequences through splicing that consecutive exons can be joined together. Furthermore, the fact that skipping of upstream exons has not been previously observed may reflect differences in the rate of splicing reactions in plants and animals.

The simultaneous availability of splice sites also provides an explanation for why the 3′ splice site of intron 11 is utilized, among *wx-G* and *wx-Stonor* transcripts, rather than the 3′ splice sites of either intron 9 or 10 (Figure 5). The 3′ acceptor of intron 11 is utilized for two of the five alternative splicing events of *wx-G* (Figure 5). Although only one splicing event of *wx-Stonor* utilized this acceptor, the fusion of exons 5 to 12 was detected in five of the 17 cDNA clones, indicating that it was a relatively frequent event. Introns 9 and 10 have the highest G + C content of all plant introns characterized to date (60 and 58%, respectively) (Goodall and Filipowicz, 1991). In fact, it was the efficient splicing of *Wx* introns 9 and 10 in monocots and not dicots that led Goodall and Filipowicz (1991) to conclude that A + T richness is not an important *cis* requirement of maize introns. Our observation that introns 9 and 10 are correctly spliced in *Wx* genes but not in some transcripts encoded by *wx-G* or *wx-Stonor* can be explained if introns 9, 10, and 11 are simultaneously available in the mutant but not the wild-type pre-mRNA. When given a choice between these introns, the maize splicing machinery may prefer the A + T rich intron 11 (44% GC) over the A + T poor introns 9 and 10.

METHODS

Strains and General Techniques

Strains carrying the alleles *wx-B5*, *wx-G* (from O. Nelson, University of Wisconsin, Madison), *wx-Stonor*, *wx-C34* (from the Maize Genetics Cooperation Stock Center, Urbana, IL), and *wx-m9* (from B. McClintock, Cold Spring Harbor Laboratory, NY) were obtained as indicated. Isolation of genomic DNA from 14-day-old seedlings (Shure et al., 1983) and DNA gel blot analysis and phage screening (Wessler et al., 1986) were as cited. RNA was extracted from immature endosperms (18 to 22 days after pollination [DAP]) as described previously (Shure et al., 1983), except that oligo(dT) column buffers contained 0.1% DEPC rather than SDS. RNA gel blot analysis was as described previously (Wessler et al., 1986).

Starch Granule Extraction and Enzymatic Assay

Starch granules were extracted from endosperm tissue collected 18 to 22 DAP (Shure et al., 1983). Four 1- to 6-mg starch granule aliquots were assayed in duplicate by measuring the incorporation of ^{14}C-UDP glucose (Amersham) into the granules. Measurement of activity was as described previously (Shure et al., 1983), except that a sixfold greater amount of radiolabeled UDP-glucose was used in assaying the mutant *wx* activity. The activity of each sample was corrected by subtracting the measured activity of *wx-C34*, a deletion of the entire gene (Wessler and Varagona, 1985).

Genomic DNA Cloning and Sequencing

The Stonor, B5, and G elements were previously localized by gel blot analysis of genomic DNA (Wessler and Varagona, 1985). DNA instability during cloning precluded the cloning of these inserts on single large restriction fragments. The 5′ half of the Stonor element was cloned on a 5-kb SstI fragment following ligation of size-fractionated genomic DNA into the λ vector ZAPII (Stratagene) and screening of the recombinant phage with radioactively labeled *Wx* probes (see Figure 4 for the position of restriction sites). The 5-kb insert was subcloned into pUC119 (Vieira and Messing, 1987), and a set of unidirectional deletions was generated (Dale et al., 1985). Approximately 2.5 kb of DNA containing the 5′ end of the Stonor element was sequenced from single-stranded templates generated from the deletion series. The 3′ half of

the Stonor element and flanking DNA was cloned on a 6.5-kb BclI fragment following ligation of size-fractionated genomic DNA into the BamHI site of the λ vector EMBL3. A 0.9-kb BamHI-SstI fragment, isolated from the larger BclI fragment, was sequenced following subcloning into pUC119.

The 5′ and 3′ halves of the G element were cloned on 6- and 1.0-kb SalI fragments, respectively, following ligation of size-fractionated genomic DNA into the XhoI site of the λ vector ZAPII (Stratagene). A 3.0-kb BamHI fragment was subcloned from the 6-kb SalI fragment into pUC118, and a set of unidirectional deletions was generated and sequenced. The 1-kb SalI insert was subcloned as a 1-kb SalI-KpnI fragment into pUC118 and pUC119. Unidirectional deletions of these clones were generated and sequenced.

The B5 element was cloned on a 6.3-kb XhoI fragment into λ ZAPII. A 1-kb SphI fragment containing the 3′ end of the insertion was subcloned into pUC119 and sequenced following generation of unidirectional deletions.

cDNA Cloning and DNA Sequencing

Three micrograms of poly(A)$^+$ RNA from *wx-Stonor* endosperms was used with a cDNA library kit (Pharmacia) to prepare inserts for cloning into λ gt10. Two different *Wx* probes, representing sequences upstream and downstream of the insertion site, were used to screen duplicate phage lifts from the cDNA library. Seventeen clones were isolated from approximately 10^6 plated phage. For most of the clones, a BglII-SstI subfragment was subcloned into pUC119 and sequenced. For those clones missing the SstI site within the cloned insert, BglII-EcoRI or SalI fragments were subcloned and sequenced.

PCR Amplification and Cloning of *Wx*, *wx-B5*, and *wx-G* cDNAs

Three micrograms of poly(A)$^+$ RNA from 18 to 22 DAP endosperms was used as substrate for single-stranded cDNA synthesis primed by oligo(dT) (Frohman et al., 1988). Approximately 10 ng of cDNA and 70 ng of each primer were added to polymerase chain reaction buffer (16.6 mM NH_4SO_4, 67 mM Tris [pH 8.8], 6.7 mM $MgCl_2 \cdot 6H_2O$, 6.7 mM EDTA) with 200 μM each deoxynucleotide triphosphate and 2.5 units of Taq polymerase (Perkin-Elmer-Cetus). The reactions were cycled 40 times for 1 min at 95°C, 2 min at 65°C, and 2 min at 72°C and then incubated for 10 min at 72°C. Due to the high G + C content of exons 1 to 4, a 3:1 mixture of 7-deaza-2′-deoxyguanosine (c^7dGTP) to dGTP was substituted for 200 μM dGTP whenever these regions were amplified (MonConlogue et al., 1988).

For *wx-G*, the multiple DNA fragments amplified between exons 4 and 10 or exons 4 and 12 were isolated from preparative agarose gels. Fragments resulting from amplification of exons 4 to 10 were digested with BglII and SstI and cloned into BamHI-SstI digest pUC119. Twenty-four clones were isolated and sequenced. Fragments resulting from amplification of exons 4 to 12 were digested with BglII and subcloned into BamHI-SmaI–digested pUC119; 32 transformants were sequenced.

For *wx-B5*, the multiple DNA fragments amplified between exons 1 and 7 were isolated as described above, their ends filled in using the Klenow fragment, and they were cloned into the SmaI site of pUC118. Fourteen clones were subsequently sequenced.

The positions of primers are shown in Figure 4. Their sequences are as follows:

exon 1: 5′-CACTGCCAGCCAGTGAAGGG-3′
exon 4: 5′-CGCGTGTTCGTTGACCACCC-3′
exon 7: 5′-GACGAACACGACGTCCTCCC-3′
exon 10: 5′-CGCCTCCTTGTTCAGCGCCT-3′
exon 12: 5′-GAGTCCACCGGTGGACGCGCAG-3′
exon 14: 5′-CTGATCTCGCGCGTGGTGCA-3′

ACKNOWLEDGMENTS

M.J.V. and M.P. contributed equally to this study. We are grateful to Cliff Weil, Ron Damiani, Jr., Sylvestre Marilonnett, Shawn White, and Thomas Bureau for constructive criticism of the manuscript. This work was supported by National Institutes of Health Grant No. GM32528 (to S.R.W).

Received May 8, 1992; accepted May 18, 1992.

REFERENCES

Arpaia, E., Dumbrille-Ross, A., Maler, T., Neote, K., Tropak, M., Troxel, C., Stirling, J.L., Pitts, J.S., Bapat, B., Lamhonwah, A.M., Mahuran, D.J., Schuster, S.M., Clarke, J.T.R., Lowden, J.A., and Gravel, R.A. (1988). Identification of an altered splice site in Ashkenazi Tay-Sachs disease. Nature **333,** 85–86.

Bingham, P.M., and Zachar, Z. (1989). Retrotransposons and the FB transposon from *Drosophila melanogaster.* In Mobile DNA, D.E. Berg and M.M. Howe, eds (Washington, DC: American Society for Microbiology), pp. 485–502.

Boeke, J. (1989). Transposable elements in *Saccharomyces cerevisiae.* In Mobile DNA, D.E. Berg and M.M. Howe, eds (Washington, DC: American Society for Microbiology), pp. 335–374.

Boeke, J.D., and Corces, V.G. (1989). Transcription and reverse transcription of retrotransposons. Annu. Rev. Microbiol. **43,** 403–434.

Camirand, A., St. Pierre, B., Marineau, C., and Brisson, N. (1990). Occurrence of a copia-like transposable element in one of the introns of the potato starch phosphorylase gene. Mol. Gen. Genet. **224,** 33–39.

Corces, V.G., and Geyer, P.K. (1991). Interactions of retrotransposons with the host genome—The case of the gypsy element of *Drosophila*. Trends Genet. **7,** 86–90.

Dale, R.M., McClure, B.A., and Houchins, J.P. (1985). A rapid single-stranded cloning strategy for producing a sequential series of overlapping clones for use in DNA sequencing. Plasmid **13,** 31–40.

Fedoroff, N., Wessler, S., and Shure, M. (1983). Isolation of the transposable maize controlling elements *Ac* and *Ds*. Cell **35,** 235–242.

Felber, B.K., Orkin, S.H., and Hamer, D.H. (1982). Abnormal RNA splicing causes one form of a thalassemia. Cell **29,** 895–902.

Finnegan, D., and Fawcett, D. (1986). Transposable elements in *Drosophila melanogaster.* Oxf. Surv. Euk. Genes **3,** 1–62.

Freeling, M. (1984). Plant transposable elements and DNA insertions. Annu. Rev. Plant Physiol. **35,** 277–298.

Frohman, M.A., Dush, M., and Martin, G.R. (1988). Rapid production of full-length cDNAs from rare transcripts: Amplification using a single gene-specific oligonucleotide primer. Proc. Natl. Acad. Sci. USA **85,** 8998–9002.

Ghosh, M.P., Ghosh, K., Simsek, M., and RajBhandary, U.L. (1982). Nucleotide sequence of wheat germ cytoplasmic initiator methionine transfer ribonucleic acid. Nucl. Acids Res. **10,** 3241–3247.

Goodall, G.J., and Filipowicz, W. (1989). The AU-rich sequences present in the introns of plant nuclear pre-mRNAs are required for splicing. Cell **58,** 473–483.

Goodall, G.J., and Filipowicz, W. (1991). Different effects of intron nucleotide composition and secondary structure on pre-messenger-RNA splicing in monocot and dicot plants. EMBO J. **10,** 2635–2644.

Grandbastien, M.-A., Spielmann, A., and Caboche, M. (1989). Tnt1, a mobile retroviral-like transposable element of tobacco isolated by plant cell genetics. Nature **337,** 376–380.

Green, M.M. (1988). Mobile DNA elements and spontaneous gene mutation. In Eukaryotic Transposable Elements as Mutagenic Agents, M.E. Lambert, J.F. McDonald, and I.B. Weinstein, eds (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory), pp. 41–50.

Haberd, N., Flavell, R., and Thompson, R.D. (1987). Identification of a transposon-like insertion in the *Glu-1* allele of wheat. Mol. Gen. Genet. **192,** 373–377.

Johns, M.A., Mottinger, J., and Freeling, M. (1985). A low copy number, *copia*-like transposon in maize. EMBO J. **4,** 1093–1102.

Kim, H.Y., Schiefelbein, J.W., Raboy, V., Furtek, D., and Nelson, O. (1987). RNA splicing permits expression of a maize gene with a defective *Suppressor-mutator* transposable element insertion in an exon. Proc. Natl. Acad. Sci. USA **84,** 5863–5867.

Levis, R., O'Hare, K., and Rubin, G.M. (1984). Effects of transposable element insertions on RNA encoded by the *white* gene of Drosophila. Cell **38,** 471–481.

Menssen, A., Hohmann, S., Martin, W., Schnable, P.S., Peterson, P.A., Saedler, H., and Gierl, A. (1990). The *En/Spm* transposable element of *Zea mays* contains splice sites at the termini generating a novel intron from a *dSpm* element in the *A2* gene. EMBO J. **9,** 3051–3057.

MonConlogue, L., Brow, M., and Innis, M. (1988). Structure-independent DNA amplification by PCR using 7-deaza-2′-deoxyguanosine. Nucl. Acids Res. **16,** 9869.

Moore, G., Lucas, H., Batty, N., and Flavell, R. (1991). A family of retrotransposons and associated genomic variation in wheat. Genomics **10,** 461–468.

Nelson, O.E. (1968). The *waxy* locus in maize. II. Location of the controlling element alleles. Genetics **99,** 275–284.

Nelson, O.E., and Rines, H.W. (1962). The enzymatic deficiency in the *waxy* mutant of maize. Biochem. Biophys. Res. Commun. **9,** 297–300.

Pereira, A., Schwarz-Sommer, Z., Gierl, A., Bertram, I., Peterson, P., and Saedler, H. (1985). Genetic and molecular analysis of the *Enhancer* (*En*) transposable element system of *Zea mays.* EMBO J. **4,** 579–583.

Pouteau, S., Spielmann, A., Meyer, C., Grandbastien, M.A., and Caboche, M. (1991). Effects of Tnt1 tobacco retrotransposon insertion on target gene transcription. Mol. Gen. Genet. **228,** 233–239.

Raboy, V., Kim, H.-Y., Schiefelbein, J.W., and Nelson, O.E. (1989). Deletions in a *dSpm* insert in a maize *bronze-1* allele alter RNA processing and gene expression. Genetics **122,** 695–703.

Ruffner, D., and Dugaiczy, K. (1988). Splicing mutation in human hereditary analbuminemia. Proc. Natl. Acad. Sci. USA **85,** 2125–2129.

Scott, M., Weiner, A., Hazelrigg, T., Polisky, B., Pirrota, V., Scalenghe, F., and Kaufman, T. (1983). The molecular organization of the Antennapedia locus of Drosophila. Cell **35,** 763–776.

Sentry, J.W., and Smyth, D.R. (1989). An element with long terminal repeats and its variant arrangements in the genome of *Lilium henryi.* Mol. Gen. Genet. **215,** 349–354.

Shure, M., Wessler, S., and Fedoroff, N. (1983). Molecular identification and isolation of the *waxy* locus in maize. Cell **35,** 225–233.

Smith, P.A., and Corces, V.G. (1991). Drosophila transposable elements: Mechanisms of mutagenesis and interactions with the host genome. Adv. Genet. **29,** 229–300.

Smyth, D.R., Kalitsis, P., Joseph, J.L., and Sentry, J.W. (1989). Plant retrotransposon from *Lilium henryi* is related to Ty3 of yeast and the gypsy group of Drosophila. Proc. Natl. Acad. Sci. USA **86,** 5015–5019.

Treisman, R., Proudfoot, N.J., Shander, M., and Maniatis, T. (1982). A single-base change at a splice site in a β-thalassemic gene causes abnormal RNA splicing. Cell **29,** 903–911.

Vieira, J., and Messing, J. (1987). Production of single-stranded plasmid DNA. Methods Enzymol. **153,** 3–11.

Voelker, R.A., Graves, J., Gibson, W., and Eisenberg, M. (1990). Mobile element insertions causing mutations in the *Drosophila suppressor of sable* locus occur in DNAse-I hypersensitive subregions of 5′-transcribed nontranslated sequences. Genetics **126,** 1071–1082.

Voytas, D.F., and Ausubel, F.M. (1988). A *copia*-like transposable element family in *Arabidopsis thaliana.* Nature **336,** 242–244.

Voytas, D.F., Konieczny, A., Cummings, M.P., and Ausubel, F.M. (1990). The structure, distribution and evolution of the *Ta1* retrotransposable element family of *Arabidopsis thaliana.* Genetics **126,** 713–721.

Wallace, M.R., Andersen, L.B., Saulino, A.M., Gregory, P.E., Glover, T.W., and Collins, F.S. (1991). A de novo Alu insertion results in neurofibromatosis type-1. Nature **353,** 864–866.

Weil, C.F., and Wessler, S.R. (1990). The effects of plant transposable element insertion on transcription initiation and RNA processing. Annu. Rev. Plant Physiol. Plant Mol. Biol. **41,** 527–552.

Wessler, S.R. (1989). The splicing of maize transposable elements from pre-mRNA—A minireview. Gene **82,** 127–133.

Wessler, S.R. (1991a). The maize transposable *Ds1* element is alternatively spliced from exon sequences. Mol. Cell. Biol. **11,** 6192–6196.

Wessler, S.R. (1991b). Alternative splicing of a *Ds* element from exon sequences may account for two forms of Wx protein encoded by the *wx-m9* allele. Maydica **36,** 317–322.

Wessler, S.R., and Varagona, M. (1985). Molecular basis of mutations at the *waxy* locus of maize: Correlation with the fine structure genetic map. Proc. Natl. Acad. Sci. USA **82,** 4177–4181.

Wessler, S.R., Baran, G., Varagona, M., and Dellaporta, S.L. (1986). Excision of *Ds* produces *waxy* proteins with a range of enzymatic activities. EMBO J. **5,** 2427–2432.

Zachar, Z., and Bingham, P.M. (1982). Regulation of *white* locus expression: The structure of mutant alleles at the *white* locus of *Drosophila melanogaster.* Cell **30,** 529–541.