Extreme Structural Heterogeneity Among the Members of a Maize Retrotransposon Family

Sylvestre Marillonnet[1] and Susan R. Wessler

Departments of Botany and Genetics, University of Georgia, Athens, Georgia 30602

Manuscript received April 25, 1998
Accepted for publication July 29, 1998

ABSTRACT

A few families of retrotransposons characterized by the presence of long terminal repeats (LTRs) have amplified relatively recently in maize and account for >50% of the genome. Surprisingly, none of these elements have been shown to cause a single mutation. In contrast, most of the retrotransposon-induced mutations isolated in maize are caused by the insertion of elements that are present in the genome at 2–50 copies. To begin to understand what limits the amplification of this mutagenic class of LTR-retrotransposons, we are focusing on five elements previously identified among 17 mutations of the maize *waxy* gene. One of these elements, *Stonor*, has sustained a deletion of the entire *gag* region and part of the protease domain. Missing sequences were recovered from larger members of the *Stonor* family and indicate that the deletion probably occurred during retrotransposition. These large elements have an exceptionally long leader of 2 kb that includes a highly variable region of ~1 kb that has not been seen in previously characterized retrotransposons. This region serves to distinguish each member of the *Stonor* family and indicates that no single element has yet evolved that can attain the very high copy numbers characteristic of other element families in maize.

RETROTRANSPOSONS belong to a class of elements distinguished by their mode of transposition through an RNA intermediate. Two types of retrotransposons have been identified in eukaryotes: the retrotransposons characterized by the presence of long terminal repeats (LTRs) and the long interspersed nuclear elements, which do not contain LTRs. LTR-retrotransposons appear to be ubiquitously distributed in plants (Flavell *et al.* 1992; Voytas *et al.* 1992; Hirochika and Hirochika 1993), where they represent the most abundant class of transposable elements. It has been estimated that 50% of the maize genome consists of retrotransposon sequences, with five large families contributing up to 25% of the genomic DNA (SanMiguel *et al.* 1996). Not all retrotransposon families, however, are present at high copy numbers. Hundreds or perhaps thousands of families containing <100 elements may also be present in maize (SanMiguel *et al.* 1996).

Interestingly, the elements found to cause mutations in maize are those that are present in the genome at relatively low copy numbers. These include *Magellan* (four to eight copies per haploid genome, Purugganan and Wessler 1994), *Hopscotch* (two to six copies, White *et al.* 1994; S. E. White and S. R. Wessler, unpublished data), and *Bs1* (one to five copies, Johns *et al.* 1985). Members of the *B5* family, with only two to four copies, are responsible for three independent mutations (*wxB5*, *wxG*, and *bm-3*; Varagona *et al.* 1992; Vignols *et al.* 1995), while *Magellan* elements have been found in two mutant alleles (*wxM* and *pl-987*, Purugganan and Wessler 1994; P. L. Cooper and K. C. Cone, personal communication). No member of any of the larger families of maize elements, such as *Opie* (30,000 copies), or *Grande*, *Ji*, or *Huck* (each with >10,000 copies), has caused any of the characterized maize mutations, despite the fact that some appear to be largely intact and capable of further retrotransposition (SanMiguel *et al.* 1996). These data led SanMiguel *et al.* (1996) to suggest that there may be a cause-and-effect relationship between element family size and the propensity to insert into genes. That is, elements from very large families may thrive because they display a target site preference for other retrotransposons. This way, they avoid mutating maize genes. Whether this trend holds true for other plant species cannot be ascertained at this time because few spontaneous mutations outside of maize have been characterized. However, it was previously demonstrated that the retrotransposon *Tos*17, which is present at only one to two copies in the rice genome, prefers insertion into gene sequences after its induction into cell culture (Hirochika *et al.* 1995).

Just as the avoidance of gene targets may favor element amplification, an inability to avoid genes may prevent significant amplification of the smaller element families. To begin to understand this apparent correlation be-

Corresponding author: Susan Wessler, Department of Genetics, University of Georgia, Athens, GA 30602.
E-mail: sue@dogwood.botany.uga.edu

[1] *Present address:* The Sainsbury Laboratory, John Innes Centre, Norwich NR4 7UH, United Kingdom.

tween the size of plant retrotransposon families and their propensity to insert into genes, we have characterized the *Stonor* family of maize. The *Stonor* element was first identified as an insertion into the intron5/exon6 junction of the maize *wxStonor* allele (Figure 1; Varagona *et al.* 1992). In this study, we have sequenced the complete *Stonor* insertion and found that it is a defective element containing a deletion of part of its coding sequence. This observation prompted us to clone larger elements from the same family to identify sequences missing from *Stonor*. Analysis of the sequences at the deletion endpoints and in the remainder of the element indicates that *Stonor* is derived from a larger, presumably active element that sustained a deletion during retrotransposition. Characterization of several family members larger than *Stonor* showed that they differ from most other retrotransposons by the presence of an unusually long, untranslated leader region that contains a domain of highly variable sequence.

MATERIALS AND METHODS

Strains: Maize strains Black Mexican Sweet, Zapalote Chico 2840, and Zapalote Chico 217413 were obtained from the North Central Regional Plant Introduction Station (Ames, IA). M14, B37, W23, W22, and GA221 are standard inbred lines. Two hybrid lines, Q60 and B70, were obtained from Lane Arthur (University of Georgia). Teosinte strains *Zea maize mexicana* and *Z. maize parviglumis* were obtained from the National Clonal Germplasm Repository (Miami, FL).

Cloning procedures: The 5′ and 3′ halves of the *Stonor* element were cloned previously (Varagona *et al.* 1992), except for 170 bp located between the cloned fragments. A fragment including the 170 bp was isolated after PCR amplification using a primer located in *Stonor* sequences contained in the 5′ end clone (primer STB: 5′ GGGTTCACACAGAGA GAAGG 3′) and a second primer located in *wx* sequences downstream of the insertion (wxexon72: 5′ TTGAGGTAG CACGAGAGAGG 3′). PCR amplification was performed in 100-μl reactions with 200 ng genomic DNA, 70 ng primer, and 2.5 units DNA polymerase (Amplitaq; Perkin Elmer, Norwalk, CT) in PCR reaction buffer (10 mm Tris-HCl pH 8.3, 50 mm KCl, 1.5 mm $MgCl_2$, and 0.01% gelatin). The reactions were cycled 40 times for 1 min at 95°, 2 min at 65°, 3 min at 72°, and then at 72° for 10 min. The amplified fragment was subcloned in pUC119 and sequenced.

Sto-12, *Sto*-14, and *Sto*-30 were cloned after PCR amplification using primer pairs ST6 (5′ CGGATTGGTATTCTAGGGAC 3′)/STC (5′ GATGTACCACTGTCTGGAGG 3′), ST5 (5′ GGA GATCGTCAAGAAGGAGG 3′)/STC, and ST5/STC2 (5′ CCA CCTGTGGTCATAGTTGG 3′), respectively. PCR buffer and amplification conditions were the same as described above. *Sto*-3 and *Sto*-4 were cloned after PCR amplification using PCR primers ST5 and ST3 (5′ GGTCGAGCGATTAGCGTTG 3′). To amplify the large sequences corresponding to entire elements (PCR product of ~6.5 kb), the Elongase enzyme mix from Bethesda Research Laboratories (BRL, Gaithersburg, MD) was used. PCR reactions (50 μl) contained 200 ng of genomic DNA, 0.2 mm of each primer, and 0.2 mm dNTPs in PCR buffer [60 mm Tris-SO_4, pH 9.1, 18 mm $(NH_4)_2SO_4$, and 1.5 mm $MgSO_4$]. Reactions were heated to 94° for 30 sec, cycled 35 times for 30 sec at 94°, 30 sec at 62°, and 7 min at 68°, and then incubated at 68° for 10 min. PCR products were cloned in the pGEM-T vector (Promega, Madison, WI).

Sto-1, *Stl*-2, *Stl*-3, *Stl*-4, *Stl*-5, *Stl*-8, *Stl*-10, *Stl*-12, *Stl*-14, *Stl*-18, *Stl*-19, *Stl*-20, *Stl*-21, *Stl*-23, *Stl*-24, *Stl*-31, *Stl*-33, *Stl*-35, *Stl*-36, *Stl*-37, *Stl*-39, and *Stl*-40 were obtained after PCR amplification from maize genomic DNA using primers ST5 and STC2 (buffer and amplification conditions were identical as described for *Sto*-30), and they were cloned in the TA cloning vector (Invitrogen, San Diego, CA).

A genomic library was constructed by cloning maize genomic DNA digested with *Bam*HI into the *Xho*I site of lambda ZAPII (Stratagene, La Jolla, CA). Insert and vector cohesive ends were rendered compatible by partial filling of the ends of the *Bam*HI-digested DNA fragments with dGTP and dATP, and the ends of the *Xho*I-digested vector with dTTP and dCTP. *Sto*-17 was obtained by screening this library with a 556-bp *Sst*I-*Pst*I fragment derived from the internal domain of *Sto*-14.

The sequences determined from these clones have been deposited into the GenBank database under the accession numbers AF082127–AF082155.

Plasmid construction and probe synthesis: Plasmid pSto-wx was made by subcloning a 509-bp *Sal*I-*Nsi*I fragment from the *wx* gene and a 556-bp *Pst*I-*Sst*-I fragment from *Sto*-14 into pUC119. A 1.05-kb fragment containing *wx* and *Sto*-14 sequences was amplified from pSto-wx using the m13-puc forward and reverse primers, and it was labeled using the Random Primers DNA Labelling System from BRL to produce probe PrSto-wx.

Sequence analysis and phylogenetic analysis: Multiple alignments were made using the program PILEUP of the University of Wisconsin Genetics Computer Group (GCG) and displayed using the GCG program BOXSHADE. An alignment of inferred amino acid sequences corresponding to the reverse transcriptase and part of the RNaseH domains of *Stonor* [sequences located between nucleotide (nt) positions 2407 and 3630], *Tnt1* (GenBank accession no. X13777), *Tto1* (D83003), *Ta1-3* (X13291), *BARE*-1 (Z17327), *Osser* (X69552), *copia* (X02599), *PREM-2* (ZMU41000), *Hopscotch* (U12626), *Tst1* (X52387), and *RIRE1* (D85597) was made using PILEUP. A phylogenetic tree was constructed using these sequences and the programs PROTPARS and SEQBOOT from the PHYLIP 3.5 package (using maximum parsimony analysis). A consensus tree was obtained after running 100 bootstraps.

DNA gel blot analysis: Hybridizations and washes (with 0.1× SSC and 0.5% SDS) were performed at 65°.

RESULTS

The *wxStonor* allele contains a *copia*-like retrotransposon: The 4542-bp insertion in the *wxStonor* allele was cloned and sequenced (see materials and methods) and found to have structural features characteristic of retrotransposons, including the following: (i) a perfect LTR of 560 bp flanking a 3422-bp internal domain, (ii) an 18-nt primer-binding site (PBS) adjacent to the 5′ LTR and homologous to the 3′ end of the wheat Met-tRNA (Ghosh *et al.* 1982), and (iii) a polypurine tract (PPT) upstream of the 3′ LTR (Figure 1). Internal sequences encode a putative 998-amino-acid open reading frame (ORF) that contains regions with similarity to several *copia*-like retrotransposons from Drosophila (*copia*, *1731*) and plants (*Tnt*1, *Tto*1, *Ta*1-3, *Tst*1, *Hopscotch*, and *BARE*-1). The most extensive identity between *Stonor* and other retrotransposons includes the

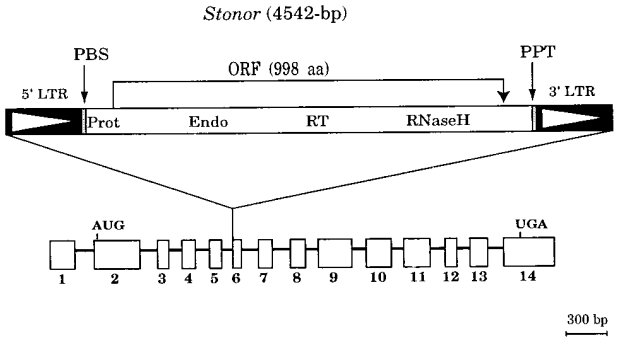


Figure 1.—Structure of the *wxStonor* allele, including the *Stonor* element. Exons (open boxes) and introns (connecting lines) of the *wx* gene are shown with the start and stop codons of translation. *Stonor* is inserted into the splice acceptor site of intron 5. The LTR of *Stonor* is represented by black boxes containing white triangles. The internal domain extends from the PBS to the PPT, and it includes the protease (Prot), endonuclease (Endo), reverse transcriptase (RT), and RNaseH domains.

endonuclease, reverse transcriptase, and RNaseH domains (Figure 2A). However, *Stonor* appears to be a defective element that is missing the entire *gag* domain (normally found between the 5′ LTR and the *pol* domain) and part of the protease domain (Figure 2B).

***Stonor* is part of a multigene family in maize, teosinte, and Tripsacum:** Southern blot analysis was undertaken to identify additional *Stonor* elements in the maize genome (including complete elements) and to ascertain the species distribution of this retrotransposon. To this end, genomic DNAs were digested with *Eco*RI (which does not cut within available *Stonor* sequences), and the blot was probed with the *Stonor* LTR. Using high-stringency washes (see materials and methods), many sequences related to *Stonor* were detected in maize, teosinte, and Tripsacum, but not in the genomes of rice, millet, soybean, and oat (Figure 3).

Identifying larger family members in the maize genome: To determine whether elements larger than *Stonor* are present in the maize genome and, thereby, identify the sequences missing in *Stonor*, PCR was performed using primers located on either side of the putative deletion endpoints (primers ST5 and STC, Figure 4E). Genomic DNA from several maize strains and from teosinte was used in conjunction with a protocol designed to amplify products of ≤10 kb (Elongase; BRL). PCR products were analyzed by Southern blot with a probe derived from the LTR. A fragment corresponding to the size expected for the *Stonor* element (2338 bp) was amplified from a strain containing the *wxStonor* allele (Figure 4A, lane 1), but not from any other maize strain. This observation suggests that the deletion in *Stonor* is not present in any other element from this family and may have been created during transposition of an active family member into the *wx* gene (see below).

The only product amplified from all strains tested

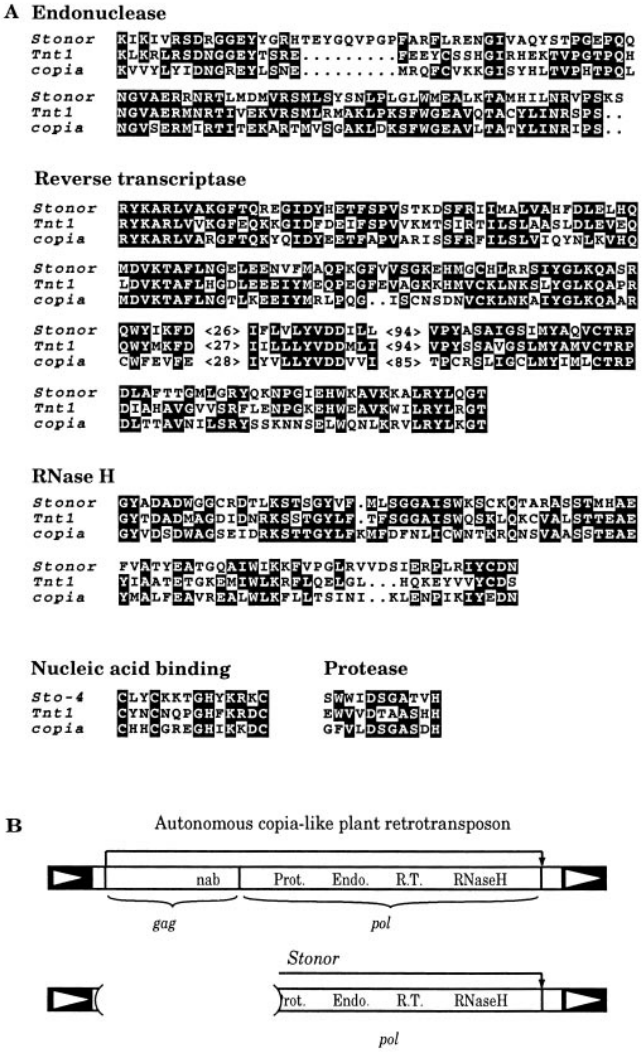


Figure 2.—Amino acid sequences shared by *Stonor* (or *Sto*-4), *Tnt1*, and *copia*. (A) Alignment of the endonuclease, reverse transcriptase, RNaseH, and nucleic acid-binding domains. (B) Components of an autonomous retrotransposon. Plant retrotransposons contain a *gag* domain, which includes a nucleic acid-binding domain, located between the 5′ LTR and *pol*. The gag and pol polyproteins are encoded in a single ORF (represented by the bent arrow). In *Stonor*, the sequences corresponding to the *gag* domain and part of the protease are deleted. Other features are as described in Figure 1.

was also the largest, at ∼5 kb (Figure 4A). Given the size of this fragment, the position of the primer within the LTR, and the size of the *Stonor* element, we estimate that a complete element should be ∼7.2 kb. To confirm the size and presence of this element in the genomes of maize and teosinte, PCR was performed using a second pair of primers located in the 5′ and 3′ LTRs (primers ST5 and ST3, Figure 4E). The positions of these primers (with the ST5 sequence downstream of the ST3 sequence) were chosen to preclude amplification of a product from a single LTR. Consistent with the existence of an element of 7.2 kb was the amplification in most strains of a product of 6.5 kb (Figure 4B). A 6.5-

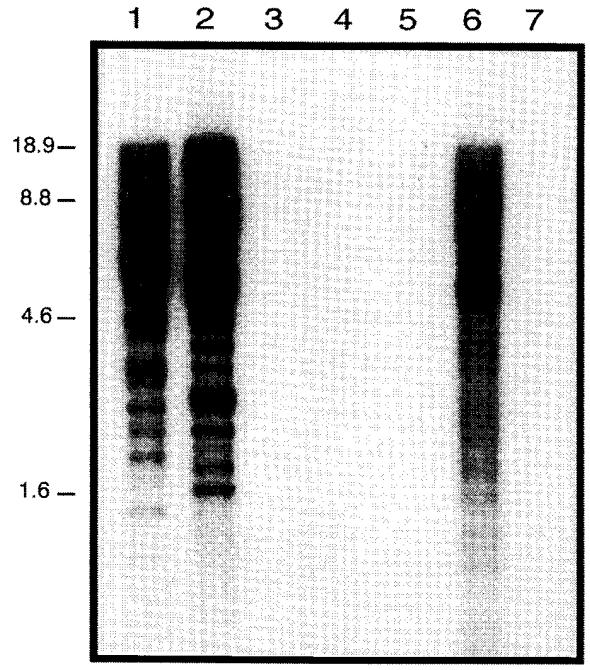


Figure 3.—DNA gel blot analysis of *Stonor* sequences in monocotyledonous and dicotyledonous plants. *Eco*RI-digested genomic DNA (10 μg) of teosinte (lane 1), maize (lane 2), rice (lane 3), millet (lane 4), soybean (lane 5), *Tripsacum dactyloides* (lane 6), and oat (lane 7) were separated by gel electrophoresis, transferred to nylon membrane, and hybridized with a probe derived from *Stonor* LTR sequences. The sizes of molecular weight standards are given in kilobases on the left.

kb fragment is not visible among the products from a strain containing the *wxStonor* allele. However, because of the efficient amplification of the PCR product corresponding to *Stonor*, a fivefold dilution of this reaction was loaded on the gel. More efficient amplification of the *Stonor* element may reflect its small size and/or the perfect match of primer sequences.

Unlike the 5′ half of *Stonor*, the 3′ half seems to be intact and representative of the other elements in the maize genome. Use of primers STB and ST3 (Figure 4E) resulted in the amplification of the expected 1.7-kb product (Figure 4C). Similarly, use of a more 5′ primer, STA (obtained from the *Sto*-4 element, Figure 5A), with ST3 also resulted in a single product of 3.8 kb (Figure 4, D and E). Taken together, these data suggest that the structural variation of family members is largely restricted to the region just downstream of the 5′ LTR.

Structural characterization of *Stonor*-like sequences: Both PCR products and fragments from genomic libraries were analyzed to define the *Stonor* deletion and to identify the putative 7.2-kb, full-length element(s) predicted to be in all maize strains analyzed. The structure of seven reconstructed elements and the *Stonor* element are shown in Figure 5A. The shaded regions of all elements, except *Sto*-17 at the top and *Stonor* at the bottom, were obtained as PCR fragments using the primer pairs shown. *Sto*-17, like *Stonor*, was isolated from a genomic library. Unshaded regions represent our reconstruction of what the rest of the element may look like in the genome, based on a comparison with the other elements and the PCR results summarized in Figure 4.

Based on a LTR length of 560 bp for *Stonor* (Varagona *et al.* 1992), complete *Sto*-3 and *Sto*-4 elements are estimated to be 7.1 and 7.2 kb, respectively. Estimates for *Sto*-17, *Sto*-30, *Sto*-1, *Sto*-12, and *Sto*-14 were based on the sizes of the cloned fragments because we had previously determined that the 3′ half of most family members was intact (Figure 4, C and D). *Sto*-4 encodes a single intact ORF of 1406 amino acids that is located 1668 bp downstream of the 5′ LTR (Figure 5A). This ORF contains all the sequence features present in active elements, including the nucleic acid-binding domain (part of *gag*, Figure 2B) and a complete *pol* domain. In contrast, both *Sto*-17 and *Sto*-3 contain a frameshift in the sequenced part of their respective ORFs (Figure 5A), suggesting that they are defective elements. The PBS of all elements, except *Stonor*, is 11 nt. The *Stonor* PBS has been extended to 18 nt by the deletion (see below).

Surprisingly, restriction sites in one region of several independently derived clones were completely different (variable region, Figure 5A). Sequence analysis of *Sto*-17, *Sto*-3, *Sto*-4, *Sto*-30, and *Sto*-1 revealed a highly variable region that includes both unique sequences (represented by G1–G6, Figure 5B) and sequences held in common among different elements (represented by A–I, Figure 5B). More detailed characterization of the variable region is described below.

Structure of the deletion derivatives: The sequences at deletion endpoints can be informative in deducing the mechanism of deletion formation. For the *Stonor* element, these endpoints are located in the PBS and the ORF, two regions that are virtually identical in all the larger family members (Figures 5B and 6A). For this reason, it was a straightforward task to compare *Stonor* with the larger elements and identify where the sequences diverged. These regions are indicated by the juxtaposition of uppercase, bold letters and lowercase letters in the large element (Figure 6A, top). For *Stonor*, the sequences downstream of the PBS, beginning with the trinucleotide GTT, are homologous with the protease domain. In the large elements, these sequences are located 2.6 kb downstream of the 5′ LTR. *Stonor* also contains 4 bp, CCAG, that cannot be aligned at either the 5′ or 3′ deletion breakpoints. DNA inserted between deletion endpoints is called filler DNA and is frequently encountered in maize deletions, where it is believed to result from double-slip mispairing during DNA replication (Wessler *et al.* 1990). According to this mechanism, filler DNA is copied from sequences flanking the deletion. However, it is unlikely that such a mechanism was involved in the generation of the deletion in *Stonor*

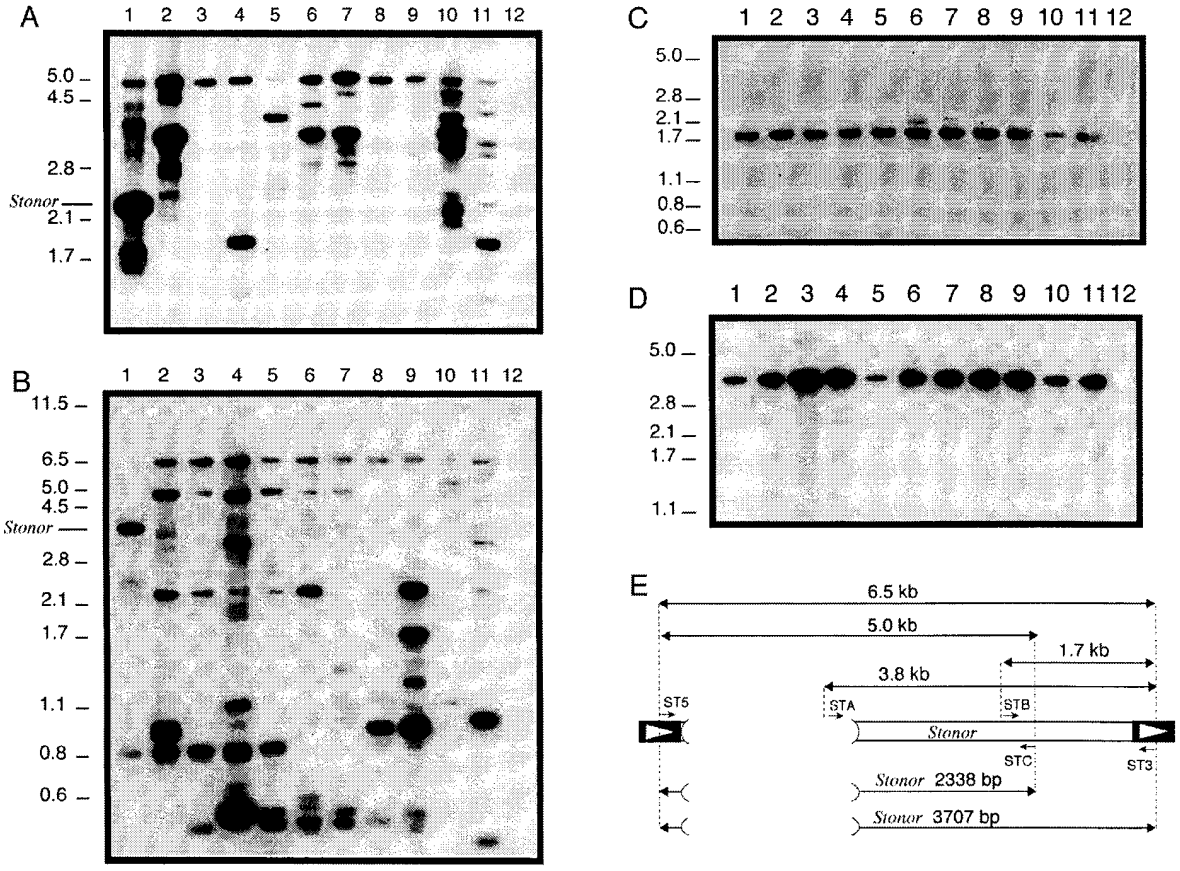


Figure 4.—DNA gel blot analysis of PCR products amplified from maize genomic DNA using primers located in LTR and internal element sequences. Blots were hybridized with a probe derived from the *Stonor* LTR. (A) Products obtained by PCR amplification using primers ST5 and STC (see E) with DNA from the following: a maize strain containing the *wxStonor* allele (lane 1), Q60 (lane 2), B70 (lane 3), Black Mexican Sweet (lane 4), M14 (lane 5), Zapalote Chico 217413 (lane 6), Zapalote Chico 2840 (lane 7), B37 (lane 8), w22 (lane 9), *Z. maize mexicana* (lane 10), *Z. maize parviglumis* (lane 11), and no DNA (lane 12). (B–D) Same as in A, except that primer pairs ST5/ST3, STB/ST3, and STA/ST3, respectively, were used for amplification. (E) Structure of some of the PCR products obtained by amplification using primer pairs ST5/STC, ST5/ST3, STB/ST3, and STA/ST3.

because the sequences of the filler DNA and of the downstream trinucleotide GTT increase the length of sequences complementary to the Met-tRNA from 11 to 18 nt (Figure 6A). Instead, this structure suggests that the deletion in *Stonor* was generated during retrotransposition (see discussion).

Like *Stonor*, both *Sto*-12 and *Sto*-14 have sustained deletions with one breakpoint downstream of the 5′ LTR and the other in the ORF (Figure 5, A and B). However, the 5′ breakpoints for *Sto*-12 and *Sto*-14 were not present in all the larger elements. For *Sto*-14, the deletion breakpoints were determined by comparison with *Sto*-1, which shares the AC motif, and were found to coincide with a 6-bp direct repeat (Figure 6B). For *Sto*-12, the 5′ breakpoint was identified by comparison with *Sto*-1, which is the only other cloned family member containing the sequences present downstream of region F (Figure 5B, region G6). The 3′ breakpoint of *Sto*-12 occurs in sequences that were not cloned in *Sto*-1, but that were cloned and sequenced in several of the other elements, including *Sto*-4, *Stonor*, and *Sto*-17 (Figure 6C). Based on such a comparison, the *Sto*-12 deletion breakpoints were also found to lie within a direct repeat, this one of 3 bp (Figure 6C).

Copy number of elements in the *Stonor* family: The size of the *Stonor* family was estimated by analyzing maize genomic DNA on Southern blots using a probe that can hybridize both to element sequences and to an unrelated single copy sequence. Upon restriction with an enzyme that cuts twice in a conserved region of the internal domain, most elements should give rise to a fragment of the same size, while the unrelated single copy sequence should give rise to a fragment of a different size. Estimation of the copy number is determined by comparing the intensity of the hybridization signals.

Probe PrSto-wx contains 556 bp from the conserved region of the internal domain of one of the members of the *Stonor* family (*Sto*-14, see materials and methods). This was then fused with a fragment of approximately the same size (509 bp) from the *wx* gene. When hybridized with a Southern blot of maize genomic DNA digested with *Eco*RI and *Sal*I, a major band of 1.1 kb is

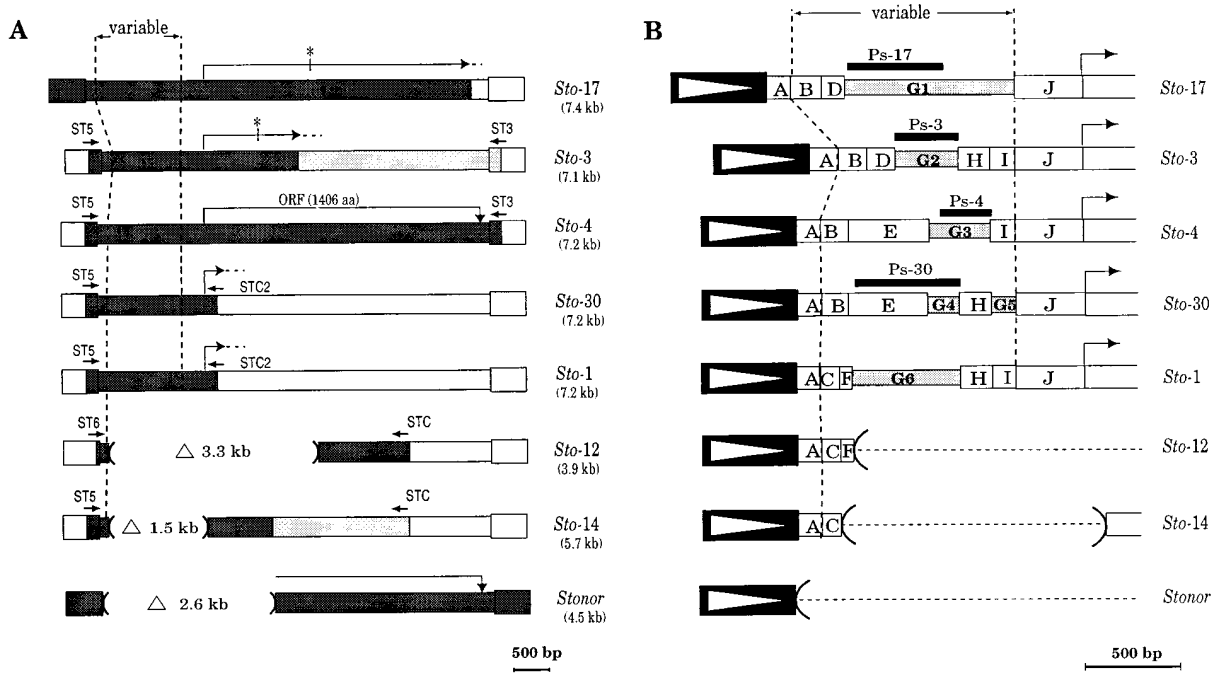


Figure 5.—Structure of the cloned family members. (A) Black areas correspond to parts of the elements that were cloned and sequenced, while gray areas represent regions cloned but not sequenced. The rectangles at the termini of each element represent the LTRs. Small arrows indicate the location of the primers used for PCR amplification. Large, bent arrows represent the extent of the ORF in the sequenced part of the elements. Asterisks indicate the location of frameshifts in the large ORFs. (B) Structure of the 5′ region of the cloned elements. The LTRs (shown as black rectangles with a white triangle) are highly conserved (93–100% identity), as is the 140-bp region A (92–99% identity). Sequences downstream of region A are highly variable among the cloned elements and contain segments that can either be aligned with some of the other elements (B–F, H, and I) or regions that cannot be aligned with any other cloned sequence (G1–G6). Sequences located downstream of the variable region and upstream of the putative ORF (region J) are highly conserved (97–99.9% identity), as are downstream sequences in the ORF.

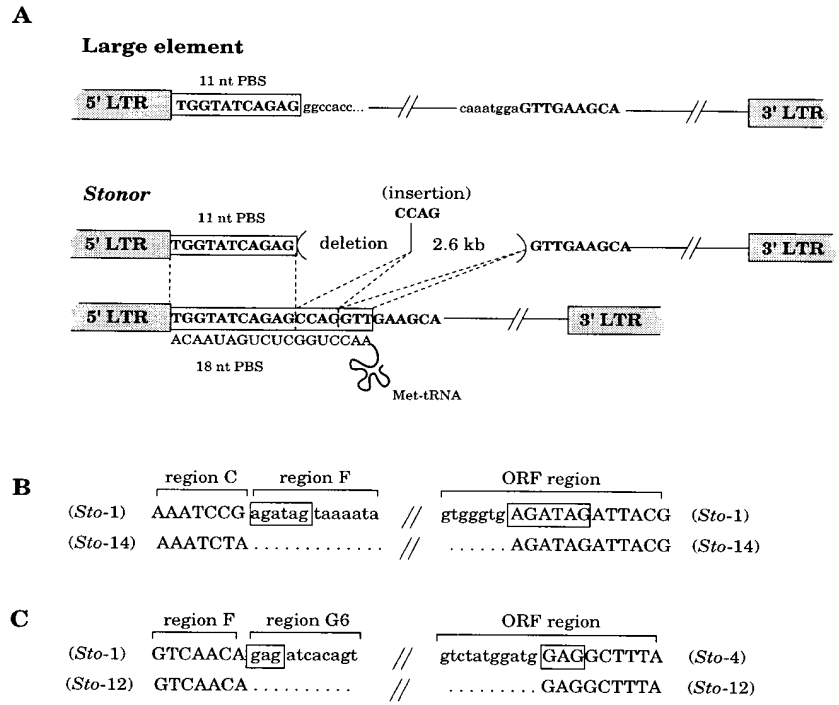


Figure 6.—Structure of the deletions in *Stonor*, *Sto*-14, and *Sto*-12. Sequences shown in lowercase letters are present in the large elements but absent in the deletion derivatives. (A) Structure of the deletion in *Stonor*. Boxed sequences are complementary to the Met-tRNA. Eighteen nucleotides of tRNA sequence complementary to the PBS of *Stonor* are shown. (B and C) The deletions in *Sto*-14 and *Sto*-12. Direct repeats at the deletion endpoints are boxed.

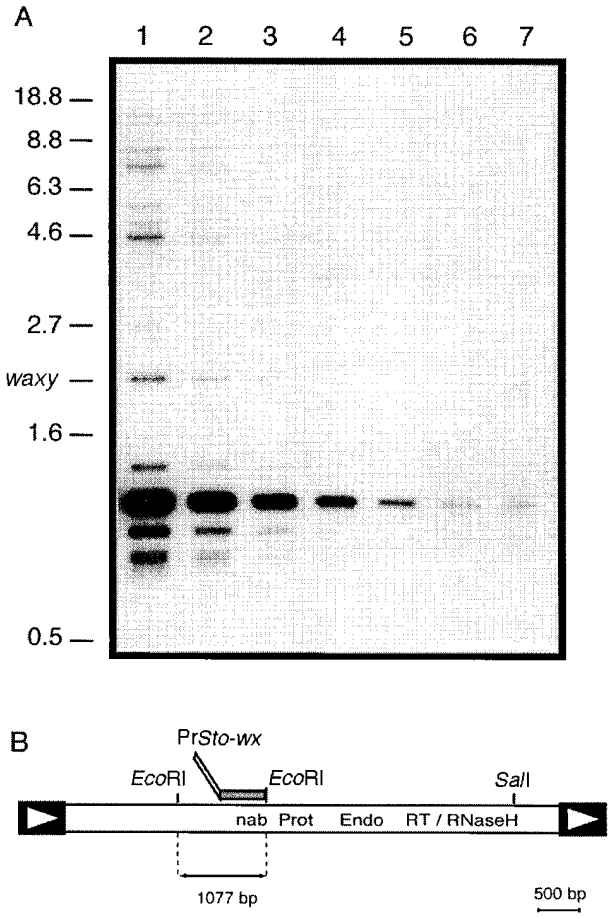


Figure 7.—DNA gel blot determination of the copy number of elements in the *Stonor* family. (A) Ten micrograms of maize genomic DNA (from inbred GA221) digested with *Eco*RI and *Sal*I were loaded in lane 1. Lanes 2–7 each contain half the amount of DNA loaded in the previous lane. The sizes of molecular weight standards are given in kilobases on the left. (B) The annealing of probe PrSto-wx to the *Stonor* family members. Annealing of the *wx* sequences present in this probe is not shown.

seen in addition to several minor bands (Figure 7), one of which (2 kb) corresponds to the *wx* gene. Quantification of the hybridization signals with a PhosphorImager indicates a copy number of 20–25 elements for the 1.1-kb *Eco*RI fragment. At least 12 other sequences, which probably represent elements that lack an *Eco*RI site or that have undergone insertions or deletions, also hybridize to probe PrSto-wx. When the elements corresponding to these bands are included, at least 32 elements are estimated to be present in the maize genome. A similar experiment using a probe from the reverse transcriptase domain indicated a copy number of 30–40 elements (data not shown).

Are the large cloned elements representative of the *Stonor* family? Analysis of PCR products (Figure 4) indicates that the *Stonor* family is comprised of large elements of ~7.2 kb with variable sequences downstream of the 5′ LTR. In addition, smaller members have sustained deletions of sequences downstream of the 5′ LTR. However, PCR products may not accurately reflect the composition of the *Stonor* family because PCR can preferentially amplify small elements or those that contain sequences more similar to the primer pairs. For these reasons, Southern blots were used to obtain an independent assessment of family composition. Ideally, this is accomplished by digesting genomic DNA with an enzyme that recognizes sites in the LTRs and by determining the size of the resulting fragments on Southern blots with a probe derived from the internal domain of the element. Unfortunately, no enzyme could be found with restriction sites only in the LTRs. Instead, we used *Asp*I, which cuts once in each LTR and twice in conserved regions of the internal domain (Figure 8C). Southern blots were then hybridized with probes homologous to each resulting fragment (probes Pr1–Pr3) from regions conserved in the four cloned large elements. The strongest hybridization signals obtained with these probes identified fragments of ~2.1 kb, 430 bp, and 4 kb, consistent with the presence of abundant elements of ~7.2 kb (Figure 8A).

Representation of the variable regions in genomic DNA: Each of the large elements contains sequences that cannot be aligned with sequences from any of the other cloned elements. To determine whether these sequences are unique or repetitive in the maize genome, they were used as probes to analyze Southern blots of maize genomic DNA (Figure 8B). Probes Ps-3, Ps-4, and Ps-17 were derived from the variable regions of *Sto*-3, *Sto*-4, and *Sto*-17, respectively (Figures 5B and 8C). Probe Ps-30 contains sequences that are shared by *Sto*-30 and *Sto*-4 (Figure 5B, region E), as well as other sequences that are unique to *Sto*-30 (Figure 5B, region G4).

Maize genomic DNA was digested with either *Asp*I or *Eco*RI; the latter should digest once in genomic sequences upstream of the element and once in element sequences downstream of the four probes. Digestion with *Asp*I should give rise to 2.1 to 2.3-kb fragments if all sequences hybridizing to the probe are part of elements with the same *Asp*I sites as in their respective cloned elements (Figure 8C). Ps-30, Ps-3, Ps-4, and Ps-17 appear to hybridize to a subset of the sequences identified by the conserved Pr-1 probe (Figure 8B). Included among these bands are the predicted 2.1- to 2.3-kb fragments in addition to larger fragments that may represent elements that have undergone mutations of the *Asp*I sites. Alternatively, these sequences may not reside in *Stonor* elements. All but one of the *Eco*RI fragments are >2.1–2.3 kb, as expected for an enzyme with sites in flanking upstream sequences. The one exception is a fragment of ~1.5 kb detected by Ps-30, a probe that also contains sequences held in common by another cloned element (Figure 5B, region E).

PCR amplification and sequence analysis of additional variable regions: Maize genomic DNA from four strains was used in conjunction with the primer pair

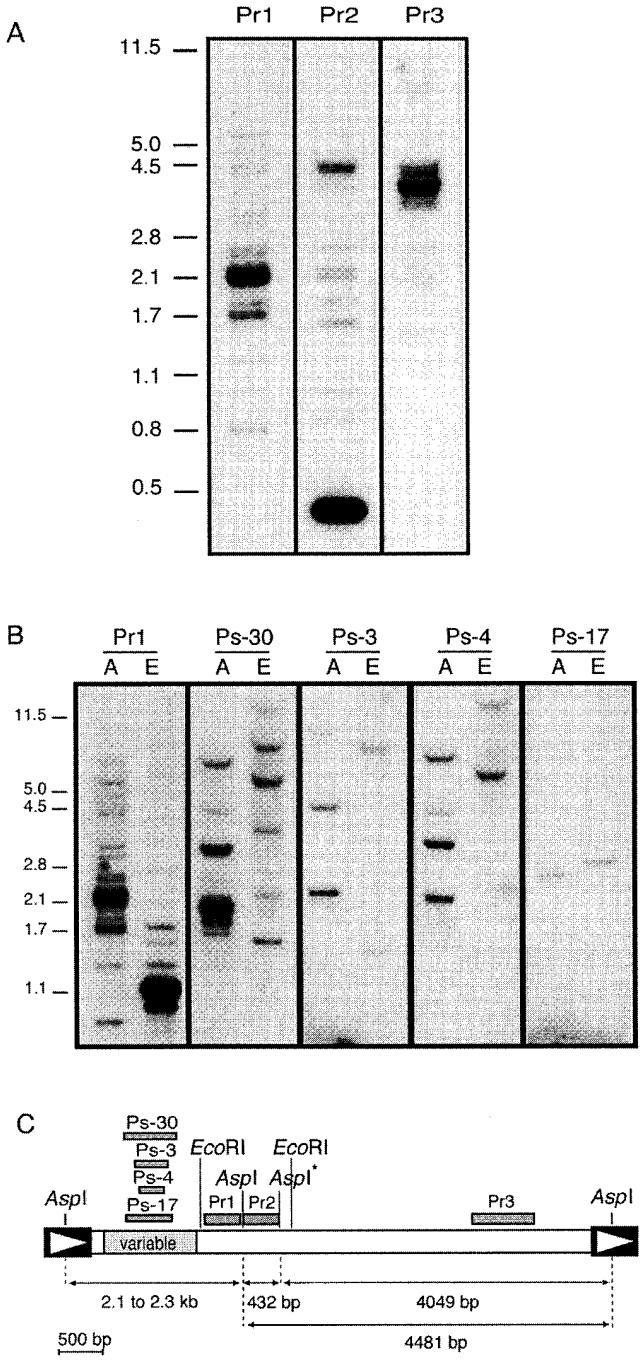
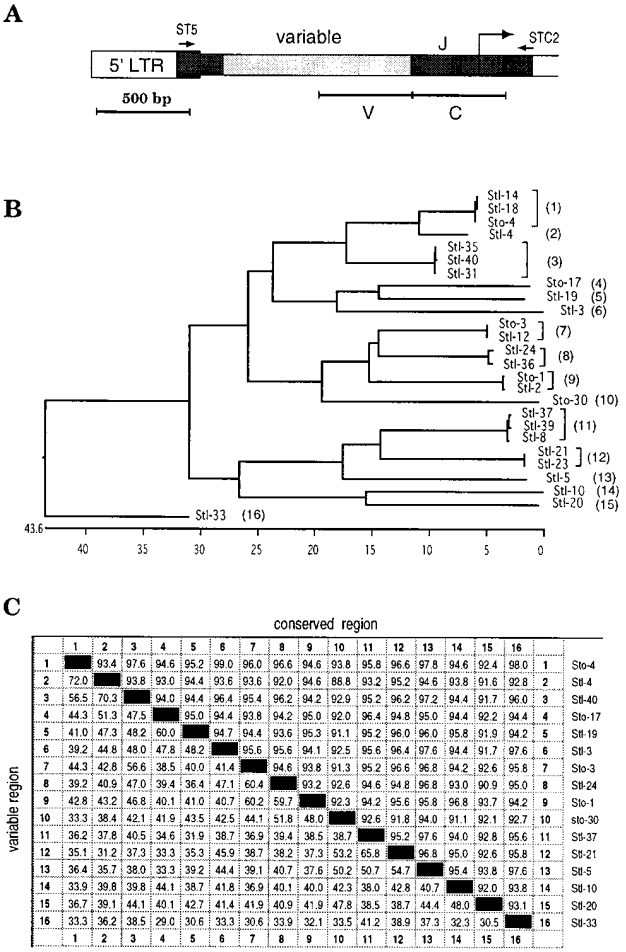


Figure 8.—DNA gel blot analysis of *Stonor* family members in the maize genome. (A) Ten micrograms of maize genomic DNA (from inbred W22) digested with *Asp*I was separated by gel electrophoresis, transferred to a nylon membrane, and hybridized with probes Pr1 (lane 1), Pr2 (lane 2), and Pr3 (lane 3). (B) Ten micrograms of maize genomic DNA was digested with either *Asp*I (noted A) or *Eco*RI (noted E), and was probed with Ps-30, Ps-3, Ps-4, and Ps-17 (see Figure 5B for the position of these probes relative to regions G1–G4 and E). (C) Positions of the probes within *Stonor* family members. The *Asp*I site noted with an asterisk may not be present in all family members.



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	
1		93.4	97.6	94.6	95.2	99.0	96.0	96.6	94.6	93.8	95.8	96.6	97.8	94.6	92.4	98.0	1 Sto-4
2	72.0		93.8	93.0	94.4	93.6	93.6	92.0	94.6	88.8	93.2	95.2	94.6	93.8	91.6	92.8	2 Stl-4
3	56.5	70.3		94.0	94.4	96.4	95.4	96.2	94.2	92.9	95.2	96.2	97.2	94.4	91.7	96.0	3 Stl-40
4	44.3	51.3	47.5		95.0	94.4	93.8	94.2	95.0	92.0	96.4	94.8	95.0	94.4	92.2	94.4	4 Sto-17
5	41.0	47.3	48.2	60.0		94.7	94.4	93.6	95.3	91.1	95.2	96.0	96.0	95.8	91.9	94.2	5 Stl-19
6	39.2	44.8	48.0	47.8	48.2		95.6	95.6	94.1	92.5	95.6	96.4	97.6	94.4	91.7	97.6	6 Stl-3
7	44.3	42.8	56.6	38.5	40.0	41.4		94.6	93.8	91.3	95.2	96.6	96.8	94.2	92.6	95.8	7 Sto-3
8	39.2	40.9	47.0	39.4	36.4	47.1	60.4		93.2	92.6	94.6	94.8	96.8	93.0	90.9	95.0	8 Stl-24
9	42.8	43.2	46.8	40.1	41.0	40.7	60.2	59.7		92.3	94.2	95.6	95.8	96.8	93.7	94.2	9 Sto-1
10	33.3	38.4	42.1	41.9	43.5	42.5	44.1	51.8	48.0		92.6	91.8	94.0	91.1	92.1	92.7	10 sto-30
11	36.2	37.8	40.5	34.6	31.9	38.7	36.9	39.4	38.5	38.7		95.2	97.6	94.0	92.8	95.6	11 Stl-37
12	35.1	31.2	37.3	33.3	35.3	45.9	38.7	38.2	37.3	53.2	65.8		96.8	95.0	92.6	95.8	12 Stl-21
13	36.4	35.7	38.0	33.3	39.2	44.4	39.1	40.7	37.6	50.2	50.7	54.7		95.4	93.8	97.6	13 Stl-5
14	33.9	39.8	39.8	44.1	38.7	41.8	36.9	40.1	40.0	42.3	38.0	42.8	40.7		92.0	93.8	14 Stl-10
15	36.7	39.1	44.1	40.1	42.7	41.4	41.9	40.9	41.9	47.8	38.5	38.7	44.4	48.0		93.1	15 Stl-20
16	33.3	36.2	38.5	29.0	30.6	33.3	30.6	33.9	32.1	33.5	41.2	38.9	37.3	32.3	30.5		16 Stl-33
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	

Figure 9.—Sequence variability in the leader region of *Stonor* family members. (A) The regions amplified with primers ST5 and STC2. Cloned sequences are shown in dark gray (conserved) or light gray (variable). The sequenced regions of the cloned PCR products are underlined and labeled for variable (V) or conserved (C). (B) Phylogenetic tree of all cloned *Stonor* family members based on an alignment of sequences from segment V of the variable region. Multiple alignments were made with the program DNASTAR using the Clustal method with a gap penalty of 8 and a gap length penalty of 8. The percent divergence between elements is indicated by the horizontal axis. Clones exhibiting <0.9% divergence were grouped together. Each group number is indicated in parentheses. (C) Similarity matrix of 16 *Stonor* family members calculated for sequence segments V and C from the alignment described above. Numbers below the diagonal represent similarity between clones for the variable region V, while numbers above the diagonal represent similarity in the conserved region C.

ST5 and STC2 to amplify variable regions that resided in members of the *Stonor* family (Figure 9A). Twenty-one independent clones containing amplified products were isolated from the following strains: B70, *Stl*-2, *Stl*-3, *Stl*-4, and *Stl*-5; BMS, *Stl*-12, *Stl*-14, *Stl*-18, *Stl*-19, and *Stl*-20; M14, *Stl*-21, *Stl*-23, and *Stl*-24; and W22, *Stl*-31, *Stl*-33, *Stl*-35, *Stl*-36, *Stl*-37, *Stl*-39, and *Stl*-40. For each clone, 1 kb of sequence was obtained from an overlapping region of variable and conserved sequences (Figure 9A, segment V includes 500 bp of variable region, and segment C includes region J and 100 bp of the ORF).

Multiple sequence alignments were performed for these regions of the 21 new sequences and the five

cloned elements containing variable domains. These 26 sequences fell into 16 distinct groups (Figure 9B). For 7 of these groups, several clones were found to be almost identical, with the exception of a few nucleotides. It is not known whether these sequence differences result from PCR errors that follow amplification of identical elements, or whether there are multiple copies of virtually identical elements in a single genome. For example, clones such as *Sto*-3 and Stl-12 or Stl-8 and Stl-37 were very similar (2- and 8-nt differences out of 1 kb, respectively) even though they were amplified from different maize strains. In contrast, most clones from different groups were found to be extremely divergent, with <60% identity (Figure 9C). For almost all sequence comparisons performed between elements of different groups, at least part of the variable region was so divergent that the sequences could not be aligned properly. In contrast to the variable region, all sequences in the conserved region were >90% conserved (Figure 9C).

DISCUSSION

The *wxStonor* allele was shown previously to contain a retrotransposon insertion in the maize *wx* gene (Varagona *et al.* 1992). The sequence of this insertion, now called the *Stonor* element, has been determined and found to contain a deletion that includes all of *gag* and a portion of the protease domain. This deletion could have arisen during or after element insertion into *wx*. As such, this element is probably no longer capable of further retrotransposition. Alternatively, retrotransposition of the defective *Stonor* element may have been mediated by the products of an autonomous element located elsewhere in the maize genome.

The *Stonor* deletion: Several lines of evidence suggest that the *Stonor* deletion was sustained during retrotransposition and probably just before insertion into the *wx* gene; *i.e.*, the *Stonor* element was recently active but is no longer capable of retrotransposition. First, what remains of the large ORF is still intact. This is expected if the *Stonor* element was recently active, but not if its movement was complemented *in trans* by the products of another family member. Second, we show that the *Stonor* element is restricted to strains carrying the *wxStonor* allele. This is consistent with the behavior of an element rendered defective upon insertion. An element that can be complemented *in trans* would have predated the *wxStonor* mutation and might be expected to be in other strains.

The strongest evidence for the generation of the *Stonor* deletion during retrotransposition is that it is reminiscent of similar structures formed during retrovirus retrotransposition. The relevant steps in both retrovirus and retrotransposon transposition are summarized in Figure 10A. According to Pulsinelli and Temin (1991), in some instances, DNA synthesis may continue past the usual site in the tRNA, resulting in the addition of tRNA-derived nucleotides at the 3′ end (Figure 10B, 6.2).

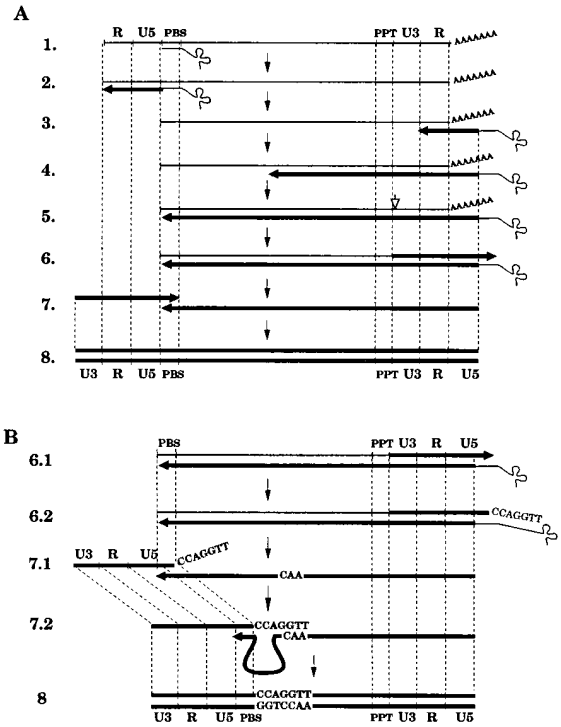


Figure 10.—Model for the generation of the deletion in the *Stonor* element. Thin lines are RNA, and thick lines are DNA. (A) Steps involved in normal reverse transcription of LTR-retrotransposons or retroviruses. U3, R, and U5 comprise the LTR. (1) Annealing of a Met-tRNA molecule (the cloverleaf structure) to the element-encoded RNA. (2) The Met-tRNA serves as a primer for synthesis of a DNA fragment corresponding to the R and U5 regions (called minus-strand, strong-stop DNA). (3) Degradation of the R and U5 regions of the template with RNaseH and transfer of the minus-strand, strong-stop DNA to the other end of the RNA template via homology to the R repeat. (4) Minus-strand DNA synthesis continues. (5) RNaseH nicks the RNA template at the PPT (location of the nick shown as a white arrow). (6) After removal of RNA downstream of PPT with RNaseH, plus-strand DNA synthesis initiates at PPT and stops in the Met-tRNA after synthesis of the plus-strand, strong-stop DNA (containing U3, R, U5, and PBS). (7) Removal of RNA by RNaseH followed by plus-strand strong-stop DNA transfer. (8) Completion of plus-strand and minus-strand DNA syntheses by reverse transcriptase. (B) Model for the generation of the deletion in *Stonor*. Steps 1–5 are not shown because they occur as described in A. (6.1) Plus-strand DNA synthesis proceeds through the point where it normally stops and (6.2) copies an additional 7 nt from the tRNA template. (7.1) Plus-strand DNA transfer. (7.2) Annealing of the extra nucleotides to downstream sequences. (8) Completion of DNA synthesis.

After strand transfer, these extra nucleotides cannot hybridize with the existing PBS and may "seek out" complementary sequences downstream (Figure 10B, 7.1–7.2). DNA synthesis then continues, resulting in the structure observed in *Stonor*.

Generation of the *Stonor* deletion by such a mechanism implies that the primer tRNA sequence was inherited during retrotransposition of *Stonor* into the *wx* gene. Such an event was thought to be associated with the propagation of full-length genomic copies of both ret-

roviruses and retrotransposons until Lauermann and Boeke (1994) demonstrated that the primer tRNA sequence was not inherited during retrotransposition of the yeast *Ty*1 element. Because *Stonor* is obviously a defective element, retention of the tRNA sequence may reflect the aberrant outcome of an unusual event. Alternatively, inheritance of tRNA sequences during reverse transcription may be a feature that distinguishes retrotransposition of the *Stonor* element, and perhaps other plant elements, from the yeast *Ty*1 element.

Other deleted members of the *Stonor* family: One byproduct of the search for larger members of the *Stonor* family was the isolation of two elements (*Sto*-14 and *Sto*-12, Figure 5A) that, like *Stonor*, have sustained deletions. However, unlike *Stonor*, both of these elements contain numerous frameshifts, stop codons, deletions, and/or insertions in the ORF, indicating that they have been inactive for a long time. Furthermore, the deletion endpoints of both *Sto*-12 and *Sto*-14 occur in a direct repeat sequence (Figure 6, B and C), a structure associated with deletion formation during both DNA replication (Nalbantoglu *et al.* 1986) and reverse transcription (Temin 1993).

The larger family members: The results of three independent experimental approaches have led us to conclude that the largest members of the *Stonor* family are ~7.1–7.4 kb. First, a PCR assay using primers located in the 5′ and 3′ LTRs generated a product of 6.5 kb from genomes of maize and teosinte. When the positions of the primers are taken into account, this corresponds to an element of ~7.2 kb. Second, when primers flanking the endpoints of the *Stonor* deletion are used, the largest PCR product obtained was 5 kb. This would also correspond to a reconstituted *Stonor* element of 7.2 kb. Finally, the largest *Stonor* family member isolated from a genomic library, *Sto*-17, would be 7.4 kb upon restoration of its missing 3′ end (Figure 5A).

It is unknown at this time whether the active members of the *Stonor* family are of this size class. Because we have not been able to identify element-encoded transcripts (S. Marillonnet and S. R. Wessler, data not shown), the *Stonor* element provides our only link to the active progenitor (if we assume that it was active before sustaining the deletion, as described above). If this scenario is correct, then identification of the *Stonor* progenitor could lead us to an active family member. Unfortunately, both the 5′ and 3′ deletion breakpoints of *Stonor* reside in sequences held in common by all the cloned large elements (Figure 5B); *i.e.*, the variable region has been deleted in *Stonor*, making it impossible to identify its direct progenitor.

***Stonor* and related elements have long 5′ leaders:** The larger elements exhibit two unusual features downstream of the 5′ LTR. First, if transcribed, they would have exceptionally long 5′ leaders. The *Sto*-4 element has a single intact ORF of 1406 amino acids that encodes all the protein domains found in active retrotransposons. This ORF has a putative translation start ~1.6–1.8 kb from the 5′ LTR, just downstream of region J (Figure 5B). A putative TATA box is located between nucleotide positions 132 and 138 in the 5′ LTR (Varagona *et al.* 1992). Transcripts that may initiate at this position would have a 2.0- to 2.2-kb leader, which is longer than all other 5′ leaders reported to date. Only the *BARE*-1 element of barley and the *RIRE1* element of wild rice have leaders of comparable length (Suoniemi *et al.* 1996b). Interestingly, an analysis based on the alignment of the conserved regions of either the reverse transcriptase/RNaseH or protease domains of *Stonor* and several plant and Drosophila retrotransposons indicates that *Stonor* is more closely related to *BARE*-1 and *RIRE1* than to other retrotransposons, including those from maize (Figure 11 and data not shown).

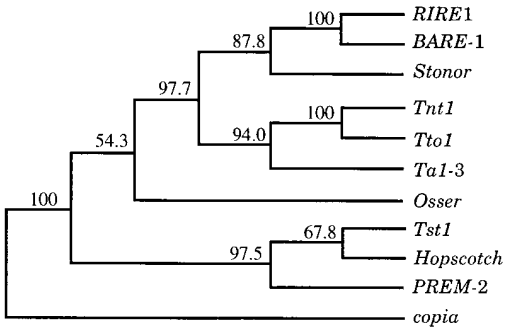


Figure 11.—Maximum parsimony tree of *copia*-like retrotransposons from plants and Drosophila using *copia* as the outgroup. Numbers next to each node give the bootstrap value supporting this node.

The presence of very long leaders in three (distantly) related families may indicate that this feature has been conserved because it has functional significance. The *BARE*-1 leader and the four putative leaders of the large *Stonor* family members contain numerous ATG codons in all reading frames (49 in *BARE*-1, 32 in *Sto*-3, 34 in *Sto*-4, 41 in *Sto*-30, and 38 in *Sto*-17). Leader AUGs have been shown to inhibit translation of downstream ORFs in both plants and animals (Damiani and Wessler 1993; Geballe 1996). However, while this may be a post-transcriptional mechanism to control retrotransposition of *Stonor* family members, it has not been very effective at regulating *BARE*-1, whose 30,000 copies comprise nearly 7% of the barley genome (Suoniemi *et al.* 1996a). Several viruses, however, including picornaviruses, can bypass the inhibitory effect of upstream AUGs by using internal ribosome entry sites (IRES) (Kaminski *et al.* 1994; Ehrenfeld 1996) that feature complex secondary structures. In picornaviruses, an IRES of 400–500 bp is required for internal initiation and is conserved within the viral species. In *Stonor* family members, ~400 bp of sequences located upstream of the large ORF are conserved and might potentially be involved in internal ribosome entry (Figure 5B, region J).

A variable domain in the larger elements: The second

unusual feature of the *Stonor* leader is a variable region that results in the structural uniqueness of each of the cloned elements. Variability of this type has not been reported previously for retrotransposons. Of interest is whether the variable sequences are an integral part of the element family, or whether they are random mutations that accumulate in inactive elements. The fact that these sequences are located at approximately the same position in all four elements and are flanked by sequences with >90% identity suggests the existence of a variable domain in the *Stonor* family.

What is the origin of this variable domain? It is unlikely to be caused by the insertion of transposable elements because the sequences do not have typical structural features of elements (*e.g.*, terminal inverted repeats and direct repeats). Furthermore, the diversity of sequences would mean that many element families would have inserted into similar but not identical sites. In contrast, mechanisms have been proposed for the acquisition of nonelement sequences during retrotransposition. The variable regions may have been acquired by a template switch mechanism like that proposed for transduction of cellular genes by oncogenic retroviruses (Swain and Coffin 1992; Zhang and Temin 1993). A similar mechanism was proposed for the acquisition of part of a maize cellular gene by the *Bs1* retrotransposon (Bureau *et al.* 1994; Jin and Bennetzen 1994). However, the presence of so many different sequences in approximately the same region of *Stonor* family members is hard to reconcile with any single mechanism.

Low- *vs.* high-copy-number families: *Stonor*, like other elements responsible for mutations in maize, is a member of a relatively low-copy-number family of ~30–40 elements. Family members display an unusual amount of structural diversity, including numerous deletions of sequences downstream of the 5′ LTR. The sequence of the *Stonor* deletion suggests that it was generated during retrotransposition. Similarly, other retrotransposon-induced mutations in maize contain defective elements, including *wxM* and *pl-987* (defective *Magellan* elements, Purugganan and Wessler 1994; P. L. Cooper and K. C. Cone, personal communication) and *bm-3* (a defective *B5* element, Vignols *et al.* 1995). Such examples may indicate that the generation of inactivating mutations may be one mechanism to minimize the copy number of these element families.

A post-transcriptional mechanism to control the *Stonor* copy number is also suggested by the identification of a long and variable leader with many upstream AUGs. However, the phylogenetically related *BARE*-1 element of barley has managed to attain very high copy numbers despite having a similar number of upstream AUGs. This could mean that the long leaders of *Stonor*, *BARE*-1, and *RIRE*1 are not conserved because they have a role in keeping copy number in check. Alternatively, the high copy number of the *BARE*-1 family may be caused by the amplification of a variant that was able to overcome the repressive effects of upstream AUGs. The finding that most of the *BARE*-1 family members appear to be full length and structurally homogeneous is consistent with this scenario (Suoniemi *et al.* 1996a). However, for such an element to be successful, it also had to evolve a mechanism to avoid inserting into genes. Although nothing is known about a target site preference for *BARE*-1, almost 60% of the 17 characterized insertion sites were found to be in other retrotransposons (Suoniemi *et al.* 1997). The existence of the *wxStonor* mutation suggests that the *Stonor* family has not (as yet) evolved a mechanism to avoid inserting into genes. The inability to evolve such a mechanism may be the most potent selective force preventing the amplification of the *Stonor* family.

We thank Drs. Kelly Dawe and Mike Scanlon for critical reading of the manuscript. This work was supported by a grant from the National Institutes of Health to S.R.W.

LITERATURE CITED

Bureau, T. E., S. E. White and S. R. Wessler, 1994 Transduction of a cellular gene by a plant retroelement. Cell **77:** 479–480.

Damiani, R. D., and S. R. Wessler, 1993 An upstream reading frame represses expression of *Lc*, a member of the *R*/*B* family of maize transcriptional activators. Proc. Natl. Acad. Sci. USA **90:** 8244–8248.

Ehrenfeld, E., 1996 Initiation of translation by picornavirus RNAs, pp. 549–573 in *Translational Control*, edited by J. W. B. Hershey, M. B. Mathews and N. Sonenberg. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Flavell, A. J., E. Dunbar, R. Anderson, S. R. Pearce, R. Hartley *et al.*, 1992 *Ty1-copia* group retrotransposons are ubiquitous and heterogeneous in higher plants. Nucleic Acids Res. **20:** 3639–3644.

Geballe, A. P., 1996 Translational control mediated by upstream AUG codons, pp.173–197 in *Translational Control*, edited by J. W. B. Hershey, M. B. Mathews and N. Sonenberg. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

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

Hirochika, H., and R. Hirochika, 1993 *Ty1-copia* group retrotransposons as ubiquitous components of plant genomes. Jpn. J. Genet. **68:** 35–46.

Hirochika, H., K. Sugimoto, Y. Otsuki and M. Kanda, 1995 Retrotransposons of rice involved in mutations induced by tissue culture. Proc. Natl. Acad. Sci. USA **93:** 7783–7788.

Jin, Y. K., and J. L. Bennetzen, 1994 Integration and nonrandom mutation of a plasma membrane proton ATPase gene fragment within the *Bs1* retroelement of maize. Plant Cell **6:** 1177–1186.

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

Kaminski, A., S. L. Hunt, C. L. Gibbs and R. Jackson, 1994 Internal initiation of mRNA in eukaryotes, pp.115–155 in *Genetic Engineering*, Vol. 16, edited by J. K. Setlow. Plenum Press, New York.

Lauermann, V., and J. D. Boeke, 1994 The primer tRNA sequence is not inherited during *Ty*1 retrotransposition. Proc. Natl. Acad. Sci. USA **91:** 9847–9851.

Nalbantoglu, J., D. Hartley, G. Phaer, G. Tear and M. Meuth, 1986 Spontaneous deletion formation at the *aprt* locus of hamster cells: the presence of short sequence homologies and dyad symmetries at deletion termini. EMBO J. **5:** 1199–1204.

Pulsinelli, G. A., and H. M. Temin, 1991 Characterization of large deletions occurring during a single round of retrovirus vector replication: novel deletion mechanism involving errors in strand transfer. J. Virol. **65:** 4786–4797.

Purugganan, M. D., and S. R. Wessler, 1994 Molecular evolution

of *magellan*, a maize *Ty3/gypsy*-like retrotransposon. Proc. Natl. Acad. Sci. USA **91:** 11674–11678.

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

Suoniemi, A., K. Anamthawat-Jonsson, T. Arna and A. H. Schulman, 1996a Retrotransposon *BARE*-1 is a major, dispersed component of the barley (*Hordeum vulgare* L.) genome. Plant Mol. Biol. **30:** 1321–1329.

Suoniemi, A., A. Narvanto and A. H. Schulman, 1996b The *BARE*-1 retrotransposon is transcribed in barley from an LTR promoter active in transient assays. Plant Mol. Biol. **31:** 295–306.

Suoniemi, A., D. Schmidt and A. H. Schulman, 1997 *BARE*-1 insertion site preferences and evolutionary conservation of RNA and cDNA processing sites. Genetica **100:** 219–230.

Swain, A., and J. M. Coffin, 1992 Mechanism of transduction by retroviruses. Science **255:** 841–845.

Temin, H. M., 1993 Retrovirus variation and reverse transcription—abnormal strand transfers result in retrovirus genetic variation. Proc. Natl. Acad. Sci. USA **90:** 6900–6903.

Varagona, M. J., M. Purugganan and S. R. Wessler, 1992 Alternative splicing induced by insertion of retrotransposons into the maize *waxy* gene. Plant Cell **4:** 811–820.

Vignols, F., J. Rigau, M. A. Torres, M. Capellades and P. Puigdomenech, 1995 The *brown midrib3* (*bm3*) mutation in maize occurs in the gene encoding caffeic acid o-methyltransferase. Plant Cell **7:** 407–416.

Voytas, D. F., M. P. Cummings, A. Konieczny, F. M. Ausubel and S. R. Rodermel, 1992 *copia*-like retrotransposons are ubiquitous among plants. Proc. Natl. Acad. Sci. USA **89:** 7124–7128.

Wessler, S. R., A. Tarpley, M. Purugganan, M. Spell and R. Okagaki, 1990 Filler DNA is associated with spontaneous deletions in maize. Proc. Natl. Acad. Sci. USA **87:** 8731–8735.

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

Zhang, J., and H. M. Temin, 1993 Rate and mechanism of nonhomologous recombination during a single cycle of retroviral replication. Science **259:** 234–238.

Communicating editor: V. Sundaresan