Sex Plant Reprod (1996) 9:65–74

ORIGINAL PAPER

Michael P. Turcich · Amana Bokhari-Riza
Douglas A. Hamilton · Caiping He · Walter Messier
Caro-Beth Stewart · Joseph P. Mascarenhas

PREM-2, a copia-type retroelement in maize is expressed preferentially in early microspores

Received: 19 September 1995 / Revision accepted: 20 November 1995

Abstract We have isolated, by screening a genomic library, a retroelement from maize designated PREM-2 (pollen retroelement maize-2), which is expressed in a tissue-specific manner. RNA transcripts of the PREM-2 family are found in the microspore but not in more mature pollen or in any of the vegetative tissues examined. The expression of PREM-2 elements in the uninucleate microspore provides an explanation for the genetic transmission of genomic rearrangements caused by the transposition of retroelements. PREM-2 elements are very abundant and are estimated to constitute about 5% of the maize genome and could possibly have played an important role in the determination of genome structure and in the generation of repetitive sequences in maize. The entire PREM-2 element is 9439 bp long. The LTRs of PREM-2 are 1307 bp in length. The internal region between the 5′ and 3′ LTRs contains 6825 bp and shares homology to the gag, pro, int, RT, and RNaseH regions of copia-type retroelements. PREM-2 elements have been found in close proximity with several maize genes registered in GenBank. The presence of PREM-2 sequence in the exact 5′ flanking position of three polygalacturonase genes expressed in pollen, has been used to examine the evolution of the polygalacturonase multigene family in maize and to estimate the time of the PREM-2 integration event.

The nucleotide sequence data reported appear in the Genebank, EMBL and DDBJ databases under the accession numbers U41000, U41079 and U41080.

M. P. Turcich · A. Bokhari-Riza · D. A. Hamilton[1]
C. He · W. Messier · C.-B. Stewart · J. P. Mascarenhas (✉)
Department of Biological Sciences
and Center for Molecular Genetics, University at Albany,
State University of New York, Albany, NY 12222, USA;
Fax 518–442–4354

Present address:
[1] Department of Biology, Hartwick College,
Oneonta, NY 13820, USA

Key words Maize · Retrotransposon · Microspore expression · Polygalacturonase genes · Genome structure

Introduction

One of the major classes of transposable elements found in eukaryotic organisms, the retroelements or retrotransposons, transpose via the reverse transcription of an RNA intermediate (Boeke and Corces 1989). These retroelements belonging to the retroviral super family contain at their ends long terminal repeats (LTRs). Internal to the LTRs are open reading frames that code for products important for replication and transposition of the element (Boeke and Corces 1989). The complete DNA sequences of retroelements from several plant species have been described (Schwarz-Sommer et al. 1987; Grandbastien et al. 1989; Jin and Bennetzen 1989; Smyth et al. 1989; Camirand et al. 1990; Konieczny et al. 1991; Manninen and Schulman 1993; White et al. 1994). In addition, many partial sequences which are very likely derived from LTR retrotransposons are known (see review, Flavell 1992; Voytas et al. 1992). The tobacco Tnt1 element is transcriptionally active. Its expression is restricted to freshly isolated mesophyll protoplasts, but it is not expressed in leaf tissue. The RNA is transcribed at low levels in the root (Pouteau et al. 1991). Transcripts of Tnt1 and other elements in tobacco are also found in established tissue cultures (Hirochika 1993). The transcription of the Tnt1 element is also greatly stimulated by different microbial elicitors of defense responses in plants (Pouteau et al. 1994). Relatively little, however, is known about the temporal or tissue expression of most plant retroelements (reviewed by Grandbastien 1992).

We recently isolated an LTR of a family of highly repetitive retroelements, designated PREM-1, which is estimated to be present in 10 000–40 000 copies in the maize genome (Turcich and Mascarenhas 1994). A diverse group of RNAs containing portions of the PREM-1

sequence are found during male gametophyte development, with highest levels of transcripts being present in the early uninucleate microspore.

While attempting to isolate a complete PREM-1 retroelement, we have identified a genomic clone that contained a complete retroelement belonging to a different family, which we designate PREM-2, which is inserted in the maize genome within a PREM-1 LTR sequence. The LTRs of PREM-2 are very different in their nucleotide sequence from those of PREM-1 but, like PREM-1, are transcribed in the young microspore although with somewhat different patterns of expression. The major characteristics of PREM-2 are presented here. These results have been briefly reported (Mascarenhas and Turcich 1994).

Materials and methods

Plant material, genomic and cDNA libraries

Seed of *Zea mays* inbred line W-22 was obtained from Illinois Foundation Seeds. The W-22 genomic library (Hamilton et al. 1989) was constructed in EMBL3. For the cDNA library, poly(A)RNA was isolated from early microspores and primed with oligo(dT); the double-stranded, linker-ligated cDNA was size selected for >0.5 kb and ligated in the pcDNAII vector (Invitrogen, San Diego, Calif.). Screening of the lambda and cDNA libraries was performed according to the methods outlined in Sambrook et al. (1989).

Isolation of RNA and genomic DNA

Microspores and pollen of various developmental stages were collected as previously described (Stinson et al. 1987) and flash-frozen in liquid nitrogen. Sporophytic tissues were collected and frozen in liquid nitrogen. Frozen tissues were ground into a fine powder (>80% cell breakage) at liquid nitrogen temperature using a mortar and pestle. RNA was isolated essentially as previously described (Willing and Mascarenhas 1984). Genomic DNA was isolated from young leaves or ears using the miniprep procedure of Dellaporta et al. (1985). DNA pellets were resuspended in TE buffer (10 mM TRIS, pH 8.0, 1 mM EDTA) and treated with 40 μg/ml RNase A for 30 min at 37°C.

DNA and RNA blot hybridizations

Restriction fragments of DNA for probe synthesis were purified by agarose gel electrophoresis and isolated using a GeneClean kit (BIO 101, Vista, Calif.). The LTR probe was a *Sal*I-*Xba*I fragment extending from nucleotide position 8422 to 8958. The RT/RNaseH probe extended from the *Sal*I site at position 5587 to the *Sac*I site at position 7282. Random primer probes were then generated as described by Feinberg and Vogelstein (1983) using α-[^{32}P]dCTP (Dupont NEN Research Products, Boston, Mass.). Digestion of the DNA and further analysis were as described in Turcich and Mascarenhas (1994).

For RNA blot hybridizations, 5 μg of RNA from each sample was analyzed by electrophoresis and hybridized as described (Turcich and Mascarenhas 1994).

DNA sequencing

Sequencing was by the dideoxy chain termination method using α-thio[^{35}S]dATP (Dupont NEN) as label and the Sequenase 2.0 sequencing kit (United States Biochemicals, Cleveland, Ohio), following the manufacturer's instructions. The sequencing strategy used synthetic oligonucleotide primers and cloning of DNA subfragments into either pBS$^+$ or pSK$^-$ vectors (Stratagene).

Evolutionary analyses

Sequences of polygalacturonase genes with flanking regions were obtained from GenBank (release 87). Nucleic acid and inferred protein sequences were aligned using the Pileup Program in the GCG package (Program manual for the Wisconsin package, version 8, 1994, Genetics Computer Group, Madison, Wis.). Parsimony trees were constructed from both the DNA and protein alignments using the program *PAUP* (Swofford 1993). Unsequenced regions of genes were coded as "missing data". All 12 known polygalacturonase genes were thus able to be included in the parsimony analysis, even though some genes contained no regions of overlap in the portions sequenced. For those polygalacturonase genes for which sufficient overlapping sequences were available to calculate distance matrices, trees were also built by the neighbor-joining method (Saitou and Nei 1987) in the computer package *PHYLIP* (Felsenstein 1993). Trees were rooted with the ZMPPOLGAL sequence as outgroup; the same topology was found by midpoint rooting. Dates of gene duplications (Fig. 6) were calculated based on the estimated rate of synonymous substitution (6×10^{-9} substitutions per site per year) in plant nuclear genes (Wolfe et al. 1989). Synonymous substitution rates were calculated for each pair of genes for which sufficient overlapping sequences were available using a computer program kindly provided by W.-H. Li (Li et al. 1985).

Results

Nucleotide sequence and other characteristics of the LTRs and internal region of PREM-2

Screening of the maize inbred line W-22 genomic library (Hamilton et al. 1989) with a PREM-1 LTR probe (Turcich and Mascarenhas 1994) resulted in the identification of a clone that appeared to contain two PREM-1 LTRs. A total of 17750 bp of maize chromosomal DNA from this genomic clone was sequenced. A map of the region sequenced together with the location of the PREM-1 LTRs and the position of the PREM-2 element are shown in Fig. 1. A complete PREM-2 retroelement has transposed into the PREM-1 3′ LTR in an orientation that is opposite to the PREM-1 sequences and 1236 bp upstream from the 3′ end of the putative 3′ PREM-1 LTR (Turcich and Mascarenhas 1994). Figure 2A shows the location of the structural features of the PREM-2 element. The entire PREM-2 element is 9439 base pairs (bp) long and has a 5-bp target site duplication (ATTAT) of maize DNA at its ends. The entire sequence as com-

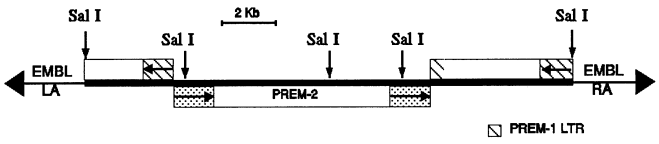

Fig. 1 Orientation and position of PREM-2 relative to a PREM-1 element in the maize genomic clone. *LA* and *RA* are the left and right arms of the vector EMBL3

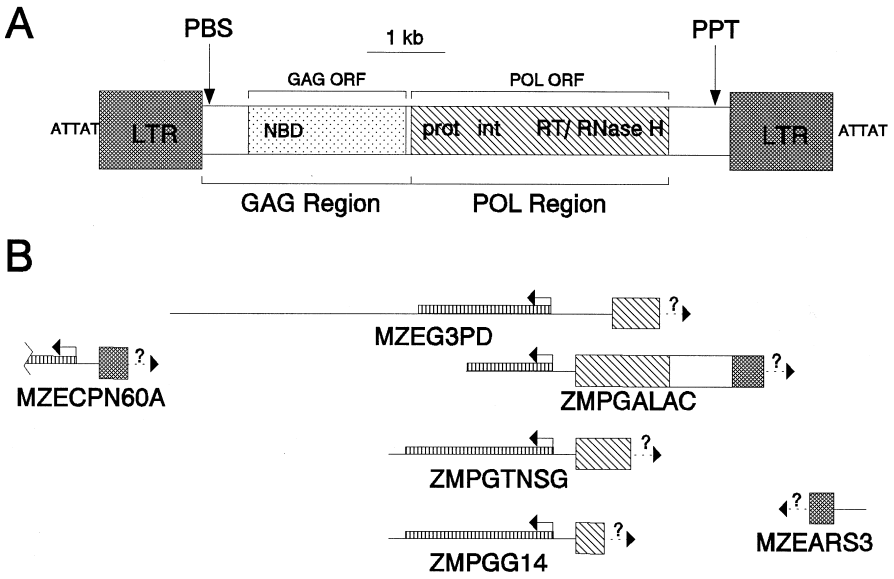

Fig. 2A, B Structural characteristics of PREM-2 and PREM-2 integrations in the maize genome. **A** Structural characteristics of the PREM-2 retroelement. *PBS* primer binding site; *PPT* polypurine tract; *prot* protease; *int* integrase; *RT* reverse transcriptase; *NBD* nucleotide binding domain. **B** The position of PREM-2 sequence in relation to six maize genes, the nucleotide sequences of which are found in Genbank. *Arrows with solid lines* denote the start and direction of translation. *Arrows with dashed lines and a question mark* indicate that nucleotide sequences are not available. Only a portion of the transcribed region for MZECPN60A is shown

municated to GenBank (accession number U41000) starts with nucleotide 1 as the first base of the 5′ target site 5-bp duplication, and ends with nucleotide 9449, the last base of the 3′ target site 5-bp duplication.

The LTRs of the PREM-2 element are 1307 bp long and contain 5′TG..3′CA terminal inverted repeats. The LTRs of the PREM-2 element appear to be consistent with the U3-R-U5 organization of retroviral LTR sequences (Boeke and Corces 1989). Two putative TATA boxes which are similar to plant consensus sequences (Joshi 1987) start at nucleotides 301 and 406, respectively. Also, two possible polyadenylation signals start at nucleotide positions 739 and 1154. Immediately interior to the 5′ LTR is a putative primer binding site (PBS) (TGGTATCAGAGCC) which has identity with 16 of 17 bases at the 3′ end of $tRNA^{met}$ from wheat (Ghosh et al. 1982) and yellow lupin (accession number X06458, Barciszewska et al. 1989). Interior to the 3′ LTR is a 13-bp polypurine tract (PPT) (AAAAAGGGGGAGA). Unlike other retroelements which have been described, however, there are several other polypurine stretches contained within a 600-bp region upstream from the polypurine tract that flanks the 3′ LTR of PREM-2. Six of these other polypurine stretches are similar, if not identical, to the sequence of the most 3′ PPT adjacent to the 3′ LTR.

The internal region between the 5′ and 3′ LTRs consists of 6825 bases and contains open reading frames which are homologous with the gag, prot (protease), int (integrase), RT (reverse transcriptase) and RNaseH regions of retroelements (Fig. 2A). The organization of these regions is similar to that of the copia element of *Drosophila* and the Ty-1 element of yeast (Boeke and Corces 1989). The gag region is unusually long (2757 bp) and consists of two overlapping open reading frames. The first ORF in the gag region starts with an ATG at nucleotide position 1361 and continues for 786 bp. This ORF overlaps with another longer ORF of 2157 bases. It is this second overlapping ORF which contains the characteristic nucleotide binding domain (NBD) of the gag polyprotein (see Fig. 2A).

The largest ORF within the internal domain of PREM-2 consists of sequences which encode a copia-like pol polyprotein. This ORF is 3243 bp in length and is in the same reading frame as the smaller gag ORF. The homologies of the conserved regions of the PREM-2 pol polyprotein with other copia type retroelements are shown in Fig. 3. As deduced from the nucleotide sequence of this particular PREM-2 element there appear to be two stop codons at positions 4054 and 4067, just between the gag ORFs and the pol polyprotein.

The amino acid sequence alignment of the NBD of the gag polyprotein region of representative copia-type retroelements, retroviruses and cauliflower mosaic virus are presented in Fig. 3A. The amino acid sequence of the PREM-2 NBD shows the greatest sequence similarity to the NBDs of the retrovirus sequences. The amino acid sequences that flank the NBD of PREM-2 also share the most identities and conservative substitutions with the amino acid sequences of the corresponding retroviral regions. In contrast, the most conserved region of the reverse transcriptase domains of PREM-2 is more similar in sequence to the copia and the plant copia-type elements than to the retroviruses or CaMV (Fig. 3B).

Developmental and tissue-related expression of the PREM-2 element

High-stringency RNA blot analyses using RNA isolated from various maize tissues including pollen at different stages of development are shown in Fig. 4. Hybridizations using both an LTR probe and a RT/RNaseH probe indicate that PREM-2 elements are transcribed in the uninucleate microspores of maize. Expression seems to be primarily in the early uninucleate microspores, with low expression in mid- to late-stage microspores, and no expression in more mature pollen or in any of the vegeta-

A

```
   PREM-2  KVCYKCGKPGHFIAKCPL
hopscotch  prCqvCsrvGHtalnCwy
    copia  vkChhCGreGHyqaqCfl
    Tnt-1  rnCYnCnqPGHFkrdCPn
   BARE-1  teCYyCkgmGHwkrnCPk
    SIVSP  qgCwKCGKaGHviAKCPe
    HIV1L  rgCwKCGKeGHqmkdCte
     CaMV  crCwiCnieGHyaneCPn
```

B

```
   PREM-2  QMDVKSAFLN GPIKEEV..Y VEQPPGFEDS EYPNHVYRLS KALYGLKQAP
hopscotch  QlDVqnAFLh GileEtV..Y mkQPPGFaDt thPNyhchLq KsLYGLKQrP
    copia  QMDVKtAFLN GtlKEEi..Y mrlPqGi..S cnsdnVckLn KAiYGLKQAa
    Tnt-1  QlDVKtAFLh GdleEEi..Y mEQPeGFEva gkkhmVckLn KsLYGLKQAP
   BARE-1  QMDVKaAFLN GllKEEl..Y mmQPeGFvDp knaNkackLq gsiYGLvQAs
   SIVagm  vlDVgdAyys iPldpnfrkY taftiptvnn qgPgirYqfn clpqGwKgsP
    HIV-1  vlDVgdAyfs vPldEdfrkY taftipsinn EtPgirYqyn vlpqGwKgsP
     CaMV  sfDcKSgFwq vlldqEsrpl taftc..... ..PqghYewn vvpfGLKQAP

   PREM-2  RAWYECLPDF LIANGFKVGK ADPTLF..TK TLENDLFVCQ IYVDDIIPGS
hopscotch  RAWYsrLsek LqslGFvpsK ADvsLF..iy nahstaiyil vYVDDIIitg
    copia  RcWfEvfeqa LkeceFvnss vDrciyildK gniNeniyvl lYVDDvviat
    Tnt-1  RqWYmkfdsF mksqtylkty sDPcvy.fkr fsENnfiill lYVDDmlivg
   BARE-1  RsWnkrfgev ikAfGFiqvv gesciy..kK vsgssvafli lYVDDIllig
   SIVagm  tifqntaasi Leeik..... .......... .rnlpaltiv qYmDDlwvGS
    HIV-1  aifqssmtki Lepfr..... .......... .kqNpdiViy qYmDDlyvGS
     CaMV  sifqrhmdea frvfr..... .......... ......kfCc vYVDDIlvfS
```

Fig. 3A,B Amino acid sequence alignment of the nucleotide binding domain of the gag region and of the most conserved region of the reverse transcriptase domains of PREM-2 and representative copia-type retroelements. **A** Amino acid sequence alignment of the nucleotide binding domain (NBD) of the gag polyprotein regions of representative copia retroelements, retroviruses, and cauliflower mosaic virus (CaMV). Conserved cysteines and histidines are in *bold*. Amino acid identities with the PREM-2 NBD are *bold upper case*. SIVSP is the PBJ/BC13 isolate of simian immunodeficiency virus (accession number P19504); the accession number for hiv1l is PO5888. **B** Amino acid sequence alignment of the most conserved region of the reverse transcriptase domains of representative Pol ORFs. *Bold letters* denote the most conserved portions in copia-type retroelements. *Upper case letters* show positions of amino acids where there is identity of at least five out of seven residues. SIVagm is the agm isolate of simian immunodeficiency virus (GenBank accession number Q02836). Alignment was generated using the pileup program of GCG

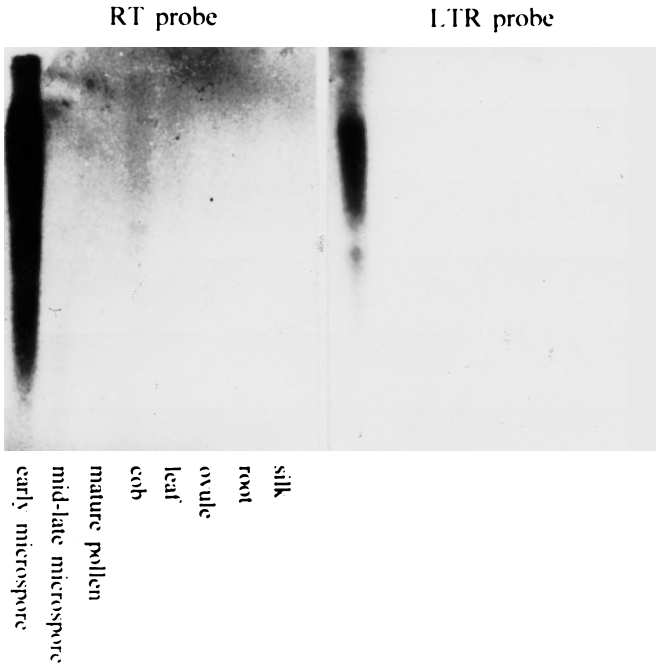

Fig. 4 RNA blot analysis with RNA (10 μg total RNA) from different stages of pollen development and different vegetative tissues, hybridized with a PREM-2 reverse transcriptase probe and a PREM-2 LTR probe. The order of the lanes is the same for both halves of the figure

tive tissues studied. This is somewhat similar but not identical to the RNA transcription pattern observed for the PREM-1 element previously described (Turcich and Mascarenhas 1994). The hybridization obtained is to a diverse population of RNA sizes ranging from very large (>10 kb) to fairly small (about 500 bp). However, the majority of transcripts appear to be large (>6 kb), consistent with the large size of this retroelement.

Further evidence for the transcription of PREM-2 elements in uninucleate microspores has been obtained by screening a cDNA library made to poly(A)RNA from early microspores with an LTR probe. Two of the obtained cDNA clones were sequenced (PREM-2cdna1, PREM-2cdna2, accession numbers U41079 and U41080, respectively). Both clones contain about 800 bp of PREM-2 LTR sequence starting about 100–200 bp from the 3′ end of the 3′ LTR. Each of the clones has about 70% identity in base sequence with the LTR of the complete PREM-2 element, although there are regions where the sequence identity is from 80 to >90%.

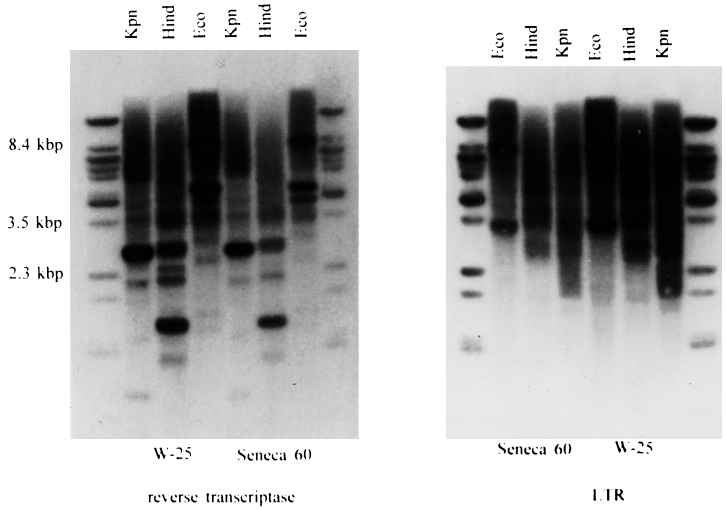

Fig. 5 Maize (inbred lines W-25 and Seneca-60) genomic DNA blot hybridizations using segments of the LTR and reverse transcriptase regions of PREM-2 as probes. The DNA was digested with various restriction endonucleases as indicated

Repetition frequency of PREM-2 elements in the maize genome

Genomic DNA blot analyses were carried out using segments of the LTR and reverse transcriptase regions of PREM-2 as probes and employing high-stringency conditions for the hybridization. These data are presented in Fig. 5. It is apparent that both the LTR sequences and the reverse transcriptase are very prevalent in the maize genome. To estimate the number of copies of the PREM-2 family present in the maize genome, the number of hybridizing plaques in the W-22 genomic library was determined. About 9% of the plaques hybridized to the PREM-2 LTR probe at high stringency. The number of copies in the genome was calculated as described (Turcich and Mascarenhas 1994) and was found to be about 30 000 per haploid genome of maize.

Identification of PREM-2 sequences in maize genes

A search of the GenBank database (Release 87) revealed six genomic sequences from maize, portions of which are comprised of sequences having a high degree of similarity to the PREM-2 LTRs and/or internal regions. These include three polygalacturonase nuclear genes expressed in pollen, ZMPGALAC (accession number X66692, Allen and Lonsdale 1993), ZMPGTNSG (accession number X64408, Barakate et al. 1993), and ZMPGG14 (accession number X65845, Barakate et al. 1993), a genomic clone containing glyceraldehyde-3-phosphate dehydrogenase subunit A (MZEG3PD, accession number M18976, Quigley et al. 1988), a nuclear-encoded mitochondrial chaperonin 60 gene (MZECPN60A, accession number L21007; P.S. Close, unpublished) and a non-coding genomic sequence from maize that has the properties of replicating autonomously in yeast (MZEARS3, accession number X12755; Berlani et al. 1988). Figure 2B summarizes the positions of the PREM-2 sequences relative to these maize genes. In all six cases the PREM-2 element appears to have integrated in an orientation opposite to the coding strand, or at least opposite to the sequence in the case of the ARS-3 clone, and in four cases is located close to the presumed promoters of the genes. In these four cases, the known sequence ends within the Pol ORF, but it is not known whether the remainder of the PREM-2 sequences are to be found downstream since further sequences are unavailable. In the case of MZECPN60A, the 5′ end of a 5′ LTR of the PREM-2 element has integrated about 460 bp upstream of the first exon of the gene. This could potentially represent the integration of a complete PREM-2 retroelement in this position, but unfortunately with the exception of about 400 bp of common sequence with PREM-2 no further sequence is available. Within this region, however, there are two deletions, one of 27 bp and another of 6 bp.

Evolution of the polygalacturonase gene family in maize

The presence of the identical PREM-2 sequence in exactly the same position upstream of the polygalacturonase genes prompted us to examine the evolution of the polygalacturonase gene family in maize. This family is estimated to contain 12–13 genes (Barakate et al. 1993). Of the six polygalacturonase genes for which 5′ flanking nucleotide sequences are available in GenBank, three have PREM-2 sequences (ZMPGTNSG, ZMPGALAC, ZMPGG14, Fig. 2B), two apparently do not contain PREM-2 sequence (ZMPOLGAL, accession number X66422; ZMPGG5, accession numbers X65846 and X65847) and one is unknown because not enough upstream sequence is available (ZMPGG6, accession number X65844). Partial coding sequences are available for several other polygalacturonase genes (Barakate et al. 1993).

To trace the evolutionary history of PREM-2 insertion(s) in this gene family, phylogenetic trees were built using both parsimony (Swofford 1993) and neighbor-

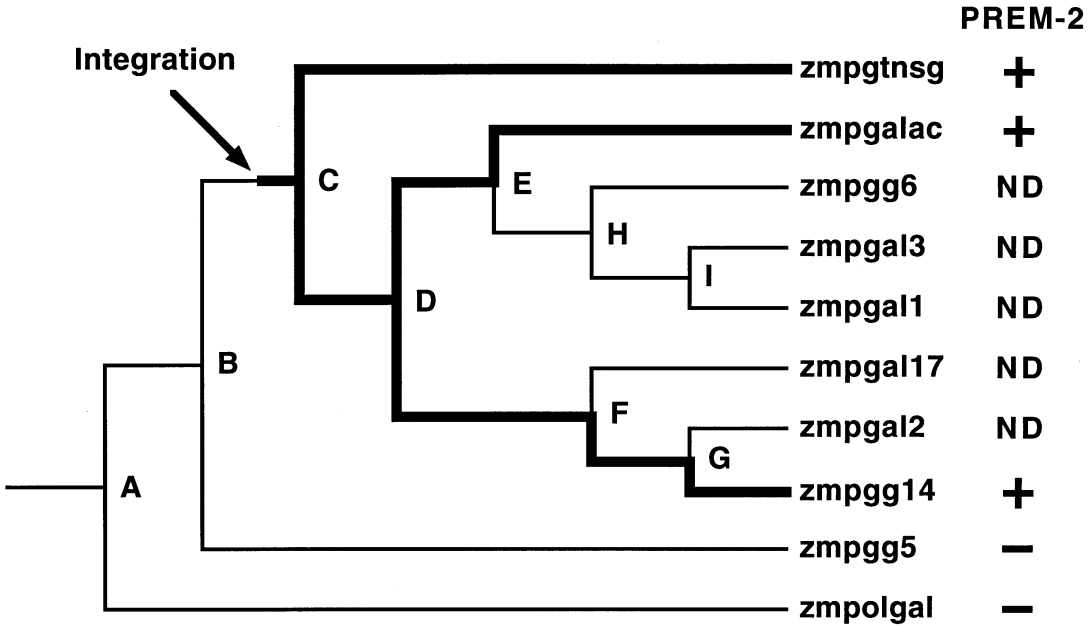

Fig. 6 Neighbor-joining tree for the polygalacturonase gene family. Gene names are listed in the text. *Internal nodes* represent putative gene duplication events; *letters beside the nodes* are for reference to calculated times of gene duplication: *A* 9.2–15 million years ago (mya); *B* 5.8–8.1 mya; *C* 2.3–3.1 mya; *D* 1.8–2.6 mya; *E* not calculated; *F* 0.62–0.84 mya; *G* 0.53–0.74 mya; *H* 0.53–0.74 mya; *I* 0.50–0.70 mya. Duplication dates *A* and *B* may not be highly reliable, as they were calculated using ZMPGG5, for which only 285 nt of common sequence is available. *Arrow* indicates the most parsimonious explanation for the time of the PREM-2 insertion. *Plus signs* indicate polygalacturonase sequences with PREM-2 sequence (see Fig. 2B). *Minus signs* indicate sequences that lack PREM-2 upstream. *ND* indicates those genes for which insufficient upstream sequence has been determined. The *bold line* traces the polygalacturonase gene duplicates with known PREM-2 sequences. Although they are not shown in this neighbor-joining tree because they do not have enough overlapping sequence for calculation of a distance matrix, ZMPGAL65 clusters with the ZMPGAL17/ZMPGAL2/ZMPGG14 clade, and ZMPGAL4 and ZMPGAL47 cluster with the ZMPGALAC/ZMPGAL1/ZMPGAL3/ZMPGG6 clade in parsimony trees

joining (Saitou and Nei 1987; Felsenstein 1993) methods from both the nucleotide and amino acid sequences of the coding regions. Because of its high divergence from the other sequences, the ZMPOLGAL gene was used as an outgroup to root these trees.

The neighbor-joining tree is shown in Fig. 6. Although there were many parsimony trees of equal length, consensus trees gave a similar picture regarding the overall evolutionary relationships to those shown in the neighbor-joining tree. Importantly, in all rooted trees the ZMPGG5 sequence, which does not contain the PREM-2 insertion, fell outside of a cluster that contained all other sequences including the three genes known to have the PREM-2 insertion (ZMPGTNSG, ZMPGALAC, ZMPGG14). The genes for which upstream sequence information is not available grouped with either ZMPGALAC or ZMPGG14. Since all of these genes are believed to be products of separate loci in maize (Barakate et al. 1993) these divergences represent gene duplication events. The calculated dates of these duplications are shown in Fig. 6.

Discussion

We have described the discovery and characterization of a new maize retroelement, which we designate PREM-2 because of its expression in immature pollen. The PREM-2 retroelement is copia-like in organization and is different from other retroelements described for maize – Cin1 (Shepherd et al. 1984); Bs1 (Jin and Bennetzen 1989); G, B5 and Stonor (Varagona et al. 1992); PREM-1 (Turcich and Mascarenhas 1994); *Hopscotch* (White et al. 1994); and *magellan* (Purugganan and Wessler 1994) – since the LTRs of the different elements are very different in sequence. The Bs1 element of maize is atypical because of its relatively small size and lack of sequence similarity to reverse transcriptase (Jin and Bennetzen 1994; Grandbastien 1992); the Cin1 element seems to be a solo LTR (Grandbastien 1992); and, except for *Hopscotch* (White et al. 1994), complete sequences for the other reported maize retroelements are not yet available. The PREM-2 family of retroelements is very prevalent in the maize genome and is present in about the same number of copies as PREM-1 (Turcich and Mascarenhas 1994). The copy number of the autonomously replicating sequence ARS3 was estimated to be about 28 000 per haploid maize genome (Berlani et al. 1988). Nucleotides 1–248 of ARS3 share 82.3% identity with a segment from the LTR of PREM-2. Based on the A+T content and strand asymmetric distribution of bases, it is likely that the portion of ARS3 that contains ARS activity extends from position 243 to 603 (Berlani et al. 1988). Since the probe used for the estimation of the number of copies of ARS3 apparently included sequences upstream of 243 which we now know make up part of the LTR of PREM-2 (Fig. 2B), the number of 28 000 copies proba-

bly reflects the number of PREM-2 copies in the maize genome in addition to the number of copies of ARS3. This is an independent confirmation of our estimate of the copy number of PREM-2 elements in the maize genome. The PREM-2 sequence, however, seems to mark the limit of the 5′ terminus of ARS3.

Based on its size and estimated copy number, PREM-2 could comprise about 5% of the maize genome, as is also true for PREM-1 (Turcich and Mascarenhas 1994). These observations provide evidence for the importance of retroelements in determining genome structure in maize, and possibly in other plants, and strengthen the suggestion that retroelements are likely involved in the generation of repetitive sequences that are responsible for the large genome size of several species (Flavell 1986; Weiner et al. 1986; McDonald 1990; Leeton and Smyth 1993).

PREM-2 appears to be the first reported case of a complete plant retroelement with microspore-specific or tissue-specific expression. Although PREM-1 is also expressed in pollen, it is not yet known whether PREM-1 is a complete element (Turcich and Mascarenhas 1994). The Tnt1 element from tobacco is transcribed in freshly isolated mesophyll-derived protoplasts but not in mesophyll cells. The induction is dependent on the presence of cell wall hydrolases or microbial elicitors of plant defense reactions, but does not seem to be developmentally regulated (Pouteau et al. 1991, 1994). McClintock (1984) suggested that the expression of transposable elements could be triggered by “genomic stress”. Clearly the activation of Tnt1 by cell wall-degrading hydrolases and microbial elicitors of plant defense reactions can be considered a reaction to genomic stress. Whether there is any similarity between such environmental stress conditions and early male gametophyte development is not known, but it is intriguing that both of these situations entail a major changeover in the preexisting genetic regulatory program.

The expression of the PREM-2 element in the early microspore (Fig. 4) is a significant finding because it provides an explanation for the genetic transmission of genomic rearrangements caused by the transposition of retroelements. Transposons and retrotransposons are hypothesized to provide important mechanisms for generating evolutionary diversity and change (McDonald 1990). For such changes to be inherited, the transposition events must occur in the cells giving rise to the sperm or egg. The uninucleate microspore is the precursor of the sperm cells and any rearrangements or transposition of DNA occurring at this time can potentially be inherited by the generative cell nucleus and later the sperm nuclei. The expression of PREM-2 elements in the early microspore might be an indication of their active transposition at this time. The male gametophyte represents a stage in the life cycle of plants that might tolerate a higher level of transposable element activity because selective pressures during and after gametogenesis could eliminate transposition events that result in deleterious effects. Changes that have positive consequences for pollen tube growth could also be selected for. The great excess of pollen grains produced relative to the numbers of embryo sacs and the resultant gametophytic competition that has been observed and studied (Mulcahy and Mulcahy 1987) might be related to retroelement activity in the microspore. Interestingly, the transposition of the yeast retrotransposon Ty1 is mostly confined to the haploid or gametic phase of the yeast life cycle (Ji et al. 1993). Alternatively, if the selfish DNA hypothesis (Doolittle and Sapienza 1980; Orgel and Crick 1980) is valid, it may be a selective advantage to the retroelement itself to have evolved to express its activity at this stage of development.

Several of the genes that lie near to where the PREM-2 element and other retroelements in maize have transposed are known to be expressed in pollen. Examples include the polygalacturonase genes (Allen and Lonsdale 1993: Barakate et al. 1993), the waxy genes (Bureau and Wessler 1992; Varagona et al. 1992), alcohol dehydrogenase (Bureau and Wessler 1992), and ADP glucose glucosyl-transferase (Bureau and Wessler 1992). If retroelements preferentially insert into transcriptionally active regions of nuclear DNA, as some indirect evidence suggests (Chalker and Sandmeye 1992), then these data may be also indicative of a heightened level of plant retroelement activity in developing pollen.

Our analyses of PREM-2 sequences in GenBank suggest that there may be possible regulatory interactions between these sequences and the genes they flank. This possibility is suggested by the fact that most of these PREM-2 sequences are intimately associated with genes at their proximal upstream locations. Retroelement insertions can give rise to altered spatial and temporal expression patterns of genes that are in close proximity (McDonald 1990; Weil and Wessler 1990; Smith and Corces 1991; Robins and Samuelson 1992; White et al. 1994). For example, the insertion of a *Stonor* retroelement in the *waxy* (*Wx*) gene of maize makes it null for *Wx* expression in pollen, yet *Wx* expression is retained in the endosperm (Varagona et al. 1992). Furthermore, control sequences known to be critical for the proper regulation of expression of eukaryotic genes have been identified within the LTRs of different retroelements (McDonald 1990). The LTR of the tobacco Tnt1 retroelement, for example, is known to contain *cis*-acting regulatory elements that lie upstream from the TATA box and which are required for the induction of its transcription in leaf protoplasts (Casacuberta and Grandbastien 1993). Similar conclusions concerning the possible regulation by retroelements of adjacent genes have been reached by White et al. (1994) from their computer-assisted database search with *Hopscotch* and other copia-like elements.

A 5′ PREM-2 LTR is present about 450 bp 5′ to the first exon of the chaperonin 60 (MZECPN60A) gene, and the 3′ end of a 3′ LTR is found just prior to the ARS3 sequence (Fig. 2B). Four of the genes (MZEG3PD, ZMPGALAC, ZMPGTNSG, ZMPGG14) have integrated into the region of the pol ORF of PREM-2 (Fig. 2B). Three of the genes which have adjacent

PREM-2 sequences are members of the multigene family of exopolygalacturonase proteins in maize. The same pol region of PREM-2 is found in the identical position at the 5′ ends of three members of the family shown in Fig. 2B; in the case of the ZMPGALAC gene this is 326 bp 5′ of the start of transcription.

The presence of PREM-2 sequence within 326 bp from the start of transcription has obviously not prevented pollen expression of these genes. Whether more subtle effects have resulted, such as effects on the strength of the promoters, cannot be stated until more quantitative studies have been carried out. The *cis*-acting elements required for pollen-specificity are located very near to the start of transcription (see review, Hamilton and Mascarenhas 1996). For example, for the pollen-specific gene Zm13, the signals for pollen specificity reside within about 118 bp from the start of transcription (Hamilton et al. 1989; Hamilton and Mascarenhas 1996; manuscript in preparation). It is thus not surprising that pollen expression of the three polygalacturonase genes does not appear to be affected even though they are so near to PREM-2.

It has recently been found that a fragment of a plasma membrane ATPase (*pma*) gene has integrated within the Bs1 element of maize (Bureau et al. 1994; Jin and Bennetzen 1994; Palmgren 1994). This is the first example of a retrotransposon that has acquired a portion of a normal cellular gene. The presence of partial PREM-2 sequence found adjacent to the polygalacturonase genes, and to the other genes described in Fig. 2, might suggest the possibility of an ancestral transduction of these genes by the PREM-2 element, similar to the situation of the *pma* gene in the Bs1 element.

The fact that none of the six PREM-2 insertions has disrupted the coding region of a gene is intriguing. In all cases, the PREM-2 sequences are found a short distance upstream of the promoter region. Retrotransposon-like sequences have also been found by White et al (1994) to lie adjacent to several plant genes. An analysis has been made of the insertion sites of over 100 insertion events of the Ty1 retroelement in the yeast chromosome III. Most of the regions where the Ty1 element was found were tRNA genes and/or LTRs of preexisting transposable elements. Almost all of the insertions were 5′ of the tRNA coding region. Although identified ORFs represent about 70% of chromosome III, only one of the sequenced insertions disrupted an ORF (Ji et al. 1993). Ji et al. (1993) have suggested that the Ty1 integration machinery is able to detect “safe havens” for insertion into the genome. It would be most interesting if PREM-2 did the same. The PREM-2 element has integrated into the PREM-1 LTR in an orientation that is opposite to that of the PREM-1 sequence. Moreover, all of the other chromosomal genes or sequences (Fig. 2B) are in opposite orientation to the PREM-2 sequence. Whether this observation has some significance with respect to the mechanism of integration of the PREM-2 element is as yet unclear.

PREM-2 sequence is known to be located in front of three separate polygalacturonase genes. Based on the evolutionary tree for the polygalacturonase family of genes (Fig. 6), the simplest explanation for this observation is that there was a single PREM-2 integration event, or a transduction event, which preceded the gene duplications that gave rise to the ZMPGTNSG, ZMPGALAC, and ZMPGG14 genes and their close relatives. This interpretation is consistent with the observation that these PREM-2 sequences are integrated at the identical nucleotide position in the three polygalacturonase genes. If this interpretation is correct we can predict that the genes ZMPGG6, ZMPGAL3, ZMPGAL1, ZMPGAL4, ZMPGAL47, ZMPGAL17, ZMPGAL2, and ZMPGAL65 (for which 5′ sequences are not known) will be found to contain PREM-2 sequences in their upstream regions, since they appear to have arisen by gene duplication events that occurred after the PREM-2 integration.

We can estimate that the earliest of these particular gene duplications occurred between about 2.3 and 3.1 million years ago (Fig. 6). Thus the PREM-2 integration event most likely occurred prior to this time but after the divergence from ZMPGG5, which we estimate at about 5.8–8.1 million years ago.

Since a majority of the members of the polygalacturonase gene family are predicted to contain a PREM-2 sequence, this raises the possibility that the presence of PREM-2 has facilitated these recent gene duplications. Whether retroelements might generally be involved in gene duplication in maize and other plants is an intriguing question that might be addressed through analyses of retroelements and their proximity to other duplicated genes.

Are PREM-2 elements presently active in the maize genome? Our data show that transcripts of the elements as determined by both LTR and reverse transcriptase probes are present in the microspore, and are of a size to include complete transcripts of the element (Fig. 4); thus, it is likely that at least some members of the PREM-2 family are functional elements. Whether the particular member of the PREM-2 family characterized here is currently active remains to be determined; however, it seems reasonable to expect that it is, for the following reasons. Although there are two stop codons just between the gag ORFs and the pol polyprotein, frame shifting and stop codon suppression could conceivably result in the translation of an active pol polyprotein, since such mechanisms are known to occur in other retroelements and in retroviruses (Boeke 1989; Lewin 1994; Voytas and Boeke 1993). Also, since there are only four nucleotide differences between the two PREM-2 LTRs (probably two changes per LTR), it seems reasonable to assume that relatively few changes have occurred within the internal domains which might render the element inactive. None of the gene sequences obtained from the DNA sequence database appear, however, to be recent integration events of PREM-2 (Fig. 2B). The fact that in the six genes involved the nucleotide identity in the regions shared with PREM-2 ranges between 68% and 91%, the fact that only a portion of the PREM-2 sequence is found adjacent to the genes, and that small de-

letions in the PREM-2 sequence are found in one of the genes, suggest that rearrangements of PREM-2 DNA and substantial sequence divergence have occurred. This would indicate that the PREM-2 integration is not very recent for any of the loci involved. With the polygalacturonase gene family, our estimates for PREM-2 integration are earlier than about 2–3 million years ago. The sequence information from the two cDNA clones isolated from a microspore cDNA library suggests, in addition, that the PREM-2 family of elements is ancient.

Further investigation of PREM-2 sequences in the maize genome and the isolation of additional high copy number retroelements and their analysis might help us better understand the evolution of genome structure and organization in plants such as maize, which contain a large amount of repeat DNA.

Acknowledgement This research was supported by grant IBN-9208337 from the National Science Foundation to J.P.M.

References

Allen RL, Lonsdale DM (1993) Molecular characterization of one of the maize polygalacturonase gene family members which are expressed during late pollen development. Plant J 3:261–271

Barakate A, Martin W, Quigley F, Mache R (1993) Characterization of a multigene family encoding an exopolygalacturonase in maize. J Mol Biol 229:797–801

Barciszewska MZ, Joachimiak A, Barciszewski J (1989) The initiator transfer ribonucleic acid from yellow lupin seeds, correction of nucleotide sequence and crystallization. Phytochemistry 28:2039–2043

Berlani RE, Walbot V, Davis RW (1988) Sequence analysis of three fragments of maize nuclear DNA which replicate autonomously in yeast. Plant Mol Biol 11:173–182

Boeke JD, Corces VG (1989) Transcription and reverse transcription of retrotransposons. Ann Rev Microbiol 43:403–434

Bureau TE, Wessler SR (1992) *Tourist*: a large family of small inverted repeat elements frequently associated with maize genes. Plant Cell 4:1283–1294

Bureau TE, White SE, Wessler SR (1994) Transduction of a cellular gene by a plant retroelement. Cell 77:479–480

Camirand A, St-Pierre B, Marineau C, 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

Casacuberta JM, Grandbastien M-A (1993) Characterization of LTR sequences involved in the protoplast specific expression of the tobacco Tnt1 retrotransposon. Nucl Acids Res 21:2087–2093

Chalker DL, Sandmeye SB (1992) Ty3 integrates within the region of RNA polymerase III transcription initiation. Gene Develop 6:117–128

Dellaporta SL, Wood J, Hicks JB (1985) Maize DNA miniprep. In: Malmberg R, Messing J, Sussex I (eds) Molecular biology of plants: a laboratory course manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp 36–37

Doolittle WF, Sapienza C (1980) Selfish genes, the phenotype paradigm and genome evolution. Nature 284:601–603

Feinberg AP, Vogelstein B (1983) A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal Biochem 132:6–13

Felsenstein J (1993) PHYLIP (Phylogenetic Inference Package), version 3.5. Computer software and manual distributed by the author, Department of Genetics, University of Washington, Seattle, Wash

Flavell AJ (1992) *Ty1-copia* group retrotransposons and the evolution of retroelements in the eukaryotes. Genetica 86:203–214

Flavell RB (1986) Repetitive DNA and chromosome evolution in plants. Philos Trans R Soc London Ser B 312:227–242

Ghosh HP, Ghosh K, Simsek M, RajBhandary UL (1982) Nucleotide sequence of wheat germ cytoplasmic initiator methionine transfer ribonucleic acid. Nucleic Acids Res 10:3241–3247

Grandbastien M-A (1992) Retroelements in higher plants. Trends Genet 8:103–108

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

Hamilton DA, Bashe DM, Stinson JR, Mascarenhas JP (1989) Characterization of a pollen-specific genomic clone from maize. Sex Plant Reprod 2:208–212

Hamilton DA, Mascarenhas JP (1996) Gene expression during pollen development. In: Sawhney VK, Shivanna KR (eds) Pollen biotechnology for crop production and improvement. Cambridge University Press (in press)

Hirochika H (1993) Activation of tobacco retrotransposons during tissue culture. EMBO J 12:2521–2528

Ji H, Moore DP, Blomberg MA, Braiterman LT, Voytas DF, Natsoulis G, Boeke JD (1993) Hotspots for unselected Ty1 transposition events on yeast chromosome III are near tRNA genes. Cell 73:1007–1018

Jin Y-K, Bennetzen JL (1989) Structure and coding properties of *Bs1*, a maize retrovirus-like transposon. Proc Natl Acad Sci USA: 86:6235–6239

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

Joshi CP (1987) An inspection of the domain between putative TATA box and translation start site in 79 plant genes. Nucleic Acids Res 15:6643–6653

Konieczny A, Voytas DF, Cummings MP, Ausubel FM (1991) A superfamily of *Arabidopsis thaliana* retrotransposons. Genetics 127:801–809

Leeton PRJ, Smyth DR (1993) An abundant LINE-like element amplifed in the genome of *Lilium speciosum*. Mol Gen Genet 237:97–104

Lewin B (1994) Genes V. Oxford University Press, Oxford

Li W-H, Wu C-I, Luo C-C. (1985) A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative liklihood of nucleotide substitution and codon changes. Mol Biol Evol 2:150–174

Manninen I, Schulman AH (1993) *BARE-1*, a *copia*-like retroelement in barley (*Hordeum vulgare* L). Plant Mol Biol 22: 829–846

Mascarenhas JP, Turcich MP (1994) The activity of retroelements during male gametophyte development. In: Kao T-h, Stephenson AG (eds) Pollen-pistil interactions and pollen tube growth. American Society of Plant Physiologists, Rockville, Md, pp 39–44

McClintock B (1984) The significance of responses of the genome to challenge. Science 226:792–801

McDonald JF (1990) Macroevolution and retroviral elements. BioScience 40:183–191

Mulcahy DL, Mulcahy GB (1987) The effects of pollen competition. Am Sci 75:44–50

Orgel LE, Crick FHC (1980) Selfish DNA: the ultimate parasite. Nature 284:604–607

Palmgren MG (1994) Capturing of host DNA by a plant retroelement – Bs1 encodes plasma membrane H^+-ATPase domains. Plant Mol Biol 25:137–140

Pouteau S, Huttner E, Grandbastien MA, Caboche M (1991) Specific expression of the tobacco Tnt1 retrotransposon in protoplasts. EMBO J 10:1911–1918

Pouteau S, Grandbastien M-A, Boccara M (1994) Microbial elicitors of plant defense responses activate transcription of a retrotransposon. Plant J 5:535–542

Purugganan MD, Wessler SR (1994) Molecular evolution of *magellan*, a maize Ty3/*gypsy*-like retrotransposon. Proc Natl Acad Sci USA 91:11674–11678

Quigley F, Martin WF, Cerff R (1988) Intron conservation across the prokaryote-eukaryote boundary: structure of the nuclear gene for chloroplast glyceraldehyde-3-phosphate dehydrogenase from maize. Proc Natl Acad Sci USA 85:2672–2676

Robins DM, Samuelson LC (1992) Retrotransposons and the evolution of mammalian gene expression. Genetica 86:191–201

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

Schwarz-Sommer Z, Leclercq L, Gobel E, Saedler H (1987) Cin4, an insert altering the structure of the *A1* gene in *Zea mays*, exhibits properties of nonviral retrotransposons. EMBO J 6:3873–3880

Shepherd NS, Schwarz-Sommer Z, Blumberg vel Spalve J, Gupta M, Wienand U, Saedler H (1984) Similarity of the Cin1 repetitive family of *Zea mays* to eukaryotic transposable elements. Nature 307:185–187

Smith PA, Corces VG (1991) Drosophila transposable elements: mechanisims of mutagenesis and interactions with the host genome. Adv Genet 29:229–300

Smyth DR, Kalitsis P, Joseph JL, Sentry JW (1989) Plant retrotransposon from *Lilium henri* is related to *Ty3* of yeast and the gypsy group of *Drosophila*. Proc Natl Acad Sci USA 86:5015–5019

Stinson JR, Eisenberg AJ, Willing RP, Pe ME, Hanson DD, Mascarenhas JP (1987) Genes expressed in the male gametophyte of flowering plants and their isolation. Plant Physiol 83:442–447

Swofford DL (1993) PAUP: phylogenetic analysis using parsimony, version 3.1. Computer program distributed by Illinois Natural History Survey, Champaign, Ill

Turcich MP, Mascarenhas JP (1994) PREM-1, a putative maize retroelement has LTR (long terminal repeat) sequences that are preferentially transcribed in pollen. Sex Plant Reprod 7:2–11

Varagona MJ, Purugganan M, Wessler SR (1992) Alternative splicing induced by insertion of retrotransposons into the maize *waxy* gene. Plant Cell 4:811–820

Voytas DF, Boeke JD (1993) Yeast retrotransposons and tRNAs. Trends Genet 9:421–427

Voytas DF, Cummings MP, Konieczny A, Ausubel FM, Rodermel SR (1992) copia-like retrotransposons are ubiquitous among plants. Proc Natl Acad Sci USA 89:7124–7128

Weil CF, Wessler SR (1990) The effects of plant transposable element insertion on transcription initiation and RNA processing. Annu Rev Plant Physiol 41:527–552

Weiner AM, Deininger PL, Efstratiadis A (1986) Nonviral retroposons: genes, pseudogenes, and transposable elements generated by the reverse flow of genetic information. Ann Rev Biochem 55:631–661

White SE, Habera LF, Wessler SR (1994) Retrotransposons in the flanking regions of normal genes: a role for copia-like elements in the evolution of gene structure and expression. Proc Natl Acad Sci USA 91:11792–11796

Willing RP, Mascarenhas JP (1984) Analysis of the complexity and diversity of mRNAs from pollen and shoots of *Tradescantia*. Plant Physiol 75:865–868

Wolfe KH, Sharp PM, Li W-H (1989) Rates of synonymous substitution in plant nuclear genes. J Mol Evol 29:208–211