Proc. Natl. Acad. Sci. USA
Vol. 91, pp. 11792–11796, December 1994
Genetics

Retrotransposons in the flanking regions of normal plant genes: A role for *copia*-like elements in the evolution of gene structure and expression

(retroelements/transposable elements/database searches/*Hopscotch*)

SHAWN E. WHITE*, LEDARE F. HABERA*, AND SUSAN R. WESSLER*†‡

Departments of *Botany and †Genetics, University of Georgia, Athens, GA 30602

Communicated by Charles S. Levings III, August 15, 1994 (received for review April 8, 1994)

ABSTRACT **The *wx-K* mutation results from the insertion of a *copia*-like retrotransposon into exon 12 of the maize waxy gene. This retrotransposon, named *Hopscotch*, has one long open reading frame encoding all of the domains required for transposition. Computer-assisted database searches using *Hopscotch* and other plant *copia*-like retroelements as query sequences have revealed that ancient, degenerate retrotransposon insertions are found in close proximity to 21 previously sequenced plant genes. The data suggest that these elements may be involved in gene duplication and the regulation of gene expression. Similar searches using the *Drosophila* retrotransposon *copia* did not reveal any retrotransposon-like sequences in the flanking regions of animal genes. These results, together with the recent finding that reverse-transcriptase sequences characteristic of *copia*-like elements are ubiquitous and diverse in plants, suggest that *copia*-like retrotransposons are an ancient component of plant genomes.**

The retroelement family is composed of transposable elements that move via an RNA intermediate (1). Included in this family are long-terminal-repeat (LTR) retrotransposons and retroviruses. Both LTR retrotransposons and retroviruses are flanked by LTRs that provide cis-regulatory sequences required for transcription of an RNA intermediate (2). The internal sequences of these elements encode proteins (Gag, protease, integrase, reverse transcriptase, and RNase H) necessary for reverse transcription and integration.

Based on the arrangement of their protein-coding domains, LTR retrotransposons can be subdivided into two groups named after the *Drosophila* retrotransposons *copia* and *gypsy* (2). The integrase domain is positioned 3′ of the reverse transcriptase domain in *gypsy*-like retrotransposons, and 5′ of reverse transcriptase in *copia*-like retrotransposons. Both groups have been found in fungi and plants in addition to *Drosophila* but have not been detected in animals other than insects and fish (3). *copia*-like reverse transcriptases have been identified in almost every plant species surveyed (4–6) and are diverse in terms of their amino acid sequences (7). Only a few plant retrotransposons, however, are responsible for recent mutations (Tnt*1* of tobacco and *Bs1*, *Stonor*, *B5*, and *G* of maize; refs. 8–10). Of these, Tnt*1* is the only retrotransposon shown to be complete and transcriptionally active in plants grown under normal conditions.

This paper presents the characterization of a second complete plant retrotransposon, *Hopscotch*.§ Use of this element in computer-based sequence similarity searches reveals that many normal plant genes have the remnants of *copia*-like retrotransposons in their upstream and downstream flanking regions. These results provide evidence that retroelements have the potential to be involved in the evolution of plant gene structure and expression by supplying genes with regulatory sequences and facilitating gene duplication. Furthermore, despite the fact that *copia*-like retrotransposons have been found in insects and fish, no element sequences were found in the flanking regions of normal animal genes.

MATERIALS AND METHODS

Cloning and Sequencing. Genomic DNA was isolated from maize seedlings homozygous for the *wx-K* mutation (11). *Sal* I fragments of 4.5–6 kb were cloned into λZAPII phage vector (Stratagene) and the resulting plaques were screened with a waxy (*wx*)-specific probe, *Sal*E (12). Both strands of a positive clone were sequenced with a Sequenase kit (United States Biochemical).

Database Searches. Computer-based amino acid similarity searches of the GenBank (version 77.0) and EMBL (version 34.0) databases were performed with the TFASTA search program of the University of Wisconsin Genetics Computer Group (GCG) software package (version 7.0) accessed through the BioScience Computing Resource at the University of Georgia. Conceptual translations of the sequences of the retrotransposons Tnt*1* of tobacco (accession no. X13777) (9), Ta*1–3* of *Arabidopsis* (X13291) (13), *PDR1* of pea (X66399) (14), Tst*1* of potato (X52387) (15), *copia* of *Drosophila* (X02599) (16), *BARE-1* of barley (Z17327) (17), and *Hopscotch* were used as query sequences. Nucleic acid-level searches of the GenBank and European Molecular Biology Laboratory databases were performed with the BLASTN (18) search program of the National Center for Biotechnology Information (Bethesda). Pairwise DNA sequence comparisons were made using the FASTA program of GCG. The GCG PILEUP and BOXSHADE programs were used to make the alignment figures. The alignments were edited to account for frameshifts.

RESULTS

The *wx-K* mutation of maize results from an ≈4.5-kb insertion in the *wx* gene (12). Analysis of the DNA sequence of this insertion revealed that it has the structure of an LTR retrotransposon. We have named this element *Hopscotch*. *Hopscotch* has identical 231-bp LTRs, is 4828 bp long, and has a single open reading frame of 4320 nt (1440 aa). A potential primer binding site with similarity (18/19 nt) to the 3′ end of wheat initiator methionine tRNA (GenBank accession no. V01383) is found adjacent to the 5′ LTR, and a polypurine tract lies next to the 3′ LTR.

Comparison of *Hopscotch* with Known Retrotransposons. The nucleotide and derived amino acid sequences of *Hop-*

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

‡To whom reprint requests should be addressed.

§The nucleotide sequence reported in this paper has been deposited in the GenBank data base (accession no. U12626).

scotch were compared with other transposable elements by computer searches of the GenBank and European Molecular Biology Laboratory databases. The searches identified significant nucleic acid (35–51%) and amino acid (23–32%) similarities between *Hopscotch* and several known *copia*-like retrotransposons. These comparisons also revealed that *Hopscotch* contains all of the amino acid domains (nucleic acid binding, protease, integrase, reverse transcriptase, and RNase H) that are found to be conserved among autonomously active retroelements (Fig. 1*A*). Both the amino acid conservation and the domain order serve to identify *Hopscotch* as a *copia*-like retrotransposon (Fig. 1*B*).

Retrotransposon-Like Sequences Flank Many Plant Genes. Surprisingly, these searches also revealed that 16 previously described plant genes have amino acid similarity to the conserved domains of *copia*-like retrotransposons (Figs. 2 and 3). Additional searches using the derived amino acid sequence of the retrotransposon Tnt*1* of tobacco as a query sequence detected 3 more plant genes with flanking regions similar to these domains (cotlea4a, cotdgala, and whtgermina) as well as 13 of the retrotransposon-like sequences identified in the *Hopscotch* searches. No additional plant genes were detected by using the amino acid sequences of other plant *copia*-like elements (Ta*1–3* of *Arabidopsis*, *PDR1* of pea, *BARE-1* of barley, and Tst*1* of potato) as query sequences.

Many of the retrotransposon-like sequences in the flanking regions of the genes probably represent ancient insertions. In several of the genes, retroelement similarity is degenerate, ends abruptly (cotmat5a, cucacc1, gmchs1, mzeg3pd, ricmtnad3a, phvarc1a, pschs1, zmpgalac, and zmpms2g), or contains internal deletions (cotmat5a, cucacc1, phvarc1a, and pschs1). In three cases, retrotransposon-like sequences are found in the same position in several members of a gene family, indicating that insertion predated gene duplication. Comparison of the 5′ end of the pea ribulose-bisphosphate carboxylase gene *rbcS-E9* with other members of the *rbcS* gene family revealed that two other *rbcS* genes (*rbcS-8.0* and *rbcS-3.6*) (37) have insertions at the same site in their upstream flanking regions. Similarly, several members of the maize 19-kDa zein gene family have elements inserted at the same position, as do members of the maize polygalacturonase gene (*PG*) family (zmpgalac, zmpgtnsg, zmpgg14) (38).

Nucleic acid-level searches using the plant retrotransposon sequences revealed two more genes with flanking regions similar to *copia*-like retrotransposons. The 3′ flanking region of the pea glyceraldehyde-3-phosphate dehydrogenase gene (*Gpb1*) (39) has 92% similarity to the *PDR1* LTR and probably represents part of an LTR from another copy of the *PDR1* retrotransposon. Likewise, the 5′ flanking sequence of a tomato gene expressed during pollen development (*LAT59*) (40) has 65% similarity to the LTR of Tnt*1*. Since Tnt*1* has been detected in the tomato genome by Southern blot analysis (9), the retrotransposon-like sequence in *LAT59* is probably the LTR of a Tnt*1*-related retrotransposon in tomato.

The amino acid- and nucleic acid-level searches combined identified 21 genes with flanking regions similar to *copia*-like retrotransposons. Retrotransposon similarity in 20 of these 21 genes had gone undetected until this study. Only the retrotransposon-like sequence at the 3′ end of the cotton 2S albumin storage-protein gene (*Mat5-A*) had been reported (22). The Tnt*1* amino acid searches detected another retrotransposon-like sequence in a 5′ flanking region of this gene previously described as repetitive (22).

Identifying Element LTRs. The flanking sequences of the genes identified in this study do not encompass complete retrotransposons, so the limits of the elements cannot be resolved by comparing LTRs. Although tRNA binding sites or polypurine tracts characteristically found immediately internal to the 5′ and 3′ LTRs, respectively, can be used to determine one end of an LTR, the other end is often unidentifiable. By comparing related elements or gene family members with and without insertions, however, we have been able to approximate the LTRs of several retrotransposon-like sequences. Alignment of the pea *rbcS-E9* gene with a closely-related pea *rbcS* gene (*rbcS-3A*) lacking the retrotransposon insertion allows the limits of the retrotransposon to be defined. The LTRs of the related elements adjacent to the cotton *Mat5a* and *Lea4a* genes and the pea *Chs1* and phenylalanine ammonia-lyase (*PAL2*) genes were defined by pairwise alignment of their nucleic acid sequences. The point at which retrotransposon sequence similarity ends was used to approximate one end of an LTR and the position of tRNA binding sites was used to define the other end. The ends of the retrotransposons in the 19-kDa zein genes were found by comparing the related upstream and downstream retrotransposons which share 90% sequence identity. The upstream element's LTR found in this manner corresponds precisely with the element end as determined by comparison with another member of the 19-kDa zein gene family lacking the insertions (*zein9*).

Retrotransposon-Like Sequences Are Not Found in *Drosophila* Genes. To identify genes harboring nearby retroelement sequences in organisms other than plants, the searches were repeated with the *Drosophila* retrotransposon *copia* as a query sequence. These searches failed to identify a single normal insect or other animal gene with *copia*-like retrotransposon sequences. They did, however, detect 15 of the plant sequences shown in Fig. 2.

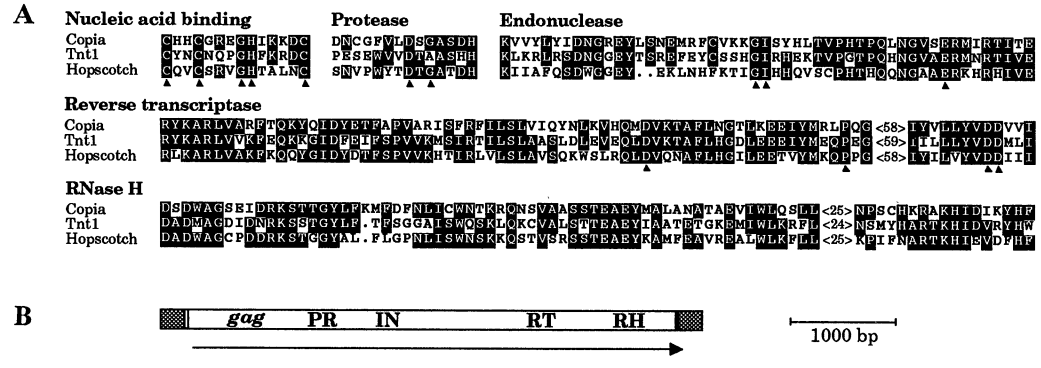

FIG. 1. (*A*) Amino acid similarity among the conserved domains of *Hopscotch*, *copia*, (16), and Tnt*1* (9). Amino acid residues invariant among retrotransposons and retroviruses (9) are indicated by triangles. (*B*) Structure of the *Hopscotch* retrotransposon. Stippled boxes represent LTRs. The *gag*, protease (PR), integrase (IN), reverse transcriptase (RT), and RNase H (RH) domains are indicated. The arrow represents the long open reading frame. A thin open bar represents a putative primer binding site with similarity to wheat initiator methionine tRNA, and a thin solid bar, a polypurine tract.

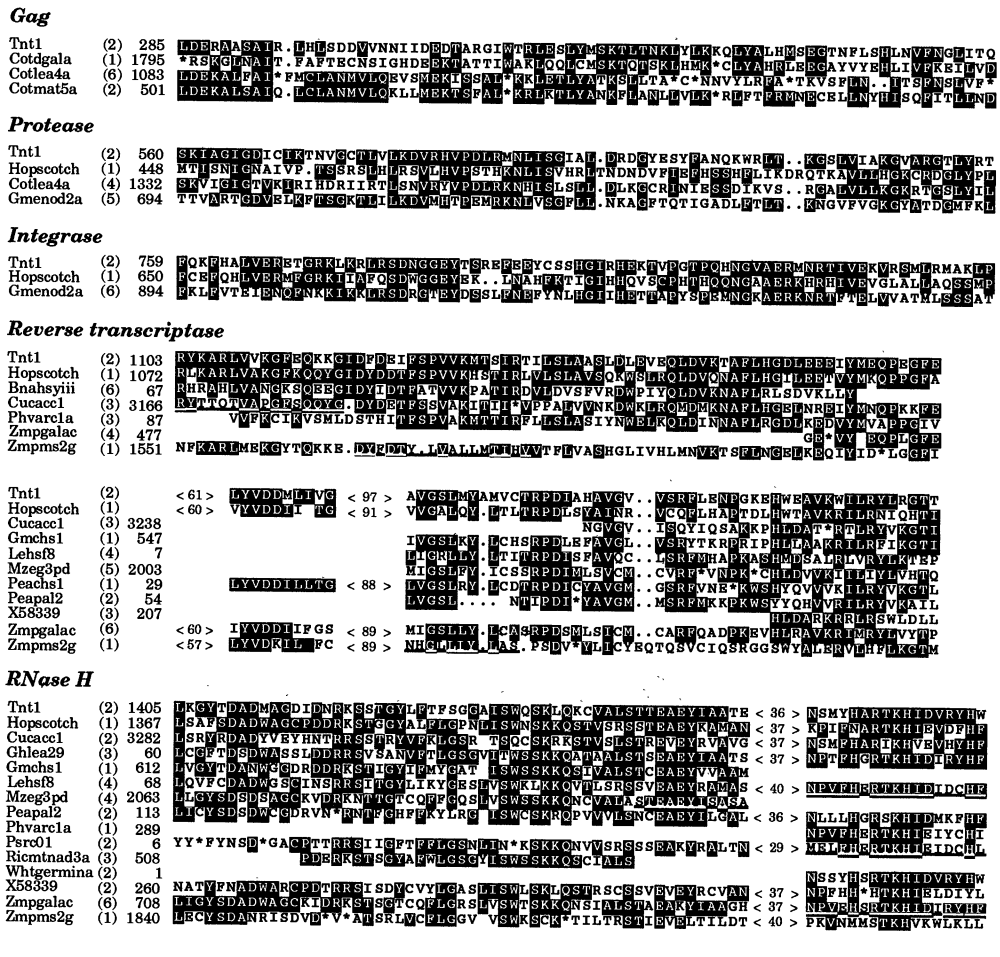

FIG. 2. Amino acid similarity among Tnt*1*, *Hopscotch*, and the retrotransposon insertions in the flanking regions of 19 plant genes. The locus names and amino acid positions refer to those used in GenBank. The reading frame of the conceptual translations used in the alignment is indicated in parentheses. Shifts in the reading frame are underlined and stars indicate stop codons. In most cases, the insertions do not include the most highly conserved nucleic acid binding site of Gag or the protease active site. Only the cotlea4a retrotransposon has similarity to the nucleic acid binding site (not shown). References: bnahsyiii (19), cotdgala (20), cotlea4a (21), cotmat5a (22), cucacc1 (23), ghlea29 (24), gmchs1 (25), gmenod2a (26), lehsf8 (accession no. X67599), mzeg3pd (27), peapal2 (28), phvarc1a (29), pschs1 (30), psrc01 (31), ricmtnad3 (32), whtgermina (33), x58339 (34), zmpgalac (35), zmpms2g (36).

DISCUSSION

Computer-assisted searches of the GenBank and European Molecular Biology Laboratory databases using three plant retrotransposons as query sequences revealed unexpected similarity to the flanking regions of 21 plant genes. The identity of these sequences as retrotransposon-like had been recognized previously in only 1 of these 21 genes. Although most regions of similarity appear to be the remnants of ancient insertions, we are confident of their retrotransposon origin for the following reasons: (*i*) the extent of amino acid similarity is striking (e.g., the ricmtnad3 element has 25 matches to either Tnt*1* or *Hopscotch* over 33 aa, and the element in the zmpms2g gene has 136 matches over 475 aa); (*ii*) the regions of amino acid similarity include conserved retrotransposon domains (i.e., Gag, protease, integrase, reverse transcriptase, or RNase H); (*iii*) several flanking sequences contain similarity to more than a single domain; and (*iv*) in many instances, other distinguishing structural features of retrotransposons such as LTRs, tRNA binding sites, and polypurine tracts can be identified. Despite the fact that retroelement similarity usually lies within 1 kb of the coding regions, 20 of the 21 published genes discussed in this paper are normal rather than mutant (the exception being the cotmat5a gene). Over 80 normal plant genes have been previously found to contain the inverted repeat transposable elements *Tourist* or *Stowaway* (41–43). Thus, the total number of plant genes harboring mobile elements or their remnants is >100.

The Retrotransposon-Like Sequences Contain Previously Identified Cis-Regulatory Elements. Four examples of ancient retroviral insertions that provide regulatory sequences to adjacent genes have been previously described. The mouse sex-limited protein gene is expressed in the presence of androgen due to a hormone-responsive enhancer in the LTR of an endogenous provirus (44). In humans, an upstream endogenous retroviral insertion has been found to be responsible for parotid-gland tissue specificity of the salivary amylase genes (45). Finally, the rat oncomodulin gene and the mouse IAP-promoted placental gene are under the control of promoters in solo LTRs of rodent intracisternal A particles (IAPs) (46, 47).

Several lines of evidence suggest that some of the retrotransposon-like sequences identified in this study may influence the expression of adjacent genes. The retrotransposon-like sequences in the maize polygalacturonase (*PG*) genes contain sequence motifs that are common among genes expressed during pollen development (35). In addition, a 501-bp fragment containing a positive regulatory region of a tomato gene expressed during pollen development (*LAT59*) (40) is composed entirely of a retrotransposon-like sequence.

The region upstream of nt −250 of the pea *rbcS-E9*, *-8.0*, and *-3.6* genes corresponds to a retrotransposon insertion that occurred prior to gene duplication. Another family member, *rbcS-3A*, lacks the insertion and, for this reason, has distinct sequences from nt −250 upstream to at least −410. Interestingly, the *rbcS-E9*, *-8.0*, and *-3.6* genes are coordinately expressed in a manner different from *rbcS-3A* (48). The combined expression of the *rbcS-E9*, *-8.0*, and *-3.6* genes in leaves is 30–50% lower than *rbcS-3A* gene expression, and their transcripts are underrepresented in pea petals and seeds when compared with *rbcS-3A* transcripts. On the basis of promoter domain-swapping experiments, the region upstream of −170 of the *rbcS-E9* gene (≈90% of which corresponds to the retrotransposon-like sequence) has been hypothesized to harbor a negative regulator of transcription (49).

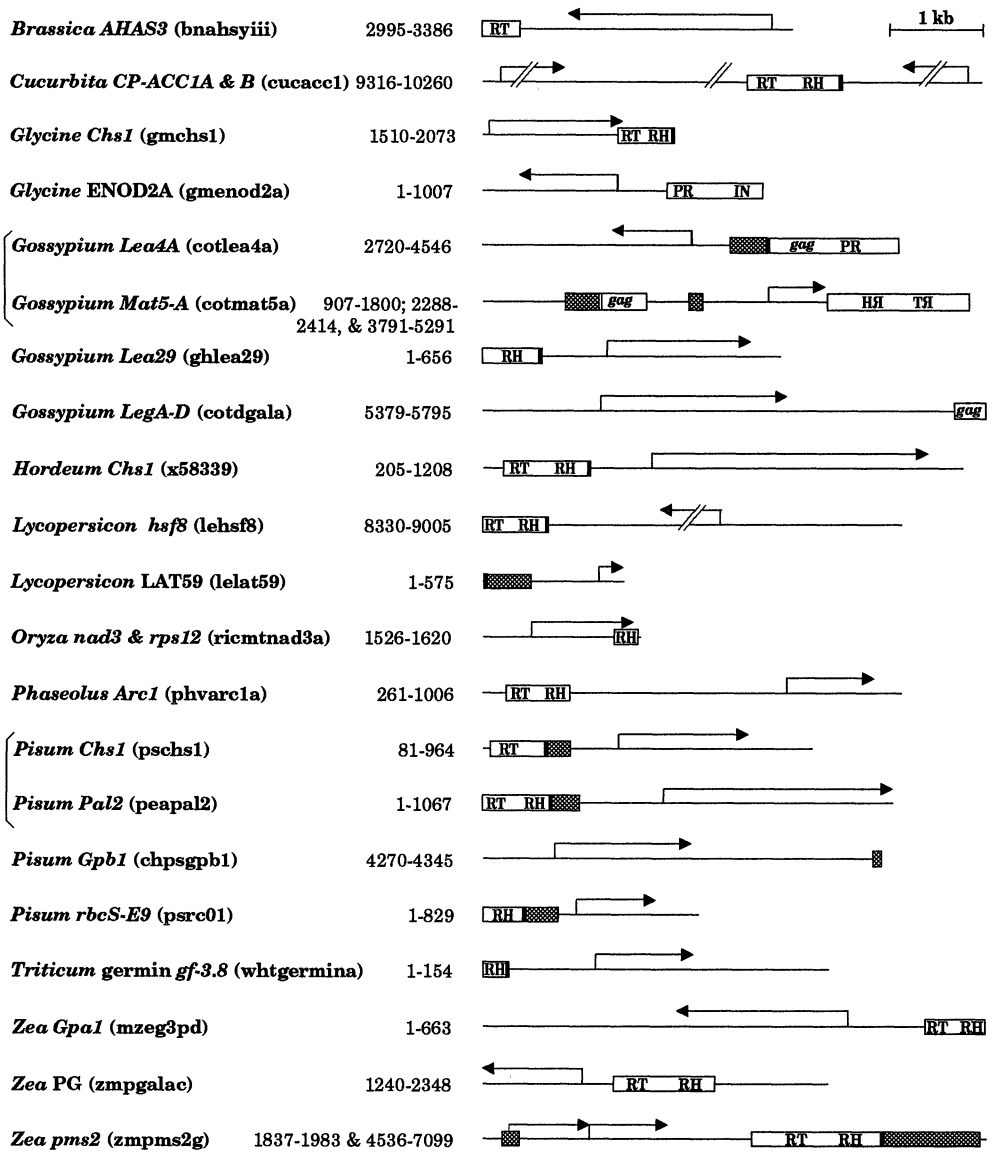

FIG. 3. Location and orientation of the retrotransposon-like sequences with respect to the direction of gene transcription. Retrotransposon-like sequences that are found at the same position in more than one member of a multigene family are represented by only one family member. The locus names of the genes as they appear in GenBank version 77.0 are in parentheses. The position of the retrotransposon-like sequences refers to the numbering used in GenBank for each locus, beginning at the 5′ end of the sequence with respect to the direction of gene transcription. Arrows indicate the direction and approximate extent of gene transcription. Stippled boxes, solid bars, and domain abbreviations are as described in Fig. 1. Open bars indicate regions of the cotlea4a and cotmat5a retrotransposons with similarity to wheat initiator methionine tRNA (14/15 and 13/15 nt, respectively). Brackets indicate retrotransposon-like sequences that appear to be members of the same element family.

The retrotransposon-like sequences flanking the 19-kDa zein genes of maize may have influenced both the expression and structure of this gene family. At least five of the seven sequenced 19-kDa zein gene family members (*pms1*, *pms2*, *ze19ba*, *zei19*, and *ze25*) have retrotransposon-like sequences at the same site in upstream flanking regions (Fig. 4). These sequences have 90% nucleic acid sequence identity to another element found in the downstream flanking regions of at least two members of this family (*pms1* and *pms2*). Many of the 19-kDa zein genes have two promoters, P1 and P2, with P1 accounting for ≈0.1% of zein gene transcripts (53). Our analysis indicates that P1 and the nearby start site of transcription are composed entirely of retrotransposon LTR sequences. The P1 promoter sequence has also been identified in the downstream flanking DNA of the *pms2* gene (36) and lies within the LTR of the downstream retrotransposon.

In addition to providing a zein promoter, the retrotransposon sequences may have facilitated the amplification of this gene family. Since the 19-kDa zein genes have been found clustered on the short arm of chromosome 7 (54), there is a possibility that they are tandemly arranged. In fact, *ze19ba* and *ze25* have been found in such a tandem arrangement (52). This organization suggests that the 19-kDa zein genes were duplicated by homologous, unequal crossing-over between retrotransposons inserted on either side of a progenitor gene.

Involvement of *copia*-Like Elements in Plant vs. Insect Evolution. Only plant gene sequences were identified as having significant similarity to either plant (Tnt*1*, *Hopscotch*) or *Drosophila* (*copia*) retrotransposons. This is surprising, since *copia*-like elements are highly expressed in *Drosophila* and have been shown to be the causative agent of many spontaneous mutations (55). In contrast, plant *copia*-like retrotransposons are transcribed at low levels under normal conditions and have been found to be responsible for only a few mutations.

The disparity between plant and *Drosophila* genes may reflect a lack of selection against retrotransposon insertions near plant genes. Alternatively, *copia*-like elements may be an older component of plant genomes and may have had a longer time frame for insertions into the flanking regions of genes to occur and become fixed. This hypothesis is consistent with the results of several recent surveys of retroelement reverse-transcriptase domains in plant genomes. These studies have revealed that reverse-transcriptase sequences characteristic of *copia*-like retrotransposons are heterogeneous and ubiquitous among plant species and were probably inherited by vertical transmission from a common ancestor (4–7). In contrast, analysis of both the codon usage of *copia* and its phylogenetic relationship to other retrotransposons has led to the hypothesis that *copia*-like elements were horizontally transmitted to *Drosophila* or one of its ancestors

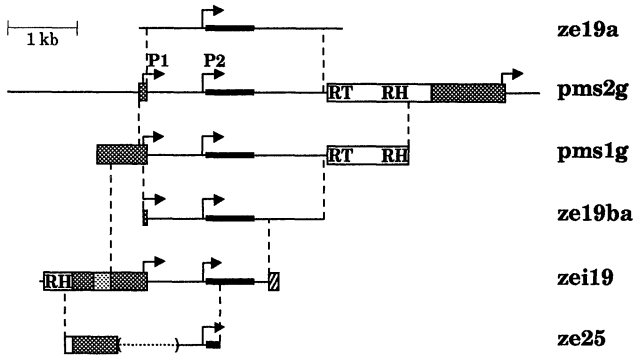

FIG. 4. The members of the 19-kDa zein gene family. Heavy black lines indicate zein coding regions, and arrows represent the two zein promoters, P1 and P2. Dark stippled boxes represent LTRs. Vertical dashed lines show the limits of >90% sequence similarity between the zein family members. Hatched box represents a previously described *CIN1* retrotransposon insertion (ref. 50 and T. E. Bureau, personal communication). Light stippled box represents a region in *zei19* that does not have similarity to LTR sequences. Horizontal dotted line in parentheses indicates a deletion in the *ze25* pseudogene. References: *ze19a* (51), *pms1* and *pms2* (36), *ze19ba* and *ze25* (52), *zei19* (53).

(56, 57). Therefore, *copia*-like insertions may have had a shorter time to become fixed in the flanking regions of insect genes than in their plant counterparts.

We are grateful to Lane Arthur, Thomas Bureau, Sylvestre Marillonnet, and Phillip Youngman for critical reading of the manuscript. Special thanks go to Michael Weise for his guidance in the use of the computer programs used in this study. This study was supported by a grant from the National Institutes of Health to S.R.W.

1. Hull, R. & Will, H. (1989) *Trends Genet.* **5**, 357–359.
2. Boeke, J. D. & Corces, V. G. (1989) *Annu. Rev. Microbiol.* **43**, 403–434.
3. Flavell, A. J. & Smith, D. B. (1992) *Mol. Gen. Genet.* **233**, 322–326.
4. Flavell, A. J., Dunbar, E., Anderson, R., Pearce, S. R., Hartley, R. & Kumar, A. (1992) *Nucleic Acids Res.* **20**, 3639–3644.
5. Voytas, D. F., Cummings, M. P., Konieczny, A. & Ausubel, F. M. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 7124–7128.
6. Hirochika, H. & Hirochika, R. (1993) *Jpn. J. Genet.* **68**, 35–46.
7. Flavell, A. J., Smith, D. B. & Kumar, A. (1992) *Mol. Gen. Genet.* **231**, 233–242.
8. Johns, M. A., Mottinger, J. & Freeling, M. (1985) *EMBO J.* **4**, 1093–1102.
9. Grandbastien, M.-A., Spielmann, A. & Caboche, M. (1989) *Nature (London)* **337**, 376–380.
10. Varagona, M. J., Purugganan, M. D. & Wessler, S. R. (1992) *Plant Cell* **4**, 811–820.
11. Dellaporta, S. L., Wood, J. & Hicks, J. B. (1983) *Plant Mol. Biol. Reporter* **1**, 19–22.
12. Wessler, S. R. & Varagona, M. J. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 4177–4181.
13. Voytas, D. F. & Ausubel, F. M. (1988) *Nature (London)* **336**, 242–244.
14. Lee, D., Ellis, T. H. N., Turner, L., Hellens, R. P. & Cleary, W. G. (1990) *Plant Mol. Biol.* **15**, 707–722.
15. Camirand, A. & Brisson, N. (1990) *Nucleic Acids Res.* **18**, 4929.
16. Emori, Y., Shiba, T., Kanaya, S., Inouye, S., Yuki, S. & Saigo, K. (1985) *Nature (London)* **315**, 773–776.
17. Mannienen, I. & Schulman, A. H. (1993) *Plant Mol. Biol.* **22**, 829–846.
18. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990) *J. Mol. Biol.* **215**, 403–410.
19. Rutledge, R. G., Quellet, T., Hattori, J. & Miki, B. L. (1991) *Mol. Gen. Genet.* **229**, 31–40.
20. Galau, G. A., Wang, H. Y.-C. & Hughes, D. W. (1991) *Plant Physiol.* **97**, 1268–1270.
21. Galau, G. A., Wang, H. Y.-C. & Hughes, D. W. (1992) *Plant Physiol.* **99**, 783–788.
22. Galau, G. A., Wang, H. Y.-C. & Hughes, D. W. (1992) *Plant Physiol.* **99**, 779–782.
23. Huang, P.-L., Parks, J. E., Rottmann, W. H. & Theologis, A. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 7021–7025.
24. Baker, J., Steele, C. & Dure, L. (1988) *Plant Mol. Biol.* **11**, 277–291.
25. Akada, S., Kung, S. D. & Dube, S. K. (1991) *Plant Mol. Biol.* **16**, 751–752.
26. Franssen, H. J., Thompson, D. V., Idler, K., Kormelink, R., van Kammen, A. & Bisseling, T. (1989) *Plant Mol. Biol.* **14**, 103–106.
27. Quigley, F., Martin, W. F. & Cerff, R. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 2672–2676.
28. Yamada, T., Tanaka, Y., Sriprasertsak, P., Kato, H., Hashimoto, T., Kawamata, S., Ichinose, Y., Kato, H., Shiraishi, T. & Oku, H. (1992) *Plant Cell Physiol.* **33**, 715–725.
29. Anthony, J. L., Vonder Haar, R. A. & Hall, T. C. (1991) *Plant Physiol.* **97**, 839–840.
30. An, C., Ichinose, Y., Yamada, T., Tanaka, Y., Shiraishi, T. & Oku, H. (1993) *Plant Mol. Biol.* **21**, 789–803.
31. Coruzzi, G., Broglie, R., Edwards, C. & Chua, N.-H. (1984) *EMBO J.* **3**, 1671–1679.
32. Suzuki, T., Kazama, S., Hirai, A., Akihama, T. & Kadowaki, K. (1991) *Curr. Genet.* **20**, 331–337.
33. Lane, B. G., Bernier, F., Dratewka-Kos, E., Shafai, R., Kennedy, T. D., Pyne, C., Munro, J. R., Vaughan, T., Walters, D. & Altomare, F. (1991) *J. Biol. Chem.* **266**, 10461–10469.
34. Rohde, W., Dörr, S., Salamini, F. & Becker, D. (1991) *Plant Mol. Biol.* **16**, 1103–1106.
35. Allen, R. L. & Lonsdale, D. M. (1993) *Plant J.* **3**, 261–271.
36. Quayle, T. J. A., Brown, J. W. S. & Feix, G. (1989) *Gene* **80**, 249–257.
37. Timko, M. P., Kausch, A. P., Hand, J. M., Cashmore, A. R., Herrera-Estrella, L., Van den Broeck, G. & Van Montagu, M. (1985) in *Molecular Biology of the Photosynthetic Apparatus*, eds. Steinback, K. E., Bonitz, S., Arntzen, C. J. & Bogorad, L. (Cold Spring Harbor Lab. Press, Plainview, NY), pp. 381–396.
38. Barakate, A., Martin, W., Quigley, F. & Mache, R. (1993) *J. Mol. Biol.* **229**, 797–801.
39. Liaud, M.-F., Zhang, D. X. & Cerff, R. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 8918–8922.
40. Twell, D., Yamaguchi, J., Wing, R. A., Ushiba, J. & McCormick, S. (1991) *Genes Dev.* **5**, 496–507.
41. Bureau, T. E. & Wessler, S. R. (1992) *Plant Cell* **4**, 1283–1294.
42. Bureau, T. E. & Wessler, S. R. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 1411–1415.
43. Bureau, T. E. & Wessler, S. R. (1994) *Plant Cell* **6**, 907–916.
44. Stavenhagen, J. B. & Robins, D. M. (1988) *Cell* **55**, 247–254.
45. Ting, C. N., Rosenberg, M. P., Snow, C. M., Samuelson, L. C. & Meisler, M. H. (1992) *Genes Dev.* **6**, 1457–1465.
46. Banville, D. & Boie, Y. (1989) *J. Mol. Biol.* **207**, 481–490.
47. Chang-yeh, A., Mold, D. E. & Huang, R. C. C. (1991) *Nucleic Acids Res.* **19**, 3667–3672.
48. Fluhr, R., Moses, P., Morelli, G., Coruzzi, G. & Chua, N.-H. (1986) *EMBO J.* **5**, 2063–2071.
49. Kuhlemeier, C., Fluhr, R. & Chua, N.-H. (1988) *Mol. Gen. Genet.* **212**, 405–411.
50. Hu, W., Das, O. P. & Messing, J. (1993) *Maize Genet. Coop. News Lett.* **67**, 94–95.
51. Pedersen, K., Devereux, J., Wilson, D. R., Sheldon, E. & Larkins, B. (1982) *Cell* **29**, 1015–1026.
52. Spena, A., Viotti, A. & Pirrotta, V. (1983) *J. Mol. Biol.* **169**, 799–811.
53. Kriz, A. L., Boston, R. S. & Larkins, B. A. (1987) *Mol. Gen. Genet.* **207**, 90–98.
54. Soave, C., Reggiani, R., Di Fonzo, N. & Salamini, F. (1981) *Genetics* **97**, 363–377.
55. Bingham, P. M. & Zachar, Z. (1989) in *Mobile DNA*, eds. Berg, D. E. & Howe, M. M. (Am. Soc. Microbiol., Washington, DC), pp. 485–502.
56. Mount, S. M. & Rubin, G. M. (1985) *Mol. Cell. Biol.* **5**, 1630–1638.
57. Konieczny, A., Voytas, D. F., Cummings, M. P. & Ausubel, F. M. (1991) *Genetics* **127**, 801–809.