UNIVERSIDAD TECNICA
FEDERICO SANTA MARIA
Centro de Biotecnología
Dr. Daniel Alkalay Lowitt

UNIVERSIDAD DE
TALCA

GOBIERNO DE CHILE
FUNDACIÓN PARA LA
INNOVACIÓN AGRARIA

Proyecto Programa de Formación

"Actualización de conocimientos, metodologías y tendencias en biología molecular de plantas"

Seminario de Difusión

"Reordenamiento genómico y promotores inducidos por estrés biotico abiótico"

Centro de Biotecnología Dr. Daniel Alkalay Lowitt
Universidad Técnica Federico Santa María

Valparaíso, 18 de Julio de 2003

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 (*ze19*).

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.

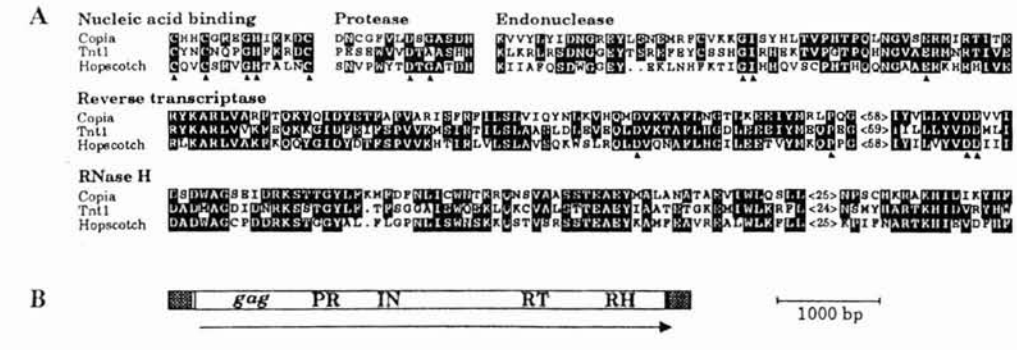


FIG. 1. (*A*) Amino acid similarity among the conserved domains of *Hopscotch*, *copia*, (16), and Tnt*1* (9). Amino acid residues invariant among retrotransposons and retroviruses (9) are indicated by triangles. (*B*) Structure of the *Hopscotch* retrotransposon. Stippled boxes represent LTRs. The *gag*, 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.

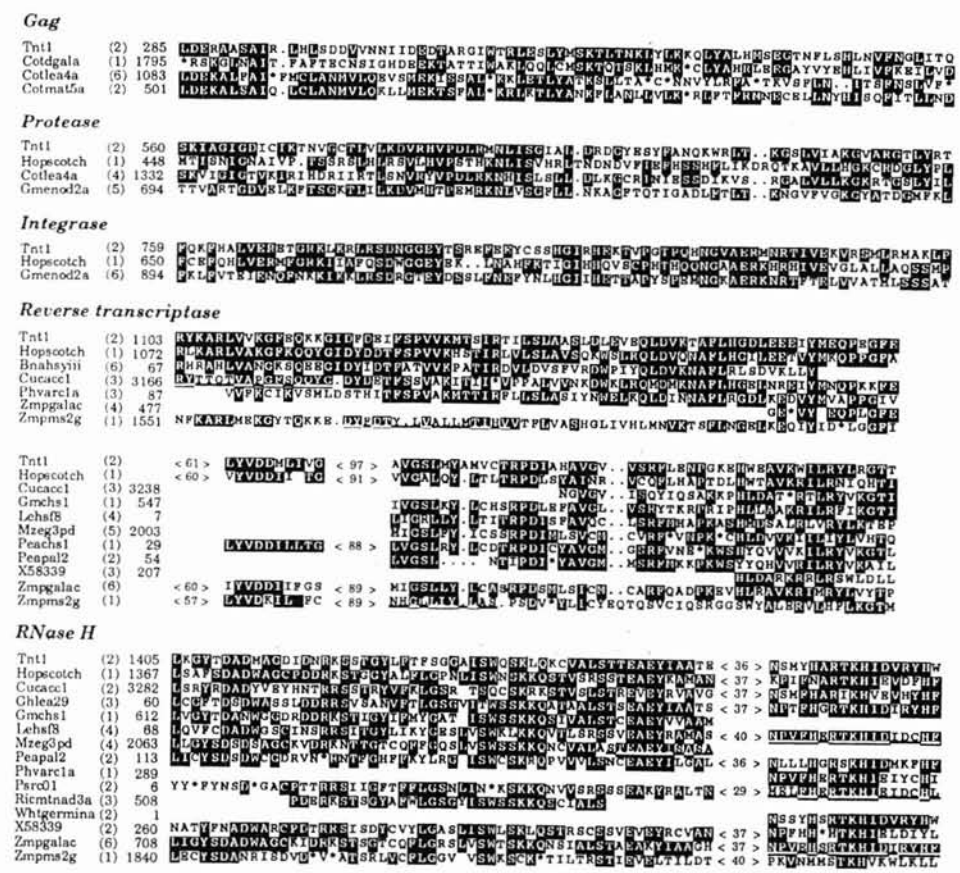


FIG. 2. Amino acid similarity among Tnt*1*, *Hopscotch*, and the retrotransposon insertions in the flanking regions of 19 plant genes. The locus names and amino acid positions refer to those used in GenBank. The reading frame of the conceptual translations used in the alignment is indicated in parentheses. 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).

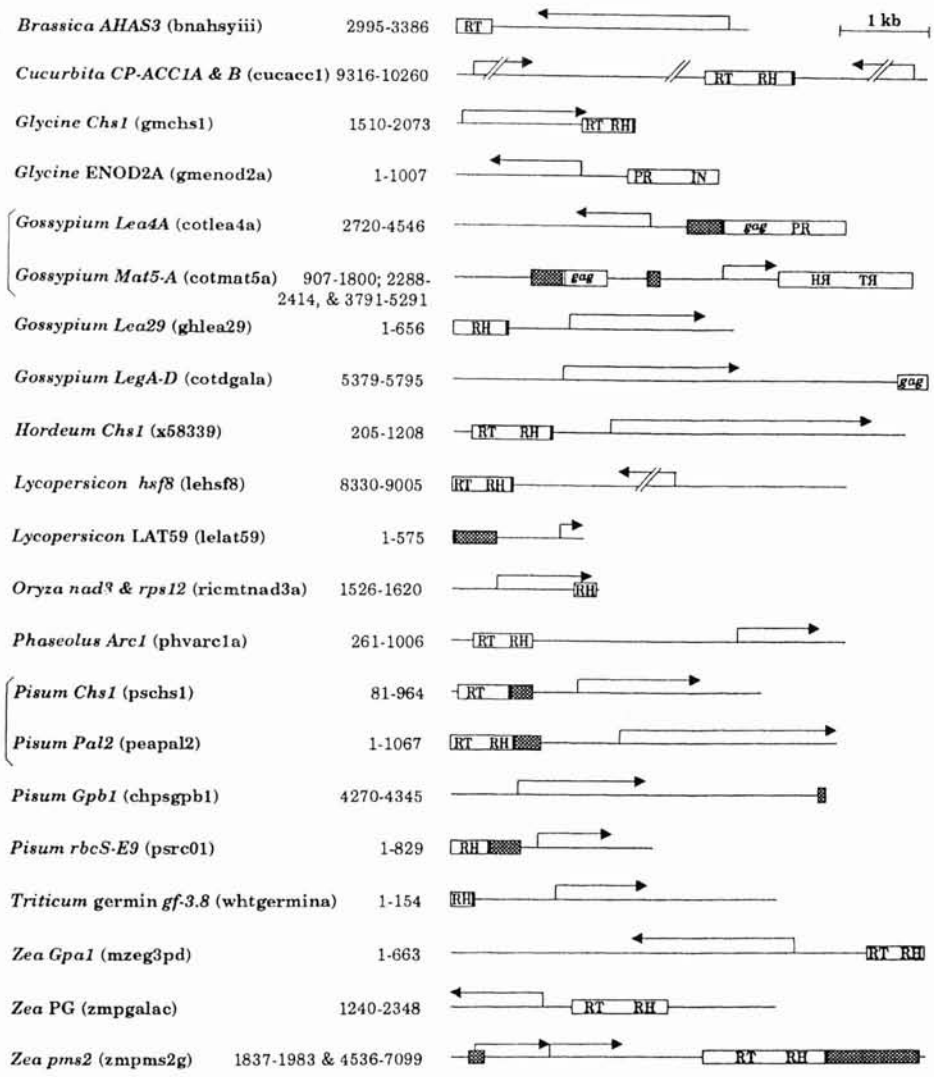


FIG. 3. Location and orientation of the retrotransposon-like sequences with respect to the direction of gene transcription. Retrotransposon-like sequences that are found at the same position in more than one member of a multigene family are represented by only one family member. The locus names of the 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

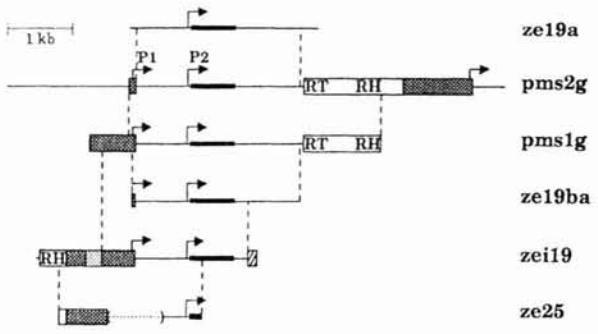


FIG. 4. The members of the 19-kDa zein gene family. Heavy black lines indicate zein coding regions, and arrows represent the two zein promoters, P1 and P2. Dark stippled boxes represent LTRs. Vertical dashed lines show the limits of >90% sequence similarity between the zein family members. Hatched box 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.

Colloquium

Maize as a model for the evolution of plant nuclear genomes

Brandon S. Gaut*, Maud Le Thierry d'Ennequin, Andrew S. Peek, and Mark C. Sawkins

Department of Ecology and Evolutionary Biology, University of California, Irvine, CA 92697-2525

The maize genome is replete with chromosomal duplications and repetitive DNA. The duplications resulted from an ancient polyploid event that occurred over 11 million years ago. Based on DNA sequence data, the polyploid event occurred after the divergence between sorghum and maize, and hence the polyploid event explains some of the difference in DNA content between these two species. Genomic rearrangement and diploidization followed the polyploid event. Most of the repetitive DNA in the maize genome is retrotransposable elements, and they comprise 50% of the genome. Retrotransposon multiplication has been relatively recent—within the last 5–6 million years—suggesting that the proliferation of retrotransposons has also contributed to differences in DNA content between sorghum and maize. There are still unanswered questions about repetitive DNA, including the distribution of repetitive DNA throughout the genome, the relative impacts of retrotransposons and chromosomal duplication in plant genome evolution, and the hypothesized correlation of duplication events with transposition. Population genetic processes also affect the evolution of genomes. We discuss how centromeric genes should, in theory, contain less genetic diversity than noncentromeric genes. In addition, studies of diversity in the wild relatives of maize indicate that different genes have different histories and also show that domestication and intensive breeding have had heterogeneous effects on genetic diversity across genes.

Genomic technologies have produced a wealth of data on the organization and structure of genomes. These data range from extensive marker-based genetic maps to "chromosome paintings" based on fluorescent *in situ* hybridization to complete genomic DNA sequences. Although genomic approaches have changed the amount and type of data, the challenges of interpreting genomic data in an evolutionary context have changed little from the challenges faced by Stebbins (1) and the coauthors of the evolutionary synthesis. The challenges are to infer the mechanisms of evolution and to construct a comprehensive picture of evolutionary change.

In this paper, we will focus on the processes that contribute to the evolution of plant nuclear genomes by using maize (*Zea mays*) as a model system. In some respects, it is premature to discuss the evolution of plant genomes, because the pending completion of the *Arabidopsis* (*Arabidopsis thaliana*) genome, with rice (*Oryza sativa*) following, is sure to unlock many mysteries about plant genome evolution. However, it must be remembered that *Arabidopsis* and rice are being sequenced, precisely because their genomes are atypically small and streamlined. Even after these genomes are sequenced, it will still be a tremendous challenge to understand the evolution of plant nuclear genomes, like the maize genome, for which entire DNA sequences will not be readily available.

Maize is a member of the grass family (Poaceae). The grasses represent a range of genome size and structural complexity, with rice on one extreme. A diploid with 12 chromosomes ($2n = 24$), rice has one of the smallest plant genomes, with only 0.9 pg of DNA per 2C nucleus (Fig. 1). Other grass species exhibit far larger genomes. Wheat, for example, is a hexaploid with 21 chromosomes ($2n = 42$) and a haploid DNA content of 33.1 pg (2). Genera like *Saccharum* (sugarcane) and *Festuca* are even more complicated, displaying wide variation in ploidy level and over 100 chromosomes in some species. As a diploid with 10 chromosomes ($2n = 20$) and a 2C genome content roughly 6-fold larger than rice, maize lies somewhere in the middle of grass genome size and structural complexity (Fig. 1).

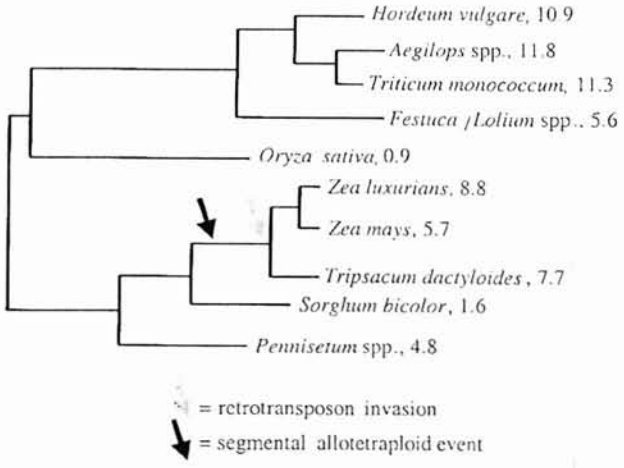


Fig. 1. A phylogeny of diploid grass species. Numerical values next to species names represent the 2C genome content of the species, measured in picograms. The phylogeny and genome content information is taken from figure 1 of ref. 51. The arrows represent the hypothesized timing of evolutionary events.

This paper focuses on the impact of chromosomal duplication, transposition, and nucleotide substitution on the evolution of the maize genome. We will discuss chromosomal duplication and transposition separately and will pay particular attention to their effects on DNA content. Nucleotide substitution will be discussed in the context of genetic diversity. Patterns of genetic diversity provide insight into the population genetic processes that act on different regions of the genome and thus uncover the evolutionary forces that act on genomes. We focus on maize throughout the paper but also generalize to other species when appropriate.

Polyploidy and Chromosomal Duplication

An Ancient Polyploid Origin. The first hints of the complex organization of the maize genome came from cytological studies.

This paper was presented at the National Academy of Sciences colloquium "Variation and Evolution in Plants and Microorganisms: Toward a New Synthesis 50 Years After Stebbins," held January 27–29, 2000, at the Arnold and Mabel Beckman Center in Irvine, CA.

Abbreviations: mya, million years ago; LTR, long terminal repeat.

*To whom reprint requests should be addressed. E-mail: bgaut@uci.edu.

Although maize is diploid, early studies by McClintock (3, 4) demonstrated the association of nonhomologous chromosomes during meiosis. Later studies documented the formation of bivalents and multivalents in maize haploids (5, 6). Altogether, cytological observations suggested that the maize genome contains extensive regions of homology, probably reflecting chromosomal duplications.

Evidence for chromosomal duplication also came from linkage information. In 1951, Rhoades (7, 8) noted that some regions of linkage maps did not contain mutants, and he proposed that the lack of mutants reflected genetic redundancy caused by chromosomal duplication. Rhoades' proposal has since been supported by molecular data. For example, isozyme studies have documented the presence of duplicated, linked loci in maize (9–12), and restriction fragment length polymorphism mapping studies have shown that many markers map to two or more chromosomal locations (13, 14). These mapping studies have established that some chromosomes—e.g., chromosomes 1 and 5 and chromosomes 2 and 7—share duplicated segments. Perhaps the most surprising information about the extent of gene duplication in maize is that 72% of single-copy rice genes are duplicated in maize (15).

Extensive chromosomal duplication in maize has been interpreted as evidence for a polyploid origin of the genome (7, 16), but until recently, there had been no estimation of the timing and mode of this polyploid event. In 1997, Gaut and Doebley (17) inferred the timing and mode of the polyploid event by studying DNA sequences from maize duplicated genes. To infer the mode of origin, Gaut and Doebley first modeled patterns of genetic divergence under three different types of polyploid formation: autopolyploidy, genomic allopolyploidy, and segmental allopolyploidy. (Briefly, allopolyploids are created by hybridization between species, with a genomic allopolyploid based on species that have fully differentiated chromosomes and a segmental allopolyploid based on species that have only partially differentiated chromosomes. Autopolyploidy refers to a polyploid event based on an intraspecific event. Stebbins contributed a great deal toward the definition and use of these terms, and precise definitions can be found in ref. 1.) The models' predictions were then compared with patterns of DNA sequence divergence in 14 pairs of maize duplicated genes. The sequence data were consistent with a segmental allotetraploid model of origin but inconsistent with the other two models of polyploid formation. Hence, the authors concluded that the maize genome was the product of a segmental allotetraploid event. They estimated the timing of the event by applying a molecular clock to the sequence data.

The hypothesized origin of the maize genome is detailed in Fig. 2 (17). Briefly, this hypothesis states that (*i*) maize is the product of a segmental allotetraploid event, (*ii*) the two diploid progenitors (or "parents") of maize diverged ≈20.5 mya, (*iii*) the tetraploid event occurred between 16.5 and 11.4 mya, sometime after the divergence of *Sorghum* from one of the progenitor lineages, and (*iv*) the genome "rediploidized" before 11.4 mya. Although valuable, there are at least three reasons to be cautious about the hypothesis. The first reason is that the hypothesis is based on a relatively small number of DNA sequences—i.e., only 14 pairs of duplicated sequences. The second reason is that some of the sequences were not mapped to a chromosomal location. Ideally, these analyses should be based on a far greater number of sequences, all of which are known to reside in regions of known chromosomal duplication. Finally, it was not possible to test molecular clock assumptions rigorously for all of the sequence data, and thus some of the clock-based time estimates are subject to an unknown amount of error. Despite the need for caution, the study of Gaut and Doebley (17) provides the first glimpse into the mode and timing of an ancient plant polyploid event, and it also proposes a hypothesis that is testable with additional data.

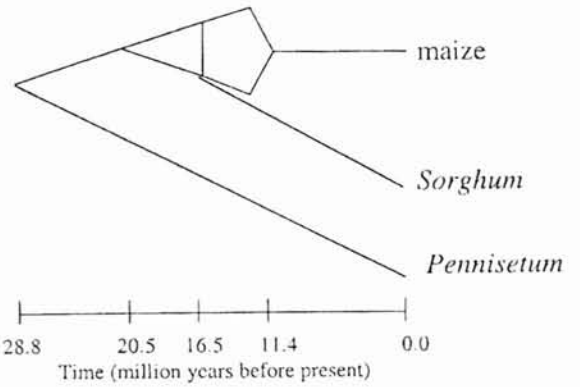


Fig. 2. A hypothesis for the origin of the maize genome (17). Under this hypothesis, *Pennisetum* and maize diverged ≈29 million years ago (mya), followed ≈9 million years later by the divergence of the two diploid progenitors of maize. Sorghum diverged from one of these progenitor lineages (≈16.5 mya) before the two diploid progenitors united to form allopolyploid maize. The polyploid event occurred sometime between 16.5 mya and 11.4 mya, with subsequent diploidization completed by 11.4 mya. Gray shading represents the period in which allotetraploidy and diploidization occurred.

The Polyploid Event and the Divergence of Maize and Sorghum. Fig. 1 places the segmental allotetraploid event in a phylogenetic context, and this context raises three important points about the comparison of maize to sorghum. First, if the allotetraploid event occurred after maize and sorghum diverged, then the maize genome should be duplicated more extensively than the sorghum genome. A corollary prediction is that maize and sorghum should not share common chromosomal duplications. Ultimately, these predictions can be tested with comparative genetic maps. At this point, however, it is unclear from comparative genetic maps as to whether the two genomes share extensive duplications in common, largely because published sorghum maps lack sufficient coverage (18–21). However, mapping information indicates that a higher proportion of markers is duplicated in maize than in sorghum. For example, Pereira *et al.* (19) found that 44% of restriction fragment length polymorphism markers detected more bands in maize than in sorghum; conversely, only 7% of markers detected more bands in sorghum than in maize. This information is consistent with the phylogenetic placement of the allotetraploid event (Fig. 1).

The second point centers on chromosome number. Maize and sorghum (*Sorghum bicolor*) have the same number of chromosomes ($2n = 20$). If maize underwent an allotetraploid event after the divergence of maize from sorghum, why do these plants have an identical number of chromosomes? At present, there is no suitable answer to this question, but there has been discussion about the evolution of chromosome number. Traditionally, it has been assumed that the basal haploid chromosome number of the tribe Andropogoneae, which encompasses maize, sorghum, and *Tripsacum*, was $n = 5$ (22, 23). More recently, it has been suggested that the basal haploid chromosome of the tribe was $n = 10$ (24). If the basal number was 10, one can hypothesize both that the chromosome number of *S. bicolor* has remained unchanged and that maize was the product of an allopolyploid event between two species with a reduced number of chromosomes ($n = 5$). This scenario is plausible, because the tribe contains diploid taxa with $n = 5$ (e.g., *Elionurus* and *Sorghum* species; ref. 24) and because comparative maps provide support that maize consists of two $n = 5$ subgenomes (25, 26).

Wilson *et al.* (27) have asserted that maize came from an

ancestor with neither 5 nor 10 chromosomes. Based on genetic map data, they argued that the chromosome number of maize before the allotetraploid event was $n = 8$. The chromosome number was doubled subsequently to $n = 16$ ($2n = 32$) during the maize allotetraploid event and then reduced further by diploidization and fusion to the current number ($n = 10$; $2n = 20$). Unfortunately, however, the argument of Wilson *et al.* contains errors regarding the timing and phylogenetic context of the allotetraploid event. For example, they suggest that the allotetraploid event occurred after the divergence of maize and *Tripsacum*, whereas most evidence suggests that the allotetraploid event occurred *before* the divergence of maize and *Tripsacum*. When these errors are taken into account, their arguments for the evolution of chromosome number seem unlikely. In short, there are no definitive answers either as to the evolution of chromosome number in this group or as to why *S. bicolor* and maize have the same number of chromosomes.

The third and final point about maize and sorghum centers on the difference in genome content between the two species. The segmental allotetraploid event predicts 2-fold variation in DNA content between sorghum and maize, but it does not account for the actual 3.5-fold variation in DNA content (Fig. 1). Based on this information, differences in DNA content probably reflect the allopolyploid event *and* additional evolutionary changes, such as the accumulation of repetitive DNA.

Genome Rearrangement After an Allopolyploid Event. It must be remembered that extant maize is a diploid, and thus the segmental allotetraploid hypothesis presumes that the maize genome rearranged and diploidized. Is this presumption reasonable? Is genome rearrangement common after allopolyploid events?

Thus far, studies of synthetic plant polyploids suggest that genomes rearrange rapidly after allopolyploid events (reviewed in ref. 28). In one study, Song *et al.* (29) created four synthetic allopolyploids. After recovery of F_2 polyploids, each line was selfed until the F_5 generation. Plants from the F_2 and each subsequent generation were subjected to Southern hybridization with a panel of 89 probes. Southern blotting revealed remarkable differences in fragment profiles from generation to generation. In one synthetic polyploid, 66% of the probes detected fragment loss, fragment gain, or a change in fragment size, demonstrating that extensive rearrangement can occur rapidly after allopolyploid formation. Feldman and coworkers (30–32) performed similar studies in *Triticum* and *Aegilops*. Their results suggest that allopolyploids lose noncoding sequences in a directed, nonrandom fashion and that coding sequences are modified extensively (30–32).

Empirical studies detect rapid rearrangement of allopolyploid genomes, but rapid rearrangement is not equivalent to a complete diploidization. However, there is growing evidence that many plant, animal, and fungal genomes are the products of ancient polyploid events that were followed by rearrangement and a reduction in ploidy level. Yeast is one example. The DNA sequence of the yeast genome contains numerous blocks of duplicated genes. The phase (or direction) of the blocks are nonrandomly associated with centromeres, suggesting that the blocks were produced by the process of chromosomal duplication (33). Altogether, the data suggest that the yeast genome is the product of an ancient tetraploid event followed by rearrangement and diploidization (34). Vertebrates are another example of diploidized ancient polyploids; it is believed that vertebrates are degenerate polyploids owing to two polyploid events before the radiation of fish and mammals (35). Similar examples come from plants; for example, both *Glycine* (soybean) (36) and *Brassica* species (37, 38) seem to be degenerate polyploids. Based on this information, one can conclude that diploidization after polyploidy is evolutionarily common.

Table 1. Duplicated chromosomes in maize and the studies that identified them

Duplicated chromosomes	Reference nos.
1–5	14, 27, 84
1–9	14, 27, 84
2–4	14
2–7	14, 27, 84
2–10	14, 15, 27, 84
3–8	14, 15, 27, 84
3–10	84
4–5	27, 84
6–8	14, 27, 84
6–9	27, 84

For maize, it should be possible to garner insights into the processes of rearrangement and diploidization from extant patterns of chromosomal duplication. Mapping studies have documented regions of chromosomal duplication in maize (Table 1). (It is important to note that Table 1 includes *only* those chromosomes that were explicitly defined as duplicated by the authors; Table 1 does *not* include all of the chromosome pairs on which markers are known to crosshybridize.) As Table 1 demonstrates, there is some disagreement among studies about chromosomal duplications, for two reasons. First, different studies use different data, leading to different conclusions. Second, and perhaps more importantly, researchers rarely denote their criteria for defining chromosomal duplications, and thus criteria likely differ among studies. Ultimately, chromosomal duplications should be defined by objective statistical criteria.

Nonetheless, there is a consensus about some chromosomal pairs. For example, it is now well established that portions of chromosome 1 are duplicated on chromosomes 5 and 9 (Table 1). The evolutionary implication for these pairings is that the process of diploidization rearranged one copy of chromosome 1. (Alternatively, chromosome 1 could be an amalgamation of regions from different parental chromosomes.) Chromosome 2 had a similar fate in that portions of chromosome 2 are also found on chromosomes 7, 10, and perhaps 4 (Table 1). More extensive evaluation of these duplications will provide an indication as to whether there has been any bias in rearrangements. For example, there is a strong bias for paracentric inversions, as opposed to translocations and pericentric inversions, between potato and tomato. It was reasoned that the bias toward paracentric inversions reflects the relatively low effect of paracentric inversions on fitness (39). Additional studies of chromosomal duplications in maize could provide additional insights into the kind of rearrangements that are most evolutionarily stable.

The Importance of Chromosomal Duplication in Genome Evolution. Is maize typical with regard to its polyploid history and prevalent chromosomal duplication? There is no doubt that polyploidy is common in plants, with up to 70% of angiosperms owing their history to polyploidy (1, 40). Furthermore, genetic maps demonstrate that a great number of species contain chromosomal duplications. Even species with streamlined genomes contain chromosomal duplications; for example, rice has a large duplication between chromosomes 11 and 12 (41) and *Arabidopsis* also has at least one large chromosomal duplication (42). Other plant genomes with chromosomal duplications include sorghum (21), cotton (43), soybean (36), and *Brassica* species (37, 38). Some of these genomes are degenerate polyploids like maize, but others may owe their chromosomal duplications to independent segmental events.

It is important to note that chromosomal duplications are

usually inferred from genetic maps, but most (if not all) genetic maps are based on low copy-number markers. Low copy-number markers are systematically biased against detecting duplicated chromosomal segments, and hence the extent of chromosomal duplication is likely grossly underestimated for most plant taxa. In addition, the resolution of most genetic maps is low, such that relatively small areas of chromosomal duplication cannot be detected. The result is that we do not have a realistic understanding of either the extent to which chromosomes are duplicated or the extent to which genomes contain functional redundancies. We can, however, look to *Arabidopsis* sequence data as preliminary examples of the extent of chromosomal duplication. Based on the sequences of chromosomes 2 and 4 (42, 44), it is estimated that 10–20% of the low-copy regions of the *Arabidopsis* genome lie within duplicated chromosomal regions (42). Given that the *Arabidopsis* genome is streamlined, this percentage is undoubtedly much higher in complex genomes. It is possible that most genes in most plant genomes reside in duplicated chromosomal regions.

Multiplication of Repeat Sequences

Extent and Identification of Repetitive DNA. Repetitive DNA constitutes a high proportion of plant genomes. This fact has been confirmed experimentally by reassociation (or C_0t) kinetics. For example, Flavell *et al.* (45) found that repetitive DNA (defined, in this case, as DNA with more than 100 copies per genome) constitutes ≈80% of genomes with a haploid DNA content >5 pg. In contrast, small genomes of <5 pg contain 62% repetitive DNA on average. Maize falls into this range; reassociation experiments indicate that the genome contains from 60% to 80% repetitive DNA (45, 46). The repetitive DNA of maize can be categorized further as 20% highly repetitive (over 800,000 copies per genome) and 40% middle repetitive (over 1,000 copies per genome; ref. 46).

It is obvious that repetitive DNA is a large component of the maize genome, and thus the proliferation of repeat sequences has had important evolutionary implications. However, reassociation studies alone cannot answer two important questions about repetitive DNA in maize: what is the repetitive DNA, and when did it arise?

To date, the most complete answers to these two questions come from studies of the maize *Adh1* region by Bennetzen and coworkers (47–50). They isolated a 280-kilobase yeast artificial chromosome clone of the *Adh1* region and characterized the composition of the repetitive intergenic DNA. Retrotransposons comprise roughly 62% of the 240 kilobases analyzed, with an additional 6% of the clone consisting of miniature inverted-repeat transposable elements, remnants of DNA transposons, and other low-copy repeats. In total, the region contained 23 retrotransposons representing 10 distinct families. Of the 23 retroelements, 10 inserted within another element, resulting in a nested or "layered" structure of intergenic DNA within maize (Fig. 3). The architecture of this region suggests that retrotransposons preferentially target other retroelements for insertion.

Perhaps the most interesting feature of the *Adh1* region is that it seems to be a representative region of the maize genome. Three observations support this contention. First, Southern blot and other analyses suggest that the retrotransposon families in the *Adh1* region comprise at least 50% of the maize genome; altogether, just three of the retroelement families found in the *Adh1* region constitute a full 25% of the genome (48). Second, 85% of repetitive DNAs from other regions were also present in the *Adh1* region (although it should be noted that the sample of repetitive DNAs from other regions was small and thus this estimate may not be robust). Finally, a more recent study suggests that retrotransposons hybridize fairly uniformly to maize bacterial artificial chromosome clones, suggesting that the

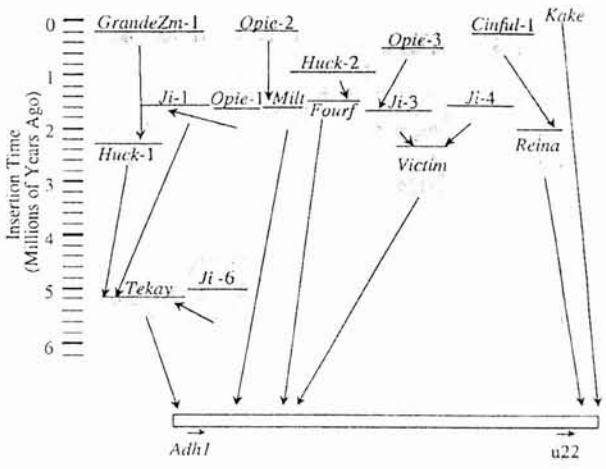


Fig. 3. The estimated insertion times of retrotransposons in the *Adh1* region (49). Each gray box represents a retrotransposon. The horizontal line through the box is the estimate of insertion time, and the height of the box represents the standard deviation of the estimate. Arrows between boxes indicate the order of insertion. For example, *Huck-2* inserted into *Fourf* ≈1 mya.

distribution of retrotransposons is reasonably homogeneous throughout the genome (B. Meyers, personal communication).

The Timing of Retrotransposon Multiplication. Maize repetitive DNA seems to be primarily retrotransposons, but the second question remains: when did these retroelements multiply? To answer this question, SanMiguel *et al.* (49) sequenced the long terminal repeat (LTR) of retrotransposons in the *Adh1* region. The rationale was as follows: when a single retrotransposon inserts into genomic DNA, both copies of the LTR are identical. Over time, the LTRs accumulate nucleotide substitutions and diverge in sequence. If the accumulation of nucleotide substitutions occurs at a regular pace, the number of nucleotide differences between the two LTRs provide insight into the date of LTR divergence and hence the date of retrotransposon insertion.

SanMiguel *et al.* (49) applied this approach to estimate the insertion time for 17 LTRs from the *Adh1* region (Fig. 3). The results show that the oldest retrotransposon insertion is ≈5.2 mya and that most (15 of 17) retrotransposons inserted within the last 3.0 million years. The question arises as to whether these time estimates are reasonable. One feature that supports the results is that the time estimates correspond to the layering of retrotransposons (Fig. 3). In other words, in most cases (10 of 11) the insertion date for a retrotransposon is less than the insertion date for the retrotransposon into which it inserted. (The one exception is an instance in which the insertion dates are statistically indistinguishable.) Another observation that supports these results is that the sorghum *Adh1* region lacks retrotransposons (50). Based on this information and ignoring the possibility of extensive retrotransposon loss in sorghum (51), retrotransposons in the maize *Adh1* region must have amassed in the ≈16 million years since the divergence of sorghum and maize.

The implications of the study are important. If the *Adh1* region is representative and the retrotransposons in this region constitute 50% of the genome, the maize genome has doubled in size in the last 5–6 million years. Like the polyploid event, retrotransposon proliferation represents a doubling of genome content over a relatively short evolutionary time scale.

Fig. 1 indicates that retrotransposon multiplication likely began in the evolutionary lineage leading to maize and *Tripsacum*, which diverged roughly ≈4.5–4.8 mya (52). Thus, most

maize retrotransposon activity postdates the divergence of genera, but the oldest retrotransposons in the maize *Adh1* region likely predate the split between *Zea* and *Tripsacum*. This discussion underscores the importance of studying *Tripsacum* to understand evolutionary events in maize better; if Fig. 1 is accurate, *Tripsacum* should share both chromosomal duplications and some retrotransposon activity in common with maize. It is known that *Zea* and *Tripsacum* share at least one low-copy retrotransposon that is absent from other closely related genera (53), but there is generally little information about chromosomal duplications or retrotransposons in *Tripsacum*.

Based on the available information, two large events differentiate the maize lineage from the sorghum lineage. The first event, segmental allotetraploidy, resulted in a 2-fold increase in maize DNA content. The second event, retrotransposon proliferation, produced another 2-fold increase in maize DNA content. Together, these events adequately explain the 3.5-fold difference in DNA content between maize and sorghum. However, it should be noted that there is also substantial variation in genomic DNA content among *Zea* and *Tripsacum* species (Fig. 1) (2, 54); this variation may reflect different amounts of retrotransposon proliferation or independent chromosomal duplications.

Remaining Questions. Studies of the *Adh1* region by Bennetzen and coworkers (47–50) have provided invaluable insight into the structure and dynamics of maize intergenic DNA, but at least three important questions remain.

Question 1. Are retrotransposons distributed homogeneously among genomic regions? The *Adh1* studies, as well as other studies (B. Meyers, personal communication), suggest that retrotransposon distribution may be roughly homogenous among regions of the maize genome. However, other lines of evidence suggest that such homogeneity is unlikely. For example, evolutionary theory predicts that transposable elements should gather in regions of low recombination, such as centromeres (55, 56). This prediction holds in *Arabidopsis*, where sequence data from chromosomes 2 and 4 indicate an increase in the frequency of transposable elements near centromeres (57).

There are other reasons to suggest that retrotransposon distribution may not be homogeneous throughout the maize genome. One obvious reason is that there are heterogeneities in chromosomal structure, such as euchromatin, heterochromatin, nucleolus organizing regions, telomeres, centromeres, and knobs. Nonetheless, recent research indicates that retrotransposons constitute a substantial fraction of both heterochromatic centromeres and heterochromatic knobs (58, 59); for one chromosome 9 knob, retroelements comprise roughly one-third of knob-specific clones (60). Many of the retrotransposons in knob and centromeric DNA belong to the element families found in the *Adh1* region. Despite these commonalties, there are also substantive differences among knobs, centromeres, and the *Adh1* region. For example, centromeres contain a centromere-specific retrotransposon (CentA; ref. 59). Similarly, chromosomal knobs associate with 180-bp and 350-bp repeat elements that are otherwise sparse in the genome (58). Altogether, the emerging picture is one in which some retroelement families are fairly ubiquitous, and other repetitive DNAs are heterogeneous in their distribution (e.g., ref. 61).

The work of Bernardi and coworkers (62, 63) is an intriguing addition to this picture. They fractionated DNA by G:C content and hybridized each G:C fraction to 38 coding-region probes. The coding genes hybridize almost exclusively to a DNA fraction of very narrow G:C content (1% of the total range), and this narrow fraction corresponds to 17% of the DNA content of the genome. To explain this hybridization pattern, Bernardi and coworkers (62, 63) reasoned that maize coding genes must be located in "gene-rich" regions and that these gene-rich regions must be flanked by DNA with highly homogeneous G:C contents. They proposed that this flanking DNA could consist of retrotransposons like those flanking the *Adh1* gene (48).

The results from G:C fractionation experiments and studies of the *Adh1* region are inconsistent. On the one hand, the study of the *Adh1* region, coupled with studies of centromeres and knobs, suggest that retrotransposon distribution is widespread, representing 50% of the genome. On the other hand, Bernardi and coworkers' work implicitly suggests that retrotransposon distributions are heterogeneous, with a higher concentration of retroelements in the 17% of the genome that represents coding DNA. Ultimately, there may be a resolution to differences implied by different studies, but such a resolution will require more sequencing of large chromosomal clones representing diverse genomic regions.

Question 2. What contributes more to the evolution of DNA content: multiplication of repetitive DNA or chromosomal duplication? The evolutionary history of maize suggests that retrotransposon multiplication and chromosomal duplication (by way of polyploidy) each have generated a 2-fold increase in DNA content within the last 16 million years. Hence, the net effect of these two evolutionary processes is similar in maize. In contrast, it seems that the multiplication of repeat sequences is the primary contributor to differences in DNA content between many taxa (45). For example, barley and rice have similar complements of low-copy genes (64) but a 12-fold difference in DNA content (Fig. 1). The difference in DNA content is thus probably attributable to differences in the amount of repetitive DNA (64).

It is premature to make the general statement that repeat proliferation contributes more to the evolution of DNA content than chromosomal duplications for two reasons. First, as mentioned previously, mapping studies are biased against the discovery of duplications, and for this reason, there is as yet no accurate indication of the extent of chromosomal duplication in complex genomes. Second, duplication and repeat proliferation are not independent. Duplication plays a role in repeat proliferation, because duplication doubles repetitive DNA as well as low-copy DNA.

Question 3. Are chromosomal duplication events correlated with an increase in the rate of transposition? This question originates from the work of Matzke, Matzke, and colleague (65, 66). They argue that polyploid genomes contain duplications of all genes and thus are relatively well buffered against mutations caused by transposon insertion. As a consequence, transposable elements multiply and are maintained in polyploid genomes. For maize, the fact that two major events (polyploidy and retrotransposon multiplication) are located on the same phylogenetic lineage gives credence to the idea that these phenomena are biologically correlated (Fig. 1), but it is not yet known whether this correlation is widely observed.

Genetic Variation in Genes Along Chromosomes

Genetic Diversity as a Function of Recombination, Natural Selection, and Chromosomal Position. Genomes are dynamic entities that can be modified extensively by polyploidy and transposon multiplication. However, ongoing evolutionary processes like mutation, recombination, natural selection, and migration also shape the genome. The effect of these extant processes on the genome can be inferred from careful study of genetic diversity.

Diversity throughout the genome is affected strongly by the interplay of recombination and natural selection. In *Drosophila*, for example, genetic diversity varies along the chromosome as a function of recombination rate (67, 68). Loci near centromeres tend to have low recombination rates and also tend to have low levels of genetic diversity, but both recombination rate and genetic diversity increase toward the tip of chromosomes. This relationship is not because recombination is mutagenic; rather,

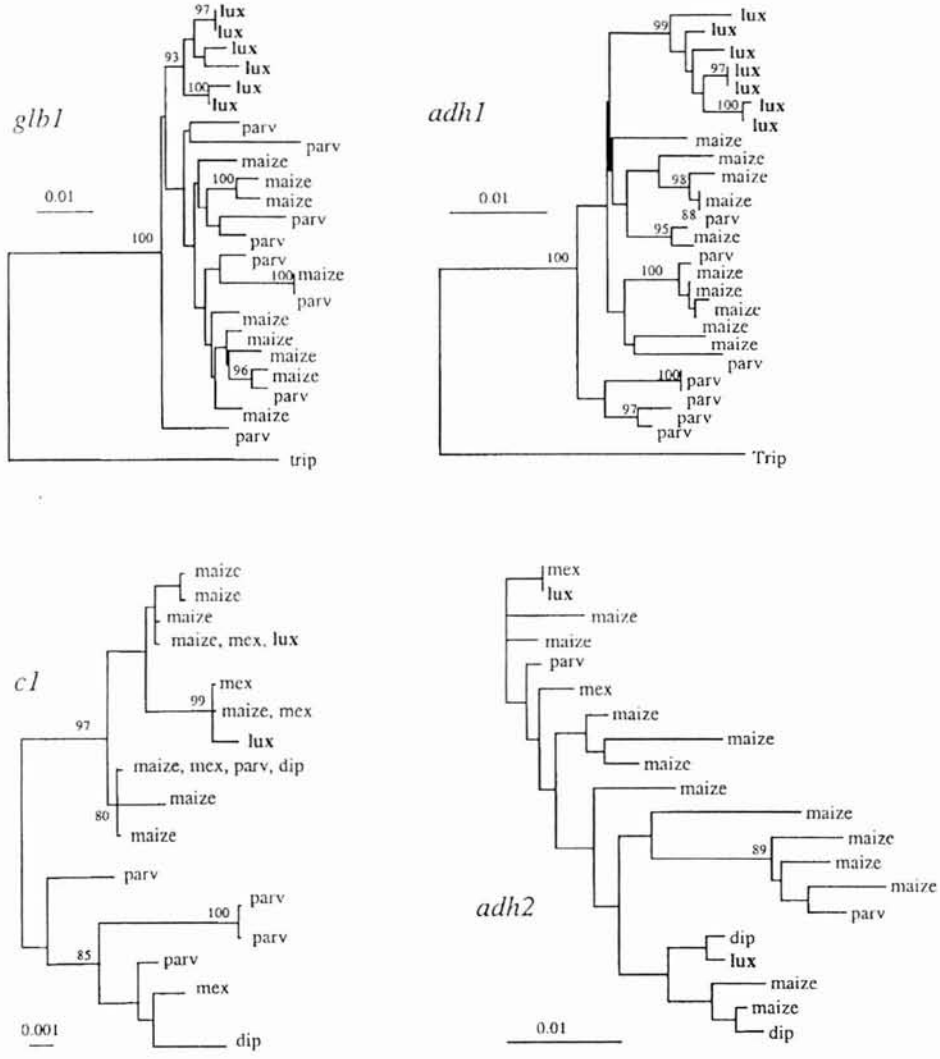


Fig. 4. Genealogies of four genes, based on the neighbor-joining method (82) with Kimura 2-parameter distances (83). Taxa are abbreviated as follows: maize, domesticated maize; parv, ancestor of domesticated maize (*Z. mays* subsp. *parviglumis*); mex, *Z. mays* subsp. *mexicana*; lux, *Zea luxurians*; dip, *Zea diploperennis*; trip, *Tripsacum dactyloides*. Sequences from *Z. luxurians* are shown in bold. The data are from refs. 52 and 76–78. Scale bars indicate level of divergence among sequences; bootstrap values >80% are shown.

it reflects an interdependence between natural selection and recombination (67, 69). In regions of low recombination, for example, linkage between nucleotide sites ensures that selection for or against a single nucleotide substitution will affect a large region of the genome. In regions of high recombination, nucleotide sites are nearly independent; thus, selection on a single site affects a much smaller region of the genome. The result of the interdependence between selection and recombination is that (*i*) levels of genetic diversity can be a function of chromosomal position and (*ii*) large chromosomal regions can be depauperate of genetic diversity.

The correlation between chromosomal position and genetic diversity has been confirmed in plants (70, 71), but it is not yet clear whether recombination in maize follows a simple pattern along chromosomes. For example, it has been documented that maize single-copy regions act as recombination hot spots, but recombination rates also vary among single-copy regions (72–74). Altogether, these studies suggest that the relationship between chromosomal position and recombination rate may not be as straightforward in maize as in *Drosophila*. More thorough elucidation of recombination rates in maize requires comparisons between genetic and physical maps; such physical maps are being produced but are not yet completed.

Nonetheless, we have a goal to quantify patterns of genetic diversity more accurately in the maize genome. To make this quantification, we have begun a long-term study of 100 maize genes along chromosomes 1 and 3. To measure genetic diversity in each gene, we will sample DNA sequences from ≈70 individuals representing maize, its progenitor, and two other wild *Zea* taxa. The project has many long-term goals, including (*i*) to investigate the relationship between chromosomal position and genetic diversity, (*ii*) to examine the impact of domestication on genetic diversity in maize, (*iii*) to compare the evolutionary history among species across genes, and (*iv*) to create a public single-nucleotide-polymorphism database.

The first stage of this ongoing project is to measure genetic diversity in 25 chromosome 1 genes from 16 maize individuals representing Mexican and South American land races and 9

individuals representing U.S. inbred lines. The results of this first stage will be reported in detail elsewhere, but we can make a preliminary contrast of diversity in centromeric vs. noncentromeric genes. Average diversity per base pair in four genes within 5 centimorgans of the centromere is $\theta = 0.0144$, as determined by using Watterson's estimator (75). This level of diversity is slightly lower than average diversity in 11 noncentromeric genes (average Watterson's $\theta = 0.0170$), but the centromeric genes do not have extremely low levels of diversity. For example, all four centromeric genes contain more diversity than 3 of the 11 noncentromeric genes. Thus, we report that there is as yet no clear evidence for a strong reduction in genetic diversity near the centromere of chromosome 1.

Discordant Evolutionary Histories Among Genes. One interesting feature of genetic diversity studies of maize and its wild relatives is that evolutionary histories differ among loci. As an example, consider Fig. 4, which summarizes sequence data from four genes. The genes *Adh1* and *Glb1* provide very similar pictures of the relationship of the wild species *Z. luxurians* to other members of the genus *Zea* (52, 76); in short, for both of these genes, *Z. luxurians* sequences comprise a separate, well defined clade. In contrast, *Z. luxurians* individuals contain sequences that are very similar (or even identical) to sequences from other *Zea* taxa for *Adh2* (77) and *c1* (78). Thus, the picture of evolutionary history from *Adh1* and *Glb1* is not consistent with information from *c1* and *Adh2*. (Fig. 4 focuses on genealogical or phylogenetic information for ease of presentation, but sequence statistics also suggest that these genes have different evolutionary histories.) One interesting feature of Fig. 4 is that *Adh1* and *Glb1* are located within a 12-centimorgan region of chromosome 1; *Adh2* and *c1* are found on chromosomes 4 and 9, respectively.

We have sampled extensively from the wild relatives of maize for only a handful of genes, but discordant patterns, such as those demonstrated in Fig. 4, continue to be identified. The challenge of these data will be to infer the evolutionary processes that contribute to discordant evolutionary histories among genes. Several possibilities exist, including differences in nucleotide substitution rates, introgression (migration) rates, and natural selection among genes. One interesting possibility is that genealogical patterns among genes may correlate with chromosomal location.

In this context, it is worth noting that studies of *Drosophila* species have also demonstrated discordant patterns of genetic diversity among loci. For example, Wang *et al.* (79) studied three loci in three *Drosophila* species. Two of the loci (*Hsp82* and *period*) yielded very similar pictures of genetic divergence among taxa. At these two loci, sequences were well differentiated among taxa. However, the pattern of genetic diversity in the third *Drosophila* locus (*Adh*) was incongruent with data from the first two loci. In this last locus, DNA sequences from different taxa were not highly diverged. Wang *et al.* (79) used population genetic tools to contrast genealogical information among *Drosophila* loci, and they concluded that introgression among species has occurred at a much higher rate at one locus (*Adh*) than at the other two loci (*Hsp82* and *period*). In short, *Drosophila* studies strongly suggest that the processes affecting genetic diversity can vary among loci and also demonstrate the importance of comparing genealogical information across species and across loci.

In crops, artificial selection can cause discordant patterns of genetic diversity among loci. Thus far, levels of nucleotide sequence diversity have been measured in maize and its wild progenitor (*Z. mays* subsp. *parviglumis*) for six genes (summarized in ref. 80). All six genes indicate that maize has reduced genetic diversity relative to its wild progenitor, probably reflecting a genetic bottleneck during domestication (52, 76). However, the level of reduction in genetic diversity varies substantially among genes. For four of the six genes, maize retains at least half of the genetic diversity of its wild progenitor. For the remaining two genes (*c1* and *tb1*), maize contains less than 20% of the level of diversity of its wild progenitor (78, 81). Low diversity in *c1* and *tb1* likely reflects artificial selection by the early domesticators of maize. The *tb1* gene was probably selected to affect morphological changes in branching pattern (81), and *c1* may have been selected for production of purple pigment in maize kernels (78).

Just as domestication has had a heterogeneous effect across loci, so has the process of maize breeding. For nine genes that we have sampled extensively thus far, U.S. inbred lines average roughly 65% the level of genetic diversity of the broader sample of maize. This level of reduction from maize land races to U.S. maize is commensurate with the original reduction in genetic diversity from wild progenitor to domesticated maize (52). Altogether, owing to reductions in diversity caused by initial domestication and subsequent intensive breeding, our initial estimates indicate that U.S. inbreds contain only ≈40% of the level of genetic diversity of the wild ancestor of maize.

Thus far, studies of genetic diversity have shown that maize genes have different levels of genetic diversity, and diversity in some genes has been affected strongly by artificial selection. In addition, studies of wild *Zea* taxa indicate that genes differ in their evolutionary histories among taxa. Our ongoing study of 100 genes will help determine whether patterns of evolutionary history among genes are, in fact, correlated with chromosomal location and will also contribute to the overall understanding of the evolutionary forces acting on plant genomes.

The authors acknowledge National Science Foundation Grants DBI-9872631 and DEB-9815855 and U.S. Department of Agriculture Grant 98-35301-6153.

1. Stebbins, G. L. (1950) *Variation and Evolution in Plants* (Columbia Univ. Press, New York).
2. Bennett, M. D. & Leitch, I. J. (1995) *Ann. Bot.* **76,** 113–176.
3. McClintock, B. (1930) *Proc. Natl. Acad. Sci. USA* **16,** 791–796.
4. McClintock, B. (1933) *Z. Zellforsch. Mikrosk. Anat.* **19,** 191–237.
5. Ting, Y. C. (1966) *Cytologia* **31,** 324–329.
6. Snope, A. J. (1967) *Chromosoma* **21,** 243–349.
7. Rhoades, M. M. (1951) *Am. Nat.* **85,** 105–110.
8. Rhoades, M. M. (1955) in *Corn and Corn Improvement*, ed. Sprague, G. F. (Academic, New York), pp. 123–219.
9. Goodman, M. M., Stuber, C. W., Newton, K. & Weissinger, H. H. (1980) *Genetics* **96,** 697–710.
10. Wendel, J. F., Stuber, C. W., Goodman, M. M. & Beckett, J. B. (1989) *J. Hered.* **80,** 218–228.
11. Wendel, J. F., Stuber, C. W., Edwards, M. D. & Goodman, M. M. (1986) *Theor. Appl. Genet.* **72,** 178–185.
12. McMillin, D. E. & Scandalios, J. G. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 4866–4870.
13. Davis, G. M., McMullen, M. D., Baysdorfer, C., Musket, T., Grant, D., Staebell, M., Xu, G., Polacco, M., Koster, L., Melia-Hancock, S. *et al.* (1999) *Genetics* **152,** 1137–1172.
14. Helentjaris, T., Weber, D. & Wright, S. (1988) *Genetics* **118,** 353–363.
15. Ahn, S. & Tanksley, S. D. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 7980–7984.
16. Anderson, E. (1945) *Chron. Bot.* **9,** 88–92.
17. Gaut, B. S. & Doebley, J. F. (1997) *Proc. Natl. Acad. USA* **94,** 6809–6814.
18. Whitkus, R., Doebley, J. & Lee, M. (1992) *Genetics* **132,** 1119–1130.
19. Pereira, M. G., Lee, M., Bramel-Cox, P., Woodman, W., Doebley, J. & Whitkus, R. (1994) *Genome* **37,** 236–243.
20. Berhan, A. M., Hulbert, S. H., Butler, L. G. & Bennetzen, J. L. (1993) *Theor. Appl. Genet.* **86,** 598–604.
21. Chittenden, L. M., Schertz, K. F., Lin, Y. R., Wing, R. A. & Paterson, A. H. (1994) *Theor. Appl. Genet.* **87,** 925–933.
22. Celarier, R. P. (1956) *Rhodora* **58,** 135–143.
23. Molina, M. D. & Naranjo, C. A. (1987) *Theor. Appl. Genet.* **73,** 542–550.
24. Spangler, R., Zaitchik, B., Russo, E. & Kellogg, E. A. (1999) *Syst. Bot.* **24,** 267–281.
25. Moore, G., Devos, K. M., Wang, Z. & Gale, M. D. (1995) *Curr. Biol.* **5,** 737–739.
26. Devos, K. M. & Gale, M. D. (1997) *Plant Mol. Biol.* **35,** 3–15.
27. Wilson, W. A., Harrington, S. E., Woodman, W. L., Lee, M., Sorrells, M. E. & McCouch, S. R. (1999) *Genetics* **153,** 453–473.

28. Wendel, J. F. (2000) *Plant Mol. Biol.* **42,** 225–249.
29. Song, K., Lu, P., Tang, K. & Osborn, T. C. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 7719–7723.
30. Liu, B., Vega, J. M., Segal, G., Abbo, S., Rodova, H. & Feldman, M. (1998) *Genome* **41,** 272–277.
31. Liu, B., Vega, J. M. & Feldman, M. (1998) *Genome* **41,** 535–542.
32. Feldman, M., Liu, B., Segal, G., Abbo, S., Levy, A. A. & Vega, J. M. (1997) *Genetics* **147,** 1381–1387.
33. Wolfe, K. H. & Shields, D. C. (1997) *Nature (London)* **387,** 708–713.
34. Seroghe, C. & Wolfe, K. H. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 4447–4452.
35. Postlethwait, J. H., Yan, Y.-L., Gates, M. A., Horne, S., Amores, A., Brownlie, A., Donovan, A., Egan, E. S., Force, A., Gong, Z. Y., *et al.* (1998) *Nat. Genet.* **18,** 345–349.
36. Shoemaker, R. C., Polzin, K., Labate, J., Specht, J., Brummer, E. C., Olson, T., Young, N., Concibido, V., Wilcox, J., Tamulonis, J. P., *et al.* (1996) *Genetics* **144,** 329–338.
37. Bohuon, E. J. R., Keith, D. J., Parkin, I. A. P., Sharpe, A. G. & Lydiate, D. J. (1996) *Theor. Appl. Genet.* **93,** 833–839.
38. Cavell, A. C., Lydiate, D. J., Parkin, I. A. P., Dean, C. & Trick, M. (1998) *Genome* **41,** 62–69.
39. Bonierbale, M. W., Plaisted, R. L. & Tanksley, S. D. (1988) *Genetics* **120,** 1095–1103.
40. Masterson, J. (1994) *Science* **264,** 421–423.
41. Harushima, Y., Yano, M., Shomura, A., Sato, M., Shimano, T., Kuboki, Y., Yamamoto, T., Lin, S.-Y., Antonio, B. A., Parco, A., *et al.* (1998) *Genetics* **148,** 479–494.
42. Mayer, K. S., Schaller, C., Wambutt, R., Murphy, G., Volckaert, G., Pohl, T., Dusterhoft, A., Stiekema, W., Entian, K. D., Terryn, N., *et al.* (1999) *Nature (London)* **402,** 769–777.
43. Remsch, A. J., Dong, J., Brubaker, C. L., Stelly, D. M., Wendel, J. F. & Paterson, A. H. (1994) *Genetics* **138,** 829–847.
44. Lin, X. Y., Kaul, S. S., Rounsley, S., Shea, T. P., Benito, M. I., Town, C. D., Fuji, C. Y., Mason, T., Bowman, C. L., Barnstead, M., *et al.* (1999) *Nature (London)* **402,** 761–768.
45. Flavell, R. B., Bennett, M. D., Smith, J. B. & Smith, D. B. (1974) *Biochem. Genet.* **12,** 257–269.
46. Hake, S. & Walbot, V. (1980) *Chromosoma* **79,** 251–270.
47. Springer, P. S., Edwards, K. J. & Bennetzen, J. L. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 863–867.
48. SanMiguel, P., Tickhonov, A., Jin, Y.-K., Melake-Berhan, A., Springer, P. S., Edwards, K. J., Avramova, Z. & Bennetzen, J. L. (1996) *Science* **274,** 765–768.
49. SanMiguel, P. J., Gaut, B. S., Tikhonov, A., Nakajima, Y. & Bennetzen, J. L. (1998) *Nat. Genet.* **20,** 43–45.
50. Tikhonov, A. P., SanMiguel, P. J., Nakajima, Y., Gorenstein, N. M., Bennetzen, J. L. & Avramova, Z. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 7409–7414.
51. Bennetzen, J. L. & Kellogg, E. A. (1997) *Plant Cell* **9,** 1509–1514.
52. Hilton, H. & Gaut, B. S. (1998) *Genetics* **150,** 863–872.
53. Vicient, C. M. & Martinez-Izquierdo, J. A. (1997) *Gene* **184,** 257–261.
54. Bennett, M. D. & Smith, J. B. (1991) *Philos. Trans. R. Soc. London Ser. B* **334,** 309–345.
55. Charlesworth, B., Langley, C. H. & Stephan, W. (1986) *Genetics* **112,** 947–962.
56. Charlesworth, B., Sniegowski, P. & Stephan, W. (1994) *Nature (London)* **371,** 215–220.
57. Copenhaver, G. N., Nickel, K., Kuromori, T., Benito, M. I., Kaul, S., Lin, X. Y., Bevan, M., Murphy, G., Harris, B., Parnell, L. D., *et al.* (1999) *Science* **286,** 2468–2474.
58. Ananiev, E. V., Phillips, R. L. & Rines, H. W. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 10785–10790.
59. Ananiev, E. V., Phillips, R. L. & Rines, H. W. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 13073–13078.
60. Ananiev, E. V., Phillips, R. L. & Rines, H. W. (1998) *Genetics* **149,** 2025–2037.
61. Zhang, Q., Arbuckle, J. & Wessler, S. R. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 1160–1165.
62. Barakat, A., Carels, N. & Bernardi, G. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 6857–6861.
63. Carels, N., Barakat, A. & Bernardi, G. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 11057–11060.
64. Saghai-Maroof, M. A., Yang, G. P., Biyashev, R. M., Maughan, P. J. & Zhang, Q. (1996) *Theor. Appl. Genet.* **92,** 541–551.
65. Matzke, M. A., Mittelsten-Scheid, O. & Matzke, A. J. M. (1999) *BioEssays* **21,** 761–767.
66. Matzke, M. A. & Matzke, A. J. M. (1998) *Trends Ecol. Evol.* **13,** 241.
67. Begun, D. J. & Aquadro, C. F. (1992) *Nature (London)* **356,** 519–520.
68. Hamblin, M. T. & Aquadro, C. F. (1999) *Genetics* **153,** 859–869.
69. Charlesworth, D., Charlesworth, B. & Morgan, M. T. (1995) *Genetics* **141,** 1619–1632.
70. Dvorak, J., Luo, M.-C. & Yang, Z.-L. (1998) *Genetics* **148,** 423–434.
71. Stephan, W. & Langley, C. H. (1998) *Genetics* **150,** 1585–1593.
72. Civardi, L., Xia, Y., Edwards, K. J., Schnable, P. S. & Nikolau, B. J. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 8268–8272.
73. Okagaki, R. J. & Weil, C. F. (1997) *Genetics* **147,** 815–821.
74. Timmermans, M. C. P., Das, O. P. & Messing, J. (1996) *Genetics* **143,** 1771–1783.
75. Watterson, G. A. (1975) *Theor. Popul. Biol.* **7,** 188–193.
76. Eyre-Walker, A., Gaut, R. L., Hilton, H., Feldman, D. L. & Gaut, B. S. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 4441–4446.
77. Goloubinoff, P., Paabo, S. & Wilson, A. C. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 1997–2001.
78. Hanson, M. A., Gaut, B. S., Stec, A. O., Fuerstenberg, S. I., Goodman, M. M., Coe, E. H. & Doebley, J. (1996) *Genetics* **143,** 1395–1407.
79. Wang, R. L., Wakeley, J. & Hey, J. (1997) *Genetics* **147,** 1091–1106.
80. White, S. E. & Doebley, J. F. (1999) *Genetics* **153,** 1455–1462.
81. Wang, R. L., Stec, A., Hey, J., Lukens, L. & Doebley, J. (1999) *Nature (London)* **398,** 236–239.
82. Saitou, N. & Nei, M. (1987) *Mol. Biol. Evol.* **4,** 406–425.
83. Kimura, M. (1980) *J. Mol. Evol.* **16,** 111–120.
84. Gale, M. D. & Devos, K. M. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 1971–1974.

Plant Transposable Elements. A Hard Act to Follow

Susan R. Wessler*

Departments of Botany and Genetics, Miller Plant Sciences Building, University of Georgia, Athens, Georgia 30602

The discovery and genetic characterization of plant transposable elements (TEs) led to a revolution in our understanding of the composition and dynamic potential of the genetic material in virtually all organisms. Most of these breakthroughs occurred between 30 and 50 years ago. It was during this time that TEs were discovered in maize (*Zea mays*) and several aspects of their genetic behavior were characterized. Through the study of spotted kernels and sectored flowers, McClintock and her contemporaries discovered: (a) the existence of multiple TE families with autonomous and nonautonomous members that are normal residents of the genome, (b) that elements can move within and between chromosomes where they can alter gene expression or serve as sites of chromosome breakage or rearrangement, (c) that excision is often imprecise and reinsertion is often to a linked locus, and (d) that elements can exist in the genome in a quiescent state that is subject to reactivation by biotic and abiotic means collectively termed "genomic stress." This era of discovery and its relevance to modern biology is reviewed by Fedoroff (9). In addition, the story of how the genomic stress hypothesis came to fruition is summarized in McClintock's Nobel lecture (18). This view of the genome as responsive and dynamic, that is, something more than a collection of genes, heralded the start of the current genomics era.

The purpose of these historical notes is to review conceptual breakthroughs that have occurred over the past quarter century. I would venture to guess that for most of us involved in the study of plant TEs during this time, the historical legacy has been a hard act to follow. However, I will argue that recent studies, especially those in the last 5 years, have raised the bar on what constitutes the dynamic genome and have placed plants once again at the forefront of transposon studies.

PHASE I: THE DNA ELEMENTS—CHARACTERIZING OUR GENETIC LEGACY

Although the historical legacy may have been a hard act to follow, the large collections of TE-induced alleles generated during that era provided most of the raw materials used by the first generation of plant molecular biologists. What I have arbitrarily called phase I in the molecular analysis of plant TEs focused on two areas: (a) the isolation of genes and their TE-induced alleles and the subsequent characterization of TE families, and (b) understanding the mechanisms underlying the diversity of unstable phenotypes. Many of the first plant genes cloned in the early 1980s were chosen because they had TE-induced alleles. These include the *sh1*, *adh1*, and *wx* genes from maize and the *nivea* gene from *Antirhinum majus*. From their mutant alleles came the first characterized *Ds*, *Ac*, and *Tam* elements (3, 11, 21). The isolation of members of the other prominent plant TE families including *Mutator*, *Spm/En*, and other *Tam* elements occurred in numerous labs around the world including those of the Burrs, Freeling, Peterson/Saedler, Starlinger, Walbot, and Carpenter/Coen (for review, see 17)

This first wave of plant elements was distinguished by their high frequency of somatic and germinal instability. All were DNA or class 2 elements, a group that is characterized by short inverted terminal repeats and transposition via a DNA intermediate. That is, the element usually excises from one site and reinserts elsewhere. Analyses of complex and diverse mutant phenotypes induced by insertion and excision of class 2 elements into plant genes revealed a myriad of ways that these elements can modify gene regulation. These included: (a) the discovery that transposon footprints are usually left behind when elements exicise (21), (b) that *Ds* and *dSpm* elements function as introns (15, 23), and (c) that promoter insertions can alter tissue-specific patterns of expression (8).

In retrospect, the most important discoveries that came out of this period were the analyses of what McClintock called changes in phase, or the reversible switch in element activity. The correlation between element activity and methylation state was first documented for the *Mutator* element (6) and foreshadowed similar correlations for other TEs and normal plant genes.

With these elements in one hand and a large collection of previously isolated TE-induced alleles in the other, transposon tagging strategies were developed (for review, see 10) that permitted the rapid isolation of many structural and regulatory genes in both maize and *Antirrhinum majus*. The importance of transposon tagging protocols for gene isolation was recognized by those working with plants that lack

* E-mail sue@dogwood.botany.uga.edu; fax 706–542–1805.

well-characterized endogenous systems. This recognition led to the introduction of the maize elements into a variety of plant species, beginning with tobacco (*Nicotiana tubacum*; 1), and has resulted in the isolation of many genes, the first being a floral color gene from petunia (*Petunia hybrida*; 7) and the most prominent being the *N* resistance gene from tobacco (24). It is unfortunate that the success of forward genetic approaches like transposon tagging is severely limited by the large size of most plant genomes and the requirement for a visible mutant phenotype. To circumvent these problems, site-selected transposon tagging protocols (also called reverse genetics) using the *Mutator* element were developed (2).

PHASE 2: TEs OF HIGH COPY NUMBER

Although the characterization of class 2 elements dominated the first 10 years of the molecular era, it soon became clear that the low copy numbers of these elements precluded their having a significant impact on genome size, structure, or evolution. One exception to this generalization may be miniature inverted repeat TEs (MITEs) which appear to be high copy number class 2 elements that, in some cases, are preferentially associated with grass genes (4).

TE studies in the 1990s have been dominated by long terminal repeat (LTR) retrotransposons, which are members of the class 1 or retro-element group. LTR retrotransposons are flanked by long terminal repeats and usually encode all of the proteins required for their transposition. For all class 1 elements, it is the element-encoded transcript, and not the element itself, that forms the transposition intermediate. It is for this reason that they can attain much higher copy number than class 2 elements. Transcription of most of the active plant elements characterized to date is largely quiescent during normal development but can be induced by biotic and/or abiotic stresses including cell culture, wounding, and pathogen attack (12). Because the element-encoded transcript is also the transposition intermediate, LTR retrotransposons may have the ability to rapidly alter genome structure in response to environmental cues (see below).

Given their large size (from 4–10 kb on average) and potential to amplify on a massive scale, it is not surprising that LTR retrotransposons comprise the largest fraction of TE-derived genomic DNA in almost all plant genomes examined to date (for review, see 16). An important series of recent experiments, led by the Bennetzen lab (20), has demonstrated that differential amplification of LTR retrotransposons largely accounts for the C-value paradox among the agronomically important members of the grass clade. The C-value paradox is the observed lack of correlation between increases in DNA content and an organism's complexity. It has been documented for both animal and plant species, but to date only appears to be "solved" for the members of the grass tribe.

The focus on high copy number elements in plants necessitated the development of new protocols to assay TE movement on a whole-genome basis. Unlike the low copy number class 2 elements discussed above, MITEs and retrotransposons rarely transpose and are not associated with mutant genes. Thus their activity could not be visualized in the traditional manner of examining spotted kernels (Fig. 1A). Instead, a modification of the gel-based amplified fragment length polymophism technique called transposon display was developed to simultaneously monitor the movement of hundreds of elements (Fig. 1B; 22). Transposon display of the stable and highly polymorphic MITE families of maize has led to their use as a new class of molecular marker that is preferentially associated with genic regions (5).

THE FUTURE OF PLANT TRANSPOSONS: POISED FOR NEW BREAKTHROUGHS

Given that a large fraction of the DNA sequence output from plant genome projects will be derived from TEs, there will be no shortage of new elements to be discovered, categorized, and exploited as potentially valuable molecular tools. However, three recent papers exemplify for me the areas where major breakthroughs are most likely to arise. The first, by Hirochiki and coworkers, reports the amplification of the tobacco retrotransposon *Tto1* in Arabidopsis plants that are methylation deficient (*ddm1*; 13). In the near future we should know how epigenetic

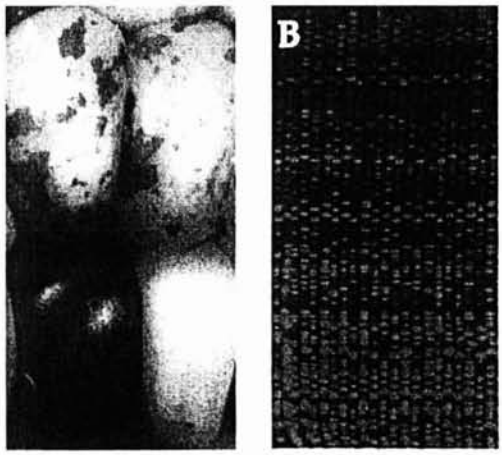


Figure 1. Assaying TE activity the old way (A) and the new way (B). A, Maize kernels displaying patterns characteristic of TE excision from a gene in the anthocyanin biosynthetic pathway. B, Computer-generated image of a transposon display where blue fluorescent bands are PCR products that are anchored in a restriction site and in a member of the MITE family *Htb2* (courtesy of Zenaida Magbanua [University of Georgia, Athens]). Red bands are M, markers. Samples were resolved on an ABI377 sequencer where the loading was staggered.

mechanisms regulate TEs, whether this control is influenced by environmental cues, how TE organization influences global chromatin structure (and in turn gene expression), and whether epigenetic regulation evolved to regulate TEs.

Two papers have raised the bar on our concept of the dynamic genome and have positioned the grass clade as a focal point for future studies. In a follow-up to their study of intergenic retrotransposons in maize, SanMiguel et al. (19) provide evidence that a burst in retrotransposon activity doubled the size of the maize genome within the past 3 million years. This result demonstrated for the first time that TEs could rapidly restructure a genome. In the second paper Kalendar et al. (14) present a dramatic example of TE-mediated genomic restructuring within populations of the wild barley *Hordeum spontaneum* growing in distinct regions of a canyon in Israel. In this case, genome restructuring takes the form of genome size variation due to retrotransposon amplification (the *BARE-1* element) and intra-element deletion. Correlation between *BARE-1* copy number, genome size, and local environmental conditions suggest for the first time a testable molecular mechanism linking habitat with TE induction in natural populations.

Taken together these two studies suggest that the grass clade is in a dynamic period of genomic restructuring and, for this reason, may be the system of choice for understanding the extent of TEs involvement in both macroevolutionary and microevolutionary processes. Given the rapid pace of recent discoveries, it may be reasonable to expect that in the not-too-distant future this line of research will provide mechanisms to explain how evolution works at the molecular level.

LITERATURE CITED

1. **Baker B, Schell J, Lorz H, Fedoroff N** (1986) Proc Natl Acad Sci USA **83:** 4844–4848
2. **Bensen RJ, Johal GS, Crane VC, Tossberg JT, Schnable PS, Meeley RB, Briggs SP** (1995) Plant Cell **7:** 75–84
3. **Bonas U, Sommer H, Saedler H** (1984) EMBO J **5:** 1015–1019
4. **Bureau TE, Wessler SR** (1992) Plant Cell **4:** 1283–1294
5. **Casa AM, Brouwer C, Nagel A, Wang L, Zhang Q, Kresovich S, Wessler SR** (2000) Proc Natl Acad Sci USA **97:** 10083–10089
6. **Chandler VL, Walbot V** (1986) Proc Natl Acad Sci USA **83:** 1767–1771
7. **Chuck G, Robbins T, Nijjar C, Ralston E, Courtney-Gutterson N, Dooner HK** (1993) Plant Cell **5:** 371–378
8. **Coen ES, Carpenter R, Martin C** (1986) Cell **47:** 285–296
9. **Fedoroff NV** (1983) *In* Mobile Genetic Elements. Academic Press, New York, pp 1–63
10. **Fedoroff NV, Furtek DB, Nelson OE** (1984) Proc Natl Acad Sci USA **81:** 3825–3829
11. **Fedoroff NV, Wessler SR, Shure M** (1983) Cell **35:** 235–242
12. **Grandbastien M-A** (1998) Trends Plant Sci **3:** 181–189
13. **Hirochika H, Okamoto H, Kakutani T** (2000) Plant Cell **12:** 357–369
14. **Kalendar R, Tanskanen J, Immonen S, Nevo E, Schulman AH** (2000) Proc Natl Acad Sci USA **97:** 6603–6607
15. **Kim HY, Schiefelbein JW, Raboy V, Furtek D, Nelson OE** (1987) Proc Natl Acad Sci USA **84:** 5863–5867
16. **Kumar A, Bennetzen JL** (1999) Annu Rev Genet **33:** 479–532
17. **Kunze R, Saedler H, Lönnig W-E** (1997) Adv Bot Res **27:** 331–470
18. **McClintock B** (1984) Science **226:** 792–801
19. **SanMiguel P, Gaut BS, Tikhonov A, Nakajima Y, Bennetzen JL** (1998) Nat Genet **20:** 43–45
20. **SanMiguel P, Tikhonov A, Jin Y-K, Motchoulskaia N, Zakharov D, Melake-Berhan A, Springer PS, Edwards KJ, Lee M, Avramova Z, Bennetzen, JL** (1996) Science **274:** 765–768
21. **Sutton WD, Gerlach WL, Schwartz D, Peacock WJ** (1984) Science **223:** 1265–1268
22. **Van den Broeck D, Maes T, Sauer M, Zethof J, De Keukeleire P, D'Hauw M, Van Montagu M, Gerats T** (1998) Plant J **13:** 121–129
23. **Wessler SR, Baran G, Varagona M** (1987) Science **237:** 916–918
24. **Whitham S, Dineshkumar SP, Choi D, Hehl R, Corr C, Baker B** (1994) Cell **78:** 1101–1115

The EMBO Journal Vol. 21 No. 17 pp. 4671–4679, 2002

Two classes of short interfering RNA in RNA silencing

Andrew Hamilton[1], Olivier Voinnet, Louise Chappell and David Baulcombe[2]

The Sainsbury Laboratory, John Innes Centre, Colney Lane, Norwich NR4 7UH, UK

[1]Present address: Department of Pathology, Glasgow University, Western Infirmary, Glasgow G11 6NT, UK

[2]Corresponding author
e-mail: david.baulcombe@sainsbury-laboratory.ac.uk

A.Hamilton and O.Voinnet contributed equally to this work

RNA silencing is a eukaryotic genome defence system that involves processing of double-stranded RNA (dsRNA) into 21–26 nt, short interfering RNA (siRNA). The siRNA mediates suppression of genes corresponding to the dsRNA through targeted RNA degradation. In some plant systems there are additional silencing processes, involving systemic spread of silencing and RNA-directed methylation/transcriptional suppression of homologous genomic DNA. We show here that siRNAs produced in plants from a green fluorescent protein (GFP) transgene are in short (21–22 nt) and long (24–26 nt) size classes, whereas those from endogenous retroelements are only in the long class. Viral suppressors of RNA silencing and mutations in *Arabidopsis* indicate that these classes of siRNA have different roles. The long siRNA is dispensable for sequence-specific mRNA degradation, but correlates with systemic silencing and methylation of homologous DNA. Conversely, the short siRNA class correlates with mRNA degradation but not with systemic signalling or methylation. These findings reveal an unexpected level of complexity in the RNA silencing pathway in plants that may also apply in animals.

Keywords: DNA methylation/double-stranded RNA/retroelements/RNA silencing/silencing suppressor/systemic signalling

Introduction

In eukaryotic cells, gene silencing operating at the RNA level has roles in adaptive protection against viruses (Voinnet, 2001), in genome defense against mobile DNA elements (Ketting *et al.*, 1999; Wu-Scharf *et al.*, 2000) and in developmental regulation of gene expression (Grishok *et al.*, 2001; Hutvagner *et al.*, 2001; Ketting *et al.*, 2001). Many of these silencing systems involve double-stranded RNA (dsRNA) that is generated by host- or virus-encoded RNA-dependent RNA polymerases (Dalmay *et al.*, 2000; Mourrain *et al.*, 2000; Sijen *et al.*, 2001a), by transcription either through inverted repeats (Grishok *et al.*, 2001; Hutvagner *et al.*, 2001; Ketting *et al.*, 2001) or from converging promoters (Aravin *et al.*, 2001). Various terms including RNA interference, post-transcriptional gene silencing and quelling have been used to refer to these examples of gene silencing. However, based on genetic and molecular analyses, it seems that their mechanisms share similarities and here we use the generic term 'RNA silencing'.

A second component of RNA silencing, in addition to dsRNA, is a 21–26 nt RNA known as short interfering RNA (siRNA) (Hamilton and Baulcombe, 1999; Elbashir *et al.*, 2001a). In *Drosophila*, the siRNA is derived from dsRNA (Zamore *et al.*, 2000) by the action of an RNaseIII-like enzyme named Dicer (Bernstein *et al.*, 2001). The siRNA guides a multi-subunit ribonuclease, referred to as RNA-induced silencing complex (RISC) (Hammond *et al.*, 2000, 2001a; Elbashir *et al.*, 2001a; Nykanen *et al.*, 2001), and ensures that it specifically degrades RNAs that share sequence similarity with the dsRNA. It is thought that the specificity is mediated by base pairing of the siRNA and the target.

In addition to mRNA degradation, RNA silencing in plants acts at several other levels, including DNA methylation and transcriptional suppression (Wassenegger and Pelissier, 1998; Mette *et al.*, 2000; Jones *et al.*, 2001), pre-mRNA processing (Mishra and Handa, 1998) and translation (VanHoudt *et al.*, 1997). Aspects of RNA silencing have also been implicated in translational control in other eukaryotes (Grishok *et al.*, 2001; Hutvagner *et al.*, 2001; Ketting *et al.*, 2001). In many of these examples the target nucleic acids share nucleotide sequence similarity with the initial dsRNA trigger, and it is likely therefore that siRNAs are the specificity determinants. However, it is not known whether RISC is involved and it remains possible that different siRNA-containing complexes act at various levels of gene expression.

As well as intracellular RNA silencing, there is also transmission of the silencing state between cells (Palauqui *et al.*, 1997; Voinnet and Baulcombe, 1997; Voinnet *et al.*, 1998). In plants, a signal of silencing moves from cell to cell through plasmodesmata and for greater distances through the vascular system. The signal is likely to incorporate a nucleic acid because it mediates a nucleotide sequence-specific effect. A similar signal may exist in *Caenorhabditis elegans*, where RNA silencing is also non-cell autonomous (Fire *et al.*, 1998; Winston *et al.*, 2002).

One possible role of the extracellular signal of silencing in plants is anti-viral. The signal would move together with, or in advance of the virus, and mediate silencing of the viral RNA in the newly infected cells. Consequently the infection would progress slowly or would be arrested (Voinnet *et al.*, 2000). Many plant viruses produce proteins that suppress RNA silencing in order to counteract this defence mechanism (Anandalakshmi *et al.*, 1998; Brigneti *et al.*, 1998; Kasschau and Carrington, 1998). These proteins share no obvious common structural

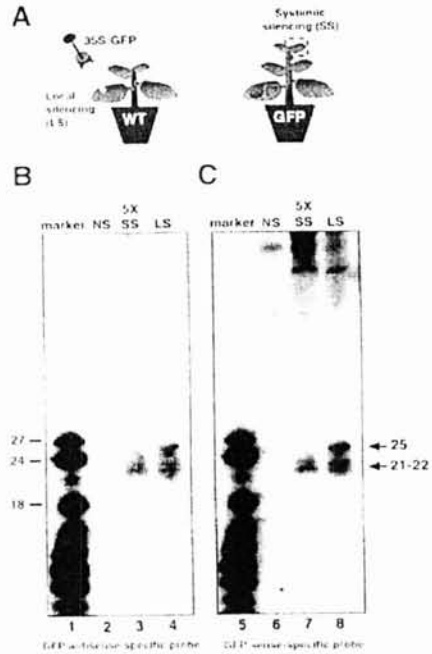


Fig. 1. Differential accumulation of long and short siRNA in local and systemic GFP silencing. (**A**) Local silencing (LS) of GFP (left) was induced in leaves of wild-type (WT) *N.benthamiana* by agro-infiltration of the 35S–GFP construct. Local silencing also occurs in GFP transgenic *N.benthamiana* (right) and precedes systemic silencing (SS), in which transgene expression is suppressed in new, emerging, non-infiltrated tissues. (B–C) Low molecular weight RNA was hybridized with GFP antisense-specific (**B**) or sense-specific (**C**) probes. Lanes 'SS' were from systemically silenced upper leaves of GFP transgenic *N.benthamiana*. '5X' indicates that the amount of RNA loaded in lanes 3 and 7 was 5-fold higher than the amount loaded in lanes 4 and 8. The samples in lanes 'LS' were from leaves of wild-type *N.benthamiana* exhibiting local silencing following infiltration with the 35S–GFP strain of *A.tumefaciens*. Control samples (lanes 'NS') were from non-infiltrated leaves of wild-type *N.benthamiana*. Both GFP siRNA classes accumulate to similar, high levels in GFP transgenic and wild-type plants (data not shown).

motifs and appear to act against different stages of the RNA silencing mechanism (Voinnet *et al.*, 1999; Anandalakshmi *et al.*, 2000; Llave *et al.*, 2000; Mallory *et al.*, 2001), including synthesis or movement of the systemic signal (Voinnet *et al.*, 2000; Guo and Ding, 2002).

Here, we characterize siRNAs produced from a transgene and from several endogenous retroelements, and we investigate the effects of viral suppressors of silencing. We show that there are two size classes of siRNAs associated with RNA silencing of a green fluorescent protein (GFP) transgene and that these siRNAs are differentially affected by the viral suppressors of RNA silencing. In contrast, siRNA from retroelements is composed of only the long class. If the abundance of the long siRNA was reduced by viral suppressors or by mutation of the *SDE4 Arabidopsis* gene, systemic silencing and methylation of genomic DNA were prevented or reduced, whereas degradation of the target RNA was unaffected. Conversely, if the short class was reduced in abundance or was absent, there was no degradation of target RNAs corresponding to the siRNA. Based on these results, we propose that the short siRNA is incorporated into RISC and is involved in degradation of the target RNA. We further propose that the long siRNA plays a separate role that is associated with systemic signalling of RNA silencing and RNA-directed DNA methylation.

Results

Agrobacterium-mediated silencing of GFP in plants

When a liquid culture of *Agrobacterium tumefaciens* is pressure-injected into leaves, the transferred (T)-DNA of the bacterial Ti plasmid is transferred into plant cells, where transient expression of the T-DNA-encoded genes procedes. Thus, in *Nicotiana benthamiana*, the 'agro-infiltration' of a GFP transgene coupled to a cauliflower mosaic virus 35S promoter (35S) results in strong green fluorescence in the infiltrated zone that contrasts with the surrounding red fluorescence from chlorophyll. GFP fluorescence and GFP mRNA reach peak levels after 2–3 days in the infiltrated patch (Voinnet and Baulcombe, 1997) and then decline as a consequence of RNA silencing activation (Johansen and Carrington, 2001; O.Voinnet, S.Rivas, P.Mestre and D.C.Baulcombe, manuscript submitted). We refer to this phenomenon as 'local silencing' of GFP.

When a 35S–GFP transgene is agro-infiltrated into plants that are already transformed with a GFP transgene, the infiltrated patch appears bright green due to the transient GFP expression superimposed on fainter green fluorescence from the resident transgene (Voinnet *et al.*, 1998). As on the non-transformed plants, transient expression of GFP peaks after 2–3 days and then declines. At later times, the tissue becomes uniformly red fluorescent (Voinnet *et al.*, 1998), indicating that the newly infiltrated transgenes and the resident GFP transgene have both become locally silenced. This local silencing precedes 'systemic silencing', in which GFP expression is suppressed in newly emerging, non-infiltrated leaves of the GFP transgenic plants (Voinnet *et al.*, 1998).

Two classes of siRNA

Local and systemic silencing of GFP in *N.benthamiana* is associated with two size classes of 21–25 nt GFP siRNA (Figure 1), corresponding to both sense and antisense strands. In the extracts of local silencing tissue, the longer siRNAs (25 nt) are as abundant as the shorter species (Figure 1B and C, lanes 4 and 8, respectively) whereas in the systemic silencing tissue the shorter siRNAs (21–22 nt) are by far the more abundant species (Figure 1B and C, lanes 3 and 7, respectively). We can rule out that the different size classes of siRNA are artefacts caused by contaminants in the RNA preparations because labelled marker RNA had the same electrophoretic mobility, irrespective of whether it was analysed alone or after mixing with the plant RNA (data not shown).

The effect of viral suppressors on local silencing

The onset of GFP silencing after agro-infiltration provides a convenient system for induction of siRNA. Furthermore, by infiltrating mixed *A.tumefaciens* cultures, the effect of viral suppressors on GFP siRNA, mRNA and systemic silencing can be assessed. In such an experiment one of the

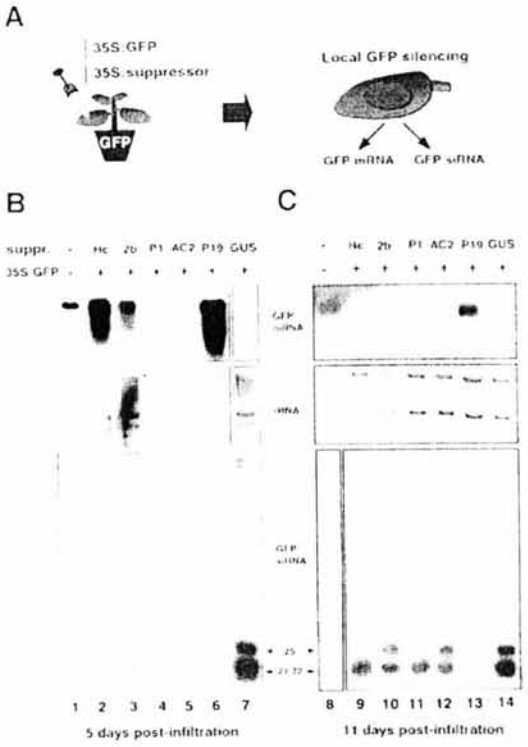


Fig. 2. The effects of viral suppressors on local RNA silencing. (**A**) Local silencing was induced in leaves of GFP transgenic *N.benthamiana* by infiltration of the 35S–GFP strain of *Agrobacterium* together with a second strain designed to express a viral suppressor of silencing. The onset of local silencing was then monitored and samples were collected for GFP mRNA and siRNA analysis. RNA was extracted from the infiltrated leaves after 5 (**B**) and 11 days (**C**). GFP mRNA was detected by hybridization with ^{32}P-labelled GFP cDNA. GFP siRNA was detected by hybridization with labelled GFP sense RNA. The length of the siRNA is indicated as 21–22 nt (short class) and 25 nt (long class). Lanes 2–6 and 9–12 are from plants in which the GFP *A.tumefaciens* was mixed 1:1 with a second culture expressing a viral suppressor. The suppressors were Hc-Pro (HC, lanes 2 and 9), 2b (lanes 3 and 10), P1 (lanes 4 and 11), AC2 (lanes 5 and 12) and P19 (lanes 6 and 13). Lanes 7 and 14 are from plants in which the second *A.tumefaciens* culture carried the 35S–GUS transgene construct. Lanes 1 and 8 correspond to non-infiltrated control leaves.

T-DNA constructs carries the 35S–GFP transgene as an initiator of silencing and the second encodes a viral suppressor (Figure 2A). The cultures are infiltrated into the leaves of GFP transgenic *N.benthamiana* and the plants are monitored for local and systemic GFP silencing. This approach has been used previously to investigate suppressors of silencing encoded in several viruses (Voinnet *et al.*, 2000; Johansen and Carrington, 2001; Dunoyer *et al.*, 2002; Guo and Ding, 2002).

The viral suppressors tested include the P1 protein of rice yellow mottle virus (RYMV), the P19 protein of tomato bushy stunt virus (TBSV), the helper component protease (Hc-Pro) of potato virus Y (PVY), the 2b protein of cucumber mosaic virus (CMV) and the AC2 protein of African cassava mosaic virus (ACMV) (Voinnet *et al.*, 1999). We used an *Agrobacterium* culture carrying a 35S-β glucuronidase (GUS) transgene as a non-suppressor control. Both GFP mRNA and GFP siRNA levels were monitored at 5 and 11 days post-infiltration (d.p.i.) in three separate experiments.

Figure 2B and C illustrates that both GFP siRNA size classes accumulated in tissues infiltrated with a GFP *Agrobacterium* strain in the absence of a viral suppressor (lanes 7 and 14, respectively). In contrast, in all GFP combinations with viral suppressor constructs, the accumulation of siRNA was reduced, although to different extents (lanes 2–6 and 8–12). The strongest effect, with P19, resulted in suppressed accumulation of both long and short siRNAs for at least 11 days (lanes 6 and 13). Hc-Pro suppressed both siRNA classes at 5 d.p.i., but by 11 d.p.i. there was accumulation of the smaller class only (lanes 2 and 9). P1 suppressed accumulation of the longer class of siRNA throughout the 11-day duration of the experiments, but caused only a moderate reduction in the shorter class (lanes 4 and 11). AC2 and 2b were the weakest suppressors and caused a similar, moderate reduction of both siRNA classes (lanes 5 and 12 and lanes 3 and 10, respectively).

The GFP mRNA levels were inversely related to the abundance of the short siRNAs. Thus, in the samples without a viral suppressor, agro-infiltration of 35S-GFP induced a high level of short siRNAs, and the GFP mRNA from the stable integrated transgene and from the transiently expressed DNA was below the limit of detection (Figure 2B and C, lanes 7 and 14, respectively). In samples with intermediate levels of the short siRNA due to the 2b, P1 and AC2 suppressors (Figure 2B, lanes 3–5), the GFP mRNA at 5 d.p.i. was at approximately the same level or was less abundant than in the non-silenced controls (Figure 2B, lane 1). However, by 11 d.p.i., with these intermediate suppressors, the GFP mRNA silencing was as strong as in the absence of viral suppressors (Figure 2C, lanes 10–12). The 11 d.p.i. sample with Hc-Pro also represented the intermediate situation in which the short siRNA class was moderately abundant and the GFP mRNA was markedly reduced (Figure 2C, lane 9). The extreme situation, in which short siRNA was reduced to levels that were at or close to the detection limit, was with Hc-Pro at 5 d.p.i. and with p19 at both time points. In these samples (Figure 2B, lanes 2 and 6, and Figure 2C, lane 13), there was strong suppression of silencing and the GFP mRNA was more abundant than in the non-silenced plants (Figure 2B and C, lanes 1 and 8, respectively).

In contrast, there was no obvious relationship between the accumulation of the long siRNA and GFP mRNA levels. Thus, GFP mRNA was suppressed if the long siRNAs were detectable, as in the 2b and AC2 samples (Figure 2B, lanes 3 and 5, and Figure 2C, lanes 10 and 12). The GFP mRNA was also suppressed if the longer siRNAs were not detected, as in the Hc-Pro (11 d.p.i.) and P1 samples (Figure 2B, lane 2, and Figure 2C, lanes 9 and 11). Taken together, these RNA analyses indicate that the short siRNA but not the long siRNA has a role in local silencing.

The effect of viral suppressors on systemic silencing

Systemic silencing of GFP in the upper leaves of the infiltrated plants (Figure 3) was strongly inhibited when the 35S–GFP trigger of local silencing was co-expressed with P1, Hc-Pro or P19. It was striking that the suppressors were those affecting production of the longer siRNA (Figure 2B, lane 2, 4 and 6, and Figure 2C, lanes 9, 11 and

13). Conversely, the suppressors that had either a slight (AC2) or moderate (2b) effect on systemic silencing (Figure 3) in our experimental conditions were those with only a slight effect on the level of the long siRNA (Figure 2B, lanes 3 and 5, and Figure 2C, lanes 10 and 12). These results were consistently reproduced in three independent experiments involving 10 plants each. They suggested a correlation between the production of the long GFP siRNA species in the infiltrated patch and the subsequent onset of systemic silencing.

To investigate this possible correlation between the longer siRNA and systemic silencing further, we exploited mutant forms of PVX incorporating a GFP gene (PVX–GFP) (Figure 4A). Replication-competent forms of PVX–GFP are able to initiate local silencing of the GFP mRNA and siRNA production in GFP transgenic *N.benthamiana*. However, only mutant forms of PVX–GFP in which the P25 movement protein gene had been deleted are able to initiate systemic silencing, indicating that P25 is a suppressor of the systemic signal but not of the local silencing triggered by virus replication (Voinnet *et al.*, 2000).

We have now extended these analyses using conditions that, unlike those used previously (Voinnet *et al.*, 2000), allow resolution of long and short siRNAs. We inoculated GFP transgenic *N.benthamiana* with forms of PVX–GFP that either did or did not encode P25, and monitored production of siRNAs and systemic silencing in the infected plants. The PVX constructs were incapable of moving out of the inoculated cells because they were defective for coat protein (Figure 4A) (Voinnet *et al.*, 2000). Consequently, any systemic silencing would be due to spread of the silencing signal rather than the virus.

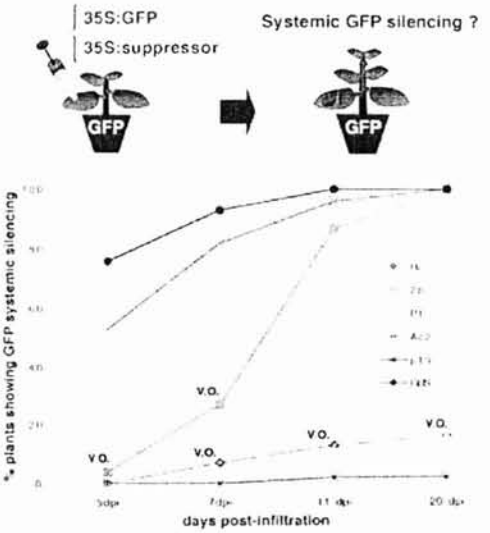


Fig. 3. The effects of viral suppressors on systemic RNA silencing. GFP transgenic *N.benthamiana* plants were infiltrated with mixed *A.tumefaciens* cultures as described in Figure 2, and were monitored for the onset of systemic silencing at 5, 7, 11 and 20 d.p.i. For each experimental treatment, a total of 30 individual seedlings (three separate experiments with 10 plants each) were tested. Abbreviations for the viral suppressors are as shown in the legend to Figure 2. V.O., veins only: this indicates that systemic silencing was incomplete and limited to the veins of a few leaves.

The results with these constructs show that there was systemic silencing and long siRNA only if the PVX–GFP had a deletion that included the P25 gene (Figure 4B, lanes 1 and 3). If the P25 gene was included as part of the PVX–GFP genome (Figure 4B, lane 2) or was provided *in trans*, there was suppression of systemic silencing and the longer siRNA was absent (Figure 4B, lane 4). These

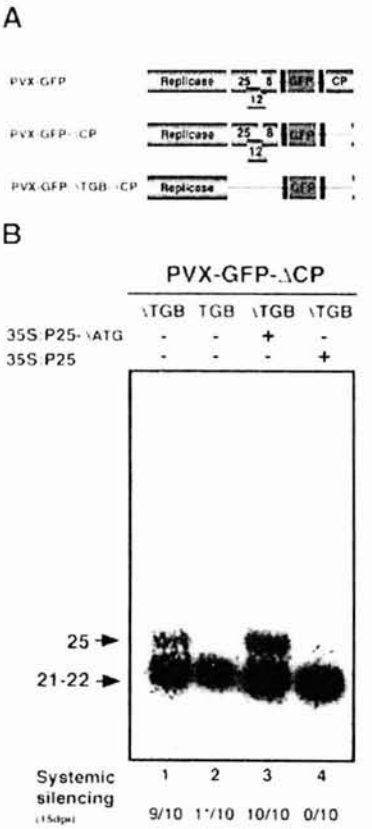


Fig. 4. The effects of PVX P25 on siRNA accumulation and systemic RNA silencing. (**A**) PVX–GFP has been described previously (Ruiz *et al.*, 1998; Voinnet *et al.*, 2000). Expression of the inserts in the PVX vector is controlled by a duplicated coat protein (CP) promoter, as indicated in blue. The replicase ORF is essential for viral replication; the 25, 12 and 8 kDa proteins are all strictly required for viral cell-to-cell movement and are collectively referred to as triple gene block (TGB) proteins. The CP is essential for encapsidation as well as cell-to-cell and systemic movement. PVX–GFP–ΔCP carries a deletion spanning the entire CP ORF. PVX–GFP–ΔTGB–ΔCP is based on PVX–GFP–ΔCP and carries a deletion spanning the entire TGB (Voinnet *et al.*, 2000). These PVX–GFP mutants were inserted between the 35S promoter and terminator of the pBin61 T-DNA, and viral inocula were provided to plants by *Agrobacterium*-mediated transient expression of the above-mentioned T-DNAs. (**B**) Four-week-old seedlings of GFP transgenic *N.benthamiana* were inoculated with PVX–GFP-based replicons. After 5 days, RNA was extracted from the inoculated leaf and antisense GFP siRNA was assayed as described in Figure 2. The onset and progression of systemic GFP silencing was monitored in 10 plants for each inoculum and the number of plants showing systemic silencing is indicated. The asterisk at the bottom of lane 2 indicates partial systemic silencing that was only restricted to a few leaf veins in this particular plant. The viral inocula were either PVX–GFP–ΔTGB–ΔCP (lanes 1, 3 and 4) or PVX–GFP–ΔCP (lane 2). The 35S–P25–ΔATG (lane 3) and 35S–P25 (lane 4) constructs were transiently co-expressed in the inoculated tissue, as described in Figure 2.

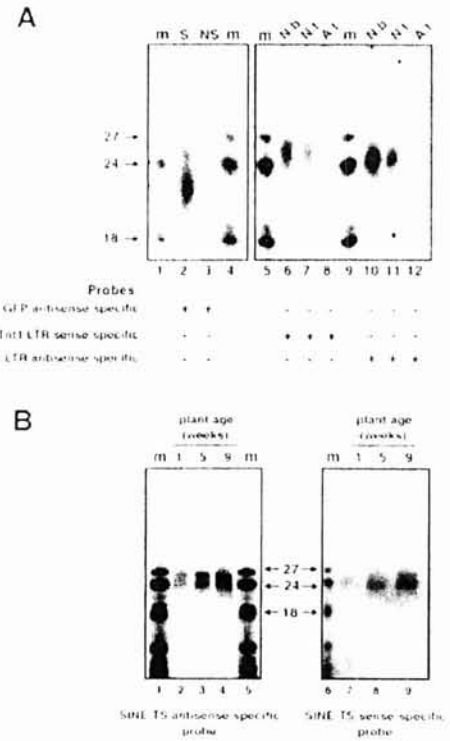
data therefore reinforce the association of the long siRNA with systemic silencing. They further suggest that P25 blocks systemic silencing by interfering with production of this long siRNA.

Retrotransposon siRNA

We also investigated the size distribution of endogenous plant siRNAs. Although there are no known natural targets of RNA silencing in plants, it was originally suggested (Flavell, 1994) that the mechanism is a natural defence against transposable elements in plants. This idea is supported by recent molecular and genetic evidence from animals and lower plants (Ketting *et al.*, 1999; Tabara *et al.*, 1999; Wu-Scharf *et al.*, 2000; Aravin *et al.*, 2001).

To find out whether plants also have transposon siRNA-like species, we probed extracts of *Arabidopsis thaliana* and *Nicotiana* species for siRNAs corresponding to three different retroelements. In each instance, as shown below, we detected RNA of both sense and antisense polarities and of a size similar to transgene siRNA (Figures 5 and 6). However, unlike transgene siRNA, in each case the small RNA corresponded only to the longer size class.

The Tnt1 element in *Nicotiana tabacum* is an active retrotransposon (Grandbastien *et al.*, 1989). Sense and antisense specific probes from the long terminal repeat



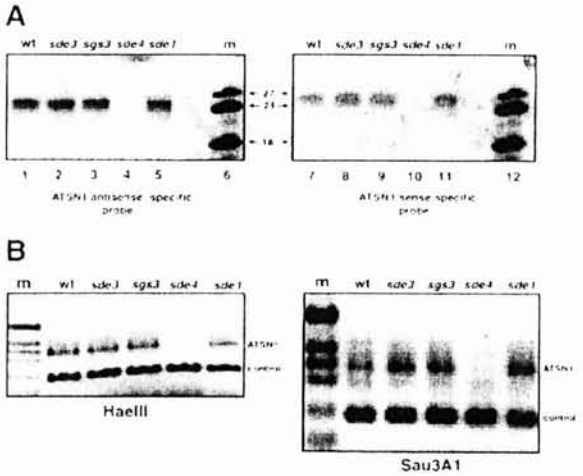


Fig. 6. Retroelement siRNA and DNA methylation in *Arabidopsis*. (**A**) RNA was extracted from wild-type and silencing-defective mutants of 6-week-old *A.thaliana* (landrace C24) seedlings. The mutations were at the *sde1/sgs2*, *sgs3*, *sde3* and *sde4* loci (Dalmay *et al.*, 2000, 2001; Mourrain *et al.*, 2000). AtSN1 siRNA was detected by hybridization with sense RNA transcribed from the cloned AtSN1 sequence. After probe stripping, the filter was reprobed with antisense RNA from the cloned AtSN1 sequence. (**B**) DNA was extracted from the lines of *A.thaliana* as in (A) and digested with the methylation-sensitive enzymes *Hae*III or *Sau*3AI. Similar extent of digestion was confirmed by electrophoresis and ethidium staining of digested DNA. Each sample of digested DNA was used as template for PCR with AtSN-specific primers and primers that would amplify an unrelated sequence (to control for variations in efficiency of individual PCR). The AtSN primers would only amplify AtSN DNA if it had remained undigested.

Fig. 5. Retroelement siRNA in *Nicotiana* sp. (**A**) Low molecular weight RNA from *N.benthamiana* (lanes 'N.b.'); 5-week-old *N.tabacum* var. Petite Havana (lanes 'N.t.') and 6-week-old *A.thaliana* landrace C24 (lanes 'A.t.') were hybridized with ^{32}P-labelled Tnt1 LTR sense or antisense RNA. Samples in lanes 'S' and 'NS' were as described in Figure 1 and the siRNA was detected by hybridization with a GFP antisense-specific probe. (**B**) RNA was extracted from the leaves of 1-, 5- and 9-week-old *N.tabacum* var. Petite Havana. Low molecular weight RNA was hybridized with a ^{32}P-labelled sense or antisense RNA corresponding to a cloned TS SINE. The two panels show hybridization to the same filter stripped of probe between hybridizations. (**C**) Leaves of *N.benthamiana* were inoculated with TRV:00 or TRV:TS and total RNA was extracted from the inoculated leaves after 2, 4 and 7 days. RNA was hybridized with a ^{32}P-labelled, 617-bp DNA fragment of TRV RNA2 generated by *Eco*RI digestion of vector pTV00. This fragment corresponds to the 3′ UTR of RNA2 and thus cross-hybridizes with RNA1. The multiple RNA species hybridizing to the probe are genomic and subgenomic RNAs that are typically associated with TRV infection (Ratcliff *et al.*, 2001). The size shift of the TRV:TS RNA species (lanes 7–9) confirms detection of the recombinant virus.

(LTR) of this sequence detected 24–26 nt RNA in *N.tabacum* (var. Samsun) and *N.benthamiana* (Figure 5A, lanes 6, 7, 10 and 11) but not in *A.thaliana* (Figure 5A, lanes 8 and 12), which does not harbour elements of the Tnt1 family. These longer siRNAs corresponded, in equal abundance, to both sense and antisense strands (Figure 5A) and were also detected using probes from the *Gag* gene of Tnt1 (data not shown). We did not detect Tnt1 siRNA of the small class (21–22 nt).

To determine whether siRNA was produced from other retroelements, we used as a probe the TS SINE element of tobacco (Yoshioka *et al.*, 1993). TS siRNAs in *N.benthamiana* and *N.tabacum* were detected with both sense and antisense probes corresponding to the entire length of a cloned TS element (Figure 5B). As with the Tnt1 element, these TS siRNAs corresponded only to the 24–26 nt long class. We infer that the TS siRNAs are abundant because they were detected with the same short fluorographic exposure times used to detect GFP siRNA in systemically silenced leaves (Figure 5B).

The siRNAs associated with transgene silencing mediate resistance against virus infection. Thus, *N.benthamiana* plants exhibiting virus-induced and transgene-induced systemic silencing of a GFP transgene are resistant against recombinant viruses carrying a GFP insert (Voinnet and Baulcombe, 1997; Ruiz *et al.*, 1998). However, the TS siRNAs did not confer resistance to *N.benthamiana* against recombinant tobacco rattle virus carrying a TS insert (TRV:TS), because this vector was able to accumulate as rapidly and as extensively as the recombinant TRV without an insert (Figure 5C, lanes 4–6 compared with lanes 7–9).

To test whether abundant siRNA production is a general feature of SINEs, we used the AtSN1 of *Arabidopsis* as a probe. As with the TS SINE, both sense and antisense siRNA (Figure 6A) were found but detection required much longer exposure time than for the tobacco TS element, suggesting that they are less abundant. The strong siRNA signal obtained with the TS probe is more likely a reflection of the high copy number of the TS element [50 000 copies per genome (Yoshioka *et al.*, 1993) compared with 70 copies of the AtSN1 (Myouga *et al.*, 2001)]. RNA blot analysis indicated the presence of sense and antisense AtSN1 siRNAs (Figure 6A) which, as with the Tnt1 and TS elements, were of the longer class. These AtSN1 siRNAs were present in wild-type plants and in *sde1/sgs2*, *sgs3* and *sde3* mutants that are defective for transgene silencing (Elmayan *et al.*, 1998; Dalmay *et al.*, 2000) (Figure 6A). However, in an *sde4* mutant background, the AtSN1 siRNAs were absent (Figure 6A, lanes 4 and 10). Based on these results it seemed likely that retroelement siRNAs are produced by mechanisms that are similar but not identical to those involved in transgene RNA silencing.

RNA silencing of transgenes and retroelements in plants is often associated with sequence-specific methylation (Ingelbrecht *et al.*, 1994; English *et al.*, 1996; Jones *et al.*, 1999; Mette *et al.*, 2000; Morel *et al.*, 2000; Miura *et al.*, 2001). Because siRNAs are strong candidates for the molecules that direct methylation, we tested whether methylation of an AtSN1 element was affected in the *sde4* mutant. The methylation analysis of AtSN1 DNA was carried out by PCR of genomic DNA from wild-type and silencing-defective mutants of *A.thaliana*. The primers were based on the *A.thaliana* (landrace Columbia) DNA that flanked an AtSN1 insertion and the PCR was carried out on DNA that had been digested independently with two separate methylation-sensitive enzymes, *Sau*3A1 and *Hae*III. These enzymes cut within the AtSN1 sequence so that if the template DNA was unmethylated, the amount of PCR product would be less than with methylated DNA. As an internal control, the PCR also included a pair of primers for a DNA fragment that did not contain *Sau*3A1 or *Hae*III sites. Figure 6B shows the results of this analysis and confirms that in the *sde4* genotype, the AtSN1 DNA was less methylated than in wild-type plants or any of the other silencing-defective mutants tested. Thus, based on this *sde4* phenotype, there is a correlation of long AtSN1 siRNA and methylation of the corresponding DNA.

Discussion

Two classes of siRNA in plants

In this paper we demonstrate that the short RNAs associated with transgene RNA silencing are heterogenous in both size and function. Thus, these RNAs are not a single class of ~25 nt, as reported previously (Hamilton and Baulcombe, 1999), but instead are two distinct species, co-migrating with 21–22 nt and 25 nt markers. The discrepancy between the present and the previous analysis is most likely due to improved electrophoretic resolution, shorter exposure times and the nature of the size markers. We previously used DNA oligonucleotide markers migrating ~10% faster than the RNA markers used here (Sambrook *et al.*, 1989). We refer to the two classes as short and long siRNAs. However, until we have sequenced these molecules and determined their length, we acknowledge that the apparent size difference may be due to RNA modifications affecting electrophoretic mobility.

In the presence of viral suppressors, the short GFP siRNAs correlated with degradation of the target GFP mRNA (Figure 2). We propose, therefore, that this 21–22 nt siRNA represents the siRNA that guides the RISC ribonuclease to the target of RNA silencing (Zamore *et al.*, 2000; Elbashir *et al.*, 2001a). The long class of siRNA is similar to the short class in that it corresponds to both RNA strands (Figure 1) and, therefore, is probably derived from dsRNA. However, it is unlikely that the two classes of siRNA have the same function because they accumulate differentially in locally and systemically silenced tissue or in the presence of viral suppressor proteins (Figures 1–3). In addition, there are two lines of evidence indicating that the long siRNA is not the guide for RISC. First, in our assay of the P1 and Hc-Pro suppressors, there was silencing of a GFP target RNA in the absence of long GFP siRNA (Figure 2B and C). The second line of evidence is based upon experiments with TRV:TS showing that targeted RNA degradation does not occur in the presence of abundant, long TS siRNA (Figure 5C). Consistent with the idea that the long siRNA is not a guide for RISC, synthetic siRNAs are inactive in an RISC assay if they are longer than 23 nt (Elbashir *et al.*, 2001b).

Our finding that long siRNA is associated with both *SDE4*-mediated DNA methylation (Figure 6B) and sys-

temic silencing (Figures 3 and 4) could mean that this class of siRNA is directly involved in these processes. For example, it could be that long siRNA is the systemic signal of silencing. This long siRNA molecule could also mediate the nucleotide sequence-specific methylation of DNA often associated with systemic RNA silencing in plants. However, our data could also accommodate the possibility that the long siRNA is either a derivative or precursor of the RNAs species that are directly involved in systemic silencing and RNA-directed DNA methylation.

To assess the precise function of the different classes of siRNA in plants, it will be necessary to establish the ways in which these RNAs differ physically and whether they are produced by the same Dicer activity. The *Arabidopsis* genome encodes at least four proteins with the helicase, PAZ and dsRNA binding domains that are characteristic of the *Drosophila* Dicer (Y.Klaue and D.Baulcombe, unpublished data), and one possibility is that different members of this protein family produce the short and long siRNAs. Loss for each of these proteins will allow us to assess this scenario. Loss of the true Dicer function would lead to loss of the primary RNA silencing mechanism. Suppression of Dicer paralogues would leave the primary silencing mechanism intact but may prevent production of the long siRNA. If systemic signalling and DNA methylation are dependent upon long siRNA, the loss of Dicer paralogue function would lead to suppression of these aspects of RNA silencing. Other RNA silencing processes such as amplification and transitivity (Sijen *et al.*, 2001a; Vaistij *et al.*, 2002) may also be dependent on the long siRNA.

RNA silencing and the control of retroelements

Mutations in *C.elegans*, *Drosophila* and *Chlamydomonas* that impair RNA silencing of transgenes or host regulatory genes also result in elevated activity of transposable elements (Ketting *et al.*, 1999; Wu-Scharf *et al.*, 2000; Aravin *et al.*, 2001). Based on these observations, there is now widespread agreement that an important natural function of RNA silencing is to restrict transposons (Plasterk and Ketting, 2000; Sharp and Zamore, 2000; Hammond *et al.*, 2001b). It is somewhat ironic that although RNA silencing was discovered in higher plants, (Napoli *et al.*, 1990), there has been no direct evidence of such a function. The data presented here provide this evidence by demonstrating the presence of siRNA corresponding to three different retrotransposons in *Nicotiana* spp. and *A.thaliana* (Figures 5 and 6). It will be interesting to determine whether the *sde4* mutant, which does not accumulate siRNA of at least one type of retroelement, has a mutator phenotype like *mut7* and other silencing-defective lines of *C.elegans*.

In *C.elegans*, there is only partial overlap in the RNA silencing-related mechanisms that control retroelement activity and mediate dsRNA interference. Some mutations suppress both processes whereas others affect only one (Ketting *et al.*, 1999; Tabara *et al.*, 1999). Clearly there are degrees of complexity in RNA silencing that are not apparent from the well studied dsRNA/Dicer/RISC-based process. In plants there is probably a similar degree of complexity because both classes of siRNA are associated with transgene silencing, whereas retrotransposon siRNA belongs only to the long class. Similarly, of the four loci identified in a mutant screen as being required for transgene-induced RNA silencing, only *sde4* has any effect on AtSN1 siRNA. Moreover, *sgs2* and *sde1*, *-3* and *-4* interfered with transgene methylation (Dalmay *et al.*, 2000), but only *sde4* interfered with methylation of AtSN1 DNA (Figure 6). Presumably, the classes of siRNA or RNA silencing genes are not involved equally in different branches of the silencing pathway. Some of the complexity may simply be a result of compartmentalization with, for example, RNA turnover degradation in the cytoplasm and DNA methylation in the nucleus (Mette *et al.*, 2001).

Retrotransposon DNA in plants is often methylated and transcriptionally silent (Hirochika *et al.*, 2000; Miura *et al.*, 2001). To account for this situation it is likely, from the data here and from prior analyses of RNA-mediated transcriptional silencing (Mette *et al.*, 2000; Sijen *et al.*, 2001b), that the transcriptional inactivation involves RNA silencing. A double-stranded form of the retroelement RNA would be processed into siRNA and there would be RNA-mediated methylation of the corresponding DNA. There is little information about both the transcriptional status and genomic organization of the three elements we tested, and we cannot predict how the putative dsRNA precursor of the siRNAs arose in each case. One possibility is that an RNA-dependent RNA polymerase converted retrotransposon RNA into double-stranded form. The dsRNA derived in this way would include promoter sequences of the retroelement and could target transcription.

A second model to account for retroelement dsRNA invokes transcription through inverted repeat retroelements. In this second model, transcription would be initiated from either an adjacent host promoter or within the retroelement, and the dsRNA would be produced independently of an RdRP. However, irrespective of the mechanism for retroelement dsRNA production, our data show that RNA silencing has the potential to protect plant genomes against foreign DNAs. Given the abundance of retroelements in many plant genomes, the activity of RNA silencing is likely to have played a major role in determining plant genome size and structure.

Materials and methods

Plant and Agrobacterium material, and growth conditions

Wild-type *N.tabacum* (var. Samsun and Petite Havana), *N.benthamiana* and the GFP transgenic 16c line of *N.benthamiana* were grown as described previously (Marano and Baulcombe, 1998; Voinnet *et al.*, 1998). *Arabidopsis thaliana* landrace C24 and transgenic and mutant lines derived from this were grown as described previously (Dalmay *et al.*, 2000). The previously described *sde2* mutant is allelic with *sgs3* (Mourrain *et al.*, 2000) (A.Herr, personal communication). GFP fluorescence was monitored visually under ultra-violet epi-illumination (Voinnet *et al.*, 1998). Recombinant *A.tumefaciens* strain C58C1 was grown to stationary phase at 29°C in L-broth with 50 μg/ml kanamycin and 5 μg/ml tetracycline, collected by centrifugation (5000 *g* for 15 min at 20°C) and re-suspended in 10 mM $MgCl_2$ and 150 μg/ml acetosyringone. Cells were left in this medium for 3 h and then infiltrated into the abaxial airspaces of 2- to 4-week-old *N.benthamiana* plants. All the *Agrobacterium* strains used harboured the pCH32 helper plasmid (Hamilton *et al.*, 1996) and were used undiluted in the experiments described here, except for the P25 strain which was maintained at an optical density (OD_{600}) of 1.0 to avoid toxicity (Voinnet *et al.*, 2000).

Constructs for transient expression, viral constructs and viral inoculation procedures

35S–GFP, 35S–25K, 35S–25KΔATG, PVX–GFP–ΔCP and PVX–GFP–ΔTGB–ΔCP constructs were described previously (Voinnet *et al.*, 2000).

The other suppressor cDNAs were amplified with primer sequences (available on request) from previously described plasmids (Brigneti *et al.*, 1998; Voinnet *et al.*, 1999) and inserted into the *Sma*I site of pBin61 (Bendahmane *et al.*, 2000), and confirmed by sequencing. The control 35S–GUS construct was from Jonathan Jones (Sainsbury Laboratory). TRV:TS is based on recombinant, bipartite tobacco rattle virus, (Ratcliff *et al.*, 2001). RNA1 encodes the viral replicase, a movement protein (MP) and a 16 kDa protein of unknown function. The replicase open reading frame (ORF) contains an intron (int). RNA2 encodes the viral coat protein (CP) and was modified to carry the TS SINE insert as a 3′ transcriptional fusion. RNA1 and RNA2 share nucleotide sequence homology in their 3′ UTRs. The cloned RNA1 of TRV:TS was inserted between the 35S promoter and the terminator of the pBin61 T-DNA. The cloned TRV:TS RNA2 was inserted between the 35S promoter and the Nos terminator of the pGreen binary vector T-DNA (Hellens *et al.*, 2000). Viral inocula were provided to plants by *Agrobacterium*-mediated transient co-expression of the above-mentioned T-DNAs.

RNA isolation

Leaves were harvested and frozen in liquid nitrogen, ground to a fine powder and mixed with RNA extraction buffer (4 M guanidine isothiocyanate, 25 mM Na Citrate pH 7, 0.1% Sarkosyl) (Figures 1, 5 and 6). The resulting slurries were extracted twice with phenol/chloroform/isoamylalcohol (equilibrated at pH 8; Sigma). Alternatively (Figures 2 and 4), the frozen leaf powder was mixed with Tri-reagent (Sigma), extracted once with 1/5 volume of chloroform and then once with an equal volume of phenol/chloroform/isoamylalcohol. RNA was precipitated with the addition of an equal volume of isopropanol and incubated at −20°C for 30 min. mRNA and rRNA (high molecular weight RNAs) were precipitated with 10% polyethylene glycol (mol. wt 8000), 0.5 M NaCl (4°C for 30 min) and redissolved in 50% formamide. Low molecular weight RNAs including siRNA were precipitated from the PEG/NaCl supernatant (3 vol. ethanol at −20°C for 2 h) and redissolved in 50% formamide. Relative quantification of the low molecular weight fraction was by ethidium bromide staining.

RNA analysis

High molecular weight RNA was analysed as described previously (Figures 1, 2 and 4). For the experiments shown in Figure 5, high molecular weight RNA in 50% formamide was denatured with 3.7% formaldehyde in 1× TAE and separated by electrophoresis in a 0.8% agarose/1× TAE gel run at 50°C. Analysis of siRNA was as described previously (Hamilton and Baulcombe, 1999). Low molecular weight RNA markers were produced by RNase T1 digestion of ^{32}P-labelled GFP sense *in vitro* transcripts yielding fragments of 27 nt, 24 nt (×2), 18 nt and 15 nt. Strand-specific hybridization controls were oligonucleotides spotted onto hybridization membranes and hybridized under the same conditions as described for northern blotting. All hybridization signals were detected by phosphorimaging.

Retroelements

Retroelements corresponding to Tnt1 LTR from *N.tabacum* (var. Samsun), the TS element (Yoshioka *et al.*, 1993) from *N.tabacum* (var. Samsun) and the AtSN1 sequence of *A.thaliana* were PCR-amplified and cloned into plasmid vectors. Direct sequence analysis showed that the Tnt1 sequence corresponded to DDBJ/EMBL/GenBank accession No. X13777, that the TS clone corresponded to the TSa subfamily of SINEs (Yoshioka *et al.*, 1993) and that the AtSN1 sequence corresponded most closely to genomic sequence on BAC clone T15B3.

DNA methylation analysis

Genomic DNA from 5-week-old plants (DNeasy plant DNA extraction kit; Qiagen) was digested with *Hae*III or *Sau*3AI (New England Biolabs). The sequence flanking the cloned AtSN1 insertion was amplified by PCR (30 cycles at 95°C for 30 s, at 55°C for 15 s and at 72°C for 30 s) with primers (ACTTAATTAGCACTCAAATTAAACAAAATAAGT; TTTAAACATAARAARAARTTCCTTTTTCATCTAC) from ~400 ng of this DNA. Internal PCR control primers based on the *Arabidopsis* (landrace Columbia) sequence *At2g19920* (A.Herr, personal communication) span a region lacking *Hae*III or *Sau*3AI sites.

Acknowledgements

We thank members of our laboratory for providing valuable comments on aspects of this manuscript, and Alan Herr and Yvonne Klaue for unpublished data. We also thank Mike Hill and his team for excellent plant care. This work was funded by the Gatsby Charitable Foundation, the Biotechnology and Biological Sciences Research Council and from the Royal Society (Dorothy Hodgkin Fellowship to O.V.). The use of genetically modified plant viruses was licensed by DEFRA [license No. PHL 24B/3654 (3/2001)].

References

Anandalakshmi,R., Pruss,G.J., Ge,X., Marathe,R., Smith,T.H. and Vance,V.B. (1998) A viral suppressor of gene silencing in plants. *Proc. Natl Acad. Sci. USA*, **95**, 13079–13084.

Anandalakshmi,R., Marathe,R., Ge,X., Herr,J.M., Mau,C., Mallory,A., Pruss,G., Bowman,L. and Vance,V.B. (2000) A calmodulin-related protein that suppresses posttranscriptional gene silencing in plants. *Science*, **290**, 142–144.

Aravin,A.A., Naumova,N.M., Tulin,A.V., Vagin,V.V., Rozovsky,Y.M. and Gvozdev,V.A. (2001) Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the *D. melanogaster* germline. *Curr. Biol.*, **11**, 1017–1027.

Bendahmane,A., Querci,M., Kanyuka,K. and Baulcombe,D.C. (2000) *Agrobacterium* transient expression system as a tool for isolation of disease resistance genes: application to the *Rx2* locus in potato. *Plant J.*, **21**, 73–81.

Bernstein,E., Caudy,A.A., Hammond,S.M. and Hannon,G.J. (2001) Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature*, **409**, 363–366.

Brigneti,G., Voinnet,O., Li,W.X., Ji,L.H., Ding,S.W. and Baulcombe,D.C. (1998) Viral pathogenicity determinants are suppressors of transgene silencing in *Nicotiana benthamiana*. *EMBO J.*, **17**, 6739–6746.

Dalmay,T., Hamilton,A.J., Rudd,S., Angell,S. and Baulcombe,D.C. (2000) An RNA-dependent RNA polymerase gene in *Arabidopsis* is required for posttranscriptional gene silencing mediated by a transgene but not by a virus. *Cell*, **101**, 543–553.

Dalmay,T.D., Horsefield,R., Braunstein,T.H. and Baulcombe,D.C. (2001) SDE3 encodes an RNA helicase required for post-transcriptional gene silencing in *Arabidopsis*. *EMBO J.*, **20**, 2069–2078.

Dunoyer,P., Pfeffer,S., Fritsch,C., Hemmer,O., Voinnet,O. and Richards,K.E. (2002) Identification, subcellular localization and some properties of a cysteine-rich suppressor of gene silencing encoded by peanut clump virus. *Plant J.*, **29**, 555–567.

Elbashir,S.M., Lendeckel,W. and Tuschl,T. (2001a) RNA interference is mediated by 21- and 22-nucleotide RNAs. *Genes Dev.*, **15**, 188–200.

Elbashir,S.M., Martinez,J., Patkaniowska,A., Lendeckel,W. and Tuschl,T. (2001b) Functional anatomy of siRNAs for mediating efficient RNAi in *Drosophila melanogaster* embryo lysate. *EMBO J.*, **20**, 6877–6888.

Elmayan,T. *et al.* (1998) *Arabidopsis* mutants impaired in cosuppression. *Plant Cell*, **10**, 1747–1757.

English,J.J., Mueller,E. and Baulcombe,D.C. (1996) Suppression of virus accumulation in transgenic plants exhibiting silencing of nuclear genes. *Plant Cell*, **8**, 179–188.

Fire,A., Xu,S., Montgomery,M.K., Kostas,S.A., Driver,S.E. and Mello,C.C. (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature*, **391**, 806–811.

Flavell,R.B. (1994) Inactivation of gene expression in plants as a consequence of specific sequence duplication. *Proc. Natl Acad. Sci. USA*, **91**, 3490–3496.

Grandbastien,M.A., Spielmann,A. and Caboche,M. (1989) TNT1, a mobile retroviral-like transposable element of tobacco isolated by plant-cell genetics. *Nature*, **337**, 376–380.

Grishok,A. *et al.* (2001) Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control *C. elegans* developmental timing. *Cell*, **106**, 23–34.

Guo,H.S. and Ding,S.W. (2002) A viral protein inhibits the long range signaling activity of the gene silencing signal. *EMBO J.*, **21**, 398–407.

Hamilton,A.J. and Baulcombe,D.C. (1999) A novel species of small antisense RNA in post-transcriptional gene silencing. *Science*, **286**, 950–952.

Hamilton,C.M., Frary,A., Lewis,C. and Tanksley,S.D. (1996) Stable transfer of intact high molecular weight DNA into plant chromosomes. *Proc. Natl Acad. Sci. USA*, **93**, 9975–9979.

Hammond,S.M., Bernstein,E., Beach,D. and Hannon,G. (2000) An

RNA-directed nuclease mediates post-transcriptional gene silencing in *Drosophila* cell extracts. *Nature*, **404**, 293–296.

Hammond,S.M., Boettcher,S., Caudy,A.A., Kobayashi,R. and Hannon,G.J. (2001a) Argonaute2, a link between genetic and biochemical analyses of RNAi. *Science*, **293**, 1146–1150.

Hammond,S.M., Caudy,A.A. and Hannon,G.J. (2001b) Post-transcriptional gene silencing by double-stranded RNA. *Nat. Rev. Genet.*, **2**, 110–119.

Hellens,R.P., Edwards,E.A., Leyland,N.R., Bean,S. and Mullineaux,P.M. (2000) pGreen: a versatile and flexible binary Ti vector for *Agrobacterium*-mediated plant transformation. *Plant Mol. Biol.*, **42**, 819–832.

Hirochika,H., Okamoto,H. and Kakutani,T. (2000) Silencing of retrotransposons in arabidopsis and reactivation by the *ddm1* mutation. *Plant Cell*, **12**, 357–368.

Hutvagner,G., McLachlan,J., Pasquinelli,A.E., Balint,E., Tuschl,T. and Zamore,P.D. (2001) A cellular function for the RNA-interference enzyme Dicer in the maturation of the *let-7* small temporal RNA. *Science*, **293**, 834–838.

Ingelbrecht,I., Van Houdt,H., Van Montagu,M. and Depicker,A. (1994) Posttranscriptional silencing of reporter transgenes in tobacco correlates with DNA methylation. *Proc. Natl Acad. Sci. USA*, **91**, 10502–10506.

Johansen,L.K. and Carrington,J.C. (2001) Silencing on the spot. Induction and suppression of RNA silencing in the *Agrobacterium*-mediated transient expression system. *Plant Physiol.*, **126**, 930–938.

Jones,L., Hamilton,A.J., Voinnet,O., Thomas,C.L., Maule,A.J. and Baulcombe,D.C. (1999) RNA–DNA interactions and DNA methylation in post-transcriptional gene silencing. *Plant Cell*, **11**, 2291–2302.

Jones,L., Ratcliff,F. and Baulcombe,D.C. (2001) RNA-directed transcriptional gene silencing in plants can be inherited independently of the RNA trigger and requires Met1 for maintenance. *Curr. Biol.*, **11**, 747–757.

Kasschau,K.D. and Carrington,J.C. (1998) A counterdefensive strategy of plant viruses: suppression of post-transcriptional gene silencing. *Cell*, **95**, 461–470.

Ketting,R., Haverkamp,T., van Luenen,H. and Plasterk,R. (1999) *mut-7* of *C. elegans*, required for transposon silencing and RNA interference, is a homolog of Werner syndrome helicase and RNaseD. *Cell*, **99**, 133–141.

Ketting,R.F., Fischer,S.E.J., Bernstein,E., Sijen,T., Hannon,G.J. and Plasterk,R.H.A. (2001) Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in *C.elegans*. *Genes Dev.*, **15**, 2654–2659.

Llave,C., Kasschau,K.D. and Carrington,J.C. (2000) Virus-encoded suppressor of posttranscriptional gene silencing targets a maintenance step in the silencing pathway. *Proc. Natl Acad. Sci. USA*, **97**, 13401–13406.

Mallory,A.C. *et al.* (2001) HC-Pro suppression of transgene silencing eliminates the small RNAs but not transgene methylation or the mobile signal. *Plant Cell*, **13**, 571–583.

Marano,M.R. and Baulcombe,D. (1998) Pathogen-derived resistance targeted against the negative-strand RNA of tobacco mosaic virus: RNA strand-specific gene silencing? *Plant J.*, **13**, 537–546.

Mette,M.F., Aufsatz,W., van der Winden,J., Matzke,M.A. and Matzke,A.J.M. (2000) Transcriptional silencing and promoter methylation triggered by double-stranded RNA. *EMBO J.*, **19**, 5194–5201.

Mette,M.F., Matzke,A.J.M. and Matzke,M.A. (2001) Resistance of RNA-mediated TGS to HC-Pro, a viral suppressor of PTGS, suggests alternative pathways for dsRNA processing. *Curr. Biol.*, **11**, 1119–1123.

Mishra,K.K. and Handa,A.K. (1998) Post-transcriptional silencing of pectin methylesterase gene in transgenic tomato fruits results from impaired pre-mRNA processing. *Plant J.*, **14**, 583–592.

Miura,A., Yonebayashi,S., Watanabe,K., Toyama,T., Shimada,H. and Kakutani,T. (2001) Mobilization of transposons by a mutation abolishing full DNA methylation in *Arabidopsis*. *Nature*, **411**, 212–214.

Morel,J.B., Mourrain,P., Beclin,C. and Vaucheret,H. (2000) DNA methylation and chromatin structure affect transcriptional and post-transcriptional transgene silencing in *Arabidopsis*. *Curr. Biol.*, **10**, 1591–1594.

Mourrain,P. *et al.* (2000) *Arabidopsis SGS2* and *SGS3* genes are required for posttranscriptional gene silencing and natural virus resistance. *Cell*, **101**, 533–542.

Myouga,F., Tsuchimoto,S., Noma,K., Ohtsubo,H. and Ohtsubo,E. (2001) Identification and structural analysis of SINE elements in the *Arabidopsis thaliana* genome. *Genes Genet. Syst.*, **76**, 169–179.

Napoli,C., Lemieux,C. and Jorgensen,R.A. (1990) Introduction of a chimeric chalcone synthase gene into Petunia results in reversible co-suppression of homologous genes *in trans*. *Plant Cell*, **2**, 279–289.

Nykanen,A., Haley,B. and Zamore,P.D. (2001) ATP requirements and small interfering RNA structure in the RNA interference pathway. *Cell*, **107**, 309–321.

Palauqui,J.-C., Elmayan,T., Pollien,J.-M. and Vaucheret,H. (1997) Systemic acquired silencing: transgene-specific post-transcriptional silencing is transmitted by grafting from silenced stocks to non-silenced scions. *EMBO J.*, **16**, 4738–4745.

Plasterk,R.H.A. and Ketting,R.F. (2000) The silence of the genes. *Curr. Opin. Genet. Dev.*, **10**, 562–567.

Ratcliff,F., Martin-Hernandez,A.M. and Baulcombe,D.C. (2001) Tobacco rattle virus as a vector for analysis of gene function by silencing. *Plant J.*, **25**, 237–245.

Ruiz,M.T., Voinnet,O. and Baulcombe,D.C. (1998) Initiation and maintenance of virus-induced gene silencing. *Plant Cell*, **10**, 937–946.

Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Sharp,P.A. and Zamore,P.D. (2000) RNA interference. *Science*, **287**, 2431–2433.

Sijen,T., Fleenor,J., Simmer,F., Thijssen,K.L., Parrish,S., Timmons,L., Plasterk,R.H.A. and Fire,A. (2001a) On the role of RNA amplification in dsRNA-triggered gene silencing. *Cell*, **107**, 465–476.

Sijen,T., Vijn,I., Rebocho,A., van Blokland,R., Roelofs,D., Mol,J.N.M. and Kooter,J.M. (2001b) Transcriptional and posttranscriptional gene silencing are mechanistically related. *Curr. Biol.*, **11**, 436–440.

Tabara,H., Sarkissian,M., Kelly,W.G., Fleenor,J., Grishok,A., Timmons,L., Fire,A. and Mello,C.C. (1999) The *rde-1* gene, RNA interference and transposon silencing in *C. elegans*. *Cell*, **99**, 123–132.

Vaistij,F.E., Jones,L. and Baulcombe,D.C. (2002) Spreading of RNA targeting and DNA methylation in RNA silencing requires transcription of the target gene and a putative RNA-dependent RNA polymerase. *Plant Cell*, **14**, 857–867.

VanHoudt,H., Ingelbrecht,I., VanMontagu,M. and Depicker,A. (1997) Post-transcriptional silencing of a neomycin phosphotransferase II transgene correlates with the accumulation of unproductive RNAs and with increased cytosine methylation of 3′ flanking regions. *Plant J.*, **12**, 379–392.

Voinnet,O. (2001) RNA silencing as a plant immune system against viruses. *Trends Genet.*, **17**, 449–459.

Voinnet,O. and Baulcombe,D.C. (1997) Systemic signalling in gene silencing. *Nature*, **389**, 553.

Voinnet,O., Vain,P., Angell,S. and Baulcombe,D.C. (1998) Systemic spread of sequence-specific transgene RNA degradation is initiated by localised introduction of ectopic promoterless DNA. *Cell*, **95**, 177–187.

Voinnet,O., Pinto,Y.M. and Baulcombe,D.C. (1999) Suppression of gene silencing: a general strategy used by diverse DNA and RNA viruses. *Proc. Natl Acad. Sci. USA*, **96**, 14147–14152.

Voinnet,O., Lederer,C. and Baulcombe,D.C. (2000) A viral movement protein prevents systemic spread of the gene silencing signal in *Nicotiana benthamiana*. *Cell*, **103**, 157–167.

Wassenegger,M. and Pelissier,T. (1998) A model for RNA-mediated gene silencing in higher plants. *Plant Mol. Biol.*, **37**, 349–362.

Winston,W.M., Molodowitch,C. and Hunter,C.P. (2002) Systemic RNA in *C.elegans* requires the putative transmembrane protein SID-1. *Science*, **295**, 2456–2459.

Wu-Scharf,D., Jeong,B.-R., Zhang,C. and Cerutti,H. (2000) Transgene and transposon silencing in *Chlamydomonas reinhardtii* by a DEAH-Box RNA helicase. *Science*, **290**, 1159–1162.

Yoshioka,Y., Matsumoto,S., Kojima,S., Ohshima,K., Okada,N. and Machida,Y. (1993) Molecular characterisation of a short interspersed repetitive element from tobacco that exhibits sequence homology to specific transfer-RNAS. *Proc. Natl Acad. Sci. USA*, **90**, 6562–6566.

Zamore,P.D., Tuschl,T., Sharp,P.A. and Bartel,D.P. (2000) RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. *Cell*, **101**, 25–33.

Received May 7, 2002; revised July 5, 2002; accepted July 16, 2002

Three Tnt1 Subfamilies Show Different Stress-Associated Patterns of Expression in Tobacco. Consequences for Retrotransposon Control and Evolution in Plants

Thierry Beguiristain, Marie-Angèle Grandbastien, Pere Puigdomènech, and Josep M. Casacuberta*

Departament Genètica Molecular, Institut de Biologia Molecular de Barcelona (Consejo Superior de Investigaciones Científicas), Barcelona, Spain (T.B., P.P., J.M.C.); and Laboratoire de Biologie Cellulaire, Institut National de la Recherche Agronomique, Versailles, France (M.-A.G.)

The genomes of most *Nicotiana* species contain three different subfamilies of the Tnt1 retrotransposon, which differ completely in their U3 sequence, whereas the rest of the sequence is relatively constant. The results presented here show that all three Tnt1 subfamilies are expressed in tobacco (*Nicotiana tabacum*) and that the U3 sequence variability correlates with differences in the pattern of expression of the Tnt1 elements. Each of the three Tnt1 subfamilies is induced by stress, but their promoters have a different response to different stress-associated signaling molecules. The Tnt1A subfamily is particularly strongly induced by elicitors and methyl jasmonate, whereas expression of the Tnt1C subfamily is more sensitive to salicylic acid and auxins. The direct relationship between U3 sequence variability and differences in the stress-associated expression of the Tnt1 elements present in a single host species gives support to our model that postulates that retrotransposons have adapted to their host genomes through the evolution of highly regulated promoters that mimic those of the stress-induced plant genes. Moreover, here we show that the analysis of the transcriptional control of a retrotransposon population such as Tnt1 provides new insights into the study of the complex and still poorly understood network of defense- and stress-induced plant signal transduction pathways.

Retrotransposons are mobile genetic elements ubiquitously present in eukaryote genomes. In some cases, such as the BARE1 element from barley, a single retrotransposon family can reach a copy number as high as 50,000 copies per haploid genome (Suoniemi et al., 1996). This high invasivity is facilitated by their replicative mechanism of transposition: the transcription of the element generates an RNA copy that is reverse transcribed into cDNA prior to re-insertion into the genome (Boeke and Corces, 1989). As retrotransposons do not excise, the copy number of retrotransposons increases exponentially with transposition. On the other hand, retrotransposition is potentially a highly mutagenic event, which makes the invasivity of these elements a hindrance for their survival in evolution. As retrotransposons are noninfective agents that cannot leave the host they inhabit, their survival depends on a fine-tuning of activity—high enough to maintain the ability to transpose, but below a threshold that would compromise the viability of the host genome. How this equilibrium is reached is an interesting and still open question that will probably have different answers for different retroelement-host genome couples. In the case of the yeast Ty elements, which are probably the most well-characterized retrotransposons, this equilibrium is achieved by a strict specificity of insertion to regions devoid of genes (Boeke and Devine, 1998), combined with a frequent intra-element long terminal repeat (LTR) recombination that eliminates newly inserted elements (Jordan and McDonald, 1999). This high turnover of Ty elements leads to the maintenance within the yeast genome of a small population of active Ty elements with a high level of sequence homogeneity (Jordan and McDonald, 1999). The case of plant genomes seems to be very different, as they contain a high copy number of retrotransposon sequences that can account for more than 50% of their DNA content (Kumar and Bennetzen, 1999). In addition, these sequences display a high degree of variability (for example, see Casacuberta et al., 1995; Marillonnet and Wessler, 1998) and, in most cases, they represent elements that have lost the ability to transpose. This is probably a consequence of the fact that although intra-element LTR recombination has been shown to occur for plant retrotransposons (Vicient et al., 1999), it does not seem to be a general or important mechanism for reducing retrotransposon copy number in most plant genomes (Bennetzen and Kellogg, 1997). Moreover, plant retrotransposons do not display target site specificity and can insert within or close to genes, creating mutations. Therefore, it seems that plant retrotransposon populations are controlled by mechanisms unlike those of the yeast Ty elements.

All the active plant retrotransposons characterized so far are only expressed under very precise stress situations, being silent during most of the plant life cycle (Wessler, 1996; Grandbastien, 1998). This sug-

* Corresponding author; e-mail jcsgmp@cid.csic.es; fax 34–93–204–59–04.

gests that transcriptional regulation is a major mechanism of control for retrotransposition in plants (Wessler, 1996; Grandbastien, 1998). This is the case for the tobacco (*Nicotiana tabacum*) retrotransposons Tnt1A and Tto1, whose promoters have been analyzed in detail (Grandbastien et al., 1997; Takeda et al., 1999). How plant retrotransposons have evolved such stress-activated promoters is an open question. We have previously shown that Tnt1 is expressed in protoplasts and roots as a heterogeneous population of RNA molecules that resembles retroviral quasispecies (Casacuberta et al., 1995). This lead us to suggest that as for retroviral quasispecies, this high sequence variability could endow Tnt1 with a high sequence plasticity that could facilitate the evolution of its promoter regions to improve its coexistence with the host (Casacuberta et al., 1997). To validate this hypothesis, we have looked for a possible correlation between sequence variability and differences in the expression patterns of Tnt1 elements. The results presented here show that the three previously defined Tnt1 subfamilies (Casacuberta et al., 1997; Vernhettes et al., 1998) are expressed in tobacco with different expression patterns. These specific patterns of expression are probably a consequence of the sequence variability of their U3 regions, which in each case contain a different stress-inducible promoter.

RESULTS

The Three Tnt1 Subfamilies Are Expressed in Tobacco Cell Cultures

Tnt1 is present in hundreds of copies in the genome of tobacco and related *Nicotiana* spp. (Casacuberta et al., 1997). The U3 region of Tnt1 is highly variable, and three different subfamilies of Tnt1 have been defined according to their U3 sequence (Casacuberta et al., 1997; Vernhettes et al., 1998). All three different subfamilies, Tnt1A, Tnt1B, and Tnt1C, are present, but they differ in relative abundance in different *Nicotiana* genomes, which suggests that all three subfamilies remain active (Vernhettes et al., 1998). However, until now only the Tnt1A subfamily has been shown to be expressed. Tnt1A is expressed in protoplasts and roots and, although we have analyzed more than 100 partial Tnt1 sequences by reverse transcriptase (RT)-PCR, we have failed to detect expression of Tnt1B and Tnt1C elements in these tissues (Casacuberta et al., 1995).

Tnt1 is transiently expressed during protoplast isolation, and its RNA level decreases rapidly when protoplasts are cultured (Grandbastien et al., 1997). However, it also has been reported that Tnt1 copy number increases slightly in cultured cells, suggesting that Tnt1 could be expressed in those cells (Hirochika, 1993). As different gene programs are activated during the different cell culture stages, we decided to analyze the expression of Tnt1 in cultured tobacco cells to look for a possible expression of Tnt1B and Tnt1C elements by RT-PCR analysis.

The results presented in Figure 1 show that whereas Tnt1 is not expressed in 1-week-old cultures (Fig. 1, 0′), there is a transient induction of Tnt1 expression 4 h after the addition of fresh media to the culture. Preliminary results suggest that this induction could be associated to the presence of auxin hormones in the fresh media (data not shown). To determine the Tnt1 subfamilies expressed under these conditions, 15 partial Tnt1 RNA sequences comprising the U3 region (see Fig. 2A) were obtained by RT-PCR amplification with RNA from tobacco cell cultures and were cloned and sequenced. A phylogenetic analysis of the sequences, including consensus sequences of the Tnt1A, Tnt1B, and Tnt1C subfamilies (Casacuberta et al., 1997) by the neighbor-joining method, is shown in Figure 2B. This analysis shows that only two of the sequences obtained are closely similar to the Tnt1A consensus, whereas most of the sequences are highly similar to the Tnt1B subfamily consensus. One of the sequences was found to be more similar to the consensus for Tnt1C than to the consensus for the two other Tnt1 subfamilies. These results show that the three Tnt1 subfamilies are expressed in freshly subcultured tobacco cells and that Tnt1B RNA is predominant in the conditions. To our knowledge, this is the first report of different subfamilies of a retrotransposon being expressed in a particular host species.

Induction of Different Tnt1 RNA Populations with Different Stress-Associated Signaling Molecules

To get an insight into the signal transduction pathways that lead to the induction of the different Tnt1 subfamilies, we infiltrated tobacco leaf discs with different elicitors and stress-associated signaling molecules. Previous results showed that cryptogein, a protein from *Phytophtora cryptogea* that elicits tobacco defense responses (Ricci et al., 1989), was able to induce the Tnt1A promoter, as was salicylic acid, the signal transduction intermediate of the plant re-

Figure 1. Induction of Tnt1 expression in tobacco cultured cells. Northern analysis of RNAs obtained from tobacco cells cultured in NK1 medium after 1-week culture (0) or after 10 min (10′), 45 min (45′), or 4 h of subculture in fresh media, hybridized with a probe corresponding to a conserved endonuclease region of Tnt1. The 5.2-kb band corresponding to the Tnt1 genomic RNA is indicated by an arrow.

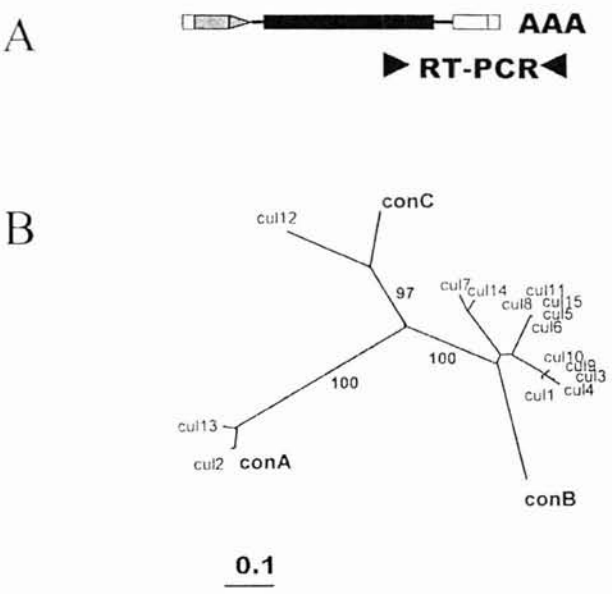


Figure 2. Analysis of the Tnt1 RNA molecules expressed in cultured cells. A, Scheme showing the region of Tnt1 amplified by RT-PCR. Coding sequences are shown in black, the U5 region is shown in dark gray, and the U3 region is shown in light gray. ▶, The approximate position of oligonucleotides used for the RT-PCR amplification. B, Neighbor-joining tree obtained with an alignment of the 15 Tnt1 partial sequences obtained from cultured cells RNA (cul#) together with the consensus sequences for the Tnt1A (conA), Tnt1B (conB), and Tnt1C (conC) subfamilies. Bootstrap values over 60 for the main branches are shown.

sponses to wounding and pathogen infection (Grandbastienet al., 1997; Vernhettes et al., 1997). Nevertheless, it has recently been proposed that another defense-related signaling molecule, methyl jasmonate, could be also part of the signaling cascade triggered by cryptogein (Rusterucci et al., 1999). On the other hand, preliminary results suggest that addition of fresh auxins to the culture medium could be responsible for the transient induction of Tnt1 we have observed in tobacco subcultured cells. Therefore, we investigated the effect on Tnt1 expression of infiltrating leaf discs with cryptogein, salicylic acid, methyl jasmonate, and 2,4-dichlorophenoxyacetic acid (2,4-D) using discs infiltrated with water as a control.

Figure 3 presents a northern-blot analysis of the RNAs obtained after each treatment. It can be seen that, as previously reported (Vernhettes et al., 1997), the 5.2-kb RNA of Tnt1 is expressed to low levels in leaf discs infiltrated with water, due to the wounding stress associated with the infiltration process. However, infiltration with chemicals results in a visible increase in induction of Tnt1. A quantitative analysis of the hybridization using a phosphoimager (Bio-Rad, Hercules, CA) shows that infiltration with cryptogein leads to a 10-fold induction of the steady-state level of Tnt1 RNA, whereas the increase in induction with salicylic acid, methyl jasmonate, and 2,4-D is 3- to 4-fold. A 6.5-kb leaf mRNA is also detected, but it has been shown that this does not correspond to the genomic transcript of the retrotransposon (Pouteau et al., 1991).

To analyze which Tnt1 subfamilies are expressed after each treatment, we amplified the 3′ end of the Tnt1 RNA, which comprises the U3 region, by RT-PCR. Two sets of 10 sequences obtained from two independent RT-PCR reactions were cloned and sequenced from RNA obtained from each treatment. These sequences were compared with the consensus of the three Tnt1 subfamilies using a phylogenetic approach. The neighbor-joining trees obtained with the sequences from each treatment are shown in Figure 4. These results show that although cryptogein infiltration induces the expression of the Tnt1A subfamily only (Fig. 4A), the rest of the treatments induce a more complex population of Tnt1 RNAs. This is particularly clear for the salicylic acid (Fig. 4B) and 2,4-D (Fig. 4D) treatments in which five and 11 out of 20 sequences, respectively, closely resemble the consensus sequence for the Tnt1C subfamily. A sequence highly similar to the Tnt1B consensus was detected in the methyl jasmonate-treated material (Fig. 4C).

All the sequences obtained from the water-infiltrated control leaf discs were closely similar to the consensus for the Tnt1A subfamily (data not shown).

The U3 Regions of the Three Tnt1 Subfamilies Contain Promoter Elements Differentially Induced by Different Stress-Associated Signaling Molecules

Our results clearly show that the stress-associated signaling molecules, salicylic acid, methyl jasmonate, and 2,4-D, as well as the fungal elicitor cryptogein, differentially induce the expression of the three Tnt1 retrotransposon subfamilies. To analyze whether the U3 regions of each Tnt1 subfamily contain different stress-inducible promoters, we bombarded leaf discs with constructs of U3 regions representative of each

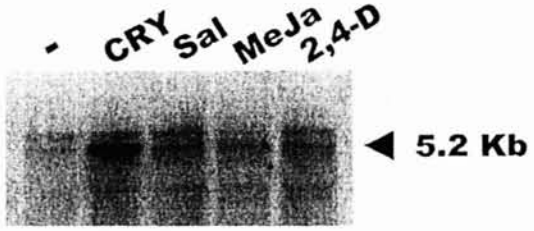


Figure 3. Induction of Tnt1 expression by infiltration of tobacco leaf discs with stress-associated signaling molecules. Northern analysis of RNAs obtained from tobacco leaf discs infiltrated with water medium (-) or 1 $\mu g\ mL^{-1}$ cryptogein (CRY), 2 mM salicylic acid (Sal), 10 mM methyl jasmonate (MeJa), and 1 mM 2,4-D, hybridized with a probe corresponding to a conserved endonuclease region of Tnt1. The 5.2-kb band corresponding to the Tnt1 genomic RNA is indicated by an arrow.

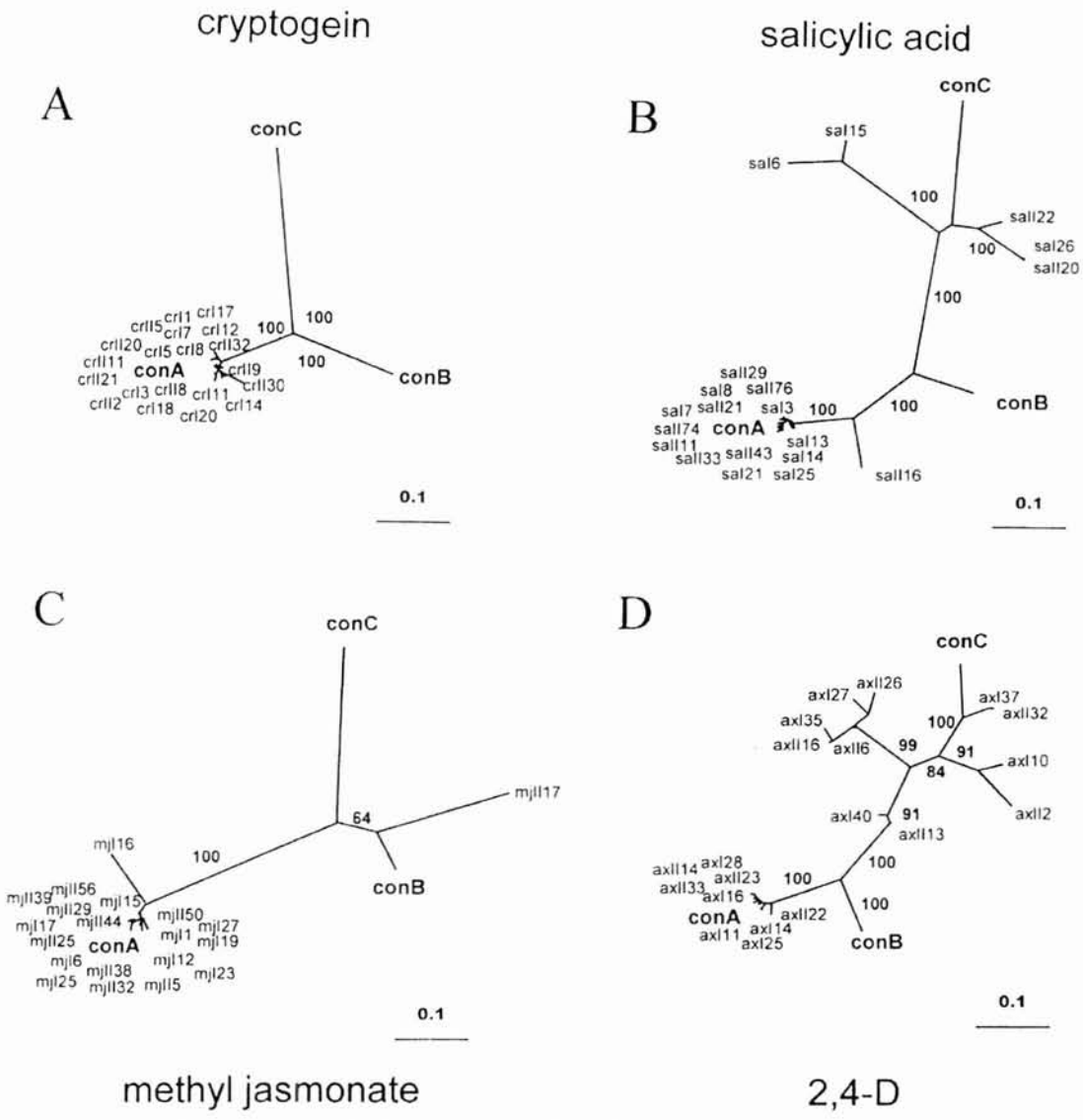


Figure 4. Analysis of the Tnt1 RNA molecules expressed in infiltrated leaf discs. Neighbor-joining trees obtained with an alignment of the 20 Tnt1 partial sequences obtained by two independent RT-PCR amplifications from infiltrated leaf disc RNA together with the consensus sequences for the Tnt1A (conA), Tnt1B (conB), and Tnt1C (conC) subfamilies. A, Sequences obtained from cryptogein-infiltrated leaf discs (crl= and crll= for the sequences obtained from the first or the second RT-PCR experiment, respectively); B, sequences obtained from salicylic acid-infiltrated leaf discs (sal= and sall=); C, sequences obtained from methyl jasmonate-infiltrated leaf discs (mjl= and mjll=); D, sequences obtained from 2,4-D-infiltrated leaf discs (axl= and axll=). Bootstrap values over 60 for the main branches are shown.

subfamily upstream of a β-glucuronidase (GUS) reporter gene, and then analyzed GUS expression after incubation with cryptogein, salicylic acid, methyl jasmonate, or 2,4-D. As secondary effects could be generated by single nucleotide differences within the conserved TATA box or transcription start region, we cloned Tnt1B and Tnt1C U3 sequences upstream of the TATA box in front of the TATA-box region of the sequence chosen as representative for the Tnt1A subfamily (see "Materials and Methods"). Leaf discs bombarded with a GUS reporter gene devoid of promoter and a GUS reporter gene driven by the 35S constitutive promoter were used, respectively, as negative and positive controls.

Results presented in Figure 5 show that the U3 region of the Tnt1B subfamily is not able to efficiently promote GUS expression in the conditions tested. On the contrary, the Tnt1A and Tnt1C U3 regions are able to drive a high level of GUS expression in tobacco leaves. It is interesting that whereas the U3

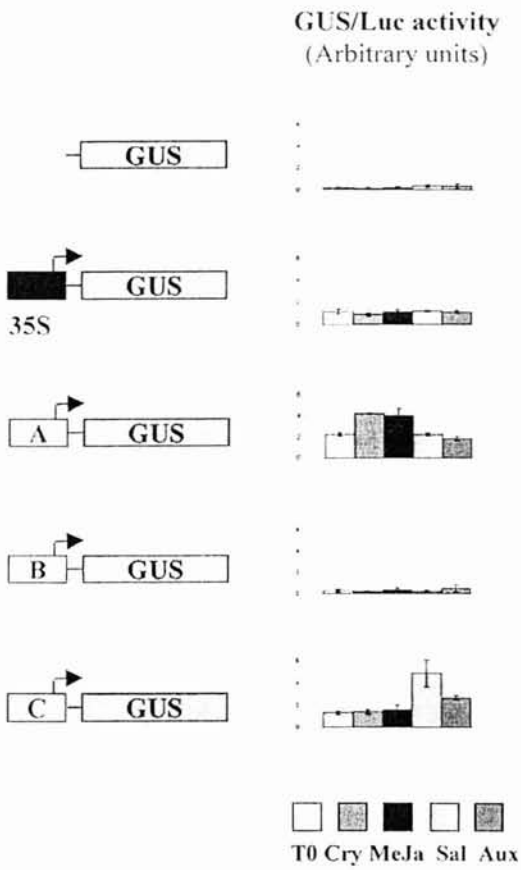


Figure 5. Transient expression analysis of Tnt1 promoters. Relative GUS/Luc activity obtained after incubating the leaf discs, bombarded with the different GUS constructs, with different chemicals.

region of Tnt1A seems to induce a particularly high level of expression after incubating the bombarded leaves with cryptogein or methyl jasmonate, the U3 region of the Tnt1C subfamily is particularly induced in leaf discs incubated with salicylic acid and 2,4-D. On the other hand, no GUS activity is detected in leaf disc bombarded with the negative control, whereas, as expected, a strong noninducible GUS activity is detected in leaf discs bombarded with the 35S-GUS construct. These results are in agreement with the RT-PCR results obtained and are consistent with a major role of the U3 region in conferring a specific pattern of expression on each Tnt1 subfamily.

We have previously shown that the Tnt1A promoter is activated by cryptogein (Vernhettes et al., 1997), and Figures 3 and 4 show that it is also highly activated by methyl jasmonate. To analyze if jasmonate induces Tnt1A through the same promoter elements as cryptogein, we compared the Tnt1 RNA sequences obtained after induction by these different chemicals. A neighbor-joining tree of the 61 Tnt1A sequences obtained failed to form groups supported by a bootstrap value of more than 10%, indicating that it is not possible to differentiate populations within them (not shown). The populations of Tnt1A sequences expressed after the different treatments here analyzed are thus indistinguishable. In addition to this, the BII elements, which were shown to be important for cryptogein induction (Vernhettes et al., 1997), are conserved in all the sequences amplified, 67% of them having four tandem repetitions of this element and only one sequence out of 61 having less than three elements (data not shown). These results suggest that methyl jasmonate and cryptogein induce Tnt1A through the same cis-acting elements.

On the other hand, our results show that the promoter of Tnt1C is particularly sensitive to 2,4-D and salicylic acid. Although the populations of Tnt1C RNA induced by 2,4-D and salicylic acid are much more variable than those of Tnt1A RNA, they cannot be differentiated from one another by phylogenetic analysis. Figure 6A presents a neighbor-joining tree of all the Tnt1C sequences obtained, and it can be seen that different groups supported by high bootstrap values can be defined. Nevertheless, sequences belonging to each of these groups were found after salicylic acid and 2,4-D treatments, and we have not been able to define any induction-specific group.

It has been shown that salicylic acid and auxins can induce plant genes through the interaction of inducible transcription factors with the as-1-related cis-elements (Chen and Singh, 1999; Niggeweg et al., 2000). The Tnt1C sequence analyzed here contains a nearly palindromic sequence within the U3 region

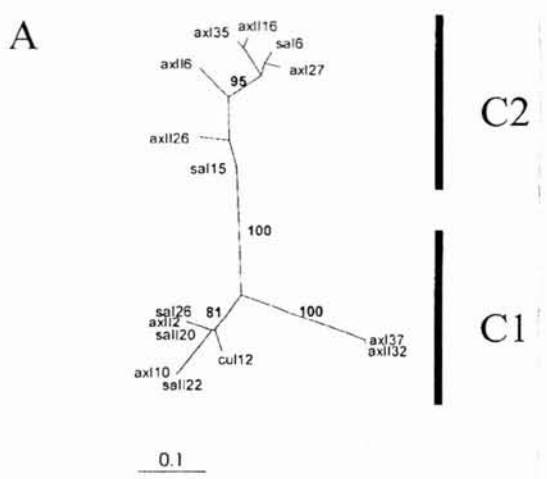


Figure 6. Analysis of the Tnt1C RNA molecules expressed in infiltrated leaf discs. A, Neighbor-joining trees obtained with an alignment of all the Tnt1C partial sequences obtained. Sequences obtained from salicylic acid-infiltrated leaf discs are shown as in Figure 4. Bootstrap values over 60 for the main branches are shown. B, Comparison of the consensus as-1 motif with those found in Tnt1C1 and Tnt1C2 sequences. Conserved nucleotides are shown in bold.

that closely resembles the *as-1* element (TGACGTCAnnnnTGACGTCA), and this could be important for the induction of this Tnt1 subfamily by auxins and SA. It is interesting that although the two Tnt1C groups of sequences differ within the U3 region, both contain an almost identical as-1-like element located 20 to 30 nucleotides upstream of the corresponding TATA boxes (see Fig. 6B), suggesting that this element is important for the expression of these retrotransposon elements.

DISCUSSION

Different Subfamilies of the Retrotransposon Tnt1 Have Maintained Their Transcriptional Activity in Tobacco

Although three different subfamilies of the tobacco retrotransposon Tnt1 have been found to be present at a high copy number in the genome of different *Nicotiana* species, until now evidence for expression activity has only been obtained for the Tnt1A subfamily (Grandbastien et al., 1997). The results presented in this work show that all three Tnt1 subfamilies are expressed in tobacco and that they have different patterns of expression. We have previously shown that the fungal elicitor cryptogein induces the expression of the Tnt1A promoter (Pouteau et al., 1994; Vernhettes et al., 1997). We show here that this treatment induces an important accumulation of RNA that corresponds to Tnt1 elements that belong to the A subfamily. This result is similar to what we have previously reported from the analysis of Tnt1 RNA present in root tips and protoplasts in which only Tnt1A elements are expressed (Casacuberta et al., 1995). In contrast, the infiltration of leaf discs with methyl jasmonate, and particularly with salicylic acid or 2,4-D, induces a more complex population of Tnt1 elements that belong to all three subfamilies. To our knowledge this is the first report of different subfamilies of a retrotransposon being expressed in a particular host genome, which makes Tnt1 a particularly interesting model for analyzing the evolution of retrotransposons within plant genomes.

The Three Different Tnt1 Subfamilies Are Differentially Induced by Stress-Associated Signaling Molecules through Specific Promoter Elements

The results presented here not only show that the three Tnt1 subfamilies are expressed in tobacco, but they also show that these subfamilies are differentially regulated. The RT-PCR analysis shows that infiltration treatments with different signaling molecules are able to induce particular subsets of Tnt1 elements. We found expression of Tnt1A RNA in leaves infiltrated with all the chemicals analyzed. Tnt1A RNA was also found in the water-infiltrated leaf discs (not shown), probably because of the wounding stress associated with the infiltration procedure (Vernhettes et al., 1997). Also, it is likely that the bombardment procedure induces a wound response on its own and, thus, we cannot rule out the possibility that the weak induction of Tnt1A detected with 2,4-D in these experiments is the result of a wounding-associated stress response. In line with this hypothesis, it has been previously reported that the promoter of Tnt1A is not inducible by 2,4-D in transgenic tobacco calli (Pauls et al., 1994). On the contrary, the clear induction of the Tnt1A promoter observed after cryptogein and methyl jasmonate incubation in the transient expression analysis reveals an elevated sensitivity of the Tnt1A promoter to both chemicals, and is in agreement with the RT-PCR results obtained.

The presence of several sequences belonging to the Tnt1C subfamily among those obtained from salicylic acid- and 2,4-D-treated leaf discs suggests that Tnt1C expression is probably induced in stress situations regulated by these signal molecules. The transient expression analysis presented here shows that the first 150 nucleotides of the U3 region of Tnt1C are sufficient to produce high levels of expression from a minimal promoter in bombarded leaves incubated with salicylic acid and 2,4-D and suggests that this region contains the promoter elements of Tnt1C that respond to these signaling molecules.

We have not been able to detect any promoter activity of the U3 region of the Tnt1B sequence analyzed. This is not completely unexpected considering that most of the Tnt1B RNA sequences obtained were amplified from subcultured cells and almost no Tnt1B RNA sequence was obtained from infiltrated leaf discs in which the transient expression experiments were performed. These results could suggest that the Tnt1B promoter is not expressed in leaf tissues. We are, at present, transforming tobacco plants with constructs similar to those used in this report to analyze in more detail the expression pattern of the different Tnt1 subfamilies.

The strong correlation of the RT-PCR results and the transient expression experiments suggest that the promoter regions that control the transcription of the Tnt1 elements in response to wounding, cryptogein, or methyl jasmonate are all contained within the U3 region of the different Tnt1 elements.

In conclusion, the results presented here show that the different Tnt1 subfamilies are expressed in tobacco with different patterns, driven by different inducible promoter elements located within their U3 regions. It is interesting to note that the expression of all three Tnt1 subfamilies is strongly regulated, with the different Tnt1 elements only being expressed under stress situations or after inoculation with stress-associated signaling molecules. This emphasizes the importance of transcriptional regulation in the control of plant retrotransposon activity and the tight relationship that exists between retrotransposition and stress situations in plants.

Retrotransposons as Models to Analyze Stress-Associated Expression in Plants

Defense-related responses of plant cells depend on a complex network of signal transduction pathways. Salicylic acid is probably the most well-characterized signaling molecule in plant defense reactions (Klessig et al., 2000), but signal transduction pathways involving methyl jasmonate and ethylene have also been characterized (Reymond and Farmer, 1998). On the other hand, auxin has been shown to activate some defense-related genes through the as-1 element, which also confers inducibility by salicylic acid (Strompen et al., 1988; Xiang et al., 1996). In some cases, these different pathways seem to be independent (Penninckx et al., 1996), but genes concomitantly activated by two of these pathways have also been reported (Strompen et al., 1988; Xiang et al., 1996; Asai et al., 2000). A recent microarray-based analysis has shown that a high number of defense-related genes are coordinately induced by different signal transduction pathways, although genes controlled by only one of these signals, and examples of signal antagonism, were also found (Schenk et al., 2000). This illustrates the complexity of the network of regulatory interactions that controls plant defense reactions, and the risk of drawing general conclusions from the analysis of a single gene promoter. The presence of retrotransposon promoter sequences within the RNA molecule, as well as the high sequence variability found for these elements in plants, allows for the combination of classical promoter analysis with population-based approaches to study transcriptional regulation. We have applied this strategy to analyze the promoter of Tnt1A (Casacuberta and Grandbastien, 1993; Casacuberta et al., 1995; Vernhettes et al., 1997), and here we used a similar approach to improve our analysis of this promoter, as well as to begin to analyze the promoters of the Tnt1B and Tnt1C subfamilies.

Our results show that the Tnt1A RNA populations induced by methyl jasmonate and cryptogein are indistinguishable, suggesting that both signaling molecules induce Tnt1A through the same cis-acting elements. Our results thus reinforce the recent hypothesis that jasmonic acid could be part of the signaling cascade triggered by cryptogein (Rusterucci et al., 1999).

On the other hand, the Tnt1C populations induced by salicylic acid and 2,4-D are also indistinguishable, suggesting that also in this case, both chemicals induce Tnt1C through the same cis-acting elements. It has been already shown that salicylic acid and 2,4-D can induce plant gene expression through *as-1*-related elements (Chen and Singh, 1999; Niggeweg et al., 2000). It is interesting that although sequence variability among the Tnt1C sequences amplified is high and two different subfamilies can be defined, all of them maintain a consensus *as-1*-related element within the U3 region. We are currently analyzing the possible implication of this element in the induction of the Tnt1C promoter by auxins and salicylic acid, and we will examine the promoters of the two Tnt1C groups here defined for differences in induction by pathogen-associated stresses.

Sequence Plasticity and Evolution of Stress-Associated Promoters: A Strategy for Retrotransposon Maintenance in Plant Genomes

We show in this paper that the sequence variability found within the U3 regions of Tnt1 elements correlates with differences in their pattern of expression. The three Tnt1 subfamilies contain different promoter elements that induce their activity in different stress situations. Such expression variability should have consequences for the evolution of these elements, as repeated exposure of the host to a particular stress situation should favor amplification of a particular subfamily and a series of different induction events will thus lead to the amplification of different populations of elements in different host genomes. We have previously shown that the three Tnt1 subfamilies are differently represented in different *Nicotiana* genomes (Vernhettes et al., 1998).

It is interesting to note that the U3 region displays a high degree of sequence variability even within a single Tnt1 subfamily. For example, two different groups of Tnt1C sequences can be defined based on their U3 sequence, so even though both groups of elements are induced by the same chemicals, slight differences in the pattern of expression might exist between them. The simultaneous expression of heterogeneous populations of Tnt1 elements with such differences in expression should allow the selection of different subsets of elements in different genomes, allowing them to evolve and to adapt to their hosts. The existence of Tnt1-related retrotransposon, Retrolyc1, in tomato and related wild species showing extensive sequence similarities to Tnt1 except for its U3 region, which also contains the promoter sequences, seems to confirm that adaptation to the host genome correlates with U3 sequence divergence (Costa et al., 1999; Araujo et al., 2001).

Thus, we propose that the high sequence variability of Tnt1 and other plant retrotransposons, which is a consequence of the infidelity of the retrotransposition process (Gabriel et al., 1996) and the high copy number of these elements in plant genomes (Casacuberta et al., 1995), could have allowed their promoter sequences to evolve the optimum pattern of expression to be maintained in each host genome. This strategy obviously would be very different from that of Ty elements in yeast and could be a consequence of the important differences of the organization of these different host genomes. The high proportion of intragenic regions and repetitive DNA in plant genomes compared with the compact genome of *Saccharomyces cerevisiae*, as well as the high frequency of

polyploidy in plant genome evolution, could have minimized the negative impact of retrotransposons being inserted randomly. In addition, this could have allowed these elements to reach the high copy number that is essential to develop such a quasispecies-like strategy of adaptation.

MATERIALS AND METHODS

Cell Culture

Tobacco (*Nicotiana tabacum* cv Wisconsin 38) cells cultured in NK1 medium (Xiang et al., 1996) were subcultured once a week by the addition of one-half-volume of fresh media to one-half-volume of cells. After subculturing, cells were harvested at different times, pelleted by centrifugation, and frozen in liquid nitrogen.

Infiltration of Leaf Discs

Tobacco (cv Xanthi, wild-type line XHFD8, Bourgin and Missonier, 1973) leaf discs were infiltrated with water, 1 μg mL^{-1} of cryptogein, 2 mM salicylic acid, and 10 μM methyl jasmonate or 1 mM 2,4-D as previously described (Vernhettes et al., 1997)

RNA Extractions, Northern-Blot Hybridizations, RT-PCR Amplifications, Cloning, and Sequencing

RNA from cultured cells and infiltrated leaves was obtained by standard methods (Casacuberta et al., 1995). RNA blotting and hybridization with a conserved Tnt1 endonuclease probe was performed as previously described (Mhiri et al., 1997). RT-PCR amplifications with the Avi and dT primers were performed as previously described (Casacuberta et al., 1995), and the PCR fragments were cloned in a pGEM-T vector (Stratagene, La Jolla, CA) and were sequenced on both strands.

Phylogenetic Analysis

Sequences were aligned using the CLUSTAL W multiple-alignment program (version 1.5; Thompson et al., 1994) with some minor refinements. DNADIST in Felsestein's PHYLIP package (Felsenstein, 1989) was used to generate a distance matrix based on the Jukes-Cantor algorithm (Jukes and Cantor, 1969). This was used to generate neighbor-joining trees (Saitou and Nei, 1987). Bootstrap analyses were performed using the Seqboot and Consense programs from Felsestein's PHYLIP package (Felsenstein, 1989).

Promoter-GUS Constructs

A previously described tab7 clone (Vernhettes et al., 1998) was chosen as representative of the U3 region of the Tnt1A subfamily. To construct the A-GUS clone, the tab7 clone was digested by *Sal*I, filled in with Klenow, and re-digested with *Bam*HI. The fragment containing the U3 of tab7 was cloned in a previously described GP plasmid (Casacuberta et al., 1993) containing a GUS reporter gene devoid of promoter and was digested with *Hin*dIII (blunt-ended with Klenow) and *Bam*HI. To obtain the B-GUS and C-GUS plasmids, the region upstream of the TATA box of the clones chosen as representatives of the Tnt1B subfamily (cul4) and Tnt1C (cul12) was amplified by PCR by standard procedures, using the universal primer located within the polylinker region of the pGEM plasmid and an oligonucleotide complementary to 25 to 30 nucleotides of each clone 10 nucleotides upstream of their respective TATA box, followed by an *Hin*dIII site (5′-GCC AAA GCT CTA CCA ACC TTG ACC-3′ for Tnt1B, 5′-GCC AAA GCT TGC ACA TAT TGA CTT ATG CAA TGA CAT C-3′ for Tnt1C). The amplified fragments were digested with *Sal*I and *Hin*dIII and this fragment was used to substitute the *Sal*I-*Hin*dIII fragment of the tab7 clone. The corresponding fragments were cloned within the GP plasmid to generate the B-GUS and the C-GUS constructs as for the A-GUS construct.

Microprojectile Bombardment and Enzyme Assays

Tobacco leaf discs were transformed by particle bombardment, with 4 μg of the different Tnt1-GUS constructs and 1 μg of a pCAMV35SLUC used as an internal standard (Marzabal et al., 1998). Enzyme assays were performed as described (Marzabal et al., 1998).

ACKNOWLEDGMENTS

We thank Pilar Fontanet for excellent technical assistance in the greenhouse, and Maria-Lluïsa Espinás and Salomé Prat (Institut de Biologia Molecular de Barcelona-Consejo Superior de Investigaciones Científicas, Barcelona) for helpful discussions and critical reading of the manuscript.

Received March 22, 2001; returned for revision May 2, 2001; accepted May 30, 2001.

LITERATURE CITED

Araujo PJ, Casacuberta JM, Costa AP, Hashimoto RY, Grandbastien MA, Van Sluys MA (2001) Retrolyc1 subfamilies defined by different U3 LTR regulatory regions in the *Lycopersicon* genus. Mol Gen Genet (in press)

Asai T, Stone JM, Heard JE, Kovtun Y, Yorgey P, Sheen J, Ausubel F (2000) Fumonisin B1-induced cell death in *Arabidopsis* protoplasts requires jasmonate-, ethylene-, and salicylate-dependent signaling pathways. Plant Cell **12:** 1823–1835

Bennetzen JL, Kellogg EA (1997) Do plants have a one-way ticket to genomic obesity? Plant Cell **9:** 1509–1514

Boeke JD, Corces V (1989) Transcription and reverse transcription of retrotransposons Annu Rev Microbiol **45:** 403–434

Boeke JD, Devine SE (1998) Yeast retrotransposons: finding a nice quiet neighborhood. Cell **93:** 1087–1089

Bourgin JP, Missonier C (1973) Vegetative propagation and cold preservation of haploid plants of *Nicotiana tabacum* and *Nicotiana paniculata*. Haploid Inf Serv **8:** 7

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

Casacuberta JM, Vernhettes S, Grandbastien M-A (1995) Sequence variability within the tobacco retrotransposon Tnt1 population. EMBO J **14:** 2670–2678

Casacuberta JM, Vernhettes S, Grandbastien M-A (1997) Quasispecies in retrotransposons: a role for sequence variability in Tnt1 evolution. Genetica **100:** 109–117

Chen W, Singh KB (1999) The auxin, hydrogen peroxide and salicylic acid induced expression of the *Arabidopsis* GST6 promoter is mediated in part by an ocs element Plant J **19:** 667–677

Costa AP, Scortecci KC, Hashimoto RY, Araujo PG, Grandbastien MA, Van Sluys MA (1999) Retrolycl-1, a member of the tntl retrotransposon super-family in the *Lycopersicon peruvianum* genome. Genetica **107:** 65–72

Gabriel A, Willems M, Mules EH, Boeke JD (1996) Replication infidelity during a single cycle of Ty1 retrotransposition. Proc Natl Acad Sci USA **93:** 7767–7771

Grandbastien M-A (1998) Activation of plant retrotransposons under stress conditions. Trends Plant Sci **3:** 181–187

Grandbastien M-A, Lucas H, Morel J-B, Mhiri C, Vernhettes S, Casacuberta JM (1997) The expression of the tobacco Tnt1 retrotransposon is linked to plant defense responses. Genetica **100:** 241–252

Felsenstein J (1989) PHYLIP: phylogeny inference package (version 3.56). Cladistics **5:** 164–166

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

Jordan IK, McDonald JF (1999) Tempo and mode of Ty element evolution in *Saccharomyces cerevisiae*. Genetics **151:** 1341–1351

Jukes TH, Cantor CR (1969) Evolution of protein molecules. *In* HN Munro, ed, Mammalian Protein Metabolism. Academic Press, New York, pp 21–132

Klessig DF, Durner J, Noad R, Navarre DA, Wendehenne D, Kumar D, Zhou JM, Shah J, Zhang S, Kachroo P et al. (2000) Nitric oxide and salicylic acid signaling in plant defense. Proc Natl Acad Sci USA **97:** 8849–8855

Kumar A, Bennetzen JL (1999) Plant retrotransposons. Annu Rev Genet **33:** 479–532

Marillonnet S, Wessler SR (1998) Extreme structural heterogeneity among the members of a maize retrotransposon family. Genetics **150:** 1245–1256

Marzabal P, Busk PK, Ludevid MD, Torrent M (1998) The bifactorial endosperm box of gamma-zein gene: characterization and function of the Pb3 and GZM cis-acting elements. Plant J **16:** 41–52

Mhiri C, Morel JB, Vernhettes S, Casacuberta JM, Lucas H, Grandbastien MA (1997) The promoter of the tobacco Tnt1 retrotransposon is induced by wounding and by abiotic stress. Plant Mol Biol **33:** 257–266

Niggeweg R, Thurow C, Kegler C, Gatz C (2000) Tobacco transcription factor TGA2.2 is the main component of as-1-binding factor ASF-1 and is involved in salicylic acid- and auxin-inducible expression of as-1-containing target promoters. J Biol Chem **275:** 19897–19905

Pauls PK, Kunert K, Huttner E, Grandbastien MA (1994) Expression of the tobacco Tnt1 retrotransposon promoter in heterologous species. Plant Mol Biol **26:** 393–402

Penninckx IA, Eggermont K, Terras FR, Thomma BP, De Samblanx GW, Buchala A, Metraux JP, Manners JM, Broekaert WF (1996) Pathogen-induced systemic activation of a plant defensin gene in *Arabidopsis* follows a salicylic acid-independent pathway. Plant Cell **8:** 2309–2323

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

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

Reymond P, Farmer EE (1998) Jasmonate and salicylate as global signals for defense gene expression. Curr Opin Plant Biol **1:** 404–411

Ricci P, Bonnet P, Huet JC, Sallantin M, Beauvais-Cante F, Bruneteau M, Billard V, Michel G, Pernollet JC (1989) Structure and activity of proteins from pathogenic fungi *Phytophtora* eliciting necrosis and acquired resistance in tobacco. Eur J Biochem **183:** 555–563

Rusterucci C, Montillet JL, Agnel JP, Battesti C, Alonso B, Knoll A, Bessoule JJ, Etienne P, Suty L, Blein JP et al. (1999) Involvement of lipoxygenase-dependent production of fatty acid hydroperoxides in the development of the hypersensitive cell death induced by cryptogein on tobacco leaves. J Biol Chem **274:** 36446–36455

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

Schenk PM, Kazan K, Wilson I, Anderson JP, Richmond T, Somerville SC, Manners JM (2000) Coordinated plant defense responses in *Arabidopsis* revealed by microarray analysis. Proc Natl Acad Sci USA **97:** 11655–11660

Strompen G, Gruner R, Pfitzner UM (1988) An as-1-like motif controls the level of expression of the gene for the pathogenesis-related protein 1a from tobacco. Plant Mol Biol **37:** 871–883

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

Takeda S, Sugimoto K, Otsuki H, Hirochika H (1999) A 13-bp cis-regulatory element in the LTR promoter of the tobacco retrotransposon Tto1 is involved in responsiveness to tissue culture, wounding, methyl jasmonate and fungal elicitors. Plant J **18:** 383–393

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res **22:** 4673–4680

Vernhettes S, Grandbastien M-A, Casacuberta JM (1997) *In vivo* characterization of transcriptional regulatory sequences involved in the defense-associated expression of

the tobacco retrotransposon Tnt1. Plant Mol Biol **36:** 673–679

Vernhettes S, Grandbastien M-A, Casacuberta JM (1998) The evolutionary analysis of the Tnt1 retrotransposon in *Nicotiana* species reveals the high variability of its regulatory sequences. Mol Biol Evol **15:** 827–836

Vicient CM, Suoniemi A, Anamthawat-Jonsson K, Tanskanen J, Beharav A, Nevo E, Schulman AH (1999) Retrotransposon BARE-1 and its role in genome evolution in the genus *Hordeum*. Plant Cell **11:** 1769–1784

Wessler SR (1996) Plant retrotransposons: turned on by stress. Curr Biol **6:** 959–961

Xiang C, Miao ZH, Lam E (1996) Coordinated activation of as-1-type elements and a tobacco glutathione S-transferase gene by auxins, salicylic acid, methyl jasmonate and hydrogen peroxide. Plant Mol Biol **32:** 415–426

Genome evolution of wild barley (*Hordeum spontaneum*) by *BARE*-1 retrotransposon dynamics in response to sharp microclimatic divergence

Ruslan Kalendar*, Jaakko Tanskanen*, Sirkka Immonen†, Eviatar Nevo‡, and Alan H. Schulman*§

*Institute of Biotechnology, University of Helsinki, Plant Genomics Laboratory, Viikki Biocenter, P.O. Box 56, FIN-00014 Helsinki, Finland; †Agricultural Research Centre, Myllytie 10, FIN-31600 Jokioinen, Finland; and ‡Institute of Evolution, University of Haifa, Mount Carmel, 31905 Haifa, Israel

Edited by Susan R. Wessler, University of Georgia, Athens, GA, and approved March 28, 2000 (received for review January 3, 2000)

The replicative spread of retrotransposons in the genome creates new insertional polymorphisms, increasing retrotransposon numbers and potentially both their share of the genome and genome size. The *BARE*-1 retrotransposon constitutes a major, dispersed, active component of *Hordeum* genomes, and *BARE*-1 number is positively correlated with genome size. We have examined genome size and *BARE*-1 insertion patterns and number in wild barley, *Hordeum spontaneum*, in Evolution Canyon, Lower Nahal Oren, Mount Carmel, Israel, along a transect presenting sharply differing microclimates. *BARE*-1 has been sufficiently active for its insertional pattern to resolve individuals in a way consonant with their ecogeographical distribution in the canyon and to distinguish them from provenances outside the canyon. On both slopes, but especially on the drier south-facing slope, a simultaneous increase in the *BARE*-1 copy number and a decrease in the relative number lost through recombination, as measured by the abundance of solo long terminal repeats, appear to have driven the *BARE*-1 share of the genome upward with the height and dryness of the slope. The lower recombinational loss would favor maintenance of more full-length copies, enhancing the ability of the *BARE*-1 family to contribute to genome size growth. These local data are consistent with regional trends for *BARE*-1 in *H. spontaneum* across Israel and therefore may reflect adaptive selection for increasing genome size through retrotransposon activity.

Retrotransposons resemble retroviruses in their structure and life cycle (1, 2). They are ubiquitous (3–5) and contribute a large proportion of the total repetitive DNA of some plant genomes (6). Retrotransposons are mobilized by a replicative mechanism that has the capacity to generate and insert many new daughter copies into the genome, thereby increasing genome size (7). The error-prone nature of their replication by reverse transcriptase (8), the mutagenic potential of their transpositional integration (9), and the effects of their accumulation and recombination (10) together suggest that active retrotransposons may be major contributors to genome diversification in the plants. Genomic changes induced by retrotransposons can be tracked by the joints between the flanking DNA and the conserved retrotransposon termini created upon integration. Marker techniques based on PCR amplification between retrotransposons and flanking DNA recently have been developed (11–13).

Accumulated data indicate that retrotransposons in plants (14–16), animals, and fungi respond to various forms of stress. When stress factors in the environment vary ecogeographically, retrotransposon prevalence and insertion patterns may vary accordingly. The immediate wild ancestor of cultivated barley (*Hordeum vulgare*), *Hordeum spontaneum*, is ideal for analyzing retrotransposon insertions and their role in the genome because of the presence of a large and active retrotransposon family and the availability of well-studied wild populations distributed in diverse habitats (17–21). The *BARE*-1 family of retrotransposons comprises on average 14×10^3 copies in the genomes of *Hordeum* species (10). Members of this family are transcriptionally (22) and translationally (23) active, encoding both a polyprotein (24, 25) and processing signals (26), which are functionally conserved. The *BARE*-1 copy number is positively correlated with both genome size and habitat aridity (10), factors that are themselves correlated (27) regionally in *H. spontaneum*.

We have examined the role of the *BARE*-1 retrotransposon in genome diversification in individuals at the Evolution Canyon microsite, Lower Nahal Oren, Mount Carmel, Israel (28–30). This 400-m-wide erosion gorge (see Fig. 4, which is published as supplementary data on the PNAS web site, www.pnas.org), dating from the Plio-Pleistocene era, presents north- and south-facing slopes (NFS and SFS, respectively) with common geologies and macroclimates but microclimates sharply differing in solar irradiation and aridity. Biotically, the NFS is Eurasian and the SFS is Afro-Asian within the Mediterranean context (28, 30). We have examined *H. spontaneum* along a north-south transect across the canyon slopes to test whether regional patterns (10) can be detected locally. The *BARE*-1 copy number and patterns of insertional polymorphism, as well as total genome size, were determined for accessions from the canyon.

Materials and Methods

Plant Materials. Spikes from individual *H. spontaneum* plants were collected at six stations located along a 300-m north-south transect across the NFS and SFS of Evolution Canyon (28). The stations previously described (29, 31) as NFS (stations 5–7) and SFS (stations 1–3) are referred to here as: NH (north high), NM (north middle), NL (north low), SL (south low), SM (south middle), and SH (south high). From each station, seeds of 10 individual plants, separated by at least 1 m from each other, were used as the samples. The seeds were grown to seedlings for preparation of DNA and nuclei.

Retrotransposen-Microsatellite Amplified Polymorphism (REMAP) Amplification. DNAs were prepared by the cetyltrimethylammonium bromide method (32). For REMAP PCR amplification, primers facing outward from the long terminal repeats (LTRs)

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: *in*, integrase; LTR, long terminal repeat; NFS, north-facing slope; NH, north high; NM, north middle; NL, north low; REMAP, retrotransposen-microsatellite amplified polymorphism; SFS, south-facing slope; SH, south high; SL, south low; SM, south middle; SSR, simple sequence repeat.

See commentary on page 6250.

§To whom reprint requests should be addressed. E-mail: alan.schulman@helsinki.fi.

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.

Article published online before print: *Proc. Natl. Acad. Sci. USA*, 10.1073/pnas.110587497
Article and publication date are at www.pnas.org/cgi/doi/10.1073/pnas.110587497

were combined with anchored simple sequence repeat (SSR) primers. The PCR conditions and the SSR primers and annealing temperatures are as specified (see Table 1, which is published as supplementary data). The cycling program was: 94°C, 2 min; 30 cycles of 94°C for 30 sec, 54–58°C, depending on the primer pair, for 30 sec, a ramp of +0.5°C (sec)$^{-1}$ to 72°C, and 72°C for 2 min, 3 sec at 72°C being added with each cycle; 72°C for 10 min; maintenance at 4°C. Primers to the *BARE*-1 LTR were LTR-Z, 5′-ctc gct cgc cca CTA CAT CAA CCG CGT TTA TT-3′, a forward primer matching bases 1993–2012 of *BARE*-1a (GenBank accession no. Z17327), the lowercase bases indicating a cloning tail, and LTR-A, 5′-gga att cat aGC ATG GAT AAT AAA CGA TTA TC-3′, a reverse primer matching 369–393 of *BARE*-1a. The products were resolved on 2% NuSieve 3:1 agarose (FMC) and detected by ethidium bromide staining.

REMAP Data Analysis. Gel banding patterns were scored, and tables of band presence and absence were created. Principal component analyses were run with GENSTAT 5, release 4.1 (NAG, Oxford, U.K.), and the plots were generated with the group average agglomerative clustering function as used previously (33). Statistical tests were carried out with SIGMAPLOT 5.0 (SPSS, Chicago). Nonparametric correlation analyses were made by Spearman Rank Order (generating the correlation coefficient r_S) as implemented in SIGMASTAT 2.01 (SPSS). Values for P in the text represent the likelihood of falsely rejecting the null hypothesis that the variables are not correlated. Data values in the text are expressed as means and SE.

***BARE*-1 Copy Number Estimation.** Copy number was determined by reconstruction from dot blot hybridizations. To control for differences in loading, 139.4 ng of lambda DNA was added to each μg of plant DNA, giving 1.2×10^4 copies of lambda per barley genome equivalent. Dot blots were prepared with multiple replicates by using 1 ng or 10 ng of genomic DNA per sample and cross-linked under UV light. Isolated plasmids (0.1–10 ng) containing the fragments for hybridization probes served as controls on each filter. LTR (*Nhe*I–*Bst*EII, 743 bp) and integrase (*in*, *Hpa*I–*Bsm*I, 589 bp) probes were subcloned from *BARE*-1a. Probes were random-primed (Rediprime or Megaprime, Amersham Pharmacia) and ^{32}P-labeled. Filters were hybridized in 50% formamide, 1.25 × standard saline phosphate/EDTA (0.18 M NaCl/10 mM phosphate, pH 7.4/1 mM EDTA), 5 × Denhardt's reagent, 0.5% SDS, and 20 μg(ml)$^{-1}$ herring sperm DNA overnight at 42°C.

Hybridized filters were washed successively with 2 × SSC, 0.1% SDS (10 min, 25°C), twice in 2 × SSC, 0.1% SDS (10 min, 65°C), and once in 0.2 × SSC (20 min, 65°C). Bound radiation was quantified by exposure of a PhosphorImager screen for 45 min followed by scanning on a Fuji PhosphorImager. The same filter was probed in series with *in*, LTR, and lambda probes. Hybridization response to the *in* and LTR probes was corrected to the average value for the lambda hybridization response and relative copy number calculated. Absolute copy number was calculated from the hybridization response of the genomic DNA compared with the control plasmids: copies(ng)$^{-1}$ = (genomic cpm)(ng)$^{-1}$ × (plasmid copies)(plasmid cpm)$^{-1}$. Copies(ng)$^{-1}$ were converted to copies(genome)$^{-1}$ by using the *Hordeum* genome size data determined here.

Preparation of Nuclei and Flow Cytometry. All leaf samples (≈50 mg each) were collected from the three-leaf stage. Protoplasts were isolated as before (34), then resuspended in 1 ml of nuclear buffer {30 mM sodium citrate, pH 7.0/45 mM $MgCl_2$/20 mM Mops [3(*N*-morpholino)propanesulfonic acid]/1% (wt/vol) Triton X-100/5 μl(ml)$^{-1}$ β-mercaptoethanol} (35, 36), to which 20 μm(ml)$^{-1}$ propidium iodide then was added for staining. The samples were stained for 15 min and then centrifuged for 10 sec at 14,000 rpm. The supernatant was discarded and the pellet was suspended in 200 μl of nuclear buffer supplemented with 2.4 μl(ml)$^{-1}$ RNase and 1 μl(ml)$^{-1}$ of an internal standard solution (37) containing chicken red blood cells (2C = 2.33 pg; ref. 36). The tubes were incubated at 37°C for 15 min and then chilled on ice before analysis. An average of three separate determinations per individual were made on a 1,023-channel flow cytometer (FACSort, Becton Dickinson) having an argon-ion laser of 488-nm excitation wavelength. The output data were processed with the CELLQUEST program supplied with the cytometer. The estimates for the nuclear DNA amount for the samples were calculated by using the median position of the plant nuclear peak.

Results

Genome size was measured by flow cytometry (see Fig. 5, which is published as supplementary material) for accessions collected at each of six stations along a north-south transect across Evolution Canyon (Fig. 4). Taking all stations together, a diploid genome size of 9.037 ± 0.027 (SE) pg was observed and is within 1.1% of the average of previous observations for a set of Israeli *H. spontaneum* accessions (38). The transect through the canyon presents two position variables, relative height (lower, middle, or upper) and orientation (NFS or SFS). With the limited precision of flow cytometry, the observed genome sizes were not distinct by sampling site within the canyon (Kruskal–Wallis ANOVA on Ranks). However, linear regression analyses indicate that genome size is weakly associated with slope orientation ($R = 0.167$, $t = 1.684$, $P = 0.095$), the SFS having larger genomes than the NFS.

Two regions of *BARE*-1, the enzyme-encoding *in* of the internal domain and the terminal LTRs, served as probes for copy number determinations. The *in* and LTR regions both are conserved (25, 26) and were used in earlier *BARE*-1 copy number determinations for *Hordeum* (10). The *in* probe is used to estimate the number of full-length *BARE*-1 elements. Copy number and genome size were estimated on the same accessions, and together gave an average of $1.40 \pm 0.04 \times 10^4$ (range, 0.83 to 2.21×10^4) *BARE*-1 copies, equivalent to 2.98 ± 0.08% (range, 1.77% to 4.70%), of the haploid genome. This was in the range seen earlier for more broadly distributed *H. spontaneum* (10).

Because *BARE*-1 and other LTR-retrotransposons contain an LTR at each end, two are expected for each internal domain. However, the LTR copy number greatly exceeds that of the internal domain in barley, *H. spontaneum*, and throughout the *Hordeum* genus (10), because of the presence of large numbers of solo LTRs, hypothesized to result primarily from intraelement recombination between the LTRs and consequent loss of the internal domain. Here, we detected an average of $7.5 \pm 0.2 \times 10^4$ LTRs per genome, 5.4 ± 0.1-fold more LTRs than internal domains. This finding indicates that the average genome measured contains 4.7×10^4 LTRs not attributable to full-length *BARE*-1 elements. These solo LTRs contribute an additional 8.4×10^7 bp or 2.03 ± 0.07% to the genome.

Each solo LTR represents a minimum of one integration event followed by recombinational loss of the internal domain and an LTR. Assuming no other changes in the repetitive DNA complement, *BARE*-1 therefore would comprise at least 11.7 ± 0.3% of the genome if none were lost through recombination. Taking all accessions together, the number of full-length *BARE*-1 elements (measured by *in* response) is positively and highly significantly correlated (Fig. 1*A*, $r_P = 0.432$, $P = 0.001$) with the number of both total LTRs and solo LTRs (Pearson Product Moment, $r_P = 0.714$, $P = 1.35 \times 10^{-9}$). The slope for the regression plotting the growth in LTR numbers (Fig. 1*A*) is 3.7 ± 0.5 ($P < 0.001$), whereas as a slope of 2 would be expected if the increase in LTR number were to come only from those remain-

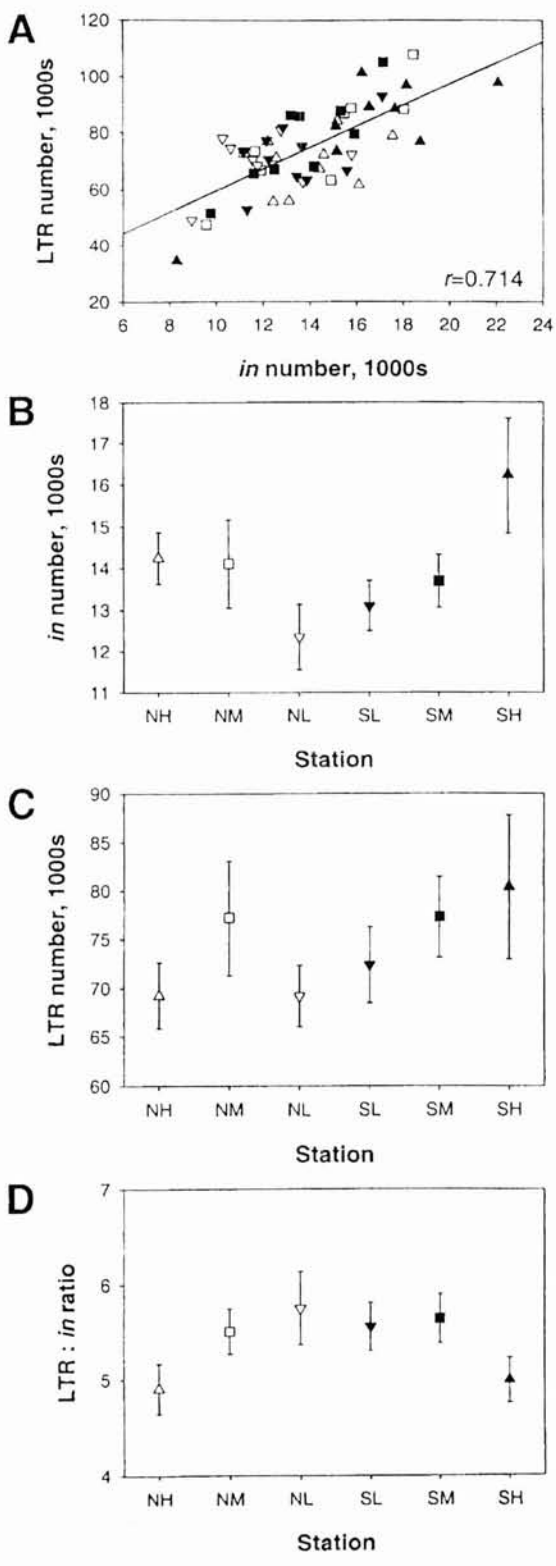


Fig. 1. BARE-1 copy number in *H. spontaneum* accessions at transect stations in Evolution Canyon. (*A*) LTR number as a function of *in* number for all accessions. (*B*) The *in* number as a function of transect station. (*C*) LTR number as a function of transect station. (*D*) Ratio of LTR number to *in* number as a function of station. For all plots, △ are samples from the NH station; □, NM; ▽, NL; ▼, SL; ■, SM; ▲, SH. For plots *B–D*, means and SE are shown.

ing in full-length elements. These calculations imply that those genomes that have more *BARE*-1 elements at present generally have also lost more through recombination.

Given that the number of LTRs, in particular the solo LTRs, rises more rapidly than the number of *BARE*-1 elements, one would expect a positive correlation between the *BARE*-1 genome share and the ratio of LTRs to full-length elements. However, in the canyon, the LTR/*in* ratio is negatively and highly significantly correlated ($r_P = -0.351$, $P = 0.009$) with the contribution of *BARE*-1 to genome size, as indeed was earlier seen for the genus *Hordeum* as a whole (10). Therefore, the higher the abundance of LTRs relative to full-length *BARE*-1 elements, the smaller the fraction of the genome occupied by *BARE*-1. Hence, variation in recombinational loss of *BARE*-1 may be linked to the genome share of the family both throughout the genus and for one species at a single geographical microsite.

In view of these correlations and the copy number and genome share variations among the *H. spontaneum* accessions, we examined whether measures of *BARE*-1 prevalence might vary with the height and orientation of the transect positions (Fig. 1*B*). The *BARE*-1 number is positively and significantly correlated ($r_P = 0.386$, $P = 0.004$) with the height of the accession site. Furthermore, the accessions from the top of the canyon (NH, SH), when considered together, have a distinctly and significantly greater (Student's *t* test, $t = 2.657$, $P = 0.01$) number of *BARE*-1 copies than found in the bottom- and mid-slope accessions. The SH station, the most stressed site in the canyon, alone is distinct from all others ($t = 3.107$, $P = 0.003$). The number of LTRs in the genome is positively but not significantly associated with height in the canyon, particularly on the SFS (Fig. 1*C*). If only the SFS is considered, the correlation between LTR number and height on the slope becomes very strong and nearly significant ($r_P = 0.991$, $P = 0.084$).

Variation in recombinational loss of *BARE*-1 with respect to insertional activity, reflected in the LTR/*in* ratio, is significantly and negatively correlated ($r_P = -0.351$, $P = 0.009$) with height of the accession on the canyon slopes (Fig. 1*D*). Furthermore, the accessions from NH and SH, taken together, have a distinctly smaller proportion of solo LTRs than all other accessions ($t = -2.985$, $P = 0.004$). These data suggest that the higher in the canyon, the more full-length *BARE*-1 elements are maintained relative to the number lost through recombination. Consistent with this, *BARE*-1 comprises a significantly increasing proportion of the genome ($r_P = 0.402$, $P = 0.0025$) with increasing height in the canyon. The SH individuals display a distinctly greater proportion of the genome than all of the rest of accessions ($t = 3.082$, $P = 0.003$), the distinction by height also being maintained when SH and NH are considered together ($t = 2.769$, $P = 0.008$) and when the top- and mid-slope accessions are compared with the two lowest stations ($t = 2.594$, $P = 0.012$). By multiple regression analyses, height in the canyon on both slopes is a good predictor of the *BARE*-1 share of the genome whether ($P = 0.01$) or not ($P = 0.008$) the SH station is included in the data set.

Using the REMAP method (12), integration joints between *BARE*-1 copies and flanking genomic sequences were detected by PCR amplification in reactions containing LTR primers in combination with primers to SSRs anchored at their 3′ end. Because *BARE*-1 elements have a tendency to insert into regions containing SSRs (12, 26, 39), many of the *BARE*-1 insertions thereby can be detected. Marker bands generated by this system are, in principal, insensitive to intraelement LTR recombination; the remaining, recombinant LTR would still serve as a priming site for PCR oriented in either direction.

Ten individuals from each of the six transect stations were typed with REMAP. Earlier results (12) showed that the REMAP products generally do not derive from amplification between pairs of SSR domains, but rather from element inser-

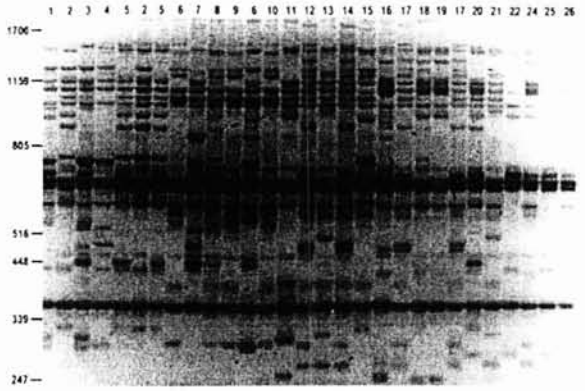


Fig. 2. Banding patterns generated by REMAP amplification. The reaction was carried out with primers LTR-A and $(CAC)_3T$. Lanes are labeled by the genotype of the sample (Table 2); two different accessions are shown for genotypes 2, 5, and 17. The products have been stained with ethidium bromide after agarose gel electrophoresis; the gel is shown as a negative image. Size markers in bp derive from a bacteriophage λ *Pst*I digest.

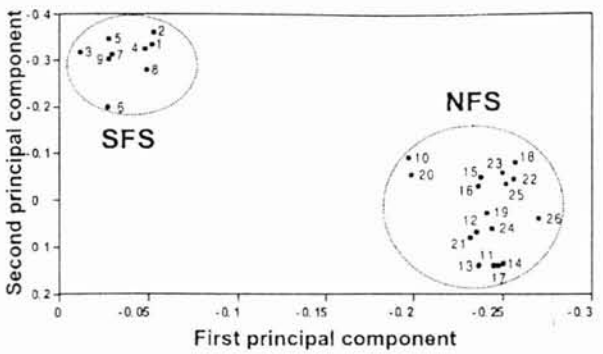


Fig. 3. Principle component analysis of Evolution Canyon *H. spontaneum* derived from variation in REMAP banding patterns. Numerals refer to the corresponding genotypes; those from the same slope have been circled.

tions. The inter-SSR products were generally longer than the REMAP products, indicating that the SSRs pairs are more widely interspersed than LTR/SSR pairs. In control experiments containing the SSR primer but not the LTR primer, none of the bands produced by amplification between SSR loci in the genome had mobilities identical to the REMAP bands (see Fig. 6, which is published as supplementary material). In the gel of Fig. 2, the generally high degree of polymorphism detected between individuals from the canyon is evident. Seven sets of REMAP primer combinations were used to generate 316 bands from the accessions (Table 1). Of these, 277 or 88% were polymorphic.

A total of 26 distinct banding genotypes were detected among the accessions (see Table 2, which is published as supplementary material). The particular primer combination (LTR-A, $(CAC)_3T$) used in Fig. 2 does not distinguish genotypes 22 and 23, although other combinations do. The genotypes distinguished all cases but one (genotype 6) within collecting stations and represented three or fewer of the individuals per genotype, excepting NFS genotype 17 with four individuals and SFS genotype 6 with 16 individuals. Genotype 6 also contained both SL and SM individuals. The slopes were clearly distinct in the number of genotypes represented, the SFS having only nine genotypes whereas the NFS had 17. A mean of 114 ± 1 bands were detected in total for each of the stations except SM, for which only 65 bands were scored (see Table 3, which is published as supplementary material). This yielded average frequencies from 0.1 to 0.7 for a given genotype and an intrastation similarity index of 0.47 ± 0.05, a value of 1 indicating all bands are shared. The banding patterns generated for each individual were used to estimate genetic distances between them. The REMAP banding data were examined by principal component analysis (Fig. 3), which allows comparison of overall genotypic similarities in the absence of phylogenetic considerations. The analyses completely separated the individuals from the NFS from those of the SFS.

Discussion

The number of full-length *BARE*-1 copies in individuals of *H. spontaneum* in Evolution Canyon, at the Lower Nahal Oren microsite, was found to range from 8.3 to 22.1 × 10^4 per haploid genome equivalent. This almost 3-fold variation in the number of full-length *BARE*-1 elements among individuals at a single microsite, within the range seen earlier for the genus *Hordeum* as a whole (10), indicates that this retrotransposon family has been highly and recently insertionally active.

An excess of LTRs, likely solo, were detected in all accessions. These are abundant across the genus *Hordeum* (10) and appear to result from intrachromosomal recombination between the LTR pairs within full-length elements. The excess in LTRs increases with the number of *BARE*-1 elements, consistent with earlier evidence that recombination is additive between elements in the genome (40). The results here furthermore confirm what was earlier shown (10) for the genus as a whole: the greater the number of solo LTRs relative to full-length *BARE*-1 elements, the smaller the part of the genome comprised by *BARE*-1. This indicates that variations in the relative rates of recombination and integration affect the success of a retrotransposon family in spreading within the genome, and that these variations may act within a single species at a single locale.

The data, moreover, suggest a linkage between *BARE*-1 numbers and the ecogeography of the Evolution Canyon microsite. More *BARE*-1 copies and proportionally fewer solo LTRs are found in the upper, drier sites within the canyon, particularly at the top of the SFS, than at lower sites. Earlier studies indicated a decrease in angiosperm species diversity (29, 31) and an increase in allozymic (30) and randomly amplified polymorphic DNA (28) diversity upward in the canyon, all correlated with increasing stress upward on both slopes, with the most stressful slope being the SFS. The upper stations on each slope are, furthermore, generally drier even during the wet season because of the movement of runoff down-slope. The local data at this single microsite mirror regional observations across Israel (10) that *BARE*-1 copy number was correlated with aridity across the range of the *H. spontaneum*, both sets of data being consistent with the presence within the *BARE*-1 promoter of abscisic acid-response elements typical for water stress-induced genes (24). The data therefore suggest that expression and propagation of *BARE*-1 may be stress induced and also that, the higher in the canyon, the lower the rate of loss of integrated copies through recombination.

Polymorphism detectable with REMAP markers yields a complete distinction between individuals growing on the NFS and SFS of Evolution Canyon, which are separated from each other by a maximum of 300 m. Because the REMAP pattern derives from short-range amplifications (hundreds of bases), the differences observed by REMAP are likely to have been generated by retrotransposon *BARE*-1 insertion, independent of other genetic changes among the individuals. Given the small percentage of the total *BARE*-1 copies visualized by the seven primer combinations, the data imply that *BARE*-1 integrational activity in the canyon has been greater than genomic homoge-

nization driven by gene flow through pollen dispersal among the largely selfing (average 98.4%, ref. 41) *H. spontaneum* or by seed dispersal. The REMAP marker data show more genotypes in the NFS than in the SFS individuals, which we interpret as being caused by the patchiness of the NFS, having open areas suitable for *H. spontaneum* interrupted by shaded, tree-growing areas, in contrast to the SFS. Under sufficiently high rates of *BARE*-1 integrational activity, the patches appear to have become genotypically distinct with regard to the *BARE*-1 insertion pattern.

Classically "selfish" self-replicating units such as retrotransposons might be expected, independent of the genome as a whole, to undergo selection for increasingly efficient propagation. However, the observed combination of decreased recombinational loss together with increases in the number of full-length copies suggests that plant-level selection is operating to increase *BARE*-1 copy number. Increasing numbers of transposon copies have been thought to be associated with decreased fitness through increasing lethality (42, 43). However, the tendency of retrotransposons to insert into repetitive DNA in barley and other cereals (6, 12, 26, 39) mitigates their deleterious potential.

The insertion and maintenance of full-length *BARE*-1 copies would marginally increase genome size, albeit against the background of fluctuations in the content of other retrotransposons and repetitive DNA. Selection for large genomes has been hypothesized to occur in the Mediterranean basin (44), a region where growth takes place primarily in the cool, wet winters and not in the dry summers. Growth is more efficient under cool conditions by increase in cell volume rather than by increase in cell number because cell division rates are decreased by low temperatures (44, 45). The potential for large cell volumes has been directly correlated with genome size and associated nuclear volume in a wide range of organisms (46, 47). Retrotransposon integrational activity, by increasing genome size, may be adaptive.

We thank Anne-Mari Narvanto for her always excellent technical assistance. This research was supported by the Academy of Finland Genome Research Program and the European Union Directorate for Biotechnology research program on Molecular Tools for Biodiversity. E.N. thanks the Israel Discount Bank Chair of Evolutionary Biology, the Ancell-Teicher Research Foundation for Genetics and Molecular Evolution, and the Israel Science Foundation (Grant 157/98) for financial support.

1. Boeke, J. D. & Corces, V. G. (1989) *Annu. Rev. Microbiol.* **43,** 403–434.
2. Doolittle, R. F., Feng, D.-F., Johnson, M. S. & McClure, M. A. (1989) *Q. Rev. Biol.* **64,** 1–30.
3. Voytas, D. F., Cummings, M. P., Konieczny, A. K., Ausubel, F. M. & Rodermel, S. R. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 7124–7128.
4. Flavell, A. J., Dunbar, E., Anderson, R., Pearce, S. R., Hartley, R. & Kumar, A. (1992) *Nucleic Acids Res.* **20,** 3639–3644.
5. Suoniemi, A., Tanskanen, J. & Schulman, A. H. (1998) *Plant J.* **13,** 699–705.
6. SanMiguel, P., Tikhonov, A., Jin, Y. K., Motchoulskaia, N., Zakharov, D., Melake-Berhan, A., Springer, P. S., Edwards, K. J., Lee, M., Avramova, Z. & Bennetzen, J. L. (1996) *Science* **274,** 765–768.
7. SanMiguel, P., Gaut, B. S., Tikhoniv, A., Nakajima, Y. & Bennetzen, J. L. (1998) *Nat. Genet.* **20,** 43–45.
8. Gabriel, A., Willems, M., Mules, E. H. & Boeke, J. D. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 7767–7771.
9. Wessler, S. R., Bureau, T. E. & White, S. E. (1995) *Curr. Opin. Genet. Dev.* **5,** 814–821.
10. Vicient, C. M., Suoniemi, A., Anamthawat-Jonsson, K., Tanskanen, J., Beharav, A., Nevo, E. & Schulman, A. H. (1999) *Plant Cell* **11,** 1769–1784.
11. Flavell, A. J., Knox, M. R., Pearce, S. R. & Ellis, T. H. N. (1998) *Plant J.* **16,** 643–650.
12. Kalendar, R., Grob, T., Regina, M., Suoniemi, A. & Schulman, A. H. (1999) *Theor. Appl. Genet.* **98,** 704–711.
13. Waugh, R., McLean, K., Flavell, A. J., Pearce, S. R., Kumar, A., Thomas, B. B. T. & Powell, W. (1997) *Mol. Gen. Genet.* **253,** 687–694.
14. Takeda, S., Sugimoto, K., Otsuki, H. & Hirochika, H. (1998) *Plant Mol. Biol.* **36,** 365–376.
15. Mhiri, C., Morel, J.-B., Vernhettes, S., Casacuberta, J. M., Lucas, H. & Grandbastien, M. A. (1997) *Plant Mol. Biol.* **33,** 257–266.
16. Wessler, S. R. (1996) *Curr. Biol.* **6,** 959–961.
17. Nevo, E., Beiles, A. & Zohary, D. (1986) *Biol. J. Linn. Soc.* **27,** 335–380.
18. Nevo, E., Zohary, D., Brown, A. H. D. & Haber, M. (1979) *Evolution* **33,** 815–833.
19. Nevo, E., Baum, B., Beiles, A. & Johnson, D. A. (1998) *Genet. Resour. Crop Evol.* **45,** 151–159.
20. Paknuyat, H., Powell, W., Baird, E., Handley, L. L., Robinson, D., Scrimgeour, C. M., Nevo, E., Hackett, C. A., Caligari, P. D. S. & Forster, B. P. (1997) *Genome* **40,** 332–341.
21. Forster, B. P., Russell, J. R., Ellis, R. P., Handley, L. L., Robinson, D., Hackett, C. A., Nevo, E., Waugh, R., Gordon, D. C., Keith, R. & Powell, W. (1997) *New Phytol.* **137,** 141–147.
22. Suoniemi, A., Narvanto, A. & Schulman, A. (1996) *Plant Mol. Biol.* **31,** 295–306.
23. Jääskeläinen, M., Mykkänen, A.-H., Arna, T., Vicient, C., Suoniemi, A., Kalendar, R., Savilahti, H. & Schulman, A. H. (1999) *Plant J.* **20,** 413–422.
24. Manninen, I. & Schulman, A. H. (1993) *Plant Mol. Biol.* **22,** 829–846.
25. Suoniemi, A., Tanskanen, J., Pentikäinen, O., Johnson, M. S. & Schulman, A. H. (1998) *Mol. Biol. Evol.* **15,** 1135–1144.
26. Suoniemi, A., Schmidt, D. & Schulman, A. H. (1997) *Genetica* **100,** 219–230.
27. Turpeinen, T., Kulmula, J. & Nevo, E. (1999) *Genome* **42,** 1094–1099.
28. Nevo, E. (1995) *Proc. R. Soc. London Ser. B* **262,** 149–155.
29. Nevo, E. (1997) *Theor. Popul. Biol.* **52,** 231–243.
30. Nevo, E., Apelbaum-Elkaher, I., Garty, J. & Beiles, A. (1997) *Heredity* **78,** 373–382.
31. Nevo, E., Fragman, O., Dafni, A. & Beiles, A. (1999) *Isr. J. Plant Sci.* **47,** 49–59.
32. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A., Struhl, K., Albright, L. M., Coen, D. M. & Varki, A. (1995) in *Current Protocols in Molecular Biology*, ed. Janssen, K. (Wiley, New York), Vol. 1, Suppl. 7, pp. 4.7.1–4.7.8.
33. Russell, J., Fuller, J., Young, G., Thomas, B., Taramino, G., Macaulay, M., Waugh, R. & Powell, W. (1997) *Genome* **40,** 442–450.
34. Valkonen, J. P. T., Nygren, M., Ylönen, A. & Mannonen, L. (1994) *Genetica* **92,** 203–207.
35. Villemont, E., Dubois, F., Sangwan, R. S., Vasseur, G., Bourgois, Y. & Sangwan-Norreel, B. S. (1997) *Planta* **201,** 160–172.
36. Galbraith, D. W., Harkins, K. R., Maddox, J. M., Ayres, N. M., Sharma, D. P. & Firoozabady, E. (1983) *Science* **220,** 1049–1051.
37. Arumuganathan, K. & Earle, E. D. (1991) *Plant Mol. Biol. Rep.* **9,** 229–241.
38. Kankaanpää, J., Mannonen, L. & Schulman, A. H. (1996) *Genome* **39,** 730–735.
39. Ramsay, L., Macaulay, M., Cardle, L., Morgante, M., degli Ivanissevich, S., Maestri, E., Powell, W. & Waugh, R. (1999) *Plant J.* **17,** 415–426.
40. Puchta, H. & Hohn, B. (1996) *Trends Plant Sci.* **1,** 340–348.
41. Brown, A. H. D., Zohary, D. & Nevo, E. (1978) *Heredity* **41,** 49–62.
42. Charlesworth, B., Sniegowski, P. & Stephan, W. (1994) *Nature (London)* **371,** 215–220.
43. Wilke, C. M., Maimer, E. & Adams, J. (1992) *Genetica* **86,** 155–173.
44. Grime, J. P. & Mowforth, M. A. (1982) *Nature (London)* **299,** 151–153.
45. Francis, D. & Barlow, P. W. (1988) *Symp. Soc. Exp. Biol.* **42,** 181–201.
46. Bennett, M. D. (1987) *New Phytol.* **106,** Suppl., 177–200.
47. Cavelier-Smith, T. (1982) *Annu. Rev. Biophys. Bioeng.* **11,** 273–302.

Short RNAs Can Identify New Candidate Transposable Element Families in Arabidopsis

M. Florian Mette, Johannes van der Winden, Marjori Matzke*, and Antonius J.M. Matzke

Institute of Molecular Biology, Austrian Academy of Sciences, Billrothstrasse 11, A–5020 Salzburg, Austria

Mining the Arabidopsis genome for transposable elements (TEs) by DNA sequence similarity searches and analysis protocols is revealing previously unidentified families of TEs and providing insights into TE structure, mobility, distribution, and diversity (Le et al., 2000). We suggest here that new putative TE families and partially diverged TE-like sequences can be identified by an alternate approach involving cloning and analyzing short RNAs, which are a hallmark of RNA silencing mechanisms.

RNA silencing is triggered by dsRNA that is cleaved to short RNAs 21 to 24 nucleotides (nts) in length by an RNase III-like enzyme termed Dicer (Matzke et al., 2001; Hutvágner et al., 2002). The short RNAs are thought to guide enzyme complexes that either degrade complementary RNAs in the cytoplasm (a process termed posttranscriptional gene silencing in plants and RNA interference [RNAi] in animals), or modify homologous DNA sequences in the nucleus (RNA-directed DNA methylation [RdDM]). In plants, RdDM can lead to transcriptional gene silencing if dsRNAs contain promoter sequences (Matzke et al., 2001). A major function of posttranscriptional gene silencing/RNAi and DNA methylation, which may result from RdDM in many cases, is to limit the proliferation of TEs (Matzke et al., 2000). The host defense role of RNA silencing is evidenced by the mobilization of some TEs in *Caenorhabditis elegans* mutants defective in RNAi (Ketting et al., 1999; Tabara et al., 1999) and in Arabidopsis mutants deficient in some aspect of DNA methylation or chromatin structure (Miura et al., 2001; Okamoto and Hirochika, 2001; Singer et al., 2001; Tomba et al., 2002). A role for RNA silencing in TE control is also supported by findings of sequences homologous to various TEs in collections of short RNAs cloned from different sources (Djikeng et al., 2001; Lagos-Quintana et al., 2001). The enrichment of known TE sequences in populations of short RNAs, which are presumably cleavage products of a Dicer-like enzyme, suggests that unidentified TEs might be detected through their presence in short RNA libraries. In an ongoing project to clone and sequence short RNAs approximately 17 to 27 nts in length from Arabidopsis leaves, we have isolated short RNAs that appear to be derived from previously unknown TE families and from TE-like sequences.

"40" FAMILY

One short RNA sequence has been isolated repeatedly and represents the most frequent nonstructural short RNA recovered in our study (11 independent clones comprising approximately 8% of total nonstructural RNAs). This group of short RNAs, designated the "40" family, ranges in size from 17 to 21 nts, with a fixed 5′ end and ragged 3′ ends. BLASTN searches revealed DNA sequence homology in three, unannotated intergenic regions of the Arabidopsis genome (Fig. 1). The only other highly similar sequences (identity in 20/21 nt) in the database are present in the *Oryza sativa* genome. Because the "40" short RNA family was exceptionally well represented in the population of cloned short RNAs, we investigated it further. An RNA folding program was used to examine whether the DNA sequences surrounding the short RNA "genes" could give rise to an RNA with a stable secondary structure. In all three cases, an approximately 200-bp imperfect RNA duplex, in which the short RNA is located in a semiconserved TIR, was generated (Fig. 1). Alignments of the three corresponding DNA sequences demonstrated that spacers internal to the TIRs of copies A and B display 79% DNA sequence identity, whereas the internal spacer of copy C, which is somewhat longer, shows no significant homology to A and B.

Although the potential RNA duplexes are quite long, only short RNAs derived from the TIR were cloned, indicating that this region is preferentially cleaved by a Dicer-like enzyme. Moreover, all 11 short RNAs originated exclusively from one side of the dsRNA (the 3′ end), which is the most conserved half of the TIR among the three copies. The presence of these short RNAs and their polarity were confirmed on northern blots: Only the antisense RNA probe produced a signal (Fig. 1). The "40" short RNA family appears to be relatively uniformly sized on the northern blot, displaying the same mobility as a 23-nt DNA oligonucleotide. The range of sizes that were cloned (17–21 nts) may indicate differences in the migration of short RNAs compared with DNA oligonucleotides or some degradation from the 3′ end during cloning procedures.

* Corresponding author; e-mail mmatzke@imb.oeaw.ac.at; fax 43-662-63961-29.

www.plantphysiol.org/cgi/doi/10.1104/pp.007047.

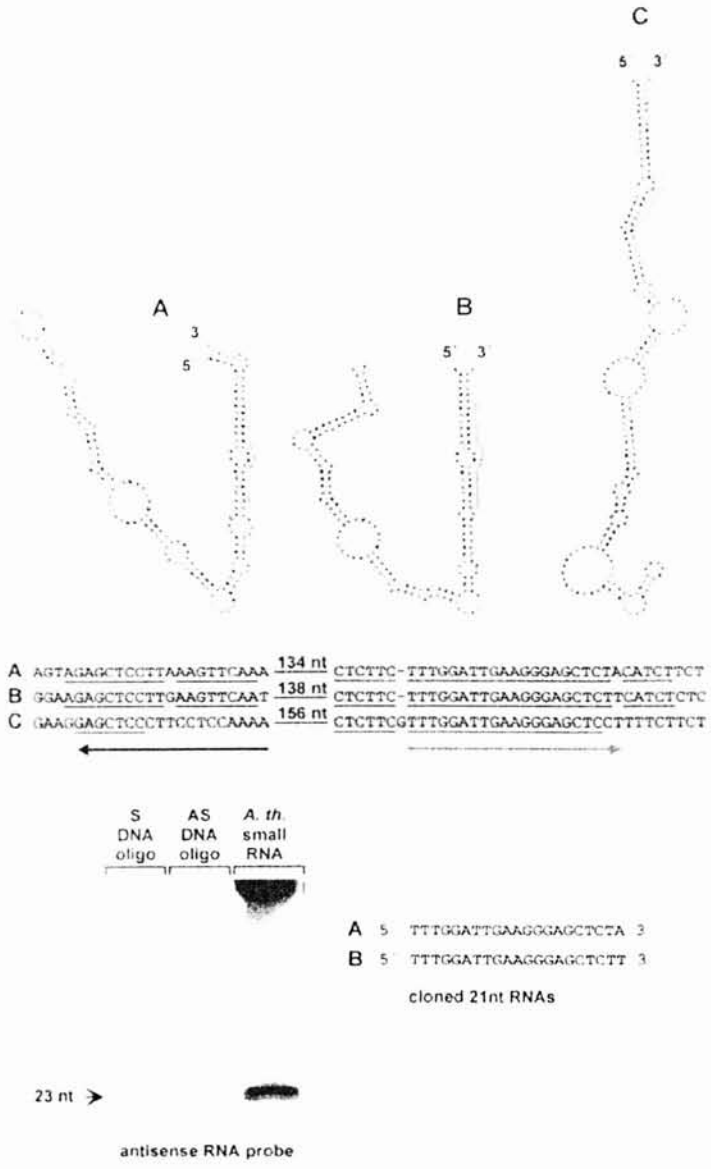


Figure 1. "40" family of short RNAs. Using the "40" short RNA sequences as queries in a BLASTN search, hits in intergenic regions in three BACs were obtained. The names of these BACs, the regions used to generate the predicted RNA secondary structures at the top, and the chromosome assignment are: A, F25P22, bases 41,361 through 41,170, chromosome 1; B, T10F20 (overlapping with T10O22), bases 26,724 through 26,919, chromosome 1; and C, T3F17, bases 28,928 through 28,714, chromosome 2. Middle, The DNA sequences conserved among copies A through C are underlined in black and the length of the spacer indicated. The long arrows below copy C denote the semiconserved terminal inverted repeats (TIRs), the right half of which gives rise to the "40" short RNA family (red arrow, red line in secondary structures at top). Bottom right, The sequences of the longest short RNA clones (21 nt) derived from copies A and B, which differ in the final nt, are shown. Bottom left, The northern blot confirms the presence of "40" short RNAs that hybridize to an antisense (AS) RNA probe in the Arabidopsis small RNA population; no hybridization signal was observed with a sense (S) RNA probe (not shown). Short RNAs were cloned according to a published procedure (Elbashir et al., 2001).

To generate the dsRNA structures that could be processed by a Dicer-like enzyme to yield the "40" family of short RNAs, the entire approximately 200-bp unit must be transcribed 5′ to 3′ from a promoter that has not yet been identified. At least copies A and B are transcribed, as exemplified by the sequences of two 21-mers, which differed in the 3′-most nt exactly according to the A and B DNA sequences (Fig. 1). Transcription of A and B must initiate either from adjacent intergenic promoters or by read-through transcription from the upstream host genes. Judging from the hybridization intensity on the northern blot, which approximates that observed with short RNAs derived from 35S promoter-driven transcripts (Mette et al., 2000), the precursor RNA for the "40" family of short RNAs is strongly transcribed.

The presence of multiple, dispersed copies of an approximately 135- to 155-bp DNA sequence flanked by relatively well conserved 20- to 30-bp TIRs in the Arabidopsis genome suggests that this small repeat family is possibly derived from a TE. Although sequence similarity generally falls off beyond the regions we have designated as TIRs, suggesting that they delimit a putative TE, we did not detect target site duplications, which would be expected from a class 2 (DNA) element. It is not yet clear whether and how members of the "40" family transpose, but their

sequence heterogeneity suggests they are degenerate relics of a previously active TE family. Whether the short RNAs derived from this putative TE family play a role in controlling transposition is not clear, but the striking conservation of both the "40" short RNA sequences and the potential secondary structures from which they are presumably derived suggests an important function.

When compared with known types of TEs, the "40" family appears similar to "neisseria miniature insertion sequences" (nemis). These are abundant, small DNA insertion sequences in the chromosome of the pathogenic bacterium *Neisseriae gonorrhoeae* (Mazzone et al., 2001). Unit length nemis (approximately 160 bp) feature TIRs (26–27 bp) and can potentially fold into a robust stem-loop structure. More than 66% of nemis are found close to cellular genes. In an intriguing parallel to the "40" family, the nemi RNAs appear to be cotranscribed with cellular genes and subsequently processed, at either one or both TIRs, by RNase III (14).

"175" FAMILY

One short RNA clone 24 nt in length was found to be homologous to regions of five BACs in the BLASTN search. In each bacteria artificial chromosome (BAC), there are two hits in inverted orientation that are separated by varying lengths of spacer DNA. For the longest copy (Fig. 2, BAC F10C21), there is one mismatch to the short RNA sequence (identity in 23/24 nts); for the other four copies, sequence identity is perfect (24/24 nts). The longest sequence is annotated as a putative MudrA transposase, suggesting a MULE family. Alignments of all five sequences demonstrated that they are related by common TIRs approximately 330 bp in length that flank internal deletions of varying sizes (Fig. 2). Each element copy is flanked by a 6- to 9-bp target site duplication.

The "175" family is distinct from MULE families described so far in Arabidopsis (Yu et al., 2000), supporting the claim that short RNA sequences can draw attention to previously unidentified TE families. The longest copy, which contains the coding region of MudrA transposase, is possibly an autonomous element that has degenerated rapidly to produce a heterogeneous group of internally deleted, nonautonomous derivatives (Fig. 2). The short RNA could originate from either the left (F14F8, T4B21) or right (F14J22, F25O24) half of the TIR of an internally deleted copy (Fig. 2). The existence of short RNAs derived from the TIR region suggests transcription through the entire element and intramolecular pairing to form a dsRNA, which would probably be produced most readily with transcripts issuing from one of the more extensively deleted copies (Fig. 2).

The "175" short RNAs are less abundant than the "40" family of short RNAs, as indicated by the recovery of fewer independent clones and a fainter signal on northern blots (Fig. 2). In addition, the "175" short RNAs are longer than those in the "40" family. The "175" short RNAs migrate as a doublet slightly above the 23-nt DNA oligonucleotide standard, which is consistent with the 24-nt length of the cloned RNA. Whether the size difference between the two short RNA families reflects the activity of different dicer-like enzymes, and/or the subcellular localization of dsRNA processing (nucleus or cytoplasm), is not yet known.

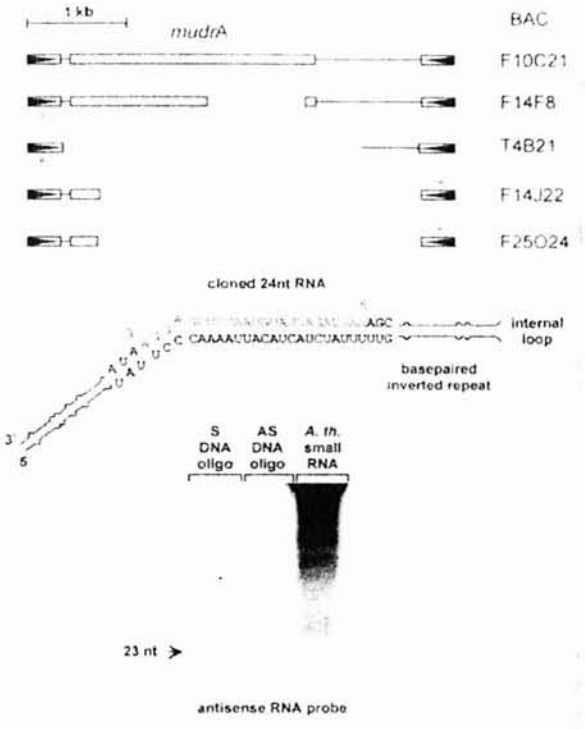


Figure 2. "175" family of short RNAs. Middle, Sequence of the 24-nt short RNA of the "175" family (red) and its position in the stem of a predicted RNA duplex generated by folding a transcript homologous to the shortest putative *Mutator*-like TE (MULE) derivative (BAC F25O24). The small red arrows indicate possible origins of the short RNA in the TIRs of the "175" MULE family. The family members are in the following regions of Arabidopsis genome: F10C21, bases 66,969 through 62,619, chromosome 1; F14F8, bases 26,965 through 30,372, chromosome 5; T4B21, bases 45,822 through 44,489, chromosome 4; F14J22, bases 50,446 through 49,368, chromosome 1; and F25O24, bases 3,075 through 4,118, chromosome 4. The northern blot confirms the presence of "175" short RNAs that hybridize to an antisense (AS) RNA probe in the Arabidopsis small RNA population. No signal was visible using a sense (S) RNA probe (not shown).

Our data suggest that investigating short RNA populations can help to identify new candidate TE families and partially diverged TE-like sequences that might be missed in conventional DNA sequence analyses. In contrast to DNA sequence similarity

searches, this approach focuses on putative TEs that are transcribed to produce dsRNA and that might be targets of transcriptional and posttranscriptional RNA silencing mechanisms. Certain short RNAs, such as those from the "40" family, are extraordinarily well represented in the short RNA population, whereas the degree of representation of standard cellular RNA genes appears considerably less (A. Matzke and M.F. Mette, unpublished results). It is striking that the short RNAs we have described in this report always originate in regions of the genome that can potentially give rise to dsRNA, indicating that they are not random products of single stranded RNA degradation. As we found with the "175" short RNA, which revealed a putative MULE family, tiny RNAs can pinpoint a widely spaced inverted repeat comprising two halves that show high DNA sequence similarity. Moreover, as shown by the "40" family, a short RNA can also reveal imperfect inverted repeats that might be undetectable from the DNA sequence alone. Because G-U pairing is allowed in RNA secondary structures, however, an RNA duplex can form from a transcript of the region.

Much remains to be learned about areas of the genome that are transcribed to produce dsRNA precursors of short RNAs, and the identity of the RNA polymerase(s) involved. In addition to their possible role in controlling transposition, it will be interesting to determine whether short RNAs derived from TEs and TE-like sequences are involved in host gene regulation. TEs flanked by TIRs, including MITEs, *Mutator* elements, and nemis, frequently integrate next to host genes, thus potentially furnishing these genes with target sites for complementary short RNAs arising from members of the TE family that produce dsRNA. Conceivably, such TEs or their derivatives might be sources of micro-RNAs (miRNAs), at least some of which are involved in developmental timing of gene expression in *C. elegans* and possibly other animals (Lagos-Quintana et al., 2001; Lau et al., 2001; Lee and Ambros, 2001; Lai, 2002). The miRNAs are derived via Dicer cleavage of an approximately 70-nt precursor that can form an imperfect RNA duplex. Similar to the "40" family described here, miRNAs usually accumulate from only one arm of the fold-back precursor. The reason for the asymmetry in short RNA accumulation is unclear, but it might indicate preferential stabilization of the copy that can base pair with the target RNA. It is also not known why short RNAs originate only from the TIR region of the putative "40" TE family, even though the predicted RNA duplexes comprise spacer sequences. Further studies on short RNAs and the intergenic regions that encode them should help answer these questions.

Received April 11, 2002; accepted May 16, 2002.

LITERATURE CITED

Djikeng A, Shi H, Tschudi C, Ullu E (2001) RNA **7:** 1522–1530

Elbashir S, Lendeckel W, Tuschl T (2001) Genes Dev **15:** 188–200

Hutvágner G, Zamore P (2002) Curr Opin Genet Dev **12:** 225–232

Ketting R, Haverkamp T, van Luenen H, Plasterk R (1999) Cell **99:** 133–141

Lagos-Quintana M, Rauhut R, Lendeckel W, Tuschl T (2001) Science **294:** 853–858

Lai E (2002) Nat Genet **30:** 363–364

Lau N, Lim L, Weinstein E, Bartel D (2001) Science **294:** 858–862

Le QH, Wright S, Yu Z, Bureau T (2000) Proc Natl Acad Sci USA **97:** 7376–7381

Lee R, Ambros V (2001) Science **294:** 862–864

Matzke MA, Matzke AJM, Kooter J (2001) Science **293:** 1080–1083

Matzke M, Mette MF, Matzke AJM (2000) Plant Mol Biol **43:** 401–415

Mazzone M, De Gregorio E, Lavitola A, Pagliarulo C, Alifano P, Nocera P (2001) Gene **278:** 211–222

Mette MF, Aufsatz W, van der Winden J, Matzke M, Matzke AJM (2000) EMBO J **19:** 5194–5201

Miura A, Yonebayashi S, Watanabe K, Toyama T, Shimada H, Kakutani T (2001) Nature **411:** 212–214

Okamoto H, Hirochika H (2001) Trends Plant Sci **6:** 527–534

Singer T, Yordan C, Martienssen R (2001) Genes Dev **15:** 591–602

Tabara H, Sarkissian M, Kelly W, Fleenor J, Grishok A, Timmons K, Fire A, Mello C (1999) Cell **99:** 123–132

Tomba R, McCallum C, Delrow J, Henikoff J, van Steensel B, Henikoff S (2002) Curr Biol **12:** 65–68

Yu Z, Wright S, Bureau TE (2000) Genetics **156:** 2019–2031

Commentary

Retrotransposon-mediated genome evolution on a local ecological scale

Jonathan F. Wendel* and Susan R. Wessler†‡

*Department of Botany, Iowa State University, Ames, IA 50011; and †Departments of Botany and Genetics, University of Georgia, Athens, GA 30602

The dynamic genome is a concept associated with the discovery of transposable elements by Barbara McClintock. Her Nobel lecture concluded with a challenge to biologists considering this issue (1). She wrote,

> We know about the components of genomes that could be made available for such restructuring. We know nothing, however, about how the cell senses danger and instigates responses to it that are often truly remarkable. (ref. 1, pp. 800–801)

Fifteen years later we still know nothing about the mechanisms underlying genome-restructuring events in response to environmental cues. This is despite numerous studies, especially in plants, suggesting a connection between growth conditions and one form of genome restructuring, a change in genome size. In many plants there is impressive variation in total genomic DNA content among individuals and populations (2) including *Helianthus annuus* [sunflower, 50% within-plant reduction (3, 4)], *Pisum sativum* [pea, 1.29-fold variation (5)], *Linum usitatissimum* [flax, 1.16-fold (6)], and *Glycine max* [soybean, 1.15-fold variation (7)]. Although this variation is correlated with environmental gradients or growth conditions in a number of species (2–4), in no cases have the specific genomic components of DNA content change been identified.

In this issue of PNAS, Kalendar *et al.* (8) document an example of genome size variation in natural populations of the wild barley *Hordeum spontaneum*. This paper shows that an abundant and active component of the barley genome, namely the *BARE-1* long terminal repeat (LTR)-retrotransposon, displays nearly a three-fold intraspecific copy number variation. Furthermore, correlations between *BARE-1* copy number, genome size, and local environmental conditions suggest, for the first time, a testable molecular mechanism linking habitat with retrotransposon induction in natural populations.

LTR retrotransposons are members of the retroelement or Class I family, which also includes retroviruses, long interspersed nuclear elements (LINEs, also known as non-LTR retrotransposons), and short interspersed nuclear elements (SINEs). LTR retrotransposons are flanked by long terminal repeats and usually encode all of the proteins required for their transposition, including a capsid (Gag), protease, integrase, reverse transcriptase, and RNase H. For all Class I elements, it is the element-encoded transcript (mRNA), and not the element itself, that forms the transposition intermediate. Transcription of most of the active plant elements characterized to date is largely quiescent during normal development but can be induced by biotic and/or abiotic stresses, including cell culture, wounding, and pathogen attack (9, 10). For two elements, the tobacco *Tnt1* and the rice *Tos17*, increased transcription is correlated with retrotransposition (11, 12).

LTR retrotransposons are the most abundant transposable element class in grass genomes, of which barley is a member (reviewed in ref. 13). In fact, differential amplification of LTR retrotransposons largely accounts for the C-value paradox in this group of organisms. The C-value paradox is the observed lack of correlation between DNA content and organismal complexity (14). It has been documented for both animal and plant species but, to date, only appears to be "solved" for the members of the grass tribe. That is, the fraction of the genome contributed by LTR retrotransposons increases with genome size from rice, the smallest characterized grass genome [430 Mbp, ≈14% LTR retrotransposons (15)], through maize [≈3,200 Mbp, 50–80% retrotransposons (16)] to barley [≈4,800 Mbp, >70% retrotransposons (17)]. For maize, SanMiguel *et al.* (18) made the remarkable discovery that the majority of the retrotransposon insertion events occurred very recently, within the last two to six million years.

As discussed above, we are beginning to understand the relative contribution and time scale of retrotransposition among different grass species. However, little is known about the dynamics of transposable element copy-number evolution within and among natural populations, or its significance with respect to natural selection. In a similar vein, it is well known that transposition events may lead to modified patterns of gene expression, but this process has rarely been demonstrated to be selectively relevant within natural populations. Thus, the possible connections between genome size variation and adaptive genic evolution (as illustrated in Fig. 1) have remained elusive.

Kalendar *et al.* (8) may have taken a first step toward intertwining these once disparate threads. In a study of natural populations of wild barley (*Hordeum spontaneum*) from a single canyon in Israel, they describe patterns of retrotransposon accumulation on a local spatial scale. Their data demonstrate a striking degree of population-level genome dynamics and suggest what well may be an example of retrotransposon-mediated adaptive evolution.

The barley plants studied derived from six natural populations distributed across a 300-m transect of a single canyon. Ten individuals were sampled from each population, which were selected to span the spectrum of local edaphic and microclimatic conditions present in the canyon, including potentially important ecological variables such as level of solar irradiation and aridity. Each individual was genetically fingerprinted, and copy number was estimated for the barley retrotransposon *BARE-1*, a relatively high copy-number (average of 14,000 copies/*Hordeum* species) family of elements that earlier was shown to be transcriptionally (19) and translationally active and assembled into virus-like particles (20). In the present study, full-length *BARE-1* retroelements were shown to comprise an average of nearly 3% of the approximately 4.5-pg haploid wild barley genome, accumulating to a mean of 14,000 copies per genome. Although this observation is not in itself especially noteworthy, the variance in

See companion article on page 6603

‡To whom reprint requests should be addressed. E-mail: sue@dogwood.botany.uga.edu

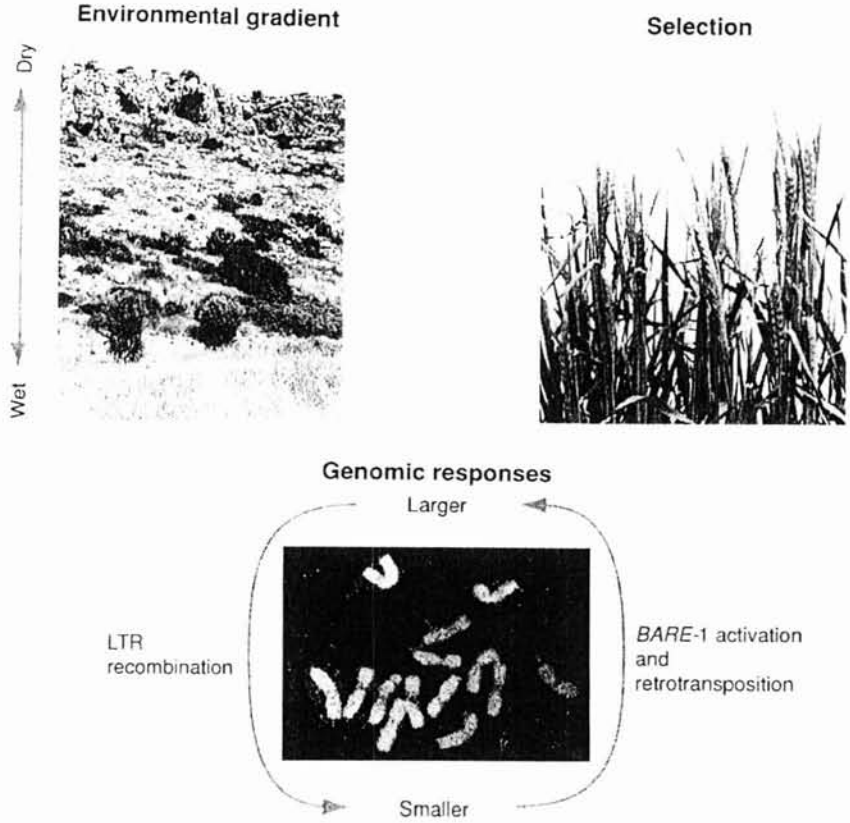


Fig. 1. Genome evolution on a local ecological scale. Wild barley plants are distributed along ecological gradients both regionally and locally, and vary nearly three-fold in copy number for the retrotransposon *BARE*-1 (*in situ* hybridization to barley chromosomes, bottom). Illustrated here are the intriguing interconnections between local adaptation to a moisture gradient (upper left) in a single canyon in Israel and the correlated distribution of *BARE*-1 copy-number (8). Local adaptation conceivably may be facilitated by direct selection on genome size (genome-level selection) or from functionally relevant physiological effects of individual *BARE*-1 insertions (gene-level selection). Credits: M Kemppinen (barley, upper right); K. Anamthawat-Jónsson (*BARE*-1 *in situ* hybridization); A. Schulman (canyon); A. Gardner (assistance with illustration).

copy number among closely spaced natural plants is unprecedented and remarkable: a nearly three-fold range in copy number (8,300–22,100) was observed among individuals of the six populations, corresponding to 1.77–4.70% of the nuclear DNA.

Such extraordinary variation in retroelement copy number among spatially adjacent plant populations implicates a history of recent transposition, a suggestion supported by "REMAP" DNA fingerprinting. This technique, which couples PCR priming sites in *BARE*-1 LTRs with those designed from simple sequence repeats, was used by Kalendar *et al.* to show that wild barley plants from this single erosion gorge have high levels of interindividual polymorphism for REMAP fragments. An important implication of these observations, when considered in light of the striking local variation in *BARE*-1 copy-number, is that retroelement proliferation may contribute to genome size evolution within and among local populations (Fig. 1). Extrapolated to a more global level, this study may provide a snapshot of the dynamics that underlie patterns of C-value evolution.

A more provocative implication of the Kalendar *et al.* study emerges from consideration of the spatial distribution of *BARE*-1 copy-number among wild barley plants. When the REMAP genetic fingerprinting data were subjected to multivariate analysis, populations from the north- and south-facing slopes of the gorge clearly were distinguished. On both slopes, but particularly on the drier, south-facing slope, there was a significant positive correlation between height in the canyon and *BARE*-1 copy-number. These data parallel regional trends observed in a broader sampling of *H. spontaneum* populations collected from across Israel (17) and suggest a relationship between retroelement accumulation and one or more ecological variable related to the sampled populations. The most obvious variable is moisture availability; higher sites and those from the south-facing slope are the driest and thus potentially the most water-stressed. By far the highest *BARE*-1 accumulation is in the highest site from the south-facing slope. Kalendar *et al.* note a remarkable connection between the presence, within the *BARE*-1 promoter, of ABA (abscissic acid)-response elements, found in water stress-induced genes (19), and *BARE-1* copy number variation, suggesting that *BARE*-1 proliferation in wild barley populations may be stress-induced. With the important caveat that the data are correlative rather than causal, it is tempting to speculate that other examples of interpopulational DNA content variation (2) will similarly be found to result from stress-induced retrotranspositional activity.

An intriguing aspect of the *BARE*-1 data of Kalendar *et al.* concerns the relative abundance of full-length elements and solo LTRs. The latter, which are relatively rare in the maize genome (21) but are common in yeast and *Hordeum* species (17, 22, 23), are thought to arise from intraelement or perhaps intrachromosomal recombination between transiently paired LTRs. Kalendar *et al.* used dot-blot reconstruction to estimate copy number for both LTRs and *BARE*-1 integrase genes and found an average of 5.4-fold more LTRs than internal domains. These data show that recombinational loss of *BARE*-1 elements is an important factor limiting element accumulation in wild barley populations. Significantly, the geographical sites with the highest *BARE*-1 copy number, i.e., those from the most stressed sites, have the highest ratio of full-length to solo LTRs, suggesting once again a connection between environmental sensing and either rates of recombinational loss (favored explanation of Kalendar *et al.*) or recent bursts of retrotranspositional activity.

A central question that emerges from this study concerns the role, if any, of *BARE*-1 element proliferation in the stress response. One might postulate, for example, that water stress-induced epigenetic modifications have led to release from suppression of *BARE*-1 retrotransposition in the higher, drier sites, but that this burst of element activity has been independent of the actual adaptively significant physiological responses. Under this scenario, local adaptation may be taking place in wild barley populations, but this adaptation is postulated to arise from genetic and/or epigenetic changes unrelated to *BARE*-1 activity. Alternatively, perhaps the relationships between *BARE*-1 activity, water-stress, and

adaptation are not only correlative but causal. To the extent that this is true, the mode of action of natural selection in the process remains mysterious, as does the organizational level on which selection might be manifested. It may be, for example, that selection is operating on one or more aspects of genome size that we presently do not perceive of as adaptively relevant. In contrast to whole-genome selection, perhaps retroelement activation has led to adaptively relevant insertions that affect drought-tolerant pathways or other ecologically relevant physiologies. Given that thousands of insertions appear to distinguish wild barley populations from adjacent sites, it may be a daunting challenge to ferret out the adaptively significant insertions.

Notwithstanding the number of remaining issues, the study of Kalendar *et al.* provides perhaps the best example yet of the dynamic nature of plant genome evolution on a local ecological scale, and hints at retrotransposon-mediated adaptive evolution. In this regard, the authors have taken the first significant step toward addressing McClintock's challenge to figure out how cells restructure their genomes in response to perceived danger.

1. McClintock, B. (1984) *Science* **226,** 792–801.
2. Price, H. J. (1988) *Evol. Trends Plants* **2,** 53–60.
3. Price, H. J. & Johnston, J. S. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 11264–11267.
4. Johnston, J. S., Jensen, A., Czechin, D. G. & Price, H. J. (1996) *Am. J. Bot.* **83,** 1113–1120.
5. Cavallini, A., Natali, L., Cionini, G. & Gennai, D. (1993) *Heredity* **70,** 561–565.
6. Evans, G. M., Durant, A. & Rees, H. (1966) *Nature (London)* **212,** 697–699.
7. Graham, M. J., Nickell, C. D. & Rayburn, A. L. (1994) *Theor. Appl. Genet.* **88,** 429–432.
8. Kalendar, R., Tanskanen, J., Immonen, S., Nevo, E. & Schulman, A. H. (2000) *Proc. Natl. Acad. Sci USA* **97,** 6603–6607.
9. Wessler, S. R. (1996) *Curr. Biol.* **6,** 959–961.
10. Grandbastien, M.-A. (1998) *Trends Plant Sci.* **3,** 181–189.
11. Hirochika, H. (1993) *EMBO J.* **12,** 2521–2528.
12. Hirochika, H., Sugimoto, K., Otsuki, Y. & Kanda, M. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 7783–7788.
13. Kumar, A. & Bennetzen, J. L. (1999) *Annu. Rev. Genet.* **33,** 479–532.
14. Thomas, C. A. (1971) *Annu. Rev. Genet.* **5,** 237–256.
15. Tarchini, R., Biddle, P., Wineland, R., Tingey, S. & Rafalski, A. (2000) *Plant Cell* **12,** 381–391.
16. SanMiguel, P. & Bennetzen, J. L. (1998) *Ann. Bot. (London)* **81,** 37–44.
17. Vicient, C. M., Suoniemi, A., Anamthawat-Jonsson, K., Tanskanen, J., Beharav, A., Nevo, E. & Schulman, A. H. (1999) *Plant Cell* **11,** 1769–1784.
18. SanMiguel, P., Gaut, B. S., Tikhonov, A., Nakajima, Y. & Bennetzen, J. L. (1998) *Nat. Genet.* **20,** 43–45.
19. Suoniemi, A., Narvanto, A. & Schulman, A. H. (1996) *Plant Mol. Biol.* **31,** 295–306.
20. Jääskelainen, M., Mykkanen, A.-H., Arma, T., Vicient, C., Suoniemi, A., Kalendar, R., Savilahti, H. & Schulman, A. H. (1999) *Plant J.* **20,** 413–422.
21. SanMiguel, P., Tikhonov, A., Jin, Y.-K., Motchoulskaia, N., Zakharov, D., Melake-Berhan, A., Springer, S., Edwards, K. J., Lee, M., Avramova, Z. & Bennetzen, J. L. (1996) *Science* **274,** 765–768.
22. Jordan, I. K. & McDonald, J. F. (1999) *Genetics* **151,** 1341–1351.
23. Shirasu, K., Schulman, A. H., Lahaye, T. & Schulze-Lefert, P. (2000) *Genome Res.*, in press.

The Plant Journal (2001) **28**(2), 159–168

The mobility of the tobacco Tnt1 retrotransposon correlates with its transcriptional activation by fungal factors

Delphine Melayah[1], Eric Bonnivard[2], Boulos Chalhoub[3], Colette Audeon and Marie-Angèle Grandbastien*
Laboratoire de Biologie Cellulaire, INRA, 78026 Versailles, France

Received 6 April 2001; revised 9 July 2001; accepted 17 July 2001.
*Correspondence (fax +33 (1) 30 83 30 99; e-mail gbastien@versailles.inra.fr).
[1]Present address: Université Claude Bernard LYON 1, UMR CNRS 5557 Ecologie Microbienne, Symbiose Mycorhizienne, 43 Boulevard du 11 Novembre 1918, 69622 Villeurbanne Cedex, France.
[2]Laboratoire de Dynamique du Génome et Evolution, Institut Jacques Monod, 2 place Jussieu – 75252 Paris cedex 05, and
[3]Unité de Recherches en Génomique Végétale, INRA, 2 rue Gaston Crémieux, CP 5708, 91057 Evry cedex, France

Summary

We have analyzed the stress-induced amplification of the tobacco Tnt1 element, one of the rare active plant retrotransposons. Tnt1 mobility was monitored using the retrotransposon-anchored SSAP strategy that allows the screening of multiple insertion sites of high copy number elements. We have screened for Tnt1 insertion polymorphisms in plants regenerated from mesophyll leaf cells, either via explant culture or via protoplast isolation. The second procedure includes an overnight exposure to fungal extracts known to induce high levels of Tnt1 transcription. Newly transposed Tnt1 copies were detected in nearly 25% of the plants regenerated via protoplast isolation, and in less than 3% of the plants derived from explant culture. These results show that Tnt1 transcription is followed by transposition, and that fungal extracts efficiently activate Tnt1 mobility. Transcription appears to be the key step to controlling Tnt1 amplification, as newly transposed Tnt1 copies show high sequence similarities to the subpopulations of transcribed Tnt1 elements. Our results provide direct evidence that factors of microbial origin are able to induce retrotransposon amplification in plants, and strengthen the hypothesis that stress modulation of transposable elements might play a role in generating host genetic plasticity in response to environmental stresses.

Keywords: fungal factor, protoplast, retrotransposon, stress, Tnt1, tobacco.

Introduction

Stress and environmental challenges are known to trigger surprisingly similar defensive mechanisms in different eukaryotic organisms (Taylor, 1998). In particular, recent evidence suggests the possible generation of genetic plasticity in response to stress through the mobilization of retrotransposons, the most widespread eukaryotic mobile elements. Stress and external challenges, including microbial attacks, are known to be major factors activating retrotransposon expression in a wide range of organisms such as yeast (Bradshaw and McEntee, 1989; Rofte *et al.*, 1986), drosophila (Strand and McDonald, 1985), mammals (Liu *et al.*, 1995) and plants (Grandbastien, 1998; Wessler, 1996). So far, the biological significance of the stress modulation of retrotransposons remains unclear. However, it is in agreement with McClintock's original model that postulates that transposable elements are involved in genome restructuring in response to environmental challenges (McClintock, 1984). A recent study indeed suggests that retrotransposon amplification has been involved in barley adaptative evolution to drought conditions (Kalendar *et al.*, 2000). An alternative, but not exclusive, hypothesis is that activation by microbial challenges might also favor horizontal transmission and allow elements to colonize new hosts. A parallel can be noted between this hypothesis and the recent proposition that stress activation of retroviral replication corresponds to an escape mechanism for the virus from damaged or stressed host cells (Andrews *et al.*, 1998).

Any debate on the putative biological impact of retrotransposon activation implies that stress induces not only expression but also subsequent transposition. However, except for yeast, direct evidence of retrotransposition in

response to stress is scarce and often controversial. This situation partly results from the difficulty in establishing experimental systems to demonstrate mobility in multicellular organisms (Arnault and Dufournel, 1994). Additionally, post-transcriptional inhibition mechanisms are often used by hosts to control potentially deleterious retrotranspositions (Curcio and Garfinkel, 1999; Menees and Sandmeyer, 1996). In particular, there is little evidence for retrotransposon mobilization by stress in plants, where experimental tests for retrotransposition are particularly difficult due to the high copy number of most plant retrotransposons. Examples of transposon insertions were reported in progenies of virus-infected maize plants (Dellaporta *et al.*, 1984; Johns *et al.*, 1985). However, a causal relationship between virus infection and transposition still remains to be established in these cases. So far, a link between transcription and retrotransposition has been demonstrated for a few low copy number elements, in response to *in vitro* tissue culture (Hirochika *et al.*, 1996; Hirochika, 1993).

The Tnt1 element of tobacco (*Nicotiana tabacum*) is one of the best characterized plant retrotransposons (Grandbastien *et al.*, 1989b). Tnt1 is not expressed in healthy tobacco tissues, except in roots (Pouteau *et al.*, 1991). Tnt1 expression is strongly induced by stresses such as pathogen attacks and factors of microbial origin, and a tight correlation has been shown between Tnt1 transcription and plant defense responses (Grandbastien *et al.*, 1997; Mhiri *et al.*, 1997; Mhiri *et al.*, 1999; Moreau-Mhiri *et al.*, 1996; Pouteau *et al.*, 1994). In contrast to many other elements, Tnt1 transcription is poorly activated by tissue culture, at least in tobacco (Grandbastien *et al.*, 1997; Hirochika, 1993; Pauls *et al.*, 1994). However, Tnt1 is highly expressed in freshly isolated tobacco protoplasts (Pouteau *et al.*, 1991). A detailed study of the different factors involved in protoplast isolation has shown that wounding, plasmolysis or hormone addition have little effect on Tnt1 transcription, and that Tnt1 expression in protoplasts is a direct response to the Onozuka solution, one of the three components of the cocktail of fungal extracts used to digest cell walls (Pouteau *et al.*, 1991). Onozuka is a crude extract prepared from the fungus *Trichoderma viride* and contains proteinaceaous elicitors that activate defense responses in tobacco, leading to necrotic hypersensitive cell death (Bailey *et al.*, 1990; Lotan and Fluhr, 1990; Yano *et al.*, 1998). Tnt1 expression is detected soon after application (Grandbastien *et al.*, 1997), and Onozuka also activates Tnt1 expression when applied to intact tobacco tissues (Pouteau *et al.*, 1994), indicating that Tnt1 is activated in direct response to elicitors contained in the Onozuka solution. The effect of the fungal extracts is transient, as Tnt1 transcript levels rapidly decrease after their removal, in the early stages of subsequent cell culture (Grandbastien *et al.*, 1997).

Tnt1 is mobile and was originally isolated after transposition into a target gene. However, in spite of a good knowledge of the conditions and mechanisms activating Tnt1 expression, no evidence was yet reported that Tnt1 transcriptional activation is correlated to subsequent Tnt1 mobility. The low level of Tnt1 expression in response to tissue culture stimuli or plasmolysis has allowed us to develop an experimental strategy aimed at testing for the effect of fungal extracts on Tnt1 mobility. We have monitored Tnt1 transposition in tobacco plants regenerated from mesophyll leaf cells, either via protoplast isolation, or as controls, via leaf explant culture. In both procedures, regenerated plants originate from cells that have been submitted to stress stimuli linked to successive steps of wounding, cell dedifferentiation, callus formation and shoot regeneration. The major difference between the two procedures consists of an early step of plasmolysis that does not induce detectable Tnt1 expression, associated with overnight exposure to fungal extracts, including the Onozuka solution known to induce very high levels of Tnt1 expression. The comparison of Tnt1 transposition between each of these two procedures is thus expected to provide an accurate estimate of the effect of the fungal extracts on Tnt1 mobility.

Since Tnt1 copy number has been estimated at several hundred copies (Grandbastien *et al.*, 1989b), we used a high resolution retrotransposon-anchored PCR strategy allowing the simultaneous detection of multiple insertion sites of high copy number elements, described as the Sequence-Specific Amplification Polymorphisms (SSAP) technique (Ellis *et al.*, 1998; Gribbon *et al.*, 1999; Waugh *et al.*, 1997) or as Transposon Display (Casa *et al.*, 2000; Van den Broek *et al.*, 1998). Newly transposed copies have been screened through the appearance of new polymorphic bands in SSAP profiles of regenerated plants. We have also characterized these new SSAP polymorphic bands, in order to analyze the sequences of the population of newly transposed Tnt1 elements and to compare them with the sequences of the population of elements known to be transcribed in the same stress conditions (Casacuberta *et al.*, 1995).

Results

Tnt1 insertion polymorphisms in regenerated plants

The SSAP procedure outlined in Figure 1 was developed to analyze Tnt1 insertion polymorphism in regenerated tobacco plants. Since most Tnt1 elements do not contain EcoRI restriction sites (Grandbastien *et al.*, 1991), EcoRI digestions were performed to avoid internal amplifications from the 3′LTR. A primer pair consisting of an EcoRI primer (E00) and a Tnt1-specific primer (LTR13) was used for SSAP amplifications. The Tnt1-specific primer, LTR13, was

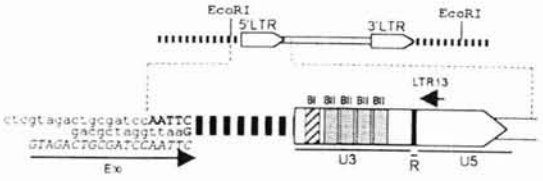


Figure 1. Schematic representation of the SSAP strategy used to analyze Tnt1 insertion polymorphism.
Top: Tnt1 general structure; bottom: enlargement of the region containing the SSAP-amplified sequence. Boxes represent Tnt1 sequence and thick dotted lines represent flanking genomic DNA. The three functional domains (U3-R-U5) of the LTR are indicated. The BI and BII regulatory motifs present in the U3 domain of Tnt1A elements are indicated by striped and shaded boxes, respectively. The sequences of the EcoRI adaptators (lower case letters) and of the E00 primer (capital italics) are indicated, together with the remaining nucleotides resulting from the EcoRI digestion (capital bolds). The positions and orientations of the SSAP primers E00 and LTR13 are indicated by arrows.

designed in the highly conserved U5 region, 250 pb downstream of the 5′ end of the element, and oriented towards the 5′ end. The E00/LTR13 primer combination thus allows amplification of variable amounts of 5′ flanking genomic DNA, together with a 5′LTR portion that includes the U3 region, previously shown to be highly variable in Tnt1 populations (Casacuberta *et al.*, 1995; Vernhettes *et al.*, 1998). Control SSAP experiments were performed on six tobacco genotypes (Figure 2a). In our experimental conditions, about 80 bands per SSAP profile were observed. These profiles are similar for all genotypes, except for the PBD6 line, which shows a low number of polymorphic bands. These results indicate that the tobacco lines used in this study are closely related. In addition, no SSAP polymorphism was found in different plants of the D8, tl and PBD6 lines and in different leaves of D8 plants, thus indicating no detectable interplant or intraplant variability (data not shown).

Following protoplast isolation or direct explant culture, 20–30 plants were regenerated per genotype, leading to totals of 156 plants and 147 plants, respectively. Tnt1 insertion polymorphism was examined in all the regenerated plants using the SSAP procedure. After protoplast isolation, new SSAP bands were detected in 11.8% to 42.9% of the regenerated plants, depending on the genotype (Table 1). SSAP profiles obtained for a few representative plants are illustrated in Figure 2b. χ^2 tests of homogeneity show that intergenotype differences are not statistically significant ($\chi^2_{df5} = 6.37$, $P = 0.27$). The presence of the *tl* somatic instability, known to increase the level of spontaneous mutations (Grandbastien *et al.*, 1989a), does not lead to a significant increase in new SSAP bands. Altogether, an average of 24.4% of all plants regenerated via protoplast isolation contain new SSAP bands. In contrast, after direct explant culture, new SSAP

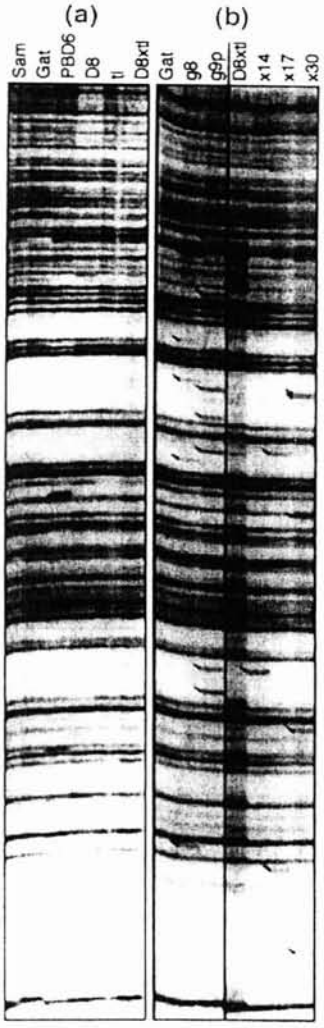


Figure 2. SSAP profiles obtained with the E00/LTR13 primer combination for the six tobacco genotypes analyzed in this study (a), and for independent regenerants obtained after protoplast isolation (b).
g8 and g9p are regenerants obtained from the Gat genotype and x14, x17 and x30 are regenerants obtained from the D8xtl hybrid genotype. Arrows indicates new SSAP bands.

bands were detected in a much lower percentage of the regenerated plants, ranging from 0% to 6.6% (Table 1). The Fisher's exact test demonstrates that intergenotype differences are not statistically significant either ($P = 0.41$). An average of 2.7% of all plants regenerated via explant culture contain new SSAP bands. Altogether, a significant difference ($\chi^2_{df1} = 29.68$, $P < 0.001$) is found between plants regenerated via protoplast isolation or via explant culture, indicating a 9-fold increase in the appearance of new SSAP bands in response to the protoplast isolation

Table 1. Tnt1 insertion polymorphisms in regenerated tobacco plants

Genotype	Plants regenerated from leaf protoplasts						Plants regenerated from leaf explants					
	Number of tested plants	Plants with new bands		New bands per profile		Transposed copies[1]	Number of tested plants	Plants with new bands		New bands per profile		Transposed copies[1]
		number	%	number	mean			number	%	number	mean	
Gat	21	9	42.9%	1–7	2.55	11/12	20	0	0%	0	0	0/0
D8	33	7	21.2%	1–4	2.43	3/5	30	0	0%	0	0	0/0
tl	29	6	20.7%	1–5	3.5	3/3	27	1	3.7%	3	3	1/1
D8 × tl	29	6	20.7%	1–5	2.67	9/9	16	1	6.2%	1	1	nt
Sam	27	8	29.6%	1–3	1.75	8/8	24	0	0%	0	0	0/0
PBD6	17	2	11.8%	1	1	1/1	30	2	6.6%	1	1	2/2
Total	156	38	24.4%		2.45	35/38	147	4	2.7%		1.5	3/3

[1]confirmed transpositions/number of tested bands; nt =-non-tested

procedure. The number of new SSAP bands revealed on E00/LTR13 profiles varied from one to seven per plant (Table 1), with a mean number of bands of 2.45 and 1.5 for plants regenerated via protoplast isolation or explant culture, respectively. Due to the low number of explant-derived plants containing new SSAP bands, it is not possible to determine if the numbers of new SSAP bands per plant statistically differ between the two populations of regenerated plants.

Thirty-eight new SSAP bands isolated from 22 plants regenerated via protoplast culture, and three new SSAP bands isolated from three plants regenerated via explant culture, were sequenced after re-amplification from the gel. The transpositional nature of each new SSAP band was tested by direct PCR performed using the LTR13 primer and primers designed in each flanking genomic sequence. Thirty-eight out of 41 new SSAP bands were confirmed to correspond to transposition events (Table 1). The remaining three bands, obtained from three different protoplast-derived regenerants, could not be identified as transposed copies. In these three cases, however, no consensus sequence was obtained from the few clones recovered from the cloning step (see Experimental procedures), suggesting that the failure to identify the transposed copy represents the background of technical problems inherent in our cloning strategy. Taken together, our results demonstrate that the vast majority (93%), and possibly all, of the new SSAP bands detected in regenerated plants correspond to newly inserted Tnt1 copies.

Sequence variability of the populations of transposed Tnt1 elements

The tobacco Tnt1 family is composed of an heterogeneous population of elements that can be grouped in subfamilies characterized by different U3 regions (Vernhettes *et al.*, 1998). We have previously shown that only the Tnt1A subfamily, characterized by the BI and BII regulatory sequences involved in Tnt1 stress-activation (Casacuberta and Grandbastien, 1993; Vernhettes *et al.*, 1997), is expressed in protoplasts (Casacuberta *et al.*, 1995). Furthermore, protoplast-specific transcripts are not a unique sequence, but a population of different, albeit very closely related, RNAs (Casacuberta *et al.*, 1995). Since the first half part of the 5′LTR is amplified by the E00/LTR13 primer pair (Figure 1), it is possible to compare the U3 regions of the Tnt1 copies that have transposed in protoplast-derived plants with the U3 regions of previously characterized Tnt1 protoplast-specific transcripts. The unexpected characterization of preexisting insertions (see Experimental procedures) also allowed the comparison of newly transposed Tnt1 copies with ancient copies that inserted earlier during tobacco evolution.

All the newly transposed copies analyzed belong to the Tnt1A1 group (Figure 3a). Furthermore, a striking correlation was found between the U3 sequences of transcribed and transposed copies. We have previously shown that two major U3 subpopulations could be defined in the protoplast-specific transcript population, the P23 and the P1 RNAs, each representing 24% of the population, and both containing four BII regulatory repeats (Casacuberta *et al.*, 1995). The P1 and P23 RNAs differ in the U3 region by four signature nucleotide changes only (G, A, A, T) (Figure 3a) and the U3 regions of the remaining RNAs differ from P1 and P23 by one or two single nucleotide changes only, as well as by the sporadic deletion of a complete BII repeat, that may be considered as a single mutational event generated during the reverse transcription process (Casacuberta *et al.*, 1995). Our results show that most of the newly transposed copies can also be grouped in P1-type (four copies) and P23-type (24 copies) sequences. Furthermore, 10 (30%) of the newly transposed Tnt1 copies are identical to the P23 sequence. A third subpopulation of newly transposed copies is closely

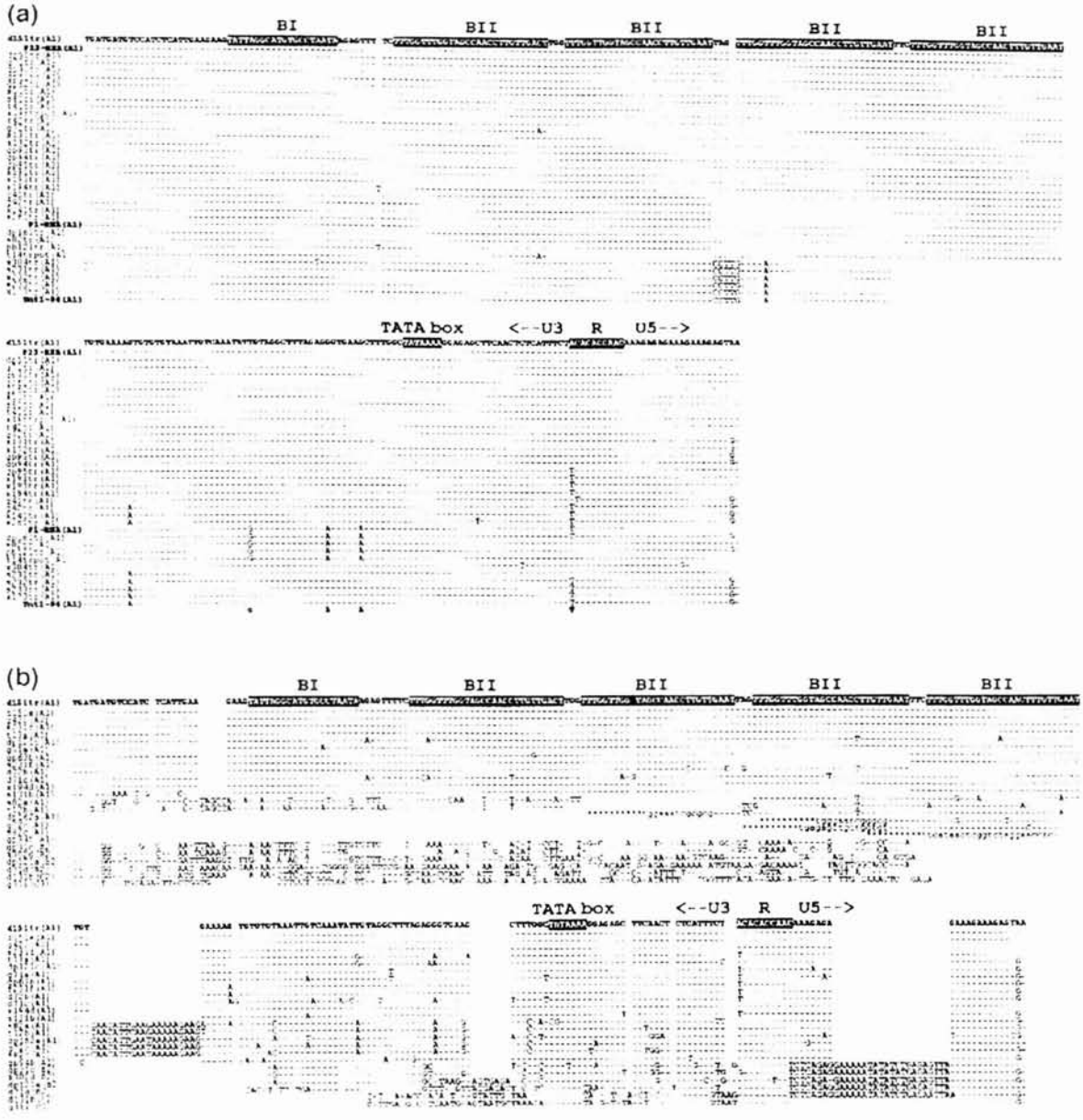


Figure 3. Sequence variability of the partial LTR sequences of newly transposed Tnt1 copies characterized in protoplast-derived plants (a) and of preexisting Tnt1 insertions (b).

Sequences have been aligned on the d161tr transposed copy, chosen as a reference because of its identity to the major P23 Tnt1 RNA species. Sequences of the P23 and P1 Tnt1 RNA species as well as of the Tnt1-94 mobile copy are included (the P23 and P1 RNA sequences terminate at the end of the R region). Dashes and blanks indicate sequence identity and deletions, respectively. Nucleotides that differ from the reference sequence are shown in capital letters. Lower case letters and stars indicate that the 5' end of the LTR could not be determined, since no homology to previously characterized Tnt1 LTRs could be detected upstream from the sequences shown by dashes. Blanks have been introduced in the reference sequence to allow for nucleotide insertions. Black boxes indicate important features of the Tnt1A LTR (BI and BII boxes, TATA box, R region). The subfamily and group of each sequence, as defined in Vernhettes *et al.* (1998) are indicated in brackets.

related to the Tnt1-94 element originally isolated after insertion into the nitrate reductase encoding gene and characterized by an atypical third BII repeat (Grandbastien *et al.*, 1989b) (Figure 3a, six lower lines). Interestingly, transcripts of this type were never characterized in protoplasts. Finally, it is noticeable that Tnt1 copies

newly transposed in the same regenerant belong usually, but not exclusively (e.g. s81tr and s82tr), to the same subpopulation type.

These data indicate that there is a good correlation, at the level of the regulatory U3 region, between the Tnt1 transcripts generated by protoplast isolation and the Tnt1 copies that have successfully completed the retrotransposition cycle. However, a significant difference between transcribed and transposed copies is found with respect to the number of BII repeats. We have previously shown that 90 percent of transcribed elements contain 4 BII repeats, indicating that BII sequences were essential for Tnt1 activation in protoplasts (Casacuberta *et al.*, 1995). Surprisingly, a significantly lower proportion of newly transposed copies containing four BII repeats (64%) has been observed in this work ($\chi^2_{df1} = 8.48$, $P < 0.001$). In addition, one of the newly transposed copies, t42tr, contains only two BII repeats, although transcripts containing fewer than three BII repeats have never been characterized (Casacuberta *et al.*, 1995).

In contrast to newly transposed Tnt1 copies, the regulatory regions of pre-existing Tnt1 insertions are much more heterogeneous (Figure 3b). While most of them (75%) belong to the Tnt1A1 group, only four of them belong to the P23-type and one of them to the P1-type. All the remaining pre-existing Tnt1A1 copies contain more than four bp changes. Only six (33%) of them contain four BII repeats, and several others display larger-scale modifications such as insertions or apparent 5′ truncations (dp162a, g72c, s82c and gp94b). The remaining ancient Tnt1 insertions belong either to the Tnt1B (17%) or to the Tnt1C (8%) subfamilies.

Discussion

Microbial factors activate retrotransposon amplification

Using the SSAP strategy, we have monitored the amplification of one of the few plant retrotransposons known to be active, the tobacco Tnt1 element. Here we report that the protoplast isolation procedure activates Tnt1 transposition efficiently in tobacco, since a nine-fold increase in Tnt1 mobility is observed in plants regenerated via protoplast isolation, compared with the Tnt1 mobility observed in control plants regenerated via explant culture. Since transcription is a prerequisite for the transposition of retroelements, this sharp increase in Tnt1 mobility results necessarily from one or several protoplast-specific factors able to strongly activate Tnt1 expression. Previous studies have shown that the major factor activating Tnt1 expression during the protoplast isolation procedure is the initial overnight application of Onozuka fungal extracts (Pouteau *et al.*, 1991). The subsequent early steps of protoplast-derived cell culture are expected to have very little impact on Tnt1 transposition, since Tnt1 transcript levels decrease strongly in the few hours following removal of the fungal extracts (Grandbastien *et al.*, 1997), and no Tnt1 transcript could be detected by Northern analysis in tobacco cell cultures (Hirochika, 1993). Furthermore, Tnt1 is not expressed in tobacco calli (Pauls *et al.*, 1994). As other factors involved in the production of the regenerated plants were also involved in the generation of the control plants from leaf explants, it can therefore be assumed that the nine-fold increase in Tnt1 mobility observed in plants regenerated from leaf protoplasts is specifically attributable to the effect of the cell wall hydrolyzing fungal extracts, most essentially the Onozuka solution. Our results therefore show that Tnt1 mobility correlates to its transcriptional activation by factors of microbial origin.

New insertions were detected in nearly 3% of the control plants regenerated from leaf explants. This demonstrates that stimuli other that microbial factors are also able to activate Tnt1 mobility. This background mobility could result from the original wounding step, since previous studies showed that, although no transcript could be detected by Northern analysis in shredded tobacco leaf tissues (Pouteau *et al.*, 1991), a low level of expression was detected in mechanically wounded tobacco leaves through the use of a reporter gene placed under control of the Tnt1 LTR (Mhiri *et al.*, 1997). Alternatively, background transpositions could result from a low level of activation during tissue culture. Although no Tnt1 transcript could be detected by Northern analysis, a small increase in Tnt1 copy number was indeed observed in established cell cultures (Hirochika, 1993). It is noticeable that new SSAP bands were usually detected in a reproducible manner in an SSAP experiment performed with DNA isolated from duplicate cuttings of the regenerated plantlets (data not shown). This suggests that the regenerated plants are probably not chimaeric for the new insertions, and that the transpositions that we detected have occured very early in the regeneration process. They might therefore originate from the original wounding step rather than from the successive rounds of transposition expected from a response to tissue-culture stimuli. Finally, we cannot exclude the possibility that background transpositions could have a somatic origin in the leaves used for explant and protoplast isolation, due to a possible low background of Tnt1 expression in mature leaves. No interplant or intraplant polymorphisms were detected in our control analyzes, but somatic transpositions occuring in a few leaf cells would not be detectable by the SSAP procedure.

Interestingly, microbial-induced transposition levels measured in this work are strikingly high, since transpositions specifically induced in protoplasts have occured in over one protoplast out of five. Due to competitive PCR amplification, the E00 primer, which contains no additional selective nucleotide in 3′, reveals only a fraction (80) of all

tobacco Tnt1 insertions. The total number of Tnt1 insertions in tobacco was indeed estimated to be more than 500 through SSAP analyzes using EcoRI primers containing two additional selective nucleotides (unpublished data). It is therefore likely that the total number of newly transposed Tnt1 copies that we observed in each plant on E00/LTR13 profiles is also underestimated. The high level of Tnt1 transposition observed in our *in vitro* experimental strategy opens the perspective to assess efficiently the genetic impact of stresses such as microbial attacks or generalized abiotic stresses applied on whole plants, since the plant germline is not sequestered early and somatic transpositions can be transmitted to the progeny in plants. Although the functional consequences of the new Tnt1 insertions remain to be established, insertions of transposable elements are thought to be a major source of genetic diversity, by generating direct gene mutation or genome structural rearrangements (Kidwell and Lisch, 2000). Our work represents the first direct demonstration that factors of microbial origin, known to activate plant defense responses, are also able to generate plant genomic diversity through retrotransposition amplification. Our results therefore strengthen the hypothesis that stress modulation of transposable elements might play a role in generating host genetic plasticity in response to environmental stresses (McClintock, 1984). Interestingly, *T. viride* does not invade tobacco tissues, but merely activates tobacco defense responses and hypersensitive cell death. A successful pathogen infection is thus not a prerequisite for Tnt1 amplification, and mobilization might also result from incompatible or non-host plant–microbe interactions, provided that tobacco defense responses are activated. Since such 'unsuccessful' microbial challenges are quite frequent in natural conditions, and since resistant or non-host plants are much more likely to produce progenies than heavily infected plants, the evolutionary impact of retrotransposon activation by microbial challenges might be quite significant.

Transcription is a key step controlling Tnt1 amplification in response to pathogen factors

Previous studies have shown that the population of tobacco Tnt1 transcripts displays a structure similar to the populations of closely related but different viral genomes referred to as 'quasi-species', and that the LTR U3 region is important for expression in protoplasts (Casacuberta *et al.*, 1995). We show here a striking similarity between the U3 sequences of Tnt1 elements transcribed in protoplasts and the U3 sequences of Tnt1 copies that have subsequently completed their retrotransposition cycle and inserted into the tobacco genome. Interestingly, the sharp contrast between the populations of newly transposed copies and the more variable pre-existing insertions allows us to determine with little error, by visual analysis of U3 regions, whether a given Tnt1 element has been recently active or is an ancient insertion. The effect of fungal extracts on Tnt1 expression is transient, since their removal after the overnight treatment is rapidly followed by a sharp decrease in Tnt1 transcript levels (Grandbastien *et al.*, 1997). This indicates that induced transpositions have mostly occured during, or shortly after, the application of the fungal extracts, at the single cell level. It can thus be assumed that transpositions in each cell have occured in a single transposition burst. Interestingly, new Tnt1 copies detected in the same regenerant can belong to different subpopulations, indicating that different master copies can be simultaneously active within the same cell.

Since the population of transposed copies is expected to reflect template RNAs that succeed in completing the downstream steps of the retrotransposition process, the tight correlation between transcribed and transposed pools suggests that there is very little or no template selection for Tnt1 reverse transcription. Transcription appears therefore to be a key step controlling Tnt1 retrotransposition, and no specific post-transcriptional control, for instance at the reverse transcription or the integration steps, seems to further select for particular Tnt1 types, at least in pathogen-related activating conditions. It is, however, interesting to note that newly transposed copies contain a significantly lower number of BII repeats, compared with RNA templates. Deletion of repeated sequences frequently occur during retroviral replication and retrotransposition (Pathak and Temin, 1990; Xu and Boeke, 1987). We have previously suggested that BII deletions during reverse-transcription could be involved in the control of active Tnt1 populations (Casacuberta *et al.*, 1995). The significant decrease in the proportion of 4-BII elements that we observe after completion of the reverse transcription cycle strengthens this hypothesis. Interestingly, an unexpected 2-BII transposed copy appears as a possible single direct deletion derivative of two adjacent BII. However, since we also observed in this work the frequent transposition of Tnt1–94-type copies, poorly represented at the transcriptional level (Casacuberta *et al.*, 1995), we cannot exclude the possibility that 3-BII elements are in fact more efficiently reverse transcribed than 4-BII elements.

SSAP is a very efficient tool for monitoring retrotransposon behavior and genomic impact

Most plant retrotransposons are present in high copy number in the genome, and information on their activity is mostly available through transcript analyzes (Grandbastien, 1998). To date, direct retrotransposition has only been monitored by characterization of insertional

mutants (Grandbastien *et al.*, 1989b; Varagona *et al.*, 1992; White *et al.*, 1994), or for low copy number elements (Hirochika *et al.*, 1996; Hirochika, 1993). We show here that the SSAP strategy is an efficient tool to detect and quantify the mobility of high copy number elements. SSAP provides an interesting alternative to reverse transcription tests based on intron splicing from indicator genes (Heidmann and Heidmann, 1991), since complete retrotransposition events, up to the insertion step, are directly detected. Consequently, results can be directly interpreted in terms of genomic impact. In addition, SSAP provides a tool for isolation of new mobile retrotransposon copies when only partial LTR data are available, since it allows the detection of transposed copies without any need for insertional mutants for a known target gene. Finally, SSAP performed with carefully designed primers allows a direct and easy access to LTR sequences, the most variable region of retroviruses and retrotransposons. In addition to polymorphism studies and retrotransposition assays, SSAP can therefore be used to correlate the activity and the population evolution of retrotransposons and endogenous retroviruses, and to study more efficiently their behavior and impact on host genomes.

Experimental procedures

Plant material

Six tobacco genotypes were used: (1) *N. tabacum* cv Xanthi: the XHFD8 line ('D8'), a homozygous diploid line obtained from an anther culture-derived haploid line (Bourgin and Missonier, 1973); the *tl*-1086 line ('tl'), carrying the somatically instable *tl* mutation (Deshayes, 1979). The *tl* mutation, originally obtained on *N. tabacum* cv Samsun (Dulieu, unpublished), was transfered to *N. tabacum* cv Xanthi by successive backcrosses (Deshayes, 1976). The F1 hybrid between the XHFD8 and tl-1086 lines was also used ('D8xtl'). (2) *N. tabacum* cv Gatersleben ('Gat'). (3) *N. tabacum* cv Samsun ('Sam') and (4) *N. tabacum* cv PBD6 ('PDB6'), both provided by the Institut du Tabac, ALTADIS, Bergerac, France.

Plants were cultured in a shaded greenhouse and fully expanded young leaves of the six genotypes were sampled to obtain leaf explants and mesophyll protoplasts. Leaf tissues were simultaneously sampled from similar plants for both procedures.

Protoplast-derived regenerants

Leaves were sterilized, wounded and placed in medium T0 (Bourgin *et al.*, 1979) containing cell wall hydrolases (Onozuka R10, Yakult Biochemicals Co. Ltd, Tokyo, Japan, 0.1%; Macerozyme R10, same origin, 0.02%; Driselase, Sigma, St Quentin Fallavier, France, 0.05%) for a standard overnight (16 h) enzymatic digestion. Protoplasts were rinsed from the digestion medium, cultured in medium T0 as described (Bourgin *et al.*, 1979) and subcultured in medium C as previously described (Muller *et al.*, 1983). Derived colonies were plated onto solidified medium R4M0 medium (Bourgin *et al.*, 1979), allowing callus formation and callus-derived shoot regeneration. Regenerated shoots were rooted on B medium (Bourgin *et al.*, 1979). Each regenerated plant was produced from a different callus, that is from a different protoplast.

Explant-derived regenerants

Leaves were sterilized, and wounded leaf pieces were placed onto solidified medium R4M0 medium (Bourgin *et al.*, 1979), allowing callus formation and callus-derived shoot regeneration at wound sites. Regenerated shoots were rooted on B medium (Bourgin *et al.*, 1979). Calli produced at different positions of the same leaf explant were separated at early stages. Each regenerated plant was obtained from a different callus.

SSAP procedure

Plant DNA was extracted from leaf material of 5–6 leaves *in vitro* plants, according to the Doyle and Doyle CTAB method (Doyle and Doyle, 1990). After an RNAse treatment, RNA-free genomic DNA was extracted with phenol:chloroform, salt-precipitated and re-suspended in TE (10 mM Tris, 0.1 mM EDTA). Genomic DNA (0.5 μg) was digested with 5 U *Eco*RI (Gibco BRL, Cergy-Pontoise, France) in a 25-μl volume of RL buffer (10 mM Tris-acetate pH 7.5, 10 mM magnesium acetate, 50 mM potassium acetate, 5 mM DTT) at 37°C overnight. The restricted DNA was diluted with one volume of ligation mix comprising 50 pmoles of *Eco*RI adaptors (Figure 1), 0.4 mM ATP, 1 U of *Eco*RI and 1 U of T4 DNA ligase (Pharmacia, Orsay, France) in RL buffer. Ligation was performed at 37°C for 3 h. Ligated DNA was finally diluted with 10 mM Tris–HCl, 0.1 mM EDTA pH 7.5 to the final concentration of 2.5 ng μL^{-1}. A 5-μl aliquot of the diluted ligation mix was amplified in a 25-μl reaction containing 35 ng of ^{33}P-labeled LTR13 primer (5′-CTTATACCTTGTCTGTGAAACC-3′, +286 to +265 of Tnt1–94 (Grandbastien *et al.*, 1989b), 50 ng of *Eco*RI primers (Figure 1), 0.24 mM of each dNTP, 1.5 mM $MgCl_2$ and 1.5 U of Taq DNA polymerase (Perkin Elmer, Courtaboeuf, France) in its reaction buffer. PCR was carried out using the following conditions: 5 min at 94°C, 13 cycles of 30 sec at 94°C, 30 sec at 65°C (– 0.7°C per cycle), 2 min at 72°C, 25 cycles of 30 sec at 94°C, 30 sec at 56°C and 2 min at 72°C, followed by a final extension step at 72°C for 10 min

After PCR, samples were diluted with one volume of loading dye (95% formamide, 0.05% xylene cyanol FF and 0.05% bromophenol blue), heat denatured at 94°C for 5 min and directly cooled on ice. For each sample, a 6–8 μl aliquot was loaded on a 6% denaturing polyacrylamide gel (20 × 40 × 0.04 cm) and pre-run at 35 W for 45 min. Samples were then run at 35 W for 4 h in 1xTBE. Gels were transferred to Whatman 3 MM paper and vacuum dried at 65°C for 1 h. Dried gels were exposed to X-ray films overnight or for 48 h, depending on the signal intensity.

Isolation and characterization of new SSAP bands

New SSAP bands that were clearly separated from non-polymorphic bands were selected, excised from the dried gels using a razor blade and re-suspended in 100 μl of water. DNA was eluted from the bands by boiling for 15 min, salt-precipitated and re-suspended in 10 μl of sterile water. A 5-μl-sample was then used as DNA template in PCR amplification using the SSAP conditions and primers described above. Amplification products were separated on 2% agarose gels. Bands with the expected molecular weight were excised from the agarose gels and cloned in pGEMT vector according to the manufacturer's recommendations

(Promega, Charbonnières, France). Sequencing was performed using the Abi prism automated sequencer. The nucleotide sequence data of these clones have been submitted to the DDBJ/EMBL/Genebank databases under the accession numbers: AF401683 to AF401739.

Five to six clones per SSAP band were usually analyzed. In some cases, minor contaminating clones were obtained in addition to a major consensus clone. Primers (sequence available upon request) were designed from the flanking genomic regions of consensus clones. Direct PCRs were performed using these primers and the LTR13 primer. In each case DNA from the regenerant, from which the new SSAP band was characterized, was tested against control DNA samples from other regenerants and/or from original genotypes. A new insertion was indicated when a PCR product of the expected molecular weight was obtained with DNA from the corresponding regenerant, but not with other control DNA samples. One regenerant was lost and the transpositional origin of two new SSAP bands could only be indirectly tested by negative PCR results on control DNA samples. For two bands isolated from protoplast-derived plants, only 2 bp of adjacent flanking sequences were found. However, these bands were assumed to be new insertions, since they each generated 5–6 identical clones. Most minor contaminating clones were also tested, and shown to correspond, as expected, to pre-existing Tnt1 insertions.

All sequenced copies were named as follow: the first letter(s) indicate(s) the genotype (g and gp = Gat; d and dp = D8; t = tl; *x* = D8xtl; s = sam; pb = PBD6); the following numbers (ranging from 1 to 23) identifie(s) the regenerant and the final number identifies the SSAP band. In the case of transposed copies, 'tr' was added to the name. In the case of pre-existing insertions (minor clones), a lower-case letter corresponding to each clone (ranging from 'a' to 'i') was added. For instance, the s231tr, s232tr and s233tr sequences correspond to three newly transposed copies, characterized from the same regenerant (numbered 23) obtained from leaf protoplasts of the cv. Samsun line, while the s231a sequence correspond to a pre-existing insertion isolated together with the s231tr new band.

Acknowledgements

We are very grateful to Josep M. Casacuberta and to Peter Hunt for critical reading of the manuscript, and to Dominique Higuet for helpful advice on statistical analyzes. This work was supported by the EEC BIOTECH program BIO4-CT96–0508 (TEBIODIV) as well as by the CNRS Génome ('Plasticité et Intégrité des Genomes') program.

References

Andrews, J.M., Oglesbee, M. and Lairmore, M.D. (1998). The effect of the cellular stress response on human T-lymphotropic virus type I envelope protein expression. *J. Gen. Virol.* **79**, 2905–2908.

Arnault, C. and Dufournel, I. (1994). Genome and stresses: reactions against aggressions, behavior of transposable elements. *Genetica*, **93**, 149–160.

Bailey, B.A., Dean, J.F.D. and Anderson, J.D. (1990). An ethylene biosynthesis-inducing endoxylanase elicits electrolyte leakage and necrosis in *Nicotiana tabacum* cv. Xanthi leaves. *Plant Physiol.* **94**, 1849–1854.

Bourgin, J.-P., Chupeau, Y. and Missonier, C. (1979). Plant regeneration from mesophyll protoplasts of several *Nicotiana* species. *Physiol. Plant.* **45**, 288–292.

Bourgin, J.-P. and Missonier, C. (1973). Vegetative propagation and cold preservation of haploid plants of *Nicotiana tabacum* and *Nicotiana paniculata*. *Haploid Inf. Serv.* **8**, 7.

Bradshaw, V.A. and McEntee, K. (1989). DNA damage activates transcription and transposition of yeast Ty retrotransposons. *Mol. Gen. Genet.* **218**, 465–474.

Casa, A.M., Brouwer, C., Nagel, A., Wang, L., Zhang, Q., Kresovich, S. and Wessler, S.R. (2000). The MITE family *Heartbreaker* (*Hbr*): molecular markers in maize. *Proc. Natl Acad. Sci. USA*, **97**, 10083–10089.

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

Casacuberta, J.M., Vernhettes, S. and Grandbastien, M.-A. (1995). Sequence variability within the tobacco retrotransposon Tnt1 population. *EMBO J.* **14**, 2670–2678.

Curcio, M.J. and Garfinkel, D.J. (1999). New lines of host defense: inhibition of Ty1 retrotransposition by Fus3p and NER/TFIIH. *Trends Genet*, **15**, 43–45.

Dellaporta, S.L., Chomet, P.S., Mottinger, J.P., Wood, J.A., Yu, S.M. and Hicks, J.B. (1984). Endogenous transposable element associated with virus infection in maize. *Cold Spring Harbor Symp Quant. Biol.* **49**, 321–328.

Deshayes, A. (1979). Cell frequencies of green somatic variations in the *tl* chlorophyll mutant of *Nicotiana tabacum* var. Samsun. *Theor. Appl. Genet.* **55**, 145–152.

Deshayes, A. (1976). Effets de cycles successifs de neoformation de bourgeons *in vitro* sur l'aptitude à la variation somatique chez un mutant chlorophyllien de *Nicotiana tabacum* var. Samsun. *Mutation Res.* **35**, 221–246.

Doyle, J.J. and Doyle, J.L. (1990). Isolation of plant DNA from fresh tissues. *Focus*, **12**, 13–15.

Ellis, T.H.N., Poyser, S.J., Knox, M.R., Vershinin, A.V. and Ambrose, M.J. (1998). Polymorphism of insertion sites of *Ty1-copia* class retrotransposons and its use for linkage and diversity analysis in pea. *Mol. Gen. Genet.* **260**, 9–19.

Grandbastien, M.-A. (1998). Activation of plant retrotransposons under stress conditions. *Trends Plant Sci.* **3**, 181–187.

Grandbastien, M.-A., Lucas, H., Mhiri, C., Morel, J.-B., Vernhettes, S. and Casacuberta, J.M. (1997). The expression of the tobacco Tnt1 retrotransposon is linked to the plant defense responses. *Genetica*, **100**, 241–252.

Grandbastien, M.-A., Missonier, C., Goujaud, J., Bourgin, J.-P., Deshayes, A. and Caboche, M. (1989a). Cellular genetic study of a somatic instability in a tobacco mutant: in vitro isolation of valine-resistant spontaneous mutations. *Theor. Appl. Genet.* **77**.

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

Grandbastien, M.-A., Spielmann, A., Pouteau, S., Huttner, E., Longuet, M., Kunert, K., Meyer, C., Rouzé, P. and Caboche, M. (1991). Characterization of mobile endogenous copia-like transposable elements in the genome of Solanaceae. In *Plant Molecular Biology 2*, (Herrmann, R.G. and Larkins, B., eds.) New York, USA: Plenum Press, pp. 333–343.

Gribbon, B.M., Pearce, S.R., Kalendar, R., Schulman, A.H., Paulin, L., Jack, P., Kumar, A. and Flavell, A.J. (1999). Phylogeny and transpositional activity of *Ty1-copia* group retrotransposons in cereal genomes. *Mol. Gen. Genet.* **261**, 883–891.

Heidmann, O. and Heidmann, T. (1991). Retrotransposition of a mouse IAP tagged with an indicator gene. *Cell*, **64**, 159–170.

Nucleic Acids Research, 1996, Vol. 24, No. 23 4825–4831

The *Saccharomyces* retrotransposon Ty5 influences the organization of chromosome ends

Sige Zou, Jin M. Kim and Daniel F. Voytas*

Department of Zoology and Genetics, 2208 Molecular Biology Building, Iowa State University, Ames, IA 50011, USA

Received June 11, 1996; Revised and Accepted October 15, 1996

DDBJ/EMBL/GenBank accession nos U19263, U19264

ABSTRACT

Retrotransposons are ubiquitous components of eukaryotic genomes suggesting that they have played a significant role in genome organization. In *Saccharomyces cerevisiae*, eight of 10 endogenous insertions of the Ty5 retrotransposon family are located within 15 kb of chromosome ends, and two are located near the subtelomeric *HMR* locus. This genomic organization is the consequence of targeted transposition, as 14 of 15 newly transposed Ty5 elements map to telomeric regions on 10 different chromosomes. Nine of these insertions are within 0.8 kb and three are within 1.5 kb of the autonomously replicating consensus sequence in the subtelomeric X repeat. This suggests that the X repeat plays an important role in directing Ty5 integration. Analysis of endogenous insertions from *S.cerevisiae* and its close relative *S.paradoxus* revealed that only one of 12 insertions has target site duplications, indicating that recombination occurs between elements. This is further supported by the observation that Ty5 insertions mark boundaries of sequence duplications and rearrangements in these species. These data suggest that transposable elements like Ty5 can shape the organization of chromosome ends through both transposition and recombination.

INTRODUCTION

Telomeres are specific protein–DNA structures found at the termini of eukaryotic chromosomes (1). Telomere sequences typically consist of tandem arrays of simple repeats synthesized by telomerase, a cellular reverse transcriptase (2). Most organisms have short and precise telomeric repeat sequences that are evolutionarily conserved. The *Saccharomyces cerevisiae* telomeric sequences, however, are atypically heterogeneous and consist of arrays of TG_{1-3} (3). A variety of middle repetitive sequences, called subtelomeric repeat sequences, are found associated with telomeres. These repeats are highly polymorphic and not well-conserved among eukaryotes.

In *S.cerevisiae*, subtelomeric repeat sequences have been studied in great detail (4–7). They are comprised of two major groups, called Y′ elements and X repeats. Y′ elements are immediately internal to the telomeres. There are two major classes that differ by size, called Y′-long (6.7 kb) and Y′-short (5.2 kb). The size differences are due to a series of small insertions and/or deletions. Y′ elements in the same class are highly conserved and typically share ~99% nucleotide identity. Y′ elements are found on most chromosomes and are highly polymorphic among different strains. At the end of any particular chromosome, Y′ elements are present in zero to four tandemly arranged copies.

X repeats are centromere-proximal when present in conjunction with Y′ elements (5). Junction sequences between X repeats and Y′ elements are normally short stretches of telomere sequences TG_{1-3}. On some chromosomes, such as chromosome III (chr III), no Y′ elements are present and X repeats are found immediately internal to the telomere. X repeats consist of a 473 bp core X sequence as well as varying numbers of short subtelomeric repeats (STR-A, STR-B, STR-C and STR-D) that range in size from 45 to 140 bp (5,6). Boundaries of X repeats vary dramatically due to the presence of STRs. The STRs are not present at all chromosome ends, while core X sequences are found in all but one subtelomeric region and share ~80% nucleotide identity.

The widespread and polymorphic distribution of X repeats and Y′ elements suggests that these subtelomeric sequences are in constant flux. Some Y′ elements have a large open reading frame (ORF) with weak homology to viral helicases, suggesting that Y′ elements may be related to transposable elements (4). However, transposition of Y′ elements has never been documented. On the other hand, Y′ recombination occurs frequently and depends on the *RAD52* gene, which is required for homologous recombination (4,8). Y′ elements preferentially recombine with members of the same size class, which results in a preponderance of one size class in any given strain. Recombination of Y′ elements can also result in exchange of sequences between ends of chromosomes. For example, gene conversion can replace sequences at one end of a chromosome with those from another end. Therefore, Y′ recombination can clearly reshape the organization of the chromosome ends. The presence of X repeats at the ends of almost all chromosomes suggests that they may also participate in homologous recombination. However, movement of X repeats by recombination has not been documented.

Ty5 is a retrotransposon identified from *S.cerevisiae* and its close relative *S.paradoxus* (9,10). The copy number and distribution of Ty5 insertions are polymorphic in different *Saccharomyces* species and strains. Characterization of eight endogenous Ty5 insertions in *S.cerevisiae* showed that two are associated with the silent mating locus *HMR* and six are located in subtelomeric

*To whom correspondence should be addressed. Tel: +1 515 294 1963; Fax: +1 515 294 0345; Email: voytas@iastate.edu

regions. None of these elements are transposition competent. In *S.paradoxus*, two of three characterized Ty5 insertions are likely subtelomeric, based on their association with a subtelomeric X repeat ( ). A Ty5 transposition assay was developed in *S.cerevisiae* using an *S.paradoxus* element ( ). Of 19 newly transposed Ty5 elements on chr III, four were inserted near the left telomere and 14 were inserted near transcriptional silencers at the *HMR* and *HML* loci. These regions are bound in a unique type of chromatin, called silent chromatin, which represses the transcription of adjacent genes ( ). The target bias for Ty5 suggests that this element recognizes silent chromatin during integration.

Transposable elements have been found associated with telomeres in other organisms. In *Drosophila melanogaster*, for example, two families of non-LTR retrotransposons have been identified that serve as telomeres (,). The telomeric location of Ty5 suggests that it may play a role in genome organization, and in particular, contribute to the dynamic nature of chromosome ends through transposition and/or recombination. To test this, we characterized the genomic location of additional Ty5 transposition events. We also used the complete nucleotide sequence of the *S.cerevisiae* genome to compare the organization of endogenous elements between *S.cerevisiae* and *S.paradoxus*.

MATERIALS AND METHODS

Strains

The following yeast strains were used in this study: *S.paradoxus* NRRL Y-17217 (Northern Regional Research Laboratory); wild-type *S.cerevisiae* strains SK1 and S288C for the characterization of endogenous Ty5 insertions (J.D. Boeke, Johns Hopkins University); *S.cerevisiae* W303-1A (*MAT***a** *ade2-1 can1-100 his3-11 leu2-3 trp1-1 ura3-1*) for the Ty5 transposition assay (A. Myers, Iowa State University). The *E.coli* strain XL1-Blue (Stratagene) was used for recombinant DNA manipulations. Transformation of *E.coli* and yeast strains was performed by electroporation as described ( ).

Mapping newly transposed Ty5 insertions

Sequences flanking newly transposed Ty5 elements were amplified by inverse PCR method as described ( ). Briefly, ~100 ng of genomic DNA was digested with *Msp*I and self-ligated in a 50 µl ligation mixture. Sequences flanking Ty5 insertions were amplified from 2 µl of the ligation mixture with Ty5-6p LTR-specific oligonucleotides DVO219 (5′-TACTGTCGGATCGGAGGT-TT-3′) and DVO220 (5′-CTGTGTACAAGAGTAGTACC-3′). PCR products were sequenced with oligo- nucleotides DVO214 (5′-CCCTCGAGCATTTACATAACATATAGAAAG-3′) or DVO243 (5′-CCTTGTCTAAAACATTACTG-3′). Ty5 integration sites were determined by comparing these sequences to the *S.cerevisiae* genome database.

DNA manipulations and analysis

Yeast genomic DNA and chromosomes were prepared as described ( ). The genomic DNA was digested with restriction enzymes and separated by agarose gel electrophoresis. Yeast chromosomes were separated by pulsed-field gel electrophoresis, and chromosome identity determined by their mobility ( ). Gels containing genomic DNA or chromosomes were transferred to nylon membranes by alkaline transfer. Filters were hybridized with DNA fragments that had been radio-labeled by random-priming (Promega). Hybridization probes included Ty5 internal sequences (probe A and B in Fig. 3A), the long terminal repeat (LTR) (Fig. 3A), as well as sequences flanking Ty5 insertions. The LTR was amplified from Ty5-5p with oligonucleotides DVO182 (5′-GGGTAATGTTTC-AGT-3′) and DVO116 (5′-TAGTAAGTTTATTGGACC-3′). Sequence flanking the 5′-end of Ty5-12p element was amplified with DVO200 (5′-CATTACCCATATCATGCT-3′) and the reverse primer, which is complementary to the vector. DNA sequences were determined with the fmol sequencing kit (Promega), or by the Nucleic Acid Facility of Iowa State University. Sequence analysis was performed using the GCG computer programs ( ). LTR sequences were identified from the complete nucleotide sequence of *S.cerevisiae* using the program BLAST ( ). Sequences were considered that had >65% nucleotide identity to the Ty5-1 left LTR.

RESULTS

Ty5 preferentially transposes to subtelomeric regions in *S.cerevisiae*

We have previously characterized 19 newly transposed Ty5 elements on chr III. Of these, 18 occurred in regions of silent chromatin, including 14 at the *HMR* and *HML* mating loci and four at the left telomere ( ). We hypothesized that this integration pattern is due to a mechanism that targets Ty5 to silent chromatin. For chromosomes other than chr III, the only known regions of silent chromatin are at the telomeres. We wanted to investigate whether Ty5 insertions on other chromosomes are near the telomeres, which would support a general role for Ty5 in the genomic organization of chromosome ends.

Fifteen strains were randomly chosen from a collection of strains with newly transposed Ty5 elements ( ). Sequences flanking these elements were amplified by inverse PCR and used directly for DNA sequencing. These insertions were found to reside on ten different chromosomes (Fig. 1). One insertion, W3, is on chr XI 152 kb from the end of the chromosome. The remaining 14 insertions are all subtelomeric and are within 15 kb of chromosome ends. The insertions show no orientation specificity with respect to the ends of the chromosomes; eight insertions are in the same 5′ to 3′ orientation as the chromosome sequence, and seven are in opposite orientation.

Ty5 preferentially inserts near the X repeat

We have previously used the autonomously replicating consensus sequences (ACS) in the X repeat as a reference point for Ty5 insertions. In this study, nine of the 15 Ty5 insertions are within 0.8 kb on either side of the ACS in the X repeat; three additional insertions are within 1.5 kb. Eight of the 11 telomeres that have Ty5 insertions also have Y′ elements. These Y′ elements separate the X repeat from the telomere by >5 kb. In all eight cases, Ty5 insertions are clustered within 1.5 kb of the ACS in the X repeat and are consequently several kilobases from the TG_{1-3} telomeric repeats. This suggests that the X repeat serves as a nucleation site for factors important for Ty5 targeting.

Features of sequences flanking endogenous Ty5 insertions in the subtelomeric regions of *S.cerevisiae*

The clear preference for Ty5 to integrate near the telomeres indicates an active role for Ty5 in shaping the organization of chromosome ends. We took advantage of the complete genome

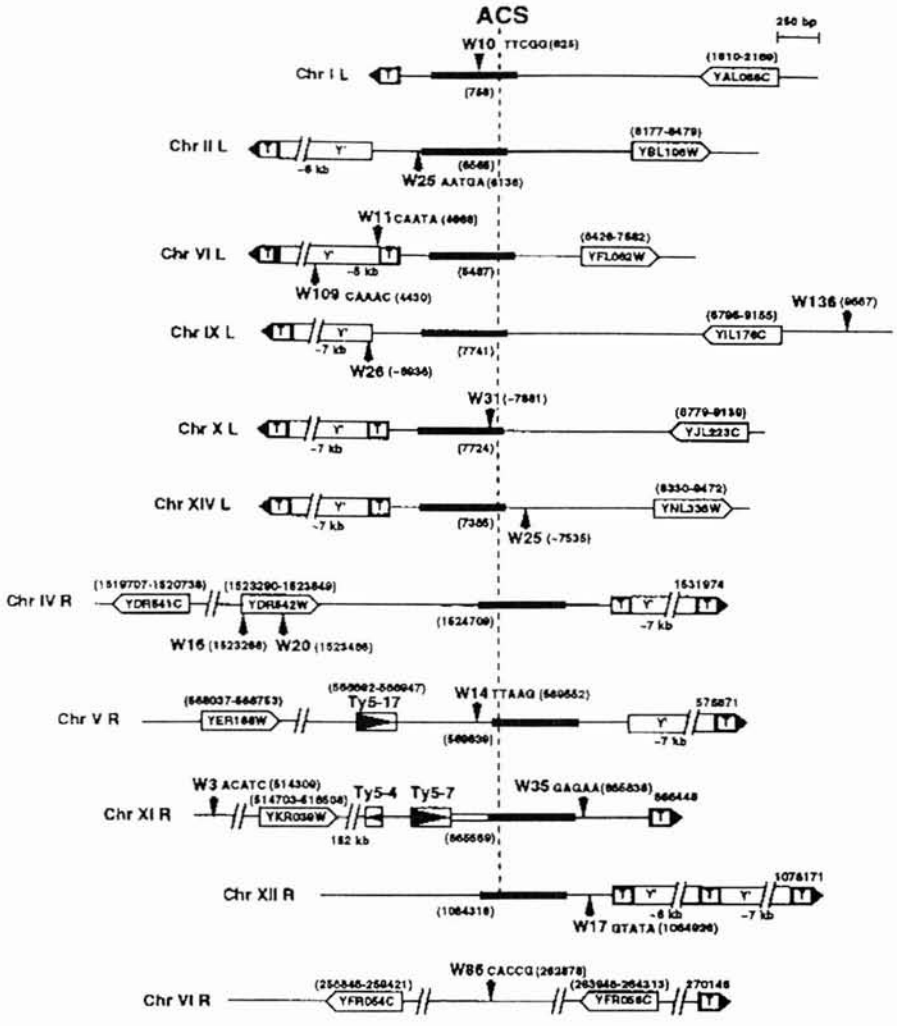


Figure 1. Location of *de novo* Ty5 transposition events. Left and right arms of chromosomes are labeled as L and R, respectively. T indicates telomeric repeat sequences, and the narrow boxes indicate X repeats, which are aligned at the ACS (the autonomously replicating consensus sequences). The sizes of Y elements are provided. Open boxes with arrowheads depict endogenous Ty5 LTRs, and open boxes with labels indicate open reading frames. The Ty5 insertions are labeled as W followed by a number, which refers to the strain from which they were isolated. Arrows pointing down indicate insertions in the same 5′ to 3′ orientation as the chromosome sequences. Arrows pointing up represent insertions in the opposite orientation. Numbers in parentheses designate base positions of open reading frames and Ty5 insertion sites. The base position of the ACS is provided adjacent to the dashed line, which marks the position of this sequence.

sequence of *S.cerevisiae* to evaluate whether endogenous Ty5 insertions have played a passive role in genome organization through recombination. We identified all Ty5 insertions in S288C, the strain used for the yeast genome project (Fig. 2A). Ten Ty5 insertions were found, including eight previously identified on chr III, VII, VIII and XI (11). The chr VII insertion (designated Ty5-15) had previously been characterized only by Southern hybridization analysis (18). Two new insertions were identified on chr V, designated Ty5-16 and Ty5-17. The chr V insertions are near the right telomere but are in opposite orientation. Ty5-17 is within 600 bp of a X ACS, and Ty5-16 is within 2.7 kb.

Target site sequences were characterized for eight Ty5 insertions with full-sized LTRs. None of these insertions have the perfect 5 bp target site duplications characteristic of newly transposed Ty5 elements (11), although Ty5-16 has flanking target sequences with four identical nucleotides out of five (Table 1). The 5′ target site of Ty5-17 is the same as the 3′ target site of Ty5-16. However, the 3′ target site of Ty5-17 is different from the 5′ target sites of these two insertions, suggesting that a gene conversion event, or two sequential reciprocal recombination events, occurred between these elements.

Some Ty5 elements mark boundaries of duplicated sequences in the *S.cerevisiae* genome. Genome sequencing efforts have identified extensive duplications between the telomeric regions of chr III and XI (19–21). Four Ty5 insertions are present in these duplicated regions (Fig. 2B). The chromosome ends, including the X repeat, are similar between the chr III left telomere and both telomeres of chr XI (region **a**). The similarity ends at the Ty5-1 insertion on chr III. Downstream of region **a**, the chr XI left-end has a unique 50 bases and both chr XI ends share a second duplicated sequence (region **b**). For the chr XI left-end, the **b** region terminates in sequences that have been duplicated from the

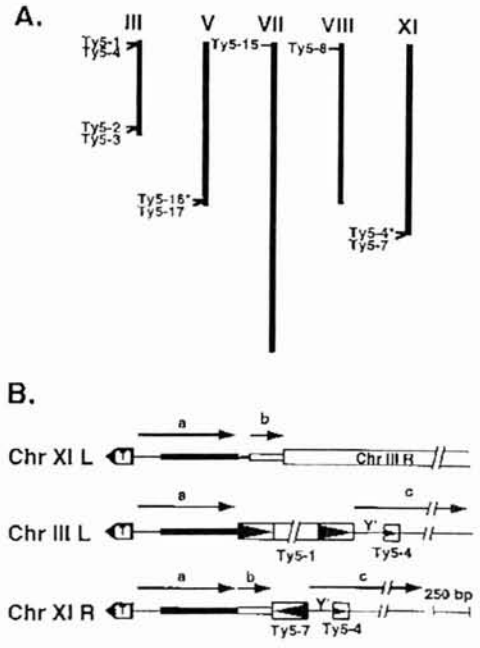


Figure 2. *Saccharomyces cerevisiae* Ty5 elements. (**A**) Chromosome location of endogenous Ty5 elements. Chromosomes are drawn to scale with the left end on top. The asterisk reflects elements on the Crick strand of the chromosome sequences. Base positions for insertions are: Ty5-1, 1172–4314; Ty5-2, 290646–290891; Ty5-3, 291015–291252; Ty5-4 on chr III, 4471–4572; Ty5-4 on chr XI, 664909–664808; Ty5-7, 665062–665300; Ty5-8, 7993–8224; Ty5-15, 863–1079; Ty5-16, 562209–562459; Ty5-17, 564300–564533. (**B**) Sequence rearrangements between chr III and chr XI of *S.cerevisiae*. Symbols are as in Figure 1. Duplicated sequence domains are indicated by arrows and designated as a, b or c. The Y′ element sequence in region c is labeled. The open box labeled chr III R represents an additional sequence duplication between chr XI L and chr III R.

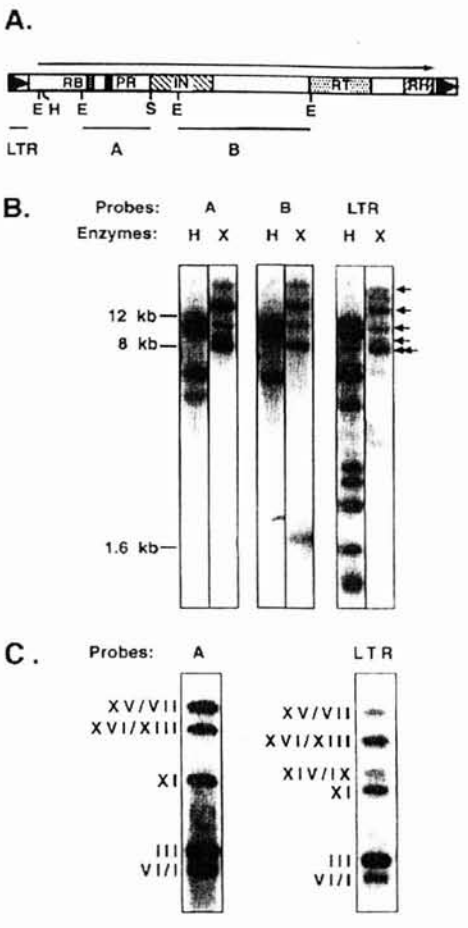


Figure 3. Copy number of Ty5 elements in strain NRRL Y-17217. (**A**) Genomic organization of Ty5-6p. Open boxes with arrows indicate the LTRs. Boxes within the internal domain depict conserved amino acid sequence domains: RB, RNA binding domain; PR, protease; IN, integrase; RT, reverse transcriptase; RH, RNase H. The arrow over the element indicates the open reading frame. E, H and S denote restriction endonuclease sites for *Eco*RI, *Hin*dIII and *Sma*I. The lines under the element depict probes used for hybridization analyses. (**B**) Southern hybridization analyses of strain NRRL Y-17217. DNA was digested with *Hin*dIII (H) or *Xho*I (X). Molecular length markers are indicated, and arrows denote *Xho*I restriction fragments that hybridize to probe A. Arrowheads indicate bands that hybridize to both internal and LTR probes. (**C**) Chromosome distribution of Ty5 in NRRL Y-17217. Filters were prepared from pulsed-field gels and hybridized with probe A and the LTR probe. Chromosome designations for hybridizing chromosomes are indicated.

right-end of chr III. The right-end of chr XI, however, has a Ty5 insertion at the end of the **b** region (Ty5-7). This insertion has different target sites from Ty5-1 and is in the opposite orientation, clearly indicating that they are different insertions. Centromere proximal to Ty5-1 and Ty5-7 are several kb of duplicated sequences (region **c**), including a Ty5 insertion (Ty5-4). It has previously been noted that the beginning of the **c** region contains 140 bp of a Y′ element (22). The location of Ty5-1 and Ty5-7 at the boundaries of rearrangements suggests that these elements have played a role in these events.

Distribution of Ty5 insertions in *S.paradoxus*

Characterization of Ty5 elements showed that transposition-competent insertions are not present in *S.cerevisiae*, but are present in its close relative *S.paradoxus*. The close relationship between these species suggests that they may be a good model further understanding the role of Ty5 in genome organization. We focused our characterization efforts on the *S.paradoxus* strain NRRL Y-17217, which harbors the most Ty5 insertions, at least one of which is transpositionally active (10,11). We first estimated copy number in this strain by Southern hybridization analysis using restriction enzymes that do not cut within the element or cut only once. Filters were hybridized with probes specific to either Ty5 internal sequences or Ty5 LTRs (Fig. 3). An example of this analysis is presented in Figure 3B, using *Hin*dIII (one internal site) and *Xho*I (no cut sites). Based on this analysis, at least six restriction fragments hybridized to both internal and LTR sequences in NRRL Y-17217. In the *Hin*dIII lanes, there are at least seven more restriction fragments that hybridized to the LTR. Some of these extra LTRs may be solo LTRs, which are derived from recombination between LTRs of full-length Ty5 elements.

The chromosomes of NRRL Y-17217 were separated by pulsed-field gel electrophoresis and transferred to nylon filters. The chromosome location of Ty5 insertions was analyzed by hybridizing the filter with either Ty5 internal or LTR sequences (Fig. 3C). Ty5 elements with internal sequences were located on at least five chromosomes, namely chr VI or I, III, XI, XVI or XIII and XIV or VII. Ty5 LTRs were located on these same chromosomes as well as chr XIV or IX.

Table 1. Sequences of Ty5 target sites

Species	Insertion (chr)	Left target sequence	Right target sequence	Target nucleotide identity	LTR nucleotide identity (%)[a]
S.cerevisiae	Ty5-1 (III)	TTTCA	TATCC	3/5	86, 89
	Ty5-2 (III)	TTCCT	TAAAA	1/5	73
	Ty5-3 (III)	ATCGC	TTTGC	3/5	55
	Ty5-7 (XI)	CGTGG	TACCG	1/5	82
	Ty5-8 (VIII)	GTATA	ATATG	3/5	77
	Ty5-15 (VII)	TTTCA	CCCAA	1/5	81
	Ty5-16 (V)	GTTAT	GTTCT	4/5	91
	Ty5-17 (V)	GTTCT	TTACA	2/5	73
S.paradoxus	Ty5-6p (XI)	TCGTA	TCGTA	5/5	100
	Ty5-5p (III/XI)	TGTCA	CTATC	0/5	100, 98
	Ty5-10p (III/XI)	AGTAT	TATAA	2/5	98
	Ty5-12p (III/XI)	AGTAT	TTTTC	1/5	97
	Ty5-14p (ND)	-	TGTCA	-	98

[a]Values are derived from comparisons with the Ty5-6p right LTR. Multiple values refer to comparisons with the left and right LTRs, respectively.

Duplication and rearrangement of sequences flanking Ty5 insertions in *S.paradoxus*

We previously isolated five of the approximately 13 Ty5 insertions present in *S.paradoxus* strain NRRL Y-17217 (10). To investigate the relationship between Ty5 and genome organization, sequences flanking these insertions were determined. The target sites of several insertions were analyzed, and only the Ty5-6p insertion was found to have target site duplications (Table 1). It is interesting to point out that the 5 bp target sequence at the 3′-end of Ty5-14p is the same as the 5′ target sequence of Ty5-5p. This suggests that Ty5-5p and Ty5-14p recombined and exchanged targets, and Ty5-14p subsequently suffered a deletion of its 5′ region (Fig. 4). Evidence for such a reciprocal translocation is also supported by sequence differences among the LTRs of these elements. The 3′ LTR of Ty5-5p has four bp that differ from the 5′ LTR. The Ty5-5p 3′ LTR, however, is identical at these four nucleotide positions to the 5′ LTR of Ty5-14p, arguing strongly that a recombination event had occurred between these elements.

Flanking sequences of several *S.paradoxus* insertions were compared to the *S.cerevisiae* genome database or used in Southern hybridization analysis. As previously reported, 5′ and 3′ sequences flanking Ty5-6p share ~90% nucleotide identity to sequences on *S.cerevisiae* chr XI (10). The 5′ flanking sequence hybridized to chr XI of *S.paradoxus* as well as *S.cerevisiae* (Fig. 5A), indicating that Ty5-6p is located on *S.paradoxus* chr XI and its flanking sequences are conserved between the two species. No evidence for a Ty5 insertion, however, was found at the corresponding region on *S.cerevisiae* chr XI, suggesting Ty5-6p transposed to this site after species divergence.

Analysis of flanking sequences of some *S.paradoxus* insertions support a role for Ty5 in genome rearrangements. Several Ty5 elements were flanked by sequences unique to *S.paradoxus*. For example, the 5′ flanking sequence of Ty5-5p shows no significant homology to any *S.cerevisiae* sequences, while the 3′ flanking sequence shares high homology with the subtelomeric X repeat. Southern analysis indicated that the unique 5′ flanking sequence hybridizes to *S.paradoxus* chr III and XI, suggesting that this

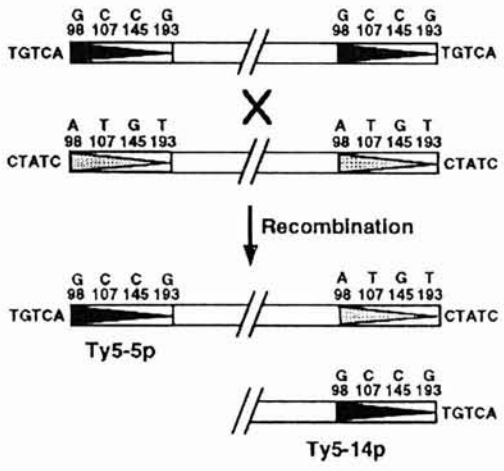


Figure 4. Proposed model for the reciprocal translocation that generated Ty5-5p and Ty5-14p. Two parental elements are shown with LTRs that differ by black or speckled arrowheads. Nucleotide differences between these LTRs are shown, with numbers indicating base positions. Sequences of target sites on either side of the elements are provided.

sequence is duplicated between these chromosomes (Fig. 5B). Since Ty5-5p is associated with a subtelomeric X repeat, the duplication may have occurred between the ends of chr III and XI. Sequence analysis of Ty5-14p indicated that it has a 5′ deletion, which includes the 5′ LTR. The 3′ flanking sequence has no significant similarity to any *S.cerevisiae* sequences.

Flanking sequences of some elements suggest that Ty5 insertions mark sites that have been rearranged between *S.paradoxus* and *S.cerevisiae*. For example, the 5′ sequence of Ty5-12p hybridized to *S.cerevisiae* chr V (Fig. 5C). This sequence, however, hybridized to *S.paradoxus* chr III and XI, indicating that it has been duplicated and rearranged between these species. Consistent with the hybridization analysis, the 5′ flanking sequence shares 90% nucleotide identity with a subtelomeric

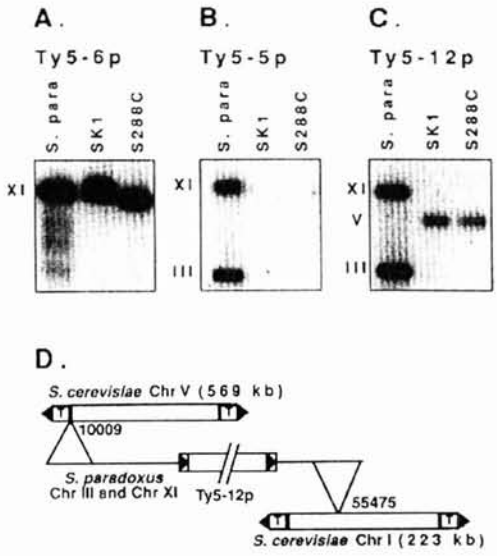


Figure 5. Organization of sequences flanking *S.paradoxus* Ty5 insertions in *S.paradoxus* and *S.cerevisiae*. S. para designates *S.paradoxus* and SK1 and S288C are wild-type strains of *S.cerevisiae*. Filters prepared from pulsed-field gels were hybridized with 5′ flanking sequences of Ty5-6p (**A**), Ty5-5p (**B**) and Ty5-12p (**C**). (**D**) Rearrangements of sequences flanking Ty5-12p. The 5′ flanking sequence of Ty5-12p is located on chr III and chr XI of *S.paradoxus*, but is located on chr V of *S.cerevisiae*. The number denotes the base position corresponding to the *S.cerevisiae* chr V nucleotide sequence. The 3′ flanking sequence shows homology to the middle of *S.cerevisiae* chr I, with the number indicating the base position within the chromosome.

region of *S.cerevisiae* chr V (Fig. 5D). The 3′ sequence of Ty5-12p shares 90% identity to the middle of the left arm of *S.cerevisiae* chr I, providing evidence for additional rearrangements. Part of the flanking sequences of insertion Ty5-11p were also determined. The 5′ sequence shares 84% identity to the subtelomeric region of *S.cerevisiae* chr XVI, but the 3′ sequence shares 84% identity to the subtelomeric region of *S.cerevisiae* chr XIII. Although the chromosome location of this element was not determined by hybridization analysis, these results indicate a sequence rearrangement between *S.paradoxus* and *S.cerevisiae* and implicate a role for Ty5 elements in genome rearrangements.

DISCUSSION

Ty5 is a subtelomeric repeat

From the available *S.cerevisiae* genomic sequence, we have identified 10 endogenous Ty5 insertions. Of these, seven are located within 10 kb of the ends of chromosomes, and one is within 15 kb. Two insertions are near the *HMR* locus. If *HMR* (~20 kb from the end of chr III) is considered a subtelomeric region, then all the endogenous Ty5 insertions are in the vicinity of chromosome ends. In addition, characterization of Ty5 in different *Saccharomyces* species and strains indicated that Ty5 distribution is polymorphic and varies in copy number and chromosome location (10). We have previously mapped the locations of 19 newly transposed Ty5 insertions on chr III (11). One insertion was in the middle of the right arm of chr III; 14 were at the silent mating loci, *HML* and *HMR*; four were within 2 kb of the left telomere. Here we mapped the locations of 15 newly transposed Ty5 insertions on 10 chromosomes other than chr III, which are not known to have silent chromatin except at the telomeres. Fourteen of fifteen are located within 15 kb from the ends of these chromosomes. These results indicate that Ty5 is a subtelomeric repetitive element based on its location and polymorphic nature.

At the subtelomeric regions of *S.cerevisiae*, the X repeats and Y′ elements are the two most abundant repetitive sequences ( ). Y′ elements are immediately adjacent to the telomere sequences and are found at the subtelomeric regions of most but not all chromosomes (1). Internal to Y′ elements are X repeats, which are found at the ends of all but one chromosome. Chromosome ends, therefore, have a relatively rigid organization; internal to the telomere sequences are Y′ elements, followed by X repeats. In contrast, the location of Ty5 is very flexible. Insertions can be found within the telomere sequences, between X repeats and the telomere sequences or centromere-proximal to X repeats or Y′ elements. Unlike Y′ elements, Ty5 insertions can occur in either orientation with respect to the chromosome end. Ty5 transposition, therefore, has a regional specificity as opposed to a site specificity.

A potential role for the X repeat in directing Ty5 transposition

In our previous study of Ty5 insertions on chr III, 14 of 19 elements were clustered near the E and I transcriptional silencers that flank *HML* and *HMR*. We have recently shown that the assembly of silent chromatin mediated by these silencers is critical for Ty5 targeting (S. Zou and D. F. Voytas, unpublished). Of the 14 telomeric insertions identified in this study, and the four telomeric insertions on chr III, 12 are located within 0.8 kb of the X repeats. Particularly notable are insertions on chromosome ends with Y′ elements, which are all >5 kb from the chromosome ends. The TG_{1-3} telomeric sequences can assemble silent chromatin in the absence of X repeats ( ). However, the clustering of Ty5 insertions near X sequences suggests that some unique feature of the X repeat directs Ty5 integration. The X repeats have binding sites for the origin recognition complex and the transcription factor ABF1. Future experiments will test the role of these binding sites in directing Ty5 transposition.

Transposable elements and the origin of subtelomeric repeats

Telomere repeat sequences are generated by reverse transcription, which is carried out by telomerase (2). Telomerase is the only known reverse transcriptase that is not associated with retroelements, and may have originated from a retrotransposon or a retrovirus. The Y′ elements have some features of transposable elements; however, transposition of Y′ elements has never been demonstrated (1). Although X repeats are conserved among *Saccharomyces* species, their origin is largely unknown (8). In contrast to the other repeat sequences, the Ty5 elements are typical LTR retrotransposons and actively transpose to subtelomeric regions. This provides direct evidence that subtelomeric repeats can originate from transposable elements. The link between transposable elements and telomeres is further substantiated by the observation that the HeT and TART transposable elements of *Drosophila melanogaster* serve as telomeres (13,14). Transposable elements, therefore, may generally contribute to the structure of chromosome ends.

Ty5 and the organization of subtelomeric regions

The Ty5 elements are unique among yeast subtelomeric repeats in that they can shape chromosome ends actively through transposition and passively through recombination. The copy number of Ty5 varies extensively in strains of both *S.cerevisiae* and *S.paradoxus* ( ). Although functional Ty5 elements do not exist in *S.cerevisiae*, they still influence genome organization through recombination, similar to the Y′ elements. In *S.paradoxus*, however, there are transposition-competent elements that can actively participate in restructuring chromosomes. The close relationship between the two yeast species makes them ideal models to more precisely evaluate the extent to which Ty5 has influenced genome organization, especially in light of the complete genome sequence of *S.cerevisiae*.

Analysis of *de novo* transposition events clearly demonstrated that Ty5 generates 5 bp target site duplications ( ). Characterization of endogenous insertions, however, showed that only one is flanked by such duplications. There are two possibilities to explain this phenomenon. First, the absence of target site duplications may be due to random mutation. We reason that if this is the case, the LTR sequences among different Ty5 insertions should be degenerate to a similar extent as the target sites. The LTRs of Ty5-5p and Ty5-12p share >98% nucleotide identity with the transpositionally functional Ty5-6p LTR, suggesting that these insertions are not ancient and their target sites should not have mutated dramatically. Nonetheless, the target sites of Ty5-5p share no similarity and the target sites of Ty5-12p have only one nucleotide in common. It is difficult to argue that mutation alone could be responsible for the extreme differences in target site sequences.

A second possibility is that recombination between elements resulted in the lack of target site duplications. There are several examples that directly support this model. The 5 bp at the 5′ target site of Ty5-5p are the same as those at the 3′ target site of Ty5-14p, suggesting these two insertions recombined, resulting in the exchange of target sites. We can not tell whether the 3′ target site of Ty5-5p is the same as the 5′ target site of Ty5-14p, because Ty5-14p has suffered a deletion of its 5′ LTR. Additional support, however, is provided by four nucleotide differences between the LTRs of Ty5-5p. These four nucleotides in the Ty5-5p 5′ LTR are shared with the 3′ LTR of Ty5-14p, suggesting that these LTRs originated from the same element. Evidence for recombination is also found among the *S.cerevisiae* elements. The 5′ target sequence of Ty5-17 shares only two nucleotides with its 3′ target sequence, but is the same as the 3′ target sequence of Ty5-16; the target sites of Ty5-16 differ by only one nucleotide. A possible explanation for this observation is that gene conversion occurred between Ty5-16 and Ty5-17. In this process, the 5′ target site of Ty5-17 was replaced by the target sequence of Ty5-16, but the other site of Ty5-17 and both target sites of Ty5-16 remained unchanged. The 5′ target site of Ty5-16 subsequently mutated, resulting in one nucleotide difference. Experiments need to be conducted to test whether subtelomeric Ty5 elements can recombine among themselves and exchange their newly acquired target sites.

Recombination between repetitive sequences has likely played an important role in restructuring chromosomes. We have obtained evidence that Ty5 has been involved in recombination events, and these events have reorganized chromosomes in *S.cerevisiae* and *S.paradoxus*. For example, the 5′ flanking sequence of Ty5-12p is located on chr V of *S.cerevisiae* but is duplicated on chr III and chr XI in *S.paradoxus*. The 5′ flanking sequence of Ty5-5p is duplicated between chr III and chr XI in *S.paradoxus*, but is completely absent from *S.cerevisiae*. Similarly, duplicated sequences between chr III and XI in *S.cerevisiae* have boundaries that are marked by Ty5 insertions. Recombination between some subtelomeric repeats, such as Y′ elements, has been well characterized ( ). Taken together, these observations clearly support the role of repetitive sequences, including transposable elements, in influencing the organization of chromosome ends. Further characterization of the genomes of closely related species such as *S.paradoxus* and *S.cerevisiae* will likely offer additional perspective on the extent to which transposable elements have shaped chromosome architecture.

ACKNOWLEDGEMENTS

We would like to thank M. Cummings for critical reading of the manuscript and David Wright for technical assistance. This work was supported by a grant from the American Cancer Society and an American Cancer Society Junior Faculty Research Award to D.F.V. This is journal paper no. J-16941 of the Iowa Agriculture and Home Economics Experiment Station (Ames); Project no. 3120.

REFERENCES

1 Zakian,V.A. (1995) *Science*, **270**, 1601–1607.
2 Blackburn,E. (1992) *Annu. Rev. Biochem.*, **61**, 113–129.
3 Shampay,J., Szostak,J. and Blackburn,E. (1984) *Nature*, **310**, 154–157.
4 Louis,E.J. and Haber,J.E. (1992) *Genetics*, **131**, 559–574.
5 Louis,E.J., Naumova,E.S., Lee,A., Naumov,G. and Haber,J.E. (1994) *Genetics*, **136**, 789–802.
6 Fiona,E.P., Tonie,C.H. and Louis,E.J. (1995) *Yeast*, **11**, 371–382.
7 Chan,C.S.M. and Tye,G.-K. (1983) *Cell*, **33**, 563–573.
8 Lundblad,V. and Blackburn,E.H. (1993) *Cell*, **73**, 347–360.
9 Voytas,D.F. and Boeke,J.D. (1992) *Nature*, **358**, 717.
10 Zou,S., Wright,D.A. and Voytas,D.F. (1995) *Proc. Natl. Acad. Sci. USA*, **92**, 920–924.
11 Zou,S., Ke,N., Kim,J.M. and Voytas,D.F. (1996) *Genes Dev.*, **10**, 634–645.
12 Laurenson,P. and Rine,J. (1992) *Microbiol. Rev.*, **56**, 543–560.
13 Biessmann,H., Mason,J.M., Ferry,K., d'Hulst,M., Valgeirsdottir,K., Traverse,K.L. and Pardue,M.L. (1990) *Cell*, **61**, 663–673.
14 Levis,R.W., Ganesan,R., Houtchens,K., Tolar,L.A. and Sheen,F.M. (1993) *Cell*, **75**, 1083–1093.
15 Ausubel,F.M., Brent,R., Kingston,R.E., Moore,D.D., Seidman,J.G., Smith,J.A. and Struhl,K. (1987) *Current Protocols in Molecular Biology*. Greene/Wiley Interscience, New York.
16 Ochman,H., Gerber,A.S. and Hartl,D.L. (1988) *Genetics*, **120**, 621–623.
17 Devereux,J., Haeberli,P. and Smithies,O. (1984) *Nucleic Acids Res.*, **12**, 387–395.
18 Aitschul,S., Gish,W., Miller,W., Myers,E. and Lipman,D. (1990) *J. Mol. Biol.*, **215**, 403–410.
19 Gromadka,R., Gora,M., Zielenkiewicz,U., Slonimski,P. and Rytka,J. (1996) *Yeast*, **12**, 583–591.
20 Oliver,S.G., van der Aart,Q.J.M., Agostoni-Carbone,M.L., Aigle,M., Alberghina,L., Alexandraki,D., Antoine,G., Anwar,R., *et al.* (1992) *Nature*, **357**, 38–46.
21 Dujon,B., Alexandraki,D., Andre,B., Ansorge,W., Baladron,V., Ballesta,J.P., Banrevi,A., Bolle,P.A., Bolotin-Fukuhara,M., Bossier,P. *et al.* (1994) *Nature*, **369**, 371–378.
22 Louis,E. (1995) *Yeast*, **11**, 1553–1573.
23 Gottschling,D.E., Aparicio,O.M., Billington,B.L. and Zakian,V.A. (1990) *Cell*, **63**, 751–762.

Final Reference	**W01-4**
Presenting author	Paula Araujo
E-mail	pgaraujo@yahoo.com magda1708@yahoo.com
Presentation type	Oral
Session	W01 - Control of transposition and its impact on plant genes and genomes

EXPRESSION PATTERN OF SUGARCANE TRANSPOSABLE ELEMENTS

Paula Araujo[1], Magdalena Rossi[1], Erika Jesus[1], Viviane Dias[1], Juliana Felix[2], Rodrigo Drummond[2], Marcelo Menossi[2], Marie-Anne Van Sluys[1]
1-Departamento de Botânica, IB-USP, Rua do Matão 277, São Paulo, CEP 055080-090, SP, Brasil
2-Centro de Biologia Molecular e Engenheria Genética,Universidade Estadual de Campinas, C.P.6010, CEP 13083-970, Campinas, SP, Brasil.

TEs are genetic units capable of movement within the genome. Present in all living organisms, these elements are mutagenic agents and their activity produces structural changes in single genes or overall genome followed by altered spatial and temporal patterns of gene expression and function. Out of the 261,609 EST sequences, produced by the Brazilian Sugarcane Expressed Sequence Tag Project (SUCEST), 220 clones highly homologous to previously reported TEs were chose for expression analysis. After full length cDNA sequencing, TE clones were assigned a family according to the best sequence alignment against a fully characterized element using BLASTX program.

Electronic Northern Analysis allowed us to visualize the frequency distribution of the 220 clones across the libraries. TE clones were mostly identified in callus, meristem tissue, flower and pathogen infected tissue.

Macro-array expression experiments with flower, leaf roll, apical meristem and calli cDNA as probes, revealed clearly that tissue culture induced drastically TEs expression in agreement with the electronic northern result. To validate macro-arrays data Northern blot experiments were carried out for some of the most representative clones.

Final Reference	**W01-12**
Presenting author	Paula Araujo
E-mail	appcosta@yahoo.com
Presentation type	Poster
Session	W01 - Control of transposition and its impact on plant genes and genomes

ASSESSMENT OF SUGARCANE GENETIC DIVERSITY USING RETROTRANSPOSONS

Ana Paula Pimentel Costa, Magdalena Rossi, Paula Araujo, Marie-Anne Van Sluys
Departamento de Botânica, IB-USP, Rua do Matão 277, São Paulo, CEP 055080-090, SP, Brasil

The Brazilian Sugarcane Expressed Sequence Tag Project (SUCEST) produced important data for characterization of the structure and organization of sugarcane genome. These data showed an abundant occurrence of transposable elements. Retrotransposons transpose via an RNA intermediate that is reverse transcribed before integration into a new location within host genome. The unique properties of retrotransposons, as their stable inheritance, have been exploited as genetic tools for plant genome analysis. Major applications are in determining phylogenic and genetic diversity in plants. We studied the sugarcane genome by analyzing SSAP (Sequence Specific Amplified Polymorphism) and IRAP (Inter-retrotransposons Amplified Polymorphism) profiles. SSAP is a multiplex amplified fragment length polymorphism (AFLP)-like technique that displays individual retrotransposon insertion as bands on a sequencing gel. IRAP displays on an agarose gel the products of a PCR amplification of sequences between two retrotransposons LTRs. Both strategies generate polymorphic band patterns in different sugarcane varieties as well S.officinarum and S. spontaneum progenitors. Our results suggest the occurrence of new insertion events during the hybridization or/and variety establishment process. The band patterns also indicate each progenitor contribution on sugarcane hybrid genome. Furthermore this polymorphism can be potentially use as a source of new molecular markers.

Final Reference	**W01-14**
Presenting author	Quang Hien Le
E-mail	hien.le@versailles.inra.fr
Presentation type	Poster
Session	W01 - Control of transposition and its impact on plant genes and genomes

DISTRIBUTION OF THE TNT1 RETROTRANSPOSON IN TOBACCO VARIES WITH AGE OF INSERTION

Quang Hien Le, Marie-Angèle Grandbastien
Biologie Cellulaire. INRA Centre de Versailles, 78026 Versailles France

Retrotransposons are mobile genetic elements that have the ability to insert into new genomic locations, increasing their copy number in the process. They can make-up a significant proportion of the repetitive sequence content of a host genome and their mobility makes them potential mutagens. Insertions into or near genes can alter expression patterns. In addition, retrotransposons can acquire and mobilise host genes and recombination between elements can generate chromosomal rearrangements. Tnt1 is a retrotransposon endogenous to tobacco and one of the few known elements that have been shown to be currently active. Demonstration that activation of Tnt1 can be induced by certain types of stress factors has lent support to the hypothesis that, in populations undergoing adverse conditions, transposition can be an important source of genetic variability upon which selection can act. Recently, sequences flanking pre-existing and novel insertions of Tnt1 were analysed in tobacco regenerants recovered from leaf mesophyll protoplasts. During protoplast isolation, Tnt1 is activated by the exposure to a fungal extract which is thought to mimic the stresses caused during pathogen attack. An analysis of the recovered sequences flanking new and pre-existing Tnt1 insertions detected a location bias towards genic regions for newly transposed copies whereas old copies tend to be found in non-coding or repetitive regions. This suggests that Tnt1 insertions into genic regions are initially favoured but that they are later counter-selected.

Final Reference	**S29-2**
Presenting author	Alan Schulman
E-mail	alan.schulman@helsinki.fi
Presentation type	Oral
Session	S29 - Biodiversity and germplasm characterization

RETROTRANSPOSONS AS EFFICIENT MOLECULAR MARKERS FOR GERMPLASM CHARACTERIZATION

Fiona Leigh[1], Ruslan Kalendar[2], Vincent Lea[1], David Lee[1], Paolo Donini[1], Alan Schulman[2]
1-Molecular Research Group, NIAB, Cambridge, UK
2-MTT/BI Plant Genomics Laboratory, Institute of Biotechnology, Viikki Biocenter, University of Helsinki, Helsinki, Finland

Retrotransposons are abundant, conserved, and dispersed components of plant genomes. Cycles of transcription, reverse transcription, and integration of these elements into the genome generate new copies and contribute to genomic diversity. These features make retrotransposons well-suited as molecular markers for germplasm characterization and molecular breeding applications. We and coworkers have developed and applied retrotransposon marker systems in a variety of crop plants, ranging from barley to banana. We have conducted a systematic comparison of the Sequence-Specific Amplification Polymorphism (S-SAP) method, together with related molecular marker techniques IRAP (inter-retrotransposon amplified polymorphism) and REMAP (retrotransposon-microsatellite amplified polymorphism) techniques in barley, and examined the utility of different retrotransposon families in producing polymorphic, scorable fingerprints. We have generated S-SAP data for three barley (Hordeum vulgare L.) varieties using primers based on sequences from six retrotransposon families (BARE-1, BAGY-1, BAGY-2, Sabrina, Nikita and Sukkula). The effect of the number of selective bases on the profiles has been examined and the profiles from eight 3-base Mse I selective primers compared for all the elements. Polymorphisms detected in the insertion pattern of all the families show that each can be used for S-SAP. The uniqueness of each transposition event and differences in the historic activity of each family suggests that a multi-retrotransposon family approach to genetic analysis will find applications for mapping, fingerprinting, and marker-assisted selection and for evolutionary studies, not only for barley and other Hordeum species and related taxa, but also more generally.

Final Reference	**W01-6**
Presenting author	Amar kumar
E-mail	akumar@scri.sari.ac.uk
Presentation type	Oral
Session	W01 - Control of transposition and its impact on plant genes and genomes

GENOMIC ORGANISATION, EVOLUTION AND IMPACT OF NON-AUTONOMOUS TRIM ELEMENTS IN THE PLANT GENOMES

Amar kumar[1], Ruslan Kalander[2], Peter Witte[1], Frank Wright[1], Arnis Druka[1], Molly Dunn[3], Alan Schulman[2], Stephen Goff[3], Hirohiko Hirochika[4]
1-Scottish Crop Research Institute, Invergowrie, Dundee, DD2 5DA, Scotland, UK
2-Plant Genomics Laboratory, Viikki Biocenter, University of Helsinki, P.O. Box 56, FIN-00014 Helsinki, FINLAND
3-Structural Genomics Group, Syngenta, USLJ, San Diego, CA 92121, USA
4-Department of Molecular Genetics, NIAR, Tsukuba, JAPAN

We have identified a new group of long terminal repeats (LTR) retrotransposons, termed terminal-repeat retrotransposons in miniature (TRIM) that are present in both monocotyledonous and dicotyledonous plants. TRIM elements have terminal direct repeat (TDR) sequences between ~100 to 250-bp in length that encompass an internal domain of ~100 to 300-bp. The internal domain contains primer binding site (PBS; complementary to methionyl initiator tRNA) and polypurine tract (PPT) motifs, but lacks the coding domains required for mobility. Thus, TRIM elements are not capable of autonomous transposition. We have now exploited recently available "almost" complete genomic sequences from several plant species (Arabidopsis thaliana, Oryza sativa (japonica and indica), Brassica oleracea, and Lotus truncatula) to perform a comprehensive study on a diverse populations of TRIM elements with respect to their genomic organization and evolution in plant genomes. The structural organization of TRIM elements suggests an evolutionary relationship to either LTR retrotransposons or retroviruses. Searches for "Mother element" have been negative to date, indicating either an ancient origin or a recent origin via horizontal transfer in the host genome. Unlike Ty1-copia and Ty3-gypsy, TRIM elements show a high level of conservation in structure, size and sequence identity levels within plants. The past mobility of TRIM elements is indicated by the presence of flanking 5-bp direct repeats typically found at LTR retrotransposon insertion sites, by the identification of related to empty sites (RESites) and by the presence of polymorphic bands using transposon display. A transcript of TRIM elements has also been detected by RT-PCR in plants. TRIM elements are actively involved in the restructuring of plant genomes and genes. In solanaceous species and in rice, TRIM elements provided target sites for further retrotransposon insertions. In Arabidopsis and in rice TRIM elements are involved in the transduction of host gene and retrotransposon sequences.

Final Reference	**W01-9**
Presenting author	wei chang
E-mail	
Presentation type	Poster
Session	W01 - Control of transposition and its impact on plant genes and genomes

REGULATION AND PROMOTER CHOICE IN RETROTRANSPOSON *BARE*-1 TRANSCRIPTION IN BARLEY

wei chang, alan H. schulman
BI-MTT Joint Genomics Laboratory, Viikki Science Park, P.O. Box 56, Helsinki, Finland and Plant Breeding Biotechnology, MTT Agrifood Research Finland, Myllytie 10, FIN-31600 Jokioionen, Finland

The *BARE*-1 retrotransposon occurs in more than 10^4 copies in the barley genome. The element is bounded by long terminal repeats (LTRs, 1829 bp) containing motifs typical of retrotransposon promoters. The demonstrated transcription, translation and VLP production by *BARE*-1 indicates that it is an active retrotransposon. Inheritance of integrated copies critically depends on cell-specific and tissue-specific expression patterns. In view of this, we looked for transcription of *BARE*-1 within different barley tissues and examined the promoter function of the *BARE*-1 LTR. Our results showed that *BARE*-1 like elements are actively transcribed in the various tissues. Transcripts beginning within the *BARE*-1 LTR downstream of TATA box 1 and TATA box 2 were found. Their relative transcription level was compared by nuclease protection assay and exact transcription start sites were mapped by RACE PCR. The LTR can drive expression of reporter genes in transiently transformed barley tissues. Furthermore, we identified regions within the LTR responsible for expression in barley leaves by deletion analyses of LTR-*luc* constructs. The differences in activity of the two promoters within different tissues were also investigated by particle bombardment. The activity of the LTR as a promoter, combined with the abundance of *BARE*-1 in the genome, suggests that *BARE*-1 may retain the potential for propagation in the barley genome.

Final Reference	**A-9**
Presenting author	Barbara Hohn
E-mail	barbara.hohn@fmi.ch
Presentation type	Invited lecture
Session	Plenary sessions invited speakers

GENETIC FLUX IN PLANTS - FROM MOLECULAR BIOLOGY TO ENVIRONMENTAL INFLUENCES

Barbara Hohn
Friedrich Miescher Institute for Biomedical Research, Basel, Switzerland

Environmental factors have played a major role in shaping plant genomes in the past and will continue to do so in the future. However, tight controls of changes of plant genomes are necessary in order to preserve existing quality and guarantee repair of mistakes in DNA.

In order to understand the balance between these apparently conflicting activities genes responsible for recognition, signal transduction and ultimate molecular responses are being analysed in plants exposed to environmental threats. In a genetic screen Arabidopsis mutants altered in their frequency of recombination at a transgenic locus were isolated. Analysis of targeted genes revealed a battery of components including DNA helicases and proteins involved in DNA repair and in signaling. Specifically, activities of chromatin remodeling proteins and proteins involved in the balancing of repair activities will be discussed. Complementary to this approach global changes in gene expression of wildtype plants grown under stress conditions known to induce homologous recombination (UV-C, bleomycin and the fungal elicitor xylanase) were investigated. Using Arabidopsis microarrays, both commonly and specifically induced or repressed genes could be identified.

Plants are constantly exposed to natural and/or artificial stresses such as UV-B, gamma radiation, high salt concentrations, heavy metals and pathogens. Plants use a range of activities to fight these influences. Concomitantly, also the rate of recombinational repair, as measured with our transgenic Arabidopsis lines, increased. This was especially dramatic for plants exposed to the environment of the Chernobyl reactor. These plants are thus sensitive bioindicators for radioactive (and other) environmental pollution. Of highest biological significance is probably the increased frequency of somatic recombination found after fungal attack of the plants or in plants subjected to simulated attacks. This may be an indication for pathogen stress mediated genomic change which, as we demonstrated, can be inherited in the next generation.

Final Reference	**B-15**
Presenting author	John Mundy
E-mail	mundy@biobase.dk
Presentation type	Invited lecture
Session	Concurrent sessions invited speakers

HORMONE SIGNALING IN STRESS & PATHOGENESIS

John Mundy
Molecular Biology Institute, Copenhagen University, Copenhagen, Denmark

Plants respond to certain pathogens by the induction of the hypersensitive response (HR) and the development of salicylate (SA)-dependent systemic acquired resistance (SAR). Loss of function mutations in genes that encode negative regulators of SAR or HR may cause plants to exhibit accelerated cell death and/or constitutive SAR.

For example, the Arabidopsis MAP kinase 4 (mpk4) mutant exhibits SAR with elevated SA levels, increased resistance to virulent pathogens, and constitutive pathogenesis-related gene expression (Petersen et al. 2000 Cell 103, 1111-20). MPK4 kinase activity is required to repress SAR as an inactive MPK4 form fails to complement mpk4. Analysis of mpk4 expressing the SA hydroxylase NahG, and of mpk4/npr1 double mutants, indicates that SAR in mpk4 is dependent upon elevated SA levels, but is independent of NPR1. PDF1.2 and THI2.1 gene induction by jasmonate was blocked in mpk4 expressing NahG, suggesting that MPK4 is required for JA-responsive gene expression.

Similarly, the accelerated-cell-death11 mutant (acd11) constitutively expresses defense-related genes and also exhibits characteristics of animal apoptosis (programmed cell death) monitored by flow cytometry (Brodersen et al. 2002 Genes & Develop. 16, 490-502). The PCD and defense pathways activated in acd11 are (SA)-dependent, but do not require intact jasmonic acid or ethylene signaling pathways. Epistatic analysis showed that the SA-dependent pathways require two regulators of SA-mediated resistance responses, PAD4 and EDS1. Furthermore, acd11 PR1 gene expression, but not cell death, depends on the SA signal tranducer NPR1, suggesting that the npr1-1 mutation uncouples resistance responses and cell death in acd11.

Final Reference	**B-45**
Presenting author	David Baulcombe
E-mail	david.baulcombe@sainsbury-laboratory.ac.uk
Presentation type	Invited lecture
Session	Concurrent sessions invited speakers

SIRNAS IN PLANTS

David Baulcombe
The Sainsbury Laboratory, Norwich, UK

RNA silencing is associated with epigenetic mechanisms at two levels. The first operates at the RNA level and provides protection against viruses. It involves a cycle in which an initiator double stranded (ds)RNA is processed into a population of primary short (21nt approx) RNAs that have two roles. Some of them are thought to guide a nuclease to target single stranded RNAs that have complete or near complete identity to the original dsRNA. The others prime the single stranded target RNA so that it can be converted into dsRNA by an RNA dependent RNA polymerase (SDE1/SGS2). This secondary dsRNA is then processed into secondary siRNAs that also have the dual guide and primer role. Once production of the secondary RNAs has been established the cycle is self maintaining as long as the target RNA is present. The second epigenetic mechanism involves a sequence-specific RNA-DNA interaction that results in methylation of the target DNA and, presumably, chromatin modification. If the target DNA is a promoter there is transcriptional gene silencing. The SDE4 gene is required for this mechanism. It is likely that short RNAs are involved although they are longer by three or four nucleotides than those associated with the SDE1/SGS2 mechanism. It is likely that this second mechanism has a role in genome protection. In my talk I shall describe recent data on the detailed aspects of these two epigenetic mechanisms and further findings about the likely biological roles.

Final Reference	**S28-23**
Presenting author	J. Enrique Cortés-Valle
E-mail	
Presentation type	Poster
Session	S28 - Gene silencing and epigenetics

A SYSTEM TO STUDY THE BIOLOGICAL ACTIVITY OF miRNAs

J. Enrique Cortés-Valle, David Baulcombe
The Sainsbury Laboratory, John Innes Centre, Norwich, UK

The mechanism of RNA silencing involves double-stranded RNA (dsRNA) degradation into short interfering RNAs (siRNAs) of 21-26 nt in length. The siRNAs are produced by Dicer, an RNase III protein, and incorporated in RNase protein complex (RISC) for degradation of homologous mRNAs. Endogenous short RNAs (microRNAs: miRNAs) have been found in different organisms, including plants, and have been implicated in regulation of development. One of the potential miRNA target genes is PHABULOSA (PHB) that encodes an HD-Zip transcription factor. Based on the molecular properties of phb mutants we propose that the PHB mRNA is downregulated on the abaxial surfaces of leaves by miRNA MIR165. To test this hypothesis we have generated transgenes producing transcripts in which the putative MIR165 target sequence is present in a GFP mRNA. The chimeric GFP/MIR165 transgenes will be introduced into Arabidopsis thaliana. We predict that the expression of GFP will be suppressed in cells with high levels of MIR165. This strategy could be a general approach to study of miRNA targeting.