Proc. Natl. Acad. Sci. USA
Vol. 89, pp. 7124–7128, August 1992
Genetics

copia-like retrotransposons are ubiquitous among plants

(transposable element/reverse transcriptase/molecular evolution)

DANIEL F. VOYTAS*†, MICHAEL P. CUMMINGS‡, ANDRZEJ KONIECZNY§, FREDERICK M. AUSUBEL§, AND STEVEN R. RODERMEL*

*Department of Botany, Iowa State University, Ames, IA 50010; ‡Museum of Comparative Zoology, Harvard University, Cambridge, MA 02138; and §Department of Molecular Biology, Massachusetts General Hospital, and Department of Genetics, Harvard Medical School, Boston, MA 02114

Communicated by Matthew S. Meselson, April 24, 1992 (received for review January 20, 1992)

ABSTRACT **Transposable genetic elements are assumed to be a feature of all eukaryotic genomes. Their identification, however, has largely been haphazard, limited principally to organisms subjected to molecular or genetic scrutiny. We assessed the phylogenetic distribution of *copia*-like retrotransposons, a class of transposable element that proliferates by reverse transcription, using a polymerase chain reaction assay designed to detect *copia*-like element reverse transcriptase sequences. *copia*-like retrotransposons were identified in 64 plant species as well as the photosynthetic protist *Volvox carteri*. The plant species included representatives from 9 of 10 plant divisions, including bryophytes, lycopods, ferns, gymnosperms, and angiosperms. DNA sequence analysis of 29 cloned PCR products and of a maize retrotransposon cDNA confirmed the identity of these sequences as *copia*-like reverse transcriptase sequences, thereby demonstrating that this class of retrotransposons is a ubiquitous component of plant genomes.**

Retrotransposons are the largest and best-characterized class of mobile genetic elements. Like retroviruses, retrotransposons encode a reverse transcriptase and replicate through an RNA intermediate (1). Retrotransposons can be divided into three major categories on the basis of their structural organization and amino acid similarities among their encoded reverse transcriptases (2, 3). Two of these classes are flanked by long terminal direct repeats (LTRs) and encode proteins similar to the retroviruses. These LTR-retrotransposons are referred to as the *gypsy*-like and *copia*-like retrotransposons after representative elements of each class in *Drosophila melanogaster*. The third class of retrotransposons, the LINE1-like or non-LTR retrotransposons, lack terminal repeats and encode proteins with significantly less similarity to those of the retroviruses.

A search for transposable elements in the flowering plant *Arabidopsis thaliana* led us to the discovery of a family of retrotransposons called Ta*1* (4). The structural organization of these elements and phylogenetic analyses based on the Ta*1* reverse transcriptase clearly placed this family of elements among the *copia*-like retrotransposons (3). At the time of their discovery in *Arabidopsis*, characterized *copia*-like retrotransposons included Ty*1* and Ty*2* of *Saccharomyces cerevisiae* (5) and two element families in *D. melanogaster*, *copia* and *1731* (6, 7). More recently we have characterized two other *copia*-like retrotransposons in *A. thaliana* (Ta*2* and Ta*3*) (8), and additional *copia*-like elements have been identified in tobacco (Tnt*1*) (9), potato (Tst*1*) (10), wheat (WIS-*2*) (11), and the slime mold *Physarum polycephalum* (Tp*1*) (12).

The presence of *copia*-like retrotransposons among plants, insects, fungi, and protists suggested that this class of elements might be a universal component of eukaryotic genomes. *copia*-like elements are too divergent at the nucleotide level to be identified by DNA hybridization techniques. However, the amino acid sequences of *copia*-like element reverse transcriptases are sufficiently conserved to permit the unambiguous classification of a reverse transcriptase as being encoded by a *copia*-like retrotransposon (2, 3). Previously, we designed degenerate oligonucleotide primers based on two highly conserved amino acid sequence domains among the reverse transcriptases of *copia* and the plant *copia*-like retrotransposons Ta*1*, Ta*2*, Ta*3*, and Tnt*1* (Fig. 1). We used these primers in the polymerase chain reaction (PCR) to identify seven additional families of *copia*-like retrotransposons in *A. thaliana* (Ta*4*–Ta*10*) (8). In this paper we report the results of a PCR survey conducted to assess the distribution of *copia*-like retrotransposons across a wide phylogenetic spectrum.¶

MATERIALS AND METHODS

DNA Samples. DNA samples were generously provided by a number of researchers or extracted as previously described (14).

DNA Manipulations. PCR amplifications were performed in 25- to 100-μl reaction mixtures with 0.1–1 μg of genomic DNA, each dNTP at 200 μM, 50 pmol of each primer, 2–5 units of *Taq* DNA polymerase, and buffer provided by the supplier of the enzyme (Perkin–Elmer/Cetus; Promega). To prevent and monitor any DNA contamination, reaction mixtures were treated with UV light prior to the addition of genomic DNA and enzyme (15) and control reactions (without genomic DNA) were included with each amplification. Temperature cycling was performed on an Ericomp thermocycler with the following profile: 94°C for 1 min, 47°C for 1 min, and 72°C for 2 min for a total of 30–50 cycles. Oligonucleotide primers for PCR were based on two highly conserved amino acid sequence domains of retrotransposon reverse transcriptases (Fig. 1). The sequences of the oligonucleotide primers have been previously reported (8) or slightly modified as follows: 5′ primer, GGAATTCGAYGTNAARACNGCNTTYYT; 3′ primer, GGGATCCAYRTCRTCNACRTANARNA, where N = A + C + G + T, R = A + G, and Y = T + C.

PCR reaction mixtures were extracted with chloroform, and products were precipitated with isopropyl alcohol and digested with *Bam*HI and *Eco*RI; restriction sites for these enzymes had been incorporated into the primers (see above). Amplification products were gel purified and inserted into M13mp18 or mp19 vectors, and their DNA sequences were determined (ref. 16, chap. 7). The reverse transcriptase sequences were hybridized to Southern filters containing *Eco*RI-digested genomic DNA for the species from which the reverse transcriptase was derived and at least one other

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 sequences reported in this paper have been deposited in the GenBank data base (accession nos. M94470–M94498).

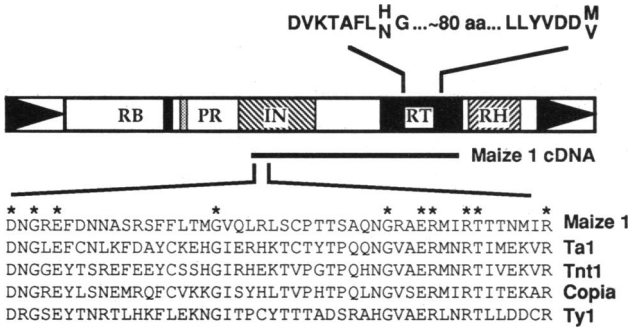

FIG. 1. Structural organization of *copia*-like retrotransposons. Boxes containing arrowheads represent the retrotransposon terminal direct repeats. Boxes within the internal portion of the element indicate conserved amino acid domains: RB, RNA-binding domain; PR, protease; IN, integrase; RT, reverse transcriptase; RH, RNase H. Consensus amino acid (aa) residues depicted above the reverse transcriptase were used to design degenerate oligonucleotides for PCR (8). The line below the element represents the Maize 1 cDNA with the amino acid sequence of its integrase domain aligned to several other *copia*-like retrotransposon integrases [Ta*1* from *A. thaliana* (4); Tnt*1* from *Nicotiana tabacum* (9), *copia* from *D. melanogaster* (6), Ty*1* from *S. cerevisiae* (13)]; asterisks indicate invariant amino acids.

species. Hybridizations were conducted as previously described (17).

A maize λ gt11 cDNA library (obtained from J. Shen and A. Schäffner, Massachusetts General Hospital) was also used as a template for PCR amplifications with the degenerate oligonucleotides. A single *copia*-like reverse transcriptase clone was identified (Maize 1) and used as a hybridization probe to screen the library (ref. 16, chap. 6). A 2.1-kilobase-pair (kbp) cDNA was identified and subcloned in M13mp18 and mp19 vectors, and the DNA sequence was obtained for both strands (ref. 16, chap. 7).

Phylogenetic Analyses. Preliminary sequence manipulations were performed by using the Genetics Computer Group programs (18), and the amino acid sequence alignment was generated by using the program TREEALIGN (19, 20) with a gap penalty of $g_k = 9 + 3k$. The aligned amino acid sequences, excluding those residues encoded in the primers, formed the basis for the phylogenetic analysis using the program PAUP, version 3.0r+4 (21). Each amino acid position was scored as a character and gaps were scored as missing data. The weighting for character state transitions (distance between character states) was based on the protein parsimony matrix in PAUP. This model accounts for the minimum number of steps between amino acids (22). The tree bisection–reconnection branch–swapping algorithm with random sequence addition was used for 898 replicates. A bootstrap analysis was also performed consisting of 100 resampling replicates. In addition to the parsimony analyses, a distance-based analysis, neighbor-joining, was conducted by using the program CLUSTALV (23) with corrected distances (24). The phylogenetic trees were rooted by using Ty*1* as an outgroup.

RESULTS

A PCR Survey of *copia*-like Retrotransposons. The presence of *copia*-like retrotransposons among various eukaryotes was tested by amplifying genomic DNAs with degenerate oligonucleotides based on conserved amino acid sequence domains of *copia*-like reverse transcriptases. The PCR primers have previously been shown to amplify characterized *copia*-like retrotransposons of *A. thaliana* (8), and in this report, the PCR product from *D. melanogaster* was shown by DNA sequence analysis to be the *copia* element reverse transcriptase (data not shown).

Amplified sequences of the expected size (≈300 bp) were visualized on agarose gels for all 64 plant species surveyed. These included representatives from 9 of 10 divisions of the plant kingdom (Table 1): mosses (Bryophyta), horsetails (Sphenophyta), lycopods (Lycophyta), ferns (Pterophyta), cycads (Cycadophyta), *Ginkgo* (Ginkgophyta), *Gnetum* (Gnetophyta), conifers (Coniferophyta), and 38 species of angiosperms (Anthophyta) representing both monocots and dicots. In addition to these plant species, positive amplification results were also obtained for the photosynthetic protist *Volvox carteri*. Amplification products could not be detected for several other species of protists (*Chlamydomonas, Trypanosoma*, and *Tetrahymena*) and animals (nematode, rotifer, several insects, fish, frog, chicken, mouse, and human).

Determining the Authenticity of the PCR Products. Several tests were carried out to verify that the amplification products represented *copia*-like reverse transcriptases and that the sequences were legitimate components of the taxa examined. First, for 23 diverse species, PCR products were cloned and 102 sequences were determined. While most of these sequences showed clear similarity to reverse transcriptase sequences (data not shown), only 29 clones were found to encode a single open reading frame uninterrupted by stop codons or translational frameshifts. An alignment of the derived amino acid sequences of these cloned DNAs is presented in Fig. 2 along with some previously determined

Table 1. Species that tested positive by the PCR assay

Division	Species (common name)
Chlorophyta	*Volvox carteri* (volvox)
Bryophyta	*Polytrichum* sp.
Sphenophyta	*Equisetum scirpoides* (equisetum)
Lycophyta	*Isoetes melanopoda, Lycopodium obscurum, Selaginella* sp.
Pterophyta	*Adiantum pedatum, Marattia* sp., *Botrychium virginiana, Osmunda cinnamomea* (osmunda)
Cycadophyta	*Cycas revoluta* (cycas), *Stangeria eriopus, Bowenia spetabilis, Ceratozamia mexicana*
Ginkgophyta	*Ginkgo biloba* (ginkgo)
Gnetophyta	*Ephedra nevadensis, Gnetum montanum* (gnetum), *Welwitschia mirabilis*
Coniferophyta	*Agathis robusta, Cephalotaxus fortuni, Callitris roei, Pinus coulteri* (pine), *Cedrus deodara, Picea glauca, Podocarpus gracilior, Metasequoia glyptostroboides, Taxus baccata*
Anthophyta	*Liriodendron chinense, Liriodendron tulipifera, Platanus occidentalis, Nelumbo lutea, Victoria* sp., *Ceratophyllum demersum, Gossypium hirsutum* (cotton), *Sarracenia pupurea, Drosera rotundifolia, Turnera ulmifolia, Brassica napus* (rape), *Brassica nigra, Hesperis matronalis, Streptanthus tortuosus, Monotropa uniflora, Pterospora andromedea, Trientalis borealis, Glycine max* (soybean); *Lupinus luteus, Prunus persica, Asclepias syriaca, Pholisma arenarium, Lycopersicon esculentum* (tomato), *Nicotiana tabacum, Solanum tuberosum* (potato), *Petunia hybrida* (petunia), *Helianthus annuus* (sunflower), *Taraxacum officinale, Avena sativa* (oat), *Hordeum vulgare* (barley), *Oryza sativa* (rice), *Sorghum bicolor, Triticum aestivum* (wheat), *Zea mays* (maize), *Uvularia sessilifolia* (lily), *Cephalanthera austinae, Corallorhiza maculata, Cypripedium acaule*

FIG. 2. Alignment of inferred amino acid sequences for *copia*-like retrotransposon reverse transcriptases. Lines at the top of the figure depict amino acids encompassed by both sets of PCR primers used in the survey (see *Materials and Methods* and ref. 8). Asterisks indicate amino acids shared by ≥30 of the 35 sequences (86%). Species names for each of the taxa are given in Table 1, and sequences for previously characterized retrotransposons not referenced in Fig. 1 include *1731* of *D. melanogaster* (7) and Ta*10* of *A. thaliana* (8).

copia-like reverse transcriptases. The sequence for this region of reverse transcriptase is distinct enough among the *copia*-like elements to unambiguously identify the clones as *copia*-like retrotransposon reverse transcriptases (2, 3). Sequence pairs ranged in similarity from 29% (oat and *Volvox* 2) to 92% (*L. tulipifera* and *L. chinense* 1) and the average similarity between sequences was 46%.

Second, sequenced DNA clones were used as hybridization probes to Southern filters containing DNA from several different plant species. With few exceptions (e.g., sequences from wheat, oat, and barley) the reverse transcriptase sequences showed little or no cross-hybridization to other plant genomes, indicating that they arose from the taxa examined. The reverse transcriptases ranged from very low to very high copy number (Fig. 3).

In a final test to determine the authenticity of the amplification products, one of the cloned maize sequences (Maize 1) was used as a hybridization probe to screen a maize cDNA library. A 2.1-kbp cDNA clone was isolated that encodes a single open reading frame of 700 amino acids (Fig. 1). While the size of this clone suggests that it is not a complete cDNA copy of a retrotransposon mRNA (Fig. 1), the 206 amino acids at the 3′ end of the clone (including the 88 amino acids amplified by the primers) encompass all of the conserved amino acid domains that characterize retrotransposon reverse transcriptases (data not shown) (3). Preceding the reverse transcriptase at the 5′ portion of the Maize 1 cDNA is a region that shows significant amino acid similarity to retrotransposon and retroviral integrases (Fig. 1). The order of the integrase and reverse transcriptase coding regions is a distinguishing characteristic of *copia*-like retrotransposons; for *copia*-like elements the integrase precedes reverse transcriptase, and this order is reversed for retroviruses and other classes of retrotransposons (2, 3). The Maize 1 cDNA thus has a distinct *copia*-like gene order.

Phylogenetic Analyses. Phylogentic analyses were conducted to assess relationships among the reverse transcriptases. The strict consensus tree based on the six equally parsimonious trees found is shown in Fig. 4. The neighbor-joining analysis produced a tree that had 20 of 36 internal branches in common with the parsimony consensus tree. The bootstrap analysis also suggests robustness for many of the relationships (Fig. 4).

DISCUSSION

***copia*-like Retrotransposons Are Ubiquitous Among Plants.** We have used the PCR to demonstrate that *copia*-like retrotransposons are ubiquitous components of plant genomes.

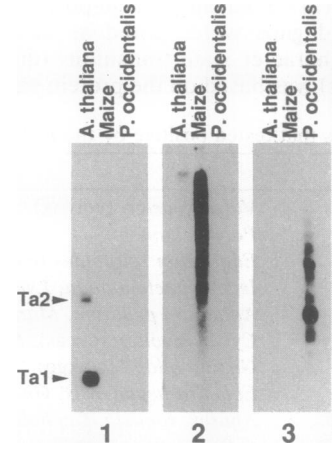

FIG. 3. Southern blot analysis of reverse transcriptases from the *A. thaliana* Ta*1* elements, the Maize 1 cDNA, and *Platanus occidentalis*. Replicate filters were prepared and hybridized with a 1.0-kbp *Eco*RI clone encompassing the Ta*1* reverse transcriptase (gel 1) (25), a 1.2-kbp *Eco*RI/*Cla* I fragment encompassing the Maize 1 reverse transcriptase (gel 2), and a 0.3-kbp fragment representing the *P. occidentalis* 1 reverse transcriptase (gel 3). The two comigrating copies of Ta*1* in the *A. thaliana* race Landsberg are labeled on the figure, as is the cross-hybridizing single-copy element Ta*2* (8, 25).

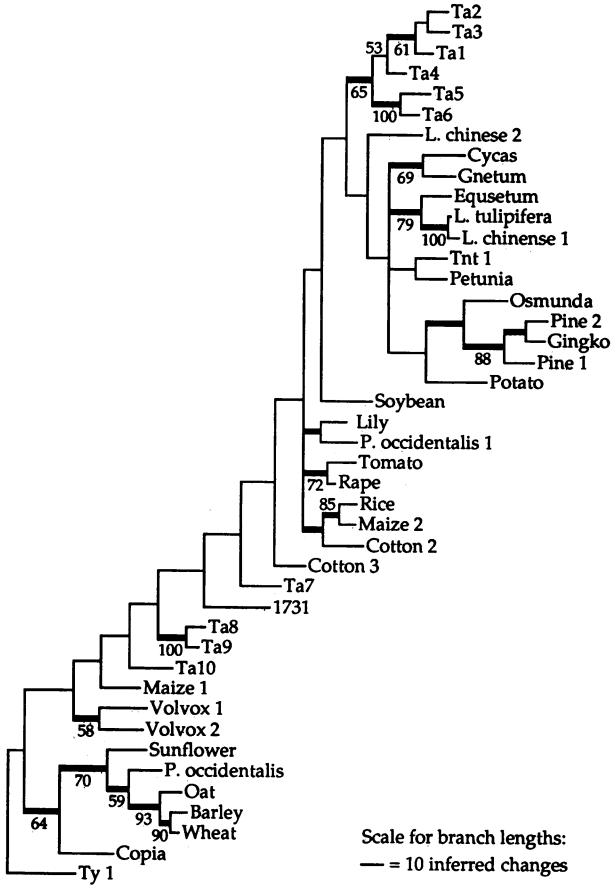

FIG. 4. Phylogenetic tree of the *copia*-like retrotransposons. A strict consensus tree is shown, derived from six equally parsimonious trees of length 1088. Branch lengths are proportional to the number of changes along each branch. Numerals adjacent to branches indicate percentage of bootstrap replicates supporting that branch, and thick branches are those common to parsimony and neighbor-joining trees.

These elements are present in all major lineages of plants, and we know of no other demonstration of the ubiquity of a given class of transposable elements across a broad phylogenetic spectrum.

The utility of the PCR assay for identifying plant *copia*-like retrotransposons has recently been corroborated by two different laboratories. Flavell *et al.* (26) have used a PCR assay based on the same conserved amino acid sequence domains to demonstrate the presence of *copia*-like elements in several species of Solanaceae, as well as in pea and barley. They have independently identified the same potato retrotransposon described in this survey. Likewise, the rice retrotransposon reported here was also identified by Hirochika *et al.* (27), using a similar PCR strategy.

For many of the plants analyzed in this survey, including a number of major crop species, the *copia*-like retrotransposons are, to our knowledge, the first transposable elements to be described. While two detailed studies of *copia*-like retrotransposons in *A. thaliana* indicated that these elements are likely no longer capable of transposition (8, 25), it is clear that retrotransposons in other plant species are both active and capable of causing mutations (9, 28). Many of the element families identified in this study are present in high copy number, suggesting transposition activity during at least part of their evolutionary history. Since transcription is the first requisite step in retrotransposition, the isolation of a maize cDNA clone can be taken as additional evidence that these elements are active and that they may play an ongoing role in shaping plant genome evolution.

In addition to the plant retrotransposons, we identified related retrotransposons in the photosynthetic protist *Volvox carteri*, and the PCR assay was successful in amplifying the *copia* element reverse transcriptase from *D. melanogaster*. Other, more completely characterized, *copia*-like retrotransposons have been identified outside of the plant kingdom [Ty*1* and Ty*2* in *S. cerevisiae* (5), Tp*1* in *Physarum polycephalum* (12), and *1731* in *D. melanogaster* (7)]. The reverse transcriptases of these elements are sufficiently diverged such that they would not be amplified by the primers used in this survey. Distinct subclasses of *copia*-like retrotransposons may be present in other lineages of organisms and thereby explain the failure of the PCR assay to detect retrotransposons in most nonplant taxa examined.

Evolution of *copia*-like Retrotransposons. Two different mechanisms of genetic transfer are frequently used to explain the distribution of transposable elements among species (e.g., see ref. 8). Like most genes, transposable elements can be transmitted vertically from generation to generation. Horizontal transfer, the transmission of genetic material by nonsexual means, has been frequently used to explain the distribution of closely related transposable elements among diverse species (2, 3). Although these mechanisms are not mutually exclusive, both can be considered in light of our present findings.

The ubiquity of *copia*-like retrotransposons throughout plants may be explained by assuming that the common ancestor to plants also had similar retrotransposons and that during subsequent speciation events the retrotransposons were transmitted vertically along with other components of the genome. The presence of *copia*-like retrotransposons in all eukaryotic kingdoms lends support to this interpretation and implies that this group of transposable elements is very ancient.

Vertical relationships among several plant *copia*-like retrotransposons are revealed through phylogenetic analyses of their reverse transcriptases. For example, the reverse transcriptases of several monocots (i.e., wheat, barley, and oats) show a high degree of similarity and a close phylogenetic relationship, implying that they shared a common ancestor prior to speciation (Fig. 4). The reverse transcriptases that share the highest degree of similarity are *L. tulipifera* and *L. chinense* 1 (92%), sequences from the two species that make up the genus *Liriodendron*. These species have been reproductively isolated since the late Miocene (10–16 million years ago) (29). The high degree of similarity between their reverse transcriptases implies that retrotransposons were present prior to the separation of these species, and they have undergone little change since that time.

In contrast to vertical relationships, relationships of many *copia*-like reverse transcriptases show little congruence with the evolutionary relationships of their hosts. For example, elements from the same family (Solanaceae: potato, tomato, tobacco, and petunia), the same genus (*L. tulipifera, L. chinense* 2), and the same species (Maize 1, Maize 2) appear to share a more recent common ancestor with elements from considerably distant taxa. While such a topology is consistent with horizontal transfer, consideration should also be given to factors that could similarly obscure relationships among these sequences, as discussed below.

Much of the incongruence of the tree may be resolved if a more exhaustive survey were undertaken to identify the full complement of *copia*-like retrotransposons among the plant species tested. *A. thaliana*, the plant with the smallest known genome of any higher plant, has 10 retrotransposon families (8), and for many of the plants surveyed we have identified more than one distinct reverse transcriptase. Furthermore, Flavell *et al.* (26) have recently characterized 31 *copia*-like reverse transcriptase sequences in potato, and these sequences fall into six related subgroups. The sampling of

elements identified in this survey, therefore, probably does not reflect the true diversity of these sequences in plants, and the incongruence of the reverse transcriptase phylogeny may simply reflect comparisons of sequences that are derived from different subclasses of elements (i.e., comparisons of elements that are not orthologous).

Relationships among retrotransposons are also likely complicated by factors intrinsic to proliferation by reverse transcription. Reverse transcriptases do not have a proofreading mechanism for DNA synthesis, and retroviral proteins have been shown to evolve at rates that far exceed those of normal cellular genes (30). Active retrotransposon families, therefore, may have undergone periods of elevated mutation and recombination typically associated with reverse transcription and transposition (30, 31). Retrotransposons, however, can also remain quiescent in the genome for extended periods of time (e.g., see ref. 32), and it is likely that different element lineages can evolve at very different rates. Finally, relationships among retrotransposons may be further complicated by the fact that extinction of retrotransposons may have occurred in some lineages, as has been shown for the *A. thaliana* retrotransposons (25). It is interesting to note in this regard that in a study of inbred mice strains, phylogenetic relationships based on vertically inherited endogenous retroviruses deviated significantly from known phylogenetic relationships of the strains examined (33).

The relationships among the plant *copia*-like retrotransposons do not unambiguously distinguish between contributions made by vertical and horizontal transfer in their distribution. However, vertical inheritance of transposable elements provides a well-established mechanism for the observed distribution. If the alternative, horizontal transfer, is to explain the present distribution of *copia*-like elements throughout all plants, as well as their presence in all major lineages of eukaryotes, it has to be invoked an indeterminate number of times. While horizontal transfer is not excluded by the present study, it is not required either.

The ability to clone *copia*-like retrotransposons by the PCR assay further provides the opportunity to directly address factors that contribute to their evolution. As an example, we have recently cloned *copia*-like reverse transcriptases from ancient DNA (18–20 million years old) which had been extracted from fossil remains of several extinct plant species, among them *Liriodendron hesperia* Berry, the likely ancestor of the modern species of *Liriodendron* (unpublished). Such sequences should prove useful in tracing rates at which elements change over given geologic time scales. In addition, the PCR assay might be used to exhaustively identify retrotransposons from plant species with well-established evolutionary histories to determine further the manner and rate in which these sequences change and the role played by horizontal transfer in their distribution.

We are particularly grateful to the many colleagues who provided DNA samples. This work was supported by a grant from the Iowa State University Biotechnology Council to S.R.R., grants from the National Institutes of Health to R. C. Lewontin and J. D. Boeke, and a grant from Hoechst A.G. to the Massachusetts General Hospital. D.F.V. is a Genentech Fellow of the Life Sciences Research Foundation.

1. Boeke, J. D., Garfinkel, D. J., Styles, C. A. & Fink, G. R. (1985) *Cell* **40**, 491–500.
2. Doolittle, R. F., Feng, D.-F., Johnson, M. S. & McClure, M. A. (1989) *Q. Rev. Biol.* **64**, 1–30.
3. Xiong, Y. & Eickbush, T. H. (1990) *EMBO J.* **9**, 3353–3362.
4. Voytas, D. F. & Ausubel, F. M. (1988) *Nature (London)* **336**, 242–244.
5. Boeke, J. D. (1989) in *Mobile DNA*, eds. Berg, D. E. & Howe, M. M. (Am. Soc. Microbiol., Washington), pp. 335–374.
6. Mount, S. M. & Rubin, G. M. (1985) *Mol. Cell. Biol.* **5**, 1630–1638.
7. Fourcade-Peronnet, F., d'Auriol, L., Becker, J., Galibert, F. & Best-Belpomme, M. (1988) *Nucleic Acids Res.* **16**, 6113–6125.
8. Konieczny, A., Voytas, D. F., Cummings, M. P. & Ausubel, F. M. (1991) *Genetics* **1**, 801–809.
9. Grandbastien, M.-A., Spielmann, A. & Caboche, M. (1989) *Nature (London)* **337**, 376–380.
10. Camirand, A. & Brisson, N. (1990) *Nucleic Acids Res.* **18**, 4929.
11. Moore, G., Lucas, H., Batty, N. & Flavell, R. B. (1991) *Genomics* **10**, 461–468.
12. Rothnie, H. M., McCurrach, K. J., Glover, L. A. & Hardman, N. (1991) *Nucleic Acids Res.* **19**, 279–286.
13. Boeke, J. D., Eichinger, D., Castrillon, D. & Fink, G. R. (1988) *Mol. Cell. Biol.* **8**, 1432–1442.
14. Doyle, J. J. & Doyle, J. L. (1990) *Focus* **12**, 13–15.
15. Sarkar, G. & Sommer, S. S. (1990) *Nature (London)* **343**, 27.
16. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. & Struhl, K. (1991) *Current Protocols in Molecular Biology* (Green/Wiley-Interscience, New York).
17. Church, G. M. & Gilbert, W. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 1991–1995.
18. Devereux, J., Haekerli, P. & Smithies, O. (1984) *Nucleic Acids Res.* **12**, 387–395.
19. Hein, J. (1989) *Mol. Biol. Evol.* **6**, 649–668.
20. Hein, J. (1989) *Mol. Biol. Evol.* **6**, 669–684.
21. Swofford, D. L. (1991) *Phylogenetic Analysis Using Parsimony, PAUP Portable Version (Unix) 3.0r+4 (Prerelease 0.4)* (Illinois Natural History Survey, Champaign).
22. Felsenstein, J. (1991) *PROTPARS—Protein Sequence Parsimony Method, PHYLIP: Phylogeny Inference Package Version 3.3* (Department of Genetics, Univ. of Washington, Seattle).
23. Higgins, D. G. & Sharp, P. M. (1989) *Comp. Appl. Biosci.* **5**, 151–153.
24. Kimura, M. (1983) *The Neutral Theory of Molecular Evolution* (Cambridge Univ. Press, Cambridge).
25. Voytas, D. F., Konieczny, A., Cummings, M. P. & Ausubel, F. M. (1990) *Genetics* **126**, 713–721.
26. Flavell, A. J., Smith, D. B. & Kumar, A. (1992) *Mol. Gen. Genet.* **231**, 233–242.
27. Hirochika, H., Fukuchi, A. & Kikuchi, F. (1992) *Mol. Gen. Genet.*, in press.
28. Weil, C. F. & Wessler, S. R. (1990) *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **41**, 527–552.
29. Parks, C. R. & Wendel, J. F. (1990) *Am. J. Bot.* **77**, 1243–1256.
30. Holland, J., Spindler, K., Horodyski, F., Grabau, E., Nichol, S. & VandePol, S. (1982) *Science* **215**, 1577–1585.
31. Hu, W.-S. & Temin, H. M. (1990) *Science* **250**, 1227–1232.
32. Dombroski, B. A., Mathias, S. L., Nanthekumar, E., Scott, A. F. & Kazazian, H. H., Jr. (1991) *Science* **254**, 1805–1808.
33. Atcheley, W. R. & Fitch, W. M. (1991) *Science* **254**, 554–558.