Proc. Natl. Acad. Sci. USA
Vol. 86, pp. 5015–5019, July 1989
Genetics

Plant retrotransposon from *Lilium henryi* is related to *Ty3* of yeast and the gypsy group of *Drosophila*

(*del*/dispersed repeat/horizontal transmission/lily/polyprotein)

D. R. SMYTH, PAUL KALITSIS, JOAN L. JOSEPH, AND JOHN W. SENTRY*

Department of Genetics, Monash University, Clayton, Victoria 3168, Australia

Communicated by Ray D. Owen, February 24, 1989 (received for review December 21, 1988)

ABSTRACT **The lily retrotransposon *del* 1-46 is 9345 base pairs (bp) long. It has long terminal repeats (LTRs) of 2406 bp (left) and 2415 bp (right), which differ in sequence by 1.4%. Sequences similar to those involved in priming DNA synthesis in retroviruses occur in the internal region. Near the left LTR is a sequence complementary to 18 residues at the 3′ end of methionine initiator tRNA of three plant species, and a run of 12 purines occurs close to the right LTR. One internal reading frame of *del* 1-46 has relatively few stop codons. The 1462-codon product from this frame has motifs, in N to C terminus order, corresponding to those identified with RNA binding, protease, reverse transcriptase, RNase H, and integrase functions in retroviruses and certain other retrotransposons. Amino acid sequence comparisons of three conserved *pol* regions show *del* to be closely related to the *Ty3* retrotransposon of yeast (37–40% identity). *del* is also related to the gypsy group of *Drosophila* (*17.6*, *297*, gypsy/*mdg4*, and *412*), showing closer identity with their reverse transcriptase (32–38%) and RNase H (36–45%) domains than with their integrase domain (21–26%). It is proposed that a gypsy group ancestor exchanged the integrase region with a more distantly related element since its divergence from a *del*/*Ty3* common ancestor. The occurrence of related retrotransposons in three different kingdoms (plants, animals, and fungi) strongly implies their horizontal transmission in recent evolutionary time.**

The genome size of eukaryotes varies over an enormous range between species. Among flowering plants, for example, *Lilium* sp. have haploid genomes of ≈35 million kilobase pairs (kbp) (1), 500 times that of *Arabidopsis thaliana*, a plant at the other end of the scale (2). The DNA of very large genomes includes many families of repeated sequences whose individual members are interspersed with each other and with single copy sequences (3). The multiple copy number and scattered distribution of dispersed repeats together imply that they are derived from mobile elements.

With the aim of defining the categories of dispersed repeat involved in the architecture of very large genomes, we have been characterizing the most abundant large element present in the genome of *Lilium henryi* (4, 5). This repeat, called *del*, occurs in excess of 13,000 copies per genome (4). It is 9.35 kbp long and has long terminal direct repeats (LTRs) of 2.4 kbp. The LTRs usually have the nucleotides TGT and ACA at their 5′ and 3′ termini, and a tandem repeat of 5 bp flanks the element (5). These properties parallel those of retrotransposons, a class of mobile element well characterized in *Drosophila melanogaster* and yeast.

Retrotransposons are closely related to retroviruses (6–9). Their LTRs are of similar design, and retrotransposons have sequences related to retroviral polyprotein (*pol*) genes, which encode protease, reverse transcriptase, RNase H, and endonuclease (integrase) activities. In the case of *Ty1* of yeast at least, retrotransposition has been shown to involve an RNA intermediate as in retroviruses (10), and integrase activity has been demonstrated *in vitro* (11).

By estimating the level of amino acid sequence identity in conserved *pol* domains, ancestral relationships have been deduced between retrotransposons including *Ty1* (12) and the recently described *Ty3* of yeast (13), and the *Drosophila* elements *copia* (14, 15) and the gypsy group, which includes *17.6* (16), *297* (17), gypsy (18) [first recorded as *mdg4* (19)], and *412* (20). These comparisons have been extended to include retroviruses and other retroid elements with coding potential for reverse transcriptase (21–23).

We now report on the sequence of one *del* repeat from *L. henryi* and deduce its relationship to previously characterized retroid elements.

METHODS

The 9.35 kbp of *del* in clone 1-46, a 15-kbp fragment of genomic DNA from *L. henryi* inserted into the λ phage vector EMBL3 (5), was fully sequenced in each direction by the chain-termination method (24). Phage 1-46 DNA was digested with *Bam*HI and *Sal* I and the resulting fragments were cloned into plasmids pBR322, pUC18, or pUC19. Four clones with inserts of 3.9, 2.0, 4.95, and 1.65 kbp, which completely span the *del* homologous region, were selected. They were further digested with *Hin*dIII (and, in several cases, other enzymes) to obtain smaller fragments of the inserts, which were then subcloned into pUC18 or pUC19. These were digested from each end for a range of times using BAL-31 exonuclease, and the digestion products as well as the full inserts were ligated into the replicative form of M13 phage vectors mp18 and mp19. Their sequence was obtained from single-stranded phage DNA preparations by conventional methods (25). Between 4 and 6 overlapping BAL-31 deletions were sequenced per kbp of insert. Finally, the sequence across each of the original *Bam*HI and *Hin*dIII subcloning sites was obtained from a further set of *Eco*RI and *Xba* I subclones of 1-46.

The DNA sequence was analyzed and homology searches were made against other sequences in the GenBank Nucleic Acid Data Base using GENEPRO. For translated sequences FASTP was used for comparisons with sequences in the Protein Identification Resource of the National Biomedical Research Foundation. The full DNA sequence and that of the putative polypeptide product have been deposited with the EMBL Data Library (accession no. X13886).

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.

Abbreviations: CaMV, cauliflower mosaic virus; CERV, carnation etched ring virus; LTR, long terminal repeat; Mo-MLV, Moloney murine leukemia virus.

*Present address: The Children's Medical Research Foundation, P.O. Box 61, Camperdown, NSW 2050, Australia.

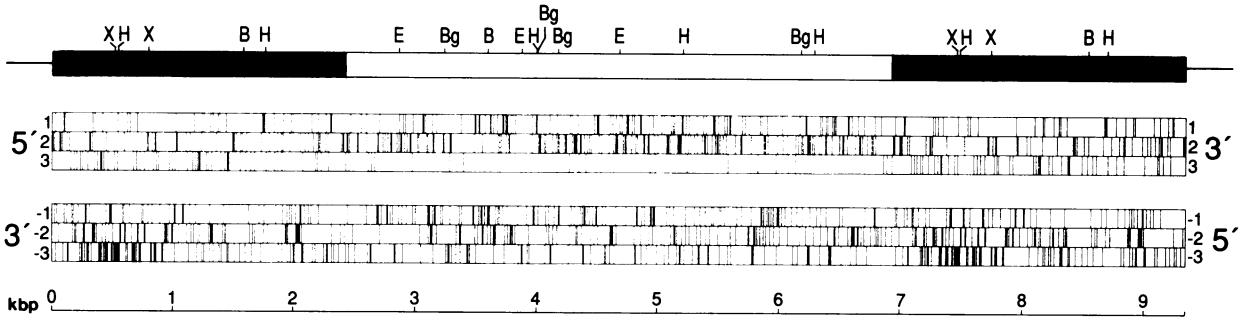

FIG. 1. Location of stop codons (vertical bars) in the six reading frames of the 9345-bp *del* sequence in clone 1-46. LTRs are shaded. B, *Bam*HI; Bg, *Bgl* II; E, *Eco*RI; H, *Hin*dIII; X, *Xba* I.

RESULTS

The full sequence of *del* 1-46 is 9345 bp long and is A+T-rich (64.2%). The left LTR is 2406 bp long and the right is 2415 bp; there are 4524 bp in the internal region. The two LTRs differ in 34 positions (1.4%), with 33 base substitutions (24 transitions and 9 transversions) and one region where the right LTR has 9 residues more than the left. Stop codons are abundant in the two LTRs, and there are no open reading frames longer than 312 bp (Fig. 1). The LTR sequences will be analyzed in more detail elsewhere

The internal sequences adjacent to the LTRs have properties indicating that they are priming sites for DNA synthesis (Fig. 2). In retroviruses minus-strand DNA synthesis is primed from a tRNA bound close to the 3′ end of the left LTR (8). A *del* sequence in this region is the complement of 17 residues at the 3′ end of methionine initiator tRNA encoded by nuclei of wheat, yellow lupine, and french bean (26). Close to the right LTR of *del* is a run of 12 purines, a characteristic of plus-strand priming sites in retroviruses (8). Immediately upstream of this is a string of 9 pyrimidines, a property shared by certain other retrotransposons (16, 17) and plant caulimoviruses (27, 28).

When the internal region is examined for open reading frames, the largest all occur in frame 3 (Fig. 1). This frame has only 21 stops, compared with between 77 (frame –3) and 144 (frame 2) in the other five frames. It may be that most of these stops represent mutational changes in the sequence of a formerly active element. To test this, we examined the consensus sequence of a 366-bp *Bgl* II/*Bam*HI fragment among five individual genomic clones (pLh205, -210, -217, -218, and -219) obtained earlier (4). In clone 1-46 this fragment covers two stop codons toward the left of the internal region (Fig. 1). In all other clones, these two TAA stop codons were represented instead by CAA, thus supporting the idea that they arose by mutation in 1-46. In analyzing the hypothetical translation product of this frame (Fig. 3), we have therefore chosen to read the full frame from the first methionine after the left LTR (nucleotide 2508) to the last stop codon before the right LTR (nucleotide 6893), realizing that further data may require these limits to be modified.

Inspection of the hypothetical product revealed six sequence motifs closely similar to those in functionally significant regions that are highly conserved in retroviruses and retrotransposons (Fig. 4). In the *gag* product of retroviruses and certain retrotransposons a cysteine-rich motif associated with RNA binding is present once or in duplicate (31). Initially, this could not be located in 1-46, but it was identified in the consensus sequence of the five other *Bgl* II/*Bam*HI fragments discussed above (Fig. 4). Motifs associated with protease (32), reverse transcriptase (16, 21), RNase H (33), and integrase (32, 33) activities of retroviral polyproteins were also identified in *del*, as was a possible HHCC zinc finger in the integrase domain (33).

Fig. 5 shows close conservation in the order and spacing of these motifs in *del* and four representative retroid elements. Differences include the presence of a tether region between the reverse transcriptase and RNase H domains in retroviruses including Moloney murine leukemia virus (Mo-MLV) (33), which is lacking in the other elements (23), and the larger reverse transcriptase region in *Ty3-2*, which is partly due to the addition of 32 amino acids by tandem duplication in this particular element (13). Also an RNA binding motif cannot be located in *17.6* or other gypsy group elements (16–18, 20), and integrase sequences are absent in cauliflower mosaic virus (CaMV) and carnation etched ring virus (CERV) (27, 28).

The most highly conserved region among retroviruses is a 300-amino acid *pol* sequence associated with reverse transcriptase (34). The equivalent region in *del* was compared with that in retrotransposons and other viruses and was found to be most similar to that in *Ty3* of yeast (37% identical) and the gypsy group of *Drosophila* (32–38%) (Table 1). When the RNase H region was compared, similar identities were observed. The only other region that could be readily aligned between elements includes that with integrase function. Here *del* shows closer identity with the sequence in *Ty3* of yeast (40%) than with any members of the gypsy group (21–26%) (Table 1; Fig. 6). Retroid elements showing the next closest relationship with *del* are the caulimoviruses CaMV and CERV, with up to 29% sequence identity, and the *Dictyostelium* element DIRS. Significant levels are also demonstrated by retroviruses typified by Mo-MLV.

DISCUSSION

The *del* element is related to both *Ty3* of yeast and the gypsy group of *Drosophila* but has closer affinities with the former. First, amino acid sequences in the integrase region of *del* and *Ty3* show closer identity (Fig. 6). Second, both *del* and *Ty3* have an RNA binding motif in the retroviral *gag* location that is absent in the gypsy group (Fig. 5). Third, both lack an

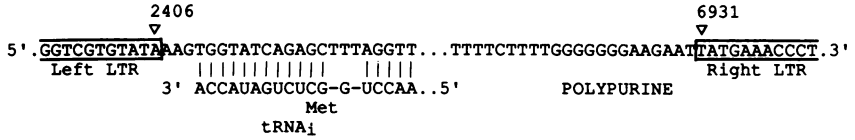

FIG. 2. DNA sequence of the internal region of *del* adjacent to the left and right LTRs showing potential priming sites for DNA synthesis. [Note that the LTR ends of 1-46 differ from the TGT . . . ACA *del* consensus (5).]

```
MVRENDSRSS TGGNVVDPFA QLLAVLQNMT NILHQ*QAAI SHQ*AEQSAR QG*ESRPRLI KEFKGLNPPI FKGDPDPLEA HRWIRHVTKI LDTLGVTDEQ  100
KVILASFQLQ GEAEFWWDAK VRSREDDTTQ IKWDEFVEVF TEKFFPDTVR DDLER*FMTL VQGSLTVAQY EAKFEELSRY APYQVDINIR KVKRFEQGLK  200
LSITKQLSSH LIKDYREVIT RALSVEKREQ REAQIMAKRS KKNRGHNHPY TRKEPHRQ*S NDKSQCPNPL SPGKVCYSYG QPGHFKSNYL TLMAPP*QQH  300
PPPYRPPQNQ PRPPQHQGEG PRVTRVNALI AQDPGVSGTM IRSILSIFSS LCHVLIDTGS THSFITPRII KMLEIPVQPL GYILSVISPI GTSTFVNQVC  400
KGCMITIGNQ ELTVDLIILD LEDPDILLGM DWLAAYHVVL DCFSKKVTFH LPGIPEFHFH GETQHTFFPT FTHQPNLSYL ASLASEINIT PSTDLSLIVR  500
EYINVFPDDL PGLPPPREIE FQINLLPGTS PISITPYHMA PSELQELKEQ LEDLLNKGFI RGSTSPWGAH VLFDPKKDDS KRMCIDY*KL NSVTVKNKYP  600
LPRIDDLFDQ LNGA*YFSKI DLRFRYHQLR IRA*DIPKTA FRTRYGHYEF LVMPFGLTNV PTAFMNLMNR VFREYLDKFI VVFVDYVLIY SRTQKDHEHH  700
LRISLQLLRN NQLYAKLSKC EFWMEKVKFL GHVVSREGIV VDPVKVKAVM NWELPKNIFE IRSFLGLAGY YRRFIKGFSK LAALMTQLTK KGENFNWTKK  800
YQNSFDELKR *LTTVPVLTI PISG*PFVVY TDASLAGLEG VLMQDGRVVA YASRQLKVHE NNYPTHDLEL AVVIFILKLW RHYLYGEDFE LYCDHKSLKY  900
ISTQKDLNLR QR*WIEVLKD FDFSIFYHPG KANVVADALS RKSQISHLIS ARHEFFVTIE GFNLLVRYDS HHTVLCNLRA KPNLINVISD AQRFDSELEA 1000
IHENIIQGKQ DKDWTIDRDN AVRF*GRLVV PLDQDIRTKV LEESHRSKFT IHPGSTKMYR NLKINFWWSG IKREVVEYVS RCLICQQVKA DHHHHSGLLQ 1100
PLPVSE*KWE HILMDFIIGF PLSKRCHDSI WVIVDRFTKS AHFIPIHTTI SGKDLA*LYI KEIIRLHGIP TTIVTDRDTK FTSRFWGSL* KSLGTELFFS 1200
TAFHPQTDG* SERTIQILED MLRSCSLDFK GNWEEHLPLV EFAYNNSYQS SIGMAPFEAL YGRPCRSPTC WAEIGEHHLI RPELIQQTTN AIEVIKRRLK 1300
AAQDRQKSYT DIRRHPLEFS VGNHIFLEVS PRKGTSYFVF KGKLSPRYTG PFEILEIIWP VAYRLALPPM LSSIHNVFHI SMLRKYEPDP SHILDWEDLR 1400
LNPDISYEEK PVQVLASESK VLRNKIILMV KVLWQHHSEE EATWELEADM QE*FPNLFSG M*
```

FIG. 3. The 1462 codons translated from reading frame 3 of the internal DNA sequence of *del* clone 1-46. Stop codons are represented by asterisks. Motifs conserved among retrotransposons and other retroid elements are underlined. 1, RNA binding; 2, protease; 3, reverse transcriptase; 4, RNase H; 5 and 6, integrase (5, possible HHCC Zn finger). The single-letter code for amino acids is used.

	RNA BINDING	PROTEASE	REVERSE TRANSCRIPTASE	RNASE H	INTEGRASE (ZN FINGER?)
	• • •• •	•	••	•• •	• • • • ••
del	CYSCGQPGHFKANC	LIDTGS	IVVFVDDVLI	TDAS	H.6.H.29.CLIC...PQTDGQSER
Ty3-2	CFYCKKEGHRLNEC	LFDSGS	VNVYLDDILI	TDAS	H.5.H.29.CVQC...PQTDGQSER
17.6	-	LIDTGS	CLVYLDDIIV	TDAS	H.4.H.29.CSIC...KTGVADIER
297	-	LLDTGS	CLVYLDDVII	TDAS	H.4.H.29.CNIC...KNGVADVER
gypsy	-	LIDTDA	CYVYVDDVII	TDAS	H.3.H.29.CRVC...SSSNGQVER
412	-	LLDTGA	AFLYMDDLIV	TDAS	H.7.H.29.CQKC...HQTVGVVER
CaMV	CWICNIEGHYANEC	FVDTGA	CCVYVDDILV	TDAS	- -
CERV	CWVCNIEGHYANEC	YVDTGS	CCVYVDDILV	TDAS	- -
DIRS-1	-	-	VIAYLDDLLI	TDAS	- -
Mo-MLV	CAYCKEKGHWAKDC	LVDTGA	LLQYVDDLLL	TDGS	H.3.H.32.CKAC...PQSSGQVER

FIG. 4. Amino acid sequences (single-letter code) of conserved motifs of *del* and 9 other retroid elements shown in N to C terminus order. Residues common to all 10 elements are indicated above with dots. The *del* sequence is from 1-46 except for (*i*) the RNA binding domain, where the Ys at positions 4 and 14 in 1-46 are shown as the consensus Cs that occur in five other independent clones (see text); (*ii*) the reverse transcriptase domain, where the Y at position 7 in clone 1-46 is shown as the usual D that occurs in clone 1-04 (P. Leeton, personal communication); and (*iii*) the integrase domain, where the stop codon in 1-46 is shown as the Q found in clones 1-04 and 4-51 (5). Other sequences are from *Ty3-2* of yeast (13), the four gypsy group elements of *Drosophila* [*17.6* (16), *297* (17), gypsy (18), and *412* (20)], CaMV (27) and CERV (28), the DIRS-1 element of *Dictyostelium* (29), and Mo-MLV (30). Note that in the proposed Zn finger of the *Ty3* integrase another histidine occurs 2 residues in the amino-terminal direction from those shown (13). This may be the N-terminal residue involved in any Zn interaction.

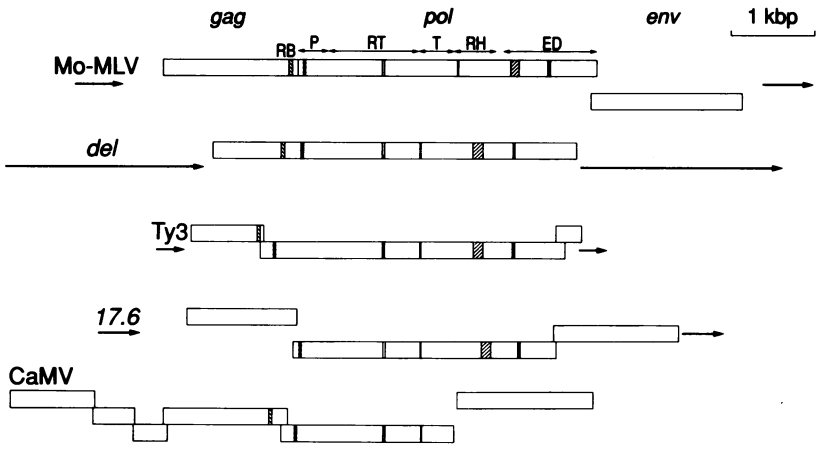

FIG. 5. Comparison of the open reading frames (N termini left, C termini right) of *del* and four other retroid elements showing the locations of conserved motifs (see Fig. 4). LTRs are shown as solid arrows and reading frames are shown as open bars. The extent of domains within the *pol* product of Mo-MLV is from McClure *et al.* (34). RB, RNA binding; P, protease; RT, reverse transcriptase; T, tether; RH, RNase H; ED, endonuclease (integrase).

Table 1. Percentage identity of amino acid sequences in *pol* regions of *del* with those in other retroid elements

Element	Reverse transcriptase, %	RNase H, %	Integrase, %
Ty3	37	37	40
17.6	37	45	21
297	38	39	23
gypsy	34	37	21
412	32	36	26
CaMV	25	29	—
CERV	21	29	—
DIRS	21	28	—
Mo-MLV	23	22	20

The domains covered are for reverse transcriptase (290 residues, from 533–822 in Fig. 3), RNase H (119 residues, from 823–941), and integrase (179 residues, from 1045–1223). Calculations are based on the alignments of Doolittle *et al.* (23) for reverse transcriptase and RNase H and of Hansen *et al.* (13) and Yuki *et al.* (20) [with modifications in the Zn finger region (see Fig. 6)] for the integrase.

additional large open reading frame in the *env* region that is present in three gypsy members (refs. 16–18; Fig. 5). Finally, the putative minus strand primer for both *del* and *Ty3* is methionine initiator tRNA, separated from the left LTR by several residues in each case (Fig. 2; see ref. 13). In the gypsy group, by contrast, the primer sequence abuts, or even overlaps, the LTR (35).

The significance of the greater divergence between the integrase regions of *del* and the gypsy group compared with their reverse transcriptase and RNase H domains requires comment. One possibility is that the integrase is subject to less intense selection pressure. Arguing against this is the observation that the three regions have diverged at similar rates in *del* and *Ty3* and also in *del* and Mo-MLV (Table 1). It is true that McClure *et al.* (34) found that the integrase region changed 1.4 times more rapidly than the reverse transcriptase domain among nine retroviruses. However, they compared the full 300 amino acids of this region, whereas we have confined our analysis to ≈180 amino acids from the N terminus that are more highly conserved.

An alternative scheme accounts for the greater integrase divergence through recombination. Recombination of retroviral sequences is well documented. The *env* region, in particular, is frequently involved, with six rearrangements indicated during the evolution of 17 retroviruses (34). We speculate that *del* and *Ty3* retain the basic *pol* sequence of a common ancestor, but that different integrase and downstream regions were captured during divergence of the gypsy group from this ancestor. This could well have happened more than once as the likely substrate specificity of the *412* integrase differs considerably from that of *17.6, 297,* and gypsy (20, 35), and *412* alone lacks a downstream open reading frame (20). Amino acid sequence comparisons of three *pol* domains also indicate that *412* is an outlier in the gypsy group (21–23).

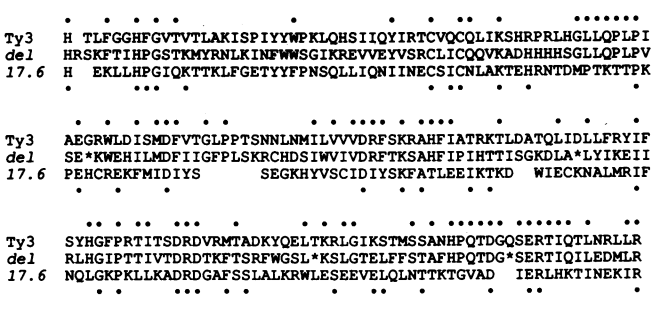

FIG. 6. Alignment of the integrase regions of *del* 1-46, *Ty3-2* (13), and *17.6* (16). Locations where residues are identical between *del* and the adjacent sequence are indicated with dots. The alignment commences at the putative HHCC zinc finger and follows that of Hansen *et al.* (13) except that gaps in the Zn finger have been restricted to the region between the two histidines.

The well-known retrotransposons *copia* of *Drosophila* and *Ty1* of yeast are only distantly related in amino acid sequence to the gypsy group (21–23) and *Ty3* (13), and the same is apparent for *del* sequences. Also, *Ty1* and *copia* have a different order of functional domains within their *pol* sequence, with the integrase region falling between the protease and reverse transcriptase regions (12, 14, 15), apparently reflecting an ancient rearrangement. Interestingly, two newly discovered plant retrotransposons, *Ta1* of *A. thaliana* (36) and *Tnt1* of *Nicotiana tabacum* (37), clearly fall into the *copia* and *Ty1* class.

Retroviruses and retrotransposons share many properties. In fact Doolittle *et al.* (23) have concluded that the only consistent difference is that retrotransposons lack envelope proteins required for autonomous movement between cells and organisms. Even so, it has been proposed that the occurrence of related retrotransposons in widely divergent host species has resulted from their horizontal transmission (21, 23). Two related sets of elements are now known from *Drosophila*, yeast, and plants—the gypsy group, *Ty3*, and *del* of *Lilium* on the one hand, and *copia*, *Ty1*, and the *Arabidopsis* and *Nicotiana* retrotransposons on the other. All these elements seem to lack the ability to escape from their host cell, so their proposed horizontal transmission either relied on internal sequences now lost or involved an independent vector. Among plants potential vectors include the insect-borne caulimoviruses, which share certain phases of the retroviral replication cycle and are themselves likely to have arisen from a retrotransposon-like ancestor in the past (23, 32). Although the direction and timing of proposed horizontal transmission events are difficult to infer, it may be relevant that the reverse transcriptase sequences of the plant viruses CaMV and CERV are more closely related to *Ty3* of yeast [36% and 33% identity (13)] than to *del* [25% and 21% identity (Table 1)].

We cannot tell whether amplification of *del* is still occurring to add to the more than 13,000 copies that already exist in the genome of *L. henryi* (4). Transcripts are required for this process, but none with homology to *del* can be detected so far in young shoots or roots of *L. henryi* (unpublished data). The element we sequenced, 1-46, seems likely to have been amplified long enough ago to have allowed the LTRs to have diverged by 1.4% and the coding region to have accumulated scattered stop codons.

It will be interesting to examine species related to *L. henryi* for the occurrence and organization of *del*. This should allow us to follow its evolutionary history and to assess its role in generating the particularly large genomes of *Lilium* sp.

We thank Marty Yanofsky for help with computer analysis, Russell Doolittle for communicating analyses prior to their publication, and Elliott Meyerowitz for hospitality during preparation of the manuscript. This work was supported by the Australian Research Grants Scheme and Monash University Special Research Grants.

1. Bennett, M. D. & Smith, J. B. (1976) *Philos. Trans. R. Soc. London Ser. B* **274**, 227–274.
2. Leutwiler, L. S., Hough-Evans, B. R. & Meyerowitz, E. M. (1984) *Mol. Gen. Genet.* **194**, 15–23.
3. Flavell, R. B. (1986) *Philos. Trans. R. Soc. London Ser. B* **312**, 227–242.
4. Sentry, J. W. & Smyth, D. R. (1985) *Chromosoma* **92**, 149–155.
5. Sentry, J. W. & Smyth, D. R. (1989) *Mol. Gen. Genet.* **215**, 349–354.
6. Temin, H. M. (1980) *Cell* **21**, 599–600.
7. Finnegan, D. J. (1983) *Nature (London)* **302**, 105–106.
8. Varmus, H. E. (1983) in *Mobile Elements*, ed. Shapiro, J. A. (Academic, New York), pp. 411–503.

9. Baltimore, D. (1985) *Cell* **40,** 481–482.
10. Boeke, J. D., Garfinkel, D. J., Styles, C. A. & Fink, G. R. (1985) *Cell* **40,** 491–500.
11. Eichinger, D. J. & Boecke, J. D. (1988) *Cell* **54,** 955–966.
12. Clare, J. & Farabaugh, P. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 2829–2833.
13. Hansen, L. J., Chalker, D. L. & Sandmeyer, S. B. (1988) *Mol. Cell. Biol.* **8,** 5245–5256.
14. Mount, S. M. & Rubin, G. M. (1985) *Mol. Cell. Biol.* **5,** 1630–1638.
15. Emori, Y., Shiba, T., Kanaya, S., Inouye, S., Yuki, S. & Saigo, K. (1985) *Nature (London)* **315,** 773–776.
16. Saigo, K., Kugimiya, W., Matsuo, Y., Inouye, S., Yoshioka, K. & Yuki, S. (1984) *Nature (London)* **312,** 659–661.
17. Inouye, S., Yuki, S. & Saigo, K. (1986) *Eur. J. Biochem.* **154,** 417–425.
18. Marlor, R. L., Parkhurst, S. M. & Corces, V. G. (1986) *Mol. Cell. Biol.* **6,** 1129–1134.
19. Bayev, A. A., Lyubomirskaya, N. V., Dzhumagaliev, E. B., Ananiev, E. V., Amiantova, I. G. & Ilyin, Y. V. (1984) *Nucleic Acids Res.* **12,** 3707–3723.
20. Yuki, S., Inouye, S., Ishimaru, S. & Saigo, K. (1986) *Eur. J. Biochem.* **158,** 403–410.
21. Yuki, S., Ishimaru, S., Inouye, S. & Saigo, K. (1986) *Nucleic Acids Res.* **14,** 3017–3030.
22. Xiong, Y. & Eickbush, T. H. (1988) *Mol. Biol. Evol.* **5,** 675–690.
23. Doolittle, R. F., Feng, D.-F., Johnson, M. J. & McClure, M. A. (1989) *Q. Rev. Biol.* **64,** 1–30.
24. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
25. Messing, J. (1983) *Methods Enzymol.* **101,** 20–78.
26. Sprinzl, M., Hartman, T., Meissner, F., Moll, J. & Vorderwülbecke, T. (1987) *Nucleic Acids Res.* **15,** Suppl. r53–r188.
27. Franck, A., Guilley, H., Jonard, G., Richards, K. & Hirth, L. (1980) *Cell* **21,** 285–294.
28. Hull, R., Sadler, J. & Longstaff, M. (1986) *EMBO J.* **5,** 3083–3090.
29. Cappello, J., Handelsman, K. & Lodish, H. F. (1985) *Cell* **43,** 105–115.
30. Shinnick, T. M., Lerner, R. A. & Sutcliffe, J. G. (1981) *Nature (London)* **293,** 543–548.
31. Covey, S. N. (1986) *Nucleic Acids Res.* **14,** 623–633.
32. Toh, H., Kikuno, R., Hayashida, H., Miyata, T., Kugimiya, W., Inouye, S., Yuki, S. & Saigo, K. (1985) *EMBO J.* **4,** 1267–1272.
33. Johnson, M. S., McClure, M. A., Feng, D.-F., Gray, J. & Doolittle, R. F. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 7648–7652.
34. McClure, M. A., Johnson, M. S., Feng, D.-F. & Doolittle, R. F. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 2469–2473.
35. Inouye, S., Saigo, K., Yamada, K. & Kuchino, Y. (1986) *Nucleic Acids Res.* **14,** 3031–3043.
36. Voytas, D. F. & Ausubel, F. M., (1988) *Nature (London)* **336,** 242–244.
37. Grandbastien, M.-A., Spielmann, A. & Caboche, M. (1989) *Nature (London)* **337,** 376–380.