Identification of genes for reverse transcriptase-like enzymes in two *Drosophila* retrotransposons, 412 and *gypsy*; a rapid detection method of reverse transcriptase genes using YXDD box probes

Shunji Yuki, Satoshi Ishimaru, Satoshi Inouye and Kaoru Saigo*

Department of Biochemistry, Kyushu University 60 School of Medicine, Fukuoka 812, Japan

Received 4 February 1986; Accepted 4 March 1986

ABSTRACT

Using synthetic oligonucleotides corresponding to the amino acid sequences best conserved among retroviral reverse transcriptases, we developed a rapid method to detect cloned DNA fragments with the genes for reverse transcriptases. By this technique followed by nucleotide sequence determination, nucleotide sequences coding for reverse transcriptase-like enzymes were identified in two *Drosophila* retrotransposons, *gypsy* and 412. Our sequence analysis suggested that (1) there are at least two major groups of retrotransposons in *Drosophila* with respect to putative reverse transcriptases and (2) both *gypsy* and 412 belong to a category of retrotransposons which have putative reverse transcriptases very similar in amino acid sequence to the counterpart of Moloney murine leukaemia virus, a typical mammalian retrovirus.

INTRODUCTION

The genome of *Drosophila melanogaster* contains a large group of repetitive, transposable genetic elements called *copia*-like elements or retrotransposons. They are considered to be related in evolution to retrovirus proviruses in vertebrates (1-13). We found reverse transcriptase (RT)-like enzyme activity associated with virus-like particles whose coat protein is encoded by *copia* and which include genome-sized, unspliced transcripts of *copia* (14, 15). In some *Drosophila* retrotransposons, linear extra-chromosomal DNA and genome-sized DNA-RNA hybrids have been identified which should be the final and intermediate reaction products, respectively, of reverse transcription (16, 17). Boeke *et al.* (18) have shown evidence that RNA of a yeast retrotransposon, *Ty*, similar in structure to *Drosophila* retrotransposons, may possibly be reverse transcribed on translocation. Furthermore, our nucleotide sequence analysis has indicated the genome of 17.6, a *Drosophila* retrotransposon,

to have the gene for RT-like enzyme (19). Thus, reverse transcription may be involved in a significant way in a life cycle of most, if not all, Drosophila retrotransposons.

To confirm and determine further implications of this notion, we developed a simple and rapid method for detecting DNA fragments having genes for RT or RT-like enzymes. By this method, we examined these genes in various Drosophila retrotransposons expected to consist of numerous kinds of species (1, 2). This method is based on the finding that mixed oligonucleotides corresponding to amino acid sequences best conserved among retroviral reverse transcriptases (20) can specifically hybridize to the DNA with RT-related sequences. The nucleotide sequences of the RT genes of gypsy and 412, newly identified here, were also determined and compared with those of the counterparts of other related elements. Retrotransposons were found to have at least two major groups of genes for RT-like enzymes, one including 17.6, 297, gypsy and 412, and the other represented by copia and Ty. Based on these results, the possible evolutionary relationship among these elements is discussed.

MATERIALS AND METHODS

Materials. Restriction enzymes, DNA polymerase, DNA ligase and polynucleotide kinase were obtained from commercial sources. ³²P-labeled nucleotides were purchased from New England Nuclear. Oligonucleotide probes were synthesized using a Model 380A synthesizer (Applied Biosystems).

Bacteria and Phage strains. The origin of the 412 containing plasmid, cDm2042 has been described (21). Our unpublished data showed the 412 in this plasmid to have a restriction map almost identical to that of the authentic 412 (22). gypsy containing phage, λbx^{34e}, (23) was obtained from C. Langley and the DNA segment covering the entire internal region of gypsy was subcloned using pUC9 as a cloning vehicle. Plasmids containing the entire regions of copia, 17.6 and 297 were subcloned into pUC9 using the transposon-containing DNA fragments derived from λcop88 (15), λhist17.6 (19, 21) and λhisC (21), respectively. After purified, these plasmid DNAs were further subcloned using the pUC9 vector as described below (see Fig.1A).

General Methods. Restriction enzyme digestions, agarose gel electrophoresis, and the subcloning of restriction segments were carried out essentially as described by Maniatis *et al.* (24). Synthetic DNA probes were labelled with ³²P by kination. The nucleotide sequences were determined using a modified dideoxy-chain termination method (15, 25) and analyzed essentially as described previously.

Dot blot hybridization. After about 0.1 ug of plasmid DNA were spotted on nitrocellulose filter, hybridization was carried out using a synthetic oligonucleotide as a probe in 6 X SSC (SSC: 0.15M NaCl + 0.015M sodium citrate) supplemented with 10 X Denhardt's solution, 0.2% sodium dodecyl sulfate, and 200 ug/ml of salmon sperm DNA at 34° C for 20 h (26). The filters were washed twice for 10 min in 6 X SSC at 18°C.

RESULTS AND DISCUSSION

Identification of genes for RTs or RT-like enzymes by dot blot hybridization

The amino acid sequences, Tyr (Met, Val or Leu) Asp Asp (hydrophobic amino acid)₃, which will be hereafter referred to as YXDD boxes, have been shown to be present in the RTs of all retroviruses and hepatitis viruses so far examined (20, 27 ; Table 1A). Similar YXDD boxes have also been identified in the

TABLE 1

(A) Amino acid sequences of YXDD boxes from various sources.

Group		A.A.Seq.	Type	Ref.
Retrovirus	Mo-MuLV	YVDDLLL	4	37
	RSV	YMDDLIL	3	38
	MMTV	YMDDILL	3	39
	HTLV-I	YMDDILL	3	40
	HTLV-II	YMDDILL	3	41
	HTLV-III	YMDDLIV	3	42
	BLV	YMDDILY	3	43
	WHV	YMDDLVL	3	44
	HBV	YMDDVVL	3	45
	CAEV	YMDDLIV	3	39
	EIAV	YMDDIYI	3	39
	IAP	YMDDILI	3	46
Plant DNA virus	CaMV	YVDDILV	4	30
Drosophila retrotransposon	297	YLDDIII	1	28
	17.6	YLDDIIV	2	19
	copia	YVDDVI	4	15,31

(B) Nucleotide sequences of mixed probes of YXDD boxes.

	A.A.Seq.	N.A.Seq.
Type 1	YLDDZ	TAY CTN GAY GAY NT
Type 2	YLDDZ	TAY TTR GAY GAY NT
Type 3	YMDDZ	TAY ATG GAY GAY NT
Type 4	YVDDZ	TAY GTN GAY GAY NT

Z;L,V or I. Y;C+T. R;A+G. N;A+G+C+T.

counterparts in a plant DNA virus , cauliflower mosaic virus (CaMV) and certain Drosophila retrotransposons (15, 19, 20, 28; see Table 1A). Thus, the YXDD box may consequently be essential to RT activity. We made four sets of mixtures of synthetic deoxyoligonucleotides corresponding to the first five amino acids of YXDD boxes (Table 1B) and tested to see if these oligomers could be used as effective probes for detecting of a DNA fragment having a YXDD box and hence a major portion of the gene for RT or RT-like enzyme. As model systems, we chose and examined five Drosophila retrotransposons already cloned. The entire nucleotide sequences of three of them 17.6 (19), 297 (28) and copia (15) have been determined and the presence of YXDD boxes in their RT-like enzymes have been verified (see Table 1A). However, whether remaining two 412 and gypsy contain similar genes remains to be clarified.

The genome DNA of each transposon was subcloned into four pieces using pUC9 as a cloning vehicle (Fig. 1A). After being spotted on nitrocellulose filters, these subcloned fragments with pUC9 sequences were hybridized with the four different types of YXDD box probes shown in Table 1B. Each probe gave one or two highly positive signal(s) (Fig. 1B). That is, YXDD box probes types 1, 2 and 3 hybridized to the fragments, number 2 of 297, 2 of 17.6 and 2 of 412, respectively, whereas the fragments number 3 of copia and 2 of gypsy were hybridized by type 4 YXDD box probe. As shown by the arrowheads in Fig. 1A and summarized in Table 1, these results are fully consistent with those expected from our nucleotide sequence data at least in the cases of 17.6 (19), copia (15) and 297 (28). Less significant hybridization signals, occasionally detectable (see Fig. 1B), were found to be attributable to cross-hybridization among YXDD box sequences. Furthermore, as shown below, the number 2 fragments of 412 and gypsy were found to include the sequences coding for types 3 and 4, respectively, of the YXDD boxes of RT-like enzymes. Thus, it is concluded that YXDD box probes are very useful hybridization probes for rapidly identifying DNA fragments including the coding regions for RTs or RT-like enzymes at least in cloned viral or retrotransposon genomes.

Fig. 1 (A) Restriction maps of five *Drosophila* retrotransposons and locations of the subdivided fragments. Arrowed horizontal lines labelled 1, 2, 3 and 4 show the size and locations of the subdivided fragments. Vertical arrowheads show fragments each containing the nucleotide sequence corresponding to the YXDD boxes of reverse transcriptase-like enzymes. (B) Dot blot hybridization patterns obtained when four sets of oligonucleotides of YXDD boxes were used as probes. Numerals 1, 2, 3 and 4 show the locations of the spots of the subdivided regions of each retrotransposon, whereas panels a, b, c and d, respectively, show the hybridization patterns obtained using types 1, 2, 3 and 4 probes.

[illegible]

GCTTCGGGAACCCAAACAGAGGTTGGTCACTTACCTTACGGAAGCTAAATGAGAAACTATTCCTGACCGGTACCGGATGCCTAGCATT 720
 A S G N P N K R L V I D F R K L N E K T I P D R Y P M P S I
 CCCATGATCTAGCGAACTCGGCAAGGCAAGTCTTCACTACCTTGATCTTAAGTCAGGGTATCATCAAAATTTACCTCGCGGAACAC 810
 P M I L A N L G K A K F F T T L D L K S G Y H Q I Y L A E H
 GACCCGGAAGACATCGTTCTCGGTGAAATGGTGGTAAATACGAGTCTTTCGGCTACCGTTCGGCTGAGAAATCGAAGCAGCATTTTT 900
 D R E K T S F S V N G G K Y E F C R L P F G L R N A S S I F
 CAAAGAGCCCTAGACGATGCTCTAGAGAGCAATCGGGAAGATATGTTACGCTATGTAGATGACGCTCAATATTTCTCTGAAACGAG 990
 Q R A L D D V L R E Q I G K I C Y V Y V D D V I I F S E N E
 b
 TCCGACCATGTCGCCACATCGATACAGTACTAAAAATGCCTGATCGATGCCAACATGAGAGTAGCGAGGAGAAAACATAGATTCTTTAAA 1080
 S G H V R H I D T V L K C L I D A N M R V S Q E K T R F F K
 GAGAGTGTAGAAATACCTCGGCTTTATTTGTCAGTACGGAGCAACTAAATCCGATCCAGAGAAAGTGGAAGCCATTCAGGAGTACCTGAA 1170
 E S V E Y L G F I V S K D G T K S D P E K V K A I Q E Y P
 CCAGACTCGTTTACAGGTAGGTCTCTCTGTTTGTAGCCAGCTACTACAGAGTCTTCAAGAGCTTTGCTGCCATAGCCGCGCGC 1260
 P D C V Y K V R S F L G L A S Y V R V F I R D F A A I A R P
 ATCACCGATATCTTAAAGGGGAAATGGTTCGGTGAGCAACACATGCTCTAAAAAATTCCTGTTAGTTTAAAGCTCAACGCAAC 1350
 I T D I L K G E N G S V S R H N S K I P V E F N E T
 CGGTCCAAAGACTCGCAACATAGTACGATCCGAGGATGCTACTCAAAATACCCGCTCTTAAAAAGCCCTTTGACCTTACTACGAT 1440
 A F Q R L R N I L A S E D V I L K Y P V F K K P F D L T T D
 GCTTCGGCAAGTGGTATCGGTGTCAGTCTATCCGAGGAGGCGCAATCACCATGATATCCGCTACCCCTTAAACAGCCGAGCAGAAC 1530
 A S A S G I G A V L S Q E G R P I T M I S R T L K Q P E Q N
 TAGCCCAACAGGAAAGGAAATGCTGGCGATGTATGGGCCCTAGGTAAAGTTCGAGAAGTCTGCTAGGCTAGGAGGATTAATATA 1620
 Y A T N E L L A I V A L G K L Q N F L Y G S R E I N I
 TTTACCGACCATCAACCCCTCACTTTCCGCTGTTCGCGACAGGAACAGAAATGCCAAGATAAAGAGGTGGAAATCTTACATAGACCCGACAT 1710
 F T D H Q P L T F A V A D R M T N A K I K R W K S Y I D Q H
 AATGGCAAGTTTCTACAAACCTGGCAAGAAATTTGCTGGCAGACGCCCTCTCTAGGCAGAAATCTGAAATGCCTTACAAAGCAACCC 1800
 N A K V F Y K F K E M F V A D A L S R Q N L N A L Q N E P
 CAATCAGACGCTCGGACCATTCACAGTAGGCTCTCCCTGACCTACACGGTTCGAGACCAACAGCAACCGTTAAATTTGCTTCAGGAACAG 1890
 Q S D A A T A T S E L S I T Y T V E T T D K P L N C F R N Q
 ATCATCTTGAAGCAGCAGCTTTCCGCTCAACGGAACCTGGTCTCTTCGAGCAAACTTCGCCACCTTGATCAGCTTTACTGATATAA 1980
 I I L E A A R F P L K R M L V L F R S K S R H L I S F T D K
 AGTTGGCTATTAACCACTTAAGGAGGTGGTAACCTGACGTCGTGAACGCTATTCAGTCCGACCTGCCCACTCTGGCAGCTTCCAA 2070
 S V L L K T L K E V V N P D V V N A I H C D L P T L A S F Q
 Hind III
 Fragment 2
 CACGACCTCATGCCCACTTTCCAGCAACCCAAATTTGTCAGTGAAGAATGTGTTAGACATAACCGACAAAAAGCAACAGATCGAA 2160
 H D L I A N F P A T Q F R H C K N V V L D I T D R K N E Q I E
 ATGCTCACTGCTGAGCACAACCGGCTCACAGAGCGGCACAAGAAACATTAACAAAGTCCTTCGGGATATTACTTTCCCAAAATTTGGC 2250
 I V T A E H N R A H R A A Q E N I K Q V L R D Y V F P K M G
 AGTTTAGTCAAGAAGTAGTACTAATTTGTAGGCTGTGACCCCAAGCAAGTATGACAGGCAACCGAAAAAGCAAGAGCTCGGGGAAACG 2340
 S L A K E V A N A N C R V C T Q A K Y D R H P K K Q E L G E T
 CCCATACCCAGCTATACAGGTGAGATGGTGTACATTTGACATTTCTCAACCGACAGGAAGCTATTCCTGAGCTGTATTGACAAATTTTCT 2430
 P I P S Y T G E N V H I D I F S T D R K L F L T C I D K F S
 AAATATGCAATAGTCAACCAAGTGGTCTAGACAAATAGTGGACATACACGACCCCTGTTGCAGATCATTAACCTTTCCCAATATC 2520
 K Y A I V Q P V V S R T I V D I T A P L L Q I M L F P M I
 AAAACGGTCTATTGTACAAATGAGCCCGCATTTAACTCAGAAACTGTCACCTCAATGCTCAAGAAGACCTTCGGCATTTGACATAGTAAAT 2610
 K T V Y C D N E P A S E T V T S M L K N T F G I D I V N
 GCGCCCGCCTCAGCAGCTCGTCCAAATGGCAAGTTGACAGGTTCCACAGCACATTCGCAGAAATTCGCGAGGTGCTGAAAGTTGGACAAA 2700
 A P P L H S S S N G G V E R C F H S T L A E I A R C L K L D K
 c
 AAAAGAGTGCACAGTAGAACTAATCTTGAAGGCGCAGATAGAATATAACAAACCGTGACACTCAGTTACTGCTGAGAACCAATTGAGG 2790
 K T S D T V E L I L R A T T I G A T T E Y N K T V S V T R E N G L R

Fig. 2 Nucleotide sequences of the DNA fragments of 412 (A) and *gypsy* (B). Amino acid sequences corresponding to the longest ORFs were depicted just beneath the nucleotide sequences. Underlines labelled a, b, and c, respectively, show the amino acid sequences similar to those best conserved of retroviral protease, reverse transcriptase, and integrase. L-shaped arrowheads indicate the size and locations of DNA fragments hybridizable to YXDD box probes. H, cutting site of *Hind*III; E, *Eco*RI cutting site.

Detection and nucleotide sequence determination of genes for RT-like enzymes in two *Drosophila* transposons, 412 and *gypsy*

To determine whether the strong hybridization signals detected in the number 2 fragments of 412 and *gypsy*, are related to the YXDD boxes of RTs or RT-like enzymes, the nucleotide sequences of these fragments along with those of neighbouring fragments were determined (Fig. 2). Both 412 and *gypsy* were each found to include a very long open reading frame (ORF) covering

Fig. 3 Partial amino acid sequence alignments of polypeptides expected to be derived from *pol* or *pol*-like genes of various elements. Line 1, Mo-MuLV (37); line 2, CaMV (30); line 3, 17.6 (19); line 4, 297 (28); line 5, gypsy (this work); line 6, 412 (this work); line 7, copia (15); line 8, Ty (33). Essential parts of amino acid sequence alignments among Mo-MuLV, CaMV, 17.6 have been described (19, 20), and the correspondence of the amino acid sequences between copia and Ty is essentially the same as that described by Mount and Rubin (31). Amino acid sequence similarity among 17.6, 297, gypsy and 412 are shown by boxes, whereas the conserved and similar amino acids in all the present elements examined are labelled by the large filled and open circles, respectively. Small filled circles indicate similarity in more than 70% of the elements. Chemically similar amino acids are defined as pairs of residues, both belonging to the same group (47). The groups are as follows: A, S, T, P and G; N, D, E and Q; H, R and K; M, L, I and V; F, Y and W. The filled

and open triangles show invariant amino acids among retroviruses and 17.6. Note that the filled triangles show the amino acids conserved in all elements examined here. The best conserved sequences of protease and reverse transcriptase in retroviruses are shown by underlines a and b, respectively. The large, vertical arrowheads A and B, respectively, indicate the amino terminus of the processed Mo-MuLV RT (29) and carboxyl end of CaMV RT (30). The L-shaped arrowheads P and RT, respectively, show the presumed protease and reverse transcriptase domains. Numerals flanked by dotted lines show the numbers of amino acid residues which could not be aligned. In the case of copia and Ty, putative pol genes were divided into at least three segments because of the presumed translocation of integrase domains (31, 32, 33, 34). The positions of the terminal amino acids of each segment are thus shown by numerals with asterisks. *1, 366 amino acids (AA) from the amino terminus of the polypeptide expected to be derived from ORFg of copia (15); *2, 866 AA from the amino terminus of copia ORFg polypeptide; *3, 111 AA from the amino terminus of the polypeptide expected to be derived from the second ORF of Ty (33); *4, 814 AA from the amino terminus of Ty pol-like ORF product. The amino and carboxyl termini of the polypeptides expected to be produced from pol of Mo-MuLV and/or pol-like genes of other elements are indicated by n and c, respectively.

the entire regions sequenced here. The longest ORFs were thus translated and their amino acid sequences are depicted just beneath the corresponding nucleotide sequences in Fig. 2.

It is evident that the amino acid sequences corresponding to the fragments, number 2 of 412 and 2 of gypsy, respectively, include type 3 and type 4 YXDD boxes, whose presence were predicted by dot blot hybridization experiments described above. (see both L-shaped arrowheads and underlines labelled b in Fig. 2). Furthermore, as shown in Fig. 3 (lines 1-6), the amino acid sequences in the vicinity of the YXDD boxes of gypsy and 412, whose locations are indicated by the underline labelled b, were found to be very similar to those of the counterparts (RT regions (19, 20)) of 17.6, 297, CaMV and Moloney murine leukaemia virus (Mo-MuLV). Similar amino acid sequences in all of 17.6, 297, gypsy and 412 are shown by boxes in Fig.3. Although the degree of homology was much reduced, a similar correspondance of amino acids was also found in copia and Ty (see lines 7 and 8). As a result, 5 of 9 amino acids invariant in all RTs of retroviruses and RT-like enzyme of 17.6 (19, 20, 27) were noted to be preserved in all the Drosophila and yeast

	297	17.6	GYPSY	412	CaMV	Mo-MuLV	COPIA	Ty	average
297	---	---	---	---	---	---	---	---	1
17.6	0.74	---	---	---	---	---	---	---	0.74
GYPSY	0.43	0.40	---	---	---	---	---	---	0.42 ± 0.01
412	0.39	0.41	0.35	---	---	---	---	---	0.39 ± 0.02
CaMV	0.30	0.30	0.25	0.27	---	---	---	---	0.28 ± 0.02
Mo-MuLV	0.23	0.24	0.20	0.19	0.20	---	---	---	0.21 ± 0.01
COPIA	0.12	0.14	0.13	0.14	0.10	0.08	---	---	0.12 ± 0.02
Ty	(0.12)	(0.09)	(0.09)	(0.10)	(0.10)	(0.10)	(0.19)*	---	(0.10) ± (0.02)

A

Fig. 4 (A) Amino acid sequence homology in the RT regions of various elements. Calculation was made using the entire RT regions defined in Fig. 3 except for small regions where deletions can be seen in at least three elements. In the case of homology to Ty, two additional regions located at the amino terminus and in the middle, where amino acid sequences of Ty could not be aligned, were omitted for the calculation. Thus, the homology to Ty appears in parenthesis. The asterisk shows the unusually high value found between copia and Ty. (B) A dendrogram of homology of RT-related sequences. The positions of 297, 17.6, gypsy, 412, CaMV, Mo-MuLV, copia and Ty were determined based on the calculation in (A). The positions of retroviruses other than Mo-MuLV and hepatitis viruses were then superimposed using the data described previously (20). Thus, their branches are labelled by dotted lines. The dotted line for Ty indicates the difference in mode of calculation as described in (A).

retrotransposons examined here (see filled and open triangles in Fig.3). Interestingly, in all cases, the degree of homology undergoes great reduction at or near the two critical points labeled A and B in Fig.3, which, respectively, correspond to the amino terminus of the processed RT of Mo-MuLV (29) and the carboxyl terminus of an ORF 5 polypeptide or RT of CaMV (30). It was thus concluded that, as with the other Drosophila

retrotransposons so far examined, both gypsy and 412 genomes include genes coding for RT-like enzymes having similar amino acid sequences and related in evolution to the RTs of CaMV and Mo-MuLV.

One additional finding from amino acid sequence analysis is that ORFs with genes encoding RT-like enzymes of gypsy and 412 have structures very similar to those of the counterparts of 17.6 and 297 and the pol gene of Mo-MuLV, since conserved sequences for protease and integrase can be detected in the flanking regions on the amino and carboxyl terminal sides of the RT-like enzyme domains of gypsy and/or 412 (Figs. 1 and 3). However, as suggested previously (31, 32, 33, 34) in contrast to these elements, the putative integrase domains of copia and Ty seem to be translocated between putative protease and RT-like enzyme domains.

Evolution of genes for RT and RT-like enzymes

Using the aligned amino acid sequences of the entire regions of RT and RT-like enzymes (the regions delimited by two points labelled A and B in Fig. 3), the mutual homology of the amino acid sequences was calculated as shown in Fig. 4A. Based on these data and those published previously (20), we constructed a dendrogram of homology possibly representing the phylogenetic relationships among RT-related sequences from Drosophila and yeast retrotransposons, Mo-MuLV and other retroviruses, hepatitis viruses and CaMV (Fig. 4B). Homology to 297 was found to decrease in the following order: 17.6 > gypsy, 412 > CaMV > Mo-MuLV > copia, Ty. 17.6, gypsy, 412, CaMV, Mo-MuLV, and copia (or Ty) may possibly have evolved almost independently from each other after diverging from 297. The diversity of RT-related sequences in retrotransposons is much greater than that in retroviruses. In fact, the RT genes of all retroviruses so far examined appeared to form a single branch in our dendrogram (Fig. 4B). In contrast, RT-related sequences in retrotransposons consist of at least two major groups, one represented by 17.6, 297, gypsy and 412 (group I), and the other by copia and Ty (group II). The homologies of group I retrotransposons to CaMV and Mo-MuLV is as high as 28 and 21%, respectively, whereas that to group II retrotransposons are only

10% (Fig. 4A), suggesting group I retrotransposons to bear a much closer relationship to retrovirus proviruses in vertebrates and CaMV than group II retrotransposons. Taking into consideration the fact that retrotransposons have genome structures similar to those of the proviruses of replication-competent retroviruses, the above findings may be evidence that retroviruses in vertebrates evolved from an element similar to present-day retrotransposons (3). Alternatively, RT genes may be introduced from retrovirus-like elements into the genomes of retrotransposons at least by two independent, evolutionary events.

The similarity in sequence between copia and Ty was originally suggested by Mount and Rubin (31). Our present results confirm this notion. Our dendrogram shows that, in spite of their presence in two different organisms, Drosophila and yeast, the divergence of copia and Ty occurred much more recently than those between copia and any other Drosophila retrotransposons so far examined. Thus, there is the possibility that Ty originated in an ancestor of copia transferred from Drosophila to yeast by an unknown mechanism during evolution.

Another interesting feature of our dendrogram is that the RT gene of 412 is intimately related in evolution to that of 297, since our recent observations showed the genes for putative primer tRNAs for reverse transcription of 412 and 297, which, respectively, turned out to be arginine and serine tRNA, to have the nucleotide sequences both similar to and located at a chromosomal loci almost identical to each other (unpublished). 17.6, 297 and gypsy have been shown to belong to a peculiar set of transposable elements whose insertion takes place in a site-specific fashion (35, 36), whereas the insertion of 412 seems to occur without apparent site-specificity (4). Thus, RT-genes themselves may not be related to the insertional steps of these retrotransposons.

ACKNOWLEDGEMENTS

The authors express their appreciation to Prof. M. Sekiguchi, Drs. T. Shiba, Y. Sano and T. Miyata for their critical comments. This work was supported in part by grants from the Ministry of Education, Science and Culture, Japan to K.S.

*To whom correspondence should be addressed

REFERENCES

1. Rubin, G.M., Broreign, W.J., Jr., Dunsmuir, P., Flavell, A.J., Levis, R., Strobil, E., Toole, J.J. and Young, E. (1981) Cold Spring dHarbor Symp. Quant. Biol. 45, 619-628.
2. Spradling, A.C. and Rubin, G.M. (1981) Ann. Rev. Genetics 15, 219-264.
3. Temin, H.M. (1980) Cell 21, 599-600.
4. Will, B.M., Bayev, A.A. and Finnegan, D.J. (1981) J. Mol. Biol. 153, 897-915.
5. Kulguslin, V.V., Ilyin, Y.V. and Georgiev, G.P. (1981) Nucl. Acids Res. 9, 3451-3464.
6. Flavell, A.J. and Ish-Horowicz, D. (1981) Nature 292, 591-595.
7. Georgiev, G.P., Ilyin, Y.V., Chemeliavskaitė, V.G., Ryskov, A.P., Kramerov, D.A., Skryabin, K.G., Krayev, A.S., Lukanidin, E.M. and Grigoryan, M.S. (1981) Cold Spring Harbor Symp. Quant. Biol. 45, 642-654.
8. Tchurikov, N.A., Ilyin, Y.V., Skryabin, K.G., Ananiev, E.V., Bayev, A.A., Jr., Krayev, A.S., Zelentsova, E.S., Kulguskin, V.V., Lyubomirskaya, N.V. and Georgiev, G.P. (1981) Cold Spring Harbor Symp. Quant. Biol. 45, 655-665.
9. Skalka, A., Ju, G., Hishinuma, F., DeBona, P.J. and Astrin, S. (1981) Cold Spring Harbor Symp. Quant. Biol. 45, 739-746.
10. Ikenaga, H. and Saigo, K. (1982) Proc. Natl. Acad. Sci. USA 79, 4143-4147.
11. Snyder, M., Kimbrell, D., Hunkapiller, M., Hill, R., Fristrom, J. and Davidson, N. (1982) Proc. Natl. Acad. Sci. USA 79, 7430-7434.
12. Scherer, G., Tschudi, C., Perera, J., Delius, H. and Pirrota, V. (1982) J. Mol. Biol. 157, 435-451.
13. Kugimiya, W., Ikenaga, H. and Saigo, K. (1983) Proc. Natl. Acad. Sci. USA 80, 3193-3197.
14. Shiba, T. and Saigo, K. (1983) Nature 302, 119-124.
15. Emori, Y., Shiba, T., Kanaya, S., Inouye, S., Yuki, S. and Saigo, K. (1985) Nature 315, 773-776.
16. Flavell, A.J. (1984) Nature 310, 514-516.
17. Arkhipova, I.R., Gorelova, T.V., Ilyin, Y.V. and Schuppe, N.G. (1984) Nucl. Acids Res. 12, 7433-7548.
18. Boeke, J.D., Garfinkel, D.J., Styles, C.A. and Fink, G.R. (1985) Cell 40, 491-500.
19. Saigo, K., Kugimiya, W., Matsuo, Y., Inouye, S., Yoshioka, K. and Yuki, S. (1984) Nature 312, 659-661.
20. Toh, H., Kikuno, R., Hayashida, H. and Miyata, T., Kugimiya, W., Inouye, S., Yuki, S. and Saigo, K. (1985) EMBO J. 4, 1267-1272.
21. Saigo, K., Millstein, L. and Thomas, C.A., Jr. (1981) Cold Spring Harbor Symp. Quant. Biol. 45, 815-827.
22. Shepherd, B.M. and Finnegan, D.J. (1984) J. Mol. Biol. 180, 21-40.
23. Bender, W., Akam, M., Karch, F., Beachy, P.A., Peifer, M., Spierer, P., Lewis, E.B. and Hogness, D.S. (1983) Science 221, 23-29.
24. Maniatis, T., Fritsch, E. F. and Sambrook, J. (1982) Molecular

- cloning, A Laboratory manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
25. Hattori, M. and Sakaki, Y. Anal. Biochem. in press.
26. Wallace, R.B., Johnson, M.J., Hirose, T., Miyake, T., Kawashima, E.H. and Itakura, K. (1981) Nucl. Acids Res. 9, 879-894.
27. Toh, H., Hayashida, H. and Miyata, T. (1983) Nature 315, 691-692.
28. Inouye, S., Yuki, S. and Saigo, K. Eur. J. Biochem. in press.
29. Tanese, N., Roth, M. and Goff, S.P. (1985) Proc. Natl. Acad. Sci. USA 82, 4944-4948.
30. Gardner, R.C., Howarth, A.J., Hahn, P., Brown-Luedi, M., Shepherd, R.J. and Messing, J. (1981) Nucl. Acids Res. 9, 2871-2888.
31. Mount, S.M. and Rubin, G.M. (1985) Mol. Cell. Biol. 5, 1630-1638.
32. Saigo, K. Adv. Biophys. 21, in press.
33. Hauber, J., Hochstetter, P.-N. and Feldman, H. (1985) Nucl. Acids Res. 13, 2745-2758.
34. Clare, J. and Farabaugh, P. (1985) Proc. Natl. Acad. Sci. USA 82, 2829-2833.
35. Inouye, S., Yuki, S. and Saigo, K. (1984) Nature 310, 332-333.
36. Freund, R. and Meselson, M. (1984). Proc. Natl. Acad. Sci. USA 81, 4462-4464.
37. Shinnick, T.M., Lerner, R.A. and Sutcliffe, J.G. (1981) Nature 293, 543-548.
38. Schwartz, D.E., Trizard, R. and Gilbert, W. (1983) Cell 32, 853-869.
39. Cui, I.-M., Yaniv, A., Dahlberg, J.E., Gazit, A., Skuntz, S.F., Tronick, S.R. and Aaronson, S.A. (1985) Nature 317, 366-368.
40. Seiki, M., Hattori, S., Hirayama, Y. and Yoshida, M. (1984) Proc. Natl. Acad. Sci. USA 80, 3618-3622.
41. Shimotono, K., Takahashi, Y., Shimizu, N., Gojobori, T., Golde, D.W., Cheu, I.S.Y., Miwa, M. and Sugimura, T. (1985) Proc. Natl. Acad. Sci. USA 81, 1079-1083.
42. Ratner, L., Haseltine, W., Patarca, R., Livak, K.J., Starcich, B., Josephs, S., Doran, E.R., Rafalski, J.A., Whitehorn, E.A., Baumeister, K., Ivanoff, L., Petteway, S.R., Jr., Pearson, M.L., Lautenberger, J.A., Papas, T.S., Ghrayeb, J., Chang, N.T., Gallo, R.C., and Wong-Staal, F. (1985) Nature 313, 277-284.
43. Sagata, N., Yasunaga, T., Tsuzuki-Kawamura, J., Ohishi, K., Ogawa, Y. and Ikawa, Y. (1985) Proc. Natl. Acad. Sci. USA 82, 677-681.
44. Galibert, F., Chen, T.N. and Mandart, E. (1982) J. Virol. 41, 51-65.
45. Ono, Y., Onda, H., Sasada, R., Igarashi, K., Sugino, Y. and Nishioka, K. (1983) Nucl. Acids Res. 11, 1747-1757.
46. Ono, M., Toh, H., Miyata, T. and Awaya, T. (1985) J. Virol. 55, 387-394.
47. Schwartz, R.M. and Dayhoff, M.O. (1978) in Atlas of protein sequence and structure, vol.5, pp353-358, National Biomedical Research Foundation, Washington.