***R1* and *R2* Retrotransposable Elements of *Drosophila* Evolve at Rates Similar to those of Nuclear Genes**

Danna G. Eickbush, Warren C. Lathe III, Maria P. Francino and Thomas H. Eickbush

Department of Biology, University of Rochester, Rochester, New York 14627

Manuscript received April 15, 1994

Accepted for publication October 20, 1994

ABSTRACT

The non-long-terminal repeat retrotransposable elements, *R1* and *R2*, insert at unique locations in the 28S ribosomal RNA genes of insects. Based on the nucleotide sequences of these elements in the eight members of the *melanogaster* species subgroup of the genus *Drosophila*, they have been maintained by vertical germline transmission for the 17–20 million year history of this subgroup. The stable inheritance of *R1* and *R2* within these species has enabled a determination of their nucleotide substitution rates. The sequence of the *R1* and *R2* elements from *D. ambigua*, a member of the *obscura* species group, has also been determined to enable an extrapolation of this rate over an estimated 45–60 million years. The mean rate of substitutions at synonymous sites (K_s) was 6.6 and 9.6 times the rate at replacement sites (K_a) in the *R1* and *R2* elements, respectively. Both elements appear to have been under selective pressure to maintain their open reading frames and thus their ability to retrotranspose for most of their evolution in these lineages. Using the rate of change at synonymous sites (K_s) as the best indicator of the nucleotide substitution rate, the mean K_s values for *R1* and *R2* were 2.3 and 2.2 times that of the alcohol dehydrogenase (*Adh*) genes. However, this faster rate is a result of the lower codon usage bias of *R1* and *R2* compared with that of *Adh*. When the K_s rates of *R1* and *R2* were compared with that of a larger number of nuclear genes available from at least two of the nine species under investigation, *R1* and *R2* were found to evolve in most lineages at rates similar to that of nuclear genes with low codon bias. The ability of *R1* and *R2* to maintain their presence in this species subgroup by retrotransposition while exhibiting rates of nucleotide evolution similar to nuclear genes suggests these transposition events are rare or not as error prone as that of retroviruses.

RETROVIRAL gene sequences evolve 10^5 – 10^6 times faster than nuclear genes (reviewed in DOOLITTLE *et al.* 1989). This high substitution rate has been correlated with nucleotide misincorporation rates of retroviral reverse transcriptases that are 10^4 – 10^5 times greater than that of DNA polymerases (PRESTON *et al.* 1988; ROBERTS *et al.* 1988; LI and LOEB 1992; DRAKE 1993). However, these rates lead to predicted sequence divergences between retroviruses that are greater than observed divergences based on estimates of their time of entrance into various host genomes. Therefore these rapid bursts of evolution must be short lived. During their extensive periods of latency, the viral genomes have probably evolved by vertical modes of transmission through the germline (DOOLITTLE *et al.* 1989).

Estimating the rates of nucleotide sequence change for retrotransposable elements should be more straightforward because their propagation is tied to the germline of the host. Unfortunately, few examples are known of retrotransposable elements that have been maintained in a species lineage by strict vertical transmission.

Corresponding author: Thomas H. Eickbush, Department of Biology, University of Rochester, Rochester, NY 14627.
E-mail: teickbush@the.biology.rochester.edu

Only in the case of the *Line 1* (*L1*) elements in mammals has extensive sequence comparisons of elements from different rodent species suggested that these elements are stable, thus allowing comparison of the rate of nucleotide sequence evolution of a retrotransposable element with that of nuclear genes. The hundreds of thousands of *L1* elements, the vast majority of which are inactive, have been estimated to evolve with an apparent nucleotide substitution rate similar to or somewhat faster than that of nuclear pseudogenes (HARDIES *et al.* 1986; PASCALE *et al.* 1993).

R1 and *R2* are non-long-terminal repeat (non-LTR) retrotransposable elements found inserted in a fraction of the 28S ribosomal RNA (rRNA) genes in most insects (JAKUBCZAK *et al.* 1991; BURKE *et al.* 1993). These elements insert into different sequences of the 28S genes, have different organizations of their open reading frames (ORFs) and show no greater sequence similarity to each other than to any other non-LTR retrotransposable elements (BURKE *et al.* 1987; XIONG and EICKBUSH 1988a; JAKUBCZAK *et al.* 1990). All available data suggest that *R1* and *R2* insert independently of one another (XIONG *et al.* 1988; JAKUBCZAK *et al.* 1992). They apparently represent separate elements that have become

FIGURE 1.—Summary of *R1* and *R2* evolution in the *melanogaster* species subgroup of *Drosophila*. Shown is the consensus species phylogeny based on the nucleotide sequences of nuclear DNA and mitochondrial DNAs, and chromosomal inversion patterns (LACHAISE *et al.* 1988). The *obscura* and *melanogaster* species groups are believed to be separated by 45–60 million years. The age of the *melanogaster* subgroup itself is 17–20 million years, with the separation of *D. melanogaster* from *Drosophila simulans*, *Drosophila mauritiana* and *Drosophila sechellia* being ~2.5 million years. The phylogeny of *R1* and *R2* follows that of the species phylogeny (EICKBUSH and EICKBUSH 1994). *R1* and *R2* elements have been stable and have evolved at a uniform rate in each lineage with the following exceptions: (1) all *R2* elements have been lost from the *D. erecta* and *D. orena* lineage; (2) the rate of sequence evolution of *R1* may be faster in the sibling species *D. sechellia*, *D. simulans* and *D. mauritiana* (see Table 3) and (3) the apparent slow rates of sequence evolution of *R1* and *R2* in the *D. teissieri* and *D. yakuba* lineages (Table 2) can probably be attributed to introgression events.

specialized for insertion into the 28S gene (JAKUBCZAK *et al.* 1990; EICKBUSH 1992). Most copies of *R1* and *R2* are located within the rDNA locus of the host, and all characterized copies of the elements that reside outside of the 28S genes are defective (ROIHA *et al.* 1981; BROWNE *et al.* 1984; XIONG *et al.* 1988; JAKUBCZAK *et al.* 1992; EICKBUSH and EICKBUSH 1995). *R2* elements have been shown to encode a sequence specific endonuclease that cleaves the 28S gene target site (XIONG and EICKBUSH 1988b). The 3' hydroxyl group on the coding strand of the 28S gene that is generated by this cleavage serves as primer to initiate the reverse transcription of the RNA template using a reverse transcriptase also encoded by the *R2* element (LUAN *et al.* 1993).

In the preceeding report (EICKBUSH and EICKBUSH 1995), we have characterized the *R1* and *R2* elements within the eight members of the *melanogaster* species subgroup of the genus *Drosophila*. Like the 28S genes in which they are inserted, the *R1* and *R2* elements of each species exist as highly uniform populations averaging <0.5% nucleotide sequence divergence. The phylogeny of *R1* and *R2*, based on their nucleotide sequence, parallels the species phylogeny, suggesting they have remained stably associated with each species during the evolution of this subgroup. The only exception to this stability is the loss of *R2* elements from the species *Drosophila orena* and *Drosophila erecta*. In this report the rates of nucleotide substitution in *R1* and *R2* have been compared with that of several nuclear genes from these same *Drosophila* species. The sequences of the *R1* and *R2* elements from *Drosophila ambigua*, a member

of the *obscura* species group, were also determined to take advantage of the many sequence comparisons that have been made between genes from the *melanogaster* and *obscura* species groups. The synonymous nucleotide substitution rates of *R1* and *R2*, which represent the best indication of the baseline nucleotide substitution rates, were found to be similar to that of nuclear genes when corrections were made for the levels of codon bias.

MATERIALS AND METHODS

Sequence of *R1* and *R2* elements from the *obscura* species group: *D. ambigua* stock 14011–0091.1 was obtained from the *Drosophila* stock center (Bowling Green University). Isolation of genomic DNA, PCR amplification of the 3' half of *R1* and *R2* elements inserted in the 28S genes of these species and cloning of the ~1.7 kb products into mp18 were conducted as described in EICKBUSH and EICKBUSH (1995). Complete sequencing of the cloned products was obtained using the Universal sequencing primer (U.S. Biochemicals), 5'-GGTCTGTCGTGAACGACGACG-3' for *R1* and 5' GATGATAGG(T/A)GTTCTGAG-3' for *R2*, which are complementary to sequences ~550 bp downstream of the PCR primer, and 5'-CATCCCGAACGCAGGCTTCG-3' for *R1* and 5' AGG-TCTCG(A/T)ACCACGAG-3' for *R2*, which are complementary to sequences ~1050 bp downstream of the PCR primer. To minimize the error associated with these sequence determinations, two *R1* and two *R2* elements were sequenced. All nucleotide positions that differed between the two copies (~0.5%) were rechecked on the original sequencing gel. We have estimated that by using this approach the error rate associated with the sequencing of only one strand is 0.06%, which is below the misincorporation rate of 0.22% associated with the PCR amplifications themselves (EICKBUSH and EICKBUSH 1995). The *D. ambigua* sequences have been submitted

TABLE 1

Nucleotide sequence divergence of the 3' untranslated regions of *R1* (below diagonal) and *R2* (above diagonal)

<i>R1</i>	<i>R2</i>								
	<i>melanogaster</i>	<i>mauritiana</i>	<i>sechellia</i>	<i>simulans</i>	<i>teissieri</i>	<i>yakuba</i>	<i>erecta</i>	<i>orena</i>	<i>ambigua</i>
<i>melanogaster</i>		0.049	0.047	0.037	0.100	0.135	n.a.	n.a.	0.418
<i>mauritiana</i>	0.061		0.009	0.039	0.104	0.116	n.a.	n.a.	0.414
<i>sechellia</i>	0.078	0.037		0.038	0.103	0.112	n.a.	n.a.	0.430
<i>simulans</i>	0.074	0.034	0.028		0.099	0.132	n.a.	n.a.	0.414
<i>teissieri</i>	0.110	0.113	0.123	0.118		0.047	n.a.	n.a.	0.414
<i>yakuba</i>	0.107	0.121	0.117	0.114	0.030		n.a.	n.a.	0.468
<i>erecta</i>	0.117	0.153	0.147	0.142	0.127	0.114		n.a.	n.a.
<i>orena</i>	0.104	0.124	0.120	0.117	0.104	0.093	0.045		n.a.
<i>ambigua</i>	0.569	0.582	0.577	0.591	0.582	0.608	0.610	0.607	

Distance values (*D*) were calculated according to the following formula $D = -\frac{3}{4} \ln(1 - \frac{4}{3}a)$, where *a* equals the number of identical nucleotides divided by the total number of positions that can be compared (JUKES and CANTOR 1969). n.a., not applicable, due to the absence of *R2* elements in *D. erecta* and *D. orena*.

to GenBank under the following accession numbers: *R1amb1*-U13639 and *R2amb1*-U13640. All other *R1* and *R2* sequences used in this report were obtained from the previous report (EICKBUSH and EICKBUSH 1994) and represent the consensus sequence of each species. Copies of the *R1* and *R2* sequence alignments are available from the authors upon request.

Sequence analysis: All DNA sequence analyses were conducted with the computer program packages MacVector (International Biotechnologies, Inc.) and the Wisconsin Genetics Computer Group (GCG). The synonymous substitution rate (K_s) and nonsynonymous (replacement) substitution rates (K_a) were determined using the GCG program DIVERGE, which is based on the method of PERLER *et al.* (1980). The standard error associated with each value was calculated as the square root of the variance as described by NEI (1987, pp. 65–66). Codon usage bias was determined according to SHIELDS *et al.* (1988) using the programs made available by E. MORIYAMA (MORIYAMA and HARTL 1993). The phylogenetic relationships of *R1* and *R2* were determined with maximum parsimony algorithms using the PAUP 3.0 package of programs (SWOFFORD 1991).

RESULTS

***R1* and *R2* elements from the *melanogaster* and *obscura* species groups:** Shown in Figure 1 is the consensus phylogenetic relationship of the eight members of the *melanogaster* species subgroup as determined by LACHAISE *et al.* (1988) based on a comprehensive survey of morphological and molecular data sets. Three distinct species complexes are recognized within this subgroup: the *melanogaster* complex (*D. melanogaster*, *Drosophila simulans*, *Drosophila sechellia* and *Drosophila mauritiana*), the *yakuba* complex (*Drosophila yakuba* and *Drosophila teissieri*), and the *erecta* complex (*D. erecta* and *D. orena*). The age of the *melanogaster* subgroup has been estimated to be 17–20 million years. The preceding paper describes the characterization of *R1* and *R2* elements found in each member of this subgroup (EICKBUSH and EICKBUSH 1995). Except for the absence of *R2* elements in *D. orena*

and *D. erecta*, each species contains uniform families of *R1* and *R2* whose sequence relationships are consistent with their species phylogeny.

We have also surveyed *R1* and *R2* elements from a number of other species groups of *Drosophila* (W. C. LATHE, W. D. BURKE and T. H. EICKBUSH, unpublished data). To provide an outgroup for the sequence comparisons within the *melanogaster* subgroup, as well as to take advantage of the large number of sequence comparisons that have been made between nuclear genes of the *melanogaster* and the *obscura* species groups, we report here the sequences of *R1* and *R2* elements from *D. ambigua*, a member of the *obscura* group. The *obscura* and *melanogaster* groups are believed to have separated 45–60 million years ago (BEVERLY and WILSON 1984; LACHAISE *et al.* 1988).

The 3' ends of *R1* and *R2* elements inserted in the 28S rRNA genes of *D. ambigua* were PCR amplified using degenerate primers to the reverse transcriptase domain of each element in combination with a primer to 28S gene sequences downstream of each insertion site (BURKE *et al.* 1993; EICKBUSH and EICKBUSH 1995). The entire PCR-amplified product, comprising 1.7 kb of DNA from the 3' half of each element, was cloned into mp18 vectors and sequenced (see MATERIALS AND METHODS). This 1.7-kb region included the carboxyl terminal 386 codons of the second ORF and a 550-bp 3' untranslated region of the *R1* element. For the *R2* element, this region included 476 codons of the ORF and a 250-bp 3' untranslated region. For both *R1* and *R2*, the sequenced region of the ORF included part of the reverse transcriptase domain and a conserved carboxyl terminal domain. This carboxyl terminal domain contains a centrally located cysteine-histidine domain similar to that found in RNA-binding proteins (JAKUBCZAK *et al.* 1990; BURKE *et al.* 1993).

TABLE 2
Synonymous and replacement substitution rates for *R1*, *R2* and *Adh*

	<i>melanogaster</i>	<i>mauritiana</i>	<i>sechellia</i>	<i>simulans</i>	<i>teissieri</i>	<i>yakuba</i>	<i>erecta</i>	<i>orena</i>	<i>ambigua</i>
<i>melanogaster</i>									
<i>Adh</i>		0.011 (0.004)	0.004 (0.002)	0.004 (0.004)	0.009 (0.004)	0.015 (0.005)	0.018 (0.006)	0.021 (0.006)	0.045 (0.009)
<i>R1</i>		0.030 (0.006)	0.035 (0.007)	0.042 (0.007)	0.050 (0.007)	0.051 (0.007)	0.064 (0.009)	0.064 (0.009)	0.259 (0.020)
<i>R2</i>		0.014 (0.004)	0.015 (0.003)	0.013 (0.004)	0.034 (0.006)	0.037 (0.006)	n.a.	n.a.	0.209 (0.016)
<i>mauritiana</i>									
<i>Adh</i>	0.076 (0.019)		0.007 (0.003)	0.007 (0.000)	0.012 (0.005)	0.019 (0.006)	0.018 (0.006)	0.020 (0.006)	0.052 (0.010)
<i>R1</i>	0.204 (0.025)		0.015 (0.004)	0.009 (0.003)	0.052 (0.008)	0.059 (0.008)	0.068 (0.009)	0.069 (0.009)	0.272 (0.021)
<i>R2</i>	0.146 (0.021)		0.007 (0.002)	0.003 (0.002)	0.040 (0.006)	0.043 (0.007)	n.a.	n.a.	0.213 (0.016)
<i>sechellia</i>									
<i>Adh</i>	0.080 (0.020)	0.026 (0.012)		0.000 (0.003)	0.009 (0.004)	0.015 (0.005)	0.014 (0.005)	0.016 (0.006)	0.048 (0.009)
<i>R1</i>	0.253 (0.029)	0.128 (0.019)		0.018 (0.005)	0.051 (0.008)	0.054 (0.008)	0.072 (0.009)	0.069 (0.009)	0.263 (0.020)
<i>R2</i>	0.158 (0.021)	0.033 (0.011)		0.003 (0.002)	0.041 (0.006)	0.044 (0.006)	n.a.	n.a.	0.217 (0.016)
<i>simulans</i>									
<i>Adh</i>	0.072 (0.019)	0.028 (0.000)	0.023 (0.012)		0.009 (0.005)	0.017 (0.006)	0.014 (0.006)	0.016 (0.006)	0.048 (0.010)
<i>R1</i>	0.256 (0.029)	0.115 (0.018)	0.140 (0.020)		0.055 (0.008)	0.059 (0.008)	0.076 (0.009)	0.076 (0.009)	0.275 (0.021)
<i>R2</i>	0.136 (0.020)	0.043 (0.012)	0.075 (0.016)		0.040 (0.006)	0.043 (0.007)	n.a.	n.a.	0.217 (0.016)
<i>teissieri</i>									
<i>Adh</i>	0.206 (0.036)	0.185 (0.035)	0.167 (0.032)	0.167 (0.035)		0.007 (0.003)	0.020 (0.006)	0.022 (0.006)	0.048 (0.010)
<i>R1</i>	0.354 (0.040)	0.329 (0.042)	0.370 (0.043)	0.414 (0.047)		0.005 (0.003)	0.045 (0.008)	0.046 (0.008)	0.270 (0.020)
<i>R2</i>	0.413 (0.043)	0.411 (0.043)	0.433 (0.045)	0.384 (0.042)		0.004 (0.002)	n.a.	n.a.	0.206 (0.015)
<i>yakuba</i>									
<i>Adh</i>	0.168 (0.032)	0.152 (0.031)	0.134 (0.020)	0.123 (0.031)	0.063 (0.019)		0.026 (0.007)	0.026 (0.007)	0.057 (0.010)
<i>R1</i>	0.303 (0.036)	0.282 (0.036)	0.324 (0.039)	0.356 (0.042)	0.075 (0.014)		0.051 (0.008)	0.048 (0.008)	0.272 (0.020)
<i>R2</i>	0.397 (0.041)	0.402 (0.042)	0.425 (0.044)	0.376 (0.041)	0.025 (0.009)		n.a.	n.a.	0.207 (0.015)
<i>erecta</i>									
<i>Adh</i>	0.156 (0.030)	0.176 (0.033)	0.161 (0.031)	0.161 (0.033)	0.187 (0.035)	0.188 (0.032)		0.009 (0.004)	0.048 (0.009)
<i>R1</i>	0.477 (0.048)	0.421 (0.045)	0.400 (0.042)	0.496 (0.049)	0.428 (0.042)	0.382 (0.040)		0.029 (0.006)	0.265 (0.020)
<i>R2</i>	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.		n.a.	n.a.
<i>orena</i>									
<i>Adh</i>	0.186 (0.034)	0.179 (0.034)	0.165 (0.032)	0.153 (0.034)	0.213 (0.036)	0.195 (0.033)	0.098 (0.023)		0.048 (0.009)
<i>R1</i>	0.451 (0.045)	0.354 (0.040)	0.350 (0.040)	0.389 (0.043)	0.391 (0.041)	0.361 (0.040)	0.234 (0.029)		0.275 (0.021)
<i>R2</i>	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.		n.a.
<i>ambigua</i>									
<i>Adh</i>	0.560 (0.080)	0.540 (0.077)	0.539 (0.077)	0.519 (0.077)	0.523 (0.074)	0.597 (0.084)	0.472 (0.068)	0.447 (0.065)	
<i>R1</i>	1.221 (0.197)	1.221 (0.222)	1.158 (0.196)	1.248 (0.248)	1.158 (0.155)	1.154 (0.158)	1.034 (0.129)	1.002 (0.129)	
<i>R2</i>	1.248 (0.219)	1.309 (0.292)	1.161 (0.383)	1.262 (0.281)	1.129 (0.167)	1.076 (0.150)	n.a.	n.a.	

Below the diagonal, synonymous substitution rate estimated by K_s (the number of synonymous substitutions per site). Above the diagonal, replacement substitution rate estimated by K_a (the number of replacement substitutions per site). Both K_s and K_a values were calculated by the method of PERLER *et al.* (1980). The alcohol dehydrogenase (*Adh*) sequences used are as follows: *D. melanogaster* (J01066); *D. mauritiana* (Z00033); *D. sechellia* (X04672); *D. simulans* (M36581); *D. teissieri* (X54118); *D. yakuba* (X54120); *D. erecta* (X54116); *D. orena* (Z00032) and *D. ambigua* (M60982).

The nucleotide sequence divergence per site of the 3' untranslated region for each pairwise comparison of *R1* and *R2* elements from the nine *Drosophila* species is shown in Table 1. The nucleotide divergence of the region encoding the ORF is shown in Table 2, where it has been separated into the number of synonymous substitutions per site (K_s) and the number of nonsynonymous (replacement) substitutions per site (K_a). For each species of the *melanogaster* subgroup, multiple copies of *R1* and *R2* were PCR amplified and sequenced. The levels of sequence divergence determined for different copies of intact elements from each species averaged <0.5%, similar to that expected from PCR misincorporation errors involved in the amplification process itself (EICKBUSH and EICKBUSH 1995). Therefore, to better estimate the level of nucleotide divergence between *R1* and *R2* elements in this subgroup, consensus *R1* and *R2* sequences have been used for each species.

3' untranslated regions: The 3' untranslated region of the sequenced *R1* elements from the nine *Drosophila*

species varied from 546 to 596 bp as a result of 43 short insertions and deletions (indels) throughout the region. The largest indels (up to 65 bp in length) were located ~100 bp from the 3' end of the element. In the case of the *R2* elements, the 3' untranslated regions were considerably shorter than that of the *R1* elements, varying from 219 to 253 bp. Again, this variation resulted from 22 indels ranging from 1 to 17 bp in length with most occurring 50–100 bp from the 3' end of the element. In spite of the large number of indels in the 3' untranslated region, the level of sequence divergence between species was only slightly higher than that of the protein coding region. Divergence within the 3' untranslated region (Table 1) averaged twice that of the divergence at replacement positions and one third that of the divergence at synonymous positions in the coding region (Table 2), suggesting that the 3' untranslated regions of *R1* and *R2* have also been under selective constraint.

Comparison of the 3' untranslated region of the *R2*

FIGURE 2.—Comparison of the nucleotide sequences of the 3' untranslated region of the *R2* elements. All sequences are compared with the *R2* element from *D. melanogaster*, with identical nucleotides shown as dots, nucleotide substitutions indicated, and gaps necessary for alignment by dashes. The eight codons at the end of the ORF are shown in triplets. The *B. mori* *R2* sequence is given for the end of the ORF and in the three regions where sequence similarity could be detected to the *Drosophila* elements (shaded regions). The numbers of nucleotides in the *B. mori* *R2* element separating these blocks of identity are given. All *Drosophila* elements end in a poly(A) sequence from 10 to 25 nucleotides in length. Sequences are from BURKE *et al.* (1987) for *B. mori*, from JAKUBCZAK *et al.* (1990) for *D. melanogaster* and from EICKBUSH and EICKBUSH (1995) for all other species.

elements is of particular interest because this sequence has been shown in the case of the *B. mori* *R2* element to be required by the reverse transcriptase for the integration reaction (LUAN *et al.* 1993; LUAN and EICKBUSH, unpublished data). The 3' untranslated region of the *B. mori* element is similar in length to the *D. melanogaster* element, but little sequence similarity between the *R2* elements from the two species had been previously detected (JAKUBCZAK *et al.* 1990). Based on a comparison of the *Drosophila* *R2* elements, three highly conserved sequences are present in the 3' untranslated region, located 170, 120 and 80 nucleotides from the poly(A) tail (Figure 2). Short regions of sequence similarity with approximately the same spacing as in the *Drosophila* elements can also be detected in the *B. mori* element. These conserved regions are prime candidates to test for their role in binding the RNA transcript to the *R2* reverse transcriptase in the in vitro integration reaction (LUAN *et al.* 1993).

Comparison of the 3' untranslated regions of the *Drosophila* *R2* elements also provides insights into the issue of how the poly(A) tail is generated at their extreme 3' ends. No sequence similarity to a polyadenylation signal is present in the region upstream of this tail (Figure 2). We have recently shown that the reverse

transcriptase encoded by the *R2* element from *B. mori* has the ability to add nontemplated T residues on the cDNA strand during the initiation of reverse transcription (LUAN and EICKBUSH, unpublished data). The addition of T's to the cDNA strand generates a poly(A) tail on the sense strand (RNA synonymous strand). The absence of a recognizable polyadenylation signal supports the model that their poly(A) tails are generated by a nontemplated process at the reverse transcription step rather than by polyadenylation during an RNA processing step.

Protein encoding regions: The 476 codons sequenced from the *D. ambigua* *R2* element could be aligned with the *R2* elements from the *melanogaster* subgroup without the need for insertions or deletions. In the case of the *R1* elements, the 386 codons of the *D. ambigua* *R1* could be aligned with *R1* elements from the *melanogaster* subgroup with the addition of a 3-bp insertion following nucleotide 495 and a 3-bp deletion 12 codons later. Compared with the large number of indels in the 3' untranslated region of these elements, 22 in the 3' region of the *R2* elements (Figure 2) and 43 in the 3' region of the *R1* elements (data not shown), the paucity of such changes in the region encoding the ORF strongly suggested that the elements

in each lineage were under selective pressure to express their proteins. A second indication of the selective pressure on the encoded protein sequence of *R1* and *R2* was the finding that in all species comparisons the rate of change at synonymous positions (K_s values) were considerably higher than the rate of change at replacement positions (K_a values) (Table 2). For all pairwise comparisons of *R1* elements from the nine species, the mean K_s value was 6.6 times higher than the K_a values (range 3.6–15.0). The mean K_s/K_a value for the *R2* elements in the seven species was 9.6 (range 4.7–25.0). Most protein encoding genes of *Drosophila* have K_s/K_a values in the range 4–20 (AKASHI 1994).

Rates of nucleotide evolution: Also included in Table 2 are the K_s and K_a values for the alcohol dehydrogenase (*Adh*) gene. The *Adh* gene has been sequenced from all eight members of the *melanogaster* subgroup and from representative members of the *obscura* species group (JEFFS *et al.* 1994). Comparison of the K_s and K_a values of *R1* and *R2* to that of *Adh* allows a determination of the rate of nucleotide substitution of *R1* and *R2* relative to that of a single copy nuclear gene from the same species. The mean rate of nucleotide substitutions at replacement sites in *R1* elements (K_a) was 4.3 times that of the *Adh* gene, whereas the mean rate of replacement substitutions in *R2* elements was 3.3 times that of the *Adh* gene. Thus the proteins encoded by both *R1* and *R2* are undergoing a more rapid rate of evolution than the *Adh* protein. The only exception to this conclusion is in comparisons of the *R1* and *R2* elements of *D. teissieri* and *D. yakuba*. The K_a values for both *R1* and *R2* in this species comparison were similar to that of the K_a for *Adh*. The higher K_a values for *R1* and *R2* for most lineages could be the result of a higher mutation rate due to retrotransposition or lower selective pressures on the *R1* and *R2* encoded protein relative to that of *Adh*. Because a comparison of the K_a values does not discriminate between these two possibilities, a better indicator of the relative rates of change for these genes *vs.* nuclear genes is the rate of substitutions at synonymous sites. Synonymous sites are least affected by selective pressures on the encoded protein sequences (SHARP and LI 1989; MORIYAMA and HARTL 1993; AKASHI 1994).

The K_s values for *R1* and *R2* have been compared with the K_s values for *Adh* in Table 3. Rather than give all pairwise comparisons, Table 3 presents the averages for all comparisons within and between the three species complexes of the *melanogaster* subgroup, as well as the average of these three complexes with that of *D. ambigua*. For most of the *Drosophila* lineages compared in Table 3, the synonymous substitution rates of *R1* and *R2* were from 1.9 to 2.5 times higher than that of *Adh*. In only two lineages were the K_s values of *R1* or *R2* notably different from these mean values. In the case of the *R1* elements from *D. simulans*, *D. sechellia* and

TABLE 3
Rate of synonymous substitution (K_s) in *R1* and *R2* elements relative to *Adh*

Species comparisons	<i>R1</i>	<i>R2</i>
<i>melanogaster</i> <i>vs.</i> <i>simulans</i> , <i>sechellia</i> and <i>mauritiana</i>	3.2 (2.7–3.6)	1.9 (1.9–2)
<i>simulans</i> <i>vs.</i> <i>sechellia</i> , <i>sechellia</i> <i>vs.</i> <i>mauritiana</i> , <i>simulans</i> <i>vs.</i> <i>mauritiana</i>	5.0 (4.1–6.1)	2 (1.3–3.3)
<i>melanogaster</i> complex <i>vs.</i> <i>yakuba</i> complex	2.2 (1.7–2.9)	2.5 (2–3.2)
<i>teissieri</i> <i>vs.</i> <i>yakuba</i>	1.2	0.4
<i>melanogaster</i> complex <i>vs.</i> <i>erecta</i> complex	2.5 (2–3.1)	n.a.
<i>yakuba</i> complex <i>vs.</i> <i>erecta</i> complex	2 (1.8–2.3)	n.a.
<i>erecta</i> <i>vs.</i> <i>orena</i>	2.4	n.a.
<i>ambigua</i> <i>vs.</i> <i>melanogaster</i> subgroup	2.2 (1.9–2.4)	2.2 (1.8–2.4)

For each species comparison the K_s value for *R1* or *R2* was divided by the K_s value for *Adh*. For those comparisons involving multiple species, the average K_s ratio is shown followed by the range of numbers in parentheses.

D. mauritiana, the synonymous substitution rates of *R1* averaged five times that of the *Adh* gene from these same species. The higher K_s values in the *D. simulans*, *D. sechellia* and *D. mauritiana* lineages has also resulted in a somewhat higher K_s value for the comparison of *D. melanogaster* to these three species (average 3.2, Table 3). The second difference found in the rate of evolution of *R1* and *R2* compared with that of *Adh* was that the K_s values for the *R1* elements from *D. teissieri* and *D. yakuba* were 1.2 times that of the *Adh* rate and only 0.4 times the *Adh* rate in the case of the *R2* elements. The low K_s values in the *D. teissieri* *vs.* *D. yakuba* comparisons were consistent with the very low K_a values described above when the *R1* and *R2* elements of these species were compared with the *Adh* gene. As will be discussed below, the slower apparent rate of *R1* and *R2* nucleotide substitutions in *D. teissieri* and *D. yakuba* may be the result of introgression of the elements rather than a slowing of the rate of nucleotide substitution.

Codon bias: Does the 2–2.5 times higher K_s values for *R1* and *R2* compared with that of *Adh* mean that approximately half of the nucleotide substitutions in *R1* and *R2* were a result of retrotransposition? The K_s values for different nuclear genes have been shown to vary with the codon utilization bias of the genes, with those genes containing higher codon preferences having lower K_s values (SHARP and LI 1989). With 0.00 representing complete random utilization of codons and 1.2 representing the highest level of bias detected in genes of *Drosophila*, *R1* and *R2* have very low codon

FIGURE 3. — Relationship between the bias in synonymous codon usage and the synonymous substitution rate (K_s) for three species comparisons in *Drosophila*. (A) Each point represents the comparison of a pair of similar gene sequences, one from the *melanogaster* species group (*D. melanogaster* in all cases) and the other from the *obscura* species group (various species described below). (B) Comparison of gene sequences from *D. melanogaster* and from *D. simulans*. (C) Comparison of gene sequences from *D. simulans* and from *D. mauritiana*. Codon usage bias is measured according to SHIELDS *et al.* (1988) with the value presented representing the average of the two species in the comparison. The synonymous substitution rate (K_s) is estimated by PERLER *et al.* (1980). The dotted line represents the least-square fit to all data points in A and B and to all data points except *R1* in C. Gene symbols and data sources are as follows: *Adh* (as in Table 2); *gart* (glycinamide ribotide transformylase), X06286 from *D. melanogaster*, X06285 from *D. pseudoobscura*; *Xdh* (xanthine dehydrogenase), Y00308 from *D. melanogaster*, M33977 from *D. pseudoobscura*; *C.s15* and *C.s19* (chorion proteins s15 and s19), X02497 from *D. melanogaster*, X53423 from *D. subobscura*; *ubx* (ultrabithorax), X05723 from *D. melanogaster*, X05179 from *D. pseudoobscura*; *hsp82* (heat shock protein 82), X03810 from *D. melanogaster*, X03812 from *D. pseudoobscura*; *tra* (transformer), M17478 from *D. melanogaster*, X66930 from *D. simulans*; *sal* (spalt), X75541 from *D. melanogaster* and M21227

biases [0.12 (range 0.08–0.16) for *R1* and 0.07 (range 0.05–0.08) for *R2*]. *Adh* genes on the other hand have a high codon bias, 0.95 (range 0.88–1.07). Thus it was necessary to compare the K_s values for *R1* and *R2* with that of nuclear genes with similar levels of codon bias. However, no gene with low codon bias has been sequenced from most of the nine *Drosophila* species analyzed in this study. As a consequence, the K_s values of *R1* and *R2* were compared with a number of genes that have been sequenced from at least two of the nine *Drosophila* species in this study.

The codon biases of *R1*, *R2* and various nuclear genes of *Drosophila* are plotted in Figure 3 vs. their K_s values for three species comparisons (*D. obscura* vs. *D. melanogaster*, panel A; *D. melanogaster* vs. *D. simulans*, panel B; *D. mauritiana* vs. *D. simulans*, panel C). For each species pair in Figure 3, the K_s rate for the various genes is inversely correlated with their level of codon bias. In the case of the most distant species (the *melanogaster* group vs. the *obscura* group, panel A), the K_s values for both *R1* and *R2* were similar to that of two other genes of low codon bias, glycineamide ribotide transformylase (*gart*) and xanthine dehydrogenase (*Xdh*). These four genes with low codon bias have K_s values that are 1.5–2 times greater than the high codon bias genes like *Adh*. In the two comparisons of more closely related genes (*D. melanogaster* vs. *D. simulans*, panel B; *D. simulans* vs. *D. mauritiana*, panel C), the K_s values for the *R2* elements are again similar to that of the genes with low codon bias. The K_s value for *R1*, however, was somewhat higher than the nuclear genes in the *D. melanogaster* vs. *D. simulans* comparison and significantly higher in the *D. simulans* vs. *D. mauritiana* comparison. High K_s values were found for the *R1* elements in all pairwise comparisons between *D. simulans*, *D. mauritiana* and *D. sechellia* (see Table 2), suggesting that the *R1* elements have evolved more rapidly in all three species.

A comparable number of nuclear genes has not been sequenced from species of the *yakuba* species complex or the *erecta* species complex. However, two genes with low codon bias, the developmentally regulated genes transformer (*tra*) and spalt (*sal*) (Figure 3B), have been sequenced from *D. erecta* and *D. oreana*, respectively, in addition to *D. melanogaster* and *D. simulans* (REUTER *et al.* 1989; O'NEIL and BELOTE 1992). The *tra* K_s values for the *D. melanogaster* vs. *D. erecta* comparison is 0.327 (0.073) and for *D. simulans* vs. *D. erecta*, 0.294 (0.069). The *sal* K_s values for the *D. melanogaster*

from *D. simulans*; *est6* (esterase 6), J04167 from *D. melanogaster*, L10670 from *D. simulans*, M84581 from *D. mauritiana*; *yp2* (yolk protein 2), L14421 from *D. melanogaster*, L14426 from *D. simulans*, L14417 from *D. mauritiana*; and *gst* (glutathione S-transferase), X14233 from *D. melanogaster*; M84577 from *D. simulans*, M84581 from *D. mauritiana*.

vs. D. orena comparison is 0.502 (0.092) and for *D. simulans vs. D. orena*, 0.584 (0.114). These values are similar to the average K_s value of 0.45 for *R1* elements in comparisons of all species from *melanogaster* and *erecta* species complexes. We conclude that the rate of synonymous site substitutions in *R1* and *R2* is similar to that of nuclear genes when controlled for their low codon bias. The only exceptions are (1) the *R1* elements of the sibling species *D. simulans*, *D. mauritiana* and *D. sechellia*, where the elements have evolved at rates over twice that of nuclear genes, and (2) both *R1* and *R2* elements in *D. teissieri* and *D. yakuba*, where their rates of evolution is similar or less than nuclear genes with higher levels of codon bias (Table 3).

DISCUSSION

In the accompanying report, we showed that the retrotransposable elements *R1* and *R2* are maintained as highly uniform populations that have been vertically transmitted during the evolution of the *melanogaster* species subgroup (EICKBUSH and EICKBUSH 1995). Based on their structure and the sequence similarity of their reverse transcriptase domains, *R1* and *R2* are no more similar to each other than they are to any other non-LTR retrotransposable element (XIONG and EICKBUSH 1990; EICKBUSH 1992). Furthermore, the level of *R1* within a species varies independently of the level of *R2*, suggesting that they function independently (XIONG *et al.* 1988; JAKUBCZAK *et al.* 1992). Thus the stability of *R1* and *R2* within the *melanogaster* subgroup provides the opportunity to compare the rates of nucleotide sequence evolution of two retrotransposable elements with that of nuclear genes sequenced from these species. To enable more distant sequence comparisons and to take advantage of the many genes that have been sequenced from members of the *obscura* species group, we have also determined the sequences of *R1* and *R2* elements from the species *D. ambigua*. Using the *Adh* gene as a reference, the sequence divergences of the *R1* and *R2* elements between *D. ambigua* and the *melanogaster* species group were consistent with that expected from the extrapolated values determined from comparisons within the *melanogaster* species subgroup (Tables 1 and 2). This would suggest that the *R1* and *R2* elements have been stably maintained for the 45–60 million years since the separation of the *melanogaster* and *obscura* groups (LACHAISE *et al.* 1988).

Surveys of Dipteran, Lepidopteran, Hymenopteran and Coleopteran species revealed that most copies of the *R1* and *R2* elements randomly selected from a species by either screening genomic libraries or by PCR amplification encode intact ORFs, that is, do not contain frameshifts or premature termination codons (BURKE *et al.* 1993; EICKBUSH and EICKBUSH 1995). The

only defective *R1* and *R2* elements frequently encountered within the 28S gene are elements with 5' truncations (DAWID and REBBERT 1981; ROIHA *et al.* 1981; KERREBROCK *et al.* 1989; JAKUBCZAK *et al.* 1990), which appear to be generated by the retrotransposition process itself (LUAN *et al.* 1993). Our interpretation of these findings has been that most full-length *R1* and *R2* elements present within the 28S genes of each species are capable of retrotransposition.

Comparison of the rates of nucleotide substitution at synonymous and replacement sites in *R1* and *R2* elements from the nine *Drosophila* species in this report greatly strengthens this interpretation. The data suggest that the *R1* and *R2* elements have been capable of retrotransposition during most, if not all, of the evolution of the *melanogaster* and *obscura* species groups. Averaged over all pairwise comparisons of the *R1* and *R2* elements within the different species, the mean K_s values were 6.6 and 9.6 times higher than the K_a values, respectively. These higher rates of nucleotide substitution at synonymous positions relative to replacement positions and the paucity of segmental changes that are prevalent in the 3' untranslated region of each element indicate that the proteins encoded by *R1* and *R2* have been under purifying selection during most of the evolution of the *melanogaster* subgroup. The protein functions known to be encoded by the 120-kD *R2* protein are associated with the retrotransposition process: an endonuclease activity specific for the 28S gene target site, a RNA binding activity specific for the 3' end of the RNA template and a reverse transcriptase activity (XIONG and EICKBUSH 1988b; LUAN *et al.* 1993). Presumably, it is this ability to retrotranspose that has enabled *R1* and *R2* to remain within the 28S genes. Those copies that have lost their ability to retrotranspose are eliminated from the rDNA locus by recombination.

The rate of nucleotide substitution in *R1* and *R2* is a function of both the accumulation of mutations during the normal rounds of chromosomal DNA replication and the cycles of *R1* and *R2* retrotransposition. Comparisons of the rate of nucleotide change in *R1* and *R2* to that of another gene also sequenced from all nine species of *Drosophila* studied in this report, the *Adh* gene, indicated that these elements evolved at a uniformly faster rate than the *Adh* gene. For most comparisons within the *melanogaster* subgroup or between the *melanogaster* group and *D. ambigua*, substitutions at replacement sites in the protein encoding regions of *R1* and *R2* were three to four times greater than in the *Adh* gene, whereas substitutions at synonymous sites were 2.0–2.5 times faster. Changes at synonymous sites are a better indicator of the baseline mutation rate because the accumulation of changes at replacement sites are affected by selection on the encoded protein and are known to vary by over an order

of magnitude (AKASHI 1994). The rate of change at synonymous sites has been shown to be constant for nuclear genes, if corrections are made for codon utilization bias (SHIELDS *et al.* 1989). *R1* and *R2* elements in *Drosophila*, like most transposable elements in *Drosophila* (SHIELDS and SHARP 1989), have very low codon biases, whereas the *Adh* genes have a high bias (Figure 3). When the rates of nucleotide substitution in *R1* and *R2* were compared with single copy nuclear genes that also have low codon biases, the synonymous nucleotide substitution rates of *R1* and *R2* in most lineages were similar to that of nuclear genes. Thus, for most lineages, potential contributions by retrotransposition to the nucleotide substitution rate of the *R1* and *R2* elements could not be detected.

As summarized in Figure 1, an increased rate of nucleotide substitution was detected over that of nuclear genes in only one lineage. Within the *D. mauritiana*, *D. sechellia* and *D. simulans* lineages, the K_s values for the *R1* elements were considerably higher than that of the *R2* elements and nuclear genes with low codon bias. Because the K_s/K_a values for the *R1* elements in these three species is also high (average 9.7) and no defects were detected in their ORFs [(EICKBUSH and EICKBUSH 1995) (Table 1)], this faster divergence does not correspond to an inactivation of the *R1* elements. Thus these sibling species represent our only evidence for the possible contribution of retrotransposition to the nucleotide substitution rate. As will be discussed below, this suggests that in these three species, *R1* elements have either undergone higher rates of retrotransposition than the *R1* elements in the other species or that these *R1* elements have a higher misincorporation rate.

The only other instance where the rate of change at synonymous sites was not 2.0–2.5 times that of the *Adh* genes was the low K_s values for both *R1* and *R2* in comparisons of *D. teissieri* and *D. yakuba* (Table 3). The K_s rate for *R1* was 1.2 times that of *Adh*, whereas the *R2* K_s rate was only 0.4 times that of the *Adh* rate. The codon bias of both *R1* and *R2* elements in *D. teissieri* and *D. yakuba* is low, and there is no reason to believe that the evolution of the *Adh* gene in these species is unusual (JEFFS *et al.* 1994). Because there is no known mechanism that would lead to such a slow synonymous substitution rate in genes with low codon bias, we suggest that there has been the introduction of these elements from one species to the other. A transfer of *R2* and *R1* between *D. teissieri* and *D. yakuba* need not have occurred by a special mechanism of horizontal transfer, because it could have occurred by introgression events when these two species were still capable of interbreeding (LACHAISE *et al.* 1988).

Few estimates have been reported of the rate at which retrotransposable elements have evolved within a lineage. Based on extensive comparisons of *L1* sequences

in several rodent species and in humans, the rate of nucleotide sequence evolution in these elements was found to be similar to that of nuclear pseudogenes (HARDIES *et al.* 1986; PASCALE *et al.* 1993). *Copia*-like retrotransposable elements in three species of wheat have been suggested to evolve at rates similar to that of nuclear genes (VANDERWIEL *et al.* 1993). However, in these systems most of the elements have been inactive for at least part of their lineage, and the sequence diversity of the elements within each species is as high as the divergence between the species. Thus the complexity of these systems complicates accurate measurements of their nucleotide substitution rates.

Is it unexpected that the rate of nucleotide substitution in *R1* and *R2* is similar to that of nuclear genes? Whether retrotransposition will increase the nucleotide substitution rate of an element is dependent on the error rate associated with each retrotransposition and the frequency with which these events occur. Although neither of these parameters are defined for the *R1* or *R2* elements, rough estimates can be made. We have argued that the stochastic loss of *R1* or *R2* from the rDNA locus would be directly proportional to the rate of recombination within this locus (JAKUBCZAK *et al.* 1992). By monitoring structural variation in the spacer region of the rDNA locus of *D. melanogaster*, COEN *et al.* (1982) have suggested that the rate of recombination that would give rise to detectable changes in the rDNA locus on genomic blots is <1 in 1,250 generations. Based on the pattern of 5' truncations in different geographical strains of *D. melanogaster*, in which many of the truncated copies of *R1* and *R2* are at a level of only one copy per genome, we have suggested that the retrotransposition frequency of *R1* and *R2* may be near this rate (JAKUBCZAK *et al.* 1992).

The second parameter needed to estimate the contribution of retrotransposition to the nucleotide substitution rate of an element, the error rate associated with each retrotransposition cycle, has only been accurately determined for the retroviruses. It has been shown that the error rate associated with retroviral transposition is 10,000- to 100,000-fold higher than normal rounds of DNA replication (LI and LOEB 1992; DRAKE 1993). The only estimate of the error rate associated with the transposition of a retrotransposable element has been with the yeast Ty1 element, where it has been estimated to be similar to that of retroviruses (A. GABRIEL, J. BOEKE, E. MULES and L. WILLEMS, personal communication). If we assume that the rate of retrotransposition is once every 1,250 generations, that there are ~ 18 germ cell divisions per generation (LINDSLEY and TOKUYASU 1980) and that the error rate of *R1* or *R2* retrotransposition is at the low end of the range determined for retroviruses, the contribution of retrotransposition to the rate of nucleotide substitution would be only a 50% increase rela-

tive to nuclear genes that do not undergo retrotransposition. Such a level would have been difficult to reliably detect in most of the lineages we studied. The increased rate of *R1* nucleotide substitution in *D. simulans*, *D. mauritiana* and *D. sechellia* may be the only cases where we have detected a contribution of retrotransposition to the nucleotide substitution rate. The expression of the reverse transcriptase of *R2* *in vitro* (LUAN *et al.* 1993) will make it possible to estimate directly the misincorporation rate of the *R2* enzyme. More accurate estimates of the frequency of retrotransposition will only be obtained when we better understand the nature of recombination within the rDNA loci and the developmental timing of *R1* and *R2* transpositions.

Evidence that *R1* and *R2* are present in most insects suggests that either they have been present in Arthropods since the divergence of insects or that they have extensively transferred between insects at a later date (JAKUBCZAK *et al.* 1991). Given the examples of horizontal transfer of mobile elements in insects (reviewed in KIDWELL 1993), we had been more inclined to believe the latter model. However, the data presented here and in the previous report (EICKBUSH and EICKBUSH 1995), indicating stability within a lineage and a slow rate of evolution, suggest that *R1* and *R2* are very well adapted to their niche in insect genomes. Clearly, models for an ancient origin and continuous maintenance of these elements need to be considered.

We thank WILLIAM BURKE for helpful comments on the manuscript and KRYSOF NEUMAN for assistance in the genomic blot analysis. This research was supported by National Science Foundation grant MCB-9219123.

LITERATURE CITED

- AKASHI, H., 1994 Synonymous codon usage in *Drosophila melanogaster*: natural selection and translational accuracy. *Genetics* **136**: 927–935.
- BEVERLY, S. M., and A. C. WILSON, 1984 Molecular evolution in *Drosophila* and the higher Diptera. II. A time scale for fly evolution. *J. Mol. Evol.* **21**: 1–13.
- BROWNE, M. J., C. A. READ, H. ROIHA and D. M. GLOVER, 1984 Site specific insertion of a type I rDNA element into a unique sequence in the *Drosophila melanogaster* genome. *Nucleic Acids Res.* **12**: 9111–9122.
- BURKE, W. D., C. C. CALALANG and T. H. EICKBUSH, 1987 The site-specific ribosomal insertion element type II of *Bombyx mori* (*R2Bm*) contains the coding sequence for a reverse transcriptase-like enzyme. *Mol. Cell. Biol.* **7**: 2221–2230.
- BURKE, W. D., D. G. EICKBUSH, Y. XIONG, J. JAKUBCZAK and T. H. EICKBUSH, 1993 Sequence relationship of retrotransposable elements *R1* and *R2* within and between divergent insect species. *Mol. Biol. Evol.* **10**: 163–185.
- COEN, E. C., J. M. THODAY and G. DOVER, 1982 Rate of turnover of structural variants in the rDNA gene family of *Drosophila melanogaster*. *Nature* **295**: 564–568.
- DAWID, I. B., and M. L. REBBERT, 1981 Nucleotide sequence at the boundaries between gene and insertion regions in the rDNA of *Drosophila melanogaster*. *Nucleic Acids Res.* **9**: 5011–5020.
- DOOLITTLE, R. F., D. F. FENG, M. S. JOHNSON and M. A. MCCLURE, 1989 Origin and evolutionary relationships of retroviruses. *Q. Rev. Biol.* **64**: 1–30.
- DRAKE, J. W., 1993 Rates of spontaneous mutation among RNA viruses. *Proc. Natl. Acad. Sci. USA* **90**: 4171–4175.
- EICKBUSH, T. H., 1992 The non-LTR class of retrotransposable elements. *New Biol.* **4**: 430–440.
- EICKBUSH, D. G., and T. H. EICKBUSH, 1995 Vertical transmission of the retrotransposable elements *R1* and *R2* during the evolution of the *Drosophila melanogaster* species subgroup. *Genetics* **139**: 671–684.
- HARDIES, S. C., S. L. MARTIN, C. F. VOLIVA, C. A. HUTCHISON and M. H. EDGELL, 1986 An analysis of replacement and synonymous changes in the rodent *L1* repeat family. *Mol. Biol. Evol.* **3**: 109–125.
- JAKUBCZAK, J. L., Y. XIONG and T. H. EICKBUSH, 1990 Type I (*R1*) and Type II (*R2*) ribosomal DNA insertions of *Drosophila melanogaster* are retrotransposable elements closely related to those of *Bombyx mori*. *J. Mol. Biol.* **212**: 37–52.
- JAKUBCZAK, J. L., W. D. BURKE and T. H. EICKBUSH, 1991 Retrotransposable elements *R1* and *R2* interrupt the rRNA genes of most insects. *Proc. Natl. Acad. Sci. USA* **88**: 3295–3299.
- JAKUBCZAK, J. L., M. K. ZENNI, R. C. WOODRUFF and T. H. EICKBUSH, 1992 Turnover of *R1* (Type I) and *R2* (Type II) retrotransposable elements in the ribosomal DNA of *Drosophila melanogaster*. *Genetics* **131**: 129–142.
- JEFFS, P. S., E. C. HOLMES and M. ASHBURNER, 1994 The molecular evolution of the alcohol dehydrogenase and alcohol dehydrogenase-related genes in the *Drosophila melanogaster* species subgroup. *Mol. Biol. Evol.* **11**: 287–304.
- JUKES, T. H., and C. R. CANTOR, 1969 Evolution of proteins, pp. 21–132 in *Mammalian Protein Metabolism*, Vol. 3, edited by H. N. MUNRO. Academic Press, New York.
- KERREBROCK, A. W., R. SRIVASTAVA and S. A. GERBI, 1989 Isolation and characterization of ribosomal DNA variants from *Sciara coprophila*. *J. Mol. Biol.* **210**: 1–13.
- KIDWELL, M. G., 1993 Lateral transfer in natural populations of eukaryotes. *Annu. Rev. Genet.* **27**: 235–256.
- LACHAISE, D., M. CARIOT, J. R. DAVID, F. LEMEUNIER, L. TSACAS *et al.*, 1988 Historical biogeography of the *Drosophila melanogaster* species subgroup. *Evol. Biol.* **22**: 159–227.
- LI, J., and L. A. LOEB, 1992 Fidelity of HIV-1 reverse transcriptase copying RNA *in vitro*. *Biochemistry* **31**: 954–958.
- LINDSLEY, D. L., and K. T. TOKUYASU, 1980 Spermatogenesis, pp. 225–249 in *The Genetics and Biology of Drosophila*, Vol. 2D, edited by M. ASHBURNER and T. R. F. WRIGHT. Academic Press, London.
- LUAN, D. D., M. H. KORMAN, J. L. JAKUBCZAK and T. H. EICKBUSH, 1993 Reverse transcription of *R2Bm* RNA is primed by a nick at the chromosomal target site—a mechanism for non-LTR retrotransposition. *Cell* **72**: 595–605.
- MORIYAMA, E. N., and D. L. HARTL, 1993 Codon usage bias and the base composition of nuclear genes in *Drosophila*. *Genetics* **134**: 847–858.
- NEL, M. 1987 *Molecular Evolutionary Genetics*. Columbia University Press, New York.
- O'NEIL, M. T., and J. M. BELOTE, 1992 Interspecific comparison of the transformer gene of *Drosophila* reveals an unusually high degree of evolutionary divergence. *Genetics* **131**: 113–128.
- PASCALE, E., C. LIU, E. VALLE, K. USDIN and A. V. FURANO, 1993 The evolution of long interspersed repeated DNA (*L1*, *LINE 1*) as revealed by the analysis of an ancient rodent *L1* DNA family. *J. Mol. Evol.* **36**: 9–20.
- PERLER, F., A. EFSTRADIADIS, P. LOMEDICO, W. GILBERT, R. KLODNER *et al.*, 1980 The evolution of genes: the chicken preproinsulin gene. *Cell* **20**: 555–566.
- PRESTON, B. D., B. J. POIESZ and L. A. LOEB, 1988 Fidelity of HIV-1 reverse transcriptase. *Science* **242**: 1168–1171.
- REUTER, D., R. SCHUH and H. JACKLE, 1989 The homeotic gene spalt (*sal*) evolved during *Drosophila* speciation. *Proc. Natl. Acad. Sci. USA* **86**: 5483–5486.
- ROBERTS, J. D., K. BEBENEK and T. A. KUNKEL, 1988 The accuracy of reverse transcriptase from HIV-1. *Science* **242**: 1171–1173.
- ROIHA, H., J. R. MILLER, L. C. WOODS and D. M. GLOVER, 1981 Arrangements and rearrangements of sequence flanking the two types of rDNA insertion in *D. melanogaster*. *Nature* **290**: 749–753.

- SHARP, P. M., and W.-H. LI, 1989 On the rate of DNA sequence evolution in *Drosophila*. *J. Mol. Evol.* **28**: 398–402.
- SHIELDS, D. C., and P. M. SHARP, 1989 Evidence that mutation patterns vary among *Drosophila* transposable elements. *J. Mol. Biol.* **207**: 843–846.
- SHIELDS, D. C., P. M. SHARP, D. G. HIGGINS and F. WRIGHT, 1988 “Silent” sites in *Drosophila* genes are not neutral: evidence of selection among synonymous codons. *Mol. Biol. Evol.* **5**: 704–716.
- Swofford, D. L., 1991 PAUP phylogenetic analysis using parsimony. Illinois Natural History Survey, Champaign.
- VANDERWIEL, P. L., D. F. VOYTAS and J. F. WENDEL, 1993 Copia-like retrotransposable element evolution in diploid and polyploid cotton (*Gossypium L.*). *J. Mol. Evol.* **36**: 429–447.
- XIONG, Y., and T. H. EICKBUSH, 1988a The site-specific ribosomal DNA insertion element *R1Bm* belongs to a class of non-long-terminal-repeat retrotransposons. *Mol. Cell. Biol.* **8**: 114–123.
- XIONG, Y., and T. H. EICKBUSH, 1988b Functional expression of a sequence-specific endonuclease encoded by the retrotransposon *R2Bm*. *Cell* **55**: 235–246.
- XIONG, Y., and T. H. EICKBUSH, 1990 Origin and evolution of retroelements based upon their reverse transcriptase sequences. *EMBO J.* **9**: 3353–3362.
- XIONG, Y., W. D. BURKE, J. L. JAKUBCZAK and T. H. EICKBUSH, 1988 Ribosomal DNA insertion elements *R1Bm* and *R2Bm* can transpose in a sequence specific manner to locations outside the 28S genes. *Nucleic Acids Res.* **16**: 10561–10573.

Communicating editor: A. G. CLARK