Proc. Natl. Acad. Sci. USA
Vol. 90, pp. 6260–6264, July 1993
Genetics

Several short interspersed repetitive elements (SINEs) in distant species may have originated from a common ancestral retrovirus: Characterization of a squid SINE and a possible mechanism for generation of tRNA-derived retroposons

(*Loligo bleekeri*/lysine tRNA)

KAZUHIKO OHSHIMA*†, RYUTA KOISHI*†, MAMI MATSUO*†, AND NORIHIRO OKADA*†‡

*Institute of Biological Sciences, University of Tsukuba, Tsukuba 305, Ibaraki, Japan; and †Faculty of Bioscience and Biotechnology, Tokyo Institute of Technology, 4259 Nagatsuta-cho, Midori-ku, Yokohama 227, Japan

Communicated by Russell F. Doolittle, March 8, 1993

ABSTRACT **Using labeled transcripts generated *in vitro* from squid total genomic DNA as a probe, we isolated and characterized a SINE that is present in the squid genome. The squid SINE appears to be derived from a $tRNA^{Lys}$. When the consensus sequences of five different SINEs with a $tRNA^{Lys}$-like structure from distantly related species, including squid, were aligned, we found in the tRNA-unrelated region two sequence motifs that were almost identical among these five SINEs. This observation suggests a common evolutionary origin for these SINEs and/or some function(s) for these motifs. Similar sequences were unexpectedly found to be present in sequences complementary to the U5 regions of several mammalian retroviruses whose primer is a $tRNA^{Lys}$. On the basis of these findings, we present a model for the generation of SINEs. We propose that they are derived from a "strong-stop DNA" with a primer $tRNA^{Lys}$ that is an intermediate in the reverse transcription of certain retroviruses. Our model suggests that a certain group of SINEs may have been generated by horizontal transmission, although it is not clear whether information was transmitted via a similar retrovirus or via an RNA or DNA of a SINE.**

In higher eukaryotes, protein-coding genes constitute 10% of the genome at most, and the remainder of the genome is composed of a variety of repetitive sequences. Repetitive sequences in eukaryotic genomes can be classified into two groups: tandemly repeated sequences and dispersed sequences (1). The dispersed sequences can be further classified into two categories on the basis of size: long interspersed elements (LINEs), which include L1 sequences, and short interspersed elements (SINEs), such as the primate *Alu* family and the rodent type 1 and type 2 families (2).

Repetitive sequences can be classified in a different way, according to the mechanisms by which they are generated. Tandemly repeated sequences are generated by gene duplication at the DNA level (1, 3), while in the case of dispersed elements another mechanism, known as retroposition, has been characterized. During retroposition, information in a cellular RNA transcript flows back into the genome via a complementary DNA intermediate (4, 5).

Retroposon is the name given to a repetitive element that is amplified by retroposition (4). Nonviral retroposons can be classified into three groups—processed retropseudogenes, SINEs, and LINEs (5). In 1985, three laboratories, including ours, reported that mammalian SINEs, with the exception of the human *Alu* family (6, 7), are derived from tRNAs (8–12). At that time, SINEs were believed to be restricted to mammals, since none had been found in the genomes of birds or *Drosophila* (4, 5). Recently, it has been shown that SINEs are widespread in the animal kingdom (for reviews, see refs. 13 and 14) and in plants (15, 16). These findings indicate that retroposition by way of SINE amplification continuously generates genetic and structural variations in the genomes of many more animal and plant species than had previously been supposed.

In general, SINEs are not simple tRNA pseudogenes but have a composite structure, consisting of a region homologous to a tRNA, a tRNA-unrelated region, and an A+T-rich region (13). Previously, it was suggested that the tRNA-related region of several SINEs may have originated from tRNA itself, rather than from tDNA, because a CCA sequence, like that present at the 3′ end of all mature tRNA species, is found in the tRNA-related region of these SINEs (13, 17). This view contrasts with that of Deininger and Daniels (18), who proposed that SINEs may have been generated from tDNAs that accumulated mutations that did not hinder the intrinsic functions of tRNAs. At present, however, the molecular mechanism by which the composite structure of SINEs has been generated during evolution remains unknown. In this paper, we present a possible model for the initial generation of SINEs.§

RESULTS

Transcription *in vitro* of total genomic DNA from squid resulted in production of a discrete transcript of about 250 nucleotides (Fig. 1, lane 1). In view of the sensitivity of this transcription to α-amanitin, this RNA was concluded to be transcribed by RNA polymerase III (Fig. 1, lanes 2 and 3). Using the transcript as a probe, we isolated six phage clones, localized the DNA loci that hybridized to the probe, and determined their sequences. Fig. 2 shows the alignment of these sequences and a consensus sequence, which is typical of a SINE sequence, composed of a tRNA-related region and a tRNA-unrelated region. The SINE is designated as the squid SK family, where S stands for squid and K stands for lysine (see below). The average sequence divergence was 7%. The six clones appear to be divided into two subfamilies: the sequence divergences of NO25, NO6, NO22, and NO28 are low (4.2% on average), whereas those of NO4 and NO31 are relatively high (12.7% on average). Four cloned DNAs belonging to the group with low sequence divergence were active as templates for transcription *in vitro* in a HeLa cell

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

‡To whom correspondence should be addressed at: Faculty of Bioscience and Biotechnology, Tokyo Institute of Technology, 4259 Nagatsuta-cho, Midori-ku, Yokohama 227, Japan.

§The sequences reported in this paper have been deposited in the GenBank data base (accession nos. D14860–D14865).

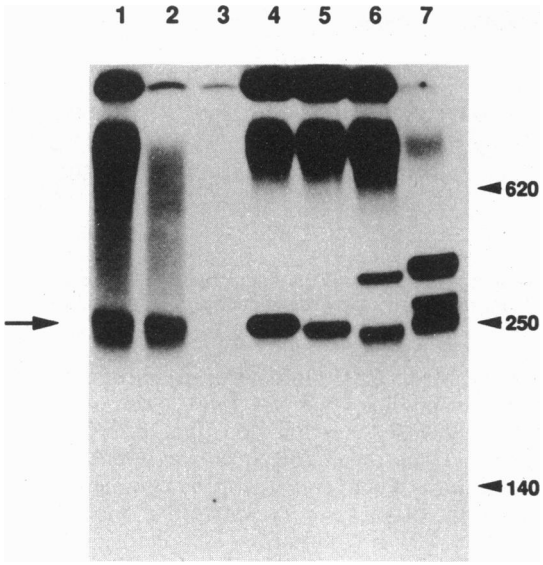

FIG. 1. *In vitro* transcription of total genomic DNA and cloned DNAs in a HeLa cell extract. Total genomic DNA from squid (*Loligo bleekeri*) was transcribed in the absence of α-amanitin (lane 1) and in the presence of α-amanitin at 2 μg/ml (lane 2) or 200 μg/ml (lane 3). The templates used were plasmid DNAs designated NO6 (lane 4), NO25 (lane 5), and NO22 (lane 6) and phage DNA designated NO28 (lane 7). Arrow shows the discrete transcript of about 250 nucleotides. Arrowheads indicate positions of markers. Electrophoresis was performed in an 8% polyacrylamide gel at 600 V for 4 hr.

extract (Fig. 1, lanes 4–7). The other two clones, NO4 and NO31, were not active as templates for transcription *in vitro*, a result that confirms the high sequence divergence among these clones.

A computer-assisted homology search revealed that the tRNA-related region of the SK family was most similar to $tRNA^{Lys}$ from rabbit (19) (Fig. 3). The extent of the similarities between the tRNA-related region of the SK family and the $tRNA^{Lys}$ sequences from mammals and squid is about the same (Fig. 3). No other similar tRNA was detected in a computer compilation of tRNA sequences (EMBL release no. 26.0). Therefore, it is very likely that $tRNA^{Lys}$ is a progenitor of the SK family, although it is possible that some unknown tRNA species from a mollusk are more similar to the SK family than the $tRNA^{Lys}$. The CCA sequence, present at the 3′ end of all mature tRNA species, is retained in the tRNA-related region of the SK family, as is also the case in several other SINEs with a $tRNA^{Lys}$-like structure (see *Discussion*).

DISCUSSION

A Superfamily of $tRNA^{Lys}$-Related SINEs. About 20 families of SINEs that are derived from tRNAs have been reported to date. Among these SINEs, 9 families, including the squid SK family, are similar to $tRNA^{Lys}$ (Table 1). Although they are similar to $tRNA^{Lys}$, it is sometimes very difficult to identify the actual parental tRNA species of several families of SINEs, because some tRNA sequences are very similar to one another and the consensus sequence of a repetitive family has often diverged far from the original tRNA sequence. In the case of the charr *Fok* I family, however, it seems likely that this family really is derived from $tRNA^{Lys}$, given its extensive similarity to this tRNA (79% homology, including the aminoacyl-stem region) (20). Since the tRNA-unrelated regions of the *Fok* I family and the salmon *Sma* I family are similar to each other, it also seems probable that the *Sma* I family is derived from a $tRNA^{Lys}$ (20). The tortoise Pol III/SINE resembles a $tRNA^{Lys}$ and a $tRNA^{Thr}$ to a similar extent, so the tRNA species that is the source of this SINE cannot yet be identified, even from an improved consensus sequence of this family (17). The origin of the *Galago* type 2 family (25) was originally proposed to be $tRNA^{Met}$ (9), but careful reexamination of the similarities between the *Galago* type 2 family and tRNAs showed that it is more likely to have originated from $tRNA^{Lys}$ (21). As for the parental tRNA species of the rodent type 2 *Alu* family (26), two groups (8, 10) have shown that this family most closely resembles $tRNA^{Lys}$, whereas another group (9) has claimed that $tRNA^{Ser}$ is the most likely parental species. Three families—namely, salmonid *Hpa* I (20), octopus OK (K.O., R.K., and N.O., unpub-

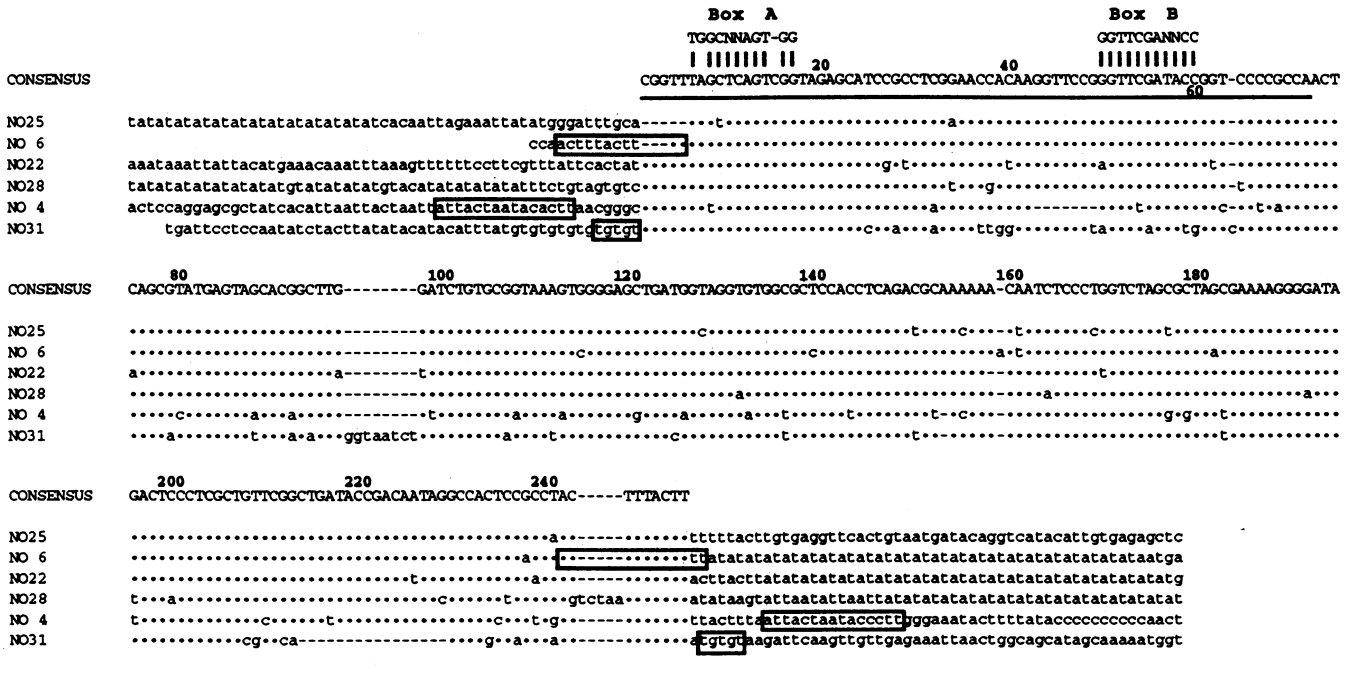

FIG. 2. Sequences of members of the squid SK family and a consensus sequence. The $tRNA^{Lys}$-related region is underlined. The two internal promoters are indicated above the consensus sequence. Direct repeats are boxed.

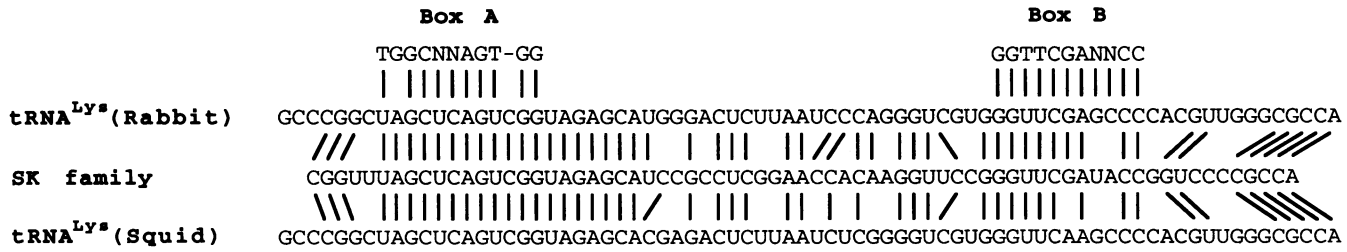

FIG. 3. Sequence comparisons between the tRNA-related segment in the consensus sequence of the squid SK family and the tRNA^{Lys} sequences from rabbit and squid. The sequence of the rabbit tRNA^{Lys} is taken from ref. 19 and the gene sequence for the squid tRNA^{Lys} is our unpublished result (M.M. and N.O.).

lished data), and tobacco TS (16)—are also included in this group, but the extent of their similarities to tRNA^{Lys} is low. We have tentatively listed these 9 families of SINEs in Table 1 as members of a superfamily of tRNA^{Lys}-related SINEs. However, it seems necessary to validate our hypothesis about the existence of a superfamily of tRNA^{Lys}-related SINEs by some other criteria before the final allocation of each family of repetitive sequences to this superfamily (see below).

Several SINEs in Distant Species Exhibit Similarities to One Another. We found that, among the families of SINEs in the superfamily of tRNA^{Lys}-related SINEs in Table 1, the tRNA-unrelated regions of five families of SINEs—rodent type 2, charr *Fok* I, salmon *Sma* I, tortoise Pol III/SINE, and squid SK families—exhibit similarities to one another (data not shown). It is noteworthy that SINEs from phylogenetically distant taxa such as rodent, tortoise, fish, and squid are similar to one another. This finding prompted us to attempt the alignment of these five SINEs, as shown in Fig. 4*a*. When at least four nucleotides are identical at the same position of the aligned sequences, they are highlighted in black. Of 78 positions in the tRNA^{Lys}-related region, 39 positions are highlighted in this way. To our surprise, we found that in the tRNA^{Lys}-unrelated region, the two sequence motifs GATCTG and TGG, at a distance of 10–11 nucleotides, were highly conserved. Although two mismatches and one deletion are present in the *Fok* I and in the *Sma* I sequences, respectively, the complete sequences of these motifs are present in the SINEs from squid, rodent, and tortoise. These results provide convincing evidence that the similarities are significant.

Table 1. Classification of three superfamilies of SINEs

SINE	Species	Ref.
Superfamily of tRNA^{Lys}-related SINEs		
Galago type 2 family	*Galago*	21
Rodent type 2 (B2) family	Mouse, rat, hamster	10
Tortoise Pol III/SINE	Tortoise	17
Charr *Fok* I family	*Salvelinus* spp.	20
Salmon *Sma* I family	Chum and pink salmon	12
Squid SK family	*Loligo bleekeri*	This paper
Salmoid *Hpa* I family	All salmonid species	20
Octopus OK family	Octopus	Unpublished
Tobacco TS family	Tobacco	16
Superfamily of tRNA^{Arg}-related SINEs		
SMα family	*Schistosoma mansoni*	22
Pig PRE-1 family	Pig	23, 24
Octopus OR1 family	Octopus	Unpublished
Octopus OR2 family	Octopus	Unpublished
Superfamily of tRNA^{Gly}-related SINEs		
Rabbit C family	Rabbit	10
Bovine and goat 73-bp repeat	Bovine, goat	10
Rice pSINE$_1$	Rice	15

A Possible Model for the Initial Generation of tRNA-Derived SINEs. It is possible that the two conserved sequences GATCTG and TGG have functions that led to their generation by convergent evolution. It is also possible, however, that these motifs were generated from a common evolutionary ancestor and have a function that has led to their conservation during evolution.

Unexpectedly, sequences similar to these motifs were detected in sequences complementary to the U5 regions of several mammalian retroviruses [SRV-1 (27), SRV-2 (28), MPMV (29), MMTV (30), HIV-1 (31), SIVcpz (32), EIAV (33), BIV127 (34), SA-OMVV (35)], and these sequences were located at distances from the respective primer-binding sites similar to the distance between the 3′ end of the tRNA-related region and these two motifs in the SINEs (Fig. 4*b*). All these retroviruses are presumed to use tRNA^{Lys} as a primer tRNA during reverse transcription because of the presence of a sequence complementary to tRNA^{Lys} at their respective primer-binding sites.

In our model for the initial generation of SINEs (Fig. 5), the 3′-terminal sequence of a tRNA^{Lys} (15–18 nucleotides), including the CCA sequence, hybridizes to the primer-binding site in the viral genome. Reverse transcription proceeds from the CCA end toward the 5′ end of the genome. The main product of reverse transcription *in vitro* is a single-stranded DNA with tRNA^{Lys} at its 5′ terminus, which is known as "strong-stop DNA." During reverse transcription *in vivo*, the transcribed DNA "jumps" to the 3′ terminus of the viral genome because of the presence of the duplicated R region. From the sequence similarities described above, we propose a model wherein the strong stop DNA with the tRNA^{Lys} becomes a SINE after several further unidentified processes. The primer tRNA is not removed, and it is copied instead into DNA or inserted directly into the genome as a covalent tRNA–DNA hybrid, thereby creating a tRNA^{Lys} pseudogene. This model explains the peculiar presence of the CCA sequence at the 3′ terminus of the tRNA-related regions of several SINEs (see Introduction). A similar mechanism for the generation of a tRNA pseudogene was originally suggested by Saigo (36), and Okada and coworkers (13, 17) discussed such a mechanism subsequently. The model presented here has been briefly described elsewhere (37).

When Saigo's model was proposed in 1986, it appeared to be very difficult to detect any similarities between SINEs and U5 regions of retroviruses, since the evolutionary rate of mutation of genes in retroviruses is 10^6 times higher than that of nuclear genes (38). We suspected at that time that no similarities would remain, even if the model were correct. However, the present study has revealed sequence similarities between SINEs and retroviruses. Nonetheless, such similarities may not be extensive enough to convince us of an evolutionary relationship. If we could isolate a pair consisting of a new animal or plant SINE and a new retrovirus in which the sequences of the regions mentioned in this study were much more similar to one another, we would be more confident of the validity of our model.

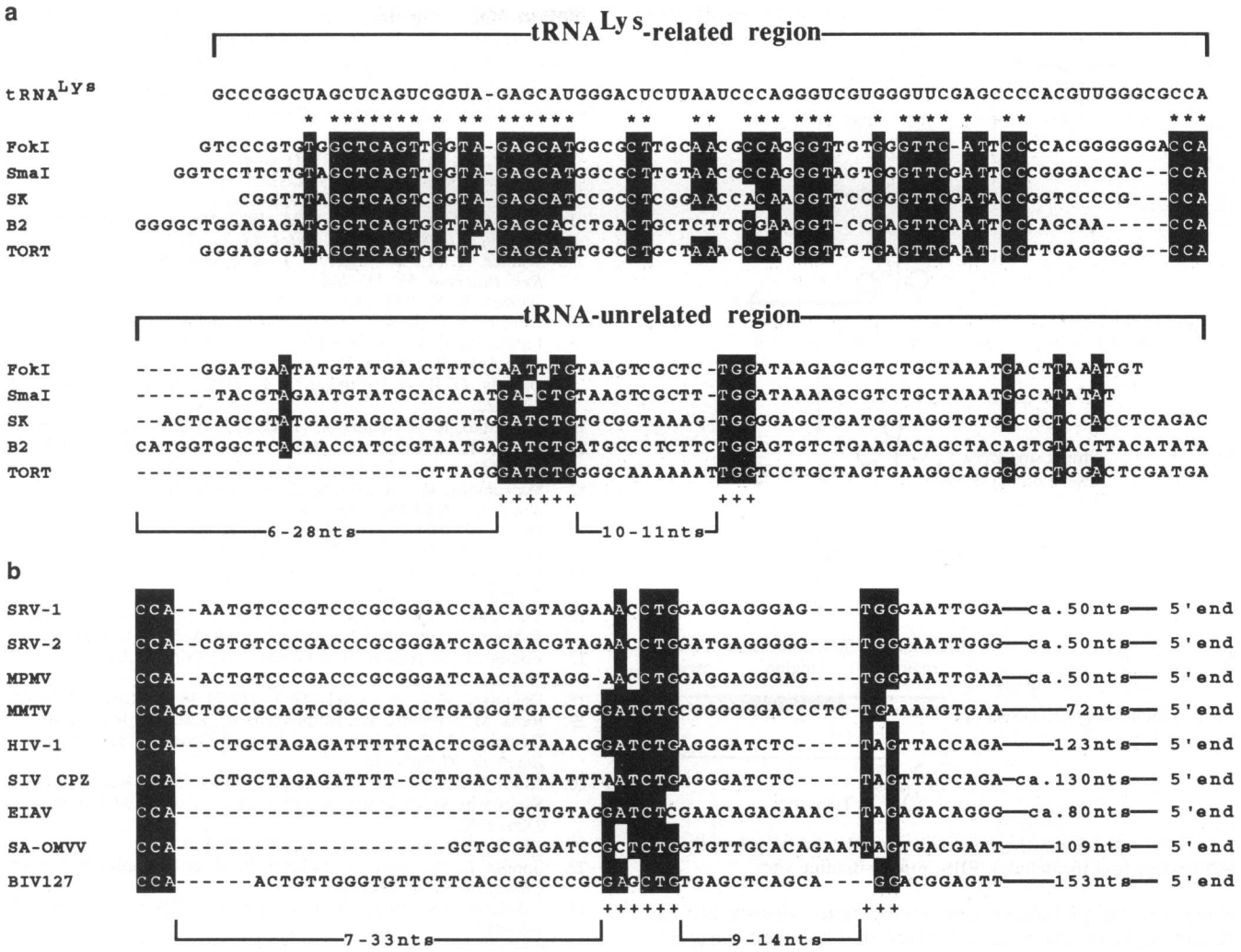

FIG. 4. SINEs may have been generated from a strong-stop DNA. (*a*) Alignment of five SINEs that appear to be derived from $tRNA^{Lys}$. Nucleotides at positions at which at least four nucleotides are identical are highlighted in black. When the highlighted nucleotides are also identical to a nucleotide in $tRNA^{Lys}$ (rabbit), the nucleotides are indicated by a star. Two conserved regions of GATCTG and TGG are indicated by plus signs. Deletions are shown by bars. Fok I, Sma I, SK, B2, and TORT stand for the charr *Fok* I family, the salmon *Sma* I family, the squid SK family, the rodent type 2 (B2) family, and the tortoise Pol III/SINE family, respectively. (*b*) Alignment of sequences complementary to the U5 regions of several retroviruses that use $tRNA^{Lys}$ as a primer. The CCA sequences at the end of each primer-binding site and the two conserved regions are highlighted in black. The two conserved regions are indicated by plus signs. References for these retroviruses are indicated in the text. Deletions are shown by bars. The distance in nucleotides to the 5′ end of the R region of the viruses is indicated.

Our model appears to support our original notion that the five families of SINEs described above, which include the rodent type 2, may have been derived from $tRNA^{Lys}$ [the progenitor of the rodent type 2 *Alu* family was proposed to be $tRNA^{Ser}$ by Daniels and Deininger (9), see above].

With regard to the origin of SINEs, $tRNA^{Lys}$ appears to be the most frequent progenitor of SINEs (13, 14). If this preponderance reflects the relative proportion of retroviruses that use $tRNA^{Lys}$ as a primer among all retroviruses, it can be concluded that such retroviruses are abundant in the biological world. If retroviruses that use $tRNA^{Lys}$ were selectively utilized in the origination of SINEs, it is likely that the sequences of SINEs with a $tRNA^{Lys}$-like structure have some function(s) and may confer some selective advantage on the respective hosts (21). If all retroviruses have the potential to generate a SINE from a strong-stop DNA, we should be able to find many kinds of SINE with a variety of tRNA-like structures that correspond to the variety of primer tRNA species used by many kinds of retrovirus in the animal kingdom. As shown in Table 1, the second most abundant SINEs are those with a $tRNA^{Arg}$-like structure, and the third are those with a $tRNA^{Gly}$-like structure. We have tentatively designated these groups as superfamilies, hoping to find homologous retroviruses with the respective primer tRNAs in the future.

On the Possible Generation of SINEs by Horizontal Transmission. Our model also suggests that SINEs in distantly related species may have been generated by horizontal transmission. With respect to the question of how several SINEs in distant species happen to have similar sequences, there are two possible answers. One possibility is that a similar retrovirus infected distantly related organisms, such as rodents and squid, during a very short evolutionary time period and that the strong-stop DNA with a primer tRNA gave rise independently to the respective SINE within each organism. Phylogenetic analysis of retrons (39), based on their reverse transcriptases, has revealed the presence of related retrotransposable elements in very different taxa, and it has been proposed that the retrotransposons have spread horizontally (40, 41). In the case of the plant *copia*-like retrotransposons, more concrete evidence for horizontal transmission has been presented recently (42). Moreover, it is well established that many endogenous retroviruses have spread to new species by horizontal transfer, and there is also

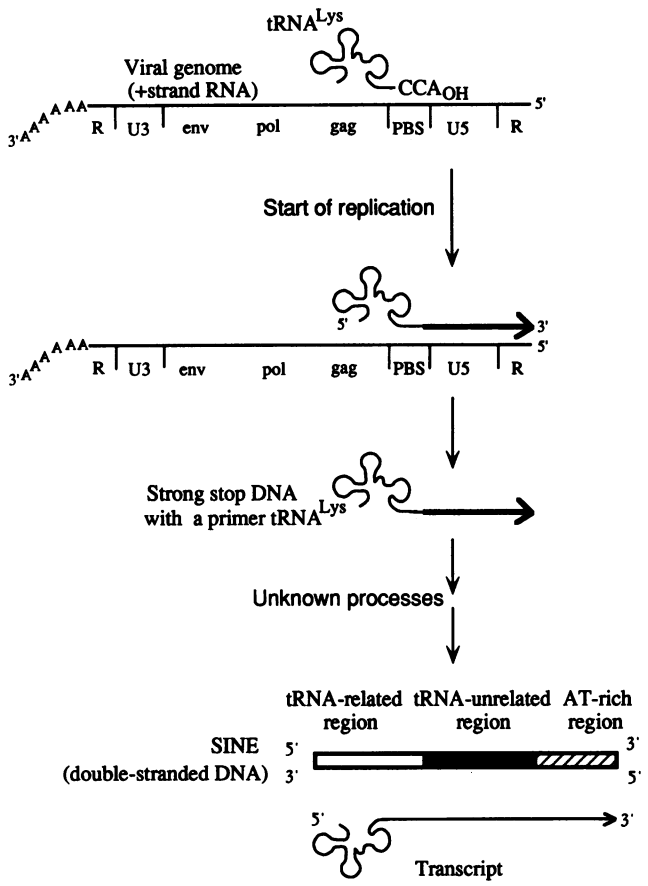

FIG. 5. Model for the mechanism of the initial generation of SINEs (see text for details). PBS, primer-binding site.

every reason to believe that endogenous viruses are the descendants of exogenous viruses that infected germ-line cells (40). The generation of SINEs by independent infection of retroviruses is not unlikely. However, it should be noted that sequences of the tRNA^{Lys}-related regions of five SINEs (highlighted nucleotides in Fig. 4*a*) are highly conserved among these SINEs. It is difficult to envision that such coincidentally highly conserved stretches arose as a result of the independent generation of SINEs within each organism unless these conserved regions were selected during a process of retroposition within a germ cell or during a fixation process in each species. The alternative possibility is that genetic information in a primordial SINE in one organism was transmitted horizontally to another organism as the transcript of this SINE or as the SINE DNA itself. The fact that the rice pSINE$_1$ is more similar to the rabbit C family than to tRNA^{Gly} (our own alignment, not shown here) can be explained only by assuming such a process. The two possibilities described above are not, however, mutually exclusive. More examples are required if we are to identify the dominant mechanism for the dispersion of SINEs throughout the biological world.

Doolittle *et al.* (40) noted that, judged from their present-day distribution, genuine retroviruses must have appeared well after the evolution of vertebrates and perhaps even after the emergence of mammals. If this were the case and if SINEs were generated from retroviruses as we propose, the SINEs evolved and have been amplified only after the time when mammals emerged. This view may explain why there are many SINEs within mammalian genomes and fewer SINEs in nonmammalian genomes (see Introduction). Since the establishment of mammals, SINEs with retroviral genetic information consisting of a long terminal repeat with a primer tRNA might have been flowing continuously into the genomes of vertebrates and also invertebrates, contributing to the remodeling of the host genomes and endowing them with genetic variability during evolution.

1. Singer, M. & Berg, P. (1991) *Genes and Genomes* (University Science Books, Mill Valley, CA).
2. Singer, M. F. (1982) *Cell* **28**, 433–434.
3. Denison, R. A. & Weiner, A. M. (1982) *Mol. Cell. Biol.* **2**, 815–828.
4. Rogers, J. (1985) *Int. Rev. Cytol.* **93**, 231–279.
5. Weiner, A. M., Deininger, P. L. & Efstratiadis, A. (1986) *Annu. Rev. Biochem.* **55**, 631–661.
6. Weiner, A. M. (1980) *Cell* **22**, 209–218.
7. Ullu, E. & Tschudi, C. (1984) *Nature (London)* **312**, 171–172.
8. Lawrence, C. B., McDonnell, D. P. & Ramsey, W. J. (1985) *Nucleic Acids Res.* **13**, 4239–4252.
9. Daniels, G. R. & Deininger, P. L. (1985) *Nature (London)* **317**, 819–822.
10. Sakamoto, K. & Okada, N. (1985) *J. Mol. Evol.* **22**, 134–140.
11. Endoh, H. & Okada, N. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 251–255.
12. Matsumoto, K., Murakami, K. & Okada, N. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 3156–3160.
13. Okada, N. (1991) *Trends Ecol. Evol.* **6**, 358–361.
14. Okada, N. (1991) *Curr. Opin. Genet. Dev.* **1**, 498–504.
15. Mochizuki, K., Umeda, M., Ohtsubo, H. & Ohtsubo, E. (1992) *Jpn. J. Genet.* **67**, 155–166.
16. Yoshioka, Y., Matsumoto, S., Kojima, S., Ohshima, K., Okada, N. & Machida, Y. (1993) *Proc. Natl. Acad. Sci. USA*, in press.
17. Endoh, H., Nagahashi, S. & Okada, N. (1990) *Eur. J. Biochem.* **189**, 25–31.
18. Deininger, P. L. & Daniels, G. L. (1986) *Trends Genet.* **2**, 76–80.
19. Raba, M., Limburg, K., Burghagen, M., Katze, J. R., Simsek, M., Heckman, J. E., RajBhandary, U. L. & Gross, H. J. (1979) *Eur. J. Biochem.* **97**, 305–318.
20. Kido, Y., Aono, M., Yamaki, T., Matsumoto, K., Murata, S., Saneyoshi, M. & Okada, N. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 2326–2330.
21. Okada, N. (1990) *J. Mol. Evol.* **31**, 500–510.
22. Spotila, L. D., Hirai, H., Rekosh, D. M. & LoVerde, P. T. (1989) *Chromosoma* **97**, 421–428.
23. Takahashi, H., Awata, T. & Yasue, H. (1992) *Animal Genet.* **23**, 443–448.
24. Sugiura, N., Hagiwara, H. & Hirose, S. (1992) *J. Biol. Chem.* **25**, 18067–18072.
25. Daniels, G. R., Fox, G. M., Loewensteiner, D., Schmid, C. W. & Deininger, P. L. (1983) *Nucleic Acids Res.* **11**, 7579–7593.
26. Krayev, A. S., Markusheva, T. V., Kramerov, D. A., Ryskov, A. V., Skryabin, K. G., Bayev, A. A. & Georgiev, G. P. (1982) *Nucleic Acids Res.* **10**, 7461–7475.
27. Power, M. D., Marx, P. A., Bryant, M. L., Gardner, M. B., Barr, P. J. & Luciw, P. A. (1986) *Science* **28**, 1567–1573.
28. Thayer, R. M., Power, M. D., Bryant, M. L., Gardner, M. B., Barr, P. J. & Luciw, P. A. (1987) *Virology* **157**, 317–329.
29. Sonigo, P., Barker, C., Hunter, E. & Wain-Hobson, S. (1986) *Cell* **45**, 375–385.
30. Fasel, N., Buetti, E., Firzlaff, J., Pearson, K. & Diggelmann, H. (1983) *Nucleic Acids Res.* **11**, 6943–6955.
31. Spire, B., Sire, J., Zachar, V., Rey, F., Barré-Sinoussi, F., Galibert, F., Hampe, A. & Chermann, J.-C. (1989) *Gene* **81**, 275–284.
32. Huet, T., Cheynier, R., Meyerhans, A., Roelants, G. & Wain-Hobson, S. (1990) *Nature (London)* **345**, 356–359.
33. Derse, D., Dorn, P. L., Levy, L., Stephens, R. M., Rice, N. R. & Casey, J. W. (1978) *J. Virol.* **61**, 743–747.
34. Garvey, K. J., Oberste, M. S., Elser, J. E., Braun, M. J. & Gonda, M. A. (1990) *Virology* **175**, 391–409.
35. Querat, G., Andoly, G., Sonigo, P. & Vigne, R. (1990) *Virology* **175**, 434–447.
36. Saigo, K. (1986) *Nucleic Acids Res.* **14**, 7815.
37. Okada, N. & Ohshima, K. (1993) *J. Mol. Evol.*, in press.
38. Gojobori, T. & Yokoyama, S. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 4198–4201.
39. Temin, H. M. (1989) *Nature (London)* **339**, 254–255.
40. Doolittle, R. F., Feng, D.-F., Johnson, M. S. & McClure, M. A. (1989) *Quart. Rev. Biol.* **64**, 1–30.
41. Xiong, Y. & Eickbush, T. H. (1990) *EMBO J.* **9**, 3353–3362.
42. Voytas, D. F., Cummings, M. P., Konieczny, A., Ausubel, F. M. & Rodermel, S. R. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 7124–7128.