論文 / 著書情報
Article / Book Information

題目(和文)	
Title(English)	Molecular biological studies on the origin of short interspersed repetitive elements (SINEs) and on the enzymatic machinery responsible for the retroposition of these elements
著者(和文)	大島一彦
Author(English)	KAZUHIKO OHSHIMA
出典(和文)	学位:博士(理学), 学位授与機関:東京工業大学, 報告番号:甲第3099号, 授与年月日:1995年12月31日, 学位の種別:課程博士, 審査員:
Citation(English)	Degree:Doctor of Science, Conferring organization: Tokyo Institute of Technology, Report number:甲第3099号, Conferred date:1995/12/31, Degree Type:Course doctor, Examiner:
学位種別(和文)	博士論文
Type(English)	Doctoral Thesis

DOCTORAL THESIS

Molecular Biological Studies on the Origin of Short Interspersed Repetitive Elements (SINEs) and on the Enzymatic Machinery Responsible for the Retroposition of These Elements

Kazuhiko Ohshima

Graduate School of Bioscience and Biotechnology

Tokyo Institute of Technology

1995

CONTENTS

Molecular Biological Studies on the Origin of Short Interspersed Repetitive Elements (SINEs) and on the Enzymatic Machinery Responsible for the Retroposition of These Elements

PART II.

PART III.

PREFACE

On the diversity of life and the unity of the principle of life

There are manifold organisms on the earth.

On the contrary, it has been verified that the principle of life is quite uniform since the formation of molecular biology and the following remarkable progress of the science. The genetic code, for example, is, in principle, ubiquitous throughout the biological world. How did life achieve the diversity using such universal principles? It is possible that there are an astronomical number of "whole systems of molecule" which are realizing respective species. However, are the number of such systems infinite? Alternatively, are these systems only variations of a limited number of prototypes?

One possible molecular basis of such prototypes has been proposed by Slack et al. (1993). They proposed the concepts of zootype and phylotypic stage. The zootype is the spatial order of anterior expression limits of Hox cluster and some other genes, and is expressed most clearly at a particular stage of embryonic development: the phylotypic stage for each individual taxon.

If we suppose the existance of such prototypic systems, it raises a fundamental question of whether the prototypic systems were constructed by the merest chance or their composition was ruled by unknown laws.

It is generally accepted that the diversity of life results from biological evolution. Following this notion, we could regard "whole systems of molecule" of respective species, as well as their phenotypes, as being evolved. Therefore, the question mentioned above should be also considered in the context of evolution. I would like to express the question in other words: whether such composition of the prototypic systems happened to be made up during biological evolution or manners of their composition are limited and these systems turn out the composition like them whatever process of evolution they pass.

To answer this question, we must elucidate the nature of the prototypic systems. However, in order to understand composition of the prototypic systems, it might be useful to know the organization of genomes from various living species because it provides information about history of its alteration. This imformation might help us to understand the logic of "system conversion" or alteration to the composition of the prototypic systems. In the present studies, I attempted to elucidate the origin of short interspersed elements (SINEs) and the enzymatic machinery responsible for the retroposition of these elements on purpose to have insight into laws of rearrangement of metazoan genomes.

Reference

Slack, J. M. W., Holland, P. W. H., and Graham, C. F. (1993). The zootype and phylotypic stage. Nature 361, 490-492.

GENERAL INTRODUCTION

The "reverse" flow of genetic information from RNA back into DNA is known as "retroposition," and each transposed information of element is known as a "retroposon" (Rogers, 1985; Weiner et al., 1986). Highly repetitive elements in eukaryotic genomes (Singer, 1982), such as short interspersed elements (SINEs) and long interspersed elements (LINEs), and processed retropseudogenes which abound in mammalian genomes are all included in this category. Retroposition of these retroelements (Temin, 1989), in addition to "classic" mechanisms that operate at the DNA level, such as mutation and recombination (Singer & Berg, 1991), is thought to be a major evolutionary force contributing to the maintenance of the remarkable fluidity of eukaryotic genomes (Rogers, 1985; Weiner et al., 1986; Okada, 1991a).

SINEs are short (approximately 80-400 bp) repetitive elements, which are often present at more 10^5 copies per genome. The best-characterized SINE is the human *Alu* family (Rubin et al., 1980). The human *Alu* family was shown to be homologous to 7SL RNA (Weiner, 1980), and it was, thus, demonstrated to be derived from this RNA during evolution in 1984 (Ullu & Tschudi). In 1985, three laboratories, including ours, reported that mammalian SINEs, with the exception of the human *Alu* family, are derived from tRNAs (Lawrence et al., 1985; Daniels & Deininger, 1985; Sakamoto & Okada, 1985; Endoh & Okada, 1986; Matsumoto et al.,1986). At that time, SINEs were believed to be restricted to mammals, since none had been found in the genomes of birds or *Drosophila* (Rogers, 1985; Weiner et al., 1986). However, it has been shown that SINEs are present in the genomes of many more animal and plant species than had previously been supposed.

The tRNA-derived SINEs are not simple pseudogenes for tRNAs but have a composite structure, with a region homologous to a tRNA, a tRNA-unrelated region, and an AT-rich region (Okada, 1991a,b). The molecular mechanism by

which the composite structure of SINEs has been generated during evolution remains to be determined.

Here we can inquire, as follows. (i) Are tRNA-derived SINEs ubiquitous in the biological world, or are those confined to some specific taxa? (ii) How has been constructed the composite structure of tRNA-derived SINEs? (iii) How have SINEs been amplified? In other words, what is the nature of enzymatic machinery of SINE amplification? (iv) What is biological significance of SINEs?

In the present thesis, I describe the results of studies about the first three problems in the following three parts, and provide possible answers. Moreover, I discuss the last problem in the chapter of General Discussion.

References

Daniels, G. R. & Deininger, P. L. (1985). Repeat sequence families derived from mammalian tRNA genes. Nature 317, 819-822.

Endoh, H. & Okada, N. (1986). Total DNA transcription *in vitro*: a procedure to detect highly repetitive and transcribable sequences with tRNA-like structures. Proc. Natl. Acad. Sci. U.S.A. 83, 251-255.

Lawrence, C. B., McDonnell, D. P. & Ramsey, W. J. (1985). Analysis of repetitive sequence elements containing tRNA-like sequences. Nucl. Acids Res. 13, 4239-4252.

Matsumoto, K., Murakami, K. & Okada, N. (1986). Gene for lysine $tRNA_1$ may be a progenitor of the highly repetitive and transcribable sequences present in the salmon genome. Proc. Natl. Acad. Sci. U.S.A. 83, 3156-3160.

Okada, N. (1991a). SINEs: Short interspersed repeated elements of the eukaryotic genome. *Trends. Ecol. Evol.* 6, 358-361.

Okada, N. (1991b). SINEs.*Curr. Opin. Genet. Develop.* 1, 498-504.

Rogers, J. H. (1985). The structure and evolution of retroposons. Int. Rev. Cytol. 93, 231-279.

Rubin, C. M., Houck, C. M., Deininger, P. L., Friedmann, T., & Schmid, C. W. (1980). Partial nucleotide sequence of the 300-nucleotide interspersed repeated human DNA sequences. Nature 284, 372-374.

Sakamoto, K. & Okada, N. (1985). Rodent type 2 Alu family, rat identifier sequence, rabbit C family, and bovine or goat 73-bp repeat may have evolved from tRNA genes. J. Mol. Evol. 22, 134-140.

Singer, M. F. (1982). SINEs and LINEs: highly repeated short and long interspersed sequences in mammalian genomes. Cell 28, 433-434.

Singer, M. & Berg, P. (1991). Genes & Genomes, University Science Books, Mill Valley, CA.

Temin, H.M. (1989). Retrons in bacteria. Nature 339, 254-255.

Ullu, E. & Tschudi, C. (1984). *Alu* sequences are processed 7SL RNA genes. Nature 312, 171-172.

Weiner, A. M. (1980). An abundant cytoplasmic 7S RNA is complementary to the dominant interspersed middle repetitive DNA sequence family in the human genome. Cell 22, 209-218.

Weiner, A. M., Deininger, P. L. & Efstratiadis, A. (1986). Nonviral retroposons: genes, pseudogenes, and transposable elements generated by the reverse flow of genetic information. Annu. Rev. Biochem. 55, 631-661.

PART I

Generality of the tRNA origin of short interspersed repetitive elements (SINEs): Characterization of three different tRNA-derived retroposons in the octopus

ABSTRACT

Transcription *in vitro* of total genomic DNA of *Octopus vulgaris* resulted in Pol III transcripts of 200 to 500 nucleotides in length. These transcripts were used as probes and, as a result, three different kinds of short interspersed element (SINE) were isolated and characterized. Two SINEs, designated the octopus OR1 and OR2 families, seem to have been derived from $tRNA^{Arg}$. The other SINE, designated the octopus OK family, have originated from tRNA, but the parental tRNA species cannot be identified due to sequence divergence from the original tRNA sequence. The OR1 and OR2 families exhibit considerable similarity to one another, in the 5' region of the tRNA-unrelated region as well as in the $tRNA^{Arg}$-related region, an observation that suggests that these two families may have had the same origin in evolution. The three SINEs together constitute at least 6% of the genome of *O. vulgaris*. Results of a dot hybridization experiment suggest that the OR1 and OR2 families are present in fewer species than the OK family and that the OK family exists in many species of octopodid. The present observations indicate that SINEs have generally been derived from tRNAs in invertebrates, as well as in vertebrates, and that retroposition is widely involved in the genetic and structural variability of invertebrate genomes. It has been demonstrated that $tRNA^{Arg}$ and $tRNA^{Lys}$ appear to be tRNA species found as progenitors of vertebrate SINEs, and the same preference for progenitor species of tRNAs seems also to be a feature of invertebrate SINEs.

INTRODUCTION

In 1985, studies in three laboratories, including ours, revealed that SINEs appear to have evolved from specific tRNAs (Lawrence et al., 1985; Daniels & Deininger, 1985; Okada et al., 1985; Sakamoto & Okada, 1985). Examples are the rodent type 2 (B2) family, the rodent identifier (ID) sequence, the *Galago* type 2 repeat, the rabbit C family, and the bovine or goat 73-bp repeat (artiodactyl C repeat). All the SINEs listed are mammalian SINEs. Initially, because of the absence of processed pseudogenes for tubulin and cytochrome *c* in chickens (Lopata et al., 1983; Limback & Wu, 1983) and pseudogenes for U2 snRNA in *Xenopus* (Mattaj & Zeller, 1983), it was believed that retroposons are not common outside mammals. However, in 1986, repeated sequences derived from tRNAs were reported to be present in the genomes of tortoise, newt and salmon (Endoh & Okada, 1986; Matsumoto et al., 1986) and it was demonstrated that those sequences in tortoise and salmon are real retroposons (Endoh et al., 1990; Matsumoto et al., 1986; Kido et al., 1991). In the case of the newt family, it was shown recently that an original unit of the newt family had been generated by retroposition and that amplification at the DNA level had occurred to produce tandem arrays of the newt repeated unit during evolution (Nagahashi et al., 1991). These studies suggested that SINEs are more common in the genomes of vertebrates than had previously been postulated. Subsequently, SINEs derived from tRNAs were characterized in several invertebrate and vertebrate species , such as silkworm (*Bombyx mori*) (Adams et al., 1986), sea urchin (Nisson et al., 1988), human blood fluke (*Schistosoma mansoni*); (Spotila et al., 1989), pig (Singer et al., 1987; Takahashi et al., 1992; Sugiura et al., 1992) and tobacco (Yoshioka et al., 1993).

The parental tRNA of each SINE appears not to have been chosen at random. The accumulated sequencing data for SINEs indicates that $tRNA^{Lys}$ is the most common progenitor of SINEs (for reviews, see Okada, 1991a,b). Several SINEs

with a $tRNA^{Lys}$-like structure have been characterized in various species from mammals to fish. The second most likely tRNA species to be chosen appears to be $tRNA^{Arg}$. Two SINEs with a $tRNA^{Arg}$-like structure, one from a vertebrate and the other from an invertebrate, namely the pig (Takahashi et al., 1992; Sugiura et al., 1992) and the human blood fluke (Spotila et al., 1989), respectively, have been reported to date. To obtain further information about the nature and origin of invertebrate SINEs, our group isolated and characterized SINEs from octopus. Our data indicate the generality of the tRNA origin of SINEs and support the hypothesis that there must be some general mechanism for the generation of tRNA-derived SINEs that is common to vertebrates and invertebrates (Ohshima & Okada, 1994).

MATERIALS AND METHODS

(a) Materials

DNA from four species of octopus (*Octopus vulgaris, O. membranaceus, O. ocellatus* and *Paroctopus dofleini*) was extracted as described elsewhere (Blin & Stafford, 1976). All sequencing reagents were obtained from Takara Shuzo Co. (Kyoto, Japan). Restriction enzymes were obtained from Takara Shuzo Co. and TOYOBO Co. (Tokyo, Japan). Oligonucleotides for use as primers were synthesized by the protocol provided by Applied Biosystems Japan Co. (Tokyo, Japan). Various labeled nucleotides and a labeling kit for the random primer method were purchased from Amersham Co. (Tokyo, Japan) and Nippon Gene Co. (Tokyo, Japan), respectively.

(b) Preparation of a HeLa cell extract and RNA synthesis *in vitro*

HeLa cells were adapted to suspension culture by Dr. F. Hanaoka (Riken Institute, Tsukuba). A cell extract was prepared as described elsewhere (Manley et al., 1980) with slight modifications (Talkington et al., 1980) for the more efficient extraction of RNA polymerases. RNA was synthesized *in vitro* essentially as described elsewhere (Manley et al., 1980). The RNA was isolated by extraction with phenol and was subjected to electrophoresis on a 10% (w/v) polyacrylamide denaturing gel (Maxam & Gilbert, 1980).

(c) Isolation of phage and plasmid DNAs and DNA sequencing

A genomic library was constructed by ligation of the genomic DNA of *O. vulgaris* that has been completely digested with *Eco* RI and λgt10 arms. In the case of the OK family, ^{32}P-labeled RNA, transcribed from total DNA, was used as the probe to isolate phage clones from the library. In the case of the OR1 family, ^{32}P-labeled RNA transcribed from NO11 plasmid DNA or the randomly ^{32}P-

labeled 1.7 kb *Hind* III DNA fragment of the same plasmid was used as the probe. In the case of the OR2 family, the randomly ^{32}P-labeled 480 bp *Hinc* II / *Eco* RI DNA fragment of NO11 plasmid DNA was used as the probe. Respective inserts of positive clones were subcloned in pUC vectors and the region of the octopus SINE was located by Southern blot hybridization (Southern, 1975). In some cases, a series of deletion mutants of the subcloned DNA was constructed, according to the previously described procedure (Henikoff, 1984), and the sequence of each was determined by the dideoxychain-termination method (Sanger et al., 1977). In other cases, the subfragment was ligated into M13 bacteriophage vectors (Messing & Vieira, 1982) and then sequenced. The sequence of each clones was determined by me except for OK-4 and OK-7. The two clones were determined by Ryuta Koishi.

(d) Estimation of the relative levels and the number of copies per genome of three SINEs of the octopus by dot hybridization

Dot hybridization was used to estimate the relative levels and the number of copies per genome of three SINEs in the genome of *O. vulgaris* and the other species of octopus. A probe specific to each SINE was synthesized by the polymerase chain reaction (PCR) by use of a pair of primers designed by reference to the sequence of each SINE and plasmid DNA that contained each SINE as template. In the case of the OK family, the sequences of the two primers were 5'-AGGAGTGGCTGAGAGGTAAG-3' and 3'-TGACTTTCTTCGGACAGCAT-5', and they were designated KFO and KR2, respectively. pOK-9 was used as template. The length of the product of PCR was 159 bp. In the case of the OR1 family, the primers were 5'-CCCAGTGGTTAGGGCAGCG-3' and 3'-AAGGCAACTAAGCCTAGTTGG-5', designated TRF1 and TRR1, respectively, and pOR1-13 was used as the template. The length of the product of PCR was 335 bp. In the case of the OR2 family, the primers were 5'-

GAGGCGCGTGGCTTAGTG-3' and 3'-GGCCCGGGTACTCGGACCGA-5', designated R2F1 and R2R1, respectively, and pOR2-1 was used as the template. The length of the product of PCR was 227 bp. The 5'-end of each fragment of DNA (5 pmol) was labeled by use of [γ-^{32}P]ATP and polynucleotide kinase. The specific activity of the three probes was adjusted to 2 x 10^6 cts per min/pmol and hybridization was performed in 50% formamide, 1 M NaCl and 1% SDS at 42 °C for 12 hours. Washing was performed twice in 2 x SSC (SSC is 0.15 M NaCl, 0.015 M trisodium citrate, pH 7) at room temperature for five minutes, twice in 2 x SSC at 55 °C for 30 minutes and finally in 0.1 x SSC at room temperature for one minute. The intensity of each spot was quantified with BAS2000 (FUJI PHOTO FILM Co.; Tokyo, Japan).

RESULTS AND DISCUSSION

(a) The octopus OR1 family may have been derived from a $tRNA^{Arg}$

Transcription *in vitro* of total genomic DNA of *O. vulgaris* resulted in production of RNA transcripts of various lengths from 100 nucleotides upwards. In the presence of α-amanitin at the low concentration of 2 μg/ml, transcription of large RNAs was inhibited, but RNAs of 200 to 500 nucleotides were still synthesized. These RNAs were shown to be transcribed by RNA polymerase III (data not shown). Using these RNAs as probe, I isolated one phage DNA, designated NO11, which had two different $tRNA^{Arg}$-like sequences with a distance of about 1 kb between them in the 2.8 kb insert of the NO11 phage. *In vitro* runoff transcription assay of the NO11 phage DNA revealed that the upstream tRNA-like structure could serve as a template for RNA polymerase III, while the downstream structure was inactive in transcription *in vitro* (data not shown). Using the transcript from the upstream tRNA-like sequence as a probe, I screened a genomic library of *O. vulgaris* and found that this sequence constitutes a SINE family, which I named the octopus OR1 family (O stands for octopus and R for arginine; see below). As described in the next section, the downstream tRNA-like sequence was found to be a member of another SINE family, designated the octopus OR2 family.

Figure 1 shows the alignment of five sequences that belong to the OR1 family. The tRNA-related segment of the OR1 family was shown to exhibit the highest degree of similarity to $tRNA^{Arg}$ (Fig. 2a). This conclusion was reached as follows. First, a computer file that contained tRNA sequences was made from EMBL data base (release no. 26.0) and the best-matching species of tRNA were chosen from the file by a homology search. These tRNAs were $tRNAs^{Arg}$ from *Drosophila*, yeast and cow, with a $tRNA^{Lys}$ from rabbit matching slightly less level. Next, the tRNA-related segment of the family and the secondary structures

of one of the tRNAs were compared. As a result, I concluded that $tRNA^{Arg}$ is the most probable candidate for the origin of the octopus OR1 family.

In Figure 3, the sequence of the NO13 phage DNA, which contains two units of the OR1 family, is shown. The upstream unit was designated OR1-13-1 and the downstream unit, OR1-13-2. The 321 bp sequence between these two units was found to be almost identical to the 321 bp sequence adjacent to the 5' end of OR1-13-1. The OR1-13-2 unit is flanked by direct repeats of 11 bp, which suggests that OR1-13-2 was originally generated by retroposition during evolution of the octopus. It is probable that the unusual structure of this locus was generated by unequal crossing-over between the 3' end of the unit of one allele and a site 320 nucleotides upstream of the unit of the other allele. It should be noted that the end of one unit (OR1-13-1, in this case) must be excised exactly to generate this unusual structure. This requirement suggests that the tail end of the OR1 family may be a hot spot for recombination.

(b) The OR2 family may also have been derived from a $tRNA^{Arg}$

To examine whether the downstream tRNA-like sequence in the NO11 phage DNA also represents a repetitive family, I screened a genomic library of *O. vulgaris* using a ^{32}P-labeled DNA fragment that contained only the downstream sequence as a probe. Several phage clones were isolated and the sequences of three clones were determined (NO1, NO2 and NO6). In the NO6 clone, two units were present in tandem in the same transcriptional orientation (the upstream unit was named OR2-6-1 and the downstream unit was named OR2-6-2). The five sequences, including the downstream sequence in NO11 that was designated OR2-11, constitute another SINE family in the octopus, designated the octopus OR2 family (Fig. 4).

In an attempt to identify the tRNA species that is most similar to the tRNA-related region of this family, the same procedure was adopted as in the case of the

OR1 family. A computer search showed that $tRNAs^{Arg}$ from various species most closely resemble the tRNA-related region of the family. Analysis of secondary structures confirmed that $tRNA^{Arg}$ is the most probable candidate for the origin of this family. Figure 2b shows sequence and structural similarities between the tRNA-related segment of the OR2 family and $tRNA^{Arg}$ from *Drosophila.*

(c) The OR1 and OR2 families share a common ancestor

Figure 5 shows the alignment of the two consensus sequences of the OR1 and OR2 families. In the 5' regions of the tRNA-unrelated regions, the sequences exhibit a high degree of similarity. This region is flanked by vertical lines of dots and, of 42 nucleotides in this region, 31 nucleotides (74%) match. Since members of an arbitrary pair of SINEs generally do not show such extensive similarity in the tRNA-unrelated region as the members of this pair of SINEs, this similarity suggests that these two families must have the same origin. Next, I compared the two tRNA-related segments of the OR1 and OR2 families with $tRNAs^{Arg}$, as shown in Figure 6. The upper two lines in Figure 6 show the tRNA-related segments of the two families, and the identical sequences are highlighted in black. Five $tRNAs^{Arg}$ selected by a computer search are aligned below these sequences, and nucleotides are highlighted in grey when, in more than three of the five $tRNAs^{Arg}$, the same nucleotide is found at that position. If a nucleotide highlighted in black in the upper two lines is identical to one highlighted in grey in the lower lines, it is marked by a star (*). As a result of this analysis, I found that the two families have in common nine nucleotides, each of which is indicated by a cross (+), that none of the five $tRNAs^{Arg}$ contain. These nine nucleotides are considered to be diagnostic for the two repetitive families. The most prominent characteristic of both SINEs is that they contain one additional insertion of U in the second promoter region. In standard tRNAs, there are two U residues in this T-loop region and no natural tRNA species is known that has three U residues in this

region. From these results, I can conclude that the OR1 and OR2 families are derived from a common ancestral sequence.

(d) Isolation and characterization of the octopus OK family

Using trsanscripts from total genomic DNA of *O. vulgaris* as a probe, I and R. Koishi characterized a third SINE family, designated the octopus OK family, where O stands for octopus and K for lysine (see below). Four phage clones were isolated and the sequences of these clones were determined (Fig. 7). The average sequence divergence of the family was 6.8%. Each member of this family has a tRNA-like structure in the 5' part of the sequence (nucleotides 5 to 76) and an $(AT)_n$-$(GT)_n$ stretch in the middle of the tRNA-unrelated region (nucleotides 165 to 221). In NO2 phage DNA, one unit of the OK family (designated OK2*) is integrated in the opposite orientation to a second unit of the OK family (designated OK2), located on the 5' side of the $(AT)_n$-$(GT)_n$ stretch. OK2* has a further insertion of 52 nucleotides in the tRNA-unrelated region.

(e) The octopus OK family may have been derived from a tRNA

In a computer-assisted homology search, the tRNA-related region of the OK family was found to be most similar to yeast $tRNA^{Trp}$, mouse $tRNA^{Ala}$ and mammalian $tRNA^{Lys}$. Alignment of the tRNA-related segment of the OK family with three tRNAs is shown in Figure 8. Figure 2c shows the sequence and structural similarities of the tRNA-related segment of the OK family and the $tRNA^{Lys}$, as an example. The putative secondary structure of the OK family (Fig. 2c, left) indicates that, in addition to the promoter sequences for RNA polymerase III, the nucleotides that are conserved or semi-conserved in all tRNAs, such as U8, C32, U33, A37, A38, G45, G46, and U47, are retained, confirming that the OK family is in fact derived from a tRNA molecule. Recently, the sequence of the $tRNA^{Lys}$ from squid was determined (Matsuo, M. et al., 1995). The extents of the

homology between the tRNA-related region of the OK family and $tRNAs^{Lys}$ from mammals and from squid are about the same. No sequences of mollusk tRNAs have been published except for the $tRNA^{Lys}$ from squid, and I cannot at present identify the tRNA species that is the actual origin of the octopus OK family. However, since the four stems in the secondary structure of the tRNA-related segment of the OK family do not include many base-pairs (Fig. 2c), I speculate that there may be no tRNA species that is much more similar to the tRNA-related region of the family than the three tRNAs shown in Figure 8, even though some tRNA species with a similar degree of homology may be found.

(f) The three tRNA-derived SINEs in *O. vulgaris* constitute at least 6% of the genome

To estimate the relative level and the number of copies per genome of these three SINEs in the genome of *O. vulgaris*, I performed dot hybridization experiments. First, I synthesized three DNA fragments that were specific to the respective SINEs by PCR. I labeled the 5'-termini of these DNAs with polynucleotide kinase and [γ-^{32}P]ATP and used the labeled DNAs, with identical specific activity, as probes. To examine the specificity of these probes, I performed a cross-hybridization experiment, as shown in Figure 9a. In rows I, II, and III, results of experiments with the labeled probes for the OK, OR1 and OR2 families, respectively, are shown. In columns 1, 2 and 3, DNAs of pOK-9, pOR1-13 and pOR2-1 were dotted, respectively. The results show that each probe hybridized specifically to the respective DNA and no cross-hybridization occurred. Using these probes, I performed the next experiment, as shown in Figure 9b. In rows I, II and III, progressively decreasing amounts of total DNA from the octopus were dotted in the upper line (from 400 ng to 12.5 ng) and, in the lower line, decreasing amounts of the cloned DNA specific to each SINE (from 4.0 ng to 125 pg) were dotted. From this experiment, I estimated the levels of these SINEs relative to each

other and to the entire genome, as shown in Table 1. The relative levels of the OK, OR1 and OR2 families in the octopus genome were calculated as 3.5%, 1.3% and 0.8%, respectively. Since each probe used in this experiment was a particular cloned DNA, which may not hybridize to all of the members of a given family, these estimates are minimum values. The total complement of these SINEs probably constitute more than 6% of the genome. Since I do not know the exact size of the octopus genome, I cannot accurately calculate the copy number of each SINE. However, if the genome of *O. vulgaris* is assumed to contain 1×10^9 bp per haploid genome, the copy numbers of the OK, OR1 and OR2 families can be estimated to be 8.5×10^4, 3.3×10^4 and 2.8×10^4, respectively.

(g) The OK family exists in many species of octopodid

To examine whether these three SINEs exist in species of octopus other than *O. vulgaris,* I performed a further dot hybridization experiment using total genomic DNA from four species of octopodid, as shown in Figure 9c. Columns I, II, and III indicate experiments with the probes specific to the OK, OR1 and OR2 families, respectively. Each probe is the same as used in the experiment shown in Figure 9a and b. This experiment showed that all the species examined contained the OK family and that *Paroctopus dofleini* did not contain the OR1 and OR2 families (see Table 2). These results indicate that amplification of the members of the OR1 and OR2 families occurred after *P. dofleini* had diverged from the other three species and that amplification of members of the OK family had probably occured before this event. This hypothesis is consistent with the values for the average sequence divergence of three SINEs: the value of the sequence divergence of the OK family is highest among the three (see Table 1). Also, it should be noted that the amounts of the OR1 and OR2 families in the genome of *O. vulgaris* are much higher than those in the other species. This observation suggests that these two families have

been amplified several times in the genome of *O. vulgaris* during evolution (Murata et al., 1993).

(h) Generality of the tRNA origin of SINEs

In the present study, we characterized three different SINEs in the genome of the octopus. They all have features common to vertebrate SINEs, as follows. First, many, if not all, units of these SINEs are flanked by direct repeats of 10 to 20 bp of flanking sequence at the 5' and 3' ends. This feature is believed to be generated by a staggered nick in the target site, with the inserted sequence being joined to the separated ends (Van Arsdell et al., 1981; Jagadeeswaran et al., 1981). Second, a tRNA-like structure is present in the 5' region of each unit, whose promoters may direct RNA synthesis from the 5' boundary of the unit (Okada, 1991a,b). To date, seven invertebrate tRNA-derived SINEs, including three octopus SINEs described here, have been reported. They are widely distributed in divergent taxa, including mollusk (octopus and squid; Ohshima & Okada, 1994; Ohshima et al., 1993; see also Part I, II), arthropod (silkworm; Adams et al., 1986), echinoderm (sea urchin; Nisson et al., 1988) and flatworm (human blood fluke; Spotila et al., 1989). The present study strongly suggests that tRNA-derived SINEs are ubiquitous in invertebrates, as well as in vertebrates. Third, as in the case of other tRNA-derived retroposons, these SINEs are not simple pseudogenes of tRNA but have a composite structure, consisting of a tRNA-related region and a tRNA-unrelated region (Okada, 1991a,b). Fourth, as in the case of vertebrate retroposons, invertebrate SINEs become a hot spot for genomic rearrangement. In the present study, we reported two examples of rearrangement mediated by invertebrate SINEs. One is the case in which the tail end of a member of the octopus OR1 family was excised exactly, followed by an unequal crossing-over to generate a tandemly duplicated structure of the OR1 family at one locus. The other is an insertion event of one member of the OK family into another pre-existing member

of the OK family. The fact that mammalian SINEs have a strong tendency to insert into the A-rich tails of each other was discussed extensively by Rogers (1985), who showed that SINEs are particularly prone to insert themselves into the tails of pre-existing SINEs. This insertion produces clusters that may generate new retroposon families and suggests a variety of novel possibilities for evolution. Clustering of SINEs at a particular locus was recently described in the case of the rabbit C family (Krane et. al., 1991). Our observations suggest that retroposition is also widely involeved in the genetic and structural variability of invertebrate genomes.

REFERENCES

Adams, D.S., Eickbush, T.H., Herrera, R.J., and Lizardi, P.M. (1986). A highly reiterated family of transcribed oligo(A)-terminated, interspersed DNA elements in the genome of *Bombyx mori.* J. Mol. Biol. 187, 456-478.

Blin, N. and Stafford, D. W. (1976). A general method for isolation of high-molecular-weight DNA from eukaryotes. Nucl. Acids Res. 3, 2303-2308.

Daniels, G. R. and Deininger, P. L. (1985). Repeat sequence families derived from mammalian tRNA genes. Nature 317, 819-822.

Endoh, H. & Okada, N. (1986). Total DNA transcription *in vitro*: a procedure to detect highly repetitive and transcribable sequences with tRNA-like structures. Proc. Natl. Acad. Sci. U.S.A. 83, 251-255.

Endoh, H., Nagahashi, S., and Okada, N. (1990). A highly repetitive and transcribable sequence in the tortoise genome is probably a retroposon. Eur. J. Biochem. 189, 25-31.

Galli, G., Hofstetter, H., and Birnstiel, M. L. (1981). Two conserved sequence blocks within eukaryotic tRNA genes are major promoter elements. Nature 294, 626-631.

Henikoff, S. (1984). Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene 28, 351-359.

Jagadeeswaran, P., Forget, B. G. , and Weissman, S. M. (1981). Short interspersed repetitive DNA elements in eukaryotes: transposable DNA elements generated by reverse transcription of RNA pol III transcripts. Cell 26, 141-142.

Kido, Y., Aono. M., Yamaki, T., Matsumoto, K., Murata, S., Saneyoshi, M., and Okada, N. (1991). Shaping and reshaping of salmonid genomes by amplification of tRNA-derived retroposons during evolution. Proc. Natl. Acad. Sci. USA 88, 2326-2330.

Krane, D.E., Clark, A.G., Cheng, J-F., and Hardison, R.C. (1991). Subfamily relationship and clustering of rabbit C repeats. Mol. Biol. Evol. 8, 1-30.

Lawrence, C. B., McDonnell, D. P., and Ramsey, W. J. (1985). Analysis of repetitive sequence elements containing tRNA-like sequences. Nucl. Acids Res. 13, 4239-4252.

Limbach, K.J. and Wu, R. (1983). Isolation and characterization of two alleles of chicken cytochrome *c* gene. Nucleic Acids Res. 11, 8931-8950.

Lopata, M.A., Havercroft, J.C., Chow, L.T., and Cleveland, D.W. (1983). Four unique genes required for β tubulin expression in vertebrates. Cell 32, 713-724.

Manley, J. L., Fire, A., Cano, A., Sharp, P. A., and Gefter, M. L. (1980). DNA-dependent transcription of adenovirus genes in a soluble whole-cell extract. Proc. Natl. Acad. Sci. USA 77, 3855-3859.

Matsumoto, K., Murakami, K., and Okada, N. (1986). Gene for lysine $tRNA_1$ may be a progenitor of the highly repetitive and transcribable sequences present in the salmon genome. Proc. Natl. Acad. Sci. USA 83, 3156-3160.

Matsuo, M., Abe, Y., Saruta, Y., and Okada, N. (1995). Mollusk genes encoding lysine tRNA (UUU) contain introns. Gene *in press.*

Mattaj, I.W. and Zeller, R. (1983). U2 small nuclear RNA genes: Tandemly repeated transcription units sharing 5' and 3' flanking homology with other RNA polymerase II transcribed genes. EMBO J. 2, 1883-1891.

Maxam, A.M. and Gilbert, W. (1980). Sequencing end-labeled DNA with base-specific chemical cleavage. Methods Enzymol. 65, 499-560.

Messing, J. and Vieira, J. (1982). A new pair of M13 vectors for selecting either DNA strand of double-digest restriction fragments. Gene 19, 269-276.

Murata, S., Takasaki, N., Saitoh, M., and Okada, N. (1993). Determination of the phylogenetic relationships among Pacific salmonids by using short interspersed elements (SINEs) as temporal landmarks of evolution. Proc. Natl. Acad. Sci. USA 90, 6995-6999.

Nagahashi, S., Endoh, H., Suzuki, Y., and Okada, N. (1991). Characterization of a tandemly repeated DNA sequence family originally deribed by retroposition of $tRNA^{Glu}$ in the newt. J. Mol. Biol. 222, 391-404.

Nisson, P.E., Hickey, R.J., Boshar, M.F., and Crain, W.R. Jr (1988). Identification of a repeated sequence in the genome of the sea urchin which is transcribed by RNA polymerase III and contains the features of a retroposon. Nucleic Acids Res. 16, 1431-1452.

Ohshima, K., Koishi, R., Matsuo, M., and Okada, N. (1993). 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. Proc. Natl. Acad. Sci. USA 90, 6260-6264.

Ohshima, K. and Okada, N. (1994). Generality of the tRNA origin of short interspersed repetitive elements (SINEs): Characterization of three different tRNA-derived retroposons in the octopus. J. Mol. Biol. 243, 25-37.

Okada, N., Endoh, H., Sakamoto, K., and Matsumoto, K. (1985). Many highly repetitive and transcribable sequences are derived from tRNA genes. Proc. Japan Acad. 61, 363-367.

Okada, N. (1991a). SINEs: Short interspersed repeated elements of the eukaryotic genome. Trends in Ecol. & Evol. 6, 358-361.

Okada, N. (1991b). SINEs. Curr. Opin. in Genet. & Dev. 1, 498-504.

Raba, M., Limburg, K., Burghagen, M., Katze, J.R., Simsek, M., Heckman, J.E., RajBhandary, U.L., and Gross, H.J. (1979). Nucleotide sequence of three isoaccepting lysine tRNAs from rabbit liver and SV40-transformed mouse fibroblasts. Eur. J. Biochem. 97, 305-318.

Rogers, J. (1985). The structure and evolution of retroposons. Int. Rev. Cytol. 93, 231-279.

Sakamoto, K. and Okada, N. (1985). Rodent type 2 Alu family, rat identifier sequence, rabbit C family, and bovine or goat 73-bp repeat may have evolved from tRNA genes. J. Mol. Evol. 22, 134-140.

Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74, 5463-5467.

Singer, D. S., Parent, L. J., and Ehrlich, R. (1987). Identification and DNA sequence of an interspersed repetitive DNA element in the genome of the miniature swine. Nucl. Acids Res. 15, 2780.

Southern, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98, 503-517.

Spotila, L. D., Hirai, H., Rekosh, D. M., and LoVerde, P. T. (1989). A retroposon-like short repetitive DNA element in the genome of the human blood fluke, *Schistosoma mansoni*. Chromosoma 97, 421-428.

Sprinzl, M., Hartmann, T., Meissner, F., Moll, J., and Vorderwülbecke, T. (1987). Compilation of tRNA sequences and sequences of tRNA genes. Nucl. Acids Res. 15, r53-r188.

Sugiura, N., Hagiwara, H., and Hirose, S. (1992). Molecular cloning of porcine soluble angiotensin-binding protein. J. Biol. Chem. 267, 18067-18072.

Takahashi, H., Awata, T., and Yasue, H. (1992). Characterization of swine short interspersed repetitive sequences. Animal Genetics 23, 413-418.

Talkington, C. A., Nishioka, Y., and Leder, P (1980). *In vitro* transcription of normal, mutant, and truncated mouse α-globin genes. Proc. Natl. Acad. Sci. USA 77, 7132-7136.

Van Arsdell, S. W., Denison, R. A., Bernstein, L. B., Weiner, A. M., Manser, T., and Gesteland, R. F. (1981). Direct repeats flank three small nuclear RNA pseudogenes in the human genome. Cell 26, 11-17.

Yoshioka, Y., Matsumoto, S., Kojima, S., Ohshima, K., Okada, N., and Machida, Y. (1993). Molecular characterization of a short interspersed repetitive element from tobacco that exhibits sequence homology to specific tRNAs. Proc. Natl. Acad. Sci. USA 90, 6562-6566.

TABLES AND FIGURES

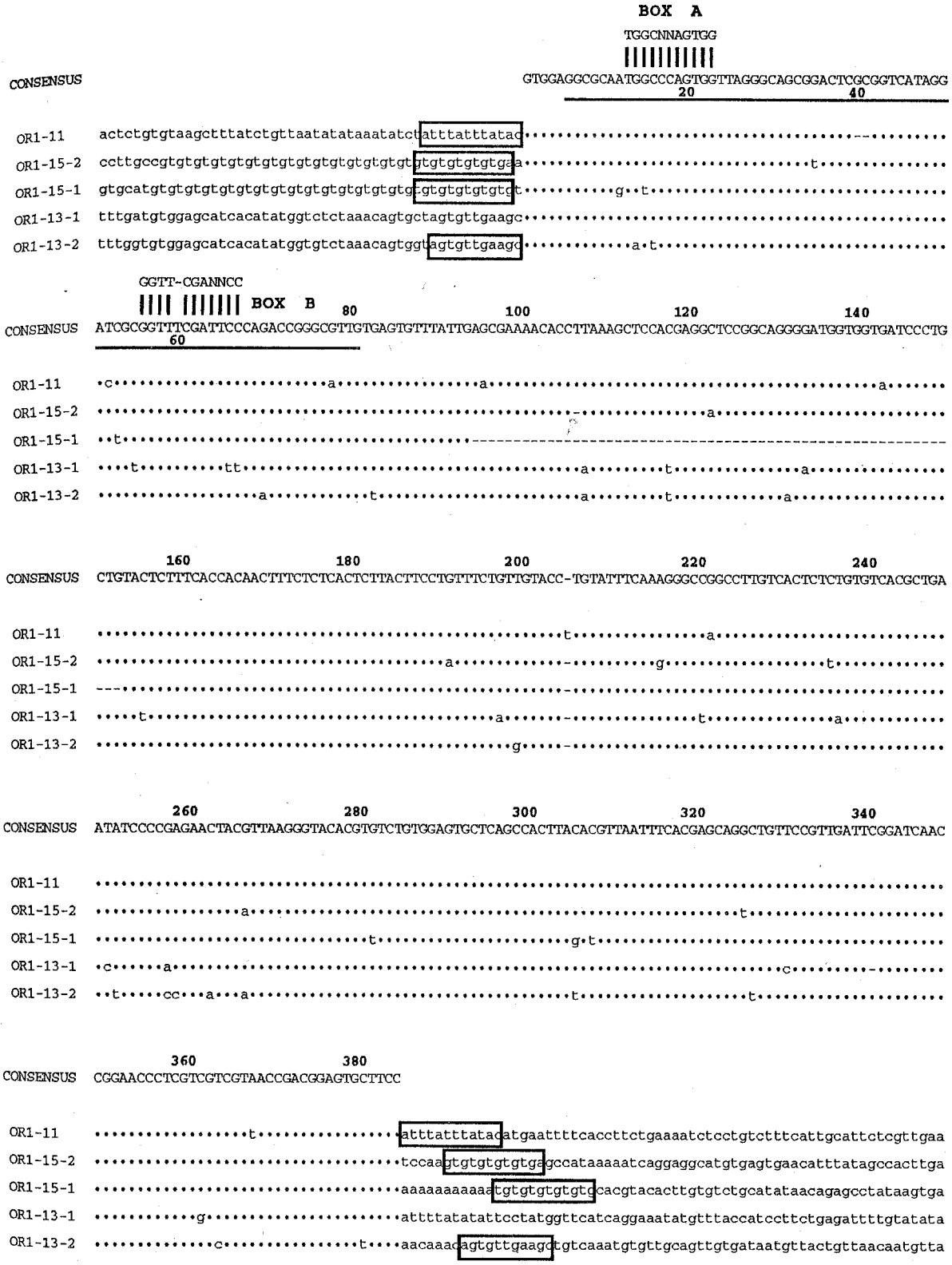


Figure 1. Sequences of members of the octopus OR1 family and the consensus sequence. The sequences of the members of the OR1 family were aligned and the consensus sequence was deduced. Nucleotides identical to those in the consensus sequence are denoted by dots, and different nucleotides are shown by lower-case letters. Gaps introduced to maximize homology are denoted by bars. The $tRNA_{Arg}$-related regionis underlined. Box A and Box B indicate consensus sequences for the promoters of RNA polymerase III (Galli et al., 1981). Direct repeats are boxed.

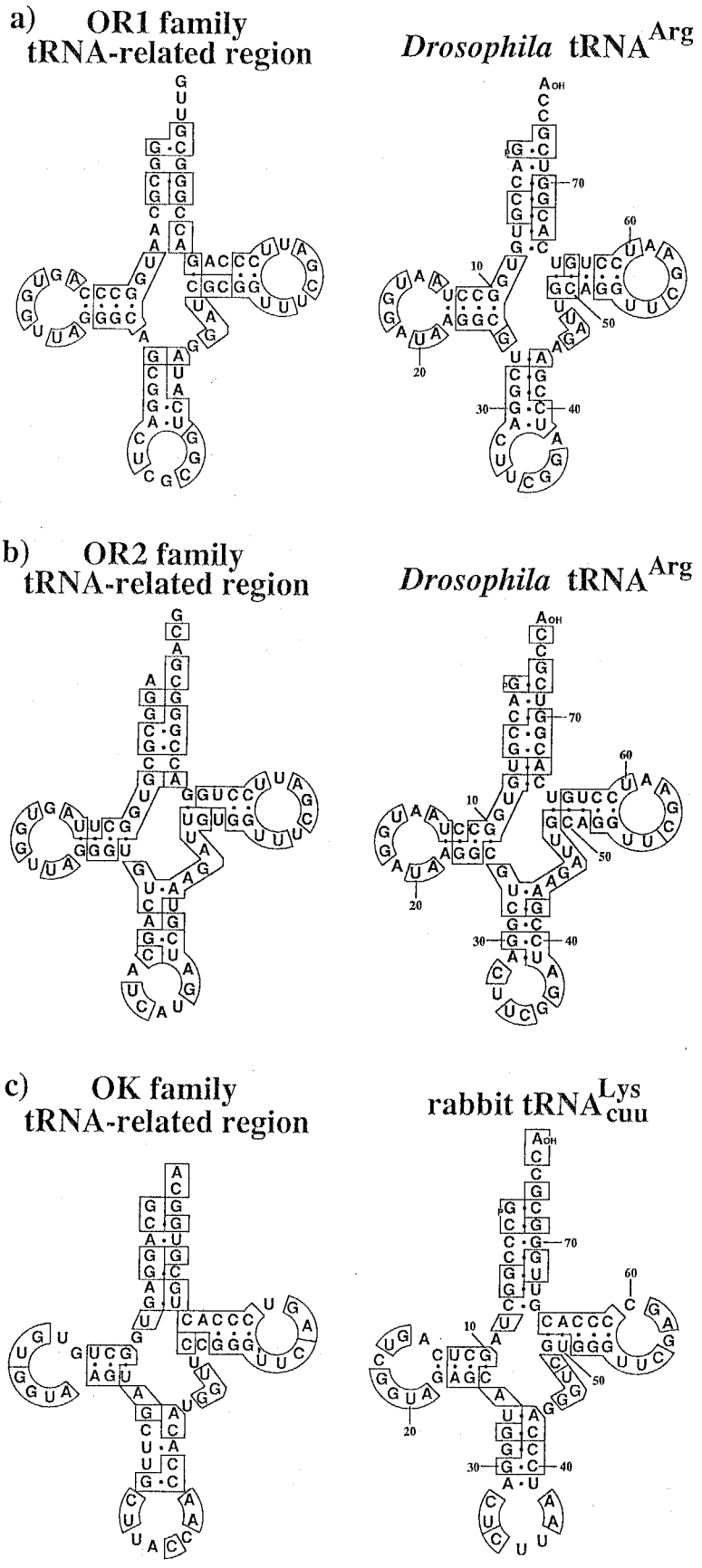


Figure 2. Sequence and structural comparisons between the tRNA-related segments of three octopus SINEs and tRNAs. (a) the tRNA ^{Arg}-related segment of the OR1 family (left) and tRNA ^{Arg} from *Drosophila* (right). (b) the tRNA ^{Arg}-related segment of the OR2 family (left) and tRNA ^{Arg} from *Drosophila* (right). (c) the tRNA-related segment of the OK family (left) and rabbit tRNA ^{Lys} (right). The sequences of tRNA ^{Arg} and tRNA ^{Lys} are taken from Yuki et al. (1986) and Raba et al. (1979), respectively. Identical sequences are boxed. Only the four standard nucleotides are shown. The numbering systems of the tRNAs are indicated.

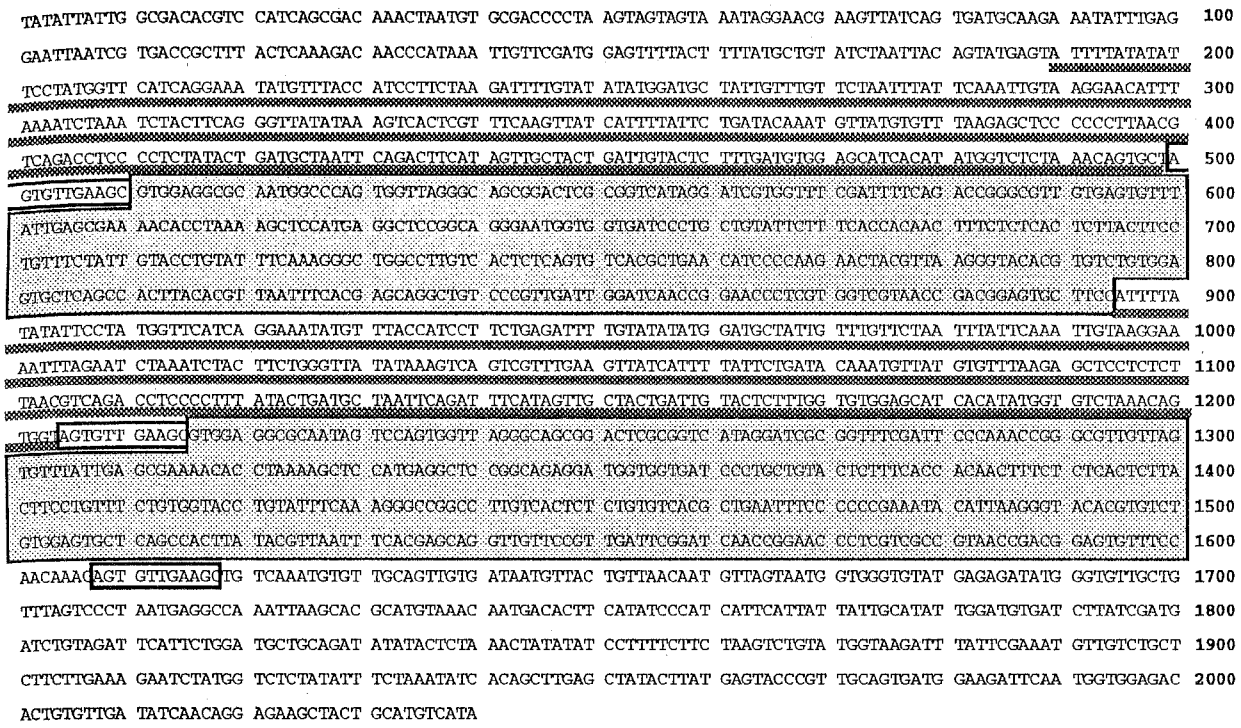


Figure 3. The sequence of the phage DNA of NO13. The phage DNA of NO13 contains two members of the OR1 family with an unusual structure. The two members of the OR1 family are boxed and shaped in grey. A sequence that is almost identical to the sequence between these two members is present in the 5'-flanking region of the upstream unit, and they are indicated by stippled bars. Direct repeats are boxed.

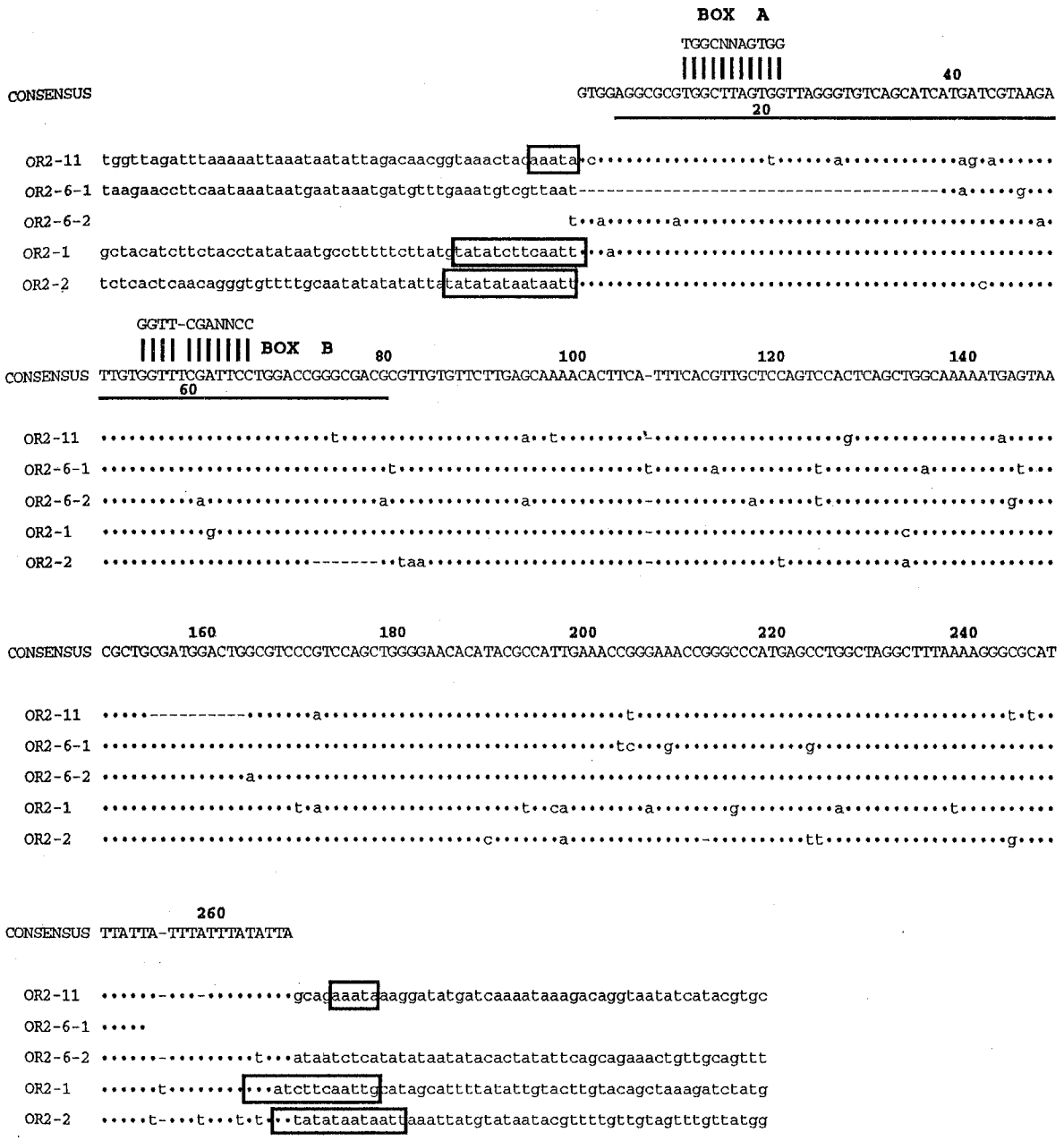


Figure 4. Sequences of members of the octopus OR2 family and the consensus sequence. The tRNA^{Arg}-related region is underlined. Box A and Box B indicate consensus sequences for the promoters of RNA polymerase III (Galli *et al.*, 1981). Direct repeats are boxed.

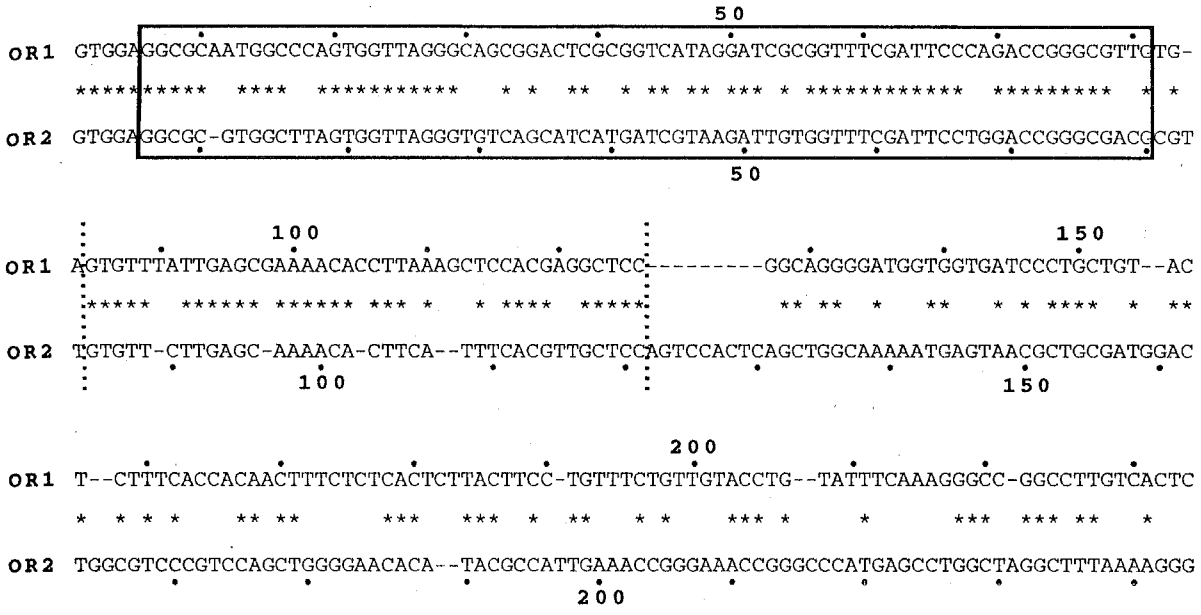


Figure 5. Alignment of the consensus sequences of the OR1 and OR2 families. Identical nucleotides are indicated by asterisks. The $tRNA^{Arg}$-related regions are boxed. Sequences similar to each other in the tRNA-unrelated regions are enclosed by vertical lines of dots.

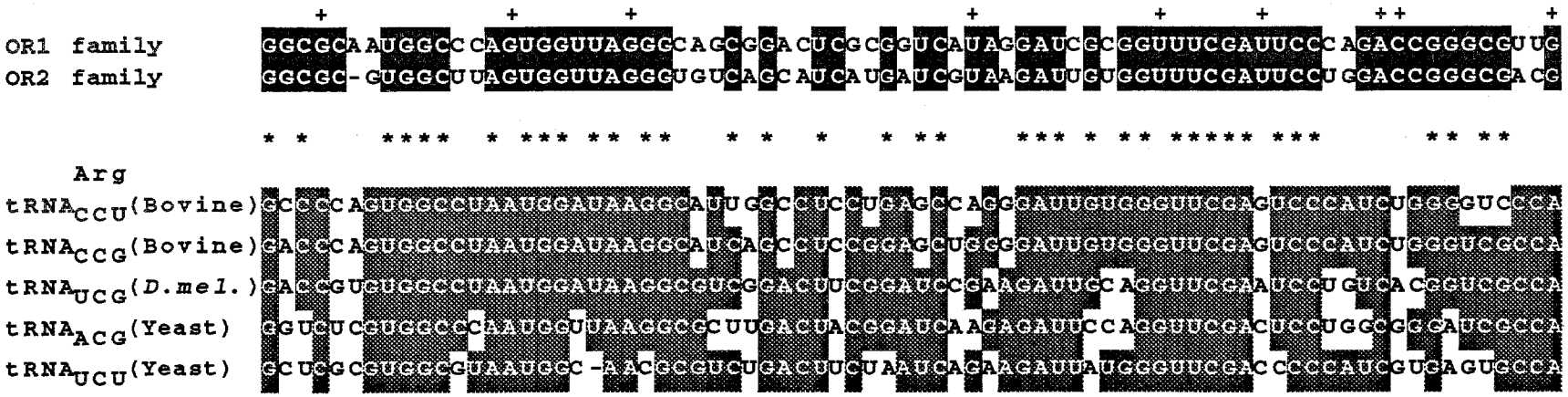


Figure 6. Sequence comparison of the tRNA^{Arg}-related segments of the OR1 and OR2 families with tRNAs^{Arg} from several eukaryotes. The upper two sequences are the tRNA^{Arg}-related segments of the OR1 and OR2 families. Identical sequences are highlighted in black. The lower sequences are those of five tRNAs^{Arg} from various eukaryotes. When, of five nucleotides at a certain position, at least three nucleotides are identical, they are highlighted in grey. When a nucleotide highlighted in black in the upper sequences is identical to that indicated in grey in the lower sequences, it is marked by an asterisk (*). The diagnostic nucleotides for the two SINEs are shown by crosses (+).

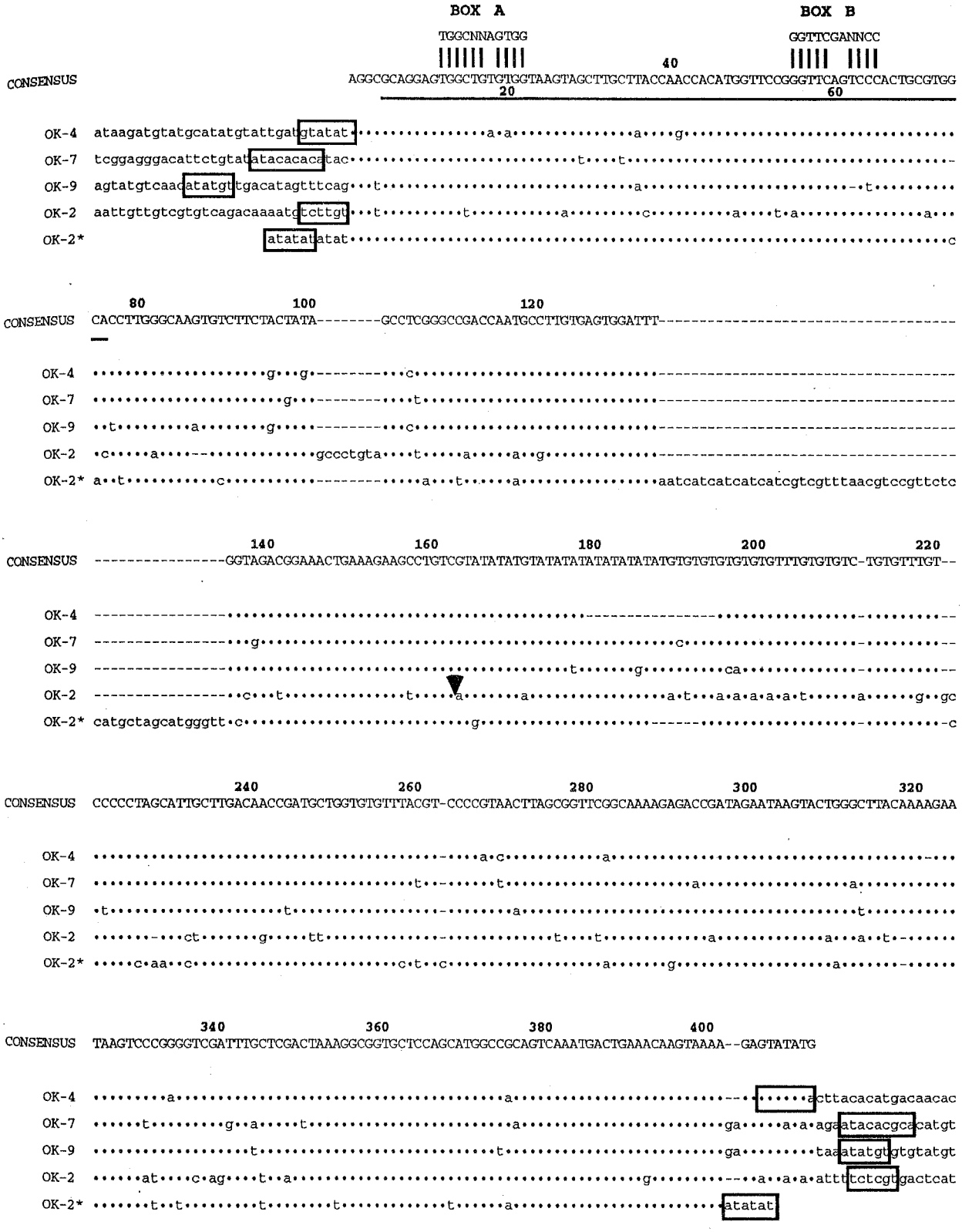


Figure 7. Sequences of members of the octopus OK family and the consensus sequence. The tRNA-related region is underlined. Box A and Box B indicate consensus sequences of promoters for RNA polymerase III (Galli *et al.*, 1981). Direct repeats are boxed. An arrow shows the site at which OK-2* is integrated in the opposite orientation.

```
OK CONS     GCAGGAGTGGCTGTGTGGTAAGTAGCTTGCTTA-CCAA-CCACATGGTTCCGGGTTCAGTCCCACTGCGTGGCA
            * **  ******   ***** * ***** *    ****  *  * **** * ****** * **  * ***  **
tRNA^Trp    GAAGCGGTGGCTCAATGGTA-G-AGCTTTCGACTCCAAATCGAAGGGTTGCAGGTTCAATTCCTGTCCGTTTCACCA

OK CONS     GCAGGA-GTGGCTGTGTGGTA-AGTAGCTTGCTTACCAACCACATGGTTCCGGGTTCAGTCCCACTGCGTGG-CA
            *  *** ** ***  ****** **  ** *****  **   *   **  ********  **** * ** *   **
tRNA^Ala    GG-GGATGTAGCTCAGTGGTAGAG-CGCATGCTTCGCATGTATGAGGCCCCGGGTTCGATCCC-CGGCATCTCCACCA

OK CONS     GCAGGAGTGGCTGTGT-GGTA-AGTAGCTTGCTTACCAACCACATGGTTCCGGGTTC-AGTCCCACTGCGTGGCA
            **  *  * ***  ** **** ** *     **    * ** ** ***   ****** ** *****   * ***
tRNA^Lys    GCCCGGCTAGCTCAGTCGGTAGAGCATGGGACTCTTAATCC-CAGGGTCGTGGGTTCGAGCCCCACGTTG-GGCGCCA
```

Figure 8. Sequence comparisons between the tRNA-related segment of the octopus OK family and three species of tRNA . The tRNA-related segment (OK CONS) in the consensus sequence of the OK family (Fig. 7) is aligned with $tRNA^{Trp}$ (yeast), $tRNA^{Ala}$ (mouse), and $tRNA^{Lys}$ (rabbit). Sequences of the tRNAs are taken from Sprinzl et al. (1987). Identical nucleotides are indicated by asterisks. Only the four standard nucleotides are shown.

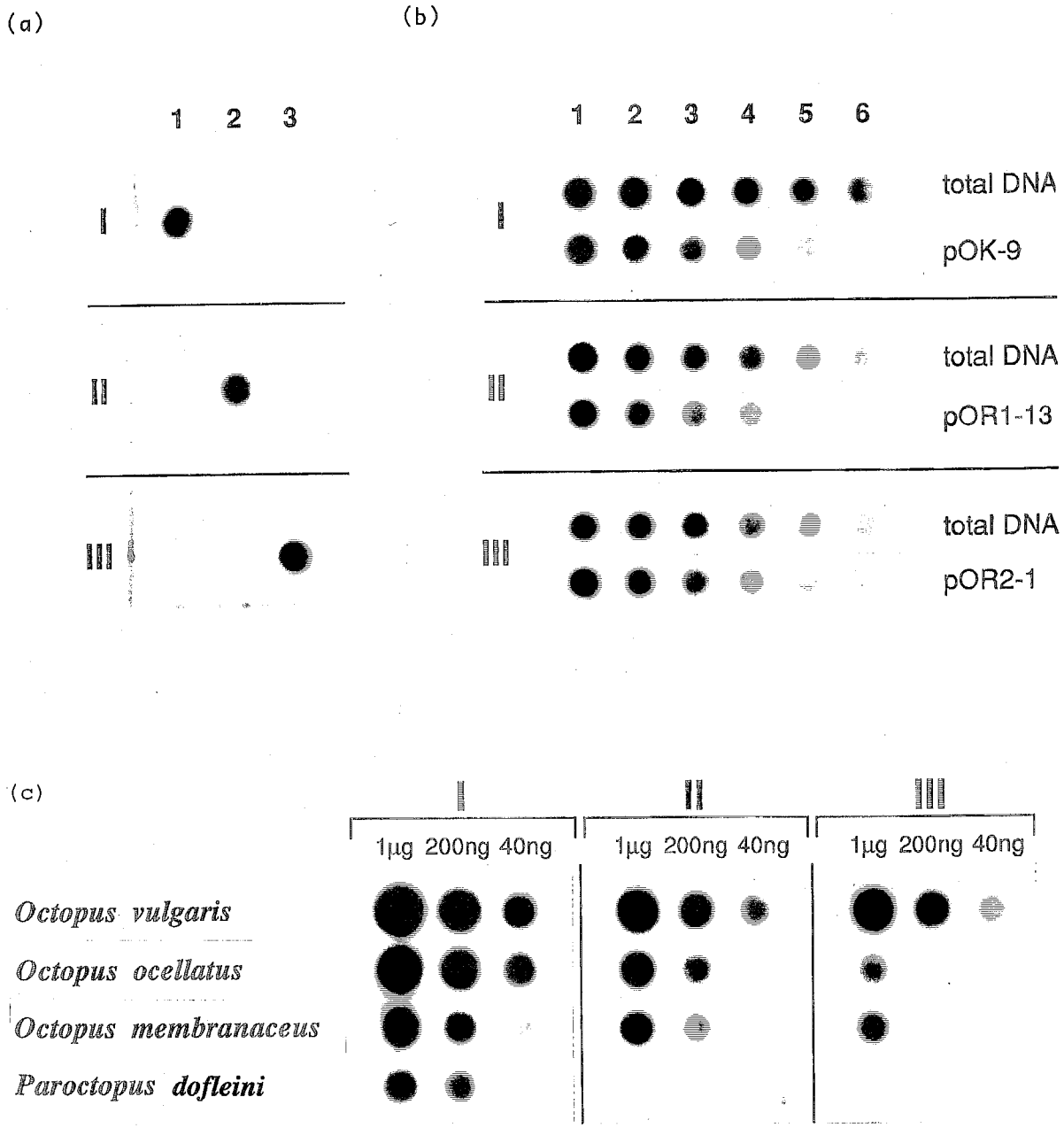


Figure 9. Estimation of relative levels and the copy numbers of the three octopus SINEs. Probes specific to each SINE with the same specific activity were used to estimate the relative copy numbers of the three SINEs. Rows I, II, and III show the results of experiments with probes specific to the OK, OR1, and OR2 families, respectively. (a) A control experiment to examine the specificity of each probe. In columns 1, 2, and 3, cloned DNAs of the OK (pOK-9), OR1 (pOR1-13), and OR2 families (pOR2-1) were dotted, respectively. (b) Estimation of the relative copy numbers of the three SINEs and their relative levels in the genome of *Octopus vulgaris*. In rows I, II, and III, progressively decreasing amounts of total DNA from *O. vulgaris* were dotted in the upper row (from 400 ng to12.5 ng). In lower lines, progressively decreasing amounts of cloned DNA of each SINE were dotted (from 4.0 ng to125 pg). (c) Estimation of the relative copy numbers of each SINE in several species of octopodid. In respective rows, progressively decreasing amounts of total DNA of the four species of octopodid were dotted.

Table 1. ***Some features of the 3 families of SINEs in*** **Octopus vulgaris**

Features / SINEs	Relative ratio	Level per genome (%)	Average sequence divergence (%)
OK family	3.04	3.5	6.8
OR1 family	1.18	1.3	2.9
OR2 family	1.00	0.8	5.4

Table 2. ***Relative copy numbers of each SINE in several species of octopodid***

Species / SINEs	*O. vulgaris*	*O. ocellatus*	*O. membranaceus*	*P. dofleini*
OK family	1.00	0.86	0.70	0.21
OR1 family	0.39	0.18	0.11	——
OR2 family	0.33	0.01	0.02	——

PART II

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

ABSTRACT

Using labeled transcripts generated *in vitro* from squid total genomic DNA as a probe, our group isolated and characterized a short interspersed repetitive element (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.

INTRODUCTION

Recently, it has been shown that SINEs are widespread in the animal kingdom (see Part I) and in plants (Mochizuki et al., 1992; Yoshioka et al., 1993). 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 AT-rich region (Okada, 1991a). 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 (Okada, 1991a; Endoh et al., 1990). This view contrasts with that of Deininger and Daniels (1986), 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 part, a possible model for the initial generation of SINEs are presented (Ohshima et al., 1993).

MATERIALS AND METHODS

(a) Materials

DNA from a squid (*Loligo bleekeri*) was extracted as described elsewhere (Blin & Stafford, 1976). All sequencing reagents were obtained from Takara Shuzo Co. (Kyoto, Japan). Restriction enzymes were obtained from Takara Shuzo Co. and TOYOBO Co. (Tokyo, Japan). Various labeled nucleotides were purchased from Amersham Co. (Tokyo, Japan).

(b) RNA synthesis *in vitro*

RNA was synthesized *in vitro* essentially as described elsewhere (Manley et al., 1980). Total genomic DNA from squid was transcribed in the absence of α-amanitin and in the presence of α-amanitin at a concentration of 2 μg/ml or 200 μg/ml. The RNA was isolated by extraction with phenol and was subjected to electrophoresis on a 8% (w/v) polyacrylamide denaturing gel (Maxam & Gilbert, 1980).

(c) Isolation of phage and plasmid DNAs and DNA sequencing

A genomic library was constructed by ligation of the genomic DNA of *L. bleekeri* that has been completely digested with *Eco* RI and λgt10 arms. This library was constracted by Ryuta Koishi. ^{32}P-labeled RNA, transcribed from total DNA, was used as the probe to isolate phage clones from the library. Respective inserts of positive clones were subcloned in pUC vectors and the region of the squid SINE was located by Southern blot hybridization (Southern, 1975). The sequence of each was determined by the dideoxychain-termination method (Sanger et al., 1977). Sequences of four clones, NO4, NO22, NO28 and NO31, were determined by me, and the others, NO6 and NO25, were determined by Koishi.

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, I and R. Koishi isolated six phage clones, localized the DNA loci that hybridized to the probe, and determined their sequences. Figure 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 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 (Raba et al., 1979) (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) 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, nine families of SINEs, 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) (Kido et al., 1991). 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} (Kido et al., 1991). 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 (Endoh et al., 1990). The origin of the *Galago* type 2 family (Daniels & Deininger, 1983) was originally proposed to be tRNA^{Met} (Daniels & Deininger, 1985), 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} (Okada, 1990). As for the parental tRNA species of the rodent type 2 *Alu* family (Krayev et al.,1982), two groups (Lawrence et al., 1985; Sakamoto & Okada, 1985) have shown that this family most closely resembles tRNA^{Lys}, whereas another group (Daniels & Deininger, 1985) has claimed that tRNA^{Ser} is the most likely parental species. Three families, namely, salmonid *Hpa* I (Kido et al., 1991), octopus OK (Ohshima & Okada, 1994; see also Part I) and tobacco TS (Yoshioka et al., 1993), are also included in this group,

but the extent of their similarities to $tRNA^{Lys}$ is low. I and Norihiro Okada have tentatively listed these nine 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).

(b) Several SINEs in distant species exhibit similarities to one another

I and N. Okada 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, namely, 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 Figure 4a. 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 of 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.

(c) 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 (Power et al., 1986), SRV-2 (Thayer et al., 1987), MPMV (Sonigo et al.,1986), MMTV (Fasel et al., 1983), HIV-1 (Spire et al., 1989), SIVcpz (Huet et al., 1990), EIAV (Derse et al., 1987), BIV127 (Garvey et al., 1990), SA-OMVV (Querat et al., 1990)], 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. 4b). 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 to 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 (Saigo, 1986), and we discussed such a mechanism several times subsequently (Okada, 1991a; Endoh et al., 1990). The model presented here has been briefly described elsewhere (Okada & Ohshima, 1993).

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 (Gojobori & Yokoyama, 1985). 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.

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 (Daniels & Deininger, 1985), see above].

With regard to the origin of SINEs, $tRNA^{Lys}$ appears to be the most frequent progenitor of SINEs (Okada, 1991a,b). 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 (Okada, 1990). 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.

(d) 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 (Temin, 1989), 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 (Doolittle et al., 1989; Xiong & Eickbush, 1990). In the case of the plant *copia*-like retrotransposons, more concrete evidence for horizontal transmission has been presented recently (Voytas et al., 1992). Moreover, it is well established that many endogenous retroviruses have spread to new species by horizontal transfer, and there is also every reason to believe that endogenous viruses are the descendants of exogenous viruses that infected germ-line cells (Doolittle et al., 1989). 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. 4a) 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. (1989) 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 was 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.

REFERENCES

Blin, N., and Stafford, D. W. (1976). A general method for isolation of high-molecular-weight DNA from eukaryotes. Nucl. Acids Res. 3, 2303-2308.

Daniels, G.R., and Deininger, P.L. (1983). A second major class of Alu family repeated DNA sequences in a primate genome. Nucleic Acids Res. 11, 7595-7610.

Daniels, G. R., and Deininger, P. L. (1985). Repeat sequence families derived from mammalian tRNA genes. Nature 317, 819-822.

Deininger, P.L., and Daniels, G.L. (1986). Trends Genet. 2, 76-80.

Derse, D., Dorn, P.L., Levy, L., Stephens, R.M., Rice, N.R., and Casey, J.W. (1987). Characterization of equine infectious anemia virus long terminal repeat. J. Virol. 61, 743-747.

Doolittle, R.F., Feng, D.-F., Johnson, M.S., and McClure, M.A. (1989). Origins and evolutionary relationships of retroviruses. Quart. Rev. Biol. 64, 1-30.

Endoh, H., Nagahashi, S., and Okada, N. (1990). A highly repetitive and transcribable sequence in the tortoise genome is probably a retroposon. Eur. J. Biochem. 189, 25-31.

Fasel, N., Buetti, E., Firzlaff, J., Pearson, K., and Diggelmann, H. (1983). Nucleotide sequence of the 5' noncoding region and part of the *gag* gene of mouse mammary tumor virus; identification of the 5' splicing site for subgenomic mRNAs. Nucleic Acids Res. 11, 6943-6955.

Galli, G., Hofstetter, H., and Birnstiel, M. L. (1981). Two conserved sequence blocks within eukaryotic tRNA genes are major promoter elements. Nature 294, 626-631.

Garvey, K.J., Oberste, M.S., Elser, J.E., Braun, M.J., and Gonda, M.A. (1990). Nucleotide sequence and genome organization of biologically active proviruses of the bovine immunodeficiency-like virus. Virology 175, 391-409.

Gojobori, T., and Yokoyama, S. (1985). Rates of evolution of the retroviral oncogene of moloney murine sarcoma virus and of its cellular homologues. Proc. Natl. Acad. Sci. USA 82, 4198-4201.

Huet, T., Cheynier, R., Meyerhans, A., Roelants, G., and Wain-Hobson, S. (1990). Genetic organization of a chimpanzee lentivirus related to HIV-1. Nature 345, 356-359.

Kido, Y., Aono. M., Yamaki, T., Matsumoto, K., Murata, S., Saneyoshi, M., and Okada, N. (1991). Shaping and reshaping of salmonid genomes by amplification of tRNA-derived retroposons during evolution. Proc. Natl. Acad. Sci. USA 88, 2326-2330.

Lawrence, C. B., McDonnell, D. P., and Ramsey, W. J. (1985). Analysis of repetitive sequence elements containing tRNA-like sequences. Nucl. Acids Res. 13, 4239-4252.

Krayev, A.S., Markusheva,T.V., Kramerov, D.A., Ryskov, A.V., Skryabin, K.G., Bayev, A.A., and Georgiev, G.P. (1982). Ubiquitous transposon-like repeats B1 and B2 of the mouse genome: B2 sequencing. Nucleic Acids Res. 10, 7461-7475.

Manley, J. L., Fire, A., Cano, A., Sharp, P. A., and Gefter, M. L. (1980). DNA-dependent transcription of adenovirus genes in a soluble whole-cell extract. Proc. Natl. Acad. Sci. USA 77, 3855-3859.

Matsuo, M., Abe, Y., Saruta, Y., and Okada, N. (1995). Mollusk genes encoding lysine tRNA (UUU) contain introns. Gene *in press.*

Maxam, A.M., and Gilbert, W. (1980). Sequencing end-labeled DNA with base-specific chemical cleavage. Methods Enzymol. 65, 499-560.

Mochizuki, K., Umeda, M., Ohtsubo, H., and Ohtsubo, E. (1992). Characterization of a plant SINE, p-SINE*1*, in rice genomes. Jpn. J. Genet. 67, 155-166.

Ohshima, K., Koishi, R., Matsuo, M., and Okada, N. (1993). 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. Proc. Natl. Acad. Sci. USA 90, 6260-6264.

Ohshima, K. and Okada, N. (1994). Generality of the tRNA origin of short interspersed repetitive elements (SINEs): Characterization of three different tRNA-derived retroposons in the octopus. J. Mol. Biol. 243, 25-37.

Okada, N. (1990). Transfer RNA-like structure of the human Alu family: implications of its generation mechanism and possible functions. J. Mol. Evol. 31, 500-510.

Okada, N. (1991a). SINEs: Short interspersed repeated elements of the eukaryotic genome. Trends in Ecol. & Evol. 6, 358-361.

Okada, N. (1991b). SINEs. Curr. Opin. in Genet. & Dev. 1, 498-504.

Okada, N., and Ohshima, K. (1993). A model for the mechanism of initial generation of short interspersed elements (SINEs). J. Mol. Evol. 37, 167-170.

Okada, N. and Ohshima, K. (1995). Evolution of tRNA-derived SINEs. in *The Impact of Short Interspersed Elements (SINEs) on the Host Genome*, ed. Maraia, R. J. (R. G. Landes Company, Austin), pp. 61-79.

Power, M.D., Marx, P.A., Bryant, M.L., Gardner, M.B., Barr, P.J., and Luciw, P.A. (1986). Nuclotide sequence of SRV-1, a type D simian acquired immune deficiency syndrome retrovirus. Science 28, 1567-1573.

Querat, G., Andoly, G., Sonigo, P., and Vigne, R. (1990). Nucleotide sequence analysis of SA-OMVV, a visna-related ovine lentivirus: phylogenetic history of lentiviruses. Virology 175, 434-447.

Raba, M., Limburg, K., Burghagen, M., Katze, J.R., Simsek, M., Heckman, J.E., RajBhandary, U.L., and Gross, H.J. (1979). Nucleotide sequence of three isoaccepting lysine tRNAs from rabbit liver and SV40-transformed mouse fibroblasts. Eur. J. Biochem. 97, 305-318.

Saigo, K. (1986). Nucleic Acids Res. 14, 7815.

Sakamoto, K., and Okada, N. (1985). Rodent type 2 Alu family, rat identifier sequence, rabbit C family, and bovine or goat 73-bp repeat may have evolved from tRNA genes. J. Mol. Evol. 22, 134-140.

Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74, 5463-5467.

Sonigo, P., Barker, C., Hunter, E., and Wain-Hobson, S. (1986). Nucleotide sequence of Mason-Pfizer monkey virus: an immunosuppressive D-type retrovirus. Cell 45, 375-385.

Southern, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98, 503-517.

Spire, B., Sire, J., Zachar, V., Rey, F., Barré-Sinoussi, F., Galibert, F., Hampe, A., and Chermann, J-C. (1989). Nucleotide sequence of HIV1-NDK: a highly cytopathic strain of the human immunodeficiency virus. Gene 81, 275-284.

Temin, H.M. (1989). Retrons in bacteria. Nature 339, 254-255.

Thayer, R.M., Power, M.D., Bryant, M.L., Gardner, M.B., Barr, P.J., and Luciw, P.A. (1987). Sequence relationships of type D retroviruses which cause simian acquired immunodeficiency syndrome. Virology 157, 317-329.

Xiong, Y., and Eickbush, T.H. (1990). Origin and evolution of retroelements based upon their reverse transcriptases. EMBO J. 9, 3353-3362.

Voytas, D.F., Cummings, M.P., Konieczny, A., Ausubel, F.M., and Rodermel, S.R. (1992). *copia*-like retrotransposons are ubiquitous among plants. Proc. Natl. Acad. Sci. USA 89, 7124-7128.

Yoshioka, Y., Matsumoto, S., Kojima, S., Ohshima, K., Okada, N., and Machida, Y. (1993). Molecular characterization of a short interspersed repetitive element from tobacco that exhibits sequence homology to specific tRNAs. Proc. Natl. Acad. Sci. USA 90, 6562-6566.

TABLES AND FIGURES

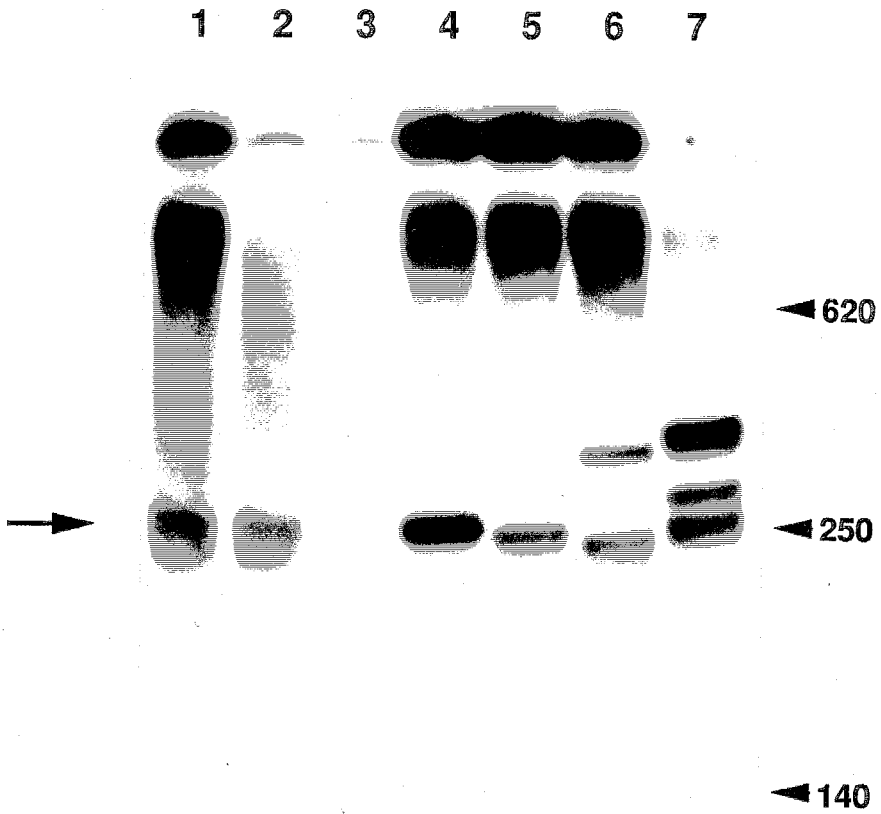


Fig. 1. *In vitro* transcription of total genomic DNA and cloned DNAs in a HeLa cell extract. Total genomic DNA from squid (*Loligo bleekeri*) was transcribed in the absence of α-amanitin (lane 1) and in the presence of α-amanitin at a concentration of 2 μg/ml (lane 2) or 200 μg/ml (lane 3). The templates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

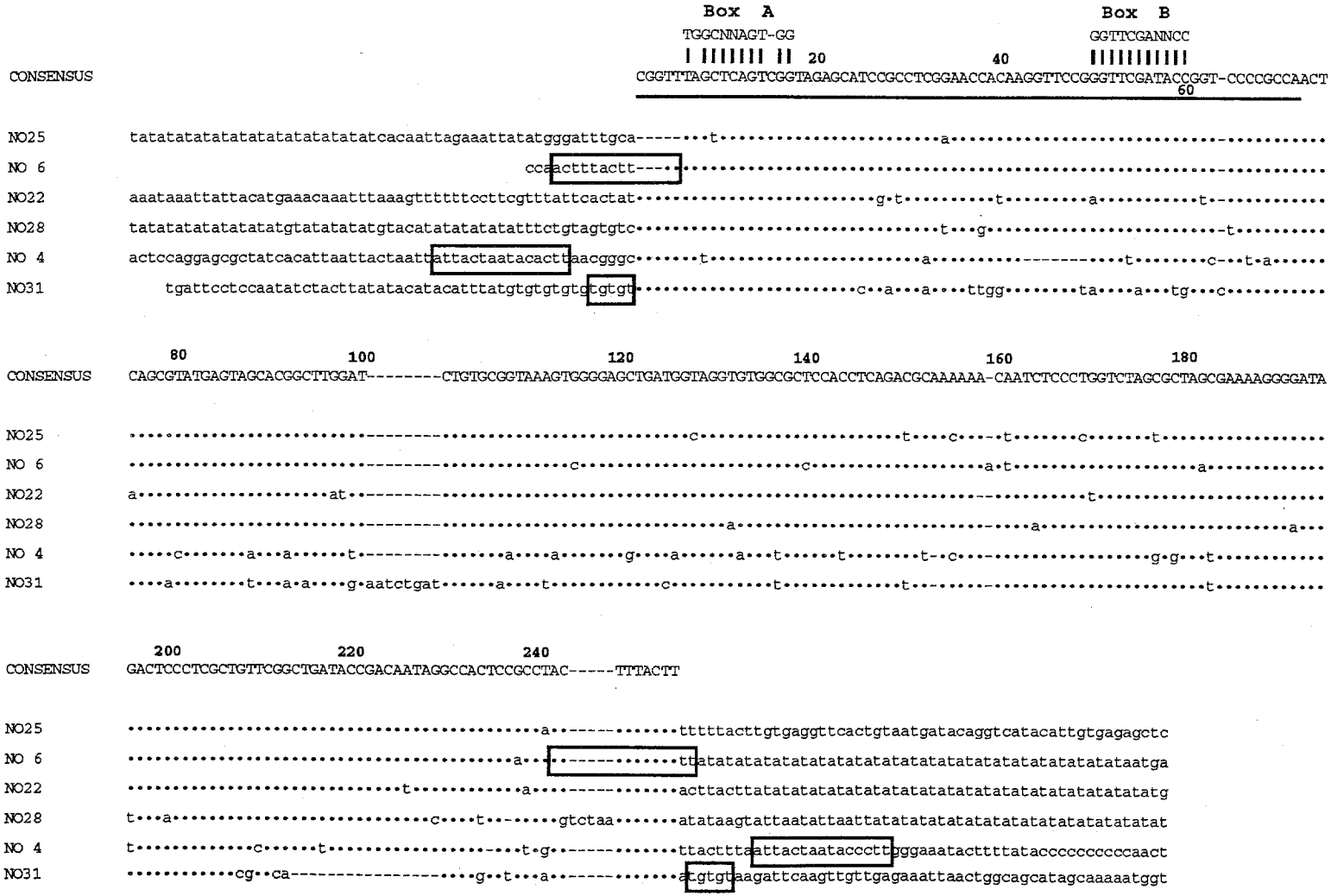


Fig. 2. Sequences of members of the squid SK family and a consensus sequence. The tRNA^{Lys}-related region is underlined. The two internal promoters are indicated above the consensus sequence. Direct repeats are boxed.

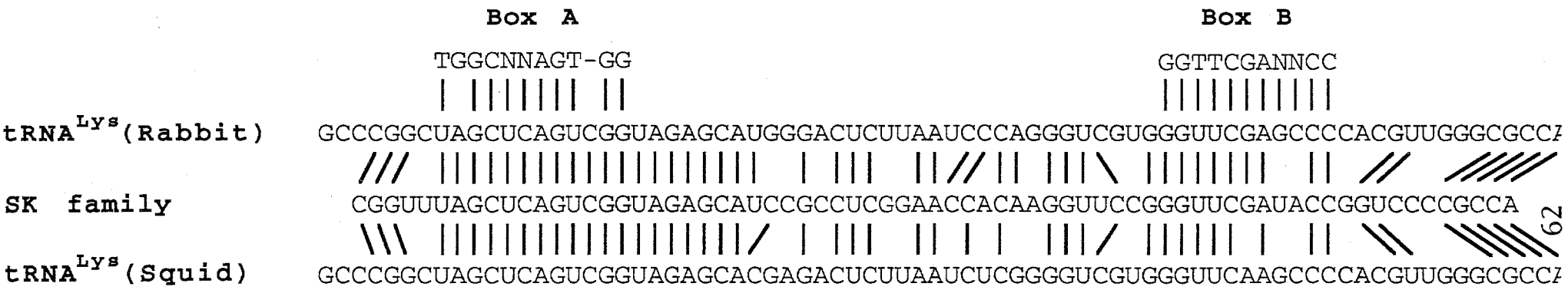
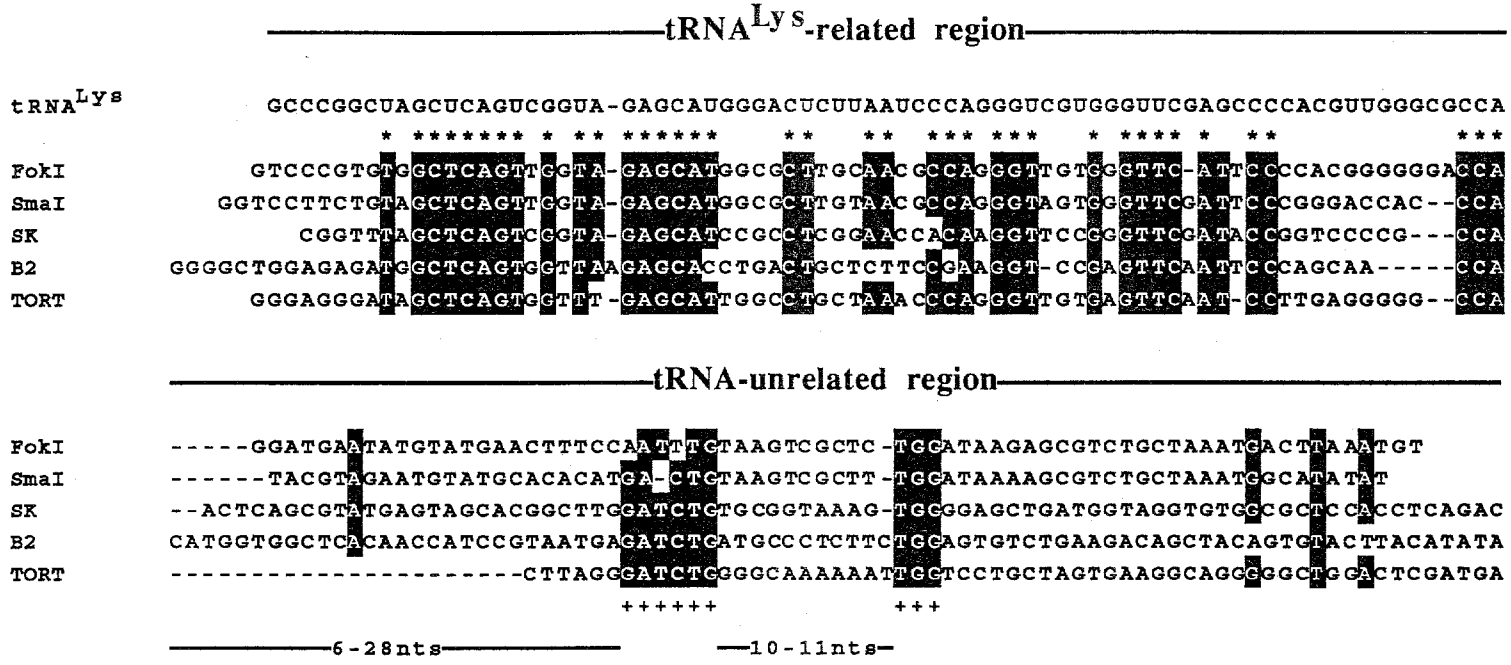


Fig. 3. Sequence comparisons between the tRNA-related segment in the consensus sequence of the squid SK family and the $tRNA^{Lys}$ sequences from rabbit and squid. The sequence of the rabbit $tRNA^{Lys}$ is taken from Raba et al. (1979) and the gene sequence for the squid $tRNA^{Lys}$ is from Matsuo et al. (1995).

a)


tRNA-related region
tRNA^Lys
GCCCGGCUAGCUCAGUCGGUA-GAGCAUGGGACUCUUAAUCCCAGGGUCGUGGGUUCGAGCCCCACGUUGGGCGCCA
FokI
SmaI
SK
B2
TORT
tRNA-unrelated region
6-28nts
10-11nts

b)

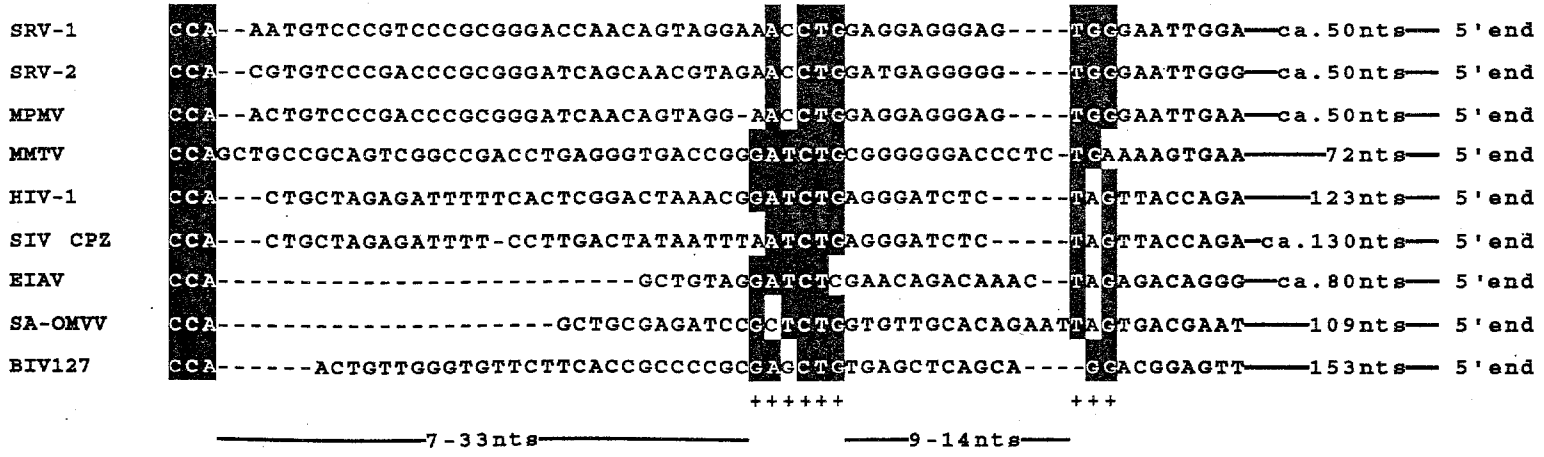

SRV-1
SRV-2
MPMV
MMTV
HIV-1
SIV CPZ
EIAV
SA-OMVV
BIV127
ca.50nts
72nts
123nts
ca.130nts
ca.80nts
109nts
153nts
5'end
7-33nts
9-14nts

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 more than 4 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 an asterisk. Two conserved regions of GATCTG and TGG are indicated by crosses. 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, 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 (PBS) and the two conserved regions are highlighted in black. The two conserved regions are indicated by crosses. 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.

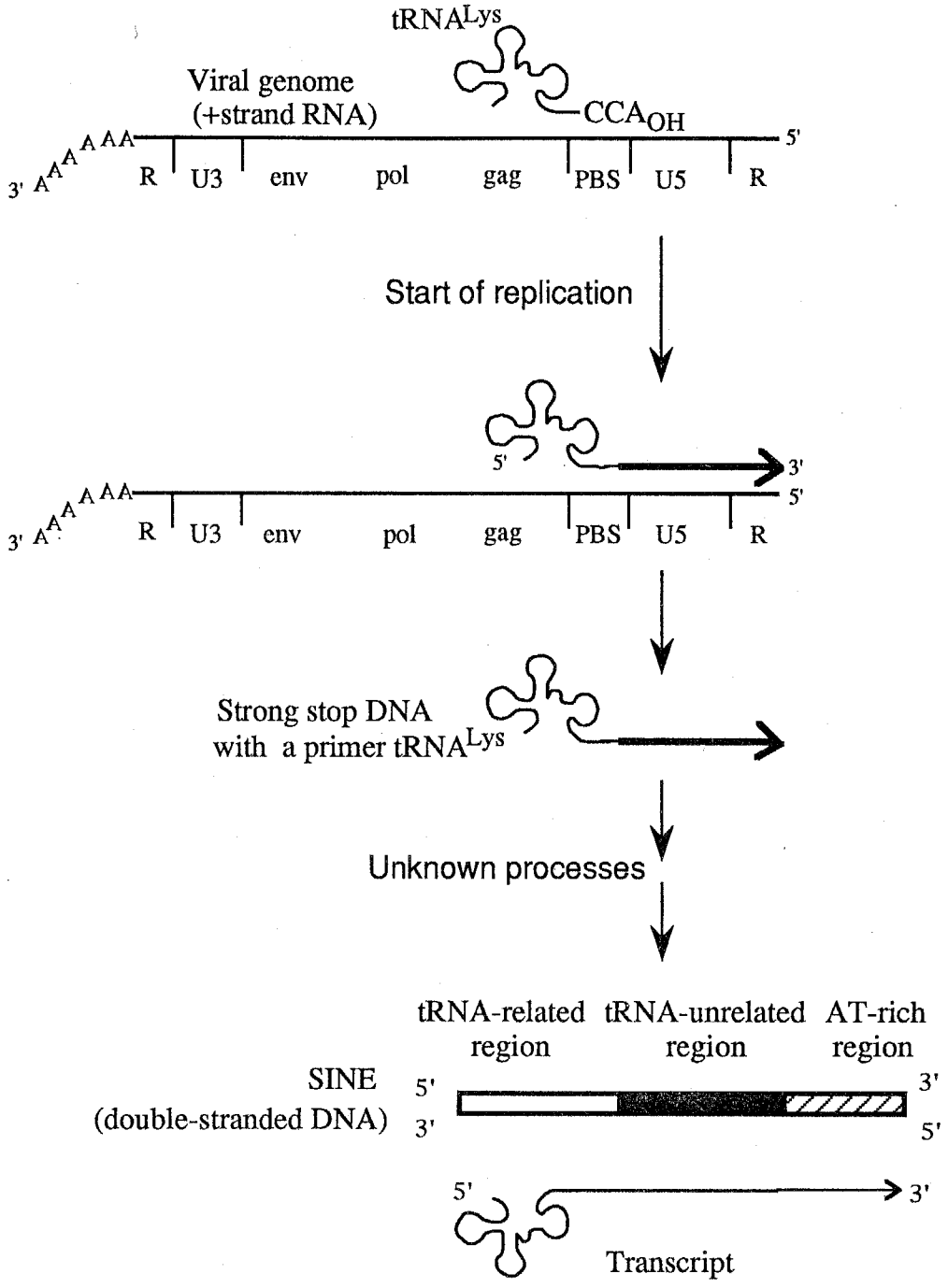


Fig. 5. A model for the mechanism of the initial generation of SINEs (see text for details)

Table 1. *Classification of three superfamilies of SINEs*

SINE	Species	Reference(#)
1. Superfamily of $tRNA^{Lys}$-related SINEs		
Rodent type 2 (B2) family	Mouse, rat, hamster	Krayev et al., 1982
Galago type 2 family	*Galago crassicaudatus*	Daniels & Deininger, 1983
Salmon *Sma* I family	Chum and pink salmon	Matsumoto et al., 1986
Tortoise Pol III/SINE	*Chinemys* (*Geoclemys*) *reevesii*	Endoh et al., 1990
Charr *Fok* I family	*Salvelinus* spp.	Kido et al., 1991
Salmonid *Hpa* I family	All salmonid spp.	Kido et al., 1991
Squid SK family	*Loligo bleekeri*	Ohshima et al., 1993
Tobacco TS family	Solanace spp.	Yoshioka et al., 1993
Can SINEs	Canoid spp.	Coltman & Wright, 1994
Salmonid *Ava* III family	All salmonid spp.	Kido et al., 1994
Octopus OK family	Octopodid spp.	Ohshima et al., 1994
2. Superfamily of $tRNA^{Arg}$-related SINEs		
Pig PRE-1 family	Pig	Singer et al., 1987
Human blood fluke SMα family	*Schistosoma mansoni*	Spotila et al., 1989
Octopus OR1 family	*Octopus* spp.	Ohshima et al., 1994
Octopus OR2 family	*Octopus* spp.	Ohshima et al., 1994
3. Superfamily of $tRNA^{Gly}$-related SINEs		
Bovine & goat 73-bp repeat	Bovid spp.	Schon et al., 1981
Rabbit C family	House rabbit	Cheng et al., 1984
Rice p-SINE*1*	*Oryza* spp.	Mochizuki et al., 1992
4. Others		
Rodent ID element	Mouse, rat, hamster, guinea pig	Milner et al., 1984
Silk worm Bm1 element	*Bombyx mori*	Adams et al., 1986
Sea urchin SURF1	*Strongylocentrotus purpuratus*	Nisson et al., 1988
Equine ERE-1 family	All *Equus* spp.	Sakagami et al., 1994
Oilseed rape $S1_{Bn}$ family	*Brassica napus*	Deragon et al., 1994

(#) References are taken from Okada & Ohshima (1995).

PART III

Short interspersed repetitive elements (SINEs) may have been generated by recombination with long interspersed repetitive elements (LINEs)

ABSTRACT

Short interspersed elements (SINEs) are one type of retroposon, a class of informational molecules that are amplified via cDNA intermediates and flow back into the host genome. By contrast to retroviruses and retrotransposons, SINEs do not encode enzymes for the amplification machinery, such as reverse transcriptases, so they are presumed to borrow these enzymes from other sources. In the present study, we isolated a retrotransposon that was not of the long-terminal repeat type (non-LTR type) from the turtle genome. Its sequence was found to be very similar to that of the avian CR1 retrotransposon. To our surprise, the sequence at the 3' end of this retrotransposon was nearly identical to that of the Pol III/SINEs of the tortoise. Since CR1-like retrotransposons are widespread in birds and in many other reptiles, including the turtle, and since the tortoise Pol III/SINEs are only found in vertical-necked turtles, it seems possible that the sequence at the 3' end of the tortoise Pol III/SINE might have been generated by recombination with the CR1-like retrotransposon in a common ancestor of vertical-necked turtles, after the divergence of side-necked turtles. We extended our observations to show that the 3'-end sequences of families of several tRNA-derived SINEs, such as the salmon *Sma* I family, the salmonid *Hpa* I family and the tobacco Ts family, might have originated from respective non-LTR-type retrotransposons. If we accept that the signals for the amplification and integration of non-LTR-type retrotransposons are located within their 3'-end sequences, these results provide a general scheme for the mechanism by which SINEs acquire retropositional activity. We propose here that tRNA-derived SINEs might have been generated by a recombination event in which a strong-stop DNA with a primer tRNA, which is an intermediate in the replication of certain retroviruses and LTR retrotransposons, was directly integrated at the 3' end of a non-LTR-type retrotransposon.

INTRODUCTION

SINEs are short (approximately 80-400 bp) repetitive elements, which are often present at more 10^5 copies per genome. Almost all SINEs reported to date, from sources as diverse as mammals and plants, are derived from tRNAs (Okada, 1991a,b; Okada and Ohshima, 1995), with the exception of the primate *Alu* and the rodent B1 families that are derived from 7SL RNA (Weiner, 1980; Ullu and Tschudi, 1984). The tRNA-derived SINEs are not simple pseudogenes for tRNAs but have a composite structure, with a region homologous to a tRNA, a tRNA-unrelated region, and an AT-rich region (Okada, 1991a,b; Okada and Ohshima, 1995). Our group proposed recently that a possible origin of the region that is homologous to the tRNA of a SINE might be a primer tRNA, attached to a 'strong-stop DNA', which is an intermediate during reverse transcription of certain retroviruses and retrotransposons (Ohshima et al., 1993; see also Part II). The source of enzymes for the reverse transcription of SINEs remains, however, to be determined.

Long interspersed repetitive elements (LINEs) (Fanning and Singer, 1987) or non-LTR retrotransposons (Eickbush, 1994) encode their own reverse transcriptases (RTases). It appears that almost all copies of such an element have identical 3' ends and their 5' regions are subject to truncation (Fanning and Singer, 1987; Eickbush, 1994). Therefore, it is believed that an RTase must recognize the 3' end of the RNA template to initiate first-strand synthesis (Eickbush, 1992). This model was verified by Luan et al. in an elegant experiment using the RTase from the R2Bm non-LTR-type retrotransposon of *Bombyx mori* (Luan et al., 1993).

The members of the CR1 (chicken repeat 1) family, found in avian genomes, are non-LTR-type retrotransposons. This family was first described as a SINE family in the chicken genome because most members of this family were severely truncated at their 5' ends and no obvious open reading frame (ORF) was observed (Stumph et al.,

1981). Subsequently, long members of the CR1 family were isolated and it was shown that they are non-LTR-type retrotransposons (Burch et al., 1993). Vandergon and Reitman (1994) detected sequences similar to the avian CR1 element in nonavians, such as lizard and ray. In addition, they pointed out the similarity of the CR1 element to the tortoise Pol III/SINE.

In this part, evidence for the sharing of the 3'-end sequence between the tortoise Pol III/SINE and a CR1-like retrotransposon from turtles is provided. Furthermore, it is demonstrated that tRNA-derived SINEs may generally have 3' ends in common with particular non-LTR-type retrotransposons. From these results, a model is proposed which explains the molecular mechanism by which tRNA-derived SINEs acquired retropositional activity during evolution (Ohshima et al., 1995).

MATERIALS AND METHODS

(a) Materials

DNA from turtles (*Platemys spixii, Eretmochelys imbricata, Apalone ferox* and *Chinemys reevessi*), caiman (*Caiman crocodilus*), chicken and human was extracted as described elsewhere (Blin & Stafford, 1976). All sequencing reagents were obtained from Takara Shuzo Co. (Kyoto, Japan). Restriction enzymes were obtained from Takara Shuzo Co. and TOYOBO Co. (Tokyo, Japan). Various labeled nucleotides were purchased from Amersham Co. (Tokyo, Japan).

(b) Dot hybridization experiments

Dot hybridization experiments were performed to examine the distribution of the tortoise Pol III/SINE in living turtles and other species. Progressively decreasing amounts of total DNA (1 μ g, 200 ng and 40 ng) from individuals of several species were dotted on a membrane (columns 1 to 7 in Fig. 1). In column 8, progressively decreasing amounts of cloned DNA (10 ng, 2 ng and 400 pg) of the tortoise Pol III/SINE (TE6; see Endoh et al., 1990) were dotted on the membrane as a control. Probe 1 (from position 1 to 154 in Fig. 3) was labeled internally with [α-^{32}P] dCTP, and the 5' end of the probe 2 (PROTO-R1; 5'-CCCCAG(A/T)T(G/C)C(T/C)TA(A/C)(G/A)TGGCCCCCTCAAGGA-3') was labeled with [γ-^{32}P] ATP.

(c) Isolation of phage and plasmid DNAs and DNA sequencing

Genomic libraries were constructed by ligation of λgt10 arms and the genomic DNA of *Platemys spixii* or *Apalone ferox* that has been completed digested with *Eco* RI. Probe 1, which was labeled internally with [α-^{32}P] dCTP, was used as the probe to isolate phage clones from the libraries. Respective inserts of positive clones were subcloned in pUC vectors and then sequenced by the dideoxychain-termination

method (Sanger et al., 1977). Sequences of clones from *Apalone ferox* (soft-shelled turtle) were determined by Takeshi Sasayama.

(d) Polymerase chain raction (PCR) experiments

PCR analysis was performed using two sets of primers specific to the CR1 element, CRF1 plus CRR0 and CRF2 plus CRR1, and two sets of primers specific to the SINE, TEF1 plus TER0 and TEF1 plus PROTO-R1, respectively. Ten ng of total DNA from each species (lanes 2 to 15 in Fig. 4), and 100 pg (Fig. 4A) or 1 ng (Fig. 4B) of cloned DNA (lanes 17 to 20) were used as templates. PCR was performed at 55 °C for 30 cycles. Nucleotide sequences of primers used in the experiments are as follows.

CRF1, 5'-GAG(G/A)T(T/C)TAGGTTGGA(T/C)ATTAGG-3';

CRR0, 5'-ATTAGGGTTGGAAGGGACCT-3';

CRF2, 5'-GAGGGTGGTGA(G/A)GCACTGGAA-3';

CRR1, 5'-TCATCTAGTCCAACCCCCTGCT-3';

TEF1, 5'-GGGAGGGATAGCTCAGTGGT-3'; and

TER0, 5'-ATAT(G/A)CC(A/T)ATCTCCTAGAAC-3'.

RESULTS

(a) A CR1-like retrotransposon is present in the genome of a primitive turtle

Living turtles can be divided into two major suborders, namely, Cryptodira, which includes vertical-necked turtles, and Pleurodira, which includes side-necked turtles (Table 1). The suborder Cryptodira consists of several superfamilies, such as Testudinoidea (tortoises), Trionychoidea (soft-shelled turtles) and Cheloniodea (sea turtles) (Gaffney and Meylan, 1988). The fossil record suggests that the suborder Pleurodira diverged from the suborder Cryptodira about 200 million years ago (Rougier et al., 1995). Short interspersed elements, designated the tortoise Pol III/SINE, were isolated from a tortoise in the superfamily Testudinoidea, and they were characterized as members of a typical tRNA-derived SINE family, each member of which consists of three regions, namely a $tRNA^{Lys}$-related region, a tRNA-unrelated region and an AT-rich region (Fig. 1A) (Endoh et al., 1990). Using almost the entire sequence of this SINE as a probe (probe 1 in Fig. 1A), I performed a dot hybridization experiment to examine the distribution of the tortoise Pol III/SINE in living turtles and other species (Fig. 1B). Signals were detected in all turtles, including a side-necked turtle, and in a caiman. In addition, a faint signal was observed in chicken. When I used just part of the SINE as the probe (Fig. 1B), signals were detected only in turtles other than the side-necked turtle (Fig. 1C).

My observations prompted me to try to explain the difference in hybridization signals. I constructed a genomic library for a side-necked turtle (*Platemys spixii*). Using probe 1, I isolated several clones from this library and determined their sequences (Fig. 2). To my surprise, these clones showed remarkable similarity to the chicken CR1 element. One clone, designated 4-2(Ps), had extensive similarity to chicken CR1 in the region of the 2.1 kbp *Eco* RI fragment (64% similarity over the entire 2.1 kbp). Two other clones, designated 2-3(Ps) and 4-1(Ps), had a common sequence at their 3' ends, and their 5' regions were truncated, as is generally the case

for several retrotransposons, including the chicken CR1 element (Fanning and Singer, 1987; Burch et al., 1993).

(b) The 3' ends of the CR1-like retrotransposon and the tortoise Pol III/SINE are similar

The above experiment also revealed that 80 bp at the 3' end of the CR1-like retrotransposon of the side-necked turtle was almost the same as the 3' end of the tortoise Pol III/SINE isolated from *Chinemys reevessi* (previously known as *Geoclemys reevessi*) (Endoh et al., 1990). To confirm the presence of the tortoise Pol III/SINE in other turtles, I and T. Sasayama isolated several clones from the genome of a soft-shelled turtle (*Apalone ferox*) in the superfamily Trionychoidea using probe 1. Figure 3 shows an alignment of a consensus sequence of the CR1-like element, designated CR1(Ps), and consensus sequences of the SINE(Cr) of tortoise and the SINE(Af) of soft-shelled turtle. The sequence of the SINE(Af) family was almost the same as that of the tortoise SINE(Cr) except for the insertion of 17 bp at position 97. The tRNA-unrelated region of the SINE from position 97 to 176 was very similar to about 80 bp at the 3' end of the CR1-like element (from -78 to -1), whereas the $tRNA^{Lys}$-related region (underlined) and the adjacent 20 bp of the tRNA-unrelated region in the SINE exhibited no obvious similarity. In particular, one member of SINE(Cr), designated TE3 (Endoh et al., 1990), exhibited considerable similarity (90%) to the CR1-like element. TE3 was shown to have diverged from other members of the SINE(Cr) family in the tortoise and, therefore, was considered to belong to an ancient group (Endoh et al., 1990).

(c) CR1-like retrotransposons are present in all living reptiles and avians, whereas the tortoise Pol III/SINE is confined to cryptodires, a dominant group of living turtles

Dot hybridization experiments, as shown in Figure 1, suggested that CR1-like elements are distributed throughout a wide range of species and no just in the species that have the tortoise Pol III/SINE. Several questions, however, remained unanswered, as follows. (i) Are the CR1-like elements present in all cryptodires? (ii) Do crocodilians have CR1-like elements? (iii) Is the tortoise Pol III/SINE completely missing from side-necked turtles? To answer these questions, I performed analyses by the polymerase chain reaction (PCR) using several sets of primers specific to the SINE and to CR1, respectively (Fig. 4). As shown in Figure 4A, DNA products of both 213 and 147 bp, specific to the CR1-like element, were detected in chicken, caiman (crocodilian), side-necked turtle, and in all cryptodires tested. This result suggested to me that CR1-like elements might be ubiquitous in all living reptiles and avians. I failed to detect any signals with human DNA, although mammals are believed to be the closest living relatives of reptiles and avians (Hedges, 1994). This result indicates that the CR1-like elements might have arisen in a common ancestor of reptiles and avians after divergence of the mammalian lineage (Fig. 5).

PCR products of 154 or 89 bp specific to the tortoise Pol III/SINE were detected only with DNA templates from cryptodires, namely, a sea turtle, a soft-shelled turtle and a tortoise, and they were not detected in the case of the side-necked turtle (Fig. 4B). These results suggest that the SINE emerged in a common ancestor of cryptodires after the divergence of pleurodires (Fig. 5). The tRNA-related region of the tortoise Pol III/SINE might not have appeared by itself in the genome until it had combined with the 3' end of the CR1-like element (see Discussion).

(d) Other SINEs, such as the salmon *Sma* I, the charr *Fok* I, the salmonid *Hpa* I and the tobacco TS families, also have a sequence at the 3' end that is common to the respective non-LTR-type retrotransposon

To examine whether members of a SINE family generally have a fused structure with a certain non-LTR-type retrotransposon at the 3' end, I searched for examples other than the tortoise Pol III/SINE in the EMBL database (release 40.0). The salmonid *Hpa* I family is a SINE family that is widespread in many salmonid species (Kido et al., 1991). The 3'-end sequence of the salmonid *Hpa* I family was found to show significant homology to the last approximately 60 bp of members of the RSg-1 family, namely, non-LTR-type retrotransposons isolated from the rainbow trout (Winkfein et al., 1988) (Fig. 6). Sequences of two TS SINE family-like sequences of tobacco (Yoshioka et al., 1993) have been deposited in the DNA database (accession nos.: X55753 and M32603) and they show significant similarity to each other. Moreover, their 3' regions are identical to the 3' part of the TS SINE (Fig. 6). In this case, however, although this sequence has not been reported to be that of a non-LTR-type retrotransposon, I suggest the presence of a family of TS-like non-LTR-type retrotransposons in the tobacco genome. Furthermore, employing the method known as genomic DNA walking, using a cassette and a primer complementary to the sequence in the unit, I and Mitsuhiro Hamada isolated a non-LTR retrotransposon-like sequence from an eel. This sequence exhibited remarkable similarity to the 3' ends of members of the salmon *Sma* I and the charr *Fok* I families. In Figure 6 these pairs of sequences are aligned to allow us to compare junction sequences between SINEs and respective non-LTR-type retrotransposons.

DISCUSSION

(a) Enzymes required for the retroposition of many tRNA-derived SINEs might be provided by respective non-LTR-type retrotransposons

Eickbush's group recently proposed a mechanism for retroposition of R2, a non-LTR-type retrotransposon in *Bombyx mori*. In this mechanism, the R2 protein, which has both sequence-specific endonucleolytic and RTase activities, make a specific nick in one of the DNA strands at the insertion site and uses the 3' hydroxyl group that is exposed by this nick to prime the reverse transcription of its RNA transcript. Thus, the R2 protein can recognize specifically the sequence near the 3' end of the RNA transcript for initiation of first-strand synthesis. These authors also proposed that the enzymes in the RNA-mediated mechanism of retrotransposition of R2 might be very similar to those responsible for the retroposition of SINEs since no particular structures required for reverse transcription, such as long terminal repeats, are present in the 3'-terminal regions of SINEs or in those of non-LTR-type retrotransposons.

In the present study, our group demonstrated that the right half of the tRNA-unrelated region of the tortoise Pol III/SINE is actually identical to the 3'-end region of the CR1-like non-LTR-type retrotransposon in the turtle genome. This coincidence suggests that the enzymatic machinery responsible for the retroposition of the tortoise Pol III/SINE might also be the same as that responsible for retrotransposition involving the CR1-like retrotransposon. During the course of our study, Vandergon and Reiman (1994) pointed out the similarity between the chicken CR1 and the tortoise Pol III/SINE (see Introduction). We have extended their observations to demonstrate that tRNA-derived SINEs might be generally composed of a chimeric structure with a tRNA-related region plus the left half of the tRNA-unrelated region and the right half of the tRNA-unrelated region that is homologous to the 3' end of a non-LTR-type retrotransposon. Examples of this organization are

provided by the tortoise Pol III/SINE, the salmonid *Hpa* I, the salmon *Sma* I, the charr *Fok* I and the tobacco TS families. The results strongly suggest that each SINE family recruited the enzymatic machinery for retroposition from the corresponding non-LTR-type retrotransposon through a common "tail" sequence.

The mechanism of retroposition of SINEs might, however, not be identical to that of non-LTR-type retrotransposons. The first step in retroposition is the generation of a nick at the target site. The second step, as shown by Luan et al. (1993), is the priming of the first-strand synthesis by use of the 3' hydroxyl group generated by a nick in one strand of the DNA target site. In the case of SINEs, AT-rich tails of variable length might be generated at this step. It should be noted, however, that the mechanism of generation of the tails seems to differ between SINEs and non-LTR-type retrotransposons. Members of the tortoise Pol III/SINE have AT-rich tails (Endoh et al., 1990), whereas those of the CR1 element do not have any AT-rich sequences (see Results and Silva and Burch, 1989). Generally, while non-LTR-type retrotransposons tend to have simple repeats of A residues at the 3' end (Eickbush, 1992), SINEs appear to have more complex, AT-rich tails, such as $(AAATGT)_n$ in the charr *Fok* I family (Kido et al., 1991) and (TTG)n in the tobacco TS family (Yoshioka et al., 1993).

The molecular mechanisms responsible for these differences are unknown. Members of the class of non-LTR-type retrotransposon are known to be transcribed by RNA polymerase II via the internal promoters that are characteristic of this group (Mizrokhi et al., 1988; Swergold, 1990), whereas SINEs are transcribed by RNA polymease III via the internal promoters of tRNA-related regions (Martigenetti and Brosius, 1995). Therefore, it is possible that distinct mechanisms for termination of transcription of these two transcriptional systems might be responsible for the differences in 3' tails between SINEs and non-LTR-type retrotransposons (Matsumoto et al., 1989; Wahle and Keller, 1992). It remains to be determined what

cellular components, in addition to an RTase from a non-LTR-type retrotransposon, are responsible for synthesis of these tail structures.

(b) A possible mechanism for the initial generation of tRNA-derived SINEs

tRNA-derived SINEs can be classified into a few superfamilies by reference to the tRNA species from which they originated, and $tRNA^{Lys}$ is the most common source of SINEs (Okada, 1991a, b; Okada and Ohshima, 1995). Thus for example, the $tRNA^{Lys}$-related SINE superfamily includes the tortoise Pol III/SINE, the rodent type 2 family, the squid SK family, the salmon *Sma* I family and the charr *Fok* I family (Okada, 1991a, b; Okada and Ohshima, 1995). Our group recently proposed a model to explain how the tRNA-related region of these SINEs might have been generated during evolution (Ohshima et al., 1993; see also Part II). Our model is based on the following observations. First, the tRNA-related regions of these five families of SINEs end with the sequence CCA. The CCA sequence is added posttranscriptionally to tRNAs and is not encoded by their genes. Accordingly, it seems plausible that the $tRNA^{Lys}$-related regions of these SINEs might have been derived from mature $tRNA^{Lys}$ itself. Therefore, we have to postulate a mechanism for the extension of DNA beyond the end of the CCA sequence of $tRNA^{Lys}$ in order to explain the origin of the composite structure of SINEs. This requirement brings to mind the LTR-based mechanism of reverse transcription of retroviruses. The second observation is that conserved DNA segments, namely, GATCTG and TGG, at a distance of 10-11 nucleotides, can be found in the tRNA-unrelated regions of these five SINEs, suggesting a common evolutionary origin for the tRNA-unrelated regions of five SINEs and/or a certain function for this region. To our surprise, we found similar sequences in the sequences complementary to the U5 regions of several retroviruses that use $tRNA^{Lys}$ as a primer for reverse transcription. On the basis of these various

observations, we proposed that a strong-stop DNA with a primer $tRNA^{Lys}$ became a primordial SINE during evolution (Ohshima et al., 1993; see also Part II).

We are now in a position to propose a more detailed possible scenario for the initial generation of SINEs, in which SINEs might have been generated by recombination between a strong-stop DNA with a primer tRNA and the double-stranded DNA of a non-LTR-type retrotransposon. As shown in the case of the *Fok* I, *Sma* I and tortoise Pol III/SINE families, recombination is presumed to have occurred at a site within two conserved motifs mentioned above (Fig. 6). This observation suggests that some preference probably exists for the choice of recombination sites and that the two conserved motifs might have been concerned with the selection. It is now necessary to characterize more SINEs and their corresponding non-LTR-type retrotransposons to reveal the nature of this preference.

All the members of the $tRNA^{Lys}$-related SINE superfamily are not, however, necessarily the direct descendants of a primordial SINE. It is possible that, after a primordial SINE was generated, the 3' tail region that was derived from a non-LTR-type retrotransposon might have been replaced by another retrotransposon. The *Hpa* I family and the *Ava* III family in salmonids might be examples of such a phenomenon. We suggest that the 3' tail of the *Hpa* I family might be derived from the RSg-1 retrotransposon in salmonids. In the *Ava* III family, this 3' tail was exactly replaced by another sequence of a different origin (Kido et al., 1994). In future, we plan to compile pairs of SINEs that have a common head and different tails and pairs of SINEs that have different heads and a common tail. In the case of the tortoise Pol III/SINE, since the tRNA-related region of this family was not detected before the appearance of this family (Fig. 1C), it seems plausible that the tortoise Pol III/SINE might be the direct descendant of the primordial SINE generated by recombination between a strong-stop DNA with a primer $tRNA^{Lys}$ and the CR1-like non-LTR-type retrotransposon.

It has been suggested that non-LTR-type retrotransposons belong to the oldest group of retroelements and that LTR-type retrotransposons emerged after acquisition of long terminal repeats by a non-LTR-type retrotransposon (Doolittle et al., 1989; Xiong and Eickbush, 1990). Then, genuine retroviruses emerged as a consequence of acquisition of env genes by LTR-type retrotransposons. Doolittle et al. (1989) noted that, from their present-day distribution, retroviruses must have appeared well after the evolution of vertebrates and perhaps even after the emergence of mammals. The present study provides a possible mechanism for creation of a novel retroelement by the combination of retroelements that were already present together into one unit. It opens up a new perspective for a better understanding of the evolution of mammalian genomes, which appear to have been endowed with considerable structural fluidity and potential for genetic variability by the presence of various retroviruses and SINEs.

REFERENCES

Blin, N., and Stafford, D. W. (1976). A general method for isolation of high-molecular-weight DNA from eukaryotes. Nucl. Acids Res. 3, 2303-2308.

Burch, J. B. E., Davis, D. L., and Hass, N. B. (1993). Chicken repeat 1 elements contain a *pol*-like open reading frame and belong to the non-long terminal repeat class of retrotransposons. Proc. Natl. Acad. Sci. USA 90, 8199-8203.

Doolittle, R.F., Feng, D.-F., Johnson, M.S., and McClure, M.A. (1989). Origins and evolutionary relationships of retroviruses. Quart. Rev. Biol. 64, 1-30.

Eickbush, T. H. (1992). Transposing without ends: The non-LTR retrotransposable elements. New Biol. 4, 430-440.

Eickbush, T. H. (1994). Origin and evolutionary relationships of retroelements. in *The Evolutionary Biology of Viruses*, eds. Morse, S. S. (Raven Press, NY), pp. 121-157.

Endoh, H., Nagahashi, S., and Okada, N. (1990). A highly repetitive and transcribable sequence in the tortoise genome is probably a retroposon. Eur. J. Biochem. 189, 25-31.

Fanning, T. G., and Singer, M. F. (1987). Biochem. Biophys. Acta 910, 203-212.

Gaffney, E. S. & Meylan, P. A. (1988). A phylogeny of turtles. in *The Phylogeny and Classification of the Tetrapods, vol. 1*, eds. Benton, M. J. (Clarendon Press, Oxford), pp. 157-219.

Hedges, S. B. (1994). Molecular evidence for the origin of birds. Proc. Natl. Acad. Sci. USA 91, 2621-2624.

Kido, Y., Aono. M., Yamaki, T., Matsumoto, K., Murata, S., Saneyoshi, M., and Okada, N. (1991). Shaping and reshaping of salmonid genomes by amplification of tRNA-derived retroposons during evolution. Proc. Natl. Acad. Sci. USA 88, 2326-2330.

Kido, Y., Himberg, M., Takasaki, N., and Okada, N. (1994). Amplification of distinct subfamilies of short interspersed elements during evolution of the salmonidae. J. Mol. Biol. 241, 633-644.

Luan, D. D., Korman, M. H., Jakubczak, J. L., and Eickbush, T. H. (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.

Martignetti, J. A. and Brosius, J. (1995). BC1 RNA: transcriptional analysis of a neural cell-specific RNA polymerase III transcript. Mol. Cell. Biol. 15, 1642-1650.

Matsumoto, K., Takii, T., and Okada, N. (1989). Characterization of a new termination signal for RNA polymerase III responsible for generation of a discrete-sized RNA transcribed from salmon total genomic DNA in a HeLa cell extract. J. Biol. Chem. 264, 1124-1131.

Mizrokhi, L. J., Georgieva, S. G., and Ilyin, Y. V. (1988). *jockey*, a mobile Drosophila element similar to mammalian LINEs, is transcribed from the internal promoter by RNA polymerase II. Cell 54, 685-691.

Ohshima, K., Koishi, R., Matsuo, M., and Okada, N. (1993). 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. Proc. Natl. Acad. Sci. USA 90, 6260-6264.

Ohshima, K., Hamada, M., Terai, Y., and Okada, N. (1995). Short interspersed repetitive elements (SINEs) may have been generated by recombination with long interspersed repetitive elements (LINEs). submitted.

Okada, N. (1991a). SINEs: Short interspersed repeated elements of the eukaryotic genome. Trends in Ecol. & Evol. 6, 358-361.

Okada, N. (1991b). SINEs. Curr. Opin. in Genet. & Dev. 1, 498-504.

Okada, N. and Ohshima, K. (1995). Evolution of tRNA-derived SINEs. in *The Impact of Short Interspersed Elements (SINEs) on the Host Genome*, ed. Maraia, R. J. (R. G. Landes Company, Austin), pp. 61-79.

Rougier, G. W., de la Fuente, M. S., and Arcucci A. B. (1995). Late Triassic turtles from South America. Science 268, 855-858.

Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74, 5463-5467.

Silva, R. and Burch, J. B. E. (1989). Evidence that chicken CR1 elements represent a novel family of retroposons. Mol. Cell. Biol. 9, 3563-3566.

Stumph, W. E., Kristo, P., Tsai, M.-J., and O'Malley, B. W. (1981). A chicken middle-repetitive DNA sequence which shares homology with mammalian ubiquitous repeats. Nucleic Acids Res. 8, 5383-5397.

Swergold, G. D. (1990). Identification, characterization, and cell specificity of a human LINE-1 promoter. Mol. Cell. Biol. 10, 6718-6729.

Ullu, E. & Tschudi, C. (1984). *Alu* sequences are processed 7SL RNA genes. Nature 312, 171-172.

Vandergon, T. L. and Reitman, M. (1994). Evolution of chicken repeat 1 (CR1) elements: evidence for ancient subfamilies and multiple progenitors. Mol. Biol. Evol. 11, 886-898.

Wahle, E. and Keller, W. (1992). The biochemistry of 3'-end cleavage and polyadenylation of messenger RNA precursors. Annu. Rev. Biochem. 61, 419-440.

Weiner, A. M. (1980). An abundant cytoplasmic 7S RNA is complementary to the dominant interspersed middle repetitive DNA sequence family in the human genome. Cell 22, 209-218.

Winkfein, R. J., Moir, R.D., Krawetz, S. A., Blanco, J., States, J. C., and Dixon, G. H. (1988). A new family of repetitive, retroposon-like sequences in the genome of the rainbow trout. Eur. J. Biochem. 176, 255-264.

Xiong, Y., and Eickbush, T.H. (1990). Origin and evolution of retroelements based upon their reverse transcriptases. EMBO J. 9, 3353-3362.

Yoshioka, Y., Matsumoto, S., Kojima, S., Ohshima, K., Okada, N., and Machida, Y. (1993). Molecular characterization of a short interspersed repetitive element from tobacco that exhibits sequence homology to specific tRNAs. Proc. Natl. Acad. Sci. USA 90, 6562-6566.

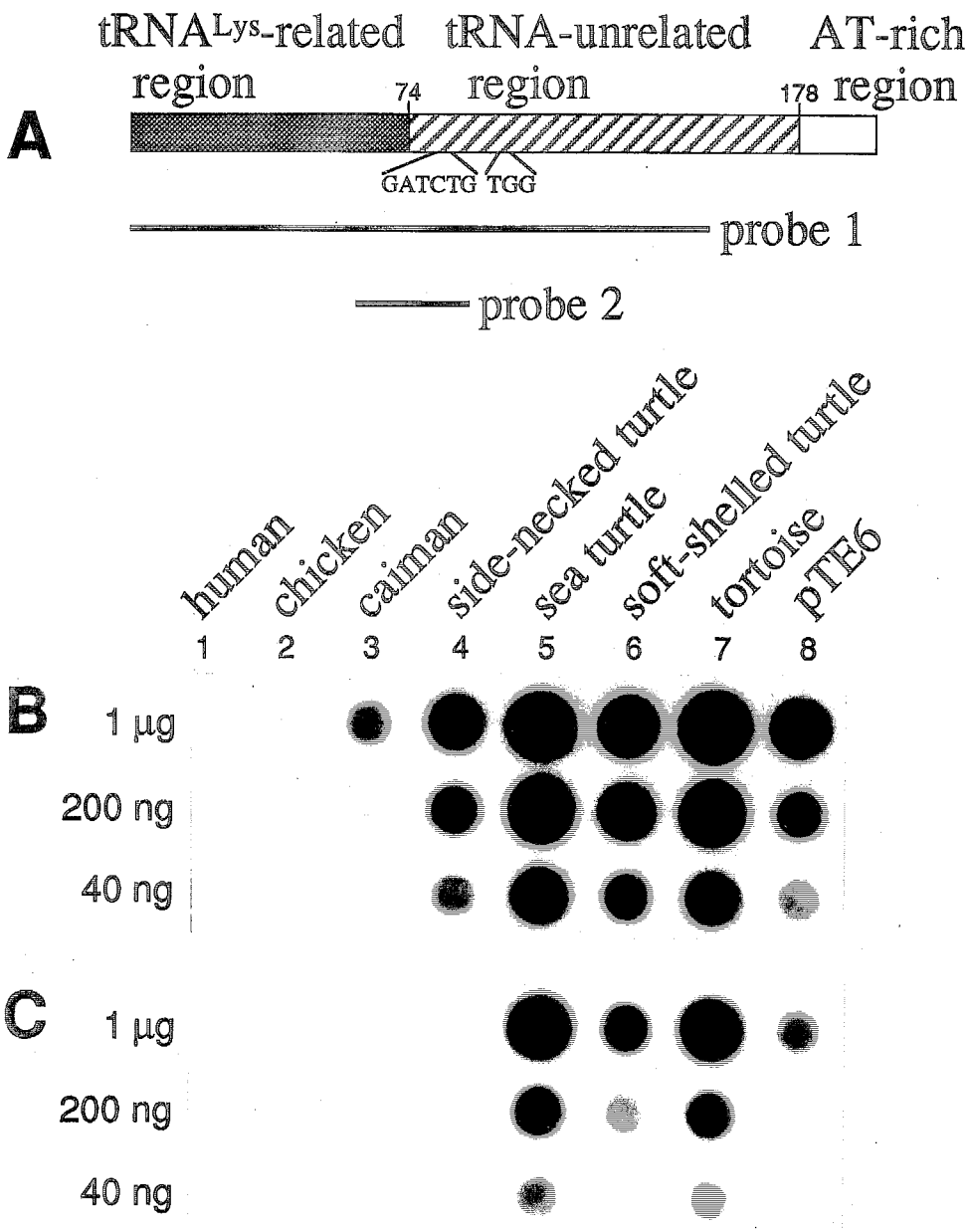


Figure 1. The distribution of the tortoise Pol III/SINE-like sequence in turtles and living relatives. (A) Schematic representation of the composite structure of the tortoise Pol III/SINE, and an indication of the regions used as probes. Dot hybridization experiments were performed using probe 1 (B) and probe 2 (C), respectively. Progressively decreasing amounts of total DNA (1 µ g, 200 ng and 40 ng) from individuals of the indicated species were dotted on a membrane (columns 1 to 7). In column 8, progressively decreasing amounts of cloned DNA (10 ng, 2 ng and 400 pg) of the tortoise Pol III/SINE (TE6; see Endoh et al., 1990) were dotted on the membrane as a control.

```
                     •         -360•         •         -340•         •         -320•         •         -300•         •         -280•
CR1 (chicken)  CCTGGAGAAGAGAAGGCTGCGCGGTGACCTCATTGCAGCCTTTCAGTACCTGAAGGGAGCCTATAAACAGGAAGGGAGTAAACTCTTTGAAAGGGTAGAT
                *** **********  ** * ** **  * ** ********* * ** ***** ** *  *  *   **** **  *** ** ***  *  *** *
CONSE(Ps)       CTGCAGAAGAGARRAATGAGGGGGGATTTGATAGCAGCCTTTAACTATCYSAAAGGGGGTTCCAGGGAGGATGGAGCTARGCTGYTTTCAGWGGTGRCA
4-2(Ps)        T••••••••••••GG•••••••••••••••••••••••••••••••••T•TG••••••••••••••••••••••••••••G••••C••••••T••••A••
4-1(Ps)         ••••••••••••AA•••••••••••••••••••••••••••••••••-•CC••••••••••••••••••••••••••••A••••T••••••A••••G••

                •         -260•         •         -240•         •         -220•         •         -200•         •         -180•
CR1 (chicken)   AACAGCAGGACAAGGGGGAACGGTTTTAAGTTGAAAGAGGGAAGATTTAGGTTGGATGTTAGGGGGAAGTTCTTTACCAGGAGAGTGGTGAGGTGCTGGA
                 *   *** ****** * ** *** * ****** *   *** ** * *********  *****   ***** ** ** ***** ******* *  *****
CONSE(Ps)       GATGACAGAACAAGGARCAATGGTCTCAAGTTGCAGTGGGGGAGGTCTAGGTTGGACATTAGGAAAAAGTTTTTCACTAGGAGGGTGGTGAAGCACTGGA
4-2(Ps)         ••••••••••••••••G•••••••••••••••••••••••••••••••••••••••••••••••••c••••••••••••••••a••••••••••••••••
4-1(Ps)         ••••••••••••••••A•••••••••••••••••••••••c••••••••••••••••••••••••••••••••••••••••••••••••••••••t••••
2-3(Ps)         tctgacagcggtctataaaatcattttgaatgaattg•t•••••••••••••••••t••••••g•••ca•••••••••••••••••••••••••••••••

                •         -160•         •         -140•         •         -120•         •         -100•         •         -80•
CR1 (chicken)   ACAGGCTGCCCAGAGAGGTTGTGGATGCTCTGTCCCTGGAGGTGTTCAAGGCCAGGTTGGATGGGGCCCTGGGCAACCTGGTCTAGTAAATGGGATGTT
                *  ** * ** ** ***** *****  ***  **  * ***** ** ***    ** * **    ****      ****  *  *  * *  ** ***
CONSE(Ps)       ATGGGTTACCTAGGGAGGTGGTGGAATCTCCATCTTTAGAGGTTTTTAAGCTGCGGCTTGACAAAACCCT-----CGCTGGGATGATTTA-GCTGAGGTT
4-2(Ps)         ••••••••••c•••••t••••••••••••••••••••••••••••••••••••••••••••••••••ta•-------••c••••••••g••-a••••-ta•
4-1(Ps)         ••a•••••••••••••••••••••c•••••t•--c•••••••••a•••••••••••••••••••••••••-----•••••••••••••••-•••ag••••
2-3(Ps)         •••t••••••••••••••••••••••••••••••••••a••••••••••••••••••••••••••g••••-----g••••••••••••••-•••••••••

                     •       -60•         •         -40•         •         -20•         •
CR1 (chicken)   GGTGGCCCTGCCCAGCAGGGGGGTTGGA-GATTCGTGATCCTCGAGGTCCCTTCCAACCCAGGCCATTCTGTGATTCTGT
                 **    **    ********* *****  *    *    **** *************** *     ***** ******* *
CONSE(Ps)       AGTCCTGCT-TTGAGCAGGGGG-TTGGACTAGATGACCTCCT-GAGGTCCCTTCCAACTC-TAATATTCTATGATTCTAT
4-2(Ps)         •••••••••-••••••••••••-•••••••••••••••••••-•••••••••••••••••-•••••••••••••••••••aaagaattc
4-1(Ps)         g••••••••-••••••••••••-••••••••••••••••t••-••••••••••••••t••-•••••••••••••••••c•gacatggcatgccacatccg
2-3(Ps)         •••••••••-••••••••••••-•••••••••••••••••••-•••••••••••••••••-•••••••••••t••••••aaagggaatagggaattgtta
```

Figure 2. A CR1-like retrotransposon was found in the side-necked turtle. Several clones [4-2(Ps), 4-1(Ps) and 2-3(Ps)] were isolated from the turtle using probe 1 and their sequences were determined. A consensus sequence of these clones [CONSE(Ps)] was aligned with the sequence of chicken CR1 (Burch et al., 1993). The nucleotides are numbered from the 3' terminus, as is the custom for retrotransposons. Asterisks (*) indicate the nucleotides of the consensus sequence of turtle that are identical to those of chicken. Dots (•) indicate the nucleotides of clones identical to those of the consensus sequence of turtle. Gaps (–) have been introduced to maximize homology.

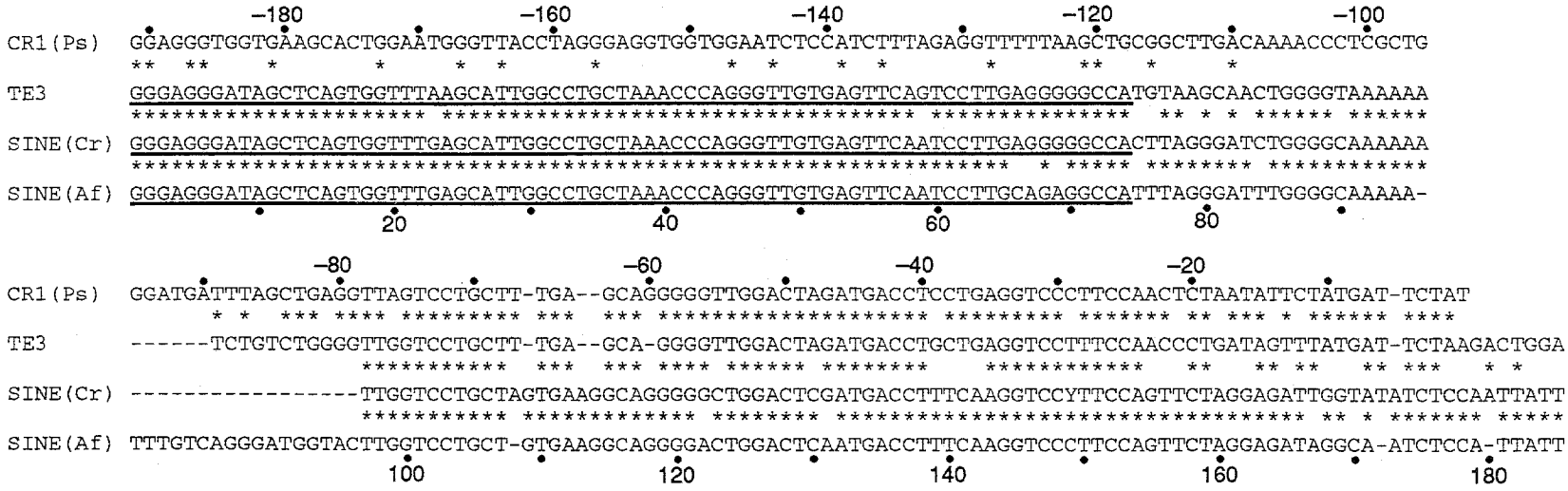


Figure 3. The 3'-end structure common to the CR1-like retrotransposon and the tortoise Pol III/SINE. The consensus sequence of the CR1-like element, designated CR1(Ps), is compared with those of the Pol III/SINE from tortoise, designated SINE(Cr), and from soft-shelled turtle, designated SINE(Af). One individual member of the SINE(Cr) family, designated TE3 is included. Identical nucleotides in two sequences are indicated by asterisks. The tRNA ^{Lys}-related regions are underlined. In the case of SINEs, nucleotides are numbered from the 5' terminus of SINE(Cr). In the case of CR1(Ps), the nucleotides are numbered from the 3' terminus.

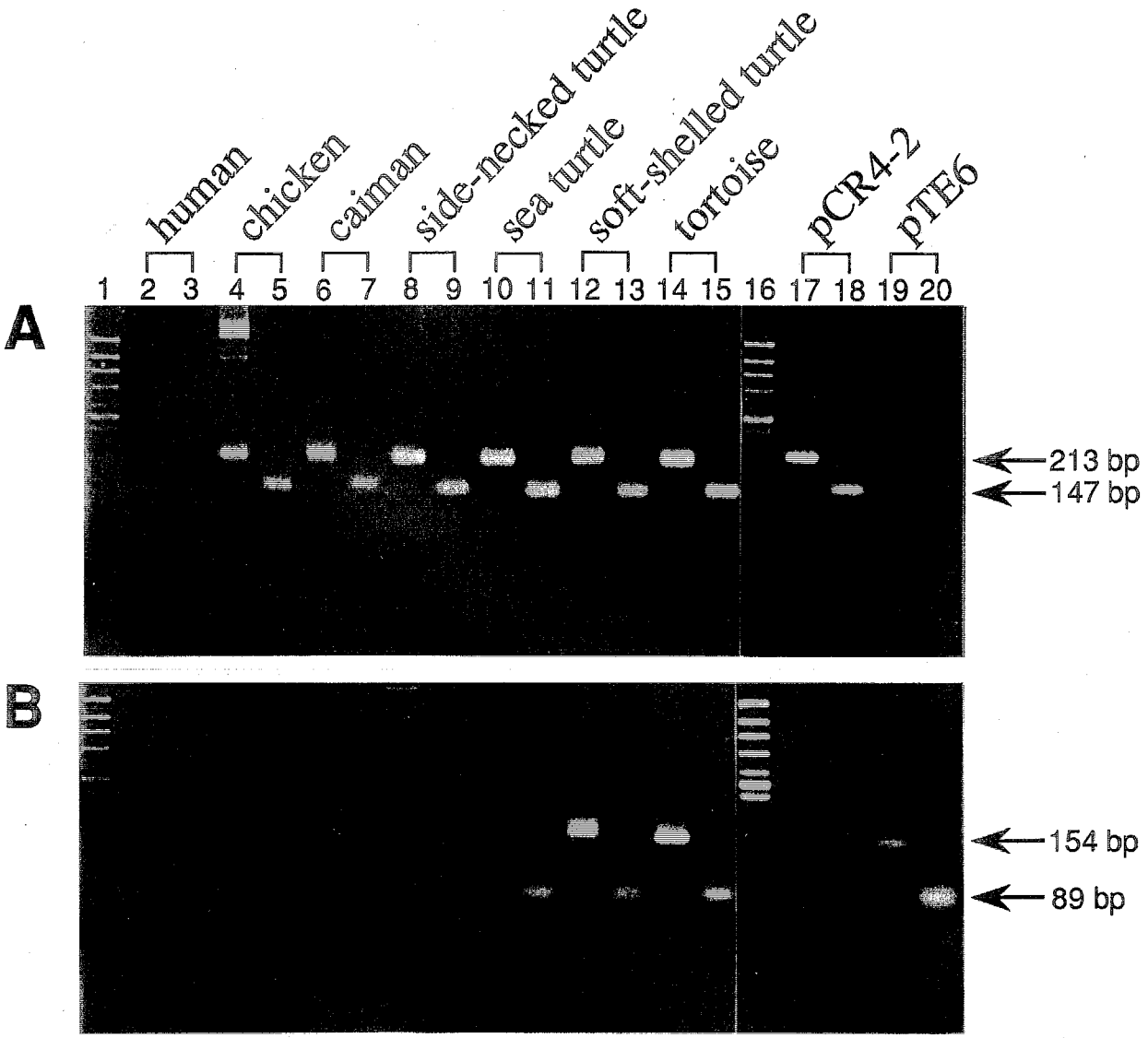


Figure 4. The CR1-like retrotransposon is present in all of the reptiles examined and in a bird, whereas the tortoise Pol III/SINE is confined to a dominant group of living turtles. PCR analysis was performed using two sets of primers specific to the CR1 element (A), and two sets of primers specific to the SINE (B), respectively. Ten ng of total DNA from each species (lanes 2 to 15), and 100 pg (A) or 1 ng (B) of cloned DNA (lanes 17 to 20) were used as templates. Products by PCR that were amplified by each set of primers are indicated by arrows. Although, in lane 10 of panel B, no product of 154 bp is visible, a product of expected size was detected when another primer, which started at 20 bp upstream of TER0, was used (data not shown). Lanes 1 and 16 show size markers.

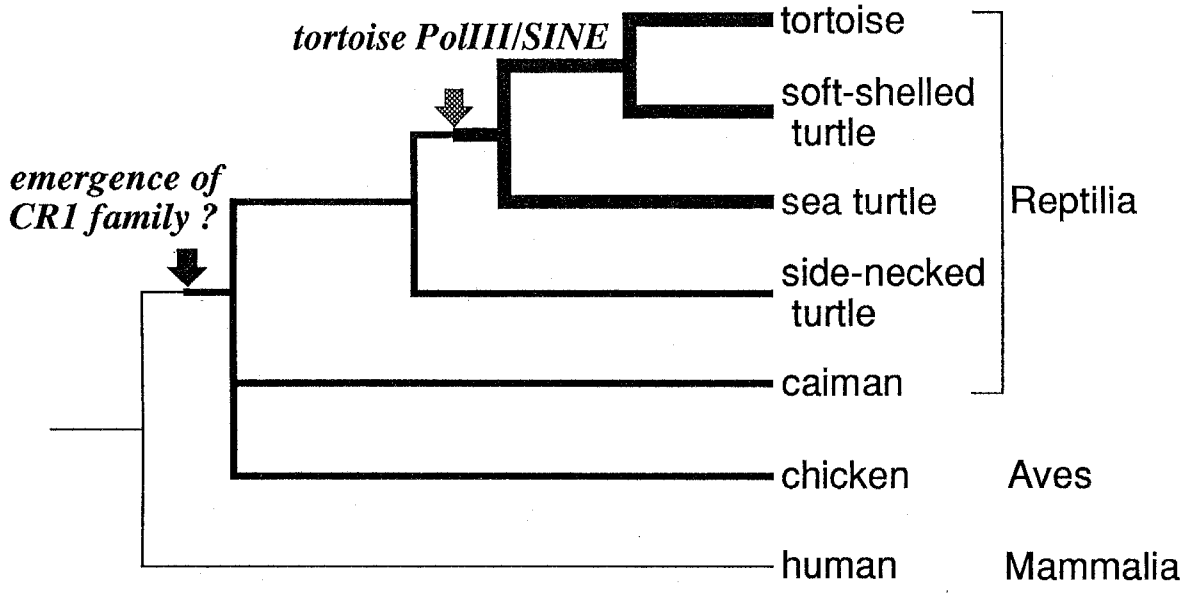


Figure 5. The timing of the appearance during evolution of the CR1-like retrotransposon and of the tortoise Pol III/SINE. Phylogenetic relationships among amniota are taken from the work of Hedges (1994).

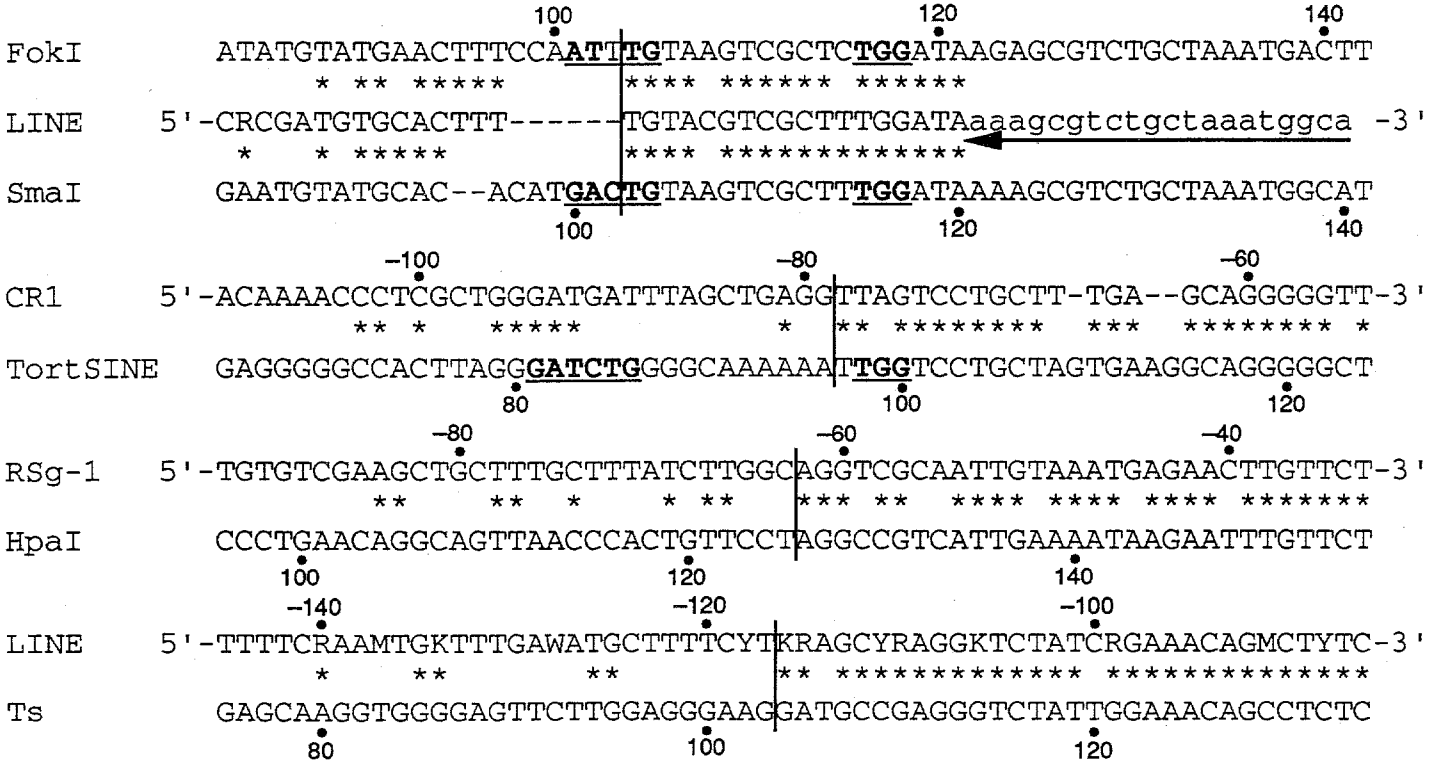


Figure 6. The junction sequences between SINEs and retrotransposons. Junctions are shown by vertical lines, and they are located within two conserved motifs, namely, GATCTG and TGG (boldface), which were detected in superfamily of tRNA Lys-related SINEs (Ohshima et al, 1993). Numbering systems are as in Figure 3. The charr *Fok* I and the salmon *Sma* I families are aligned togather with a putative retrotransposon because they have almost identical tails. The sequence of the primer used to isolate the retrotransposon is indicated by an arrow. R=G/A, Y=C/T, W=A/T, M=A/C, K=T/G.

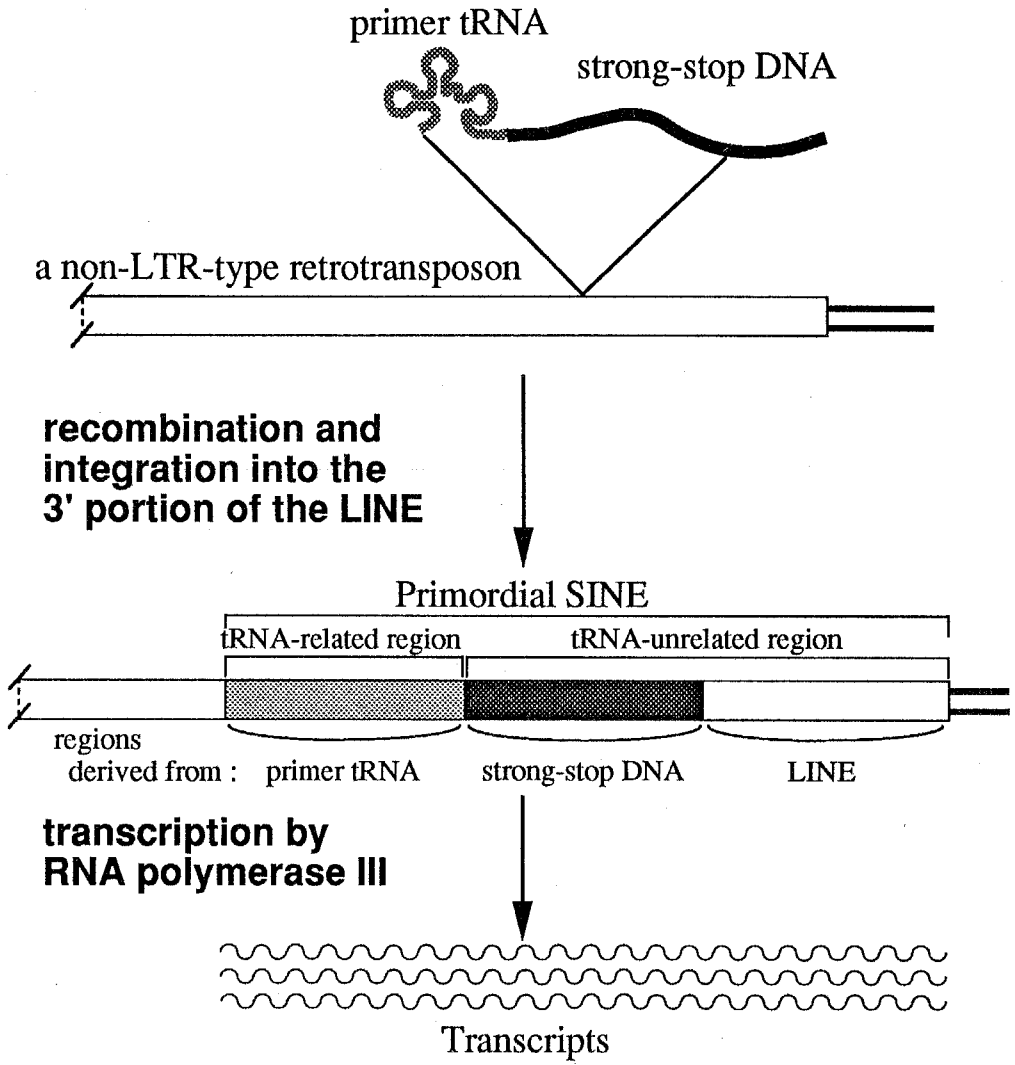


Figure 7. A possible model for the initial generation of tRNA-derived SINEs.

Table 1. ***Classification of living turtles***

Order	Megaorder	Superfamily
Testudines (turtles)	Cryptodira (vertical-necked turtles)	Testudinoidea (tortoises) Trionychoidea (soft-shelled turtles) Cheloniodea (sea turtles)
	Pleurodira (side-necked turtles)	

GENERAL DISCUSSION

I am now in a position to answer the problems mentioned in the chapter of General Introduction. The first question is whether tRNA-derived SINEs are ubiquitous in the biological world or those are confined to some specific taxa. It is possible for me to answer this problem more clearly than before. Namely, tRNA-derived SINEs are probably ubiquitous in invertebrates, as well as in vertebrates. To date, fourteen families of SINEs that are derived from tRNAs have been reported from vertebrates. By contrast, only three examples of SINEs from invertebrate phyla—arthropod, echinoderm and flatworm—had been reported before this work. In this thesis, I described four mollusk SINEs. They all have structural features common to vertebrate SINEs that are derived from tRNA (see Part I, II). In addition, the same preference as vertebrates for progenitor species of tRNA might be present. These results strongly suggest that tRNA-derived SINEs are ubiquitous in metazoans and that there must be some general mechanism for the generation of tRNA-derived SINEs.

Second, how has been constructed the composite structure of tRNA-derived SINEs? To answer this problem, our group proposed that a possible origin of the region that is homologus to the tRNA of a SINE might be a primer tRNA, attached to a 'strong-stop DNA', which is an intermediate during reverse transcription of certain retroviruses and retrotransposons (see Part II).

Third, what is the nature of enzymatic machinery of SINE amplification? Our group demonstrated that tRNA-derived SINEs might be generally composed of a chimeric structure with a tRNA-related region plus the left half of the tRNA-unrelated region and the right half of the tRNA-unrelated region that is homologous to 3' end of a non-LTR-type retrotransposon (see Part III). Examples of this organization are provided by the tortoise Pol III/SINE, the salmonid *Hpa* I, the salmon *Sma* I, the charr *Fok* I and the tobacco TS families. The results strongly

suggest that each SINE family recruited the enzymatic machinery for retroposition from corresponding non-LTR-type retrotransposon through a common "tail" sequence.

Horizontal Transmission of SINEs and the Enzymes for Their Retroposition

With respect to the question of how several SINEs in distant species happen to have similar sequences (see Part II), 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, or that a retrotransposon with LTR transferred horizontally between such organisms and that the "strong stop DNA" with a primer tRNA gave rise independently to the respective SINE within each organism. 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. In this case, the transmitted sequence might have exchanged its 3' tail sequence between a non-LTR-type retrotransposon of new host genome and the recombinant might have acquired retropositional activity through a common tail sequence in the new host genome. It is now necessary to characterize more non-LTR-type retrotransposons which provide enzymatic source for retroposition to tRNA-derived SINEs if we are to identify the dominant mechanism for the dispersion of SINEs throughout the biological world.

SINEs and Evolution of Retroelements

The results of this thesis may open up a new perspective for a better understanding of the evolution of retroelements. It has been suggested that non-LTR-type retrotransposons belong to the oldest group of retroelements and that LTR-type retrotransposons emerged after acquisition of long terminal repeats by a non-LTR-type retrotransposon (Xiong and Eickbush, 1990). Then, genuine

retroviruses emerged as a consequence of acquisition of env genes by LTR-type retrotransposons. The present thesis provides a possible mechanism for creation of a novel retroelement by the combination of retroelements that were already present together into one unit (see Part III). If this is the case, we can place this peculiar retroelement on the context of the evolution of retrotransposons. Namely, the evolution of SINEs might be a kind of roundabout way for the evolution of retrotransposons. Efficient promoters from a primer tRNA, attached to a 'strong-stop DNA' of LTR-type retrotranposons or of retroviruses and an effective signal for reverse transcription from non-LTR-type retrotransposons might have combined to form a small, but an efficient retroelement. Although SINEs do not encode their own reverse transcriptases, in recompense for it, their size is small. This advantage of size for efficiency at reverse transcription or for a limited capacity of host genomes might have allowed them to be extremely abundant in the host genomes.

Biological Significance of SINEs (Makalowski, 1995; Labuda et al., 1995; Brosius, 1991; Brosius and Gould, 1992)

During the evolution of these elements, the host genomes might have been endowed with structural fluidity and genetic variability. In the following essay, I will discuss the last problem, biological significance of SINEs.

Retroelements are major components of eukaryotic genomes (Singer and Berg, 1991). Among them, the most abundant are SINEs and non-LTR-type retrotransposons. These elements often make up the tenth or more of a genome. Are such sequences necessary to the organism? On the contrary, are those merely molecular parasites? They are often quoted as selfish or junk DNA. According to Orgel and Crick (1980), to be considered selfish a piece of DNA must merely be able to generate additional copies of itself and to make no specific contribution to organism's phenotype. The idea of selfish DNA and ambiguous or even derogatory

names such as "pseudo"gene and "junk"DNA arose at the time when our knowledge of biology of repetitive elements was very narrow. Our view of the entire phenomenon of repetitive elements has now be revised in light of data on their biology and evolution as rapidly growing sequence data. For simplicity, I will focus on SINEs although the following view point of this problem is applicable to other retroelements, such as retropseudogenes.

Does the Majority of Members of a SINE Have Essential Function for the Host Cell?

One possible answer to this problem is that the majority of members of a SINE might have some fundamental cellular function(s). Several researchers have argued that primate *Alu* family must perform specific function in the host genomes. R. J. Maraia proposed a possible role of *Alu*s in the regulation of protein synthesis (Maraia and Sarrowa, 1995). Moreover, It is suggested by C. W. Schmid that *Alu*s might be required in cell stress response (Liu et al., 1995). According to B. H. Howard, a substantial fraction of members of *Alu* family might possess the capacity to modify the local organization of chromatin (Howard et al., 1995). Although these arguments are suggestive, it seems that more extensive research is necessary to verify these opinions.

The Impact of Local Members of a SINE on the Host Genome

Even if we cannot define the specific functions of SINEs, we can nonetheless show another possibility that SINEs might have multiple different functional properties depending on their local context within the genome. Accumulation of sequence data on eukaryotic genomes, especially on human's, provides us many examples of the impact of respective members of *Alu* family on human genome (Makalowski, 1995; Labuda et al., 1995). We can divide such examples into the following four categories.

First, SINEs serve as recombination hot spots. The repetitive elements play an important role in the unequal recombination events. *Alu* family appear to have facilitated the genomic rearrangements leading to changes in the number and the order of genes as well as their sequence identity through homologus recombination and/or gene conversion, thus promoting the creation of gene families. The human glycophorin gene family, for example, evolved through several duplication and recombination steps (Onda et al., 1993). The most important step in this respect seems to be a recombination event between an *Alu* within the intron of precursor glycophorin gene and another *Alu* element located more than 9 kb downstream of the glycophorin gene. This recombination event created the unique 3' end of the glycophorin B and E gene.

Second, SINEs can constitute translated parts of host genes. It has been shown that *Alu* elements nested in the opposite orientation with respect to the direction of the host gene transcription are predisposed to become regular exons because of the internal potential splicing sites. By activation of these splicing sites, a part of an *Alu* element can be introduced into the mRNA as an additional or alternative exon. For example, about 10% of the human decay accelerating factor (DAF) mRNA contains the *Alu* cassette (Caras et al., 1987). DAF is a cell membrane glycoprotein that binds activated complement. The introduction of the *Alu* cassette into DAF mRNA has been expected to create a hydrophilic carboxy-terminal region in the peptide, which would inhibit the migration of DAF into the cell membrane. Caras et al. (1987) observed that DAF translated from a wild-type message was membrane-bound, while the DAF peptide expressed from *Alu*-containing cDNA was not. They hypothesized that a fraction of the *Alu*-containing DAF mRNA in a normal cell probably accounts for the soluble form of DAF.

Third, SINEs can be a source of regulatory elements for gene expression. Analysis of the region upsream of the Θ-1 globin locus revealed one full and one truncated *Alu* element (Kim et al., 1989). The latter, oriented in the opposite

direction to the transcription of the Θ-1 globin, serves as a source of the canonical CCAAT signal (positions 137-133 according to the *Alu* consensus). This signal, along with the TATA motif just downstream of the *Alu* sequence, acts as a functional promoter for the nearby Θ-1 globin gene.

Fourth, SINEs can be recruited for new cellular functions. Two neuronal small RNAs, rodent BC1 and primate BC200 that are derived from ID elements and *Alu* family, respectively, show very specific and similar patterns of tissue-specific expression (DeChiara and Brosius, 1987; Tiedge et al., 1993). Although the specific function of these RNAs is not known, the experimental data suggest that they might be functional (Tiedge et al., 1993; Kobayashi et al., 1992).

These observations suggest that SINEs have provided novel functions to the host genome over a long evolutionary time period.

Generality of the Impact of SINEs on the Host Genome

These observations mainly come from human *Alu* family. Here I would like to inquire whther the impact of *Alu* family on the human genome is also the case of invertebrate SINEs. In this thesis, I described two examples of genomic rearrangement mediated by invertebrate SINEs (see Part I). These results suggest that SINEs may promote the generation of novel structure of the genome and contribute to new cellular functions depending on their local context within the genome throughout the biological world.

Potonuons or Seeds of Evolution (Brosius, 1991; Brosius and Gould, 1992)

J. Brosius (1991) referred to this problem with intriguing metaphor. " ... Apart from the ability of retroposon to keep the genome in flux, thus favoring genetic diversity, they can be considered a shotgun approach of nature wherein the majority of these genetic elements are inactive and left to rot in the genomic soil. Nevertheless, some seeds will integrate near a fertile genomic environment, giving

rise (usually after mutational alterations) to new genes or gene domains (the second and the fourth categories of the impact of SINEs mentioned above: the author of this thesis annotated) and complementing the conventional gene duplication that is essential to evolution (the first category: the author). Retroposition may also match existing genes with new regulatory elements (the third category: the author) ..."

I am expecting to find cases where these "seeds" gave rise to evolutionary progress on the morphological level as well as on the molecular level in the near future.

References

Brosius, J. (1991). Retroposons–seeds of evolution. Science 251, 753.

Brosius, J. and Gould, S. J. (1992). On "genomenclature": a comprehensive (and respectful) taxonomy for pseudogenes and other "junk DNA". Proc. Natl. Acad. Sci. USA 89, 10706-10710.

Caras, I. W., et al. (1987). Cloning of decay-accelerating factor suggests novel use of splicing to generate two proteins. Nature 325, 545-549.

DeChiara, T. M., Brosius, J. (1987). Neural BC1 RNA: cDNA clones reveal nonrepetitive sequence content [published erratum appears in Proc. Natl. Acad. Sci. USA (1987) 84, 4895.] Proc. Natl. Acad. Sci. USA 84, 2624-2628.

Howard, B. H., Russanova, V. R., and Englander, E. W. (1995). Alu silencing mechanisms: implications for the modulation of local chromatin structure. in *The Impact of Short Interspersed Elements (SINEs) on the Host Genome*, ed. Maraia, R. J. (R. G. Landes Company, Austin), pp. 133-142.

Kim, J. H., Yu, C. Y., Bailey, A., et al. (1989). Unique sequence organization and erythroid cell-specific nuclear factor-binding of mammalian Θ1 globin promoters. Nucleic Acids Res. 17, 5687-5700.

Kobayashi, S., et al. (1992). The 10 S BC-1 ribonucleoprotein particle contains identifier sequence- binding proteins that interact with an array of GCAAG/CTTGC motifs between split promoter sequences for RNA polymerase III. J. Biol. Chem. 267, 18291-18297.

Labuda, D., Ziçtkiewicz, E., and Mitchell, G. A. (1995). Alu elements as a source of genomic variation: deleterious effects and evolutionary novelties. in *The Impact of Short Interspersed Elements (SINEs) on the Host Genome*, ed. Maraia, R. J. (R. G. Landes Company, Austin), pp. 1-24.

Liu, W.-M., Chu, W.-M., Choudary, P. V., and Schmid, C. W. (1995). Cell stress and translational inhibitors transiently increase the abundance of mammalian SINE transcripts. Nucleic Acids Res. 23, 1758-1765.

Makalowski, W. (1995). SINEs as a genomic scrap yard: an essay on genomic evolution. in *The Impact of Short Interspersed Elements (SINEs) on the Host Genome*, ed. Maraia, R. J. (R. G. Landes Company, Austin), pp. 81-104.

Maraia, R. J. and Sarrowa, J. (1995). Alu-family SINE RNA: interacting proteins and pathways of expression. in *The Impact of Short Interspersed Elements (SINEs) on the Host Genome*, ed. Maraia, R. J. (R. G. Landes Company, Austin), pp. 163-196.

Onda, M., Kudo, S., Rearden, A., et al. (1993). Identification of a precursor genomic segment that provided a sequence unique to glycophorin B and E genes. Proc. Natl. Acad. Sci. USA 90, 7220-7224.

Orgel, L. E. and Crick, F. H. (1980). Selfish DNA: the ultimate parasite. Nature 284, 604-607.

Singer, M. & Berg, P. (1991). Genes & Genomes, University Science Books, Mill Valley, CA.

Tiedge, H., Chen, W., and Brosius, J. (1993). Primary structure, neural-specific expression, and dendritic location of human BC200 RNA. J. Neurosci. 13, 2382-2390.

Xiong, Y., and Eickbush, T.H. (1990). Origin and evolution of retroelements based upon their reverse transcriptases. EMBO J. 9, 3353-3362.

ACKNOWLEDGMENT

First I would like to express my sincere gratitude to professor Norihiro Okada of Tokyo Institute of Technology for his valuable guidance and useful discussion throughout the course of my research.

I am greatly indebted to Mr. Ryuta Koishi for his earlier work toward characterization of mollusk SINEs.

I also would like to thank Mr. Shigehisa Nagahashi for his teaching about the experimental techniques and for his constant encouragement.

I also thank the other members of Dr. Okada's laboratory for their exciting discussions as well as practical advice.

Finally, I thank my parents, Toshihiko and Noriko.