Retropositional Parasitism of SINEs on LINEs: Identification of SINEs and LINEs in Elasmobranchs

Ikuo Ogiwara,* Masaki Miya,† Kazuhiko Ohshima,* and Norihiro Okada*

*Faculty of Bioscience and Biotechnology, Tokyo Institute of Technology, Yokohama, Japan; and †Department of Zoology, Natural History Museum and Institute, Chiba, Japan

Some previously unidentified short interspersed repetitive elements (SINEs) and long interspersed repetitive element (LINEs) were isolated from various higher elasmobranchs (sharks, skates, and rays) and characterized. These SINEs, members of the HE1 SINE family, were tRNA-derived and were widespread in higher elasmobranchs. The 3′-tail region of this SINE family was strongly conserved among elasmobranchs. The LINEs, members of the HER1 LINE family, encoded an amino acid sequence similar to that encoded by the chicken CR1 LINE family, and they contained a strongly conserved 3′-tail region in the 3′ untranslated region. This tail region of the HER1 LINE family was almost identical to that of the HE1 SINE family. Thus, the HE1 SINE family and the HER1 LINE family provide a clear example of a pair of SINEs and LINEs that share the same tail region. Conservation of the secondary structures of the tail regions, as well as of the nucleotide sequences, between the HE1 SINE family and HER1 LINE family during evolution suggests that SINEs utilize the enzymatic machinery for retroposition of LINEs through the recognition of higher-order structures of the conserved 3′-tail region. A discussion is presented of the parasitism of SINEs on LINEs during the evolution of these retroposons.

Introduction

Mobile genetic elements are major components of the genomes of higher eukaryotes (Boeke and Devine 1998). The mobile elements that are amplified by reverse transcription of an RNA intermediate are called retroposons (Rogers 1985; Weiner, Deininger, and Efstratiadis 1986). Retroposons can be divided into two groups: those that do not encode a reverse transcriptase (RTase) and those that do. The former group includes short interspersed repetitive elements (SINES), processed pseudogenes, and other pseudogenes for small nuclear RNAs (snRNAs; Singer 1982; Weiner, Deininger, and Efstratiadis 1986). The latter group can be further divided into two subgroups on the basis of structure: long terminal repeat (LTR) retrotransposons and non-LTR retrotransposons. Non-LTR retrotransposons are also known as long interspersed repetitive elements (LINEs).

SINES range from 100 to 500 bp in length and each contains an internal promoter for RNA polymerase III (Okada 1991a, 1991b; Schmid and Maraia 1992; Deininger and Batzer 1993; Schmid 1996, 1998; Shendelle and Okada 1999). Since SINES lack open reading frames (ORFs), it is likely that the enzymes required for their retroposition are provided in trans. With the exception of the primate Alu family and the rodent B1 family, which are derived from 7SL RNA (Weiner 1980; Ullu and Tscharudi 1984), all families of SINES examined to date appear to be derived from tRNAs (Okada 1991a, 1991b; Okada and Ohshima 1995; Shendell and Okada 1999). The tRNA-derived SINES are not simple pseudogenes for tRNAs, but have a composite structure that consists of a tRNA-related region and a tRNA-unrelated region.

Key words: SINES, LINEs, retroposition, evolution.

Address for correspondence and reprints: Norihiro Okada, Faculty of Bioscience and Biotechnology, Tokyo Institute of Technology, 4259 Nagatsuta-cho, Midori-ku, Yokohama 226-8501, Japan. E-mail: nokada@bio.titech.ac.jp.

Mol. Biol. Evol. 16(9):1238–1250. 1999
© 1999 by the Society for Molecular Biology and Evolution. ISSN: 0737-4038

LINEs range from about 3 to 7 kb in length, and each includes an ORF that encodes an RTase (Xiong and Eickbush 1990) and, usually, an endonuclease (ENase; Martín et al. 1995, 1996; Feng et al. 1996). The best current model for the mechanism of retroposition of LINEs comes from work on the insect R2Bm element (Luan et al. 1993; Luan and Eickbush 1995). The R2Bm element includes a single ORF that encodes a protein with both ENase activity and RTase activity. The R2 ENase nicks the target site in DNA and the R2 RTase uses the 3′-OH group generated at the target site to prime the reverse transcription of the R2 template. About 250 nucleotides in the 3′ untranslated region (UTR) of the R2 transcript are responsible for recognition by the RTase during retroposition. This mechanism, known as target-primed reverse transcription (TPRT), is also believed to operate in retroposition by other LINEs and group II introns. (Schwarz-Sommer et al. 1987; Bucheton 1990; Levis et al. 1993; Zimmerly et al. 1995; Feng et al. 1996; Moran et al. 1996).

The molecular mechanism for retroposition of SINEs during evolution is unknown since no known SINE encodes a protein. Recently, Jurka’s group and our group suggested, from different pieces of evidence, that LINEs might be a source of the enzymatic machinery that is required for the retroposition of SINES (Jurka and Kl Fogowski 1996; Ohshima et al. 1996; Jurka et al. 1997). Jurka’s group analyzed the sequences of target site duplications of the primate Alu SINE, the rodent ID SINE, the rodent B1 SINE, and the rodent B2 SINE, and they suggested that the mammalian L1 family and these SINE families might share the same consensus target site duplication (Feng et al. 1996; Jurka and Kl Fogowski 1996; Jurka 1997; Cost and Boeke 1998). The association of these SINES and L1 suggests the possibility that the L1 ENase and RTase might be involved in the retroposition of these SINES (Boeke 1997; Jurka 1997).

We have identified several clear examples of the phenomenon whereby the 3′-tail region of certain
tRNA-derived SINEs is identical to the 3’ portion of the 3’ UTRs of certain LINEs. The tortoise Pol III/SINE and the turtle PsCR1 LINE were first described as a pair of SINEs and LINEs with the same 3’-tail region (Ohshima et al. 1996; Kajikawa, Ohshima, and Okada 1997). Then, other examples were identified, such as the ruminant BovA SINE and BovB LINE (Szmraj et al. 1995; Smit 1996; Okada and Hamada 1997), the mammalian MIR SINE and LINE2 (Smit 1996), and the cichlid AFC SINE and CiLINE2 (Terai, Takahashi, and Oka- mura et al. 1996; Kajikawa, Ohshima, and Okada 1997). Recently, Gilbert and Labuda (1999) reported that two MIR-like SINEs share their 3’-tail regions with those of the CR1 LINE and BovB LINE, respectively. These findings suggest that several families of SINEs were generated by recombination between a tRNA-derived precursor and a preexisting LINE at a certain stage of evolution (Ohshima et al. 1996; Okada et al. 1997). Although we had no supporting experimental evidence, we proposed that an RTase encoded by a LINE might recognize the 3’-tail region of SINEs, just as the RTase recognizes the 3’-tail region of LINEs during the retro- position of LINEs.

In the present study, we identified the HE1 SINES and HER1 LINEs in the genomes of elasmobranchs. They provide the clearest and most convincing example to date of a pair of SINES and LINEs that share the same 3’-tail region.

**Materials and Methods**

**Transcription In Vitro of Total Genomic DNA**

Transcription of total genomic DNA in vitro in an extract of HeLa cells has previously been used for the isolation of SINEs (Endoh and Okada 1986), and we used this method to isolate SINE sequences from the genome of the gummy shark.

**Construction and Screening of Genomic Libraries**

Genomic libraries were constructed by ligation of λgt10 arms with genomic DNA from Mustelus manazo (gummy shark), Scyliorhinus torazame (cat shark), and Heterodontus japonicus (bullhead shark) that had been completely digested with EcoRI. For isolation of DNA clones with a SINE sequence from the library of gummy shark, we used in vitro labeled RNA, prepared as described by Endoh and Okada (1986), as the probe. For isolation of DNA clones with a SINE sequence from the libraries of cat shark and bullhead shark, we used [α-32P] dCTP-labeled DNA that corresponded to nucleotide positions 10–193 of the gummy shark HE1 SINE (fig. 1A) as the probe. For isolation of DNA clones with a LINE sequence from the library of cat shark, we used [α-32P] dCTP-labeled DNA that corresponded to nucleotide positions −1542 to −1449 or −1805 to −1571 of the HER1 LINE as the probe.

**Dot Blot Hybridization**

Genomic DNA isolated from each of 15 species of elasmobranchs from 12 families (marked by asterisks in table 1) and linearized plasmid DNA, used as positive and negative controls, were blotted on membranes. As the probe for detection of members of a SINE family in each genome, we used [γ-32P] ATP-labeled oligonucleotides that corresponded to the nucleotide positions indicated in figure 2. Hybridization was carried out at 37°C for 15 h in a solution of 6 × SSC and 1% SDS, with subsequent washing in a solution of 2 × SSC and 1% SDS at 55.0°C. Hybridization signals were detected by autoradiography. The nucleotide sequences of the oligonucleotides used as probes in these experiments were as follows: probe 1, 5’-TGTGTTCGTTGTTTCCTCC-3’; probe 2, 5’-GACCCACATCCACCCAGGC-3’; and probe 3, 5’-TCGAATCCCSCRGCRTCCTG-3’.

**Genomic DNA Walking**

To determine the 5’ upstream sequence from a breakpoint at the EcoRI site of the HER1 LINE, we employed the method of genomic DNA walking that is known as cassette polymerase chain reactions (PCR; Kajikawa, Ohshima, and Okada 1997). PCR was performed with the following primers, which were specific to the HER1 LINE: SHCR3R (5’-GGGAACCTTAT-CAAATGCCAT-3’) and SHCR2R (5’-ACATCCACGGTTACTTACATC-3’). Fifty nanograms of cat shark genomic DNA which had been completely digested with HindIII were ligated with 3.3 pmol of a HindIII cassette (Takara, Kyoto, Japan). The first PCR was performed with one fifth of the sample as template and oligonucleotides C1 (Takara) and SHCR3 as primers. The second PCR was performed with 1/100 of the products of the first PCR as template and C2 (Takara) and SHCR2 as primers.

**Phylogenetic Analysis**

A phylogenetic tree was constructed and bootstrap values were calculated using programs in the PHYLIP package (Felsenstein 1996).

**Accession Numbers**

The nucleotide sequences of the HE1 SINES and the HER1 LINEs have been deposited in the DDBJ, EMBL, and GenBank nucleotide databases under the accession numbers AB27716–AB27741.

**Results**

**Isolation of a Previously Unrecognized Family of SINES from Higher Elasmobranchs**

To isolate SINE sequences from the genome of Mustelus manazo (gummy shark), we performed in vitro transcription analysis using total genomic DNA from gummy shark as a template. On the basis of the sensitivity of transcription to α-amanitin, RNAs of about 300–800 nt were assumed to be in vitro transcripts generated by pol III (data not shown). We then screened a gummy shark genomic library using these transcripts as probes. We isolated several phase clones and determined the sequences of five of them. Figure 1A shows a sequence alignment of the members of the newly identified SINE family, designated the HE1 SINE family (higher elasmobranch SINE family 1). The HE1 SINE in gummy shark was about 350 bp in length. It included a region that was homologous to a tRNA (underlined in fig. 1A).
FIG. 1.—Alignments of nucleotide sequences of members of the HE1 SINE family. A, HE1 SINEs from the gummy shark. B, HE1 SINEs from the cat shark. C, HE1 SINEs from the bullhead shark. D, HE1 SINEs from the electric ray. The sequence shown on each top line is the consensus sequence (Cons.) of each SINE family. Each tRNA-related region is underlined, and promoters for RNA polymerase III are italicized (Galli, Hofstetter, and Birnstiel 1981). Dots denote identical nucleotides, hyphens indicate gaps introduced to maximize homology, and lowercase letters indicate nucleotides that are different from those in the respective consensus sequence. Nucleotide positions in each consensus sequence are indicated above the sequence.
Elasmobranch SINEs and LINEs

Table 1. The Species of Elasmobranchs Included in This Study

<table>
<thead>
<tr>
<th>Subclass</th>
<th>Superorder</th>
<th>Order</th>
<th>Suborder</th>
<th>Family</th>
<th>Species (common name)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Galeu</td>
<td>Elasmobranchii</td>
<td>Carcharhiniformes</td>
<td>Carcharhinidae</td>
<td>Galeocerdo cuvier (tiger shark)*</td>
<td>Triakididae Musiulus manaza (gummy shark)*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sphyrnidae</td>
<td>Sphyra zygaena (smooth hammerhead)*</td>
<td>Carcharhinus plumbeus (sandbar shark)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Scyliorhinidae</td>
<td>Orectolobidae</td>
<td>Orectolobus japonicus (Japanese wobbegong)*</td>
<td>Sphyraena torazame (cloudy cat shark)*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Centrophoridae</td>
<td>Centrophorus</td>
<td>Heterodontidae Heterodontus japonicus (Japanese bullhead shark)</td>
<td>Dasyatis fuscus (kitefin shark)*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Dalatidiformes</td>
<td>Dalatidae</td>
<td>Squalidae</td>
<td>Squalus mitsukurii (shortspine spurdog)*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Squaliformes</td>
<td>Squalidae</td>
<td>Torpedinidae</td>
<td>Torpedo californica (Pacific electric ray)*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Rajiformes</td>
<td>Rajidae</td>
<td>Rajidae</td>
<td>Raja schmidti (Japanese skate)*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Rajoidae</td>
<td>Rhinobatidae</td>
<td>Rhinobatos schlegelii (brown guitarfish)*</td>
<td>Raja sp.*</td>
</tr>
</tbody>
</table>
Comparison of HE1 SINEs Among Elasmobranchs

As shown in figure 1A and B, the primary sequences of the 5′ and 3′ portions of the tRNA-unrelated region (positions 81–211 and 318–365; fig. 2) of the HE1 SINEs were strongly conserved in the gummy shark and the cat shark, respectively, while the middle of the tRNA-unrelated region (positions 212–317; fig. 2) exhibited considerable divergence. This nature of sequence conservation is also evident in the case of a comparison of the consensus sequences of the HE1 SINEs of the various elasmobranchs (fig. 2).

It was also noteworthy that the 3′-terminal tandemly repeated sequences of the members of the HE1 SINE family were not simple AT-rich repeats of the type usually observed in most SINEs at their 3′ termini (figs. 1D).
Fig. 3.—Distribution of HE1 SINEs among elasmobranchs. Dot hybridization experiments were performed with probes 1, 2, and 3 (as indicated in 2B). Conditions for hybridization and washing are given in Materials and Methods. NT = Not tested.

In particular, the repeated sequences in the HE1 SINEs from the gummy shark and the cat shark were nearly identical to the atypical 3' terminal repeat NATCTAT (GATTCTAT)\(_n\) (where N is any nucleotide) of the chicken and turtle CR1 LINE (Silva and Burch 1989; Kajikawa, Ohshima, and Okada 1997).

The 5' end of the HE1 SINE (positions 1–80) retained a tRNA-like cloverleaf structure (data not shown). However, we were unable to identify the ancestral tRNA species from which the HE1 SINE was derived because of the accumulation of nucleotide substitutions within the sequence of the tRNA during evolution.

Distribution of Members of the HE1 SINE Family in Higher Elasmobranchs

We examined the distribution of members of the HE1 SINE family in higher elasmobranchs. Total genomic DNA from 15 species of elasmobranchs from 11 families (marked by asterisks in table 1) was isolated and subjected to dot blot hybridization analysis. Probe 1, which was derived from the 5' half of the tRNA-unrelated region of the HE1 SINEs (see fig. 2B), hybridized to the DNA from all the elasmobranchs examined (fig. 3, panel 1). Although the copy number of SINEs in the family Rajidae is apparently small, the existence of SINEs with similar sequences were confirmed by PCR analysis (data not shown). These results suggested that the HE1 SINEs might be widespread in higher elasmobranchs and that they might constitute from about 10% to 0.01% of the genome of each contemporary species. Since the fossil record indicates that the hypothetical common ancestor of living elasmobranchs existed in the Lower Jurassic (Shirai 1996), this wide distribution suggests that the age of the HE1 SINE might be about 200 Myr. To examine the distribution of the HE1 SINEs in detail among elasmobranchs, we performed dot blot hybridization analysis with probes that were specific for a particular type of HE1 SINE. Probe 2, which was specific for the HE1 SINEs from the gummy shark (see fig. 2B), hybridized specifically to the DNA from three sharks that belong to the order Chondrichthyes (fig. 3, panel 2). Probe 3, which was specific for the HE1 SINEs from the electric ray (see fig. 2B), hybridized only to the DNA from Narke japonica, which belongs to the family Torpedinidae (as does T. californica; fig. 3, panel 3). Judging from their distribution, it seems likely that the members of the HE1 SINE family have been transmitted vertically from the common ancestor of higher elasmobranchs that might have existed 200 MYA and, moreover, that the HE1 SINEs diverged into several types with the development of the diversity of elasmobranchs.

It is generally accepted that each family of SINEs was created relatively recently on the evolutionary timescale, in view of the fact that the distribution of a certain family of SINEs can be confined to a few species, to a genus, or to a family. As indicated above, the wide distribution of the HE1 SINE suggests that it might be one of the oldest known families of SINEs (200 Myr old). Other examples of ancient families of SINEs are the tortoise Pol III/SINE, the mammalian MIR SINE, and the cichlid AFC SINE. These families of SINEs are estimated to be of about the same age as the HE1 SINE family (Jurka, Zietkiewicz, and Labuda 1995; Smit and Riggs 1995; Ohshima et al. 1996; Terai, Takahashi, and Okada 1998).

Isolation of CR1-like LINEs from the Genomes of Elasmobranchs

In the present study, we also identified the CR1-like LINE family in elasmobranchs. During our efforts...
to isolate SINEs from various elasmobranchs, we found that two clones, TORA 4 (cat shark) and HJ 3 (bullhead shark), contained CR1-like LINE sequences in addition to the sequence of an HE1 SINE. We then screened the GenBank database with the CR1-like LINE sequence of TORA 4 using the BLAST program (Altschul et al. 1990). We identified CR1-like LINE sequences in the genomes of *Carcharhinus plumbeus* (sandbar shark) and *T. californica* (Burch, Davis, and Haas 1993). Figure 4A shows an alignment of the amino acid sequences encoded by members of the CR1-related LINE family, which we designated the HER1 LINE (higher elasmobranch CR1-related LINE 1), together with those of carboxy-terminal regions of ORF2 proteins of the chicken and
turtle CR1 LINE (Haas et al. 1997; Kajikawa, Ohshima, and Okada 1997). The sequences of the HER1 LINEs isolated from the bullhead shark, the sandbar shark, and the electric ray were extensively 5'-truncated, as is usually the case for members of most families of LINEs. The HER1 LINE sequence from the cat shark encoded the amino acids of domains III–VII of the RTase domain that is typical of LINEs (see fig. 5).

Figure 4B shows an alignment of the 3'-UTR of the HER1 LINEs. The 3'-end tail region of the 3'-UTR of the HER1 LINE was strongly conserved among the HER1 LINEs, while the rest of the 3'-UTR exhibited considerable divergence. Since the 3'-end region of a LINE transcript is thought to serve as a recognition site for the RTase encoded by the LINE (Luan and Eickbush 1995; Moran et al. 1996), the strongly conserved tail region of the HER1 LINE might be critical for retroposition of the HER1 LINE.

The HER1 LINEs ended in atypical tandemly repeated sequences that were nearly identical to the atypical 3'-terminal repeat NATTCTAT(GATTCTAT)ₙ of the chicken and turtle CR1 LINE. This observation suggests a close phylogenetic relationship between the HER1 LINE and the CR1 LINE.

Further Characterization of the HER1 LINE

To characterize the HER1 LINE in further detail, we screened a cat shark genomic library using a fragment of TORA 4 (positions -1542 to -1449; fig. 5B) as the probe, and we determined the sequences of five phage clones. All clones isolated, including TORA 4, were found to contain the 3' terminus of the HER1 LINE, and none included any useful information in the region, upstream in the 5' direction, from the breakpoint at the EcoRI site (TORA 4 and TORAL 1–5; fig. 5A).

To extend the sequence in the 5' direction, we employed the method known as PCR genome walking (Kajikawa, Ohshima, and Okada 1997) and obtained two HER1 elements (TORAL 6 and 7; fig. 5A). Finally, using the clone designated TORAL 6 (positions -1805 to -1571; fig. 5B) as the probe, we isolated four phage clones and determined their sequences (TORAL 8 to TORAL 11; fig. 5A). As shown in figure 5, the HER1 LINE from the cat shark included an ORF that encoded an RTase domain (Xiong and Eickbush 1990) and an ENase domain (Martín et al. 1995, 1996; Feng et al. 1996), as do members of most of the other LINE families reported to date.

Phylogenetic Analysis of the HER1 LINE

To examine the phylogenetic relationships of the HER1 LINE within CR1-related families of LINEs, we constructed a phylogenetic tree of LINEs based on the amino acid sequences of the RTase domains encoded by the LINEs (fig. 5C). According to the phylogenetic tree, the HER1 LINE appeared to be closely related to the chicken and reptile CR1 LINE. This result was consistent with our observations that (1) the HER1 LINE encoded an amino acid sequence with strong homology to that of the carboxy-terminal region of the ORF2 protein of the CR1 LINE, and (2) the HER1 LINE contained a tandemly repeated sequence nearly identical to that in the CR1 LINE (see fig. 4).

Our phylogenetic tree also showed that the phylogeny of vertebrate CR1-related LINEs, such as the chicken and reptile CR1 LINE and the frog X1CR1 LINE, was not congruent with that of the host species. This observation suggests two possibilities: It is possible that CR1-related LINEs might have been transmitted horizontally among vertebrates. Alternatively, several types of CR1-related LINEs might have been present in the genome of a common ancestor of vertebrates, and we analyzed in this study molecular species of LINEs that are not orthologous to one another. Since only a few sequences of vertebrate CR1-related LINEs are available at the present time, there are no strong arguments that favor one possibility over the other. Further analysis of CR1-related LINEs from additional vertebrates is clearly required to resolve this issue.

Discussion

Retropositional Parasitism of SINEs on LINEs with Identical Tail Regions

Several examples of the phenomenon whereby the 3'-end regions of certain SINEs are identical to those of certain LINEs were identified recently (see Introduction; for review, see Okada et al. 1997). Here, we added another example of this category. Figures 4B and C show comparisons of nucleotide sequences and secondary structures between the tail regions of the HE1 SINEs and those of the HER1 LINEs. Three observations clearly indicate that the HE1 SINE and the HER1 LINE constitute another pair of SINEs and LINEs that share the same tail region. First, the sequences of the tail region of the HE1 SINE and the tail region of the HER1 LINE were almost identical (fig. 4B). Second, the secondary structures of the tail regions of the HE1 SINEs and those of the HER1 LINEs were very similar (fig. 4C). Compensatory nucleotide changes, namely from 5'-TG-CA-3' to 5'-CA-TG-3' (or 5'-CA-TG-3' to 5'-TG-CA-3') in the tail region, provide strong evidence for the validity of our proposal (marked by asterisks in fig. 4B and by boxes in fig. 4C). Finally, as was the case for the HE1 SINEs of the gummy shark and the cat shark, the HER1 LINEs ended in atypical 3'-terminal tandem repeats that were nearly identical to the 3'-terminal repeats that are characteristic of the chicken and turtle CR1 LINE (fig. 4B; Silva and Burch 1989; Kajikawa, Ohshima, and Okada 1997).

According to the current model for the molecular mechanism of retroposition of LINEs, the 3'-end region of a LINE transcript serves as a recognition site for the RTase encoded by the LINE during retroposition (Luan and Eickbush 1995; Moran et al. 1996). The presence of the highly conserved motif in the 3'-end regions of the elasmobranch SINEs and LINEs suggests that the RTase encoded by the HER1 LINE recognizes the 3'-tail regions of transcripts of the HER1 LINEs, as well as those of the HE1 SINEs, during retroposition. Moreover, we showed simultaneous conservation of both the tail region of the HE1 SINE and that of the HER1 LINE.
in each elasmobranch. The HE1 SINE and the HER1 LINE provided the clearest and most convincing example among the pairs of SINEs and LINEs proposed to date in terms of conservation of primary sequences and secondary structures, and they add strong support to our model of the way in which nonautonomous SINEs might have recruited the enzymatic machinery for retroposition during evolution.

There May Be Two Types of LINE: a Relaxed Type and a Stringent Type

As noted in the Introduction, Jurka’s group analyzed the sequence of the target site duplications of several mammalian SINE families, such as the Alu, ID, B1, and B2 SINEs, proposing that these SINEs might utilize the enzymatic machinery that is used for retroposition of the mammalian L1 LINE (Jurka and Kloneowski 1996; Jurka 1997, 1998). For retroposition of L1, the 3’-terminal poly(A) tract appears to be critical, whereas the 3’ UTR appears not to be (Holmes et al. 1994; Moran et al. 1996). It seems likely that in retroposition of these various SINEs the L1 RTase recognizes the 3’-terminal AT-rich sequences of the transcripts of the SINEs (Boeke 1997). It is also likely that the L1 RTase is involved in formation of retropseudogenes via its recognition of the poly(A) tracts of mRNAs. Therefore, we propose that the L1 be designated the “relaxed” type of LINE that can drive the retroposition of a large variety of templates (Okada et al. 1997).

The families of LINEs that have 3’-tail regions identical to those of the respective families of SINEs appear to require the 3’-tail region during retroposition.
We propose to designate such LINEs the “stringent” type of LINE (Okada et al. 1997). It will be of interest to determine the extent of distribution of each type among LINEs.

Possible Mechanisms for the Generation of the 3′-Terminal Tandem Repeats

The HE1 SINEs and HER1 LINEs end in an unusual tandem repeat, whereas most SINEs and LINEs have simple AT-rich repeats at their 3′ termini (see fig. 4B). Since the tandem repeats of some of the HE1 SINEs and those of HER1 LINEs are almost identical to those of CR1-related LINEs, these findings provide convincing evidence that the HE1 SINE and the HER1 LINE underwent retroposition that was engineered by machinery similar to that exploited by the CR1-related LINEs.
There are, however, several examples of pairs of SINEs and LINEs that probably use the same RTase for their retroposition but end in different tandem repeats. Our findings suggest that LINEs might have the flexibility to generate various 3'-terminal repeats in target primed reverse transcription.

We shall discuss here two possible explanations for the generation of the tandem repeats at the 3' termini of non-LTR retroposons. One explanation is as follows. Since the RTases encoded by LINEs are closely related, both phylogenetically and mechanistically, to the telomere RTase (TERT; Eickbush 1997; Nakamura et al. 1997), it is possible that the mechanism for terminal addition of repeating units that is operative in telomeres and the mechanism that operates at the 3' termini of LINEs might have the same origin. In the reaction catalyzed by telomerase, a ribonucleoprotein complex containing telomere RNA functions to create telomeric tandem repeats by using the RNA as template (Greider and Blackburn 1989). Similarly, a ribonucleoprotein consisting of the RTase encoded by a LINE and an RNA template might gain access to the 3'-OH group at the target site of nicked DNA, and then this RTase might discontinuously add a few repeating units, using the 3'-terminal repeating sequence as template in the initial step of reverse transcription. The differences among the repeated sequences in members of the same family of retroposons can be readily explained by postulating the presence of several mutations in the 3'-terminal repeat sequences of the source genes of the retroposons.

The second possibility involves “nontemplate addition,” a mechanism proposed by Luan and Eickbush (1995, 1996), who provided evidence that the R2Bm RTase has the ability to add extra “nontemplated” nucleotides to the target site before the enzyme begins reverse transcription with various R2Bm RNAs as templates. However, the added nucleotide sequences vary according to the extreme 3'-end sequence of the R2 RNA template. Therefore, it is possible, in general, that the sequences of the tail regions of retroposons or their secondary structures might affect the nucleotides added to the DNA at target sites.

The HER1 LINE of the Cat Shark Contains Two Tail Regions in its 3' UTR

It is noteworthy that a unit of the HE1 SINE sequence was located within the 3' UTR of the cat shark HER1 LINE (positions −390 to −50; fig. 5A and B). Moreover, all members of the cat shark HER1 LINE family that we isolated and characterized contained a SINE unit within the 3' UTR (fig. 5A and B; see TORA 4 in fig. 1B). This observation suggests the presence of a master source gene for the SINE-containing HER1 LINE that has the ability to be actively retroposed. Moreover, these observations suggest the existence of a mechanism for the preferential choice of the extreme 3'-tail region from among the two possible tail regions as the recognition site for the RTase that is encoded by the HER1 LINE. Since each unit of the HER1 LINE apparently lacks a poly(A) signal, RNA polymerase II might read through a LINE unit, adding some nucleotides in the 3'-flanking region to its transcript. It is possible that the 3' termini of a template RNA with a redundant sequence might bind the HER1 RTase and that the RTase might scan the template in the 5' direction until it recognizes the tail region and initiates reverse transcription.

Our hypothesis calls to mind the detection by Holmes et al. (1994) of a readthrough transcript of an L1 in the human genome. The existence of this transcript, which contained a unique sequence between the poly(A) tract at the 3' end of the L1 element and the poly(A) tract at the true 3' terminus, suggests that a mechanism might exist for recognition of the 3' terminus of the RNA template, rather than of an internal poly(A) tract, as the site for initiation of reverse transcription. At present, however, the mechanism by which transcription of LINEs is terminated and the way in which the initiation site is recognized by the RTase remain largely unknown. Detailed characterization of the HE1 SINEs and HER1 LINEs should provide clues to the general mechanism of the termination of transcription and the initiation of the reverse transcription of non-LTR retroposons.

The Ages of SINEs

Two of the oldest known families of SINEs, the HE1 SINE and the tortoise Pol III/SINE, are derived from two CR1-related LINEs, namely, the HER1 LINE and the CR1 LINE, respectively (see fig. 5C). This finding suggests that LINE families within the same lineage of a LINE phylogeny created their own respective SINE families independently in the genomes of different hosts. If the LINE families within a given group are considered siblings that have the same parent (namely, a common ancestral LINE of these LINEs), the SINE families descended from such LINE families can be considered cousins. Therefore, we refer to such SINE families as “SINE cousins” (Terai, Takahashi, and Okada 1998). According to this definition, the elasmobranch HE1 SINE and tortoise Pol III/SINE are SINE cousins, as in the case of the cichlid AFC SINE and the mammalian MIR SINE (Terai, Takahashi, and Okada 1998).

Our findings raise several questions about the evolution of SINEs and LINEs. Did the HER1 LINE and the CR1 LINE acquire the ability to provide their 3'-end tail regions for the creation and retroposition of new families of SINEs recently on the evolutionary timescale (during the past 200 Myr)? Alternatively, did the common ancestor of these LINE families have this ability (about 2,000 Myr ago, perhaps)? Although it appears more likely that this ability represents an intrinsic prop-
property of most LINEs, it is necessary to examine many LINEs from vertebrates and nonvertebrates more extensively from a phylogenetic perspective. If a LINE were to have existed with the ability to create a partner SINE in the remote past, it is possible that a large number of members of an ancient SINE family might have been created during evolution. Such SINES might have been buried in genomes because of the accumulation of mutations and would thus no longer be recognizable as SINES. We should be able to determine whether a considerable fraction of the genomes of multicellular organisms is composed of such debris of SINES and LINEs if an extensive search for SINES and LINEs is performed in conjunction with the complete sequencing of the genomes of such organisms in the future.

Acknowledgment

This work was supported by a Grant-in-Aid for Specially Promoted Research from the Ministry of Education, Science, Sports and Culture of Japan.

LITERATURE CITED


Schwarz-Sommer, Z., L. Leclercq, E. Göbel, and H. Sae


Terai, Y., K. Takahashi, and N. Okada. 1998. SINE cousins: the 3’-end tails of the two oldest and distantly related families of SINES are descended from the 3’ ends of LINEs with the same genealogical origin. Mol. Biol. Evol. 15:1460–1471.


Pierre Capy, reviewing editor

Accepted June 3, 1999