# Short Retroposons in Eukaryotic Genomes

Dimitri A. Kramerov and Nikita S. Vassetzky
Engelhardt Institute of Molecular Biology, Russian Academy of Sciences
Moscow 119991, Russia

Short retroposons (SINEs) are repetitive elements amplified in the genome via an RNA intermediate, using the enzymatic machinery of autonomous retroposons (LINEs). SINEs are widely distributed in eukaryotes; for instance, all tested mammalian genomes contain $10^4$–$10^6$ SINE copies. Although several SINE families such as primate Alu or rodent B1 have long been recognized, the more recent discovery of many SINEs in various eukaryotes, as well as progress in understanding the mechanisms of LINE replication and genome functioning as a whole, shed light on the biology and evolution of SINEs and their significance for the cell.

**KEY WORDS:** Repetitive elements, Mobile elements, Transposons, Transposable elements, Retroposons, SINEs, RNA polymerase III. 

## I. Introduction

The genomes of higher eukaryotes are crowded with DNA elements that are repeated thousands or millions times. For instance, they amount to ~40–45% of human and mouse genomes. Many of these repeats were generated through the activity of transposable elements or transposons that can insert their copies into new chromosomal locations. Transposons are divided into two classes according to whether their replication is mediated by RNA (class I) or DNA (class II). Both classes include autonomous and nonautonomous elements. Autonomous transposons have open reading frames (ORFs) encoding proteins essential for transposition, whereas nonautonomous transposons encode no proteins and rely on the replication machinery of the autonomous transposons. Integration of nearly all transposons results in



0074-7696/05 $35.00
DOI: 10.1016/S0074-7696(05)47004-7

duplication of a short genomic sequence at the insertion point (target site duplication).

Class I elements (or retroelements) are further divided into three groups by their replication strategy and structure: LTR-transposons (or retrotransposons), long interspersed elements (LINEs), and short interspersed elements (SINEs). LTR-transposons resemble retroviruses: they have long terminal repeats (LTRs) with transcription control sequences and open reading frames encoding retropositional activities but lack those for forming a viral capsid.

LINEs (also called long retroposons) encode similar activities but lack LTRs; instead, they carry a promoter sequence preceding the open reading frames and mediating synthesis of polyadenylated RNA by RNA polymerase II, and are usually terminated by poly(A).

SINEs (short retroposons) are nonautonomous retroposons; the internal promoter at the 5′ end provides for their transcription by RNA polymerase III (Pol III). Similar to LINEs, they end with A-rich or other simple repeat sequence.

Nonautonomous elements of DNA transposons, LTR-transposons, and LINEs are usually derived from the corresponding autonomous elements via internal deletion; the origin of SINEs is not so straightforward. Their 5′ part descends from a cellular RNA (usually tRNA); the origin of the 3′ part is unclear, although the extreme 3′ part of many (but not all) SINEs is similar to the 3′ end of conspecific LINEs.

Although SINEs are only one of many types of transposons, the number of their copies makes up half of the total number of repeated elements (at least in humans and mice). In this article, we review the available data on the structure, replication, evolution, and biological significance of SINEs.

SINEs were discovered more than 25 years ago with the cloning (Jelinek *et al.*, 1980; Kramerov *et al.*, 1979) and sequencing of rodent B1 and B2 elements (Haynes and Jelinek, 1981; Haynes *et al.*, 1981; Krayev *et al.*, 1980, 1982) as well as primate Alu elements (Daniels and Deininger, 1983; Deininger *et al.*, 1981). Although SINEs were initially believed to occur only in mammals, later studies demonstrated their presence in the genomes of reptiles, fish, ascidians, insects, and flowering plants (Table I). Progress in LINE research demonstrated that their proteins execute retroposition. The finding of similar nucleotide sequences at the 3′ ends of nonmammalian SINEs and LINEs (Okada and Hamada, 1997) was the starting point for the concept of involvement of LINE machinery in the retroposition of SINEs (Jurka, 1997; Kajikawa and Okada, 2002). For many years, SINEs were considered to be selfish or parasitic DNA; however, the available data indicate that SINEs can mediate certain cellular processes. Moreover, they play an important role in the evolution of individual genomic loci and genome as a whole.

TABLE I
Eukaryotic SINEs[a]

| Classification | SINE | Ancestral RNA | Structure | Length (nt) | Associated LINE | Tail | Species range | Number of copies | References[b] |
|---|---|---|---|---|---|---|---|---|---|
| | | | | **Animals** | | | | | |
| **Phylum Vertebrata** | | | | | | | | | |
| | Ther-1 (MIR) | tRNA | Monomeric; CORE-SINE | 270 | L2 | $(TTA)_n$ | Mammals, birds, and reptiles | $4 \times 10^5$ (human) $1 \times 10^5$ (mouse, rat) | 1, 2 |
| Class Mammalia | | | | | | | | | |
| Subclass Theria | Ther-2 (MIR3) | tRNA | Monomeric; CORE-SINE | 220 | L3 | | Marsupials and placentals | $7.5 \times 10^4$ (human) | (2, 3) |
| Subclass Metatheria | Mar-1 | tRNA | Monomeric; CORE-SINE | 240 | Bov-B | $(AAC)_n$ | Marsupials | | 2 |
| | Opo-1 | tRNA | Monomeric; CORE-SINE | 190 | | | North American marsupials | | 2 |
| Subclass Prototheria | Mon-1 | tRNA | Monomeric; CORE-SINE | 270 | L2 | $(TTA)_n$ | Monotremes | | 2 |
| Order Primates | Alu | 7SL RNA | Homodimeric | 282 | L1 | A-rich | Primates | $1.1 \times 10^6$ (human); $1.5 \times 10^5$ (galago) | 4, 5 |
| | SINE type II | $tRNA^{Ile}$ + 7SL RNA | Dimeric | 260 | L1 | A-rich | Bush babies and lorises (Lorisiformes) | $3.2 \times 10^5$ | 6, 7 |
| | SINE type III | $tRNA^{Ile}$ | Monomeric | 100 | L1 | A-rich | Bush babies and lorises (Lorisiformes) | $2.0 \times 10^5$ | 7, 8 |
| Order Rodentia | B1 | 7SL RNA | Monomeric | 135 | L1 | A-rich | Rodents | $5.6 \times 10^5$ (mouse) | 9, 10 |
| | ID | $tRNA^{Ala}$ | Monomeric | 75 | L1 | A-rich | Rodents | $10^3$–$10^5$ | 11 |
| | B4 (RSINE2) | $tRNA^{Ala}$ + 7SL RNA | Dimeric ID + B1 | 275 | L1 | $(CA)_n$ | Mouse and rat | $4 \times 10^5$ (mouse) $3.6 \times 10^5$ (rat) | 12 |

(*continued*)

TABLE I (*continued*)

| Classification | SINE | Ancestral RNA | Structure | Length (nt) | Associated LINE | Tail | Species range | Number of copies | References[b] |
|---|---|---|---|---|---|---|---|---|---|
| | B2 | $tRNA^{Ala}$ | Monomeric | 185 | L1 | A-rich | Muridae, Cricetidae, Spalačidae, and Rhizomyidae | $3.5 \times 10^5$ | 13 |
| | DIP | $tRNA^{Ala}$ | Monomeric | 190 | L1 | A-rich | Dipodidae and Zapodidae | $10^5$ | 14 |
| | MEN | $tRNA^{Ala}$ + 7SL RNA | Dimeric tRNA + B1 | 259 | L1 | A-rich | Squirrels (*Menetes and Callosciurus*) | $10^5$ | 14 |
| | B1-dID | 7SL RNA + $tRN A^{Ala}$ | Dimeric B1 + ID | 200 | L1 | A-rich | Sciuridae and Gliridae | $10^5$ | 15 |
| | IDL-Geo | $tRNA^{Ala}$ + tRNA | Dimeric | 192 | L1 | A-rich | Geomyidae and Heteromyidae | $10^5$ | 16 |
| Order Dermoptera | CYN (t-SINE) | $tRNA^{Ile}$ | Mono-, di-, and trimeric | 90, 160, 220 | L1 | A-rich | Dermoptera (*Cynocephalus variegatus*) | $1.5 \times 10^5$ | 17, 18 |
| Order Lagomorpha | C | $tRNA^{Gly}$ | Monomeric | 309 | L1 | A-rich | Rabbit (*Oryctolagus cuniculus*) | $1.7 \times 10^5$ | 19 |
| Order Cetartiodactyla | Bov-tA | $tRNA^{Gly}$ | Dimeric tRNA + Bov-A | 210 | Bov-B | AT-rich | Bovidae (cattle, goats, and sheep) | $2 \times 10^5$ | 20 |
| | CHRS (CHR-1; CHRS-S) | $tRNA^{Glu}$ | Monomeric | 120–160 | L1 | A-rich | Cetaceans, hippopotamuses, ruminants, and suiforms | $5 \times 10^4$ (sperm whale) $2 \times 10^5$ (pig) | 21 |
| | CHR-2 | $tRNA^{Glu}$ | Monomeric | 270–330 | L1 | A-rich | Cetaceans, hippopotamuses, and ruminants | $2 \times 10^5$ (sperm whale) | 22 |
| | PRE-1 | $tRNA^{Arg}$ | Monomeric | 246 | L1 | A-rich | Pigs and peccaries | $10^6$ (cow and pig) | 23 |
| | Vic-1 | $tRNA^{Ala}$ | Monomeric | 117 | L1 | A-rich | Camelidae | $10^5$ | 24 |
| Order Perissod actyla | ERE-1 | $tRNA^{Ser}$ | Monomeric | 212 | L1 | A-rich | Horses (*Equus* spp.) | $5 \times 10^4$ | 25 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Order Chiroptera | VES | $tRNA^{Tyr}$ | Monomeric | 190 | L1 | A-rich | Bat families Vespertilionidae, Molossidae, Phyllostomidae, and Emballonuridae | $10^5$ | 26 |
| | Rhin-1 | $tRNA^{Ile}$ | Monomeric | 190 | L1 | A-rich | Bat families Rhinolophidae and Hipposideridae | | 27 |
| Order Insectivora | SOR | $tRNA^{Lys}$ | Monomeric | 157 | L1 | A-rich | Shrews (Soricidae) | $10^5$ | 28 |
| | TAL | tRNA | Monomeric | 237 | L1 | A-rich | Moles (Talpidae) | $10^5$ | 28 |
| | ERI-1 | $tRNA^{Lys}$ | Monomeric | 126 | L1 | A-rich | Hedgehogs (Erinaceidae) | $10^5$ | 28 |
| | ERI-2 | tRNA | Monomeric | 186 | L1 | A-rich | Hedgehogs (Erinaceidae) | $10^5$ | 28 |
| Order Carnivora | CAN | tRNA | Monomeric | 160 | L1 | A-rich | Carnivores | $2 \times 10^5$ | 29, 30 |
| Order Scandentia | Tu type I | tRNA + 7SL RNA | Dimeric tRNA + B1 | 190 | L1 | A-rich | Tree shrews | $10^2$ | 31 |
| | Tu type II | tRNA + 7SL RNA | Trimeric tRNA + B1 + B1 | 290 | L1 | A-rich | Tree shrews | $10^2$ | 31 |
| | Tu type III | tRNA | Monomeric | 260 | L1 | A-rich | Tree shrews | $2 \times 10^5$ | 31 |
| Orders Hyracoidea, Sirenia, Proboscidea, Tubulidentata, Macroscelidea, Insectivora | AfroSINE | tRNA | Monomeric | 230 | | $(TTG)_n$ | Afrotherians | $8 \times 10^5$ | 32 |
| Order Xenarthra | DAS-I | $tRNA^{Ala}$ | Monomeric | 90 | L1 | A-rich | Nine-banded armadillo (*Dasypus novemcinctus*) | $2.9 \times 10^4$ | 27, 33 |
| | DAS-II | $tRNA^{Ala}$ | Homodimeric | 190 | L1 | A-rich | Nine-banded armadillo (*Dasypus novemcinctus*) | $6.5 \times 10^4$ | 33 |
| | DAS-III | $tRNA^{Ala}$ | Monomeric | 440 | L1 | A-rich | Nine-banded armadillo (*Dasypus novemcinctus*) | $2.2 \times 10^5$ | 33 |

(*continued*)

TABLE I (*continued*)

| Classification | SINE | Ancestral RNA | Structure | Length (nt) | Associated LINE | Tail | Species range | Number of copies | References[b] |
|---|---|---|---|---|---|---|---|---|---|
| Class Reptilia | Cry I/Cry II (pol III/SINE) | $tRNA^{Lys}$ | Monomeric | 200 | PsCR1 | AT-rich | Tortoises | | 34, 35 |
| | P.s. 1/SINE | $tRNA^{Lys}$ | Monomeric | 116 | Lucy 1 (CR1) | AT-rich | Lizard (*Podarcis sicula*) | | 36 |
| Class Actinopterygii (ray-finned fishes) | Sma I (Sma I-div, Sma I-cor, Hirt, Pol III/SINE) | $tRNA^{Lys}$ | Monomeric | 150 | RSg-1 | AT-rich | Salmonidae | $2.6 \times 10^4$ (chum and pink salmons) | 37 |
| | Fok I | $tRNA^{Lys}$ | Monomeric | 150 | RSg-1 | AT-rich | Chars (*Salvelinus* spp.) | | 38 |
| | Hpa I | tRNA | Monomeric; CORE-SINE | 200 | Rsg-1 | AT-rich | Salmonidae | $1–2 \times 10^4$ (Salmoninae) | 38 |
| | Ava III | tRNA | Monomeric; CORE-SINE | 280 | | AT-rich | Salmonidae | $\sim 10^2$ | 39 |
| | AFC | tRNA | Monomeric; CORE-SINE | 320 | CiLINE2 | $(ATT)_n$ | Cichlids | $2 \times 10^3–2 \times 10^4$ | 40 |
| | DANA | $tRNA^{Val}$ | Monomeric; V-SINE | 370 | ZfL3 | $(TGAA)_n$ | Zebrafish (*Danio*) | $4–5 \times 10^5$ | 41 |
| | SINE3 | 5S rRNA | Monomeric | 560 | CR1-like | $(ACATT)_n$; $(ATT)_n$ | Zebrafish (*Danio rerio*) | $10^4$ (0.4% of genome) | 42 |
| | Ras1 | $tRNA^{Val}$ | Monomeric; V-SINE | 300 | | $(YAAA)_n$ | Rasbora (*Rasbora pauciperforata*) | $2 \times 10^2$ | 43 |
| | AC1 | $tRNA^{Val}$ | Monomeric; V-SINE | 380 | | $(TGAG)_n$ | Fugu (*Fugu rubripes*) | $2 \times 10^2$ | 43 |
| | | | | | | $(TGAA)_n$ | Medaka (*Oryzias latipes*) | $2.8 \times 10^3$ | 43 |
| | | | | | | | Other Percomorpha | | 43 |
| | UnaSINE1 | tRNA | Monomeric | 300 | UnaL2 | $(TGTAA)_n$ | Eel (*Anguilla japonica*) | | 44 |
| Class Chondrichthyes (cartilaginous fishes) | HE1 | $tRNA^{Val}$ | Monomeric; V-SINE | 340 | HER1 | $(NATTCTAT)_n$ | Higher elasmobranchs | $10^3–10^6$ | 45 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Class Dipnoi (lungfishes) | Lun1 | $tRNA^{Val}$ | Monomeric; V-SINE | 300 | LfR1 | $(GAAC-CTAT)_n$ | Lungfish (*Lepidosiren paradoxa*) | $3 \times 10^5$ | 43 |
| Class Petromyzontidae (lampreys) | Lam1 | $tRNA^{Val}$ | Monomeric; V-SINE | 230 | | $(GCA)_n$ | Lamprey (*Lethenteron reissneri*) | $2.4 \times 10^3$ | 43 |
| **Phylum Tunicata** | | | | | | | | | |
| Class Ascidiacea (sea squirts) | Cisc-1 | tRNA | Monomeric; CORE-SINE | 290 | | $(CATT)_n$ | Sea squirt (*Ciona intestinalis*) | | 46 |
| **Phylum Mollusca** | | | | | | | | | |
| Class Octopoda | SK | $tRNA^{Lys}$ | Monomeric | 260 | | AT-rich | Squid (*Loligo bleekeri*) | | 47 |
| | OK | tRNA | Monomeric | 410 | | AT-rich | *Octopus vulgaris* | $8.5 \times 10^4$ | 48 |
| | OR1 | $tRNA^{Arg}$ | Monomeric; CORE-SINE | 390 | | AT-rich | *Octopus* spp. | $3.3 \times 10^4$ | 48 |
| | OR2 | $tRNA^{Arg}$ | Monomeric; CORE-SINE | 270 | | AT-rich | *Octopus* spp. | $2.8 \times 10^4$ | 48 |
| **Phylum Platyhelminthes** | | | | | | | | | |
| Class Trematoda | Smα (T2), Sjα | $tRNA^{Arg}$ | Monomeric | 330 | | AT-rich | Schistosomatidae | | 49–51 |
| **Phylum Echinodermata** | | | | | | | | | |
| Class Echinoidea | SURF1 | tRNA | Monomeric | ~330 | | AT-rich | Sea urchin (*Strongylocentrotus purpuratus*) | $8 \times 10^2$ | 52 |
| **Phylum Arthropoda** | | | | | | | | | |
| Class Insecta | | | | | | | | | |
| Order Lepidoptera | Bm1 | tRNA | Monomeric | 430 | | A-rich | Silkworm (*Bombyx mori*) | $2 \times 10^4$ | 53 |
| Order Orthoptera | Lm1 (SGRP1) | $tRNA^{Leu}$ | Monomeric | ~200 | | T-rich | African migratory locust (*Locusta migratoria*) *Schistocerca Chorthippus*) | $6 \times 10^5$ (locust) | 54 |

(*continued*)

TABLE I (*continued*)

| Classification | SINE | Ancestral RNA | Structure | Length (nt) | Associated LINE | Tail | Species range | Number of copies | References[b] |
|---|---|---|---|---|---|---|---|---|---|
| Order Diptera | Feilai | $tRNA^{Ser}$ | Monomeric | 275 | Juan-A | $(GAA)_n$ | Yellow fever mosquito (*Aedes aegypti*) | $6 \times 10^4$ | 55 |
| | Sine200 | Type 2 Pol III promoter (not tRNA) | Monomeric | 210 | | $(GAA)_n$ | *Anopheles gambiae* | $2 \times 10^3$ | 56 |
| | Twin | $tRNA^{Arg}$ | Homodimeric | 250 | | A-rich | Mosquito (*Culex pipiens*) | $5 \times 10^2$ | 57 |
| **Phylum Nematoda** | | | | | | | | | |
| Subclass Rhabditia | CELE45 | tRNA (Lys) | Dimeric | 270 | RTE1 | AT-rich | *Caenorhabditis elegans* | Several hundreds | 3; and our unpublished data |
| | | | | **Plants** | | | | | |
| **Phylum Magnoliophyta (dicoteledons)** | | | | | | | | | |
| Order Poales | Au | tRNA | Monomeric | 180 | | T-rich | Poaceae (grass family) Fabaceae (pea family) Solanaceae (nightshade family) | | 58; and our unpublished data |
| | p-SINE1 | $tRNA^{Gly}$ (Ser) | Monomeric | 130 | | T-rich | Rice (*Oryza* genus) | $6 \times 10^3$ | 59, 40 |
| | F524 | tRNA | Monomeric | 290 | | A-rich | Rice (*Oryza sativa*) | | 3 |
| | CASINE | tRNA | Monomeric | 190 | | A-rich | Rice (*Oryza sativa*) | | 3; and our unpublished data |
| Order Solanales | TS | $tRNA^{Lys}$ (Arg) | Monomeric | 210 | | T-rich | Solanaceae and Convolvulaceae | $5 \times 10^4$ (tobacco) | 61 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Order Brassicales | S1 | $tRNA^{Pro}$ (Pro) | Monomeric | 180 | Bali1 | A-rich | Brassicaceae (mustard family) | $\sim 10^3$ | 62 |
| | RAthE1 (AtSN2, AtSINE3, SL1) | $tRNA^{Cys}$ (Lys) | Monomeric | 150 | | A-rich | *Arabidopsis thaliana* | $1.5 \times 10^2$ | 63, 64 |
| | RAthE2 (SL2) | $tRNA^{Gly}$ | Monomeric | 310 | | A-rich | *Arabidopsis thaliana* | 60 | 63 |
| | RBolE2 | tRNA | Monomeric | 300 | | A-rich | *Brassica* | | Our unpublished data |
| | AtSINE1 (AtSN1) | tRNA | Monomeric | 170 | AtLINE1-3A | A-rich | *Arabidopsis thaliana* | | 64, 65 |

*Abbreviations:* LINE, long interspersed element (autonomous retroposon); SINE, short interspersed element (short retroposon).

[a]*Note:* Many repeats described as SINEs do not belong to this class of repetitive elements: Bov-A2, ARE (cetartiodactyls), ELA (carnivores), Hy/Pol III (reptiles), RANA/Pol III (amphibians), D88I (echinoderms), Bm1b, Cp1 (insects), ZmSINE1, XC1, XC2, TSCL (plants), MGSR1, Mg-SINE, Ch-SINE (fungi), EHINV1/2, EhLSINE1, EhLSINE2, and *Scal* (protists).

[b]*References:* (1) Smit and Riggs, 1995; (2) Gilbert and Labuda, 2000; (3) Jurka, 2000; (4) Deininger *et al*., 1981; (5) International Human Genome Sequencing Consortium, 2001; (6) Daniels and Deininger, 1983; (7) Roos *et al*., 2004; (8) Daniels and Deininger, 1991; (9) Krayev *et al*., 1980; (10) Vassetzky *et al*., 2003; (11) Milner *et al*., 1984; (12) Lee *et al*., 1998; (13) Krayev *et al*., 1982; (14) Serdobova and Kramerov, 1998; (15) Kramerov and Vassetzky, 2001; (16) I. K. Gogolevsky and D. A. Kramerov, unpublished data; (17) Piskurek *et al*., 2003; (18) Schmitz and Zischler, 2003; (19) Cheng *et al*., 1984; (20) Lenstra *et al*., 1993; (21) Shimamura *et al*., 1999; (22) Nikaido *et al*., 2001; (23) Singer *et al*., 1987; (24) Lin *et al*., 2001; (25) Sakagami *et al*., 1994; (26) Borodulina and Kramerov, 1999; (27) Borodulina and Kramerov, 2005; (28) Borodulina and Kramerov, 2001; (29) Lavrent'eva *et al*., 1989; (30) Vassetzky and Kramerov, 2002; (31) Nishihara *et al*., 2002; (32) Nikaido *et al*., 2003; (33) Churakov *et al*., 2005; (34) Endoh and Okada, 1986; (35) Sasaki *et al*., 2004; (36) Fantaccione *et al*., 2004; (37) Hamada *et al*., 1997; (38) Kido *et al*., 1991; (39) Kido *et al*., 1994; (40) Takahashi *et al*., 1998; (41) Izsvak *et al*., 1996; (42) Kapitonov and Jurka, 2003; (43) Ogiwara *et al*., 2002; (44) Kajikawa and Okada, 2002; (45) Ogiwara *et al*., 1999; (46) Simmen and Bird, 2000; (47) Ohshima *et al*., 1993; (48) Ohshima and Okada, 1994; (49) Spotila *et al*., 1989; (50) Ferbeyre *et al*., 1998; (51) Laha *et al*., 2000; (52) Nisson *et al*., 1988; (53) Adams *et al*., 1986; (54) Bradfield *et al*., 1985; (55) Tu, 1999; (56) Holt *et al*., 2002; (57) Feschotte *et al*., 2001; (58) Y. Yasuo, N. Shuhei, and Y. Matsuoka, unpublished data; (59) Mochizuki *et al*., 1992; (60) Motohashi *et al*., 1997; (61) Yoshioka *et al*., 1993; (62) Deragon *et al*., 1994; (63) Lenoir *et al*., 2001; (64) Myouga *et al*., 2001; (65) Kapitonov and Jurka, 1999.

Although excellent reviews of various aspects of SINEs regularly appear (Deininger and Batzer, 2002; Smit, 1999; Weiner, 2002), no fundamental reviews have been published more recently (Maraia, 1995; Rogers, 1985; Weiner *et al*., 1986). This article is an attempt to fill this gap.

## II. Structure of SINEs

As can be deduced from their name, SINEs are short relative to other transposons: 80–500 bp (typically 150–200 bp). In the majority of cases, SINEs are flanked by short (5–15 nucleotides) direct repeats of host DNA. A typical short retroposon is composed of three parts: the 5′ *head*, also called the tRNA-related region; the central *body* (tRNA-unrelated region); and the 3′ *tail*, usually AT-rich or composed of a simple repeat. Certain SINEs, however, deviate from this pattern: their head can be derived from another cellular RNA; the body or tail may be short or absent; they can form dimers or even trimers with other SINEs; and so on. Some examples of SINE structure are given in Fig. 1 and are discussed later.

SINE copies in the genome are stably inherited over time. They accumulate point mutations and can be quite variable, usually 5–35%, depending on the time of their appearance in the genome. Such copies together form a SINE family. However, some regions of SINE sequences are more conserved than others (commonly, their tail is most variable), which may be associated with the retroposition machinery.

SINEs are often enriched in CpG (e.g., Alu, CAN, and S1). These dinucleotides are primary targets for methylation. Indeed, most (but not all) SINEs are partially or completely methylated in somatic cells (Hellmann-Blumberg *et al.*, 1993). Moreover, deamination of 5-methylcytosine and subsequent DNA replication introduce TG or CA instead of CG, providing high rates of such transition (Bird, 1980). Indeed, an ~10-fold higher mutation rate is observed at CpG sites, compared with non-CpG sites, within human Alu (Batzer *et al.*, 1990; Labuda and Striker, 1989).

### A. Head

The heads of all SINEs share at least one common character: a promoter for Pol III (Jagadeeswaran *et al.*, 1981). This eukaryotic enzyme is responsible for synthesis of small nuclear and cytoplasmic RNAs (tRNA, 5S rRNA, 7SL RNA, U6 RNA, etc.). It can utilize three types of promoters: the type 2 promoter specific for tRNA (and SINEs) is internal (i.e., it lies downstream

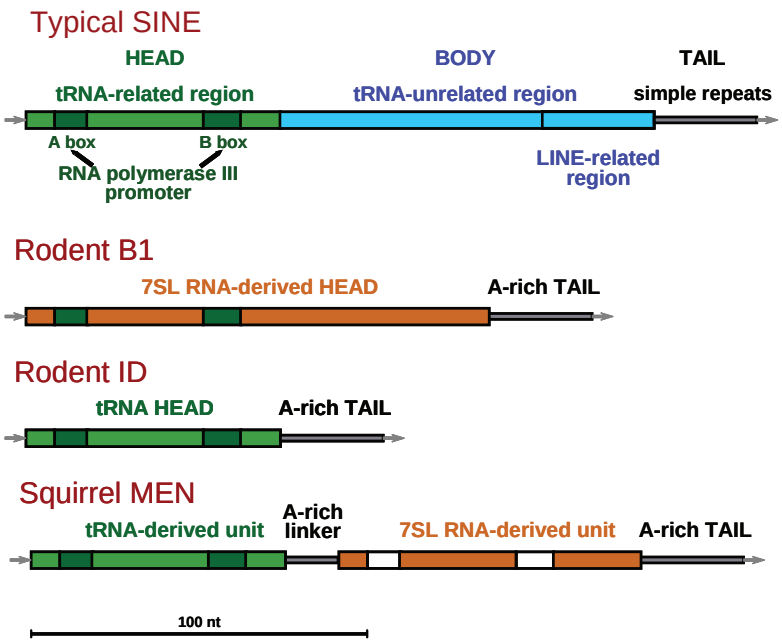


FIG. 1 Examples of SINE structure. A typical SINE includes a tRNA-derived head, a body with the 3' end derived from a partner LINE, and a tail (simple repeat sequence). Other examples include 7SL RNA-derived SINE (B1), bodyless SINE (ID), and dimeric SINE (MEN). (See also color insert.)

of the transcription initiation site) and includes two short (~11 bp) sequences typically spaced by 30–35 bp; these sequences are called A and B boxes (see Section III.A.1).

SINEs share a typical consensus of the Pol III promoter: TGGCNNAGTGGN$_{30–35}$GGTTCGANNCC. It usually starts 10 –15 nucleotides from the 5' end. The distance between the A and B boxes can be longer (e.g., ~45 bp in CAN and ERE 1). Accordingly, this (as well as various duplications within this region) can extend the head length beyond the length of tRNA.

As already mentioned, heads of most SINEs are derived from cellular tRNAs as deduced from considerable sequence similarity (Daniels and Deininger, 1985; Lawrence et al., 1985; Sakamoto and Okada, 1985). In many cases, a particular tRNA that gave rise to a SINE can be easily found, which substantiated attempts to classify SINEs as tRNA$^{Lys}$-related, tRNA$^{Arg}$-related, and so on (Okada and Ohshima, 1995). At the same time, some SINEs have deviated considerably from the original sequences and their reliable attribution to a particular tRNA species is hardly possible. The

cloverleaf structure of tRNA is not necessarily preserved in the derived SINEs (e.g., in ID SINE; Rozhdestvensky *et al.*, 2001), although it was proposed to mediate their transport to the cytoplasm by the tRNA nuclear export system (Weiner, 2002). In any case, the highest similarity with tRNA covers the A and B boxes and the region between them.

The main function of the head is initiation and regulation of SINE transcription; in addition, this region can mediate SINE transport to the cytoplasm (at least in 7SL-derived SINEs) and provide for RNA stability.

## B. Body

Nucleotide sequences of a typical SINE body are 50–200 bp long and are usually unique for each SINE family. The 3′ part of a typical SINE body is similar to the 3′ end of a partner LINE (Ohshima *et al.*, 1996; Okada *et al.*, 1997). This region is essential for reverse transcription of a SINE RNA; apparently, the reverse transcriptase complex uses it instead of the similar 3′ end of the LINE RNA (see Section III.A.4). Accordingly, its length (typically 50–100 nucleotides) depends on recognition requirements of reverse transcriptase of the partner LINE.

However, there usually is another region between the tRNA- and LINE-derived regions. We know neither the origin nor the function of it. This region can be conserved in a broad range of SINEs even with unrelated tRNA- and LINE-derived parts. Two such "core" regions are presently known (Gilbert and Labuda, 1999; Ogiwara *et al.*, 2002). Although their function is unclear, such conservation in quite different SINEs suggests functional significance of these cores.

One more structure that can be present in the body of certain SINEs is a $(CT)_n$ or simply CT-rich stretch of variable length just before the tail (Borodulina and Kramerov, 2001). Such variability can be maintained by reverse transcriptase slippage (see Section III.A.5) or by DNA-mediated mechanisms specific for microsatellite sequences. Its significance remains unclear.

However, no similarity between SINE body and the 3′ end of a partner LINE is observed for many SINEs, for example, most mammalian SINEs. In this case, the function of the LINE-like region can be adopted by the tail (see later discussion).

## C. Tail

The 3′ end of SINEs is usually an A- or AT-rich tail. It can vary greatly in both length and sequence; for instance, it can be $(A)_n$ or $(CA)_n$ in individual SINEs of the same family. In many SINEs these tails end with a run of

T residues, which is a termination signal for Pol III (Borodulina and Kramerov, 2001; Haynes and Jelinek, 1981). Other SINEs lack such a terminator and Pol III continues transcription until a terminator is encountered in the genomic sequence outside the transcribed SINE copy. Alternatively, SINE tails can be composed of short tandem repeats of 3–5 bp (nonmammalian vertebrate SINEs) or T-rich sequence (e.g., p-SINE1 from rice).

In addition to transcription termination in some SINEs, the tail can be responsible for SINE RNA delivery to the LINE reverse transcriptase complex. Sometimes, their sequence is important (human Alu; Dewannieux *et al.*, 2003); in other cases the tail length rather than the sequence is important (eel UnaSINE1; Kajikawa and Okada, 2002). In addition to these functions, the SINE tail can be significant for RNA stability and transport between the nucleus and cytoplasm.

## D. SINEs with Atypical Structure

Many SINEs deviate from the previously described typical structure. Some SINEs are derived from 7SL RNA or 5S rRNA; others lack or have a very short body or are composed of two or three SINE units, and so on (Fig.1). Actually, the first discovered SINEs, human Alu (Deininger *et al.*, 1981) and mouse B1 (Krayev *et al.*, 1980), are related not to tRNA but rather to the beginning and end of cellular 7SL RNA (Ullu and Tschudi, 1984), a component of cytoplasmic ribonucleoprotein (RNP) called the *signal recognition particle* involved in translation of secreted proteins in all eukaryotes. The 7SL-derived region is ~100–160 bp long, depending on the size of internal deletion. Although 7SL-derived SINEs have no LINE-related region, their A-rich tail, essential for retroposition (Dewannieux *et al.*, 2003), seems to be its functional substitute.

Primate Alu is composed of two similar but not identical units. Such dimerization can also be found in other SINEs. In some cases SINEs with quite different structure can be fused, for example, 7SL RNA- and tRNA-derived SINEs; trimeric SINEs are also known. Although several SINEs with both units derived from tRNA are known (e.g., CYN and DAS-II); most composite SINEs have at least one 7SL RNA-derived unit. We do not know the significance of dimerization; however, dimeric SINEs usually outnumber their monomeric ancestors. Although rodent B1 SINEs are monomeric, they have an ~30-nucleotide internal duplication in their body (called ''quasi-dimer'' by analogy with Alu). Such internal duplications can be found in other SINEs as well.

To date, we know one SINE derived from the third class of cellular RNA, 5S ribosomal RNA, in the zebrafish genome (Kapitonov and Jurka, 2003). Discovery of this SINE was not too surprising (Weiner, 2002), because,

similar to 7SL RNA and tRNAs, 5S rRNA is an abundant Pol III transcript with internal initiation (see Section III.A.1).

In addition to SINEs with complicated structures, there are SINEs with reduced structure, such as rodent ID, which has a very short tRNA-derived body (75 bp) and an A-rich tail (Sutcliffe *et al.*, 1982). Vic-1 from camels (Lin *et al.*, 2001) and DAS-I from armadillo (Borodulina and Kramerov, 2005; Churakov *et al.*, 2005) are other examples of this kind. It is of interest that all these SINEs derive from $tRNA^{Ala}$.

Because SINE structure depends on both the cellular machinery (which is relatively uniform in eukaryotes) and the machinery of partner LINEs, unusual SINE forms can be expected in species with unusual LINEs (e.g., in lower eukaryotes).

## III. Replication of SINEs

### A. Amplification Mechanism

Unlike autonomous transposons that transcribe their genes to produce mRNA and proteins mediating their replication, nonautonomous SINEs completely rely on the cell machinery and autonomous retrotransposons for their replication. First, active genomic SINE is transcribed by the cellular Pol III, this RNA is then delivered to the reverse transcriptase complex in the cytoplasm, and finally it is reverse transcribed and the resulting DNA is integrated in the genome (Fig. 2).

#### 1. Transcription

Autonomous transposons express proteins and, hence, use the RNA polymerase II machinery (which is also the case for the derived nonautonomous transposons). SINE RNA is present in two RNA pools in the cell: high molecular weight RNA in the nucleus, and low molecular weight RNA in both nucleus and cytoplasm (Elder *et al.*, 1981; Kramerov *et al.*, 1982). High molecular weight SINE RNA is transcribed by RNA polymerase II within introns, with most of the transcripts being eliminated by splicing; low molecular weight SINE RNA is synthesized by Pol III, as indicated for several mammalian and nonmammalian SINEs in α-amanitin inhibition experiments (Elder *et al.*, 1981; Haynes and Jelinek, 1981; Kramerov *et al.*, 1985a; Matsumoto *et al.*, 1984), and is involved in further replication.

Generally, three types of eukaryotic Pol III promoters are recognized (Fig. 3A) (Schramm and Hernandez, 2002). The type 1 promoter (specific for 5S RNA) is internal (i.e., it resides within the transcribed region) and

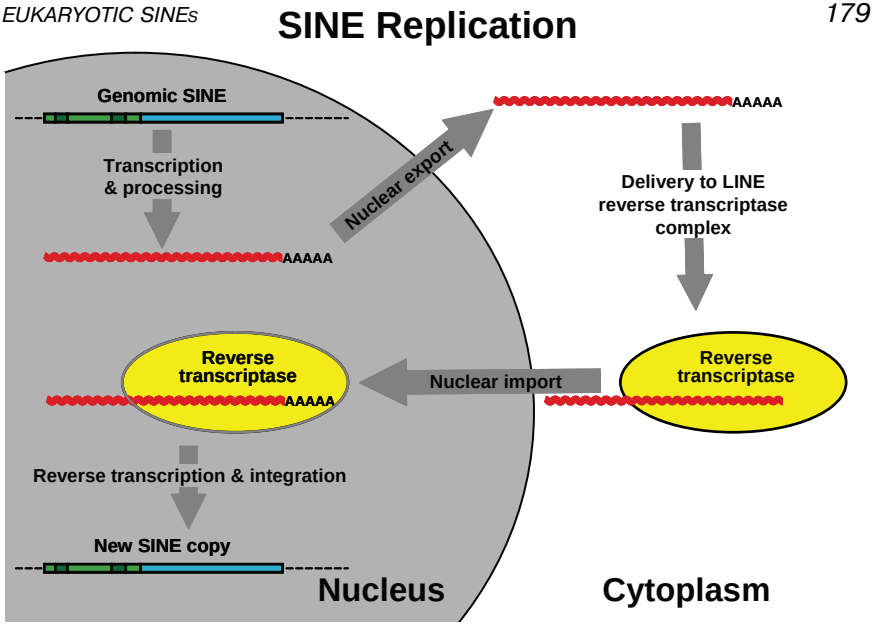


FIG. 2 Scheme of SINE replication (see text for explanation). (See also color insert.)

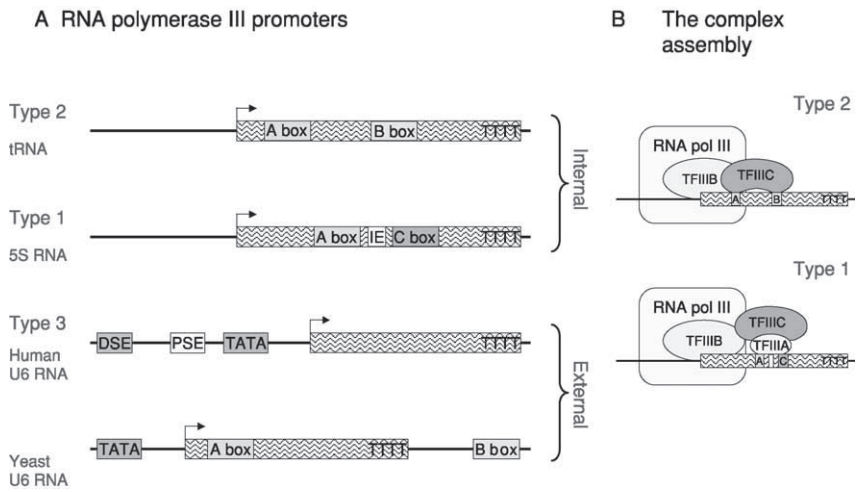


*FIG. 3 RNA polymerase III promoters and initiation of transcription. (A) Various types of promoters. (B) Schematic initiation of transcription. (See also color insert.)*

includes a conserved A box, an intermediate element (IE), and a C box. The type 2 promoter (specific for tRNAs) is also internal and includes conserved A and B boxes. In contrast, the type 3 promoter (U6 small nuclear RNA [snRNA] and 7SK RNA) is external and includes a TATA box, a proximal sequence element (PSE) (both are better known as elements of RNA polymerase II promoters), and a distal sequence element (DSE). In addition, there are hybrid promoters combining these internal and external elements (e.g., the promoter of 7SL RNA has an upstream element in addition to internal A and B boxes).

The general pattern of transcription initiation includes binding of the transcription factor TFIIIB to the DNA sequence, which allows recruitment of the Pol III complex (Fig. 3B). In the case of type 2 promoters, the A and B boxes are recognized by TFIIIC, which is followed by recruitment of TFIIIB and Pol III. Type 1 promoters are recognized by TFIIIA, which allows sequential binding of TFIIIC, TFIIIB, and Pol III. Finally, in type 3 promoters, the PSE is recognized by the $SNAP_C$ complex, and the TATA box is recognized by TATA-binding protein (a component of the TFIIIB-like complex), which allows recruitment of Pol III.

Unlike other RNA polymerases, Pol III efficiently terminates at a simple stretch of four or more Us without any additional factors (Paule and White, 2000). All transcription factors remain bound to DNA during transcription, which allows efficient recycling of the process on the same template. However, both termination and recycling can be facilitated by phosphoprotein La (Maraia, 2001). This protein can bind the oligo(U) (termination signal) at the 3′ end of nascent RNA and protects it from degradation; in addition, it facilitates further RNA processing and/or their assembly into specific RNP complexes.

All known SINEs have internal promoters, most use the type 2 promoter, and zebrafish SINE3 uses the type 1 promoter (Kapitonov and Jurka, 2003). The promoter sequences are essential for SINE transcription by Pol III *in vitro* (Fuhrman *et al.*, 1981; Perez-Stable *et al.*, 1984; Wilson *et al.*, 1988); likewise, the promoter sequences are intact in most transcripts (Shaikh *et al.*, 1997).

In addition to these factors, transcription of SINEs can depend considerably on the upstream flanking sequences (Chesnokov and Schmid, 1996; Martignetti and Brosius, 1995; Roy *et al.*, 2000). The upstream *cis* factors can include a TATA box, a PSE, a CAA motif, and an E box (Arnaud *et al.*, 2001; Kobayashi and Anzai, 1998; Martignetti and Brosius, 1995; Roy *et al.*, 2000). Conversely, nonoptimal context can decrease or block SINE transcription, which may be responsible for the transcriptional inactivity of the majority of SINE copies (Shaikh *et al.*, 1997).

As with other class III genes, transcription of SINEs is terminated at a $T_4$ stretch, except that there may be no such signal within a SINE; in such cases a signal randomly located in the unique downstream sequence is used. Such a

situation is typical for many but not all SINEs; for instance, rodent B2 and equine ERE-1 have their own terminators (Borodulina and Kramerov, 2001). Finally, noncanonical termination was shown for fish Sm2 *in vitro* (Matsumoto *et al.*, 1989).

The context of a terminator such as a preceding short palindrome can significantly improve the efficiency of SINE transcription (at least *in vitro*), apparently, through accelerated recycling (Chu *et al.*, 1997). The involvement of La protein in termination, recycling, and nascent RNA protection from degradation was also shown for SINEs (Goodier and Maraia, 1998; and references therein).

## 2. Posttranscriptional Modification

RNA of at least some SINEs is processed after transcription by Pol III. Such processing includes polyadenylation specific for rodent B2, and possibly some other SINEs with polyadenylation signals (AATAAA) and a Pol III terminator at their 3′ end (Borodulina and Kramerov, 2001; Kramerov *et al.*, 1990). Likewise, B2 RNA seems to be posttranscriptionally polyadenylated: the mean length of the A-rich tail in mouse B2 RNA is nearly twice that of genomic copies (Bachvarova, 1988; Kramerov *et al.*, 1985b, 1990). However, such processing is missing in SINEs with an A-rich tail and no Pol III terminator (e.g., Alu); in this case, SINE tail is maintained by read-through transcription and, possibly, slippage mechanism of reverse transcriptase (see Section III.A.5).

Although there is no direct evidence of involvement of poly(A) polymerase in polyadenylation of B2 RNA, it is so far the only candidate. Normally, this enzyme complex recognizes the AAUAAA signal, cleaves the downstream mRNA, and adds poly(A) to the emerged 3′ end; however, cleavage-independent polyadenylation is possible for RNAs with an AAUAAA sequence near its 3′ end (Wahle and Ruegsegger, 1999). Indeed, the sequences of B2 (and other SINEs that are polyadenylated) carry several AAUAAA signals upstream of the Pol III terminator, so that nascent RNAs have such signals at their 3′ ends and thus can be polyadenylated (Borodulina and Kramerov, 2001; and references therein).

Unlike mRNAs carrying a 7-methylguanosine cap at their 5′ end, typical Pol III transcripts have a free triphosphate group (pppN–). However, capping was shown in at least one SINE. Murine B2 (but not B1) RNA proved to have an unusual caplike structure at the 5′ end (Shumyatsky *et al.*, 1990). This cap is a methyl group blocking γ-phosphate ($CH_3$pppN–); it is also specific for U6 snRNA, 7SK RNA, and plant U3 RNA (Shimba *et al.*, 1992; Shumyatsky *et al.*, 1990; Singh and Reddy, 1989). Such a cap proved to increase RNA stability (Shumyatsky *et al.*, 1993) and to reduce the ability of these RNAs to bind La protein (Bhattacharya *et al.*, 2002). This protein (at

least in humans) can bind the 5′ end of nascent RNA with oligo(U) at the 3′ end and protect the RNA from processing until La is phosphorylated, suggesting its involvement in RNA quality control (Maraia, 2001). Hence, thus modified SINE RNA can bypass this cellular check.

In addition, 7SL RNA is known to be processed by a special 3′-adenylating enzyme that removes three U residues and adds a single A residue at the 3′ end (Chen *et al.*, 1998). Alu RNA can also be processed by this enzyme *in vitro*. We do not know whether RNA of Alu (and other SINEs) is modified in this way *in vivo*; if so, removing the Pol III terminator could affect SINE replication (see Section III.A.1). A posttranscriptional addition of C or CA to the 3′ end was also observed for plant S1 (Pelissier *et al.*, 2004), which resembles enzymatic addition of CCA to tRNA.

Small forms of Alu and B1 RNA are found in the cytoplasm (Adeniyi-Jones and Zasloff, 1985; Maraia *et al.*, 1993). They lack a poly(A) tail and, in the case of Alu, are shorter than the full-length sequence. These small cytoplasmic RNAs (scRNAs) seem to arise by processing of the full-length RNAs (Adeniyi-Jones and Zasloff, 1985) but can also represent transcripts of mutant SINEs with a terminator signal in the middle (Shaikh *et al.*, 1997). Similar scRNAs are observed for insect Bm1 (Kimura *et al.*, 1999), plant S1 (Pelissier *et al.*, 2004), and rodent B2 (Bladon *et al.*, 1990; Kramerov *et al.*, 1990) and ID (McKinnon *et al.*, 1987).

Finally, *in vitro* transcripts of Sma I SINE from salmon were modified to contain pseudouridylic acid residues at the same positions as the ancestor $tRNA^{Lys}$; we do not know whether this modification is functionally significant or just reflects structural similarity between Sma I and $tRNA^{Lys}$ (Matsumoto *et al.*, 1984).

### 3. Transport to Cytoplasm

Although SINE RNAs are synthesized and later integrate into the genome in the nucleus, they also enter the cytoplasm. For instance, Pol III-transcribed RNAs of various SINEs are basically located in the cytoplasm (Kramerov *et al.*, 1982; Liu *et al.*, 1994; Pelissier *et al.*, 2004). Hence, they must be transported to the cytoplasm. Still, little is known about this process and the mechanisms of SINE nucleocytoplasmic transport discussed later are speculative.

There are several pathways of RNA nuclear export (Cullen, 2003). The most straightforward hypothesis is that tRNA-derived SINE RNAs are transported by the exportin-t/RanGTP pathway of tRNA export. However, only mature tRNA molecules with correct ends that can be aminoacylated are transported (Lund and Dahlberg, 1998), which is definitely not the case for SINEs; if they still use this pathway there should be a way to overcome the tRNA proofreading.

A possible pathway of RNA export can be proposed, considering the caplike structure found in mouse B2 SINE (Shumyatsky *et al.*, 1990). The nuclear export of uridine-rich small nuclear RNAs (snRNAs) is mediated by specific binding of their cap (Izaurralde *et al.*, 1995). At the same time, a caplike structure has so far been found only in B2 SINEs, which restricts any possible cap-specific Crm1/RanGTP pathway to cap-bearing SINEs.

There is another pathway of 5S rRNA nuclear export in *Xenopus* oocytes. This RNA is transcribed by Pol III and is exported by another Crm1/RanGTP pathway (Nakielny *et al.*, 1997). Because this pathway is active in oocytes but not in somatic cells, it could be adopted by SINEs with the same time pattern of transcription derived both from 5S rRNA and other RNAs.

Polyadenylated Pol III SINE transcripts could be exported from the nucleus via a RanGTP-independent mRNA pathway. This assumption is supported by the following lines of evidence: (1) mRNA polyadenylation is required for its nuclear export and many SINEs have an A-rich tail; (2) SINE RNA has no introns, which is a requirement for mRNA export; while (3) the presence of a cap (missing from at least some SINEs) is not critical for mRNA export (Cullen, 2003).

Finally, slightly more is known about 7SL-derived SINEs nuclear export. 7SL RNA is neither polyadenylated nor capped and is transported to the cytoplasm in a complex with signal recognition particle (SRP) proteins via the Xpo1-dependent Crm1/RanGTP pathway specific for large rRNAs (Gadal *et al.*, 2001). According to the *Xenopus* injection assay, Alu nuclear export is similarly mediated by specific binding to SRP9 and SRP14 (He *et al.*, 1994).

Still, particular RNA domains, an A-rich tail, or a cap are not specific for all SINEs, suggesting that there may be no universal pathway of SINE nuclear export and that individual SINE families find their own ways to the cytoplasm. Moreover, export of SINE RNAs can be unrelated to their replication; and certain SINE RNAs can remain in the nucleus to meet the reverse transcriptase complex and enter it there.

## 4. Delivery to Reverse Transcriptase

SINEs depend on LINEs in the retroposition process, as indicated by the (1) considerable similarity of the 3′ ends of some SINEs and LINEs (Okada *et al.*, 1997) and (2) similar preferred integration sites of SINEs and LINEs (at least in some mammals) (Jurka, 1997). Efficient retroposition of a SINE by a LINE reverse transcriptase was demonstrated (Dewannieux *et al.*, 2003; Kajikawa and Okada, 2002).

Reverse transcriptases of certain LINEs (human L1) function best *in cis*, that is, they process the RNA molecule that encoded them (Esnault *et al.*,

2000; Wei *et al.*, 2001). This can be true for I factor in *Drosophila* as well (see Boeke, 1997).

In the case of 7SL-derived SINEs (Alu and B1), RNA can be specifically bound by two SRP proteins, SRP9 and SRP14 (Weichenrieder *et al.*, 2000). Normally, SRP recognizes and binds a signal sequence of secretory and membrane proteins emerging from the ribosome, which inhibits further translation until the complex diffuses to the endoplasmic reticulum (ER) membrane (Walter and Johnson, 1994).

Thus, the Alu–SRP9/14 complex can also bind the ribosome translating the L1 RNA and present the A-rich tail of Alu to the reverse transcriptase, instead of the A-rich tail of L1 RNA (Boeke, 1997). Although proposed for a particular SINE–LINE pair (7SL-derived Alu and L1), this model can be extended to SINEs related to tRNA or rRNA, which are also components of the translation machinery.

One more factor, poly(A)-binding protein (PABP) associated with RNA of some SINEs in the cytoplasm (West *et al.*, 2002), can also mediate delivery of the RNA to a nascent reverse transcriptase, considering the ability of PABP to interact with the translation machinery proteins (Roy-Engel *et al.*, 2002a).

However, it remains unclear whether the *cis* preference is specific for reverse transcriptases of other LINEs as well, in particular those recognizing the 3′-terminal sequence of their RNA. It looks probable that LINEs elaborated two mechanisms to resist replication of foreign RNA: recognition of the 3′-terminal structure and *cis* preference. Accordingly, the partner SINEs had to acquire such a 3′ structure or find another way to present their RNA to the reverse transcriptase instead of the LINE RNA. Thus, most mammalian SINEs belong to the ''relaxed'' group, whereas fish SINEs represent the ''stringent'' group (Okada and Hamada, 1997).

## 5. Nuclear Import, Reverse Transcription, and Integration

LINE reverse transcriptase complex seems to capture RNA template for retroposition (either LINE or SINE RNA) in the cytoplasm (Wei *et al.*, 2001). The extremely rare occurrence of retropseudogenes with unexcised introns, as compared with intronless ones, further supports cytoplasmic localization of the functioning LINE reverse transcriptase complex (Weiner, 2002). Hence, the reverse transcriptase–RNA complex formed in the cytoplasm needs to be transported to the nucleus. Such transport is commonly mediated by special proteins, importins (karyopherins), that bind amino acid motifs called nuclear localization signals. Alternatively, the reverse transcriptase can be delivered to the nucleus during nuclear breakdown in dividing cells.

Progress in LINE research shed light on the retroposition process and proposed a model of target-primed reverse transcription (Luan *et al.*, 1993; Ostertag and Kazazian, 2001b). LINEs encode a protein with three enzyme

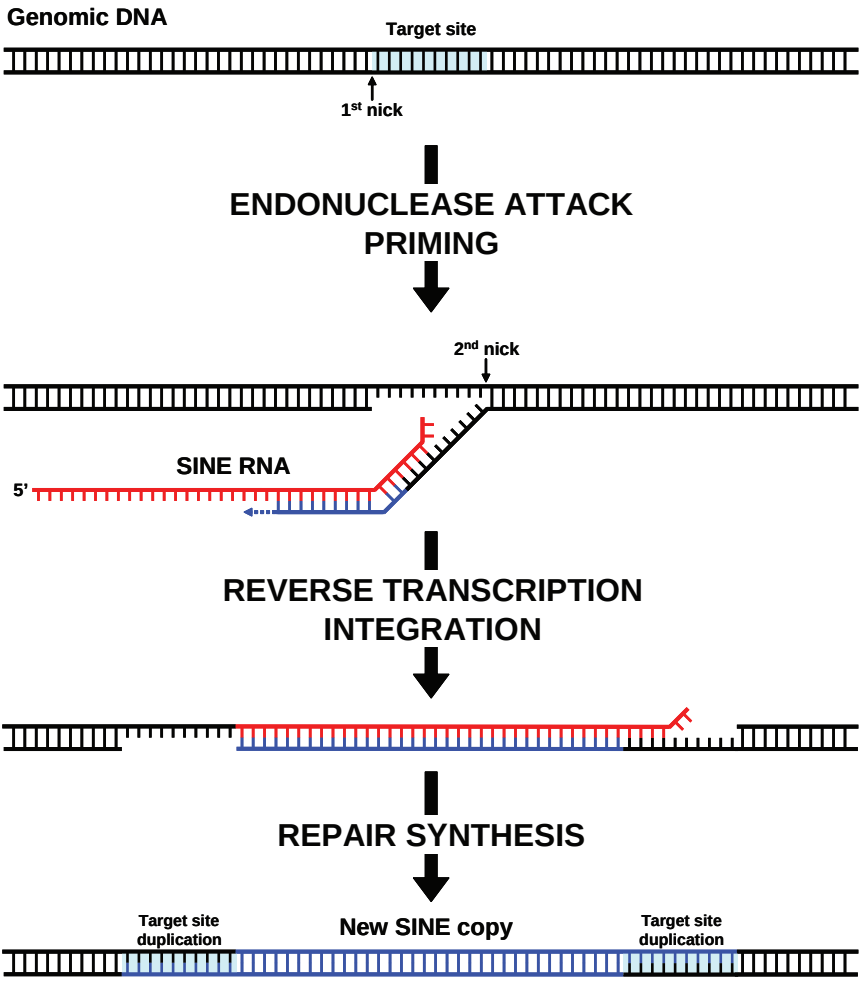


FIG. 4 Target-primed reverse transcription of SINEs by the reverse transcriptase complex of a partner LINE. (Model proposed for LINEs by Luan et al., 1993.) (See also color insert.)

activities: endonuclease, reverse transcriptase, and RNase H. Initially, the endonuclease cleaves one strand of the target site, generating a 3'-OH DNA, so that the reverse transcriptase can use it as a primer for reverse transcription of the retroposon RNA (Fig. 4). Although there are no direct data on the mechanisms of second strand synthesis and nick repair, it is likely done by the cellular DNA repair machinery and the RNase H.

There are two types of LINE endonucleases: rare, strictly sequence-specific endonucleases (e.g., in R2) similar to bacterial restriction endonucleases

(Yang *et al.*, 1999) and apurinic/apyrimidinic endonucleases (e.g., in L1), usually with low or no sequence specificity (Feng *et al.*, 1996; Jurka, 1997). If the second nick is formed "downstream" of the first nick, this leads to duplication of the target site (Fig. 4). For instance, the preferred site of the 5′ nick is TT$^{\downarrow}$AAAA for human Alu, whereas the nick in the other strand typically occurs 15–16 nucleotides downstream at the preferred site, TYTN$^{\downarrow}$ (Jurka, 1997). Other SINEs can have a similar site preference for the first nick; no specificity of the second nick has been revealed for rodent ID or plant S1 (Jurka, 1997; Tatout *et al.*, 1998). Although endonuclease activity is required for retroposition *in vivo* (Feng *et al.*, 1996), *in vitro* it can proceed on prenicked DNA in the absence of endonuclease (Cost *et al.*, 2002).

The reverse transcriptases of LINEs can also be divided into two groups by their sequence specificity (Okada and Hamada, 1997). The stringent group can initiate reverse transcription only for specific 3′-terminal sequences (Kajikawa and Okada, 2002; Luan and Eickbush, 1995) with specific secondary structure (Baba *et al.*, 2004; Mathews *et al.*, 1997). In addition, the presence of a short stretch of simple tri- or pentanucleotide repeats (apparently, generated by template slippage in a manner resembling another reverse transcriptase, telomerase) rather than their sequence may be crucial for retroposition mediated by the stringent reverse transcriptases. The relaxed group has considerably less stringent requirements for the 3′-terminal structure of their templates (Esnault *et al.*, 2000) but utilize the mechanism of *cis* preference to avoid processing "wrong" templates such as mRNA (Wei *et al.*, 2001).

Applicability of these schemes to SINEs has been confirmed by two breakthrough publications (Dewannieux *et al.*, 2003; Kajikawa and Okada, 2002). They demonstrated the involvement of LINE machinery in reverse transcription and integration of SINE RNAs: retroposition of eel UnaSINE1 was mediated by UnaL2 (stringent LINE) and retroposition of human Alu was mediated by L1 (relaxed LINE).

The length of the A-rich tail of SINEs can be an important factor of L1-mediated retroposition (Roy-Engel *et al.*, 2002a); moreover, such elongation of the tail (apparently, by the slippage mechanism) was demonstrated for both UnaSINE1 (Kajikawa and Okada, 2002) and Alu (Dewannieux *et al.*, 2003).

## B. Regulation of SINE Activity

Eukaryotic genomes can carry huge numbers of SINE copies; however, just a few active copies are involved in replication while the great majority of SINE copies are excluded from replication. What underlies such exclusion?

SINE replication includes two processes, transcription and retroposition, and an active SINE copy should be able to be transcribed by Pol III and delivered and processed by a LINE retroposition complex. In other words, an active SINE should at least reside in a favorable genomic environment and preserve an efficient Pol III promoter as well as other structures required for transcription and reverse transcription.

## 1. Control Factors

***a. SINE Sequence*** Because SINEs are transcribed by Pol III, the promoter sequence is essential for their transcription (Fuhrman *et al.*, 1981; Perez-Stable *et al.*, 1984; Wilson *et al.*, 1988) and only sequences with an intact promoter are transcribed *in vivo* (Shaikh *et al.*, 1997). Most mutations in Alu sequence beyond the promoter had little effect on its transcription levels (Aleman *et al.*, 2000). At the same time, addition of a Pol III terminator at the end of an Alu sequence enhanced its transcription considerably (Aleman *et al.*, 2000; Goodier and Maraia, 1998).

The Alu subfamilies representing ~80% of *de novo* inserted elements constitute less than 1% of the Alu transcripts (Shaikh *et al.*, 1997), which clearly indicates that SINE sequences are selected not only on the basis of efficient transcription. The importance of the 3′-terminal structure of the corresponding LINE as well as a simple repeat region at its very 3′ end has been confirmed for an eel SINE retroposition (Baba *et al.*, 2004; Kajikawa and Okada, 2002). The 3′ end of SINEs associated with the relaxed group of LINEs is also important; for instance, the length of the A-rich tail seems critical for Alu retroposition (Roy-Engel *et al.*, 2002a).

Conservation in some SINEs of other structures, such as tRNA-like folding (Okada and Ohshima, 1995) or a "core" region maintained in quite distant SINEs (Gilbert and Labuda, 1999), suggests their importance for yet unknown functions in SINE activity. Conserved Alu domain folding in 7SL-derived SINEs suggests the significance of these SINEs bound by the SRP9/14 complex; indeed, the long-term evolutionary decrease in Alu activity correlates with a decrease in their ability to be bound by SRP9/14 (Sarrowa *et al.*, 1997).

***b. Flanking Sequences*** Although the presence of A and B boxes of the Pol III promoter suffices to provide for their transcription *in vitro*, upstream sequences can contribute to transcription control *in vivo* (see Section III.A.1). Indeed, the upstream sequences of certain SINEs proved to modulate their transcription, making it tissue specific (Kobayashi and Anzai, 1998; Martignetti and Brosius, 1995; Roy *et al.*, 2000) or subject to p53 regulation (Chesnokov *et al.*, 1996). Moreover, such "fine-tuning" with the upstream

sequences can be mediated by a conserved mutation in the B box (Martignetti and Brosius, 1995).

There are no data confirming the effect of the downstream flanking sequences on SINE activity; however, because these sequences can be transcribed in the case of terminatorless SINEs, they can be important as well. For instance, preterminator sequences may contribute to recognition by La protein (Wolin and Cedervall, 2002) and thus affect transcription and stability of SINE RNA, or the distance between the poly(A) and terminator can be important for recognition by the reverse transcriptase complex.

***c. Chromatin Context*** Although nucleosomal repression of Pol III transcription is template dependent (Paule and White, 2000), many SINEs are susceptible to it. For instance, positioning of histone octamers on the promoter and transcriptional repression in Alu (Englander *et al.*, 1993) and considerable activation of B2 transcription in H1 histone-free cells (Russanova *et al.*, 1995) were shown. Hence, the chromatin context of a SINE can be an important factor in its developmental and tissue-specific regulation. Indeed, transcription of a transfected SINE correlated with that of a neighboring reporter gene (Slagel and Deininger, 1989). Moreover, because endonuclease activity of some LINEs depends on the chromatin structure (Cost *et al.*, 2001; Ye *et al.*, 2002), retroposition can be targeted to the chromatin contexts optimal for transcription at the time of integration.

***d. DNA Methylation*** Methylation of DNA template can repress Pol III transcription *in vitro*, which particularly applies to SINEs that are often CG-rich (CpG dinucleotide is the methylation target; Bird, 1980). For instance, most Alus are highly methylated and repressed in differentiated cells (but not in male germ line cells) (Kochanek *et al.*, 1993; Rubin *et al.*, 1994), their hypomethylation increases the level of Alu transcription, and vice versa, their methylation represses it (Kochanek *et al.*, 1993; Li *et al.*, 2000; Liu *et al.*, 1994). This pattern seems to be true even for plant SINEs (Arnaud *et al.*, 2001).

***e. RNA Stability*** RNA stability is another common issue of activity control. Most commonly, RNAs are protected from degradation by specifically bound proteins. In the nucleus, nascent SINE RNA is bound and protected by La protein (see Section III.A.1).

In the cytoplasm, RNAs of some SINEs proved to be associated with poly (A)-binding protein (PABP) (Muddashetty *et al.*, 2002; West *et al.*, 2002). Indeed, shortening of the 3′ poly(A) tail in mRNA is a signal for its degradation (Tourriere *et al.*, 2002). Similarly, this can be true for SINE RNA; likewise, the most recently inserted SINEs have longer A tails (Roy-Engel *et al.*, 2002a; Odom *et al.*, 2004).

The 5′-terminal protection can be mediated by the tRNA-like structure that seems to be maintained in some tRNA-derived SINEs (Okada and Ohshima, 1995). The 5′ end of 7SL-related SINEs seems to be protected by two SRP proteins (SRP9/14) (Bovia *et al.*, 1997; Chang *et al.*, 1996; Hsu *et al.*, 1995), whereas the 5′ RNA of BC1 (ID) can be bound by other yet unidentified protein(s) (West *et al.*, 2002). By analogy with mRNA, the 5′ end of rodent B2 can be protected by a caplike structure (Shumyatsky *et al.*, 1990, 1993).

At the same time, the retroposition-competent (full-length) cytoplasmic RNA of many SINEs seems to be short-lived (Bladon *et al.*, 1990; Chu *et al.*, 1995), particularly in comparison with their ancestors such as tRNA or 7SL RNA (Li and Schmid, 2004).

## 2. Regulation of SINE Activity in Organisms

SINEs were shown to be expressed by Pol III during early embryogenesis, but their expression rapidly decreases with development (Bachvarova, 1988; Taylor and Piko, 1987; Vasseur *et al.*, 1985). Similarly, transcription of SINEs is highly active in tumor cells but is low or absent in normal differentiated tissues (Grigoryan *et al.*, 1985; Kramerov *et al.*, 1982; Murphy *et al.*, 1983; Singh *et al.*, 1985). Note that transcriptional activity of these SINEs coincides with the germ line pattern of L1 expression (Branciforte and Martin, 1994; Ostertag *et al.*, 2002). This is consistent with a parsimonious retroposition strategy of SINEs: LINE-dependent replication in the genome requires integration into genomes of the germ line rather than somatic cells.

DNA methylation is a possible mechanism responsible for the developmental pattern of mammalian SINE activity (Bird, 1997), because DNA is demethylated at the early stages of zygote development (Jaenisch, 1997), which can activate SINE transcription (see Section III.B.1.d).

Rodent ID elements are specifically expressed in the testes at early stages of spermatogenesis (Kim *et al.*, 1995; Muslimov *et al.*, 2002) and in neural cells (this also applies to BC1 RNA, a functional RNA in rodents but also a master copy of ID) (Martignetti and Brosius, 1995; Sutcliffe *et al.*, 1984).

Finally, expression of certain SINEs was shown to increase under conditions of cell stress such as heat shock (Fornace and Mitchell, 1986; Kimura *et al.*, 1999; Liu *et al.*, 1995) and infection with adenovirus (Panning and Smiley, 1993), herpesvirus (Jang and Latchman, 1989), human immunodeficiency virus (HIV) (Jang and Latchman, 1992), simian virus 40 (SV40) (Singh *et al.*, 1985), and parvovirus (Williams *et al.*, 2004), as well as treatment with ethanol (Li *et al.*, 1999), cycloheximide (Liu *et al.*, 1995), and DNA-damaging agents (Rudin and Thompson, 2001). These findings, shared for many SINEs (human Alu, rodent B1 and B2, rabbit C, and silkworm Bm1), suggest that they can mediate the stress response (see Section V.B.2.b).

## C. Distribution of SINEs

### 1. Distribution in the Genome

Distribution of SINEs in the genome depends primarily on specificity of the endonuclease in the partner LINE reverse transcriptase complex (see Section III.A.5). Endonucleases of some LINEs are sequence specific, and these LINEs are predominantly inserted into specific genomic locations such as repeated ribosome genes (e.g., insect R2; Eickbush *et al.*, 2000). The endonucleases of most other LINEs are less specific and their integration occurs at a large scale, randomly. Moreover, *in vitro* retroposition can proceed at preformed nicks even in the absence of endonuclease activity (Cost *et al.*, 2002), suggesting that such integration can occasionally occur *in vivo* even for retroposons with sequence-specific endonucleases. Because SINEs utilize the retroposition machinery of LINEs, their insertion site preferences should follow that of their partners. This seems to be true for most if not all SINEs (e.g., Jurka, 1997); although no SINEs integrating into specific loci have been found so far.

The process of integration can also be responsible for a certain specificity of SINE distribution at the gene level. For instance, LINE endonuclease activity can depend on the chromatin structure (Cost *et al.*, 2001; Ye *et al.*, 2002) and integration of *Brassica* S1 SINE demonstrates preferential targeting to matrix attachment regions (Tikhonov *et al.*, 2001). In terms of the replication strategy, it could be advantageous to avoid integration into inactive genomic regions (with repressed transcription). Indeed, more than half of *Arabidopsis* SINEs are located within 0.5 kb of genes or in their introns (Lenoir *et al.*, 2001).

Local density of SINEs on chromosomes is also uneven. For instance, the local concentration of SINEs in the centromeric regions is quite high in *Arabidopsis thaliana* (Kapitonov and Jurka, 1999; Lenoir *et al.*, 2001). Despite a more even distribution of SINEs over human, mouse, and rat chromosomes, there are regions with high or low local densities of SINEs (Mouse Genome Sequencing Consortium, 2002; Rat Genome Sequencing Project Consortium, 2004). Moreover, a striking similarity of densities of lineage-specific SINEs is observed along orthologous human, mouse, and rat chromosomes; in other words, at the megabase scale, different SINEs tend to accumulate in the same genomic loci of different organisms (Mouse Genome Sequencing Consortium, 2002; Rat Genome Sequencing Project Consortium, 2004).

Another pattern of SINE distribution concerns local GC content in the genome. Thus, most human and rodent SINEs appear in a GC-rich context, whereas their partner LINEs are in GC-poor regions (Smit, 1999). In addition, a negative correlation was observed between SINE density and

nucleotide substitution rate; that is, SINEs tend to accumulate in slowly changing DNA (Yang *et al.*, 2004). This is not surprising considering that GC-rich regions represent active chromatin enriched in expressed (and thus conserved) genes. At the same time, the youngest Alu repeats (but not the rest) deviate from this pattern and, similar to L1, are commonly found in an AT-rich context (International Human Genome Sequencing Consortium, 2001).

SINEs, which can be targets for *de novo* methylation (Arnaud *et al.*, 2000; Hasse and Schulz, 1994; Yates *et al.*, 1999), seem to be excluded from imprinted regions of human genome (Greally, 2002) whose control is linked to methylation. Active young subfamilies of Alu are more abundant on the human Y chromosome (Jurka *et al.*, 2002), which can also be related to developmental methylation of the genome (see Section III.B.2).

Such preferred accumulation of SINEs can be explained by positive or negative selection. In the case of positive selection, SINE insertion into active chromatin can be beneficial, for example, through promotion of segmental duplications of gene-rich regions (Bailey *et al.*, 2003) or stress-induced control of gene expression (see Section V.B.2.b). In the case of negative selection, SINEs can be lost from AT-rich regions as a result of different fixation in a population or by a yet unknown mechanism of SINE excision (International Human Genome Sequencing Consortium, 2001; Pavlicek *et al.*, 2001). Alternatively, such preferred accumulation can reflect insertional preferences of SINEs, which requires additional factors that invert the insertional preferences of LINE retroposition machinery (Yang *et al.*, 2004).

## 2. Distribution Among Organisms

At present SINEs are known in many higher eukaryotes including plants, vertebrates, and invertebrates (Table I). No SINEs have yet been identified in fungi or protists (some repeated elements reported as SINEs in these organisms are not transcribed by Pol III and, rather, represent fragments of autonomous elements). No SINEs have been described in organelles.

Because SINEs rely on the machinery of LINEs, the species lacking LINEs should also lack SINEs (e.g., *Saccharomyces cerevisiae*). In addition, we know that SINEs are missing from several genomes with LINEs: the fruit fly *Drosophila* (while SINEs are known in at least some insects), and the diatom *Thalassiosira pseudonana* (our unpublished data) and are likely to be missing from the genomes of many other lower eukaryotes. It appears that SINEs are not found in small-genome eukaryotes (the genome sizes of the fruit fly and diatom are ~3.6 and 1% that of mammals, respectively), suggesting mechanisms opposing repeat expansion in their genomes (indeed, such mechanisms are known in *Saccharomyces* and *Neurospora*) (Bestor, 1990). Alternatively, there may be more specific limitations such as the inability of

Pol III to direct transcription using an internal promoter alone in *Schizosaccharomyces pombe*. Initiation of Pol III in this yeast species requires a TATA box outside the transcript (Hamada *et al.*, 2001), which is hardly compatible with the scheme of SINE transcription.

In general, the greater the size of the genome, the larger the fraction of transposons (and SINEs, in particular); for instance, the fraction of SINEs is 0.02% in the *Arabidopsis* genome ($1.25 \times 10^8$ bp) or 13.6% in the human genome ($3.3 \times 10^9$ bp). It is difficult to determine what is cause and what is effect in this relationship, but it is tempting to speculate that active SINEs together with other retroposons can considerably increase the size of the genome.

In mammals, usually more than one SINE family (typically two to four) can be found in a species (Table I). This can be true for other organisms as well (e.g., fish or mollusks; Table I). Moreover, in some cases more than one SINE family can be active at the same time (e.g., B1 and B2 in mouse or B2 and ID in rat; Rat Genome Sequencing Project Consortium, 2004).

## IV. Evolution of SINEs

### A. Origin of SINEs

The first insight into the origin of SINEs was provided by the similarity between Alu/B1 and 7SL RNA sequences (Ullu and Tschudi, 1984). Soon after, an analogous similarity was shown between many other SINEs and tRNA (Daniels and Deininger, 1985; Lawrence *et al.*, 1985; Sakamoto and Okada, 1985). Indeed, tRNA pseudogenes and other small nuclear RNAs can be found in eukaryotic genomes (Weiner *et al.*, 1986). The internal Pol III promoter can provide for their transcription (although it had to be modified for 7SL RNA to be transcribed without the external promoter element specific for this RNA gene); however, tRNA pseudogenes are quite rare, indicating low efficiency of their retroposition.

The tail of SINEs had to be modified to be efficiently processed by the LINE machinery. Indeed, Okada *et al.* demonstrated that the 3′ ends in some SINEs and LINEs are similar (Ohshima *et al.*, 1996; Okada *et al.*, 1997) and later confirmed the relevance of this region for reverse transcription (Kajikawa and Okada, 2002). Thus, the tail of such SINEs seems to be derived from partner LINEs. The origin of the A-rich tail, which can play a similar role in SINEs whose partner is L1 LINE (Roy-Engel *et al.*, 2002a), is not so clear.

Gilbert and Labuda (1999, 2000) revealed a short "core" region conserved in many SINEs from a broad range of organisms (e.g., molluscan OR1 and

mammalian Ther-2), whereas the 5′ tRNA- and 3′ LINE-related sequences could be unrelated. Another, even more conserved core region was found in a different set of SINEs from fish and amphibians (V-SINEs; Ogiwara *et al.*, 2002). Although we presently do not know the function of these core regions, such conservation seems significant. Little is known about the origin of these cores; the core of V-SINEs is similar to a fragment of human DNA transposon MER6 (Ogiwara *et al.*, 2002).

The precise mechanism of SINE generation is not known; however, we can speculate that SINEs appeared by combination of these modules (in the general case, tRNA gene, core, and 3′ LINE-derived region). Such an event can be illustrated by the "family" of U6 snRNA–3′ L1 terminus chimeras identified in the human genome. These chimeras seem to have originated independently (Buzdin *et al.*, 2002) and were not as successful as SINEs, apparently, because U6 snRNA transcription is initiated at an external promoter. Another example of this kind (5S rRNA–3′ LINE terminus) was found in the *Dictyostelium* genome (Szafranski *et al.*, 2004).

The appearance of LINEs made possible the emergence of pseudogenes from mRNA, which are delivered to a LINE reverse transcriptase complex instead of the LINE RNA (as was demonstrated experimentally; Esnault *et al.*, 2000; Wei *et al.*, 2001). Such Pol II pseudogenes are not likely to be transcribed unless integrated near an active promoter (apparently disturbing transcription of a cellular gene so that such integrations will be negatively selected). Such an event can be the first stage in SINE origin if a Pol III transcript (rather than mRNA) is integrated into a favorable genomic environment, because such a pseudogene already contains an internal Pol III promoter and can be transcribed.

The generation of an RNA–LINE 3′ end hybrid can be the second stage. This event, increasing the efficiency of the template for reverse transcription, could be an accidental genetic event placing the 3′ end of a LINE adjacent to an RNA pseudogene. We can further speculate that, if this event was related to retroposition, it could be facilitated by sequence preferences of LINE reverse transcription: for example, a tRNA pseudogene could integrate into a target site of a previously integrated truncated LINE or into the 3′ end of a LINE.

Alternatively, this could be mediated by a template switch mechanism (Gilbert and Labuda, 1999; Weiner, 2002). Indeed, Kazazian and co-workers described quite frequent inversions in the L1 sequence after integration (Goodier *et al.*, 2000) and proposed a model explaining this event by a preliminary nick in the second strand so that it could be used as a second primer for reverse transcription (twin priming) (Ostertag and Kazazian, 2001a). A similar mechanism with a "pre-SINE" RNA as the second template can generate a transcriptionally and retropositionally competent SINE.

It is difficult to determine when SINEs appeared. Because short retroposons are active for a certain time and are subject to mutation, we can easily reveal the SINE families that are active now or were active recently. SINEs that became inactive long ago (such as Ther-1) are difficult to find and, if older short retroposons existed, their remains have diverged too much to be revealed. For instance, fossil SINEs older that 100–200 MYA cannot be recognized in mouse and human genomes (Mouse Genome Sequencing Consortium, 2002). In any case, because SINEs cannot amplify without partner LINEs, their appearance should follow the appearance of these autonomous retroposons (apparently, dating back to the origin of eukaryotes; Malik *et al.*, 1999).

In contrast to LINEs that share a common ancestor (Malik *et al.*, 1999), different SINEs seem to be generated many times *de novo* in different lineages from the available cellular and LINE material, although SINE families can share similar modules, such as the core (see also Section IV.D).

## B. SINE Activity Over Time

SINE amplification activity has changed during evolution. For instance, the human Ther-1 element is no longer active whereas Alu is still replicating (International Human Genome Sequencing Consortium, 2001). We do not know exactly how long SINEs can remain active; however, this period can considerably vary. For instance, human Alu remained active for at least 65 million years (Batzer and Deininger, 2002), whereas CHR-2 of cetaceans, hippopotamuses, and ruminants was active for twice as long (Nomura *et al.*, 1998). Another example is the different fate of certain SINEs in mouse and rat: B1 is active and ID is inactive in mouse, and vice versa in rat (Rat Genome Sequencing Project Consortium, 2004). Moreover, this pattern is also true for SINE subfamilies that can replace each other over time. For instance, the oldest Alu subfamilies are not replicating now; they were active for different time periods and yielded different numbers of copies in the genome (Batzer and Deininger, 2002; Ohshima *et al.*, 2003); a similar pattern is observed for rat ID (Kim and Deininger, 1996).

Although we do not know the processes that lead to SINE inactivation, there is one clear relationship. Because SINEs depend on LINEs as a source of reverse transcriptase, inactivation of a partner LINE inevitably leads to their inactivation. Such coordinated dynamics are clearly seen for LINE2–Ther-1 activity in human and mouse genomes (International Human Genome Sequencing Consortium, 2001; Mouse Genome Sequencing Consortium, 2002).

One can also expect such coordinated changes between families of partner SINEs and LINEs. Thus, amplification of Alu SINEs is attributed to the

activity of just a few L1 subfamilies (Odom *et al.*, 2004). Alternatively, SINE family dynamics can reflect the evolution of cellular functions associated with SINE replication. For instance, such coordinated changes between Alu replication activity and structural changes in SRP9/14 protein have been demonstrated (Sarrowa *et al.*, 1997).

## C. Patterns and Mechanisms of SINE Evolution

Once a SINE capable of amplification appears in the genome, it starts to evolve. In some cases we can find a fossil SINE with limited activity replaced by a highly active successor; for instance, ancestor B1 (pB1) has ~100 times fewer copies than descendant B1 in mouse genome (Mouse Genome Sequencing Consortium, 2002; Vassetzky *et al.*, 2003). Some patterns can be noted in such "tuning" of SINE activity, including fusion of SINE monomers, large-scale (tens of nucleotides) insertions, deletions, and duplications, as well as point mutations.

Although most SINEs are monomeric, dimeric, and even trimeric, short retroposons are not uncommon; for instance, dimeric Alu in humans is probably the best known SINE (Deininger *et al.*, 1981). Usually such dimers are more successful than their monomers, although opposite examples are known. Thus, dimeric B1 and dID are much more abundant in squirrels and dormice than monomeric B1–dIDs, whereas the opposite is observed in guinea pig (Kramerov and Vassetzky, 2001). Significantly, dimeric B4 with an inverse arrangement of units ("ID–B1") is found in the mouse genome, where it is almost as numerous as monomeric B1 and considerably outnumbers ID (Mouse Genome Sequencing Consortium, 2002). The benefit of dimerization is clearly seen for 7SL RNA-derived SINEs: most of their successful variants are dimeric.

Although we do not know the mechanisms of such dimerization, it could be mediated by SINE integration at the site of previous integration. Because of context preferences of LINE reverse transcriptase complexes, such nested integrations are probably not uncommon. We can further speculate that the absence of a transcription terminator observed in many SINEs (see Section III.A.1) can facilitate dimerization in this case.

Another typical modification of SINEs is internal duplication resulting in tandem repeats up to 30 nucleotides long. Some of these duplications are fixed in the descendants whereas others are present in single sequences; sometimes, more than two repeat units are present (Borodulina and Kramerov, 2001; Vassetzky and Kramerov, 2002). Akin to the above-mentioned dimerization of 7SL RNA-derived SINEs, independent internal duplications (also called quasi-dimerizations) occurred in roughly the same region of B1 sequences in two rodent lineages, a 29-bp duplication and a

19-bp duplication in myomorph and sciuromorph rodents, respectively (Kramerov and Vassetzky, 2001). As with true dimerization, SINEs with quasi-dimerizations can be much more successful, which suggests its (yet unknown) functional significance.

Such duplications and deletions are quite common in SINEs, which may point to a specific mechanism for their generation. Experiments on eel SINE/LINE retroposition demonstrated that reverse transcriptase replicating short 3- to 5-nucleotide tandem repeats can reuse one of them as a template; such template slippage becomes evident if the repeats are not perfect (Kajikawa and Okada, 2002). Likewise, many internal duplications in SINEs arise from a sequence flanked by short direct repeats (Vassetzky *et al.*, 2003), suggesting that such duplications were generated through reverse transcriptase slippage at the short repeats (Fig. 5). Although they are not tandem in this case, it is not unusual for reverse transcriptases to jump between templates. For instance, replication of retroviruses requires two template switches of their reverse transcriptase (a homolog of LINE reverse transcriptase) at direct repeats (Coffin *et al.*, 1997).

A similar mechanism can be true for short deletions. Retroviral reverse transcriptase activity results in deletion of regions between short direct repeats at a high frequency (Pathak and Temin, 1990). Accordingly, short direct repeats frame deletions in many cases in SINEs (Vassetzky *et al.*, 2003).

Likewise, a high rate of short deletions and tandem duplications associated with 2- to 13-nucleotide direct repeats was observed in different families of LTR-retrotransposons in *Arabidopsis* (Devos *et al.*, 2002), which was attributed to illegitimate recombination. Of course, RNA-unrelated mechanisms can underlie these events; for instance, the association between deletions/insertions and short direct repeats is known for nonhomologous DNA end joining after a double-strand break (Puchta, 2005).

Finally, a fraction of point mutations in SINE sequences can also be mediated by RNA-related mechanisms. Indeed, both Pol III and reverse transcriptase have no editing function and are error-prone (Katz and Skalka, 1990). Even in "conventional" DNA-mediated mutagenesis there is one mechanism with particular significance for SINEs. Many SINE sequences are rich in (CG) dinucleotide, which is a target for methylation. Deamination of 5-methylcytosine and subsequent DNA replication introduce TG or CA instead of CG, providing high rates of such transition (Bird, 1980) to create hypervariable sites in SINEs (Batzer *et al.*, 1990; Labuda and Striker, 1989).

Another unusual feature of SINEs is their "mosaic evolution." Because only a few SINE copies can replicate, there are subfamilies that share similar structural traits (diagnostic mutations). At the same time, these traits may be shuffled in rare copies (Lenoir *et al.*, 1997; Terai *et al.*, 2003; Zietkiewicz and Labuda, 1996). A study on segmental duplications demonstrated such mosaic elements at the junction sites of nonallelic Alu–Alu recombination

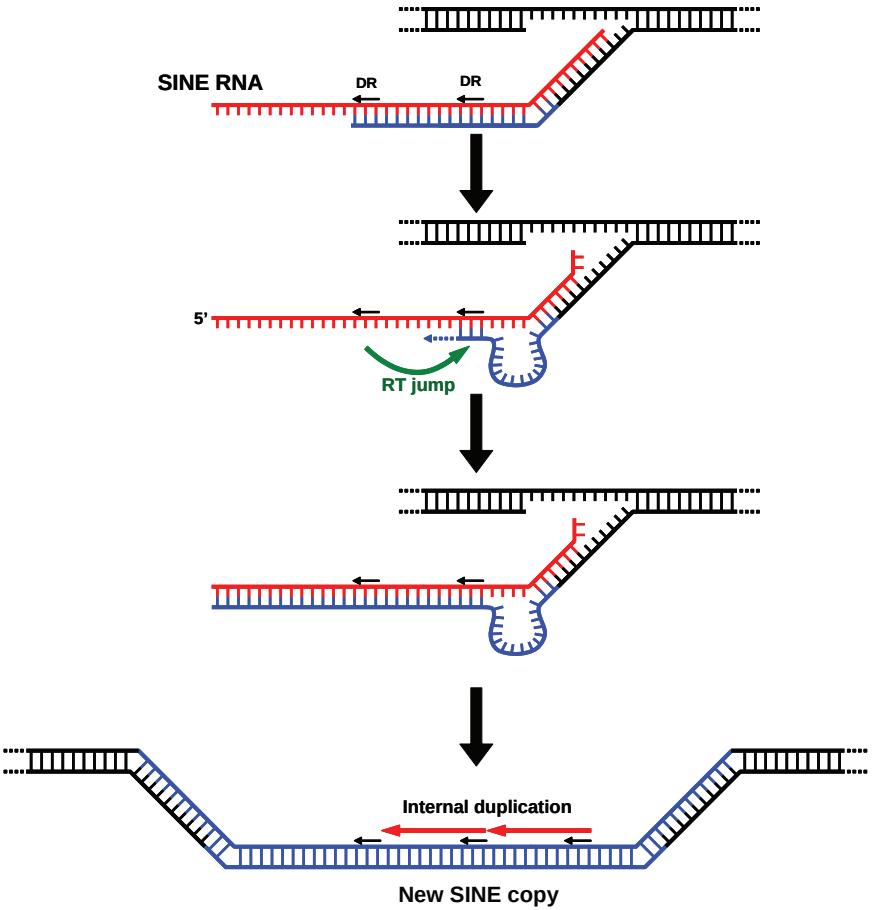


FIG. 5 Proposed mechanism of internal duplication in SINEs during reverse transcription. A backward reverse transcriptase jump leads to a duplication (as shown), whereas a forward jump causes a deletion. DR, short direct repeat; RT, reverse transcriptase. (See also color insert.)

(Bailey et al., 2003). Similarly, several cases of ''gene conversion'' were reported for Alu sequences (Maeda et al., 1988; Roy-Engel et al., 2002b). In this case, a fragment within a SINE is replaced with that from another SINE, which involves two recombination events. Although such events seem to be repressed for short (SINE-sized) sequences (Cooper et al., 1998), they occasionally occur, probably because of SINE abundance. It is also possible that a template switch mechanism during reverse transcription can contribute to the formation of mosaic SINEs in a manner similar to recombination in retroviruses (Negroni and Buc, 2001).

## D. SINEs as Phylogenetic Markers of Their Host Evolution

Once inserted a SINE copy remains in the genomes of all descendants indefinitely, which provokes the use of SINEs as landmarks of their host evolution (Hillis, 1999; Miyamoto, 1999; Ryan and Dugaiczyk, 1989; Shedlock and Okada, 2000).

Vertical transmission is the only known form of SINE distribution. There are no confirmed cases of SINE horizontal transfer and it is severely restricted by their nonautonomous amplification. Although elimination of particular SINE copies occasionally occurs, it usually involves the neighboring genomic sequences. Importantly, there seem to be no specific mechanisms for SINE elimination from the genome (in contrast to some DNA transposons) (Shedlock and Okada, 2000).

SINE insertions are not entirely random and thus can occur independently in the same locus (one example of this kind was reported by Hillis, 1999). Indeed, several independent insertions of Alu into paralogous regions of primate genomes were reported (Roy-Engel *et al.*, 2002b). Although these "parallel insertions" did not occur at identical sites, there is one example of ID integration into precisely the same site in rats (Rothenburg *et al.*, 2002). Still, such events are extremely rare and SINE insertions can be considered as practically homoplasia-free.

All these properties make SINEs a convenient tool for phylogenetic analysis. Thus, Okada and co-workers used the presence of SINEs in specific genomic loci for such analysis and developed a polymerase chain reaction (PCR)-based technique to resolve evolutionary relationships among salmonid fishes (Murata *et al.*, 1993) and to demonstrate amazing relations of even-toed ungulates to whales rather than to other ungulate orders (Nikaido *et al.*, 1999; Shimamura *et al.*, 1997).

Likewise, SINE insertional polymorphism is used in population genetics and the relations between human populations are being resolved with increasing accuracy (Antunez-de-Mayolo *et al.*, 2002; Batzer *et al.*, 1994; Watkins *et al.*, 2003).

Another way to use SINEs in systematics is to consider the presence or absence of a particular SINE family as a character (Fig. 6). Different SINE families are thought to have originated independently from the available cellular and LINE modules (see Section IV.A). Hence, independent SINE formation from the same modules (e.g., tRNA, core, and LINE-derived regions) is theoretically possible in a lineage with the same LINE being active; however, the same length and arrangement of the modules and the same pathway of the subsequent "fine-tuning" of their sequence should be an incredibly improbable event and has never been observed so far. The

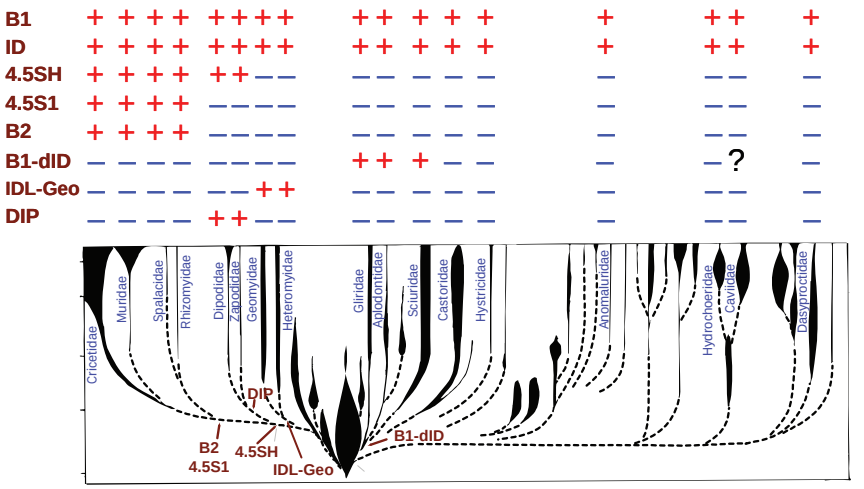


FIG. 6 SINE families as a phylogenetic marker. Evolutionary tree of rodent families and the distribution of SINEs and SINE-derived 4.5SH and 4.5SI RNAs. Plus and minus signs indicate the presence and absence of a SINE family, respectively; arrows show putative time of emergence of particular SINEs. (See also color insert.)

presence of a tRNA-unrelated sequence of unknown origin unique for each SINE family makes this approach particularly reliable. Thus, it was used to resolve evolutionary relationships between dormice (Gliridae) and myomorph rodents (Kramerov et al., 1999), among carnivores (Vassetzky and Kramerov, 2002), as well as among rodents, primates, and related mammalian taxa (Vassetzky et al., 2003). The monophyly of the clade Afrotheria, comprising elephants, hyraxes, sea cows, aardvarks, golden moles, tenrecs, and elephant shrews, was also confirmed by this approach (Nikaido et al., 2003).

At the same time, there are some limitations to using SINEs in phylogenetic analysis. First of all, it is applicable only to organisms with SINEs. Although this approach is perfect for tree topology, it can hardly be used for branch length evaluation (because of the sporadic rather than regular pattern of SINE insertions). Because SINE families have a finite lifetime, only the periods of their activity can be resolved by SINE analysis. In the case of unusually fast speciation events (e.g., for African cichlid fishes), SINE copies may not be fixed and their analysis can be difficult to interpret (Terai et al., 2003). Overall, SINEs provide an additional and reliable source of data for phylogenetic analysis.

## V. Functions of SINEs for the Host Genome

In the 1970s, before the first short retroposons were cloned and sequenced, extensive studies of repeated DNA suggested that the interspersed repeats are the main regulatory regions of genes (Davidson and Britten, 1979; Georgiev, 1969). Further studies introduced an antipodal concept of "selfish DNA" that propagates in the genome and makes no specific contribution to the phenotype (Doolittle and Sapienza, 1980; Orgel and Crick, 1980). Thus, selfish DNA resembles a not too harmful parasite. Most known short retroposons are consistent (or do not conflict) with this hypothesis.

Later the term "junk DNA" became popular; it implied that the genomic repeats are useless. At the same time, the original concept of selfish DNA suggested that, despite the absence of a specific function, it can be significant and beneficial for the organism. Orgel and Crick (1980) claimed that "it would be surprising if the host organism did not occasionally find some use for particular selfish DNA sequences" and, continuing the analogy to parasitism, "slightly harmful infestation may ultimately be transformed into a symbiosis."

Now we know that the impact of SINEs on eukaryotic genome involves many specific and general mechanisms, such as mutagenesis, regulation of gene activity, and genome expansion, which can shed light on their significance for the life and evolution of eukaryotes.

### A. Mechanisms of SINE-Mediated Mutagenesis

Owing to their capacity to integrate new copies in various genomic sites, SINEs can affect the functioning of genes. Although many SINE copies integrate into nonfunctional regions, other copies appear within or near genes and can affect their expression. Clearly, integration of a short retroposon into an exon should considerably disturb the structure and activity of the encoded protein. At present, we know of such events in humans and animals. SINE integration upstream of a gene into the transcription control region can affect gene expression, whereas its integration into an intron can disturb splicing and result in the formation of mRNA with, for example, extra nucleotide sequences. Insertional mutagenesis of Alu in humans is responsible for ~0.1% of genetic diseases; many examples of this kind have been provided by Deininger and Batzer (1999).

Another SINE-mediated mechanism of genetic changes is unequal homologous recombination. In this case, pairing of nonallelic SINEs is followed by breakage and rejoining of chromatid fragments, which gives rise to large-scale deletions or insertions (in the range of thousands to hundreds of

thousands of base pairs). Particular SINE pairs sometimes become recombination hot spots, which can cause genetic diseases (e.g., familial hypercholesterolemia). Preferential SINE-mediated recombination can occur in somatic cells and such events were recorded in association with acute myelogenous leukemia (Jeffs *et al.*, 2001). Moreover, analysis of segmental duplications in the human genome demonstrated Alu sequences at their boundaries in about one-third of cases (Bailey *et al.*, 2003). Overall, SINEs do not seem to be the only factor of such genetic events but make a great contribution to them; thus, ~0.3% of human genetic diseases result from Alu-mediated unequal recombination (Deininger and Batzer, 1999).

## B. Genome Shaping

Although integrations of short retroposons into a genome can be deleterious (e.g., the inability of humans to synthesize vitamin C is attributed to Alu insertion; Challem and Taylor, 1998) and even lethal, in some cases it can be creative. Below we summarize examples of this kind.

### 1. Appearance of New Genes

The most straightforward variant of creative impact of SINEs is probably genomic duplication resulting from SINE–SINE unequal crossover. Such an event can introduce a copy of a functional genomic element such as a gene, which can be further modified by evolution without disturbing the function of the original protein. Thus, Alu–Alu recombination introduced two different transmembrane proteins of erythrocytes in humans and apes, glycophorins B and E (Kudo and Fukuda, 1989; Rearden *et al.*, 1993). Duplication of the primordial human growth hormone gene mediated by Alu gave rise to an ancestor of human chorionic somatomammotropin (Hirt *et al.*, 1987). A similar event mediated by B2 elements flanking a lysozyme gene provided for the presence of two lysozyme genes with different functions and expression in mouse (Cross and Renkawitz, 1990).

SINEs can give rise to certain small cellular RNAs. Such RNAs have a relatively narrow distribution (Fig. 6) and the nucleotide sequence similarity suggests that rodent 4.5SH, 4.5SI, and BC1 RNAs originated from B1, B2, and ID, whereas primate BC200 descended from free left Alu monomer. BC1 and BC200 RNAs are synthesized almost exclusively in nervous tissue. BC1 RNA is a specific translation factor in dendrites (Wang *et al.*, 2002) and the knockout of its gene has some impact on mouse behavior (Lewejohann *et al.*, 2004). The role of 4.5SH and 4.5SI RNAs remains unclear; however, the conservation of their sequences points to their functional significance (Gogolevsky and Kramerov, 2002; Gogolevsky *et al.*, 2005). These RNAs

can be involved in translational control and/or cellular protection specific for myomorph rodents.

## 2. Delivery of Functional Sequences

Short retroposons can carry functionally active sequences, for example, binding sites for control factors. In such cases, SINE integration could confer a different regulation pattern and even change the function of proteins.

***a. Impact on Transcription*** Studies on individual genes demonstrated that some SINE copies are involved in the gene transcription control as *cis*-positive (enhancer) and *cis*-negative elements. For instance, Alus seem to be involved in transcription of human $\epsilon$-globin (Wu *et al.*, 1990), adenosine deaminase (Aronow *et al.*, 1992), BRCA1 (Norris *et al.*, 1995), BRCA2 (Sharan *et al.*, 1999), type 1 deiodinase (Zhang *et al.*, 1998), glycoprotein hormone $\alpha$ subunit (Scofield *et al.*, 2000), and nicotinic acetylcholine receptor $\alpha_3$ and $\alpha_6$ subunits (Ebihara *et al.*, 2002; Fornasari *et al.*, 1997). Sometimes different parts of the same SINE copy can serve as an enhancer and a silencer, as in the Alu upstream of the gene encoding the $\gamma$ chain of the IgE receptor ($Fc_\epsilon RI$) (Brini *et al.*, 1993). In other cases SINEs can be involved in complex regulatory mechanisms; for instance, integration of a second Alu upstream of the CD8 $\alpha$ gene made possible the formation of a cruciform structure that suppressed a T-cell-specific enhancer in the first Alu copy (Hambor *et al.*, 1993; Hanke *et al.*, 1995).

Analysis of SINE nucleotide sequences reveals binding sites for proteins involved in the regulation of RNA polymerase II-mediated transcription (Tomilin, 1999). For instance, Lyf, MEP-1, PPAR, YY1, SIF, RAR/RXR, CREB/ATF, and Sp1 binding sites can be found in most or at least many Alu copies, whereas the sites for transcription factors AP1, ER, REF/HAP1, the GATA family, and the bHLHzip family are found in a small fraction ($<10\%$) of Alu copies. Thus, involvement of SINEs in such regulation of transcription can require mutations in their integrated copies. This involvement of SINEs in transcriptional regulation was documented for many cases. For instance, the above-mentioned T-cell-specific enhancer of the CD8 $\alpha$ gene consists of Lyf-1 and GATA-3 sites in the first Alu (Hambor *et al.*, 1993). SINEs (human Alu and mouse B1) proved to be involved in the distribution of the binding sites for Pax6, a transcription factor critical for the development of the eye, pancreas, and central nervous system (Zhou *et al.*, 2002). The presence of binding sites for the superfamily of nuclear hormone receptors (including retinoic acid receptor) was convincingly demonstrated (Babich *et al.*, 1999; Norris *et al.*, 1995; Piedrafita *et al.*, 1996; Vansant and Reynolds, 1995). Alu likely played an important role in the distribution of

binding sites for nuclear hormone receptors and some other transcription factors in primates.

SINEs can serve not only as enhancers and silencers but also as promoters for RNA polymerase II. For instance, $\theta_1$-globin, an $\alpha$-globin-like gene, remained silent (e.g., in prosimians and rabbit) until Alu integration introduced a CCAAT box and restored its expression in higher primates (Kim *et al.*, 1989). Likewise, mouse B2 provided a functional promoter for *Lama3* gene (Ferrigno *et al.*, 2001).

SINEs can also mediate gene transcription by a mechanism called transcriptional interference. For instance, active Alu transcription upstream of the $\epsilon$-globin gene suppresses its low-level transcription from an alternative promoter, which allows $\epsilon$-globin transcription from the major promoter (Wu *et al.*, 1990). A similar effect was described for the human keratin 18 gene: violation of transcription of an upstream Alu made this gene sensitive to transcriptional interference (Willoughby *et al.*, 2000). Amazingly, a mouse homolog of this gene has a B1 instead of Alu in a similar upstream position (Ichinose *et al.*, 1988). Such an "insulator" effect of SINEs can underlie a considerable increase in the level of LacZ expression when the gene is flanked with B1 and B2 SINEs (Kang *et al.*, 2000). Thus, utilization of SINEs with active Pol III promoters as insulators can be an attractive mechanism for the genome.

It is also possible that SINEs can affect transcription of neighboring genes through their methylation. Although DNA methylation is believed to lock genes in a silent state (at least in vertebrates), another concept considers methylation as a defense mechanism against genomic repeats and SINEs in particular (Yoder *et al.*, 1997). Indeed, most SINEs are heavily methylated in almost all cell types, which represses their transcription and, hence, reproduction (see Sections III.B.1.d and III.B.2). Moreover, new SINE copies can attract *de novo* methylation to the neighboring regions, as was demonstrated for plant SINE S1 (Arnaud *et al.*, 2000) and mouse B1 (Yates *et al.*, 1999). Still, it remains unclear whether methylation is used to repress SINEs or whether they are used as factors of methylation and silencing (or both).

***b. Posttranscriptional Impact*** Many protein-coding sequences include SINEs or their fragments. Most SINE copies carry many stop codons; at the same time, antisense Alu has fewer stops than its sense sequence and, thus, has a better chance to be included into an ORF. SINE integration into the translated part of a gene commonly results in a shortened (because of stop codons) and nonfunctional protein. However, there exists an elegant mechanism of SINE introduction into an ORF.

The presence of sequences similar to donor and acceptor splice sites in SINE sequences was noted long ago (Krayev *et al.*, 1980, 1982). For instance,

antisense Alu sequence includes 12 and 7 potential donor and acceptor sites, respectively (Makalowski *et al.*, 1994; Sorek *et al.*, 2002). Thus, SINE integration into an intron can alter the splicing pattern and this process requires only minimal base substitutions (Sorek *et al.*, 2004). For instance, mouse haplotype *H-2*$^k$ originated as a result of B2-mediated alternative splicing (Pattanakitsakul *et al.*, 1992). Alu-mediated alternative splicing introduced a minor soluble form of hydrophobic human decay-accelerating factor (Caras *et al.*, 1987). Bovine SINE CHR-I made possible alternative splicing of prostaglandin $E_2$ receptor EP3, whose isoforms are involved in the activation of various signaling pathways. In this case, the translated SINE sequences present in the resulting protein seem to be responsible for its different functional activity (Shimamura *et al.*, 1998). Alternative splicing is an attractive evolutionary mechanism because the original protein is still synthesized and is thus less deleterious, and the contribution of SINEs to this process cannot be overstated (Kreahling and Graveley, 2004): for instance, at least 5% of all human alternatively spliced proteins are derived from Alu (Sorek *et al.*, 2002).

Many SINE families also contain the potential polyadenylation signal (AATAAA) in their tail and some of them are indeed polyadenylated (see Section III.A.2). Although such a site is not sufficient for polyadenylation in the wrong context, SINEs can be a source of an alternative functional polyadenylation signal. Thus, such signals originated from B2 in mouse cytochrome *P*-450 (Ryskov *et al.*, 1983), D and L genes of the major histocompatibility complex class I (Kress *et al.*, 1984), the $\gamma$ subunit of muscle phosphorylase kinase (Maichele *et al.*, 1993), leukemia inhibitory factor receptor (Michel *et al.*, 1997), and rat glutathione *S*-transferase (Rothkopf *et al.*, 1986), whereas C SINE provided active polyadenylation signals to rabbit isozyme 4 of cytochrome *P*-450, the major apoprotein of pulmonary surfactant, and a major histocompatibility complex gene (Krane and Hardison, 1990).

## C. *Trans* Effects of SINEs

All functions of SINEs previously considered affect DNA or RNA molecules where they reside (in *cis*), but short retroposons can have *trans* activities as well. The level of Alu and other SINEs considerably increases after heat shock, viral infection, and treatment with cellular poisons (see Section III.B.2). The diverse range of SINEs involved suggests that this response is evolutionarily conserved and may have biological significance.

Several mechanisms were proposed to explain it. High levels of Alu RNA proved to inhibit kinase PKR involved in antiviral repression of translation (Chu *et al.*, 1998); thus, Alu transcription is a possible inducer of protein

synthesis. Later the same group demonstrated PKR-independent stimulation of reporter translation (but not total protein translation) by RNA of human Alu and mouse B1 and B2 (Rubin *et al.*, 2002). Finally, B2 (but not B1) RNA proved to bind RNA polymerase II and to repress its activity *in vitro*, which can mediate cellular heat shock response (Allen *et al.*, 2004).

Overall, to date the available data allow no definite conclusion concerning whether SINE transcription during stress points to their beneficial cellular function or just reflects stress-induced perturbations in the cell.

Experiments with B2-targeted ribozymes suggest a direct role for this SINE in the cell. Rat hepatocytes with the active ribozyme demonstrated decelerated growth rate and prolonged S phase, which was attributed to degradation of B2 RNA (Crone *et al.*, 1999) (although it could also result from degradation of certain mRNAs containing B2).

Finally, indirect evidence points to the involvement of SINEs in the higher order structure of DNA, for example, matrix/scaffold formation (Jackson *et al.*, 1996), heterochromatin nucleation (directly, similar to other repeats [Hsieh and Fire, 2000], or through DNA methylation they can attract; see Section V.B.2.a), or sister chromatid cohesion (Hakimi *et al.*, 2002).

## VI. Concluding Remarks

Only now are we starting to explore the genome as an integrated system, and the role of SINEs should become apparent when we better understand how this system functions. The accumulation of data on the involvement of SINEs in the genome functioning is gradually making the concept of selfish DNA less popular. The integral function hypothesis proposed instead considers genomic repeats as flexible multitask units of the genome system as important as regulatory RNAs or introns (Shapiro, 1999; von Sternberg, 2002). This concept assumes that SINEs are integral genomic/epigenetic tools, which is hardly compatible with their absence in at least some higher organisms.

It seems probable that the evolution of our views on the significance of SINEs from selfish to functional is a brief and rapid recapitulation of their phylogenetic history. SINEs emerged as genomic parasites and gradually invaded the genomes of most higher eukaryotes, but later became an integral part of the genome and were used for the benefit of the organism.

### Acknowledgments

This research was supported by the Russian Foundation for Basic Research (project nos. 05–04–49553 and 02–04–48644).

## References

Adeniyi-Jones, S., and Zasloff, M. (1985). Transcription, processing and nuclear transport of a B1 Alu RNA species complementary to an intron of the murine α-fetoprotein gene. *Nature* **317,** 81–84.

Aleman, C., Roy-Engel, A. M., Shaikh, T. H., and Deininger, P. L. (2000). *Cis*-acting influences on Alu RNA levels. *Nucleic Acids Res.* **28,** 4755–4761.

Allen, T. A., Von Kaenel, S., Goodrich, J. A., and Kugel, J. F. (2004). The SINE-encoded mouse B2 RNA represses mRNA transcription in response to heat shock. *Nat. Struct. Mol. Biol.* **11,** 816–821.

Antunez-de-Mayolo, G., Antunez-de-Mayolo, A., Antunez-de-Mayolo, P., Papiha, S. S., Hammer, M., Yunis, J. J., Yunis, E. J., Damodaran, C., Martinez de Pancorbo, M., Caeiro, J. L., Puzyrev, V. P., and Herrera, R. J. (2002). Phylogenetics of worldwide human populations as determined by polymorphic Alu insertions. *Electrophoresis* **23,** 3346–3356.

Arnaud, P., Goubely, C., Pelissier, T., and Deragon, J. M. (2000). SINE retroposons can be used *in vivo* as nucleation centers for *de novo* methylation. *Mol. Cell. Biol.* **20,** 3434–3441.

Arnaud, P., Yukawa, Y., Lavie, L., Pelissier, T., Sugiura, M., and Deragon, J. M. (2001). Analysis of the SINE *S1* Pol III promoter from *Brassica*: Impact of methylation and influence of external sequences. *Plant J.* **26,** 295–305.

Aronow, B. J., Silbiger, R. N., Dusing, M. R., Stock, J. L., Yager, K. L., Potter, S. S., Hutton, J. J., and Wiginton, D. A. (1992). Functional analysis of the human adenosine deaminase gene thymic regulatory region and its ability to generate position-independent transgene expression. *Mol. Cell. Biol.* **12,** 4170–4185.

Baba, S., Kajikawa, M., Okada, N., and Kawai, G. (2004). Solution structure of an RNA stem–loop derived from the 3′ conserved region of eel LINE *UnaL2*. *RNA* **10,** 1380–1387.

Babich, V., Aksenov, N., Alexeenko, V., Oei, S. L., Buchlow, G., and Tomilin, N. (1999). Association of some potential hormone response elements in human genes with the Alu family repeats. *Gene* **239,** 341–349.

Bachvarova, R. (1988). Small B2 RNAs in mouse oocytes, embryos, and somatic tissues. *Dev. Biol.* **130,** 513–523.

Bailey, J. A., Liu, G., and Eichler, E. E. (2003). An Alu transposition model for the origin and expansion of human segmental duplications. *Am. J. Hum. Genet.* **73,** 823–834.

Batzer, M. A., and Deininger, P. L. (2002). Alu repeats and human genomic diversity. *Nat. Rev. Genet.* **3,** 370–379.

Batzer, M. A., Kilroy, G. E., Richard, P. E., Shaikh, T. H., Desselle, T. D., Hoppens, C. L., and Deininger, P. L. (1990). Structure and variability of recently inserted Alu family members. *Nucleic Acids Res.* **18,** 6793–6798.

Batzer, M. A., Stoneking, M., Alegria-Hartman, M., Bazan, H., Kass, D. H., Shaikh, T. H., Novick, G. E., Ioannou, P. A., Scheer, W. D., Herrera, R. J., and Deininger, P. L. (1994). African origin of human-specific polymorphic Alu insertions. *Proc. Natl. Acad. Sci. USA* **91,** 12288–12292.

Bestor, T. H. (1990). DNA methylation: Evolution of a bacterial immune function into a regulator of gene expression and genome structure in higher eukaryotes. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **326,** 179–187.

Bhattacharya, R., Perumal, K., Sinha, K., Maraia, R., and Reddy, R. (2002). Methylphosphate cap structure in small RNAs reduces the affinity of RNAs to La protein. *Gene Expr.* **10,** 243–253.

Bird, A. P. (1980). DNA methylation and the frequency of CpG in animal DNA. *Nucleic Acids Res.* **8,** 1499–1504.

Bird, A. (1997). Does DNA methylation control transposition of selfish elements in the germline? *Trends Genet.* **13,** 469–472.

Bladon, T. S., Fregeau, C. J., and McBurney, M. W. (1990). Synthesis and processing of small B2 transcripts in mouse embryonal carcinoma cells. *Mol. Cell. Biol.* **10,** 4058–4067.

Boeke, J. D. (1997). LINEs and Alus: The polyA connection. *Nat. Genet.* **16,** 6–7.

Borodulina, O. R., and Kramerov, D. A. (2001). Short interspersed elements (SINEs) from insectivores: Two classes of mammalia SINEs distinguished by A-rich tail structure. *Mamm. Genome* **12,** 779–786.

Borodulina, O. R., and Kramerov, D. A. (2005). PCR-based approach to SINE isolation: Simple and complex SINEs. *Gene* **349,** 197–205.

Bovia, F., Wolff, N., Ryser, S., and Strub, K. (1997). The SRP9/14 subunit of the human signal recognition particle binds to a variety of Alu-like RNAs and with higher affinity than its mouse homolog. *Nucleic Acids Res.* **25,** 318–326.

Branciforte, D., and Martin, S. L. (1994). Developmental and cell type specificity of LINE-1 expression in mouse testis: Implications for transposition. *Mol. Cell. Biol.* **14,** 2584–2592.

Brini, A. T., Lee, G. M., and Kinet, J. P. (1993). Involvement of Alu sequences in the cell-specific regulation of transcription of the $\gamma$ chain of Fc and T cell receptors. *J. Biol. Chem.* **268,** 1355–1361.

Buzdin, A., Ustyugova, S., Gogvadze, E., Vinogradova, T., Lebedev, Y., and Sverdlov, E. (2002). A new family of chimeric retrotranscripts formed by a full copy of U6 small nuclear RNA fused to the 3′ terminus of l1. *Genomics* **80,** 402–406.

Caras, I. W., Davitz, M. A., Rhee, L., Weddell, G., Martin, D. W., Jr., and Nussenzweig, V. (1987). Cloning of decay-accelerating factor suggests novel use of splicing to generate two proteins. *Nature* **325,** 545–549.

Challem, J. J., and Taylor, E. W. (1998). Retroviruses, ascorbate, and mutations, in the evolution of *Homo sapiens*. *Free Radic. Biol. Med.* **25,** 130–132.

Chang, D. Y., Hsu, K., and Maraia, R. J. (1996). Monomeric scAlu and nascent dimeric Alu RNAs induced by adenovirus are assembled into SRP9/14-containing RNPs in HeLa cells. *Nucleic Acids Res.* **24,** 4165–4170.

Chen, Y., Sinha, K., Perumal, K., Gu, J., and Reddy, R. (1998). Accurate 3′ end processing and adenylation of human signal recognition particle RNA and Alu RNA *in vitro*. *J. Biol. Chem.* **273,** 35023–35031.

Chesnokov, I., and Schmid, C. W. (1996). Flanking sequences of an Alu source stimulate transcription *in vitro* by interacting with sequence-specific transcription factors. *J. Mol. Evol.* **42,** 30–36.

Chesnokov, I., Chu, W. M., Botchan, M. R., and Schmid, C. W. (1996). p53 inhibits RNA polymerase III-directed transcription in a promoter-dependent manner. *Mol. Cell. Biol.* **16,** 7084–7088.

Chu, W. M., Liu, W. M., and Schmid, C. W. (1995). RNA polymerase III promoter and terminator elements affect Alu RNA expression. *Nucleic Acids Res.* **23,** 1750–1757.

Chu, W. M., Ballard, R. E., and Schmid, C. W. (1997). Palindromic sequences preceding the terminator increase polymerase III template activity. *Nucleic Acids Res.* **25,** 2077–2082.

Chu, W. M., Ballard, R., Carpick, B. W., Williams, B. R., and Schmid, C. W. (1998). Potential Alu function: Regulation of the activity of double-stranded RNA-activated kinase PKR. *Mol. Cell. Biol.* **18,** 58–68.

Churakov, G., Smit, A. F., Brosius, J., and Schmitz, J. (2005). A novel abundant family of retroposed elements (DAS-SINEs) in the nine-banded armadillo (*Dasypus novemcinctus*). *Mol. Biol. Evol.* **22,** 886–893.

Coffin, J. M., Hughes, S. H., and Varmus, H. (1997). ''Retroviruses.'' Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Cooper, D. M., Schimenti, K. J., and Schimenti, J. C. (1998). Factors affecting ectopic gene conversion in mice. *Mamm. Genome* **9,** 355–360.

Cost, G. J., Golding, A., Schlissel, M. S., and Boeke, J. D. (2001). Target DNA chromatinization modulates nicking by L1 endonuclease. *Nucleic Acids Res.* **29,** 573–577.

Cost, G. J., Feng, Q., Jacquier, A., and Boeke, J. D. (2002). Human *L1* element target-primed reverse transcription *in vitro*. *EMBO J.* **21,** 5899–5910.

Crone, T. M., Schalles, S. L., Benedict, C. M., Pan, W., Ren, L., Loy, S. E., Isom, H., and Clawson, G. A. (1999). Growth inhibition by a triple ribozyme targeted to repetitive B2 transcripts. *Hepatology* **29,** 1114–1123.

Cross, M., and Renkawitz, R. (1990). Repetitive sequence involvement in the duplication and divergence of mouse lysozyme genes. *EMBO J.* **9,** 1283–1288.

Cullen, B. R. (2003). Nuclear RNA export. *J. Cell Sci.* **116,** 587–597.

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.

Davidson, E. H., and Britten, R. J. (1979). Regulation of gene expression: Possible role of repetitive sequences. *Science* **204,** 1052–1059.

Deininger, P. L., and Batzer, M. A. (1999). Alu repeats and human disease. *Mol. Genet. Metab.* **67,** 183–193.

Deininger, P. L., and Batzer, M. A. (2002). Mammalian retroelements. *Genome Res.* **12,** 1455–1465.

Deininger, P. L., Jolly, D. J., Rubin, C. M., Friedmann, T., and Schmid, C. W. (1981). Base sequence studies of 300 nucleotide renatured repeated human DNA clones. *J. Mol. Biol.* **151,** 17–33.

Devos, K. M., Brown, J. K., and Bennetzen, J. L. (2002). Genome size reduction through illegitimate recombination counteracts genome expansion in *Arabidopsis*. *Genome Res.* **12,** 1075–1079.

Dewannieux, M., Esnault, C., and Heidmann, T. (2003). LINE-mediated retrotransposition of marked Alu sequences. *Nat. Genet.* **35,** 41–48.

Doolittle, W. F., and Sapienza, C. (1980). Selfish genes, the phenotype paradigm and genome evolution. *Nature* **284,** 601–603.

Ebihara, M., Ohba, H., Ohno, S. I., and Yoshikawa, T. (2002). Genomic organization and promoter analysis of the human nicotinic acetylcholine receptor α6 subunit (CHNRA6) gene: Alu and other elements direct transcriptional repression. *Gene* **298,** 101–108.

Eickbush, D. G., Luan, D. D., and Eickbush, T. H. (2000). Integration of *Bombyx mori R2* sequences into the 28S ribosomal RNA genes of *Drosophila melanogaster*. *Mol. Cell. Biol.* **20,** 213–223.

Elder, J. T., Pan, J., Duncan, C. H., and Weissman, S. M. (1981). Transcriptional analysis of interspersed repetitive polymerase III transcription units in human DNA. *Nucleic Acids Res.* **9,** 1171–1189.

Englander, E. W., Wolffe, A. P., and Howard, B. H. (1993). Nucleosome interactions with a human Alu element: Transcriptional repression and effects of template methylation. *J. Biol. Chem.* **268,** 19565–19573.

Esnault, C., Maestre, J., and Heidmann, T. (2000). Huma LINE retrotransposons generate processed pseudogenes. *Nat. Genet.* **24,** 363–367.

Feng, Q., Moran, J. V., Kazazian, H. H., Jr., and Boeke, J. D. (1996). Human *L1* retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell* **87,** 905–916.

Ferrigno, O., Virolle, T., Djabari, Z., Ortonne, J. P., White, R. J., and Aberdam, D. (2001). Transposable B2 SINE elements can provide mobile RNA polymerase II promoters. *Nat. Genet.* **28,** 77–81.

Fornace, A. J., Jr., and Mitchell, J. B. (1986). Induction of B2 RNA polymerase III transcription by heat shock: Enrichment for heat shock induced sequences in rodent cells by hybridization subtraction. *Nucleic Acids Res.* **14,** 5793–5811.

Fornasari, D., Battaglioli, E., Flora, A., Terzano, S., and Clementi, F. (1997). Structural and functional characterization of the human α3 nicotinic subunit gene promoter. *Mol. Pharmacol.* **51,** 250–261.

Fuhrman, S. A., Deininger, P. L., La Porte, P., Friedmann, T., and Geiduschek, E. P. (1981). Analysis of transcription of the human Alu family ubiquitous repeating element by eukaryotic RNA polymerase III. *Nucleic Acids Res.* **9,** 6439–6456.

Gadal, O., Strauss, D., Kessl, J., Trumpower, B., Tollervey, D., and Hurt, E. (2001). Nuclear export of 60S ribosomal subunits depends on Xpo1p and requires a nuclear export sequence-containing factor, Nmd3p, that associates with the large subunit protein Rpl10p. *Mol. Cell. Biol.* **21,** 3405–3415.

Georgiev, G. P. (1969). On the structural organization of operon and the regulation of RNA synthesis in animal cells. *J. Theor. Biol.* **25,** 473–490.

Gilbert, N., and Labuda, D. (1999). CORE-SINEs: Eukaryotic short interspersed retroposing elements with common sequence motifs. *Proc. Natl. Acad. Sci. USA* **96,** 2869–2874.

Gilbert, N., and Labuda, D. (2000). Evolutionary inventions and continuity of CORE-SINEs in mammals. *J. Mol. Biol.* **298,** 365–377.

Gogolevsky, I. K., and Kramerov, D. A. (2002). Evolutionary history of *4.5SI* RNA and indication that it is functional. *J. Mol. Evol.* **54,** 354–364.

Gogolevsky, I. K., Koval, A. P., and Kramerov, D. A. (2005). Evolutionary history of *4.5SH* RNA. *Mol. Biol. Evol.* **22,** 1546–1554.

Goodier, J. L., and Maraia, R. J. (1998). Terminator-specific recycling of a B1–Alu transcription complex by RNA polymerase III is mediated by the RNA terminus-binding protein La. *J. Biol. Chem.* **273,** 26110–26116.

Goodier, J. L., Ostertag, E. M., and Kazazian, H. H., Jr. (2000). Transduction of 3′-flanking sequences is common in L1 retrotransposition. *Hum. Mol. Genet.* **9,** 653–657.

Greally, J. M. (2002). Short interspersed transposable elements (SINEs) are excluded from imprinted regions in the human genome. *Proc. Natl. Acad. Sci. USA* **99,** 327–332.

Grigoryan, M. S., Kramerov, D. A., Tulchinsky, E. M., Revasova, E. S., and Lukanidin, E. M. (1985). Activation of putative transposition intermediate formation in tumor cells. *EMBO J.* **4,** 2209–2215.

Hakimi, M. A., Bochar, D. A., Schmiesing, J. A., Dong, Y., Barak, O. G., Speicher, D. W., Yokomori, K., and Shiekhattar, R. (2002). A chromatin remodelling complex that loads cohesin onto human chromosomes. *Nature* **418,** 994–998.

Hamada, M., Huang, Y., Lowe, T. M., and Maraia, R. J. (2001). Widespread use of TATA elements in the core promoters for RNA polymerases III, II, and I in fission yeast. *Mol. Cell. Biol.* **21,** 6870–6881.

Hambor, J. E., Mennone, J., Coon, M. E., Hanke, J. H., and Kavathas, P. (1993). Identification and characterization of an Alu-containing, T-cell-specific enhancer located in the last intron of the human CD8 α gene. *Mol. Cell. Biol.* **13,** 7056–7070.

Hanke, J. H., Hambor, J. E., and Kavathas, P. (1995). Repetitive Alu elements form a cruciform structure that regulates the function of the human CD8 α T cell-specific enhancer. *J. Mol. Biol.* **246,** 63–73.

Hasse, A., and Schulz, W. A. (1994). Enhancement of reporter gene *de novo* methylation by DNA fragments from the α-fetoprotein control region. *J. Biol. Chem.* **269,** 1821–1826.

Haynes, S. R., and Jelinek, W. R. (1981). Low molecular weight RNAs transcribed *in vitro* by RNA polymerase III from Alu-type dispersed repeats in Chinese hamster DNA are also found *in vivo*. *Proc. Natl. Acad. Sci. USA* **78,** 6130–6134.

Haynes, S. R., Toomey, T. P., Leinwand, L., and Jelinek, W. R. (1981). The Chinese hamster Alu-equivalent sequence: A conserved highly repetitious, interspersed deoxyribonucleic acid sequence in mammals has a structure suggestive of a transposable element. *Mol. Cell. Biol.* **1,** 573–583.

He, X. P., Bataille, N., and Fried, H. M. (1994). Nuclear export of signal recognition particle RNA is a facilitated process that involves the Alu sequence domain. *J. Cell Sci.* **107,** 903–912.

Hellmann-Blumberg, U., Hintz, M. F., Gatewood, J. M., and Schmid, C. W. (1993). Developmental differences in methylation of human Alu repeats. *Mol. Cell. Biol.* **13,** 4523–4530.

Hillis, D. M. (1999). SINEs of the perfect character. *Proc. Natl. Acad. Sci. USA* **96,** 9979–9981.

Hirt, H., Kimelman, J., Birnbaum, M. J., Chen, E. Y., Seeburg, P. H., Eberhardt, N. L., and Barta, A. (1987). The human growth hormone gene locus: Structure, evolution, and allelic variations. *DNA* **6,** 59–70.

Hsieh, J., and Fire, A. (2000). Recognition and silencing of repeated DNA. *Annu. Rev. Genet.* **34,** 187–204.

Hsu, K., Chang, D. Y., and Maraia, R. J. (1995). Human signal recognition particle (SRP) Alu-associated protein also binds Alu interspersed repeat sequence RNAs: Characterization of human SRP9. *J. Biol. Chem.* **270,** 10179–10186.

Ichinose, Y., Morita, T., Zhang, F. Y., Srimahasongcram, S., Tondella, M. L., Matsumoto, M., Nozaki, M., and Matsushiro, A. (1988). Nucleotide sequence and structure of the mouse cytokeratin endoB gene. *Gene* **70,** 85–95.

International Human, Genome Sequencing, Consortium (2001). Initial sequencing and analysis of the human genome. *Nature* **409,** 860–921.

Izaurralde, E., Lewis, J., Gamberi, C., Jarmolowski, A., McGuigan, C., and Mattaj, I. W. (1995). A cap-binding protein complex mediating U snRNA export. *Nature* **376,** 709–712.

Jackson, D. A., Bartlett, J., and Cook, P. R. (1996). Sequences attaching loops of nuclear and mitochondrial DNA to underlying structures in human cells: The role of transcription units. *Nucleic Acids Res.* **24,** 1212–1219.

Jaenisch, R. (1997). DNA methylation and imprinting: Why bother? *Trends Genet.* **13,** 323–329.

Jagadeeswaran, P., Forget, B. G., and Weissman, S. M. (1981). Short interspersed repetitive DNA elements in eucaryotes: Transposable DNA elements generated by reverse transcription of RNA Pol III transcripts? *Cell* **26,** 141–142.

Jang, K. L., and Latchman, D. S. (1989). HSV infection induces increased transcription of Alu repeated sequences by RNA polymerase III. *FEBS Lett.* **258,** 255–258.

Jang, K. L., and Latchman, D. S. (1992). The herpes simplex virus immediate-early protein ICP27 stimulates the transcription of cellular Alu repeated sequences by increasing the activity of transcription factor TFIIIC. *Biochem. J.* **284,** 667–673.

Jeffs, A. R., Wells, E., and Morris, C. M. (2001). Nonrandom distribution of interspersed repeat elements in the BCR and ABL1 genes and its relation to breakpoint cluster regions. *Genes Chromosomes Cancer* **32,** 144–154.

Jelinek, W. R., Toomey, T. P., Leinwand, L., Duncan, C. H., Biro, P. A., Choudary, P. V., Weissman, S. M., Rubin, C. M., Houck, C. M., Deininger, P. L., and Schmid, C. W. (1980). Ubiquitous, interspersed repeated sequences in mammalian genomes. *Proc. Natl. Acad. Sci. USA* **77,** 1398–1402.

Jurka, J. (1997). Sequence patterns indicate an enzymatic involvement in integration of mammalian retroposons. *Proc. Natl. Acad. Sci. USA* **94,** 1872–1877.

Jurka, J., Krnjajic, M., Kapitonov, V. V., Stenger, J. E., and Kokhanyy, O. (2002). Active Alu elements are passed primarily through paternal germlines. *Theor. Popul. Biol.* **61,** 519–530.

Kajikawa, M., and Okada, N. (2002). LINEs mobilize SINEs in the eel through a shared 3′ sequence. *Cell* **111,** 433–444.

Kang, Y. K., Park, J. S., Lee, C. S., Yeom, Y. I., Han, Y. M., Chung, A. S., and Lee, K. K. (2000). Effect of short interspersed element sequences on the integration and expression of a reporter gene in the preimplantation-stage mouse embryos. *Mol. Reprod. Dev.* **56,** 366–371.

Kapitonov, V. V., and Jurka, J. (1999). Molecular paleontology of transposable elements from *Arabidopsis thaliana*. *Genetica* **107,** 27–37.

Kapitonov, V. V., and Jurka, J. (2003). A novel class of SINE elements derived from 5S rRNA. *Mol. Biol. Evol.* **20,** 694–702.

Katz, R. A., and Skalka, A. M. (1990). Generation of diversity in retroviruses. *Annu. Rev. Genet.* **24,** 409–445.

Kim, J., and Deininger, P. L. (1996). Recent amplification of rat ID sequences. *J. Mol. Biol.* **261,** 322–327.

Kim, J., Kass, D. H., and Deininger, P. L. (1995). Transcription and processing of the rodent ID repeat family in germline and somatic cells. *Nucleic Acids Res.* **23,** 2245–2251.

Kim, J. H., Yu, C. Y., Bailey, A., Hardison, R., and Shen, C. K. (1989). Unique sequence organization and erythroid cell-specific nuclear factor-binding of mammalian $\theta_1$ globin promoters. *Nucleic Acids Res.* **17,** 5687–5700.

Kimura, R. H., Choudary, P. V., and Schmid, C. W. (1999). Silk worm Bm1 SINE RNA increases following cellular insults. *Nucleic Acids Res.* **27,** 3380–3387.

Kobayashi, S., and Anzai, K. (1998). An E-box sequence acts as a transcriptional activator for BC1 RNA expression by RNA polymerase III in the brain. *Biochem. Biophys. Res. Commun.* **245,** 59–63.

Kochanek, S., Renz, D., and Doerfler, W. (1993). DNA methylation in the Alu sequences of diploid and haploid primary human cells. *EMBO J.* **12,** 1141–1151.

Kramerov, D. A., and Vassetzky, N. S. (2001). Structure and origin of a novel dimeric retroposon B1–dID. *J. Mol. Evol.* **52,** 137–143.

Kramerov, D. A., Grigoryan, A. A., Ryskov, A. P., and Georgiev, G. P. (1979). Long double-stranded sequences (dsRNA-B) of nuclear pre-mRNA consist of a few highly abundant classes of sequences: Evidence from DNA cloning experiments. *Nucleic Acids Res.* **6,** 697–713.

Kramerov, D. A., Lekakh, I. V., Samarina, O. P., and Ryskov, A. P. (1982). The sequences homologous to major interspersed repeats B1 and B2 of mouse genome are present in mRNA and small cytoplasmic poly(A)$^+$ RNA. *Nucleic Acids Res.* **10,** 7477–7491.

Kramerov, D. A., Tillib, S. V., Ryskov, A. P., and Georgiev, G. P. (1985a). Nucleotide sequence of small polyadenylated B2 RNA. *Nucleic Acids Res.* **13,** 6423–6437.

Kramerov, D. A., Tillib, S. V., Lekakh, I. V., Ryskov, A. P., and Georgiev, G. P. (1985b). Biosynthesis and cytoplasmic distribution of small poly(A)-containing B2 RNA. *Biochim. Biophys. Acta* **824,** 85–98.

Kramerov, D. A., Tillib, S. V., Shumyatsky, G. P., and Georgiev, G. P. (1990). The most abundant nascent poly(A)$^+$ RNAs are transcribed by RNA polymerase III in murine tumor cells. *Nucleic Acids Res.* **18,** 4499–4506.

Kramerov, D., Vassetzky, N., and Serdobova, I. (1999). The evolutionary position of dormice (Gliridae) in Rodentia determined by a novel short retroposon. *Mol. Biol. Evol.* **16,** 715–717.

Krane, D. E., and Hardison, R. C. (1990). Short interspersed repeats in rabbit DNA can provide functional polyadenylation signals. *Mol. Biol. Evol.* **7,** 1–8.

Krayev, A. S., Kramerov, D. A., Skryabin, K. G., Ryskov, A. P., Bayev, A. A., and Georgiev, G. P. (1980). The nucleotide sequence of the ubiquitous repetitive DNA sequence B1 complementary to the most abundant class of mouse fold-back RNA. *Nucleic Acids Res.* **8,** 1201–1215.

Krayev, A. S., Markusheva, T. V., Kramerov, D. A., Ryskov, A. P., 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.

Kreahling, J., and Graveley, B. R. (2004). The origins and implications of Aluternative splicing. *Trends Genet.* **20,** 1–4.

Kress, M., Barra, Y., Seidman, J. G., Khoury, G., and Jay, G. (1984). Functional insertion of an Alu type 2 (B2 SINE) repetitive sequence in murine class I genes. *Science* **226,** 974–977.

Kudo, S., and Fukuda, M. (1989). Structural organization of glycophorin A and B genes: Glycophorin B gene evolved by homologous recombination at Alu repeat sequences. *Proc. Natl. Acad. Sci. USA* **86,** 4619–4623.

Labuda, D., and Striker, G. (1989). Sequence conservation in Alu evolution. *Nucleic Acids Res.* **17,** 2477–2491.

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

Lenoir, A., Cournoyer, B., Warwick, S., Picard, G., and Deragon, J. M. (1997). Evolution of SINE *S1* retroposons in Cruciferae plant species. *Mol. Biol. Evol.* **14,** 934–941.

Lenoir, A., Lavie, L., Prieto, J. L., Goubely, C., Cote, J. C., Pelissier, T., and Deragon, J. M. (2001). The evolutionary origin and genomic organization of SINEs in *Arabidopsis thaliana*. *Mol. Biol. Evol.* **18,** 2315–2322.

Lewejohann, L., Skryabin, B. V., Sachser, N., Prehn, C., Heiduschka, P., Thanos, S., Jordan, U., Dell' Omo, G., Vyssotski, A. L., Pleskacheva, M. G., Lipp, H. P., Tiedge, H., Brosius, J., and Prior, H. (2004). Role of a neuronal small non-messenger RNA: Behavioural alterations in *BC1* RNA-deleted mice. *Behav. Brain Res.* **154,** 273–289.

Li, T. H., and Schmid, C. W. (2004). Alu's dimeric consensus sequence destabilizes its transcripts. *Gene* **324,** 191–200.

Li, T., Spearow, J., Rubin, C. M., and Schmid, C. W. (1999). Physiological stresses increase mouse short interspersed element (SINE) RNA expression *in vivo*. *Gene* **239,** 367–372.

Li, T. H., Kim, C., Rubin, C. M., and Schmid, C. W. (2000). K562 cells implicate increased chromatin accessibility in Alu transcriptional activation. *Nucleic Acids Res.* **28,** 3031–3039.

Lin, Z., Nomura, O., Hayashi, T., Wada, Y., and Yasue, H. (2001). Characterization of a SINE species from vicuna and its distribution in animal species including the family Camelidae. *Mamm. Genome* **12,** 305–308.

Liu, W. M., Maraia, R. J., Rubin, C. M., and Schmid, C. W. (1994). Alu transcripts: Cytoplasmic localisation and regulation by DNA methylation. *Nucleic Acids Res.* **22,** 1087–1095.

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

Luan, D. D., and Eickbush, T. H. (1995). RNA template requirements for target DNA-primed reverse transcription by the R2 retrotransposable element. *Mol. Cell. Biol.* **15,** 3882–3891.

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.

Lund, E., and Dahlberg, J. E. (1998). Proofreading and aminoacylation of tRNAs before export from the nucleus. *Science* **282,** 2082–2085.

Maeda, N., Wu, C. I., Bliska, J., and Reneke, J. (1988). Molecular evolution of intergenic DNA in higher primates: Pattern of DNA changes, molecular clock, and evolution of repetitive sequences. *Mol. Biol. Evol.* **5,** 1–20.

Maichele, A. J., Farwell, N. J., and Chamberlain, J. S. (1993). A B2 repeat insertion generates alternate structures of the mouse muscle γ-phosphorylase kinase gene. *Genomics* **16,** 139–149.

Makalowski, W., Mitchell, G. A., and Labuda, D. (1994). Alu sequences in the coding regions of mRNA: A source of protein variability. *Trends Genet.* **10,** 188–193.

Malik, H. S., Burke, W. D., and Eickbush, T. H. (1999). The age and evolution of non-LTR retrotransposable elements. *Mol. Biol. Evol.* **16,** 793–805.

Maraia, R. J. (1995). "The Impact of Short Interspersed Elements (SINEs) on the Host Genome," R.G. Landes, Austin, TX.

Maraia, R. J. (2001). La protein and the trafficking of nascent RNA polymerase III transcripts. *J. Cell Biol.* **153,** F13–F18.

Maraia, R. J., Driscoll, C. T., Bilyeu, T., Hsu, K., and Darlington, G. J. (1993). Multiple dispersed loci produce small cytoplasmic Alu RNA. *Mol. Cell. Biol.* **13,** 4233–4241.

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.

Mathews, D. H., Banerjee, A. R., Luan, D. D., Eickbush, T. H., and Turner, D. H. (1997). Secondary structure model of the RNA recognized by the reverse transcriptase from the R2 retrotransposable element. *RNA* **3,** 1–16.

Matsumoto, K., Murakami, K., and Okada, N. (1984). Pseudouridylic modification of a 6S RNA transcribed *in vitro* from highly repetitive and transcribable (Hirt) sequences of salmon total DNA. *Biochem. Biophys. Res. Commun.* **124,** 514–522.

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.

McKinnon, R. D., Danielson, P., Brow, M. A., Bloom, F. E., and Sutcliffe, J. G. (1987). Expression of small cytoplasmic transcripts of the rat identifier element *in vivo* and in cultured cells. *Mol. Cell. Biol.* **7,** 2148–2154.

Michel, D., Chatelain, G., Mauduit, C., Benahmed, M., and Brun, G. (1997). Recent evolutionary acquisition of alternative pre-mRNA splicing and 3′ processing regulations induced by intronic B2 SINE insertion. *Nucleic Acids Res.* **25,** 3228–3234.

Miyamoto, M. M. (1999). Molecular systematics: Perfect SINEs of evolutionary history? *Curr. Biol.* **9,** R816–R819.

Mouse Genome, Sequencing Consortium (2002). Initial sequencing and comparative analysis of the mouse genome. *Nature* **420,** 520–562.

Muddashetty, R., Khanam, T., Kondrashov, A., Bundman, M., Iacoangeli, A., Kremerskothen, J., Duning, K., Barnekow, A., Huttenhofer, A., Tiedge, H., and Brosius, J. (2002). Poly(A)-binding protein is associated with neuronal BC1 and BC200 ribonucleoprotein particles. *J. Mol. Biol.* **321,** 433–445.

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.

Murphy, D., Brickell, P. M., Latchman, D. S., Willison, K., and Rigby, P. W. (1983). Transcripts regulated during normal embryonic development and oncogenic transformation share a repetitive element. *Cell* **35,** 865–871.

Muslimov, I. A., Lin, Y., Heller, M., Brosius, J., Zakeri, Z., and Tiedge, H. (2002). A small RNA in testis and brain: Implications for male germ cell development. *J. Cell Sci.* **115,** 1243–1250.

Nakielny, S., Fischer, U., Michael, W. M., and Dreyfuss, G. (1997). RNA transport. *Annu. Rev. Neurosci.* **20,** 269–301.

Negroni, M., and Buc, H. (2001). Mechanisms of retroviral recombination. *Annu. Rev. Genet.* **35,** 275–302.

Nikaido, M., Rooney, A. P., and Okada, N. (1999). Phylogenetic relationships among cetartiodactyls based on insertions of short and long interpersed elements: Hippopotamuses are the closest extant relatives of whales. *Proc. Natl. Acad. Sci. USA* **96,** 10261–10266.

Nikaido, M., Nishihara, H., Hukumoto, Y., and Okada, N. (2003). Ancient SINEs from African endemic mammals. *Mol. Biol. Evol.* **20,** 522–527.

Nomura, O., Lin, Z. H., Muladno, Wada, Y., and Yasue, H. (1998). A SINE species from hippopotamus and its distribution among animal species. *Mamm. Genome* **9,** 550–555.

Norris, J., Fan, D., Aleman, C., Marks, J. R., Futreal, P. A., Wiseman, R. W., Iglehart, J. D., Deininger, P. L., and McDonnell, D. P. (1995). Identification of a new subclass of Alu DNA repeats which can function as estrogen receptor-dependent transcriptional enhancers. *J. Biol. Chem.* **270,** 22777–22782.

Odom, G. L., Robichaux, J. L., and Deininger, P. L. (2004). Predicting mammalia SINE subfamily activity from A-tail length. *Mol. Biol. Evol.* **21,** 2140–2148.

Ogiwara, I., Miya, M., Ohshima, K., and Okada, N. (2002). V-SINEs: A new superfamily of vertebrate SINEs that are widespread in vertebrate genomes and retain a strongly conserved segment within each repetitive unit. *Genome Res.* **12,** 316–324.

Ohshima, K., Hamada, M., Terai, Y., and Okada, N. (1996). The 3′ ends of tRNA-derived short interspersed repetitive elements are derived from the 3′ ends of long interspersed repetitive elements. *Mol. Cell. Biol.* **16,** 3756–3764.

Ohshima, K., Hattori, M., Yada, T., Gojobori, T., Sakaki, Y., and Okada, N. (2003). Whole-genome screening indicates a possible burst of formation of processed pseudogenes and Alu repeats by particular L1 subfamilies in ancestral primates. *Genome Biol.* **4,** R74.

Okada, N., and Hamada, M. (1997). The 3′ ends of tRNA-derived SINEs originated from the 3′ ends of LINEs: A new example from the bovine genome. *J. Mol. Evol.* **44,** 52–56.

Okada, N., and Ohshima, K. (1995). Evolution of tRNA-derived SINEs. *In* "The Impact of Short Interspersed Elements (SINEs) on the Host Genome" (R. J. Maraia, Ed.), p. 61. R.G. Landes, Austin, TX.

Okada, N., Hamada, M., Ogiwara, I., and Ohshima, K. (1997). SINEs and LINEs share common 3′ sequences: A review. *Gene* **205,** 229–243.

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

Ostertag, E. M., and Kazazian, H. H., Jr. (2001a). Twin priming: A proposed mechanism for the creation of inversions in L1 retrotransposition. *Genome Res.* **11,** 2059–2065.

Ostertag, E. M., and Kazazian, H. H., Jr. (2001b). Biology of mammalian L1 retrotransposons. *Annu. Rev. Genet.* **35,** 501–538.

Ostertag, E. M., De Berardinis, R. J., Goodier, J. L., Zhang, Y., Yang, N., Gerton, G. L., and Kazazian, H. H., Jr. (2002). A mouse model of human L1 retrotransposition. *Nat. Genet.* **32,** 655–660.

Panning, B., and Smiley, J. R. (1993). Activation of RNA polymerase III transcription of human Alu repetitive elements by adenovirus type 5: Requirement for the E1b 58–kilodalton protein and the products of E4 open reading frames 3 and 6. *Mol. Cell. Biol.* **13,** 3231–3244.

Pathak, V. K., and Temin, H. M. (1990). Broad spectrum of *in vivo* forward mutations, hypermutations, and mutational hotspots in a retroviral shuttle vector after a single replication cycle: Substitutions, frameshifts, and hypermutations. *Proc. Natl. Acad. Sci. USA* **87,** 6019–6023.

Pattanakitsakul, S., Zheng, J. H., Natsuume-Sakai, S., Takahashi, M., and Nonaka, M. (1992). Aberrant splicing caused by the insertion of the B2 sequence into an intron of the complement C4 gene is the basis for low C4 production in H-$2^k$ mice. *J. Biol. Chem.* **267,** 7814–7820.

Paule, M. R., and White, R. J. (2000). Survey and summary: Transcription by RNA polymerases I and III. *Nucleic Acids Res.* **28,** 1283–1298.

Pavlicek, A., Jabbari, K., Paces, J., Paces, V., Hejnar, J. V., and Bernardi, G. (2001). Similar integration but different stability of Alus and LINEs in the human genome. *Gene* **276,** 39–45.

Pelissier, T., Bousquet-Antonelli, C., Lavie, L., and Deragon, J. M. (2004). Synthesis and processing of tRNA-related SINE transcripts in *Arabidopsis thaliana*. *Nucleic Acids Res.* **32,** 3957–3966.

Perez-Stable, C., Ayres, T. M., and Shen, C. K. (1984). Distinctive sequence organization and functional programming of an Alu repeat promoter. *Proc. Natl. Acad. Sci. USA* **81,** 5291–5295.

Piedrafita, F. J., Molander, R. B., Vansant, G., Orlova, E. A., Pfahl, M., and Reynolds, W. F. (1996). An Alu element in the myeloperoxidase promoter contains a composite SP1–thyroid hormone–retinoic acid response element. *J. Biol. Chem.* **271,** 14412–14420.

Puchta, H. (2005). The repair of double-strand breaks in plants: Mechanisms and consequences for genome evolution. *J. Exp. Bot.* **56,** 1–14.

Rat Genome, Sequencing Project, Consortium (2004). Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature* **428,** 493–521.

Rearden, A., Magnet, A., Kudo, S., and Fukuda, M. (1993). Glycophorin B and glycophorin E genes arose from the glycophorin A ancestral gene via two duplications during primate evolution. *J. Biol. Chem.* **268,** 2260–2267.

Rogers, J. H. (1985). The origin and evolution of retroposons. *Int. Rev. Cytol.* **93,** 187–279.

Rothenburg, S., Eiben, M., Koch-Nolte, F., and Haag, F. (2002). Independent integration of rodent identifier (ID) elements into orthologous sites of some RT6 alleles of *Rattus norvegicus* and *Rattus rattus*. *J. Mol. Evol.* **55,** 251–259.

Rothkopf, G. S., Telakowski-Hopkins, C. A., Stotish, R. L., and Pickett, C. B. (1986). Multiplicity of glutathione *S*-transferase genes in the rat and association with a type 2 Alu repetitive element. *Biochemistry* **25,** 993–1002.

Roy, A. M., West, N. C., Rao, A., Adhikari, P., Aleman, C., Barnes, A. P., and Deininger, P. L. (2000). Upstream flanking sequences and transcription of SINEs. *J. Mol. Biol.* **302,** 17–25.

Roy-Engel, A. M., Salem, A. H., Oyeniran, O. O., Deininger, L., Hedges, D. J., Kilroy, G. E., Batzer, M. A., and Deininger, P. L. (2002a). Active Alu element "A-tails": Size does matter. *Genome Res.* **12,** 1333–1344.

Roy-Engel, A. M., Carroll, M. L., El-Sawy, M., Salem, A. H., Garber, R. K., Nguyen, S. V., Deininger, P. L., and Batzer, M. A. (2002b). Non-traditional Alu evolution and primate genomic diversity. *J. Mol. Biol.* **316,** 1033–1040.

Rozhdestvensky, T. S., Kopylov, A. M., Brosius, J., and Huttenhofer, A. (2001). Neuronal BC1 RNA structure: Evolutionary conversion of a $tRNA^{Ala}$ domain into an extended stem–loop structure. *RNA* **7,** 722–730.

Rubin, C. M., Vande Voort, C. A., Teplitz, R. L., and Schmid, C. W. (1994). Alu repeated DNAs are differentially methylated in primate germ cells. *Nucleic Acids Res.* **22,** 5121–5127.

Rubin, C. M., Kimura, R. H., and Schmid, C. W. (2002). Selective stimulation of translational expression by Alu RNA. *Nucleic Acids Res.* **30,** 3253–3261.

Rudin, C. M., and Thompson, C. B. (2001). Transcriptional activation of short interspersed elements by DNA-damaging agents. *Genes Chromosomes Cancer* **30,** 64–71.

Russanova, V. R., Driscoll, C. T., and Howard, B. H. (1995). Adenovirus type 2 preferentially stimulates polymerase III transcription of Alu elements by relieving repression: A potential role for chromatin. *Mol. Cell. Biol.* **15,** 4282–4290.

Ryan, S. C., and Dugaiczyk, A. (1989). Newly arisen DNA repeats in primate phylogeny. *Proc. Natl. Acad. Sci. USA* **86,** 9360–9364.

Ryskov, A. P., Ivanov, P. L., Kramerov, D. A., and Georgiev, G. P. (1983). Mouse ubiquitous B2 repeat in polysomal and cytoplasmic poly(A)$^+$ RNAs: Uniderectional orientation and 3′-end localization. *Nucleic Acids Res.* **11,** 6541–6558.

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.

Sarrowa, J., Chang, D. Y., and Maraia, R. J. (1997). The decline in human Alu retroposition was accompanied by an asymmetric decrease in SRP9/14 binding to dimeric Alu RNA and increased expression of small cytoplasmic Alu RNA. *Mol. Cell. Biol.* **17,** 1144–1151.

Schramm, L., and Hernandez, N. (2002). Recruitment of RNA polymerase III to its target promoters. *Genes Dev.* **16,** 2593–2620.

Scofield, M. A., Xiong, W., Haas, M. J., Zeng, Y., and Cox, G. S. (2000). Sequence analysis of the human glycoprotein hormone α-subunit gene 5′-flanking DNA and identification of a potential regulatory element as an Alu repetitive sequence. *Biochim. Biophys. Acta* **1493,** 302–318.

Shaikh, T. H., Roy, A. M., Kim, J., Batzer, M. A., and Deininger, P. L. (1997). cDNAs derived from primary and small cytoplasmic Alu (scAlu) transcripts. *J. Mol. Biol.* **271,** 222–234.

Shapiro, J. A. (1999). Transposable elements as the key to a 21st century view of evolution. *Genetica* **107,** 171–179.

Sharan, C., Hamilton, N. M., Parl, A. K., Singh, P. K., and Chaudhuri, G. (1999). Identification and characterization of a transcriptional silencer upstream of the human BRCA2 gene. *Biochem. Biophys. Res. Commun.* **265,** 285–290.

Shedlock, A. M., and Okada, N. (2000). SINE insertions: Powerful tools for molecular systematics. *Bioessays* **22,** 148–160.

Shimamura, M., Yasue, H., Ohshima, K., Abe, H., Kato, H., Kishiro, T., Goto, M., Munechika, I., and Okada, N. (1997). Molecular evidence from retroposons that whales form a clade within even-toed ungulates. *Nature* **388,** 666–670.

Shimamura, M., Nikaido, M., Ohshima, K., and Okada, N. (1998). A SINE that acquired a role in signal transduction during evolution. *Mol. Biol. Evol.* **15,** 923–925.

Shimba, S., Buckley, B., Reddy, R., Kiss, T., and Filipowicz, W. (1992). Cap structure of U3 small nucleolar RNA in animal and plant cells is different: γ-Monomethyl phosphate cap structure in plant RNA. *J. Biol. Chem.* **267,** 13772–13777.

Shumyatsky, G., Wright, D., and Reddy, R. (1993). Methylphosphate cap structure increases the stability of 7SK, B2 and U6 small RNAs in *Xenopus* oocytes. *Nucleic Acids Res.* **21,** 4756–4761.

Shumyatsky, G. P., Tillib, S. V., and Kramerov, D. A. (1990). B2 RNA and 7SK RNA, RNA polymerase III transcripts, have a cap-like structure at their 5′ end. *Nucleic Acids Res.* **18,** 6347–6351.

Singh, K., Carey, M., Saragosti, S., and Botchan, M. (1985). Expression of enhanced levels of small RNA polymerase III transcripts encoded by the B2 repeats in simian virus 40-transformed mouse cells. *Nature* **314,** 553–556.

Singh, R., and Reddy, R. (1989). γ-Monomethyl phosphate: A cap structure in spliceosomal U6 small nuclear RNA. *Proc. Natl. Acad. Sci. USA* **86,** 8280–8283.

Slagel, V. K., and Deininger, P. L. (1989). *In vivo* transcription of a cloned prosimian primate SINE sequence. *Nucleic Acids Res.* **17,** 8669–8682.

Smit, A. F. (1999). Interspersed repeats and other mementos of transposable elements in mammalian genomes. *Curr. Opin. Genet. Dev.* **9,** 657–663.

Sorek, R., Ast, G., and Graur, D. (2002). Alu-containing exons are alternatively spliced. *Genome Res.* **12,** 1060–1067.

Sorek, R., Lev-Maor, G., Reznik, M., Dagan, T., Belinky, F., Graur, D., and Ast, G. (2004). Minimal conditions for exonization of intronic sequences: 5′ splice site formation in Alu exons. *Mol. Cell* **14,** 221–231.

Sutcliffe, J. G., Milner, R. J., Bloom, F. E., and Lerner, R. A. (1982). Common 82-nucleotide sequence unique to brain RNA. *Proc. Natl. Acad. Sci. USA* **79,** 4942–4946.

Sutcliffe, J. G., Milner, R. J., Gottesfeld, J. M., and Lerner, R. A. (1984). Identifier sequences are transcribed specifically in brain. *Nature* **308,** 237–241.

Szafranski, K., Dingermann, T., Glockner, G., and Winckler, T. (2004). Template jumping by a LINE reverse transcriptase has created a SINE-like 5S rRNA retropseudogene in *Dictyostelium*. *Mol. Genet. Genomics* **271,** 98–102.

Tatout, C., Lavie, L., and Deragon, J. M. (1998). Similar target site selection occurs in integration of plant and mammalian retroposons. *J. Mol. Evol.* **47,** 463–470.

Taylor, K. D., and Piko, L. (1987). Patterns of mRNA prevalence and expression of B1 and B2 transcripts in early mouse embryos. *Development* **101,** 877–892.

Terai, Y., Takahashi, K., Nishida, M., Sato, T., and Okada, N. (2003). Using SINEs to probe ancient explosive speciation: "Hidden" radiation of African cichlids? *Mol. Biol. Evol.* **20,** 924–930.

Tikhonov, A. P., Lavie, L., Tatout, C., Bennetzen, J. L., Avramova, Z., and Deragon, J. M. (2001). Target sites for SINE integration in *Brassica* genomes display nuclear matrix binding activity. *Chromosome Res.* **9,** 325–337.

Tomilin, N. V. (1999). Control of genes by mammalian retroposons. *Int. Rev. Cytol.* **186,** 1–48.

Tourriere, H., Chebli, K., and Tazi, J. (2002). mRNA degradation machines in eukaryotic cells. *Biochimie* **84,** 821–837.

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

Vansant, G., and Reynolds, W. F. (1995). The consensus sequence of a major Alu subfamily contains a functional retinoic acid response element. *Proc. Natl. Acad. Sci. USA* **92,** 8229–8233.

Vassetzky, N. S., and Kramerov, D. A. (2002). CAN: A pan-carnivore SINE family. *Mamm. Genome* **13,** 50–57.

Vassetzky, N. S., Ten, O. A., and Kramerov, D. A. (2003). B1 and related SINEs in mammalian genomes. *Gene* **319,** 149–160.

Vasseur, M., Condamine, H., and Duprey, P. (1985). RNAs containing B2 repeated sequences are transcribed in the early stages of mouse embryogenesis. *EMBO J.* **4,** 1749–1753.

von Sternberg, R. (2002). On the roles of repetitive DNA elements in the context of a unified genomic–epigenetic system. *Ann. N.Y. Acad. Sci.* **981,** 154–188.

Wahle, E., and Ruegsegger, U. (1999). 3′-End processing of pre-mRNA in eukaryotes. *FEMS Microbiol. Rev.* **23,** 277–295.

Walter, P., and Johnson, A. E. (1994). Signal sequence recognition and protein targeting to the endoplasmic reticulum membrane. *Annu. Rev. Cell Biol.* **10,** 87–119.

Wang, H., Iacoangeli, A., Popp, S., Muslimov, I. A., Imataka, H., Sonenberg, N., Lomakin, I. B., and Tiedge, H. (2002). Dendritic BC1 RNA: Functional role in regulation of translation initiation. *J. Neurosci.* **22,** 10232–10241.

Watkins, W. S., Rogers, A. R., Ostler, C. T., Wooding, S., Bamshad, M. J., Brassington, A. M., Carroll, M. L., Nguyen, S. V., Walker, J. A., Prasad, B. V., Reddy, P. G., Das, P. K., Batzer, M. A., and Jorde, L. B. (2003). Genetic variation among world populations: Inferences from 100 Alu insertion polymorphisms. *Genome Res.* **13,** 1607–1618.

Wei, W., Gilbert, N., Ooi, S. L., Lawler, J. F., Ostertag, E. M., Kazazian, H. H., Boeke, J. D., and Moran, J. V. (2001). Human L1 retrotransposition: *Cis* preference versus *trans* complementation. *Mol. Cell. Biol.* **21,** 1429–1439.

Weichenrieder, O., Wild, K., Strub, K., and Cusack, S. (2000). Structure and assembly of the Alu domain of the mammalian signal recognition particle. *Nature* **408,** 167–173.

Weiner, A. M. (2002). SINEs and LINEs: The art of biting the hand that feeds you. *Curr. Opin. Cell Biol.* **14,** 343–350.

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

West, N., Roy-Engel, A. M., Imataka, H., Sonenberg, N., and Deininger, P. (2002). Shared protein components of SINE RNPs. *J. Mol. Biol.* **321,** 423–432.

Williams, W. P., Tamburic, L., and Astell, C. R. (2004). Increased levels of B1 and B2 SINE transcripts in mouse fibroblast cells due to minute virus of mice infection. *Virology* **327,** 233–241.

Willoughby, D. A., Vilalta, A., and Oshima, R. G. (2000). An Alu element from the K18 gene confers position-independent expression in transgenic mice. *J. Biol. Chem.* **275,** 759–768.
Wilson, E. T., Condliffe, D. P., and Sprague, K. U. (1988). Transcriptional properties of BmX, a moderately repetitive silkworm gene that is an RNA polymerase III template. *Mol. Cell. Biol.* **8,** 624–631.
Wolin, S. L., and Cedervall, T. (2002). The La protein. *Annu. Rev. Biochem.* **71,** 375–403.
Wu, J., Grindlay, G. J., Bushel, P., Mendelsohn, L., and Allan, M. (1990). Negative regulation of the human ϵ-globin gene by transcriptional interference: Role of an Alu repetitive element. *Mol. Cell. Biol.* **10,** 1209–1216.
Yang, J., Malik, H. S., and Eickbush, T. H. (1999). Identification of the endonuclease domain encoded by R2 and other site-specific, non-long terminal repeat retrotransposable elements. *Proc. Natl. Acad. Sci. USA* **96,** 7847–7852.
Yang, S., Smit, A. F., Schwartz, S., Chiaromonte, F., Roskin, K. M., Haussler, D., Miller, W., and Hardison, R. C. (2004). Patterns of insertions and their covariation with substitutions in the rat, mouse, and human genomes. *Genome Res.* **14,** 517–527.
Yates, P. A., Burman, R. W., Mummaneni, P., Krussel, S., and Turker, M. S. (1999). Tandem B1 elements located in a mouse methylation center provide a target for *de novo* DNA methylation. *J. Biol. Chem.* **274,** 36357–36361.
Ye, J., Yang, Z., Hayes, J. J., and Eickbush, T. H. (2002). R2 retrotransposition on assembled nucleosomes depends on the translational position of the target site. *EMBO J.* **21,** 6853–6864.
Yoder, J. A., Walsh, C. P., and Bestor, T. H. (1997). Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet.* **13,** 335–340.
Zhang, C. Y., Kim, S., Harney, J. W., and Larsen, P. R. (1998). Further characterization of thyroid hormone response elements in the human type 1 iodothyronine deiodinase gene. *Endocrinology* **139,** 1156–1163.
Zhou, Y. H., Zheng, J. B., Gu, X., Saunders, G. F., and Yung, W. K. (2002). Novel PAX6 binding sites in the human genome and the role of repetitive elements in the evolution of gene regulation. *Genome Res.* **12,** 1716–1722.
Zietkiewicz, E., and Labuda, D. (1996). Mosaic evolution of rodent B1 elements. *J. Mol. Evol.* **42,** 66–72.

## Further Reading

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,** 465–478.
Borodulina, O. R., and Kramerov, D. A. (1999). Wide distribution of short interspersed elements among eukaryotic genomes. *FEBS Lett.* **457,** 409–413.
Bradfield, J. Y., Locke, J., and Wyatt, G. R. (1985). An ubiquitous interspersed DNA sequence family in an insect. *DNA* **4,** 357–363.
Cheng, J. F., Printz, R., Callaghan, T., Shuey, D., and Hardison, R. C. (1984). The rabbit C family of short, interspersed repeats: Nucleotide sequence determination and transcriptional analysis. *J. Mol. Biol.* **176,** 1–20.
Daniels, G. R., and Deininger, P. L. (1991). Characterization of a third major SINE family of repetitive sequences in the galago genome. *Nucleic Acids Res.* **19,** 1649–1656.
Deragon, J. M., Landry, B. S., Pelissier, T., Tutois, S., Tourmente, S., and Picard, G. (1994). An analysis of retroposition in plants based on a family of SINEs from *Brassica napus*. *J. Mol. Evol.* **39,** 378–386.

Endoh, H., and 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. USA* **83,** 251–255.

Fantaccione, S., Russo, C., Palomba, P., Rienzo, M., and Pontecorvo, G. (2004). A new pair of CR1-like LINE and tRNA-derived SINE elements in *Podarcis sicula* genome. *Gene* **339,** 189–198.

Ferbeyre, G., Smith, J. M., and Cedergren, R. (1998). Schistosome satellite DNA encodes active hammerhead ribozymes. *Mol. Cell. Biol.* **18,** 3880–3888.

Feschotte, C., Fourrier, N., Desmons, I., and Mouches, C. (2001). Birth of a retroposon: The Twin SINE family from the vector mosquito *Culex pipiens* may have originated from a dimeric tRNA precursor. *Mol. Biol. Evol.* **18,** 74–84.

Hamada, M., Kido, Y., Himberg, M., Reist, J. D., Ying, C., Hasegawa, M., and Okada, N. (1997). A newly isolated family of short interspersed repetitive elements (SINEs) in coregonid fishes (whitefish) with sequences that are almost identical to those of the Sma I family of repeats: Possible evidence for the horizontal transfer of SINEs. *Genetics* **146,** 355–367.

Holt, R. A., Subramanian, G. M., Halpern, A., Sutton, G. G., Charlab, R., Nusskern, D. R., Wincker, P., Clark, A. G., Ribeiro, J. M., Wides, R., Salzberg, S. L., Loftus, B., Yandell, M., Majoros, W. H., Rusch, D. B., Lai, Z., Kraft, C. L., Abril, J. F., Anthouard, V., Arensburger, P., Atkinson, P. W., Baden, H., de Berardinis, V., Baldwin, D., Benes, V., Biedler, J., Blass, C., Bolanos, R., Boscus, D., Barnstead, M., Cai, S., Center, A., Chatuverdi, K., Christophides, G. K., Chrystal, M. A., Clamp, M., Cravchik, A., Curwen, V., Dana, A., Delcher, A., Dew, I., Evans, C. A., Flanigan, M., Grundschober-Freimoser, A., Friedli, L., Gu, Z., Guan, P., Guigo, R., Hillenmeyer, M. E., Hladun, S. L., Hogan, J. R., Hong, Y. S., Hoover, J., Jaillon, O., Ke, Z., Kodira, C., Kokoza, E., Koutsos, A., Letunic, I., Levitsky, A., Liang, Y., Lin, J. J., Lobo, N. F., Lopez, J. R., Malek, J. A., McIntosh, T. C., Meister, S., Miller, J., Mobarry, C., Mongin, E., Murphy, S. D., O'Brochta, D. A., Pfannkoch, C., Qi, R., Regier, M. A., Remington, K., Shao, H., Sharakhova, M. V., Sitter, C. D., Shetty, J., Smith, T. J., Strong, R., Sun, J., Thomasova, D., Ton, L. Q., Topalis, P., Tu, Z., Unger, M. F., Walenz, B., Wang, A., Wang, J., Wang, M., Wang, X., Woodford, K. J., Wortman, J. R., Wu, M., Yao, A., Zdobnov, E. M., Zhang, H., Zhao, Q., Zhao, S., Zhu, S. C., Zhimulev, I., Coluzzi, M., della Torre, A., Roth, C. W., Louis, C., Kalush, F., Mural, R. J., Myers, E. W., Adams, M. D., Smith, H. O., Broder, S., Gardner, M. J., Fraser, C. M., Birney, E., Bork, P., Brey, P. T., Venter, J. C., Weissenbach, J., Kafatos, F. C., Collins, F. H., and Hoffman, S. L. (2002). The genome sequence of the malaria mosquito *Anopheles gambiae*. *Science* **298,** 129–149.

Izsvak, Z., Ivics, Z., Garcia-Estefania, D., Fahrenkrug, S. C., and Hackett, P. B. (1996). DANA elements: A family of composite, tRNA-derived short interspersed DNA elements associated with mutational activities in zebrafish (*Danio rerio*). *Proc. Natl. Acad. Sci. USA* **93,** 1077–1081.

Jurka, J. (2000). Repbase update: A database and an electronic journal of repetitive elements. *Trends Genet.* **16,** 418–420.

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.

Laha, T., McManus, D. P., Loukas, A., and Brindley, P. J. (2000). Sjα elements, short interspersed element-like retroposons bearing a hammerhead ribozyme motif from the genome of the oriental blood fluke *Schistosoma japonicum*. *Biochim. Biophys. Acta* **1492,** 477–482.

Lavrent'eva, M. V., Rivkin, M. I., Shilov, A. G., Kobets, M. L., Rogozin, I. B., and Serov, O. L. (1989). B2-like repetitive sequence in the genome of the American mink. *Dokl. Akad. Nauk. SSSR* **307,** 226–228.

Lee, I. Y., Westaway, D., Smit, A. F., Wang, K., Seto, J., Chen, L., Acharya, C., Ankener, M., Baskin, D., Cooper, C., Yao, H., Prusiner, S. B., and Hood, L. E. (1998). Complete genomic sequence and analysis of the prion protein gene region from three mammalian species. *Genome Res.* **8,** 1022–1037.

Lenstra, J. A., van Boxtel, J. A., Zwaagstra, K. A., and Schwerin, M. (1993). Short interspersed nuclear element (SINE) sequences of the Bovidae. *Anim. Genet.* **24,** 33–39.

Milner, R. J., Bloom, F. E., Lai, C., Lerner, R. A., and Sutcliffe, J. G. (1984). Brain-specific genes have identifier sequences in their introns. *Proc. Natl. Acad. Sci. USA* **81,** 713–717.

Mochizuki, K., Umeda, M., Ohtsubo, H., and Ohtsubo, E. (1992). Characterization of a plant SINE, p-SINE1, in rice genomes. *Jpn. J. Genet.* **67,** 155–166.

Motohashi, R., Mochizuki, K., Ohtsubo, H., and Ohtsubo, E. (1997). Structures and distribution of p-SINE1 members in rice genomes. *Theor. Appl. Genet.* **95,** 359–368.

Myouga, F., Tsuchimoto, S., Noma, K., Ohtsubo, H., and Ohtsubo, E. (2001). Identification and structural analysis of SINE elements in the *Arabidopsis thaliana* genome. *Genes Genet. Syst.* **76,** 169–179.

Nikaido, M., Matsuno, F., Abe, H., Shimamura, M., Hamilton, H., Matsubayashi, H., and Okada, N. (2001). Evolution of CHR-2 SINEs in cetartiodactyl genomes: Possible evidence for the monophyletic origin of toothed whales. *Mamm. Genome* **12,** 909–915.

Nishihara, H., Terai, Y., and Okada, N. (2002). Characterization of novel Alu- and tRNA-related SINEs from the tree shrew and evolutionary implications of their origins. *Mol. Biol. Evol.* **19,** 1964–1972.

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.

Ogiwara, I., Miya, M., Ohshima, K., and Okada, N. (1999). Retropositional parasitism of SINEs on LINEs: Identification of SINEs and LINEs in elasmobranchs. *Mol. Biol. Evol.* **16,** 1238–1250.

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.

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.

Piskurek, O., Nikaido, M., Boeadi, Baba, M., and Okada, N. (2003). Unique mammalian tRNA-derived repetitive elements in dermopterans: The t-SINE family and its retro-transposition through multiple sources. *Mol. Biol. Evol.* **20,** 1659–1668.

Roos, C., Schmitz, J., and Zischler, H. (2004). Primate jumping genes elucidate strepsirrhine phylogeny. *Proc. Natl. Acad. Sci. USA* **101,** 10650–10654.

Sakagami, M., Ohshima, K., Mukoyama, H., Yasue, H., and Okada, N. (1994). A novel tRNA species as an origin of short interspersed repetitive elements (SINEs): Equine SINEs may have originated from $tRNA^{Ser}$. *J. Mol. Biol.* **239,** 731–735.

Sasaki, T., Takahashi, K., Nikaido, M., Miura, S., Yasukawa, Y., and Okada, N. (2004). First application of the SINE (short interspersed repetitive element) method to infer phylogenetic relationships in reptiles: An example from the turtle superfamily Testudinoidea. *Mol. Biol. Evol.* **21,** 705–715.

Schmid, C., and Maraia, R. (1992). Transcriptional regulation and transpositional selection of active SINE sequences. *Curr. Opin. Genet. Dev.* **2,** 874–882.

Schmitz, J., and Zischler, H. (2003). A novel family of tRNA-derived SINEs in the colugo and two new retrotransposable markers separating dermopterans from primates. *Mol. Phylogenet. Evol.* **28,** 341–349.

Serdobova, I. M., and Kramerov, D. A. (1998). Short retroposons of the B2 superfamily: Evolution and application for the study of rodent phylogeny. *J. Mol. Evol.* **46,** 202–214.

Shimamura, M., Abe, H., Nikaido, M., Ohshima, K., and Okada, N. (1999). Genealogy of families of SINEs in cetaceans and artiodactyls: The presence of a huge superfamily of $tRNA^{Glu}$-derived families of SINEs. *Mol. Biol. Evol.* **16,** 1046–1060.

Simmen, M. W., and Bird, A. (2000). Sequence analysis of transposable elements in the sea squirt, *Ciona intestinalis*. *Mol. Biol. Evol.* **17,** 1685–1694.

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. *Nucleic Acids Res.* **15,** 2780.

Smit, A. F., and Riggs, A. D. (1995). MIRs are classic, tRNA-derived SINEs that amplified before the mammalian radiation. *Nucleic Acids Res.* **23,** 98–102.

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

Takahashi, K., Terai, Y., Nishida, M., and Okada, N. (1998). A novel family of short interspersed repetitive elements (SINEs) from cichlids: The patterns of insertion of SINEs at orthologous loci support the proposed monophyly of four major groups of cichlid fishes in Lake Tanganyika. *Mol. Biol. Evol.* **15,** 391–407.

Tu, Z. (1999). Genomic and evolutionary analysis of Feilai, a diverse family of highly reiterated SINEs in the yellow fever mosquito, *Aedes aegypti*. *Mol. Biol. Evol.* **16,** 760–772.

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.