



## V-SINEs: A New Superfamily of Vertebrate SINEs That Are Widespread in Vertebrate Genomes and Retain a Strongly Conserved Segment within Each Repetitive Unit

Ikuo Ogiwara, Masaki Miya, Kazuhiko Ohshima, et al.

*Genome Res.* 2002 12: 316-324

Access the most recent version at doi:[10.1101/gr.212302](https://doi.org/10.1101/gr.212302)

---

### References

This article cites 46 articles, 14 of which can be accessed free at:  
<http://genome.cshlp.org/content/12/2/316.full.html#ref-list-1>

### License

### Email Alerting Service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

---

To subscribe to *Genome Research* go to:  
<https://genome.cshlp.org/subscriptions>

---

Cold Spring Harbor Laboratory Press

# V-SINEs: A New Superfamily of Vertebrate SINEs That Are Widespread in Vertebrate Genomes and Retain a Strongly Conserved Segment within Each Repetitive Unit

Ikuo Ogiwara,<sup>1,3</sup> Masaki Miya,<sup>2</sup> Kazuhiko Ohshima,<sup>1</sup> and Norihiro Okada<sup>1,4</sup>

<sup>1</sup>Graduate School of Bioscience and Biotechnology, Tokyo Institute of Technology, 4259 Nagatsuta-cho, Midori-ku, Yokohama 226-8501, Japan; and <sup>2</sup>Department of Zoology, Natural History Museum and Institute, 955-2 Aoba-cho, Chuo-ku, Chiba 260-0852, Japan

We have identified a new superfamily of vertebrate short interspersed repetitive elements (SINEs), designated V-SINEs, that are widespread in fishes and frogs. Each V-SINE includes a central conserved domain preceded by a 5'-end tRNA-related region and followed by a potentially recombinogenic (TG)<sub>n</sub> tract, with a 3' tail derived from the 3' untranslated region (UTR) of the corresponding partner long interspersed repetitive element (LINE) that encodes a functional reverse transcriptase. The central domain is strongly conserved and is even found in SINEs in the lamprey genome, suggesting that V-SINEs might be ~550 Myr old or older in view of the timing of divergence of the lamprey lineage from the bony fish lineage. The central conserved domain might have been subject to some form of positive selection. Although the contemporary 3' tails of V-SINEs differ from one another, it is possible that the original 3' tail might have been replaced, via recombination, by the 3' tails of more active partner LINEs, thereby retaining retropositional activity and the ability to survive for long periods on the evolutionary time scale. It seems plausible that V-SINEs may have some function(s) that have been maintained by the coevolution of SINEs and LINEs during the evolution of vertebrates.

[The sequences reported in this paper have been deposited in the DDB/GenBank database under accession nos. ABO72981–ABO73004. Supplemental figures are available online at <http://www.genome.org>.]

Retroposons are genetic elements that have moved within and among genomes via an RNA intermediate (Weiner et al. 1986). They include SINEs, LINEs, LTR-retrotransposons, and retroviruses. L1 and Alu are representative of the LINEs and SINEs, respectively, in the human genome (Weiner et al. 1986; Eickbush 1994; Smit 1996). An analysis of the entire human genome revealed that retroposons constitute up to 40% of the genome and that protein-coding sequences account for only about 1.5% (International Human Genome Sequencing Consortium 2001). The above mentioned 40% of the human genome includes several clearly discernible groups of retroposons, such as L1, LINE2, Alu, and MIR, and these easily recognizable sequences represent the results of relatively recent amplifications over the past 200 Myr (International Human Genome Sequencing Consortium 2001). Accordingly, the proportion of the genome that originated from retroposons may increase up to as much as 60% to 70%, if we could identify and include ancient retroposons, which might have been amplified in the very distant past, are now barely recognizable as such, and are buried in the genome as debris. Such a putative extensive contribution of retroposons to the

construction of the contemporary human genome has not been anticipated.

Typical SINEs vary from 100 to 500 bp in length. They have an internal RNA polymerase III (pol III) promoter but no open reading frames (Okada 1991; Schmid and Maraia 1992; Deininger and Batzer 1993; Okada and Ohshima 1995; Schmid 1998). Many families of SINEs have been identified in multicellular eukaryotes, such as invertebrates (Ohshima et al. 1993; Ohshima and Okada 1994), plants (Yoshioka et al. 1993; Deragon et al. 1994), fishes (Kido et al. 1991; Takasaki et al. 1994; Takahashi et al. 1998; Ogiwara et al. 1999), and mammals (Shimamura et al. 1999; Gilbert and Labuda 2000; see also Shedlock and Okada 2000 for review). Specific families of SINEs are found only in closely related species; thus, it has been postulated that each family of SINEs was created relatively recently on the evolutionary time scale. For example, the SmaI SINE family is present in the genomes of chum and pink salmon but not in genomes of related species in the genus *Oncorhynchus* and related genera (Kido et al. 1991). The FokI SINE family is found exclusively in the genomes of species of *Salverinus* but not in those of related genera (Kido et al. 1991). The HpaI SINE family is apparently present in the genomes of all members of the family *Salmonidae* (Kido et al. 1991). However, recent examination of the distribution of SINEs indicates some similarity between the HpaI family in salmonids and the Thr-1 family in mammals (Gilbert and Labuda 1999). Possible evolutionary relationships among different families of SINEs in distantly related

<sup>3</sup>Present address: Department of Biochemistry, University of Washington, Seattle, WA 98195-7350, USA.

<sup>4</sup>Corresponding author.

E-MAIL [nokada@bio.titech.ac.jp](mailto:nokada@bio.titech.ac.jp); FAX 81-45-924-5835.

Article and publication are at <http://www.genome.org/cgi/doi/10.1101/gr.212302>.

species and the mechanism of the apparently episodic appearance of SINEs are poorly understood.

Retroposons were initially considered to be genomic parasites or junk DNA because of their redundancy in the genome and their restricted distribution among species. However, in many cases, a retroposon can become functional upon introduction of a new poly(A) signal, an enhancer signal, or a new exon (Makalowski 1995; Britten 1997; Brosius 1999). Schmid's group recently suggested a general function for vertebrate SINEs, proposing that SINE RNAs may promote translation under stress (Chu et al. 1998). Moreover, analysis of the draft sequence of the human genome supports possibility of positive selection for human Alu SINEs, which might be related to Schmid's hypothesis, in gene-rich regions of the genome (International Human Genome Sequencing Consortium 2001). The functional significance of SINEs in other mammals and other vertebrates, such as fishes, remains to be confirmed.

In this study, we characterized a new superfamily of SINEs that are widespread in vertebrates. Each SINE contains a strongly conserved region of unknown origin. Our results suggest that the members of the superfamily may have some function(s) providing new insight into a large part of the vertebrate genome.

## RESULTS

### Characterization of V-SINEs

We characterized five new SINE families from distantly related fishes, namely, lamprey, lungfish, rasbora, medaka, and fugu. Figure 1 shows an alignment of the consensus sequences of these families of SINEs, as well as those of previously characterized SINEs from shark, ray (Ogiwara et al. 1999), and zebrafish (Izsvak et al. 1996; Shimoda et al. 1996; Izsvak et al. 1997). The 5' end of each SINE family was clearly derived from a tRNA, and it was homologous to *Xenopus* valyl tRNA (CAC) in sequence and structure (see Supplemental Fig. 1, available online at <http://www.genome.org>). The conservation of the various tRNA-related regions presumably reflects the requirements for transcription by pol III (Okada and Ohshima 1995). The tRNA-related region was followed in every case by a strongly conserved region that was almost identical in the different SINE families. We designated this region the central conserved domain. Each family had a different 3' tail, downstream from the central conserved domain, with potentially recombinogenic (TG)<sub>n</sub> tracts (Boehm et al. 1989; Wahls et al. 1990) between the central conserved domain and the 3' tail region. The extent of identity between the various SINE families ranged from 86% to 93% in the central conserved domain of 83 bp (nt 107–nt 189 in Fig. 1). In the lamprey and zebrafish, the extent of identity was 88% in this region, whereas it was only 70% even in the most highly conserved segment of the tRNA-related region of 64 bp (nt 13–nt 76 in Fig. 1). Conservation of the central conserved domain of these SINE families was surprising to us in view of the timing of the divergence of the various hosts (see Discussion). We chose the designation V-SINEs (SINEs from vertebrates; Fig. 2) for these SINE families.

### Distribution of V-SINEs

To examine the distribution of V-SINEs among vertebrates, we isolated genomic DNA from 77 vertebrate species and analyzed it by PCR using the DNA as template and two oligonucleotide primers that were specific for the tRNA-related re-

gion and the central conserved domain (Fig. 2). The results are summarized in Figure 3. V-SINEs were distributed extensively among vertebrates but appeared to be absent in salmonids and amniotes. The distribution of V-SINEs suggests that they might have been generated in a common ancestor of vertebrates and might then have survived in most vertebrates. We estimated the copy number for each family of V-SINEs by calculating the frequencies of SINE sequences isolated from the genomic libraries. The copy numbers varied greatly among species (Table 1), probably because of differences among retropositional activities of V-SINEs in the various host species.

### Isolation of Partner LINEs That Share 3' Tails with V-SINEs

To identify the origins of the variable 3' tails of V-SINEs, we attempted to isolate partner LINEs that share 3' tails with the V-SINEs (see Discussion). We successfully isolated two novel LINE families and characterized their 3'-half sequences. The zebrafish family of ZfL3 LINEs encoded an RTase domain that is typical of LINEs (Fig. 4A). The 3' tail (but not the extreme 3' end of the 3' UTR) was almost identical to that of the zebrafish DANA SINE (Fig. 4B). The lungfish LfR1 family of LINEs also retained a 3' tail that was almost identical to that of the lungfish Lun1 SINE (Fig. 4C). We previously reported the sharing of the 3'-tail sequence between the shark HER1 LINE and HE1 SINE, which we identified as a V-SINE in this study. These three examples of pairs of V-SINEs and LINEs suggest that other variable 3' tails of V-SINEs may correspond to the 3' tails of uncharacterized LINEs in a given organism.

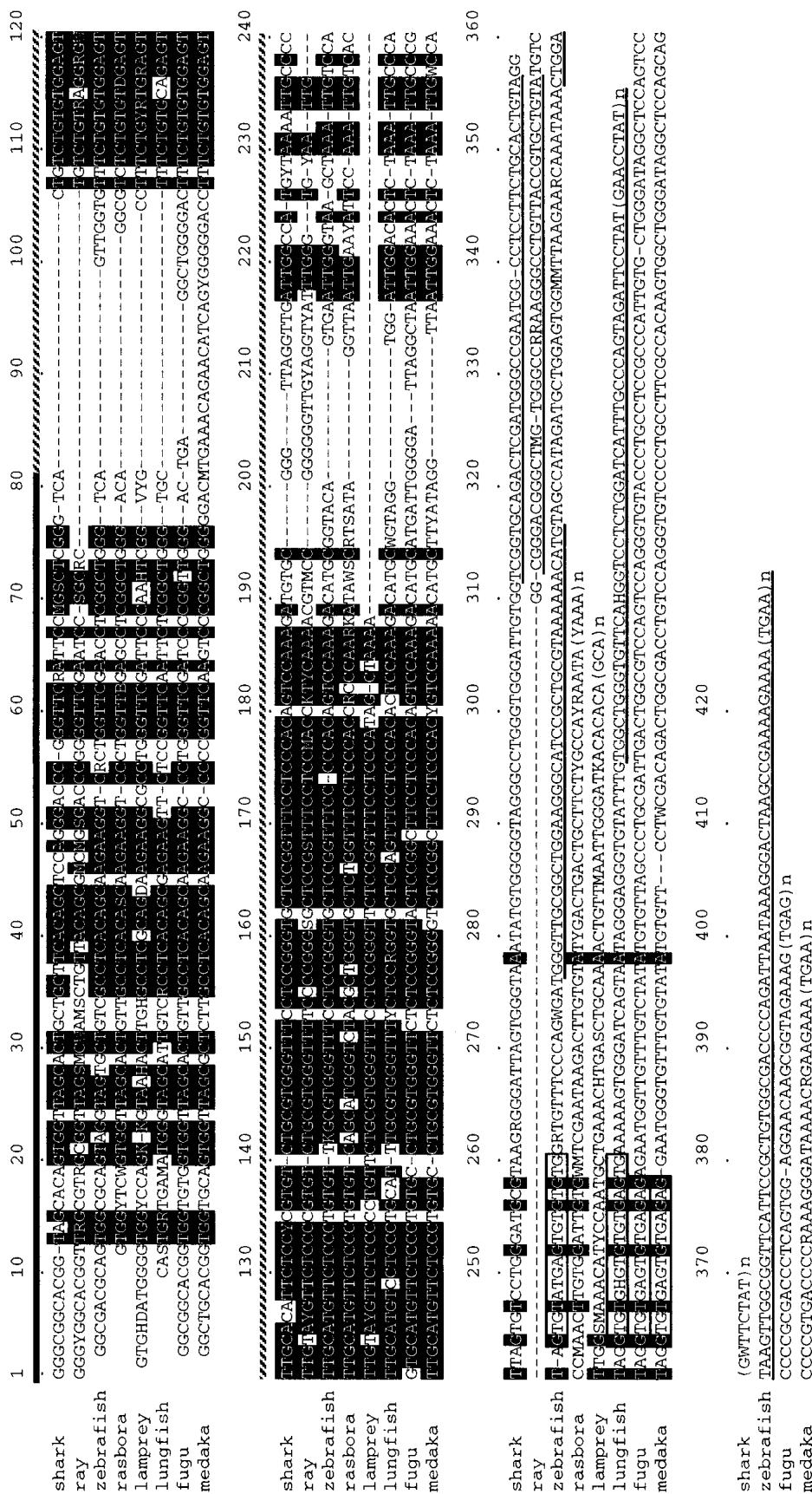
### Phylogenetic Relationships among Partner LINEs of V-SINEs

We analyzed the phylogeny of LINEs that share 3' UTRs with respective V-SINEs using the amino acid sequences of the RTases encoded by the LINEs (Figure 5). The phylogenetic tree of LINEs was divided into 11 major clades, reflecting the results of Malik et al. (1999). The zebrafish ZfL3 LINE was located in a clade that included human LINE2, several LINE2-like fish LINEs (Terai et al. 1998), and zebrafish ZfL2. We designated this clade the "LINE2 clade". The lungfish LfR1 LINE and shark and ray HER1 LINEs fell into the "CR1 clade" (Malik et al. 1999), which appears to have a sibling relationship with the LINE2 clade. The zebrafish ZfL3 LINEs and lungfish LfR1 LINE (and the shark and ray HER1 LINEs) were separated by invertebrate LINEs, such as nematode SAM1 and mosquito T1 and Q, so it is reasonable to propose that they are in different clades. Within the CR1 clade, the lungfish LfR1 LINE and shark and ray HER1 LINEs appeared to be distantly related, with lungfish LfR1 being separated by schistosome SR1 from shark and ray HER1. Considering the "vertical descent" of LINEs (Malik et al. 1999), we can speculate that the partner LINEs of V-SINEs described here are paraphyletic families of LINEs rather than a monophyletic family of LINEs in the genomes of respective species (see Discussion).

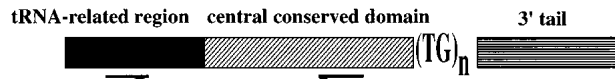
## DISCUSSION

### Extensive Conservation and Possible Function(s) of V-SINEs

We found V-SINEs in the genomes of the lamprey, cartilaginous fishes, bony fishes, and amphibians. The most primitive fishes, the Agnathans, which include lampreys and hagfishes,



**Figure 1** Characterization of V-SINEs. Consensus sequences of several V-SINEs from vertebrates are aligned and compared. The consensus sequences are from the gummy shark (*Mustelus manazo*) HE1 SINE (Ogiwara et al. 1999), electric ray (*Torpedo californica*) HE1 SINE (Ogiwara et al. 1999), zebrafish (*Danio rerio*) DANA SINE (Izsvak et al. 1996), rasbora (*Rasbora pauciperforata*) Ras1 SINE (this study), lamprey (*Lethenteron reissneri*) Lam1 SINE (this study), lungfish (*Lepidosiren paradoxa*) Lum1 SINE (this study), fugu (*Fugu rubripes*) Ac1 SINE (this study), and medaka (*Oryzias latipes*) Ac1 SINE (this study). Consensus sequences were deduced from six (zebrafish), three (rasbora), five (lamprey), five (lungfish), three (fugu), and two (medaka) sequences, respectively (see Supplemental Fig. 2, available online at <http://www.genome.org>). Nucleotides are numbered above the sequences. Under the numbers, the tRNA-related regions and the central conserved domains are indicated by a bold line and by a dotted line, respectively. (TG)<sub>n</sub> tracts are boxed. Nucleotide sequences in 3' tails that are nearly identical to those of partner LINEs are underlined. Nucleotides are highlighted in black when at least six are identical at a given position.

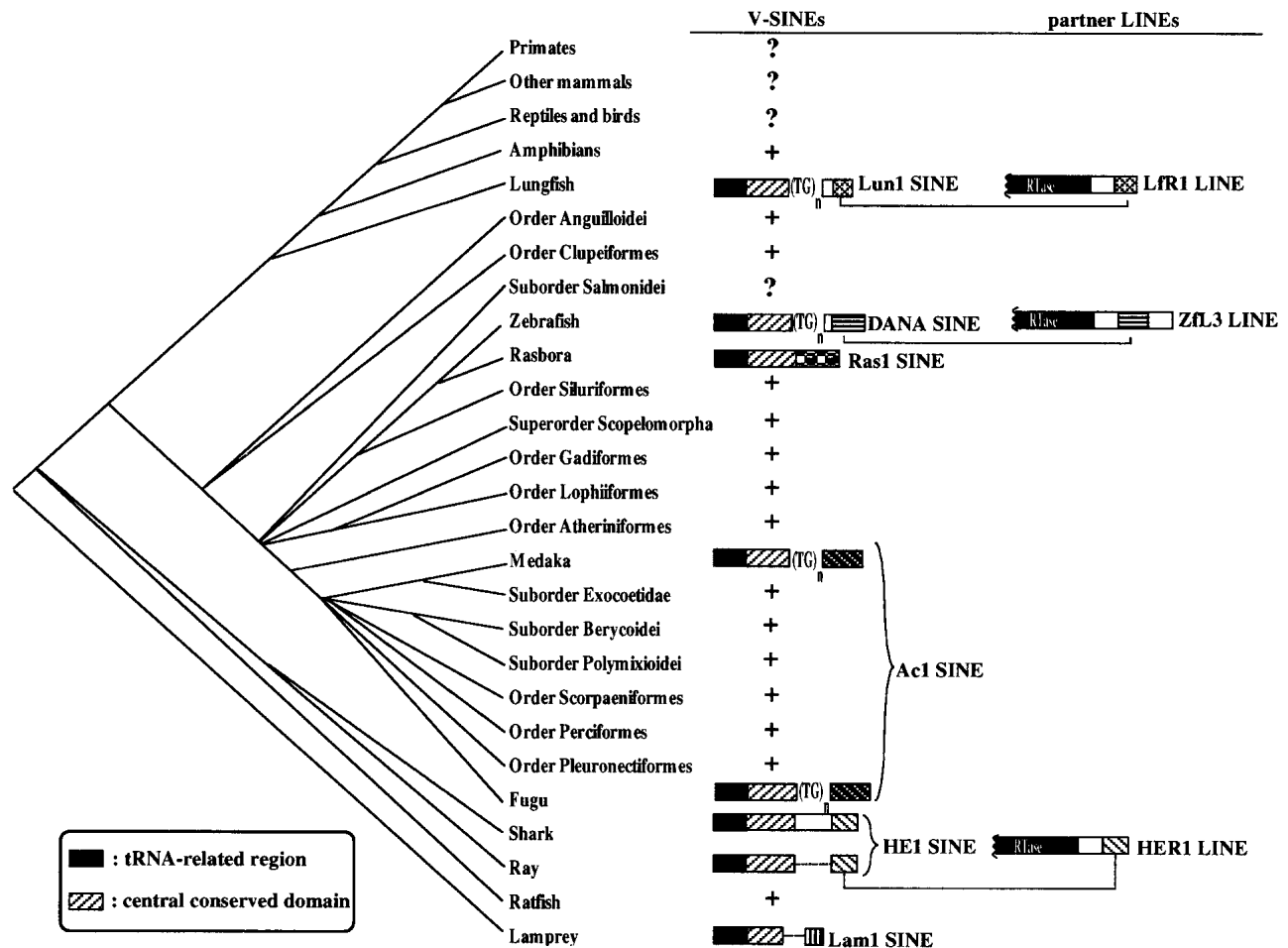


**Figure 2** A schematic representation of V-SINEs. Each V-SINE has a composite structure, namely, a tRNA-related region and a tRNA-unrelated region. The latter can be further divided into the central conserved domain, a  $(TG)_n$  tract and a 3' tail that is characteristic of each V-SINE family but differs considerably among families. Short bars with half-arrowheads indicate positions and directions of oligonucleotide primers used for PCR.

appeared in the Cambrian era (544–510 Myr ago), whereas the ancestor of cartilaginous fishes and bony fishes seems to have appeared in the Ordovician era (510–439 Myr ago; Carroll 1997; Kumar and Hedges 1998). The extensive horizontal transfer of these SINEs cannot be strictly ruled out, but, nonetheless, this wide distribution of V-SINEs suggests that they might have been generated in the genome of a common ancestor of vertebrates about 540 Myr ago or even more and might then have survived in most vertebrates until the present day.

Gilbert and Labuda (1999; 2000) described another group of SINEs, the CORE-SINEs, that might have survived in eukaryotic genomes for 550 Myr. They identified a common sequence motif of ~65 bp within the SINEs. The identity among such “core” sequences ranged from 50% to 74%. For example, the identity between the consensus sequences from human Ther-1 and salmonid HpaI was 65% in the core region of 65 bp (Gilbert and Labuda 1999). Thus, the identity between such sequences was somewhat limited. In contrast, the identity among V-SINEs was much more marked, ranging from 86% to 93% in the central conserved domain of 83 bp. In the consensus sequences of the V-SINEs from lamprey and zebrafish, the identity in this region was 88%, whereas it was only 70% even in the strongly conserved tRNA-related region of 64 bp, which provides promoters for RNA polymerase III.

The extensive identity among V-SINEs is surprising in view of the timing of divergence of their hosts, even if we accept a possible low rate of sequence divergence in lower vertebrates (Britten 1986; Martin and Palumbi 1993). It is possible that the central conserved domain of the V-SINEs has undergone a recent widespread gene conversion or is subject



**Figure 3** The widespread distribution of V-SINEs among vertebrates. Analysis by PCR was performed using the DNA from 77 vertebrate species as template and two oligonucleotides specific for the tRNA-related region and the central conserved domain of V-SINEs as primers. The Lun1, Ras1, Ac1, and Lam1 SINE families were characterized for the first time in this study. A plus sign indicates a species in which the presence of V-SINEs was strongly suggested by our analysis. A question mark indicates the absence of strong evidence for V-SINEs. Partner LINES with the same 3' tails as V-SINEs are depicted on the right.

**Figure 4** (See figure on facing page.) Characterization of partner LINEs that share 3' tails with respective V-SINEs. The zebrafish ZfL3 and lungfish LfR1 each belong to a LINE family with an open reading frame (ORF) that encodes a typical RTase (A). The RTase domains are highlighted in black (Xiong and Eickbush 1990; Malik et al. 1999). A comparison between zebrafish ZfL3 LINEs and the DANA SINE (B). The top line shows the consensus sequence of the DANA SINE. Asterisks indicate nucleotides in the consensus sequence of the zebrafish ZfL3 LINE that are identical to those in the consensus sequence of the DANA SINE. Nucleotides are highlighted in black when they predominate at a given position. Among four nucleotides, when two nucleotides are identical and the other two nucleotides differ from them but are identical, they are not highlighted but are indicated by asterisks when either pair is identical to the consensus nucleotide of V-SINEs. Dashes indicate gaps inserted to improve the alignment. The 3'-terminal repeats of the DANA SINE are underlined with bold half-arrowheads and those of ZfL3 LINEs are similarly underlined with dots. The junction between the SINE and LINEs is shown by a vertical line. A comparison between lungfish LfR1 LINEs and the Lun1 SINE (C). The top line shows the consensus sequence of the Lun1 SINE. Asterisks indicate identical nucleotides in the consensus sequences of the lungfish LfR1 LINE and the Lun1 SINE. Nucleotides are highlighted in black when at least two are identical at a given position. Dashes indicate gaps inserted to improve the alignment. The 3'-terminal repeats are underlined with bold half-arrowheads. The junction between the SINE and LINEs is shown by a vertical line.

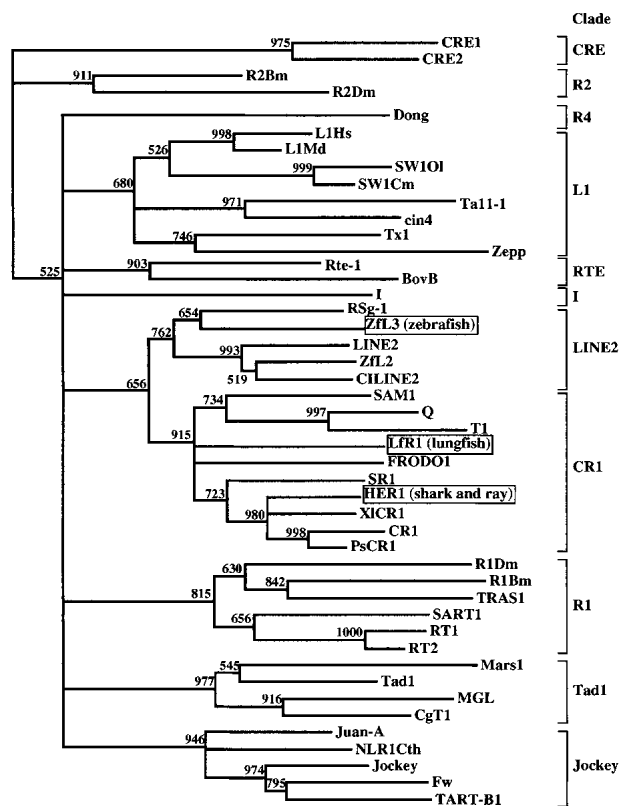
to ongoing gene conversion from some limited set of drivers giving rise to the very homogeneous structure of this region (Roy et al. 2000).

Another possibility is that the central conserved domain might have been subject to some form of positive selection. The central conserved domain may be beneficial to the retropositional activity of the SINEs. For example, Gilbert and Labuda (2000) proposed that the conserved segment of CORE-SINE may serve to anchor the transcribed CORE-element to a LINE-encoded reverse transcriptase-containing ribonucleoprotein particle. Extensive conservation of V-SINEs can, however, be more easily explained by the hypothesis that the central conserved domain may somehow "earn its keep" in the genome. The different degrees of conservation of V-SINEs, as compared with CORE-SINEs, may reflect their different contributions to host genomes. It will be of great interest to examine whether mutations at certain positions in the repetitive units of V-SINEs might have been suppressed during evolution (Britten 1994).

**Table 1. The Relative Contributions of V-SINEs to the Genomes of Various Organisms**

Organism	Percent of genome	Reference
Sharks and rays	~10–0.01	Ogiwara et al. 1999
Zebrafish	~10	Izsvák et al. 1996 Shimoda et al. 1996 Izsvák et al. 1997
Catfish	~0.3	Kim et al. 2000
Rasbora	a ~ 0.004	This study
Lamprey	b ~ 0.04	This study
Lungfish	c ~ 0.15	This study
Medaka	d ~ 0.1	This study Shimoda et al. 1996 Izsvák et al. 1997
Fugu	e ~ 0.02	This study Izsvák et al. 1997

Copy numbers were estimated as follows: (a) When 10,000 clones, which represented about  $3 \times 10^4$  kbp of rasbora genomic DNA, were screened with the Ras1 SINE, four clones, which represented about 1.2 kbp of Ras1 SINEs, were isolated by hybridization. (b) 5,000 clones, which represented about  $1.5 \times 10^4$  kbp of lamprey genomic DNA, were screened with the Lam1 SINE, and 25 clones, which represented about 6.3 kbp of Lam1 SINEs, were isolated by hybridization. (c) 5,000 clones, which represented about  $1.5 \times 10^4$  kbp of lungfish genomic DNA, were screened with the Lun1 SINE, 75 clones, which represented about 23 kbp of Lun1 SINEs, were isolated by hybridization. (d) The number of Ac1 SINEs in the medaka genome was estimated from the intensity of signals on a Southern blot (data not shown). (e) In fugu, about 3.5 kbp were identified as Ac1 SINEs in the  $2 \times 10^4$  kbp of the fugu genome that have been sequenced to date (<http://fugu.hgmp.mrc.ac.uk>).



**Figure 5** A phylogeny of families of LINEs, including LINEs that form pairs with corresponding V-SINEs. The phylogenetic tree was constructed by the neighbor-joining method (Saitou and Nei 1987), with programs in the PHYLIP package (Felsenstein 1996), using the sequences of the seven subdomains of the RTase domains of LINEs (Xiong and Eickbush 1990). Each number above a branch indicates the bootstrap value per 1000 replications and provides an indication of the statistical significance of the node. Branches with bootstrap values below 50% are reduced to polytomies. The LINEs paired with V-SINEs are boxed. Clade names (Malik et al. 1999) are shown on the right. Sources of sequences are as follows (notations refer to accession numbers unless otherwise indicated): CRE1, M33009; CRE2, U19151; Ta11-1, L47193; cin4, Y00086; Zepp, AB008896; Fw, M17214; I, M14954; Jockey, M22874; R1Dm, X51968; R2Dm, X51967; TART-B1, U14101; NLR1Cth, S59870; Q, U03849; RT1, M93690; RT2, M93691; T1, M93689; Juan-A, M95171; R1Bm, M19755; R2Bm, M16558; SART1, D85594; Dong, L08889; TRAS1, D38414; Rte-1, U00034; Tx1, M26915; L1Hs, M80340; LINE2, Repbase Update 1997 on World Wide Web URL: <http://www.girinst.org/~server/rebase.html>; L1Md, M13002; LfR1, this study; SW1OI, AF055642; SW1Cm, AF055643; ZfL2, Ogiwara et al. unpublished; ZfL3, AA497301 and AA497226 and this study; CILINE2, Ogiwara et al. unpublished; RSg-1, Ogiwara et al. unpublished; HER, Ogiwara et al. 1999; SR1, U66331; CR1, U88211; PsCR1, AB005891; SAM1, U13643; FRODO1, Z70755; BovB, Z25531; XICR1, AF027962; MGL, AF018033; Mars1, X99081; CgT1, L76169; and Tad1, L25662.



The functions of V-SINEs are unknown. However, an intriguing possible function has been proposed for other SINEs. The proposal is based on the observation that, in many species, SINEs are transcribed under stress conditions, with the resultant RNAs binding specifically to a particular protein kinase (PKR), thereby blocking its ability to inhibit translation (Schmid 1998, Chu et al. 1998). SINE RNAs would, thus, promote translation of proteins under stress. Furthermore, an analysis of the density of Alu sequences and the GC content of individual human chromosomes suggested the possibility of positive selection for Alu sequences in the GC-rich chromatin in which genes are concentrated (International Human Genome Sequencing Consortium 2001). However, because this function of SINEs would not necessarily require particular sequences (Schmid 1998), such may not be the case for V-SINEs. Double-stranded RNA interference or cosuppression is known to suppress the movement of transposable elements in *Caenorhabditis elegans* (Tabara et al. 1999) and *Drosophila* (Jensen et al. 1999). The high-order structure of the central conserved domain of V-SINE RNA may act beneficially to suppress expression of retroposons in an epigenetic manner (Whitelaw and Martin 2001).

V-SINEs are widely distributed among vertebrates but appear to be absent in amniotes. There are many sequences that are quite similar to the central conserved domain of V-SINEs in the human genome. They are a component of a DNA transposon-like element referred to as MER6 (Smit and Riggs 1996; data not shown). These sequences have no tRNA-related region and are severely divergent. Thus, such sequences may be merely genomic debris derived from V-SINEs that had been inserted in the DNA transposon. Alternatively, the central segment of V-SINEs might have been maintained in times past in the primate genome because of some benefit to the host. Further efforts are required to clarify the function(s) of the central conserved domain of V-SINEs.

### Variations among the 3' Tails of V-SINEs

The 5' end of each V-SINE was clearly derived from a tRNA and was followed by a central conserved domain that was nearly identical in all V-SINEs from various species. However, 3' tail sequences downstream of the central conserved domain were quite different. The three different 3' tail sequences of V-SINEs from three species were, however, almost identical to the respective 3' UTRs of particular LINEs. What does this phenomenon mean?

There are many examples of a 3' tail sequence being shared by a noncoding SINE and a retrotranspositionally active LINE (Ohshima et al. 1996; Smit 1996; Kajikawa et al. 1997; Terai et al. 1998; Gilbert and Labuda 1999; Ogiwara et al. 1999; see Okada et al. 1997 for review). Such sharing suggests that the LINE-encoded RTase mobilizes the passive SINE by recognizing the 3' tail of the SINE RNA as a template for reverse transcription (Luan et al. 1993; Luan and Eickbush 1995; Ohshima et al. 1996; see Okada et al. 1997 for review).

The transmission of LINEs in eucaryotic genomes is generally considered to be "vertical" (Malik et al. 1999). There is no reliable evidence for the horizontal transfer of LINEs between eukaryotes, with one exception (Kordis and Gubensek 1997), and it is, therefore, considered that the phylogeny of LINEs corresponds basically to the phylogeny of species. Each SINE is apparently dependent on a partner LINE, with the same 3' end sequence, for its retroposition, and the introduction of even a few mutations in the 3' tail of the partner LINE

abolishes its retropositional activity (M. Kajikawa and N. Okada, in press). Thus, it is possible that a SINE becomes inactive when the SINE is transferred horizontally to another species that does not include the specific partner LINE within its genome. Therefore, extensive transfer of SINEs between species does not seem to provide a suitable explanation for the presence of the central conserved domain and for the wide distribution of V-SINEs.

If SINEs and the majority of LINEs were inherited only vertically, it is possible that SINEs and LINEs became inactive upon the accumulation of mutations in the respective lineages. When SINEs become inactive, they are buried as debris in the genome. When a family of LINEs becomes inactive, the partner SINEs should also be inactivated because there is no longer a source of RTase for retroposition. However, a SINE would be able to survive inactivation of its partner LINE if the 3' tail of the SINE were to be replaced, via a mechanism such as gene conversion, by that of an active non-partner LINE that could supply an RTase for retroposition. The present distribution of the variable 3' tails of V-SINEs probably reflects many such events that occurred frequently during evolution and allowed the survival of V-SINEs that included the central conserved domain. These V-SINEs would otherwise have been lost but maintenance of V-SINEs within the genome might have been associated with an as yet unidentified selective advantage.

## METHODS

### Construction and Screening of Genomic Libraries

Genomic libraries were constructed by ligation of  $\lambda$ gt10 arms or EMBL4 arms with genomic DNA from zebrafish (*Danio rerio*), rasbora (*Rasbora pauciperforata*), lamprey (*Lethenteron reissneri*), and lungfish (*Lepidosiren paradoxa*) that had been completely digested with Eco RI. For isolation of DNA clones with a SINE sequence from the zebrafish library, we used [ $\alpha$ - $^{32}$ P]dCTP-labeled DNA that corresponded to nucleotide (nt) positions 10–191 of the gummy shark (*Mustelus manazo*) HE1 SINE (Fig. 1) as probe. For isolation of DNA clones with SINE sequences from the rasbora, lamprey, and lungfish libraries, we used [ $\alpha$ - $^{32}$ P]dCTP-labeled DNA that corresponded to nt 33–191 of the respective V-SINEs (Fig. 1) as probes (see below). For isolation of DNA clones with LINE sequences from the zebrafish and lungfish libraries, we used [ $\alpha$ - $^{32}$ P]dCTP-labeled DNA that corresponded to the reverse transcriptase (RTase) domain of the zebrafish ZfL3 LINE and of the lungfish LfR1 LINE as probe, respectively (see below).

### Search of the DDBJ, EMBL, and GenBank Nucleotide Databases

To isolate the V-SINEs from zebrafish, fugu (*Fugu rubripes*), and medaka (*Oryzias latipes*) we screened nucleotide databases using the nucleotide sequence of the gummy shark HE1 SINE and the BLAST program (Altschul et al. 1990). To isolate the RTase domain of the ZfL3 LINE from zebrafish, we screened the databases using the sequences of the RTase domains of various LINEs and the BLAST program (Altschul et al. 1990).

### Analysis by Polymerase Chain Reaction

We amplified sequences by polymerase chain reaction (PCR) using the DNA from 77 vertebrate species as template and two primers that were specific for the tRNA-related region and the core region of the gummy shark HE1 SINEs (nt 10–191 in Fig. 1). Reactions were performed with 50 ng of total genomic DNA and 1.25 U of Taq DNA polymerase (Takara, Kyoto, Ja-

pan), in the presence of 2.0 mM MgCl<sub>2</sub> under the following conditions: 30 cycles of 94°C for 1 min, 50°C for 1 min, and 72°C for 30 sec. The nucleotide sequence of the 5' primer was 5'-GGTGGCACAGTGGTTAGCAC-3' and that of the 3' primer was 5'-CATCTTTGGACTGTGGGA-3'. For amplification of sequences from rasbora, lamprey, and lungfish, we used two primers that were specific for the tRNA-related region of the zebrafish DANA SINE and the core region (nt 33–191 in Fig. 1). Reaction conditions were the same as described above with the exception that the concentration of MgCl<sub>2</sub> was reduced to 1.5 mM. Amplified products were used to screen genomic libraries for the isolation of SINEs. The nucleotide sequences of the 5' and 3' primers were 5'-TCGCCTCACAGCAAGAAG-3' and 5'-CATCTTTGGACTGTGGGA-3', respectively.

### Isolation of a Lungfish LINE

To isolate the lungfish LfR1 LINE, we performed PCR with two primers that were specific for ORF2 of the shark HER1 LINE and the 3' tail of lungfish Lun1 SINEs, respectively. Products of appropriate sizes were cloned and their nucleotide sequences were determined (data not shown). The nucleotide sequences of the 5' and 3' primers were 5'-GGTAAGGC CACATTTGGAGTAC-3' and 5'-CATGACACCCAGCCACA-3', respectively. To determine the amino acid sequence of the RTase domain encoded by the lungfish LfR1 LINE, we performed PCR with a 5' primer specific for the RTase domains of various LINES and a 3' primer specific for ORF2 of LfR1 LINES. Products of appropriate sizes were cloned and their nucleotide sequences were determined (data not shown). The nucleotide sequences of the 5' and 3' primers were 5'-TRGAYDYNRAN AANGCMTT-3' and 5'-CCCCTCTCAGGAGTCTGTAGG-3', respectively.

### ACKNOWLEDGMENTS

We thank Dr. Alan Weiner (University of Washington) for his comments. This work was supported by a Grant-in-Aid from the Ministry of Education, Culture, Sports, Science, and Technology.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

### REFERENCES

- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215**: 403–410.
- Boehm, T., Mengle-Gaw, L., Kees, U.R., Spurr, N., Lavenir, I., Forster, A., and Rabbitts, T.H. 1989. Alternating purine-pyrimidine tracts may promote chromosomal translocations seen in a variety of human lymphoid tumours. *EMBO J.* **8**: 2621–2631.
- Britten, R.J. 1986. Rates of DNA sequence evolution differ between taxonomic groups. *Science* **231**: 1393–1398.
- . 1994. Evolutionary selection against change in many Alu repeat sequences interspersed through primate genomes. *Proc. Natl. Acad. Sci.* **91**: 5992–5996.
- . 1997. Mobile elements inserted in the distant past have taken on important functions. *Gene* **205**: 177–182.
- Brosius, J. 1999. RNAs from all categories generate retrosequences that may be exapted as novel genes or regulatory elements. *Gene* **238**: 115–134.
- Carroll, R.L. 1997. *Patterns and processes of vertebrate evolution*. Cambridge University Press, Cambridge, NY.
- 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.
- Deininger, P.L. and Batzer, M.A. 1993. Evolution of retrospoons. *Evol. Biol.* **27**: 157–196.
- 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.
- Eickbush, T.H. 1994. Origin and evolutionary relationships of retroelements. In *The evolutionary biology of viruses* (ed. S.S. Morse), pp. 121–157. Raven Press, Ltd., New York.
- Felsenstein, J. 1996. PHYLIP (phylogeny inference package). Version 3.572. University of Washington, Seattle, WA.
- Gilbert, N. and Labuda, D. 1999. CORE-SINEs: Eukaryotic short interspersed retroposing elements with common sequence motifs. *Proc. Natl. Acad. Sci.* **96**: 2869–2874.
- . 2000. Evolutionary inventions and continuity of CORE-SINEs in mammals. *J. Mol. Biol.* **298**: 365–377.
- International Human Genome Sequencing Consortium. 2001. Initial sequencing and analysis of the human genome. *Nature* **409**: 860–921.
- Izsvák, 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.* **93**: 1077–1081.
- Izsvák, Z., Ivics, Z., and Hackett, P.B. 1997. Repetitive elements and their genetic applications in zebrafish. *Biochem. Cell Biol.* **75**: 507–523.
- Jensen, S., Gassama, M.P., and Heidmann, T. 1999. Taming of transposable elements by homology-dependent gene silencing. *Nat. Genet.* **21**: 209–212.
- Kajikawa, M., Ohshima, K., and Okada, N. 1997. Determination of the entire sequence of turtle CR1: The first open reading frame of the turtle CR1 element encodes a protein with a novel zinc finger motif. *Mol. Biol. Evol.* **14**: 1206–1217.
- 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.* **88**: 2326–2330.
- Kim, S., Karsi, A., Dunham, R.A., and Liu, Z. 2000. The skeletal muscle alpha-actin gene of channel catfish (*Ictalurus punctatus*) and its association with piscine specific SINE elements. *Gene* **252**: 173–181.
- Kordis, D. and Gubensek, F. 1997. Bov-B long interspersed repeated DNA (LINE) sequences are present in *Vipera ammodytes* phospholipase A2 genes and in genomes of Viperidae snakes. *Eur. J. Biochem.* **246**: 772–779.
- Kumar, S. and Hedges, S.B. 1998. A molecular timescale for vertebrate evolution. *Nature* **392**: 917–920.
- 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.
- 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.
- Makalowski, W. 1995. SINEs as a genomic scrap yard: an essay on genomic evolution. In *The impact of short interspersed elements (SINEs) on the host genome* (ed. R.J. Maraia), pp. 81–104. R.G. Landes Company, Austin, Texas.
- 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.
- Martin, A.P. and Palumbi, S.R. 1993. Body size, metabolic rate, generation time, and the molecular clock. *Proc. Natl. Acad. Sci.* **90**: 4087–4091.
- 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., 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., 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.* **90**: 6260–6264.
- Ohshima, K. and Okada, N. 1994. Generality of the tRNA origin of short interspersed repetitive elements (SINEs). Characterization of three different tRNA-derived retroposons in the octopus. *J. Mol. Biol.* **243**: 25–37.

- Okada, N. 1991. SINEs: Short interspersed repeated elements of the eukaryotic genome. *Trends Ecol. Evol.* **6**: 358–361.
- Okada, N. and Ohshima, K. 1995. Evolution of tRNA-derived SINEs. In *The impact of short interspersed elements (SINEs) on the host genome* (ed. R.J. Maraia), pp. 61–79. R.G. Landes Company, Austin, Texas.
- Okada, N., Hamada, M., Ogiwara, I., and Ohshima, K. 1997. SINEs and LINEs share common 3' sequences: a review. *Gene* **205**: 229–243.
- Roy, A.M., Carroll, M.L., Nguyen, S.V., Salem, A.-H., Oldridge, M., Wilkie, A.O.M., Batzer, M.A., and Deininger, P.L. 2000. Potential gene conversion and source genes for recently integrated Alu elements. *Genome Res.* **10**: 1485–1495.
- Saitou, N. and Nei, M. 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**: 406–425.
- Schmid, C.W. 1998. Does SINE evolution preclude Alu function? *Nucleic Acids Res.* **26**: 4541–4550.
- Schmid, C. and Maraia, R. 1992. Transcriptional regulation and transpositional selection of active SINE sequences. *Curr. Opin. Genet. Dev.* **2**: 874–882.
- Shedlock, A.M., and Okada, N. 2000. SINE insertions: Powerful tools for molecular systematics. *Bioessays* **22**: 148–160.
- 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<sup>Glu</sup>-derived families of SINEs. *Mol. Biol. Evol.* **16**: 1046–1060.
- Shimoda, N., Chevrette, M., Ekker, M., Kikuchi, Y., Hotta, Y., and Okamoto, H. 1996. Mermaid, a family of short interspersed repetitive elements, is useful for zebrafish genome mapping. *Biochem. Biophys. Res. Commun.* **220**: 233–237.
- Smit, A.F.A. 1996. The origin of interspersed repeats in the human genome. *Curr. Opin. Genet. Dev.* **6**: 743–748.
- Smit, A.F.A. and Riggs, A.D. 1996. Tiggers and DNA transposon fossils in the human genome. *Proc. Natl. Acad. Sci.* **93**: 1443–1448.
- Tabara, H., Sarkissian, M., Kelly, W. G., Fleenor, J., Grishok, A., Timmons, L., Fire, A., and Mello, C.C. 1999. The rde-1 gene, RNA interference, and transposon silencing in *C. elegans*. *Cell* **99**: 123–132.
- 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.
- Takasaki, N., Murata, S., Saitoh, M., Kobayashi, T., Park, L., and Okada, N. 1994. Species-specific amplification of tRNA-derived short interspersed repetitive elements (SINEs) by retroposition: A process of parasitization of entire genomes during the evolution of salmonids. *Proc. Natl. Acad. Sci.* **91**: 10153–10157.
- Terai, Y., Takahashi, K., and Okada, N. 1998. SINE cousins: the 3'-end tails of the two oldest and distantly related families of SINEs are descended from the 3' ends of LINEs with the same genealogical origin. *Mol. Biol. Evol.* **15**: 1460–1471.
- Wahls, W.P., Wallace, L.J., and Moore, P.D. 1990. The Z-DNA motif d(TG)<sub>30</sub> promotes reception of information during gene conversion events while stimulating homologous recombination in human cells in culture. *Mol. Cell. Biol.* **10**: 785–793.
- 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.
- Whitelaw, E. and Martin, D.I.K. 2001. Retrotransposons as epigenetic mediators of phenotypic variation in mammals. *Nat. Genet.* **27**: 361–365.
- Xiong, Y. and Eickbush, T.H. 1990. Origin and evolution of retroelements based upon their reverse transcriptase sequences. *EMBO J.* **9**: 3353–3362.
- Yoshioka, Y., Matsumoto, S., Kojima, S., Ohshima, K., Okada, N., and Machida, Y. 1993. Molecular characterization of a short interspersed repetitive element from tobacco that exhibits sequence homology to specific tRNAs. *Proc. Natl. Acad. Sci.* **90**: 6562–6566.

Received August 2, 2001; accepted in revised form November 30, 2001.