A Novel Class of SINE Elements Derived from 5S rRNA

Vladimir V. Kapitonov and Jerzy Jurka

Genetic Information Research Institute, Mountain View, California

Eukaryotic genomes are colonized by different retroposons, including short interspersed repetitive elements (SINEs). All currently known SINEs are derived from tRNA and 7SL RNA genes and exploit their type 2 internal pol III promoters. We report here a novel class of SINE elements, called SINE3, derived from 5S rRNA. SINE3s are transcribed from the type 1 internal pol III promoter. Approximately 10,000 copies of SINE3 elements are present in the zebrafish genome, they constitute approximately 0.4% of the genomic DNA. Some elements are as little as 1% diverged from each other, indicating that the retrotransposition of SINE3 in zebrafish is an ongoing process. The 3′-tail of SINE3 is significantly similar to that of CR1-like non-LTR retrotransposons, represented by numerous subfamilies in the zebrafish genome. Analogously to CR1-like elements, SINE3 copies are not flanked by target site duplications, and their 3′ termini are composed of $(ACATT)_n$ and $(ATT)_n$ microsatellites, specific for different subfamilies of SINE3. Given the common structural features, it is highly likely that the enzymatic machinery encoded by CR1-like elements powers proliferation of SINE3.

Downloaded from https://academic.oup.com/mbe/article/20/5/694/985403 by U.S. Department of Justice user on 17 August 2022

Introduction

Transposable elements (TEs) constitute a major component of the eukaryotic genomes. TEs relying on reverse transcription of their mRNA copies are called retroposons (Weiner, Deininger, and Efstratiadis 1986). Typically, retroposons are divided into two groups depending on whether or not they encode reverse transcriptase. The first group is composed of long terminal repeat (LTR) retrotransposons and non-LTR retrotransposons. The second group includes processed pseudogenes, pseudogenes for small nuclear RNA, and short interspersed repeats (SINEs).

SINE elements are 100-bp to 500-bp long and contain internal promoters for RNA polymerase III (Singer 1982; Okada and Ohshima 1995; Schmid 1996). The unusual feature of the internal pol III promoters is that their control regions are located downstream of the transcription start site (Paule and White 2000). Therefore, a retrotransposed cDNA copy of a RNA molecule transcribed by the polymerase III preserves the internal promoter. There are three general types of pol III promoters (Paule and White 2000). The first two types are formed from internal promoters, whereas the type 3 represents external promoters. Type 1 promoters are composed of the ~15-bp A box, the ~5-bp intermediate element IE, and the ~18-bp C box, which form ~50-bp internal control region (ICR) running between +50 and +90. The 5S rRNA gene, encoding a highly conserved 120-bp RNA component of the large ribosomal subunit, is the most prominent gene employing the type 1 pol III promoter (Paule and White 2000; Barciszewska et al. 2001). Type 2 promoters are composed of the ~15-bp A box and ~10-bp B box that form an ICR running from +10 to +65. Typical type II promoters are present in tRNA and animal 7SL RNA genes. All SINEs characterized to date can be divided into two classes. The major class includes SINEs derived from tRNA molecules (Okada and Ohshima 1995; Ogiwara et al. 1999). The minor one is composed the Alu and B1 families derived from 7SL RNA and present in the primate and rodent genomes, respectively (Weiner 1980; Ullu and Tschudi 1984). Therefore, all currently known SINEs utilize pol III promoters that belong to the type 2. Given the high abundance of 5S rRNA in eukaryotic cells, the apparent lack of SINEs derived from 5S rRNA remained a puzzle (Weiner 2002).

Here, we report a new class of SINE elements, called SINE3, derived from 5S rRNA and utilizing the type 1 pol III promoter in the zebrafish genome. We also show that a 70-bp 3′ tail of SINE3 is similar to the 3′ tail of CR1-like non-LTR retrotransposons. Furthermore, SINE3 elements are not flanked by the target site duplications (TSD). Therefore, we suggest that retrotransposition of SINE3 is catalyzed by the enzymatic machinery encoded by CR1-like retrotransposons.

Key words: SINE, non-LTR retrotransposon, CR1 clade, LINE, 5S rRNA, pol III, transposable element.

E-mail: vladimir@ulam.girinst.org.

Mol. Biol. Evol. 20(5):694–702. 2003
DOI: 10.1093/molbev/msg075
© 2003 by the Society for Molecular Biology and Evolution. ISSN: 0737-4038

Materials and Methods

Copies of the unclassified repetitive Dr000031 element found recently in the zebrafish genome (Bao 2002) were identified by using CENSOR (Jurka et al. 1996) and were expanded ±1400-bp at their ends. Dr000031 copies created by long chromosomal duplications or redundant sequencing were discarded based on the more than 90% identities between the 1400-bp flanks. Using the majority rule applied to a set of the multiple aligned expanded sequences, we reconstructed the consensus sequence.

Using the SINE3 consensus sequence as a CENSOR query, we identified all SINE3-like elements present in GenBank sequences and in sequences from the Ensembl Zebrafish release 4.06.1, which comprises approximately 1% of the 1.7-Gb zebrafish genome (http://www.ensembl.org/Danio_rerio/). To identify insertions of SINE3 elements into copies of other known TEs, we expanded the GenBank-derived copies of SINE3 ±700-bp at both termini. Next, the internal portions of the expanded sequences similar to the SINE3 consensus sequence were masked out, and the remaining sequences were aligned to the sequences of transposable elements from the zebrafish section of Repbase Update (http://www.girinst.org/Repbase_Update [Jurka 2000]). We selected flanking

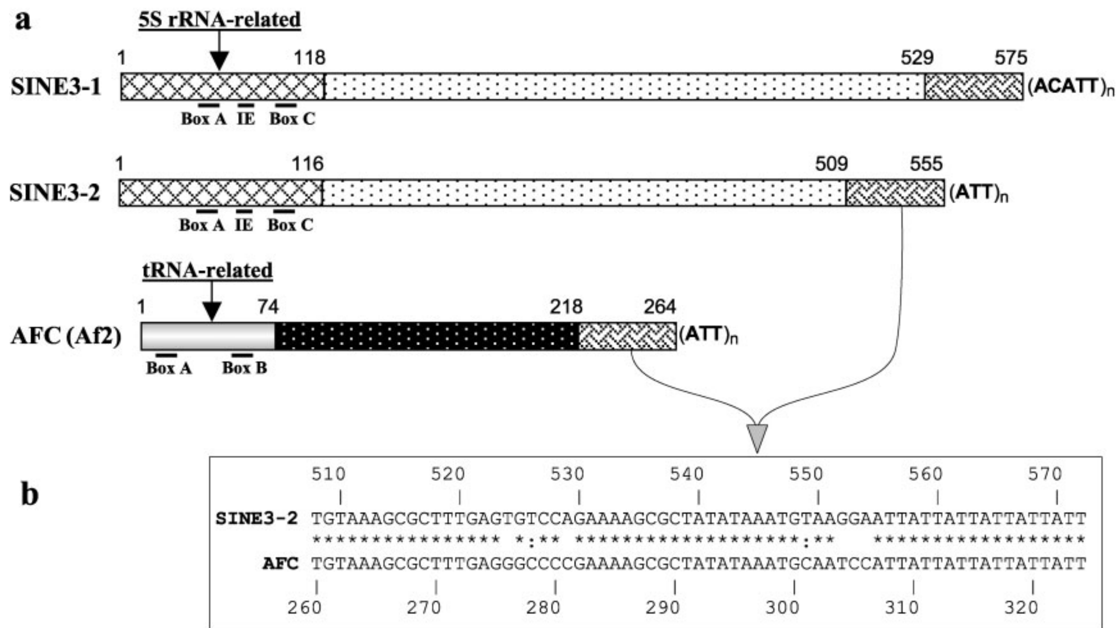

FIG. 1.—Schematic structure of SINE3 and AFC families of SINEs. (*a*) The SINE3-1 and SINE3-2 consensus sequences are 85% identical to each other and represent two subfamilies of SINE3 identified in the zebrafish genome. The 5′ terminal fragments of SINE3-1 and SINE3-2, which are approximately 74% identical to the 5S rRNA zebrafish genes, central regions, unrelated to any known transposable elements or genes, and the 3′ tails of the SINE3-1 SINE3-2, are marked by different geometrical patterns. The AFC SINE element (Takahashi et al. 1998) is composed of a tRNA-derived region and an internal portion that are unrelated to SINE3. The A, IE, C, and B boxes of the pol III promoters are underlined. The 3′ termini of SINE3 and AFC are composed of microsatellites whose repeat units are shown in parentheses. (*b*) Alignment of the SINE3-2 and AFC 3′ tails. Asterisks and semicolons indicate identical nucleotides and transitions. Positions of the aligned nucleotides in the SINE3-2 and AFC consensus sequences are shown in the first and last rows, respectively.

regions similar to DNA sequences of known TEs in order to analyze target sites associated with SINE3 insertions.

Results

We have found that an ~100-bp 3′ portion of the unclassified 600-bp Dr000031 repetitive zebrafish element (Bao 2002) was related to the 120-bp 5S RNA gene, based on comparison with sequences from Repbase Update (Jurka 2000). Using this observation as a starting point, we built a new 590-bp consensus sequence, schematically represented in figure 1. It is composed of three regions: a 118-bp 5′ region similar to 5S rRNA, a 390-bp internal region (IR), and ~60-bp 3′ terminus related to the 3′ end of a CR1-like non-LTR retrotransposon. The 5S rRNA genes in higher eukaryotes are generally arranged in tandem arrays composed of 5S rRNA and nontranscribed intergenic spacers (NTS) (Little and Braaten 1989). Since SINE3 elements are interspersed in the genome and as such do not form tandem repeats, their IR and 3′ termini cannot be classified as NTS.

General Features of SINE3

We identified approximately 100 copies of SINE3 elements in the public set of DNA sequences. Given the clear interspersed nature of SINE3 and random sequences representing the ~1.7×10^9-bp zebrafish genome, the estimated number of the SINE3 elements is 10^4 copies per the haploid genome. The average length of SINE3 elements is about 600 bp; therefore, they are accountable for as much as ~0.4% of the whole zebrafish genome.

We identified 14 insertions of SINE3 into 14 genomic copies of nine different DNA transposons and retrotransposons harbored by the zebrafish genome (table 1). Therefore, SINE3 is a transposable element. Analysis of DNA sequences harboring the identified insertions strongly indicates (data not shown) that SINE3 elements are not flanked by target site duplications (TSD), usually generated upon integration of most DNA transposons and retrotransposons into genomes. There are also no apparent open reading frames (ORFs), terminal inverted repeats, or direct repeats present in SINE3. Numerous SINE3 elements have been transposed recently (several MYA) because they are only 5% divergent from each other and are inserted into different nonhomologous regions. We identified two young subfamilies of SINE3 elements, called SINE3-1a and SINE3-2a (fig. 2). They are represented by 14 ~606-bp and 10 ~575-bp copies, respectively, identified in the current public set of zebrafish sequences. Therefore, the whole zebrafish genome is expected to harbor ~1,400 and ~1,000 copies of SINE3-1a and SINE3-2a elements, respectively. Their consensus sequences are 605-bp and 575-bp long, respectively. There is no significant similarity between the ~30-bp 5′ and 3′ termini of SINE3-1a and SINE3-2a. Whereas the 3′ tail of SINE3-1a is composed of the ACATT microsatellite, the 3′ tail of SINE3-2a is composed of the ATT microsatellite. Excluding the termini, there is an 84% identity between the SINE3-1a

Downloaded from https://academic.oup.com/mbe/article/20/5/694/985403 by U.S. Department of Justice user on 17 August 2022

Table 1
Insertions of SINE3 Elements into Copies of Known Transposable Elements

GenBank Accession Number	X1	X2	Transposable Element	Y1	Y2	O	S
AL591671	46091	46161	HE1_DR1	329	398	C	0.80
AL591671	46164	46574	SINE3-1	183	590	D	0.94
AL591671	46579	46838	HE1_DR1	10	328	C	0.78
AL592204_27	4718	4813	HE1_DR1	290	388	C	0.81
AL592204_27	4821	5011	SINE3-1	396	587	C	0.93
AL592204_27	5017	5215	HE1_DR1	74	289	C	0.68
AL590146	53883	54113	TDR7	1	240	D	0.94
AL590146	54175	54724	SINE3-2a	2	576	D	0.97
AL590146	54731	54878	TDR7	392	541	D	0.92
AL592289_1	35909	36027	ANGEL	163	303	C	0.79
AL592289_1	36028	36584	SINE3-1	3	584	D	0.90
AL592289_1	36586	36744	ANGEL	1	162	C	0.79
AL603743	63505	63673	DNA9NNN1_DR	64	232	C	0.92
AL603743	63712	64069	SINE3-2a	216	576	D	0.96
AL603743	64079	64140	DNA9NNN1_DR	1	63	C	0.78
AL626804_1	2297	2927	DNA9NNN1_DR	438	1160	C	0.82
AL626804_1	2927	3002	SINE3-2a	393	503	C	0.94
AL626804_1	2984	3173	DNA9NNN1_DR	245	437	C	0.78
AL627126_7	21692	22427	CR1-1_DR	4097	4982	C	0.85
AL627126_7	22443	22817	SINE3-1	216	590	C	0.93
AL627126_7	22826	22886	CR1-1_DR	4037	4096	C	0.89
AL627129_2	27658	28034	TC1DR3	568	944	D	0.89
AL627129_2	28040	28435	SINE3-2a	1	532	D	0.88
AL627129_2	28436	28683	TC1DR3	945	1222	D	0.83
AL596027_2	35845	36172	TC1DR3	106	400	C	0.89
AL596027_2	36284	36859	SINE3-2a	1	576	D	0.96
AL596027_2	37110	37212	TC1DR3	4	105	C	0.92
AL645795_6	6377	6434	TDR3	137	200	C	0.77
AL645795_6	6435	6964	SINE3-2	2	575	C	0.86
AL645795_6	6970	7090	TDR3	1	136	C	0.78
AL645800_5	699	1035	TDR2	1	352	D	0.90
AL645800_5	1036	1101	SINE3-2	513	577	D	0.91
AL645800_5	1105	1289	TDR2	353	544	D	0.87
AL662861_10	16705	17103	CR1-1_DR	4097	4489	C	0.86
AL662861_10	17119	17493	SINE3-1	216	590	C	0.93
AL662861_10	17502	17562	CR1-1_DR	4037	4096	C	0.89
AL645822_9	13451	13709	DNA21TA1_DR	269	537	C	0.83
AL645822_9	13710	13888	SINE3-2a	398	576	C	0.94
AL645822_9	13892	14067	DNA21TA1_DR	87	268	C	0.81
AL627254_7	30391	30611	DNA21TA1_DR	166	386	C	0.92
AL627254_7	30612	30960	SINE3-1	217	589	D	0.96
AL627254_7	30974	31139	DNA21TA1_DR	1	165	C	0.96

NOTE.—Column 1 reports accession numbers of 14 GenBank sequences where we have identified 14 copies of SINE3 elements inserted into 14 copies of nine different transposable elements (their names are listed in column 4). Nucleotide sequences of the consensus sequences of these TEs are collected in the zebrafish section of Repbase_Update (http://girinst.org/Repbase_Update). Each SINE3 insertion and the TE copy harboring it are shown as a group of three horizontal lines. X1 and X2 indicate the positions of the TE and SINE3 copies in the GenBank sequences. Y1 and Y2 show positions of fragments of the TE consensus sequences that are similar to the corresponding copies. In the column O, letters D and C mark the direct or reverse orientations of the TEs copies, respectively. S is an identity between the corresponding TE consensus sequence and its copy.

(positions 33 to 573) and SINE3-2a (positions 32 to 550) consensus sequences. However, each consensus sequence is ~98% identical to the elements from the same subfamily. Such a high intrasubfamily identity suggests that the retrotransposition of SINE3 elements is an ongoing process. For example, two SINE3-2a elements identified in the GenBank sequences AL831789 (positions 30523 to 31098) and AL844197 (positions 51654 to 51300) are only 1% different from each other.

There is a significant difference between base compositions of SINE3 and the zebrafish genome. Whereas the randomly sequenced ~1% of the genome is AT-rich (64% A+T and 36% G+C), the SINE3 elements are GC-rich (46% A+T and 54% G+C). In this regard, SINE3 is similar to the GC-rich Alu and S1 (Deragon et al. 1994) SINE elements in primates and plants, respectively. There is no significant target site specificity that can be attributed to SINE3 elements. Locally, the AT content of

Downloaded from https://academic.oup.com/mbe/article/20/5/694/985403 by U.S. Department of Justice user on 17 August 2022

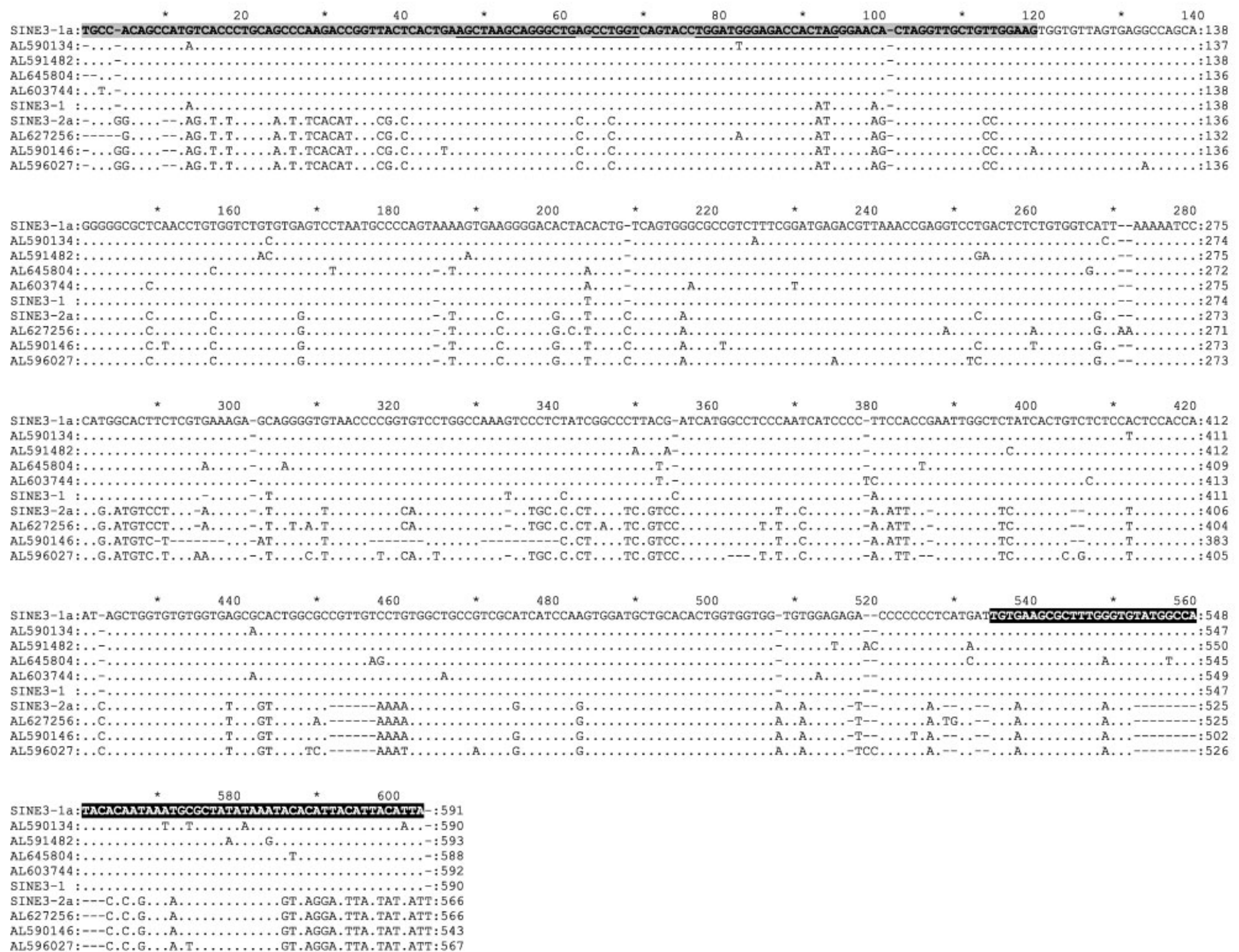

FIG. 2.—Multiple alignment of members of the SINE3-1, SINE3-1a, and SINE3-2a subfamilies of SINE3. The consensus sequence of the SINE3-1a elements is shown in the first line of the alignment. Dots indicate nucleotides identical to those in the consensus sequence. Dashes indicate insertions/deletions. The 5′ region derived from the 5S rRNA is highlighted in gray. The A, IE, and C boxes, which constitute the type 1 pol III promoter, are underlined. The 3′ tail, which is similar to those in the AFC SINEs and CR1-like non-LTR retrotransposons, is highlighted in black. Accession numbers of GenBank sequences that the SINE3-1 elements were extracted from are shown in the first column.

the SINE3 flanking regions is approximately 70% (±50 bp on each side of the SINE3 insertion position). However, the average AT content of the 1-kb flanking regions is very close to the AT content of the genome. Many SINE3 elements are present in zebrafish EST sequences. It indicates that some SINE3 are either not silenced in the genome or are inserted downstream of host gene promoters, contributing to evolution of alternative splice signals and 3′ untranscribed regions.

The 5′ End of SINE3 Is a Former 5S rRNA Gene

The very 5′ end of SINE3 is ~75% identical to the 5S rRNA gene (fig. 3). The 120-bp 5S rRNA is highly conserved in different species. It is believed that tRNA and 5S rRNA belong to the most abundant classes of RNA molecules in eukaryotic cells. Usually, the 5S rRNA gene is present in multiple copies tandemly repeated in the genome. The zebrafish 5S rRNA gene was not studied and characterized in depth, but it was mapped recently to the long arm of chromosome 3 (Gornung et al. 2000; Phillips and Reed 2000). We identified 145 copies of 5S rRNA in the AL645691 HTGS GenBank sequence. The sequence is composed of two contigs, the order of which has not yet been determined (AL645691, positions 1 to 145784 and 145885 to 158047). Sequence data show that the zebrafish 5S RNA genes form at least two separate clusters. The first cluster is sequenced completely and is composed of 76 copies of 5S rRNA (AL645691, positions 97294 to 78116). The second cluster is sequenced only partially and it is represented by the whole second contig composed of 69 tandemly repeated copies of 5S rRNA. Flanking regions of the last cluster are not known. In both clusters, the repeat unit is 177-bp long and is composed of the 120-bp 5S rRNA gene and a 57-bp NTS. There is ~99% identity between these 5S rRNA genes. Similar identity characterizes the spacer copies. The GCTT-3′ terminus of the 5S rRNA gene and the 5′-TTCG terminus of the spacer (fig. 3) form the GCTTTTCG signal that works presumably as a terminator of the pol III transcription

Downloaded from https://academic.oup.com/mbe/article/20/5/694/985403 by U.S. Department of Justice user on 17 August 2022

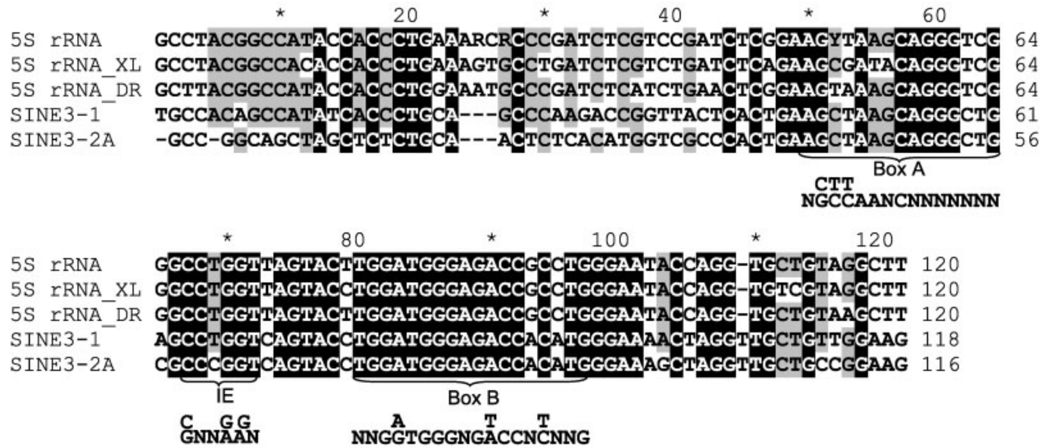

FIG. 3.—Alignment of 5S rRNA genes and SINE3 elements. The consensus sequence of eukaryotic 5S rRNA genes, the *X. laevis* and *D. rerio* 5S rRNA genes, and the 5′ regions of the SINE3-1 and SINE3-2 consensus sequences are compared. Nucleotides conserved in all five or four sequences at a given positions are highlighted in black and gray, respectively. The Box A, IE, and Box C, which constitute internal control regions of the 5S rRNA pol III promoter, are underlined. The consensus sequences of these regions, obtained for the eukaryotic 5S rRNA genes (Pieler, Hamm, and Roeder 1987) are shown beneath the marked boxes. Capital letters above each sequence are allowed alternate bases (Sajdak, Reed, and Phillips 1998).

(Sajdak, Reed, and Phillips 1998; Paule and White 2000). Interestingly, this signal is missing in SINE3 (figs. 2 and 3). All three functional sites, Box A, IE, and Box C, which constitute the pol III internal promoter of the 5S rRNA gene (Sajdak, Reed, and Phillips 1998), are surprisingly well preserved in SINE3 (fig. 3). Given this conservation, we assume that the type 1 internal promoter of pol III is involved in the SINE3 transcription, which is necessary for its transposition. Due to the loss of the 5S rRNA pol III terminator (fig. 3), the pol III transcription does not stop at the 3′ end of the 5S rRNA-related region and goes further to the 3′ terminus of SINE3.

The 3′ End of SINE3 Is Related to the 3′ End of CR1-Like Non-LTR Retrotransposons

A 44-bp region of SINE3, followed by the 3′-terminal microsatellites, is 90% identical with the corresponding region of repetitive elements from the AFC family (fig. 1). The AFC family is a classical SINE element derived from tRNA and retrotransposed in cichlids (Takahashi et al. 1998; Terai, Takahashi, and Okada 1998). Transposition of AFC is mediated by the reverse transcriptase encoded by a CR1-like non-LTR retrotransposon (Terai, Takahashi, and Okada 1998). Similarly to SINE3, AFC elements are not flanked by any target site duplications, and their 3′ termini are composed of microsatellites. Moreover, the 3′ terminal portion of AFC, which we found related to the 3′ end of SINE3, is similar to the 3′ end of CR1-like retrotransposons (Terai, Takahashi, and Okada 1998). Terai et al. (1998) suggested that a small stem formed by the 3′ end of AFC is a functional element important for recognition of the AFC mRNA by a CR1-like reverse transcriptase. Surprisingly, the stem structure is also present in SINE3. Finally, we have found that the SINE3 3′ end is significantly similar to the 3′ end of CR1-4_DR non-LTR retrotransposon present in the zebrafish genome (fig. 4). CR1-4_DR is one of at least several young CR1-like families of non-LTR retrotransposons identified recently in the zebrafish genome (Kapitonov and Jurka 2002).

The Internal Region of SINE3

The 400-bp internal region (IR) of SINE3 (fig. 1) is presumably a noncoding DNA accidentally recruited into SINE3. We found only four GenBank sequences similar to the SINE3 IR and encoded by fish species other than *D. rerio*. Two sequences (U08043, positions 5025 to 4877, 82% identical with IR; AF131253, positions 99 to 1, 81% identical with IR) are from the *Latimeria chalumnae* genome. One is from *Anguillla japonica* (AB023960, positions 4602 to 4872, 69% identical with IR) and another is from *Ictalurus furcatus* (EST BQ097237, position 380 to 1, 81% identical with IR). None of the regions flanking the IR-like elements is similar to known transposable elements, rRNA, tRNA, or other functional elements. In *A. japonica*, the IR-like SINE3 element is present in a reverse orientation in the 3′ UTR of mRNA encoding androgen receptor alpha. In latimeria, one element is present in the 3′UTR of a gene coding for the 10 MHC class I protein (U08043), and the second IR-like element is present approximately 800-bp upstream of exon 1 in the rhodopsin gene (AF131253). It is unlikely that the IR element is a novel class of functional elements conserved in fishes, because it is found in different orientation in 5′ UTR, 3′ UTR, and intron and because it is not present in the recently sequenced *Fugu rubripes* genome (Aparicio et al. 2002). Therefore, the IR element is likely a transposable element. The most surprising is an observation of the latimeria 148-bp and 99-bp sequences approximately 80% identical to the SINE3 IR. Latimeria is a "living fossil," a single surviving species of the old lineage of crossopterygian fishes that originated approximately 400 MYA (Cloutier and Ahlberg 1996). Given a very high divergence of latimeria and zebrafish, which

Downloaded from https://academic.oup.com/mbe/article/20/5/694/985403 by U.S. Department of Justice user on 17 August 2022

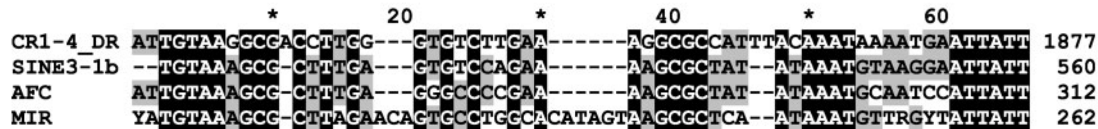

FIG. 4.—An alignment of the 3′ tails of the CR1-4_DR non-LTR retrotransposon, SINE3, AFC, and MIR SINEs. Nucleotides conserved in all five or four sequences at a given positions are highlighted in black and gray, respectively. Dashes indicate putative insertions/deletions. Numbers in the last column are positions of the 3′ termini in the corresponding consensus sequences.

corresponds to 400 Myr of separate evolution of these species, it is highly unlikely the IR-like transposable element was active in their common ancestor. Presumably, this element was transferred horizontally to one of these species.

Origin of SINE3

Figure 5 illustrates a putative model of the origin of SINE3. According to this model, SINE3 has emerged accidentally as a composite retroelement, whose 5′ and 3′ ends were derived from a 5S rRNA gene/pseudogene and a CR1-like non-LTR retrotransposon's 3′ terminal portion, respectively, separated by a short fragment of the host genome. If so, the SINE3 internal region is a copy of the last genomic fragment (fig. 5). The model does not assume too many unlikely events necessary for the origin of SINE3. Typically, retrotransposed copies of CR1-like elements are 5′ truncated and preserve their 3′ termini, which are necessary for retrotransposition, given their transcription and availability of reverse transcriptase expressed by corresponding full-length CR1-like non-LTR retrotransposons. It is believed that the CR1-like reverse transcriptase recognizes specifically DNA sequences similar to the 3′ end of the CR1-like retrotransposon encoding the reverse transcriptase. Due to the deleted promoter, further transcription and retrotransposition of the 5′ truncated copy would be suppressed (fig. 5*b*), unless this copy is close to a pol III promoter carried by 5S rRNA (fig. 5*c*). Presumably, the pol III transcription of zebrafish 5S rRNA genes produces a mature RNA because the pol III terminator is formed at the border of the 120-bp 5S rRNA coding sequence and NTS. However, as seen for SINE3, the transcription terminator can be inactivated by a few mutations. Alternatively, it can be inactivated by an abnormal 3′ processing of the 5S rRNA or by integration of 5S rRNA into some random sites. As a result, the read-through transcription of the 5S rRNA pseudogene can be extended to the 3′ termini of the CR1-like element only, forming a proto-SINE3 mRNA (fig. 5*d*). Given presence of the cognate CR1-like reverse transcriptase, the new SINE3 mRNA can be retrotransposed and give rise to a new SINE family.

Discussion

Most enzymes involved directly in proliferation of SINEs are still unknown. However, there are convincing indirect data suggesting that the proliferation is catalyzed in *trans* by the reverse transcriptase encoded by non-LTR retrotransposons. This is based on studies of human retroposons, including Alu (SINE) and L1 (LINE). Both elements are flanked by ~15-bp target site duplications, have the poly(A) 3′ termini, and are preferentially inserted into AT-rich regions (Weiner, Deininger, and Efstratiadis 1986). Similar features have been observed in processed pseudogenes (Moos and Gallwitz 1983; Vanin 1985). Based on these premises and the fact that L1 encodes the reverse transcriptase (Hattori et al. 1986), researchers have expected that Alu and processed pseudogenes are mobilized by the L1-encoded reverse transcriptase (Eickbush 1992; King 1992). Discovery of the TTTTAA-like target site shared by L1, Alu, and processed pseudogenes (Jurka 1997) provided additional strong arguments in favor to the direct link between the L1-encoded enzymatic machinery, Alu, and processed pseudogenes. Finally, it was shown directly that L1s generate processed pseudogenes in the human genome (Esnault, Maestre, and Heidmann 2000). Independently, the Okada group has shown 6 years ago that the 3′ tails of certain tRNA-derived SINEs are similar to the 3′ tails of certain CR1-like elements (Ohshima et al. 1996; Okada et al. 1997). In addition, reconstruction of ancient MIR (tRNA-derived SINE) and LINE2 (CR1-like non-LTR retrotransposons) from the human genome has shown that these elements also share similar 3′ tail regions (Smit and Riggs 1995; Smit 1996).

As reported in this manuscript, SINE3 and zebrafish CR1-like non-LTR retrotransposons share common structural features. Neither CR1-like elements nor SINE3 are flanked by target site duplications, theirs 3′ termini are composed of 3-bp to 5-bp microsatellites, and they share common 3′ ends. All these features suggest that transpositions of SINE3 depend on the enzymatic machinery encoded by CR1-like elements.

Different classes of transposable elements, including endogenous retroviruses and LTR retrotransposons (Jin and Bennetzen 1989; Kapitonov and Jurka 1999; Witte et al. 2001), "cut and paste" DNA transposons (Fedoroff, Wessler, and Shure 1983; Smit and Riggs 1996; Kapitonov and Jurka 1999; Jurka and Kapitonov 2001), and "rolling-circle" DNA transposons (Kapitonov and Jurka 2001) are composed of autonomous and nonautonomous elements. An autonomous element encodes a complete set of enzymes catalyzing its transpositions, whereas some or all these enzymes are not encoded by a nonautonomous element, which can be mobilized only when the necessary enzymes are provided in trans by the autonomous element. Presumably, the "autonomous-nonautonomous" *dyad* (the A-N *dyad*) is a characteristic common to all classes of transposable elements. Consequently, all SINEs can be considered as nonautonomous non-LTR retrotransposons.

Downloaded from https://academic.oup.com/mbe/article/20/5/694/985403 by U.S. Department of Justice user on 17 August 2022

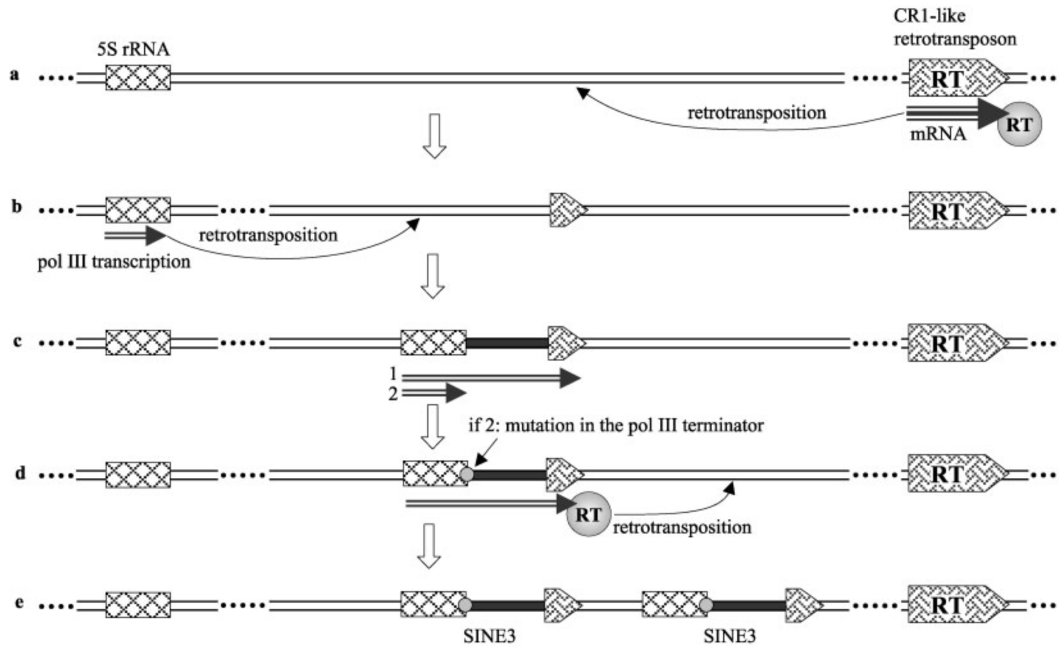

FIG. 5.—A model of the SINE3 origin. (*a*) An ancestral zebrafish genome harbors an expressed 5S rRNA gene/pseudogene, denoted by the hatched rectangle, and an active CR1-like non-LTR retrotransposon encoding reverse transcriptase (RT). The RT polyprotein, a gray circle, binds the 3′ end of the mRNA copy of the retrotransposon. As a result, its 5′ truncated cDNA copy is retrotransposed at a random site. (*b*) The mRNA copy of 5S rRNA is retrotransposed at a site close to the CR1 retrotransposon. (*c*) The black rectangle marks an internal region surrounded by the recently retrotransposed 5S RNA and CR1-like retrotransposon. It is possible (1) that the 5S rRNA pol III terminator is not functional because of the target site specificity or an abnormal processing of the 5S rRNA 3′ end. Another possibility (2) is that the pol III transcription of the retrotransposed 5S rRNA is terminated at its 3′ end. (*d*) If the pol III terminator is intact, a random mutation at the 3′ end of 5S rRNA abolishes it. As a result of the pol III read-through transcription, a SINE3 mRNA is formed. It is composed of 5S rRNA, the internal region and the CR1-like 3′-end. The 5S rRNA internal promoter regulates the transcription of SINE3. The RT polyprotein binds the SINE3 3′ end and catalyzes both the reverse transcription and the DNA cleavage, which are the main stages of the SINE3 retrotransposition (*e*) in the genome.

Since, as reported in this manuscript, the zebrafish genome harbors nearly 10,000 SINE3 elements, the "mystery" of the lack of SINEs derived from 5S rRNA (Weiner 2002) does not exist anymore. However, given the similar size and expression level of tRNA and 5S rRNA, the observed difference in a diversity of species colonized by SINEs derived from these molecules continues to be puzzling. SINEs derived from tRNA are present in mammals (Daniels and Deininger 1985; Smit and Riggs 1995; Shimamura et al. 1999), vertebrates (Kido et al. 1991; Ohshima et al. 1996; Ogiwara et al. 2002), invertebrates (Ohshima et al. 1993), and plants (Yoshioka et al. 1993; Deragon et al. 1994). However, SINEs derived from 5S rRNA are present in the zebrafish genome only. It is known that transcription of pol III internal promoters can be significantly modulated by DNA signals juxtaposed upstream of the transcribed region (Paule and White 2000). Presumably, the type 1 promoters in 5S rRNAs depend much more on these upstream signals than do type 2 promoters in tRNAs. As a result, the pol III promoter in a retroposed 5S rRNA copy remains silent or is expressed at a low level. If so, it is quite unlikely that 5S rRNA-derived SINE elements will be very common in eukaryotic species.

Although the last explanation is fairly simple and can be easily verified in experiments, there are other interesting biological considerations that may well apply. Efficient retrotransposition of tRNA-derived and 5S rRNA-derived SINEs would be unlikely unless the corresponding RNAs escaped their basic role in translation and avoided being bound by proteins and other RNA molecules interacting with and involved in regulation of functional tRNA/5S rRNA. Therefore, the observed common occurrence of tRNA-derived SINEs and scarcity of 5S rRNA-derived SINES can be explained by a presumption that tRNA molecules are much more capable "escapists and evaders" than 5S rRNA. In other words, tRNA molecules are not bound by as many different translation-related factors as 5S rRNAs, nor are they as tightly regulated in cells as 5S rRNAs. Interestingly, tRNA and 5S rRNA utilize two different pathways for export out of the nucleus into the cytoplasm (Izaurralde and Mattaj 1995; Pasquinelli et al. 1997; Arts et al. 1998; Ullman et al. 1999). It is possible that the tRNA-related pathway can be accidentally evaded more easily than the alternative one. Noteworthy, nucleocytoplasmic transport of 5S rRNA is mediated by L5 ribosomal protein that binds 5S rRNA at positions 27 to 44 (Scripture and Huber 1995). This region is highly modified in SINE3 (fig. 3). Presumably, it was necessary for evading the 5S rRNA nucleocytoplasmic transport.

Acknowledgments

We thank Alan M. Weiner and the anonymous referee for helpful comments on the manuscript. We are grateful to Jolanta Walichiewicz for help with illustrations.

Downloaded from https://academic.oup.com/mbe/article/20/5/694/985403 by U.S. Department of Justice user on 17 August 2022

We also thank Aneta Drazkiewicz for editing the manuscript. This work was supported by National Institutes of Health (Grant 2 P41 LM06252-04A1).

Literature Cited

Aparicio, S., J. Chapman, E. Stupka et al. (41 co-authors). 2002. Whole-genome shotgun assembly and analysis of the genome of fugu rubripes. Science **297**:1301–1310.

Arts, G. J., S. Kuersten, P. Romby, B. Ehresmann, and I. W. Mattaj. 1998. The role of exportin-t in selective nuclear export of mature tRNAs. EMBO J. **17**:7430–7441.

Bao, Z. 2002. Dr000031. (http://www.genetics.wustl.edu/fish_lab/repeats/).

Barciszewska, M. Z., M. Szymanski, V. A. Erdmann, and J. Barciszewski. 2001. Structure and functions of 5S rRNA. Acta. Biochim. Pol. **48**:191–198.

Cloutier, R., and P. E. Ahlberg. 1996. Interrelationships of fishes. Pp. 445–479 *in* M. L. Stiassny, L. R. Parenti, and G. D. Johnson, eds. Academic Press, San Diego.

Daniels, G. R., and P. L. Deininger. 1985. Repeat sequence families derived from mammalian tRNA genes. Nature **317**: 819–822.

Deragon, J. M., B. S. Landry, T. Pelissier, S. Tutois, S. Tourmente, and G. Picard. 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. 1992. Transposing without ends: the non-LTR retrotransposable elements. New Biol. **4**:430–440.

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

Fedoroff, N., S. Wessler, and M. Shure. 1983. Isolation of the transposable maize controlling elements Ac and Ds. Cell **35**: 235–242.

Gornung, E., S. De Innocentiis, F. Annesi, and L. Sola. 2000. Zebrafish 5S rRNA genes map to the long arms of chromosome 3. Chromosome Res. **8**:362.

Hattori, M., S. Kuhara, O. Takenaka, and Y. Sakaki. 1986. L1 family of repetitive DNA sequences in primates may be derived from a sequence encoding a reverse transcriptase-related protein. Nature **321**:625–628.

Izaurralde, E., and I. W. Mattaj. 1995. RNA export. Cell **81**: 153–159.

Jin, Y. K., and J. L. Bennetzen. 1989. Structure and coding properties of Bs1, a maize retrovirus-like transposon. Proc. Natl. Acad. Sci. USA **86**:6235–6239.

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

———. 2000. Repbase update: a database and an electronic journal of repetitive elements. Trends Genet. **16**:418–420.

Jurka, J., and V. V. Kapitonov. 2001. PIFs meet Tourists and Harbingers: a superfamily reunion. Proc. Natl. Acad. Sci. USA **98**:12315–12316.

Jurka, J., P. Klonowski, V. Dagman, and P. Pelton. 1996. CENSOR—a program for identification and elimination of repetitive elements from DNA sequences. Comput. Chem. **20**:119–121.

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

———. 2001. Rolling-circle transposons in eukaryotes. Proc. Natl. Acad. Sci. USA **98**:8714–8719.

———. 2002. CR1-4_DR, a family of CR1-like non-LTR retrotransposons in zebrafish. Repbase Reports **2**:6.

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

King, C. C. 1992. Modular transposition and the dynamical structure of eukaryote regulatory evolution. Genetica **86**:127–142.

Little, R. D., and D. C. Braaten. 1989. Genomic organization of human 5S rDNA and sequence of one tandem repeat. Genomics **4**:376–383.

Moos, M., and D. Gallwitz. 1983. Structure of two human beta-actin-related processed genes one of which is located next to a simple repetitive sequence. EMBO J. **2**:757–761.

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

———. 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., M. Hamada, Y. Terai, and N. Okada. 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., R. Koishi, M. Matsuo, and N. Okada. 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.

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

Okada, N., and K. Ohshima. 1995. Evolution of tRNA-derived SINEs. Pp. 61–79 *in* R. J. Maraia, ed. The impact of short interspersed elements (SINEs) on the host genome. R. G. Landes Company, Austin, Texas.

Pasquinelli, A. E., M. A. Powers, E. Lund, D. Forbes, and J. E. Dahlberg. 1997. Inhibition of mRNA export in vertebrate cells by nuclear export signal conjugates. Proc. Natl. Acad. Sci. USA **94**:14394–14399.

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

Phillips, R. B., and K. M. Reed. 2000. Localization of repetitive DNAs to zebrafish (*Danio rerio*) chromosomes by fluorescence in situ hybridization (FISH). Chromosome Res. **8**:27–35.

Pieler, T., J. Hamm, and R. G. Roeder. 1987. The 5S gene internal control region is composed of three distinct sequence elements, organized as two functional domains with variable spacing. Cell **48**:91–100.

Sajdak, S. L., K. M. Reed, and R. B. Phillips. 1998. Intraindividual and interspecies variation in the 5S rDNA of coregonid fish. J. Mol. Evol. **46**:680–688.

Schmid, C. W. 1996. Alu: structure, origin, evolution, significance and function of one-tenth of human DNA. Prog. Nucleic Acid Res. Mol. Biol. **53**:283–319.

Scripture, J. B., and P. W. Huber. 1995. Analysis of the binding of Xenopus ribosomal protein L5 to oocyte 5 S rRNA: the major determinants of recognition are located in helix III-loop C. J. Biol. Chem. **270**:27358–27365.

Shimamura, M., H. Abe, M. Nikaido, K. Ohshima, and N. Okada. 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.

Downloaded from https://academic.oup.com/mbe/article/20/5/694/985403 by U.S. Department of Justice user on 17 August 2022

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

Smit, A. F. 1996. The origin of interspersed repeats in the human genome. Curr. Opin. Genet. Dev. **6**:743–748.

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

———. 1996. Tiggers and DNA transposon fossils in the human genome. Proc. Natl. Acad. Sci. USA **93**:1443–1448.

Takahashi, K., Y. Terai, M. Nishida, and N. Okada. 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.

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

Ullman, K. S., S. Shah, M. A. Powers, and D. J. Forbes. 1999. The nucleoporin nup153 plays a critical role in multiple types of nuclear export. Mol. Biol. Cell **10**:649–664.

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

Vanin, E. F. 1985. Processed pseudogenes: characteristics and evolution. Annu. Rev. Genet. **19**:253–272.

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

———. 2002. SINEs and LINEs: the art of biting the hand that feeds you. Curr. Opin. Cell. Biol. **14**:343–350.

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

Witte, C. P., Q. H. Le, T. Bureau, and A. Kumar. 2001. Terminal-repeat retrotransposons in miniature (TRIM) are involved in restructuring plant genomes. Proc. Natl. Acad. Sci. USA **98**:13778–13783.

Yoshioka, Y., S. Matsumoto, S. Kojima, K. Ohshima, N. Okada, and Y. Machida. 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.

Pierre Capy, Associate Editor

Accepted December 10, 2002

Downloaded from https://academic.oup.com/mbe/article/20/5/694/985403 by U.S. Department of Justice user on 17 August 2022