Mol Gen Genomics (2004) 271: 98–102
DOI 10.1007/s00438-003-0961-9

ORIGINAL PAPER

K. Szafranski · T. Dingermann · G. Glöckner
T. Winckler

Template jumping by a LINE reverse transcriptase has created a SINE-like 5S rRNA retropseudogene in *Dictyostelium*

Received: 15 June 2003 / Accepted: 10 November 2003 / Published online: 2 December 2003

Abstract Short interspersed nuclear elements (SINEs) are non-autonomous retroelements that mimic the 3′ ends of so-called long interspersed nuclear elements (LINEs) to ensure their propagation by proteins encoded by autonomous LINEs. The *Dictyostelium discoideum* genome contains a family of LINE-like retrotransposons that specifically target tRNA genes for integration (TRE elements). We describe here a retrotransposed ribosomal 5S RNA pseudogene in the *D. discoideum* genome that contains at its 3′ end an 8-bp sequence derived from the 3′ end of a TRE and a polyadenine tail. The r5S "retropseudogene" is flanked by target-site duplications that are characteristic for TREs, and is inserted upstream of a tRNA gene, just like a typical TRE. The *D. discoideum* r5S retropseudogene has structural features of a SINE, but has not been amplified, probably due to the 5′-truncation that occurred upon its initial retrotransposition. The discovery of this *D. discoideum* r5S retropseudogene reveals that SINEs can be created de novo during reverse transcription of LINE transcripts, if the LINE-encoded reverse transcriptase dissociates from the LINE RNA and jumps to other cellular RNAs—particularly genes transcribed by RNA polymerase III—to create continuous mixed cDNAs.

Keywords *Dictyostelium* · Retrotransposon · Long interspersed nuclear elements (LINEs) · Short interspersed nuclear elements (SINEs) · RNA polymerase III

Communicated by G. P. Georgiev

K. Szafranski · G. Glöckner
Department of Genome Analysis, IMB Jena, Beutenbergstrasse 11, 07745 Jena, Germany

T. Dingermann · T. Winckler (✉)
Institut für Pharmazeutische Biologie, Biozentrum, Universität Frankfurt/M., Marie-Curie-Strasse 9, 60439 Frankfurt am Main, Germany
E-mail: winckler@em.uni-frankfurt.de
Tel.: +49-69-79829648
Fax: +49-69-79829662

Indroduction

Almost every organism harbors a multitude of molecular parasites known as transposable elements in its genome. Transposable elements can influence their host genomes in many ways. Thus, they may expand genome size and complexity, rearrange genomic DNA, mutagenize host genes, and alter the expression levels of nearby genes (Kazazian and Moran 1998; Smit 1999; Esnault et al. 2000; Kazazian 2000; Ostertag and Kazazian 2001). Class I elements, or retroelements, amplify by reverse transcription of RNA intermediates into DNA by means of an RNA-directed DNA polymerase (reverse transcriptase, RT). The Class II elements (DNA transposons) move from DNA to DNA locus by a cut-and-paste mechanism catalyzed by a transposase.

The retrotransposons are divided into two large families on the basis of phylogenetic analysis of their encoded RTs (Xiong and Eickbush 1990) and structural features such as the presence or absence of long terminal repeats (LTRs) flanking the retrotransposon-encoded genes. Non-LTR retrotransposons are also referred to as long interspersed nuclear elements (LINEs). Transposable elements are termed autonomous if they code for the proteins required for transposition and have limited requirements for host factors. Non-autonomous transposable elements, on the other hand, do not code for proteins and seem to borrow required enzyme functions among those encoded by autonomous transposable elements present in the same cells. Examples of non-autonomous Class I and Class II elements are short interspersed nuclear elements (SINEs) and miniature inverted-repeat transposable elements (MITEs), respectively (Okada 1991; Feschotte et al. 2002). SINEs are usually 100–400 bp long and are mostly derived from tRNA gene transcripts synthesized by RNA polymerase III (polIII). The human *Alu* element that is derived from the gene coding for the signal recognition particle component 7SL is also classified as a SINE element. SINEs

harbor internal promoters that permit their transcription by polIII.

LINEs use an ancient retrotransposition mechanism referred to as target-primed reverse transcription (TPRT; Luan et al. 1993; Cost et al. 2002). The TPRT model states that a LINE-encoded endonuclease activity nicks the genomic target DNA and the resulting 3′ hydroxyl group is then used to prime reverse transcription by RT. Experimental data suggest that the recognition of RNA sequences at the 3′ end of the LINE transcript is important for efficient reverse transcription (Luan and Eickbush 1995; Moran et al. 1996). Many SINEs have 3′ ends that are similar to those of LINEs found in the same cell. This observation has led to the speculation that SINEs may be mobilized by TPRT by borrowing the required enzyme functions from LINEs (Okada 1991; Okada et al. 1997). Recently direct evidence has been obtained in support of this hypothesis (Kajikawa and Okada 2002).

Dictyostelium discoideum is a single-cell eukaryote with a very compact genome. The average spacing between coding regions is less than 1 kb (Glöckner et al. 2002), leaving limited space within which TEs can move and amplify without causing deleterious mutations. Nevertheless, about 10% of the *D. discoideum* genome is composed of transposable elements (Glöckner et al. 2001). A special adaptation of *D. discoideum* transposable elements that allows them to avoid insertion mutagenesis of functional genes is specific integration close to tRNA genes (reviewed in Winckler et al. 2002). A family of seven LINE-like tRNA gene-targeted retrotransposons (TREs) has evolved mechanisms that ensure their integration ~50 bp upstream and 100 bp downstream of tRNA genes, respectively (these are therefore named TRE5 and TRE3). We have inspected the recently completed genomic sequence of chromosome 2, which represents about 25% of the total *D. discoideum* genome, for the distribution of TREs. We found that TREs are scattered all over the chromosome, closely following the distribution of tRNA genes (Winckler et al. 2002).

The ribosomal genes of *D. discoideum* are found not on any of the six normal chromosomes, but rather are encoded on an ~88-kb palindromic minichromosome that is present in about 100 copies per cell (Sucgang et al. 2003). We have previously shown that the r5S gene of *D. discoideum* is most probably transcribed by polIII as in other eukaryotes (Hofmann et al. 1993). Here we present evidence which indicates that a copy of the *D. discoideum* r5S gene was mobilized from the minichromosome and moved to chromosome 2 by a LINE-like TRE5 element. A fusion cDNA made up of the r5S gene sequence and the 3′ end of the TRE5 element was created by this retrotransposition event, suggesting that template jumping from LINE transcripts to cellular polIII-transcripts may provide a route for the generation of SINEs in eukaryotic genomes.

Materials and methods

The methods used for library construction and sequencing have been described previously (Glöckner et al. 2001). An error-prone assembly of the transposon copy locus from genome-wide shotgun sequences was performed by using information from copy-specific sequence morphology and paired reads that originated from both ends of clone inserts with lengths of up to 5000 bp. Target-site duplications that flank the transposon copies confirmed the consistency of the resulting assembly. The *Dictyostelium* Genome Project can be accessed at http://genome.imb-jena.de/dictyostelium or http://www.uni-koeln.de/dictyostelium/.

Results and discussion

A single 5S rRNA-like sequence is present on chromosome 2 of D. discoideum

D. discoideum TREs specifically target tRNA genes for integration. We wanted to determine whether ribosomal 5S genes, which, like tRNA genes, are also transcribed by polIII, are targets for TRE integration. We analyzed the sequence data from the *Dictyostelium* genome project, and found no evidence for natural insertions of TREs near r5S genes. However, we found a single r5S gene copy that was not located in the extrachromosomal ribosomal gene palindrome. This r5S derivative, located on chromosome 2, spans 104 bp and contains only two mismatches relative to the original 119 bp ribosomal 5S gene (the first 15 bp are missing; Fig. 1A). The coding sequence of a canonical *D. discoideum* r5S gene is followed by the sequence TTATTTATTTTTT (Fig. 1A); the T_7 stretch probably acts as the transcription terminator. However, the r5S gene derivative on chromosome 2 lacks the T_7 stretch, suggesting that it represents part of a reverse transcription product derived from a ribosomal 5S precursor RNA. We therefore refer to the r5S derivative on chromosome 2 as a 'retropseudogene'. Immediately downstream of the r5S retropseudogene is a 9-bp sequence followed by a long polyadenine stretch (Fig. 1A). We interpreted this sequence as the remains of an ancient TRE5 element that retains only limited similarity with modern TRE5s (Fig. 1B). Importantly, a 13-bp target-site duplication (TSD) was found upstream of the r5S retropseudogene and downstream of the polyadenine tract (Fig. 1A), suggesting that the r5S retropseudogene resulted from a retrotransposition event that led to the fusion of the r5S cDNA with the 3′ end of a TRE5 element.

If the r5S retropseudogene was indeed copied onto chromosome 2 by a TRE5, we would expect to find the r5S retropseudogene about 50 bp upstream of a tRNA gene. Inspection of the chromosomal sequence upstream of the r5S retropseudogene revealed the presence of a full-length DGLT-A retrotransposon that, like the TRE5s, is known to specifically target the flanking regions upstream of tRNA genes (reviewed in Winckler et al. 2002). DGLT-A elements insert ~30 bp upstream of tRNA genes (Hofmann et al. 1991). Their

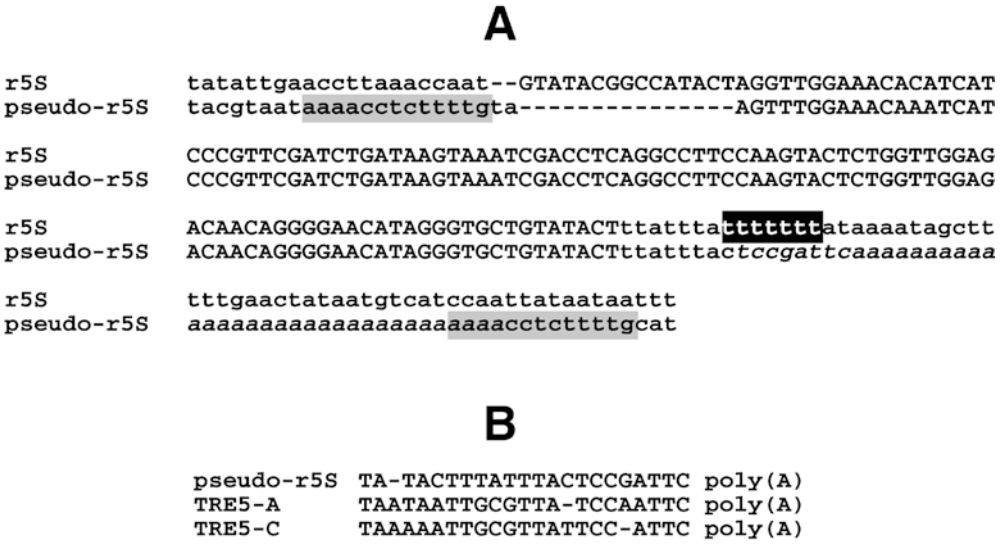

Fig. 1A, B Structure of the r5S retropseudogene. **A** Comparison of the *D. discoideum* ribosomal 5S gene sequence with that of the r5S retropseudogene. Flanking sequences are shown in *lower case letters*. The TRE5-like 3′ end of the r5S retropseudogene including the poly(A) tail is shown in *lower case italics*. The putative transcription termination signal of the original r5S gene is shown in a *black box*. Target site duplications are shown in *grey boxes*. The locus containing the r5S retropseudogene is located on chromosome 2 between HAPPY markers DH2917 and DH3211 (http://genome.imb-jena.de/dictyostelium) and detailed under GenBank Accession No. AY293825. **B** Comparison of the 3′ ends of the r5S retropseudogene and TRE5

LTRs may be interrupted by subsequent integration of a TRE5 about 50 bp upstream of a tRNA gene flanked by a DGLT-A (i.e. 20 bp inside the DGLT-A LTR). This situation is also present upstream of the r5S retropseudogene locus, where one DGLT-A LTR was interrupted by the insertion of a TRE5-A.2 (not shown). Importantly, a Leu(UAA) tRNA gene is located 49 bp upstream of this TRE5-A.2 element. Thus, we can conclude that the integration of the r5S retropseudogene was the first of three sequential integration events that occurred upstream of the Leu(UAA) tRNA gene.

The r5S retropseudogene gene lacks the first 15 bp of the normal r5S gene sequence, suggesting abortive reverse transcription of the r5S RNA. Two additional, non-templated, nucleotides (TA) are inserted between the 5′ end of the r5S retropseudogene and the TSD (Fig. 1A). The generation of 5′ deleted TRE5 copies and insertion of non-templated nucleotides at the 5′ ends of integrated TRE5 copies are typical for TRE5 retrotransposition in *D. discoideum* (Beck et al. 2002).

Creation of a retropseudogene by a LINE RT?

Taken together the observations described above strongly support the view that the r5S retropseudogene was created by retrotransposition in *trans* by a TRE5-encoded RT. Yet it is unusual that a short sequence derived from the 3′ end of a TRE5 should have been fused to the reverse transcript of the 5S RNA. The r5S retropseudogene has structural features reminiscent of those of a SINE. Most SINEs consist of the remains of a polIII-transcribed gene and a sequence with similarity to the 3′ end of a resident LINE. The r5S retropseudogene consists of an almost complete r5S gene and a 3′ tail derived from a LINE-like retrotransposon, although the similarity to TRE5 elements is restricted to 8 bp plus a poly(A) tract (Fig. 1B). The *D. discoideum* r5S retropseudogene is a single-copy gene that has not been amplified like a typical SINE. The reason why the r5S retropseudogene did not give rise to a *Dictyostelium* r5S-derived SINE family may be the fact that 5′ truncation occurred upon its very first retrotransposition (i.e. upon its creation). The internal promoter elements of the r5S gene were retained and may allow transcription of the r5S retropseudogene. However, the r5S retropseudogene lacks the first 15 bp of the normal ribosomal 5S gene, which may be important for double-strand RNA formation in the secondary structure of the mature r5S RNA (stem I). Mutations in stem I are known to drastically inhibit the maturation of precursor 5S transcripts, preventing packaging of the 5S RNA into ribosomes and eventually leading to the degradation of the entire 5S RNA (Levinger et al. 1992). Thus, the r5S retropseudogene may be transcribed into a highly unstable RNA that is not accessible to TRE5 proteins for reverse transcription and integration. Alternatively, lack of the first 15 bp of the r5S retropseudogene gene may prevent its transcription by interfering with binding of polIII transcription factors, such that no SINE transcripts are formed. Unfortunately each of these hypotheses is difficult to test experimentally since the consequence of either of the two alternatives would be very low or undetectable levels of transcripts.

A proposed mechanism for the generation of the SINE-like sequence by template switching during reverse transcription of a LINE RNA

How are SINEs generated? Okada and coworkers (Ohshima et al. 1993) have discussed evidence which suggests that SINEs may be derived from "strong-stop DNAs", which are intermediates of reverse transcription of retroviruses and retrovirus-like LTR retrotransposons. However, the authors failed to explain how the tRNA-derived strong-stop DNAs could become fused to LINE-like 3′ ends including the characteristic poly(A) tails. On the other hand, similarities between SINE 3′ ends and LINEs residing in the same cell prompted the same authors to speculate that LINE proteins may be hijacked by SINE RNAs for reverse transcription and integration (Okada et al. 1997; Kajikawa and Okada 2002). Malik and Eickbush (1998) have suggested that a class of SINEs that lack polIII-derived sequences can be understood as extremely 5′-truncated LINEs that retained only a few bases of the original LINE including a poly(A) tail. These atypical SINEs, however, would be unable to further retrotranspose unless they integrated by chance downstream of a promoter that could facilitate their transcription. Hence one could imagine that a tRNA-derived SINE would originate when a 5′-truncated LINE integrated downstream of a tRNA gene. Transcription of this tRNA gene beyond the polIII transcription termination signal could theoretically create a SINE-like transcript that could be retrotransposed in *trans* by an autonomous LINE. However, this seems unlikely since tRNA-derived SINEs are created from processed tRNA gene transcripts (Okada 1991). This hypothesis also fails to explain the generation of the *Dictyostelium* r5S retropseudogene, because (1) no natural r5S-targeted TRE5 insertions have yet been found, (2) TRE5 usually targets the 5′ flanks of polIII genes (i.e. tRNA genes) rather than their 3′ ends, and (3) the oligo(T) termination signal downstream of the *D. discoideum* r5S gene is missing in the r5S retropseudogene, arguing against read-through transcription of a r5S gene to create a tRNA-TRE fusion RNA.

The newly discovered *Dictyostelium* r5S retropseudogene provides some hints that SINEs may originate from co-reverse transcription of LINE transcripts with other cellular RNAs. Bibillo and Eickbush (2002) have recently shown in elegant in vitro experiments that the RT encoded by the non-LTR retrotransposon R2 of *Bombyx mori* is able to switch between different RNA templates to produce continuous fusion cDNAs. The ability to switch templates seems to be an intrinsic property of the RTs encoded by non-LTR retrotransposons that is required to join the newly synthesized cDNA to the upstream genomic sequences (Bibillo and Eickbush 2002). RTs occasionally lose contact with their retroelement-derived RNA templates, resulting in abortive reverse transcription of the mobile retroelement. One could therefore imagine that a second RNA might compete with the upper DNA strand at the integration site for binding to the active site of the RT, such that the RT can continue cDNA synthesis on the second RNA template (Fig. 2). After completion of reverse transcription of the second RNA, the RT could jump to the upper DNA strand to join the cDNA to the upstream cleavage site. It is tempting to speculate that co-reverse transcription of LINE RNAs and polIII transcripts may occur at low levels in any TPRT reaction, thereby providing a means for the generation of SINEs in many organisms. Creation of SINEs by template jumping may be favored in cellular compartments that are enriched in polIII transcripts, as the efficiency of template jumping correlates with increasing RNA concentrations (Bibillo and Eickbush 2002). Ribosomal 5S transcripts should be enriched in the nucleolus, the site of ribosome assembly. Notably, a recent report has suggested that tRNA processing is localized in the nucleolus (Bertrand et al. 1998). Although it is unknown whether LINE retrotransposition occurs in the

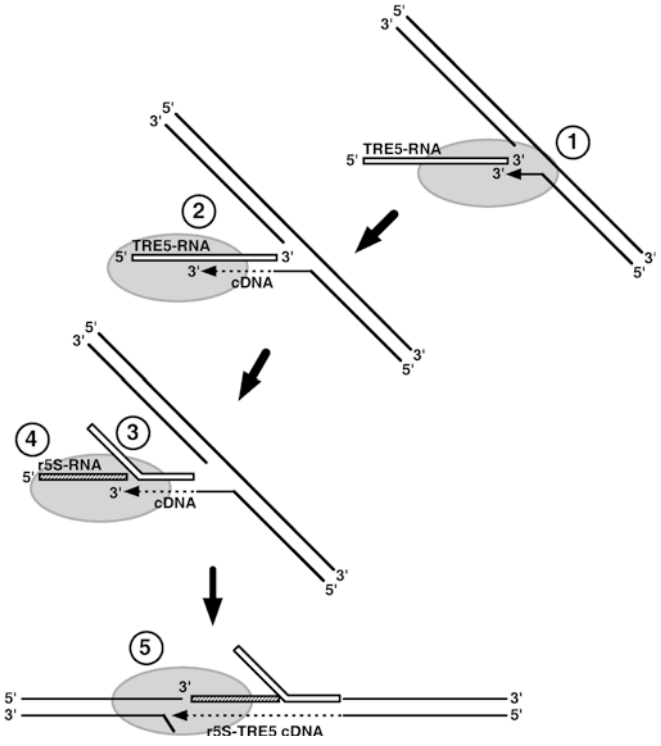

Fig. 2 Model for template jumping which accounts for the creation of fusion cDNAs during the TPRT reaction. The TRE5 transcript bound to TRE5 ORF2 protein (*grey circle*) is carried to an appropriate integration target site upstream of a tRNA gene. The lower strand of the genomic DNA is cleaved (step 1) and reverse transcription starts using the free 3′-OH of the genomic DNA strand as primer (step 2). Eventually the ORF2 protein may lose contact with the TRE5 template RNA, resulting in abortive reverse transcription (step 3). Usually this would promote template switching by the RT onto the upper DNA strand for 5′ end joining after second-strand cleavage (step 5). At this step a second RNA may compete with the genomic DNA template for binding to the RT active site, allowing the RT to further extend cDNA synthesis by jumping to the new RNA template (step 4). Run-off reverse transcription of the second RNA is then followed by a second template switch to the genomic DNA for 5′ end joining (step 5)

nucleolus, we propose that the high concentrations of polIII gene transcripts in the nucleolus may favor template switching by the RT during reverse transcription to create SINEs. Our hypothesis not only suggests how SINE may be generated, it also helps to explain why many SINEs share conserved 3′ ends with fellow LINEs: they received them upon their creation and then conserved them under selection to maintain active *trans*-mobilized populations. cDNA fusions of LINE-derived and other cellular transcripts may also occur and copy pieces of genomic information to new chromosomal locations and add new features to preexisting genes. Hence the discovery of naturally occurring cDNA fusions in the *D. discoideum* genome adds yet another feature to the list of effects that non-LTR retrotransposons have on the evolution of eukaryotic genomes.

Acknowledgements We are grateful to an anonymous referee for comments that helped to significantly improve the manuscript. Our work in the international *Dictyostelium* Genome Sequencing Consortium is supported by grants from the Deutsche Forschungsgemeinschaft.

References

Beck P, Dingermann T, Winckler T (2002) Transfer RNA gene-targeted retrotransposition of *Dictyostelium* TRE5-A into a chromosomal UMP synthase gene trap. J Mol Biol 318:273–285

Bertrand E, Houser-Scott F, Kendall A, Singer RH, Engelke DR (1998) Nucleolar localization of early tRNA processing. Genes Dev 12:2463–2468

Bibillo A, Eickbush TH (2002) The reverse transcriptase of the *R2* non-LTR retrotransposon: continuous synthesis of cDNA on non-continuous RNA templates. J Mol Biol 316:459–473

Cost GJ, Feng Q, Jacquier A, Boeke JD (2002) Human L1 element target-primed reverse transcription in vitro. EMBO J 21: 5899–5910

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

Feschotte C, Zhang X, Wessler SR (2002) Miniature inverted-repeat transposable elements and their relationships to established DNA transposons. In: Craig NL, Craigie R, Gellert M, Lambowitz AM (eds) Mobile DNA II. ASM Press, Washington, D.C., pp 1147–1157

Glöckner G, Szafranski K, Winckler T, Dingermann T, Quail M, Cox E, Eichinger L, Noegel AA, Rosenthal A (2001) The complex repeats of *Dictyostelium discoideum*. Genome Res 11:585–594

Glöckner G, et al (2002) Sequence and analysis of chromosome 2 of *Dictyostelium discoideum*. Nature 418:79–85

Hofmann J, Schumann G, Borschet G, Gosseringer R, Bach M, Bertling WM, Marschalek R, Dingermann T (1991) Transfer RNA genes from *Dictyostelium discoideum* are frequently associated with repetitive elements and contain consensus boxes in their 5′-flanking and 3′-flanking regions. J Mol Biol 222: 537–552

Hofmann J, Winckler T, Hanenkamp A, Bukenberger M, Schumann G, Marschalek R, Dingermann T (1993) The *Dictyostelium discoideum* 5S rDNA is organized in the same transcriptional orientation as the other rDNAs. Biochem Biophys Res Comm 191:558–564

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

Kazazian HH (2000) L1 retrotransposons shape the mammalian genome. Science 289:1152–1153

Kazazian HH, Moran JV (1998) The impact of L1 retrotransposons on the human genome. Nat Genet 19:19–24

Levinger L, Vasisht V, Greene V, Arjun I (1992) The effects of stem I and loop A on the processing of 5 S rRNA from *Drosophila melanogaster*. J Biol Chem 267:23683–23687

Luan DD, Eickbush TH (1995) RNA template requirements of target DNA-primed reverse transcription by the *R2* retrotransposable element. Mol Cell Biol 15:3882–3891

Luan DD, Korman MH, Jakubczak JL, Eickbush TH (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

Malik HS, Eickbush TH (1998) The RTE class of non-LTR retrotransposons is widely distributed in animals and is the origin of many SINEs. Mol Biol Evol 15:

Moran JV, Holmes SE, Naas TP, DeBerardinis RJ, Boeke JD, Kazazian HH (1996) High frequency retrotransposition in cultured mammalian cells. Cell 87:917–927

Ohshima K, Koishi R, Matsuo M, 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

Okada N (1991) SINEs. Curr Opin Genet Dev 1:498–504

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

Ostertag EM, Kazazian HH (2001) Biology of mammalian L1 retrotransposons. Annu Rev Genet 35:501–538

Smit AFA (1999) Interspersed repeats and other momentos of transposable elements in mammalian genomes. Curr Opin Genet Dev 9:657–663

Sucgang R, Chen G, Liu W, Lindsay R, Lu J, Muzny D, Shaulsky G, Loomis W, Kuspa A (2003) Sequence and structure of the extrachromosomal palindrome encoding the ribosomal RNA genes in *Dictyostelium*. Nucleic Acids Res 31:2361–2368

Winckler T, Dingermann T, Glöckner G (2002) *Dictyostelium* mobile elements: strategies to amplify in a compact genome. Cell Mol Life Sci 59:2097–2111

Xiong Y, Eickbush T (1990) Origin and evolution of retroelements based upon their reverse transcriptase sequences. EMBO J 9:3353–3362