The origin of interspersed repeats in the human genome

Arian FA Smit

Over a third of the human genome consists of interspersed repetitive sequences which are primarily degenerate copies of transposable elements. In the past year, the identities of many of these transposable elements were revealed. The emerging concept is that only three mechanisms of amplification are responsible for the vast majority of interspersed repeats and that with each autonomous element a number of dependent non-autonomous sequences have co-amplified.

Addresses
Department of Molecular Biotechnology, University of Washington, Box 352145, Seattle, Washington 98195, USA; e-mail: asmit@u.washington.edu

Current Opinion in Genetics & Development 1996, 6:743–748

 ISSN 0959-437X

Abbreviations

Ac	*Activator*
ERVs	endogenous retroviruses
HERV	human ERV
LINEs	long interspersed nuclear elements
LTRs	long terminal repeats
MaLRs	mammalian LTR-retrotransposons
MIR	mammalian-wide interspersed repeat
MT	mouse transposon
ORR	origin region repeat
RLEs	retrovirus-like elements
SINEs	short interspersed nuclear elements

Introduction

Transposable elements are exemplary selfish genes [1]. These sequences may not be selectively advantageous for their host genome but nevertheless can survive over evolutionary time through the efficient heritable production of accurate copies (replicative transposition); when the original source becomes transpositionally inactive (e.g. through genetic drift), one or more of its progeny may continue its existence. In the process, the genome becomes littered with copies which — in the absence of a specific way to delete these sequences — decay over time until they blend into the background. Originally detected by hybridization experiments, the detection level of these interspersed repeats has increased dramatically with the accumulation of genomic sequence data and computer-aided analysis. By using a database of interspersed repeat consensus sequences and the computer program RepeatMasker, 36% of a human genomic sequence is found, on average, to be derived from mobile elements (Table 1). It is likely that much more of our DNA has accumulated in this fashion but has decayed too much to be recognized as such at present. In this review, I discuss recent revelations about the identity and mechanism of these transposable elements and categorize them by the three main transpositional systems known to exist in eukaryotes.

SINEs and LINEs

Transposition can occur via reverse transcription of an RNA intermediate, retro(trans)position, or via excision and reintegration of the DNA itself — DNA transposition. Most, if not all, retroposed repeats are either short interspersed nuclear elements (SINEs), long interspersed nuclear elements (LINEs), or retrovirus-like elements, and the first two form by far the biggest fraction of human interspersed repeats (Table 1). Although introduced merely to distinguish mammalian repeats by length, the names SINE and LINE now denote well defined elements present in all eukaryotes (Fig. 1).

The 100–400 bp long SINEs are characterized by an internal (mobile) polymerase III promoter that ensures a fair chance for transcriptional activity of new copies. This promoter is present in a tRNA-derived region in all SINEs [2–4], except in the primate *Alu* and rodent *B1* SINEs, which have a common origin in a 7SL RNA derived proto-*Alu* element [5]. The preponderance of tRNA-derived SINEs is at odds with the fact that pseudogenes of many other small structural RNAs are more common than those of tRNAs; however, it may be explained by the appearance that tRNA expression, unlike that of other small RNAs, is truly independent of upstream promoter elements and, thus, is self-sufficient [6,7]. This notion is supported by the observation that the key event in the emergence of the *Alu*/*B1* family was a 2 bp mutation in a 7SL RNA gene creating a tRNA-like polymerase III promoter [5].

The 6–8 kb LINE1 has been active in mammalian genomes since before the marsupial–eutherian split and has created copies that form at least 15% of our genome (Table 1); it encodes a reverse transcriptase and other proteins necessary for retrotransposition. LINE1 is a member of a widespread monophyletic group of retrotransposons [8], copies of which are generally characterized by variably truncated 5′ ends. In a model for LINE transposition [9], substantiated for a LINE-like element in insects [10,11•], the reverse transcriptase recognizes the 3′ end of the LINE transcript and initiates reverse transcription simultaneously with integration by using nicked genomic DNA as a primer. Loss of 3′ terminal sequences during transposition is thus avoided, whereas the 5′ end is often deleted through degradation or recombination. An internal polymerase II promoter, present in at least some LINE-like elements, assures expression that is relatively position-independent.

SINEs and LINEs share a variable length insertion duplication site and a poly A or simple repeat tail, suggesting that SINE mobility is dependent on proteins provided by LINEs rather than by retroviruses (the only

Table 1

Interspersed repeat composition of the human genome.

CG level and total size database entries		SINEs *Alu*	SINEs MIR	LINEs LINE1	LINEs LINE2	Elements with LTRs HERVs	Elements with LTRs MaIRs	Elements with LTRs others	DNA transposons *mariner*	DNA transposons others	Unclassified elements	Total
36–43% GC	Number:	885	494	939	351	80	261	48	15	283	80	
(4102 kb)	Fraction:	5.7%	1.6%	20.5%	2.1%	1.7%	3.1%	0.8%	0.16%	1.7%	1.0%	38.5%
43–52% GC	Number:	1182	290	252	178	21	75	19	2	101	38	
(1724 kb)	Fraction:	17.9%	2.1%	6.1%	2.6%	0.7%	1.9%	0.7%	0.7%	1.3%	0.7%	34.1%
52–63% GC	Number:	996	110	168	90	12	42	13	0	44	15	
(1225 kb)	Fraction:	20.2%	0.9%	4.6%	1.4%	0.5%	1.1%	0.7%	0.0%	0.5%	0.4%	30.3%
Extrapolation to a 3 billion bp genome												
Copy number (in thousands)		1188	402	593	271	50	167	34	8	192	60	2969
Fraction of total genome		10.0%	1.7%	14.6%	2.1%	1.3%	2.6%	0.7%	0.1%	1.5%	0.8%	35.5%

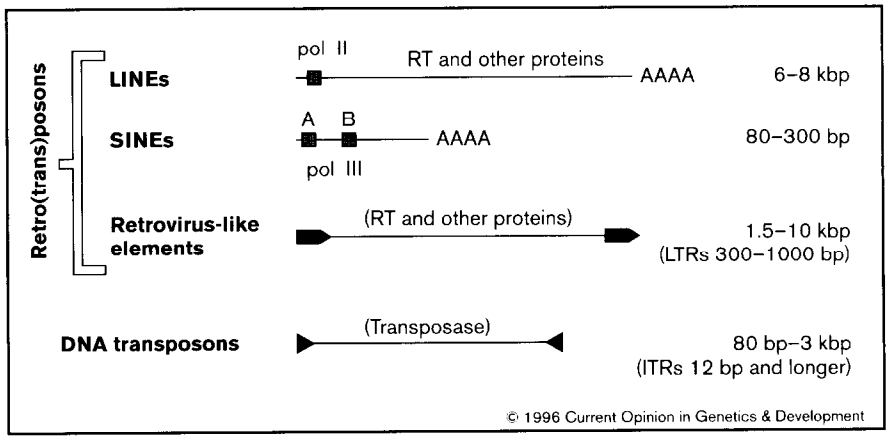
The data in this table are based on the analysis of all unique human genomic GenBank entries >40 kb as of June 1996; a total of 7051 kb derived from 40 distinct loci. For this we used an extended version of the database of human interspersed repeat consensus sequences [50] and the program RepeatMasker [51]. In calculating the number of repeats, fragmented sequences were counted as one. The sequences have been pooled by the GC content to show the differential repeat distribution in AT- and GC-rich DNA (see Fig. 2). As ~60% of human DNA is <43% GC, 30% is 43–52% GC, and 3–5% is >52% GC [52] (the rest is satellite DNA relatively devoid of interspersed repeats) and these fractions comprise 58%, 24.5% and 17.5% of the analyzed sequences respectively, adjustments were made to calculate the genome-wide numbers (GC-rich DNA is probably overrpresented in the database as a consequence of its higher gene density [52]). Except for *Alu*, probably all numbers are underestimates – especially those of MIR and LINE2 – as very old copies/members probably escape detection.

other known source of reverse transcriptase in mammalian cells; see below). This hunch was boosted recently by the observation that the 3′ 60–80 bp of two separate SINEs found in turtles and salmons correspond to the 3′ end of LINE-like elements in the same genomes [12••]. The origin of such SINEs is easily explained by the integration of a tRNA (derived) reverse transcript near the end of a LINE element; new transcripts from the polymerase III promoter will contain a 3′ end recognized by the LINE1 reverse transcriptase and a mobile element is born. The first indication for such a SINE–LINE relationship actually came from a report by Szemraj *et al.* [13•], who found that the bovine *Bov-B* (also known as *Art-2* or *Pst*) SINE could be extended to include a LINE-like reverse transcriptase coding region. The 3′ end of *Bov-B* is shared by the tRNA-derived SINE *Bov-tA* [14]. *Bov-B* and *Bov-tA* are detected in true ruminants only and in no other mammals [15,16], whereas repetitive sequences 75% similar to *Bov-B* are present in snakes [17•], strongly suggesting an invasion of this LINE-like element ~30 million years ago in the ruminant germline.

Besides more than a million *Alu*s, the human genome contains about half a million recognisable tRNA-derived SINEs named MIR (for mammalian-wide interspersed repeat) [18•]. MIRs amplified before the eutherian radiation and copies can even be found in marsupials and monotremes [19]. Extension of a repetitive sequence sharing the 3′ 50 bp with MIR (named MIR2 in [18•]) to ~3 kb revealed it to be a LINE-like element (LINE2 in Table 1) most closely related to elements in reptilian and amphibian genomes (A Smit, unpublished data), confirming the generality of the above model for SINE origins.

Figure 1

Schematic representation of the types of transposable elements that have produced high copy number mammalian interspersed repeats. The shaded boxes denote internal promoter sites; names in parentheses indicate that only autonomous elements code for these proteins. ITR, inverted terminal repeat; RT, reverse transcriptase.



As LINE1 and *Alu/B1* do not share a 3′ end and this sequence has been extremely divergent in the evolution of LINE1 [20•], the transpositional mechanism of the best-studied mammalian elements remains ironically enigmatic. Another mystery to solve—if *Alu* uses the LINE1 transpositional machinery—is the near-complementary distribution of *Alu*s and LINE1 in our genome (Fig. 2). Fascinating developments in *Alu* research (reviewed recently in [21]) have been its use in human population studies [22] and studies on its unusual methylation pattern [23•] and the processing of the transcript [24•]. The impact of LINE1 on mammalian evolution and the importance of continued LINE1 research is emphasized further by observations that processed pseudogene formation could be dependent on LINE1 activity too [25•].

Retrovirus-like elements

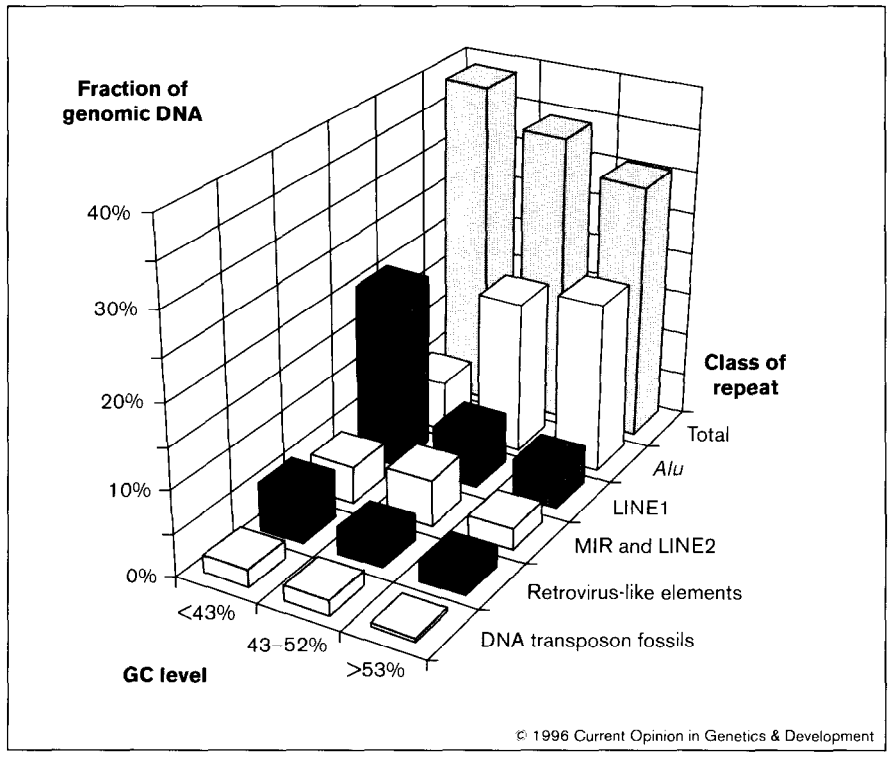
Retrovirus-like elements (RLEs) are retroposons characterized by long terminal repeats (LTRs; see Fig. 1). These LTRs carry the transcriptional regulatory sequences and are reproduced partly from each other in a complex reverse transcription process, thereby solving the retroposon's inherent problem of mobilising the promoter region and maintaining its full length. In the sequences analyzed to create Table 1, on average six out of seven RLE copies are fixed as solitary LTRs, presumably through internal homologous recombination. Solo LTRs can be recognized by their 5′ TG/CA 3′ termini, 4–6 bp flanking repeats, and polyadenylation signal. The high conservation of these and other features may reflect the constraints of the intricate retrotransposition mechanism.

Analogous to LINEs and SINEs is the existence of autonomous and non-autonomous RLEs, the autonomous represented by LTR retrotransposons and retroviruses. The vertebrate-specific infectious retroviruses are monophyletic and probably evolved from a purely vertically transmitted LTR retrotransposon, perhaps resembling a member of the closely related *gypsy/Ty3* retrotransposon family or the more distantly related *Ty1–copia* group [8]. Despite the wide distribution of both types of retrotransposons in eukaryotes—including fish [26–28] and, for *Ty1/copia* like elements, amphibians and reptiles [29•]—searches for related sequences in mammalian genomes have been negative. Considering the large amount of human sequence data available, any such retrotransposon sequences in human are either highly diverged or of a low copy number.

The human genome contains at least 50 000 copies (Table 1) of a variety of endogenous retroviruses (ERVs) [30]. ERVs most likely originated from retroviral germline infections, with subsequent loss of infectivity but retention of transpositional activity. In their turn, contemporary retroviruses may all have evolved from endogenous relatives [31]. ERVs are known to be transpositionally active in some other mammals but not in humans, although at least two primate ERV families were active as recently as 10–15 million years ago [32,33]. ERV

Figure 2

Unequal distribution of repeats over the genome. This chart presents data from Table 1, highlighting the almost complementary dispersal of *Alu*s and LINE1s in our genome. As with LINE1, most repeats are less common in GC-rich regions of the genome, possibly correlated to the high density of highly expressed genes and the relatively high burden of transposition into these regions. I suspect that the reported longer lengths of introns and genes in AT-rich DNA [53] can be explained by a biased dispersal of repeats, although most of these may remain unrecognized. The maverick distribution of *Alu* could be related to a need for high expression levels for retroposition, dependent on external promoter elements [54], although it is unclear how the integration bias could be established, especially if *Alu* depends on LINE1 for reverse transcription and integration.



families often show little similarity to each other and many may still be discovered, as exemplified by the recent description of HERV-L, a family related to the foamy (retro)viruses [34•]. Although the authors estimate that only 100–200 full-length copies exist in our genome, HERV-L-related sequences in fact make up over half of all retroviral sequences identified (Table 1). The HERV-L LTRs — recognized as LTRs and named MLT2 in the human repeat reference library since 1994 — are of special interest as, more often than not, they have spawned a potentially polymorphic CA dinucleotide repeat from their putative polyadenylation site [7].

Of even greater interest is the potential relationship of HERV-L with the mammalian LTR-retroposons (MaLRs), which form the largest family of RLEs in mammals with >150 000 copies estimated to cover at least 2.5% of the human genome (Table 1). MaLRs have been amplified in mammals since well before the eutherian radiation but appear to be extinct in higher primates [35,36]. The open reading frame that comprises most of the 1.3–1.6 kb internal sequences of primate-specific and mammalian-wide MaLRs does not encode a reverse transcriptase (no similarity to any protein is apparent) [7]. Unlike the primate lineage, MaLRs in mouse — represented by origin region repeat (ORR1) and mouse transposon (MT) elements — may still continue to expand, as indicated by the 98.5% similarity of several sequenced members [7]. MT elements appear to be a mosaic of ORR1 and MLT2/HERV-L-like fragments [7]. Mosaics are commonly observed in RLEs and are usually explained by recombinations between co-packaged transcripts in a virus-like particle. This suggests a dependence of MaLR transposition on HERV-L, a theory that is strengthened by similarity of the primer-binding regions at the ends of the internal sequence and of the LTR termini, a similar sequence bias in a 5 bp target site duplication, the mammalian-wide distribution of both, and the observation that HERV-Ls like MaLRs have recently amplified in the mouse but not in the human genome [7,34•].

Most of the >30 000 other RLEs in Table 1 belong to a second group of apparently non-autonomous elements, among which are elements with MER4 (or Spm) LTRs [7,37]. The consensus internal sequences for these elements are up to 5 kb long but do not appear to contain long open reading frames (A Smit, unpublished data). One other RLE without apparent coding regions is known in rodents: *Myr* appears to be transpositionally active in white-footed mouse populations [38] and the study of this element may elucidate the retroposition mechanism of non-autonomous RLEs in general.

DNA transposon remains in the human genome

Until recently, mammalian interspersed repeats all appeared to have been amplified by retroposition but publications this year have shown that a significant fraction (>1.5%; Table 1) of our genome consists of the remains of DNA transposons. These elements, characterized by short inverted terminal repeats, move by excision and reintegration. This process is not inherently replicative but can lead to element duplication, such as when the gap at the site of excision is repaired using the sister chromatid that still contains the element as a template. It is conceivable that interruption of this gap-repair can give rise to the internal deletion products that often outnumber full-length elements [39]. On the basis of similarity between the transposases, most eukaryotic DNA transposons fall into two classes: the *Ac/hobo* class, characterized by an 8 bp insertion site duplication, and the *Tc1/mariner* class, with a TA dimer duplication.

The presence of *mariner*-like elements in the mammalian genome — announced by Hugh Robertson at the Keystone Meeting on Transposition and Site-specific Recombination in 1994 — was published recently by several groups [40•,41,42,43••]. Besides the relatively low copy number *mariner* relics, our genome contains >150 000 short sequences resembling internally deleted members of the *Ac/hobo* and *Tc1/mariner* group [43••]. Three of the autonomous elements responsible for the accumulation of these short repeats have been characterized: two *Tc1*-type elements related to *pogo* [43••] and an element distantly related to *hobo* (A Smit, unpublished data). The level of sequence divergence suggests that activity of all identified elements predates human evolution.

The mammalian *mariner* 'fossils' closely resemble members of three subfamilies identified in insects [40•,41,42,43••], adding to the already extensive evidence that horizontal transfer between genomes has been important in *mariner* evolution [44]. Other human DNA transposon remains also show high similarity to sequences in distantly related organisms [7,43••]. Internal deletion products and most other mutated elements co-amplify with the autonomous DNA transposon, as transposases necessarily work *in trans* in eukaryotes and recognize only the terminal base pairs of a transposon (e.g. the final 26 bp of *Tc1* are sufficient for transposition [45••]). An autonomous sequence will eventually lose the competition for its transposase from the accumulating non-autonomous elements and face extinction; thus, it could be argued that evolutionary survival of a DNA transposon may depend on regular invasion of (horizontal transfer to) a new, 'naive' genome. To allow this, transposition has to be independent of host factors, a feature confirmed for *Tc1* [45••] and *mariner* [46•] this year.

The abundant presence of DNA transposon fossils puts mammalian genomic evolution in a new light as the 'cut and paste' mechanism of DNA transposition may have been responsible for many chromosomal rearrangements. In plants and *Drosophila*, transposase activity is correlated with increased recombination, especially around the transposon sites (e.g. [47,48]). Indeed, a *mariner* fossil

has already been implicated in a recombination hotspot responsible for two inherited neuropathies [49], although there is no indication for concurrent *mariner* activity.

Conclusions

Almost all human interspersed repeats belong to a few well defined classes and the time and mechanism of distribution have been inferred for many. This growing knowledge benefits many fields, not least being the human genome project — 36% of newly sequenced DNA can immediately be 'explained' — but also mammalian phylogenetics and studies on the evolution of specific loci. Improvements in computer analysis methods should allow us to probe even further back in the evolutionary history of our non-coding DNA. Meanwhile, the mechanism of *Alu* transposition still is remarkably elusive and its resolution as well as an explanation for the reciprocal genomic distribution of *Alu* and LINE remain prime targets for research. The discovery that DNA transposons have been active in mammals raises the possibility that these elements, such useful tools in invertebrate and plant genetics, may be used in mammalian genetics too.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Doolittle RF, Sapienza C: **Selfish genes, the phenotype paradigm and genomic evolution.** *Nature* 1980, **284**:601–603.

2. Okada N: **SINEs.** *Curr Opin Genet Dev* 1991, **1**:498–504.

3. Yoshioka Y, Matsumoto S, Kojima S, Ohshima K, Okada N, Machida Y: **Molecular characterization of a short interspersed repetitive element from tobacco that exhibits sequence homology to specific tRNAs.** *Proc Natl Acad Sci USA* 1993, **90**:6562–6566.

4. Kachroo P, Leong SA, Chattoo BB: **Mg-SINE: a short interspersed nuclear element from the rice blast fungus, *Magnaporthe grisea*.** *Proc Natl Acad Sci USA* 1995, **92**:11125–11129.

5. Quentin Y: **Emergence of master sequences in families of retroposons derived from *7sl* RNA.** *Genetica* 1994, **93**:203–215.

6. Willis IM: **RNA polymerase III. Genes, factors and transcriptional specificity.** *Eur J Biochem* 1993, **212**:1–11.

7. Smit AFA: **Structure and evolution of mammalian interspersed repeats [PhD thesis].** Los Angeles: University of Southern California, Los Angeles; 1996.

8. Eickbush TH: **Origin and evolutionary relationships of retroelements.** In *The Evolutionary Biology of Viruses*. Edited by Morse SS. New York: Raven Press; 1994:121–157.

9. Schwarz-Sommer Z, Leclercq L, Gobel E, Saedler H: ***Cin4*, an insert altering the structure of the *A1* gene in *Zea mays*, exhibits properties of non-viral retrotransposons.** *EMBO J* 1987, **6**:3873–3880.

10. Luan DD, Korman MH, Jakubczak JL, Eickbush TH: **Reverse transcription of *R2Bm* RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition.** *Cell* 1993, **72**:595–605.

11. • Luan DD, Eickbush TH: **RNA template requirements for target DNA-primed reverse transcription by the *R2* retrotransposable element.** *Mol Cell Biol* 1995, **15**:3882–3891.
In a continuation of their work [10], the authors find that the last 250 bp of the *R2* transcript is sufficient for reverse transcription and that the reverse transcriptase can add the poly A or simple repeat tail to *R2* transcripts.

12. •• Ohshima K, Hamada M, Terai Y, Okada N: **The 3′ ends of tRNA-derived short interspersed repetitive elements are derived from the 3′ ends of long interspersed repetitive elements.** *Mol Cell Biol* 1996, **16**:3756–3764.
The authors find that some LINEs and tRNA-derived SINEs in the same genome share the same 3′ sequence. As LINE reverse transcriptases are believed to bind the 3′ end of the transcript, this observation suggests an obvious model for the origin and mobilization of SINEs.

13. • Szemraj J, Plucienniczak G, Jaworski J, Plucienniczak A: **Bovine *Alu*-like sequences mediate transposition of a new site-specific retroelement.** *Gene* 1995, **152**:261–264.
By finding that Bov-B – which shares its 3′ end with two SINEs – extends at least to include a LINE-like reverse transcriptase coding region, the authors could have come to the same conclusion as Ohshima *et al.* [12••], but interpret the data differently.

14. Lenstra JA, Van Boxtel JA, Zwaagstra KA, Schwerin M: **Short interspersed nuclear element (SINE) sequences of the Bovidae.** *Anim Genet* 1993, **24**:33–39.

15. Jobse C, Buntjer JB, Haagsma N, Breukelman HJ, Beintema JJ, Lenstra JA: **Evolution and recombination of bovine DNA repeats.** *J Mol Evol* 1995, **41**:277–283.

16. Modi WS, Gallagher DS, Womack JE: **Evolutionary history of highly repeated DNA families among the artiodactyla (mammalia).** *J Mol Evol* 1996, **42**:337–349.

17. • Kordis D, Gubensek F: **Horizontal SINE transfer between vertebrate classes [letter].** *Nat Genet* 1995, **10**:131–132.
This note reports a high similarity of sequences in snakes and the Bov-B LINE in ruminants, providing the first evidence for horizontal relationships of LINEs in vertebrates (at that time, Bov-B was still believed to be a SINE, posing the problem of how such a short, non-coding element could amplify in a newly invaded genome).

18. • Smit AFA, Riggs AD: **MIRs are classic, tRNA-derived SINEs that amplified before the mammalian radiation.** *Nucleic Acids Res* 1995, **23**:98–102.
The authors of this paper describe a transposable element that amplified to several 100 000 copies mostly or entirely before the mammalian radiation some 80 million years ago. The element fits the description of a classic SINE as it has a tRNA-like internal polymerase III promoter and a simple repeat tail. SINEs are thus not an evolutionary novelty, as once proposed to explain the species (cluster) specificity of previously described SINEs. Furthermore, MIRs are remarkable for the uneven conservation of their sequence in the genome.

19. Jurka J, Zietkiewicz E, Labuda D: **Ubiquitous mammalian-wide interspersed repeats (MIRs) are molecular fossils from the mesozoic era.** *Nucleic Acids Res* 1995, **23**:170–175.

20. • Smit AFA, Tóth G, Riggs AD, Jurka J: **Ancestral, mammalian-wide subfamilies of LINE-1 repetitive sequences.** *J Mol Biol* 1995, **246**:401–417.
Many previously detected but uncharacterized interspersed repeats are shown to be part of ancient LINE1 elements and an attempt is made to construct a complete evolutionary history of LINE1 elements in the human genome. The 40 described subfamilies with an associated approximate age of dispersal form a valuable tool in the timing of evolutionary events in the mammalian genome.

21. Schmid CW: ***Alu*: structure, origin, evolution, significance, and function of one-tenth of human DNA.** In *Progress in Nucleic Acid Research and Molecular Biology*. Edited by Cohn WE, Moldave K. New York: Academic Press; 1996, **53**:283–319.

22. Batzer MA, Arcot SS, Phinney JW, Alegria Hartman M, Kass DH, Milligan SM, Kimpton C, Gill P, Hochmeister M, Ioannou PA *et al.*: **Genetic variation of recent *Alu* insertions in human populations.** *J Mol Evol* 1996, **42**:22–29.

23. • Chesnokov IN, Schmid CW: **Specific *Alu* binding protein from human sperm chromatin prevents DNA methylation.** *J Biol Chem* 1995, **270**:18539–18542.
Fully methylated in somatic cells and oocytes – where *Alu*s contain a third of the genome's 5-methylcytosines – *Alu*s are largely unmethylated in sperm DNA. Here, evidence is provided that this is caused by sequence-specific protection by a sperm-specific protein.

24. • Chang DY, Sasaki Tozawa N, Green LK, Maraia RJ: **A trinucleotide repeat-associated increase in the level of *Alu* RNA-binding protein occurred during the same period as the major *Alu* amplification that accompanied anthropoid evolution.** *Mol Cell Biol* 1995, **15**:2109–2116.

Alu transcripts are processed (inactivated) by a protein of the signal recognition complex. The authors find that, during the heydays of *Alu* activity, the gene for this *Alu*-binding protein was duplicated and has since been overexpressed, either suggesting an *Alu*-suppression mechanism or a solution to overcome the titration of this functional protein by the abundant *Alus*.

25. • Maestre J, Tchénio T, Dhellin O, Heidmann T: **mRNA retroposition in human cells: processed pseudogene formation.** *EMBO J* 1995, **14**:6333–6338.
Using a retroposition reporter gene, the authors show that processed pseudogenes are formed in HeLa cells in the absence of retroviral infection. As with naturally occurring ones, these pseudogenes have LINE1-like characteristics, such as 5′ truncation, variable length target site duplication and poly A tails.

26. Flavell AJ, Smith DB: **A *Ty1–copia* group retrotransposon sequence in a vertebrate.** *Mol Gen Genet* 1992, **233**:322–326.

27. Britten RJ, McCormack TJ, Mears TL, Davidson EH: ***Gypsy/Ty3*-class retrotransposons integrated in the DNA of herring, tunicate, and echinoderms.** *J Mol Evol* 1995, **40**:13–24.

28. Tristem M, Kabat P, Herniou E, Karpas A, Hill F: ***Easel*, a *gypsy* LTR-retrotransposon in the Salmonidae.** *Mol Gen Genet* 1995, **249**:229–236.

29. • Flavell AJ, Jackson V, Iqbal MP, Riach I, Waddell S: ***Ty1–copia* group retrotransposon sequences in amphibia and reptilia.** *Mol Gen Genet* 1995, **246**:65–71.
Ty1–copia LTR retrotransposon sequences are shown to exist in the genomes of several amphibians and reptiles. As retroviruses have been found in some of these species as well as in some fishes including salmons, this report and [28] refute an old notion that the host ranges of retroviruses and retrotransposons do not overlap. The authors suggest that that retroviruses may have arisen from retrotransposons in one of these host species.

30. Wilkinson DA, Mager DL, Leong JC: **Endogenous Human Retroviruses.** In *The Retroviridae*. Edited by Levy JA. New York: Plenum Press; 1994.

31. Doolittle RF, Feng DF, Johnson MS, McClure MA: **Origins and evolutionary relationships of retroviruses.** *Q Rev Biol* 1989, **64**:1–30.

32. Goodchild NL, Wilkinson DA, Mager DL: **Recent evolutionary expansion of a subfamily of RTVL-H human endogenous retrovirus-like elements.** *Virology* 1993, **196**:778–788.

33. Di Cristofano A, Strazzullo M, Parisi T, La Mantia G: **Mobilization of an ERV9 human endogenous retroviral element during primate evolution.** *Virology* 1995, **213**:271–275.

34. • Cordonnier A, Casella JF, Heidmann T: **Isolation of novel human endogenous retrovirus-like elements with foamy virus-related *pol* sequence.** *J Virol* 1995, **69**:5890–5897.
The authors describe a novel but heterogeneous (i.e. probably inactive) endogenous retrovirus in human DNA, HERV-L. It is closely related to and could be the ancestor of the infectious foamy viruses, a separate branch of retroviruses. HERV-L-related sequences were detected in all mammals. Mouse but not rat DNA contains homogeneous sequences closely related to the primate-specific form, suggesting recent horizontal transfer into the mouse lineage.

35. Paulson KE, Deka N, Schmid CW, Misra R, Schindler CW, Rush MG, Kadyk L, Leinwand L: **A transposon-like element in human DNA.** *Nature* 1985, **316**:359–361.

36. Smit AFA: **Identification of a new, abundant superfamily of mammalian LTR-transposons.** *Nucleic Acids Res* 1993, **21**:1863–1872.

37. Erickson LM, Maeda N: **A new family of retroviral long terminal repeat elements in the human genome identified by their homologies to an element 5′ to the spider monkey haptoglobin gene.** *Genomics* 1995, **27**:531–534.

38. Sawby R, Wichman HA: **Analysis of orthologous retrovirus-like elements in the white-footed mouse, *Peromyscus leucopus*.** *J Mol Evol* 1996, in press.

39. Engels WR, Johnson Schlitz DM, Eggleston WB, Sved J: **High-frequency *P* element loss in *Drosophila* is homolog-dependent.** *Cell* 1990, **62**:515–525.

40. • Auge-Gouillou C, Bigot Y, Pollet N, Hamelin MH, Meunier-Rotival M, Periquet G: **Human and other mammalian genomes contain transposons of the *mariner* family.** *FEBS Lett* 1995, **368**:541–546.
This paper was the first of five independent publications [41,42,43••,49] on the presence of the DNA transposon remains in mammalian genomes, whereas heretofore only retroposed repeats repeats had been identified. By PCR amplifications with primers from conserved sites in the *mariner* DNA transposases, the authors find *mariner* sequences in human, ruminant and other species.

41. Morgan GT: **Identification in the human genome of mobile elements spread by DNA-mediated transposition.** *J Mol Biol* 1995, **254**:1–5.

42. Oosumi T, Belknap WR, Garlick B: ***Mariner* transposons in humans [letter].** *Nature* 1995, **378**:672.

43. •• Smit AFA, Riggs AD: **Tiggers and other DNA transposon fossils in the human genome.** *Proc Natl Acad Sci USA* 1996, **93**:1443–1448.
Evidence is presented that both *Tc1/mariner*-type and *Ac/hobo*-type DNA transposons have been active during mammalian evolution, leaving >150 000 degenerate copies in their trail; also, a new subgroup of *Tc1/mariner* type DNA transposons is defined that includes mammalian elements, *Drosophila pogo* and nematode *Tc2*, *Tc4* and *Tc5*.

44. Capy P, Langin T, Bigot Y, Brunet F, Daboussi MJ, Periquet G, David JR, Hartl DL: **Horizontal transmission versus ancient origin: *mariner* in the witness box.** *Genetica* 1994, **93**:161–170.

45. •• Vos JC, De Baere I, Plasterk RH: **Transposase is the only nematode protein required for *in vitro* transposition of *Tc1*.** *Genes Dev* 1996, **10**:755–761.
The authors achieve *Tc1* transposition *in vitro* using only recombinant transposase, explaining how *Tc1* horizontal transfer can occur. They also show that only terminal sequences are required for transposition *in trans* and that *Tc1* has sequence-dependent target site specificity beyond the TA insertion site.

46. • Lampe DJ, Churchill MEA, Robertson HM: **A purified *mariner* transposase is sufficient to mediate transposition *in vitro*.** *EMBO J* 1996, **15**:5470–5479.
Similar to the experiments described in [45••] and leading to the same conclusions, these authors find that the purified transposase of a *mariner* element is sufficient for an *in vitro* interplasmid transposition of a non-autonomous marked *mariner* element.

47. McCarron M, Duttaroy A, Doughty G, Chovnick A: ***Drosophila P* element transposase induces male recombination additively and without a requirement for *P* element excision or insertion.** *Genetics* 1994, **136**:1013–1023.

48. Lim JK, Simmons MJ: **Gross chromosome rearrangements mediated by transposable elements in *Drosophila melanogaster*.** *Bioessays* 1994, **16**:269–275.

49. Reiter LT, Murakami T, Koeuth T, Pentao L, Muzny DM, Gibbs RA, Lupski JR: **A recombination hotspot responsible for two inherited peripheral neuropathies is located near a *mariner* transposon-like element.** *Nat Genet* 1996, **12**:288–297.

50. Jurka J, Smit AFA: Reference Collection of Human Repetitive Elements available by FTP from: ftp://ncbi.nlm.nih.gov/repository/repbase/REF

51. University of Washington Genome Server; RepeatMasker on World Wide Web URL: http://ftp.genome.washington.edu

52. Bernardi G: **The human genome: organization and evolutionary history.** *Annu Rev Genetics* 1995, **29**:445–476.

53. Duret L, Mouchiroud D, Gautier C: **Statistical analysis of vertebrate sequences reveals that long genes are scarce in GC-rich isochores.** *J Mol Evol* 1995, **40**:308–317.

54. Chesnokov I, Schmid CW: **Flanking sequences of an *Alu* source stimulate transcription *in vitro* by interacting with sequence-specific transcription factors.** *J Mol Evol* 1996, **42**:30–36.