J Mol Evol (1997) 45:412–423

JOURNAL OF MOLECULAR EVOLUTION

Rapid Evolution of a Young L1 (LINE-1) Clade in Recently Speciated *Rattus* Taxa

Eric L. Cabot,* Barbara Angeletti,† Karen Usdin, Anthony V. Furano

Section on Genomic Structure and Function, Laboratory of Biochemical Pharmacology, NIDDK, NIH, Bethesda, MD 20892-0830

Received: 6 November 1996 / Accepted: 12 April 1997

Abstract. L1 elements are retrotransposons that have been replicating and evolving in mammalian genomes since before the mammalian radiation. *Rattus norvegicus* shares the young $L1_{mlvi2}$ clade only with its sister taxon, *Rattus* cf *moluccarius*. Here we compared the $L1_{mlvi2}$ clade in these recently diverged species and found that it evolved rapidly into closely related but distinct clades: the $L1_{mlvi2\text{-}rm}$ clade (or subfamily), characterized here from *R.* cf *moluccarius,* and the $L1_{mlvi2\text{-}rn}$ clade, originally described in *R. norvegicus*. In addition to other differences, these clades are distinguished by a cluster of amino acid replacement substitutions in ORF I. Both rat species contain the $L1_{mlvi2\text{-}rm}$ clade, but the $L1_{mlvi2\text{-}rn}$ clade is restricted to *R. norvegicus*. Therefore, the $L1_{mlvi2\text{-}rm}$ clade arose prior to the divergence of *R. norvegicus* and *R.* cf *moluccarius,* and the $L1_{mlvi2\text{-}rn}$ clade amplified after their divergence. The total number of $L1_{mlvi2\text{-}rm}$ elements in *R.* cf *moluccarius* is about the same as the sum of the $L1_{mlvi2\text{-}rm}$ and $L1_{mlvi2\text{-}rn}$ elements in *R. norvegicus*. The possibility that L1 amplification is in some way limited so that the two clades compete for replicative supremacy as well as the implications of the other distinguishing characteristic of the $L1_{mlvi2\text{-}rn}$ and $L1_{mlvi2\text{-}rm}$ clades are discussed.

Key words: Molecular sequence data — Retrotransposon — L1 — LINE-1 — Repeated DNA — Rodent — Evolution

Introduction

L1 (LINE-1) transposable elements are ubiquitous in mammals (Fanning and Singer 1987b; Hutchison III et al. 1989). L1 elements consist of a 5′ untranslated region (UTR), two open reading frames (ORF I and ORF II) and a 3′ UTR (top diagram, Fig. 1). The 5′ UTR contains a regulatory sequence (Nur et al. 1988; Swergold 1990; Minakami et al. 1992; Severynse et al. 1992), and ORF II encodes a reverse transcriptase (Mathias et al. 1991). ORF I encodes an ~40 kD protein also thought to be involved in retrotransposition since it can be found together with L1 transcripts in ribonucleoprotein particles (Martin 1991; Kolosha and Martin 1995; Hohjoh and Singer 1996).

L1 replication occurs by retrotransposition (Voliva et al. 1984; Hattori et al. 1986; Fanning and Singer 1987a; Casavant et al. 1988; Kazazian et al. 1988; Singer 1995), whereby transcripts of existing elements are reverse transcribed into DNA which is then integrated into the genome. Retrotransposition generates mostly defective copies which remain in the genome and accumulate base substitutions at the pseudogene rate (Fanning 1983; Voliva et al. 1983; Hardies et al. 1986; Pascale et al. 1993). Variant, replicatively active elements are also produced and generate novel L1 subfamilies (or clades) some of which can attain many thousands of copies (Martin et al. 1985; Hardies et al. 1986; Pascale et al. 1993; Casavant and Hardies 1994b). It is not known whether the repli-

* *Present address:* Genetics Computer Group., Inc. University Research Park, 575 Science Drive, Suite B, Madison, WI 53711, USA, e-mail cabot@gcg.com.

† *Present address:* Istituto di Medicina Sperimentale, CNR, Viale Marx 43, 00137 Roma, Italy

Correspondence to: A.V. Furano

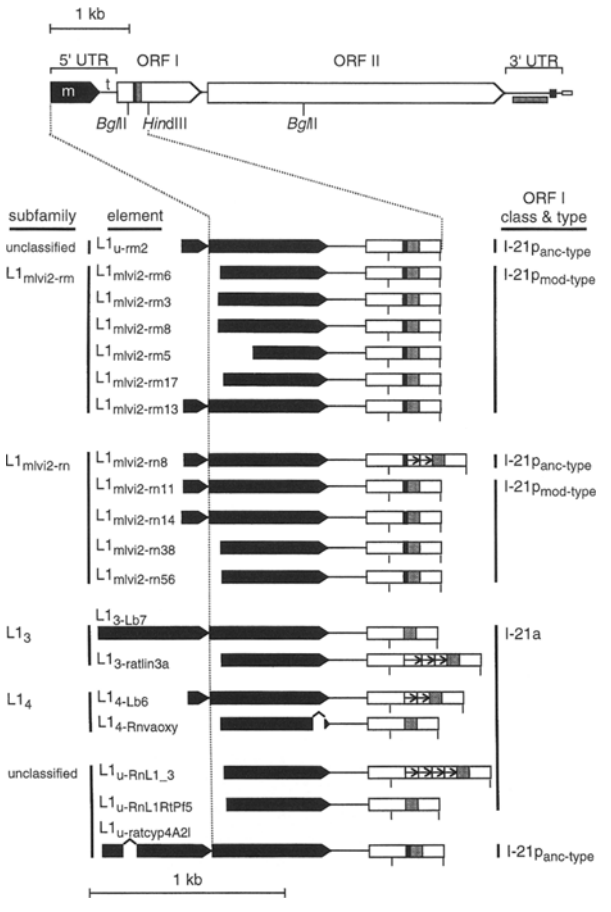


Fig. 1. L1 elements. The top diagram represents a generic rodent L1 element. The 5′ UTR has two regions: a sequence with regulatory properties labeled "m" for monomer (Padgett et al. 1988) since it can be present in tandem multiple copies in both rat and mouse L1 elements. The second region is labeled "t" for tether. The narrow black box within ORF I represents the 21-bp sequence whose presence or absence defines the ORF I I-21p and I-21a classes, and the contiguous shaded box designates the 66-bp sequence that is tandem repeated in some L1 elements. The small black and open boxes in the 3′ UTR represent the G-rich polypurine stretch and A-rich 3′ terminus, respectively. The filled box below the 3′ UTR corresponds to the PCR-generated fragment used as the generic modern L1 probe. The lower diagrams, at 2.5 times the scale of the top one, represent the 5′ UTR and the 5′ one-third of ORF I of various L1 elements. The top seven were isolated and sequenced in this study, and next five were from another study (Hayward et al. 1997). The lower seven were from Genbank. The names given to these elements are the same as their Genbank locus names. These sequences are grouped according to their 3′ UTR subfamily type [$L1_{mlvi2}$, $L1_4$, $L1_3$, and $L1_u$, for unclassified, (Hayward et al. 1997)] and their ORF I class, I-21a, I-21p_{ancest} or I-21p_{mod}. The elements, $L1_{u\text{-}rm2}$, $L1_{u\text{-}RnRL13}$, $L1_{u\text{-}RnL1RtP5F}$, and $L1_{u\text{-}Ratcyp4a21}$, are unclassified because their 3′-UTR sequences are not known. The boxes with an *arrow* represent the tandem repeated 66-bp sequence or versions thereof. The tick marks beneath ORF I represent conserved *Bgl*II and *Hind*III sites, respectively.

cative success of a variant L1 element is merely a stochastic event or represents an adaptive response of the element to the host.

Given the amount of L1 DNA in mammalian genomes, the lack of information about the interaction between L1 elements and their hosts represents a significant gap in our knowledge of mammalian biology. A potentially useful approach to this problem is provided by comparing L1 subfamilies both between and within species. Such studies have identified features of L1 elements that are at least compatible with, if not responsible for, replicative success. For example, repeated acquisition of a novel 5′ UTR has been an important feature of L1 evolution (Scott et al. 1987; Wincker et al. 1987; Furano et al. 1988; Padgett et al. 1988; Jubier-Maurin et al. 1992; Schichman et al. 1993; Adey et al. 1994a), and this has been proposed as a means of restoring transcriptional activity to otherwise moribund elements (Furano et al. 1988; Padgett et al. 1988; Schichman et al. 1992; Adey et al. 1994a, b).

The *Rattus* $L1_{mlvi2}$ clade emerged less than 0.5 Mya and is only in *R. norvegicus* and its sister taxon, *Rattus* cf *moluccarius* (Furano and Usdin 1995; Usdin et al. 1995). The $L1_{mlvi2}$ clade of *R. norvegicus* is quite unusual in that it has a dual ancestry (Hayward et al. 1997). Although its 3′ UTR is descendant from its chronologically antecedent $L1_4$ subfamily, the $L1_{mlvi2}$ ORF I sequence was not derived from an $L1_4$ ORF I, but was recruited from a more ancestral ORF I. This ancestral ORF I contains a 21-bp sequence that was deleted from the ORF I of $L1_4$ elements. The $L1_4$ clade appeared and amplified during the radiation of *Rattus sensu stricto* about 2 Mya, well before the speciation of *R. norvegicus* and *R.* cf *moluccarius*. Figure 2 shows a schematic of this scenario.

Consequently, *R. norvegicus* L1 elements contain two classes of ORF I: the I-21p class (ORF I 21 bp present) in $L1_{mlvi2}$ (and ancestral) elements, and the I-21a class (21 bp absent) in $L1_4$ elements (and other modern subfamilies such as the $L1_3$ clade, see Fig. 2). To further complicate matters, two types of I-21p ORF I sequences are present in the *R. norvegicus* $L1_{mlvi2}$ subfamily: one resembles an ancestral type of I-21p ORF I sequence (I-21$p_{anc\text{-}type}$), and the second is a modern derivative of it, I-21$p_{mod\text{-}type}$. This modern type sequence differs from the ancestral type mainly in and around the 21 bp sequence (Fig. 2 and Hayward et al. 1997). Thus, this particular region of the $L1_{mlvi2}$ ORF I underwent evolutionary change quite soon after the ancestral ORF I was recruited.

Here, we further analyzed the $L1_{mlvi2}$ clade by examining $L1_{mlvi2}$ elements from *R.* cf *moluccarius*. These $L1_{mlvi2}$ elements are distinct from those isolated from *R. norvegicus* ($L1_{mlvi2\text{-}rn}$) and constitute a closely related but separate clade, $L1_{mlvi2\text{-}rm}$. Although *R.* cf *moluccarius* contains only $L1_{mlvi2\text{-}rm}$, *R. norvegicus* contains both this clade and $L1_{mlvi2\text{-}rn}$. Therefore, $L1_{mlvi2\text{-}rm}$ arose prior to the divergence of the two hosts, while $L1_{mlvi2\text{-}rn}$ amplified after the speciation of *R. norvegicus*. Interestingly, $L1_{mlvi2\text{-}rn}$ and $L1_{mlvi2\text{-}rm}$ amplified to about the same extent in *R. norvegicus* as did the single $L1_{mlvi2\text{-}rm}$ clade in *R.* cf *moluccarius*. A major difference between these clades involves ORF I, which harbors six diagnos-

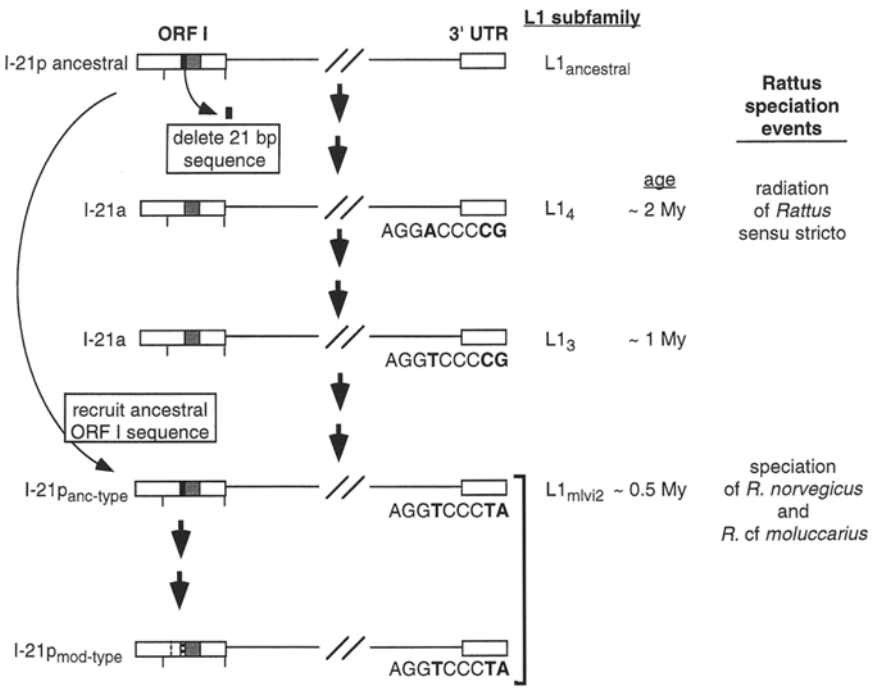


Fig. 2. Evolutionary changes in the 3′-UTR and ORF I of some rat L1 subfamilies (clades). In this highly simplified version only part of the indicated regions of the L1 elements are shown and not to scale. The $L1_4$, $L1_3$, and $L1_{mlvi2}$ 3′ UTR are distinguished by a number of diagnostic base differences (Usdin et al. 1995; Hayward et al. 1997) only a few of which are shown here, in bold type. The age of the indicated subfamilies were calculated from the extent divergence of members of each of the indicated clades and are reasonably congruent with the estimated times of the indicated *Rattus* speciation events (Usdin et al. 1995; Hayward et al. 1997). The differences between the ORF I DNA sequences are shown in Figure 3. The small black rectangle and larger gray one represent the 21- and 66-bp sequences, respectively, and the tick marks below the ORF I sequence indicate the location of the conserved *Bgl*II and *Hind*III sites (see Fig. 1 and its legend).

tic substitutions. All produce amino acid replacement (nonsynonymous base substitutions), and five are within or near the 21-bp sequence. This region of ORF I is also uniquely susceptible to tandem repetition of a 66-bp sequence and thus can be considered hypervariable. The possible biological relevance of the hypervariable region of ORF I and the implications of other aspects of the evolutionary dynamics of the $L1_{mlvi2}$ clade are discussed.

Materials and Methods

General Methods

We used standard molecular techniques as described in (Ausubel et al. 1989) or earlier (see Usdin et al. 1995), except as noted below.

Isolation of $L1_{mlvi2}$ Clones

Clones containing the 5′ ends of $L1_{mlvi2}$ elements were isolated from *R. cf moluccarius* as follows. Rat genomic DNA was digested with *Hind*III and the ends of the resulting fragments were half-filled using Klenow polymerase to leave a 5′ AG overhang. The fragments were ligated into pUC19 that had been digested with *Xba*I and half-filled leaving a compatible 5′ CT overhang. Colonies were screened by hybridization with bh19, a 35-nt oligonucleotide that is homologous to the $I\text{-}21p_{mod\text{-}type}$ subfamily of $L1_{mlvi2}$ elements from *R. norvegicus* (see below). Clones obtained by this procedure contain L1 fragments bounded by the conserved *Hind*III site in ORF I (see Fig. 1), and a *Hind*III site located in 5′ flanking non-L1 rat genomic DNA. The L1-specific parts of the inserts of seven such clones ranged from 1306 to 948 bp in length due to truncation or partial tandem duplication of the 5′ UTR monomer region (Fig. 1).

DNA Sequencing and Sequence Analysis

The DNA sequence was determined on both strands of the $L1_{mlvi2}$ clones using Sequenase Version 2.0 (U.S. Biochemical Corp.). All sequences were manipulated and aligned using either the ESEE DNA sequence editor (Cabot and Beckenbach 1989) or the programs provided in the Wisconsin Package (Program Manual for the Wisconsin Package, Version 8, September 1994, Genetics Computer Group, 575 Science Drive, Madison, Wisconsin 53711, USA). A diagram of the aligned sequences is shown in Figure 1 and is available by request from A.V.F. Phylogenetic analyses were performed using the PAUP (Swofford 1993), PHYLIP 3.5 (Felsenstein 1993), and MEGA (Kumar et al. 1993) computer programs. All of the sequences determined here have been submitted to GENBANK, and their accession numbers are: $L1_{u\text{-}rm2}$, AF000199; and $L1_{mlvi2\text{-}rm3,\ -rm5,\ -rm6,\ -rm8,\ -rml3}$, and $_{-rm17}$. AF000200–AF000205, respectively.

Determination of L1 Subfamily Copy Number

Genomic DNA was digested with various restriction enzymes and electrophoresed in duplicate on 1.25% agarose gels in 0.5 × TBE (1× is 89 mM Tris base, 89 mM boric Acid, 2 mM Na_2 EDTA) along with serially diluted DNA of known concentration from cloned representatives of the appropriate L1 subfamily. The DNA was transferred to Bio-Rad Zeta-Probe GT nylon membranes using the manufacturer's protocol for alkaline capillary transfer and then hybridized to either oligonucleotide probes specific for the particular L1 subfamily and region (i.e., ORF I or 3′ UTR) or to generic ORF I or 3′ UTR probes to measure total L1 copy number. The genomic DNA was digested with *Hind*III and *Bgl*II or *Hae*III, and *Nla*III, for the ORF I or 3′ UTR probes, respectively. The amounts of radioactivity hybridized to the genomic digests were determined and compared to the amounts hybridized to the cloned subfamily standard using a Molecular Dynamics phosphoimager and ImageQuant software. Genomic DNA concentrations were determined after restriction digestion by comparison to a dilution series of digested calf thymus DNA of known concentration. Photographs of ethidium bromide stained agarose gels containing these DNA samples were scanned and converted to TIFF files by a PDI analyzer. The TIFF files were then processed using the NIH Image analysis software (Wayne Rasband, NIH).

The hybridization probes, competitors, and the L1 DNA used for positive and negative controls are listed in Table 1. The latter DNAs were generated from either ORF I or the 3′ UTR by the PCR from the indicated cloned L1 elements and then gel purified. The ORF I frag-

ment began just 5′ of the initiation codon and extended to the first three bases of the conserved *Hind*III site in ORF I (see Fig. 1 and the highlighted regions of the appropriate oligonucleotides below). The region of the 3′ UTR that was amplified is indicated in Figure 1. The concentration of the PCR fragments was determined using appropriately sized DNA standards of known concentration as described above for genomic DNA. PCR fragments used as hybridization probes were radiolabeled using [^{32}P]-dCTP by the random hexamer primer method (Feinberg and Vogelstein 1983) using a Strategene Prime-It kit. Oligonucleotides were phosphorylated with [^{32}P] γ-ATP using T4 polynucleotide kinase. The oligonucleotide hybridization reactions contained a 100-fold molar excess of nonradioactive, noncognate oligonucleotides to prevent cross-hybridization of the oligonucleotide probe to noncognate subfamilies.

Oligonucleotide Sequences

Hybridization Probes and Competitors:

ORF I:

mlvimol	GGGAAACACAGGAAATCATT
mlvirn	GGGAAGCACAGGAAAACATT
bh19 (I-21p$_{mod}$)	GAACACACTTAGGGAAGCACAGGAAAACATTAATA
bh18 (I-21p$_{ancest}$)	GAACTCCCTTAGAGAAACACAGGAAAACATTAA
I-21a	CTCCCTTAGAG<----21 bp absent--->AACAAGTAGAAG

3′ UTR:

L1a2 (L1$_{mlvi2}$)	AGGTCCCCTATTGAAGGA
L3.alt (L1$_3$)	GAATAGGTCCCCCGTTGAA
L4 (L1$_4$)	GAGAATAGGACCCCCGTTG
Lafp (L1$_{afp}$)	AGGTCCCCTGTTGAAGGAA

PCR primers:

ORF I:

5′ ORF I-7	CACTAGAGATAATCTGATGGCG
3′ bh15	CTTGTATCTACAGCTCCTTG

3′ UTR:

5′ E21	AACAGAGACTGAAGGAA
3′ E23	AGTCTAGTTCACTGGGG

Table 1. Oligonucleotide probes and competitors, and cloned L1 DNA used for positive and negative controls

	Oligonucleotide		Cloned DNA from:	
Region probed/subfamily	Probe	Competitors	Positive Control	Negative Control
ORF I				
L1$_{mlvi2\text{-}rn}$	mlvirn	bh18 mlvimol	L1$_{mlvi2\text{-}rn14}$	L1$_{mlvi2\text{-}rn5}$
L1$_{mlvi2\text{-}rm}$	mlvimol	bh18 mlvirn	L1$_{mlvi2\text{-}rm5}$	L1$_{mlvi2\text{-}rn14}$
I-21a[a]	I-21a		L1$_{3\text{-}ratlin3a}$	L1$_{mlvi2\text{-}rm5}$ L1$_{mlvi2\text{-}rn14}$ L1$_{mlvi2\text{-}rn8}$[c]
L1Rn (modern)	L1$_{mlvi2\text{-}rn14}$[b]	none	L1$_{mlvi2\text{-}rn14}$	None
3′ UTR				
L1$_{mlvi2}$	L1$_{a2}$	L3.alt Lafp L4	L1$_{mlvi2\text{-}rn14}$	L1$_{3\text{-}ratlin3a}$ L1$_{4\text{-}ratlin4a}$
L1$_3$	L3.$_{alt}$	L1a2 Lafp L4	L1$_{3\text{-}ratlin3a}$	L1$_{mlvi2\text{-}rn14}$ L1$_{4\text{-}ratlin4a}$
L1$_4$	L$_4$	L3.alt Lafp L1a2	L1$_{4\text{-}ratlin4a}$[d]	L1$_{3\text{-}ratlin3a}$ L1$_{mlvi2\text{-}rn14}$
L1Rn (modern)	L1$_{3\text{-}ratlin3a}$[b]	none	L1$_{3\text{-}ratlin3a}$[c]	None

[a] Includes all subfamilies lacking the 21-bp insert in ORF I such as members of the L1$_4$ and L1$_3$ subfamilies

[b] Probe was a PCR fragment generated from a clone of this element

[c] This contains an I-21a element (see Fig. 1)

[d] See (Usdin et al. 1995)

Results

We characterized seven L1 elements from *R.* cf *moluccarius* that had been selected for the presence of the 21 bp-containing sequence typical of the *R. norvegicus* $L1_{mlvi2}$ ORF I sequence. The DNA sequence of the 5′ UTR and ORF I [to the conserved *Hind*III site (Fig. 1)] of six ($L1_{mlvi2\text{-}rm3\ \ldots\ rm17}$) were almost identical, and their ORF I sequence was of the $I\text{-}21p_{mod\text{-}type}$ (see next section). However, these *R.* cf *moluccarius* elements differed from the $I\text{-}21p_{mod\text{-}type}$ $L1_{mlvi2\text{-}rn}$ elements of *R. norvegicus* at 13 diagnostic sites: seven in the 5′ UTR and six in ORF I. Therefore, *R.* cf *moluccarius* contains a distinct version of the $L1_{mlvi2}$ family, which we call $L1_{mlvi2\text{-}rm}$.

The ORF I Region of R. *cf* moluccarius $L1_{mlvi2\text{-}rm}$ Elements

The $L1_{mlvi2\text{-}rm3\ \ldots\ rm17}$ ORF I and $L1_{mlvi2\text{-}rn11\ \ldots\ rn56}$ ORF I sequences are identical at six positions that distinguish them from all or most of the $I\text{-}21p_{anc\text{-}type}$ or I-21a elements. Figure 3 shows five of these positions including the three bp deletion that is a hallmark of the $I\text{-}21p_{mod\text{-}type}$ ORF I sequence. Since the $I\text{-}21p_{mod\text{-}type}$ ORF I sequence is characteristic only of the $L1_{mlvi2}$ subfamily in *R. norvegicus* (Fig. 3 and Hayward et al. 1997), we conclude that the $L1_{mlvi2\text{-}rm3\ \ldots\ rm17}$ elements belong to the $L1_{mlvi2}$ subfamily, even though our cloning strategy precluded isolation of the 3′ UTR which contains the original diagnostic sequence that defined the $L1_{mlvi2}$ clade. Phylogenetic analysis of both the ORF I and 5′ UTR confirms this conclusion (see below). Since the $I\text{-}21p_{anc\text{-}type}$ $L1_{mlvi2\text{-}rn8}$ element is a member of the $L1_{mlvi2}$ subfamily, we also consider this an $L1_{mlvi2\text{-}rn}$ element. The classification of L1 elements is summarized in Figure 1.

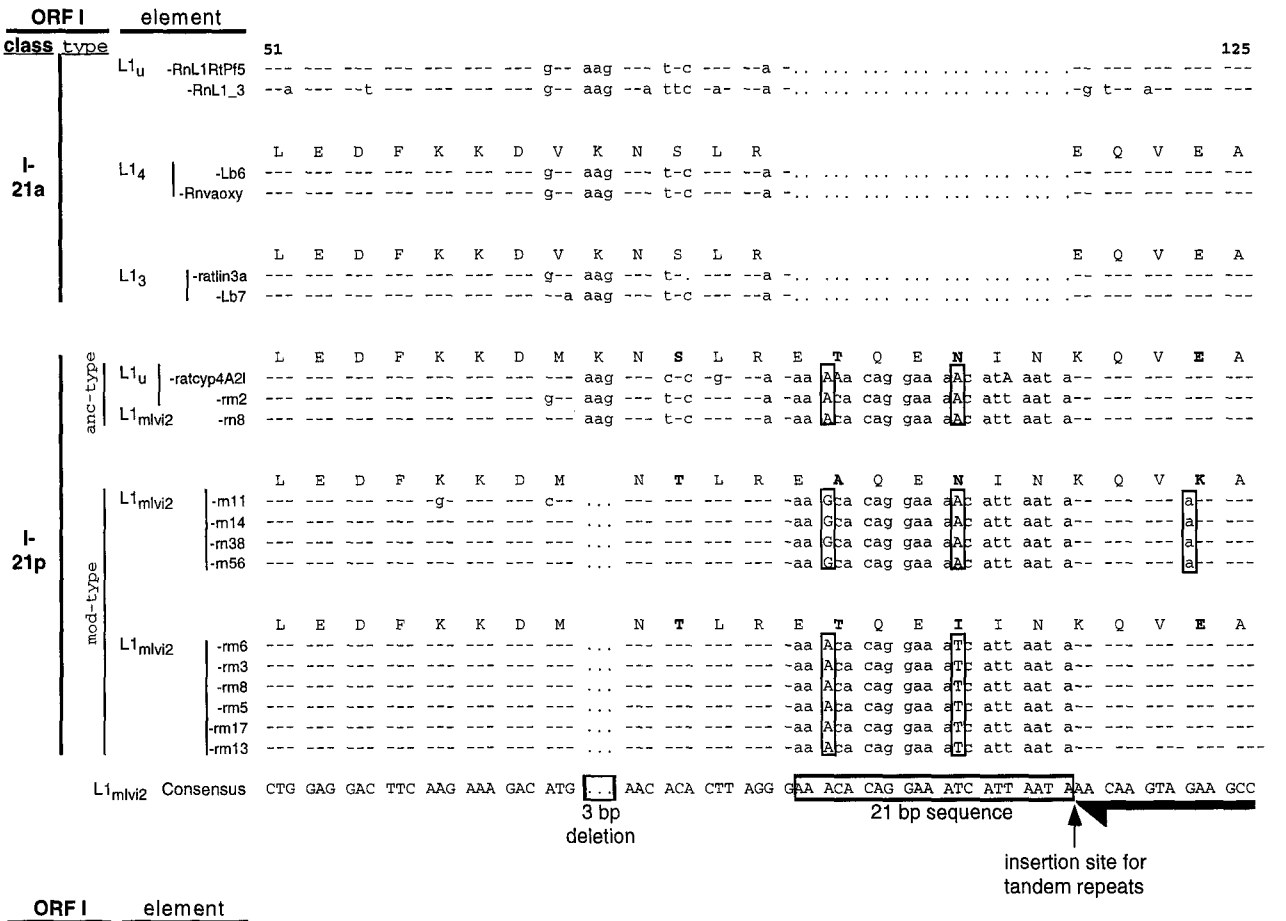
The $L1_{u\text{-}rm2}$, $L1_{u\text{-}ratcyp4a21}$, and $L1_{mlvi2\text{-}rn8}$ elements each contain an $I\text{-}21p_{anc\text{-}type}$ ORF I sequence (Fig. 3). This ORF I class is phylogenetically quite old since it is present throughout *Rattus sensu stricto* (Hayward et al. 1997). These $I\text{-}21p_{anc\text{-}type}$ sequences lack the 3-bp deletion that typifies $I\text{-}21p_{mod\text{-}type}$ elements and share an almost identical (2 mismatches) 21-bp sequence which is intermediate in sequence between that of the $L1_{mlvi2\text{-}rn}$ and $L1_{mlvi2\text{-}rm}$ $I\text{-}21p_{mod\text{-}type}$ elements (Fig. 3). Comparison of the 5′ UTR sequences and the phylogenetic analysis (see below) shows that, in contrast to $L1_{mlvi2\text{-}rn8}$, $L1_{u\text{-}rm2}$ and $L1_{u\text{-}ratcyp4a21}$ are not members of the $L1_{mlvi2}$ clade.

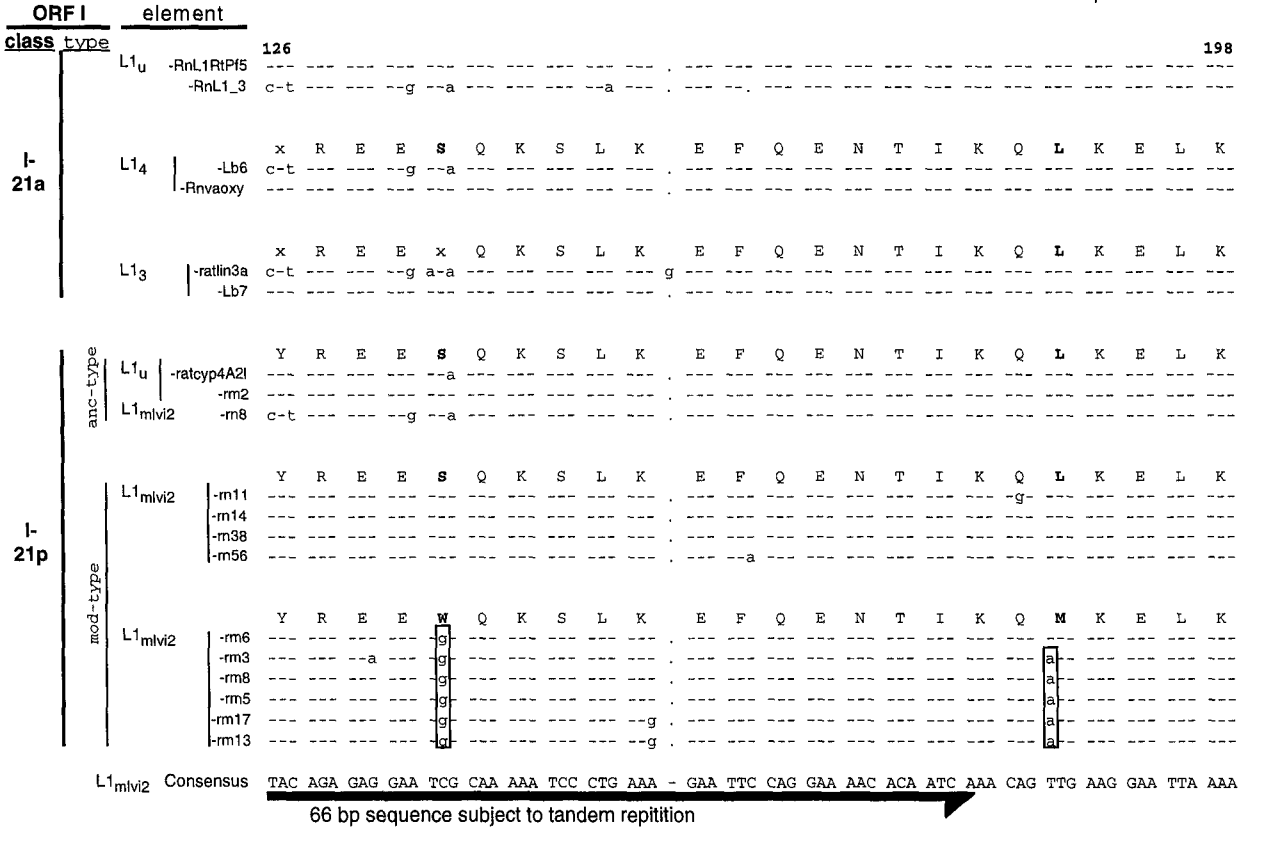
The 21-bp sequence is present in a hypervariable region of ORF I. One characteristic of this region is the tendency for the 66 bp just 3′ of the 21 bp sequence to undergo tandem repetition. While tandem repetition is characteristic of the older $I\text{-}21p_{anc\text{-}type}$ and I-21a families, none of the cloned or genomic $I\text{-}21p_{mod\text{-}type}$ $L1_{mlvi2\text{-}rn}$ members contained tandem repeated 66-bp sequences (Hayward et al. 1997). This is also the case for both the cloned and genomic members of the $L1_{mlvi2\text{-}rm}$ elements (Figure 1 and see section on Determination of $L1_{mlvi2}$ subfamily copy number, pp xxx–xxx). A second characteristic of the hypervariable region is its susceptibility to nonsynonymous base substitutions. Our results here confirm this. Each of the six diagnostic differences between the ORF I sequence of the $L1_{mlvi2\text{-}rm}$ and $L1_{mlvi2\text{-}rn}$ subfamilies produces an amino acid replacement. As Figure 3 shows, four of these are either in the 21- or 66-bp sequence and a fifth one is just 3′ of the 66-bp sequence.

The 5′ UTR of R. *cf* moluccarius $L1_{mlvi2\text{-}rm}$ Elements

Because of the repeated replacement of the 5′ UTR during L1 evolution (see Introduction), we used a cloning strategy for isolating *R.* cf *moluccarius* $L1_{mlvi2}$ elements that would be indifferent to the sequence of the 5′ UTR (see Materials and Methods). Sequence alignments (not shown) and phylogenetic analysis (see below) showed that the 5′-UTR of the *R. norvegicus* and *R.* cf *moluccarius* $L1_{mlvi2}$ elements, and of the phylogenetically older I-21a and $I\text{-}21p_{anc\text{-}type}$ elements are homologous. We also found that although the monomer region (Fig. 1) is highly variable both within and between subfamilies, the contiguous "tether" is not. For example, the base substitution rate for the monomer region was 2.9 ± 1.0 greater than that for the tether region. This was the average for four different rat L1 subfamilies ranging in age from the ~2 My old $L1_4$ subfamily to the ~0.5 My old $L1_{mlvi2}$ clades. We found a similar ratio for two different

Fig. 3. Alignment of the ORF I hypervariable region. The two residues in the 21-bp sequence that distinguish the *R. norvegicus*, the *R.* cf *moluccarius*, and the ancestral I-21p ORF I types are boxed. These are nonsynonymous substitutions as are the four remaining diagnostic differences between the L1mlvi2 clades, three of which reside in the region of ORF I shown here and are also boxed. The sixth diagnostic difference is at position 47 (not shown). All but one of the diagnostic differences between the ~1-kb ORF I of the $L1_{mlvi2}$ clades and ORF I of the other L1 subfamilies reside in the region shown here. The other diagnostic difference resides at position 52, and this difference also is nonsynonymous. Translations into amino acids are shown for the consensus DNA sequence of each of the four subfamilies: $L1_3$, $L1_4$, $L1_{mlvi2\text{-}rn}$, and $L1_{mlvi2\text{-}rm}$. Where there was no consensus possible, the amino acid is indicated by an "x." The dashes indicate sites that are identical to the consensus sequence, and the dots indicate gaps. The numbers indicate the position relative to the initiation codon of ORF I. For clarity, none of the tandem repeated copies of the 66-bp sequence are shown (compare with Fig. 1).


ORF I
element
class type
I-21a
I-21p
anc-type
mod-type
L1u -RnL1RtPf5
-RnL1_3
L1L4 -Lb6
-Rnvaoxy
L1L3 -ratlin3a
-Lb7
L1u -ratcyp4A2l
-rm2
L1mlvi2 -rm8
L1mlvi2 -rm11
-rm14
-rm38
-rm56
L1mlvi2 -rm6
-rm3
-rm8
-rm5
-rm17
-rm13
L1mlvi2 Consensus CTG GAG GAC TTC AAG AAA GAC ATG ... AAC ACA CTT AGG GAA ACA CAG GAA ATC ATT AAT AAA CAA GTA GAA GCC
3 bp deletion
21 bp sequence
insertion site for tandem repeats


ORF I
element
class type
I-21a
I-21p
anc-type
mod-type
L1u -RnL1RtPf5
-RnL1_3
L1L4 -Lb6
-Rnvaoxy
L1L3 -ratlin3a
-Lb7
L1u -ratcyp4A2l
-rm2
L1mlvi2 -rm8
L1mlvi2 -rm11
-rm14
-rm38
-rm56
L1mlvi2 -rm6
-rm3
-rm8
-rm5
-rm17
-rm13
L1mlvi2 Consensus TAC AGA GAG GAA TCG CAA AAA TCC CTG AAA - GAA TTC CAG GAA AAC ACA ATC AAA CAG TTG AAG GAA TTA AAA
66 bp sequence subject to tandem repitition

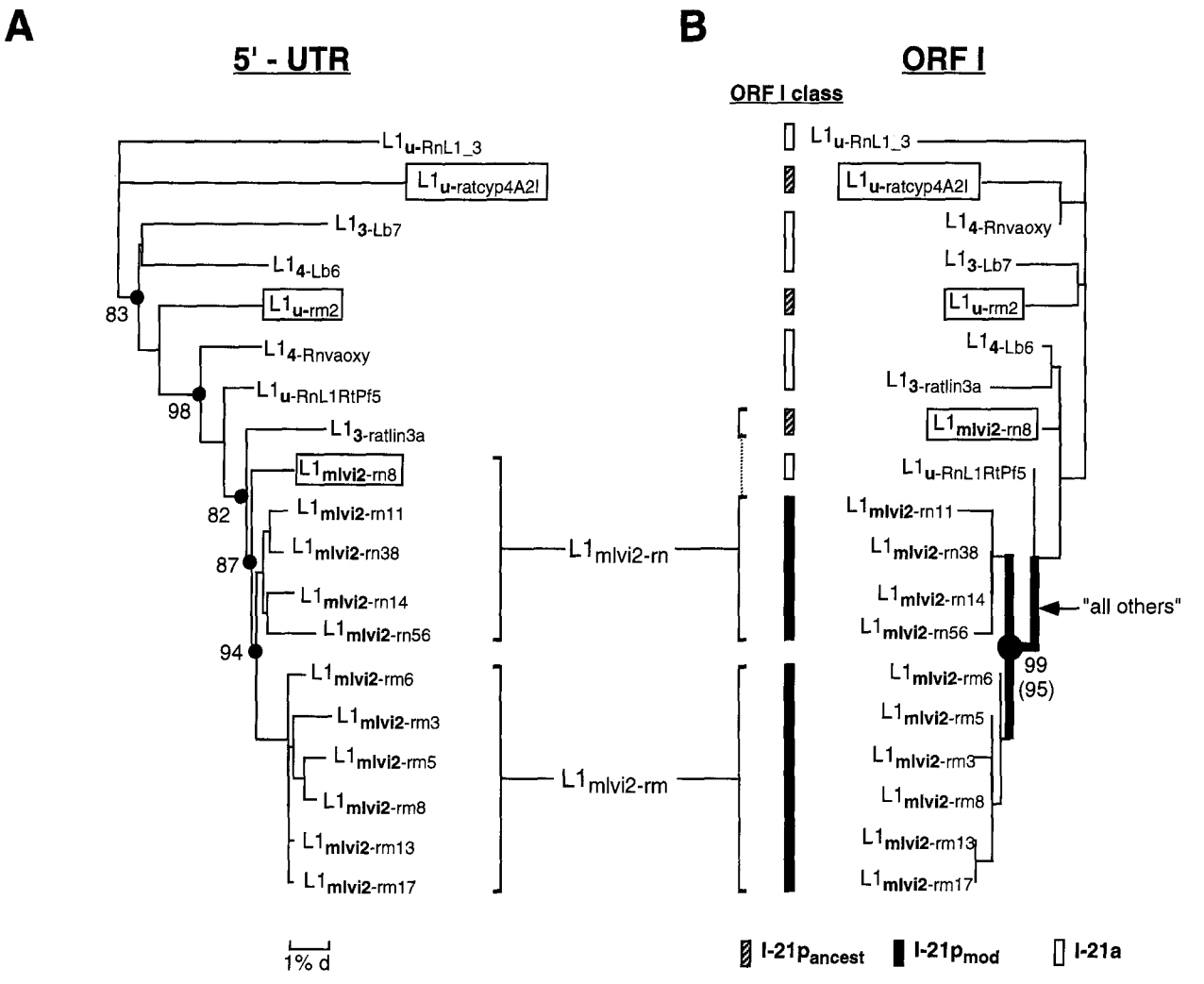


Fig. 4. Unrooted maximum likelihood trees for the 5′-UTR and ORF I of *Rattus* L1 elements. The trees shown were generated using PHYLIP 3.5 (see Materials and Methods) from the L1 elements shown in Figure 1. (**A**) Tree for the 5′-UTR. These sequences were also examined using neighbor joining on 1000 bootstrap replicates of the data using PHYLIP 3.5. This analysis generated a consensus tree topologically equivalent to the one shown and all the nodes supported by bootstrap values ≥80% are indicated with their values. Maximum likelihood on bootstrapped replicates and parsimony (using PAUP, see Materials and Methods) analysis of these sequences were computationally prohibitive. (**B**) 5′ portion of ORF I shown in Figure 1. These sequences were also analyzed by the neighbor joining method or parsimony on 1000 bootstrap replicates of the data. For the later analysis the presence or absence of the 21 bp sequence or AAG triplet (see Fig. 3) were scored using a weight of 2. Both methods supported only one node (heavy lines) with the bootstrap values given.

aged subclades of the modern $L1Md_A$ L1 family in mouse (E. Cabot and A.V. Furano, unpublished observations). Like the monomer, the tether is not conserved between *Rattus* and *Mus* (Scott et al. 1987; Furano et al. 1988; Adey et al. 1994a). Therefore, the high degree of sequence variability exhibited by the monomer does not appear to be directly related to the process of turnover of the 5′ UTR.

Phylogenetic Analysis of L1 Elements

The trees generated by the maximum likelihood, neighbor joining, or maximum parsimony methods with either the 5′ UTR or the ORF I sequence grouped the four I-21$p_{mod\text{-}type}$ $L1_{mlvi2\text{-}rn}$ elements and the six I-21$p_{mod\text{-}type}$ $L1_{mlvi2\text{-}rm}$ elements into two well-defined and statistically supported clusters. Figure 4A shows the tree generated from the 5′ UTR sequence by the maximum likelihood method annotated with the frequency with which a topologically equivalent tree was found by the neighbor joining method in 1000 bootstrap replicates of the data (see legend to Fig. 4). Figure 4B shows the maximum likelihood tree generated from the ORF I sequence. Since this method ignores gaps, the 21-bp sequence, the I-21$p_{mod\text{-}type}$ specific 3-bp deletion, and the tandem repeated copies of the 66-bp sequence were not included in the analysis. Including the 21-bp sequence and scoring its presence or absence and that of the I-21$p_{mod\text{-}type}$, specific 3-bp deletion in the parsimony analysis (see legend to Fig. 4) did not materially change the ORF I phylogeny. In no case did the $L1_{mlvi2\text{-}rn8}$ ORF I sequence share the same node as the I-21$p_{mod\text{-}type}$ $L1_{mlvi2\text{-}rn}$ and

I-21p$_{mod\text{-}type}$ L1$_{mlvi2\text{-}rm}$ clusters (results not shown, see the legend to Fig. 4). This reflects the fact that L1$_{mlvi2\text{-}rn8}$ ORF I is an I-21p$_{anc\text{-}type}$ sequence and shares numerous characters with I-21a elements (Fig. 3, and Hayward et al. 1997).

Only one node (highlighted in Fig. 4B) was consistently supported by all three methods using the ORF I sequence. This joined the following three branches: the "I-21p$_{mod\text{-}type}$ L1$_{mlvi2\text{-}rm}$" branch, the "I-21p$_{mod\text{-}type}$ L1$_{mlvi2\text{-}rn}$" branch, and the "all others" branch. The consensus trees produced from 1000 bootstrap replicates of the "all others" L1 elements by the neighbor joining and parsimony methods were inconsistent with each other, contained few if any statistically supported nodes, and were different from the maximum likelihood tree shown in Figure 4B. Such results would be explained if recombination between the ORF I regions of the older L1 elements occurred since this would intermix the various lineages. Some evidence for recombination between the older elements is apparent from the sequence alignments of ORF I (data not shown).

With the 5′-UTR sequence, all of the tree building methods placed the I-21p$_{anc\text{-}type}$ L1$_{mlvi2\text{-}rn8}$ element as a branch of the same node that is shared by the I-21p$_{mod\text{-}type}$ L1$_{mlvi2\text{-}rn}$ and L1$_{mlvi2\text{-}rm}$ clusters (Fig. 4A). By contrast, the other I-21p$_{anc\text{-}type}$ sequences, L1$_{u\text{-}rm2}$ and L1$_{u\text{-}Ratcyp4a21}$, were branches of far more distant nodes. This result strongly suggests that the latter two I-21p$_{anc\text{-}type}$ elements do not belong to the L1$_{mlvi2}$ clade. By contrast, the phylogenetic placement of L1$_{mlvi2\text{-}rn8}$ 5′ UTR is expected from our earlier results which showed that this element, despite its I-21p$_{anc\text{-}type}$ ORF I sequence, is a bonafide member of the L1$_{mlvi2}$ family (Hayward et al. 1997). Some of the results in Figure 4A also suggest that recombination occurred between the 5′ UTRs of the older L1 elements. For example, two members of the L1$_3$ subfamily, L1$_{3\text{-}Lb7}$ and L1$_{3\text{-}ratlin3a}$, were placed on quite separate branches of the tree. The former shared a node with an older L1$_4$ subfamily member, and the latter shared the same well-supported branch as the much younger L1$_{mlvi2}$ clade (Fig. 5B). Again, this recombination is also apparent from the sequence alignments (data not shown).

Copy Number of L1 Clades

We determined the copy number of the L1$_{mlvi2}$, L1$_3$, and L1$_4$ clades in *R. norvegicus* and *R.* cf *moluccarius*. For comparison, we also estimated the total copy number of modern L1 elements using generic hybridization probes derived from both the 5′ region of ORF I and the 3′ UTR (see Materials and Methods and Table 1). Control experiments showed that all of the probes are specific for their cognate subfamilies. For example, Figure 5 shows that the I-21p$_{mode\text{-}type}$ probes hybridize only to their cognate ORF I sequences (see lane 5c in panels A and B and the legend to Fig. 5). Experiments with the other probes produced similar results (data not presented or published elsewhere, Pascale et al. 1993; Usdin et al. 1995; Hayward et al. 1997).

Comparison of Figure 5A and B shows that while *R.* cf *moluccarius* contains only the L1$_{mlvi2\text{-}rm}$ subfamily, *R. norvegicus* contained both L1$_{mlvi2\text{-}rm}$ and L1$_{mlvi2\text{-}rn}$ elements. Thus, L1$_{mlvi2\text{-}rm}$ existed in the common ancestor of *R. norvegicus* and *R.* cf *moluccarius*. By contrast, L1$_{mlvi2\text{-}rn}$ either failed to amplify in *R.* cf *moluccarius* or emerged in *R. norvegicus* only after the two species diverged. Panels C and D of Figure 5 show respectively the hybridization of genomic DNA with PCR fragments derived from the 5′ terminus of ORF I and from a region of the 3′ UTR (see Materials and Methods and Fig. 1, upper diagram). The 3′ UTR probe detected a major 280-bp fragment which is the expected size from the location of the highly conserved *Hae*III and *Nla*III sites in this region of the 3′ UTR.

The ORF I probe hybridized to three major bands in the *Bgl*II/*Hind*III digests of genomic DNA. The smaller two represent different classes of ORF-I sequence (see Fig. 1). The 270-bp fragment is derived from I-21p elements and the 400-bp fragment is derived from I-21p elements that contain two extra copies of the 66-bp sequence (A.V. Furano, unpublished observations, Hayward et al. 1997). Not all the possible size classes of ORF I illustrated in Figure 1 are resolved on this gel. See the legend to Figure 4 for additional details. In other experiments, blots like these were hybridized with oligonucleotides specific for the I-21a ORF I lineage as well for the 3′ UTR region of other subfamilies and the quantitative analysis of these blots is summarized in Table 2.

L1$_{mlvi2\text{-}rm}$ elements are about 1.6 times as abundant in *R.* cf *moluccarius* as in *R. norvegicus* (Table 2, line 2). Therefore, despite the fact that *R.* cf *moluccarius* lacks L1$_{mlvi2\text{-}rn}$ elements, the total number of L1$_{mlvi2}$ elements in *R.* cf *moluccarius* is about equal to the total of both types in *R. norvegicus* (Table 2, lines 1 + 2). This equivalence was also observed with the 3′ UTR probe (Table 2, line 5). In both species, the L1$_{mlvi2}$ copy number based on 3′-UTR probes was about twice that determined with the ORF I probes. The actual ratio is probably lower than this because our ORF I measurements would not include I-21p$_{anc\text{-}type}$ L1$_{mlvi2}$ elements like L1$_{mlvi2\text{-}rn8}$. The ratio of 3′ UTR to ORF I using the generic L1 PCR probes is somewhat higher being 2.8 for *R. norvegicus* and 2.4 for *R.* cf *moluccarius*. Both values are similar to our earlier published estimates (D'Ambrosio et al. 1986; Nur et al. 1988). Most likely, this is due to the production of truncated elements during transposition.

Table 2 shows that the copy number of the I-21a L1 elements is about the same in both *R. norvegicus* and *R.* cf *moluccarius* (Table 2, line 3) and in each case is about 0.6 that of the I-21p$_{mod\text{-}type}$ L1$_{mlvi2}$ elements. So far, we know the subfamily classification of four I-21a elements;

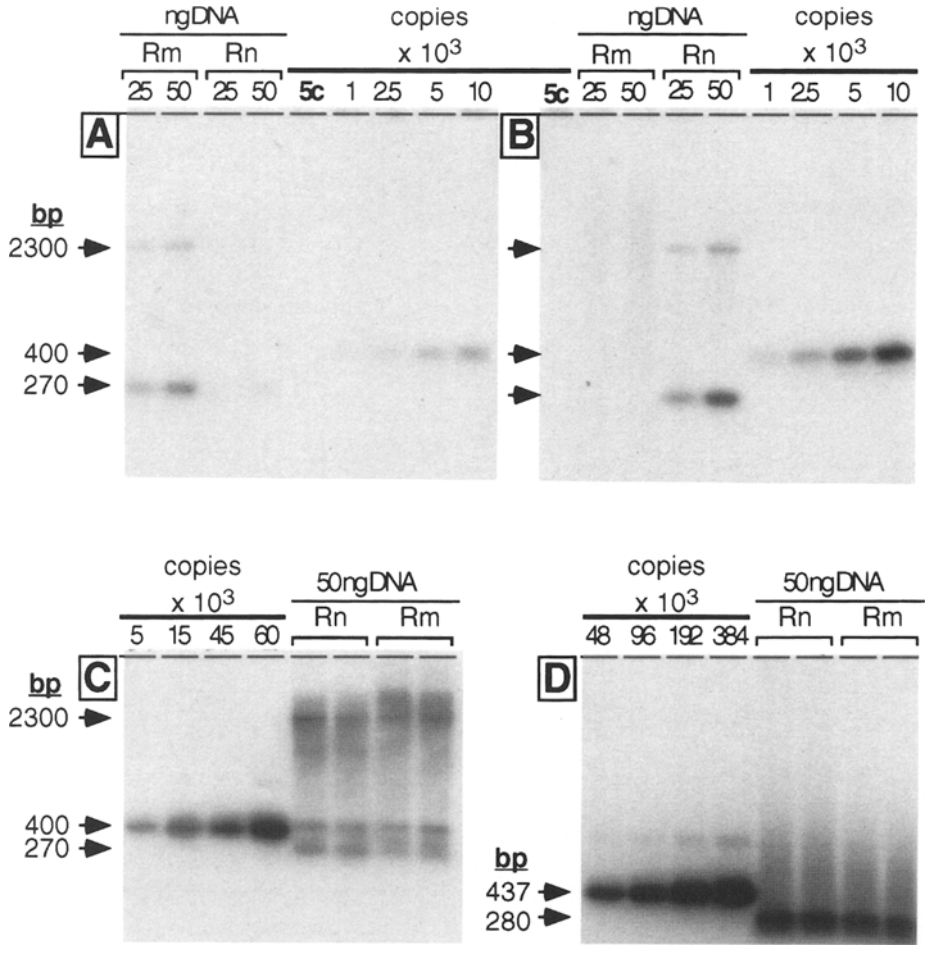


Fig. 5. Hybridizations of various L1 subfamily probes to blots of *R. norvegicus* and *R.* cf *moluccarius* DNA. Each panel shows a hybridization with a different probe: **(A)** The mlvimol oligonucleotide, specific for the I-21p$_{mod\text{-}type}$ L1$_{mlvi2\text{-}rn}$ elements. **(B)** The mlvirn oligonucleotide, specific for the I-21p$_{mod\text{-}type}$ L1$_{mlvi2\text{-}rn}$ elements. **(C)** The first 400 bp of ORF I of L1$_{mlvi2\text{-}rn14}$, a generic probe for all modern L1 subfamilies. **(D)** 437 bp of the 3′-UTR region of L1$_{3\text{-}lin3a}$ (see Fig. 1, top). This probe hybridizes to all modern *Rattus* L1 families, but not to the ancestral murine L1$_{mur\text{-}1}$ family (previously called Lx, Pascale et al. 1990, 1993). In **(A)** and **(B)** 25 and 50 ng of DNA from each species, digested with *Hind*III and *Bgl*II, were electrophoresed. In **(C)** and **(D)** 50 ng of *Hind*III/*Bgl*II or *Hae*III/*Nla*III digested DNA, respectively, from each species was electrophoresed in duplicate. The lanes labeled 5c in **(A)** and **(B)** contained 5000 genome equivalents (copies) of either the I-21p$_{mod\text{-}type}$ L1$_{mlvi2\text{-}rn}$ or L1$_{mlvi2\text{-}rm}$ ORF I DNA which served as negative controls for these hybridizations. The other lanes contained the indicated genome equivalents of L1$_{mlvi2\text{-}rm}$ **(A),** L1$_{mlvi2\text{-}rn}$ **(B and C),** or the 3′-UTR from L1$_{3\text{-}lin3a}$ **(D).** Control experiments, (not presented) showed that the ORF I probe hybridized to the same extent with ORF I fragments of the I-21p$_{anc\text{-}type}$, I-21p$_{mod\text{-}type}$, or I-21a. The 252 bp I-21a *Bgl*II/*Hind*III fragment is not resolved from the 270 bp I-21p fragment on these gels. Likewise, I-21a fragments that contain repeated 66-bp sequences or I-21p fragments that contain fewer than two extra 66-bp sequences are not resolved from the 270- and 400-bp fragments on these gels. The 2300-bp fragment is a *Bgl*II fragment derived from L1 elements that do not contain the *Hin*dIII site in ORF I (see upper diagram of Fig. 1). See Materials and Methods and Table 1 for additional details regarding the probes and positive and negative controls.

two are L1$_4$ and two are L1$_3$. The ratio of the L1$_4$ plus L1$_3$ 3′ UTR sequences to that of the I-21a sequence is about five (in both species) which is higher than the 3′ UTR/ORF I ratio of ~2 for the L1$_{mlvi2}$ subfamily or of ~2.6 for the total population of L1 elements. Under estimation of the I-21a sequence due to sequence variability of the older L1$_3$ and L1$_4$ elements could explain this difference (see Fig. 3). It is also possible that more members of these subfamilies are truncated than is the case for L1$_{mlvi2}$ or the general population of modern L1 elements. The 3′ UTR copy number of the combined L1$_{mlvi2}$, L1$_3$, and L1$_4$ subfamilies account for about one half the total (120,000–150,000) determined by the generic L1 probe. These genomes also contain older subfamilies, L1$_{4a}$ and L1$_5$ (Hayward et al. 1997), which we did not measure here that could account for some of this difference. However, these genomes may also contain yet to be detected major L1 subfamilies or numerous low copy number clades.

Although there are about 1.5 as many L1$_{mlvi2\text{-}rm}$ elements as L1$_{mlvi2\text{-}rn}$ elements in *R. norvegicus,* none of the five clones that we earlier selected from *R. norvegicus* using the L1$_{mlvi2}$-specific 3′ UTR oligonucleotide were of the L1$_{mlvi2\text{-}rm}$ type. This result could be due to chance. However, it may also mean that some L1$_{mlvi2\text{-}rm}$ elements contain a 3′ UTR other than the one cognate to the

Table 2. L1 copy number in *Rattus norvegicus* and *Rattus* cf *moluccarius*

	Thousands ± SE	
	R. norvegicus	*R.* cf *moluccarius*
ORF I		
I-21$_p$		
$L1_{mlvi2\text{-}rn}$	5.6 ± 0.2	0
$L1_{mlvi2\text{-}rm}$	8.7 ± 0.2	13.8 ± 3.1
I-21$_a$[a]	8.9 ± 1.3	7.9 ± 0.1
L1Rn[b]	53.3 ± 5.0	51.0 ± 2.8
3′ UTR		
$L1_{mlvi2}$[c]	29.7 ± 0.7	27.4 ± 0.0
$L1_3$	19.4 ± 1.6	13.0 ± 1.8
$L1_4$	28.0 ± 0.5	27.3 ± 0.1
L1Rn[b]	148.8 ± 1.1	121.1 ± 1.4

[a] Includes L1 elements that lack the 21-bp sequence such as members of the $L1_3$ and $L1_4$ subfamilies

[b] This category includes all modern L1 subfamilies but not the murine $L1_{mur\text{-}1}$ (Lx) family (Materials and Methods). We estimated earlier ~75,000 (Pascale et al. 1993) or ~120,000 (Witney and Furano 1984) copies of modern L1 elements in *R. norvegicus* from the percentage of λ genomic DNA clones (two different libraries) that hybridized to a modern L1 3′-UTR probe. The values here are more direct measurements, and from them we estimate that about 600,000 kb of L1 DNA has accumulated in these species since the murine radiation 10–15 Mya. The length of the elements detected with the ORF I probe is about 6 kb, and we assumed that the remaining ~85,000 copies detected with the 3′ UTR probe are on average ~2 kb. We also included the ~60,000 copies of the $L1_{mur\text{-}1}$ families using an average length of 2 kb.

[c] Includes $L1_{mlvi2\text{-}rn}$, $L1_{mlvi2\text{-}rm}$, and an unknown, but presumably small, number of I-21p_{ancest} $L1_{mlvi2}$ elements such as $L1_{mlvi2\text{-}rn8}$.

oligonucleotide that was used to screen for $L1_{mlvi2}$ elements. *R. norvegicus* contains a modern subfamily, $L1_{afp}$, that differs from the $L1_{mlvi2}$ subfamily by one base in a 290-bp region of the 3′ UTR, and this base difference is in the region cognate to the oligonucleotide used to select the $L1_{mlvi2}$ clones from *R. norvegicus*. However, the $L1_{afp}$ and $L1_{mlvi2}$ subfamilies share three diagnostic nucleotides in the 290-bp region that distinguish both from most members of the next closest subfamily, $L1_3$ (Hayward et al. 1997). These similarities and the nature of the base substitutions between the $L1_3$, $L1_{afp}$, and $L1_{mlvi2}$ subfamilies suggests that $L1_{afp}$ is an early version of $L1_{mlvi2}$. Therefore, the 3′ UTR of some $L1_{mlvi2\text{-}rm}$ elements, which amplified earlier than $L1_{mlvi2\text{-}rn}$ elements, may be the putative $L1_{afp}$ precursor version. However, since we found an equivalent amount of 5′ truncation for the total complement of $L1_{mlvi2}$ elements from both species, the $L1_{mlvi2}$ 3′-UTR oligonucleotide probe presumably detects a significant number of $L1_{mlvi2\text{-}rm}$ elements in both genomes.

Discussion

Here we further analyzed the rat $L1_{mlvi2}$ clade by examining L1 elements from *R.* cf *moluccarius* that contained the ORF-I 21-bp sequence that distinguishes the $L1_{mlvi2}$ subfamily in *R. norvegicus* from its antecedent $L1_3$ and $L1_4$ subfamilies. One of the *R.* cf *moluccarius* elements contained an I-21$p_{anc\text{-}type}$ ORF I sequence and proved not to be a member of the $L1_{mlvi2}$ subfamily. The remaining six were typical $L1_{mlvi2}$ I-21$p_{mod\text{-}type}$ elements. While these six were almost identical to each other, they differed from the $L1_{mlvi2\text{-}rn}$ I-21$p_{mod\text{-}type}$ elements at 13 diagnostic sites including two in the 21 bp sequence of ORF I (Fig. 3). Therefore, *R.* cf *moluccarius* contains a distinct version (or clade) of the $L1_{mlvi2}$ family, $L1_{mlvi2\text{-}rm}$. Using an oligonucleotide specific for just the 21-bp sequence of $L1_{mlvi2\text{-}rm}$, we detected $L1_{mlvi2\text{-}rm}$ elements in *R. norvegicus*. However, we have yet to characterize any of $L1_{mlvi2\text{-}rm}$ elements from this species. Using an oligonucleotide specific for the 21-bp sequence of the $L1_{mlvi2\text{-}rn}$ elements we found that this clade was limited to *R. norvegicus*.

A simple scenario that would account for the relationship between these $L1_{mlvi2}$ variants and their distribution in the two species would be that only the $L1_{mlvi2\text{-}rm}$ type element existed in the ancestor of *R. norvegicus* and *R.* cf *moluccarius*, and that $L1_{mlvi2\text{-}rn}$ was derived from the original $L1_{mlvi2\text{-}rm}$ some time after the divergence of the two species. The original $L1_{mlvi2\text{-}rm}$ element might have arisen from an I-21$p_{anc\text{-}type}$ $L1_{mlvi2}$ element like the *R. norvegicus* $L1_{mlvi2\text{-}rn8}$ element. However, a second interpretation is that both the $L1_{mlvi2\text{-}rm}$ and $L1_{mlvi2\text{-}rn}$ clades existed prior to the divergence of the two hosts, but that the $L1_{mlvi2\text{-}rn}$ lineage did not amplify in *R.* cf *moluccarius*. In this case one might imagine that both the $L1_{mlvi2\text{-}rm}$ and $L1_{mlvi2\text{-}rn}$ clades were derived from an I-21$p_{anc\text{-}type}$ $L1_{mlvi2}$ element. In fact, a comparison of the 21-bp sequence that distinguishes these three $L1_{mlvi2}$ variants (Figure 3) shows that the most parsimonious path between these sequences is the one that leads from the I-21$p_{anc\text{-}type}$ sequence to either of the other two. However, parsimony analysis of the remainder of the ORF I sequence does not convincingly support the latter model (Fig. 4B), and, at this point, the phylogenetic relationship between the I-21$p_{mod\text{-}type}$ ORF I sequences in $L1_{mlvi2\text{-}rm}$ and $L1_{mlvi2\text{-}rn}$ elements remains unresolved.

Since the $L1_{mlvi2\text{-}rn}$ clade amplified in *R. norvegicus* after it diverged from *R.* cf *moluccarius*, we can estimate a minimal amplification rate for this clade by dividing the copy number of $L1_{mlvi2\text{-}rn}$ (5,600, Table 2) by the time since the two species diverged. We calculated the later using mitochondrial DNA D-loop (control region) sequences from seven *R. norvegicus* and three *R.* cf *moluccarius* specimens (Usdin et al. 1995). Their average pairwise divergence after correction using the 2-parameter method of Kimura (1980) was 7.7%. Assuming that the divergence rate for the *Rattus* mtDNA control region is similar to the 20% per My for mice (Prager et al. 1993), the mitochondrial DNA divergence corresponds to a mean divergence time between the two *Rat-*

tus species of 0.385 My, with a maximum of 0.452 My. Thus the net rate of $L1_{mlvi2\text{-}rn}$ accumulation is about 1.23 new copies per 100 years, not counting the truncated copies not detected by our oligonucleotide for the I-21p insert. However, this rate might be higher if the amplification began well after the species diverged or if it ended well before the present.

The combined copy number of the $L1_{mlvi2\text{-}rn}$ and $L1_{mlvi2\text{-}rm}$ subfamilies in *R. norvegicus* is about the same as that of the latter subfamily alone in *R.* cf *moluccarius.* Although this could be coincidental, it may mean that the $L1_{mlvi2\text{-}rn}$ subfamily began amplifying in *R. norvegicus* at the expense of the $L1_{mlvi2\text{-}rm}$ clade. Casavant and Hardies (1994a) also have suggested that actively replicating L1 subfamilies compete with each other. Although the accumulation of inactivating mutations (Adey et al. 1994b) or the development of repressive measures by the host would reduce the replicative activity of older L1 subfamilies, it is possible that competition between L1 elements may be a major factor in determining which L1 variants emerge as successful subfamilies.

The hypervariable region of ORF I (Fig. 3) contains five of the six diagnostic differences between the $L1_{mlvi2\text{-}rn}$ and $L1_{mlvi2\text{-}rm}$ ORF I sequences. Each would result in an amino acid substitution. This region of ORF I is also uniquely subject to deletions and tandem repetition (Fig. 3, and Hayward et al. 1997). Whether these are adaptive changes that were favored by the possibly different selective "environments" provided by *R. norvegicus* and *R.* cf *moluccarius,* or merely mark a noncritical region of ORF-I can not be decided without further data. Studies on the coevolution of viruses and their hosts indicate that viral hypervariable regions may be the targets of host defensive measures and that the hypervariablity reflects adaptive responses by the virus to bypass these measures (e.g., Kilbourne 1994). It is also possible that the changes in ORF I are a response to changes in other parts of the L1 element. For example, the 5′ UTR of the $L1_{mlvi2\text{-}rn}$ and $L1_{mlvi2\text{-}rm}$ clades differ at seven positions.

Although numerous diagnostic changes differentiate the 5′ UTR of the 0.5 My $L1_{mlvi2}$ clades from the 2 My $L1_4$ subfamily (results not shown and Hayward et al. 1997), wholesale replacement of the 5′ UTR was not necessary for the replicative success of these subfamilies. This is consistent with functional studies that showed that the regulatory monomer from an old $L1_4$ element ($L1_{4\text{-}Lb6}$) is 20–50% as active as the more modern $L1_{mlvi2\text{-}rn14}$ monomer in a transient expression assay in an *R. norvegicus* cell line (B. Hayward and A.V. Furano, unpublished observations and Nur et al. 1988). Therefore, either the $L1_{4\text{-}Lb6}$ monomer has suffered from debilitating mutations as was demonstrated for the monomer of the mouse F subfamily (Adey et al. 1994a), or the modern *R. norvegicus* cells used for this experiment are more congenial to the newer $L1_{mlvi2}$ monomer than the older $L1_4$ monomer. In this regard, it would be interesting to compare the activity of $L1_{mlvi2\text{-}rm}$ and $L1_{mlvi2\text{-}rn}$ monomers in *R. norvegicus* and *R.* cf *moluccarius* cells.

The fixation of L1 elements in a species depends on such factors as the germline transposition rate, the effective population size of the host, and the fitness cost associated with L1 transposition. The latter may not be insignificant. For example, in less than 0.5 My of amplification, the $L1_{mlvi2}$ clade alone has contributed at least 84,000 kb of DNA to the genomes of *R. norvegicus* and *R.* cf *moluccarius,* including only the 14,000 copies that extend to the beginning of ORF I. If we include all of the modern L1 DNA measured in Table 2 as well as the ancestral $L1_{mur\text{-}1}$ family (previously called Lx), then about 600,000 kb of L1 DNA has been accumulated by these *Rattus* species in the 10–15 My since the murine radiation (see footnote b, Table 2). This represents more than 20% of the present mass of these genomes. In spite of this, L1 amplification appears to go on seemingly unabated at least if judged by the recent replicative success of the $L1_{mlvi2}$ family.

Acknowledgments. We acknowledge the National Cancer Institute for allocation of computing time and staff support at the Frederick Biomedical Supercomputing Center of the Frederick Cancer Research and Development Center.

References

Adey NB, Schichman SA, Graham DK, Peterson SN, Edgell MH, Hutchison III CA (1994a) Rodent L1 evolution has been driven by a single dominant lineage that has repeatedly acquired new transcriptional regulatory sequences. Mol Biol Evol 11:778–789

Adey NB, Tollefsbol TO, Sparks AB, Edgell MH, Hutchison III CA (1994b) Molecular resurrection of an extinct ancestral promoter for mouse L1. Proc Natl Acad Sci USA 91:1569–1573

Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA (1989) Current protocols in molecular biology. Wiley, New York

Cabot EL, Beckenbach AT (1989) Simultaneous editing of multiple nucleic acid and protein sequences with ESEE. Comput Applic Biosci 5:233–234

Casavant NC, Hardies SC (1994a) The dynamics of murine LINE-1 subfamily amplification. J Mol Biol 241:390–397

Casavant NC, Hardies SC (1994b) Shared sequence variants of *Mus spretus* LINE-1 elements tracing dispersal to within the last 1 million years. Genetics 137:565–572

Casavant NC, Hardies SC, Funk FD, Comer MB, Edgell MH, Hutchison III CA (1988) Extensive movement of LINES ONE sequences in beta-globin loci of *Mus caroli* and *Mus domesticus.* Mol Cell Biol 8:4669–4674

D'Ambrosio E, Waitzkin SD, Witney FR, Salemme A, Furano AV (1986) Structure of the highly repeated, long interspersed DNA family (LINE or L1Rn) of the rat. Mol Cell Biol 6:411–424

Fanning T, Singer MF (1987a) The LINE-1 DNA sequences in four mammalian orders predict proteins that conserve homologies to retrovirus proteins. Nucleic Acids Res 15:2251–2260

Fanning TG (1983) Size and structure of the highly repetitive BAM HI element in mice. Nucleic Acids Res 11:5073–5091

Fanning TG, Singer MF (1987b) LINE-1: a mammalian transposable element. Biochem Biophys Acta 910:203–212

Feinberg AP, Vogelstein B (1983) A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal Biochem 132:6–13

Felsenstein J (1993) PHYLIP (Phylogeny Inference Package). University of Washington, Seattle

Furano AV, Robb SM, Robb FT (1988) The structure of the regulatory region of the rat L1 (L1Rn, long interspersed repeated) DNA family of transposable elements. Nucleic Acids Res 16:9215–9231

Furano AV, Usdin K (1995) DNA "fossils" and phylogenetic analysis. J Biol Chem 270:25301–25304

Hardies SC, Martin SL, Voliva CF, Hutchison III CA, Edgell MH (1986) An analysis of replacement and synonymous changes in the rodent L1 repeat family. Mol Biol Evol 3:109–125

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

Hayward BE, Zavanelli M, Furano AV (1997) Recombination creates novel L1 (LINE 1) elements in *Rattus norvegicus*. Genetics 146:641–654

Hohjoh H, Singer MF (1996) Cytoplasmic ribonucleoprotein complexes containing human LINE-1 protein and RNA. EMBO J 15: 630–639

Hutchison III CA, Hardies SC, Loeb DD, Shehee WR, Edgell MH (1989) LINEs and related retroposons: long interspersed repeated sequences in the eucaryotic genome. Mobile DNA. Washington, D.C., American Society for Microbiology, 593–617

Jubier-Maurin V, Cuny G, Laurent A-M, Paquereau L, Roizes G (1992) A new 5′ sequence associated with mouse L1 elements is representative of a major class of L1 termini. Mol Biol Evol 9:41–55

Kazazian HH Jr, Wong C, Youssoufian H, Scott AF, Phillips DG, Antonarakis SE (1988) Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man. Nature 332:164–166

Kilbourne ED (1994) Host determination of viral evolution: a variable tautology. The evolutionary biology of viruses. Raven Press, New York, pp 253–271

Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16:111–120

Kolosha VO, Martin SL (1995) Polymorphic sequences encoding the first open reading frame protein from LINE-1 ribonucleoprotein particles. J Biol Chem 270:2868–2873

Kumar S, Tamura K, Nei M (1993) MEGA: Molecular evolutionary genetics analysis. The Pennsylvania State University, University Park

Martin SL (1991) Ribonucleoprotein particles with LINE-1 RNA in mouse embryonal carcinoma cells. Mol Cell Biol 11:4804–4807

Martin SL, Voliva CF, Hardies SC, Edgell MH, Hutchison III CA (1985) Tempo and mode of concerted evolution in the L1 repeat family of mice. Mol Biol Evol 2:127–140

Mathias SL, Scott AF, Kazazian HH Jr, Boeke JD, Gabriel A (1991) Reverse transcriptase encoded by a human transposable element. Science 254:1808–1810

Minakami R, Kurose K, Etoh K, Furuhata Y, Hattori M, Sakaki Y (1992) Identification of an internal *cis*-element essential for the human L1 transcription and a nuclear factor(s) binding to the element. Nucleic Acids Res 20:3139–3145

Nur I, Pascale E, Furano AV (1988) The left end of rat L1 (L1Rn, long interspersed repeated) DNA which is a CpG island can function as a promoter. Nucleic Acids Res 16:9233–9251

Padgett RW, Hutchison III CA, Edgell MH (1988) The F-type 5′ motif of mouse L1 elements: a major class of L1 termini similar to the A-type in organization but unrelated in sequence. Nucleic Acids Res 16:739–749

Pascale E, Liu C, Valle E, Usdin K, Furano AV (1993) The evolution of long interspersed repeated DNA (L1, LINE 1) as revealed by the analysis of an ancient rodent L1 DNA family. J Mol Evol 36:9–20

Pascale E, Valle E, Furano AV (1990) Amplification of an ancestral mammalian L1 family of long interspersed repeated DNA occurred just before the murine radiation. Proc Natl Acad Sci USA 87:9481–9485

Prager EM, Sage RD, Gyllensten U, Thomas WK, Hübner R, Jones CS, Noble L, Searle JB, Wilson AC (1993) Mitochondrial DNA sequence diversity and the colonization of Scandinavia by house mice from East Holstein. Bio J Linnean Soc 50:85–122

Schichman SA, Adey NB, Edgell MH, Hutchison III CA (1993) L1 A-monomer tandem arrays have expanded during the course of mouse L1 evolution. Mol Biol Evol 10:552–570

Schichman SA, Severynse DM, Edgell MH, Hutchison III CA (1992) Strand-specific LINE-1 transcription in mouse F9 cells originates from the youngest phylogenetic subgroup of LINE-1 elements. J Mol Biol 224:559–574

Scott AF, Schmeckpeper BJ, Abdelrazik M, Comey CT, O'Hara B, Rossiter JP, Cooley T, Heath P, Smith KD, Margolet L (1987) Origin of the human L1 elements: proposed progenitor genes deduced from a consensus DNA sequence. Genomics 1:113–125

Severynse DM, Hutchison III CA, Edgell MH (1992) Identification of transcriptional regulatory activity within the 5′ A-type monomer sequence of the mouse LINE-1 retroposon. Mammalian Genome 2:41–50

Singer MF (1995) Unusual reverse transcriptases. J Biol Chem 270: 24623–24626

Swergold GD (1990) Identification, characterization, and cell specificity of a human LINE-1 promoter. Mol Cell Biol 10:6718–6729

Swofford DL (1993) PAUP: Phylogenetic analysis using parsimony. Illinois Natural History Survey, Champaign

Usdin K, Chevret P, Catzeflis F, Verona R (1995) L1 (LINE) retrotransposable elements provide a "Fossil" record of the phylogenetic history of murid rodents. Mol Biol Evol 12:73–82

Voliva CF, Jahn CL, Comer MB, Hutchison III CA, Edgell MH (1983) The L1Md long interspersed repeat family in the mouse: Almost all examples are truncated at one end. Nucleic Acids Res 11:8847–8859

Voliva CF, Martin SL, Hutchison III CA, Edgell MH (1984) Dispersal process associated with the L1 family of interspersed repetitive DNA sequences. J Mol Biol 178:795–813

Wincker P, Jubier-Maurin V, Roizès G (1987) Unrelated sequences at the 5′ end of mouse LINE-1 repeated elements define two distinct subfamilies. Nucleic Acids Res 15:8593–8606

Witney FR, Furano AV (1984) Highly repeated DNA families in the rat. J Biol Chem 259:10481–10492