Copyright © 1997 by the Genetics Society of America

Recombination Creates Novel L1 (LINE-1) Elements in *Rattus norvegicus*

Bruce E. Hayward,[1] Mary Zavanelli[2] and Anthony V. Furano

Section on Genomic Structure and Function, Laboratory of Molecular and Cellular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892-0830

Manuscript received November 11, 1996
Accepted for publication February 21, 1997

ABSTRACT

Mammalian L1 (long interspersed repeated DNA, LINE-1) retrotransposons consist of a 5′ untranslated region (UTR) with regulatory properties, two protein encoding regions (ORF I, ORF II, which encodes a reverse transcriptase) and a 3′ UTR. L1 elements have been evolving in mammals for >100 million years and this process continues to generate novel L1 subfamilies in modern species. Here we characterized the youngest known subfamily in *Rattus norvegicus*, $L1_{mlvi2}$, and unexpectedly found that this element has a dual ancestry. While its 3′ UTR shares the same lineage as its nearest chronologically antecedent subfamilies, $L1_3$ and $L1_4$, its ORF I sequence does not. The $L1_{mlvi2}$ ORF I was derived from an ancestral ORF I sequence that was the evolutionary precursor of the $L1_3$ and $L1_4$ ORF I. We suggest that an ancestral ORF I sequence was recruited into the modern $L1_{mlvi2}$ subfamily by recombination that possibly could have resulted from template strand switching by the reverse transcriptase during L1 replication. This mechanism could also account for some of the structural features of rodent L1 5′ UTR and ORF I sequences including one of the more dramatic features of L1 evolution in mammals, namely the repeated acquisition of novel 5′ UTRs.

LONG interspersed repeated DNA (LINE-1, L1) elements are mammalian non-LTR retrotransposons that are ancient in origin and ubiquitous in nature (HUTCHISON *et al.* 1989; EICKBUSH 1994). A full length mammalian L1 element (Figure 1) is 6–7 kb and contains four regions: a 5′ untranslated region (UTR), which includes a regulatory region (NUR *et al.* 1988; SWERGOLD 1990; MINAKAMI *et al.* 1992; SEVERYNSE *et al.* 1992; KUROSE *et al.* 1995); two protein encoding regions (open reading frames, ORFs I and II), the latter of which is a reverse transcriptase (MATHIAS *et al.* 1991); and a 3′ UTR that contains a G-rich polypurine tract capable of forming unusual nucleic acid structures (R. HOWELL and K. USDIN, personal communication; USDIN and FURANO 1988; USDIN and FURANO 1989). The 3′ UTR terminates in an A-rich region, which may have originated as the polyadenylated tail of an L1 transcript (FANNING 1983; VOLIVA *et al.* 1984) or perhaps as a nontemplated product of the L1 reverse transcriptase (LUAN and EICKBUSH 1995).

L1 is believed to replicate by the production of an RNA copy that is then reverse transcribed and inserted in the genome (for a recent review see SINGER 1995). These insertions cause polymorphisms at individual loci (BURTON *et al.* 1985; ECONOMOU-PACHNIS *et al.* 1985; LAKSHMIKUMARAN *et al.* 1985; BELLIS *et al.* 1987), some of which were detected as genetic defects in humans and mice (KAZAZIAN *et al.* 1988; MIKI *et al.* 1992; HOLMES *et al.* 1994; KINGSMORE *et al.* 1994). Most L1 copies are defective (*e.g.*, incomplete, rearranged, etc.) and are not capable of further replication (FANNING 1983; VOLIVA *et al.* 1983; MARTIN *et al.* 1985; HARDIES *et al.* 1986; SCOTT *et al.* 1987). They are neither excised nor homogenized by recombination with intact elements, and diverge from each other with time due to the accumulation of random mutations, thereby becoming DNA "fossils" (CASAVANT *et al.* 1988; PASCALE *et al.* 1990; HARDISON and MILLER 1993; PASCALE *et al.* 1993; VANLERBERGHE *et al.* 1993; CASAVANT and HARDIES 1994b; SMIT *et al.* 1995).

L1 has been replicating and evolving in mammals for ~100 million years (BURTON *et al.* 1986; HARDISON and MILLER 1993; SMIT *et al.* 1995). L1 evolution is quite rapid and novel subfamilies rapidly succeed each other or even coexist (present work; MARTIN *et al.* 1985; HARDIES *et al.* 1986; PADGETT *et al.* 1988; JUBIER-MAURIN *et al.* 1992; DOMBROSKI *et al.* 1993; PASCALE *et al.* 1993; ADEY *et al.* 1994a; CASAVANT and HARDIES 1994a,b; FURANO *et al.* 1994; HOLMES *et al.* 1994). Comparisons between L1 subfamilies have revealed distinctions that may be relevant to both the evolution and biological properties of L1 elements. For example, the 5′ $^1/_3$ of ORF I of modern mammalian L1 families can be quite divergent, in marked contrast to the rest of ORF I and all of ORF II (SCOTT *et al.* 1987; FURANO *et al.* 1988;

Corresponding author: Anthony V. Furano, Buildg. 8, Room 203, NIH-NIDDK, 8 Center Dr. MSC 0830, Bethesda, MD 20892-0830.
E-mail: avf@helix.nih.gov
[1] *Present address:* B. E. Hayward, Human Genetics Unit, Molecular Medicine Centre, Western General Hospital, University of Edinburgh, Edinburgh EH4 2XU, United Kingdom.
[2] *Present address:* M. Zavanelli, Biology Department, Sinsheimer Laboratories, University of California, Santa Cruz, CA 95064.

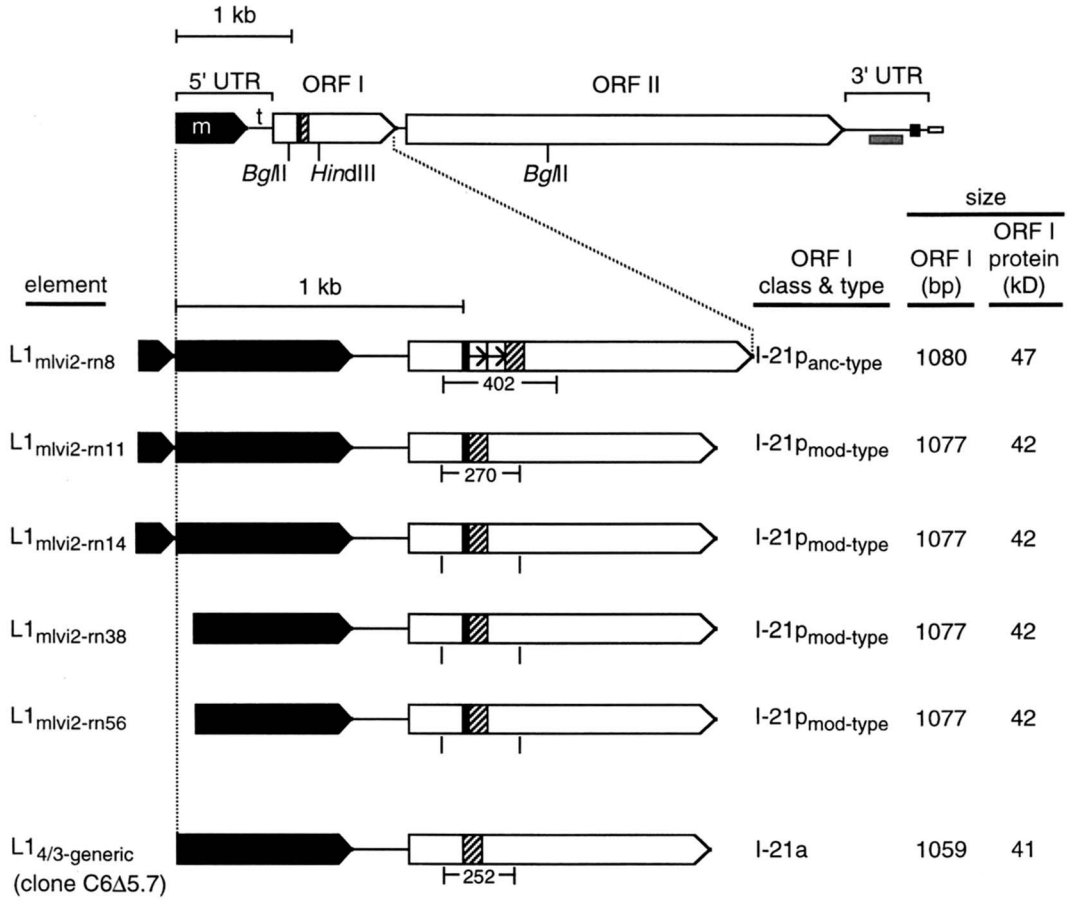


FIGURE 1.—Diagrammatic representation of different L1 elements. The top diagram represents a generic rodent L1 element. The 5′ UTR is divided into two regions: a sequence with regulatory properties labeled m for monomer (PADGETT *et al.* 1988) since it can be present in tandem multiple copies in both rat and mouse L1 elements; the remainder is labeled t for tether. The narrow black box within ORF I represents the 21-bp sequence, the presence or absence of which defines the I-21p and I-21a classes of ORF I, respectively (see text). The contiguous cross-hatched box designates the 66-bp sequence that is tandem repeated in some L1 elements. The small black and open boxes at the 3′ end represent the G-rich polypurine stretch and A-rich 3′ terminus, respectively. The filled box below the 3′ UTR corresponds to the aligned DNA sequence shown in Figure 2. The bottom diagrams, at 2.5 times the scale of the top one, represent the 5′ UTR and ORF I region of the various L1 elements examined in this study. The boxes with the arrow represent the tandem repeated 66-bp sequence found in the $L1_{mlvi2\text{-}rn8}$ element. A generic I-21a L1 element is at the bottom and the ORF I sequence of such an element is in clone C6Δ5.7. The tick marks beneath ORF I represent conserved *Bgl*II and *Hin*dIII sites, respectively, and their length in bp is given. The three columns to the right of the lower six diagrams indicate, respectively, the ORF I class and type (of I-21p sequence), length of the ORF I sequence and the calculated size of the encoded ORF I protein.

DEMERS *et al.* 1989). In addition, acquisition of a novel 5′ UTR has occurred repeatedly during L1 evolution (SCOTT *et al.* 1987; WINCKER *et al.* 1987; FURANO *et al.* 1988; JUBIER-MAURIN *et al.* 1992; ADEY *et al.* 1994a,b). In some cases the 5′ UTR of even closely related L1 subfamilies, such as the A and F subfamilies in *Mus domesticus*, are not homologous (FANNING 1983; WINCKER *et al.* 1987; PADGETT *et al.* 1988). Therefore, novelty in the 5′ UTR or the first third of the ORF I protein, or both, may be required to maintain the replication competence of L1 elements.

Here we characterized the most recently emerged L1 subfamily that we have so far identified in *R. norvegicus*, $L1_{mlvi2}$ (FURANO and USDIN 1995; USDIN *et al.* 1995). The $L1_{mlvi2}$ 3′ UTR is of the same lineage as the older predecessor $L1_4$ and $L1_3$ subfamilies (D'AMBROSIO *et al.* 1986; FURANO *et al.* 1988). However, the $L1_{mlvi2}$ ORF I sequence did not descend from the ORF I of the $L1_{4/3}$ subfamilies but represents an ancestral class of ORF I sequence from which the $L1_{4/3}$ ORF I sequences were derived. We emphasize that we use ancestral not to refer to $L1_{mlvi2}$ elements but only to the class of ORF I sequences that such elements contain. The $L1_{mlvi2}$ class of ORF I sequence contains a 21-bp sequence that was deleted upon formation of the $L1_{4/3}$ ORF I class. Therefore, we call the ancestral 21-bp-containing class of ORF-I sequences, I-21p (ORF I 21 bp present), and the ORF I sequences in the $L1_{4/3}$ subfamilies that lack the 21-bp sequence, the I-21a class (ORF I 21 bp absent). We suggest that the recruitment of the ancestral class

of I-21p ORF I sequence to form the modern $L1_{mlvi2}$ subfamily was the result of a recombination event.

MATERIALS AND METHODS

General methods: Genomic clones were isolated from a commercial λGEM library (Promega) prepared from a partial *Sau3A*1 digest of male *R. norvegicus* DNA. The desired clones were identified (by hybridization with different oligonucleotides, see RESULTS), isolated, propagated and subcloned using standard techniques (AUSUBEL *et al.* 1989). Electrophoresis, nucleic acid blots [capillary or electroblot (Trans-Blot SD, Bio-Rad) to nylon membranes (Zeta-Probe, Life Technologies)], hybridizations (in 7% NaSDS, 0.25 M $NaPO_4$ pH 7.0, 0.025 M NaEDTA), PCR, 5′ radiolabeling of oligonucleotides with [^{32}P]ATP, and other molecular techniques followed standard protocols (AUSUBEL *et al.* 1989) or have been described earlier (PASCALE *et al.* 1993; USDIN *et al.* 1995). In the case of commercially supplied kits or apparatus the supplier's recommendations were followed.

DNA sequence determination and analysis: In all cases DNA sequence determination was carried out on both strands. The λ clones or subclones derived from them were sequenced using the cycle sequencing kit from Life Technologies. In some cases DNA sequences were determined using the Sequenase Version 2.0 kit from U.S. Biochemicals.

DNA sequence manipulations and alignments were carried out using the suite of programs provided in the Wisconsin Package (Program Manual for the Wisconsin Package, Version 8, September 1994, Genetics Computer Group, 575 Science Drive, Madison, Wisconsin, USA 53711). Phylogenetic analysis was carried out using the neighbor joining method (SAITOU and NEI 1987) as implemented in the Wisconsin Package or using a parsimony method (SWOFFORD 1993). All of the sequences reported here have been submitted to GenBank and their accession numbers are as follows: $L1_{mlvi2\text{-}rn8}$ 5′ UTR and ORF I, U87598; $L1_{mlvi2\text{-}rn8}$ 3′ UTR, U87599; $L1_{mlvi2\text{-}rn11}$ 5′ UTR and ORF I, U87600; $L1_{mlvi2\text{-}rn11}$ 3′ UTR, U87601; $L1_{mlvi2\text{-}rn14}$ 5′ UTR and ORF I, U87602; $L1_{mlvi2\text{-}rn14}$ 3′ UTR, U87603; $L1_{mlvi2\text{-}rn38}$ 5′ UTR and ORF I, U87604; $L1_{mlvi2\text{-}rn38}$ 3′ UTR, U87605; $L1_{mlvi2\text{-}rn56}$ 5′ UTR and ORF I, U87606; $L1_{mlvi2\text{-}rn56}$ 3′ UTR, U87607; $L1_{4\text{-}rnL17}$ 3′ UTR, U87608; $L1_{mlvi2\text{-}rnL18}$ 3′ UTR, U87609; $L1_{mlvi2\text{-}rnL6}$ 3′ UTR, U87610; $L1_{mlvv2\text{-}rnL7}$ 3′ UTR, U87611; $L1_{3\text{-}rnL9}$ 3′ UTR, U87612.

***In vitro* transcription and translation:** The ORF I sequence of selected $L1_{mlvi2}$ elements was amplified either directly from the λ clones or from subclones thereof by the PCR. The 5′ PCR primer, Sp6-ORF I, has the following sequence: acg cgt ATT TAG GTG ACA CTA TAG AAG ccg cca ccA TGG CGA GAG GCA AGC GCA. The first six nucleotides comprise an *Mlu*I site, followed by the Sp6 promoter (upper case) a "Kozak" sequence (KOZAK 1987) and then the first 18 bases of ORF I (upper case). The sequence of the 3′ primer, ORF I-6, is CTG AGT CCC GGG CTA TTA GTT TCT TGC TTC TTC TAG GGT. The last 24 nucleotides of this oligonucleotide correspond to ORF I sequence; the first 15 nucleotides consist of a hexamer of arbitrary sequence followed by a *Sma*I site followed by an additional termination codon. The PCR products were either used directly or after purification (treatment with proteinase K and passage through Sepharose CL-6B) in either a RiboMAX (Promega) *in vitro* transcription kit or in a 50-μl reaction containing 15 units of Sp6 RNA polymerase (Life Technologies), 40 units of RNAse block (Stratagene), 0.4 mM NTP, 40 mM Tris-HCl pH 7.9, 6 mM $MgCl_2$, 2 mM spermidine, and 2 mM DTT. The reactions were incubated for 1–2 hr at 37°. One fifth of the product was used directly in a 30 μl rabbit reticulocyte lysate *in vitro* translation system (either from Life Technologies, or the Flexi system from Promega) and the products were electrophoresed on a polyacrylamide gel that was dried and autoradiographed. Reticulocyte lysates from Life Technologies contained significantly less nonspecific protease activity than those from Promega. The protease inhibitor, aprotonin, was supplied by ICN Biomedicals.

RESULTS

Modern L1 subfamilies in *R. norvegicus*: Figure 2 shows an alignment of a portion of the 3′ UTR sequence of various *R. norvegicus* L1 elements determined here or obtained from the Genbank data base. A number of subfamilies can be distinguished by shared sequence similarities in the compared region of the 3′ UTR. These subfamilies range in age from <0.5 my for $L1_{mlvi2}$ to 10 my for $L1_5$ (Figure 2). Rattus sensu stricto and Rattus sensu lato diverged ~3.5 mya and species within Rattus sensu stricto began diverging from each other ~2.5 mya (see USDIN *et al.* 1995). The phylogenetic distribution of the $L1_4$ subfamily is limited to Rattus sensu stricto, which is consistent the ~2 my age of the subfamily (USDIN *et al.* 1995). By contrast, the older $L1_{4a}$ and $L1_5$ subfamilies are present in both Rattus

FIGURE 2.—L1 subfamilies in *Rattus norvegicus*. The sequence aligned corresponds to the box shown under the 3′ UTR of Figure 1 (top). Sequences $L1_{4\text{-}rnL17}$, $L1_{3\text{-}rnL9}$, $L1_{mlvi2\text{-}rnL6}$, $L1_{mlvi2\text{-}rnL7}$, $L1_{mlvi2\text{-}rnL18}$, $L1_{mlvi2\text{-}rn8}$, $L1_{mlvi2\text{-}rn11}$, $L1_{mlvi2\text{-}rn14}$, $L1_{mlvi2\text{-}rn38}$, and $L1_{mlvi2\text{-}rn56}$ are cloned L1 elements selected from a genomic library using subfamily-specific oligonucleotides as described in MATERIALS AND METHODS. All of the others are from GenBank. The Rnmlvi2r sequence is from the first described $L1_{mlvi2}$ element (ECONOMOU-PAGHNIS *et al.* 1985; FURANO *et al.* 1986). The sequence of the various oligonucleotides used here are either given or boxed. Note that there are two $L1_{mlvi2}$-specific oligonucleotides: one used for hybridization, L1-mlvi2, and one used for PCR amplification, L1a2.2. The dashes indicate identity with the consensus sequence and the dots indicate deletions. Missing data is indicated by spaces. The lower right corner shows the phylogenetic analysis of these sequences using the neighbor joining method (left) or a parsimony method (right) as described in MATERIALS AND METHODS. For the latter a consensus sequence was derived from the members of each major subfamily (indicated by the heavy vertical bars). The $L1_{mlvi2}$ subfamily was subdivided into those that lacked (■) or contained $L1_{mlvi2\text{-}t}$ (□), a T at position 122. The tree shown on the right was rooted using the $L1_5$ subfamily as an outgroup and was the single shortest one generated by the exhaustive search option. The numbers above the branches are the branch length and those below are the percent times that the indicated node appeared in 1000 bootstrap replicates. The low bootstrap values for the $L1_{afp}/L1_{mlvi2}$ or $L1_{mlvi2}/L1_{mlvi2\text{-}t}$ nodes would be expected given the very few characters that distinguish these subfamilies from each other. We consider the $L1_{afp}$ and $L1_{mlvi2\text{-}t}$ groups to be variants of the $L1_{mlvi2}$ subfamily (see text). The only members of the $L1_{mlvi2}$ and $L1_{mlvi2\text{-}t}$ subfamilies that were not grouped within their cohort by the neighbor joining method were those where significant amount of sequence data was missing (see alignment).

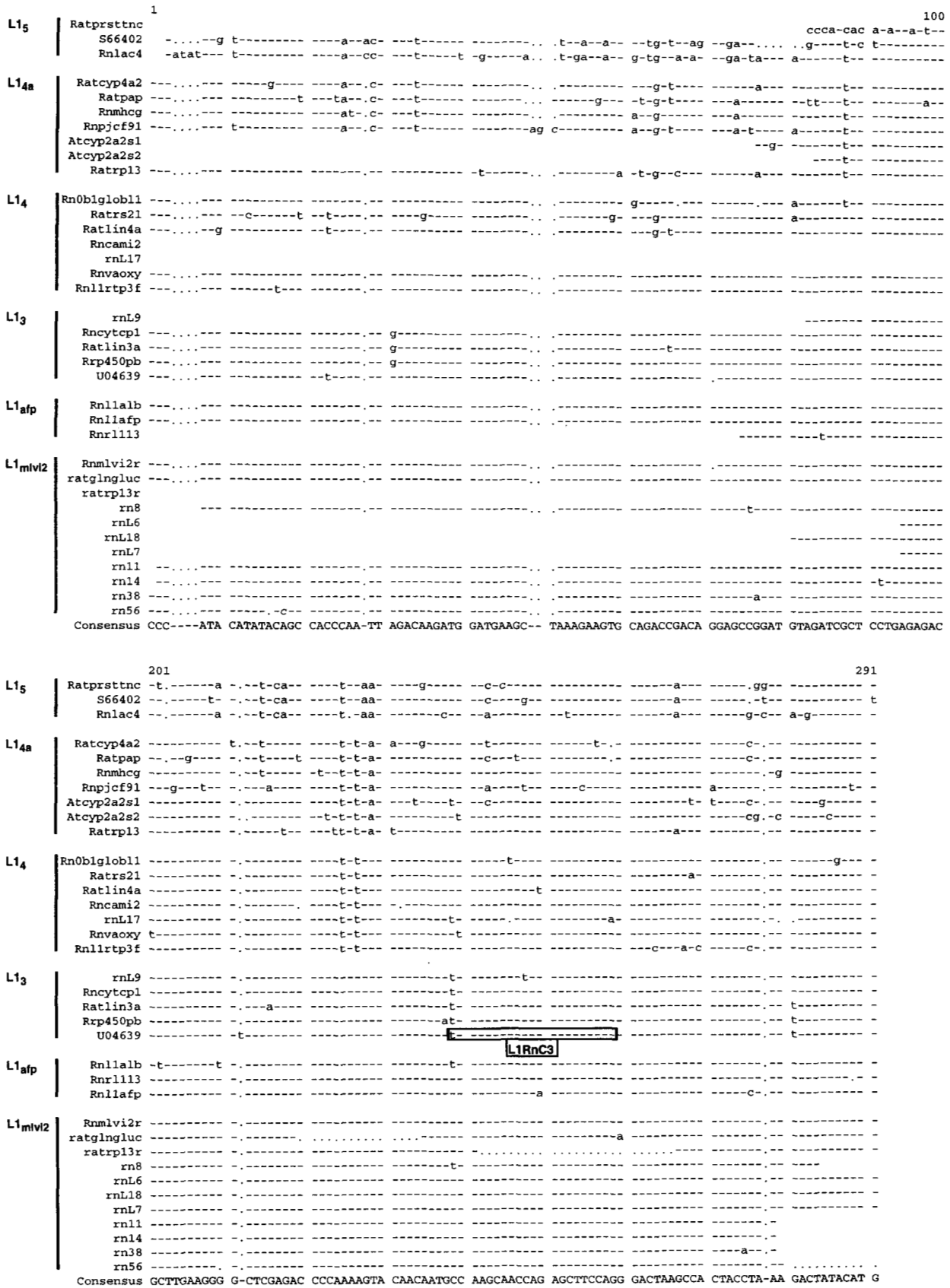
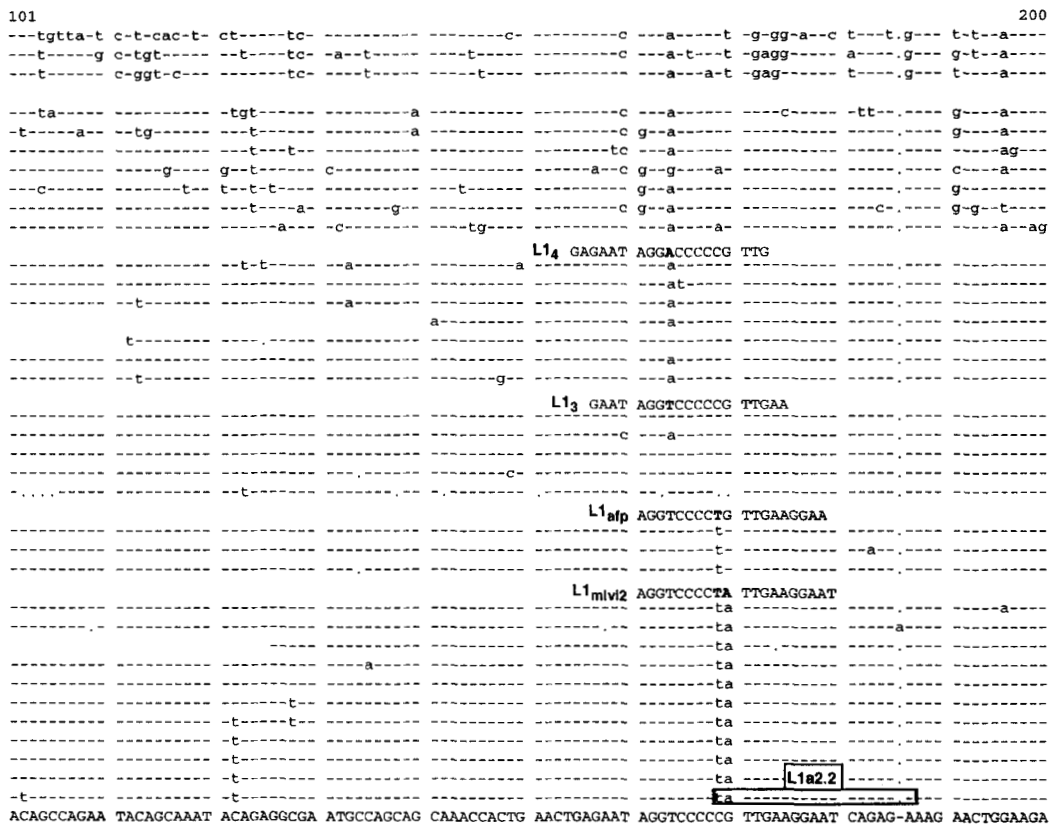


FIGURE 2



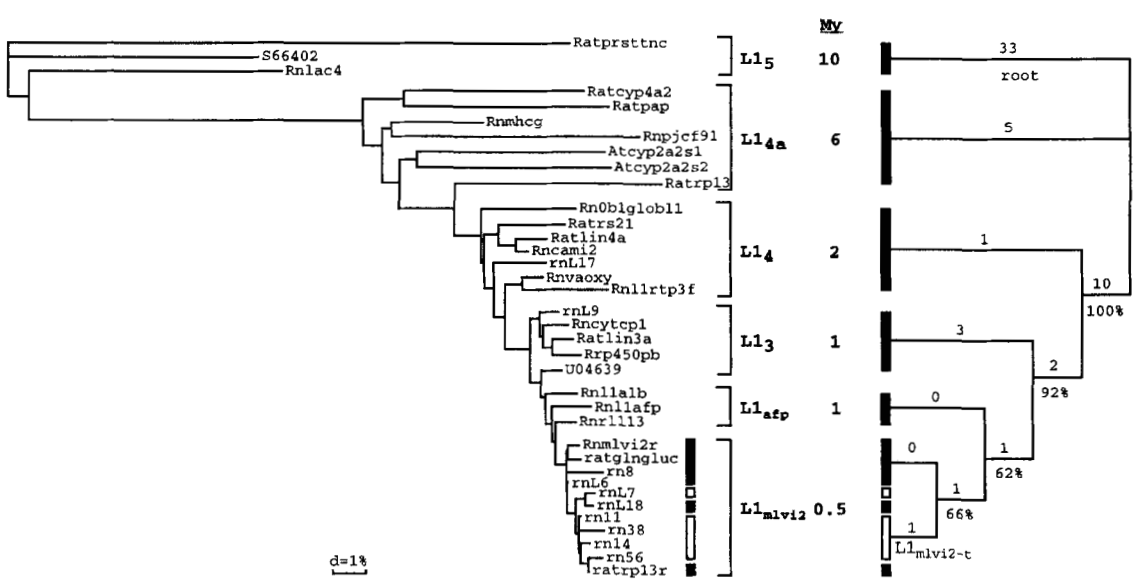


FIGURE 2.—*Continued*

sensu stricto and Rattus sensu lato (VERNEAU *et al.* 1997). Here we consider only the L1 subfamilies that are limited to Rattus sensu stricto as modern *R. norvegicus* L1 subfamilies.

Hybridization with the L1-mlvi2 oligonucleotide, which is cognate to the diagnostic L1$_{mlvi2}$ nucleotides at positions 169, 170 (see Figure 2), showed that the L1$_{mlvi2}$ subfamily is limited to just *R. norvegicus* and its sister taxon, *R.* cf *moluccarius* (USDIN *et al.* 1995). This limited phylogenetic distribution is consistent with the

recent emergence of the $L1_{mlvi2}$ subfamily (FURANO and USDIN 1995; USDIN *et al.* 1995). In addition, the Mlvi-2 (Murine leukemia virus integration site 2) locus, which is the integration site for the first discovered $L1_{mlvi2}$ element (Rnmlvi2r in Figure 2), is dimorphic in various strains of the laboratory rat (*R. norvegicus*) for the presence or absence of the element (ECONOMOU-PACHNIS *et al.* 1985). The fact that the L1 insertion at this locus has not yet been fixed in the *R. norvegicus* population also indicates that it occurred relatively recently.

Although we grouped the three $L1_{afp}$ elements separately from the $L1_{mlvi2}$ members, the $L1_{afp}$ subfamily may just be an early version of the $L1_{mlvi2}$ subfamily. In the compared region of the 3′ UTR, the three $L1_{afp}$ elements differ from the $L1_{mlvi2}$ elements only at diagnostic position 170; the $L1_{afp}$ elements have the G of the older subfamilies rather than the A of $L1_{mlvi2}$. On the other hand, the $L1_{afp}$ and $L1_{mlvi2}$ elements share three diagnostic differences (positions 31, 231, and 280) that distinguish both from most members of the next closest $L1_3$ subfamily. Figure 2 also shows a phylogenetic analysis of the aligned sequences, using either the neighbor joining method (which generates a tree based on the pair-wise distances between these sequences) or parsimony (which generates a tree based on character differences). The $L1_{4a}$, $L1_4$, $L1_3$, $L1_{afp}$, and $L1_{mlvi2}$ elements belong to a single major lineage, and the $L1_{afp}$ and $L1_{mlvi2}$ elements share a well supported node in the parsimony-generated tree (Figure 2). As judged from quantitative blots of genomic DNA, the $L1_4$, $L1_3$, and $L1_{mlvi2}$ subfamilies contain 20,000–30,000 members each (CABOT *et al.* 1997).

Isolation of $L1_{mlvi2}$ elements: We identified full length $L1_{mlvi2}$ elements by rescreening plaques that hybridized to the $L1_{mlvi2}$ 3′ UTR oligonucleotide (L1-mlvi2, Figure 2) with an oligonucleotide cognate to a conserved region of the 5′ UTR (oligonucleotide A1, Figure 4). Positive plaques were arbitrarily selected ($L1_{mlvi2\text{-}rn8}$, $L1_{mlvi2\text{-}rn11}$, $L1_{mlvi2\text{-}rn14}$, $L1_{mlvi2\text{-}rn38}$, $L1_{mlvi2\text{-}rn56}$) and DNA was prepared from them. To initially characterize these elements, we determined whether the L1 ORFs could be translated into their expected protein products. We examined ORF I first, amplifying this sequence using the PCR and primers cognate to the 5′ and 3′ sequence of ORF I as described in MATERIALS AND METHODS. The 5′ primer also contained an SP6 RNA polymerase promoter site. Each SP6-ORF I fragment was transcribed *in vitro* and the RNA was translated in a rabbit reticulocyte lysate as described in MATERIALS AND METHODS. Figure 3 shows that different maximal sized ORF I products (white dots) and a series of smaller bands was produced from the ORF-I transcripts. The maximal sized product of the $L1_{mlvi2\text{-}rn11...rn56}$ clones was ~43 kD; a larger one was produced from the $L1_{mlvi2\text{-}rn8}$ element, and a somewhat smaller one from clone C6Δ5.7. On this gel the mobility of all of the C6Δ5.7 translation products was

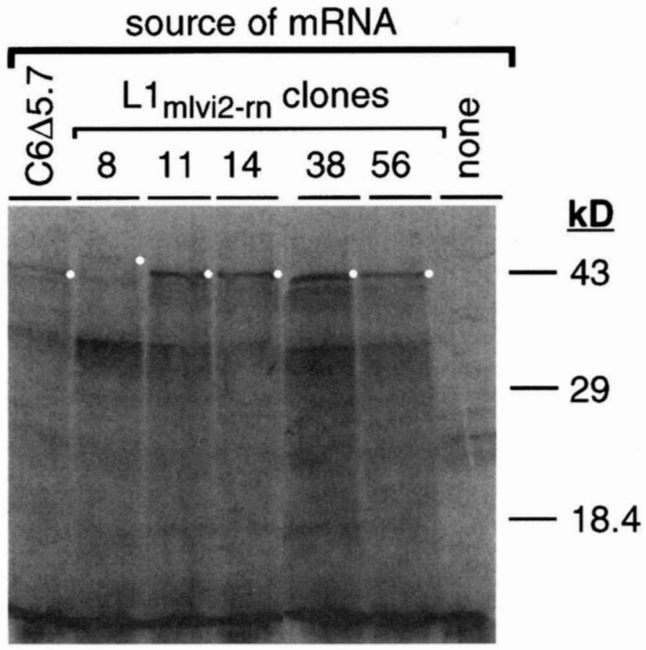


FIGURE 3.—Arbitrarily selected $L1_{mlvi2}$ members contain an intact ORF I sequence. RNA transcripts synthesized from templates prepared from the indicated cloned L1 elements were used to program protein synthesis *in vitro* using a reticulocyte system as described in MATERIALS AND METHODS. The C6Δ5.7 clone is an I-21a element that contains no internal ORF I tandem repeats (see text and the bottom most diagram of Figure 1 for additional details). The lane labeled none contained no mRNA.

retarded. Therefore, the small but distinct difference between the mobility of the maximal $L1_{mlvi2\text{-}rn11...rn56}$ and C6Δ5.7 bands, which is easily seen on other gels, is not so evident here. The C6Δ5.7 clone contains an intact ORF I constructed from two previously described rat L1 elements (bottom-most diagram Figure 1). These results indicated that different class sizes of ORF I sequence are represented in these L1 clones, and this was confirmed by DNA sequence determination (see below).

The smaller translation products are probably not the translation products of degraded mRNA but the degradation products of the full length ORF I protein, which, being very basic, contains numerous protease recognition sites. We base this conclusion on a series of control experiments that we do not present here. First, formaldehyde gel electrophoresis showed that the ORF I transcripts were almost entirely full length. Second, the transcripts are not extensively degraded by the reticulocyte extracts, since chloramphenicol acetyl transferase (CAT) mRNA produced a single band of the expected size. Third, aprotonin, a protease inhibitor, increased the yield of the full length ORF I product but had no effect on the yield of the CAT protein. No other protease inhibitor was as efficacious. Similar experiments using ORF II transcripts did not produce

the predicted full length-sized ORF II product; only degradation products were seen, even in the presence of aprotonin.

The $L1_{mlvi2}$ elements contain a distinctive ORF I sequence: We determined the DNA sequence of the 5′ UTR of $L1_{mlvi2\text{-}rn8...rn56}$ and of part ($L1_{mlvi2\text{-}rn8}$, $\text{-}rn11$, $\text{-}rn38$) or all ($L1_{mlvi2\text{-}rn14}$, $\text{-}rn56$) of ORF I. Figure 4 shows that a major difference between the $L1_{mlvi2}$ elements and previously characterized $L1_4$ or $L1_3$ elements is that the $L1_{mlvi2}$ ORF I contains a 21-bp sequence that is lacking in the $L1_4$ or $L1_3$ ORF I (Figures 1 and 4). Therefore, different classes of ORF I characterize the $L1_{mlvi2}$ and the $L1_4$ and $L1_3$ subfamilies: the I-21p class (ORF I 21 bp present, $L1_{mlvi2}$) and the I-21a class (21 bp absent, $L1_4$, $L1_3$). In other ways however, the 5′ UTR and ORF I regions of the $L1_{mlvi2}$ elements are similar to previously described rat L1 elements. First, the monomer region of the 5′ UTR is often either partially tandem duplicated or truncated (FURANO *et al.* 1988). Second, the $L1_{mlvi2\text{-}rn8}$ element contains a tandem repetition of two copies of the 66-bp sequence that is immediately 3′ of the 21-bp insertion (Figures 1 and 4). Some I-21a class elements also contain multiple copies of the 66-bp sequence (SOARES *et al.* 1985; D'AMBROSIO *et al.* 1986; FURANO *et al.* 1988).

The C6Δ5.7 ORF I sequence, which produced the smallest full length *in vitro* translation product (Figure 3), contains an I-21a ORF I that lacks 66-bp repeats and therefore is the smallest ORF I sequence (Figure 1, bottom-most diagram). By contrast the I-21p $L1_{mlvi2\text{-}rn8}$ ORF I contains two extra copies of the 66-bp sequence, encodes the largest protein, and produces the largest *in vitro* translation product (Figure 3). The I-21p $L1_{mlvi2\text{-}rn11...rn56}$ ORF I lack 66-bp repeats and each encode a protein of intermediate size, and each produced a product of intermediate size (Figure 3). Therefore, the results of the *in vitro* translation experiment are congruent with the difference between the various ORF I DNA sequences of the elements used.

Relationship between the different L1 ORF I DNA sequences: Although the $L1_{mlvi2\text{-}rn8}$ ORF I contains the 21-bp sequence, it is otherwise almost identical to the I-21a ORF I sequences and clearly distinct from the ORF I sequence of the other $L1_{mlvi2}$ elements. Compare positions 971–989 and 1011–1041 of $L1_{mlvi2\text{-}rn8}$ with the corresponding region of the $L1_4$, $L1_3$, or $L1_{mlvi2\text{-}rn11...rn56}$ sequences in Figure 4. The $L1_{mlvi2\text{-}rn8}$ ORF I sequence actually resembles a hypothetical ancestor of both the I-21a class and the I-21p ORF I sequences in the $L1_{mlvi2\text{-}rn11...rn56}$ elements. We are not saying that the $L1_{mlvi2\text{-}rn8}$ ORF I sequence (or the $L1_{mlvi2\text{-}rn8}$ element) is the ancestor of the aforementioned L1 elements, only that the sequence of the indicated region of the $L1_{mlvi2\text{-}rn8}$ ORF I has the characteristics predicted for a common ancestor of the corresponding region of the $L1_{4/3}$ and $L1_{mlvi2\text{-}rn11...rn56}$ ORF I sequences. Before examining this idea further we wanted to be sure that the putative ancestral type $L1_{mlvi2\text{-}rn8}$ ORF I sequence was part of a modern $L1_{mlvi2}$ element as defined by its 3′ UTR.

The putative ancestral type ORF I sequence of $L1_{mlvi2\text{-}rn8}$ is present on a modern $L1_{mlvi2}$ element: We determined the nature of the ORF I sequence that was colinear with $L1_{mlvi2}$ 3′ UTR in each of the clones as follows. We carried out the PCR on the various $L1_{mlvi2}$ elements with a 3′ primer, L1a2.2, that would prime DNA synthesis only with the $L1_{mlvi2}$ 3′ UTR (see Figure 2) and a 5′ primer cognate to all of the ORF I sequences (ORF I-5, see Figure 4). As a negative and positive control we carried out the PCR on the $L1_{3\text{-}lin3a}$ element with the ORF I primer and either the $L1_{mlvi2}$ specific L1a2.2 PCR primer (negative control) or one that specifically primes the $L1_3$ 3′ UTR, L1RnC3 (see Figure 2). Aliquots of the PCR products were electrophoresed in duplicate on an agarose gel and each half of the blot of this gel was hybridized to [^{32}P]-oligonucleotides specific for either the $L1_{mlvi2\text{-}rn8}$ ORF I (bh18, Figure 4) or for the ORF I of the other $L1_{mlvi2}$ elements (bh19, Figure 4).

The top of Figure 5 shows that, with the exception of $L1_{mlvi2\text{-}rn38}$, the expected ~5.5-kb PCR fragment was produced from each of the $L1_{mliv2}$ clones. The lack of a PCR product from the $L1_{mlvi2\text{-}rn38}$ element is not due to mismatch between it and either of the priming oligonucleotides, ORF I-5 and L1a2.2 (see Figures 2 and 4). We have not investigated the structure of the $L1_{mlvi2\text{-}rn38}$ clone sufficiently to account for the absence of a PCR product from this template. Although the ~5.5-kb PCR fragment was obtained from the $L1_{3\text{-}lin3a}$ element (ab-

FIGURE 4.—Sequence alignment of the 5′ UTR and ORF I sequences of various L1 elements. The first four sequences are those of $L1_4$ and $L1_3$ subfamily members that are present in GenBank. The accession numbers of the these sequences are as follows: X59496, X07686, M13100, and X07687, respectively. The X59496 and M13100 sequences are the only GenBank entries that contained both the diagnostic region of 3′ UTR and the 5′ end of the element. The X07686 and X07687 sequences were from clones originally isolated by us and we determined the DNA sequence of the diagnostic region of the 3′ UTR to assign these elements to their respective subfamilies (results not shown). The remaining sequences were determined in this study and their accession numbers are given in MATERIALS AND METHODS. The beginning of the monomer, tether and ORF I are marked as are several conserved restriction endonuclease sites (*cf.* with Figure 1). The 21-bp insert and the contiguous 66-bp region that is subject to tandem repetition are also boxed. The 66-bp tandem repeats of the $L1_{4\text{-}Lb6}$, $L1_{3\text{-}lin3a}$ and the $L1_{mlvi2\text{-}rn8}$ elements have been removed from the alignment to simplify the presentation. The sequences of various oligonucleotides are either given or indicated with a bracket. The bh18 oligonucleotide is specific for the ancestral version of the I-21p ORF I class ($\text{I-21p}_{anc\text{-}type}$), and the bh19 sequence is specific for the modern version of the I-21p class ($\text{I-21p}_{mod\text{-}type}$, see text for additional details). The dashes indicate identity with the consensus sequence, the dots indicate deletions, and missing data is indicated by spaces.

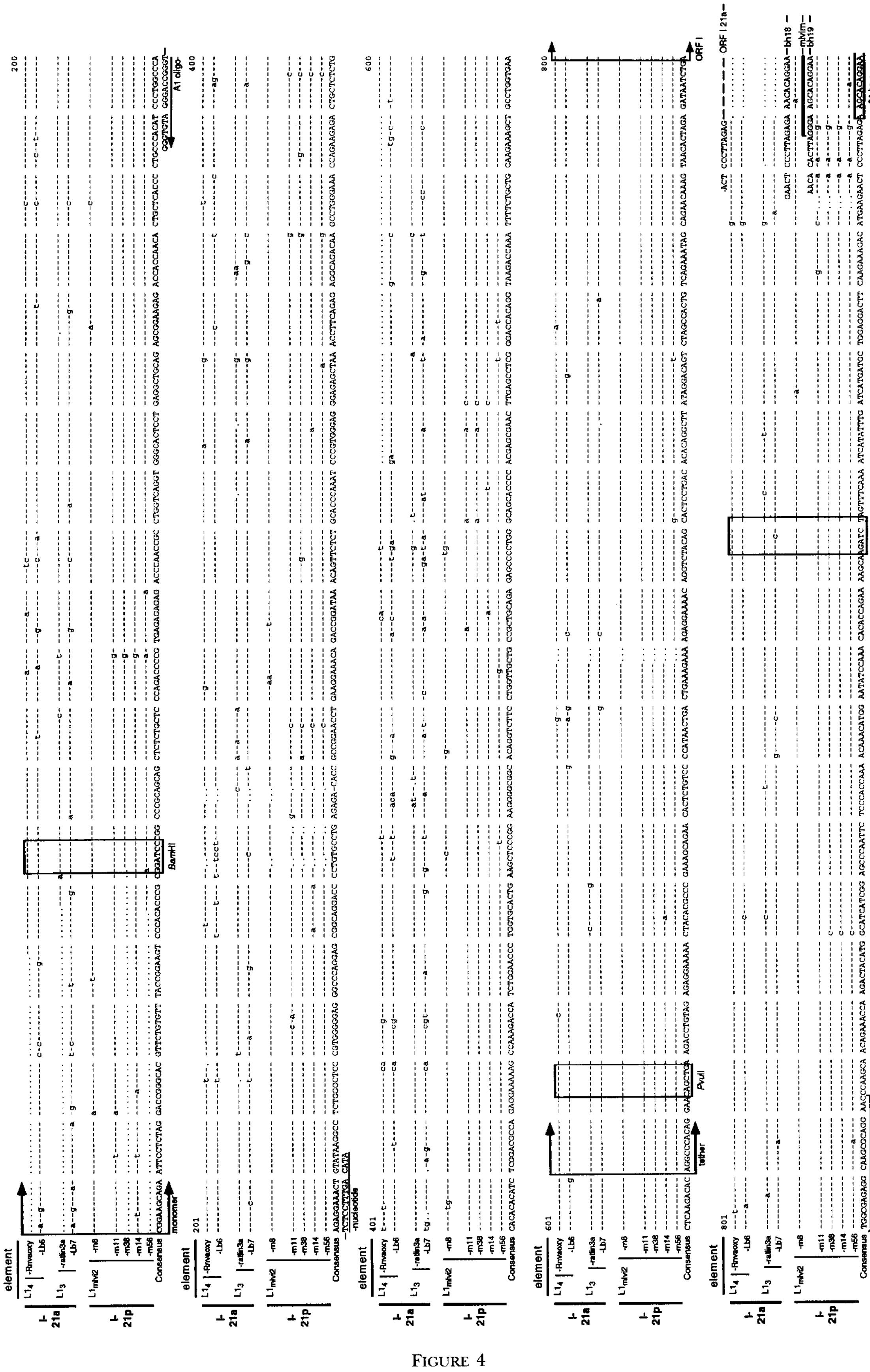


FIGURE 4

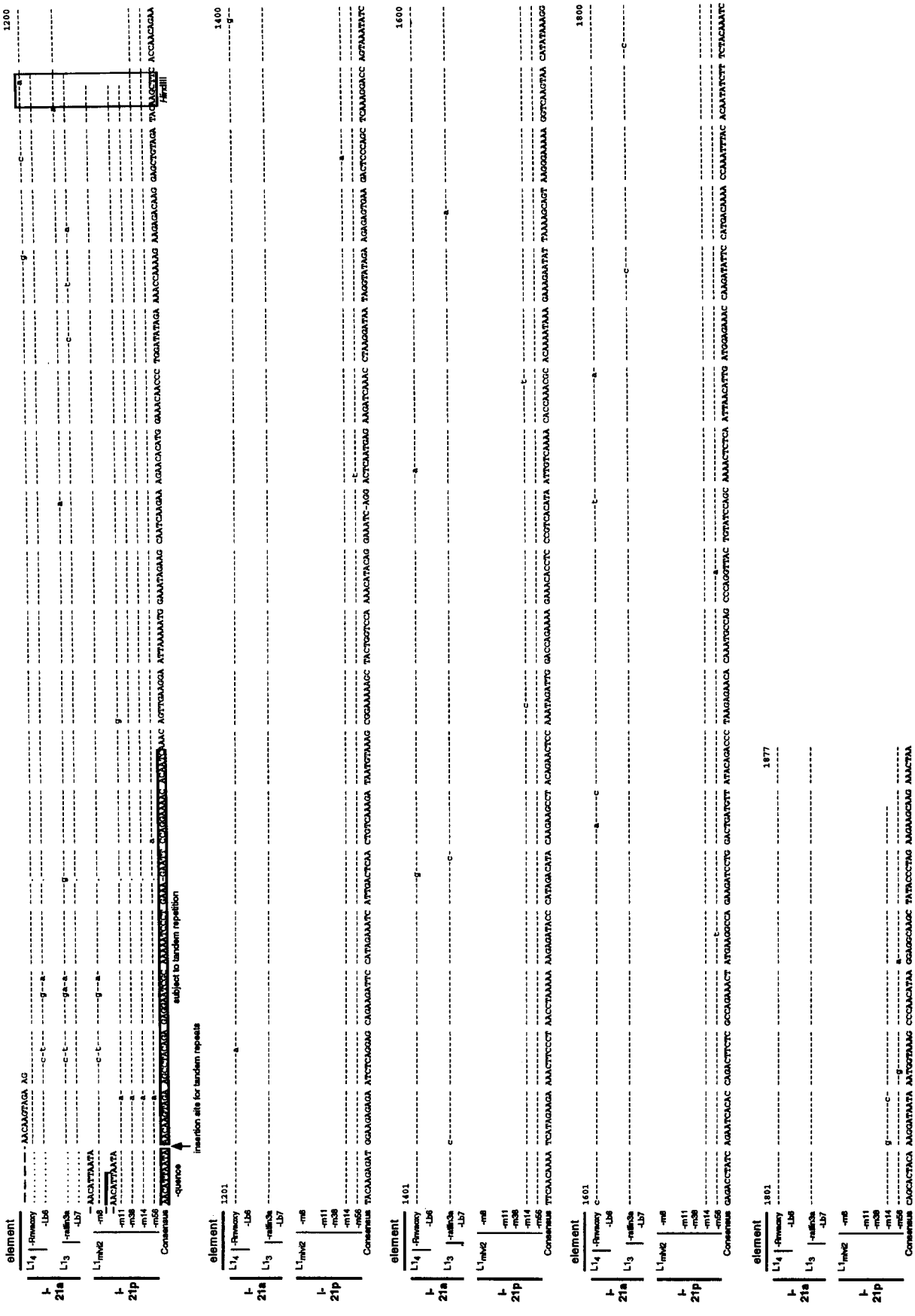


FIGURE 4.—*Continued*

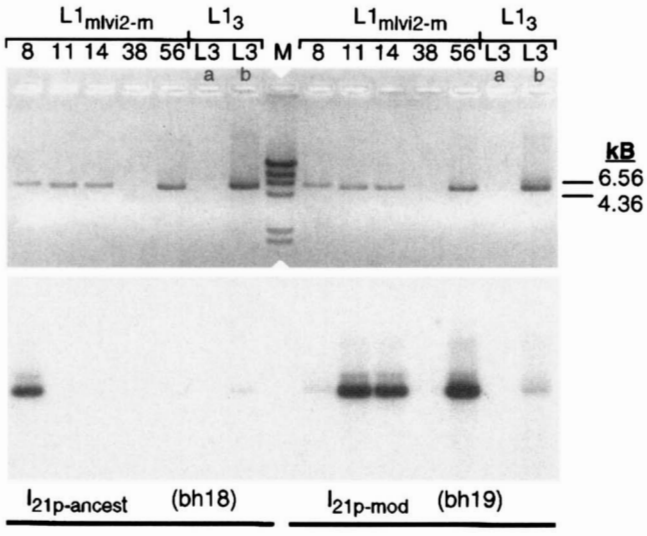


FIGURE 5.—The $L1_{mlvi2}$ ORFs and $L1_{mlvi2}$-specific 3′ UTRs are contained on the same L1 element. The top shows the ethidium bromide-stained gel of the PCR products generated from the indicated L1 clones using as a 3′ PCR primer either the $L1_{mlvi2}$-specific 3′ UTR oligonucleotide, L1a2.2 (lanes 8–56 and L3a), or one specific for the $L1_3$ subfamily, L1RnC3 (lane L3b), and a 5′ primer cognate to the beginning of ORF I, ORF I-5. See Figures 2 and 4 for the location of these primers. The bottom shows an autoradiogram of the blot of each half of the gel hybridized at 62° with the indicated radioactive oligonucleotide. The bh18 ($I\text{-}21p_{anc\text{-}type}$) hybridization was carried out in the presence of a 100-fold molar excess of bh19, and the bh19 ($I\text{-}21p_{mod\text{-}type}$) hybridization with a 100-fold excess of bh18.

breviated L3 in Figure 5) using its cognate 3′ primer (lane L3b, Figure 5), no fragment was obtained from it using the $L1_{mlvi2}$-specific 3′ primer (lane L3a, Figure 5). This requirement for a cognate 3′ primer indicates that each of the PCR products amplified from the $L1_{mlvi2}$ elements terminates in a bona fide $L1_{mlvi2}$ 3′ UTR.

The bottom left of Figure 5 shows that only the $L1_{mlvi2\text{-}rn8}$ PCR product hybridized to the bh18 oligonucleotide, which is cognate to the $L1_{mlvi2\text{-}rn8}$ ORF I sequence (Figure 4). Therefore, the putative ancestral type of ORF I sequence in the $L1_{mlvi2\text{-}rn8}$ clone (Figure 4) and the $L1_{mlvi2\text{-}rn8}$ 3′ UTR sequence (Figure 2) are part of the same L1 element. Figure 5 (bottom right) shows that each of the PCR products from the $L1_{mlvi2\text{-}rn11...rn56}$ clones hybridize specifically to the bh19 oligonucleotide, which is cognate to the ORF I of these $L1_{mlvi2}$ elements.

Therefore, elements with a typical $L1_{mlvi2}$ 3′ UTR can contain two types of the I-21p class of ORF I sequence: the $L1_{mlvi2\text{-}rn8}$ type, which, as mentioned in the previous section, resembles an ancestral type of sequence ($I\text{-}21p_{anc\text{-}type}$); and a second, modern type, $I\text{-}21p_{mod\text{-}type}$, which is found in the other $L1_{mlvi2}$ elements. If the $I\text{-}21p_{anc\text{-}type}$ ORF I sequence in the modern $L1_{mlvi2\text{-}rn8}$ element is actually an ancestral type ORF I sequence, then it should be phylogenetically older than the other ORF I sequences. To determine this, we compared the phylogenetic distribution of the $I\text{-}21p_{anc\text{-}type}$, the $I\text{-}21p_{mod\text{-}type}$, and I-21a sequences.

Phylogenetic distribution of different ORF I sequences: We determined the phylogenetic distribution of the various ORF I sequences by hybridizing DNA from species of Rattus sensu stricto and Rattus sensu lato with one of three oligonucleotides: (1) bh18, specific for the $I\text{-}21p_{anc\text{-}type}$ sequence; (2) mlvi2rn, specific for an *R. norvegicus* version of the $I\text{-}21p_{mod\text{-}type}$ sequence; (3) ORF I-21a, specific for the I-21a class. Various genomic DNAs were digested with *Hin*dIII and *Bgl*II. Conserved sites for these endonucleases flank the 21-bp sequence of ORF I (Figures 1 and 4) in *R. norvegicus* L1 elements. The digested DNA was electrophoresed and then blotted to a nylon membrane that was sequentially hybridized with the above probes. The results are shown in Figure 6.

Control experiments using cloned representatives of the ORF I classes showed that all of the hybridization probes are specific for their cognate ORF I sequences. The $I\text{-}21p_{anc\text{-}type}$ probe, bh18, hybridizes only to the ancestral type I-21p sequence of $L1_{mlvi2\text{-}rn8}$ (Figure 6A, lane rn-8) and to neither the $I\text{-}21p_{mod\text{-}type}$ $L1_{mlvi2}$ elements from *R. norvegicus* (A, lane rn-14) nor to the distinct $I\text{-}21p_{mod\text{-}type}$ $L1_{mlvi2}$ elements from *R.* cf *moluccarius* (A, lane rm-5; CABOT *et al.* 1997). The I-21a probe hybridized to none of the I-21p ORF I sequences (B, lanes rn-8, rm-5, rn-14) but very well to cloned I-21a elements (results not shown). The mlvi2rn probe hybridized only to the $I\text{-}21p_{mod\text{-}type}$ $L1_{mlvi2}$ ORF I found in *R. norvegicus* (C, lane rn-14).

Figure 6A shows that the $I\text{-}21p_{anc\text{-}type}$-specific oligonucleotide, bh18, hybridizes to all of the species of Rattus sensu stricto tested but not to two representatives of Rattus sensu lato (*Leopoldamys sabanus,* lane Ls, *Berylmys bowersi,* lane Bb) nor to *M. domesticus* (lane Md). By contrast, the more modern ORF I classes, I-21a and $I\text{-}21p_{mod\text{-}type}$, were limited to either a subset of Rattus sensu stricto species (B, I-21a) or to just *R. norvegicus* (C, $I\text{-}21p_{mod\text{-}type}$). Therefore, the $I\text{-}21p_{anc\text{-}type}$ sequence in $L1_{mlvi2\text{-}rn8}$ represents a phylogenetically old class of sequences. In fact, in the region cognate to the bh18 oligonucleotide, the ORF-I sequence of $L1_{mlvi2\text{-}rn8}$ may almost be the same as, if not identical to, the ancestral I-21p sequence that was the ultimate progenitor of the present day $I\text{-}21p_{anc\text{-}type}$, $I\text{-}21p_{mod\text{-}type}$, and I-21a sequences. The $I\text{-}21p_{mod\text{-}type}$ element could have evolved from an $I\text{-}21p_{anc\text{-}type}$ element soon after the $L1_{mlvi2}$ family emerged (see DISCUSSION). The fact that the bh18 oligonucleotide does not distinguish the truly ancestral I-21p ORF I sequences present in old L1 families from the $I\text{-}21p_{anc\text{-}type}$ that now resides in modern $L1_{mlvi2}$ elements does not imply that the latter elements are actually ancestral elements, but only that some modern

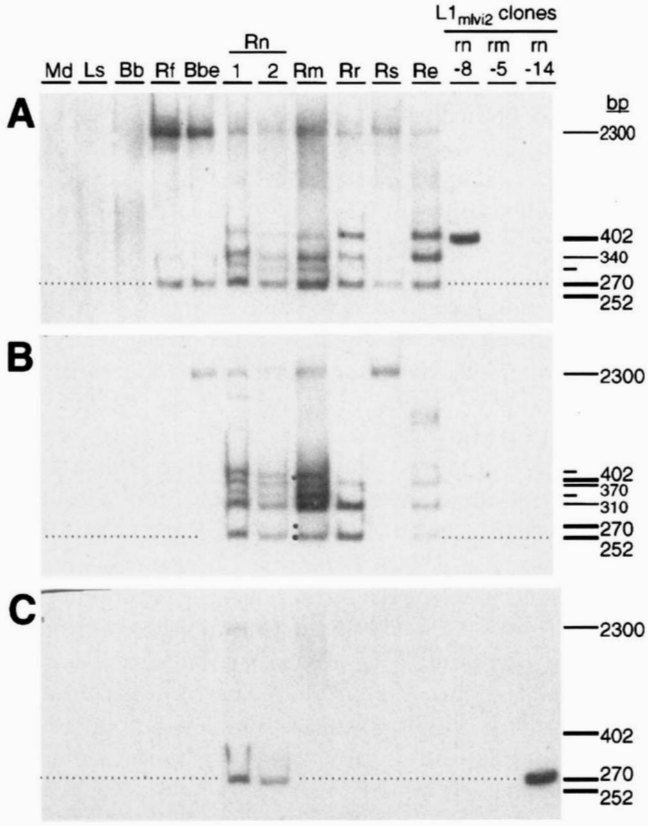


FIGURE 6.—The phylogenetic distribution of various ORF I classes. A blot of the indicated genomic DNAs that had been digested with *Bgl*II and *Hin*dIII were prepared as described in MATERIALS AND METHODS. The blot was sequentially hybridized with the following [^{32}P]-oligonucleotides: (A) bh18 (I-21p$_{anc\text{-}type}$) at 62° in the presence of a 100-fold molar excess each of bh19 and mlvimol, an oligonucleotide cognate to the I-21p$_{mod\text{-}type}$ version of the L1$_{mlvi2}$ ORF I in *R.* cf *moluccarius* (see text); (B) ORF I-21a at 50°; (C) mlvirn (cognate to the I-21p$_{mod\text{-}type}$ version of the L1$_{mlvi2}$ ORF I in *R. norvegicus*, see Figure 4) at 53° in the presence of a 100-fold molar excess each of bh18 and mlvimol (see legend to Figure 5). The heavy tick marks indicate the position and size of *Bgl*II/*Hin*dIII fragments from sequenced L1 clones; the light tick marks indicate the position of the other bands seen in the digests of *R. norvegicus* and *R.* cf *moluccarius* and the assigned sizes were extrapolated from the above fragments and stained DNA size standards.

L1$_{mlvi2}$ elements contain an ancestral-type ORF I sequence.

There is a difference between the size of the *Hin*dIII/*Bgl*II fragment derived from the L1$_{mlvi2\text{-}rn8}$ (Figure 6A, lane rn-8, 402 bp) and L1$_{mlvi2\text{-}rn14}$ (Figure 6C, lane rn-14, 270 bp) elements. This is because the L1$_{mlvi2\text{-}rn8}$ ORF I sequence contains three 66-bp sequences, the basal 66-bp sequences and two additional copies, whereas the L1$_{mlvi2\text{-}rn14}$ element contains only the basal 66-bp sequence (see Figure 1). The pattern of fragments in the digests of genomic DNA indicates that most genomic I-21p$_{anc\text{-}type}$ and truly ancestral I-21p elements (Figure 6A) or I-21a elements (Figure 6B) contain more than one copy of the 66-bp sequence. With the I-21p$_{anc\text{-}type}$ and truly ancestral I-21p elements (Figure 6A) the three major bands seen in most of the digests correspond to elements with no tandem repetition of the 66-bp sequence (270 bp) or with one or two additional copies. For the I-21a elements (Figure 6B) the major bands in *R. norvegicus* and *R.* cf *moluccarius* correspond to one, ~1.5, two and ~2.5 copies. The results also show that tandem repetition is compatible with replicative success. For example, each of the four size classes of tandem repeat-containing I-21a elements in either *R. norvegicus* or *R.* cf *moluccarius* contains ~1700 members (based on a total number of ~8500 I-21a elements in either genome, CABOT *et al.* 1997).

None of the I-21p$_{mod\text{-}type}$ elements in *R. norvegicus* (Figure 6C) contained extra copies of the 66-bp sequence. This is also true of the I-21p$_{mod\text{-}type}$ elements in *R.* cf *moluccarius* (results not shown). Therefore, tandem repetition of the 66-bp sequence appears to be limited to the older L1 families. Restriction site mapping of genomic L1 elements in *R. norvegicus* and *R.* cf *moluccarius* showed that the ~2.3-kb band (size determined on 1.25% agarose gels) is the *Bgl*II fragment derived from elements that lack the *Hin*dIII site (data not shown, see Figure 1).

DISCUSSION

An ancestral ORF I sequence was recruited to generate a modern L1 element: Phylogenetic analysis using a region of the 3′ UTR showed that the L1$_{mlvi2}$ subfamily of *R. norvegicus* is the latest version of a single major lineage of L1 elements that goes back at least 6 my (Figure 2). However, two classes of ORF I are represented in this lineage. The older I-21p ORF I class contains a 21-bp sequence that was deleted in the formation of the younger I-21a ORF I class (Figures 1 and 4). The I-21p class is at least as old as the murine radiation (~12 mya, for references see FURANO *et al.* 1994), since various modern mouse ORF I sequences (SCHICHMAN *et al.* 1992) belong to the I-21p class (alignments not shown). The I-21a class emerged much more recently, during the radiation of Rattus sensu stricto (Figure 6), which began ~2.5 mya (see USDIN *et al.* 1995). The I-21a class is represented in the L1$_4$ and L1$_3$ subfamilies that are antecedent to the L1$_{mlvi2}$ subfamily (Figure 2). The latter subfamily is limited to just two species of Rattus: *R. norvegicus* and *R.* cf *moluccarius* (USDIN *et al.* 1995).

Although the I-21p class is ancestral to the I-21a class, it is the I-21p class that was recruited to form the L1$_{mliv2}$ subfamily. We suggest the I-21a ORF-I class was bypassed by a recombinational event that recruited ancestral I-21p ORF I sequence into the progenitor L1$_{mlvi2}$ element. The alternate explanation would have an I-21p-containing element evolving exactly in parallel with the

I-21a class of elements so as to generate a 3′ UTR of exactly the same lineage. This seems very unlikely to us. The $L1_{mlvi2}$ subfamily now contains two types of I-21p ORF I sequences: an ancestral type, $I\text{-}21p_{anc\text{-}type}$, such as the one in the $L1_{mlvi2\text{-}rn8}$ element, and a modern type, $I\text{-}21p_{mod\text{-}type}$, such as that in the $L1_{mlvi2\text{-}rn11...rn56}$ elements. The particular ORF I sequence in the $L1_{mlvi2\text{-}rn11...rn56}$ elements is limited to *R. norvegicus*. Therefore, these $I\text{-}21p_{mod\text{-}type}$ elements (or more precisely, $I\text{-}21p_{mod\text{-}rn}$ elements) either emerged or amplified in *R. norvegicus* only after this species split from *R.* cf *moluccarius* (Figure 6C).

Another modern version of the I-21p ORF I sequence, $I\text{-}21p_{mod\text{-}rn/rn}$, also emerged. But this occurred before the divergence of *R. norvegicus* and *R.* cf *moluccarius* since $I\text{-}21p_{mod\text{-}rn/rn}$ elements are in both species (CABOT *et al.* 1997). Studies on the parallel evolution of the $L1_{mlvi2}$ subfamily in *R. norvegicus* and *R.* cf *moluccarius* (CABOT *et al.*, in press) suggest that both of the modern type I-21p ORF I sequences (*i.e.*, $I\text{-}21p_{mod\text{-}rn}$ and $I\text{-}21p_{mod\text{-}rn/rn}$) could have been derived from an $I\text{-}21p_{anc\text{-}type}$ element.

Generation of novel L1 elements by recombination: A novel L1 subfamily will emerge from the background of preexisting L1 elements only if it is replicatively successful. However, the generation of a novel, replication competent L1 element by chance recombination between the appropriate combination of biologically intact genomic L1 DNA sequences, though possible, seems quite problematic. This is because almost all of the tens of thousands of genomic L1 copies that have accumulated in mammalian genomes over many millions of years are defective. Most were defective when generated and have accumulated additional genetic defects as they evolved as pseudogenes (see Introduction).

The generation of a novel replicatively successful L1 element by recombination is much more easily envisioned if the recombination occurs during L1 element replication. In this case, the recombination necessarily involves a replication competent element and a novel element would emerge if the replicating L1 element recombines with any sequence that either preserves or enhances replication competence. Recombination could occur during L1 replication if the L1 reverse transcriptase can switch from replicating its own RNA template to copying another RNA. This mechanism implies that the RNA templates be in close proximity, which would be the case if they were packaged into an RNP particle. Particles containing L1 RNA and ORF I protein have been identified in human and mouse cells (MARTIN 1991; KOLOSHA and MARTIN 1995; MARTIN 1995; HOHJOH and SINGER 1996) and a particulate form of a putative L1 reverse transcriptase activity has been described in human teratocarcinoma cells (DERAGON *et al.* 1990). In the strand switching model, the recruitment of ancestral I-21p ORF I sequence would have occurred if an RNA encoding this sequence became copackaged in the RNP of a replicating progenitor of the $L1_{mlvi2}$ element.

Little is known about the requirements, if any, for packaging L1 RNA into RNP particles. It is possible that the packaging process favors full length RNAs from biologically competent L1 elements that contain appropriate packaging signals. If this were true then the recruitment of ancestral I-21p sequence would imply that some L1 families may retain active members long after they first arose. The recent retrotransposition of a putative member of what was thought to be the extinct F L1 subfamily in mouse indicates that old L1 families may retain replicatively competent members (PADGETT *et al.* 1988; ADEY *et al.* 1994b; KINGSMORE *et al.* 1994), even though they may not be as transcriptionally active overall as the newer families (SKOWRONSKI *et al.* 1988; SCHICHMAN *et al.* 1992). However, it is also conceivable that packaging requirements are not stringent and that the ancestral I-21p ORF I sequence could have resided on just a fragment of L1 RNA that was adventitiously transcribed as part of a non-L1 transcriptional unit.

The strand switching mechanism has ample precedent in retrovirus replication where it generates genetic novelty including tandem repeats (HAJJAR and LINIAL 1993; KAMEDA *et al.* 1993; ZHANG and TEMIN 1993). The strand switching mechanism could account for the tandem repeats typical of the 5′ UTR and ORF I sequence of rodent L1 elements if reinitiation on the second template occurs from a "slipped" position, *i.e.*, one that had been previously copied on the first template. (Of course, reinitiation from a slipped position on the same strand could also generate tandem repeats.) Strand switching between different L1 subfamilies could have generated the mosaic 5′ UTR of certain mouse (ADEY *et al.* 1991) or rat L1 elements (CABOT *et al.* 1997). In addition, by analogy to the retroviral packaging of nonretroviral RNAs (*e.g.*, see HAJJAR and LINIAL 1993; ZHANG and TEMIN 1993), non-L1 RNAs might also by chance be trapped in L1 RNP particles. This could account for the keratin-like sequence that comprises the 5′ half of the rabbit L1 ORF I (DEMERS *et al.* 1989) and for the repeated acquisition of novel 5′ UTRs, which characterizes L1 evolution (see Introduction). The self-selection for replicative efficiency that is built into the strand switching model of L1 recombination would seem to guarantee a propensity toward genetic novelty that is a hallmark of L1 evolution.

We thank Dr. ERIC CABOT for providing some of the DNA sequence of the *R. norvegicus* $L1_{mlvi2}$ clones and Mr. BRIAN GOOD for his help in sequencing the 3′ UTR sequence of the $L1_{mlvi2\text{-}rn8...rn56}$ clones. We are especially grateful to Dr. KAREN USDIN for numerous and critical discussions of this work. 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.

LITERATURE CITED

ADEY, N. B., S. A. SCHICHMAN, C. A. HUTCHISON III and M. H. EDGELL, 1991 Composite of A and F-type 5′ terminal sequences defines a subfamily of mouse LINE-1 elements. J. Mol. Biol. **221:** 367–373.

ADEY, N. B., S. A. SCHICHMAN, D. K. GRAHAM, S. N. PETERSON, M. H. EDGELL *et al.*, 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, N. B., T. O. TOLLEFSBOL, A. B. SPARKS, M. H. EDGELL and C. A. I. HUTCHISON, 1994b Molecular resurrection of an extinct ancestral promoter for mouse L1. Proc. Natl. Acad. Sci. USA **91:** 1569–1573.

AUSUBEL, F. M., R. BRENT, R. E. KINGSTON, D. D. MOORE, J. G. SEIDMAN *et al.*, 1989 *Current Protocols in Molecular Biology*. John Wiley & Sons, New York.

BELLIS, M., V. JUBIER-MAURIN, B. DOD, F. VANLERBERGHE, A. M. LAURENT *et al.*, 1987 Distributions of two recently inserted long interspersed elements of the L1 repetitive family at the Alb and beta h3 loci in wild mice populations. Mol. Biol. Evol. **4:** 351–363.

BURTON, F. H., D. D. LOEB, S. F. CHAO, C. A. HUTCHISON III, and M. H. EDGELL, 1985 Transposition of a long member of the L1 major interspersed DNA family into the mouse beta globin gene locus. Nucleic Acids Res. **13:** 5071–5084.

BURTON, F. H., D. D. LOEB, C. F. VOLIVA, S. L. MARTIN, M. H. EDGELL *et al.*, 1986 Conservation throughout mammalia and extensive protein-encoding capacity of the highly repeated DNA long interspersed sequence one. J. Mol. Biol. **187:** 291–304.

CABOT, E., B. ANGELETTI, K. USDIN and A. V. FURANO, 1997 Rapid evolution of a young L1 (LINE 1) clade in recently speciated Rattus taxa. J. Mol. Evol. (in press).

CASAVANT, N. C., and S. C. HARDIES, 1994a The dynamics of murine LINE-1 subfamily amplification. J. Mol. Biol. **241:** 390–397.

CASAVANT, N. C., and S. C. HARDIES, 1994b Shared sequence variants of Mus spretus LINE-1 elements tracing dispersal to within the last 1 million years. Genetics **137:** 565–572.

CASAVANT, N. C., S. C. HARDIES, F. D. FUNK, M. B. COMER, M. H. EDGELL *et al.*, 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., S. D. WAITZKIN, F. R. WITNEY, A. SALEMME, and A. V. FURANO, 1986 Structure of the highly repeated, long interspersed DNA family (LINE or L1Rn) of the rat. Mol. Cell. Biol. **6:** 411–424.

DEMERS, G. W., M. J. MATUNIS, and R. C. HARDISON, 1989 The L1 family of long interspersed repetitive DNA in rabbits: sequence, copy number, conserved open reading frames, and similarity to keratin. J. Mol. Evol. **29:** 3–19.

DERAGON, J. M., D. SINNETT, and D. LABUDA, 1990 Reverse transcriptase activity from human embryonal carcinoma cells NTera2D1. EMBO J. **9:** 3363–3368.

DOMBROSKI, B. A., A. F. SCOTT, and H. H. J. KAZAZIAN, 1993 Two additional potential retrotransposons isolated from a human L1 subfamily that contains an active retrotransposable element. Proc. Natl. Acad. Sci. USA **90:** 6513–6517.

ECONOMOU-PACHNIS, A., M. A. LOHSE, A. V. FURANO, and P. N. TSICHLIS, 1985 Insertion of long interspersed repeated elements at the Igh (immunoglobulin heavy chain) and Mlvi-2 (Moloney leukemia virus integration 2) loci of rats. Proc. Natl. Acad. Sci. USA **82:** 2857–2861.

EICKBUSH, T. H., 1994 Origin and evolutionary relationships of retroelements, pp. 121–157 in *The Evolutionary Biology of Viruses*, edited by S. S. MORSE. Raven Press, New York.

FANNING, T. G., 1983 Size and structure of the highly repetitive BAM HI element in mice. Nucleic Acids Res. **11:** 5073–5091.

FURANO, A. V., and K. USDIN, 1995 DNA fossils and phylogenetic analysis. Using L1 (LINE-1, long interspersed repeated) DNA to determine the evolutionary history of mammals. J. Biol. Chem. **270:** 25301–24304.

FURANO, A. V., C. C. SOMERVILLE, P. N. TSICHLIS, and E. D'AMBROSIO, 1986 Target sites for the transposition of rat long interspersed repeated DNA elements (LINEs) are not random. Nucleic Acids Res. **14:** 3717–3727.

FURANO, A. V., S. M. ROBB, and F. T. ROBB, 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, A. V., B. E. HAYWARD, P. CHEVRET, F. CATZEFLIS, and K. USDIN, 1994 Amplification of the ancient murine Lx family of long interspersed repeated DNA occurred during the murine radiation. J. Mol. Evol. **38:** 18–27.

HAJJAR, A. M., and M. L. LINIAL, 1993 A model system for nonhomologous recombination between retroviral and cellular RNA. J. Virol. **67:** 3845–3853.

HARDIES, S. C., S. L. MARTIN, C. F. VOLIVA, C. A. HUTCHISON III and M. H. EDGELL, 1986 An analysis of replacement and synonymous changes in the rodent L1 repeat family. Mol. Biol. Evol. **3:** 109–125.

HARDISON, R., and W. MILLER, 1993 Use of long sequence alignments to study the evolution and regulation of mammalian globin gene clusters. Mol. Biol. Evol. **10:** 73–102.

HOHJOH, H., and M. F. SINGER, 1996 Cytoplasmic ribonucleoprotein complexes containing human LINE-1 protein and RNA. EMBO J. **15:** 630–639.

HOLMES, S. E., B. A. DOMBROSKI, C. M. KREBS, C. D. BOEHM, and H. H. J. KAZAZIAN, 1994 A new retrotransposable human L1 element from the *LRE2* locus on chromosome 1q produces a chimaeric insertion. Nat. Genet. **7:** 143–148.

HUTCHISON III, C. A., S. C. HARDIES, D. D. LOEB, W. R. SHEHEE and M. H. EDGELL, 1989 LINEs and related retroposons: long interspersed repeated sequences in the eucaryotic genome, pp. 593–617 in *Mobile DNA*, edited by D. E. BERG and M. M. HOWE. American Society for Microbiology, Washington, DC.

JUBIER-MAURIN, V., G. CUNY, A. M. LAURENT, L. PAQUEREAU, and G. ROIZES, 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.

KAMEDA, T., A. AKAHORI, M. H. SONOBE, T. SUZUKI, and T. ENDO *et al.*, 1993 JunD mutants with spontaneously acquired transforming potential have enhanced transactivating activity in combination with Fra-2. Proc. Natl. Acad. Sci. USA **90:** 9369–9373.

KAZAZIAN, H. H. J., C. WONG, H. YOUSSOUFIAN, A. F. SCOTT, D. G. PHILLIPS *et al.*, 1988 Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man. Nature **332:** 164–166.

KINGSMORE, S. F., B. GIROS, D. SUH, M. BIENIARZ, M. G. CARON *et al.*, 1994 Glycine receptor beta-subunit gene mutation in spastic mouse associated with LINE-1 element insertion. Nat. Genet. **7:** 136–141.

KOLOSHA, V. O., and S. L. MARTIN, 1995 Polymorphic sequences encoding the first open reading frame protein from LINE-1 ribonucleoprotein particles. J. Biol. Chem. **270:** 2868–2873.

KOZAK, M. 1987 An analysis of 5′-noncoding sequences from 699 vertebrate messenger RNAs. Nucleic Acids Res. **15:** 8125–8148.

KUROSE, K., K. HATA, M. HATTORI, and Y. SAKAKI, 1995 RNA polymerase III dependence of the human L1 promoter and possible participation of the RNA polymerase II factor YY1 in the RNA polymerase III transcription system. Nucleic Acids Res. **23:** 3704–3709.

LAKSHMIKUMARAN, M. S., E. D'AMBROSIO, L. A. LAIMINS, D. T. LIN, and A. V. FURANO, 1985 Long interspersed repeated DNA (LINE) causes polymorphism at the rat insulin 1 locus. Mol. Cell. Biol. **5:** 2197–2203.

LUAN, D. D., and T. H. EICKBUSH, 1995 RNA template requirements for target DNA-primed reverse transcription by the R2 retrotransposable element. Mol. Cell. Biol. **15:** 3882–3891.

MARTIN, S. L., 1991 Ribonucleoprotein particles with LINE-1 RNA in mouse embryonal carcinoma cells. Mol. Cell. Biol. **11:** 4804–4807.

MARTIN, S. L., 1995 Characterization of a LINE-1 cDNA that originated from RNA present in ribonucleoprotein particles: implications for the structure of an active mouse LINE-1. Gene **153:** 261–266.

MARTIN, S. L., C. F. VOLIVA, S. C. HARDIES, M. H. EDGELL, and C. A. HUTCHISON III, 1985 Tempo and mode of concerted evolution in the L1 repeat family of mice. Mol. Biol. Evol. **2:** 127–140.

MATHIAS, S. L., A. F. SCOTT, H. H. J. KAZAZIAN, J. D. BOEKE, and A.

GABRIEL, 1991 Reverse transcriptase encoded by a human transposable element. Science **254:** 1808–1810.

MIKI, Y., I. NISHISHO, A. HORII, Y. MIYOSHI, J. UTSUNOMIYA *et al.*, 1992 Disruption of the APC gene by a retrotransposal insertion of L1 sequence in a colon cancer. Cancer Res. **52:** 643–645.

MINAKAMI, R., K. KUROSE, K. ETOH, Y. FURUHATA, M. HATTORI *et al.*, 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., E. PASCALE, and A. V. FURANO, 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, R. W., C. A. HUTCHISON III, and M. H. EDGELL, 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., E. VALLE, and A. V. FURANO, 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.

PASCALE, E., C. LIU, E. VALLE, K. USDIN, and A. V. FURANO, 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.

SAITOU, N., and M. NEI, 1987 The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4:** 406–425.

SCHICHMAN, S. A., D. M. SEVERYNSE, M. H. EDGELL, and C. A. I. HUTCHISON, 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, A. F., B. J. SCHMECKPEPER, M. ABDELRAZIK, C. T. COMEY, B. O'HARA *et al.*, 1987 Origin of the human L1 elements: proposed progenitor genes deduced from a consensus DNA sequence. Genomics **1:** 113–125.

SEVERYNSE, D. M., C. A. HUTCHISON III, and M. H. EDGELL, 1992 Identification of transcriptional regulatory activity within the 5′ A-type monomer sequence of the mouse LINE-1 retroposon. Mamm. Genome **2:** 41–50.

SINGER, M. F., 1995 Unusual reverse transcriptases. J. Biol. Chem. **270:** 24623–24626.

SKOWRONSKI, J., T. G. FANNING, and M. F. SINGER, 1988 Unit-length line-1 transcripts in human teratocarcinoma cells. Mol. Cell. Biol. **8:** 1385–1397.

SMIT, A. F. A., G. TÓTH, A. D. RIGGS, and J. JURKA, 1995 Ancestral, mammalian-wide subfamilies of LINE-1 repetitive sequences. J. Mol. Biol. **246:** 401–417.

SOARES, M. B., E. SCHON, and A. EFSTRATIADIS, 1985 Rat LINE1: the origin and evolution of a family of long interspersed middle repetitive DNA elements. J. Mol. Evol. **22:** 117–33.

SWERGOLD, G. D., 1990 Identification, characterization, and cell specificity of a human LINE-1 promoter. Mol. Cell. Biol. **10:** 6718–6729.

SWOFFORD, D. L., 1993 PAUP: Phylogenetic Analysis Using Parsimony, Version 3.1.1. Computer Program distributed by the Illinois Natural History Survey. Champaign, IL.

USDIN, K., and A. V. FURANO, 1988 Rat L (long interspersed repeated DNA) elements contain guanine-rich homopurine sequences that induce unpairing of contiguous duplex DNA. Proc. Natl. Acad. Sci. USA **85:** 4416–4420.

USDIN, K., and A. V. FURANO, 1989 The structure of the guanine-rich polypurine:polypyrimidine sequence at the right end of the rat L1 (LINE) element. J. Biol. Chem. **264:** 15681–15687.

USDIN, K., P. CHEVRET, F. M. CATZEFLIS, R. VERONA, and A. V. FURANO, 1995 L1 (LINE-1) retrotransposable elements provide a fossil record of the phylogenetic history of murid rodents. Mol. Biol. Evol. **12:** 73–82.

VANLERBERGHE, F., F. BONHOMME, C. HUTCHISON III, and M. H. EDGELL, 1993 A major difference between the divergence patterns within the lines-1 families in mice and voles. Mol. Biol. Evol. **10:** 719–731.

VERNEAU, O., F. M. CATZEFLIS and A. V. FURANO, 1997 Determination of the evolutionary relationships in Rattus sensulato (Rodentia: Muridae) using L1 (LINE 1) amplification events. J. Mol. Evol. (in press).

VOLIVA, C. F., C. L. JAHN, M. B. COMER, C. A. HUTCHISON III, and M. H. EDGELL, 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, C. F., S. L. MARTIN, C. A. HUTCHISON III, and M. H. EDGELL, 1984 Dispersal process associated with the L1 family of interspersed repetitive DNA sequences. J. Mol. Biol. **178:** 795–813.

WINCKER, P., V. JUBIER-MAURIN, and G. ROIZES, 1987 Unrelated sequences at the 5′ end of mouse LINE-1 repeated elements define two distinct subfamilies. Nucleic Acids Res. **15:** 8593–8606.

ZHANG, J., and H. M. TEMIN, 1993 Rate and mechanism of nonhomologous recombination during a single cycle of retroviral replication. Science **259:** 234–238.

Communicating editor: W-H. LI