J Mol Evol (1993) 36:9–20

Journal of
Molecular Evolution

The Evolution of Long Interspersed Repeated DNA (L1, LINE 1) as Revealed by the Analysis of an Ancient Rodent L1 DNA Family

Esterina Pascale,* Christine Liu,† Eulalia Valle,‡ Karen Usdin, and Anthony V. Furano

Section on Genomic Structure and Function, Laboratory of Biochemical Pharmacology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Building 8, Room 203, Bethesda, MD 20892, USA

Summary. All modern mammals contain a distinctive, highly repeated (≥50,000 members) family of long interspersed repeated DNA called the L1 (LINE 1) family. While the modern L1 families were derived from a common ancestor that predated the mammalian radiation ~80 million years ago, most of the members of these families were generated within the last 5 million years. However, recently we demonstrated that modern murine (Old World rats and mice) genomes share an older long interspersed repeated DNA family that we called Lx. Here we report our analysis of the DNA sequence of Lx family members and the relationship of this family to the modern L1 families in mouse and rat. The extent of DNA sequence divergence between Lx members indicates that the Lx amplification occurred about 12 million years ago, around the time of the murine radiation. Parsimony analysis revealed that Lx elements were ancestral to both the modern rat and mouse L1 families. However, we found that few if any of the evolutionary intermediates between the Lx and the modern L1 families were extensively amplified. Because the modern L1 families have evolved under selective pressure, the evolutionary intermediates must have been capable of replication. Therefore, replication-competent L1 elements can reside in genomes without undergoing extensive amplification. We discuss the bearing of our findings on the evolution of L1 DNA elements and the mammalian genome.

Key words: L1 family evolution — LINE family — Long interspersed repeated DNA — Ancient L1 family — L1 family amplification

Introduction

As much as 10% of the genomes of all mammalian genera studied to date consists of 50,000–100,000 dispersed copies of long interspersed repeated DNA (L1, LINE 1) elements (Rogers 1985; Burton et al. 1986; Weiner et al. 1986). Unlike other highly repeated mammalian DNA elements, L1 is apparently self-replicating. The element contains both a transcriptional regulatory sequence and two protein encoding sequences, ORF I and ORF II, that have evolved under strong selective pressure, indicating that the putative ORF I and ORF II products are required for L1 activity (Martin et al. 1984; Hattori et al. 1985; D'Ambrosio et al. 1986; Hardies et al. 1986; Loeb et al. 1986; Scott et al. 1987; Nur et al. 1988; Swergold 1990; Severynse et al. 1992).

Although the structure of ORF I does not suggest its possible function, ORF II encodes a protein that is homologous to known reverse transcriptases (Hattori et al. 1986; Loeb et al. 1986; Xiong and Eickbush 1988), and the human ORF II product has been shown to function as one (Mathias et al. 1991). Therefore, it is quite likely that L1 replicates via an RNA intermediate as was first suggested by Fanning (1983) and Voliva et al. (1983). These and other studies since then have shown that the highly am-

* *Present address:* Cattedra di Chimica e Microscopia Clinica, Universita degli Studi, L'Aquila, Italy
† *Present address:* Columbia University, New York, NY 10027, USA
‡ *Present address:* Department of Functional Biology, Faculty of Medicine, c/Julian Claveria S/N, 33071, Oviedo, Spain
Offprint requests to: A.V. Furano

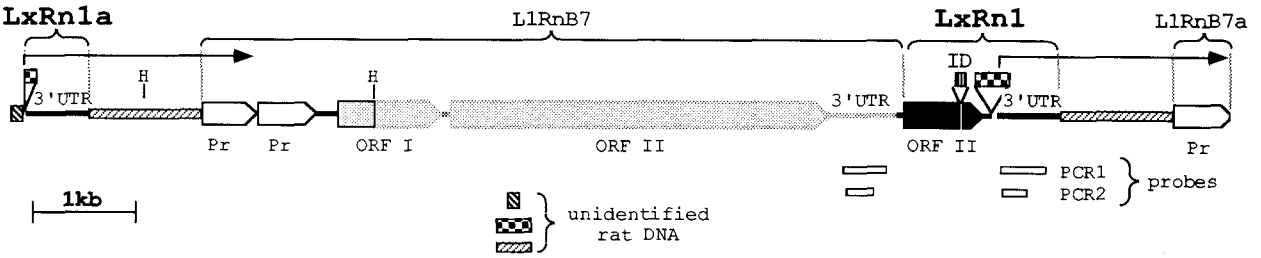

Fig. 1. Diagram of an L1Rn- and LxRn-containing rat genomic clone, γB7. The *solid lines* and *boxes* indicate the regions that were sequenced. The structure of the *stippled region* was determined by blot hybridization of restriction endonucleolytic fragments of the clone with various L1Rn probes. *Pr* designates the tandem repeated regulatory regions of L1RnB7. *ORF I* or *ORF II* and *3′ UTR* indicate the open reading frames and the untranslated region 3′ of ORF II, respectively. The two *H*s demarcate a *Hind*III fragment, the sequence of which was previously published [GENBANK locus RatLb7, accession number, X07687 (Furano et al. 1988)] *ID* indicates a member of the rat ID SINE (short interspersed repeated DNA) family (Sutcliffe et al. 1982). The sequence between the 3′ end of LxRn1 and the 5′ end of L1RnB7a (*crosshatched region*) is not highly repeated (results not shown). Hybridization of this sequence to nitrocellulose blots of genomic DNA that had been digested with various restriction endonucleases indicated that the tandem repeated L1Rn elements (L1RnB7 and L1RnB7a) are not part of an extensive tandem array (results not shown). Most of the DNA sequence at the left end of the diagram (includes LxRn1a and part of the Pr region of L1RnB7, see *bent arrow*) is almost identical to that at the right end of the clone (see *bent arrow*). The region of L1 or Lx DNA corresponding to the *PCR 1* and *PCR 2* probes is shown by the *open rectangles* below the diagram.

plified modern L1 families consist mostly of truncated or otherwise defective members (i.e., pseudogenes) and presumably just a few functional, or active, members (Martin et al. 1985; Hardies et al. 1986; Scott et al. 1987; Casavant et al. 1988). Therefore, as is the case with other elements that replicate using an RNA intermediate (Pathak and Temin 1990, for example, and references therein), L1 replication is apparently quite error prone. In addition to generating large numbers of defective L1 members, this process could also produce a relatively rapid rate of evolutionary change in L1 elements (Hardies et al. 1986).

The evolution of L1 families presents an apparent enigma. Although L1 DNA existed in the mammalian lineage before the mammalian radiation ~80 million years ago (Burton et al. 1986), the L1 families in modern mammals appear to be of very recent origin. Each is now quite distinct and seems to consist mainly of members which were inserted in the genome within just the last 5 million years (Martin et al. 1985; Hardies et al. 1986; Scott et al. 1987; Casavant et al. 1988). Two ideas were proposed to account for these results: One was that as new functional L1 variants evolved and amplified themselves, the members of the older L1 families were lost from the genome (Martin et al. 1985; Hardies et al. 1986; Scott et al. 1987). The second was that the extensive amplification that typifies the replicative process of modern L1 families is a relatively recent and singular event in L1 evolution (Deininger and Daniels 1986; Weiner et al. 1986).

However, we recently showed that an extensive amplification of an ancient L1 family, which we call Lx, occurred in the progenitor of modern murine rodents [e.g., rats (*Rattus*), mice (*Mus*), etc.] and that relics of this amplification have persisted in these species (Pascale et al. 1990). Here we determined and compared the DNA sequence of a number of Lx elements. The extent of nucleotide substitutions between them indicates that the amplification of Lx occurred ~12 million years ago, which is about the time of the murine radiation (Catzeflis et al. 1987). Parsimony analysis showed that both the modern rat L1 family in *R. norvegicus* (L1Rn) and the modern mouse L1 family in *M. domesticus* (L1Md) originated from the Lx family. However, few if any of the evolutionary intermediates between the Lx and either the rat or mouse L1 family were extensively amplified. The implications of these results for the evolution of these elements and mammalian genomes are discussed.

Materials and Methods

Nomenclature. Analogous to the use of L1, L2, . . . etc., for the naming of the modern long interspersed repeated DNA families (Voliva et al. 1983), we propose the use of Lx, Ly, . . . etc., for naming ancestral or ancient long interspersed repeated DNA families. Therefore, just as L1Rn, L1Md, and L1AF refer to the long interspersed repeated DNA families in *R. norvegicus, M. domesticus,* and *Apodemus flavicollis,* respectively, LxRn, LxMd, and LxAf would refer to Lx family members in these species. An important point here is that although the L1 families in different murine genera are now distinct, the Lx family is not. In addition, L alone (or LINE alone) could be used when not distinguishing between modern and ancestral families.

Isolation of the λB7 Genomic Clone. The λB7 clone shown in Fig. 1 was isolated from a rat genomic library [prepared by Dr. T. Sargent (N.I.H.) from a partial *Hae*III digest of rat DNA] by sequentially hybridizing nitrocellulose blots of phage plaques with oligonucleotide probes specific for the right or left ends of L1Rn. This particular clone contains both a full-length element

(L1RnB7) and the 5′ region of another L1 (L1RnB7a). Various restriction fragments of this clone were then subcloned into plasmid or M13 vectors using standard techniques (Ausubel et al. 1989). DNA sequence determination (see below and Results) showed that this clone contained Lx DNA in addition to L1 sequences.

Hybridization Probes. Figure 1 shows the location of the hybridization probes PCR1 and PCR2. These probes contained only DNA from the 3′ untranslated region (3′ UTR) since preliminary experiments showed that only 3′ UTR DNA was specific for L1 or Lx. The appropriate region of L1 DNA was amplified from the previously described L1Rn4 clone, LINE 4 (D'Ambrosio et al. 1986), and the corresponding region of Lx DNA was amplified from LxRn1 by use of the polymerase chain reaction (PCR) (Saiki et al. 1985). The oligonucleotides used were as follows: L1 PCR1, AACAGAGACTGAAGGAA and AGTCTAGTTCACTGGGG; L1 PCR2, ATA**GGATCC**GCCCCACAGGTGGCCCAT and ATA**GGATCC**AGTGGCTTAGTCCCTGGA; Lx PCR1, AACAGAGACTGAAGGAA and TGCAATTCCCTACACTG; Lx1 PCR2 ATA**GGATCC**CATCCAGAGACTACCTCACCT and ATA**GGATCC**TCTTCCTATGGGGTTGAAAAC. The appropriate region of mouse L1 DNA was amplified by PCR from mouse genomic DNA using the following oligonucleotides: ATA**GGATTC**ATACACTAGCAAGATTTT and ATA**GGATTC**GTCAAGAGCTCCGGGGTA. The product amplified with these primers is about the same length and represents the same region of the mouse L1 family as the L1 PCR2 probe. (See Fig. 1.) The first nine nucleotides of the PCR2 and the mouse L1 primers are not complementary to either L1 or Lx and contain the bases ATA followed by a *Bam*HI site.

PCR reactions were carried out for 25 or 30 cycles using Taq DNA polymerase (Perkin Elmer Cetus) in 100-μl reactions using the conditions suggested by supplier. The PCR-amplified DNA was then purified and radiolabeled by the random priming method with [^{32}P]dCTP. (See below.)

Preparation of Genomic Libraries. Genomic clones of *M domesticus* and *R. norvegicus* were prepared by ligating a partially filled (using dGTP and dATP only) *Sau*3A partial digest of mouse or rat DNA to partially filled (using dTTP and dCTP only) *Xho*I sites of a λ cloning vector (Zabarovsky and Allikmets 1986). The vector, lambdaGEM™-11, and λ packaging mixture were purchased from Promega and the recombinant phages were propagated in *E. coli* MB406, which is an *mcr*A1, *mcr*B1 strain and therefore does not discriminate against methylated mammalian DNA (Woodcock et al. 1988).

Preparation of Lx Clones from Various Murine Species. Since we wished to clone and determine the DNA sequence of the same discrete region of Lx from several murine species, we elected to use the PCR to amplify the desired region of Lx from genomic DNA. However, this method could produce chimeric sequences from a divergent repeated DNA family in either of two ways: First, incompletely extended products synthesized from one family member could prime DNA synthesis on a different family member in a subsequent round of DNA synthesis. This has been found to occur at a frequency of about 1% (Jansen and Ledley 1990). Therefore, we did not think that this would be a serious problem, and our control experiment (see below) supported this supposition. A second and far more likely source of chimeras would result from the cloning of hybrid DNA duplexes (Jansen and Ledley 1990). This would occur if the amplified products of different family members annealed to each other rather than serving as templates for primer extension. To eliminate this problem we decided to clone single strands of the amplified Lx DNA by first transcribing the PCR products and then making cDNA clones of the transcripts. To do this we modified the Lx1 PCR2 primers as follows: The first nine bases of the first PCR2 primer were replaced with a strong T7 promoter followed by an *Eco*RI site to give CGAAATTAATACGACTCACTATAGGG**AGAATTC**CATCCAGAGACTACCTCACCT (transcription starts at base 24). The first nine bases of the second PCR2 primer were replaced with nine bases containing a *Sal*I site to give ATA**GTCGAC**TCTTCCTATGGGGTTGAAAAC.

After the final round of the PCR reaction the products were purified and used for RNA synthesis using the T7 RNA polymerase and 1 unit RNAse–Block II (Stratagene) in a 25-μl reaction using the conditions suggested by the supplier. The DNA template was destroyed by DNAse and the RNA product was purified and copied into DNA using the above-mentioned version of the second PCR2 primer with an RNAse H$^-$ Moloney murine virus reverse transcriptase (M-MLV H$^-$ RT [Superscript]™, GIBCO BRL) using the conditions suggested by the supplier. The RNA template was destroyed by alkali (20 mM NaEDTA, 50 mM NaOH, 60 min at 70°C) and the complementary strand of DNA was synthesized using Sequenase 2.0 (United States Biochemicals) (Tabor and Richardson 1989) and the primer GGA**GAATTC**CATCCAGAGACT. The DNA was digested with *Eco*RI and *Sal*I, purified, and ligated to the *Eco*RI and *Sal*I sites of pUC19. Isolation of plasmid clones was carried out using standard techniques (Ausubel et al. 1989).

To test the efficacy of this method we determined the DNA sequence of seven clones that were derived from the amplification of a 1:1 mixture of two rat Lx sequences, LxRn1 and LxRn3. (See Results.) LxRn1 and LxRn3 differ by about 20%; the DNA sequence of four of the clones was identical to LxRn3 and the sequences of the other three were identical to LxRn1 (results not shown).

DNA Sequence Determination and Analysis. The DNA sequences (both strands) were determined by the dideoxy chain termination method (Sanger et al. 1977), using the modified T7 DNA polymerase (Tabor and Richardson 1987, 1989) supplied as Sequenase™ 1.0 and 2.0, respectively, by United States Biochemicals. General DNA sequence manipulation and database searches were performed using the programs of the Genetic Computer Group (GCG) originally developed at the University of Wisconsin (Devereux et al. 1984). We are indebted to the National Cancer Institute for computing resources and staff support at the Advanced Scientific Computing Laboratory of the Frederick Cancer Research and Development Center. The multiple sequence alignments used to generate the data in Table 2 and Fig. 4 are available from the authors. Parsimony analysis was carried out using the PAUP program of Swofford (1990). We used a step matrix to differentially weight transitions and transversions based on the nucleotide substitution data of Gojobori et al. (1982) and Li et al. (1984). We employed the heuristic search method and used the suggested (Swofford 1990) two-step procedure to generate trees: First, the sequences were added in the order in which they appeared in the alignment (simple addition). Then we carried out ten trials of tree building using random addition, saving only those trees of length no greater than one plus the length of the shortest tree obtained using simple addition. Finally, a strict consensus tree was produced from the total of 36 equally parsimonious trees that were generated.

Miscellaneous Procedures. Restriction enzyme digestions, gel electrophoresis, nitrocellulose blots, preparation of [^{32}P]DNA hybridization probes, nucleic acid purification, and

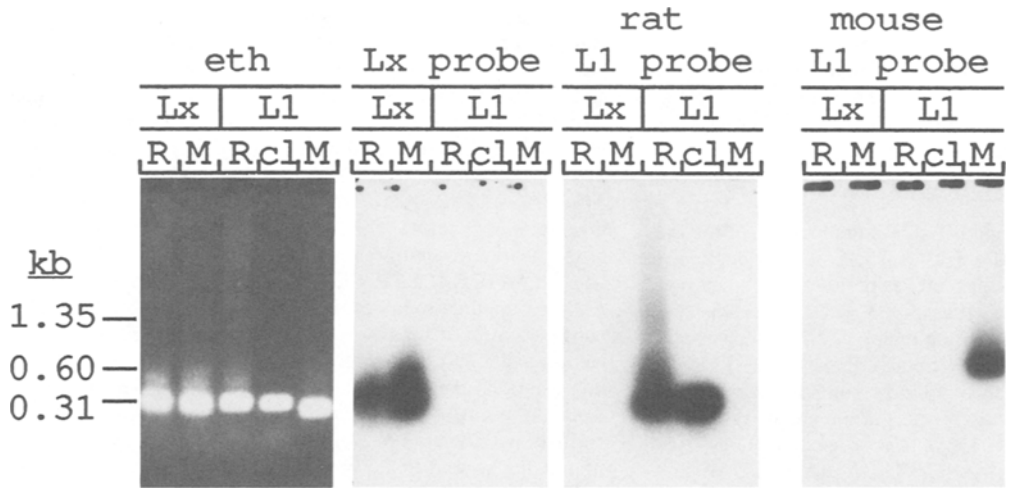

Fig. 2. Hybridization of Lx and L1 sequences. Lx or L1 sequences were amplified from rat (*R*) or mouse (*M*) genomic DNA, or from an L1Rn clone (lane *cl*) by the PCR using the cognate PCR2 primers. (See Fig. 1.) A portion of the products was subjected to electrophoresis on an agarose gel, blotted to nitrocellulose, and hybridized at 65°C as described in Materials and Methods to the PCR1 Lx probe or the PCR1 rat L1 probe. (See Fig. 1.) The mouse L1 probe is from a region of L1Md that corresponds to the L1 or Lx PCR2 probes. (See Fig. 1 and Materials and Methods.) The panel labeled *eth* shows the ethidium-bromide-stained gel.

the various steps in the RNA mad cDNA snythesis described above were carried out using standard techniques (Ausubel et al. 1989). Hybridization reactions were incubated at 55°C or 65°C in a solution containing 0.2 M sodium phosphate buffer, pH 6.8; 0.125% (w/v) sodium dodecylsulfate; 0.5 mM NaEDTA; 50 mg/ml Ficoll 400 (Pharmacia); 50 mg/ml polyvinylpyrrolidone; 50 μg/ml denatured salmon sperm DNA. The filters were washed at room temperature twice in 0.3 m NaCl, 0.03 M Na citrate (pH 7.4) , 0.1% sodium dodecylsulfate, followed by two washes with 3 mM Tris base.

GENBANK Accession Numbers. All of the DNA sequences have been deposited in GENBANK and they have the following accession numbers: LB7 LF (contains the DNA sequence, including LxRn1a, that flanks the left end of L1RnB7, see Fig. 1), M60810; LB7 RF (contains the DNA sequence including LxRn1 but not L1RnB7a, that flanks the right end of L1RnB7, see Fig. 1), M60811; R3, M60824 (contains LxRn3); LxRn5a, M60821; LxRnx2, M60822; LxRnx3, M60825; LxMdx3, M60814; LxMdx4, M60815; LxMsx1, M60819; LxMsx3, M60820; LxMpx1, M60817; LxMpx4, M60818; LxMmx1, M60823; LxMmx3, M60816; LxCpx2, M60812; and LxCpx3, M60813.

Results

Discovery of Lx

We discovered Lx by chance while examining the DNA sequence between two closely spaced L1Rn elements. The DNA sequence analysis showed that L1RnB7 had inserted directly into an L1-like sequence which we call LxRn1 (Fig. 1). LxRn1 also contains two other insertions. The first is a member of the highly repeated rat ID short interspersed repeated DNA (SINE) family (Sutcliffe et al. 1982), and the second is a sequence of unknown origin.

Preparation of Lx- or L1-Specific Hybridization Probes

To further examine the Lx family against the background of the modern L1 families we prepared Lx- or L1-specific hybridization probes. These were generated by the polymerase chain reaction (PCR) from the 3′ UTR of Lx or L1 DNA since only this region is sufficiently divergent to be distinguished by hybridization. (See Materials and Methods.) Figure 2 shows that the Lx and L1 hybridization probes were specific for the families from which they were derived: Panel 2 (second from left) shows that the Lx probe did not hybridize with either rat (R) or mouse (M) genomic L1 sequences but hybridized equally well to rat or mouse genomic Lx sequences. Panel 3 shown that the rat L1 probe did not hybridize to either rat or mouse genomic Lx sequences or to genomic L1Md sequences. Panel 4 shows that the mouse L1 probe, which was a mixture of L1Md sequences amplified from mouse DNA by the PCR, did not hybridize to Lx or rat L1 sequences. In an experiment not shown here we found that a similarly prepared L1Rn probe did not hybridize to genomic Lx sequences.

The results in Fig. 2 also reveal that neither the rat nor mouse Lx or L1 family contains a large number of sequences that are intermediate between Lx and either modern L1 sequence. The 3′ UTR of Lx DNA is about 65% similar to either modern L1 sequence (Pascale et al. 1990, and results not shown) and since DNA sequences that are about 80% similar will hybridize (e.g., between Lx sequences, see Table 2), sequences intermediate between Lx and

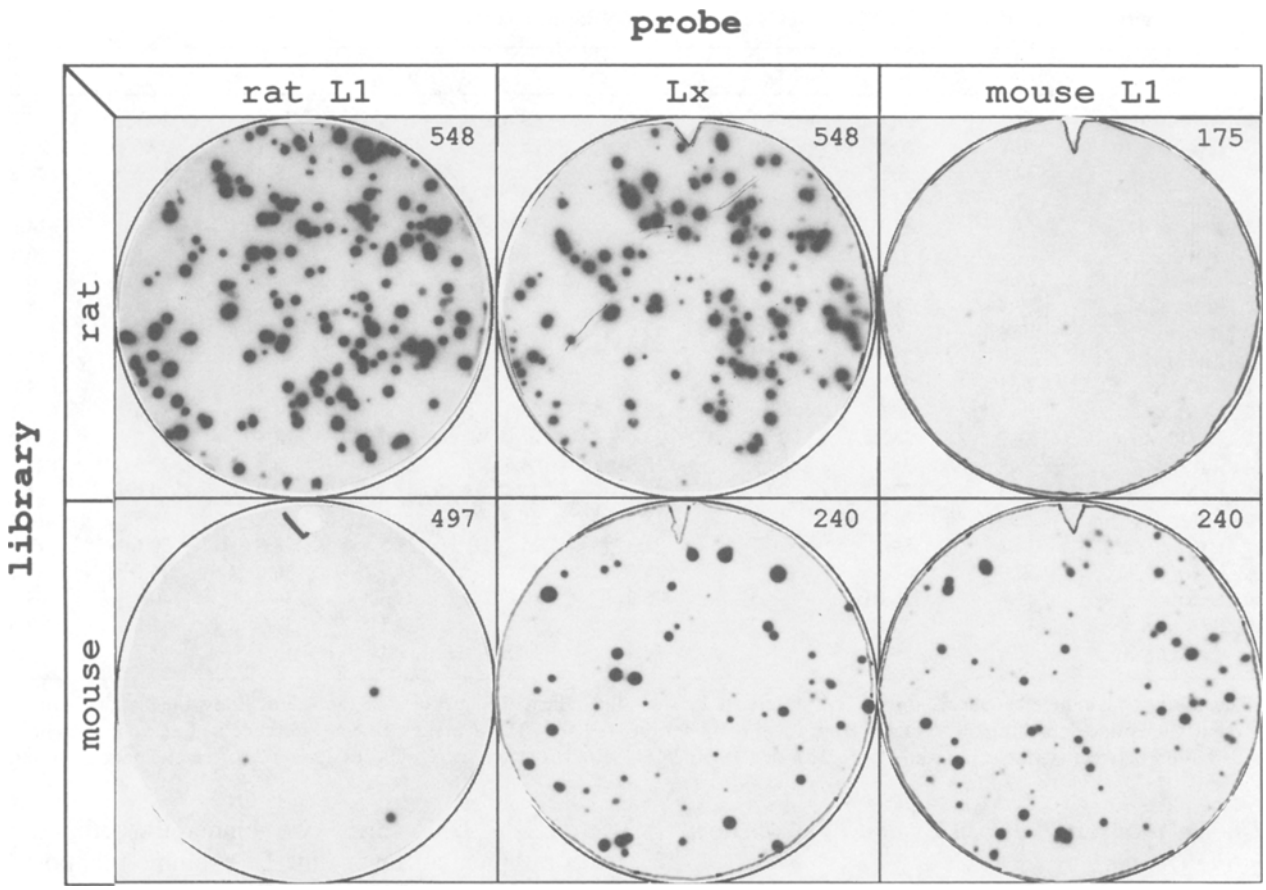

Fig. 3. Hybridization of *Lx* or *L1* sequences to genomic libraries of rat or mouse DNA. Nitrocellulose blots of λ phage plaques of rat or mouse genomic libraries, prepared as described in Materials and Methods, were hybridized at 65°C with the Lx PCR1 probe, the rat L1 PCR1 probe, or the mouse L1 probe (see Fig. 1 and the Materials and Methods). The number in the upper right-hand corner of each panel is the total number of plaques on the plate.

Table 1. The number of L1- or Lx-positive recombinant clones in rat or mouse genomic libraries

	Positive clones				
	Total		Observed (expected)[a]		
Animal	L1	Lx	L1 only	Lx only	L1 + Lx
Rat					
Number	180	139	98 (136)	57 (92)	82 (46)
% Of total clones	33	25			
Genomic copies[b] × 10^{-3}	74	57			
Mouse					
Number	97	69	58 (68)	30 (42)	39 (28)
% Of total clones	40	29			
Genomic copies[b] × 10^{-3}	91	65			

[a] The values expected are given in the parentheses and were calculated from the percentage of L1- or Lx-positive plaques assuming that the distribution of L1 or Lx sequences in the genome were independent of each other

[b] This value equals (% positive clones) (2,700,000 kb/haploid genome ÷ 12 kb/clone)

either of the modern L1 DNA sequences would have been detected by the experiment shown in Fig. 2. These putative sequence intermediates should have been amplified by the PCR since target sequences with only 50% complementarity to a primer will be efficiently amplified providing there is no mismatch with the three 3′ nucleotides of the primer (Sommer and Tautz 1989).

Table 2. Pairwise comparison between the 3′ untranslated region of Lx members[a]

	1	2	3	4	5	6	7	8	9
1. LxMmx3	****	24.5	18.1	19.1	19.6	20.6	21.7	24.1	15.4
2. LxMpx4	30.2	****	25.0	27.0	25.5	25.3	26.1	29.6	21.5
3. LxRn3	20.9	30.9	****	15.0	15.5	10.9	18.4	15.5	13.8
4. LxRn1	22.1	33.5	16.7	****	19.8	15.9	22.3	22.3	16.7
5. Musonctlm	23.2	31.9	17.5	23.3	****	15.9	19.4	19.9	14.3
6. Ratil6g1	24.7	31.6	11.9	18.0	18.5	****	18.9	16.9	16.1
7. Musigkfi2	25.9	32.8	21.3	26.6	23.0	22.1	****	20.4	18.3
8. Ratcyp2a1	29.4	38.0	17.4	26.5	23.3	19.2	24.1	****	18.8
9. LxMsx1	17.4	25.7	15.3	19.0	16.1	18.4	21.2	21.7	****
10. LxMdx3	25.7	31.2	25.3	27.4	30.1	29.2	32.4	35.2	23.0
11. LxCpx2	26.3	35.6	20.9	21.0	32.4	26.5	30.6	30.9	24.8
12. LxRn5a	28.7	36.2	26.0	28.7	26.3	29.5	25.5	32.7	23.1
13. LxMdx4	27.7	36.1	25.0	25.0	27.0	28.0	28.4	33.9	26.2
14. Rattran	25.3	38.9	22.6	18.6	24.3	27.2	30.3	32.1	23.3
15. LxMsx3	25.3	32.6	20.7	22.1	24.4	25.5	30.8	29.0	21.2
16. LxMpx1	22.0	31.0	17.5	18.8	21.9	20.0	25.4	28.7	18.5
17. LxCpx3	24.2	34.3	20.7	22.0	25.2	23.5	28.0	30.6	17.7
18. LxRnx2	32.0	42.4	28.3	26.7	33.7	28.6	36.6	36.0	30.1
19. LxRnx3	43.6	56.8	38.5	36.9	40.7	37.6	44.4	40.4	43.4
20. LxMmx1	34.5	49.8	36.0	37.2	38.8	37.5	37.7	42.8	39.5
21. Consensus	11.7	20.9	8.2	10.4	12.9	11.4	15.2	17.5	7.8

[a] The values above the diagonal are the percent difference between the indicated sequences and the values below the diagonal are the percent nucleotide substitution calculated from them (Gojobori et al. 1990). The average percent difference and percent nucleotide substitution between sequences 1 and 18 are 21.6 ± 4.1 and 26.5 ± 5.6, respectively. Sequences 19 and 20 were excluded from this

The Relationship Between Lx and L1 Sequences in the Genome

To learn if L1 was invariably physically associated with Lx in the genome, as was the case for LxRn1 (Fig. 1), we determined the percentage of rat or mouse genomic DNA clones that hybridized to L1 or Lx (or both). Figure 3 shows the results of hybridizing replicate nitrocellulose blots of genomic libraries of rat or mouse DNA with the Lx or either L1 PCR probe. Table 1 shows that both genomes contain about the same number of Lx elements and somewhat higher copies of L1 sequences, and that L1 and Lx sequences are physically associated to a somewhat greater extent than would occur by chance given their respective copy numbers. The relative amounts of Lx and L1 sequences agree with what we found earlier using blot hybridization of genomic DNA (Pascale et al. 1990) and the copy number of L1 in the two genera is similar to earlier estimates using probes from comparable regions of these L1 elements (Gebhard et al. 1982; Fanning 1983; D'Ambrosio et al. 1986).

In addition, Fig. 3 shows that each library contained a few clones that hybridized to the heterologous L1 probe. Since the rat and mouse L1 probes are only 58% similar and do not hybridize to each other or to Lx (Fig. 2), these results indicate that each genome contains members of other L1 families in addition to Lx and its own L1 family. We are now investigating the nature of these families.

Since the L1RnB7 element was inserted directly into LxRn1 DNA (Fig. 1) we wondered whether this was a typical arrangement. To examine this possibility we isolated 10 more Lx-containing clones from the rat library. By restriction enzyme analysis and blot hybridization seven proved to contain L1 DNA (results not shown). From the results of Table 1 we would have expected that six of the Lx-containing clones would also contain L1 DNA; i.e., 82, or ~60% of the 139 Lx-containing clones, also contain L1. Four of the clones that contained both Lx and L1 were examined at the DNA sequence level and in only one case was the L1 element inserted directly into what was probably a rearranged Lx element (results not shown). Therefore, L1 and Lx sequences are not necessarily physically contiguous, as is the case for LxRn1.

Lx DNA Sequences from Different Murine Genera Are Diverged Members of the Same Family

The melting temperature of reannealed Lx family members in different modern murine genomes was similar (Pascale et al. 1990) and indicated that in each case the Lx family was about 22% divergent. If the Lx members are relics of an amplification event that occurred at about the same time, then we would expect that any pair of Lx DNA sequences should be similarly divergent irrespective of their source. To test this prediction we determined the DNA sequence of more Lx members including two

Table 2. Extended

10	11	12	13	14	15	16	17	18	19	20	21
21.7	21.9	23.6	23.0	21.1	21.2	18.8	20.5	25.4	32.8	27.4	10.8
25.2	28.0	28.6	28.5	29.6	26.2	25.0	27.1	31.5	39.4	35.3	18.1
21.5	18.2	21.8	21.2	19.3	18.0	15.6	17.9	23.1	30.0	28.4	7.8
22.9	18.2	23.8	21.2	16.3	19.0	16.6	18.9	22.1	29.0	28.9	9.7
24.4	25.6	21.8	22.2	20.3	20.5	18.6	20.9	26.1	31.0	29.9	11.7
24.0	22.1	24.0	23.1	22.3	21.4	17.4	19.9	23.3	29.4	29.3	10.5
26.0	24.8	21.5	23.4	24.4	24.9	21.2	23.0	28.3	33.2	29.6	13.7
27.9	25.2	26.3	27.1	25.9	23.9	23.7	25.0	28.3	31.2	32.5	15.6
19.7	20.9	19.8	22.0	19.7	18.3	16.2	15.7	24.1	32.7	30.3	7.4
****	24.3	27.0	26.5	25.0	23.5	20.2	24.0	27.4	35.7	36.1	14.6
29.4	****	26.2	25.1	19.2	19.3	18.8	21.7	23.1	29.6	29.9	13.3
33.7	33.0	****	23.4	24.4	24.0	23.7	23.0	28.3	35.2	36.0	16.1
32.8	31.1	28.1	****	22.7	22.9	20.2	24.5	29.3	34.4	32.1	14.2
30.7	22.5	29.9	27.8	****	18.5	17.0	17.9	21.1	30.8	30.2	11.9
28.5	22.4	29.2	27.6	21.7	****	19.6	22.0	20.6	33.2	30.2	12.7
23.6	21.8	29.0	23.9	19.7	23.1	****	16.9	23.0	30.9	28.1	9.1
29.2	26.1	27.8	30.3	21.1	26.8	19.6	****	26.4	35.4	30.0	12.4
34.6	28.1	36.6	38.7	25.8	24.5	28.6	34.4	****	30.8	30.8	17.4
48.7	37.9	48.1	47.3	40.3	44.2	40.2	49.6	40.1	****	40.5	26.4
50.0	38.7	49.5	42.8	40.0	39.7	35.8	38.9	41.1	59.9	****	23.8
16.3	14.8	18.2	15.8	13.2	14.0	9.7	13.6	20.3	32.7	28.9	****

calculation because the mean differences between them and the others were 31.3 ± 3.3 and 32.9 ± 3.2, respectively, which is almost three standard deviations from the mean of all the others

additional rat genomic clones (LxRn3, and LxRn5a) and 14 Lx members cloned by the cDNA method (Materials and Methods) from several murine genera. The cDNA clones contained the region of the 3′ UTR corresponding to the PCR2 probe. (See Fig. 1.) We used this region of the LxRn1 clone to screen the GENBANK nucleic acid database for additional Lx sequences.

An initial pairwise comparison of the cloned Lx sequences and the output of the database search was carried out using the PILEUP program. (See Materials and Methods.) This program generates a dendrogram (Fig. 4) in which the grouping of the sequences is based on the similarity found during the initial pairwise comparison. As Fig. 4 shows, we distinguished three major clusters of sequences: L1Rn, L1Md, and Lx. Since the database search yielded sequences as distant from Lx as the modern rat and mouse L1 families, we feel that the list of sequences in Fig. 4 is quite comprehensive for Lx and related sequences.

The sequences included in the L1Rn or L1Md clusters are ~90% or more similar to the canonical members of these modern L1 families (L1Rn3 and L1MdA2, respectively). We included in the Lx cluster all of the sequences that were grouped with the cloned Lx members by the PILEUP program. (See Fig. 4.) The pairwise comparison of all of the Lx sequences is shown in Table 2. Because the length of the branches in the dendrogram (Fig. 4) includes a gap penalty, the distance of particular pairs from their node will not necessarily correspond to the percent divergence in Table 2. With the exception of LxRnx3 and LxMmx1, the Lx sequences from different murine species are, on average, 21.6 ± 4.1% divergent, which agrees with the DNA melting data (Pascale et al. 1990, and additional unpublished results).

The results in Table 2 can also be used to estimate when the Lx members were generated. Since any pair of Lx sequences has accumulated 26.6% ± 5.6 nucleotide substitutions, each has accumulated 26.6% ÷ 2 or 13.3% nucleotide substitutions since the time of their divergence. If the Lx amplification was similar to the amplification of modern L1 families, then most of the members were defective and have been diverging from each other at the pseudogene or neutral rate since their generation. The ratio of replacement to synonymous substitutions in the ORF II region of various Lx members indicates that this is so (data not shown). Using 1.1% substitutions per site per million years as the neutral nucleotide substitution rate estimated for rodents (Li et al. 1987), we calculate that the Lx family was amplified ~12.1 million years ago (13.3% ÷ 1.1%/million years).

Interestingly, LxRn3 and Ratil6gl differ by only 11.9% nucleotide substitutions per site, which is far lower than the difference between these sequences and the other Lx members. Therefore, LxRn3 and Ratil6g1 apparently have been diverging for only about 5.4 million years, well after the wave of Lx amplification supposedly ceased. Since the non-Lx sequence flanking the LxRn3 or the Ratil6g1 Lx

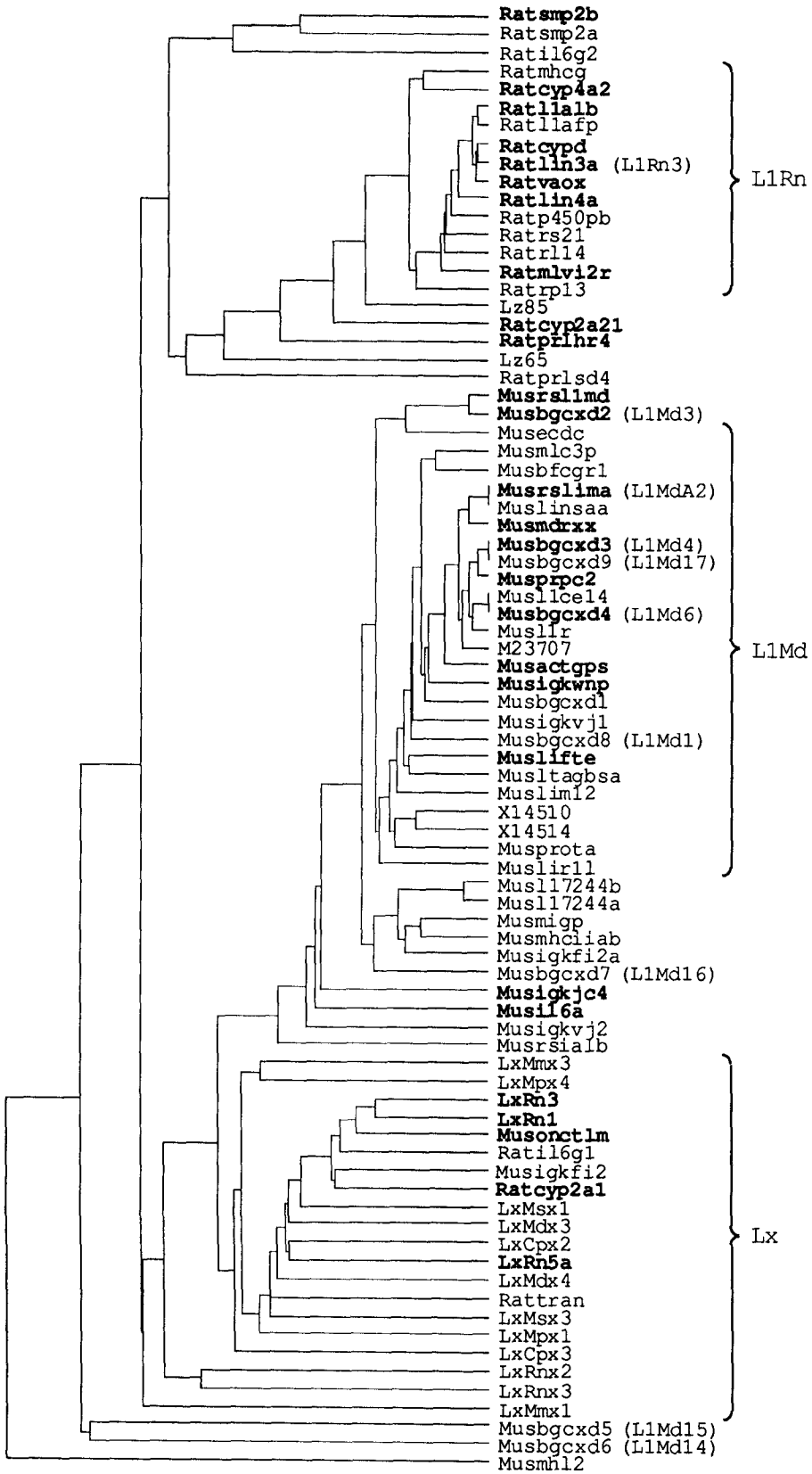

Fig. 4. Dendrogram of Lx and L1 sequences. This dendrogram was generated by the GCG multiple sequence alignment program, PILEUP, and represents the initial alignment of the sequences based on their similarity to each other. The sequences labeled *M23707, X14510,* and *X14514* and all those beginning with *Mus* or *Rat* were obtained from GENBANK using the region of the LxRn1 sequence that corresponds to the PCR 2 probe (see Fig. 1) as a query sequence for the GCG program WORDSEARCH. We named these L1 sequences by their GENBANK locus name. The Lz85 and Lz65 sequences are about 85% and 65% homologous, respectively, to L1Rn sequences and were taken from Zullo et al. (1991). All of the others were either selected from a λ rat genomic library using the LxRn1 PCR 1 probe (LxRn3, LxRn5a) or represent cDNA clones of the PCR-amplified Lx families in various murine genomes prepared as described in Materials and Methods. The curly brackets labeled *L1Rn* and *L1Md* indicate sequences that were about 90%, or more, similar to the canonical rat and L1 family members L1Rn3 and L1MdA2, respectively. The curly bracket labeled *Lx* includes all of the sequences that clustered with cloned *Lx* sequences. The sequences named in boldface type were used for the parsimony analysis shown in Fig. 5. The other names in parentheses (L1Md1, L1Md3, . . .) indicate L1 or L1-like sequences located in the mouse β-globin complex (Shehee et al. 1989).

sequence are not homologous (results not shown), gene duplication of one of these sequences about 5.5 million years ago does not account for their low rate of divergence.

LxRnx3 and LxMmx1 are only about 60% similar to each other or to either L1Rn or L1Md, and therefore each represents distinct L1-like families. We have not studied these sequences further.

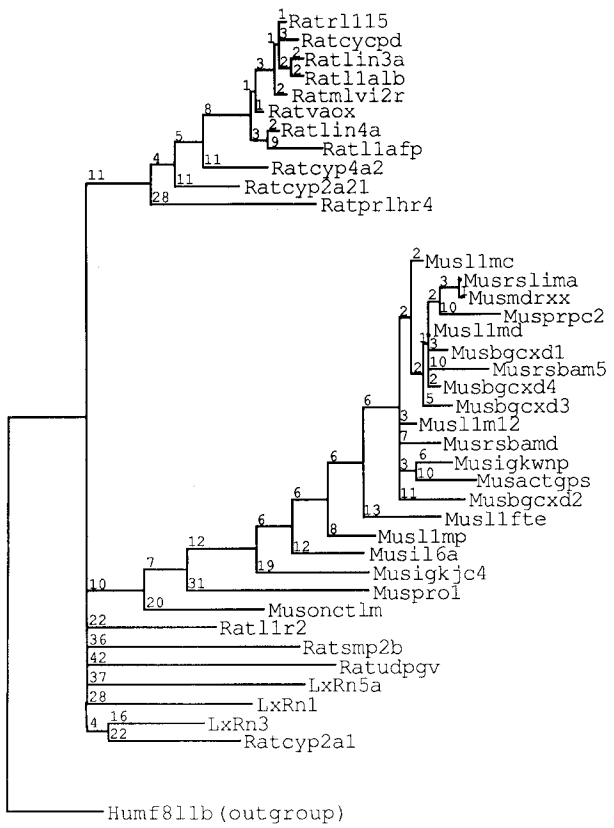

Fig. 5. Parsimony analysis of Lx and L1 elements. Parsimony analysis using the last 312 bp of ORF II was performed with the PAUP program (Swofford 1990) as described in Materials and Methods. We used all of the sequences shown in bold type in Fig. 4. These include representative members of the modern L1 families (i.e., those that are included in the L1Rn or L1Md brackets shown in Fig. 4) and all of the other sequences listed in Fig. 4 that contained this region of ORF II. We also included any other nonmodern rodent L1 sequences that were found in GENBANK using the last 312 bp of the ORF II of the LxRn1 sequence as a query sequence. The human L1Hs sequence, Humf8l1b, was used to root the tree. Branch lengths are the number of nucleotide changes between a terminus and a node or between nodes. The sequence names are either GENBANK locus names or the names of sequences determined in the present work.

Phylogenetic Relationship Between Lx and L1 Sequences

We determined the phylogenetic relationship between Lx and the modern L1 families in rat and mouse using parsimony analysis as described in the Materials and Methods (Swofford 1990). To ensure the best possible sequence alignment we used the protein-encoding region of the sequences, and in particular, the last 312 bp of ORF II, which was previously used for the phylogenetic analysis of the *Mus* L1 families (Martin et al. 1985; Hardies et al. 1986).

Figure 5 shows the results of the PAUP analysis, which are consistent with the previously established phylogenetic relationships between the mouse L1 sequences. For example, the ancestral relationship between the families in *M. pahari*, (Musl1mp), *M. caroli* (Musl1mc), and *M. domesticus* (Musl1md) as well as the fact that the F-type L1Md family (Musl1fte) is of earlier vintage than the A-type L1Md family (e.g., Musrslima or L1MdA2) (Adey et al. 1991) is evident in Fig. 5.

Both modern L1 rodent families were derived from the node shared by all but one of the sequences present in the Lx cluster indicated in Fig. 4. Therefore, the parsimony analysis supports our earlier suggestion based on the degree of sequence similarity between Lx and either L1Rn or L1Md (Pascale et al. 1990) that Lx is ancestral to the modern rat and mouse L1 families. One of the sequences, Ratsmp2b, that also shared the Lx node was not clustered with the Lx sequences in Fig. 4. Pairwise comparisons of the 3′ UTR of this sequence with the corresponding region of members of the Lx and rodent L1 members indicated in Fig. 4 showed that Ratsmp2b was considerably more similar to rat L1 then to mouse L1 (data not shown). Thus, while Ratsmp2b may be an evolutionary intermediate between Lx and the modern rat L1 sequences, it apparently diverged from the evolutionary pathway that gave rise to the modern L1 family exemplified by the L1Rn3-type sequences.

The parsimony analysis could not establish a lineage for most of the Lx sequences as well as some of the modern L1 sequences. These results would be expected if the sequences in question are the diverged progeny of the same (or a group of minimally differentiated) sequence(s). This idea was also proposed to account for the phylogenetic relationship between some *Mus* L1 sequences (Hardies et al. 1986).

Discussion

The extent of DNA sequence divergence between Lx family members (Table 2) shows that this family contains the relics of an ancient L1 family that was extensively amplified around the time of the murine radiation, ~12 million years ago. The confinement of Lx sequences to murine rodents (Pascale et al. 1990, and additional unpublished results now extending our analysis to a total of 20 murine and 16 nonmurine species) supports this conclusion. Parsimony analysis showed that the modern L1 families, L1Rn and L1Md, were derived from functional L1 elements related to Lx. However, the lack of cross-hybridization between the Lx family and either of the modern L1 families (Fig. 2) indicates that none of these families contains significant numbers of DNA elements that are intermediate in sequence between the Lx and modern families. The latter results are consistent with our DNA melting experi-

ments of reannealed L1Rn, L1Md, or various murine Lx family members which showed that each of these families consists of a discrete, though divergent in the case of Lx, cohort of sequences (Pascale et al. 1990).

Our interpretation of the cross-hybridization results of Fig. 2 depends on the assumption that the PCR amplification would not discriminate against the putative evolutionary intermediates between the Lx and either modern L1 family. Although these intermediates would be 10–15% more divergent from the canonical members of these respective families than their typical members, it would be very surprising if they were not amplified. As mentioned earlier, Sommer and Tautz (1989) showed that amplification by the PCR requires no more than 50% homology between the primer and the target DNA provided there is no mismatch with the three 3′ bases of the primer. This result is borne out by the results given in Table 2 which show that sequences ranging from 15% to as much as 40% divergent from each other were amplified from genomic DNA by our single set of Lx PCR primers.

The fact that the Lx and either the modern rat or mouse L1 families are not loci on a continuum of L1 DNA sequences but rather represent discrete episodes of L1 DNA amplification indicates that the functional evolutionary intermediates between Lx and L1Rn or L1Md were not extensively amplified. Figure 5 shows that the modern rat L1 family [e.g., Ratlin3a (L1Rn3)] and the modern mouse family [e.g., Musrslima (L1MdA2)] are separated from the Lx node by several intermediate elements. However, none of these is sufficiently divergent from either modern L1 family or the Lx family to be distinguished from them by hybridization, and so their genomic copy number could not be unambiguously determined. For example, pairwise comparisons between the 3′ UTR of the Ratprlhr4 or Musigkjc4 elements showed them to be about 80% homologous to modern rat or mouse L1 elements, respectively (results not shown). The Muspro1 sequence lacked a 3′ UTR and the Musonctlm sequence is about 80% similar to Lx sequences (Table 2).

Studies on the amplification of modern L1 families support our contention that replication-competent L1 elements can reside in genomes without undergoing extensive amplification. For example, the large-scale amplification of both L1Rn and L1Md began only about 5 million years ago (Martin et al. 1985; Hardies et al. 1986; Casavant et al. 1988; and unpublished pairwise comparisons of L1Rn members carried out in this laboratory) even though the rat and mice lineages have been separate for at least 12 million years (Catzeflis et al. 1987). Second, in studies to be reported elsewhere, we found that whereas the Ratmlvi2r subfamily (Fig. 5) of rat L1 is present in both *R. norvegicus* and *R. rattus*, it was extensively amplified in *R. norvegicus* but in only one of nine isolates of *R. rattus* (defined morphologically) collected from around the world. This suggests that although the Ratmlvi2r subfamily is functional in *R. rattus*, by and large, it has remained quiescent. And finally, a recently identified functional human L1 element which has apparently generated a copy of itself within the past 20 years has resided in the same position in the primate genome since before gorillas, chimps, and humans diverged 6 million years ago, but appears to have generated only about 10 copies of itself (Dombroski et al. 1991).

If the integrity of L1 elements depends solely on selection for self-replication, then it is apparent that the extensive amplification that typifies the Lx and modern L1 families is not an inevitable by-product of L1 amplification. This may mean that replication and extensive amplification occur by different mechanisms, or that factors other than the functionality of the L1-encoded products determine whether large-scale amplification will occur. One possibility is that residence in some genomic locations may be more conducive to large-scale amplification than others.

Together, Lx and L1 sequences make up 10–20% of the rat and mouse genomes. Relics of L1 DNA amplifications that occurred more than 30 million years ago would now be 65% divergent and difficult to distinguish from random DNA sequence even by DNA sequence determination. Therefore, the fraction of murine genomes that is derived from L1 DNA may even be >20% and, in the absence of compensatory deletions, the repeated amplification of L1 would have increased the size of murine genomes. In addition, recurrent episodes of L1 amplification could have provided, in some cases, the reverse transcriptase that catalyzed the retroposition of distinctive SINE families (and subfamilies) into mammalian genomes (Deininger and Daniels 1986; Weiner et al. 1986; Willard et al. 1987; Britten et al. 1988; Jurka and Smith 1988). For example, the rat ID family (Sutcliffe et al. 1982) became highly repeated only in rat and was amplified at about the same time as the L1Rn family (Deininger and Daniels 1986).

Since modern mammalian L1 elements are apparently still undergoing evolutionary change, we would assume that extensive amplification of L1 DNA could occur again. Besides causing insertion mutations (Kazazian et al 1988; Morse et al. 1988) we have shown in other studies that L1 DNA can effect the structural and regulatory properties of the DNA into which it inserts (Usdin and Furano 1989;

and unpublished results). Therefore, L1 elements have had and most likely will continue to have a major effect on the structure and function of the mammalian genome.

Acknowledgments. The authors are indebted to Dr. Bruce Hayward, of this laboratory, for suggesting the use of a transcription and cDNA step to ensure the faithful cloning of individual family members from a mixture of a multigene family that was amplified by the PCR.

References

Adey NB, Comer MB, Edgell MH, Hutchison III CA (1991) Nucleotide sequence of a mouse full-length F-type L1 element. Nucleic Acids Res 19:2497

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

Britten RJ, Baron WF, Stout DB, Davidson EH (1988) Sources and evolution of human Alu repeated sequences. Proc Natl Acad Sci USA 85:4770–4774

Burton FH, Loeb DD, Voliva CF, Martin SL, Edgell MH, Hutchison III CA (1986) Conservation throughout mammalia and extensive protein-encoding capacity of the highly repeated DNA long interspersed sequence one. J Mol Biol 187: 291–304

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

Catzeflis FM, Sheldon FH, Ahlquist JE, Sibley CG (1987) DNA-DNA hybridization evidence of the rapid rate of muroid rodent DNA evolution. Mol Biol Evol 4:242–253

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

Deininger PL, Daniels GR (1986) The recent evolution of mammalian repetitive DNA elements. Trends Genet 2:76–80

Devereux J, Haeberli P, Smithies O (1984) A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res 387–395

Dombroski BA, Mathias SL, Nanthakumar E, Scott AF, Kazazian HJ (1991) Isolation of an active human transposable element. Science 254:1805–1808

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

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

Gebhard W, Meitinger T, Hochtl J, Zachau HG (1982) A new family of interspersed repetitive DNA sequences in the mouse genome. J Mol Biol 157:453–471

Gojobori T, Li W-H, Graur D (1982) Patterns of nucleotide substitution in pseudogenes and functional genes. J Mol Evol 18:360–369

Gojobori T, Moriyama EN, Kimura M (1990) Statistical methods for estimating sequence divergence. Methods Enzymol 183: 531–550

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, Hidaka S, Sakaki Y (1985) Sequence analysis of a KpnI family member near the 3′ end of human β-globin gene. Nucleic Acids Res 13:7813–7827

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

Jansen R, Ledley FD (1990) Disruption of phase during PCR amplification and cloning of heterozygous target sequences. Nucleic Acids Res 18:5153–5156

Jurka J, Smith T (1988) A fundamental division in the Alu family of repeated sequences. Proc Natl Acad Sci USA 85:4775–4778

Kazazian HHJ, 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

Li W-H, Tanimura M, Sharp PM (1987) An evaluation of the molecular clock hypothesis using mammalian DNA sequences. J Mol Evol 25:330–342

Li W-H, Wu C-I, Luo C-C (1984) Nonrandomness of point mutation as reflected in nucleotide substitutions in pseudogenes and its evolutionary implications. J Mol Evol 21:58–71

Loeb DD, Padgett RW, Hardies SC, Shehee WR, Comer MB, Edgell MH, Hutchison III CA (1986) The sequence of a large L1Md element reveals a tandemly repeated 5′ end and several features found in retrotransposons. Mol Cell Biol 6:168–182

Martin SL, Voliva CF, Burton FH, Edgell MH, Hutchison III CA (1984) A large interspersed repeat found in mouse DNA contains a long open reading frame that evolves as if it encodes a protein. Proc Natl Acad Sci USA 81:2308–2312

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 HJ, Boeke JD, Gabriel A (1991) Reverse transcriptase encoded by a human transposable element. Science 254:1808–1810

Morse B, Rotherg PG, South VJ, Spandorfer JM, Astrin SM (1988) Insertional mutagenesis of the myc locus by a LINE-1 sequence in a human breast carcinoma. Nature 333:87–90

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

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

Pathak VK, Temin HM (1990) Broad spectrum of in vivo forward mutations, hypermutations, and mutational hotspots in a retroviral shuttle vector after a single replication cycle: deletions and deletions with insertions. Proc Natl Acad Sci USA 87:6024–6028

Rogers JH (1985) The origin and evolution of retroposons. Int Rev Cytol 93:187–279

Saiki RK, Scharf S, Faloona F, Mullis KB, Horn GT, Erlich HA, Arnheim N (1985) Enzymatic amplification of β-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. Science 230:1350–1354

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74: 5463–5467

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 MG (1992) Identification of transcriptional regulatory activity within the 5′ A-type monomer sequence of the mouse LINE-1 retroposon. Mamm Genome 2:41–50

Shehee WR, Loeb DD, Adey NB, Burton FH, Casavant NC, Cole P, Davies CJ, McGraw RA, Schichman SA, Severynse DM, Voliva CF, Weyter FW, Wisely GB, Edgell MH, Hutchison III CA (1989) Nucleotide sequence of the BALB/c mouse beta-globin complex. J Mol Biol 205:41–62

Sommer R, Tautz D (1989) Minimal homology requirements for PCR primers. Nucleic Acids Res 17:6749

Sutcliffe JG, Milner RJ, Bloom FE, Lerner RA (1982) Common 82-nucleotide sequence unique to brain RNA. Proc Natl Acad Sci USA 79:4942–4946

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

Swofford DL (1990) PAUP: Phylogenetic analysis using parsimony, version 3.0. Computer Program distributed by the Illinois Natural History Survey. Champaign, Illinois, Illinois Natural History Survey

Tabor S, Richardson CC (1987) DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. Proc Natl Acad Sci USA 84:4767–4771

Tabor S, Richardson CC (1989) Selective inactivation of the exonuclease activity of bacteriophage T7 DNA polymerase by in vitro mutagenesis. J Biol Chem 264:6447–6458

Usdin K, Furano AV (1989) Insertion of L1 elements into sites that can form non-B DNA. Interactions of non-B DNA-forming sequences. J Biol Chem 264:20736–20743

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

Weiner AM, Deininger PL, Efstratiadis A (1986) Nonviral retroposons: genes, pseudogenes, and transposable elements generated by the reverse flow of genetic information. Annu Rev Biochem 55:631–661

Willard C, Nguyen HT, Schmid CW (1987) Existence of at least three distinct Alu subfamilies. J Mol Evol 26:180–186

Woodcock DM, Crowther PJ, Diver WP, Graham M, Bateman C, Baker DJ, Smith SS (1988) RglB facilitated cloning of highly methylated eukaryotic DNA: the human L1 transposon, plant DNA and DNA methylated in vitro with human DNA methyltransferase. Nucleic Acids Res 16:4465–4482

Xiong Y, Eickbush TH (1988) Similarity of reverse transcriptase-like sequences of viruses, transposable elements, and mitochondrial introns. Mol Biol Evol 5:675–690

Zabarovsky ER, Allikmets RL (1986) An improved technique for the efficient construction of gene libraries by partial filling-in of cohesive ends. Gene 42:119–123

Zullo S, Sieu LC, Slightom JL, Hadler HI, Eisenstadt JM (1991) Mitochondrial D-loop sequences are integrated in the rat nuclear genome. J Mol Biol 221:1223–1235

Received May 8, 1992/Revised July 2, 1992