Two Persistent LINE-1 Lineages in *Peromyscus* Have Unequal Rates of Evolution

N. Carol Casavant, Amy N. Sherman[1] and Holly A. Wichman

Department of Biological Sciences, University of Idaho, Moscow, Idaho 83844

ABSTRACT

LINE-1, the major family of long, interspersed repeats in the mammalian genome, moves via an RNA intermediate and encodes its own reverse transcriptase. Comparative sequence analysis was used to reconstruct the phylogenetic history of LINE-1 dynamics in the deer mouse, *Peromyscus*. As is the case in *Mus* and *Rattus*, a very small number of active templates produce the majority of LINE-1 copies in *Peromyscus*. However, in contrast to the single LINE-1 lineage seen in the muroid rodents, *Peromyscus* has at least two LINE-1 lineages whose most recent common ancestor probably existed before the peromyscine radiation. Species-specific variants of Lineage 1, and intact open reading frames in the youngest elements of both Lineages 1 and 2, suggest that both lineages have remained active within the same genome. The higher number of shared-sequence variants in Lineage 1 relative to Lineage 2 suggests that Lineage 1 has replaced its master template much more frequently than Lineage 2 or that the reverse transcriptase Lineage 1 is more error prone. The implications of the method used to acquire LINE-1 sequences for analysis are discussed.

LINE-1 is the major family of long, interspersed repeats in the mammalian genome. It is dispersed via retrotransposition and is present in tens of thousands of copies per haploid genome in all species examined to date (for reviews, see ROGERS 1985; SKOWRONSKI and SINGER 1986; EDGELL *et al.* 1987; HUTCHISON *et al.* 1989). LINE-1 has been detected by Southern blot analysis throughout seven orders of mammals including Marsupialia (BURTON *et al.* 1986), suggesting that it was present in the common ancestor of subclass Theria. It has been characterized at the sequence level in a narrower range of species, including house mice (*Mus*; MARTIN *et al.* 1985; RIKKE *et al.* 1991; CASAVANT and HARDIES 1994), rats (*Rattus*; D'AMBROSIO *et al.* 1986), voles (*Microtus* and *Arvicola*; VANLERBERGHE *et al.* 1993), deer mice (*Peromyscus*; KASS *et al.* 1992), rabbits (*Oryctolagus cuniculus*; PRICE *et al.* 1992), and primates (*Homo*; JURKA 1989; SMIT *et al.* 1995). However, these sequence analyses vary not only in the number of elements and amount of sequence examined, but also in the method used to select elements for analysis.

A full-length LINE-1 consists of a 5' untranslated region that includes the promoter, two open reading frames (ORFs), a 3' untranslated region, and an A-rich tail. The second ORF, which encodes reverse transcriptase, is the most conserved portion of the element. However, in *Mus* only 10% of the elements are full length (HUTCHISON *et al.* 1989), and many of these have accumulated debilitating mutations within the ORFs, so the vast majority of LINE-1 elements in the these genomes are pseudogene copies incapable of propagation.

Because there are tens or hundreds of thousands of LINE-1 elements per genome, it is difficult to identify active elements. As a result, most of what is understood about LINE-1 movement comes from comparative sequence analysis of elements that may not be capable of producing copies, but can nevertheless yield information about the active element that gave rise to them. A major conclusion of such analyses is that most LINE-1s are produced by one or a few closely related templates termed "master" elements, and that over evolutionary time, these masters are replaced by a small number of their own progeny to form a lineage of "sequential masters" or "molecular drivers" (DEININGER *et al.* 1992). This model predicts that phylogenetic analysis of randomly selected elements will yield a tree with a single major lineage rather than a highly branched tree with many independent clades (CLOUGH *et al.* 1996). Analyses of LINE-1 sequences from mouse and rat yield trees with a single major lineage (RIKKE *et al.* 1991; PASCALE *et al.* 1993; ADEY *et al.* 1994; CASAVANT and HARDIES 1994; FURANO *et al.* 1994), although examination of the youngest elements in a lineage may reveal a more branched tree suggestive of multiple simultaneously active but closely related master elements (MARTIN *et al.* 1985; CASAVANT and HARDIES 1994; HOLMES *et al.* 1994; DOMBROSKI *et al.* 1993).

Phylogenetic analysis also provides a means of discriminating older inserts from younger inserts. Master elements can accumulate mutations; when this occurs, all subsequent progeny inherit these changes. Thus, retrotransposon subfamilies can be defined by specific shared variants as deduced by phylogenetic analysis

Corresponding author: Holly A. Wichman, Department of Biological Sciences, University of Idaho, Moscow, ID 83844.
E-mail: hwichman@uidaho.edu

[1] *Present address:* Department of Genetics, Stanford University Medical Center, Stanford, CA 94305.

(WILLARD *et al.* 1987; BRITTEN *et al.* 1988; RIKKE *et al.* 1991; SHEN *et al.* 1991; DEININGER *et al.* 1992; DEININGER and BATZER 1993; CASAVANT and HARDIES 1994). Furthermore, these diagnostic markers can be sequentially ordered, in which case the older elements will share only the sequence variants that occurred earliest in the evolutionary history of the lineage while the youngest elements will share those older variants as well as more recently derived diagnostic differences. Subfamilies (clades) can be identified on the basis of shared restriction sites or single basepair variants. Subfamilies or clades within a single lineage will be nested (*i.e.*, a subfamily defined by recent sequence variants will be included in an older subfamily because the younger elements will contain all of the older variants that define the whole lineage in addition to the youngest sequence variants not found in the older elements). Because elements accumulate changes as they sit in the genome, sequence divergence between elements within a subfamily is greater for older subfamilies and very small for the youngest subfamilies. Thus there are several ways to recognize recently active lineages: members will share sequence variants accumulated over time by the sequential master elements, the youngest elements will have very limited sequence divergence from each other, and very young elements might be expected to have intact open reading frames.

Mus, Rattus, and human LINE-1s have been more extensively characterized than the elements of any other species; in fact, characterization in other species is generally limited to a single study. Most of what is known about LINE-1 comes from studies of a limited number of species, and it is not clear how applicable the findings may be to other species. For example, our current model for the dynamics of LINE-1 amplification is based on data from *Mus* species, but this model may not apply to all mammalian taxa (see for example VANLERBERGHE *et al.* 1993). Additionally, most (but not all) studies in *Mus* and *Rattus* have been carried out on inbred strains. If inbreeding alters the fixation rate of polymorphic inserts or the activity of master elements, it would be important to study LINE-1 in natural populations.

Here we report further characterization of LINE-1 in mice of the genus *Peromyscus*. *Peromyscus* is a genus of New World rodents that diverged from *Mus* ~35 mya (BROWNELL 1983). *Peromyscus* has been used extensively as a model system for mammalian evolution and population genetics (KING 1968; KIRKLAND and LAYNE 1989). The two species used in the present study (*Peromyscus leucopus* and *P. maniculatus*) are extremely wide ranging and are the most intensely studied members of the genus. These studies complement ongoing analysis of *mys* (WICHMAN *et al.* 1985; BAKER and WICHMAN 1990; Lee *et al.* 1996), a retrovirus-like element found in *Peromyscus* but absent in *Mus*.

MATERIALS AND METHODS

DNA: Clones (pDK55 and pDK62) containing L1Pm55 and L1Pm62 were generous gifts from W. DAWSON. *P. manicu-*

latus (TK29798; Orono, Maine) and *P. leucopus* (TK43952; Orono, Maine) tissues were generous gifts from R. BAKER. An additional *P. leucopus* specimen (20-3-143234; Browning, Texas) was also used.

Mouse DNA preparation: DNA was extracted from the tissues either by the standard phenol extraction or by the method of LONGMIRE *et al.* 1988. The DNA concentration was determined on an LS30 Luminescence Spectrometer (Perkin Elmer, Norwalk, CT) using a Hoechst staining method (LABARCA and PAIGEN 1980).

Construction of the LINE-1 libraries: Two libraries were constructed by isolating a 3.0-kb *XbaI* fragment from *P. leucopus* (TK43952) and *P. maniculatus* genomic DNA. DNA was extracted from the agarose gel using a Bio 101 GeneClean kit. The purified fragments were ligated into both pUC19 and Bluescript vectors from Stratagene and transformed into competent *Escherichia coli* DH5α cells without an outgrowth. Screening for single colonies was performed using probes made from the inserts of both pDK55 and pDK62 clones. Single clones were isolated, purified, and sequenced.

The LINE-1 pDK55 and pDK62 inserts were labeled using the USB random priming kit and $[\alpha^{32}P]$ dCTP (800 Ci/mmol) from NEN. Hybridizations were performed overnight in 5× SSCP (1× is 120 mM sodium chloride, 15 mM sodium citrate, and 20 mM sodium phosphate), 2 mg denatured salmon sperm and 1× Denhardt's. Hybridization was carried out overnight at 55°. Filters were washed three times for 30 min each with 5× SSCP at 55°.

A third library contained PCR derived fragments from *P. leucopus* 20-3-143234 genomic DNA. Clones containing these fragments are designated Leu*. The fragments were amplified with primers PEROL1F (AAGGATCCGCAGGATACAAGATCAACTCA) and PEROL1R (AAAGGATCCCAATTCGATTCCATTGGT) that are specific for conserved sequences within the collection of rodents in Figure 1. The reaction mixture for amplification included: 200 μM dNTPs, 1.5 mM MgCl$_2$, 1× Amplitaq reaction buffer, 50 pmoles of each primer and approximately 100 ng of genomic DNA in a total volume of 50 μl. The conditions for amplification included an initial denaturation at 95° for 5 min followed by 25 cycles of the following: 30 sec at 95°, 30 sec at 50°, and 1 min at 72°. The fragments were ligated into pGEMT vector (Promega). Fragments generated from the PCR derived method were screened to determined if the insert was either a Lineage 1 or Lineage 2 representative. The probes were oligonucleotides specific to Lineage 1 (Clade 1-3′, TTGTCATATAGGTCC) and to Lineage 2 (Clade 2.2 GATCTCCCATGCTCA). End labeling of the oligonucleotides was done according to standard procedures (AUSUBEL *et al.* 1989) in a 20 μl volume with 10 pmoles of oligonucleotides and 15 pmoles of $[\gamma^{32}P]$ ATP (6000 Ci/mmol). DE52 columns were used to remove the unincorporated ATP (AUSUBEL *et al.* 1989). Blots were prehybridized for 1 hr at 25° in 50 ml hybridization buffer [6× SSC, 10× Denhardt's, 0.3% (w/v) SDS] augmented with 2 mg of sheared, denatured salmon sperm. Blots were hybridized with probe overnight at 25° in 20 ml of hybridization buffer. Blots were washed three times for 30 min each in 6× SSC at 42° and exposed to film overnight.

Sequencing the LINE-1 elements: Individual clones were sequenced by the dideoxy chain termination method of SANGER *et al.* 1977, as per USB sequencing kit or Epicentre Technologies SequiTherm Cycle sequencing kit instructions. The sequences reported were verified by sequencing the complementary strand. The individual readings were assembled into contiguous sequences using the GeneWorks Sequencing Project program.

Sequence collection: Previously published sequences used in the analysis were: *P. maniculatus*, GenBank accession No. M97518, L1Pm55, coordinates 676–808, 929–1196; and

1228–1419, and M9751 (KASS *et al.* 1992), L1Pm62 coordinates 708–837 and 880–1356 (KASS *et al.* 1992); *Cricetulus griseus*, EMBL accession No. X15465, L1Cg, coordinates 157–721 (MILES and MEUTH 1989); *Mus domesticus*, L1MdA2, coordinates 5020–5631 (LOEB *et al.* 1986); and *Rattus norvegicus*, GenBank Accession No. M13100, L1Rn3A, coordinates 4665–5277 (D'AMBROSIO *et al.* 1986). The 19 sequences were assembled into an alignment with the GeneWorks alignment program, and subsequently adjusted by hand with the GeneWorks editor.

The phylogenetic trees: The DNA and Protein trees were derived using PAUP, version 3.0s (SWOFFORD 1990) using the Bootstrap 1000 replication heuristic algorithm. The parameters within the heuristic search options included: only minimal trees were kept, stepwise addition was random with 10 replications, and the branch swapping was by tree bisection-reconnection (TBR). The distance tree was derived using neighbor joining (SAITOU and NEI 1987) in PHYLIP, version 3.5C, using the Kimura two-parameter method (KIMURA 1980). Pairwise differences were based on the distance matrix derived in PAUP.

RESULTS

Collecting LINE-1 elements specific to Peromyscus: Because we wanted to collect elements that would reflect the dynamics of LINE-1 amplification in *Peromyscus*, we adopted a strategy that would bias against both the selection of only young elements and the probability of selecting mainly relic elements. We chose to collect elements contained in a major hybridizing restriction band seen in several diverse species of *Peromyscus* when probed with either of the previously characterized LINE-1 sequences from *P. maniculatus*, L1Pm55 and L1Pm62. KASS *et al.* (1992) showed that both of these probes hybridize strongly to a 3.0-kb *Xba*I band in a number of *Peromyscus* species. We found strong hybridization of these two probes to the above-mentioned *Xba*I band in *Peromyscus* (*leucopus, maniculatus, truei, californicus, difficilis,* and *polionotus*). We supplemented this with three PCR amplified LINE-1s from *P. leucopus.*

Our approach was to collect LINE-1 elements whose shared-sequence variants would define the lineage(s) existing within the *P. maniculatus-leucopus* species complex. We therefore isolated a number of LINE-1 elements from the 3.0-kb region of *P. leucopus* and *P. maniculatus.* We have sequenced a 600-bp region within ORF 2, the reverse transcriptase domain, of 14 different *Peromyscus* LINE-1 elements (eight from *P. maniculatus* and six from *P. leucopus,* Figure 1) to construct a phylogenetic history of LINE-1 within the genus *Peromyscus.*

Division of young peromyscine LINE-1 elements into two major groups based on shared-sequence variants: The 14 LINE-1 sequences we isolated were compared with two other *P. maniculatus* sequences (L1Pm55 and L1Pm62), another cricetid sequence (L1Cg, *C. griseus*) and two muroid rodent sequences (L1MdA2, *M. domesticus*; L1Rn3A, *R. norvegicus*), giving a total of 19 sequences. These 19 sequences are divided into three groups (Figure 1), with sequences from both *P. leucopus* and *P. maniculatus* represented in all three groups. The bottom two groups, containing 13 of the 14 newly col-

lected sequences, are distinguished from each other by their mutually exclusive shared-sequence variants. These variants are indicated in the alignment. The third group, consisting of the first six sequences in Figure 1, does not share these specific variants. This group contains the outgroup sequences and was utilized to determine the ancestral state of the shared-sequence variants in the other two groups. To further understand the evolutionary relationship within and among these three groups, phylogenetic trees were derived from both their DNA and protein sequences.

Phylogenetic trees further delineate the relationship of the Peromyscus elements: To derive the most parsimonious tree, the two muroid sequences, L1MdA2 (*Mus*) and L1Rn3A (*Rattus*) were used to root the cricetid sequences. L1Cg (*C. griseus,* Chinese hamster) could not be used as an outgroup because some of the other cricetid sequences appear to be as old as, or older than, this element.

The most parsimonious tree (Figure 2) revealed that the sequences were divided among two distinct lineages and their ancestors. Lineage 1 includes: Man29, Man108, Man110, Leu2-22, Leu*1-2, Man105, Man27, and Man109. Lineage 2 includes: Man28, Leu*2-1, and Leu*2-2, Leu4-5, and Man106. The sequences ancestral to these two lineages include L1Pm55, and possibly L1Cg, L1Pm62, and Leu1-18. The order and percent bootstrap confidence levels were taken from PAUP. The lineage-specific shared-sequence variants were initially determined using MacClade (MADDISON and MADDISON 1992), and when possible, ambiguous sites were resolved manually. The lineage-specific variants are listed at the node where they were acquired by the master template.

Although this is the most parsimonious tree, there are some branches with relatively low confidence scores, thus their order in the tree is not well resolved. The most significant unresolved placement is whether Leu1-18, L1Pm62, and L1Cg inserted within the early history of either lineage (Figure 2) or before the split between the two lineages. A tree (Figure 3a) derived by bootstrap analysis of the translated sequence (not shown) of those elements that contained an intact reading frame (L1MdA2, L1Rn3A, Man109, Man27, Leu2-22, Leu*1-2, Man106, and Leu*2-1) plus sequences where frameshift mutations could be easily deduced (Leu4-5, Man29, L1Cg, and Leu1-18) places Leu1-18 and L1Cg within Lineage 1, but with a confidence value of only 52%. The DNA tree in Figure 3b, derived by a distance method known as neighbor joining, places Leu1-18, L1Pm62, and L1Cg early in the history of Lineage 2.

All of these trees support three conclusions. First, there are at least two major LINE-1 lineages in *Peromyscus.* Second, Leu1-18, L1Pm62, and L1Cg inserted into the genome approximately at the time of the split between the two lineages or early in the history of one of these lineages. The sites that are involved in determining the position of these elements are shown in the
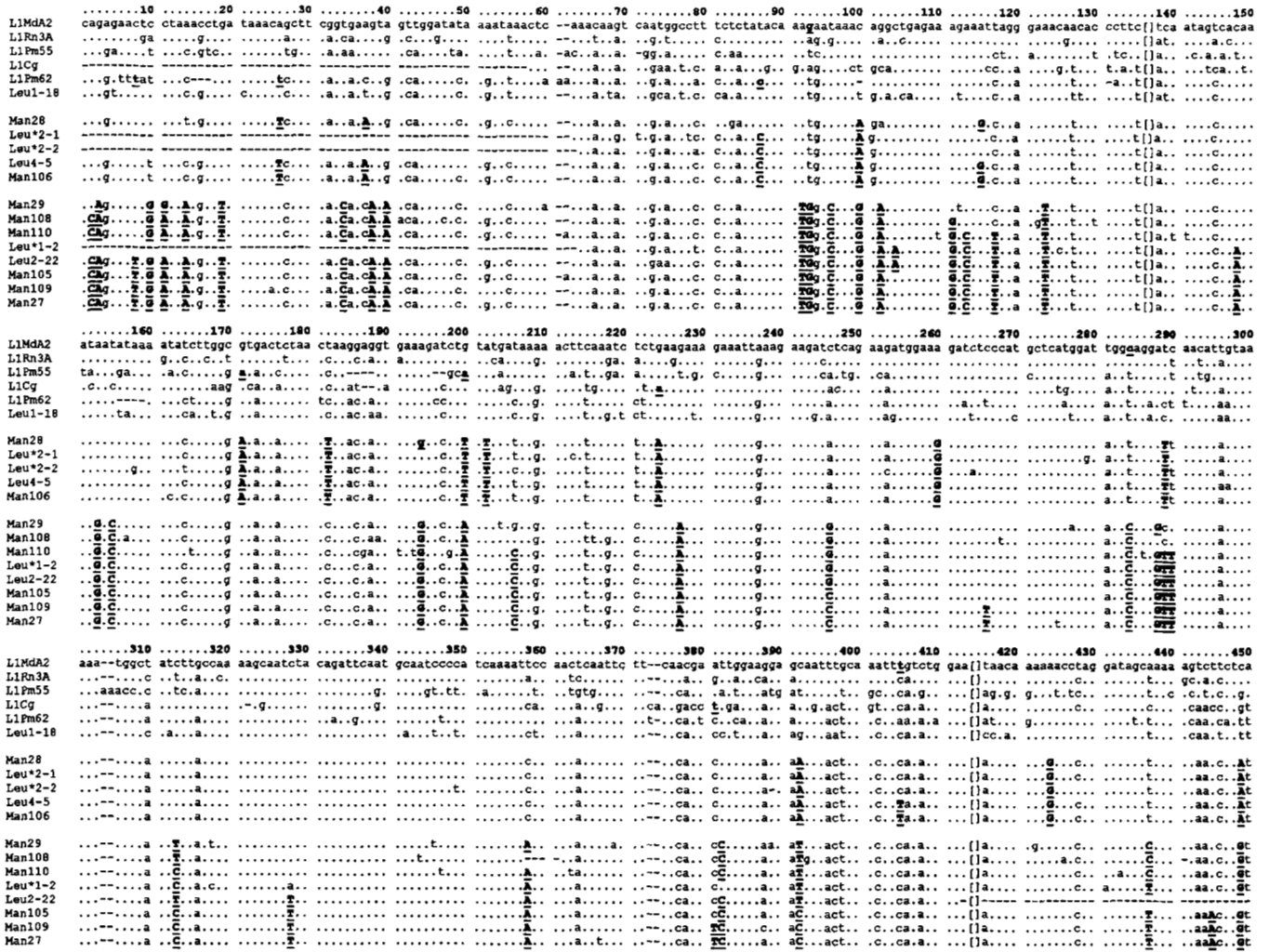
```
            .......10  .......20  .......30  .......40  .......50  .......60  .......70  .......80  .......90  ......100  ......110  ......120  ......130  ......140  ......150
L1MdA2    cagagaaactc ctaaacctga taaacagctt cggtgaagta gttggatata aaataaactc --aaacaagt caatggcctt tctctataca aagaataaac aggctgagaa agaaattagg gaaacaacac ccttc[]tca atagtcacaa
L1Rn3A    ........ga .....g.... ......a... .a.ca...g .c...g.... ....t.... --...t..a. ..g.t..... c........ ...aq.g... .a..c.... .......... .......g... ....[]at. ...a.c...
L1Pm55    ...ga....t ...c.gtc.. ......tg.. .a.aa.... .ca....ta. ....t...a. -ac..a..a. -gg.a.... c.aa..... ...tc..... .......... .....ct.. a.......t .tc..[]a. .c.a.a.t..
L1Cg      .......... .......... .......... .......... .......... .....a..a. ..gaa.t.c. a..a...g. g.ag....ct gca...... ......cc..a ...g.t... t.a.t[]a. ...tca..t.
L1Pm62    ...g.tttat ...c---... ......tc.. .a..a.c..g .ca.....c. .g..t....a aa...a... .g.a...a. c.a..o... ...tg....- .......... .c...a ...t.... -a.t[]a. ...c.....
Leu1-18   ...gt..... ...cg.... c........ .a..a.t..g .ca.....c. .g..t..... --...a.ta. ..gca.t.c. ca.a..... ...tg....t g.a.ca... .t....c..a ...tt.... t[]at. ...c.....

Man28     ...g..... ...t.g.... .tc... .a..a.A..g .ca.....c. .g..c..... --..a..a. ..g.a...c. c..a..... ...tg....A ga....... .o.c..a ...t.... ....t[]a. ...c.....
Leu*2-1   .......... .......... .......... .......... .......... ....a..g. t.g.a..tc. c..a.c... ...tg....A.g .......... ...c..a ...t.... t[]a. ...c.....
Leu*2-2   .......... .......... .......... .......... .......... ...a..a. c..a.c... ...tg....A g....... .......... ...c..a ...t.... t[]a. ...c.....
Leu4-5    ...g..... ...t ...c.g.... .tc... .a..a.A..g .ca.....c. .g..c..... --..a..a. ..g.a...c. c..a.c... ...tg....A g....... .o.c..a ...t.... t[]a. ...c.....
Man106    ...g..... ...t ...c.g.... .tc... .a..a.A..g .ca.....c. .g..c..... --..a..a. ..g.a...c. c..a.c... ...tg....A g....... .g.c..a ...t.... t[]a. ...c.....

            .......160 .......170 .......180 .......190 .......200 .......210 .......220 .......230 .......240 .......250 .......260 .......270 .......280 .......290 .......300
L1MdA2    ataatataaa atatcttggc gtgactctaa ctaaggaggt gaaagatctg tatgataaaa acttcaaatc tctgaagaaa gaaattaaag aagatctcag aagatggaaa gatctcccat gctcatggat tgggaggatc aacattgtaa
L1Rn3A    .......... g..c..c..t .......t... .c...c..a.. a......... ..ca...g. .......ga. a...g..... ......g... .......c... .......... .......... ........t ..tg.....
L1Pm55    ta...ga... .a.c....g a.a..c... .c.,---.. ....--gca .a....... ..a.t..ga. a...t.g. ...g... ..ca.tg. .ca...... .......... a..t.... t ..tg.....
L1Cg      .c..c.... ....aag .ca..a... .c..at--a ....c... ...ag..g. ...tg.... t.a...... ....ac... ..a...... .......... ..tg.... a..t.... t ..g.....
L1Pm62    .....----. ...ct...g .a....... tc..ac.a.. ....cc.... .......c.g. ....t.... ct........ .......g... .......a... .......... a..t.... t ..t.....
Leu1-18   .....ta... ...ca.t.g .a....... .c..ac.aa. ....c.... ....c.g. ...t.g.t ct......t .......g... ..g.a... ..ag...... .......c... a..t...a.ct t...aa...

Man28     .......... ...c....g A.a..a.. .I..ac.a. ..g..c..I I..t.g. ...t.... t.A........ .......a... ...g....... ..o....... a..t.... It .....a....
Leu*2-1   .......... ...c....g A.a..a.. .I..ac.a. .....c..I I..t.g. ..c.t.... t.A........ .......g... ..a....... ..o....... g. a..t.... I. .....a....
Leu*2-2   .......g... .......t... ...g A.a..a.. .I..ac.a. .....c..I I..c.g. ...t.... t.A........ .......g... ..a....c... ..o..a.. a..t.... I. .....a....
Leu4-5    .......... ...c....g A.a..a.. .I..ac.a. .....c..I I..t.g. ...t.... t.A........ .......g... ..a....... ..o....... a..t.... It .....aa...
Man106    .......... .c.c....g A.a..a.. .I..ac.a. .....c..I I..t.g. ...t.... t.A........ .......a... .......... ..o....... a..t.... It .....a....
```

legend of Figure 1. Finally, L1Pm55 (KASS et al. 1992) inserted into the genome long before the division into the two major lineages. L1Pm55 contains sequence variants that match the muroid rodent LINE-1 elements at positions 79, 120, 127, 135, and 258. L1Pm55 appears not only to be the oldest sequence retrieved from P. maniculatus but even older than the hamster sequence, since it, unlike the hamster element L1Cg, inserted long before the split of the two lineages.

A possible third lineage: A third possible lineage that includes Leu1-18 and L1Pm62, appeared within all of the trees derived, although the DNA bootstrap confidence was only 38%. There are six shared sequence variants at positions 206, 287, 297, 428, 449, and 538

that link these two elements into a distinct group (Figure 2). In the protein parsimony and DNA neighbor joining trees, this group and L1Cg were the first offshoots of Lineage 1 or 2. The implications of this third lineage are discussed below.

Shared sequence variants that define the two lineages: Any base within an element can be classified into one of four states: invariant base, ancestral state, shared-sequence variant, or private mutation. Shared-sequence variants are changes that are common to two or more elements within a clade and reflect changes that occurred within the master template before the transposition events that gave rise to the copies being examined. Private mutations are those changes unique
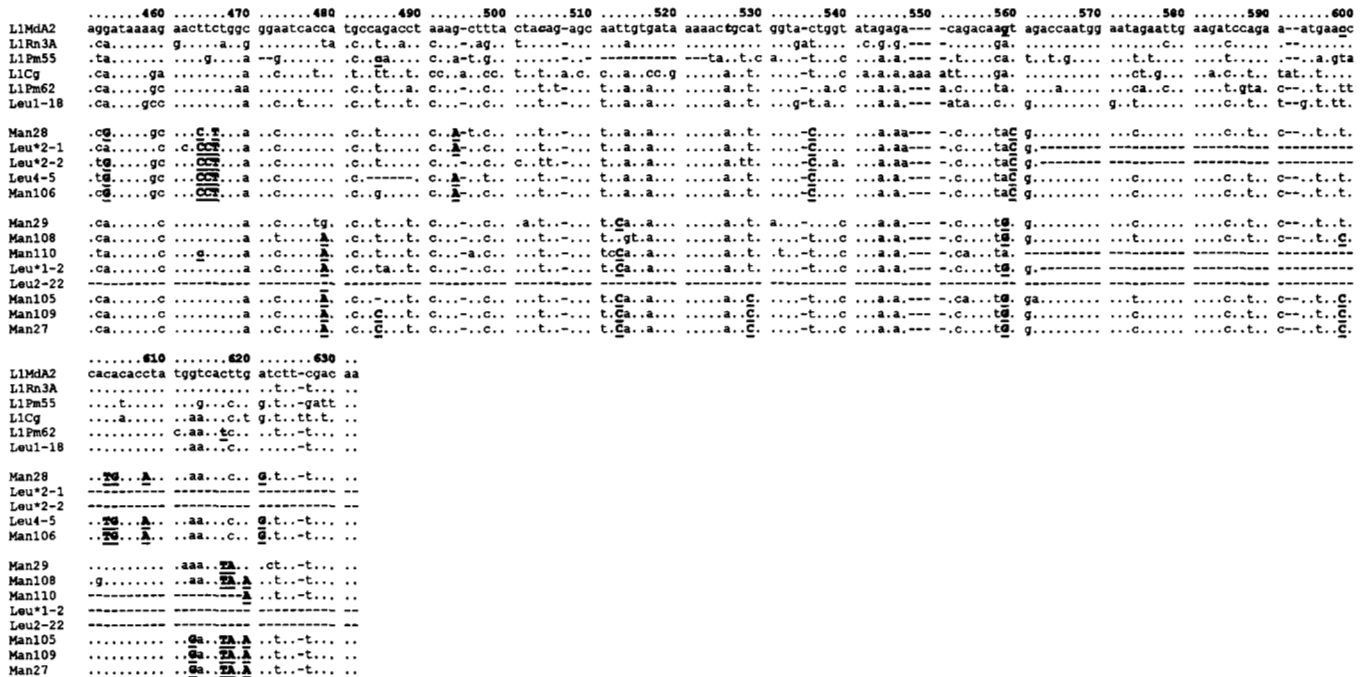
```
         .......460 .......470 .......480 .......490 .......500 .......510 .......520 .......530 .......540 .......550 .......560 .......570 .......580 .......590 .......600
L1MdA2   aggataaaag aacttctggc ggaatcacca tgccagacct aaag-cttta ctacag-agc aattgtgata aaaactgcat ggta-ctggt atagaga--- -cagacaagt agaccaatgg aatagaattg aagatccaga a--atgaagc
L1Rn3A   .ca...... g....a..g .........ta .c..t..a.. c..-.ag.. t....-... ...a..... ......... ...gat... .c.g.g.--- -......ga. ......... .......c.. ...c..... .-.....--
L1Pm55   .ta...... ....g....a --g....... .c..aa.... c..a-t.g.. ......-.. ---------- ---ta..t.c a...-t..c .a.a.--- -.t....ca. t..t.g.... ...t.t.... ...t..... .--..a.gta
L1Cg     .ca....ga .........a ..c...t... .t..tt..t cc..a..cc. t..t..a.c. c..a..cc.g .....a..t. ....-..c .a.a.a.aaa att....ga. ......... ...ctg.... ..a.c..t. tat..t....
L1Pm62   .ca....gc .........aa ......... .c..t...a. .c...-..c. ...t-.... t..a..a... ......a..t. ...-..a.c .a.a.--- a.c....ta. ...a..... ...ca..c.. ...t.gta. c--..t..tt
Leu1-18  .ca....gcc .........a ..c..t.... .c..t...t. c...-..c.. ...t.-... t..a..a... ......a..t. ...g-t.a. ...a.a.--- -ata...c.. g......... g..t..... ...c..t.. t--g.t.tt.

Man28    .cg....gc ..c.t..a. ..c...... .c..t.... c..a-t.c. ...t.-... t..a..a... ......a..t. ...c....... ...aa--- -.c....tac g......... ...c...... ...c..t.. c--..t..t.
Leu*2-1  .ca......c .c.cct..a. ..c...... .c..t.... c..a-.... ...t.-... t..a..a... ......a..t. ...c....... ...aa--- -.c....tac g.......... ...c...... ...c..t.. c--..t..t.
Leu*2-2  .tg....gc ...cct..a. ..c...... .c..t.... c...-..c. c..tt.-.. t..a..a... ......a.tt. ...c..a. ...aa--- -.c....tac g.-------- ---------- ---------- ----------
Leu4-5   .tc......gc ...cct..a. ..c...... .c-------. c..a-..t. ...t.-... t..a..a... ......a..t. ...c....... ...a.a.--- -.c....tac g......... ...c...... ...c..t.. c--..t..t.
Man106   .cg....gc ...cct..a. ..c...... .c..g.... c..a-.... ...t.-... t..a..a... ......a..t. ...c....... ...a.a.--- -.c....tac g......... ...c...... ...c..t.. c--..t..t.

Man29    .ca......c .........a ..c...tg. .c..t..t. c...-..c. .a.t.-... t.ca..a... ......a..t. a...-...c ...a.a.--- -.c....tg. g......... ...c...... ...c..t.. c--..t..t.
Man108   .ca......c .........a ..t.....a. .c..t..t. c...-..c. ...t.-... t..gt.a... ......a..t. ...-t..c .a.a.--- -.c....tg. g......... ...t...... ...c..t.. c--..t..c.
Man110   .ta......c ...c.....a ..c....a. .c..t..t. c...-a.c. ...t.-... tcca..a... ......a..t. .t..c ...ca..ta. ---------- ---------- ---------- ----------
Leu*1-2  .ca......c .........a ..c....a. .c..ta..t. c...-..c. ...t.-... t.ca..a... ......a..t. ...-t..c .a.a.--- -.c....tg. g......... ---------- ---------- ----------
Leu2-22  ---------- ---------- ......... --------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
Man105   .ca......c .........a ..c..-..t. .c..t..t. c...-..c. ...t.-... t.ca..a... ......a..c. ....-t..c .a.a.--- -.ca..ta. ga........ ...t...... ...c..t.. c--..t..c.
Man109   .ca......c .........a ..c....c. .c..c..t. c...-..c. ...t.-... t.ca..a... ......a..c. ....-t..c .a.a.--- -.c....tg. g......... ...c...... ...c..t.. c--..t..c.
Man27    .ca......c .........a ..c....a. .c..c..t. c...-..c. ...t.-... t.ca..a... ......a..c. ....-t..c .a.a.--- -.c....tg. g......... ...c...... ...c..t.. c--..t..c.

         .......610 .......620 .......630 ..
L1MdA2   cacacaccta tggtcacttg atctt-cgac aa
L1Rn3A   .......... .......... ..t..-t.... ..
L1Pm55   ...t..... ...g...c.. g.t..-gatt ..
L1Cg     ....a..... ..aa...c.t g.t..tt.t. ..
L1Pm62   .......... c.aa...tc. ..t..-t.... ..
Leu1-18  .......... ..aa...c.. ......-t.... ..

Man28    ..tg...a.. ..aa...c.. g.t..-t.... ..
Leu*2-1  ---------- ---------- ---------- --
Leu*2-2  ---------- ---------- ---------- --
Leu4-5   ..tg...a.. ..aa...c.. g.t..-t.... ..
Man106   ..tg...a.. ..aa...c.. g.t..-t.... ..

Man29    .......... ..aaa..ta. .ct..-t.... ..
Man108   .g........ ..aa..ta.a ..t..-t.... ..
Man110   ---------- ........a ..t..-t.... ..
Leu*1-2  ---------- ---------- ---------- --
Leu2-22  ---------- ---------- ---------- --
Man105   .......... ..ga..ta.a ..t..-t.... ..
Man109   .......... ..ga..ta.a ..t..-t.... ..
Man27    .......... ..ga..ta.a ..t..-t.... ..
```

FIGURE 1.—*Continued*

to one element in the clade and are generally considered to occur after an element is inserted into the genome; however, it is important to remember that some
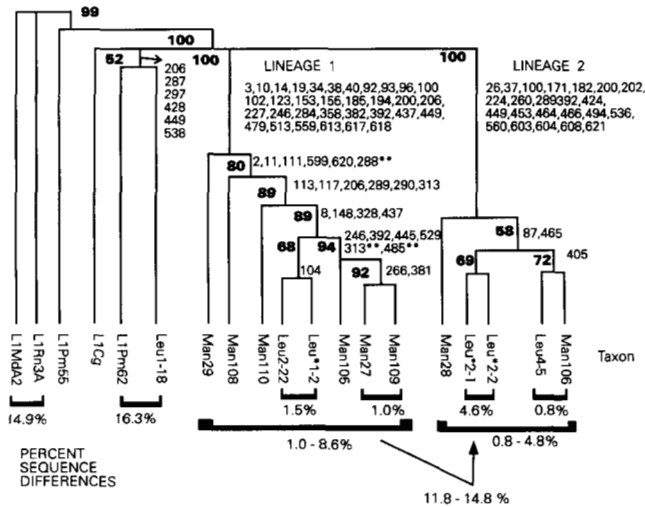

FIGURE 2.—Phylogenetic tree of *Peromyscus* LINE-1 elements depicting a bootstrap analysis. Two lineages of LINE-1 sequences are shown with their separate shared-sequence variants. The bold numbers at the nodes indicate the percent confidence value of the branch according to 1000 bootstrap replicates in PAUP. The numbers listed on the descending branches indicate the shared-sequence variants. The numbers at the bottom of the tree indicate the percent differences between or among the elements bracketed. Positions involved in the placement of one or more of Leu1-18, L1Pm62, and/or L1Cg within: Lineage 1; 37, 221, 289, and 489: Lineage 2; before the split; 55, 176, 202, 222, 101, 185, 313, and 459: before the split; 55, 176, 202, 222, 442, and 561: **; positions 288, 313 and 485 in Lineage 1 and 115, 453, and 494 in Lineage 2 are shared-sequence variants whose placement is unclear due to possible coincidental private mutations, See Figure 2 alignment.

private mutations may reflect changes in a master template that have only been sampled in one element. The muroid sequences, L1MdA2 and L1Rn3A, were used to determine the ancestral state because at many sites they reveal the state of LINE-1 before the muroid-cricetid split. However, because these muroid rodent sequences are relatively young and very closely represent their own active templates, they contain muroid-specific shared-sequence variants and private mutations that differ from the ancestral sequence. L1Cg, L1Pm55, L1Pm62, and Leu1-18 were used to clarify the ancestral state at sites with private mutations and to identify cricetid-specific shared-sequence variants.

By using sequences from both cricetids and muroids, sites with multiple changes in the master template have a high probability of being resolved. However, even though there are six different elements used in determining the ancestral sequence, there were a few positions where the ancestral state could not be determined. There were also positions where both of the lineages were different from all six ancestral sequences. For example, at position 100 in Figure 1, the ancestral base in L1MdA2, L1Rn3A, and L1Pm55 is a C, that changed to a T in the cricetids and further mutated to an A in Lineage 2 and a G in Lineage 1. The legend of Figure 1 lists the shared-sequence variants within each lineage as well as those changes that are ambiguous.

The order in which the master templates acquired the shared-sequence variants is shown on the tree in Figure 2. For both Lineage 1 and Lineage 2, the majority of shared-sequence variants were common to all elements sampled for that lineage. This suggests that the collected set of sequences is young relative to the time when the lineages split, which is probably a conse-
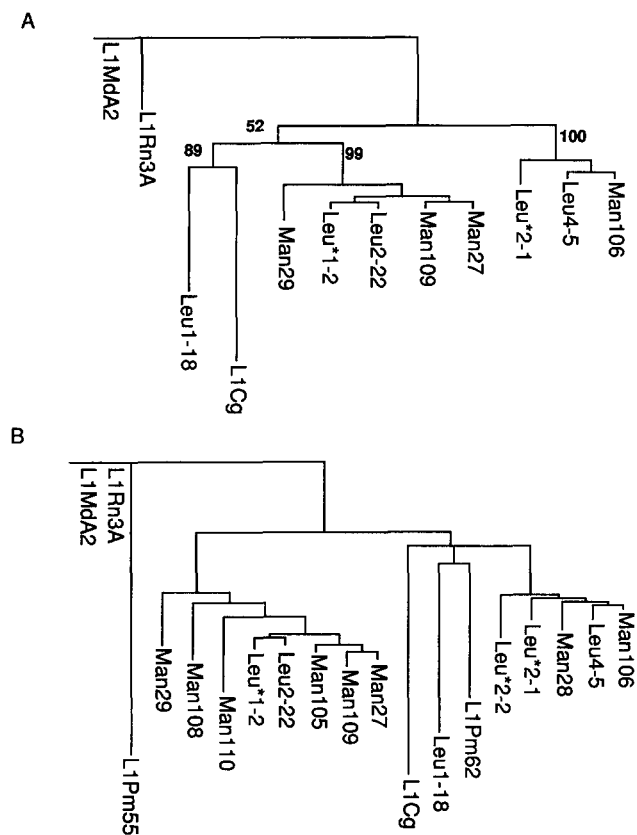
A

B

FIGURE 3.—Phylogenetic trees of *Peromyscus* LINE-1 sequences. (A) Phylogenetic tree of amino acid sequence of *Peromyscus* LINE-1 elements. Two major lineages are depicted with L1Pm62 and L1Cg as members of Lineage 1. The bold numbers at the nodes indicate the percent confidence value of the branch according to 1000 bootstrap replicates in PAUP. L1MdA2 and L1Rn3A were outgroups for rooting the tree. (B) Phylogenetic tree based on neighbor joining. Two major lineages are depicted with Leu1-18, L1Pm62, and L1Cg as members of Lineage 2. L1MdA2 was used as the outgroup to root the tree.

TABLE 1

Distribution of base changes that occurred in each sequence

Element	No. shared[a]	No. private[b]	Total no.[c]
Lineage 2			
Man28	24 (3.9)	6 (0.9)	613
Leu*2-1	18 (3.7)	10 (2.2)	484
Leu*2-2	18 (3.7)	13 (2.7)	483
Leu4-5	27 (4.4)	2 (0.3)	607
Man106	27 (4.4)	3 (0.5)	613
Lineage 1			
Man29	32 (5.2)	17 (2.8)	613
Man108	34 (5.6)	16 (2.6)	608
Man110	38 (6.8)	22 (3.9)	557
Leu*1-2	30 (6.2)	6 (1.2)	485
Leu2-22	37 (9.2)	2 (0.5)	403
Man105	48 (7.8)	3 (0.5)[d]	612
Man109	51 (8.3)	4 (0.7)	613
Man27	51 (8.3)	2 (0.3)	613

Values in parentheses are percentages.

[a] The shared sites are positions that have changed within the parent elements. Because there are sites that have changed several times, the numbers were based on counting the changes in Figure 1.

[b] The private changes are those changes that occurred after the element inserted into the genome and are not representative of the parent. Gaps were not counted.

[c] Total number of bases analyzed.

[d] Additionally there is a one base insertion and another base deletion.

quence of the strategy we used to select the elements. Further evidence that the elements sampled are young relative to the split between lineages is the distinct differences between *P. leucopus* elements and *P. maniculatus* elements within Lineage 1 (Figure 1), three of which result in replacements in the protein sequence. This suggests that the Lineage 1 master template has been active within each species.

Lineage 1 has acquired more shared-sequence variants than Lineage 2: Almost twice as many shared-sequence changes were acquired in Lineage 1 as in Lineage 2 (Figures 1 and 2). Inclusion of L1Cg, L1Pm62, and Leu1-18 in Lineage 2 does not markedly affect this discrepancy, nor does inclusion of positions for which the ancestral state could not be determined. Within a lineage, the youngest elements are expected to have more shared variants than the oldest elements, because shared variants are added as the master elements accumulate new changes. In comparing these lineages, it is important to note that the youngest elements of each lineage are approximately the same age: they have open

reading frames, there is low divergence between young elements, and there are very few private mutations. Nevertheless, the youngest elements of Lineage 1 have many more shared variants than the youngest elements of Lineage 2, and even the oldest elements in Lineage 1 have 31 shared variants while the youngest elements in Lineage 2 have 27 (Table 1). Because sampling two young elements will reveal all of the shared-sequence variants above that node, we do not believe that this difference is an artifact of the number of elements sampled in each lineage. Lineage 1 and Lineage 2 obviously diverged from their common ancestor at the same time, therefore, the increase in the number of shared-sequence variants in Lineage 1 is not due to an age difference between the two lineages.

The percent differences between and within the lineages: To better understand the extent of divergence between these two lineages, we compared the sequences of a young element from each lineage, Man27 and Man106, with young *Rattus* and *Mus* sequences, L1Rn3A and L1MdA2 (Table 2). The *Rattus* and *Mus* elements are part of a single lineage (ADEY *et al.* 1994); thus the distance between the two young elements, L1Rn3A and L1MdA2, reflects divergence because *Rattus* and *Mus* last shared a common ancestor ~10 mya (CATZEFLIS *et al.* 1992). For both the DNA and protein comparisons, the distances between the young elements of Lineage 1 and Lineage 2 were equivalent to the distances be-

TABLE 2

Comparison between Mus/Rattus and Peromyscus LINE-1 lineages

	(Mus) L1MdA2	(Rattus) L1Rn3A	(Lin 1) Man27	(Lin 2) Man106
(Mus)L1MdA2	—	14.9	21.2	19.9
(Rattus)L1Rn3A	15.2	—	20.3	21.1
(Lin 1)Man27	24.0	22.5	—	12.9
(Lin 2)Man106	22.1	23.0	16.2	—

Above the diagonal: percent nucleotide sequence differences; below the diagonal: percent amino acid sequence differences.

tween the *Rattus* and *Mus* elements (Table 2). The uneven tempo of movement and the accumulation of mutations during retrotransposition preclude using this comparison in any calculation of absolute time. However, this comparison is consistent with the earlier suggestion that Lineage 1 and Lineage 2 may have split from each other at or before the peromyscine radiation.

DISCUSSION

Peromyscine Line-1 model: We have described two LINE-1 lineages whose most recent common ancestor probably existed before the peromyscine radiation and whose master templates have remained active within the same genome. The divergence between these lineages is comparable with the divergence between the LINE-1s of *Rattus* and *Mus*, elements that last shared a common ancestor 10 mya (BROWNELL 1983; CATZEFLIS *et al.* 1992). In addition to these two lineages, the clade derived from their common ancestor includes three elements, Leu1-18, L1Pm62 and L1Cg, that straddle the split between the lineages. Because all three of these sequences represent ancient insertions, and L1Cg is from another cricetid genus, these elements strengthen the hypothesis that the split occurred within the cricetine radiation, but before the peromyscine radiation.

Mutually exclusive shared-sequence variants, as observed in the two lineages, are interpreted to be evidence of divergent active master templates. The longer the two active lineages have been separated, the more divergent the copies become due to the accumulation of changes in the master templates. The large divergence between the collected sequences of the two lineages is due primarily to the shared-sequence changes (Table 1). Master templates can also diverge due to speciation of the host species, as was seen in *Mus* (MARTIN *et al.* 1985). The mutually exclusive shared-sequence variants specific to a species are those that accumulated in the master template after the species last shared a common ancestor. There are species-specific variants in Lineage 1 for both *P. leucopus* and *P. maniculatus* (Figure 2), indicating that Lineage 1 has been active in both species since they last shared a common ancestor.

Although there are no species-specific variants for

Lineage 2, the number of private mutations within the youngest members of Lineage 2 is comparable to those found in the youngest members of Lineage 1. Private mutations are those changes that are not shared in a phylogenetically informative manner by more than one member of a data set. They include both mutations that occurred after an individual element inserted into the genome, and potential shared-sequence variants that were only sampled once. Thus, a low number of private mutations is an indicator that the individual copy recently inserted into the genome. The species-specific variants of Lineage 1 and the relative youth of the elements in both Lineages 1 and 2 suggest that the master templates of both lineages have remained active during the peromyscine radiation.

In *M. domesticus*, there appears to be a single lineage of active LINE-1 master templates that have acquired a series of new promoters; each successor may overlap with its predecessor before the predecessor died (ADEY *et al.* 1994). For the four species of *Mus* that have been examined to date, the sequence difference between the youngest elements in the single *Mus* lineage is small compared with the difference between the youngest members of Lineages 1 and 2 in *Peromyscus* but similar to the distance between young members of each *Peromyscus* lineage (MARTIN *et al.* 1985; HARDIES *et al.* 1986; CASAVANT and HARDIES 1994). This suggests that in Mus, the youngest master templates diverged after speciation, but are all successors of the same LINE-1 lineage. In contrast, the divergence of the two LINE-1 lineages in *Peromyscus* is consistent with the existence of multiple propagating master templates within species.

There have been few reported cases of multiple master lineages within a single species. A number of scenarios have been proposed for SINE evolution (JURKA and SMITH 1988; DEININGER *et al.* 1992; LEEFLANG *et al.* 1992; JURKA 1995), and alternatives to a master lineage model are discussed by DEININGER and BATZER (1993). In muroid rodents there is evidence for an extinct LINE-1 lineage in addition to the currently active one, suggesting that at some point in evolutionary history there were two concurrently active lineages (PASCALE *et al.* 1993; ADEY *et al.* 1994). There is also evidence for minor coexisting active lineages, generally of recent evolutionary origin (DEININGER and BATZER 1993; CASAVANT and HARDIES 1994). In humans, for example, KAZAZIAN and his colleagues have identified several very recent LINE-1 insertions associated with human genetic disease (KAZAZIAN *et al.* 1988; HOLMES *et al.* 1994). In two cases, the parent loci have been characterized and found to differ by 0.7% (HOLMES *et al.* 1994), suggesting that while these recent transposition events did not arise from a single master element, they arose from two closely related masters.

Possible third lineage splitting at the division of the other two: Within all of the derived phylogenetic trees of the *Peromyscus* LINE-1 sequences, there appears to be a third lineage consisting of Leu1-18 and L1Pm62.

This lineage appears to branch off at about the time Lineage 1 and Lineage 2 diverged (Figures 2 and 3). There are seven shared-sequence sites that unite Leu1-18 and L1Pm62, of which an undefined number may be convergent *in situ* mutations. Leu1-18 and L1Pm62 have more private mutations, insertions, and deletions than the oldest members of Lineages 1 and 2, indicating that both are relics. Because no young elements were sampled, this lineage may have been short lived; however, due to the small sample size of this collection, it can not be formally ruled out that there is still an active lineage. The placement of these sequences would be important in determining the time of the split between the two LINE-1 lineages relative to the peromyscine radiation.

Lineage 1 and Lineage 2 have unequal rates of evolution: Shared-sequence variants indicate changes that occur in the master template for that lineage. For a lineage to persist, each master template must replicate its successor (HARDIES and RIKKE 1990). Because replication of a successor involves an RNA intermediate that is reverse transcribed into DNA, a process that has a high error rate, the number of changes in the sequence is greater than would be accounted for from the usual neutral mutation rate. The assumption is that the more frequently the master template is replaced, the more shared-sequence variants are acquired in the lineage. Lineage 1 LINE-1 sequences have acquired almost twice as many shared-sequence variants as Lineage 2 LINE-1 sequences. The higher number of shared-sequence variants in Lineage 1 relative to Lineage 2 could suggest that Lineage 1 has replaced its master template much more frequently than Lineage 2. Alternatively, the reverse transcriptase may be more error prone in Lineage 1. Implicit in either explanation, and in the general hypothesis of a master lineage, is the assumption that the LINE-encoded proteins are acting largely in *cis*.

The increased number of shared-sequence variants in Lineage 1 does not necessarily indicate a general increase in the number of copies produced. Lineage 1 master templates may produce more total LINE-1 copies, or they may be more efficient at making masters without necessarily making more pseudogene copies than the Lineage 2 master templates. The greater number of Lineage 1 elements isolated compared with Lineage 2 elements could reflect a real difference in copy number in each lineage or could be a sampling artifact. (Note that the elements with names marked by an asterisk were not randomly chosen.)

There are mechanisms other than those involving reverse transcription that could result in an apparent unequal rate of mutation in the two lineages. One suggestive aspect of these data is that 40% of the shared-sequence variants and 60% of the shared-amino acid replacements occur within the first 25% of the sequenced region of Lineage 1. This could be explained by recombination, gene conversion, diversifying selection, or random variation. It is important to note, how-

ever, that the remaining shared-sequence variants also support the existence of two lineages. Resolution of the mechanism accounting for the unequal rate of evolution will require characterization of Lineage 1 elements closer to the split between the lineages to determine if these shared-sequence variants can be sequentially ordered or occurred simultaneously.

The inherent biases of different sequence selection strategies: The two traditional methods of acquiring LINE-1 elements for phylogenetic analysis include: isolating clones from genomic libraries and cloning hybridizing restriction fragments. Depending on the probe and stringency used, screening a genomic library should give a collection of elements that reflects the age structure of elements in the genome. If the genome has a large number of relic sequences, it will be difficult to sift through these to reconstruct the history of LINE-1 evolution, especially because at least two relatively young elements are required to define the shared-sequence variants in a lineage. In contrast, cloning a hybridizing restriction fragment enriches for elements that have retained two internal restriction sites, and thus biases against the collection of relic elements. However, if the restriction fragment cloned has a narrow phylogenetic distribution, then it is likely to be enriched for very young elements. Because we wanted to collect elements that would reflect the dynamics of LINE-1 amplification in *Peromyscus*, we chose a strategy that would bias against both the selection of only young elements and the probability of selecting mainly relic elements. Therefore, we selected a restriction fragment that was a major hybridizing band in several phylogenetically diverse species of *Peromyscus* (*leucopus, maniculatus, truei, californicus, difficilis,* and *polionotus*) (KASS *et al.* 1992).

Different methods of element selection can lead to apparently contradictory conclusions. For example, VANLERBERGHE *et al.* (1993) selected LINE-1 elements from libraries constructed from sonicated genomic DNA of two species of voles. They found that none of the 13 elements sequenced had intact ORFs and that the intraspecific pairwise differences were similar to the interspecific pairwise differences. From this they concluded that the vole elements they sampled were between 3 and 13 million years old and that the rate of transposition of any currently active element must be very low. In contrast, MODI (1996) used restriction-site subfamilies defined by genomic Southern blot analysis to examine the activity of LINE-1 in voles. A restriction-site defined subfamily is a subset of elements yielding a discrete fragment in genomic Southern blot analysis because all members of the subfamily share the two defining restriction sites; it usually contains at least a few hundred copies (LEE *et al.* 1996). New subfamilies would arise as master elements acquire shared-sequence variants that change restriction sites. After examining 30 species of voles with eleven restriction endonucleases, MODI (1996) found that almost half of the restriction-site defined subfamilies were unique to a sin-

gle species. This suggests (to us) that LINE-1 has been highly active in the recent evolutionary history of voles. Although it should be noted that the two species of voles used by VANLERBERGHE *et al.* (1993) were not included in MODI's analysis, we believe that the different conclusions of these studies largely reflect a difference in methods.

Previous analysis of LINE-1 in *P. maniculatus* employed a different approach than was used in the present study (KASS *et al.* 1992). KASS *et al.* (1992) isolated elements by screening a genomic library with a *Mus* LINE-1 probe. They subcloned and sequenced two LINE-1s and carried out genomic Southern blot analysis using his characterized *Peromyscus* elements as probes. He concluded that L1Pm55 and L1Pm62 were members of two distinct LINE-1 lineages that separated prior to the divergence of cricetine and murine rodents. In our analysis, L1Pm55 and L1Pm62 appear to be ancestral to the two lineages described in this paper (Figures 2 and 3) and L1Pm62 appears to have arisen within the cricetine radiation. Neither L1Pm55 nor L1Pm62 contains the lineage-specific variants that differentiate between Lineages 1 and 2. Thus the lineages of LINE-1 defined here are distinct from those proposed by KASS *et al.* (1992). L1Pm55 and L1Pm62 may represent two additional lineages; alternatively, these two sequences may both be ancestral to the lineages defined herein and differentiated from them and each other only by age.

The limited number of active templates relative to the production of pseudogene copies: The small number of LINE-1 lineages that survived through long evolutionary periods in *Mus* and *Peromyscus* indicates that only one or a few master templates are producing the majority of copies at any one period of evolutionary time. This raises the question of what is happening to the other potential master elements produced within each generation of successive master templates. Because 10% of the LINE-1 copies in *Mus* are full length, and an estimated 50 of these should be viable (HARDIES and RIKKE 1990), it appears that a very select minority of the potential master templates have produced viable offspring to continue the lineage. The percentage of full-length copies in *Peromyscus* is not known. The question that these data raise is not why there are two LINE-1 lineages in *Peromyscus* but only one in *Mus*, but rather what constrains the number of LINE-1 master templates in these genera of rodents, and whether this constraint is seen in all mammals.

LITERATURE CITED

ADEY, N. B., S. A. SCHICHMAN, D. K. GRAHAM, S. N. PETERSON, M. H. EDGELL *et al.*, 1994 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.

AUSUBEL, F. M., R. BRENT, R. E. KINGSTON, D. D. MOORE, J. G. SEIDMAN *et al.* (Editors), 1989 *Current Protocols in Molecular Biology*, Vol. 1. Green Pub./Wiley-Interscience, New York.

BAKER, R. J., and H. A. WICHMAN, 1990 Retrotransposon *mys* is concentrated on the sex chromosomes: implications for copy number containment. Evolution **44:** 2083–2088.

BRITTEN, R. J., W. F. BARON, D. B. STOUT and E. H. DAVIDSON, 1988 Sources and evolution of human Alu repeated sequences. Proc. Natl. Acad. Sci. USA **85:** 4770–4774.

BROWNELL, E., 1983 DNA/DNA hybridization studies of muroid rodents: symmetry and rates of molecular evolution. Evolution **37:** 1034–1051.

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.

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

CATZEFLIS, F. M., J.-P. AGUILAR and J.-J. JAEGER, 1992 Muroid rodents: phylogeny and evolution. TREE **7:** 122–126.

CLOUGH, J. E., J. A. FOSTER, M. BARNETT and H. A. WICHMAN, 1996 Computer simulation of transposable element evolution: random template and strict master models. J. Mol. Evol. **42:** 52–58.

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.

DEININGER, P. L., and M. A. BATZER, 1993 Evolution of retroposons. Evol. Biol. **27:** 157–196.

DEININGER, P. L., M. A. BATZER, C. A. HUTCHISON III and M. H. EDGELL, 1992 Master genes in mammalian repetitive DNA amplification. Trends Genet. **8:** 307–311.

DOMBROSKI, B. A., A. F. SCOTT and H. H. KAZAZIAN, JR., 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.

EDGELL, M. H., S. C. HARDIES, D. D. LOEB, W. R. SHEHEE, R. W. PADGETT *et al.*, 1987 The L1 family in mice. Prog. Clin. Biol. Res. **251:** 107–129.

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.

HARDIES, S. C., and B. A. RIKKE, 1990 A selfish retrotransposition model for rodent LINE-1, pp 127–134 in *Molecular Evolution: UCLA Symposia on Molecular and Cellular Biology; New Series*, Vol. 122, edited by M. T. CLEGG and S. J. O'BRIEN. Alan R. Liss, New York.

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.

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

HUTCHISON, C. A., III, S. C. HARDIES, D. D. LOEB, W. R. SHEHEE and M. H. EDGELL, 1989 LINEs and related retroposons, pp 593–617 in *Mobile DNA*, edited by D. E. BERG and M. M. HOWE. Am. Soc. Microbiol., Washington, DC.

JURKA, J., 1989 Subfamily structure and evolution of the human L1 family of repetitive sequences. J. Mol. Evol. **29:** 496–503.

JURKA, J., 1995 Origin and evolution of Alu repetitive elements, pp 25–41 in *The Impact of Short Interspersed Elements (SINES) on the Host Genome* edited by R. J. MARAIA. R. G. Landes Company, Austin, TX.

JURKA, J., and T. SMITH, 1988 A fundamental division in the Alu family of repeated sequences. Proc. Natl. Acad. Sci. USA **85:** 4775–4778

KASS, D. H., F. G. BERGER and W. D. DAWSON, 1992 The evolution of coexisting highly divergent LINE-1 subfamilies within the rodent genus *Peromyscus*. J. Mol. Evol. **35:** 472–485.

KAZAZIAN, H. H., JR., C. WONG, H. YOUSSOUFIAN, A. F. SCOTT, D. G. PHILLIPS *et al.*, 1988 Haemophilia A resulting from *de novo* inser-

tion of L1 sequences represents a novel mechanism for mutation in man. Nature **332:** 164–166.

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

KING, J. A., 1968 *Biology of Peromyscus (Rodentia).* American Society of Mammalogists, Provo, UT, 593 pp.

KIRKLAND, G. L., JR, and J. N. LAYNE, 1989 *Advances in the Study of Peromyscus (Rodentia).* Texas Tech University Press, Lubbock, TX. 366 pp.

LABARCA, C., and K. PAIGEN, 1980 A simple, rapid, and sensitive DNA assay procedure. Anal. Biochem. **102:** 344–352.

LEE, R. N., J. C. JASKULA, R. A. VAN DEN BUSSCHE, R. J. BAKER and H. A. WICHMAN, 1996 The retrotransposon Mys was active during evolution of the *Peromyscus leucopus-maniculatus* complex. J. Mol. Evol. **42:** 44–51.

LEEFLANG, E. P., W. M. LIU, C. HASHIMOTO, P. V. CHOUDARY and C. W. SCHMID, 1992 Phylogenetic evidence for multiple Alu source genes. J. Mol. Evol. **35:** 7–16.

LOEB, D. D., R. W. PADGETT, S. C. HARDIES, W. R. SHEHEE, M. B. COMER *et al.,* 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.

LONGMIRE, J. L., A. K. LEWIS, N. C. BROWN, J. M. BUCHINGHAM, L. M. CLARK *et al.,* 1988 Isolation and molecular characterization of a highly polymorphic centromeric tandem repeat in the family Falconidae. Genomics **2:** 14–24.

MADDISON, W. P., and D. R. MADDISON, 1992 MacClade: analysis of phylogeny and character evolution. Version 3.0. Sinauer Associates, Sunderland, MA.

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.

MILES, C., and M. MEUTH, 1989 Hamster line and ALU-equivalent sequences are present in the small polydispersed circular DNA population of CHO cells. Nucleic Acids Res. **17:** 5846.

MODI, W. S., 1996 Phylogenetic history of LINE-1 among arvicolid rodents. Mol. Biol. Evol. (In press).

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 ancient rodent L1 DNA family. J. Mol. Evol. **36:** 9–20.

PRICE, D. K., J. A. AYRES, D. PASQUALONE, C. H. CABELL, W. MILLER *et al.,* 1992 The 5' ends of LINE-1 repeats in rabbit DNA define subfamilies and reveal a short sequence conserved between rabbits and humans. Genomics **14:** 320–331.

RIKKE, B. A., L. D. GARVIN and S. C. HARDIES, 1991 Systematic identification of LINE-1 repetitive DNA sequence differences having species specificity between *Mus spretus* and *Mus domesticus.* J. Mol. Biol. **219:** 635–643.

ROGERS, J. H., 1985 The origin and evolution of retroposons. Int. Rev. Cytol. **93:** 187–279.

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

SANGER, F., S. NICKLEN and A. R. COULSON, 1977 DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:** 5463–5467.

SHEN, M. R., M. A. BATZER and P. L. DEININGER, 1991 Evolution of the master *Alu* gene(s). J. Mol. Evol. **33:** 311–320.

SKOWRONSKI, J., and M. F. SINGER, 1986 The abundant LINE-1 family of repeated DNA sequences in mammals: genes and pseudogenes. Cold Spring Harbor Symp. Quant. Biol. **51:** 457–464.

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

SWOFFORD, D. L., 1990 PAUP: phylogenetic analysis using parsimony (PAUP), Version 3.0s. Illinois Natural History Survey, Champaign.

VANLERBERGHE, F., F. BONHOMME, C. A. 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.

WICHMAN, H. A., S. S. POTTER and D. S. PINE, 1985 Mys, a family of mammalian transposable elements isolated by phylogenetic screening. Nature **317:** 77–81.

WILLARD, C., T. NGUYEN and C. W. SCHMID, 1987 Existence of at least three distinct *Alu* subfamilies. J. Mol. Evol. **26:** 180–186.