The Evolution of Coexisting Highly Divergent LINE-1 Subfamilies Within the Rodent Genus *Peromyscus*

D.H. Kass, F.G. Berger, and W.D. Dawson

Department of Biological Sciences, University of South Carolina, Columbia, SC 29208, USA

Summary. Two distinct members of the LINE-1 (L1) family in *Peromyscus* were characterized. The two clones, denoted L1Pm55 and L1Pm62, were 1.5 kb and 1.8 kb in length, respectively, and align to the identical region of the L1 sequence of *Mus domesticus*. Sequence similarity was on the order of 70% between L1Pm55 and L1Pm62, which approximates that between either *Peromyscus* sequence and *Mus* L1. L1Pm62 represents a more prevalent subfamily than L1Pm55. L1Pm62 exists in about 500 copies per haploid genome, while L1Pm55 exists in about 100 copies. The existence of major and minor subpopulations of L1 within *Peromyscus* is in contrast to murine rodents and higher primates, where L1 copy number is on the order of 20,000 to 100,000, and where levels of intraspecific divergence among L1 elements are typically less than 15–20%. Additional *Peromyscus* clones are similarly divergent from both L1Pm62 and L1Pm55, implying the existence of more than two distinct L1 subfamilies. The highly divergent L1 subfamilies in *Peromyscus* apparently have been evolving independently for more than 25 million years, preceding the divergence of cricetine and murine rodents. Investigations of the evolution of L1 within *Peromyscus* by restriction and Southern analysis was performed using species groups represented by the partially interfertile species pairs *P. maniculatus-P. polionotus, P. leucopus-P. gossypinus,* and *P. truei-P. difficilis* of the nominate subgenus and *P. californicus* of the *Haplomylomys* subgenus. Changes in L1 and species group taxonomic boundaries frequently coincided. The implications for phylogeny are discussed.

Key words: LINE-1 (L1) — *Peromyscus* — Repetitive elements — Molecular drive

Introduction

Mammalian LINEs are long interspersed repetitive sequences usually greater than 5 kilobases (kb) in length. They are generally found in about 10^4 copies per genome (Singer 1982), though 3' regions have been found approximating 10^5 copies (Gebhard et al., 1982; Hwu et al., 1986). This phenomenon is due to truncation of the elements at the 5' end (Voliva et al. 1983). L1 elements, as a unit, exhibit changes that correspond to mammalian phylogeny (Burton et al. 1986), attributed principally to proliferative transposition to novel sites, presumably accompanied by clearance of preexisting sequences (Casavant et al. 1988). Additional properties and suggested mechanisms of propagation of LINEs are reviewed by Hutchison et al. (1989).

Partial sequences of LINEs demonstrate a strong conservation in the mammalian orders Carnivora, Lagomorpha, Rodentia, and Primates (Fanning and Singer 1987). By sequence analysis of L1 elements in *Mus* species, Martin et al. (1985) observed less

Present address and offprint requests to: D.H. Kass, Department of Biochemistry and Molecular Biology, University of Texas M.D. Anderson Cancer Center, 1515 Holcombe, Houston, TX 77030, USA

than 5% divergence among elements within a species, and 5.4–9% divergence between species, indicating extensive maintenance of intraspecies homology. Although L1 is highly conserved in the genus *Mus,* concerted changes in restriction sites accumulate with time (Jubier-Maurin et al. 1985). Dover (1982) proposed the theory of molecular drive, a collective term for the processes involved, based on the correlation of concerted changes in repetitive elements with species boundaries in diverse organisms. A molecular phylogeny led Hardies et al. (1986) to conclude that most L1 members in *Mus* are pseudogenes; a few functional L1 members, termed molecular drivers, e.g., source genes, give rise to new elements.

Minor variants of L1 sequences, indicated by the presence of novel restriction fragments on Southern blots, imply the existence of L1 subfamilies within the mammalian genomes (Jubier-Maurin et al. 1985). Subfamilies are also suggested from 5' motifs in *Mus,* known as A and F (Shehee et al. 1987). These 5' tandem repeats are similar in organization but unrelated in sequence (Padgett et al. 1988). Skowronski and Singer (1986) and Jurka (1989) have defined subfamilies of the human L1 on the basis of diagnostic nucleotides shared among a subpopulation of elements. These diagnostic base pairs constitute less than 3% of the bases within the compared sequences.

More recently, Pascale et al. (1990) identified an L1-related subfamily (Lx) in murine rodents. This subfamily is maintained in high copy number (>60,000) and is 15% divergent within a region of the open reading frame. Extensive amplification of Lx may have predated the murine radiation (Pascale et al. 1990).

The taxonomically well-characterized cricetine rodent genus *Peromyscus* (Osgood 1909; Hooper 1968; Carleton 1989) provides a singular opportunity to further examine the evolution of LINEs and its role, if any, in speciation. This genus is traditionally classified into two or more subgenera with numerous species groups containing over 50 species (Hall 1981). The nominate subgenus (*Peromyscus*) contains about 45 species further classified into 11 species groups. Members of a species group sometimes have the capacity to interbreed in captivity (Blair 1943), although they do not hybridize in nature (Dice 1968). Species assigned to different species groups or subgenera are probably reproductively isolated from one another, since all attempted crosses between individuals of separate species groups have been unsuccessful, but subspecies within a species are generally interfertile in captivity (Dice 1968). Correlation between changes in L1 and taxon boundaries would be of interest.

Materials and Methods

Specimens. Deermice (*P. maniculatus*), oldfield mice (*P. polionotus*), and white-footed mice (*P. leucopus*) were obtained through the *Peromyscus* Genetic Stock Center, University of South Carolina. Two subspecies of *P. californicus* were donated by Susan Hoffman, Mammal Division, Museum of Zoology, University of Michigan. Pinyon mouse (*P. truei*) and rock mouse (*P. difficilis*) livers were obtained courtesy of T.L. Yates and J. Cook of the Museum of Southwestern Biology at the University of New Mexico. Wild-caught cotton mice (*P. gossypinus*) and their first-generation progeny were used. Laboratory house mouse (*Mus domesticus*) of the C57BL strain were obtained through the University of South Carolina Animal Resource Facility.

Isolation of Peromyscus L1 Sequences. A partial *Sau*3A/*Bam*HI *P. maniculatus* lambda Charon 30 genomic library (provided by Dr. M. Edgell; Padgett et al. 1987) was screened with a MIF-1 probe (provided by Dr. M. Edgell; Voliva et al. 1984). Two randomly selected hybridizing phages were purified; the DNA was isolated using a DE52 resin (Whatman) slurry (Benson and Taylor 1984), digested with *Eco*RI, and analyzed by Southern hybridization to MIF-1. The series of *Eco*RI fragments, as observed on an ethidium-bromide-stained agarose gel, inclusive of the single L1-hybridizing fragment per phage clone (detected by autoradiography), were subcloned by random insertion into the *Eco*RI site of plasmid pT7/T3-18 (BRL, Inc.), followed by verification of isolated L1-containing clones using Southern hybridization.

DNA Analysis. Southern blotting was performed by standard protocols (Southern 1975). Genomic DNA was isolated by modification of the method of Flamm et al. (1966) incorporating proteinase K digestion and phenol-chloroform extractions described by Blin and Stafford (1976). Livers were removed from animals fasted for 24 h prior to sacrifice by cervical dislocation and were homogenized in 10 ml cold buffer (100 mM EDTA, 1% SDS, 10 mM Tris pH 8.0). Spleens from adult and livers from immature *P. californicus* were utilized, since liver DNA from this species exhibits excessive degradation upon isolation (Kass and Hoffmann unpublished). Genomic DNA (10 μL) was cleaved by incubations with excess restriction enzyme (*Eco*RI, *Hind*III, *Bgl* II, or *Xba*I) 12–16 h in 40 μl reaction mixtures (Maniatis et al. 1982). Digested DNA was concentrated by ethanol precipitation and fragments were separated by electrophoresis on 1% agarose gels in 1 × TBE (89 mM Tris-borate, 89 mM boric acid, 2 mM EDTA) for 19–20 h at 8–10 V. For Southern analysis, 6 μg of DNA was cleaved and subjected to gel electrophoresis in 1 × TAE buffer (400 mM Tris-acetate, 1 mM EDTA); gels were denatured in 1.5 M NaCl, 0.5 M NaOH and neutralized in 1 M Tris-HCl, 1.5 M NaCl pH 5.0 for 1 h each and DNA was transferred to Hybond (Amersham) nylon filters. Prehybridization of filters was done in 4 × SSC, 0.2% SDS, (0.1 M NaH$_2$PO$_4$, 10 mM sodium pyrophosphate pH 6.5), herring sperm DNA (200 μg/ml), and 10 × Denhardt's solution for at least 1 h at 65°C. Filters were hybridized to approximately 2 × 10^6 cpm of denatured probe in 4 × SSC, 0.2% SDS (0. 1 M NaH$_2$PO$_4$, 10 mM sodium pyrophosphate pH 6.5), herring sperm DNA (200 μg/ml), and 1 × Denhardt's solution at 65°C. Filters were washed in 2 × SSC (or 1 × SSC for quantitative Southerns), 0.2% SDS three times for 1 h each at 65°C. Filters were placed in autoradiography cassettes with Kodak XAR film. Hybond filters were rehybridized after stripping the first probe in a basic solution (0.2 N NaOH, 0.1 × SSC, 0.1% SDS) at 65°C. DNA sequencing was done using Sequenase 2.0 (US Biochemical). Nested deletions for sequencing

were generated as described by Henikoff (1984). Sequences were analyzed using programs from the University of Wisconsin Genetics Package (Devereux et al. 1984). Evolutionary rates were estimated as described by Kimura (1977).

Results

Sequence Analysis of Isolated Peromyscus L1 Clones

Two clones, pDK62 and pDK55, were isolated from a *Peromyscus maniculatus* genomic library with a MIF-1 probe and subcloned into plasmid vectors; they contain L1 elements within *Eco*RI fragments of 1.8 kb and 1.5 kb, respectively. The individual elements were designated L1Pm62 and L1Pm55, in accordance with conventional nomenclature (Voliva et al. 1983). Sequences of the L1 elements L1Pm62 and L1Pm55 were obtained to characterize and determine their relationship to each other and to L1 elements of *Mus domesticus* (L1Md-A2; Loeb et al. 1986), *Rattus norvegicus* (L1Rn; Soares et al. 1985), and the human (L1Hs; Skowronski et al. 1988). The two *P. maniculatus* fragments are homologous to the MIF-1 region of L1Md, L1Rn, and L1Hs (Fig. 1A). The homology is 76% between L1Pm62 and L1Md, 72% between L1Pm55 and L1Md, and 71% between the two *Peromyscus* sequences. This demonstrates that L1Pm55 and L1Pm62 are not adjacent regions of L1 (Fig. 1B), but rather are members of distinct L1 families in *P. maniculatus*. A comprehensive search (GCG-wordsearch) through the GenBank data base confirmed that these fragments are more closely related to L1 elements of other species than to any other DNA sequence. Sequence comparisons indicate neither fragment exhibits homology to other transposonlike elements, such as the THE-1 element in human (Paulson et al. 1985) or the Mys element in *P. leucopus* (Wichman et al. 1985).

Analysis of Prevalent Repetitive Sequences in Peromyscus

The 1.8-kb fragment (L1Pm62) was anticipated, as *Eco*RI-cleaved genomic DNA from seven *Peromyscus* species yield a band this size amongst a smear of DNA observed by agarose gel electrophoresis (Fig. 2A). The previously known 1.35-kb *Eco*RI highly repetitive L1 fragment (MIF-1) in *M. domesticus* was clearly recognized and more intensely staining as compared to the *Peromyscus* fragment indicative of higher copy number. It is also likely that L1Pm62 represents a more populous L1 sub-

```
L1Md   GAGTTCTATC AGACCTTCAA AGAAGATCTA ATTCCAATTC TGCACAAACT ATTTCACAAA
L1Rn   ..A....... .A......T .......C..C ..A.....AT .ATC...... ...C......
62     ..A....C. T..AT..... ......GT.. ..A...T.A. .C.T....T. G..C....C.
55     ..A....... ..G....-.. ......AT.. .C....T.A. .C.T....T. ...A..A...
L1Hs   ..A....C. ...GG.A... G..G..A..G G.A...T.CT .T.TG..... ...C..ATC.

L1Md   ATAGAAGTAG AAGGTACTCT ACCCAACTCA TTTTATGAAG CCACTATTAC TCTGATACCT
L1Rn   ..T...AC.. .T..AG.A.. ...G..T..C ..C......T ....A..... ...T......
62     G.....AC.. ....A..AT. G..A.....T ..CC....G. .T..AG.... ......T..C
55     ......ACT. ....A..AT. GTAT..T..T ......T... ...A.....T C..A.A.A..
L1Hs   ......AA.. .G..A.TC.. C..T...... ........G. ...GC..C.T .........A

L1Md   AAACCACAGA AAGAT---CC AACAAAGATA GAGAACTTCA GACCAATTTC TCTTATGAAT
L1Rn   C.......C. .T..A---.. ........A. .......... .......... C.........
62     ........C. .....GCAA. ......A.A. ..A..T.A.. ..A...C.. A..C...G.G
55     .........-- ---------- ----..T.A. .GA..T.AA. ......G...G C..T.....G
L1Hs   ..G..GGGC. G...C---A. ...C..A.A. .....T..T. ........A.. CT.G.....C

L1Md   ATCGATGCAA AAATCCTCAA TAAAATTCTC GCTAACCGAA TCCAAGAACA CATTAAAGCA
L1Rn   .....C.... ....A..... ....C....- .GA....... ........G. ..A....A..
62     G.T....... ..G.A..... ......A..G ..A....A.. ........TT. T.-....AA.
55     ..A.C..G.. T...T..... C...TACT.A AAA....A.. .T...ACG.. T.--...AAG
L1Hs   ..T....... .......... ......A..G ..A....A.. ....GC.G.. ...C...AAG

L1Md   ATCATCCATC CTGACCAAGT AGGTTTTATT CCAGGGATGC AGGGATGGTT TAATATACGA
L1Rn   .........C. A...T..... ...C..C..C T....C.... .......... .-....T.G
62     .........A A...T..... T..C...... C....A...T. .A....T... C..C..GT..
55     ..T....... A...T..... TA.CA..... ..T..A..CAT G......... C..C..GG.T
L1Hs   C.T.....C. A...T..... G..C..C..C ..T....... .A..C..... C........C

L1Md   AAATCCATCA ATGTAATCCA TTATATAAAC AAACTCAAAG --ACAAAAAC CACATGATCA
L1Rn   .C.A...... .C..G..... .........A .....G..... --.AC..... ..........
62     .....T.... .--....T.. .C.C.C.... .......G... AA.A.....A ..........
55     .....A.... ........GA. CC.....T.. .....G.... ---.C....T G.T...G...
L1Hs   .....A..A. .......... GC........ .G.GC..... --......... ........TG

L1Md   TCTCGTTAGA TGCAGAAAAA GCATTTGACA AGATCCAACA CCCATTCATG ATAAAAGTTT
L1Rn   .T..A..... C..T..G... ......... ...T...... ...C...... .........CC
62     ....AA.... CA.TA..G.. ..C....... .A......A. T..C...... .....GA.CC
55     C...A..... .........G A.CC..TG.. .A........ .T.C...C.. .---------
L1Hs   ....AA.... ......... ..C....... ..A..T.... A..C...... C.....AC.A

L1Md   TGGAAAGATC AGGAATTCAA GGCCCATACC TAAACATGAT AAAAGCAATC TACAGCAAAC
L1Rn   ........AT ........-. ....A..... ...--..AG. ......C..A ..........
62     ....G.A... ......A..G ..AA...... .......A.. ...G....T .........G.
55     -----...GT T..T..A.C. A.GA.....A .TG....A.. ...GATG.CT .....A..T.
L1Hs   .CA.T.A.AT ...T...G.T ..GATG..TT .C..A..A.. ..G...T... ..TGA....A

L1Md   CAGTAGCCAA CATCAAAGTA AATGGAGAGA AGCTGGAAGC AATCCCACTA AAATCAGGGA
L1Rn   ....T..T.. ...T...C.. .......... ..A..T.... .......... ......G....
62     ..AC.A..-- A..A...T.. ......A... ..AA.CA...T G..T...... ......TAA.
55     .CAC...... ..G...CC.. ....T..... .A.CCA.... .T.T....C. .....A..A.
L1Hs   .CAC...... T....T.C.G .....GC.A. .A........ .T....TT.G ...A.C..C.

L1Md   CTAGACAAGG CTGCCCACTT TCTCCCTACC TTTTCAACAT AGTACTTGAA GTATTAGCCA
L1Rn   .......... .........C .........T .A.....T.. ...TT..... ..TC......
62     .A...-.... ...TG....C ..C.TA.... .A.....T.. ....A.C... ..TC...T.
55     .AG.....AA A..T...... .....A.... .A.....T.. .C........ ...C....A.
L1Hs   .A.....G.. A.....T..C ..A..GCT.. .A........ ...GT.G... T.TC.G....

L1Md   GAGCAATTCG ACAACAAAAG GAGATCAAGG GGATACAAAT TGGAAAAGAG GAAGTCAAAA
L1Rn   .......CA. .......G... ...G.G...T ........G.. ........--- ..........
62     A......AAA .......... ....T...T .G........ G.....A... ..........C
55     ..A....AA. .....TG... ....C..... .......TG..C A.....G.-- --.......G
L1Hs   .G.....CA. G..GG.G... ..A..A.... .T..T...T. A......... .........T
```

A

Fig. 1. Alignment of cloned *Peromyscus* L1 sequences to other mammalian LINEs. A Sequence alignment of L1Md-A2 (*Mus*) beginning with nucleotide 4310 (Loeb et al. 1986), the 1.8-kb *Eco*RI insert of pDK62, and the 1.5-kb insert within pDK55, rat L1Rn (Soares et al. 1985) and human L1Hs (Skowronski et al. 1988). *Eco*RI recognition sequences are underlined. Dots indicate nucleotides identical to L1Md. Dashes are inserted for maximal alignment. Orientation is from 5' to 3' based on the open reading frame in L1Md. Continued on next page.

family compared with L1Pm55. A variation was observed between members of the *maniculatus* and *leucopus* species groups when DNA was digested with *Bam*HI and analyzed (Fig. 2b). A 1.7-kb repetitive fragment was observed in *P. leucopus* and *P. gossypinus* which was not seen in either species from the *maniculatus* group, demonstrating a homogenous variation of a repetitive fragment had occurred in the ancestry of one or the other species group.

```
L1Md    TATCACTTCT TGCAGATGAT ATGATAGTAT ATATAAGTGA CCCTAAAAAT TCTACCAGAG        L1Md    TAAAAGAACT TCTGGCGGAA TCACCATGCC AGACCTAAAG CTTTACTACA GAGCAATTGT
L1Rn    .......AT. .......... .....T.... ..T...-... T..C....G. ..C.......        L1Rn    C.....G... ..A..G.... ....T..C.. T..A..C... .AG..T.... .......A..
62      .T.TG..AT. ..G...A... ......A.-. .C.C..... T..-.C.... .........G.        62      ....GC.... ...AA..... ...C.. T...A.C... ..C.....T. T...T..A..
55      ....T..A.. .......... .....TA... .C....A.T. .......... ...CT....GA        55      .......... G....A..-A .......C.. CA....C..A T.G....... ...T.TG.AC
L1Hs    .G..C..GT. .........C .....T..TG .C,.GAAA. ...C.TCGTC ..AG..CA.A        L1Hs    A..G.AC.AA G....A..C. ....AC.A.. T...T.C..A ..A....... AG..T.CA..

L1Md    AACTCCTAAA CCTGATAAAC AGCTTCGGTG AAGTAGTTGG ATATAAAATA AACT--CAAA        L1Md    GATAAAAACT GCATGGTACT GGTAT-AGAG ACAGACAAGT AGACCAATGG AATAGAATTG
L1Rn    ....A..... G......... .A....A.CA ....G.C... G.........T ....--....        L1Rn    ...-...... ......T... ....C-.... ....GA. ....T..... ..C......
62      TTTAT...-- -......... TC....A..A .C..G.CA.. ...C..G..T ....AAA....        62      A........A ..T....... A.C..A.A.A ..C....TA. ....A..... ...CA..C..
55      ....T...C. GTC....... .TG...A.AA ......CA.. ..TA.....T ...-..A..C.        55      .G..------ --------. T.GCAT.A.A ..T....CA. ...T.G.... ...T.T....
L1Hs    .T.....T.. G........G. .A...A.CA ..N.CTCA.. ...C.....C ..TG--T.C.        L1Hs    A.CC.....A .......... ...C-CA.A ....G.TA. ...T.C.... .......CA.

L1Md    CAAGTCAATG GCCTTTCTCT ATACAAAGAA TAAACAGGCT GAGAAAGAAA TTAGGGAAAC        L1Md    AAGATCCAGA AATGAACCCA CACACCTATG GTCACTTGAT CTT------- -CGACAAGGG
L1Rn    T..A.A.G.A .....C.... .C....AG. G.C.G.A... .-......G. .A......        L1Rn    ....C..... .......... .........G. ......... T..------- -T.....A..
62      A..A...G.A ...A.C..A. .C....TG. ....-..... .......... .C..A....        62      ......GTA. C..T..TT.. .........C. AA..TC.... T..------- -T.....A.A
55      A..AATGG.A ......C.AA. .T....TC. .......... .......... ...CT..A....        55      .......... ..A.GTA.. ..T....... .G....C..G. T..GATTAAA AAA.A..AAA
L1Hs    A..A...CAA ..A..C..A. .C..C..C.. C.G...AA.A .....GCC... .C.T.AGTGA        L1Hs    .GCCCT.... ...A.CGNNG .TT.....CA ACT.TC..... ...------- -T.....ACC

L1Md    AACAC----- ---------- ---------- ---------- ---------- ----------        L1Md    AGCTAAAACC ATCCAGTGGA AGAAAGACAG CATTTTCAAC AATTGGTGCT GGCACAACTG
L1Rn    ...T.----- ---------- ---------- ---------- ---------- ----------        L1Rn    ...C...... ...A.A.... .A.....T.. .........G. ..A....... ..TT......
62      .T...AT--- ---------- ---------- ---------- ---------- ----------        62      ...C.....T G.A..A.... .A.....A.. ...C...... ..A....A.A ..T.T.....
55      ....TCTCTC AGAGACCTTG CCTTGGAGAA GGTGGGTTGG AGGAAGGCTG AGGGGCAGGA        55      ...C.G..AT .CA..A.T.. .A.....A.. ....C..... ...A.AC.... ..TCA......
L1Hs    .CTC.----- ---------- ---------- ---------- ---------- ----------        L1Hs    T.AG....A. .AG..A...G GA..G..TTC .C.A..T..T ..A....... ..G.A.....

L1Md    ---------- ---------- ---------- ---------- ---------- ----------        L1Md    GTTGTTATCG TGTAGAAGAA TGCGAATCGA TCCATACTTA TCTCCTTGTA CTAAGGTCAA
L1Rn    ---------- ---------- ---------- ---------- ---------- ----------        L1Rn    .AG..C.A.A .......... ...A...A.. ..T..G.... ..A..C.... .A...C.T..
62      ---------- ---------- ---------- --------T TACAATATTT ACAATTATAA TATTTACAAT      62      .A.------. .C.......T ...A...A.. ....---C.. ..AT.A..C. GA..AC....
55      GAAGAGAAGA CAGAAGAATC TGTGGTTGGT ATATAAAATG AATAGAAAAA TCTCTTTAAT        55      .A.------A GC..C.T.T. AAA....TC
L1Hs    ---------- ---------- ---------- ---------- ---------- ----------        L1Hs    .C.AGCCATA .......AGC ..AA.G.G.. ...C.T.C.T A.A.....A.. .A..AA....

L1Md    -------CC TTCTCAATAG TCACAAAT-A ATATAAAATA TCTTGGCGTG ACTCTAACTA        L1Md    ATCTAAGTGG ATCAAGGAAC TTCACATAAA ACCAGAGACA CTGAAACTTA TAGAGGAGAA
L1Rn    --------. ...AT..A.. A.C....... .......G.. C..C..T... ..AA....C.        L1Rn    G..C...... ........C. .C..T..... ......T... ..C....CA. ....A..A..
62      ATTACAAA.A ..TA...... C.......-. .------... CT....G..A .......TC.        62      ......... ...A..C. GCA....... T...CCT.... .....C..G. ..........
55      AATAATAA.A ACAA.....A ..A.T....G. .A-----.A. C.....GA.A .C..-----.        55      T..A.GA... ..T..A..CT .AA..G.T.G ...TA.A..C A.A...ACCC ....A..A..
L1Hs    --------.A ..A....T. CTT....G-. GA........ C..A..AA.C CAA..T..A.        L1Hs    

L1Md    AGGAGGTGAA AGATCTGTAT GATAAAAACT TCAAATCTCT GAAGAAAGAA ATTAAAGAAG        L1Md    AGTGGGGAAA AGCCTTGAAG ATATGGGCAC AGGGGAAAAA TTCCTGAACA GAACAGCAA-
L1Rn    ..T.A.A.. ....T....C A.....G.... ....GA..A. .........G.....        L1Rn    .C.A.....G CAT..G...C .C........ T..AA....T .......... A....C...-
62      .AC.A..... ...CC...... ..C..G.... .T.....CT. .......... ...G.....        62      .A.A..A.GT .....A...A GC..TC.... .T.A..TC.C C.......T. T....C...G
55      .CC.A..... ...GC--A... A......A .T..GA.A.. ..T.G.C. ...G.....        55      
L1Hs    G...T..... G..C..C.TC A.GG.G..... A....C.A.. .CTC..G.... ..A.....G.        L1Hs    CC.A..C.TT .C.A..C.G. .C..A...GT G..CA.GG.C ...A..TC.. A....C...-

L1Md    ATCTCAGAAG ATGGAAAGAT CTCCCATGCT CATGGATTGG CAGGATCAAC ATTG--TAAA        L1Md    TGGCTTGTGC TGTAAGATCG AGAATCGACA AATGGGACCT AATGAAACTG CAAAGTTTCT
L1Rn    TC..AT.... .......... ..G.....A. ...A...... G....AT..T ....-....        L1Rn    ......A.AT .C.......A ...G.A.... .......T.. C..A....... ...C....
62      ..A....... ....A. ..T...... ..A....A. T..A.CTT.. ...-AA....        62      .A..ACAGA. AC.G...GT. .T...TA.T. .......T.. CC.A...... AG...C....
55      .CA.TG..CA .......... ........C. .......... T....TT..T G...TAA...        L1Hs    AA..AATG.. AAC..A.G.C .A...T.... .......T.. ...T.....A A.G..C....
L1Hs    ..ACA.A..A ......GA.C ANT....... .....G.A. A..A....T ..C.--.G..

L1Md    AATGGCTATC TTGCCAAAAG CAATCTACAG ATTCAATGCA ATCCCCATCA AAATTCCAAC        L1Md    GCAAGGCAAA AGACACTGTC AATAAGACAA AAAGACCACC AACAGATTGG GAAAGGATCT
L1Rn    ......C..T ..A....... .C........ .......T.. .......... ....A....T        L1Rn    .T........ G.........G GT..G..... .C.G.A..... .......... ....A.....
62      ......A... ..A....... .........A ..G....... ...T...... .........        62      .T...A.... G.....AA.A .......... G.....AG.. T.T...A.... .G..A.....
55      ..ACC.C..T C.A....... .......... .....G.... .GT.TT..A. .....T...T        L1Hs    ...CA..... ...A...AC. .TC.GAGTG. .C..G.A... T...ACA... ...G.CA..T.
L1Hs    ......C..A C.....C..G. T...T..... .........C .......... .GC.A....T

L1Md    TCAATTC-TT CAACGAATTG GAAGGAGCAA TTTGCAAATT TGTCTGGAA- ----------        L1Md    TTACCTATCC TAAATCAGAT AGGGGACTAA TATCCAACAT ATATAAAGAA CTCAAGAAGG
L1Rn    A.......-.. A..A..G..A ..CA..A... .......... CA..A...G- ----------        L1Rn    .....A.... ..C.A..... ..A..C..T. .......A.. ...C...... .......T.T
62      A......T.. ..CA.TC..C A..A..A... ..ACT...C. AA.A.----- ----------        62      .C...A.... C.C...T..C ..A..G..G. .C.....A.. .......... .A.-------
55      GTG...-... ..CA...A.T ...ATGAT.. ...T..GC.. CA.G.....A CAGAGAGAAA        L1Hs    .C..AACCTA CTC...T..C .AA..G.... ......GA.. C..C..T.... ...AC.AA
L1Hs    GACT...-.. ..CA...... ...AA.A.T. C..TA..G.. CA.A.....- ----------

L1Md    ---------- ---------- -TAACAAAAA ACCTAGGATA GCAAAAAGTC TTCTCAAGGA        L1Md    TGGACTTCAG AAAATCAAAT AACCCCATTA AAAAATGGGG CTCAGAACTG AACAAAGAATTC
L1Rn    ---------- ---------- -.......... ...C...... ..T....C.A .C.....CA.        L1Rn    .A....G... GG.GA..... ...G.T.... ....G..... T.....G..A ......
62      ---------- -------AGA AAT...G... .......... T.T....CAA .CA.TT.CA.        62      .A...AA..A ..T.C....C .GT..A...T .........C TA....T... ......
55      GGGAGGGGGG GGGAGGGAGA G.G.G.G... T.TC...... ...T..C.C.T .C..G..TA.        L1Hs    .TT..AAG.A ....A....C .........C. ....G....C GAAG..CA.. ....G.C.C...
L1Hs    ---------- ---------- -CC.A....G .G.CC.C..T ..C..GTCAA .C..A.GCC.
```

Fig. 1A. Continued.

Fig. 1B. Positional relationship to cloned *Peromyscus* L1 sequences to a full-length *Mus domesticus* L1 element (L1Md-A2; Loeb et al. 1986). Boxes on the left refer to a 208-bp tandemly repeated sequence. ORF = open reading frame, B = *Bam*HI restriction site, E = *Eco*RI restriction site

Copy Number of L1 Subfamilies

Subfamily copy number was determined in several *Peromyscus* species on Southern blots of *Eco*RI-cleaved DNA; band intensities were compared to standards of known copy number (Fig. 3A). Approximately 500 copies of the L1Pm62 subfamily exist per haploid genome, as demonstrated by hybridization to pDK62, which identified a 1.8-kb fragment on the blots. When the same filter was

hybridized to pDK55 (Fig. 3B), 1.8-kb and 1.5-kb bands were identified. The 1.8-kb fragment observed with the pDK55 probe represents cross-hybridization to the L1Pm62 subfamily as evidenced by washing a filter of EcoRI-digested DNA, probed with pDK55, under increasing stringencies. A proportionately greater loss of probe that had hybridized to the 1.8-kb fragment was observed relative to the 1.5-kb fragment (Fig. 4). The 1.5-kb fragment was estimated to be about 100 copies per haploid genome within members of the P. maniculatus species group (Fig. 3B). In other Peromyscus species, a less-intense band (0.8 kb in P. leucopus and P. gossypinus, 1.6 kb in P. truei and P. difficilis) is seen by hybridization to pDK55 (Fig. 3B) indicative either of low copy number (approx. 8–9) of a highly homologous fragment or a greater number of copies of a more divergent fragment. A weak signal representing the 1.35-kb MIF-1 fragment of Mus domesticus was observed with the pDK62 and pDK55 probes (Fig. 3).

Estimated Divergence Rates of L1 Sequences

The rate of divergence of L1 sequences was estimated from the expected proportion of nucleotide sites which differ between two sequences after their evolutionary separation (Kimura 1977). The rates of divergence of various L1 sequences between rodents and primates are, in general, lower than rates within rodents (Table 1). This may reflect low rates of gene evolution in hominoid primates as compared to other mammalian orders (Bailey et al. 1991; Li et al. 1987). However, the average value obtained from the various rodent comparisons (4.15 × 10^{-9} changes/site-year) is consistent with the value obtained by Martin et al. (1985) for species of Mus; this is less than that for pseudogenes (4.6 × 10^{-9}; Li et al. 1981), reinforcing the idea of a constant molecular clock for L1 in rodents. Using the average divergence rate for rodent L1 sequences, it can be estimated that the two Peromyscus L1 subfamilies diverged about 44 million years ago. This value may be biased, though, as an alignment of the two Peromyscus sequences to L1Hs and the corresponding L1Hs amino acid sequence demonstrates these sequences contain both selected and nonselected mutations. Comparisons between L1Pm55 and L1Pm62 third-codon-position vs middle-codon-position changes for 429 codons resulted in a 1.5:1 ratio, indicating that these sequences inherited enough selected differences to support the view that each arose from diverse source genes.

Identification of Additional L1 Subfamilies

The L1Pm55 and L1Pm62 subfamilies together total from 600 L1 copies per haploid genome in P. mani-

culatus, as opposed to 20,000 copies of Mus. To detect the possible presence of additional subfamilies, the P. maniculatus genomic library was rescreened using the 1.8-kb EcoRI fragment of L1Pm62 and the 1.5-kb fragment of L1Pm55 as probes. Similar patterns of hybridization of the two probes to several plaque lifts of the library indicated that the same clones hybridized to both probes, though, in each case, longer exposures were necessary to visualize plaques hybridizing to the L1Pm55 probe. Signals, though variably intense, were observed in approximately 8% of the plaques. Based on an average insert of 15 kb, this estimates to approximately 16,000 copies (Bennett et al. 1984); a value approaching that of Mus. This value is greatly reduced with the omission of less-intense signals. This finding indicates the presence of additional L1Pm subfamilies.

Several phage clones that hybridized to different degrees with the probes were selected and purified, and the L1-containing EcoRI fragments were subcloned into the PT7/T3-18 plasmid vector (see Materials and Methods) generating pDK144, pDK145, and pDK150. Restriction maps are unique to each the L1 clones. Therefore, to further determine the relationship among the clones, cross-hybridization studies were performed. For each clone, the EcoRI fragment was isolated, labeled by nick-translation, and hybridized to electrophoretically separated EcoRI-cleaved DNAs representing the other clones. The pDK55 and PDK144 fragments hybridized poorly to the other subclones (Table 2). The 1.8-kb insert of pDK62 showed a greater similarity to the inserts of pDK145 and pDK150 than to those of pDK55 and pDK144, as indicated by the relative intensities of hybridization signals (Table 2). The 3.8-kb insert of pDK150 hybridized relatively strongly to the 2.2-kb insert of pDK145. These results indicate pDK55 and pDK144 each contain rather distinct L1 inserts, while pDK145 and pDK150 are more closely related. A less noticeable, though significant, relationship is shared between pDK62 and pDK145 or pDK150 (Table 2). These results indicate that L1Pm62, L1PM145, and L1Pm150 comprise the same or closely related subfamily; the elements represented by pDK55, pDK144, and pDK62 appear to be three very distinct subfamilies. Sequence analysis of these clones will be necessary for more detailed classification of additional L1 subfamilies.

Evolution of L1 Subfamilies Among Peromyscus Species

Restriction-site variation in the L1-repetitive elements of P. maniculatus (subspecies bairdii, rufinus,
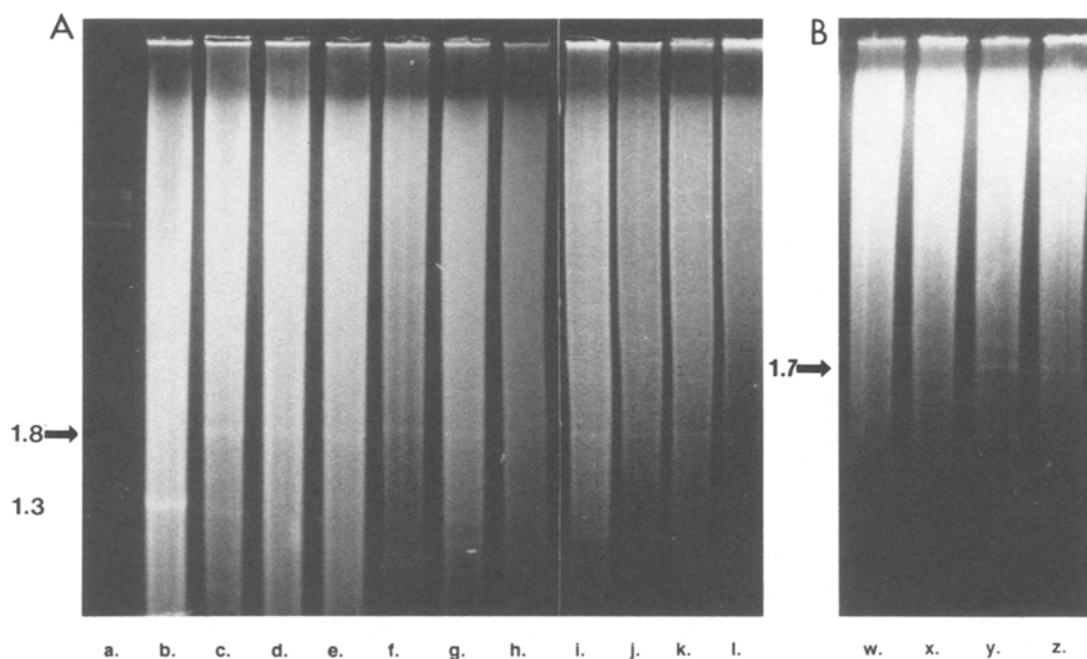
Fig. 2. Species comparison of restriction endonuclease-cleaved DNA by 1% agarose gel electrophoresis. **A.** *Eco*RI-cleaved genomic DNA. Lanes: a. Molecular weight marker. b. *Mus domesticus.* c. *Peromyscus maniculatus bairdii.* d. *P. m. rufinis.* e. *P. m. sonoriensis.* f. *P. polionotus.* g. *P. leucopus.* h. *P. gossypinus.* i. *P. m. bairdii.* j. *P. truei.* k. *P. difficilis.* l. *P. californicus.* **B.** *Ban*HI-cleaved genomic DNA. Lanes: w. *P. maniculatus.* x. *P. polionotus.* y. *P. leucopus.* z. *P. gossypinus.*
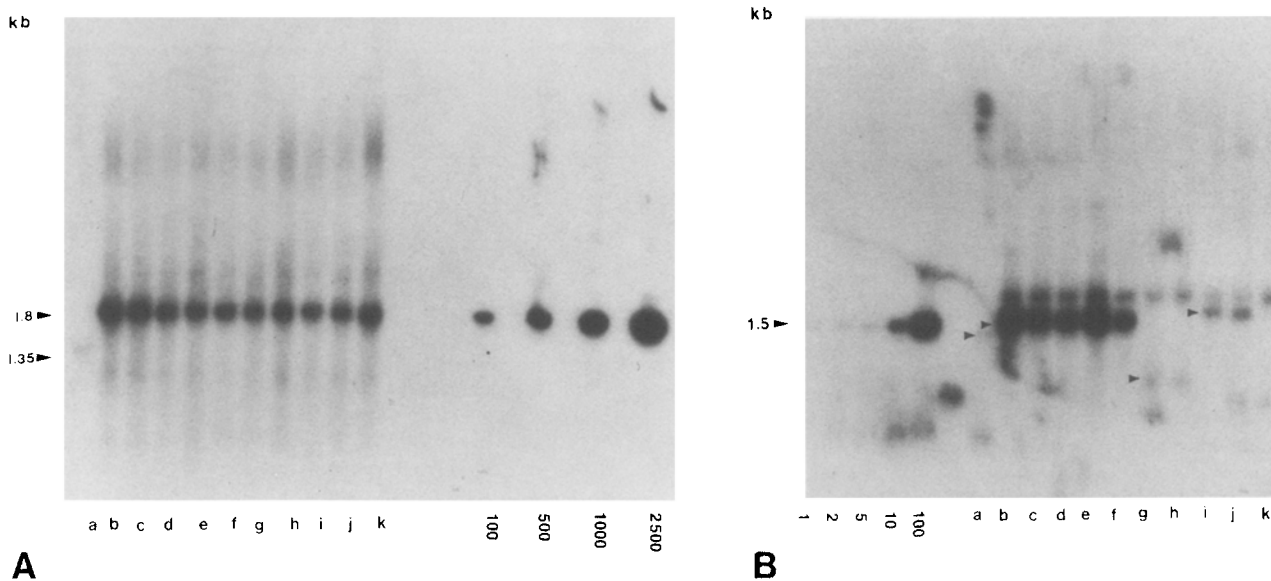
Fig. 3. Determination of copy number by Southern analysis. Genomic DNAs (1.0 μg) were digested with *Eco*RI, separated by agarose gel electrophoresis, and transferred to nylon. Amounts of plasmid equivalent to the indicated copy numbers were analyzed in parallel lanes. **A.** Hybridization to nick-translated 1.8-kb insert of pDK62. Lanes: a. *Mus musculus.* b. *P. m. bairdii.* c. *P. m. rufinus.* d. *P. m. sonoriensis.* e. *P. m. nubiterrae.* f. *P. po-*

lionotus. g. *P. leucopus.* h. *P. gossypinus.* i. *P. truei.* j. *P. difficulis.* k. *P. californicus.* Numbers refer to copies of pDK62 per haploid genome. **B.** Hybridization to nick-translated 1.5-kb insert of pDK55. Lanes a–k are identical to A. Numbers refer to copies of pDK55 per haploid genome. *Arrowheads* refer to corresponding fragments among species groups. (See text).

sonoriensis, and *nubiterrae*), *P. polionotus, P. leucopus, P. gossypinus, P. truei, P. difficilis,* and *P. californicus* was detected by Southern analysis. When DNA of these species was digested with *Eco*RI and hybridized to the nick-translated pDK62

(1.8-kb insert) probe, a 1.8-kb band appeared uniformly in all species (Fig. 5A). This was expected from earlier observations of ethidium-bromide-stained digests. Variant patterns were apparent among the species upon digestion with *Hind*III. A

A   **B**
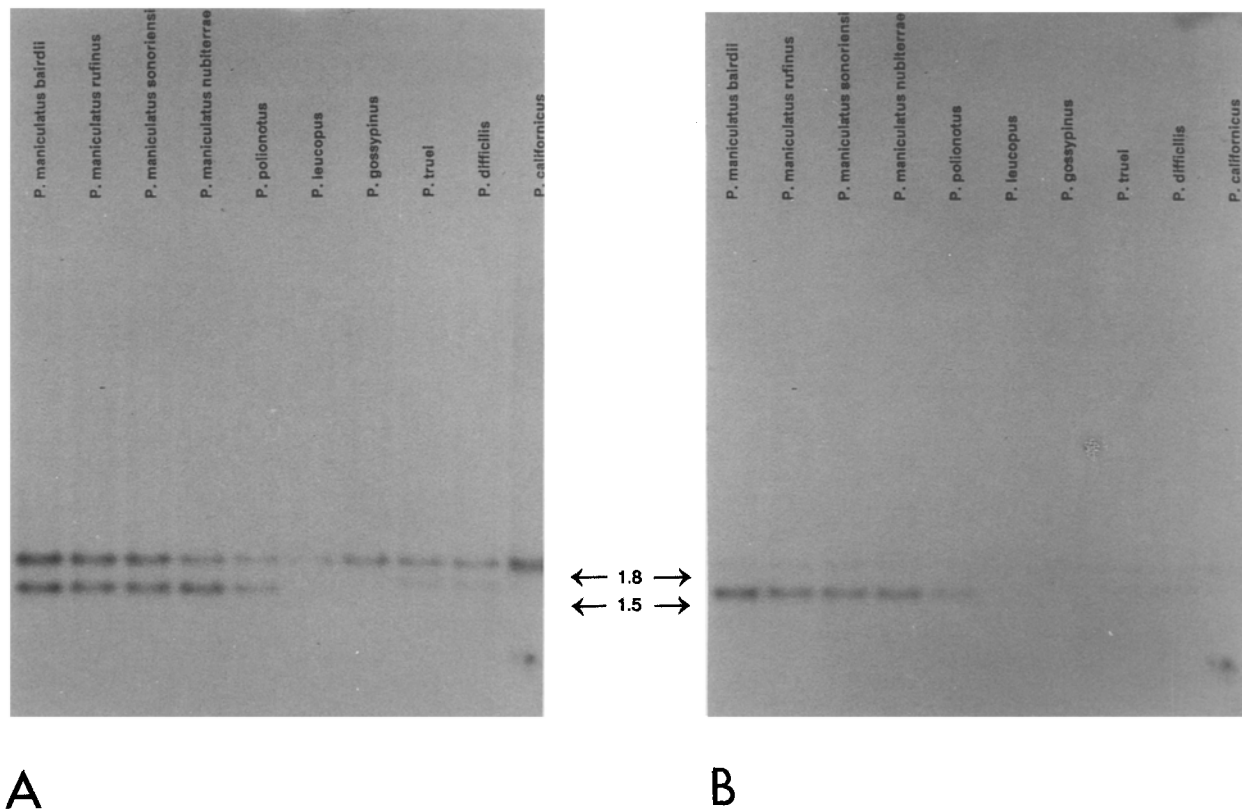
Fig. 4. Comparison of a Southern blot hybridized to pDK55 and washed under increasing stringencies. **A.** Blot of EcoRI-cleaved genomic DNA from seven *Peromyscus* species hybridized to a nick-translated 1.5-kb insert of pDK55. Wash conditions: 2 × SSC, 0.2% SDS, 65°C. **B.** The same southern blot washed under more stringent conditions: 0.5 × SSC. 0.2% SDS, 65°C, three washes at 1 h each.

Table 1. Relationship of LINEs[a]

	Human	Rat	L1Md	L1Pm62	L1Pm55
Human	—	0.326	0.327	0.343	0.377
Rat	2.67	—	0.171	0.267	0.293
L1Md	2.68	3.88	—	0.244	0.282
L1Pm62	2.86	4.12	3.69	—	0.289
L1Pm55	3.27	4.64	4.42	nd	—

[a] Above diagonal: Divergence (changes/site) among L1 elements was determined by comparing MIF-1 regional sequences (Fig. 1) using the gap/lim program (Devereux *et al.* 1984). Human L1 is from Skowronski *et al.* (1988), rat L1 is from Soares *et al.* (1985), *Mus* L1 is from Shehee *et al.* (1987), and *Peromyscus* L1 (62, 55) is from this study. Below diagnonal; Estimated rates of L1 divergence recorded as changes/site/10^9 years using 80 million years (my) as the divergence time between rodents and primates, 40 my between *Peromyscus* and murine rodents (*Rattus* and *Mus*) and 25 my for *Rattus* and *Mus*. nd = not determined

*Hind*III fragment of approximately 3.6 kb was present in members of the *maniculatus* and *leucopus* species groups, but was absent from species in the *truei* group and *P. californicus* (Fig. 5B), indicating that the variant arose prior to the divergence of the *leucopus* and *maniculatus* groups, but after separation from the *truei* group and *P. californicus*. Additionally, a 1.2-kb *Hind*III fragment of greater intensity appears in the *leucopus* species group relative to other species of *Peromyscus*, indicating the greater proportion of this particular variant has been maintained. Members of the *truei* species group have an *Xba*I variant (2.2 kb) not found in the

other species (Fig. 5C); therefore, it arose specifically within the *truei* group lineage. The presence of a *Xba*I band of approximately 3.8 kb is found in *P. californicus*, which lacks the 4.7-kb fragment found in other *Peromyscus* species, demonstrating a pattern which distinguishes a species of subgenus *Haplomylomys* from those of subgenus *Peromyscus*. A variant, approximately 0.5 kb in size, was detected in *Bg*/II DNA digests (Fig. 5D) and is restricted to members of the *maniculatus* group. A single band of high intensity is seen in autoradiograms of *Eco*RI and *Xba*I digests, while a series of bands of nearly equal intensities occurs with *Hind*III and *Bg*/II di-

Table 2. Comparative level of cross-hybridization among subcloned *Peromyscus* L1 fragments by Southern analysis[a]

Labeled probe	L1Pm55	L1Pm62	L1Pm144	L1Pm145	L1Pm150
L1Pm55	+ + + +	+	+	+	+
L1Pm62	+	+ + + +	+	+ + +	+ +
L1Pm144	+	+	+ + + +	+	+
L1Pm150	+	+ +	+ +	+ + +	+ + + +

[a] The number of pluses (+) corresponds to relative intensities visualized on autoradiographs. Four +'s correspond to the standard (self-hybridization); three +'s refer to nearly intense signals; two +'s refers to visually distinct bands and one plus to very low intensities. By contrast, the 3.8-kb fragment of pDK145 was not observed to hybridize to any of the labeled probes under these hybridization conditions (Materials and Methods). The 2.2-kb insert of pDK145 could not be separated from the equivalently sized plasmid vector.

gests indicating that a greater proportion of the *Eco*RI and *Xba*I sites of the repetitive element are conserved.

Rehybridization of washed filters with pDK55 (1.5-kb insert) DNA was performed. A Southern blot of *Eco*RI-digested DNA revealed distinctive species-group variants (Fig. 6A). A major 1.5-kb variant was present only in members of the *maniculatus* species group. The *leucopus* group members contained a minor fragment (0.8 kb), while a 1.6-kb band was present in the *truei* group; *P. californicus* may have an additional *Eco*RI fragment about 1.8 kb in size, but which could not be resolved into a distinct size fragment. *Hind*III-digested DNA probed with pDK55 (Fig. 6B) showed the same pattern observed with pDK62 (Fig. 5B), but an additional 6.0-kb fragment in the *maniculatus* group was present, and an independent less-intense band, also approximately 6.0 kb, occurred in species of the *truei* group. A 2.5-kb *Hind*III fragment appeared to hybridize more intensely in the *truei* group and *P. californicus* than in the others. Analysis of *Xba*I-digested DNA (Fig. 6C) revealed differences in addition to those observed with the pDK62 probe (Fig. 5C). A 4.4-kb *Xba*I fragment exists in the *maniculatus* group, while a less-intensely hybridized fragment of approximately the same size exists in the *truei* group. An additional 7.0-kb fragment distinguishes certain *P. maniculatus* subspecies from *P. polionotus*, in which it is absent. *Bg*/II fragments of 4.6 kb and 0.9 kb appear when probed with pDK55 (Fig. 6D), which were not found with pDK62 (Fig. 5D). This represents a variant that occurs only within the *maniculatus* species group in which it is polymorphic. In *P. m. rufinus* the predominant fragment detected was the 4.6-kb *Bg*/II fragment which is also prevalent in *P. m. nubiterrae* indicative of a closer relationship between these two subspecies than with *P. m. bairdii* and *P. m. sonoriensis,* which lack this variant. This fragment is also absent in *P. polionotus.*

A phylogeny showing points at which restriction-fragment-length variants (RFLVs) occurred is given in Fig. 7A and Fig. 7B as detected with the pDK62 and pDK55 probes, respectively. Variations in the *Peromyscus* L1 family detected with pDK62 all occurred at points of divergence at the species group level, rather than at the level of taxonomic species (Fig. 7A). Numerous major and minor RFLVs identified with pDK55 also predominantly occurred at the species group level of divergence (Fig. 7B). One RFLV of less intensity, a minor band in the *Xba*I digests, was present in *P. maniculatus,* but not in *P. polionotus.*
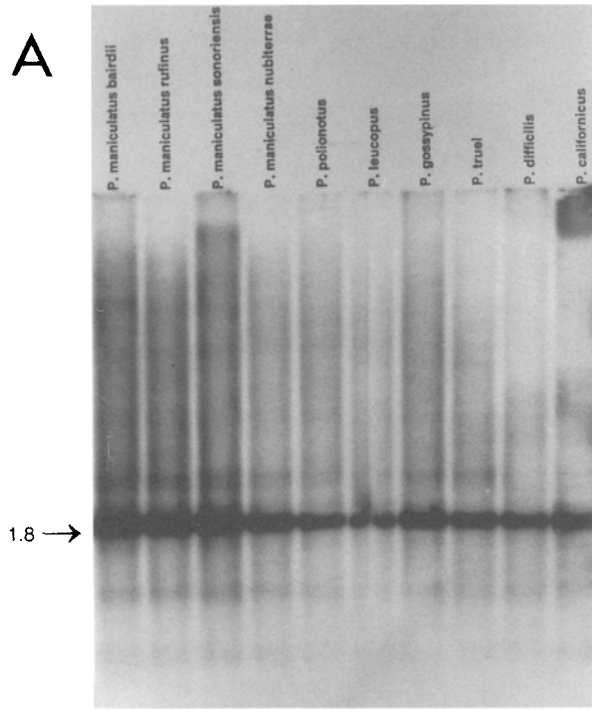
Discussion

Peromyscus is a major mammalian genus of more than 55 species inhabiting a wide variety of ecological situations, but limited to the North American continent. The animals are generally abundant where they occur. The fossil record of the group is well documented; many aspects of the biology of these rodents are known and they are amenable to laboratory conditions. Hence, *Peromyscus* has long been considered an excellent model for evolutionary study at the morphological, biochemical and, more recently, mitochondrial DNA levels. This study is the first to examine LINE-1 evolution within the genus.

Two highly divergent LINE-1 subfamilies were documented in *Peromyscus,* one more prevalent than the other. Though major and minor subfamilies of L1 have previously been identified in other mammalian species (Brown and Dover 1981; Jubier-Maurin et al. 1985), subfamily classification of L1 has been determined only by diagnostic nucleotides (Skowronski and Singer 1986; Jurka 1989). The level of divergence observed between L1Pm55 and L1Pm62 far exceeds that of individual intraspecies elements, with the exception of a few aberrant L1 copies (Soares et al. 1985), and exceeds the divergence between the Lx and L1 families in murine rodents (Pascale et al. 1990). L1 evolution in *Peromyscus,* therefore, is rather unique.

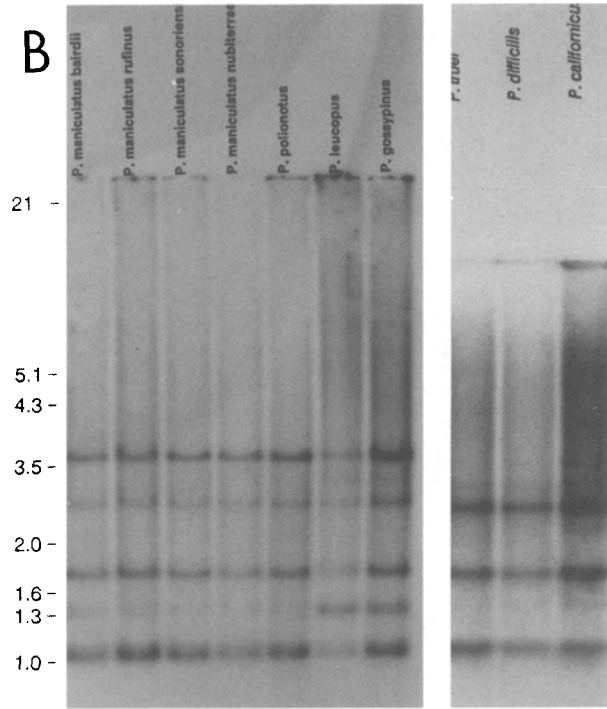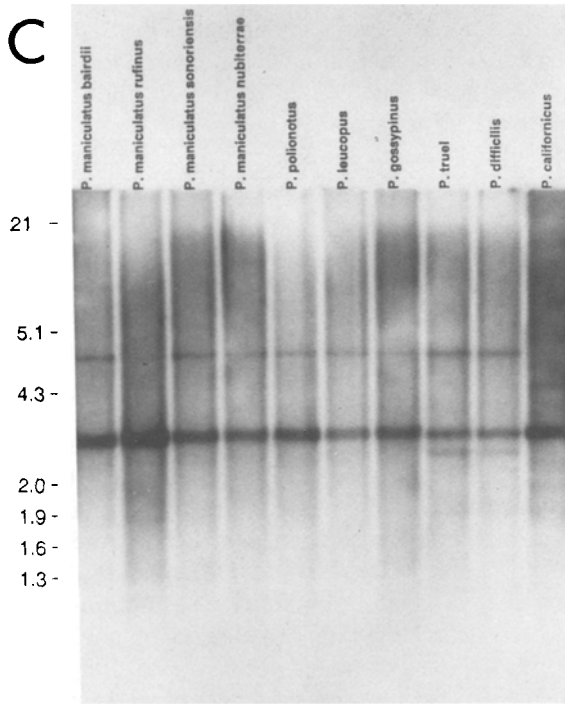Open-reading-frame sequences of L1 in mouse,

480

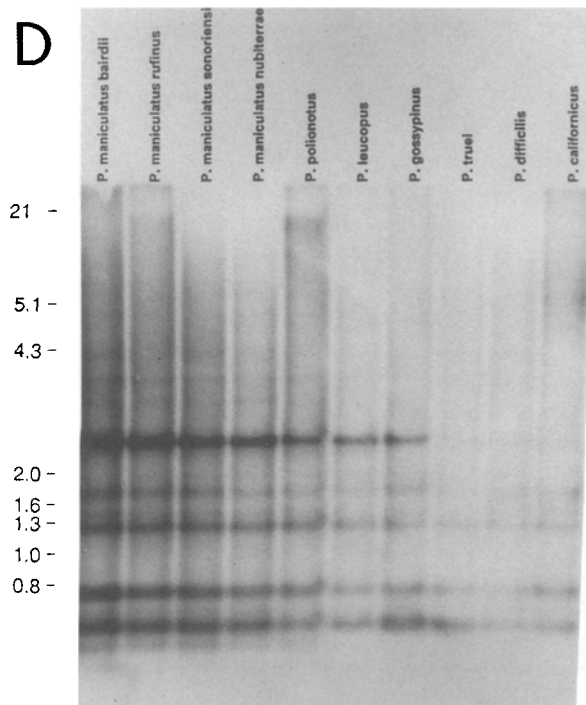Fig. 5. Southern blot of restriction endonuclease-digested genomic DNA of seven species of *Peromyscus* run on a 1% agarose gel and hybridized to a nick-translated 1.8-kb insert of pDK62. **A.** *Eco*RI. **B.** *Hind*III. **C.** *Xab*I. **D.** *Bg*/II.
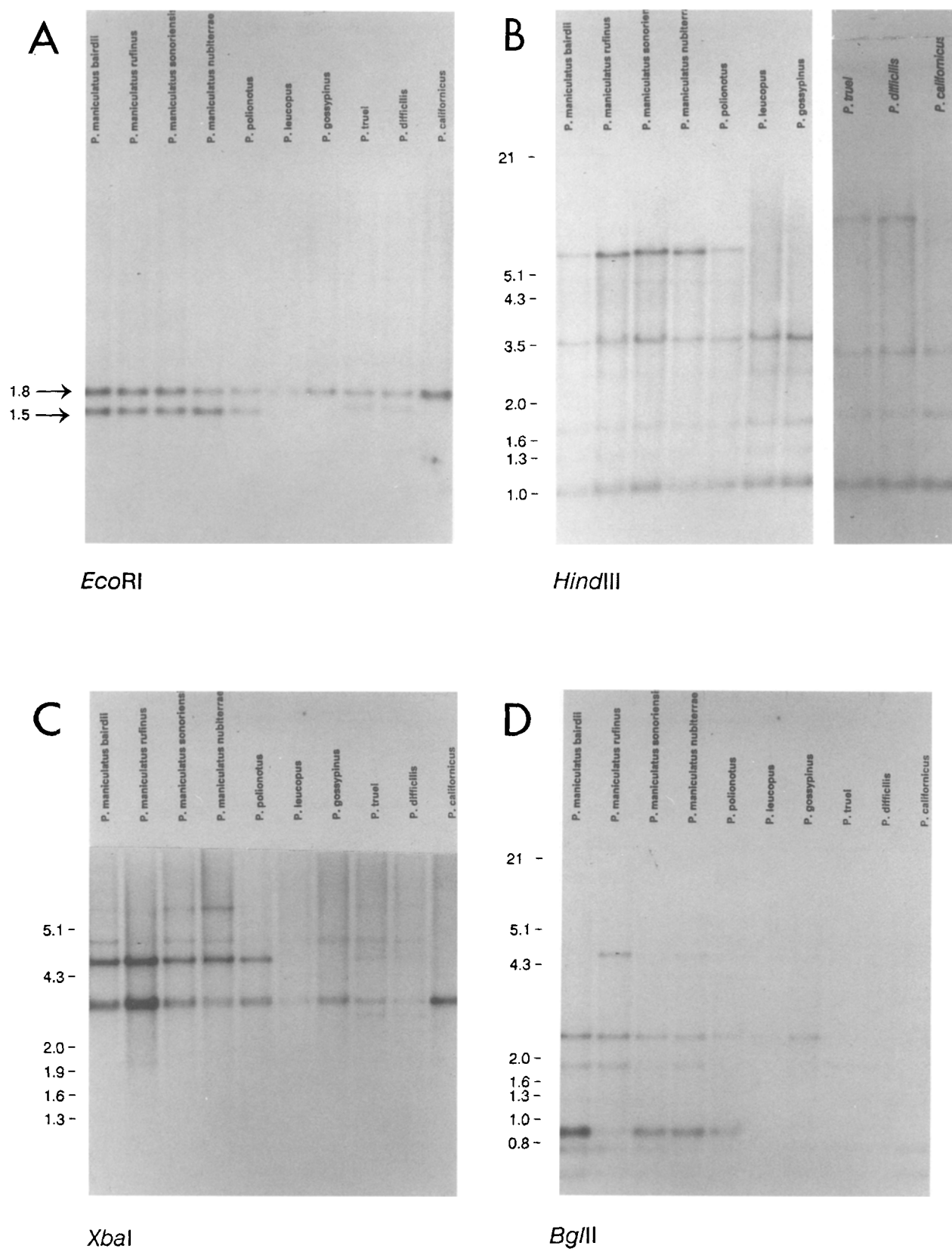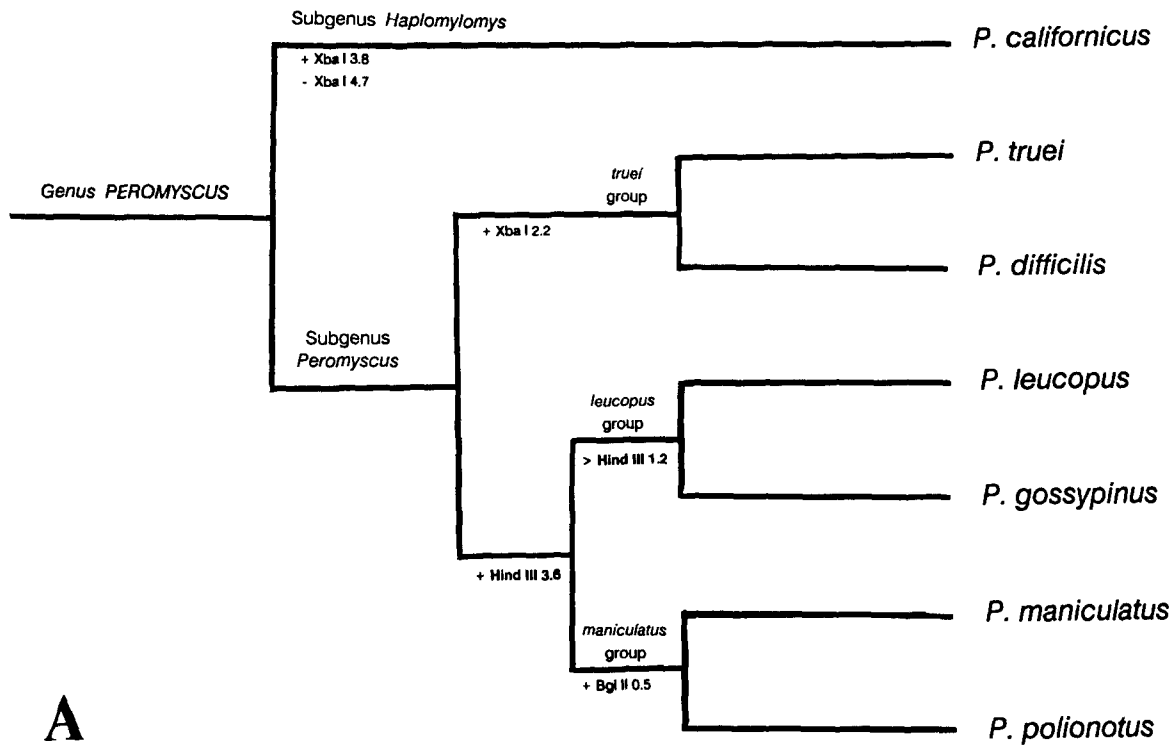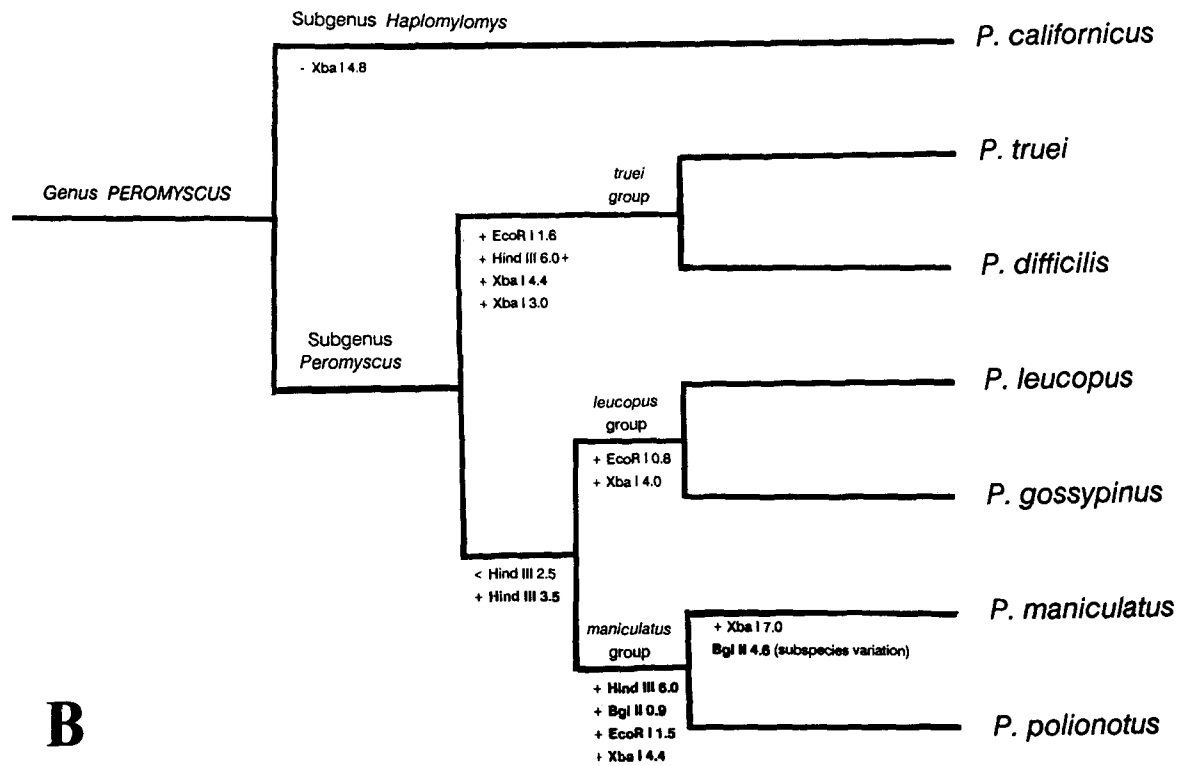
Fig. 6. Southern blot of restriction endonuclease-digested genomic DNA of seven species of *Peromyscus* run on an agarose gel and hybridized to a nick-translated 1.5-kb insert of pDK55. **A.** *Eco*RI. **B.** *Hind*III. **C.** *Xba*I. **D.** *Bg*/II.

Subgenus *Haplomylomys*

+ Xba I 3.8
- Xba I 4.7

Genus PEROMYSCUS

P. californicus

Subgenus
Peromyscus

truei
group

+ Xba I 2.2

P. truei

P. difficilis

leucopus
group

> Hind III 1.2

P. leucopus

P. gossypinus

+ Hind III 3.6

maniculatus
group

+ Bgl II 0.5

P. maniculatus

P. polionotus

A

Subgenus *Haplomylomys*

- Xba I 4.8

Genus PEROMYSCUS

P. californicus

Subgenus
Peromyscus

truei
group

+ EcoR I 1.6
+ Hind III 6.0+
+ Xba I 4.4
+ Xba I 3.0

P. truei

P. difficilis

leucopus
group

+ EcoR I 0.8
+ Xba I 4.0

P. leucopus

P. gossypinus

< Hind III 2.5
+ Hind III 3.5

maniculatus
group

+ Xba I 7.0
Bgl II 4.6 (subspecies variation)

P. maniculatus

+ Hind III 6.0
+ Bgl II 0.9
+ EcoR I 1.5
+ Xba I 4.4

P. polionotus

B

Fig. 7. Restriction-fragment-length variation of L1 in *Peromyscus* superimposed on a consensus phylogenic diagram. Key to symbols: + = addition of a variant; − = loss of a variant; > = increased copy number of variant. Major restriction variants are indicated in bold. Fragment size in kb. **A.** Variants detected wth pDK 62. **B.** Variants detected with pDK55.

rat, and human have diverged in a manner consistent with a constant rate of accumulation of mutational changes (Soares et al. 1985). The sequence comparisons obtained here confirmed this relationship, though rates differ in primates and rodents. However, in *Peromyscus*, L1 sequences (i.e., L1Pm62, L1Pm55) are as divergent from one another as either is from murine L1 members. The inference is that these subfamilies have been evolving independently for 25–35 million years, or approximately the time since murines (*Mus* and *Rattus*) and cricetines (*Peromyscus*) diverged.

Peromyscus is distinctive in having low copy numbers of each of the L1 subfamilies. This may explain the occurrence of a significantly lower total copy number in *Peromyscus* (approx. 500) relative to *Mus* (approx. 20,000) for the MIF-1 region (Martin et al. 1984; Brown and Dover 1981). Perhaps other, more divergent, L1 subfamilies exist in *Peromyscus;* alternatively, the *Peromyscus* genome contains a smaller number of L1 repeats. The identification of additional L1 subfamilies (Table 2) favors the former rather than the latter interpretation. Preliminary partial sequence data (Kass unpublished) corroborates the cross-hybridization patterns (Table 2). L1Pm144 apparently represents another subfamily (Kass unpublished) of about 25 copies. Sequence homology to L1 disappears at the 5' ends of L1Pm145 and L1Pm150, consistent with known L1 truncations (Voliva et al. 1983), and explains the inconsistency between the high degree of homology among the L1Pm62-like subfamily members with their variable restriction patterns. This may also explain the significantly-more-intense self-hybridization patterns (Table 2) than patterns of hybridization to other subfamily members.

Mechanisms by which coexisting divergent subfamilies evolve in *Peromyscus* are not fully understood. Subfamily formation in the human possibly occurs either by long periods of steady L1 evolution interspersed with rapid periods of retroposition or by short periods of rapid evolution separating periods of cumulative retroposition (Jurka 1989). Different conserved progenitor (source) genes have been proposed to explain subfamily structure in human *Alu* sequences (Jurka and Milosavljevic 1991). By alignment of L1Pm55 and L1Pm62 with L1Hs and the corresponding L1Hs amino acid sequence, the middle codon position was more conserved than the third codon position. This demonstrates enough selected-vs-random differences to support the view that these sequences have arisen from diverse source genes.

The pattern of L1 variation is consistent with the phylogeny of *Peromyscus* as currently understood (Stangl and Baker, 1984), and therefore has potential utility as an adjunct to other criteria for phylo-genetic reconstruction in other taxa. The observed variation corresponds primarily to the species group level of differentiation. The species group in *Peromyscus* is the level at which reproductive isolation is fully established. The concerted restriction-site changes of L1 in the genus *Mus* (Jubier-Maurin et al. 1985), considered with regard to the degree of interfertility (Chapman et al. 1974; Thaler et al. 1981; Bonhomme et al. 1984), also correspond to the outer limits of reproductive compatibility rather than to the taxonomic species definition.

Ohta and Dover (1984) speculate that "homogenization" of the LINE family of repeats may produce significant fitness differences between populations. Species discontinuities could originate by chromosomal mispairing due to divergent DNA compositions, by disruption of coordinated gene regulation by the insertion of repeat elements into regulatory sequences, and by hybrid dysgenesis (Rose and Doolittle 1983; Ginzburg et al. 1984). While it is clear that an association between patterns of L1 variation and physiological barriers to hybridization occurs in *Peromyscus*, the role, if any, of L1 in speciation in speculative.

The amount of L1Pm55-hybridizing DNA varies considerably among *Peromyscus* species; this may be the result of pronounced divergence among repeats, or of variation in the copy number of the L1Pm55 subfamily. Variation due to sequence divergence implies a long period of L1 evolution followed by rapid retroposition events. Alternatively, if there are simply variable numbers of L1Pm55 copies in the different species of *Peromyscus*, then L1 may be considered to be in a transitional period wherein this subfamily is gradually being gained or lost. An insight into the mechanism is possible by means of chromosomal analysis—an approach used by Baker and Wichman (1991) to demonstrate a correlation between loss of *Mys* element copy number in *Peromyscus* and frequency of chromosomal meiotic exchange resulting from unequal crossing over. Preliminary investigation by in situ hybridization (Baker and Kass unpublished) indicates the less-abundant subfamily (L1Pm55) is localized to a single chromosomal site.

Examination of specific genes, with known divergence times, that carry L1 insertions may clarify which subfamilies contain active source genes based on the retroposition model, by ascertaining approximate times of insertions (Casavant et al. 1988). Additionally, determining the level of change of the flanking direct repeats of individual L1 members can be useful in estimating comparative ages of specific L1 members.

We have presented a detailed analysis concerning the evolution of L1 elements in *Peromyscus*. We documented an atypical example of the existence of

two intraspecific highly divergent, low abundant, L1 subfamilies represented by L1Pm55 and L1Pm62, arising from different source genes. It is also evident that the evolution of L1 frequently corresponds to reproductive barriers.

References

Bailey WJ, Fitch DHA, Tagle DA, Czelusniak J, Slightom JL, Goodman M (1991) Molecular evolution of the ψη-globin gene locus: gibbon phylogeny and the hominoid slowdown. Mol Biol Evol 8:155–184

Baker RJ, Wichman HA (1991) Retrotransposition *Mys* is concentrated on the sex chromosomes: implications for copy number containment. Evolution 44:2081–2088

Bennett KL, Hill RE, Pietras DF, Woodworth-Gutai M, Kane-Haas C, Houston JM, Heath JK, Hastie ND (1984) Most highly repeated dispersed DNA families in the mouse genome. Mol Cell Biol 4:1561–1571

Benson SA, Taylor RK (1984) A rapid small-scale procedure for isolation of phage lambda DNA. Biotechniques 2:126–127

Blair WF (1943) Criteria for species and their subdivisions from the point of view of genetics. NY Acad Sci 44:179–188

Blin N, Stafford DW (1976) A general method for the isolation of high molecular weight DNA from eukaryotes. Nucl Acids Res 3:2303–2308

Bonhomme EF, Catalan J, Britton-Davidian J, Chapman K, Moriwaki K, Nevo E, Thaler L (1984) Biochemical diversity and evolution in the genus *Mus*. Biochem Genet 22:275–303

Brown SDM, Dover G (1981) Organization and evolutionary progress of a dispersed repetitive family of sequences in widely separated rodent genomes. J Mol Biol 150:441–466

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 1987:291–304

Carleton MD (1989) Systematics and evolution. In: Kirkland GL Jr, Layne JN (eds) Advances in the study of *Peromyscus* (Rodentia). Texas Tech University Press, Lubbock, pp 7–142

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

Chapman VM, Nichols EA, Ruddle RH (1974) Esterase-8 (Es-8): characterization, polymorphism, and linkage of an erythrocytic esterase locus on chromosome 7 of *Mus musculus*. Biochem Genet 11:347–358

Devereux J, Haeberli P, Smithies O (1984) A comprehensive set of sequence analysis programs for the VAX. Nucl Acids Res 12:387–395

Dice LR (1968) Speciation. In: King JA (ed) Biology of *Peromys-*

cus (Rodentia). Spec Publ 2, American Society of Mammalogists, pp 75–97

Dover G (1982) Molecular drive: A cohesive mode of species evolution. Nature 299:111–117

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

Flamm WG, Bond HE, Burr HE (1966) Density-gradient centrifugation of DNA in a fixed-angle rotor. A higher order of resolution. Biochim Biophys ACTA 129:310–319

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

Ginzburg LR, Bingham PM, Yoo S (1984) On the theory of speciation induced by transposable elements. Genetics 107:331–341

Hall ER (1981) The mammals of North America, 2nd ed. John Wiley and Sons, New York

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

Heinkoff S (1984) Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene 28:351–359

Hooper ET (1968) Classification. In: King JA (ed) Biology of *Peromyscus* (Rodentia). Spec Publ 2, American Society of Mammalogists, pp 27–74

Hutchison CA III, Hardies SC, Loeb DD, Shehee WR, Edgell MH (1989) LINEs and related retroposons: long interspersed repeated sequences in the eucaryotic genome. In: Berg DE, Howe MM (eds) Mobile DNA. American Society of Microbiologists, Washington DC, pp 594–617

Hwu HR, Roberts JW, Davidson EH, Britten RJ (1986) Insertion and/or deletion of many repeated DNA sequences in human and higher ape evolution. Proc Natl Acad Sci USA 83:3875–3879

Jubier-Maurin V, Dod BJ, Bellis M, Piechaczyk M, Roizes G (1985) Comparative study of the L1 family in the genus *Mus:* Possible role of retroposition and conversion events in its concerted evolution. J Mol Biol 184:547–564

Jurka J (1989) Subfamily structure and evolution of the human L1 family of repetitive sequences. J Mol Evol 29:496–503

Jurka J, Milosavljevic A (1991) Reconstruction and analysis of human Alu genes. J Mol Evol 32:105–121

Kimura M (1977) Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. Nature 267:275–276

Li W-H, Gojobori T, Nei M (1981) Pseudogenes as a paradigm of neutral evolution. Nature 292:237–239

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

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

Maniatis T, Fritsch EF, Sambrook J (1982) Molecular cloning. A laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

Martin SL, Voliva CF, Burton FH, Edgell MH, Hutchison III CA (1984) A large interspersed repeat found in mouse DNA contains a long ORF 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

Ohta T, Dover GA (1984) The cohesive population genetics of molecular drive. Genetics 108:501–521

Osgood WH (1909) Revision of the mice of the American genus *Peromyscus*. N Am Fauna 28:1–285

Padgett RW, Loeb DD, Snyder LRG, Edgell MH, Hutchison III CA (1987) The molecular organization of the beta-globin complex of the deer mouse, *Peromyscus maniculatus*. Mol Bio Evol 4:31–45

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

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

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

Rose MR, Doolittle WF (1983) Molecular biological mechanisms of speciation. Science 220:157–162

Shehee WR, Chao S-F, Loeb DD, Comer MB, Hutchison III CA, Edgell MH (1987) Determination of a functional ancestral sequence and definition of the 5' end of A-type mouse L1 elements. J Mol Biol 196:757–767

Singer MF (1982) SINEs and LINEs: highly repeated short and long interspersed sequences in mammalian genomes. Cell 28: 433–434

Skowronski J, Fanning TG, Singer MF (1988) Unit-length LINE-1 transcripts in human teratocarcinoma cells. Mol Cell Biol 8:1385–1397

Skowronski J, Singer MF (1986) The abundant LINE-1 family of repeated DNA sequences in mammals: genes and pseudogenes. Cold Spring Harbor Symp Quant Biol 51:457–464

Soares MB, Schon E, Efstratiadis A (1985) Rat LINE 1: the origin and evolution of long interspersed middle repetitive DNA elments. J Mol Evol 22:117–133

Southern EM (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. J Mol Biol 98:503–517

Stangl FB Jr, Baker RJ (1984) Evolutionary relationships in *Peromyscus:* congruence in chromosomal, genic and classical data sets. J Mamm 65:643–654

Thaler L, Bonhomme F, Britton-Davidian J (1981) Processes of speciation and semi-speciation in the house mouse. Symp Zool Soc Lond 47:27–41

Voliva CF, Jahn CL, Comer MB, Hutchison III CA, Edgell MH (1983) The L1 long interspersed repeat family in the mouse: almost all examples are truncated at one end. Nucl Acids Res 11:8847–8859

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

Wichman HA, Potter SS, Pine DS (1985) *Mys,* a family of mammalian transposable elements isolated by phylogenetic screening. Nature 317:77–81