Molecular Evolution of Two Lineages of L1 (LINE-1) Retrotransposons in the California Mouse, *Peromyscus californicus*

N. Carol Casavant,[1] Rhonda N. Lee,[2] Amy N. Sherman[3] and Holly A. Wichman

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

Manuscript received August 7, 1997
Accepted for publication May 28, 1998

ABSTRACT

The large number of L1 [long interspersed elements (LINE)-1] sequences found in the genome is due to the insertion of copies of the retrotransposon over evolutionary time. The majority of copies appear to be replicates of a few active, or "master" templates. A continual replacement of master templates over time gives rise to lineages distinguishable by their own unique set of shared-sequence variants. A previous analysis of L1 sequences in deer mice, *Peromyscus maniculatus* and *P. leucopus*, revealed two active L1 lineages, marked by different rates of evolution, whose most recent common ancestor predates the expansion of the Peromyscus species. Here we exploit lineage-specific, shared-sequence variants to reveal a paucity of Lineage 2 sequences in at least one species, *P. californicus*. The dearth of Lineage 2 copies in *P. californicus* suggests that Lineage 2 may have been unproductive until after the most recent common ancestor of *P. californicus* and *P. maniculatus*. We also show that Lineage 1 appears to have a higher rate of evolution in *P. maniculatus* relative to either *P. californicus* or *P. leucopus*. As a phylogenetic tool, L1 lineage-specific variants support a close affinity between *P. californicus* and *P. eremicus* relative to the other species examined.

LONG interspersed elements (LINEs) were originally defined in the mammalian genome on the basis of high copy number and a size longer than 5 kb; it is now known that LINEs encode functions required for their own retrotransposition, but lack long terminal repeats. A phylogenetic analysis of reverse transcriptases shows that the major family of mammalian LINEs, designated L1, is distinct from the retroviruses and more closely related to a class of transposable elements that includes several Drosophila elements, as well as elements from amphibians, plants, protists (Xiong and Eickbush 1990), and fish (Duvernell *et al.* 1996). L1 is present in tens of thousands of copies per haploid genome in all species of mammals examined to date (for reviews see Rogers 1985; Skowronski and Singer 1986; Edgell *et al.* 1987; Hutchison *et al.* 1989; Martin 1991). The detection of L1 by Southern blot analysis throughout seven orders of mammals including Marsupialia (Burton *et al.* 1986) has been interpreted as evidence that L1 was present in the common ancestor of subclass Theria. L1 has been characterized at the sequence level in a much narrower range of species, including several rodents, rabbits, bovids, and primates.

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 Zoology and Genetics, Iowa State University, Ames, IA 50011.

[2]*Present address:* L. J. Roberts Center for Alzheimer's Research, Sun Health Research Institute, Sun City, AZ 85372.

[3]*Present address:* Pathology Department, Stanford University, Palo Alto, CA 94305.

A full-length L1 is over 5 kb and consists of a 5′-untranslated region that includes the promoter, two open reading frames (ORFs), a 3′-untranslated region, and an A-rich tail. ORF1 encodes an RNA-binding protein that complexes with L1 RNA in the cell (Martin 1991; Holmes *et al.* 1992; Kolosha and Martin 1995; Hohjoh and Singer 1996). The second ORF encodes reverse transcriptase, the most conserved portion of the element (Xiong and Eickbush 1990), and additionally encodes for an endonuclease (Feng *et al.* 1996). Active elements are thought to be transcribed and then reverse transcribed by their own reverse transcriptase (Hattori *et al.* 1986; Loeb *et al.* 1986) before insertion into the genome. However, in Mus and human, most inserts are truncated at the 5′ end (Hutchison *et al.* 1989), and many of the full-length elements have accumulated debilitating mutations within the ORFs, so the vast majority of L1 elements in these genomes are pseudogene copies incapable of propagation.

The tens or hundreds of thousands of L1 elements per genome can be thought of as molecular "fossils" (Jurka *et al.* 1995; Usdin *et al.* 1995), so that each genome contains a record of the evolutionary history of L1 activity. Much of what is understood about L1 movement comes from comparative sequence analysis of "fossils" 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 the L1 evolutionary analyses is that most L1s 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. This model manifests itself in a phylogenetic analysis of selected elements as a tree with one or a few major lineages, rather than a highly branched tree with many independent clades (Clough *et al.* 1996).

The youngest L1 sequences in *Mus domesticus* (Rikke *et al.* 1991), *M. spretus* (Rikke *et al.* 1991; Casavant and Hardies 1994), and *Rattus norvegicus* (Cabot *et al.* 1997) are grouped into two clades; the clades appear to emerge from a single common ancestor either during or immediately before their respective host's speciation. A phylogenetic analysis of both older and younger sequences from Mus reveals two briefly overlapping L1 lineages: F as the putatively extinct clade, and A as the "replacement" (Adey *et al.* 1994). However, a recent insertion in Mus (Mulhardt *et al.* 1994) of a recombinant product between the A and F clades (S. Martin, personal communication) refutes the extinction of the F clade in *M. domesticus.* In Peromyscus, as in Mus, the number of active lineages appears to be few (Casavant *et al.* 1996).

Retrotransposon subfamilies, including the L1 lineages, are defined by specific shared variants that can be deduced by phylogenetic analysis (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; Casavant *et al.* 1996), and these diagnostic markers can be sequentially ordered. Variants affecting restriction enzyme recognition sites can alter the sizes of bands detected by genomic Southern blot analysis. For example, observation of restriction fragment length polymorphisms has revealed the rapid evolution of L1 in arvicolid rodents (Modi 1996). Variants that do not affect restriction enzyme recognition sites can also be detected by genomic Southern blot analysis using lineage-specific probes. For example, oligonucleotides can be designed as species-specific L1 probes (Rikke *et al.* 1991) as well as probes that are informative above the species level (Usdin *et al.* 1995; Verneau *et al.* 1997).

In this study, we expand our characterization of L1 in deer mice of the genus Peromyscus. Six Peromyscus species-groups are represented: *maniculatus*, *leucopus* (based primarily on results from the previous study), *truei* (*P. truei* and *difficilis*), *crinitus*, *californicus*, and *eremicus.* A species from a closely related genus, *Osgoodomys banderanus*, was included in one aspect of this study.

The taxonomic relationships between species included in the study are shown in Table 1. Peromyscus has been divided into two subgenera (Nowak 1991) composed of 13 species-groups that include 53 species (Carleton 1989). Six additional subgenera, including *Osgoodomys* (see below), were sometimes included in Peromyscus but are now recognized as full genera (Carleton 1989; Nowak 1991). Although there is good agreement on the close relationship between species within each species-group, the relationship between the groups has not been as well established. *P. crinitus* and *O. banderanus* have the primitive karyotype for the group, and *P. maniculatus* has the most derived karyotype (Baker *et al.* 1987). *P. eremicus* and *P. californicus* are classified together in the subgenus *Haplomylomys* (Carleton 1989; Nowak 1991), but appear to be karyotypically distinct and derived independently from the primitive karyotype (Baker *et al.* 1987). Although *P. difficilis* is a member of the *truei* species-group, the two species are also karyotypically distinct.

Recently active L1 lineages have been characterized in *P. leucopus* and *P. maniculatus* (Casavant *et al.* 1996). We suggested that the divergence of two L1 lineages active in both species preceded the radiation of Peromyscus, and that these lineages have different rates of evolution. L1 lineages are similar in these two closely related species, and this species-complex will be represented here by *P. maniculatus* except where there are differences between them.

Here we demonstrate that *P. californicus* is lacking restriction-site defined L1 subfamilies shared by species of the subgenus Peromyscus, and that these subfamilies are made up largely of sequences from Lineage 2. We confirm the reduction of Lineage 2 copies in *P. californicus* by characterization of PCR-amplified L1 elements from this species. This observation supports previous reports of the episodic activity of L1s (Pascale *et al.* 1990, 1993). There is no evidence for additional, highly active L1 lineages in the species examined. We infer that Lineage 1 and Lineage 2 produce the majority of L1 copies in the Peromyscus species examined to date.

Hybridization of L1 lineage and species-specific oligonucleotides to DNA from these species demonstrates the potential usefulness of these markers for investigating host sytematics. *P. californicus* and *eremicus* both have a scarcity of Lineage 2 copies and share a unique Lineage 1 subfamily, but with a restriction-site polymorphism between them. These data support the proposed relationship between the *californicus* and *eremicus* species groups.

MATERIALS AND METHODS

DNA: *P. maniculatus* (TK25740 and TK29798), *P. crinitus* (TK26309), *P. truei* (TK21858), *P. difficilis* (TK32041), *P. eremicus* (TK26234), and *O. banderanus* (TK19658) tissues were from The Museum, Texas Tech University. *P. californicus* (ISC134 and ISC137) tissues were from the Peromyscus Genetic Stock Center. DNA was prepared as previously described (Casavant *et al.* 1996).

Labeling and probing: The lineage-specific oligonucleotides were labeled and hybridizations carried out as previously described (Casavant *et al.* 1996).

L1GEN, a 252-bp subclone of Man109, recognizes sequences from both Lineage 1 and Lineage 2 and is thus considered a generic L1 probe. It was prepared by PCR-amplification using the following primers: AAGAAGTCAAGCTTTCCC, which recognizes a site 70 bases upstream from the region of Man109 sequence reported in Casavant *et al.* (1996), and

GGTTAGTGTTACCCCAA, which ends at base 182 of Man109. When used as a probe, L1GEN was PCR-amplified and labeled by random priming. The filters were hybridized with the labeled L1GEN in 5× SSCP (1× SSCP is 120 mm sodium chloride, 15 mm sodium citrate, and 20 mm sodium phosphate), 2 mg of denatured salmon sperm, and 1× Denhardt's overnight at 55°, and were washed three times for 30 min each with 5× SSCP at 55°, then exposed to film.

Lineage-specific L1 screening of Peromyscus species: Two micrograms of DNA from six individuals representing four different species-groups of Peromyscus were digested separately with *Eco*RV and *Bgl* II, electrophoresed into a 1% agarose gel, blotted essentially according to Southern (1975) and hybridized consecutively with two lineage-specific oligonucleotide probes (LIN1-3′ and LIN2-3′, Figures 1 and 3) and the generic probe L1GEN. The individual animals included *P. maniculatus* (TK29798), *P. crinitus* (TK26309), *P. truei* (TK-21858), *P. difficilis* (TK32041), and *P. californicus* (ISC134 and ISC137). The blots were stripped between each oligonucleotide hybridization until no detectable counts were observed on the blots.

Construction and screening of PCR-derived L1 libraries: PCR-derived libraries of *P. californicus* and *P. maniculatus* were constructed as previously described (Casavant *et al.* 1996) using genomic DNA from *P. maniculatus* (TK25740) and *P. californicus* (ISC134) as templates. Colonies from both *P. maniculatus* and *P. californicus* libraries were gridded, lifted, and probed serially with lineage-specific oligonucleotides. The membranes containing the gridded libraries from both *P. maniculatus* and *P. californicus* were first hybridized with LIN1-3′ (TTGTCATATAGGTCC). Ten *P. californicus* clones negative for LIN1-3′ and three *P. californicus* clones positive for LIN1-3′ were sequenced. Because no *P. maniculatus* PCR-derived clones were sequenced, the membranes were hybridized to additional lineage-specific oligonucleotides. The additional oligonucleotides included: LIN2-3′ (TTATCAAAAAGGTCT), LIN1 (CAA TGGACAAAGAAG), and LIN2.2 (GATAAAAGGGCTGAG).

Sequencing the L1 clones: Individual *P. californicus* PCR-derived clones were sequenced as previously described (Casavant *et al.* 1996). The sequence of each clone was verified by sequencing 95–100% of the complementary strand. GenBank accession numbers for DNA sequences generated for this study are U70828 for CAl26B; U70829 for Cal24B; U70830 for Cal31D; U70831 for Cal15D; U70832 for Cal38D; U70833 for Cal28B; U70834 for CalCC7; U70835 for CalCC13; U70836 for Cal37A; U70837 for Cal21D; U70838 for Cal32A; U70839 for Cal44A; and U70840 for CalCC2. Accession numbers for the *P. maniculatus* and *P. leucopus* included in the phylogenetic trees are U70925 for Leu1-18; U70926 for Leu*2-1; U70927 for Leu*2-2; U70928 for Leu4-5; U70931 for Leu4-5; U70932 for Leu2-22; U70924 for Man29; U70929 for Man28; U70930 for Man106; U70933 for Man108; U70934 for Man110; and U70935 for Man27.

The phylogenetic trees: Phylogenetic trees were derived using PAUP, version 3.0s (Swofford 1990), using the Bootstrap 100 replication heuristic algorithm. The final tree was based on using the single peromyscine ancestral sequence, as well as different combinations of ancestral sequences, and both peromyscine lineage sequences. Due to the large number of sequences in the total collection, smaller samples were used to attempt to resolve polytomies.

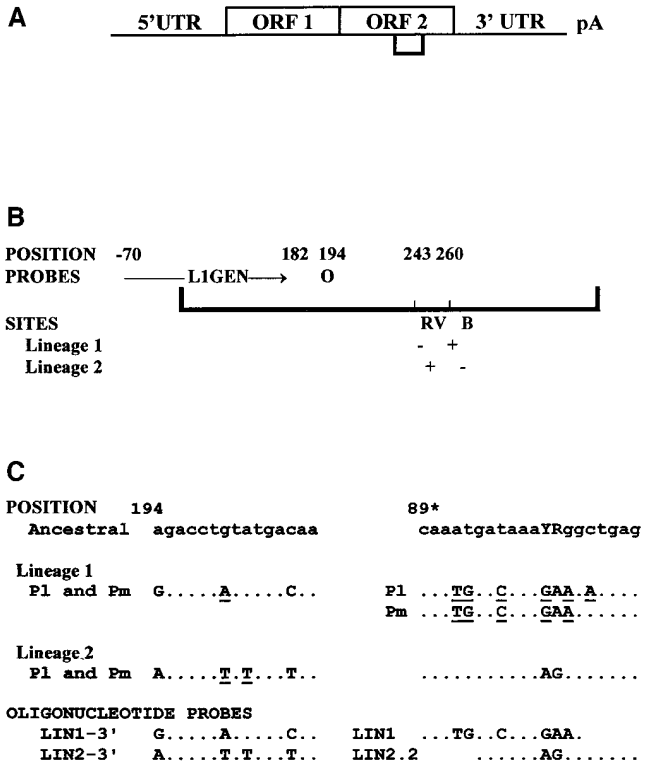


Figure 1.—Lineage-specific differences between Peromyscus L1s. (A) L1 map. The bracket underneath the general physical map of a rodent L1 indicates the region pertinent to the previous study on *P. leucopus* and *P. maniculatus* L1s in Casavant *et al.* (1996) and the present study on *P. californicus* L1s. ORF, open reading frame; UTR, untranslated region; pA, polyadentylated region. (B) Consensus restriction site map of Lineage 1 and 2. Restriction sites *Eco*RV and *Bgl* II differ between the two Peromyscus lineages. Only *Eco*RV and *Bgl* II sites relevant to the current study (Casavant *et al.* 1996) are illustrated. The presence or absence of a site in the lineage is indicated below the bracket. Numbers indicate the position of the restriction sites or probes (used in Figure 2) and correspond to those shown in Figure 1 (Casavant *et al.* 1996) and Figure 3, this study. L1GEN, the 252-bp subclone of Man109; O, location of lineage-specific oligonucleotides LIN1-3′ and LIN2-3′; B, *Bgl* II; RV, *Eco*RV; +, presence of site; −, absence of site. (C) Lineage-specific oligonucleotide probes. Oligonucleotides LIN1-3′ and LIN2-3′ (position indicated by O in Figure 1B) are specific to Lineage 1 and Lineage 2, respectively, and are used to illustrate the presence or absence of Lineage 1 and Lineage 2 in Figure 2. The complement for each oligonucleotide binding site was synthesized. Oligonucleotides LIN1 and LIN2.2, also specific to Lineage 1 and Lineage 2 (respectively), are used for corroboration. The ancestral sequence (designated as CRI_ANC in Figure 3) was derived using parsimony from L1MdA2, L1Rn3A, L1Pm55, L1Pm62, Leu1–18, L1Cg, Ory3, and Ory4. 194 and 89, position of the most 5′ end base (Figure 3); dot (.), same as in the ancestral sequence; boldface underlined letters, shared-sequence variants (see Figure 3); boldface letters, differences between lineages; asterisk (*), position not indicated in Figure 1B; Y, pyrimidine; R, purine; Pl, *P. leucopus*; Pm, *P. maniculatus*.

RESULTS

Differentiation between Lineages 1 and 2 by lineage-specific restriction sites: Lineage-specific variants within a 614-bp region from within ORF 2, the reverse transcriptase gene of L1, have been defined for two peromyscine L1 lineages (Figure 1B; Figure 1 in Casavant *et al.* 1996). Examination of L1 sequences reveals restriction

TABLE 1

Taxonomic relationship between species used in this study, based on Carleton (1989) and Nowak (1991)

Family: *Muridae*
Subfamily: *Sigmodontinae*
Genus: Peromyscus
Subgenus: *Peromyscus*
Species-group: *maniculatus*
P. maniculatus
Species-group: *leucopus*
P. leucopus
Species-group: *truei*
P. truei
P. difficilis
Species-group: *crinitus*
P. crinitus
Subgenus: *Haplomylomys*
Species-group: *californicus*
P. californicus
Species-group: *eremicus*
P. eremicus
Genus: Osgoodomys
O. banderanus

recognition site differences between Lineages 1 and 2. Comparison to the ancestral sequences further shows that at some point in evolution, Lineage 1 lost a *Eco*RV site (GATATC) due to a transition at position 246, and Lineage 2 lost a *Bgl*II site (AGATCT) due to a transition at position 260. In addition, Lineage 1 acquired two *Bgl*II sites in *P. maniculatus* due to a transversion at position 246 and a transition at 381 after *P. maniculatus* and *P. leucopus* diverged. These two *P. maniculatus*-specific sites are irrelevant for this study: *Bgl*II fragments based on differences at positions 246 and 260 are too similar in size to be differentiated by standard Southern blot analysis, and any fragments produced 3′ of position 246 (including those produced by the change at position 381) would not be observed because they are distal to the probes.

To examine the history of two L1 lineages in Peromyscus, individuals representing four distinct species-groups (*crinitus*, *californicus*, *truei* and *maniculatus*; Table 1) were examined. DNA from *P. crinitus*, two *P. californicus* individuals, *P. truei* and *difficilis*, and *P. maniculatus*, digested separately with the enzymes *Eco*RV and *Bgl*II, was Southern blotted and hybridized (Figure 2A) with the L1 subclone designated L1GEN (Figure 1B). Lineage-specific restriction sites result in restriction fragment length differences made evident by genomic Southern blot analysis. Additional *Eco*RV and *Bgl*II sites located outside of the region characterized are necessary to generate a restriction fragment; their number, their exact positions, and their changes in distribution during L1 evolution are unknown. Assuming no other restriction site changes, two restriction fragments would be predicted. The ancestral fragment would contain both Lineage 1 and 2 sequences, and the second fragment would contain either Lineage 1- or Lineage 2-specific sequences inserted after the loss of the *Eco*RV or *Bgl*II site (respectively). However, Figure 2A reveals more than the two predicted bands for each enzyme. The sizes of hybridizing bands common among the four different species-groups in the *Eco*RV-digested DNA are 2.5, 2.25, and 1.65 kb; in the *Bgl*II-digested DNA, the sizes of the common hybridizing bands are 2.6, 2.1, 1.75, and 1.0 kb. These additional bands must be due to polymorphisms in restriction sites outside of the characterized region and could reflect undefined lineages.

The intensity of these bands after only short exposure of the autorads suggests that these hybridizing bands are substantial multicopy families. Restriction fragments containing low copy numbers would only be detected after longer exposure time.

Differentiation between Lineages 1 and 2 by lineage-specific variants: Further differentiation of the banding pattern for both the *Eco*RV and *Bgl*II digests is necessary to resolve the identity of the bands, including the possibility of additional undefined lineages. Shared-sequence variants for both peromyscine lineages have been defined, and oligonucleotides have been synthesized to contain lineage-specific variants. The alignment of Lineage 1, Lineage 2, and the inferred peromyscine ancestral sequence in Figure 1C shows the differences contained in the oligonucleotides Lin1-3′ and Lin2-3′. Both oligonucleotides are complementary to the sense strand shown and differ from each other at four bases and from the ancestral sequence at two or more bases.

Hybridization with lineage-specific oligonucleotides (LIN1-3′ and LIN2-3′) reveals the relationship between the two lineages and the individual restriction bands in the *Eco*RV and *Bgl*II hybridizations (Figure 2, B and C). *Eco*RV 2.5- and 2.25-kb bands and *Bgl*II 2.1-, 1.75-, and 1-kb bands hybridized exclusively with LIN1-3′ (Figure 2B). *Eco*RV 1.65-kb and *Bgl*II 2.6-kb bands hybridized predominantly to LIN2-3′ (Figure 2C), but also hybridize slightly with LIN1-3′. LIN1 (Figure 1C) also hybridizes faintly to these bands, suggesting that both lineages have subfamilies within the *Eco*RV 1.65-kb and *Bgl*II 2.6-kb bands. Other lineage-specific oligonucleotides, LIN1 and LIN2.2 shown in Figure 1C, produce results consistent with the data presented (data not shown). Loss of a restriction site during the evolution of a lineage is expected to produce a larger, lineage-specific band. Thus, the presence of Lineage 1 sequences in the *Eco*RV 1.65-kb band is not unexpected, and probably reflects those ancestral sequences amplified before the loss of the *Eco*RV site in the lineage. However, the presence of Lineage 1 sequences in the *Bgl*II 2.6-kb fragment is less easily explained. An important caveat about the use of oligonucleotides to associate these bands with their respective lineages is that the temporal order of the changes represented in the oligonucleotides and their relationship in time to the restriction site changes are unknown. An oligonucleotide containing changes, all

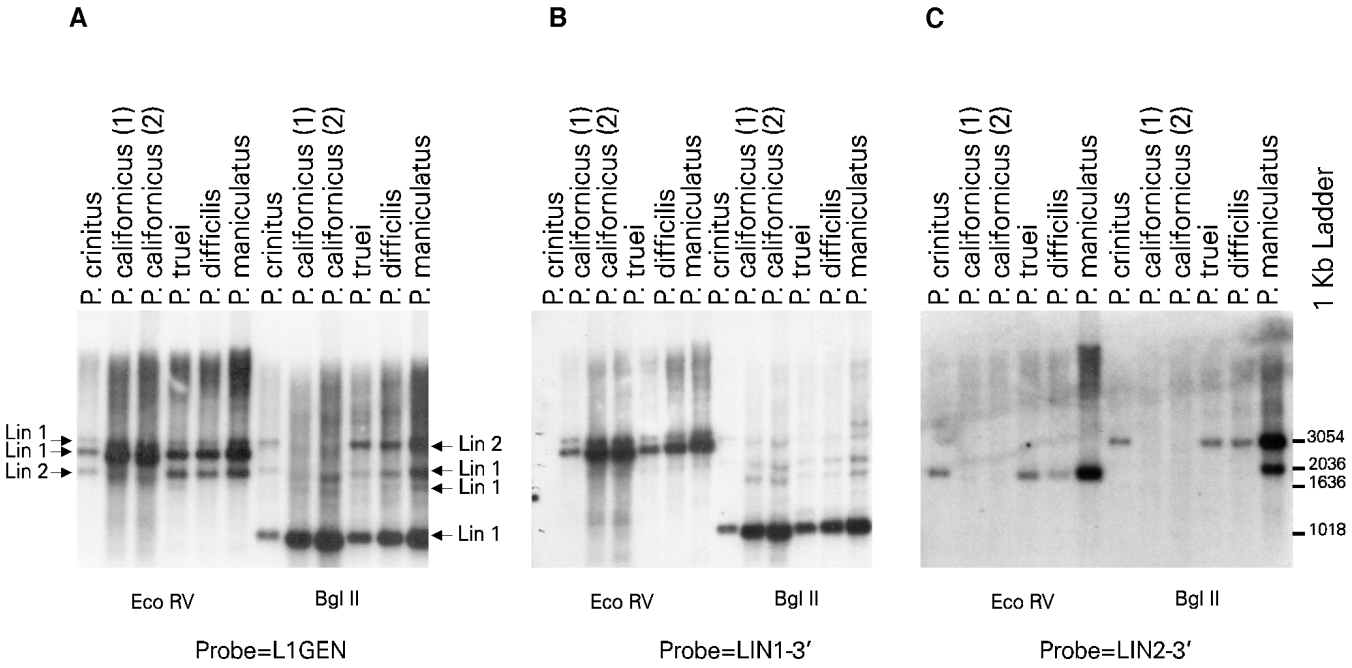


Figure 2.—Genomic Southern blot analysis of L1 in Peromyscus. (A) L1 bands detected with a generic L1 probe. DNA from four Peromyscus species-groups was hybridized with a 252-bp L1 probe, L1GEN (Figure 1A). Three of the species-groups show similar banding patterns when hybridized with L1GEN, but one band is almost undetectable in *P. californicus* when digested with either *Bgl*II or *Eco*RV. The arrow and designated lineage (Lin1, Lineage 1; Lin2, Lineage 2) mark each band based on the information in B and C. (B) L1 bands detected with LIN1-3′ (a Lineage 1-specific oligonucleotide). The banding pattern observed in A is further differentiated by the probe LIN1-3′ (Figure 1C) to show those bands that contain Lineage 1 sequences. Although not completely absent, one of the bands in each lane is greatly reduced in intensity. (C) L1 bands detected with LIN2-3′ (a Lineage 2-specific oligonucleotide). The banding pattern observed in A is differentiated by the probe LIN2-3′ (Figure 1C) to show bands that contain Lineage 2 sequences. *P. californicus* sequences have no hybridizing band.

of which were acquired after the restriction site loss, would hybridize only to "lineage-specific" bands. Oligonucleotides containing changes acquired before the restriction site loss would hybridize to lineage-specific bands and those containing both lineages. This suggests that loss of the *Bgl*II site occurred early in the evolution of Lineage 2, before the changes reflected in the oligonucleotides. All of the L1GEN-hybridizing bands hybridized with either LIN1-3′ or LIN2-3′; thus no new major, uncharacterized L1 lineages are revealed by these experiments, although lineages that do not differ at *Eco*RV or *Bgl*II sites could be hidden within the observed bands.

Absence of Lineage 2 in *P. californicus*: The hybridization of the *Eco*RV 1.65-kb and *Bgl*II 2.6-kb bands in *P. californicus* was substantially reduced in intensity relative to these same bands in other species even when DNA from the two *P. californicus* individuals was loaded in 5- to 10-fold excess (Figure 2A). Although the *Eco*RV 1.65-kb and *Bgl*II 2.6-kb bands hybridize to LIN2-3′ in all of the species tested except *P. californicus* (Figure 2C) they were not detected in the *P. californicus* DNA even with the additional LIN2.2 oligonucleotide. Similar reduced hybridizations patterns were observed for *P. eremicus* both with L1GEN and with the Lineage 2-specific oligonucleotides (data not shown).

L1 sampling strategy: To determine if *P. californicus* contains Lineage 2 copies, we devised a strategy to identify L1 sequences from either Lineage 2 or from other, unidentified active lineages. To select for copies from any active lineage, and against individual relic copies, PCR primers were synthesized to well-conserved regions identified in the peromyscine and murine L1 alignment (Casavant *et al.* 1996). These primers are sufficiently conserved to amplify a fragment from genomic DNA of distantly related rodent genera including Microtus and Mus. The primers were used to amplify L1 from *P. californicus*, and cloned PCR-derived fragments were screened with the oligonucleotide LIN1-3′. The remaining clones potentially represent members of Lineage 2 or other active lineages. Of the 180 colonies gridded and transferred to nitrocellulose, 143 hybridized with LIN1-3′. Ten of the nonhybidizing clones and three of the hybridizing clones (CalCC7, CalCC2, and CalCC13) were sequenced.

The alignment of L1 copies from *P. californicus*: These *P. californicus* L1 sequences were compared to the previously collected (Casavant *et al.* 1996) *P. maniculatus* and *P. leucopus* L1 copies and are shown in Figure 3. The 19 sequences in the alignment include the 13 from *P. californicus*, 4 inferred sequences, and 2 sequences from Oryzomys, another genus of cricetid rodents included here for comparison. The sequences are divided

into three groups based on the previously identified lineage-specific shared-sequence variants (Casavant *et al.* 1996): the top sequences include an inferred ancestral sequence of cricetid L1 before the split between Lineage 1 and 2 (CRI_ANC) and two Oryzomys (Ory3 and Ory4), the middle group contains Lineage 2 sequences including the youngest inferred *P. leucopus* and *P. maniculatus* master sequences of that lineage (MAN2-MAST and LEU2MAST), and the bottom group contains Lineage 1 sequences including the youngest inferred *P. leucopus* and *P. maniculatus* master sequences (MAN1MAST and LEU1MAST). The lineage-specific variants described in Casavant *et al.* (1996) have been refined based on this more complete data set. Ambiguous characters are listed in the legend of Figure 3. The inferred ancestral L1 sequence is derived, based on parsimony, using the following sequences: Mus (L1Md2), Rattus (L1Rn3A), Cricetus (L1Cg), three peromyscine copies that inserted into the genome before the split between the two lineages (L1Pm55 and L1Pm62, and Leu1-18), and two Oryzomys sequences (Ory3 and Ory4). (See Casavant *et al.* 1996 for GenBank accession numbers and sequence coordinates.) Ambiguous positions are denoted with IUPAC abbreviations. MAN1-MAST, LEU1MAST, MAN2MAST, and LEU2MAST include the most recently acquired shared-sequence variants and thus represent the most recently active master that can be inferred for each lineage.

To be consistent with the previous analysis, the position numbers have remained the same. Where any individual *P. californicus* L1 sequence contains an insertion, a space or hyphen has been inserted into the remaining sequences and numbers to maintain the alignment and numbering. The start and end positions in Figure 3 reflect the region amplified in *P. californicus*, which begins at base 80 and ends at base 519 of the previous alignment (Casavant *et al.* 1996).

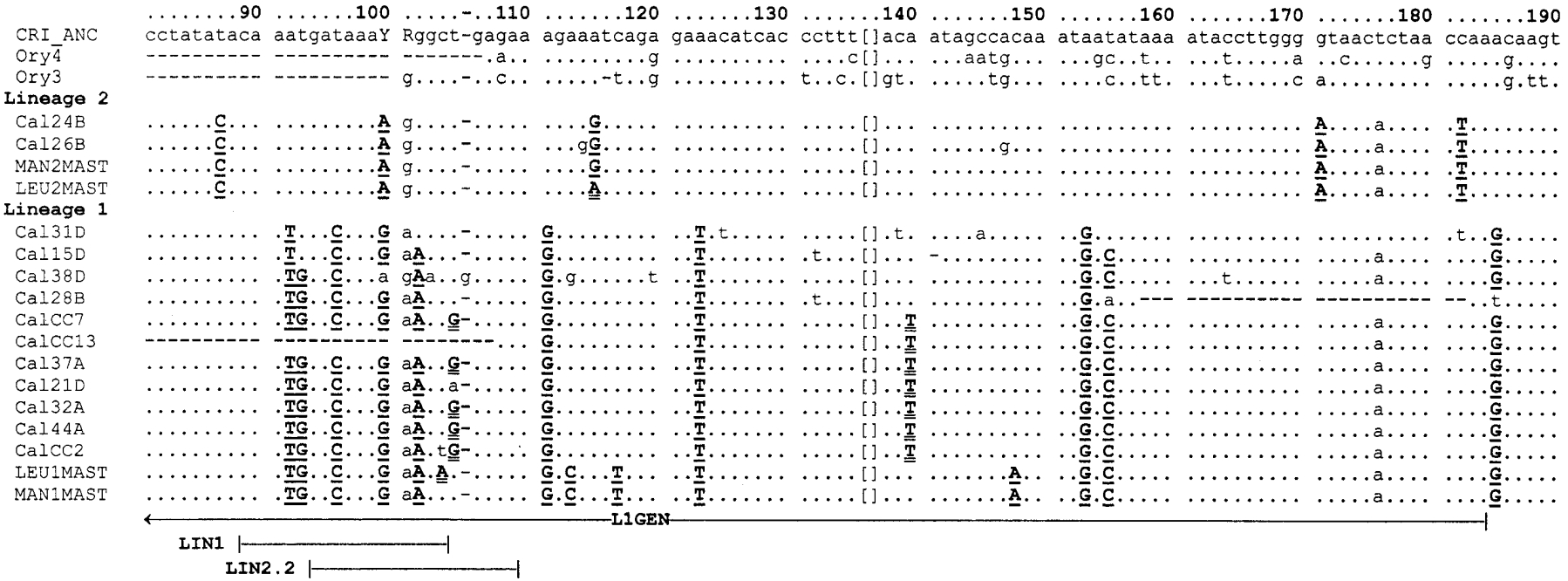
Of the 13 *P. californicus* sequences, 11 contain Lineage 1-specific variants and 2 (Cal26B and Cal24B) contain Lineage 2-specific variants including the loss of the *Bgl* II site. The eight Lineage 1 clones that do not hybridize with LIN1-3′ contain single base differences from the oligonucleotide. The two *P. californicus* Lineage 2 sequences contain the shared-sequence variants in LIN2-3′ but differ in the 5′ extreme base. This difference may prevent or reduce hybridization of the oligonucleotide. However, this does not account for the reduced amount of Lineage 2 detected by genomic Southern blot hybridization (Figure 2), because hybridizing with LIN2.2, an oligonucleotide that does match in all 15 bases, also failed to detect the band in *P. californicus*. *P. californicus* does contain Lineage 2 sequences; however, there is an absence of a discrete hybridizing band with lineage-specific oligonucleotides.

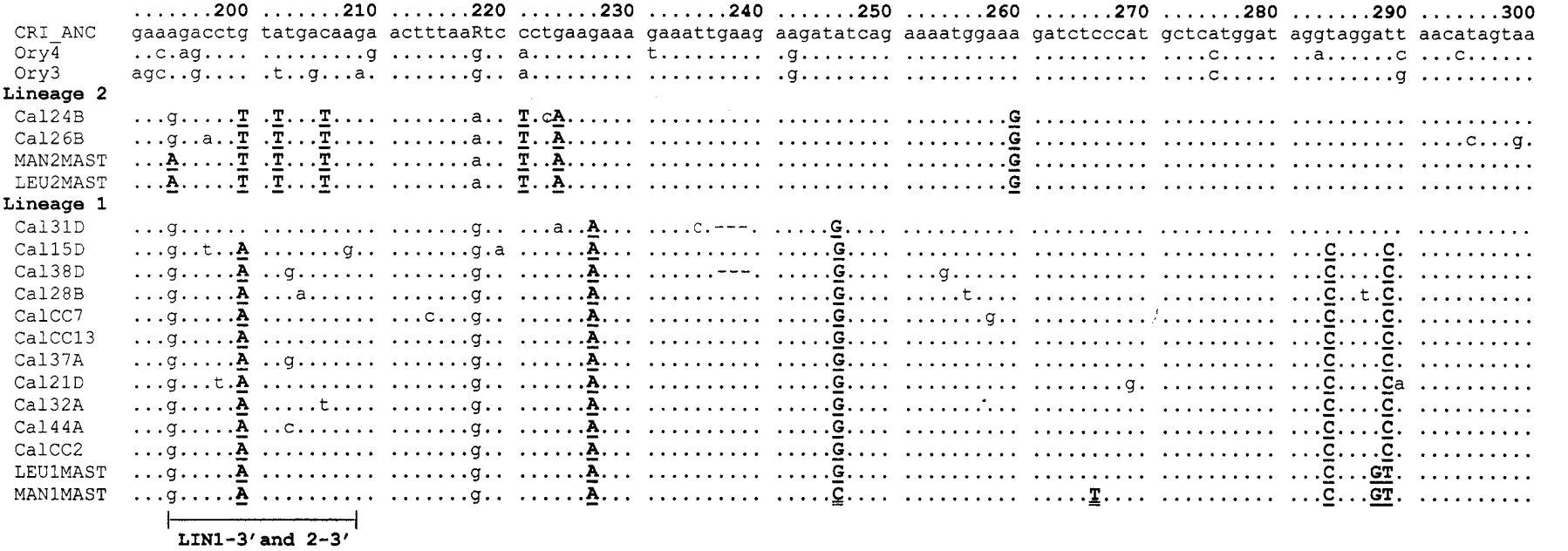
On the basis of this analysis, it appears that *P. californicus* L1 has a preponderance of elements belonging to Lineage 1: 154 Lineage 1 sequences determined by either hybridizing or sequencing, and only 2 Lineage 2 sequences. (The remaining 24 were not assayed for the presence of an insert or analyzed further.) This is in substantial agreement with the results of genomic Southern blot analysis (Figure 2). However, in our hands PCR amplification selects for a younger subset of elements than library screening, so an additional experiment was carried out to allow a more direct interpretation of these results. A PCR-derived library from the *P. maniculatus* specimen TK25740 was similarly constructed. Because no clones were to be sequenced, the library was probed with two different Lineage 1 oligonucleotides, and two Lineage 2 oligonucleotides. Of the 290 colonies gridded, 50 hybridized with the Lineage 1 oligonucleotides and 34 hybridized with the Lineage 2 oligonucleotides. The remaining colonies were not examined for inserts or further characterized. Thus, of those clones from the PCR libraries that hybridized with lineage-specific oligonucleotides, 41% were from Lineage 2 in *P. maniculatus*, compared to less than 1% in *P. californicus*. However, direct comparison of the copy number of Lineage 1 and Lineage 2 in these two speices cannot be made from these data because the within-species estimates are ratios.

L1 phylogenetic tree of three species of Peromyscus: The phylogenetic tree derived from the L1 DNA sequences divides the elements into two distinct lineages (Figure 4). In trees of L1 sequences derived using parsimony, the shared-derived characters correspond to the changes that have been acquired by the master element and observed in the pseudogene copies. The phylogenetic tree reveals the order in which these changes accumulated in the master(s). The inferred ancestral sequence (CRI_ANC) was used to root the tree.

The *P. californicus* sequences divide between Lineage

Figure 3.—Alignment of *P. californicus* L1 sequences with other Peromyscus, Oryzomys and consensus L1 sequences. The numbering of positions in the alignment is consistent with the alignment in Casavant *et al.* (1996). Both lineages are ordered by age, with the oldest sequences at the top. Boldface capital underlined letters indicate lineage-specific variants; boldface capital double underlined letters indicate the youngest and possibly species-specific variants; gaps (-) were introduced to maintain the alignment; dot (.) indicates the same base as in the inferred ancestral sequence CRI_ANC; Y, pyrimidine; R, purine; V, not T. See text for the derivation of the inferred ancestral sequences. Ambiguous positions for CRI_ANC are 100, 101, 218, 382, 384, and 460. There is a possible clade-specific variant for Lineage 2 at site 452 (but see text). Three hypervariable sites (313, 452, and 464) are indicated with an asterisk (*); in some cases the status of the master could not be resolved at these sites, so those changes were not included in either shared-variant or private mutation counts. Heavy black lines underneath the alignment indicate the position of the relevant probe sequence: L1GEN, LIN1, LIN2.2, LIN1-3′, and LIN2-3′.


.....90 .....100 .....110 .....120 .....130 .....140 .....150 .....160 .....170 .....180 190
CRI_ANC cctatataca aatgataaaY Rggct-gagaa agaaatcaga gaaacatcac ccttt[]aca atagccacaa ataatataaa ataccttggg gtaactctaa ccaaacaagt
Ory4 ---------- ---------- -------a.. ....g ....c[]... ...aatg... ...gc..t.. ...t.....a ..c......g .....g....
Ory3 ---------- ---------- g....-..c.. .....-t..g .... t..c.[]gt. .....tg... ....c..tt. ...t.....c a......... .....g.tt.
Lineage 2
Cal24B
Cal26B
MAN2MAST
LEU2MAST
Lineage 1
Cal31D
Cal15D
Cal38D
Cal28B
CalCC7
CalCC13
Cal37A
Cal21D
Cal32A
Cal44A
CalCC2
LEU1MAST
MAN1MAST
L1GEN
LIN1
LIN2.2


.....200 .....210 .....220 .....230 .....240 .....250 .....260 .....270 .....280 .....290 300
CRI_ANC gaaagacctg tatgacaaga actttaaRtc cctgaagaaa gaaattgaag aagatatcag aaaatggaaa gatctcccat gctcatggat aggtaggatt aacatagtaa
Ory4 ..c.ag.... ....g ....g.. a......... t.......... .g........ ....c..... ..a......c ...c......
Ory3 agc..g.... .t..g...a. ....g.. a......... .... .g........ ....c..... ....g ....
Lineage 2
Cal24B
Cal26B
MAN2MAST
LEU2MAST
Lineage 1
Cal31D
Cal15D
Cal38D
Cal28B
CalCC7
CalCC13
Cal37A
Cal21D
Cal32A
Cal44A
CalCC2
LEU1MAST
MAN1MAST
LIN1-3'and 2-3'

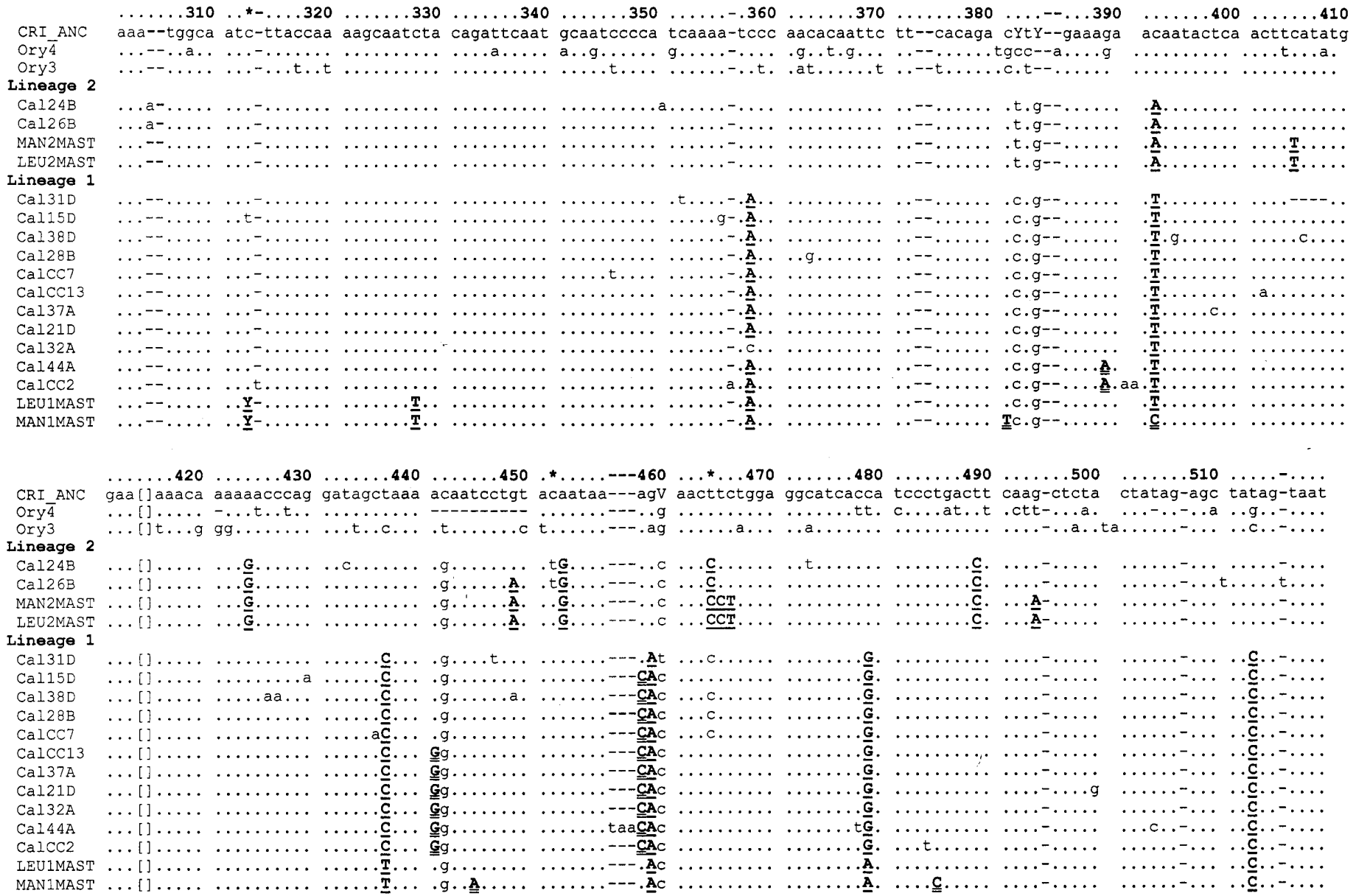


Figure 3.—*Continued.*

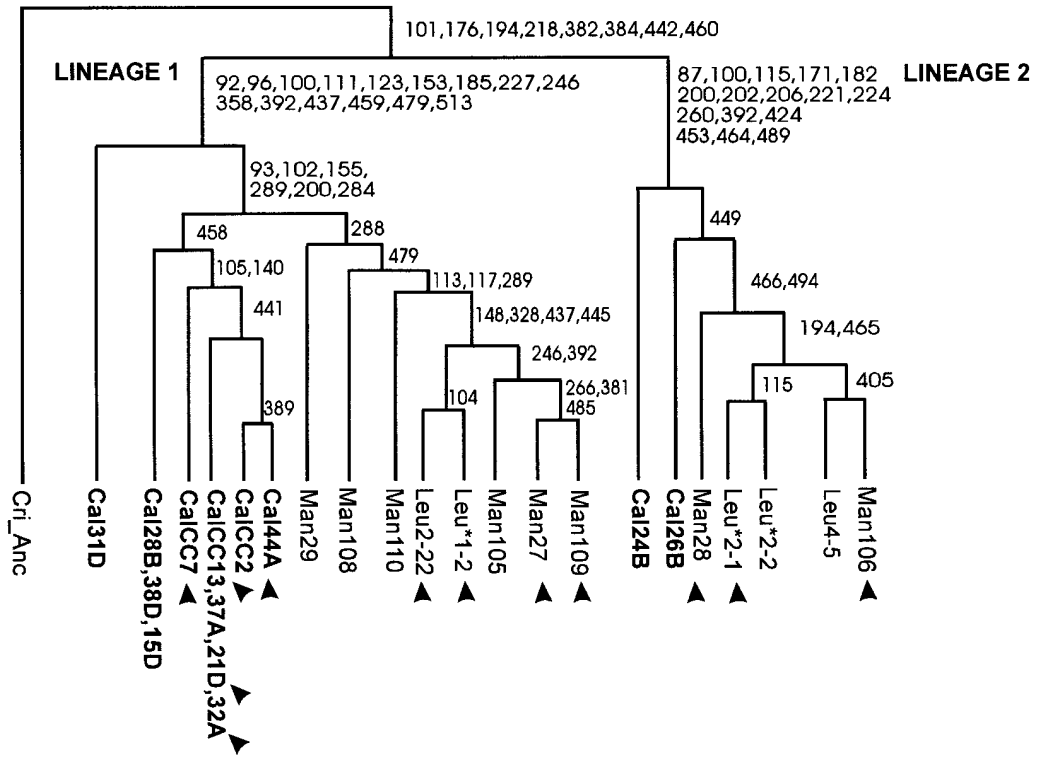


Figure 4.—Phylogenetic tree of L1 elements from three species of Peromyscus. The shared-sequence variants are shown for two L1 lineages. The numbers listed in descending branches indicate the shared-sequence variants. The ambiguous positions (Figure 3) are not included. Arrowheads indicate those sequences that have an open reading frame in the sequenced region. Taxa with the prefix Cal, indicated in boldface, are *P. californicus* sequences, Man are *P. maniculatus* sequences, and Leu are *P. leucopus* sequences. The exact time of change at 101, 218, 382, 384, and 460 relative to the divergence between the lineages cannot be determined, so they are tentatively placed before the divergence. Sites 100 and 392 changed from the inferred cricetid ancestral state early in the history of both lineages and are thus indicated as changes in both lineages. Master elements changed at sites 115, 194, 200, 246, 289, 392, 437, and 479 more than once since the divergence of the two lineages.

1 and Lineage 2 as predicted from the alignment (Figure 4). The *P. californicus* Lineage 1 sequences include Cal44A, CalCC2, CalCC13, Cal37A, Cal21D, Cal32A, CalCC7, Cal28B, Cal38D, Cal15D, and Cal31D. Many of the sampled sequences inserted at different time points in the Lineage 1 history, with Cal31D as the oldest fossil sequence sampled. Such sequences are useful in determining the order in which the master elements acquired changes. The *P. californicus* Lineage 2 sequences include Cal24B and Cal26B. There is no evidence within the collected sequences to suggest that either a divergent Lineage 2 master or a replacement lineage is propagating a significant number of L1 copies in *P. californicus*.

All of the sequences were analyzed to determine the relative time since their insertion into the genome. The number of private mutations, or differences between the individual sequence and its parental template, is an indication of the length of time since that individual element inserted into the genome. Table 2 tabulates changes that occurred in the sequence after insertion into the genome. Neither of the *P. californicus* Lineage 2 sequences (Cal26B and Cal24B) contain open reading frames because of stop codons due to frameshifts caused by base insertions and deletions. Cal26B contains a younger shared-sequence variant than Cal24B and, therefore, may have inserted more recently, but Cal26B contains more private mutations, deletions and insertions. The number of private changes in Cal24B and Cal26B (three and five, respectively) is comparable to the number of changes in the youngest *P. maniculatus* and *P. leucopus* elements (Man106 and Leu4-5). Assuming a neutral mutation rate for rodents (She *et al.* 1990), and not counting the indels, Cal24B, Man106, and Leu4-5 probably inserted within the last two myr, suggesting that the Cal24B inserted after the *P. californicus* species-group diverged. Thus it is likely that *P. californicus* has an inefficient Lineage 2 master template.

Divergence of the L1 master correlates with speciation of the host organism: The division of Lineage 1 into several clades appears to correlate with the speciation of the host organism. The first major division separates *P. maniculatus*/*P. leucopus* sequences from *P. californicus* sequences. Shared-sequence variants at positions 105, 140, 389, 441, and 458 define the *P. californicus* Lineage 1 clade, whereas shared-sequence variants at positions 113, 117, 148, 246, 266, 288, 289, 328, 381, 392, 437, 445, 479, and 485 define the *P. maniculatus* Lineage

TABLE 2

Private sequence changes and insertions/deletions occurring in each L1 copy

Element	No. private	In/del	Open	Total length[a]
		Lineage 1		
Cal31D	9	2	−	422
Cal15D	7	1	−	428
Cal38D	13	2	−	437
Cal28B	7	1	−	404
CalCC7	4	0	+	429
CalCC13	1	0	+	402
Cal37A	2	0	−	429
Cal21D	5	0	+	429
Cal32A	2	0	+	429
Cal44A	3	1	+	432
CalCC2	2	3	−	433
Man29	13	0	−	429
Man108	15	2	−	424
Man110	18	0	−	429
Leu*1-2	6	0	+	429
Leu2-22	0	0	+	321
Man105	0	1	−	428
Man109	2	0	+	429
Man27	2	0	+	429
		Lineage 2		
Cal24B	3	2	−	432
Cal26B	5	3	−	433
Man28	5	0	+	429
Leu*2-1	5	0	+	429
Leu*2-2	9	1	−	430
Leu4-5	2	1	−	423
Man106	2	0	+	429

(+) An open reading frame in the region sequenced; (−) lack of an intact open reading frame in this region.

[a]Only the positions in the *P. maniculatus* and *P. leucopus* sequences corresponding to the region sequenced in *P. californicus* were tabulated, including insertions and deletions. Changes at hypervariable positions 313, 452, and 464 were not included in the counts.

1 clade. A second division of Lineage 1 separates the youngest *P. maniculatus* and *P. leucopus* sequences. A shared-sequence variant at position 104 defines the *P. leucopus* Lineage 1 clade, whereas the shared-sequence variants at positions 246, 266, 381, 392, and 485 define the *P. maniculatus* Lineage 1 clade. In each species-specific clade, the shared-sequence variants are accumulated sequentially so that only the youngest elements share all of these changes.

The sample size is insufficient to determine where Lineage 2 diverges with respect to either *P. californicus*, *P. maniculatus*, or *P. leucopus*. The two *P. californicus* Lineage 2 sequences Cal24B and Cal26B may contain a species-specific variant at position 452; however, the T shared by both sequences might only be a parallel mutation due to the methylation of the C in the CG pair in the master. Additional sequences would be required to verify this change as a *P. californicus*-specific mutation.

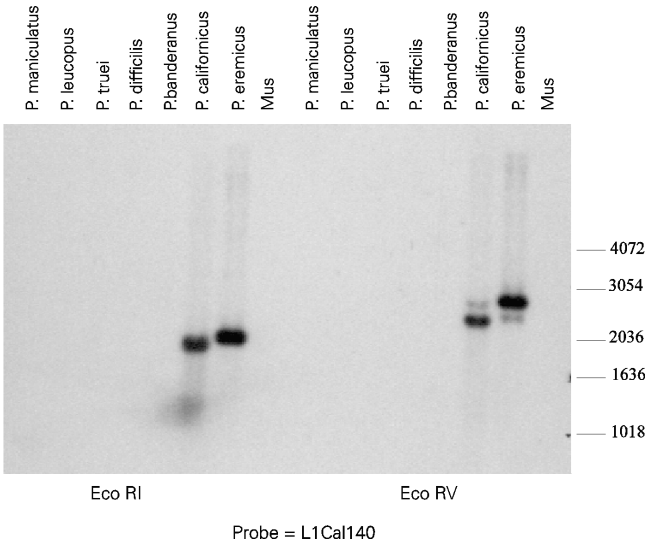


Figure 5.—Differentiating between species with a Lineage 1-specific oligonucleotide probe. The DNA from eight species was digested with *Eco*RV and *Eco*RI and probed with L1Cal140. The L1 oligonucleotide differentiates those that have the shared-sequence variant (*P. californicus* and *P. eremicus*) and also reveals species-specific polymorphisms.

To further investigate the correlation between the divergence of an L1 master and host speciation, an oligonucleotide was synthesized to contain a *P. californicus*-specific L1 shared-sequence variant (L1Cal140: TGG CTATAGTAAAGG). The change at position 140 appears to have been acquired by the Lineage 1 master after the separation of the *maniculatus* and *californicus* species-groups (Figure 4). DNA from eight different species, *P. californicus*, *P. eremicus*, *P. truei*, *P. difficilus*, *O. banderanus*, *P. maniculatus*, *P. leucopus*, and Mus was digested with *Eco*RI and *Eco*RV, blotted, and probed with L1Cal140 (Figure 5). *P. californicus* and *P. eremicus*, a member of a closely related species-group, have discrete hybridizing bands, whereas *P. maniculatus*, *P. leucopus*, *O. banderanus*, *P. truei*, and *P. difficilus* do not. Thus, the change at position 140 occurred before the divergence between *P. eremicus* and *P. californicus*, but after the divergence of this group from the common ancestor of the remaining species. However, after the divergence of *P. eremicus* and *P. californicus*, additional changes in restriction sites were acquired in this subfamily as can be observed from differences in the sizes of the hybridizing band between the two species (Figure 5). *P. eremicus*, like *P. californicus*, lacks the hybridizing Lineage 2 band in both *Eco*RV and *Bgl*II digests (not shown).

Differences in the acquisition of variants among the three Peromyscus species: We reported previously that the Lineage 1 master(s) had acquired twice as many changes as the Lineage 2 master(s), and suggested that the Lineage 1 master(s) were either twice as active or had a less efficient reverse transcriptase than the Lineage 2 master (Casavant *et al.* 1996). The estimation number of shared-sequence variants in *P. maniculatus*

in this larger data set (but with a decreased length of sequence examined) is 35 in Lineage 1, versus 22 in Lineage 2. Because Lineage 2 is not well resolved, the point of divergence between *P. californicus* and *P. maniculatus* cannot be determined for Lineage 2 , and therefore it is difficult to compare the acquisition of changes between the two lineages in *P. californicus*.

In addition to this difference in the rate of molecular evolution between lineages, we have also observed differences among species within a lineage. The number of shared-sequence variants in Lineage 1, and thus the number of accumulated changes in the masters, differs among the three species of Peromyscus examined. Since the divergence of *P. maniculatus* and *P. californicus*, 14 changes have accumulated in Lineage 1 *P. maniculatus* but only 5 changes have accumulated in *P. californicus* (Figures 3 and 4). Similarly, there are many more changes in *P. maniculatus* Lineage 1 since the divergence between *P. maniculatus* and *P. leucopus* (5 changes in *P. maniculatus* and only a single change in *P. leucopus*). The youngest sequences, from which the shared-sequence variants were deduced, appear to be approximately equivalent in age; they have intact reading frames and few private changes (Figure 4 and Table 2). In addition to the more rapid evolution of the master elements in *P. maniculatus*, individual *P. maniculatus* sequences within Lineage 1 appear to have accumulated a greater number of private changes relative to the individual sequences from *P. californicus*. Thus the rate of change appears to be greater in *P. maniculatus* than in either of the other species. A larger sample of L1 and non-L1 sequences will be needed to better evaluate the basis for this rate difference.

DISCUSSION

Increase in transposition rate of Lineage 2 after the divergence of *P. californicus*: In this study we have shown dramatically lower numbers of L1 Lineage 2 copies in *P. californicus* (and *eremicus*) relative to *P. maniculatus*, *crinitus*, *truei*, and *difficilis*. This could be explained by a real scarcity of Lineage 2 elements in *P. californicus* or, alternatively, by our failure to detect elements because of their rapid divergence from other members of the lineage. Several lines of evidence suggest that we are not simply failing to detect Lineage 2 copies in *P. californicus* due to rapid sequence divergence: (1) the absence of Lineage 2-hybridizing bands for two different restriction sites in genomic Southern blot analysis, despite close sequence similarity of the amplified Lineage 2 sequences between *P. californicus* and *maniculatus*; (2) the absence of uncharacterized bands in these same analyses; (3) the presence in *P. californicus* Lineage 2 copies of the same Lineage 2 shared-sequence variants included in the lineage-specific oligonucleotides LIN2-3′ and LIN2.2; and (4) the presence in *P. californicus* Lineage 2 copies of restriction sites common to other Lineage 2 copies. The absence of *P. californicus*-specific shared-sequence variants further suggests that the Lineage 2 master is not rapidly diverging. Therefore, we conclude that our failure to detect many Lineage 2 copies in *P. californicus* reflects the true scarcity of these sequences in the genome, perhaps due to a relatively unproductive Lineage 2 master.

Our data suggest there has been at least one major change in the rate of transposition in Lineage 2 since its divergence from Lineage 1. The difference in transposition rate between *P. californicus* and *P. maniculatus* is inferred from the significant difference in copy number. Although our current data cannot completely differentiate between a slowdown in the rate of transposition in *P. californicus* after divergence from the common ancestor with the *P. maniculatus*/*P. leucopus* clade and an increase in the rate of transposition in the line giving rise to the *P. maniculatus*/*P. leucopus* clade, there is evidence for the latter. The low number of private mutations in Cal24B and Cal26B suggests that these elements inserted into the genome long after the divergence of the two Lineages, which argues against an extinction of Lineage 2 in *P. californicus*. However, failure to detect Lineage 2 by genomic Southern blot analysis suggests that the copy number of Lineage 2 is low in this species. Accordingly, our hypothesis is that species formed before the increase(s) of Lineage 2 transposition rate contain a paucity of these sequences, and those formed after contain a significantly higher copy.

LINE-1 as a tool to study host systematics: The peromyscine L1 lineages may provide phylogenetic information about the systematics of Peromyscus. The low copy number of Lineage 2 in *P. californicus* and *eremicus* has been interpreted here to suggest that these species diverged before the burst(s) of Lineage 2 transposition leading to higher Lineage 2 copy number in *P. crinitus*, *truei*, *maniculatus*, and *leucopus*. However, the absence of Lineage 2 copies is a primitive character. Because more than one burst of Lineage 2 transposition may have occurred during the evolution of Peromyscus, the low copy number of Lineage 2 cannot be used as a character to unite *P. eremicus* and *californicus* to the exclusion of species with higher copy number. Oligonucleotides designed to detect changes in master lineages provide a more definitive phylogenetic tool. The shared-sequence variant at site 140 in Lineage 1 unites *P. eremicus* and *californicus*; a probe made from this variant also exposes polymorphic restriction site differences between these two species (Figure 5). While a close phylogenetic affinity between *P. eremicus* and *californicus* was not recognized in a recent treatment of the genus (Carleton 1989) and is not consistent with a phylogeny of Peromyscus based on karyotypic data (Baker *et al.* 1987), it is consistent with Osgood's 1904 assignment of both species to the subgenus *Haplomylomys* (Nowak 1991). Taxonomic classifications should be supported by more than a single character, and resolution of these conflicting data sets is clearly beyond the scope of this article. However, oligonucleotide hybridization with lin-

eage-specific probes could prove to be a powerful systematic tool for understanding relationships between the host species. This tool must be applied with a full appreciation for L1 dynamics, and may be most useful when applied to high copy number subfamilies. It is important to be cognizant of the limitations of these techniques. Radiolabeled probes, either oligonucleotides or small fragments, may fail to detect restriction bands containing only small numbers of copies. Furthermore, the reduced sequence specificity characteristic of small sequence fragments generates a background smear obscuring less intense restriction bands.

Unequal rates of L1 evolution: Casavant *et al.* (1996) previously showed that Lineages 1 and 2 appear to have unequal rates of L1 evolution. In this article, we further show that evolution appears to change even within a single lineage. Since the divergence of *P. maniculatus* and *P. californicus*, Lineage 1 in *P. maniculatus* has acquired nearly three times as many variants as either Lineage 2 in *P. maniculatus* or Lineage 1 in *P. californicus* (Figure 4). This comparison between Lineages 1 and 2 assumes that changes in the masters at positions 93 and 449 (Lineages 1 and 2, respectively) occurred at the roughly equivalent times. Furthermore, Lineage 1 in *P. maniculatus* has acquired five changes in Lineage 1 while *P. leucopus* has acquired a single change since these species shared a common ancestor (Figure 4). These rate differences represent shared-sequence variants, not private mutations in the individual L1 copies, and therefore the differences in the number of variants reflect changes in the rate of evolution of the master templates.

There are at least four explanations for the large differences in the number of shared-sequence variants in Lineage 1 between *P. californicus* and *P. maniculatus*: (1) sampling error; (2) an increased rate of evolution of master elements due to frequent replacement; (3) an increased rate of evolution of master elements due to a differential reverse transcriptase error rate; or (4) an increased or decreased rate of mutations in a host genome. Errors in sampling could occur if very different age groups are compared. Lineage 1 sequences from *P. maniculatus* and *P. californicus* appear to be close in age based on open reading frames and the low number of private mutations in their youngest sequences, therefore explanation 1 is not likely. Explanation 2 is also unlikely because frequent replacement of the master should lead to divergence of the lineage into multiple clades (Clough *et al.* 1996). The magnitude of change in reverse transcriptase fidelity required to explain these results also makes 3 an unlikely explanation. Although we cannot definitively rule out any of the above explanations, we have developed a statistical method to examine the effects of sampling and possible differences in genomic mutation rates (Joyce *et al.* 1998).

We thank Robert Baker and The Museum of Texas Tech University in Lubbock, Texas, for donating the tissues or DNA for this research. We thank LuAnn Scott for technical assistance and help with manuscript preparation. We especially thank two anonymous reviewers for extensive comments on an earlier version of this manuscript. This work was supported by a grant from the National Institutes of Health (GM39727 to H.A.W.).

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.

Baker, R. J., M. B. Qumsiyeh and C. S. Hood, 1987 Role of chromosomal banding patterns in understanding mammalian evolution, pp. 67–96 in *Current Mammalogy*, Vol. 1, edited by H. H. Genoways. Plenum Publishing Corp, New York.

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.

Burton, F. H., D. D. Loeb, C. F. Voliva, S. L. Martin, M. H. Edgell *et al.*, 1986 Conservation throughout Mammalia and extensive protein-encoding capacity of the highly repeated DNA long interspersed sequence one. J. Mol. Biol. **187:** 291–304.

Cabot, E. L., B. Angeletti, K. Usdin and A. V. Furano, 1997 Rapid evolution of a young L1 (LINE-1) clade in recently speciated Rattus taxa. J. Mol. Evol. **45:** 412–423.

Carleton, M. D., 1989 Systematics and evolution, pp. 7–14 in *Advances in the Study of Peromyscus (Rodentia)*, edited by G. L. Kirkland, Jr. and J. N. Layne. Texas Tech University Press, Lubbock, TX.

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

Casavant, N. C., A. N. Sherman and H. A. Wichman, 1996 Two persistent LINE-1 lineages in Peromyscus have unequal rates of evolution. Genetics **142:** 1289–1298.

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.

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.

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

Duvernell, D. D., B. J. Turner and H. A. Wichman, 1996 An active L1-like LINE mobile sequence family in fish genomes: its variation and divergence in Death Valley pupfishes. Abstr. Am. Soc. Ichthyol. Herpetol., 295–296.

Edgell, M. H., S. C. Hardies, D. D. Loeb, W. R. Shehee, R. W. Padgett *et al.*, 1987 The L1 family in Mice, pp. 107–129 in *Developmental Control of Globin Gene Expression*, edited by G. Stamatoyannopoulos and A. W. Nienhuis. Progress in Clinical and Biological Research, Vol. 251. Alan R. Liss, Inc., New York.

Feng, Q., J. V. Moran, H. H. Kazazian, Jr. and J. D. Boeke, 1996 Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. Cell **86:** 905–916.

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

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

Holmes, S. E., M. F. Singer and G. D. Swergold, 1992 Studies on p40, the leucine zipper motif-containing protein encoded by the first open reading frame of an active human LINE-1 transposable element. J. Biol. Chem. **267:** 19765–19768.

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. American Society for Microbiology, Washington, DC.

Joyce, P., L. Fox, N. C. Casavant and H. A. Wichman, 1998 *Linear*

Estimators for the Evolution of Transposable Elements (Institute for Mathematical Statistics Monograph Series, Vol. 33), edited by F. Seillier-Moiseniwitch. Institute for Mathematical Statistics, Hayward, CA (in press).

Jurka, J., E. Zietkiewicz and D. Labuda, 1995 Ubiquitous mammalian interspersed repeats (MIRS) are molecular fossils from the Mesozoic era. Nucleic Acids Res. **23:** 170–175.

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

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.

Martin, S. L., 1991 LINEs. Curr. Opin. Genet. Dev. **1:** 505–508.

Modi, W. S., 1996 Phylogenetic history of LINE-1 among Arvicolid rodents. Mol. Biol. Evol. **13:** 633–641.

Mulhardt, C., M. Fischer, P. Gass, D. Simon-Chazottes, J. L. Guenet *et al.*, 1994 The spastic mouse: aberrant splicing of glycine receptor beta subunit mRNA caused by intronic insertion of L1 element. Neuron **13:** 1003–1015.

Nowak, R. M., 1991 White-footed mice, or deer mice, pp. 659–662 in *Walker's Mammals of the World*, Ed. 5. Johns Hopkins University Press, Baltimore and London.

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

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

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.

She, J. X., F. Bonhomme, P. Boursot, L. Thaler and F. Catzeflis, 1990 Molecular phylogenies in the genus Mus: Comparative analysis of eletrophoretic, scnDNA hybridization, and mtDNA RFLP data. Biol. J. Linn. Soc. **41:** 83–103.

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.

Southern, E. M., 1975 Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98:** 503–517.

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

Usdin, K., P. Chevret, F. M. Catzeflis, R. Verona and A. V. Furano, 1995 L1 (LINE-1) retrotransposable elements provide a "fossil" record of the phylogenetic history of muroid rodents. Mol. Biol. Evol. **12:** 73–82.

Verneau, O., F. Catzeflis and A. V. Furano, 1997 Determination of the evolutionary relationships in Rattus sensu lato (Rodentia : Muridae) using L1 (LINE 1) amplification events. J. Mol. Evol. **45:** 424–436.

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

Xiong, Y., and T. H. Eickbush, 1990 Origin and evolution of retroelements based upon their reverse transcriptase sequences. EMBO J. **9:** 3353–3362.

Communicating editor: W.-H. Li