SINE extinction preceded LINE extinction in sigmodontine rodents: Implications retrotranspositional dynamics and mechanisms

[1,2]Timothy A. Rinehart, [1,3]Robert A. Grahn and [1]Holly A. Wichman

[1]University of Idaho
Department of Biological Sciences, 443051
Moscow, ID 83844-3051

[2]Current Address: USDA Small Fruit Research Station
306 South High Street
P.O. Box 287
Poplarville, Mississippi 39470

[3]Current address: College of Veterinary Medicine
Department of Population Health and Reproduction
University of California
Davis, California 95616

Running Head:

SINE and LINE extinction

Keywords:

LINE-1
Sigmodontinae
transposable element
B1
SINE
Corresponding author:

Holly A Wichman
Dept. of Bio. Sci.
Univ. of Idaho
Moscow ID 83844-3051
(208) 885-7805
(208) 885-7905 (fax)
hwichman@uidaho.edu

Abstract

Short Interspersed Nuclear Elements, or SINEs, retrotranspose despite lacking protein-coding capability. It has been proposed that SINEs utilize enzymes produced in *trans* by Long Interspersed Nuclear Elements, or LINEs. Strong support for this hypothesis is found in LINE and SINE pairs that share sequence homology; however, LINEs and SINEs in primates and rodents are only linked by an insertion site motif. We have now profiled L1 LINE and B1 SINE activity in twenty-four rodent species including candidate taxa for the first documented L1 extinction. As expected, there was no evidence for recent activity of B1s in species that also lack L1 activity. However, B1 silencing appears to have preceded L1 extinction, since B1 activity is also lacking in the genus most closely related to those lacking active L1s despite the presence of active L1s in this genus. A second genus with active L1s but inactive B1s was also identified.

Introduction

SINE retrotransposition must be dependent upon enzymes produced in *trans* because SINEs lack protein-coding capability. A generally accepted hypothesis is that LINEs, a widespread family of retrotransposons and potential source of reverse transcriptase, drive SINE retrotransposition at the molecular level. Support for this postulate has grown with the identification of LINE counterparts to some SINE families. Sequence homology between LINEs and tRNA-derived SINEs is found at their 3' ends, the initiation site for

reverse transcription (Ohshima *et al.*, 1996). While the reverse transcriptase encoded by LINEs has been shown to have *cis* preference during LINE retrotransposition, it is utilized in *trans* during pseudo-gene formation and, presumably, during SINE retrotransposition (Esnault *et al.*, 2000). LINE-SINE pairs are widespread and have been found in mammals, salmonids, reptiles, insects, fungi and plants (Okada *et al.*, 1997). Some pairs such as HE1 SINEs and HER1 LINEs in elasmobranchs are estimated to be over 100 million years old, suggesting this retrotranspositional duet is a long-term arrangement (Ogiwara *et al.*, 1999).

L1 LINEs and *Alu* or B1 SINEs, the predominant LINEs and SINEs in primates and rodents respectively, do not share sequence homology. However, similarities between L1 and B1 insertion sites support the hypothesis that L1 encoded proteins may be necessary for B1 retrotransposition. These sites were previously characterized as A/T rich regions, implying non-random insertion into fortuitous DNA nicks as chromosomal "Band-Aids" (Moore and Haber, 1996; Teng *et al.*, 1996). Refined alignments of L1 insertion sites have identified an upstream consensus sequence "TTAAAA" at 53% of the sites, with approximately 80% of the sites having some similarity to this sequence (Jurka, 1997). This target site implies that there is an enzymatic requirement for insertion. The identification of an L1 encoded apurinic/apyrimidinic endonuclease domain and its *in vitro* nicking activity at L1 consensus sites further refines the molecular model of L1 retrotransposition (Feng *et al.*, 1996). Mammalian SINEs such as *Alu* in primates and B1, B2 and ID in rodents share this insertion site motif at approximately the same frequency as L1 (Jurka and Klonowski, 1996). In this paper we offer evidence based upon activity profiles in rodents that further strengthens the association between L1 and B1 retrotransposition.

L1 amplification accounts for as much as 10-20% of the rodent genome (Furano, 2000). The full length L1 is over six kilobases and includes two open reading frames (ORFs) and an internal RNA polymerase II promoter. The first ORF encodes a single-strand nucleic acid binding protein, while ORF2 encodes a reverse transcriptase with an endonuclease domain (Feng *et al.*, 1996). Many L1 sequences are truncated at the 5' end suggesting that reverse transcription of a full length L1 is rare and most retrotransposed copies are non-functional pseudogenes. These copies are scattered throughout the genome by the tens of thousands and serve as molecular fossils of past retrotransposition events (Furano, 2000).

B1 retrotransposons are also widespread and prevalent with more than 80,000 copies detected in voles, rats, mice and deer mice (Deininger *et al.*, 1996). They are short, with an average size of 148 basepairs, and contain an internal RNA polymerase III promoter. Much like L1, B1 RNA copies are reverse transcribed at or before insertion. tRNA-derived SINEs such as B2 and ID, which have a composite structure consisting of a tRNA homologous region, a potentially LINE related region and a terminal AT rich region. However, *Alu* and B1 are derived from 7SL RNA genes (Zietkiewicz and Labuda, 1996) and lack L1 related regions. B1 copies accumulate mutations rapidly after insertion, partly due to the high CpG content; that, along with their short length, makes phylogenetic reconstruction difficult. However, B1 sequences are well-characterized in mice and rats and have been surveyed in hamsters, voles, deer mice, chipmunks and guinea pigs (Kass *et al.*, 2000; Quentin, 1989; Zietkiewicz and Labuda, 1996).

Sigmodontine rodents appear to lack recent L1 activity and are ideal for a comparative analysis of linked L1 and B1 retrotransposition (Casavant *et al.*, 2000; Grahn *et*

al., submitted). By profiling retrotransposition within a phylogenetic framework, we compared taxa in which L1s are inactive to sister taxa in which they are active in order to test the hypothesized connection between L1 and B1 activity, as well as to define the boundaries of L1 and B1 extinction events. Although Sigmodontine rodents contain more than one SINE family, we have chosen to profile B1 retrotransposons because they share the same gene of origin as *Alu* elements in primates, and because their slightly larger size makes them more amenable to analysis. In addition to providing a model system for studying retrotransposon evolution, these species are of systematic interest because of their rapid South American radiation and unresolved phylogeny, and because unusually high karyotypic variation has also been observed in four genera of sigmodontine rodents, *Oryzomys*, *Holochilus*, *Akodon* and *Rhipidomys* (Barros *et al.*, 1992; Fagundes *et al.*, 1998; Koop *et al.*, 1983; Nachman, 1992; Nachman and Myers, 1989; Silva and Yonenaga-Yassuda, 1999).

To test B1 dependence upon L1 activity, we reconstructed the evolutionary history of retrotransposition by sampling elements dispersed in the genome. Three comparative and relatively independent criteria were used to estimate the recent activity of L1 and B1 populations. First, sequence divergence was determined for the aligned retrotransposon sequences of each species. Low sequence divergence indicates recent duplication, while high sequence divergence suggests ancient insertion and the accumulation of private mutations. Minimum divergence, in this case the average number of independent mutations between the four or five youngest sequences, reflects the time of most recent insertions because both L1s and B1s retrotranspose from a few or small group of closely related "master" elements (Furano *et al.*, 1994; Quentin, 1989).

Second, relative copy number was determined by genomic dot blot hybridization to L1 and B1 probes. Retrotransposition over evolutionary time should increase copy number within the genome because mechanisms for actively removing retrotransposons have not been observed. However, as elements sit in the genome, they will accumulate mutations and eventually become unrecognizable by sequence similarity and undetectable by molecular probes. Thus low levels of hybridization imply reduced retrotransposition, while high copy number suggests continued recent retrotransposition. Because retrotransposition rates are not expected to be constant over evolutionary time, this measure does not discriminate between a long period of low-level retrotransposition and a single, recent burst of activity.

Finally, L1 and B1 clones were analyzed for potential functionality in order to detect sequences of high similarity that were duplicated by a mechanism other than retrotransposition. A portion of the L1 ORF2 was analyzed for protein coding integrity. Impaired or interrupted ORF2 suggests a loss of activity while an intact ORF2 was used as a marker of potential function. Measures of function for B1 retrotransposons were based upon the conserved RNA secondary structure critical for B1 retrotransposition (Labuda *et al.*, 1991; Labuda and Zietkiewicz, 1994).

Singularly, these criteria are broad, but taken together we find exceptional consistency when classifying retrotransposon populations as active or non-active in the twenty-four species examined. Recently active, or "young," L1 and B1 populations were characterized by low sequence divergence, high copy number and a high percentage of potentially functional clones. Conversely, species exhibiting high sequence divergence, low copy number and a paucity of potentially functional clones were classified as having "old,"

or non-active retrotransposon populations. Five of the twenty-four species examined show high levels of L1 and B1 activity while sixteen apparently lack L1 and B1 activity, supporting B1 dependence on L1 activity. Interestingly, three species show L1 activity in the absence of B1 activity.

Materials and Methods

Species Identification and Genomic DNA Extraction Tissue samples were graciously donated by the Museum of Texas Tech University (TK numbers), Jack Sullivan at the University of Idaho (JS numbers), the Museum of Natural History collection at University of New Mexico (NK numbers), Texas A&M University Texas Coop Wildlife Collection TCWC (AK and MUR numbers) and the University of Michigan (UUMZ numbers). Species include *Phyllotis xanthopygus* (AK13012), *Peromyscus maniculatis* (TK24121), *Oryzomys albigularis* (MUR17), *Oryzomys palustris* (TK28621), *Sigmodon hispidus* (MUR15), *Oligoryzomys fornesi* (NK22527), *Holochilus brasiliensis* (UMMZ166480 and NK13055), *Oryzomys nitidus* (NK13451), *Reithrodontomys fulvescens* (TK21614), *Peromyscus leucopus* (TK27127), *Microtus arvalis* (TK44790), *Sigmodon mascotensis* (JS2013), *Thomasomys baeops* (NK27679), *Rhipidomys nitela* (NK21695), *Akodon boliviensis* (NK11561), *Oxymycterus paramensis* (NK22836), *Nectomys squamipes* (NK13407), *Microryzomys minutus* (NK25822), *Neacomys spinosis* (NK25265), *Oecomys bicolor* (NK12701), *Calomys callosus* (NK37800), *Callomys tener* (NK21054), *Nyctomys sumichrasti* (NK4306) and *Peromyscus nudipes* (NK17807). Genomic DNA extractions were performed using a standard protocol (Longmire *et al.*, 1988).

Because species misidentification or human error during tissue sampling and labwork could greatly confound our analysis, taxon identification was confirmed by sequencing the first 750 basepairs of the *cytochrome b* gene (Sullivan *et al.*, 1996). Sequences were compared to published or deposited sequences where applicable. For species where *cytochrome b* had not been previously characterized, phylogenetic analysis (not shown) was used to confirm that specimens had the expected taxonomic affinities. Seven samples including *Oryzomys nitidus* (NK13451), *Microtus arvalis* (TK44790), *Rhipidomys nitela* (NK21695), *Akodon boliviensis* (NK11561), *Microryzomys minutus* (NK25822), *Oecomys bicolor* (NK12701) and *Peromyscus nudipes* (NK17807) could not be confirmed because the DNA supply was exhausted during L1 and B1 analysis. Nevertheless, the positions of L1 and B1 sequences from these species were as expected on phylogenetic trees.

B1 PCR Amplification B1 retrotransposons were PCR amplified from genomic DNA. PCR reactions used the following conditions: 94° for 2 minutes then 25 cycles at 94°, 56°, 72° for 30 seconds each, followed by 72° for 6 minutes. The B1-forward 5'GCCGGGCGTGGTGGCG3' and B1-reverse 5'TTGGTTTTTCGAGACAGGGTTTCT3' primers are modified consensus primers (Labuda *et al.*, 1991). For each species, amplified B1 sequences from at least three independent PCR reactions were mixed in equal amounts and cloned into the pGemT-Easy vector (Promega, Madison, WI), and at least twenty random colonies were isolated for sequence analysis.

L1 ORF screen A degenerate PCR and color screening technique (Cantrell *et al.*, 2000) was used to isolate part of the ORFII of the L1 retrotransposons from position 4969 to

5583 This cloning system is designed to produce an L1 / *lac*Z fusion protein if the cloned sequence has an intact open reading frame. Thus, colonies with a blue phenotype should contain intact PCR-amplified open reading frames while white colonies are predicted to contain PCR products with premature stop codons. In this manner, we could enrich our sample for clones having intact ORFs by sequencing the inserts from more than 10 random blue colonies from each species. Several random white colonies were also picked and sequenced for comparative purposes.

Sequence divergence and RNA folding Clones were sequenced on both strands using a 377 ABI automated sequencer (Perkin Elmer, Norwalk, CT). Duplicate clones, non-L1 and non-B1 containing clones, and null clones were eliminated from the remaining analysis. The number of L1 or B1 sequences for each species varied from 6 to 42 (Table 1). L1 and B1 sequence data within each species were aligned and compared using the *MegAlign* program (DNASTAR, Madison, WI). The final alignment was refined manually. The minimum sequence divergence within a species represents the average percent difference between the four most similar L1 sequences from the alignment or the average percent difference between the five most similar B1 sequences. B1 number was increased to diminish the effects of a single sequence due to their short length. This reduces the total dataset of 458 L1 and 441 B1 to a subset of 96 L1 and 120 B1 sequences when determining minimum divergence.

All B1 sequences were subjected to minimum energy RNA folding using *GeneSys* by DNASTAR. Two structures, one that contains guanine-uracil pairing and one that does not contain guanine-uracil pairing, were generated at 37° for each sequence. Minimum energy RNA folding predictions were compared to the empirically derived B1 and 7SL

structures (Figure 1). Structures were scored for two of the three domains; Domain I could not be examined as it contains the primer sequence used to extract B1 clones from the genome. Clones that folded into structures similar to Domains II and III were scored as potentially functional. Comparisons were initially made by nucleotide position with Domain II defined as a hairpin between nucleotides 28 to 42. However, all secondary structures were checked by eye to compensate for insertions and deletions that might shift the structure outside the defined nucleotide positions. Domain III was scored positive even when it included combinations of local pairing not present in the empirically derived structure as long as the 3' end of the RNA returned to the 5' end in a single loop. Bulges and hairpins larger than 12 basepairs were considered disruptive. Some of the significantly smaller B1 sequences folded into structures scored as potentially functional but this was rare and did not change the overall results. The number of clones that scored positive for both domains is expressed as a percent of the total clones for that species.

Relative copy number by quantification of dot blots Genomic DNAs were quantified using Hoescht dye and an LS30 Luminesence Spectrophotometer (Perkin Elmer, Norwalk, CT). DNA for each species was dot-blotted onto a charged nylon membrane using a 96 well vacuum apparatus (BIORAD, Hercules, CA). Three dilutions for each DNA sample (1000ng, 200ng, 50ng) were blotted along with positive and negative controls. Replicate filters were hybridized independently with B1 and L1 probes. Radiolabeled probes were generated by mixing equivalent amounts of B1 PCR product or L1 plasmid DNA from all species and random-prime labeling (RadPrime Kit by GIBCO, Rockville, MD). In this manner, species and lineage specific hybridization was reduced. Hybridization and washes were carried out at low stringency (55°, 6mM Sodium Citrate,

0.6mM NaCl and 0.05% Sodium Dodecyl Sulfate). Filters were quantified by phosphor-imaging. Radioactive counts were converted to a percentage, called relative copy number, by dividing hybridization counts for a species by the spot with maximum counts on the filter (*Reithrodontomys fulvescens*).

Statistical analysis Correlation matrices and statistical values were generated under a Pearson product-moment correlation (Statistica, Tulsa, OK). Regressions were generated for all pairwise comparisons of the three criteria within the B1 and L1 datasets (Figure 2), and to compare the minimum sequence divergence of B1 elements within each species examined to that of L1 elements (Figure 5).

GenBank accession numbers Cytochrome b sequences (AY041185 to AY041206), L1 sequences (AY041207 to AY041643), and B1 sequences (AY041644 to AY042081) are available in GenBank. L1 and B1 sequences are labeled by a shortened species name that includes the first letter of the genus and the first three letters from the species, a unique clone identification number and a colony phenotype. For example, the twenty-third L1 sequence from *Akodon boliviensis* that had a blue colony phenotype would be labeled "Abol23b". B1 sequences are labeled similarly except that the shortened species name is followed by a tissue sample number and unique letter for each clone.

Results

Minimum sequence divergence, relative copy number and functional predictions for L1 populations from 24 rodent species were examined as indicators of L1 activity (Table 1). Significant correlation was found between the three L1 activity criteria. Minimum sequence divergence is negatively correlated with relative copy number and negatively correlated with

potential functionality ($r = -0.85$ and -0.90 respectively). Relative copy number is positively correlated with functionality (r–value = 0.81; $p \ll .001$ for all comparisons). Thus, all three L1 activity criteria give a consistent picture and can all be considered rough indicators of the last time of L1 activity. Furthermore, the regression plots show a bimodal distribution, with data points clustering in two distinct groups (Figure 2a, b, and c). This demonstrates a clear split between L1 active and L1 inactive taxa rather than a continuum of L1 activity.

All three criteria can be plotted on a single graph (Figure 3). As predicted, species with high sequence divergence exhibit relatively low copy number and a low percentage of potentially functional clones (distribution skewed to the left). Low sequence divergence corresponds with high copy number and a high percentage of potentially functional clones (distribution skewed to the right). Thus, data points clustered on the right side of the graph indicate species with active L1 populations while data points clustered on the left suggest species in which L1 retrotransposons are inactive. The recovery of intact L1 ORFs, shown as darkened diamonds for emphasis, appears to be a good measure of activity, as no species lacking an intact ORF met the criteria for having an active L1 population.

Profiles of B1 activity also consisted of three criteria: minimum sequence divergence, relative copy number and predicted RNA folding as a surrogate for functionality. The three measures for B1 activity also demonstrate significant correlations. B1 sequence divergence is negatively correlated with copy number ($r = -0.82$) and with potential B1 functionality ($r = -0.81$). B1 copy number and functionality are positively correlated ($r = 0.73$; $p \ll .001$ for all comparisons). Thus, species with recent B1

amplifications contain many, closely related and potentially functional elements, and these criteria demonstrate linear relationships as we look at older amplification events.

The bimodal distribution between putatively active and non-active B1 populations is also clearly defined (Figure 2d, e, and f). When all three criteria are plotted on a single graph (Figure 4), points skewed to the left correspond to species with inactive B1 populations, while data points skewed to the right indicate species with active B1 retrotransposons. The single exception may be *Nyctomys sumichrasti* where the measures for all three criteria appear intermediate. While these data do not indicate recent activity when compared with *Peromyscus*, *Reithrodontomys* and *Microtus* genera, they are not entirely consistent with the proposed B1 extinction in the other South American rodents such as *Oryzomys* and *Sigmodon* genera. The remaining five species from the *Peromyscus*, *Reithrodontomys* and *Microtus* genera have active B1 populations.

To reconstruct their retrotransposon history, the activity profiles for L1 and B1 populations were overlaid on a phylogenetic framework that shows the currently accepted relationships between the rodents examined. The phylogenetic tree in Figure 6 represents a conservative estimate of the rodent phylogeny and is based on multiple phylogenetic studies using morphological and genetic data (Dickerman and Yates, 1995; Engel *et al.*, 1998; Smith and Patton, 1999; Steppan, 1995; Minin *et al.*, 2003). While there is strong support for several key nodes, many of the relationships remain unresolved, particularly within the South American rodents. As expected, all taxa with inactive L1s also lack B1 activity. However, the overlay clearly shows the unexpected result that B1 extinction preceded L1 extinction in the Sigmodontinae.

Discussion

Evolutionary history of the Sigmodontinae The subfamily Sigmodontinae represents one of the largest recent radiations in mammals and includes over 300 species that occupy a wide variety of habitats in South America (Steppan, 1996). The group clearly originated in North America, but there is considerable debate about whether a single ancestral form arrived in South America prior to dispersal and speciation, or whether several or many derived lineages crossed into South America before radiating outward (reviewed by Engel *et al.*, 1998)). Either way, the phylogenetic ambiguities seen among the sigmodontines are thought to be due to a massive, rapid radiation event after the introduction of the ancestral forms into South America. The hypotheses for an "early" or "late" South American radiation have two commonalities. First, Sigmodontinae remain a monophyletic group under either hypothesis and the many diverse species present today share a common ancestor (Smith and Patton, 1999; Steppan, 1996). Second, the genus *Sigmodon* is the most basal genus in Sigmodontinae (Engel *et al.*, 1998; Steppan, 1996). The population dynamics and speciation patterns in these rodents may have played an important role in the dynamics and extinction patterns in their genetic parasites.

Extinction of L1s and B1s The sigmodontine rodents are the only group of mammals in which L1 extinction has been documented (Casavant *et al.*, 2000; Grahn *et al.*, submitted). Because SINEs are though to parasitize the retrotranspositional machinery of LINEs, we hypothesized that B1 SINEs should also be extinct in this group. We characterized retrotransposon activity in a broad range of sigmodontine rodents – 18 species from 14 genera found in five of the nine tribes – and in representatives of three related

rodent subfamilies (Tylominae, Neotominae, and Arvicolinae; Figure 6). Three criteria were used to assess both L1 and B1 activity – minimum sequence divergence, relative copy number and potential functionality. For both L1s and B1s these three criteria were highly correlated and gave a consistent picture of retrotranposon activity. Activity profiles were bimodally distributed, with a clear group of species containing active elements and another group of species containing inactive elements in each case (Figures 2, 3 and 4). For the most part, the groupings of active and inactive elements were consistent between L1s and B1s. As expected, all species with inactive L1s also lacked active B1s. These data are consistent with the insertion site evidence for conjoined LINE/SINE activity and identify rodent taxa lacking both L1 and B1 activity. However, three species with active L1s also lacked active B1s (Figure 5). Thus the extinction events for these two retrotransposons are not entirely coincident.

To understand the evolutionary dynamics of L1 and B1 extinction, the activity profile results were overlaid on a phylogenetic tree of the host species (Figure 6). The typical hierarchical structure missing from the sigmodontine tree does not affect interpretation of the L1 results because the critical node is strongly supported in the literature. The L1 extinction identified here appears to have occurred as a single event within the Sigmodontinae, after the split between the lineage giving rise to *Sigmodon* and the common ancestor of the rest of the subfamily (Figure 6,♦; (Grahn *et al.*, submitted)). However, interpretation of the B1 results is more problematic. Because *Sigmodon hispidus* and *S. mascotensis* contain active L1 but inactive B1 populations, the disappearance of B1 activity appears to have preceded L1 extinction. This is consistent with the directional interaction between the LINE/SINE pairs described above – a species can have active L1s

without active B1s, but the reverse situation is not possible if B1s require active L1s for their own movement. B1s are also inactive in *Nyctomys sumichrasti.* This species is a member of the rodent subfamily Tylominae which, in contrast to the Sigmodontinae, contains only four genera and a total of only ten species. The position of the Tylominae relative to the remaining rodent subfamilies examined is unresolved, so it is not clear if B1 extinction occurred twice, as indicated by the ◊ in Figure 6, or once, in the common ancestor of the Tylominae and Sigmodontinae. Of course, even if Tylominae and Sigmodontinae are sister taxa, the B1 extinction event could have occurred independently in each group. The intermediate position of *N. sumichrasti* for all three measures of B1 activity (Table 1; Figure 2, ◊; Figure5) could simply represent noise in the data, could be due to a lower mutation rate in Tylominae than in Sigmodontinae, or could suggest that cessation of B1 transposition may have occurred more recently in Tylominae than in Sigmodontinae. While the time of B1 extinction and the number of events involved cannot be resolved with the current data, it is clear that B1 extinction preceded L1 extinction in the Sigmodontinae.

Possible explanations for differences between L1 and B1 extinctions Our prediction going into this study was that L1 and B1 activity or extinction would be strictly correlated because B1 extinction would follow from L1 extinction. How then do we explain the observation that B1 extinction appears to have preceded L1 extinction? We entertain several possibilities: 1) the events are not causally related, 2) dependence of L1s on B1s, 3) an arms race involving a third genetic parasite, or 4) gradual L1 extinction with recovery in some species.

Hypothesis 1: The L1 and B1 extinction events are not causally related. It is possible that changes in L1 and B1 populations are attributable to independent random events. Stochastic processes such as genetic drift might effectively eliminate master sequences from the host population, resulting in extinction of the retrotransposon family in that species and in all species that arise from it over the course of evolution. Such stochastic events appear to be relatively rare in mammals since SINEs, and to a greater extent L1s, have long histories of persistence across evolutionary time. However, the population dynamics of the host species during the massive radiation event in sigmodontine rodents may include severe population bottlenecks that could facilitate stochastic loss of master elements.

Hypothesis 2: L1 transposition is dependent on B1s. The alternate explanation for the extinction of B1s before L1s is that SINE activity is needed for L1 retrotransposition, so that B1 extinction is causal to L1 loss. Under this hypothesis another SINE would have replaced this function of B1s in *Sigmodon* and *Nyctomys*, allowing L1 function to recover. Our current understanding of transposon biology does not support this hypothesis. Furthermore, individual SINE families are more narrowly distributed in mammals than are L1s: SINE families tend to be limited to a single mammalian order while L1s have been found in all mammals examined to date. This suggests that SINEs can turn over without jeopardizing the persistence of L1s.

Hypothesis 3: L1 and B1 extinction was due to an arms race involving a third genetic parasite. B1s are not the only SINE family in rodents. It is possible that a variant of B2, ID or another SINE family became so successful that it out-competed B1s (and perhaps even L1s) for cellular factors or the L1 transpositional machinery, leading to the extinction

of B1s and then of L1s in some lineages. The periodic cycling of retrotransposon activity, or patterns of bursts and quiescence, may have profound effects on retrotransposon life history, especially of the parasitic SINE populations. Selection for *de novo*, highly-active RNA structures could occur during the retrotransposition process where small advantages would confer large benefits to competing SINE masters. The predicted result is rapid cycling of B1 masters and independent evolution of SINE populations. This has been observed for B1 populations in murids where sub-families are defined by diagnostic substitutions but often share mutations, suggesting a mosaic evolution (Zietkiewicz and Labuda, 1996).

Hypothesis 4: Gradual L1 extinction with recovery in some species. Perhaps the most plausible explanation for the observed data is that L1 extinction was gradual. In this manner, a reduction in L1 activity below some threshold level might have eliminated B1 activity early on, presumably because SINE master sequences turnover more quickly than LINEs. A subsequent increase in L1 activity in the *Sigmodon* (and perhaps *Nyctomys*) ancestor after it diverged from the ancestor of the remaining Sigmodontinae tribes would not necessarily result in B1 recovery. Recovery of L1 activity did not occur in the ancestor of the Akodontini, Thomasomyini, Phyllotini and Oryzomyini tribes. It is worthwhile to note that an inactivated L1 might be reactivated by recombination. It is significant that even species lacking L1 activity are not completely devoid of L1 sequences. The residual sequences present in the genome could provide the raw material for L1 resurrection especially before they were hammered with point mutations, insertions and deletions. Recombination is known to be an important component of L1 life history (Cabot *et al.*, 1997; Hayward *et al.*, 1997; Saxton and Martin, 1998).

We cannot currently differentiate between these four hypotheses, although we have weak support for Hypothesis 4. We sampled L1s from 16 species in which L1s were inactive and identified 137 that inserted into the genome after that group last shared a common ancestor with *Sigmodon* and calculated the mean divergence from a common ancestor to estimate the time of the L1 extinction at approximately 8.8 million years ago (Grahn *et al.*, submitted). If L1 extinction was instantaneous, we would expect the mean and the variance to be equal under a Poisson distribution. However, the variance was 3.2-fold higher than expected. One factor that could account for the high variance of the estimate is an extended period of L1 quiescence before final extinction. Although Hypothesis 1 is not directly testable, the remaining three hypotheses would all leave molecular signatures in the genome that could help elucidate the dynamics and mechanism of this extinction event.

Conjoined LINE/SINE activity in mammals Our data are consistent with SINE dependence upon LINEs and emphasize the unique relationship between L1 and B1 retrotransposons. If L1 and B1 share enzymatic components, they do so without sharing sequence similarity. This would make the relationship between L1 and B1 unique among the LINE and SINE pairs expected to share retrotransposon machinery. However, this is not altogether unexpected given that L1 retrotransposons with foreign DNA at the 3' end can insert at L1 endonuclease consensus sites (Moran *et al.*, 1996). The less stringent L1 retrotransposition mechanism that allows different retrotransposons to insert at L1 consensus sites may also encourage the widespread and long-term success of LINE and SINE populations. Approximately thirty percent of the L1, *Alu*, B1 and B2 insertion sites, however, do not match the consensus sequence (Jurka, 1997, 1998). This suggests that

LINEs and SINEs may also act as agents of DNA repair. Interestingly from that perspective, unusually high karyotypic variation has been observed in four South American rodent genera that lack L1 and B1 activity *Oryzomys*, *Holochilus*, *Akodon* and *Rhipidomys* (Barros *et al.*, 1992; Fagundes *et al.*, 1998; Koop *et al.*, 1983; Nachman, 1992; Nachman and Myers, 1989; Silva and Yonenaga-Yassuda, 1999).

Based on these results, we predict that rodent tRNA-derived SINEs such as ID and B2 will also be affected by the loss of L1-encoded reverse transcriptase. Preliminary dot blot hybridization of seven of the species lacking L1 activity and three containing L1 activity using a B2 probe suggests reduced copy number in species lacking L1 activity (data not shown). The short length of B2 and ID, less than 100 basepairs, and the lack of a method to measure potential function complicates our assessment of activity under the criteria established here. Surprisingly, there are no known LINEs with sequence homology to ID or B2, suggesting that not all tRNA-derived SINEs have LINE partners as predicted by Okada *et al.*, 1997.

Overall, L1 activity is not the only factor influencing rodent SINE retrotransposition because B1, B2 and ID SINEs show remarkably different evolutionary histories. B1 populations are found at consistently high copy number in mice, rats and deer mice. B2 amplification appears high in both rats and mice (Kass *et al.*, 1997; Serdobova and Kramerov, 1998). ID elements, however, are found in drastically varying copy number due to massive amplification events in rats but not mice (Kim and Deininger, 1996). This amplification was most likely preceded by duplication of the conserved source gene, BC1, which generates ID transcripts (Kim *et al.*, 1994). Thus, L1 may be necessary but not sufficient for SINE retrotransposition and evolutionary histories must consider other factors

such as prerequisite SINE transcription in addition to the availability of retrotransposition enzymes.

Impact of retrotransposon extinction on the genome Several researchers have implied that there could be a biological function of benefit to the host and thus positive selective pressure for LINE activity. L1 function is indirectly supported by evidence of induced L1 transcription in response to cell stress and by host SRY family transcription factors (Moran *et al.*, 1996; Tchenio *et al.*, 2000). Cultured mammalian cells under stress show increased SINE expression (Liu *et al.*, 1995). Transient increases in B1 transcription have also been shown *in vivo* for mice subjected to heat and alcohol stress, suggesting physiological function (Li *et al.*, 1999). Moreover, *Alu* SINEs dispersed in the human genome contain binding sites for a protein that inhibits methylation, a key component in epigenetic control of gene expression and parental imprinting (Chesnokov and Schmid, 1995; Liu and Schmid, 1993). *Alu* expression also inhibits the activity of a double-stranded RNA-activated protein kinase *in vivo* (Chu *et al.*, 1998). Further support for the proposition that L1s are involved in X-inactivation (Lyon, 2000) recently came from analysis of large sequence datasets, showing a higher density of L1s on regions of the X chromosome that undergo X inactivation (Bailey *et al.*, 2000). Importantly, selfish replication of an element and phenotypic benefit for the host are not mutually exclusive ideas (Schmid, 1998). Given their ubiquitous nature and interesting evolutionary dynamics, it would be surprising if transposable elements do not occasionally provide the raw material for evolution. If these retrotransposons are important for organismal fitness, then loss should be detrimental to the host. The South American sigmodontine rodents used in this study would be ideal systems

to evaluate questions about retrotransposon load, levels of regulation, possible host function and genome evolution.

Acknowledgements

We would like to thank The Museum of Texas Tech University, Museum of Natural History collection at the University of New Mexico, Jack Sullivan, Scott Davis and Stacia Engle for tissue and DNA donation. Special thanks are extended to Jack Sullivan, Terry Yates, Margaret Smith and Scott Steppan for taxonomic consultation, Darin Rokyta for help with phylogenetic analysis, Melodie Armstrong for assistance in manuscript preparation and LuAnn Scott and Michael Cantrell for helpful comments. This work was supported by a grant from the National Institutes of Health (GM38737 to HAW).

Table 1. Summary of L1 and B1 Data. **N** is the number of L1 or B1 sequences in the complete dataset. **Minimum sequence divergence** is the average percent divergence based on the four most closely related L1 sequences or 5 most closely related B1 sequences from each species; **% ORFs** is the percentage of L1 sequences with intact reading frames over the region examined and is considered a surrogate estimation of potential functionality; **% RNA folding** is the percentage of B1 sequences that retain the potential for RNA folding and is considered a surrogate estimation of potential functionality; **Relative copy number** is the percentage of hybridization relative to *R. fulvescens* in a dot blot analysis.

	L1				**B1**			
	N	Minimum sequence divergence	% ORFs	Relative Copy Number	N	Minimum sequence divergence	% RNA folding	Relative Copy Number
Calomys callosus	10	14.98	0	39	10	20.44	10.00	37
Calomys tener	16	19.22	0	21	20	17.69	5.00	34
Phyllotis xanthopygus	18	13.70	0	18	42	17.91	26.19	13
Akadon boliviensis	14	12.28	0	49	26	24.63	25.00	41
Oxymycterus paramensis	21	10.07	0	32	16	20.61	18.75	28
Microryzomys minutus	16	18.97	0	17	18	31.06	27.78	21
Neacomys spinosus	16	10.58	0	26	8	29.79	0.00	20
Oligoryzomys fornesi	10	12.97	0	34	11	19.98	18.18	41
Nectomys squamipes	22	9.67	0	24	14	25.64	26.67	28

Oryzomys nitidus	27	16.68	0	30	7	31.07	14.29	33
Oryzomys albigularis	16	17.10	0	20	11	29.00	9.09	16
Oryzomys palustrus	16	10.55	0	21	16	17.19	31.25	19
Holochilus brasiliensis	7	19.62	0	30	24	19.27	12.50	26
Oecomys bicolor	17	17.07	0	16	11	28.94	27.27	14
Thomasomys baeops	20	13.67	0	54	13	19.98	15.38	36
Rhipidomys nitela	11	13.40	0	46	8	24.38	37.50	33
Sigmodon hispidus	16	0.85	68.75	92	39	17.92	28.57	18
Sigmodon mascotensis	31	0.78	54.84	78	17	25.72	11.11	17
Nyctomys sumichrasti	6	2.68	66.67	49	13	15.54	42.86	42
Reithrodontomys fulvescens	23	0.87	47.83	100	14	4.63	50.00	100
Peromyscus nudipes	30	0.60	55.17	58	7	2.94	57.14	89
Peromyscus maniculatis	23	0.33	69.57	84	11	2.84	72.73	78
Peromyscus leucopus	17	1.70	23.53	79	32	2.08	71.88	71
Microtus arvalis	35	0.97	68.57	82	10	3.39	70.00	57

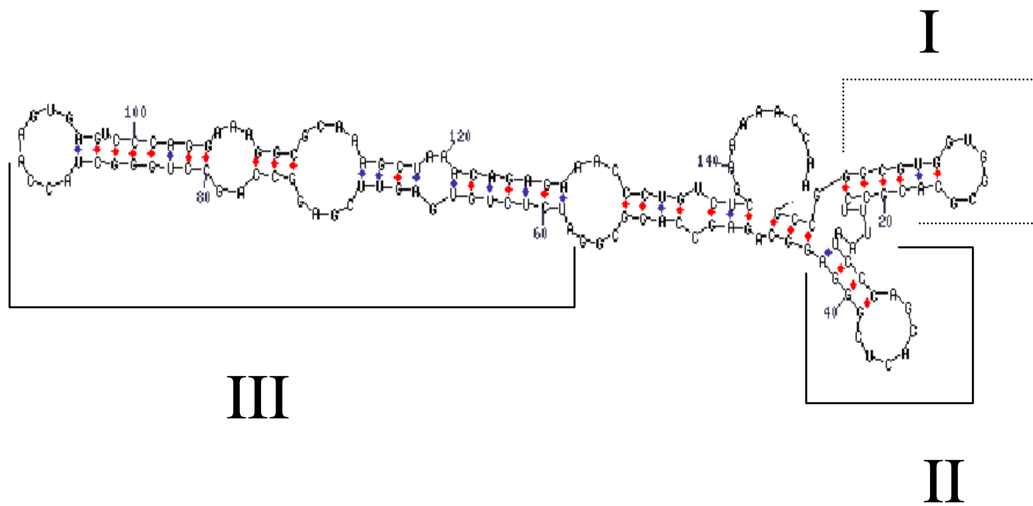

Figure 1. RNA folding structure of the B1 consensus sequence. Domains I, II and III are labeled according to the empirical work by Labuda *et al*, 1991. Domain I and II include hairpin structures while Domain III is a helix which brings the 3’ end of the RNA in close proximity to the 5’ end.

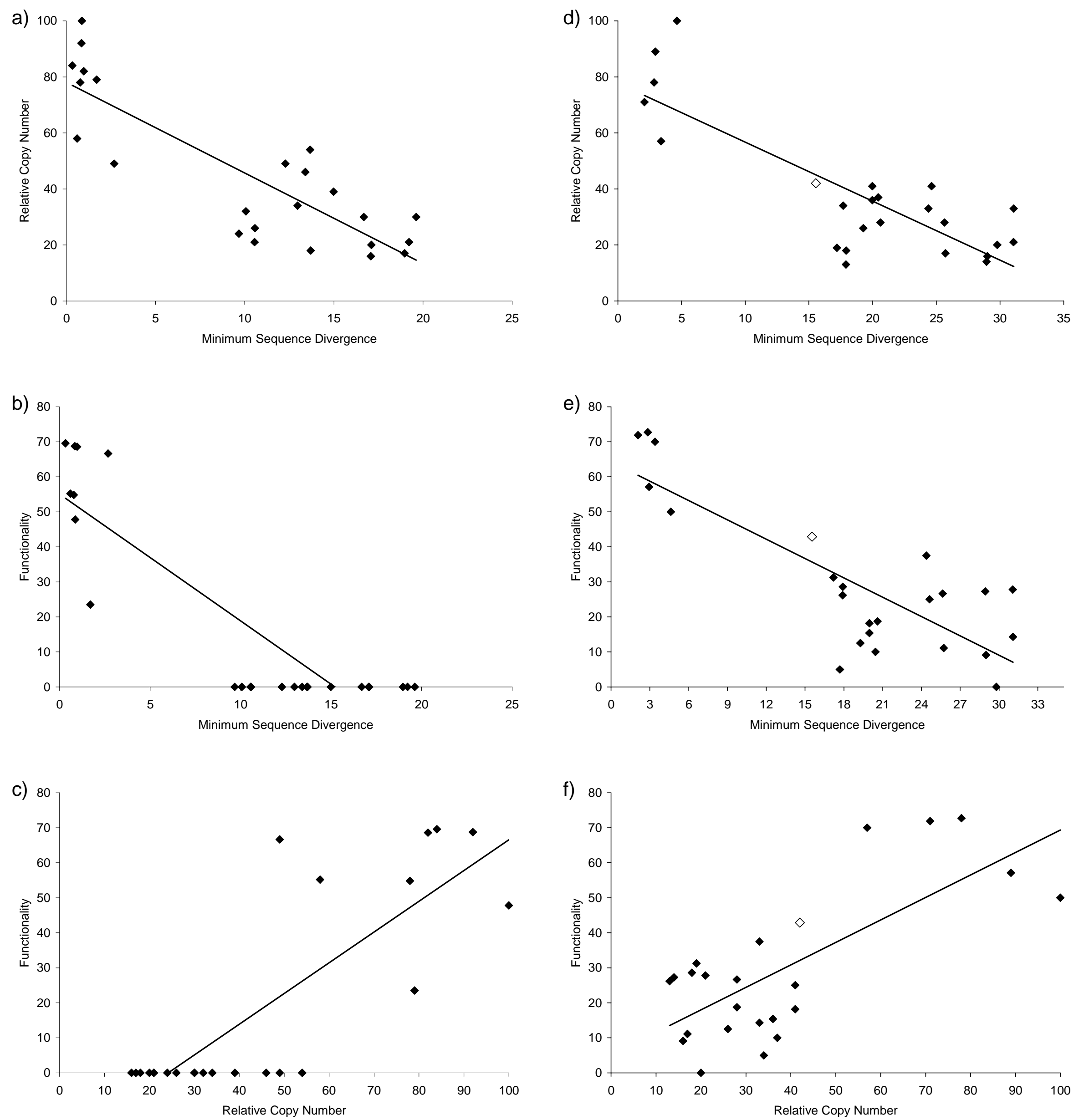

Figure 2. Relationship between minimum sequence distance, potential functionality and relative copy number as criteria for L1 and B1 activity. Regression analysis between L1 criteria are shown in panels a, b and c; regression analysis between **B**1 criteria are shown in panels d, e and f. To highlight its intermediate position, *Nyctomys sumichrasti* is shown as an open diamond in the B1 graphs.

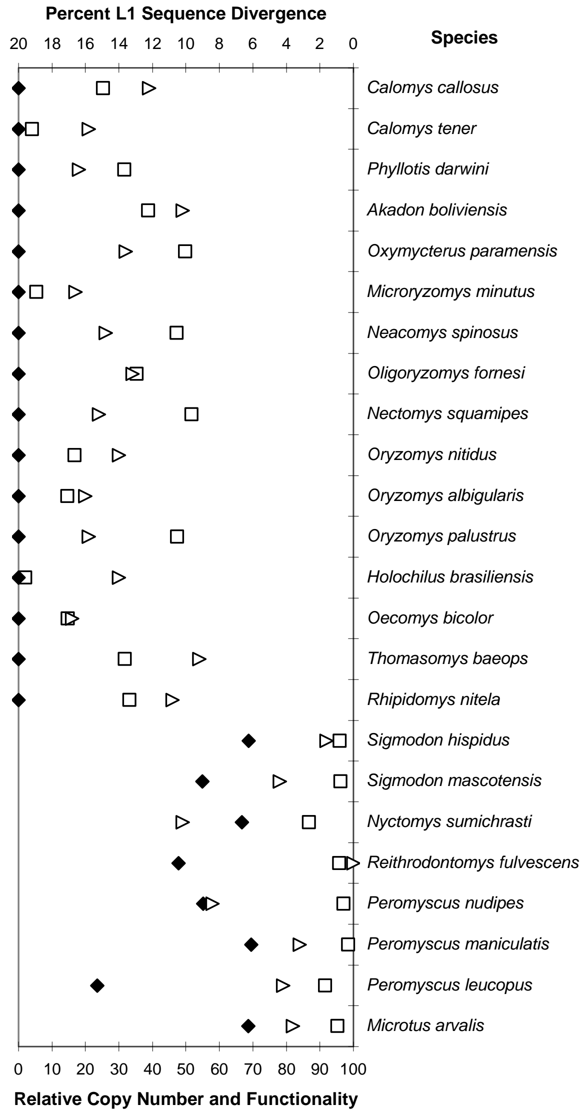

Figure 3. L1 activity profile. L1 activity criteria consist of three measures that can be plotted on the same graph. Minimum percent sequence divergence is shown as open squares (upper X axis). Relative copy number is shown as open triangles and potential functionality is shown as filled diamonds (lower X axis). The order of the species names corresponds with the phylogenetic topology in Figure 6.

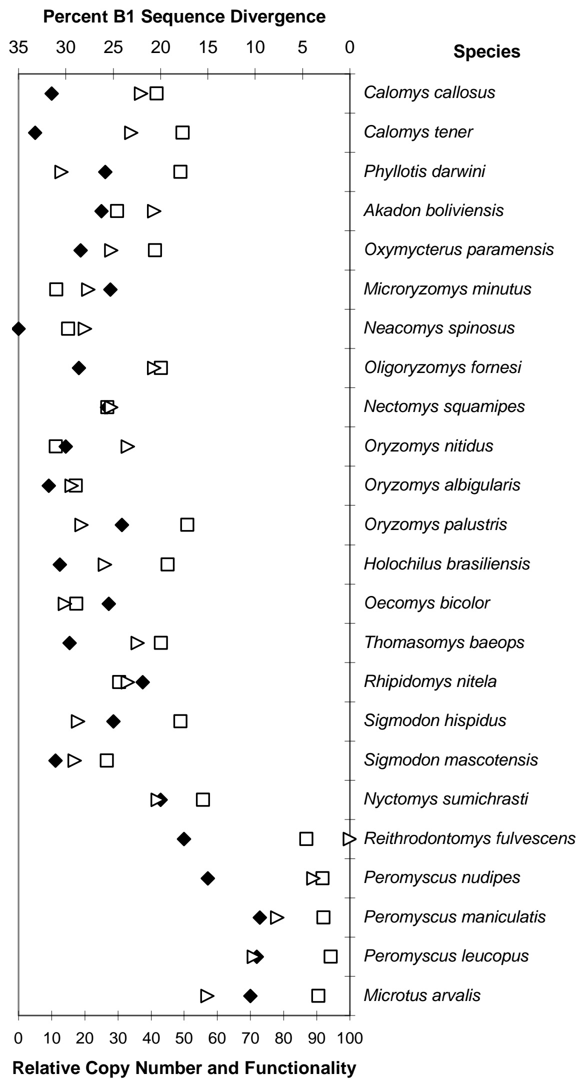

Figure 4. B1 activity profile. B1 activity criteria consist of three measures that can be plotted on the same graph. Minimum percent sequence divergence is shown as open squares (upper X axis). Relative copy number is shown as open triangles and potential functionality is shown as filled diamonds (lower X axis). The order of the species names corresponds with the phylogenetic topology in Figure 6.

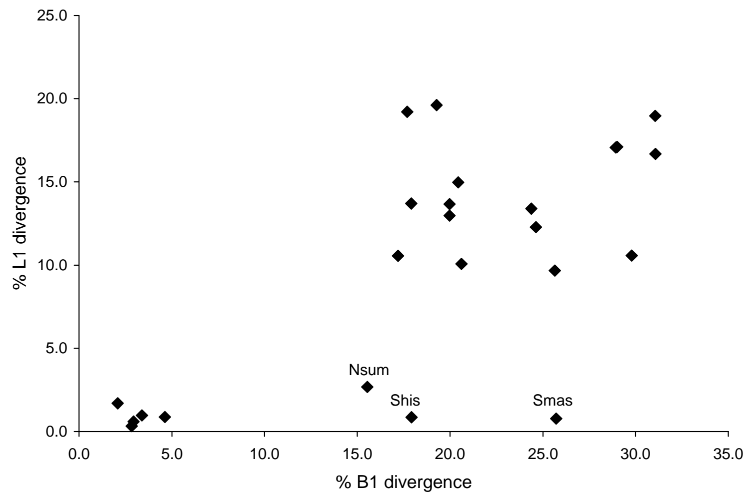

Figure 5. Correlation between divergence of L1 and B1 sequences. Species in which both L1s and B1s are active fall into the lower left corner of the graph. Species in which both are inactive fall in the upper right. In three species, L1s are active but B1s appear to be inactive – Nsum (*Nyctomys sumichrasti*), Shis (*Sigmodon hispidus*) and Smas (*S. mascotensis*).

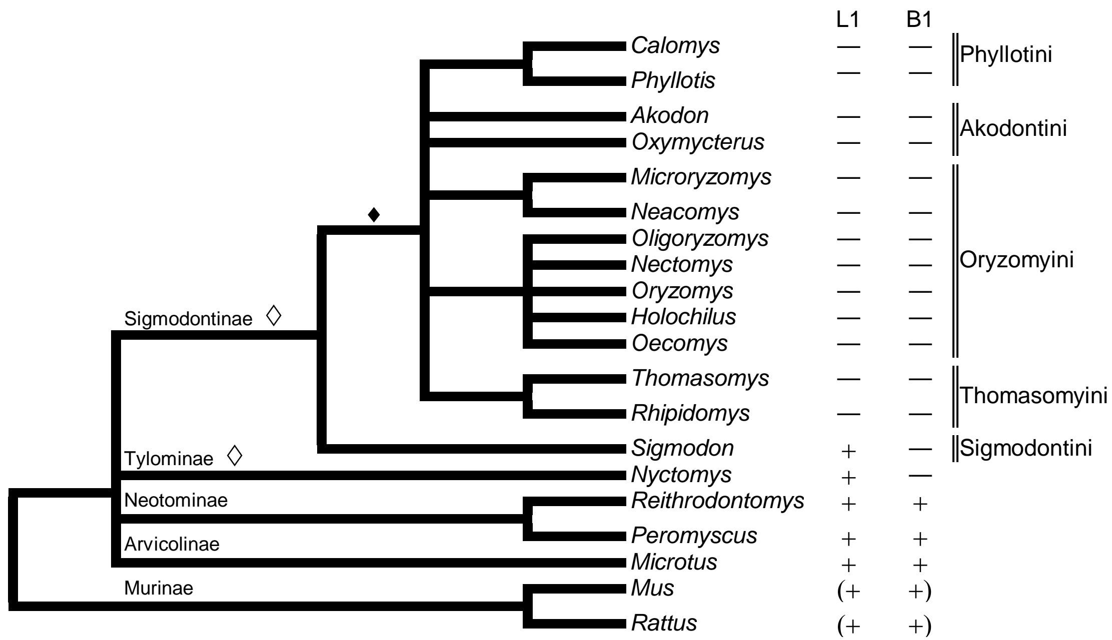

Figure 6. Reconstruction of L1 and B1 activity in sigmodontine rodents. Twenty genera of rodents, including 14 representatives of the Sigmodontinae are shown in a phylogenetic framework. B1 and L1 populations are categorized as active or non-active based upon the three *a priori* criteria. The presumed time of the single L1 extinction (♦) and the two possible B1 extinction events (◊) are marked on the tree. Activity of L1s and B1s in *Mus* and *Rattus* were not demonstrated in this study and thus are marked with positive symbols inside parentheses (+).

References

Bailey JA, Carrel L, Chakravarti A, Eichler EE: Molecular evidence for a relationship between LINE-1 elements and X chromosome inactivation: The Lyon repeat hypothesis. Proc Natl Acad Sci USA 97:6634-6639. (2000).

Barros MA, Reig OA, Perez-Zapata A: Cytogenetics and karyosystematics of South American oryzomyine rodents (Cricetidae: Sigmodontinae). IV. Karyotypes of Venezuelan, Trinidadian, and Argentinian water rats of the genus *Nectomys*. Cytogenet Cell Genet 59:34-38 (1992).

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

Cantrell MA, Grahn RA, Scott L, Wichman HA: Isolation of markers from recently transposed LINE-1 retrotransposons. Biotechniques 29:1310-1316. (2000).

Casavant NC, Scott L, Cantrell MA, Wiggins LE, Baker RJ, Wichman HA: The end of the LINE? Lack Of recent L1 activity in a group of South American rodents. Genetics 154:1809-1817 (2000).

Chesnokov IN, Schmid CW: Specific Alu binding protein from human sperm chromatin prevents DNA methylation. J Biol Chem 270:18539-18542 (1995).

Chu WM, Ballard R, Carpick BW, Williams BR, Schmid CW: Potential Alu function: regulation of the activity of double-stranded RNA-activated kinase PKR. Mol Cell Biol 18:58-68 (1998).

Deininger PL, Tiedge H, Kim J, Brosius J: Evolution, expression, and possible function of a master gene for amplification of an interspersed repeated DNA family in rodents. Prog Nucleic Acid Res Mol Biol 52:67-88 (1996).

Dickerman AW, Yates TL: Systematics of *Oligoryzomys*: Protein-electrophoretic analyses. Journal of Mammalogy 76:172-188 (1995).

Engel SR, Hogan KM, Taylor JF, Davis SK: Molecular systematics and paleobiogeography of the South American sigmodontine rodents. Molecular Biology and Evolution 15:35-49 (1998).

Esnault C, Maestre J, Heidmann T: Human LINE retrotransposons generate processed pseudogenes. Nat Genet 24:363-367 (2000).

Fagundes V, Christoff AU, Yonenaga-Yassuda Y: Extraordinary chromosomal polymorphism with 28 different karyotypes in the neotropical species *Akodon cursor* (Muridae, Sigmodontinae), one of the smallest diploid number in rodents (2n = 16, 15 and 14). Hereditas 129:263-274 (1998).

Feng Q, Moran JV, Kazazian HH, Jr., Boeke JD: Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. Cell 87:905-916 (1996).

Furano AV: The biological properties and evolutionary dynamics of mammalian LINE-1 retrotransposons. Prog Nucleic Acid Res Mol Biol 64:255-294 (2000).

Furano AV, Hayward BE, Chevret P, Catzeflis F, Usdin K: Amplification of the ancient murine Lx family of long interspersed repeated DNA occurred during the murine radiation. J Mol Evol 38:18-27 (1994).

Grahn RA, Rinehart TA, Cantrell MA, Wichman HA: Extinction of LINE-1 activity coincident with a major mammalian radiation in rodents. Cytogenetics and Genome Research (submitted).

Hayward BE, Zavanelli M, Furano AV: Recombination creates novel L1 (LINE-1) elements in *Rattus norvegicus*. Genetics 146:641-654 (1997).

Jurka J: Sequence patterns indicate an enzymatic involvement in integration of mammalian retroposons. Proc Natl Acad Sci USA 94:1872-1877 (1997).

Jurka J: Repeats in genomic DNA: mining and meaning. Curr Opin Struct Biol 8:333-337 (1998).

Jurka J, Klonowski P: Integration of retroposable elements in mammals: selection of target sites. J Mol Evol 43:685-689 (1996).

Kass DH, Kim J, Rao A, Deininger PL: Evolution of B2 repeats: the muroid explosion. Genetica 99:1-13 (1997).

Kass DH, Raynor ME, Williams TM: Evolutionary history of B1 retroposons in the genus *Mus*. J Mol Evol 51:256-264 (2000).

Kim J, Deininger PL: Recent amplification of rat ID sequences. J Mol Biol 261:322-327 (1996).

Kim J, Martignetti JA, Shen MR, Brosius J, Deininger P: Rodent BC1 RNA gene as a master gene for ID element amplification. Proc Natl Acad Sci USA 91:3607-3611 (1994).

Koop BF, Baker RJ, Genoways HH: Numerous chromosomal polymorphisms in a natural population of rice rats (Oryzomys, Cricetidae). Cytogenet Cell Genet 35:131-135 (1983).

Labuda D, Sinnett D, Richer C, Deragon JM, Striker G: Evolution of mouse B1 repeats: 7SL RNA folding pattern conserved. J Mol Evol 32:405-414 (1991).

Labuda D, Zietkiewicz E: Evolution of secondary structure in the family of 7SL-like RNAs. J Mol Evol 39:506-518 (1994).

Li T, Spearow J, Rubin CM, Schmid CW: Physiological stresses increase mouse short interspersed element (SINE) RNA expression in vivo. Gene 239:367-372 (1999).

Liu WM, Chu WM, Choudary PV, Schmid CW: Cell stress and translational inhibitors transiently increase the abundance of mammalian SINE transcripts. Nucleic Acids Res 23:1758-1765 (1995).

Liu WM, Schmid CW: Proposed roles for DNA methylation in Alu transcriptional repression and mutational inactivation. Nucleic Acids Res 21:1351-1359 (1993).

Longmire JL, Lewis AK, Brown NC, Buckingham JM, Clark LM, Jones MD, Meincke LJ, Meyne J, Ratliff RL, Ray FA, Wagner RP, Moyzis RK: Isolation and molecular characterization of a highly polymorphic centromeric tandem repeat in the family Falconidae. Genomics 2:14-24 (1988).

Lyon MF: LINE-1 elements and X chromosome inactivation: a function for "junk" DNA? Proc Natl Acad Sci USA 97:6248-6249. (2000).

Minin V, Abdo Z, Joyce P, Sullivan J: Performance-based selection of likelihood models for phylogeny estimation. Syst Biol 52:674-683 (2003).

Moore JK, Haber JE: Capture of retrotransposon DNA at the sites of chromosomal double-strand breaks. Nature 383:644-646 (1996).

Moran JV, Holmes SE, Naas TP, DeBerardinis RJ, Boeke JD, Kazazian HH, Jr.: High frequency retrotransposition in cultured mammalian cells. Cell 87:917-927 (1996).

Nachman MW: Geographic patterns of chromosomal variation in South American marsh rats, *Holochilus brasiliensis* and *H. vulpinus*. Cytogenet Cell Genet 61:10-16 (1992).

Nachman MW, Myers P: Exceptional chromosomal mutations in a rodent population are not strongly underdominant. Proc Natl Acad Sci USA 86:6666-6670 (1989).

Ogiwara I, Miya M, Ohshima K, Okada N: Retropositional parasitism of SINEs on LINEs: identification of SINEs and LINEs in elasmobranchs. Mol Biol Evol 16:1238-1250 (1999).

Ohshima K, Hamada M, Terai Y, Okada N: The 3' ends of tRNA-derived short interspersed repetitive elements are derived from the 3' ends of long interspersed repetitive elements. Mol Cell Biol 16:3756-3764 (1996).

Okada N, Hamada M, Ogiwara I, Ohshima K: SINEs and LINEs share common 3' sequences: a review. Gene 205:229-243 (1997).

Quentin Y: Successive waves of fixation of B1 variants in rodent lineage history. J Mol Evol 28:299-305 (1989).

Saxton JA, Martin SL: Recombination between subtypes creates a mosaic lineage of LINE-1 that is expressed and actively retrotransposing in the mouse genome. J Mol Biol 280:611-622 (1998).

Schmid CW: Does SINE evolution preclude Alu function? Nucleic Acids Res 26:4541-4550 (1998).

Serdobova IM, Kramerov DA: Short retroposons of the B2 superfamily: evolution and application for the study of rodent phylogeny. J Mol Evol 46:202-214 (1998).

Silva MJ, Yonenaga-Yassuda Y: Autosomal and sex chromosomal polymorphisms with multiple rearrangements and a new karyotype in the genus *Rhipidomys* (Sigmodontinae, Rodentia). Hereditas 131:211-220 (1999).

Smith MF, Patton JL: Phylogenetic relationships and the radiation of Sigmodontine rodents in South America: evidence from cytochrome *b*. Journal of Mammalian Evolution 6:89-128 (1999).

Steppan SJ: Revision of the tribe *Phyllotini* (Rodentia: Sigmodontinae), with a phylogenetic hypothesis for the Sigmodontinae. Fieldiana: Zoology 80:1-112 (1995).

Steppan SJ: Sigmodontinae: Neotropical mice and rats. The Tree of Life Project (1996).http://tolweb.org/tree?group=Sigmodontinae&contgroup=Muridae.

Sullivan J, Holsinger KE, Simon C: The effect of topology on estimates of among-site rate variation. J Mol Evol 42:308-312 (1996).

Tchenio T, Casella JF, Heidmann T: Members of the SRY family regulate the human LINE retrotransposons. Nucleic Acids Res 28:411-415 (2000).

Teng SC, Kim B, Gabriel A: Retrotransposon reverse-transcriptase-mediated repair of chromosomal breaks. Nature 383:641-644 (1996).

Zietkiewicz E, Labuda D: Mosaic evolution of rodent B1 elements. J Mol Evol 42:66-72 (1996).