

Extinction of LINE-1 activity coincident with a major mammalian radiation in rodents

^{1,2}Robert A. Grahn, ^{1,3}Timothy A. Rinehart, ¹Michael A. Cantrell and ¹Holly A. Wichman

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

²Current address: College of Veterinary Medicine
Department of Population Health and Reproduction
University of California
Davis, California 95616

³Current Address: USDA Small Fruit Research Station
306 South High Street
P.O. Box 287
Poplarville, Mississippi 39470

Running Head:

LINE-1 extinction

Keywords:

LINE-1

Sigmodontinae

transposable element

evolution

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

LINE-1 transposable elements (L1s) are ubiquitous in mammals and are thought to have remained active since before the mammalian radiation. Only one L1 extinction event, in South American rodents in the genus *Oryzomys*, has been convincingly demonstrated. Here we examine the phylogenetic limits and evolutionary tempo of that extinction event by characterizing L1s in related rodents. Fourteen genera from five tribes within the Sigmodontinae family were examined. Only the Sigmodontini, the most basal tribe in this group, demonstrate recent L1 activity. The Oryzomyini, Akodontini, Phyllotini, and Thomasomyini contain only L1s that appear to have inserted long ago: their L1s lack open reading frames, have mutations at conserved amino acid residues, and show numerous private mutations. They also lack restriction site-defined L1 subfamilies specific to any species, genus or tribe examined, and fail to form monophyletic species, genus or tribal L1 clusters. We determine here that this L1 extinction event occurred roughly 8.8 million years ago, near the divergence of *Sigmodon* from the remaining Sigmodontinae species. These species appear to be ideal model organisms for studying the impact of L1 inactivity on mammalian genomes.

Introduction

The long interspersed element LINE-1 (L1) makes up a large component of the genome in all mammalian species examined to date, with tens of thousands of copies present per genome (Furano, 2000). Sequence analysis of mammalian genomes shows that they make up 17 % or more of the human and mouse genomes (Lander *et al.*, 2001; Waterston *et al.*, 2002).

Hybridization of an L1 probe to genomic Southern blots from representatives of several orders of mammals revealed a widespread distribution of these elements and lead to the suggestion that L1s predate the mammalian radiation (Burton *et al.*, 1986). L1s have been characterized at the sequence level in numerous rodents (Martin *et al.*, 1985; Hardies *et al.*, 2000; Casavant and Hardies, 1994b; Martin, 1995; Naas *et al.*, 1998; D'Ambrosio *et al.*, 1986; Casavant *et al.*, 1996; Casavant *et al.*, 1998; Kass *et al.*, 1992; Vanlerberghe *et al.*, 1993; Kholodilov *et al.*, 1993; Mayorov *et al.*, 1999), in several primates, especially humans (Fanning and Singer, 1987b; Kazazian *et al.*, 1988; Skowronski *et al.*, 1988; Salem *et al.*, 2003), in two carnivores (Choi *et al.*, 1999; Fanning and Singer, 1987a), rabbits (Fanning and Singer, 1987a), cattle (Plucienniczak and Plucienniczak, 1999) and a marsupial (Dorner and Paabo, 1995).

Full length L1s range from 6.0 to 7.5 kb. (Loeb *et al.*, 1986). They have 5' and 3' untranslated regions that bound two open reading frames (ORFs), and a poly A tract of varying length following the 3' untranslated region. The first open reading frame (ORF 1) encodes a nucleic acid binding protein (Kolosha and Martin, 1997; Hohjoh and Singer, 1996). The second open reading frame (ORF 2) not only encodes reverse transcriptase (RT) but also has a region near the 5' end of the gene that functions as an endonuclease (Feng *et al.*, 1996). L1 elements proliferate via autonomous duplicative retrotransposition. Transcription of a full length, replication competent element is followed by translation of the two ORFs, which appear to

normally function in cis (Kimberland *et al.*, 1999), reverse transcription, and insertion of a cDNA copy into a new location in the genome. Most newly inserted elements are ‘dead on arrival’ due to 5’ truncation or deleterious mutations, but new active templates are occasionally generated by retrotransposition. Over time, active elements (masters) are rendered inactive by the accumulation of mutations and are replaced by their active progeny. One would assume that duplicative retrotransposition would lead to an accumulation and divergence of active templates, but in those species for which detailed analysis has been carried out, L1 elements appear to belong to one or a few closely related long-term lineages (Cabot *et al.*, 1997; Casavant and Hardies, 1994a; Casavant *et al.*, 1996; Naas *et al.*, 1998; Martin *et al.*, 1985; Mayorov *et al.*, 1999; Hardies *et al.*, 2000; Furano, 2000). The presence of one or a few lineages as opposed to many divergent lineages could be explained if, at any point in evolutionary time, active elements are present in low copy number per genome. This does not appear to be the case. It has been suggested that there are as many as 100 potentially active elements in the average diploid genome of humans (Sassaman *et al.*, 1997; Brouha *et al.*, 2003), and up to 3,000 in mice (DeBerardinis *et al.*, 1998). A small number of active lineages might also be explained by an arms race between the host, which attempts to suppress transposition, and the elements, which attempt to escape host suppression, but aspects of L1 biology that regulate the number and divergence of active copies remain poorly understood.

The widespread occurrence of L1s in mammals has been interpreted to suggest that L1s are maintained by selection because they serve some function for the host. It has been hypothesized that L1s may function in DNA break repair (Hutchison III *et al.*, 1989; Teng *et al.*, 1996), and more recently, that they may be involved in X chromosome inactivation (Lyon, 2000; Bailey *et al.*, 2000). While a precise function of L1s in the genome remains unclear, their impact

on the genome is significant. In humans, L1s constitute about 13% of the genomic mass of autosomes and 26% of X chromosomes (Bailey *et al.*, 2000). In *Rattus*, it is estimated that the genome has increased by 20% over the last 10 million years due to an increase in L1 insertion (Pascale *et al.*, 1990). The replication and random placement of so many L1 copies throughout the genome provides numerous potential sites for both equal and unequal recombination events (Schwartz *et al.*, 1998; Furano, 2000). L1s may also provide the reverse transcriptase necessary for short interspersed nuclear element (SINE) activity and the generation of other pseudogenes (Dewannieux *et al.*, 2003; Esnault *et al.*, 2000; Jurka, 1997; Rinehart *et al.*, submitted). De novo L1 insertions are also a direct source of genetic mutation. Cases of hemophilia were determined to be the result of L1 insertion events in the human factor VIII gene (Kazazian *et al.*, 1988; Van de Water *et al.*, 1998), and a novel L1 insertion has been implicated in one case of breast cancer (Morse *et al.*, 1988). Given the impact of L1s on the genome, the identification of species without active L1s could be extremely useful for elucidating L1 function and effect.

Previous L1 analyses have identified three groups of mammals that appeared to have quiescent or inactive L1s – the deer mouse (*Peromyscus*), (Kass *et al.*, 1992), voles (*Microtus* and *Arvicola*) (Vanlerberghe *et al.*, 1993), and rice rats and marsh rats (*Oryzomys* and *Holochilus*) (Casavant *et al.*, 2000). However, subsequent research revealed relatively young L1s in deer mice (Casavant *et al.*, 1996; Casavant *et al.*, 1998) and voles (Kholodilov *et al.*, 1993; Mayorov *et al.*, 1999; Modi, 1996). Only *Oryzomys* and *Holochilus* remain as viable candidates for sustained L1 quiescence.

We initiated a large-scale phylogenetic analysis to determine whether the *Oryzomys* / *Holochilus* case represents quiescence or an actual extinction of L1s. The goal of this study was to determine both the phylogenetic distribution and tempo of this event. A study such as this

requires consideration of the host phylogeny. Both *Oryzomys* and *Holochilus* are rodents in the family Muridae, subfamily Sigmodontinae, which includes over 300 species encompassing one of the greatest recent radiations in mammals (Smith and Patton, 1999; Engel *et al.*, 1998). Genera within the Sigmodontinae can be consistently grouped into tribes, but many of the relationships among the tribes within this subfamily remain unclear (Smith and Patton, 1999). This is probably a result of the divergence of these tribes over a short evolutionary time period during the South American radiation. Using the published work of others (Smith and Patton, 1999; Stepan, 1995; Engel *et al.*, 1998; Dickerman and Yates, 1995; Minin *et al.*, 2003; Stepan, 1996), we constructed a composite tree that approximates the relationships of the genera used in our analysis (Figure 1). We characterized L1 sequences from eighteen South American Sigmodontine species representing five of seven tribes; we were unable to obtain samples for the other two tribes. Because the original observation was in the Oryzomyini, this tribe was heavily sampled with nine species representing seven genera (*Oryzomys*, *Oligoryzomys*, *Holochilus*, *Nectomys*, *Oecomys*, *Neacomys*, and *Microryzomys*) and including a more extensive analysis of the two previously characterized species (Casavant *et al.*, 2000). The remaining four tribes are each represented by at least two species. Two species were sampled from *Sigmodon*, the sole genus in the Sigmodontini tribe, which is considered basal to the remaining tribes. An additional seven species of related rodents outside the South American Sigmodontines were also evaluated.

Here we demonstrate that L1s are inactive not only in the Oryzomyini tribe, but also in all other tribes evaluated in the subfamily Sigmodontinae with the exception of the Sigmodontini. This was accomplished by phylogenetic analysis of L1 sequences, along with evaluation of L1 reading frames, pairwise genetic distances, insertions, deletions, and conserved

amino acid changes in the youngest elements from the species under analysis. We demarcate the L1 extinction event uniting the subset of Sigmodontinae rodents by Southern blot analysis and estimate a time point for the actual extinction event by sequence comparison.

Materials and methods

Specimens and DNA isolation Specimens for this study were obtained both as tissues and as purified DNA. Species with their sample designators are shown in Table 1. LBJ and AK sample designators originate from the Texas Cooperative Wildlife Collection at Texas A&M, TK from The Museum at Texas Tech, and NK from New Mexico. Tissues from *Sigmodon hispidus* were provided by the Savannah River Ecology Laboratory, and from *Phodopus sungorus* by Wesleyan University. *Rattus norvegicus* DNA was obtained from the Sprague-Dawley laboratory strain. When tissues were provided, DNA was extracted using the method of Longmire (Longmire *et al.*, 1988).

Genomic Southern blot analysis Genomic Southern blot analysis was carried out for all species for which sufficient DNA was available. High molecular weight genomic DNA from 19 species was digested with *Rsa* I. 1.4 µg of digested DNA was loaded on a 1.3% agarose gel, electrophoresed and transferred to a nylon membrane. A DNA probe from a *M. arvalis* L1 element corresponding to bases 4969 to 5583 of *Mus* L1 (GenBank M13002) was random prime labeled with ³²P. Hybridization was carried out under conditions of low stringency as previously described (Casavant *et al.*, 1996).

PCR and sequencing Degenerate PCR and a screening technique designed to enrich for L1 fragments with an intact reading frame (Cantrell *et al.*, 2000) were used to isolate a portion of ORF 2 for a number of L1 elements from each species. With this technique, degenerate primers

containing restriction sites were designed to conserved regions within ORF 2 of mammalian L1s. Following restriction digestion, the amplified product is ligated in-frame into a *lacZ* reporter vector. Clones containing intact L1 open reading frames primarily generate blue colonies while L1s with stop codons primarily generate white colonies. This technique has proven successful in selection for young elements in such taxonomically diverse groups as marsupials, rodents, bats, equids, and primates. L1s were sequenced from 23 species of Sigmodontinae, representing a broad taxonomic survey of that subfamily, as well as representatives of the Arvicolinae, Cricetinae, Tylominae, and Neotominae (see Figure 1). Double strand sequence was obtained using an ABI 377 automated sequencer (Perkin).

Initially, elements from at least six blue colonies were sequenced for each species. If open reading frames were identified, L1s were considered to be active in that species. No further sequencing was performed on L1s from two such species: *N. sumichrasti* and *S. teguina*. In all other cases, a minimum of 10 L1s from blue colonies and 10 from white colonies were sequenced. When identical clones were found, only one was included in the final dataset. Species from which multiple L1s with intact reading frames were isolated are referred to as belonging to the L1-active group; species from which no L1s were found to have intact reading frames are referred to as members of the L1-inactive group. L1s from previously characterized species were also included in the analysis (*M. arvalis* and *P. maniculatis*) and are members of the L1-active group.

Alignments and analysis Sequences were aligned in MegAlign (DNASTAR), and alignments were modified by hand. Phylogenetic analyses were performed using PAUP* (Swofford, 1998). Gaps were coded as missing data.

Independent phylogenetic analyses were initially carried out on each species to select a subset of elements to be included in the final dataset. For these analyses, two L1 sequences from *S. hispidus*, *P. maniculatus* and *M. musculus* with intact reading frames over the region of analysis were included to aid in alignment and to serve as outgroups. For species from the L1-active group, four recently inserted elements with intact ORFs were selected. For species from the L1-inactive group, all elements were selected that appeared to have inserted into the genome after the divergence of the common ancestor leading to the Sigmodontini tribe. This was done by constructing neighbor-joining trees and identifying elements which were monophyletic with respect to *S. hispidus*. These sequences were then used for further analyses. L1 trees for each species in the L1-inactive group were also examined to assure that no divergent L1 lineage had been recently active in the ancestor of the L1-inactive group. No evidence for divergent L1 lineages was observed.

To determine how many mutations affected conserved amino acids or gave rise to stop codons, all selected elements from species in the L1-inactive group were returned to the correct reading frame by the removal of insertions and filling of gaps with Ns. To detect possible recombinants, the dataset was partitioned into three separate data blocks of about 190 bp each. Optimal maximum likelihood parameters were determined and distance trees were generated for each data block. Elements that radically changed relative position on the three output trees were further evaluated to determine if they were recombinants. Recombinants could arise either in the genome or during PCR. Elements demonstrating high levels of divergence in one block, but nearly identical sequences in another block were judged to be recombinants. Putative recombinants were eliminated from the analysis.

The tree topology for the dataset was generated using the neighbor-joining algorithm with maximum likelihood corrected distances. The GTR+ γ model was selected for correction with Modeltest version 3.06 (Posada and Crandall, 1998), and parameter values were estimated on a neighbor-joining tree generated with logdet distances. One hundred neighbor-joining bootstrap replicates were performed. All analysis was done with PAUP* 4.0b10 (Swofford, 1998). Within the L1-inactive group, clades with bootstrap values above 70 were examined for evidence of L1 transposition after the divergence of the common ancestor leading to the Sigmodontini tribe. Six such clades were identified. Four *O. albigularis* elements were identified that shared a 3 bp insertion, 2 stop codons, and 6 changes at conserved amino acid sites. Because four elements from a single animal had this mutational profile, they cannot all represent alleles from a single locus. Therefore, we hypothesize that duplication of these elements must have occurred via a mechanism other than typical L1 replication (e.g. unequal crossing over or gene conversion). We retained Oalb56w and removed Oalb59w, Oalb55w and Oalb17b from subsequent analyses. Four additional clades contained sequences that we judged to be orthologous loci on the basis of shared deletions, stop codons and changes in conserved amino acids. Sequences from nine different species were represented among the five putative orthologs. In each clade, one sequence was arbitrarily selected for retention in the final dataset and the others were excluded. Ccal12b was retained and Cten11b excluded. Hbra14b was retained and Nspi06b and Oalb18b excluded. Mmin05b was retained and Oalb16b excluded. Nspi63w was retained and Obic07b and Onit63w excluded. Mmin07b was retained and Ofor15b was excluded. Tree topology for the final dataset was generated as described above.

Construction of consensus sequences Because there is a lack of phylogenetic signal among the L1 sequences from species in the L1-inactive group, we estimated the most recent

active L1 ancestor by constructing consensus sequences for each species. In addition, a group consensus was constructed for the same sequences from the entire L1-inactive group. These consensus sequences were corrected at three CpG dinucleotides that showed very high rates of mutation.

Results

Initial classification of L1s within each species was based on the presence or absence of clones with intact reading frames over the 575 bp region characterized for this study. Species from which multiple L1s with intact reading frames were isolated are referred to as belonging to the L1-active group; species from which no L1s were found to have intact reading frames are referred to as members of the L1-inactive group (Table 1). Detailed analysis was carried out to confirm that this measure gives an accurate assessment of L1 activity or extinction and to determine the timing and tempo of the putative L1 extinction event.

Southern blot analysis supports lack of recent activity in species from the L1-inactive group. To determine the phylogenetic limits of the L1 inactivity in South American sigmodontine rodents (Casavant *et al.*, 2000), genomic Southern blot analysis was carried out on 19 rodent species (Figure 2). In this type of analysis, a robust band indicates a high copy-number L1 subfamily (i.e., L1 restriction site-defined subfamily). To minimize hybridization bias, the probe was a young L1 from *M. arvalis* specific to the region of sequence analysis. Thus, the probe was separated from each of the remaining species (except *Rattus*) by a similar evolutionary distance. One to six strongly hybridizing bands were identified in *R. fulvescens*, *P. nudipes*, *N. sumichrasti*, *S. hispidus* and *R. norvegicus*. *R. fulvescens* and *P. nudipes* are closely related and shared all hybridizing subfamilies, while *N. sumichrasti*, *S. hispidus* and *R.*

norvegicus each contained unique restriction site-defined subfamilies. The remaining 14 species evaluated were members of the L1-inactive group. Hybridization profiles for these species were very similar, with low levels of hybridization as previously seen in *Oryzomys* and *Holochilus* (Casavant *et al.*, 2000). The single identifiable restriction site-defined subfamily present in this group is also seen in *S. hispidus*, albeit in much higher copy number, suggesting that this L1 subfamily arose before the divergence of the Sigmodontinae, and that it has been more active in *Sigmodon* than in the other sigmodontine rodents since their divergence. Slight differences in hybridization intensity among these 14 species may be due to minor differences in DNA loading.

Species-specific clustering of L1s in the L1-inactive group was not observed. Degenerate PCR and an assay to enrich for intact ORFs were used to clone a portion of L1 from each species of interest. Multiple clones were sequenced from each species and these elements were aligned. The initial dataset included 315 sequences from the L1-inactive group. Non-L1 sequences and identical sequences were removed, and all elements that appeared monophyletic with respect to *Sigmodon* were identified. Likely recombinants and putative orthologous alleles were removed as described in Materials and Methods, yielding a final dataset in the L1-inactive group of 136 sequences. The number of sequences included for each species is shown in Table 1. For each species from the L1-active group, four sequences with an intact ORF over the region examined were included. Two *Mus* sequences from GenBank were also added (*M. musculus* (AF081109), *M. domesticus* (M13002) to produce a final dataset of 174 sequences.

Aligned sequences were analyzed in a phylogenetic framework. A neighbor-joining tree generated under maximum likelihood parameters is depicted in Figure 3. Ten thousand fastboot bootstrap replicates were performed and values greater than 70 % are indicated on internal nodes. Each genus outside of the L1-inactive group forms a monophyletic cluster with 99 %

support or greater. L1 elements from the Sigmodontinae subfamily form a monophyletic group with 100 % bootstrap support. L1 elements from species within the L1-inactive group form a giant polytomy and terminal branch lengths are noticeably longer than in other species, indicating a large number of private mutations since these elements inserted into the genome. Three pairs of sequences within this group had bootstrap support greater than 70 %. Each of these pairs included two sequences from two different genera; internal branches are short relative to terminal branches in each case. It is unlikely that these groupings indicated recent L1 activity within the clade. L1s from species in the L1-active group have short terminal branch lengths connected by long internal nodes. Although *P. sungorus* elements fit into this latter group and are united by a long internal node, they are somewhat anomalous in that they contain slightly longer terminal branches indicative of numerous private mutations. Only three of the 32 L1 sequences from *P. sungorus* have intact reading frames.

Mutation profiles exhibit dramatic differences between the L1-inactive group and the remaining species. Table 1 shows the mutation profile of L1 sequences from each species. The uppermost 16 species are the L1-inactive group, followed by two species of *Sigmodon* and seven non-Sigmodontines from the L1-active group. Major differences delineate the L1-inactive group sequences from all the other L1s. The L1-inactive group has more inserted and deleted bases per sequence. To count stop codons due to point mutations while ignoring frame shift mutations, insertions were removed and each sequence was returned to the correct reading frame. As with the indel profiles, there is a dramatic difference between species in the L1-inactive group and the L1-active group. The fourth mutational observation presented was derived from the work of Casavant *et al.* (2000). In that analysis, 64 amino acid residues that were conserved in all mammalian L1s were identified within this region of L1 ORF 2 and then characterized. Our

laboratory has since expanded the mammalian L1 dataset to include representatives from 17 orders of placental mammals and two orders of marsupials (manuscript in preparation). In this expanded dataset, 38 amino acid residues are conserved in the youngest L1s from all 19 orders. Changes in conserved sites were calculated as a percentage of amino acid substitutions among the 38 residues present in the sequence after accounting for deletions and stop codons affecting these sites. The dramatic difference observed previously for *Oryzomys* and *Sigmodon* is upheld – species in the L1-inactive group have many more substitutions at these conserved residues than L1s with intact ORFs from other related rodents.

The most recently active L1s are very similar in all species in the L1-inactive group.

Consensus sequences were constructed for the most recently inserted elements from each species as described above. These consensus sequences approximate the sequence of the active L1s at the time they were last active. Consensus sequences were compared to each other and to the group consensus. If transposition ceased in the common ancestor of these species, consensus sequences should be closely related to each other and to the group consensus. On the other hand, persistence of L1 activity beyond this split should leave its signature on the consensus sequences. Specifically, consensus sequences should be more similar for taxa that cluster together phylogenetically. Nine of the twelve species-specific consensus sequences are identical, and the others differ from the group consensus by only 0.5 to 2.1 %. Two species (*Ccal* and *Rnit*) have somewhat more divergent consensus sequences (2.1% from the group consensus and 4.3% from each other). However, these are the two species from which we obtained the fewest usable sequences (five each), so these may represent less reliable consensus sequences. These data support a rapid silencing of L1s in the ancestor of these twelve genera with little or no subsequent activity in any lineage.

To determine the approximate time that L1s were last active in this group, we estimated the divergence of each element from its putative common ancestor. The common ancestor was estimated by producing a consensus for the entire group, as described above. The mean number of mutations from a common ancestor was 50.6 (\pm 12.6 standard deviation), corrected for the probability of same site mutations and for length variation. The mean percent divergence was 8.80 % with species estimates ranging from 7.18 to 11.18 %. If a neutral mutation rate of 1 % per million years for rodents is assumed (She *et al.*, 1990) and references therein), then this extinction event occurred approximately 8.8 million years ago (MYA).

Discussion

LINE-1 transposable elements are ubiquitous throughout the class Mammalia, yet many aspects of their biology remain an enigma. While it is easy to consider them as well adapted genomic parasites, L1 replication dynamics and phylogenetic persistence may be difficult to explain under a simple parasite model. It has been alternatively hypothesized that they serve a necessary function for their host by providing reverse transcriptase activity for the cell, aiding in DNA break repair, or providing raw material for recombination (Hutchison III *et al.*, 1989; Schwartz *et al.*, 1998), and it has recently been proposed that they may be a component of the machinery for X chromosome inactivation (Lyon, 2000; Bailey *et al.*, 2000). One key to elucidating the biology of L1s is the identification of an organism that does not possess active elements. Though all mammalian genomes examined contain L1s, it is necessary to differentiate between recently transposed sequences and ancient fossil sequences to determine if L1s are in a period of quiescence or even extinction.

Recent technical advances allow us to rapidly isolate markers from young L1 sequences. One of the most efficient methods is the degenerate PCR and ORF enrichment technique

(Cantrell *et al.*, 2000). This procedure has successfully identified young L1s in both *Peromyscus* and *Microtus*, which were previously proposed to be in a period of L1 quiescence or extinction, and from such taxonomically diverse groups as marsupials, bats, equids and primates.

Previously, we proposed an L1 extinction event in the rice rat, *Oryzomys palustris*, and an additional tribe member, *Holochilus brasiliensis* (Casavant *et al.*, 2000). Despite using the ORF enrichment technique, no evidence of recent L1 activity was found in the rice rat, *O. palustris*. Sequence data were supported by *in situ* hybridization and Southern blot analysis. Thus, these related South American rodents were the first viable candidates for L1 extinction among all mammals yet examined. We have extended that work here, delineating the range of sigmodontine rodents that appear to have been affected by that L1 extinction event, and suggest that it occurred in the common ancestor of the group prior to their massive radiation in South America.

Several lines of evidence indicate a lack of recent L1 retrotransposition in species of the L1-inactive group. First, Southern blot hybridization failed to reveal any strongly hybridizing L1 subfamilies among members of the L1-inactive group, or any evidence for species-specific or genus-specific L1 subfamilies. All species evaluated from outside the L1-inactive group contain strongly hybridizing subfamilies. Secondly, there is a marked difference in the mutation patterns between sequences from the L1-inactive group and from the remaining species, including more insertions, deletions, stop codons and changes at conserved amino acid residues. Furthermore, sampling properties of clones from our PCR reactions suggested that younger elements were not present in these species and, in fact, that amplifiable L1 template was limited. This was evidenced by the low frequency of ‘blue’ clones in our blue-white screening assay (Cantrell *et al.*, 2000), the relatively frequent isolation of clones with identical sequences, which is rare in

species with active L1s, and the isolation of putative orthologous alleles from different species. Finally, phylogenetic analysis resulted in a clustering of L1s from all species in the L1-inactive group into one giant polytomy with long terminal branches, consistent with the hypotheses that these elements transposed in some common ancestor of this group and have been sitting in the genome acquiring mutations for millions of years. There was no evidence for species or genus-specific clusters within this group, or of species or genus-specific changes in the active L1 templates from which they arose.

Are L1s extinct or simply quiescent? It is always difficult to offer convincing evidence for a recent extinction event, whether it is at the organismal or genomic level. In this case, it may be hard to differentiate between complete loss of L1 activity with no chance of recovery and a long transpositional quiescence from which the L1s might eventually emerge. Here we make the case that this event represents an extinction event rather than a simple quiescence. Extinction would be indicated if a large, monophyletic group of species all lacked evidence of recent L1 activity and if the cessation of transposition appeared to occur at about the same time in all these species. Also, L1 activity would have to be absent for long enough to allow all of the active copies to acquire enough disabling mutations so that reemergence by back mutation or recombination was no longer possible. Quiescence, on the other hand, would be indicated if L1 activity had re-emerged in one or more species or if the most recent L1 activity dated to drastically different times within the species group, especially if some of those dates were recent.

Among the sigmodontines, evidence for recent L1 activity was found only in the two species of *Sigmodon* examined. *Sigmodon* is the sole member of the tribe Sigmodontini, and molecular data place this genus as the earliest extant genus of sigmodontine rodents to arise (Engel *et al.*, 1998 and references therein). Thus L1 activity seems to have ceased at or after the

separation of *Sigmodon* from most of the remaining sigmodontine rodents. No evidence of a later reemergence of active L1s was seen among any of the remaining 13 genera of sigmodontine rodents examined as might have been expected if this event was a simple quiescence.

Furthermore, the consensus sequences of all 16 species are remarkably similar, and sequence analysis provides further evidence that L1s have remained inactive since before these 13 genera emerged from a common ancestor.

We assayed only 9 % of a full-length L1 and not a single element from the L1-inactive group was found with an intact reading frame of this short region. Most had numerous insertions, deletions and stops, and all had at least one of these fatal mutations. Previous estimates of the half-life of an active L1 range from 40,000 to 200,000 years, assuming that as many as 50 % and as few as 10 % of mutations would be strongly deleterious (See Casavant *et al.*, 2000 for a further discussion of these estimates.) Taking the more conservative estimate of 200,000 years, this would suggest that L1s have been extinct in these South American rodents for about 44 half-lives. If there were 3000 potentially active L1s in the genome at the time of extinction, as has been estimated for extant *Mus* (DeBerardinis *et al.*, 1998), fewer than one active element would survive after 13 half-lives (2.6 million years). Thus we argue that this represents a true extinction event and not simply a transpositional quiescence.

Data from *Phodopus sungorus* is anomalous and requires further discussion. The L1 data from *P. sungorus* are unique among the L1-active group, suggesting what might be considered an intermediate level of L1 activity. Terminal branches are longer than those seen in other species in the L1-active group but not as long as those seen in the L1-inactive group. Although few sequences with open reading frames were identified, *P. sungorus* L1s do form a monophyletic cluster. These data suggest that this species has had reduced L1 activity in the

recent past. *P. sungorus* may be a species whose L1s are in a period of quiescence between transpositional bursts or have recently gone extinct. The unique nature of the L1 complement of *P. sungorus* deserves further evaluation.

How accurate is our estimate of the time of extinction? We have estimated the time since L1 extinction by estimating the number of mutations acquired by the youngest detectable set of L1 elements from 16 species of sigmodontine rodents. Assuming that neutral mutations accumulate at a rate of 1% per million years in rodents (She *et al.*, 1990), we estimate the time of extinction at about 8.8 million years ago. The standard error of this estimate is small, but we have refrained from putting error bounds on the estimate because we know that sampling error is only one source of uncertainty in our methods. Furthermore, our variance is higher than expected. The mean number of substitutions per sequence is 50.6, and variance is 163.9 (3.2-fold higher than expected for a Poisson distribution). What factors might account for error in our estimate? 1) Active templates tend to be very closely related, but not identical. Thus even in a very active L1 lineage, not all recent insertions are identical. Pre-existing variation between sequences at the time of the extinction event would increase the mean and perhaps affect the variance of our estimate. 2) Extinction probably was not instantaneous. Low rates of transposition before the final cessation of activity would affect both the mean and the variance of our estimate. 3) PCR-based sampling is not random. This sampling scheme may not retrieve the most divergent samples from the genome, and will certainly not retrieve elements with very large insertions or deletions. 4) The molecular clock, upon which our estimate is implicitly based, assumes that mutation rates are uniformly distributed. We know this is not actually the case – mutation rate on the autosomes is higher than on the X and lower than on the Y chromosome. This within-genome variation could increase the variance of sampled elements. It is also

possible that there are differences in mutation rates between species. 5) There is error associated with other parameters: our calculation of the ancestral sequence, the estimate of the rodent mutation rate, etc.

Our data suggest that L1 extinction occurred after the divergence of *Sigmodon* but before the massive radiation of species in South America. An independent estimate of these events based on mitochondrial sequence puts the origin of the South American sigmodontines at 11.9 ± 2.9 MYA, the divergence of *Sigmodon* at 10.0 ± 2.4 MYA, the origin of the oryzomines at 6.8 ± 1.6 MYA, and the radiation of the other South American sigmodontines at 6.6 ± 1.6 MYA (Engel *et al.*, 1998). Thus despite the inexactness of our estimate of 8.8 MYA, it is reassuringly consistent with these estimates of speciation patterns within the sigmodontines.

How does L1 extinction relate to speciation and the biology of the L1-inactive group? It is a commonly held view that transposition may contribute to speciation either as a source of genetic variation (Furano and Usdin, 1995) or by affecting chromosome pairing during meiosis (Hutchison III *et al.*, 1989). Indeed, in *Rattus* there appears to be a correlation between the expansion of novel L1 clades and a number of speciation / extinction events (Furano *et al.*, 1994). It is interesting, then, that the single L1 extinction event documented here affects so many species. Even though this extinction event occurred relatively recently, it was coincident with one of the greatest recent mammalian radiations and thus may have been passed on to over 300 species of rodents. Had some of those species been developed as the model genetic species for rodents, we would have had an entirely different view of mammalian genome dynamics than has come from our studies of *Mus* and *Rattus*.

If L1 insertion is an important mechanism of chromosome repair, one might expect species in the L1-inactive group to be particularly susceptible to genetic or environmental agents

that increase the incidence of chromosomal breaks. Indeed, karyotypic variation in the sigmodontines is exceptional. There are several examples in the literature of extreme karyotypic variation within species in this subfamily. High levels of karyotypic polymorphism within a single population have been reported for *Oryzomys* (Koop *et al.*, 1983), *Holochilus* (Nachman and Myers, 1989; Nachman, 1992), *Nectomys* (Barros *et al.*, 1992), *Akodon* (Fagundes *et al.*, 1998) and *Rhipidomys* (Silva and Yonenaga-Yassuda, 1999), lending at least correlative support to the hypothesis that L1s may play a role in DNA-break repair.

These species, which show no evidence for recent L1 activity, appear to be ideal model organisms for studying the impact of L1 inactivity on mammalian genomes. In particular, it will be interesting to examine X chromosome inactivation and chromosomal break repair in this group. *P. sungorus* may also be useful as an example of a species in a period of L1 quiescence.

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 and Scott Davis for tissue and DNA donation. We also appreciate taxonomic consultation from Scott Stepan, Margaret Smith, Terry Yates, Robert Baker and Jack Sullivan. We thank Darin Rokyta for help with phylogenetic analysis and LuAnn Scott and Melodie Armstrong for assistance with manuscript preparation. Dave Rodriguez performed preliminary *P. sungorus* characterization. This work was supported by a grant from the National Institutes of Health (GM38737 to HAW).

Table 1. Mutational profile of sampled L1 elements. Members of the L1-inactive group are listed above, followed by members of the L1-active group. Abbreviations listed are used in other figures. Specimen numbers are listed and further described in Methods. The numbers of sequences from each species used in the final data set are indicated. Two additional sequences from GenBank were also included, bringing the total to 174. Means are shown for each species; group means are unweighted. The number of base pairs inserted and the number deleted were determined for all sequences included in the final dataset. Insertions were removed and deletions were replaced by Ns to permit counts of stop codons and changes to alternate amino acids at 38 highly conserved positions. The percent change at conserved sites was corrected to reflect the number of sites removed by deletions and stops. For sequences in the L1-inactive group, divergence from an active ancestor was calculated as divergence from the modified group consensus sequence and was corrected for multiple hits. For sequences in the L1-active group, divergence from an active ancestor was calculated from private mutations on terminal branches of a phylogenetic tree.

Species	abbrv.	Source	# of seqs	inserted bases	deleted bases	Stops	% change at conserved sites	divergence from active ancestor
L1-inactive group								
<i>Calomys callosus</i>	Ccal	NK37800	5	2.0	27.6	2.0	19.5	8.26
<i>Calomys tener</i>	Cten	NK21054	7	2.3	9.1	3.0	15.6	9.36
<i>Phyllotis xanthopygus</i>	Pxan	AK13012	7	5.3	6.1	2.3	18.5	8.57
<i>Akodon boliviensis</i>	Abol	NK11561	7	2.9	10.4	2.1	16.3	9.80
<i>Oxymycterus paramensis</i>	Opar	NK22836	11	4.1	3.1	1.9	11.8	7.18
<i>Microryzomys minutus</i>	Mmin	NK25822	8	3.3	15.3	3.1	16.7	10.35
<i>Neacomys spinosus</i>	Nspi	NK25265	10	4.1	14.5	2.1	14.1	8.89
<i>Oligoryzomys fornesi</i>	Ofor	NK22527	6	0.7	15.3	1.5	13.2	8.84
<i>Nectomys squamipes</i>	Nsqu	NK13407	15	3.6	14.0	2.2	12.3	7.54
<i>Oryzomys nitidus</i>	Onit	NK13451	13	3.5	15.1	1.9	17.3	9.44
<i>Oryzomys albigularis</i>	Oalb	LBJ1380	7	2.6	34.0	3.1	20.5	9.3
<i>Oryzomys palustris</i>	Opal	TK28621	11	2.9	15.6	2.5	16.1	8.79
<i>Holochilus brasiliensis</i>	Hbra	NK13055	6	11.7	13.7	2.8	20.1	9.8
<i>Oecomys bicolor</i>	Obic	NK12701	7	2.9	22.7	1.7	13.2	11.18
<i>Thomasomys baeops</i>	Tbae	NK27679	11	5.5	13.5	1.8	14.7	7.99
<i>Rhipidomys nitela</i>	Rnit	NK21695	5	5.0	13.4	2.0	14.8	8.52
Mean of L1-inactive group			136	3.9	14.6	2.2	15.82	8.80
L1-active group								
				1.9	16.6	1.7	18.9	
<i>Sigmodon hispidus</i>	Shis	SREL	4	0	0	0	0	0.39
<i>Sigmodon mascotensis</i>	Smas	JS2013	4	0	0	0	1.3	0.70
<i>Nyctomys sumichrasti</i>	Nsum	AK006235	4	0	0	0	1.3	1.40
<i>Scotinomys teguina</i>	Steg	LBJ1953	4	0	0	0.3	2.0	1.88
<i>Reithrodontomys fulvescens</i>	Rful	TK21614	4	0	0	0	0.0	0.44
<i>Peromyscus nudipes</i>	Pnud	NK17807	4	0	0	0	1.3	0.39
<i>Peromyscus maniculatis</i>	Pman	TK28643	4	0	0	0	0.7	0.35
<i>Microtus arvalis</i>	Marv	TK44790	4	0	0	0	0.0	0.48
<i>Phodopus sungorus</i>	Psun	Wesleyan	4	0	2.8	0.5	5.4	3.07
Mean of L1-active group			36	0	0.3	0.1	1.5	1.01

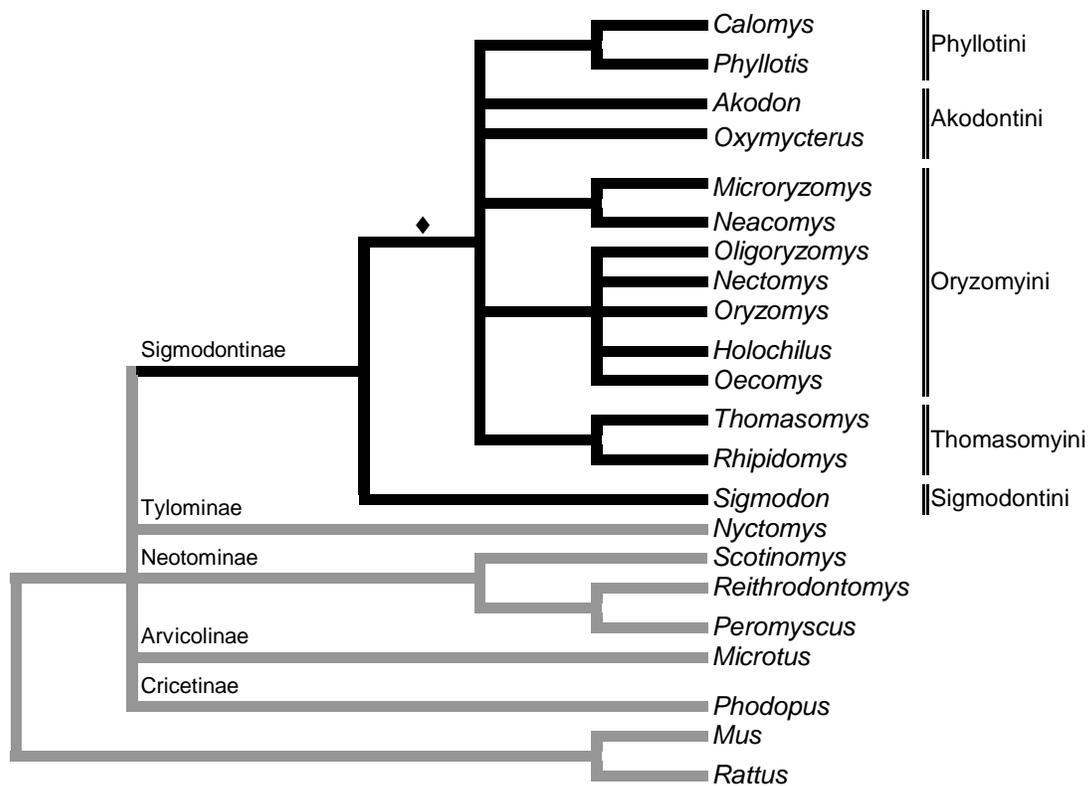


Figure 1. Phylogenetic relationships among genera. This composite tree is based on the previously reported work of other investigators (Smith and Patton, 1999; Stepan, 1995; Engel *et al.*, 1998; Dickerman and Yates, 1995), including analysis of our cytochrome b dataset (Sullivan *et al.*, 1996). The backbone of the tree for genera within the Sigmodontinae is shown in black; backbone for the outgroup genera is shown in grey. Tribe assignments within the Sigmodontinae are shown on the right. ◆ indicates loss of L1 activity.

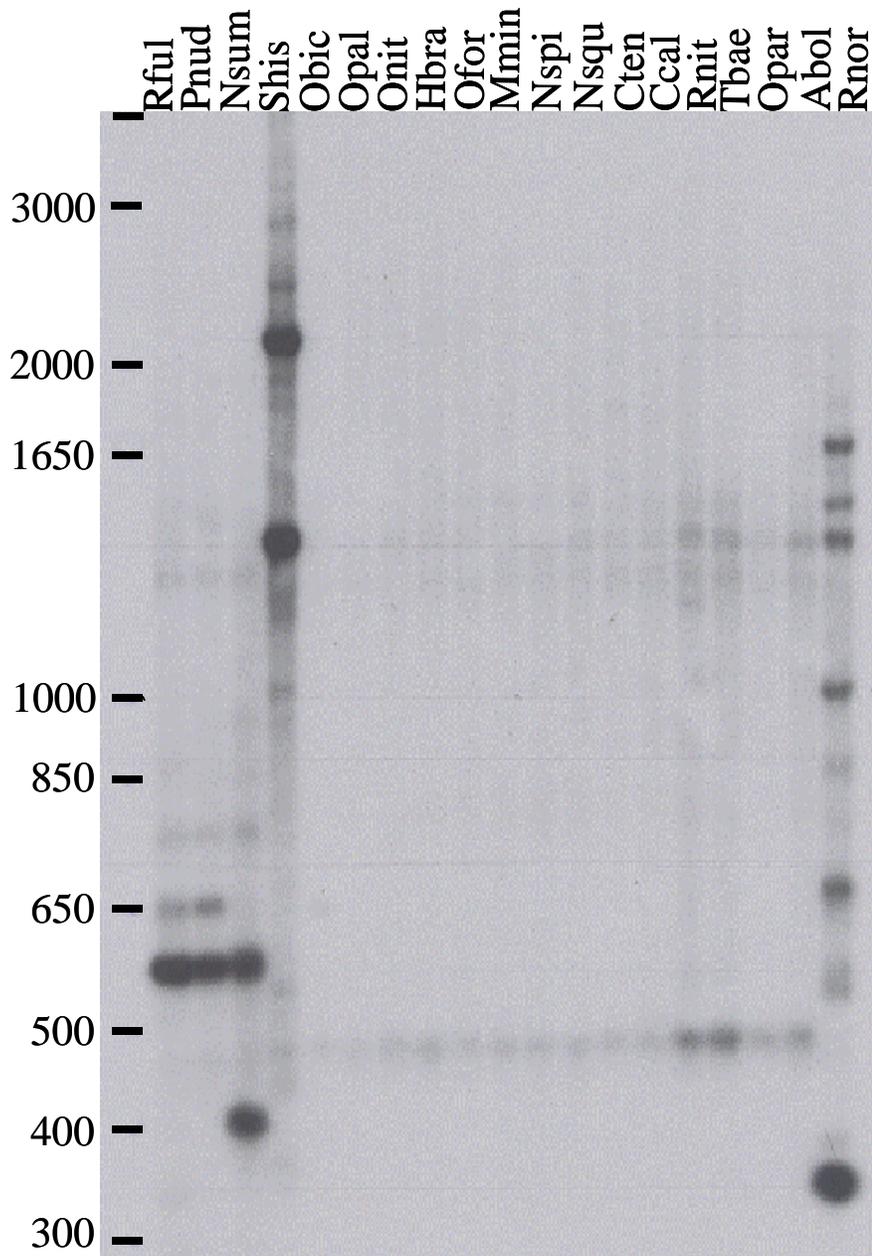


Figure 2. Southern hybridization of *Rsa* I digested genomic DNA probed with a *Microtus* LINE-1 probe. Bands represent restriction site-defined subfamilies. The first four lanes are species with recently active L1s, and the next 14 lanes are species from the L1-inactive group. Abbreviations above the lanes correspond to those listed in Table 1. *Rattus norvegicus* (Rnor) DNA was used in the final lane to provide a more distant outgroup as a control for hybridization.



Figure 3. Neighbor-joining tree of four closely related elements from species with open reading frames and all elements from the Sigmodintinae subfamily species that were monophyletic with respect to *Sigmodon*. Individual names have been removed for ease of viewing; species abbreviations are as shown in Table 1. Elements were returned to frame prior to analysis by removing insertions and replacing gaps with N (excluded from analysis). All four elements from each L1-active group formed monophyletic groups except *S. mascotensis* and *S. hispidus*, which formed a monophyletic *Sigmodon* clade. For those species with greater than one active clade, only the most active clade is represented.

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).
- Brouha B, Schustak J, Badge RM, Lutz-Prigge S, Farley AH, Moran JV, Kazazian HH, Jr.: Hot L1s account for the bulk of retrotransposition in the human population. *Proc Natl Acad Sci USA* 100:5280-5285 (2003).
- Burton FH, Loeb DD, Voliva CF, Martin SL, Edgell MH, Hutchison CA, 3rd: Conservation throughout mammalia and extensive protein-encoding capacity of the highly repeated DNA long interspersed sequence one. *J Mol Biol* 187:291-304 (1986).
- 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, Hardies SC: The dynamics of murine LINE-1 subfamily amplification. *J Mol Biol* 241:390-397 (1994a).
- Casavant NC, Hardies SC: Shared sequence variants of *Mus spretus* LINE-1 elements tracing dispersal to within the last 1 million years. *Genetics* 137:565-572 (1994b).

- Casavant NC, Lee RN, Sherman AN, Wichman HA: Molecular evolution of two lineages of L1 (LINE-1) retrotransposons in the californian mouse, *Peromyscus californicus*. *Genetics* 150:345-357 (1998).
- 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).
- Casavant NC, Sherman AN, Wichman HA: Two persistent LINE-1 lineages in *Peromyscus* have unequal rates of evolution. *Genetics* 142:1289-1298 (1996).
- Choi Y, Ishiguro N, Shinagawa M, Kim CJ, Okamoto Y, Minami S, Ogihara K: Molecular structure of canine LINE-1 elements in canine transmissible venereal tumor. *Anim Genet* 30:51-53. (1999).
- D'Ambrosio E, Waitzkin SD, Witney FR, Salemme A, Furano AV: Structure of the highly repeated, long interspersed DNA family (LINE or L1Rn) of the rat. *Mol Cell Biol* 6:411-424 (1986).
- DeBerardinis RJ, Goodier JL, Ostertag EM, Kazazian HH, Jr.: Rapid amplification of a retrotransposon subfamily is evolving the mouse genome. *Nat Genet* 20:288-290 (1998).
- Dewannieux M, Esnault C, Heidmann T: LINE-mediated retrotransposition of marked Alu sequences. *Nat Genet* 35:41-48 (2003).
- Dickerman AW, Yates TL: Systematics of Oligoryzomys: Protein-electrophoretic analyses. *Journal of Mammalogy* 76:172-188 (1995).
- Dorner M, Paabo S: Nucleotide sequence of a marsupial LINE-1 element and the evolution of placental mammals. *Mol Biol Evol* 12:944-948 (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).
- Fanning T, Singer M: The LINE-1 DNA sequences in four mammalian orders predict proteins that conserve homologies to retrovirus proteins. *Nucleic Acids Res* 15:2251-2260 (1987a).
- Fanning TG, Singer MF: LINE-1: A mammalian transposable element. *Biochim Biophys Acta* 910:203-212 (1987b).
- 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).

Furano AV, Usdin K: DNA "fossils" and phylogenetic analysis. Using L1 (LINE-1, long interspersed repeated) DNA to determine the evolutionary history of mammals. *J Biol Chem* 270:25301-25304 (1995).

Hardies SC, Wang L, Zhou L, Zhao Y, Casavant NC, Huang S: LINE-1 (L1) lineages in the mouse. *Mol Biol Evol* 17:616-628. (2000).

Hohjoh H, Singer MF: Cytoplasmic ribonucleoprotein complexes containing human LINE-1 protein and RNA. *Embo J* 15:630-639 (1996).

Hutchison III CA, Hardies SC, Loeb DD, Shehee WR, Edgell MH: LINEs and related retroposons: Long interspersed repeated sequences in the eucaryotic genome. *Mobile DNA*. Berg DE, Howe MM (eds) Washington DC, American Society for Microbiology pp 593-617 (1989).

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

Kass DH, Berger FG, Dawson WD: The evolution of coexisting highly divergent LINE-1 subfamilies within the rodent genus *Peromyscus*. *J Mol Evol* 35:472-485 (1992).

Kazazian HH, Jr., Wong C, Youssoufian H, Scott AF, Phillips DG, Antonarakis SE: Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man. *Nature* 332:164-166 (1988).

Kholodilov NG, Mayorov VI, Mullokandov MR, Cheryaukene OV, Nesterova TB, Rogozin IB, Zakian SM: LINE-1 element in the vole *Microtus subarvalis*. *Mamm Genome* 4:624-626 (1993).

Kimberland ML, Divoky V, Prchal J, Schwahn U, Berger W, Kazazian HH, Jr.: Full-length human L1 insertions retain the capacity for high frequency retrotransposition in cultured cells. *Hum Mol Genet* 8:1557-1560 (1999).

Kolosha VO, Martin SL: In vitro properties of the first ORF protein from mouse LINE-1 support its role in ribonucleoprotein particle formation during retrotransposition. *Proc Natl Acad Sci USA* 94:10155-10160. (1997).

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).

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC *et al.*: Initial sequencing and analysis of the human genome. *Nature* 409:860-921 (2001).

Loeb DD, Padgett RW, Hardies SC, Shehee WR, Comer MB, Edgell MH, Hutchison III CA: 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 (1986).

Longmire JL, Lewis AK, Brown NC, Buchingham JM, Clark LM, *et al.*: 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).

Martin SL: Characterization of a LINE-1 cDNA that originated from RNA present in ribonucleoprotein particles: Implications for the structure of an active mouse LINE-1. *Gene* 153:261-266 (1995).

Martin SL, Voliva CF, Hardies SC, Edgell MH, Hutchison CA: Tempo and mode of concerted evolution in the L1 repeat family of mice. *Mol Biol Evol* 2:127-140 (1985).

Mayorov VI, Rogozin IB, Adkison LR: Characterization of several LINE-1 elements in *Microtus kirgisorum*. Mamm Genome 10:724-729 (1999).

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

Modi WS: Phylogenetic history of LINE-1 among arvicolid rodents. Mol Biol Evol 13:633-641 (1996).

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

Naas TP, DeBerardinis RJ, Moran JV, Ostertag EM, Kingsmore SF, Seldin MF, Hayashizaki Y, Martin SL, Kazazian HH: An actively retrotransposing, novel subfamily of mouse L1 elements. EMBO J 17:590-597 (1998).

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).

Pascale E, Valle E, Furano AV: 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 (1990).

Plucienniczak G, Plucienniczak A: Fragments of LINE-1 retrotransposons flanked by inverted telomeric repeats are present in the bovine genome: Homology with human LINE-1 elements. Acta Biochim Pol 46:873-878 (1999).

Posada D, Crandall KA: MODELTEST: Testing the model of DNA substitution. Bioinformatics 14:817-818 (1998).

- Rinehart TA, Grahn RA, Wichman HA: SINE extinction preceded LINE extinction in sigmodontine rodents: Implications retrotranspositional dynamics and mechanisms. Cytogenetics and Genome Research (submitted).
- Salem AH, Myers JS, Otieno AC, Scott Watkins W, Jorde LB, Batzer MA: LINE-1 preTa elements in the human genome. *J Mol Biol* 326:1127-1146 (2003).
- Sassaman DM, Dombroski BA, Moran JV, Kimberland ML, Naas TP, DeBerardinis RJ, Gabriel A, Swergold GD, Kazazian HH, Jr.: Many human L1 elements are capable of retrotransposition. *Nat Genet* 16:37-43 (1997).
- Schwartz A, Chan DC, Brown LG, Alagappan R, Pettay D, Disteche C, McGillivray B, de la Chapelle A, Page DC: Reconstructing hominid Y evolution: X-homologous block, created by X-Y transposition, was disrupted by Yp inversion through LINE-LINE recombination. *Hum Mol Genet* 7:1-11 (1998).
- She JX, Bonhomme F, Boursot P, Thaler L, Catzeflis FM: Molecular phylogenies in the genus *Mus*: comparative analysis of electrophoretic, scnDNA hybridization and mtDNA RFLP data. *Biol J Linnean Soc* 41:83-103 (1990).
- 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).
- Skowronski J, Fanning TG, Singer MF: Unit-length line-1 transcripts in human teratocarcinoma cells. *Mol Cell Biol* 8:1385-1397 (1988).
- 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* I-VI, 1-112 (1995).
- Steppan SJ: Sigmodontinae: Neotropical mice and rats. The Tree of Life Project <http://tolweb.org/tree?group=Sigmodontinae&contgroup=Muridae>. (1996).
- Sullivan J, Holsinger KE, Simon C: The effect of topology on estimates of among-site rate variation. *J Mol Evol* 42:308-312 (1996).
- Swofford DL: PAUP*. Phylogenetic Analysis Using Parsimony (* and Other Methods). Sinauer Associates, Sunderland, Massachusetts (1998).
- Teng SC, Kim B, Gabriel A: Retrotransposon reverse-transcriptase-mediated repair of chromosomal breaks. *Nature* 383:641-644 (1996).
- Van de Water N, Williams R, Ockelford P, Browett P: A 20.7 kb deletion within the factor VIII gene associated with LINE-1 element insertion. *Thromb Haemost* 79:938-942 (1998).
- Vanlerberghe F, Bonhomme F, Hutchison CA, Edgell MH: A major difference between the divergence patterns within the LINEs-1 families in mice and voles. *Mol Biol Evol* 10:719-731 (1993).
- Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF *et al.*: Initial sequencing and comparative analysis of the mouse genome. *Nature* 420:520-562 (2002).