

## Comparison of chromosomal distribution of a retroposon (LINE) and a retrovirus-like element *mys* in *Peromyscus maniculatus* and *P. leucopus*

R. J. Baker & D. H. Kass

Received 19 January 1993; received in revised form 17 December 1993;  
Accepted for publication by J. S. (Pat) Heslop-Harrison 20 December 1993

**Chromosomal distribution for two interspersed elements (LINEs and *mys*) that are thought to have established their chromosomal position primarily by transposition was compared between two species of deer mice (*Peromyscus leucopus* and *P. maniculatus*). Both LINEs and *mys* generally produced an autosomal banding pattern reflective of G-bands and both hybridized preferentially to the sex chromosomes. The pattern on the long arm of the X was unique for each, with *mys* reflecting the G-bands (four bands with the telomeric most prominent) and LINE producing five equally spaced bands of equal intensity. LINE also preferentially hybridized to the short arm of the longest autosomal pair. Some aspects of these patterns are explained adequately with proposed mechanisms that would produce a non-random pattern of chromosomal distribution (i.e. both reflect autosomal G-bands and both preferentially insert into AT-rich regions characteristic of G-bands). However, other aspects such as the differences observed on the long arm of the X do not appear to fit any predictions of proposed mechanisms.**

**Key words:** copy number containment, LINE, *mys*, *Peromyscus*, transposable elements

### Introduction

Transposable elements (TEs) are not randomly distributed on chromosomes (Manuelidis & Ward 1984, Montgomery *et al.* 1987, Korenberg & Rykowski 1988, Baker & Wichman 1990, Boyle *et al.* 1990). Undoubtedly, several different mechanisms interact to produce such non-random patterns. Some mechanisms (such as specificity of insertion site) would be expected to produce distinct chromosomal distributions for different TEs even within the same species or individual. Other mechanisms, such as excision of copies through ectopic crossing over (Langley *et al.* 1989, Baker & Wichman 1990), might be expected to produce similar chro-

mosomal distributions for different TEs in the same species or individuals. Therefore, the extent to which chromosomal distribution of different TEs is the same in a given species is a critical observation in assessing the role of various mechanisms (Wichman *et al.* 1992) in explaining the non-random distribution of TEs on chromosomes.

The non-random distribution of the retrotransposon *mys* was described for the chromosomes of *Peromyscus maniculatus* and *P. leucopus* (deer mice) (Baker & Wichman 1990). *mys* preferentially accumulated on the X and Y elements, and in G-bands, and appeared excluded from the most common type of tandemly repeated heterochromatin. This study examined the chromosomal distribution of long interspersed elements (LINEs) in the deer mice, *Peromyscus maniculatus* and *P. leucopus*, and compared these results to the patterns observed with *mys*.

*mys* is a 2.8 kb retrovirus-like element with long-terminal repeats (LTRs) of 343 bp, a 6 bp target site duplication, a polypurine tract and a lysine tRNA binding site at the LTR junctions. It has two open reading frames (ORF), one of which has amino acid similarity to reverse transcriptase. In *P. leucopus*, *mys* elements with this conserved structure and size have been estimated to occur at a rate of 500–1000 copies per haploid genome (Wichman *et al.* 1985). Based on the frequency of *mys* in clones of a cosmid library, Janacek *et al.* (1993) estimated that there were 4700 *mys*-1-related elements in the haploid genome of *P. leucopus*. The methods used by Janacek *et al.* (1993) could not estimate the relative completeness of these *mys* elements.

LINE elements are approximately 7 kb in rodents (Fanning 1983, Voliva *et al.* 1983) and are flanked by short, direct repeats rather than LTRs (Hutchison *et al.* 1989). Although LINE (L1) has not been demonstrated to have a specific target site it has been shown to preferentially insert in A-T rich regions (Korenberg & Rykowski 1988). The consensus L1 structure has a

R. J. Baker (corresponding author) at the Department of Biological Sciences and the Museum, Texas Tech University, Lubbock, TX 79409. Tel: (+1) 806 742 2702; Fax: (+1) 806 742 2963; E-mail: bjrjb@ttacs1.ttu.edu. D. H. Kass is at the University of South Carolina, Columbia, SC 29208, USA.

poly-A tail at the 3' end and L1 contains two large open reading frames, one of the ORFs containing amino acid similarity to reverse transcriptase (Hattori *et al.* 1986, Loeb *et al.* 1986). Kass *et al.* (1992) estimated copy number of LINEs at 16 000 copies per haploid genome in *P. maniculatus*. Janacek *et al.* (1993) estimated 41 000 copies per haploid genome in *P. leucopus*. Both estimates include severely truncated copies.

## Materials and methods

*In situ* hybridization using biotinylation primarily following published procedures (Moyzis *et al.* 1987, 1988, Baker & Wichman 1990) was employed to examine the genomic distribution of LINE-1 in *Peromyscus maniculatus* and *P. leucopus*. Metaphase chromosomes were denatured for 2 min at 70°C with 70% formamide in 2 × SSC (1 × SSC is 0.15 M NaCl, 0.015 M Na citrate) and hybridized for 12–18 h with approximately 1–3 µg/ml biotinylated plasmid DNA (labelled by nick translation as described by Clontech, Palo Alto, (A) in 2 × SSC, 500 µg/ml *Escherichia coli* carrier DNA, and 30% formamide. Hybridization was maintained at 37°C in moist chambers under sealed coverslips. Following hybridization, the slides were washed for 2 min each in five changes of 2 × SSC, pH 7.0 at 40–42°C. Hybridization was detected by binding fluorescein-labelled avidin (Vector Laboratories, Burlingame, CA) to biotin-labelled probes. The preparations were amplified once with biotinylated goat anti-avidin antibody (Vector Laboratories): in some instances two amplifications were required for precise visualization. Slides were counterstained with propidium iodide (Sigma, St Louis, MO) and 4,6-diamidino-2-phenylindole (DAPI; Sigma). Hybridization under these conditions (conditions used to produce the hybridization seen in Figure 1) should detect sequences of >70% similarity (Meinkoth & Wahl 1984, Moyzis *et al.* 1987). In some instances, we altered the concentration of formamide in the hybridization mixture in order to increase or decrease the stringency of hybridization. This gave us three relative hybridization stringencies which we categorized as low, medium and high. Alteration of stringency methods failed to produce detectable variation in hybridization patterns.

The two L1 probes used for these experiments were a 1.8 kb *Eco*R1 and a 6.1 kb *Eco*R1 fragment isolated from a genomic library of *Peromyscus maniculatus* (Kass *et al.* 1992). These fragments were subcloned into the plasmid pT7/T3-18 (Gibco BRL, Gathersburg, MD) and are named L1Pm62 and L1Pm143, respectively. Copy number for L1Pm62 is 500 per genome (Kass *et al.* 1992) and the copy number for L1Pm143 has not been determined. The portions of the LINE elements (L1Md-A2) that have been described for *Mus domesticus* (Loeb *et al.* 1986) covered by these two probes do not overlap; L1Pm143 includes sequences further upstream to L1Pm62 (Kass *et al.* 1992) plus part of the single copy flanking region.

Based on Kass *et al.*'s (1992) estimates of copy numbers, the medium stringency level (which is equivalent to hybridization at 70% or greater sequence similarity) would allow us to compare the chromosome distribution of LINEs to results described for *mys* (Baker & Wichman 1990).

*In situ* hybridization of the plasmid pT7/T3-18 employing the same methods used to detect the distribution of the LINE probe did not reveal visible hybridization to the chromosomes of *Peromyscus*, therefore the plasmid was not excised from the probes before hybridization for patterns of LINE.

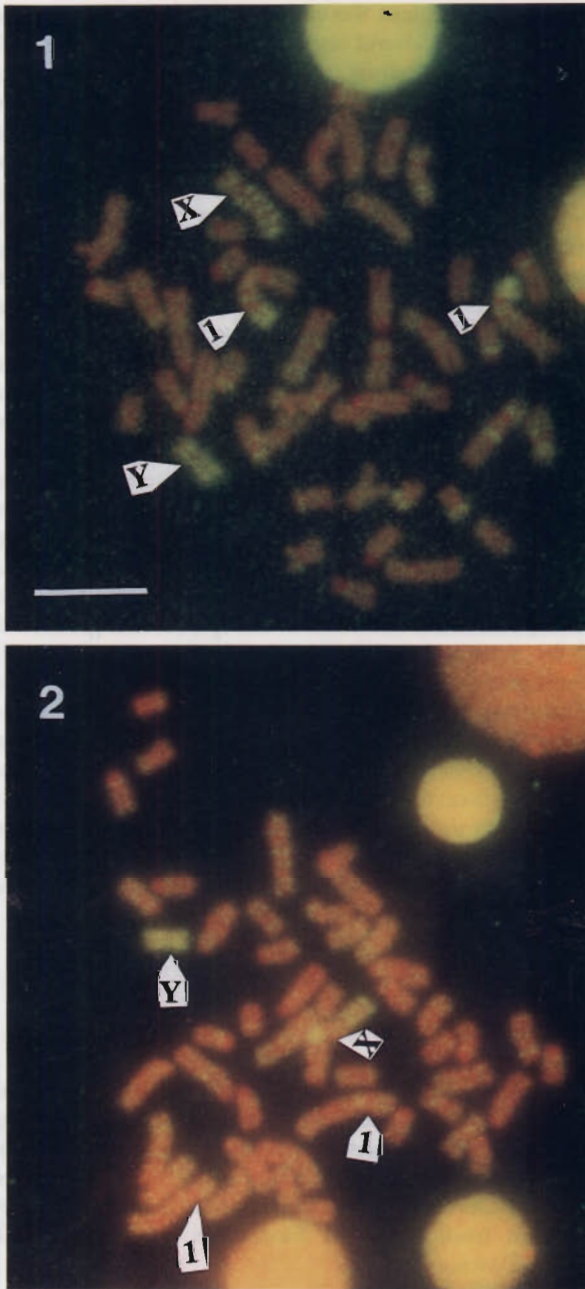
## Specimens examined

*Peromyscus maniculatus*: Texas: Castro Co., 5.5 miles S, 2.5 miles W Dimmit (1 female, TK 32216; 4 males, TK 32268, TK 32276, TK 32277, TK 32300); Mexico: Nuevo Leon, Ejido San Francisco, 1 mile N by road (1 male, TK 25761). *Peromyscus leucopus*: Oklahoma: Pottawatomie Co., 5.9 miles E, 2.5 miles N Tecumseh (5 males, TK 30188, TK 30198, TK 30202, TK 31501, TK 31593). At least five metaphase spreads were examined for each individual and photographic comparisons were made for each probe for each individual. Voucher specimens were deposited in the Collection of Mammals of the Museum, Texas Tech University.

## Results

Using the same stringency methods and the same individual of *P. maniculatus*, hybridization for L1Pm143 is shown in Figure 1 and for *mys* in Figure 2. The patterns described below for the chromosomal distribution of LINEs in *P. maniculatus* and *P. leucopus* were present in preparations for both L1Pm143 and L1Pm62 at low, medium and high hybridization stringencies. Hybridization was detectable on all chromosomes and did not appear to be random. The greatest amount of hybridization from the perspective of an entire chromosome was to the Y in males with the intensity of hybridization to the X chromosome(s) being slightly less prominent. The magnitude of difference in intensity between the sex chromosomes and the autosomes was sufficient to allow the X and Y to be distinguished easily from the remainder of the chromosomes.

The distribution of LINE on the long arm of the X is as follows. There are five evenly spaced fluorescent bands of approximately equal intensity (Figure 1). It is unclear whether two of these five bands closely align with the two major G-bands on the X, but clearly the five bands produced by *in situ* hybridization of LINE probes do not match the characteristic G-band pattern of the X of *P. maniculatus* and *P. leucopus* (Baker *et al.* 1983, Stangl & Baker 1984). For *mys* there are four major bands on the long arm, the telomeric band being the most intense (Figure 2). In *mys*, two of the four



**Figure 1.** Karyotype of a male *Peromyscus maniculatus* *in situ* hybridized with a LINE element (L1PM143) showing the non-random accumulation of this transposable element (see Methods for specific stringency methods used). Notice that the long arm of the X chromosome has five equally spaced areas of LINE accumulation. The short arm of chromosome 1 (Committee for Standardization of Chromosomes of *Peromyscus*, 1977) also shows a preferential accumulation of LINE relative to that observed for other autosomes.

**Figure 2.** Karyotype of the same male individual shown in Figure 1 *in situ* hybridized with the transposable element, *mys*. Notice that the long arm of the X chromosome has a different banding pattern than that observed for LINE in Figure 1. Also notice the difference in intensity of the short arm of the autosomal pair, chromosome 1.

bands appear identical in position to the two major G-bands that are generally present on the *Peromyscus* X (Committee for Standardization of Chromosomes of *Peromyscus* 1977).

A second feature of hybridization with L1 was their preferential detection in G-band regions of autosomes. There was general agreement between the pattern of L1 bands and that described for G-bands, but the level of distinction in the autosomes was not sufficient in our preparations to easily prepare a G-band karyotype for all chromosomes as has been described for *Mus* (Boyle *et al.* 1990).

Another prominent feature of the LINE *in situ* hybridization was the short arm of chromosome 1. The intensity of hybridization signal present on this arm equalled that present on the X chromosome; however, the intensity of hybridization on the long arm was typical of that seen in other autosomes. Finally, hybridization of L1 to chromosomes in *Peromyscus* was characterized by the absence of a detectable signal to the heterochromatic centromeric regions and to heterochromatic short arms.

## Discussion

For the observed chromosomal distribution of LINE and *mys* there are several major areas of concordance. First is the non-random pattern of chromosomal distribution. Second, autosomally, both show preferential hybridization to prominent G-bands and reduced hybridization intensity to the R-bands. Third, both probes fail to hybridize to the autosomal heterochromatin, as described by Baker & Wichman (1990, Figure 1). Fourth, both preferentially hybridize to the sex elements, with the Y having the greatest fluorescent signal intensity of any chromosome in the karyotype.

However, there are several aspects of the hybridization pattern that distinguish LINE from *mys*. LINE is clearly different from *mys* in its intensity of hybridization to the short arm of chromosome 1. This increased intensity does not correspond to the presence of G-bands. The short arm of chromosome 1 has been problematic in chromosome banding studies in that it sometimes appears heterochromatic and intermediate in staining between euchromatic and other heterochromatic regions. Originally, this arm was described as heterochromatic (Pathak *et al.* 1973); however, more recent work has described the arm as euchromatic (Committee for Standardization of Chromosomes of *Peromyscus*, 1977). Perhaps the exceptional accumulation of retroposons such as LINE may, in part, account for the observed variation in C-banding in the short arm of chromosome 1. A close examination of *mys* hybridized to the short arm of chromosome 1 does reveal a greater intensity of signal than is present in the long arm, but the magnitude of difference is slight and not comparable to that observed for LINE.

*In situ* hybridization of LINE is also distinct from *mys* in that for LINE the long arm of the X has five evenly spaced bands of equal hybridization intensity, whereas for *mys* there are four unevenly spaced bands, and in most preparations the telomeric band of the long arm exhibits the most intense signal. In hybridization experiments using *mys*, the bands on the long arm of the X are less distinct than those observed in hybridization using LINES.

Four potential mechanisms (sequence specific insertion, S-phase insertion, ectopic excision, recombinational editing) to produce non-random variation in copy number of transposable elements were proposed by Wichman *et al.* (1992). They outlined certain predictions for empirical data on TE distribution. A fifth mechanism to increase copy number of TEs was described by Nasir *et al.* (1991). This involved a tandemly repeated element that is amplified after insertion of a TE. Increase in copy numbers of TEs resulting from this mechanism could be detected by identification of the tandemly repeated element and should reflect copy number increase of the tandemly repeated element. Do any of these mechanisms adequately explain the *in situ* hybridization patterns observed in *Peromyscus* for LINE and *mys*?

That LINE and *mys* both accumulate in the G-bands is probably the result of sequence specific insertion. *mys* has an AT-rich target site (Wichman *et al.* 1985, Pine *et al.* 1988) which may explain its preferential occurrence in G-bands, which are AT-rich relative to the R-bands. LINES also have a preference for insertion in AT-rich regions, but no specific insertion sequence site has been observed (Korenberg & Rykowski 1988). That both LINES and *mys* occur in the same region of DNA is supported by cosmid genome library studies (Janacek *et al.* 1993). They examined over 2000 genomic fragments averaging 35 kb in length, and found that LINE and *mys* co-occurred in fragments at a frequency greater than expected by chance. These data suggest that the presence of either *mys* or LINE in a 35 kb region of DNA does not inhibit the insertion of the other TE (Furano *et al.* 1986, Sandmeyer *et al.* 1990).

From the perspective of understanding the significance of different mechanisms that regulate copies of TEs, a critical difference in chromosomal distribution between LINE and *mys* involves the X chromosome. Langley *et al.* (1989) proposed recombination and ectopic excision as a means of containing retrotransposon copy number in the genome. Baker & Wichman (1990) described a pattern of chromosomal distribution for *mys* that was compatible with the prediction of such an hypothesis, because the major portion of the X would encounter less recombination (and therefore less possible ectopic excision) and would be predicted to have a greater abundance of *mys*. This is what was observed (Baker & Wichman 1990).

LINE shares with *mys* the characteristic of preferential accumulation on the X, but some mechanism other than biased G-band insertion and ectopic excision is

required to explain the observed patterns. If insertion was biased to G-band regions then the pattern on the X would reflect G-bands, as is observed in most regions of the autosomes. Even though the effectiveness of the removal might vary for LINES compared with *mys*, the basic G-band pattern of the *Peromyscus* X should still be present. What is observed is two discretely different patterns on the long arm of the X (Figures 1, 2), one of which (*mys*) reflects G-band characteristics of the long arm of the X and the other (LINE) which does not. Might one of the other two mechanisms described in Wichman *et al.* (1992) account for the observed differences? We do not understand how either the S-phase insertion or the recombinational editing mechanisms would result in the different patterns observed on the long arm of the X.

Another feature that needs to be accounted for is the accumulation of LINE on the short arm of chromosome 1. Perhaps chromosome 1 is subject to little or no meiotic crossing over, but if so why does it not accumulate *mys* in the same proportions as LINE? Janacek *et al.* (1993) estimated the number of copies in the genome of *P. leucopus* to be 4700 for *mys* and 41 000 for LINE. Perhaps this difference in copy number is critical, but how this would favour only the short arm of chromosome 1 is unclear. Wichman *et al.* (1992) concluded that no single mechanism can adequately explain the empirical data on the chromosomal distribution of TEs. We agree with this conclusion, and the patterns observed for LINE and *mys* should be added to the bank of poorly understood data required to explain the nonrandom distribution of TEs.

## Acknowledgments

We thank Holly A. Wichman, Meredith J. Hamilton, Laura L. Janacek and Ira F. Greenbaum for reviewing drafts of this manuscript. We thank Mary Maltbie for technical assistance. This study was supported by the National Science Foundation (to R.J.B) and by a grant from the NIH Biomedical Research to Dr W. D. Dawson.

## References

- Baker RJ, Wichman HA (1990) Retrotransposon *mys* is concentrated on the sex chromosomes: implications for copy number containment. *Evolution* **44**: 2083–2088.
- Baker RJ, Robbins LW, Stangl FB Jr, Birney E (1983) Chromosomal evidence for a major subdivision in *Peromyscus leucopus*. *J Mamm* **64**: 356–359.
- Boyle AL, Ballard SG, Ward DC (1990) Differential distribution of long and short interspersed element sequences in the mouse genome—chromosome karyotyping by fluorescence *in situ* hybridization. *Proc Natl Acad Sci USA* **87**: 7757–7761.

- Committee for Standardization of Chromosomes of *Peromyscus* (1977) Standardized karyotype of deer mice, *Peromyscus* (Rodentia). *Cytogenet Cell Genet* **19**: 38–43.
- Fanning TG (1983) Size and structure of the highly repetitive BAM HI element in mice. *Nucleic Acids Res* **11**: 5003–5013.
- Furano AV, Somerville CC, Tschlis PN, D'Ambrosio E (1986) Target sites for the transposition of rat lung interspersed repeated DNA elements (LINEs) are not random. *Nucleic Acids Res* **14**: 3717–3727.
- Hattori M, Kuhara S, Takenaka T, Sakaki Y (1986) Sequence analysis of a KPN I family member near the 3' end of human beta-globin gene. *Nucleic Acids Res* **13**: 2745–2758.
- Hutchison CA, Hardies SC, Loeb DD *et al.* (1989) LINEs and related retroposon: long interspersed repeated sequences in the eucaryotic genome. In Berg DE, Howe MM, eds. *Mobile DNA*, American Society for Microbiology, Washington, DC, pp 593–617.
- Janacek LL, Longmire JL, Wichman HA, Baker RJ (1993) Genome organization of repetitive elements in the rodent, *Peromyscus leucopus*. *Mammalian Genome* **4**: 374–381.
- Kass DH, Berger FG, Dawson WD (1992) The evolution of coexisting highly divergent LINE-1 subfamilies within the rodent genus *Peromyscus*. *J Mol Evol* **33**: 472–485.
- Korenberg JR, Rykowski MC (1988) Human genome organization: alu, lines and the molecular structure of metaphase chromosome bands. *Cell* **53**: 391–400.
- Langley CH, Montgomery E, Hudson R, Kaplan N, Charlesworth B (1989) On the role of unequal exchange in the containment of transposable element copy number. *Genet Res* **52**: 223–235.
- Loeb DD, Padgett RW, Hardies SC *et al.* (1986) The sequence of a large L1Md element reveals a tandemly repeated S' end and several features found in retrotransposons. *Mol Cell Biol* **6**: 168–182.
- Manuelidis L, Ward DC (1984) Chromosomal and nuclear distribution of the HindIII 1.9 kb human DNA repeat. *Chromosoma* **91**: 28–38.
- Meinkoth J, Wahl G (1984) Hybridization of nucleic acids immobilized on solid supports. *Anal Biochem* **138**: 267–284.
- Montgomery E, Charlesworth B, Langley CH (1987) A test for the role of natural selection in the stabilization of transposable element copy number in a population of *Drosophila melanogaster*. *Genet Res* **49**: 31–41.
- Moyzis RK, Albright KL, Bartholdi MF *et al.* (1987) Human chromosome-specific repetitive DNA sequences; novel markers for genetic analysis. *Chromosoma* **95**: 375–386.
- Moyzis RK, Buckingham JM, Cram LS *et al.* (1988) A highly conserved repetitive DNA sequence (TTAGGG)<sub>n</sub> present at the telomeres of human chromosomes. *Proc Natl Acad Sci USA* **85**: 6622–6626.
- Nasir J, Maconochie MK, Brown SDM (1991) Co-amplification of L1 line elements with localised low copy repeats in Giemsa dark bands: implications for genome organization. *Nucleic Acids Res* **19**: 3255–3260.
- Pathak S, Hsu TC, Arrighi FE (1973) Chromosomes of *Peromyscus* (Rodentia, Cricetidae). IV. The role of heterochromatin in karyotypic evolution. *Cytogenet Cell Genet* **12**: 31–36.
- Pine DS, Bourekas EC, Potter SS (1988) *Mys* retrotransposon in *Peromyscus leucopus* and transgenic *Mus musculus*. *Nucleic Acids Res* **16**: 3359–3373.
- Sandmeyer SB, Hansen LJ, Chalker DL (1990) Integration specificity of retrotransposons and retroviruses. *Annu Rev Genet* **24**: 491–518.
- Stangl JR, FB, Baker RJ (1984) Evolutionary relationships in *Peromyscus*: congruence in chromosomal, genic, and classical data sets. *J Mamm* **65**: 643–654.
- Voliva CF, Jahn CL, Comer MB, Hutchison CA III, Edgell MH (1983) The L1Md long interspersed repeat family in the mouse: almost all examples are truncated at one end. *Nucleic Acids Res* **11**: 8847–8859.
- Wichman HA, Pine DS, Potter SS (1985) *Mys*, a family of mammalian transposable elements isolated by phylogenetic screening. *Nature* **317**: 77–81.
- Wichman HA, Van Den Bussche RA, Hamilton MJ, Baker RJ (1992) Transposable elements and the evolution of genome organization in mammals. *Genetica* **86**: 287–293.