Mammalian Genome 4, 374–381 (1993).

35

Mammalian Genome

© Springer-Verlag New York Inc. 1993

Genome organization of repetitive elements in the rodent, *Peromyscus leucopus*

Laura L. Janecek,[1,*] Jonathan L. Longmire,[1,3] Holly A. Wichman,[4] Robert J. Baker[1,2]

[1]Department of Biological Sciences, Box 43131, Texas Tech University, Lubbock, Texas 79409, USA,
[2]The Museum, Texas Tech University, Lubbock, Texas 79409, USA
[3]Life Sciences Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA
[4]Department of Biological Science, University of Idaho, Moscow, Idaho 83843, USA

Received: 7 December 1992 / Accepted: 16 March 1993

Abstract. To document the frequency and distribution of repetitive elements in *Peromyscus leucopus,* the white-footed mouse, a cosmid genomic library was examined. Two thousand thirteen randomly chosen recombinants, with an average insert size of 35 kb and representing 2.35% of the haploid genome of *P. leucopus,* were screened with probes representing microsatellites, tandem repeats, and transposable elements. Of the four dinucleotides, $(GT)_n$ was present in 87% of the clones, $(CT)_n$ was present in 59% of the clones, and $(AT)_n$ and $(GC)_n$ each was represented in our sample by a single clone (0.05%). $(TCC)_n$ was present in 8% of the clones. Of the tandem repeats, the 28S ribosomal probe and the $(TTAGGG)_n$ telomere probe were not represented in the library, whereas a heterochromatic fragment was present in 9% of the clones. A transposable element, *mys,* was estimated to occur in 4700 copies, whereas a long interspersed element (LINE) was estimated to occur in about 41,000 copies per haploid genome. LINE and *mys* occurred together in the same clones more frequently than expected on the basis of chance. Hybridizing the library to genomic DNA from *P. leucopus, Reithrodontomys fulvescens, Mus musculus,* and human produced general agreement between phylogenetic relatedness and intensity of hybridization. However, dinucleotide repeats appeared to account for a disproportionately high number of positive clones in the more distantly related taxa.

Introduction

Peromyscus is the most widely distributed and intensively studied genus of North American rodent. This genus comprises over 50 species and 200 subspecies and is found in nearly every terrestrial habitat, from the subarctic of Canada to tropical regions of Panama (Kirkland and Layne 1989). *Peromyscus* is of considerable value to studies of mammalian behavior, ecology, physiology, and development (Kirkland and Layne 1989, and citations therein). Additionally, the importance of this genus to biological research is reflected by the establishment of the NSF-supported *Peromyscus* Stock Center at the University of South Carolina, where researchers can obtain individuals representing wild types, mutants, and inbred lines of several species of *Peromyscus*.

In recent years, *P. leucopus* has been the subject of studies investigating the dynamics and maintenance of hybrid zones (Baker et al. 1983b, 1991; Nelson et al. 1985; Stangl 1986; Adkins et al. 1991), effects of petrochemical exposure on structural aberrations of chromosomes (McBee et al. 1987; McBee 1991), effects of genes of the major histocompatibility complex (MHC) on maximum life span and rate of aging (Smith et al. 1989; Crew et al. 1989, 1990), chromosomal distribution of satellite DNA (Hamilton et al. 1992), and documentation of the presence and chromosomal location of repetitive elements such as the $(TTAGGG)_n$ telomeric sequence (Meyne et al. 1990) and transposable elements such as the retrotransposon *mys* (Wichman et al. 1985; Baker and Wichman 1990).

The recent interest in the structure, function, and chromosomal location of genes and repetitive elements in *P. leucopus* suggests that this species will be an excellent model for investigations into the frequency, organization, and distribution of repetitive elements within a genomic cosmid library constructed from an individual of this highly successful native mammal. We examined the distribution within the library of four general classes of probes: (1) di- and tri-nucleotide microsatellites [$(AT)_n$, $(GC)_n$, $(CT)_n$,

**Present address:* Savannah River Ecology Laboratory, Drawer E, Aiken, South Carolina 29802, USA

Correspondence to: R.J. Baker

$(GT)_n$, $(TCC)_n$]; (2) tandemly repeated sequences such as the 28S ribosomal subunit, the $(TTAGGG)_n$ sequence characteristic of vertebrate telomeres, and an element identified as N4, which is characteristic of centromeric regions of *Peromyscus* chromosomes; (3) the transposable element LINE L1Pm55 and the retrotransposon *mys*-1; and (4) genomic DNAs isolated from four mammals (*P. leucopus, Reithrodontomys fulvescens, Mus musculus,* and *Homo sapiens*). Hybridizations with mammalian genomic DNAs were conducted to give a measure of the extent to which repetitive elements in the *P. leucopus* genome would cross-hybridize with those of other mammals. Because previous studies have documented the chromosomal distribution of many of the repetitive elements used in this study, we will be able to make correlations between the representation of these elements in the genomic library and their location and distribution on chromosomes of *P. leucopus*, as evidenced by in situ hybridization.

Materials and methods

Construction of cosmid genomic library

High-molecular-weight genomic DNA was isolated from a wild-caught male *P. leucopus* from Garza County, Texas (TK27500), according to the procedure of Longmire and coworkers (1988), with the modification that minced testicular tissue, rather than blood, was used as a source of cellular DNA. After partial digestion with *Sau*3AI and dephosphorylation with calf intestinal alkaline phosphatase, approximately 1.0 μg of *P. leucopus* genomic DNA was ligated into 2.0 μg of *Bam*HI cloning arms from the cosmid vector sCos-1 (Evans et al. 1989). In vitro packaging was carried out in Giga Pack Gold packaging extracts (Stratagene). Primary infection of *E. coli* host strain DH5αMCR yielded 1.37×10^5 independent recombinants, giving an approximately twofold statistical representation of *P. leucopus* genomic sequences. Average size of inserts in the cosmid vector was determined by digestion of 20 random clones with *Eco*RI followed by electrophoresis within a 0.7% agarose gel run at 35V for 26 h. Fragment sizes were determined from photographs of the ethidium-stained gel with the computer programs NCSA Gel Reader 2.02 and Adobe Photoshop 2.0 for the Macintosh.

Molecular characterization of repetitive elements in the genomic library

Two thousand sixteen independent clones from the primary library were picked, grown, and archived in 96-well microtiter plates. A replica plater (Sigma Chemical Co.) was used to inoculate nylon membranes (MIS Magna NT 0.45 micron; Biodyne B 0.45 micron) with clones from the microtiter plates. Membranes were incubated at 37°C for 7 h on LB agar containing kanamycin (30 μg/ml) and then were transferred to LB agar containing kanamycin and chloramphenicol (170 μg/ml; Sambrook et al. 1989) and grown at 37°C overnight. DNA was fixed onto membranes by placing membranes sequentially on blotting pads soaked in 0.4 *M* NaOH (5 min), 0.5 *M* Tris/1.5 *M* NaCl, pH 7.5 (5 min), and 2× SSC (5 min), followed by baking at 80°C for 1–3 h.

Membranes were hybridized with a variety of radiolabeled repetitive elements to determine the distribution and relative abundance of these elements in the genome of *P. leucopus*. Membranes were washed prior to hybridization at 65°C for 1 h in 0.1× SSC, 0.1% SDS. Prehybridization was carried out at 65°C for 1 h in 6× SSC, 40% formamide (Kodak), 1% SDS, 0.005 M EDTA (pH 8.0), and 0.005 g/ml Carnation evaporated milk. Membranes were hybridized overnight at 42°C in fresh prehybridization solution containing approximately 1×10^6 cpm/ml probe. With the exception of the $(TCC)_n$ microsatellite, all probes were labeled with [α-^{32}P]dCTP or [α-^{32}P]dATP by nick translation. The $(TCC)_n$ microsatellite was labeled with [α-^{32}P]dCTP by primer extension (Feinberg and Vogelstein 1983). Nonincorporated label was removed by spin column chromatography (Sambrook et al. 1989). Prior to hybridization, probes were denatured for 10 min at 37°C in 0.1 M NaOH. Following hybridization, membranes were washed once for 15 min in 2× SSC, 0.1% SDS at room temperature and twice for 15 min in 0.1× SSC, 0.1% SDS at 50°C. Washed membranes were autoradiographed at −80°C with Kodak XAR-5 film and Lightning Plus intensifying screens.

Source of repetitive elements and genomic DNAs

Genomic DNAs were isolated from tissues of a male *P. leucopus* from Kiowa Co., OK (TK30023) and a male *Reithrodontomys fulvescens* from Seminole Co., OK (TK23409), and from placental tissue of a male *Homo sapiens*, with techniques modified from Longmire and colleagues (1988). Dinucleotides $(GT)_n \cdot (AC)_n$, $(CT)_n \cdot (AG)_n$, $(AT)_n \cdot (AT)_n$, $(GC)_n \cdot (GC)_n$ were purchased from Pharmacia LKB. The $(TTAGGG)_8$ telomere repeat (Moyzis et al. 1988) and $(TCC)_{n[n\sim30]}$ trinucleotide were synthesized with a Beckman System I DNA synthesizer. The 28S ribosomal subunit (provided by N. Arnheim), *mys*-1 (Wichman et al. 1985), N4 (Hamilton et al. 1992), and LINE L1Pm55 (Kass et al. 1992) probes were supplied as fragments cloned into plasmid vectors. Because plasmid vectors are complementary to the sCOS-1 cosmid vector used for construction of the genomic library, these probes were prepared by removing the inserts with appropriate restriction enzymes. Inserts were gel-purified twice according to the Prep-a-Gene protocol (Bio-Rad). The portion of *mys*-1 used for the probe was a 2.5 kb *Pst*I fragment, which contained all the internal sequence and one LTR (Wichman et al. 1985). The L1Pm55 LINE probe was a 1.5-kb fragment isolated from a lambda Charon 30 genomic library from *P. maniculatus* screened with a MIF-1 LINE probe (Kass et al. 1992).

Results

Characterization of the P. leucopus genomic library

Only 12 of the 2016 cosmid clones screened in this study did not hybridize to any of the 14 probes examined. Of these 12 clones (0.6% of those screened), three produced no bands when they were digested with the restriction endonuclease *Eco*RI, electrophoresed on a 0.7% agarose gel, and stained with ethidium bromide. These three clones either did not contain the cosmid or failed to grow to a density that allowed detection of the cosmid DNA by a standard miniprep procedure, bringing the actual number of cosmids screened to 2,013. The remaining nine clones were visible on an ethidium-stained gel; when digested with *Eco*RI, each clone was verified to have the 6.7-kb DNA fragment characteristic of the sCOS-1 vector, as well as additional bands resulting from restriction of the insert DNA. For each of the 14 probes, each clone was scored on a scale of 0 (no detectable hybridization) to 3, a completely black spot on an autoradiograph (maximum detectable hybridization).

The size of *P. leucopus* DNA inserted into 20 randomly selected cosmid clones ranged from 18 to 49 kb, with a mean of 35 kb. On the basis of this mean insert size, the 2013 positive clones represented 70.5×10^6

bp, or 2.35% of the *P. leucopus* genome, assuming a genome size of 3×10^9 bp.

Representation of dinucleotide and trinucleotide microsatellites

The four dinucleotides used in this study represent all possible combinations of two different dinucleotides, because $(GT)_n$ is $(AC)_n$ and $(CT)_n$ is $(AG)_n$ on the other strand, and, for example, $(AT)_n$ is equivalent to $(TA)_n$. The four dinucleotide microsatellites varied greatly in their representations among the 2013 clones examined. Dinucleotides $(GT)_n$ and $(CT)_n$ each hybridized to more than half of the clones that were screened, whereas $(AT)_n$ and $(GC)_n$ each was found to hybridize only to single clones (Table 1). The trinucleotide $(TCC)_n$ was used because it has proven useful in a variety of vertebrates as a DNA-fingerprinting and gender-identification probe (Epplen et al. 1991). $(TCC)_n$ was detected in 8% of the clones screened. Of the clones positive for $(GT)_n$ and $(CT)_n$, 87% and 68% respectively were scored as maximally hybridizing (score = 3), whereas the majority (97%) of the clones positive for $(TCC)_n$ hybridized submaximally (score = 1 or 2); only 4 clones out of 160 positives for $(TCC)_n$ (3%) were screened as maximally hybridizing.

Representation of transposable elements

The two transposable elements, *mys*-1 and LINE L1Pm55, also varied in their representations among the clones that were screened. *Mys*-1 was detected in approximately 5% of the clones, whereas LINE L1Pm55 hybridized to almost half (Table 1). Of the clones positive for *mys*-1 and LINE L1Pm55, 1.8% and 1.0% respectively hybridized maximally (score = 3).

Representation of tandem repeats

Two of the tandem repeats examined, the 28S ribosomal subunit and the $(TTAGGG)_n$ telomeric repeat, were not detected in any of the clones screened in this study. The N4 tandem repeat was detected in 9% of the clones examined (Table 1). This centromeric repeat produced a hybridization pattern in which the majority (90%) of the N4-positive clones hybridized maximally to the probe (score = 3).

Table 1. Representation of 14 probes in 2013 independent clones from a cosmid library constructed from *Peromyscus leucopus* genomic DNA. Percentages are shown in parentheses.

Probe	Negative clones	Positive clones
$(AT)_n$	2015 (99.95)	1 (0.05)
$(GC)_n$	2015 (99.95)	1 (0.05)
$(CT)_n$	821 (40.72)	1195 (59.28)
$(GT)_n$	253 (12.55)	1763 (87.45)
$(TCC)_n$	1856 (92.06)	160 (7.94)
mys-1	1905 (94.49)	111 (5.51)
LINE L1Pm55	1058 (52.48)	958 (47.52)
N4	1834 (90.97)	182 (9.03)
28S	2016 (100.00)	0 (0.00)
Telomere	2016 (100.00)	0 (0.00)
P. leucopus genomic DNA	120 (5.95)	1896 (94.05)
Reithrodontomys fulvescens genomic DNA	256 (12.70)	1760 (87.30)
Mus musculus genomic DNA	372 (18.45)	1644 (81.55)
Homo sapiens genomic DNA	1105 (54.81)	911 (45.19)

Hybridization to genomic DNAs

Hybridization of the *P. leucopus* library to genomic DNA from *P. leucopus* resulted in some degree of hybridization to 94% of the clones examined. Of these positive clones, 8.6% were scored as maximally hybridizing (score = 3). In contrast, hybridization to *Reithrodontomys fulvescens, Mus musculus,* and *Homo sapiens* genomic DNAs resulted in decreasing levels of hybridization as the species used for the probe became more distantly related to *P. leucopus* (Table 1). Hybridization of the *P. leucopus* library to *Reithrodontomys* genomic DNA resulted in about 87% positive clones, whereas positive hybridization decreased with *Mus* to about 81% and with human to about 45%. The numbers of clones scored as maximally hybridizing also showed a phylogenetic pattern, with 2.7% of the *R. fulvescens* clones, 0.3% of the *Mus* clones, and none of the human clones being scored as maximally hybridizing.

Genomic organization of repetitive elements

The co-occurrence of pairs of repetitive element probes in the *P. leucopus* genomic cosmid clones is presented in Table 2. Most pairs of elements were detected together at levels expected on the basis of their representation in individual clones. However, several pairs of probes were detected together less often than expected [$(CT)_n/(TCC)_n$, N4/$(CT)_n$, N4/$(GT)_n$, N4/$(TCC)_n$, N4/*Reithrodontomys* genomic, N4/*Mus* genomic, N4/human genomic]. Other probe combinations were detected together more often than expected [$(CT)_n/(GT)_n$, *mys*-1/LINE L1Pm55, $(CT)_n$/*Mus* genomic, $(CT)_n$/human genomic, $(GT)_n$/*Reithrodontomys* genomic, $(GT)_n$/*Mus* genomic, $(GT)_n$/human genomic, $(TCC)_n$/*Mus* genomic, $(TCC)_n$/human genomic, LINE L1Pm55/human genomic].

Discussion

Prevalence and distribution of sequences in the P. leucopus genome

There are sufficient published observations on the distribution of repetitive sequences in the mammalian genome to permit comparisons. Most recent analyses of the distribution of repetitive elements within mammalian genomes have been conducted on human genomic DNA by direct experimental methods as well as analyses of published DNA sequences from GenBank. Our methods sample random clones from a cosmid genomic library to estimate the abundance of specific se-

Table 2. Pairwise comparisons of the 12 probes detected in 2013 cosmid clones from the genomic library of *Peromyscus leucopus*. Two probes (28S, telomeric sequence) that failed to hybridize to any library clones are not included. Above diagonal: expected number for each pair of probes based on total occurrence in the 2013 clones examined. Chi-square values are shown in parentheses. Below diagonal: observed number for each pair of probes; "+" indicates pairs of probes observed more often than expected, "−" indicates pairs of probes observed less often than expected.

Probe	$(AT)_n$	$(GC)_n$	$(CT)_n$	$(GT)_n$	$(TCC)_n$	*mys*-1	LINE L1Pm55	N4	*P. leucopus* genomic DNA	*R. fulvescens* genomic DNA	*Mus musculus* genomic DNA	*Homo sapiens* genomic DNA
$(AT)_n$	- - - - -	0.00 (0.00)	0.60 (0.60)	0.88 (0.02)	0.08 (0.08)	0.06 (0.06)	0.48 (0.56)	0.09 (0.09)	0.95 (0.003)	0.88 (0.02)	0.82 (0.04)	0.46 (0.46)
$(GC)_n$	0	- - - - -	0.60 (0.27)	0.88 (0.02)	0.08 (0.08)	0.06 (0.06)	0.48 (0.56)	0.09 (0.09)	0.95 (0.003)	0.88 (0.02)	0.82 (0.04)	0.46 (0.63)
$(CT)_n$	0	1	- - - - -	1045.5 (11.89)[a]	94.9 (6.02)[a]	65.8 (1.02)	567.6 (0.32)	107.6 (88.53)[a]	1123.98 (1.09)	1143.31 (1.45)	974.59 (24.46)[a]	540.06 (25.76)[a]
$(GT)_n$	1	1	1157 +	- - - - -	140.0 (1.61)	97.1 (0.09)	837.4 (0.07)	158.7 (123.0)[a]	1658.09 (0.29)	1539.09 (17.46)[a]	1437.72 (19.93)[a]	796.70 (13.92)[a]
$(TCC)_n$	0	0	71 −	155	- - - - -	8.8 (1.16)	76.0 (2.22)	14.4 (10.68)[a]	150.55 (0.04)	139.74 (1.67)	130.54 (3.86)[a]	72.34 (12.99)[a]
mys-1	0	0	74	100	12	- - - - -	52.7 (5.04)[a]	10.0 (3.6)	104.47 (0.02)	96.97 (0.09)	90.59 (0.21)	50.20 (0.29)
LINE L1Pm55	1	1	581	830	89	69 +	- - - - -	86.5 (0.07)	901.00 (0.81)	836.34 (0.19)	781.25 (1.06)	432.92 (38.49)[a]
N4	0	0	10 −	19 −	2 −	4	84	- - - - -	171.21 (0.68)	158.93 (126.75)[a]	148.46 (119.98)[a]	82.27 (70.71)[a]
Peromyscus leucopus genomic DNA	1	1	1159	1680	153	103	928	182	- - - - -	1655.25 (1.38)	1546.23 (3.15)	856.82 (3.30)
Reithrodontomys fulvescens genomic DNA	1	1	1184	1703 +	155	100	849	17 −	1703	- - - - -	1435.25 (27.52)[a]	795.33 (16.53)[a]
Mus musculus genomic DNA	1	1	1129 +	1607 +	153 +	95	810	15 −	1616	1634 +	- - - - -	742.95 (37.11)[a]
Homo sapiens genomic DNA	0	1	658 +	902 +	103 +	54	562 +	6 −	910	910 +	909 +	- - - - -

[a] Indicates a significant difference between observed and expected values ($p < 0.05$).

quences for which probes are available and thereby provide an alternative means of estimating the extent and distribution of these sequences. One strength of using different methods to provide estimates of copy number and genomic organization is that when estimates are corroborating, the probability of having an accurate estimate becomes more likely. Because our method examines clones with DNA inserts as large as 40 kb, more than one copy of a given repetitive element may be present in any given positive clone; these copy number estimates must be considered as minimal estimates. However, the use of large inserts allows us to detect any nonrandom association of elements within the genome.

Microsatellite sequences

Recent investigations have revealed that, in human, $(GT)_n$ sequences occur in 50,000–100,000 copies per haploid genome (Stallings et al. 1991). This estimate corroborates earlier investigations by Hamada and associates (1982) that identified $(GT)_n$ repetitive sequences in a variety of eukaryotes and estimated that the number of copies of $(GT)_n$ sequences ranged from about 100,000 in the *Mus* haploid genome to about 50,000 in human and about 30,000 in the cow. Examination of a cosmid library constructed from flow-sorted human Chromosome (Chr) 16 DNA resulted in the estimation that about 50% of cosmid-sized recombinants (~40 kb) cloned from human DNA should hybridize to a synthetic $(GT)_{25}$ sequence, resulting in occurrence of a $(GT)_n$ repeat every 50–100 kb (Stallings et al. 1990). Other studies have agreed with these values, with Stallings and coworkers (1991) estimating that $(GT)_n$ sequences are under-represented in human centric heterochromatin but occur every 30 kb in human euchromatic regions. Similar results were obtained by Pardue and colleagues (1987), who used in situ hybridization to document that $(GT)_n$ sequences were dispersed over euchromatin in several species of *Drosophila*, with an especially heavy concentration of $(GT)_n$ sequences on the X Chr. However, Pardue and associates (1987) found $(GT)_n$ sequences to be conspicuously absent from regions of centromeric β-heterochromatin and on the dot chromosome of all *Drosophila* species examined except *D. virilis*. Moyzis and coworkers (1989) found $(GT)_n$ repetitive elements to comprise approximately 0.09% of 29 human DNA sequences analyzed from GenBank; these authors estimated that $(GT)_n$ sequences averaged 40 bp in length and occurred, on average, every 54 kb in human DNA. $(GT)_n$ repetitive sequences are more prevalent in those rodents that have been examined. Stallings and associates (1991) found that 77.8% of cosmid clones constructed from *Mus* genomic DNA contained at least one $(GT)_n$ sequence, as compared with 63% of human cosmid clones. Analysis of GenBank sequences also allowed Stallings and coworkers (1991) to estimate that $(GT)_n$ repetitive sequences ($n > 6$) occurred every 18 kb in *Mus* and every 21 kb in *Rattus*, corroborating earlier studies (Hamada et al. 1982) that found $(GT)_n$

sequences to be more prevalent in rodent than in human genomic DNA.

Hybridization of the *P. leucopus* cosmid genomic library to a $(GT)_n$ probe resulted in positive hybridization to 87% of the clones (Table 1); 76% of these clones were scored as maximally hybridizing (score = 3). Assuming that 2.35% of the genome was examined, we estimate that the *P. leucopus* haploid genome contains at least 75,000 $(GT)_n$ repetitive sequences. This estimate agrees well with that of Stallings and colleagues (1991) that mammalian genomes contain 50,000–100,000 copies of $(GT)_n$ sequences and that $(GT)_n$ sequences are more prevalent in rodent than in human DNA.

Although there have been many studies of the $(GC)_n$ repetitive sequence because this dinucleotide repeat has been documented to have Z-DNA-forming potential, few studies have examined the distribution in genomes of the $(GC)_n$ microsatellite. Hamada and associates (1982) looked at the distribution of $(GC)_n$ sequences in DNA from several eukaryotes. Their results suggested that $(GC)_n$ sequences were moderately repetitive in the genomes of human, *Mus*, and salmon, but were not documented in the genomes of cow or yeast. In contrast, Stallings (1992) surveyed more than 27 Mb of GenBank sequence and found the frequency of $(GC)_n$ was about 20% lower than expected and was the rarest dinucleotide repeat in vertebrate genomes. Although a few $(GC)_n$ repeats were found in GenBank sequences, all were short ($n \leq 12$); all other dinucleotide repetitive sequences could be found where the length of the repeat exceeded $n = 35$ (Stallings 1992). Our survey of the *P. leucopus* cosmid library identified a single clone that hybridized moderately (score = 2) to $(GC)_n$, providing a tentative estimate of 40 copies in the *P. leucopus* genome. Our results and those of Stallings (1992) suggest that, unlike $(GT)_n$, $(GC)_n$ repetitive regions are rare in vertebrate genomes.

The $(CT)_n$ dinucleotide repeat was the second most common microsatellite present in the clones screened, and we estimated $(CT)_n$ to be represented by 50,000 copies in the *P. leucopus* haploid genome. In contrast, only about 40 $(AT)_n$ sequences and about 6,800 $(TCC)_n$ sequences were estimated to be present. The relative abundance of $(AT)_n$ and $(CT)_n$ repetitive sequences in the *P. leucopus* library agrees well with the frequency of these microsatellites in GenBank sequences. Stallings (1992) found $(CT)_n$ sequences to be second in frequency to $(GT)_n$ in mammals in general and in rodents in particular. $(AT)_n$ was the third most common dinucleotide repeat in rodent DNA, being two to three times less frequent than $(CT)_n$ and about 25 times more common than $(GC)_n$ sequences.

Previous studies that examined human DNA sequences from GenBank (Stallings et al. 1991) and in situ hybridization of *Drosophila* chromosomes (Pardue et al. 1987) found $(GT)_n$ repetitive sequences to be under-represented in centromeric regions of chromosomes. Although we did not use in situ hybridization to examine directly the location of $(GT)_n$ sequences on the chromosomes of *P. leucopus*, comparisons of the pairwise occurrence of probes in specific cosmid clones (Table 2) led to the conclusion that $(GT)_n$ sequences also are under-represented in centromeric regions of *P. leucopus*. The centromeric tandem repeat N4, which hybridized to 9% of the *P. leucopus* clones (Table 1), occurred significantly less often than expected in combination with the repetitive sequences $(GT)_n$, $(CT)_n$, and $(TCC)_n$ (Table 2). The N4 repeat is localized to all centromeric regions and a few heterochromatic short arms, including their telomeric regions (Hamilton et al. 1992). The conspicuous lack in *P. leucopus* of $(GT)_n$, $(CT)_n$, and $(TCC)_n$ sequences in combination with N4 suggests that these microsatellite sequences are under-represented in centromeric regions and possibly the telomeric regions as well. The observation that over 87% of the cosmid clones hybridized positively to $(GT)_n$ suggests that $(GT)_n$ sequences may be fairly evenly distributed over euchromatic regions in *P. leucopus*, just as they appear to be in human (Stallings et al. 1991) and *Drosophila* (Pardue et al. 1987). However, more specific conclusions concerning the distribution of microsatellite repeats on the chromosomes of *P. leucopus* await future in situ hybridization studies.

Transposable elements

The two mobile DNA elements used as probes in this study varied significantly in their representation in the *P. leucopus* cosmid library. *Mys* is a retrotransposon originally isolated from *P. leucopus* by phylogenetic screening (Wichman et al. 1985). Sequences complementary to *mys* are present in the genomes of all species of *Peromyscus* and related cricetid rodents examined to date, but are absent from *Mus* and other mammals (Baker and Wichman 1990). It has been estimated that there are 500–1000 2.8-kb copies of *mys* with *Eco*RV sites in both LTRs in the genome of *P. leucopus* and an unknown number of other elements which cross-hybridize (Wichman et al. 1985). *Mys*-1 (Wichman et al. 1985) hybridized to 111 cosmid clones from the *P. leucopus* library (Table 1), permitting us to estimate that there are about 4700 *mys*-1 related elements in the haploid genome of *P. leucopus*. This frequency of *mys* in the *P. leucopus* genome is higher than the 500–1000 copies previously estimated by Wichman and coworkers (1985), suggesting that a large number of *mys*-related sequences such as M-9 and lone LTRs (Wichman et al. 1985) occur in the genome. We currently are characterizing the cosmid clones positive for *mys*-1 to determine the nature of *mys*-related sequences that exist in the genome of *P. leucopus*.

In contrast, the repetitive LINEs (L1) exist at a frequency of 10,000–100,000 copies per haploid genome in all mammals (Burton et al. 1986) and are found in human DNA once every 30–60 kb (Moyzis et al. 1989). Full-length LINEs may be 6–7 kb long, but approximately 95% of them are heterogeneously truncated from their 5′ end. Many LINEs are rearranged

and individual LINEs within a species may share greater than 80% sequence identity (Belmaaza et al. 1990; Kass et al. 1992). LINE L1Pm55 hybridized to over 47% of the cosmid clones and was estimated to be present at about 41,000 copies per haploid genome in *P. leucopus*. However, only 1% of the positive clones hybridized maximally to LINE L1Pm55, which may indicate that only about 400 full-length copies of LINE L1Pm55 are present in the *P. leucopus* genome, if one assumes that maximal hybridization can be equated to the detection of full-length LINEs in the genomic library. Kass and colleagues (1992) estimated that LINE L1Pm55 exists in about 100 copies in the *Peromyscus* genome, but these authors also documented that LINE L1Pm55 cross-hybridizes with another LINE (L1Pm62) that exists as about 500 copies in *Peromyscus*. Thus, the 1% of the cosmid clones that hybridized maximally to the LINE L1Pm55 probe may contain full-length LINEs that are homologous to both LINE L1Pm55 and LINE L1Pm62. Alternatively, some of these maximally hybridizing clones may contain multiple copies of LINEs, or the maximally hybridizing clones may be those for which the nucleotide sequence identity of the probe and the LINE in the genome was greatest. The majority of the clones positive for LINE L1Pm55 hybridized faintly, which may indicate that repetitive elements that share only minimal sequence identity with LINE L1Pm55 exist in the genome of *P. leucopus*.

Both *mys*-1 and LINE L1Pm55 occurred together in cosmid clones significantly more often than expected on the basis of their individual frequencies (Table 2). This association was not unexpected, given that both repetitive elements have been documented to occur predominantly in A + T-rich regions that are characteristic of chromosomal G/Q bands (Wichman et al. 1985, 1992; Korenberg and Rykowski 1988; Pine et al. 1988).

Tandem repeats

A large portion of the mammalian genome may consist of blocks of tandemly repeated DNA sequences (reviewed by Singer 1982). We examined the frequency of three tandemly repeated sequences: the 28S ribosomal subunit, which exists in mammalian genomes as a few hundred copies usually located in nucleolus-organizer regions of several chromosomes (Wellauer and Dawid 1979; Gerbi 1985), the $(TTAGGG)_n$ sequence characteristic of vertebrate telomeres (Moyzis et al. 1988), and the N4 tandem repeat characteristic of all centromeric and some telomeric regions of *Peromyscus* chromosomes (Hamilton et al. 1992).

The 28S and telomeric sequences did not hybridize to any of the 2013 clones examined from the *P. leucopus* library. The cosmid library was constructed from a *Sau*3AI partial digest of *P. leucopus* genomic DNA, and because the $(TTAGGG)_n$ telomeric sequence is distributed in *P. leucopus* only on telomeres and not intrachromosomally (Meyne et al. 1990), we did not expect this probe to hybridize to any of the library clones.

The 28S ribosomal tandem repeats are located on two pairs of chromosomes at centromeric sites and on three pairs of chromosomes at telomeric sites in the diploid karyotype of *P. leucopus* (unpublished data, R.J. Baker and M.J. Hamilton). Given the restricted chromosomal distribution of this tandem repeat, we expected our screening of the *P. leucopus* genomic library to produce at most one or two positive clones. Therefore, the fact that none of the 2013 clones hybridized to the 28S probe was not surprising.

The N4 tandem repeat was isolated originally from *P. leucopus* by phylogenetic screening of a genomic library (Wichman et al. 1985, 1990) to identify rapidly evolving repetitive DNA sequences. N4 is about 7.3 kb long and consists of a tandem arrangement of four restriction fragments of similar, but not identical, size (Hamilton et al. 1992). In situ hybridization of N4 to the chromosomes of *P. leucopus* resulted in strong hybridization of the probe to all centromeric regions as well as to the short arms of six pairs of chromosomes (Hamilton et al. 1992). The N4 probe hybridized to 9% of the *P. leucopus* cosmid clones screened (Table 1), with about 90% of the positive clones hybridizing maximally (score = 3). This pattern of hybridization to the cosmid clones is exactly what is expected for a large cluster tandem repeat: the clone either contains the repeat and hybridizes maximally, or the repetitive element is absent from the clone and no detectable hybridization occurs. N4 has been in situ hybridized extensively to a large number of individuals of *P. leucopus* as part of ongoing studies in our laboratory, and the hybridization frequency of 9% to the cosmid library is compatible with the extent of hybridization of N4 to the karyotype of *P. leucopus*.

Comparisons of hybridization patterns of N4 in combination with other probes used in this study indicate that *P. leucopus* cosmid clones that hybridized to N4 also hybridized significantly less often to probes made from a closely related genus, *Reithrodontomys* (Carleton 1980, 1989; Baker et al. 1983a), or more distantly related taxa such as *Mus* or human genomic DNA (Table 2). This is expected given that the N4 tandem repeat has been shown to be specific for the genus *Peromyscus* (Hamilton et al. 1992). The few clones that did cross-react with N4 and non-*Peromyscus* genomic DNA probes (Table 2) hybridized submaximally (score = 1 or 2) with the genomic DNA probes.

Extent of sequences shared between genomes

The extent to which repetitive sequences of various types are shared between mammalian genomes was examined by probing the *P. leucopus* cosmid library with genomic DNAs from other rodents and human. Under these conditions, only clones containing repetitive sequences common to both species are expected to be detected because each single or low-copy sequence makes up such a small proportion of each probe. As expected, the greatest hybridization to the *P. leucopus* library was observed with the *P. leucopus*

genomic DNA probe, with 94% of all clones hybridizing to some degree (Table 1). Some library clones did not hybridize to the *P. leucopus* DNA probe, indicating that unique DNA sequences were not detected under these hybridization conditions (Crampton et al. 1981). The intensity of hybridization of library clones to these genomic DNA probes made from *Reithrodontomys, Mus,* and human also showed a phylogenetic pattern, with hybridization intensity decreasing as the phylogenetic distance increased between *P. leucopus* and the three other species used as probes.

Which types of repetitive sequences account for hybridization between the genomic DNA of one species and the *P. leucopus* library? Almost the entire difference between *P. leucopus* and *Reithrodontomys* can be accounted for by the failure of *Reithrodontomys* genomic DNA to detect the major *Peromyscus* satellite DNA (N4). Thus, the majority of the interspersed sequences appear to be very similar between these two species, within the limits of this technique. Further insight into this pattern can be gained by examining the nature of the clones that hybridized to genomic DNA. The greatest proportion of probes that hybridized to genomic DNA were positive for one of the four dinucleotides or the trinucleotide $(TCC)_n$. As the phylogenetic distance increased, the percentage accounted for by dinucleotides increased, and most of the hybridization observed between *P. leucopus* and human may be accounted for by microsatellites. With the exception of 25 clones [one $(TCC)_n$ positive, seven $(CT)_n$ positive, and 17 $(GT)_n$ positive], all clones identified as positive for dinucleotides or $(TCC)_n$ also hybridized to genomic DNA from *P. leucopus*. For the LINE L1Pm55 probe, only nine clones hybridized weakly to LINE and did not hybridize to genomic DNA from *P. leucopus*. The most probable explanation for clones being positive for a specific probe but not hybridizing to genomic DNA is that the probe was underrepresented in the genome. However, this explanation is inadequate to explain all observed variation. Only two probes hybridized to *P. leucopus* genomic DNA that did not hybridize to one or more of the other probes used.

With greater phylogenetic distance between *P. leucopus* and the taxon used as a source of genomic DNA, the presence of microsatellites accounts for a greater percentage of sequences that hybridize to a specific genomic DNA. As the level of hybridization of clones decreases for the entire library, the percentage of positive clones that are also microsatellite positive increases. With genomic DNA from more divergent taxa, such as human, used as a probe, the microsatellite component of the genome probably accounts for a disproportionately high number of positive clones identified in the *P. leucopus* library.

Another correlation was found more often than expected between *P. leucopus* clones positive for LINE L1Pm55 and human genomic DNA (Table 2). Because genomic DNAs from taxa more closely related to *P. leucopus* did not hybridize significantly more often with the LINE probe, there may simply be more sequences homologous to LINEs in the human genome than in the genomes of *Reithrodontomys* or *Mus*. Alternatively, there may exist in the human genome a repetitive sequence that cross-hybridizes with some as yet unidentified sequence that is located near LINEs in the genome.

DNA–DNA hybridization of single-copy DNA has been exploited extensively for the construction of phylogenetic trees. That work attempts to remove all of the repetitive sequences and to use only single-copy DNA (Sibley and Ahlquist 1990). However, data resulting from DNA-DNA hybridization also has come under considerable criticism (Marks et al. 1989; Sarich et al. 1989). Although the methods used here focus on repetitive DNA, there is, at least from a broad perspective, some phylogenetic information available. Of the taxa used in this study, the greatest evidence of hybridization was demonstrated between *P. leucopus* and *Reithrodontomys,* which are phylogenetically the most closely related taxa. The least hybridization was observed between *P. leucopus* and human, which are the most distantly related taxa. However, the microsatellites that exist in mammalian genomes may confuse these results because blocks of simple nucleotide repeats such as $(GT)_n$ and $(CT)_n$ may obscure phylogenetic relationships.

P. leucopus is an evolutionarily successful species that occupies an extensive geographic range and a wide variety of habitats. Investigations into the organization of repetitive elements within the genome of the white-footed mouse will complement studies that are examining similar phenomena in laboratory species and humans. Studies of this type will permit comparison of the distributions, types, and frequencies of repetitive elements in a variety of mammalian taxa and ultimately broaden our knowledge of the structural and functional constraints affecting organization of the mammalian genome.

Acknowledgments. We thank A.D. Simmons and N.C. Brown for technical assistance and R.L. Ratliff for synthesizing the $(TTAGGG)_n$ and $(TCC)_n$ oligonucleotides. The 28S ribosomal and LINE L1Pm55 probes kindly were provided by N. Arnheim and D.H. Kass, respectively. We also thank R.A. Van Den Bussche and M.J. Hamilton for providing comments on an earlier version of this paper, and C.E. Hildebrand for helpful discussions. A portion of this work was supported by National Science Foundation grant DEB-9207597 to R.J. Baker.

References

Adkins, R.M., McBee, K., Baker, R.J., Porter, C.A. (1991). Chromosomal hybrid zone of *Peromyscus leucopus:* test of the recombinational breakdown model. J. Mammal. 72, 535–541.

Baker, R.J., Wichman, H.A. (1990). Retrotransposon *Mys* is concentrated on the sex chromosomes: implications for copy number containment. Evolution 44, 2083–2088.

Baker, R.J., Koop, B.F., Haiduk, M.W. (1983a). Resolving systematic relationships with G-bands: a study of five genera of South American cricetine rodents. Syst. Zool. 29, 239–253.

Baker, R.J., Robbins, L.W., Stangl, F.B., Jr., Birney, E.C. (1983b). Chromosomal evidence for a major subdivision in *Peromyscus leucopus*. J. Mammal. 64, 356–359.

Baker, R.J., Porter, C.A., Hanks, B.G., Bickham, J.W. (1991).

Increased variation in cellular DNA content at a hybrid zone: hybrid breakdown in *Peromyscus leucopus*. J. Hered. 82, 27–30.

Belmaaza, A., Wallenburg, J.C., Brouillette, S., Gusew, N., Chartrand, P. (1990). Genetic exchange between endogenous and exogenous LINE-1 repetitive elements in mouse cells. Nucleic Acids Res. 18, 6385–6391.

Burton, F.H., Loeb, D.D., Voliva, C.F., Martin, S.L., Edgell, M.H., Hutchison, C.A. III. (1986). Conservation throughout mammalia and extensive protein encoding capacity of the highly repeated DNA L1. J. Mol. Biol. 187, 291–304.

Carleton, M.D. (1980). Phylogenetic relationships in neotomine-peromyscine rodents (Muroidea) and a reappraisal of the dichotomy within New World Cricetinae. Misc. Publ. Mus. Zool. Univ. Michigan 157, 1–146.

Carleton, M.D. (1989). Systematics and evolution. In *Advances in the Study of Peromyscus* (*Rodentia*), G.C. Kirkland, Jr., J.N. Layne, eds. (Lubbock, Texas: Texas Tech University Press), pp. 7–141.

Crampton, J.M., Davies, K.E., Knapp, T.P. (1981). The occurrence of families of repetitive sequences in a library of cloned cDNA from human lymphocytes. Nucleic Acids Res. 9, 3821–3834.

Crew, M.D., Smith, G.S., Zeller, E.C., Walford, R.L. (1989). Polymorphism in the major histocompatibility complex class II genes of *Peromyscus leucopus*. Immunogenetics 30, 214–217.

Crew, M.D., Filipowsky, M.E., Zeller, E.C., Smith, G.S., Walford, R.L. (1990). Major histocompatibility complex class I genes of *Peromyscus leucopus*. Immunogenetics 32, 371–379.

Epplen, J.T., Ammer, H., Epplen, C., Kammerbauer, C., Mitreiter, R., Roewer, L., Schwaiger, W., Steimle, V., Zischler, H., Albert, E., Andreas, A., Beyermann, B., Meyer, W., Buitkamp, J., Nanda, I., Schmid, M., Nürnberg, P., Pena, S.D.J., Pöche, H., Sprecher, W., Schartl, M., Weising, K., Yassouridis, A. (1991). Oligonucleotide fingerprinting using simple repeat motifs: a convenient, ubiquitously applicable method to detect hypervariability for multiple purposes. In *DNA Fingerprinting: Approaches and Applications*, T. Burke, G. Dolf, A.J. Jeffries, and R. Wolff, eds. (Basel, Switzerland: Birkhäuser Verlag), pp. 50–69.

Evans, G.A., Lewis, K., Rothenberg, B.E. (1989). High efficiency vectors for cosmid microcloning and genomic analysis. Gene 79, 9–20.

Feinberg, A.P., Vogelstein, B. (1983). A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. 132, 6–13.

Gerbi, S.A. (1985). Evolution of ribosomal DNA. In *Molecular Evolutionary Genetics*, R.J. MacIntyre, ed. (Plenum Publishing Corp.), pp. 419–517.

Hamada, H., Petrino, M.G., Kakunaga, T. (1982). A novel repeated element with Z-DNA-forming potential is widely found in evolutionarily diverse eukaryotic genomes. Proc. Natl. Acad. Sci. USA 79, 6465–6469.

Hamilton, M.J., Hong, G., Wichman, H.A. (1992). Intragenomic movement and concerted evolution of satellite DNA in *Peromyscus:* evidence from in situ hybridization. Cytogenet. Cell Genet. 60, 40–44.

Kass, D.H., Berger, F.G., Dawson, W.D. (1992). The evolution of coexisting highly divergent Line-1 subfamilies within the rodent genus *Peromyscus*. J. Mol. Evol. 35, 472–485.

Kirkland, G.L., Jr., Layne, J.N. (eds.) (1989). *Advances in the Study of Peromyscus* (*Rodentia*). (Lubbock, Texas: Texas Tech University Press).

Korenberg, J.R., Rykowski, M.C. (1988). Human genome organization: Alu, Lines, and the molecular structure of metaphase chromosome bands. Cell 53, 391–400.

Longmire, J.L., Lewis, A.K., Brown, N.C., Buckingham, J.M., Clark, L.M., Jones, M.D., Meincke, L.J., Meyne, J., Ratliff, R.L., Ray, F.A., Wagner, R.P., Moyzis, R.K. (1988). Isolation and molecular characterization of a highly polymorphic centromeric tandem repeat in the family Falconidae. Genomics 2, 14–24.

Marks, J., Schmid, C.W., Sarich, V.M. (1989). DNA hybridization as a guide to phylogeny: relations of the Hominoidea. J. Hum. Evol. 17, 769–786.

McBee, K. (1991). Chromosomal aberrations in native small mammals (*Peromyscus leucopus*) at a petrochemical waste disposal site. II. Inherited aberrations and fragile sites indicated by G-band analysis. Environ. Toxicol. Chem. 10, 1321–1329.

McBee, K., Bickham, J.W., Brown, K.W., Donnelly, K.C. (1987). Chromosomal aberrations in native small mammals (*Peromyscus leucopus* and *Sigmodon hispidus*) at a petrochemical waste disposal site. I. Standard karyology. Arch. Environ. Contam. Toxicol. 16, 681–688.

Meyne, J., Baker, R.J., Hobart, H.H., Hsu, T.C., Ryder, O.A., Ward, O.G., Wiley, J.E., Wurster-Hill, D.H., Yates, T.L., Moyzis, R.K. (1990). Distribution of nontelomeric sites of the $(TTAGGG)_n$ telomeric sequence in vertebrate chromosomes. Chromosome 99, 3–10.

Moyzis, R.K., Buckingham, J.M., Cram, L.S., Dani, M., Deaven, L.L., Jones, M.D., Meyne, J., Ratliff, R.L., Wu, J.R. (1988). A highly conserved repetitive DNA sequence, $(TTAGGG)_n$ present at the telomeres of human chromosomes. Proc. Natl. Acad. Sci. USA 85, 6622–6626.

Moyzis, R.K., Torney, D.C., Meyne, J., Buckingham, J.M., Wu, J.-R., Burks, C., Sirotkin, K.M., Goad, W.B. (1989). The distribution of interspersed repetitive DNA sequences in the human genome. Genomics 4, 273–289.

Nelson, K., Baker, R.J., Honeycutt, R.L. (1985). Mitochondrial DNA and protein differentiation between hybridizing cytotypes of the white-footed mouse, *Peromyscus leucopus*. Evolution 41, 864–872.

Pardue, M.L., Lowenhaupt, K., Rich, A., Nordheim, A. (1987). $(dC\text{-}dA)_n \cdot (dG\text{-}dT)_n$ sequences have evolutionarily conserved chromosomal locations in *Drosophila* with implications for roles in chromosome structure and function. EMBO J. 6, 1781–1789.

Pine, D.S., Bourekas, E.C., Potter, S.S. (1988). *Mys* retrotransposons in *Peromyscus leucopus* and transgenic *Mus musculus*. Nucleic Acids Res. 16, 3359–3373.

Sambrook, J., Fritsch, E.F., Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual*. 2nd ed. (Plainview, NY: Cold Spring Harbor Laboratory Press).

Sarich, V.M., Schmid, C.W., Marks, J. (1989). DNA hybridization as a guide to phylogeny: a critical analysis. Cladistics 5, 3–32.

Sibley, C.G., Ahlquist, J.E. (1990). *Phylogeny and Classification of Birds. A Study in Molecular Evolution*. (New Haven, Conn.: Yale Univ. Press).

Singer, M.F. (1982). Highly repeated sequences in mammalian genomes. Int. Rev. Cytol. 76, 67–112.

Smith, G.S., Crew, M.D., Walford, R.L. (1989). *Peromyscus* as a gerontologic animal: aging and the MHC. In *Genetic Effects on Aging II*, D.E. Harrison, ed.). (Caldwell, N.J.: Telford Press), pp. 457–472.

Stallings, R.L. (1992). CpG suppression in vertebrate genomes does not account for the rarity of $(CpG)_n$ microsatellite repeats. Genomics 17, 890–891.

Stallings, R.L., Torney, D.C., Hildebrand, C.E., Longmire, J.L., Deavin, L.L., Jett, J.H., Doggett, N.A., Moyzis, R.K. (1990). Physical mapping of human chromosomes by repetitive sequence fingerprinting. Proc. Natl. Acad. Sci. USA 87, 6218–6222.

Stallings, R.L., Ford, A.F., Nelson, D., Torney, D.C., Hildebrand, C.E., Moyzis, R. (1991). Evolution and distribution of $(GT)_n$ repetitive sequences in mammalian genomes. Genomics 10, 807–815.

Stangl, F.B., Jr. (1986). Aspects of a contact zone between two chromosomal races of *Peromyscus leucopus* (Rodentia: Cricetidae). J. Mammal. 67, 465–473.

Wellauer, P.K., Dawid, I.B. (1979). Isolation and sequence organization of human ribosomal DNA. J. Mol. Biol. 128, 280–303.

Wichman, H.A., Potter, S.S., Pine, D.S. (1985). *Mys*, a family of mammalian transposable elements isolated by phylogenetic screening. Nature 317, 77–81.

Wichman, H.A., Payne, C.T., Reeder, T.W. (1990). Intragenomic variation in repetitive sequences isolated by phylogenetic screening of mammalian genomes. In *Molecular Evolution*, M.T. Clegg and S.J. O'Brien, eds., (New York: Alan R. Liss), pp. 153–160.

Wichman, H.A., Van Den Bussche, R.A., Hamilton, M.J., Baker, R.J. (1992). Transposable elements and the evolution of genome organization in mammals. Genetica 86, 287–293.