Volume 11 Number 20 1983 Nucleic Acids Research

Construction of a small *Mus musculus* repetitive DNA library: identification of a new satellite sequence in *Mus musculus*

D.F.Pietras[2], K.L.Bennett[3], L.D.Siracusa[1], M.Woodworth-Gutai, V.M.Chapman[1], K.W.Gross[1], C.Kane-Haas and N.D.Hastie[4]*

Departments of Cell and Tumor Biology and [1]Molecular Biology, Roswell Park Memorial Institute, Buffalo, NY 14203, USA

Received 8 August 1983; Accepted 23 September 1983

ABSTRACT

We report the construction of a small library of recombinant plasmids containing Mus musculus repetitive DNA inserts. The repetitive cloned fraction was derived from denatured genomic DNA by reassociation to a C_0t value at which repetitive, but not unique, sequences have reannealed followed by exhaustive S_1 nuclease treatment to degrade single stranded DNA. Initial characterizations of this library by colony filter hybridizations have led to the identification of a prevously undetected M. musculus minor satellite as well as to clones containing M. musculus major satellite sequences. This new satellite is repeated 10-20 times less than the major satellite in the M. musculus genome. It has a repeat length of 130 nucleotides compared wih the M. musculus major satellite with a repeat length of 234 nucleotides. Sequence analysis of the minor satellite has shown that it has a 29 base pair region with extensive homology to one of the major satellite repeating subunits. We also show by in situ hybridization that this minor satellite sequence is located at the centromeres and possibly the arms of at least half the M musculus chromosomes. Sequences related to the minor satellite have been found in the DNA of a related Mus species, Mus spretus, and may represent the major satellite of that species.

INTRODUCTION

The genomes of all eukaryotes consist of mixtures of single copy and repeated DNA sequences. Repeated DNA has traditionally been subdivided into two categories: satellite sequences and so called "middle repetitive" sequences. Satellite sequences are generally organized as long tandem arrays of simple sequences (reviewed in 1,2), whereas middle repetitive elements are generally interspersed with single copy sequences in the genome (reviewed in 1,3,4,5).

Britten and Davidson and colleagues (6-14) have made a systematic study of sea urchin repetitive DNA sequences. They cloned repetitive DNA sequences and studied the base sequence, organization, repetition frequency, transcriptional properties and evolution of selected cloned sequences. Several classes of repetitive DNA sequences have been

identified in the sea urchin genome. These families consist of either long or short elements whose members exhibit varying degrees of sequence divergence and repetition frequency. Hybridizing transcripts have been found for each class of sequences and there is evidence for tissue specific transcription. The potential role of these elements in gene regulation and the evolutionary significance of these elements has been discussed by these investigators.

We have constructed a small library of mouse repetitive DNA sequences to determine the abundance, organization and transcriptional properties of these elements in mammals. This paper describes the initial charaterization of this library and focuses upon the repetitive elements of mouse satellite DNA.

In subsequent papers, we will describe the properties of the highly repeated, dispersed sequences in the library (15). In this paper we describe the identification of a previously undetected, secondary satellite DNA sequence in the laboratory mouse, Mus musculus. We have characterised this satellite in terms of repeat length, partial DNA sequence, genetic divergence and chromosomal distribution.

MATERIAL AND METHODS

Animals

Mus musculus mice of the Swiss Ha/ICR strain, Mus spretus mice originally trapped in France and Spain (gifts of Drs. R. Sage and F. Bonhomme), and Mus caroli mice originally trapped in Thailand (gifts of Dr. J. Marshall) were obtained from randomly bred colonies maintained at Roswell Park Memorial Institute by Dr. V.. M. Chapman. M. musculus, M. spretus and M. caroli are separate Mus species. M. spretus is a field mouse, sympatric with M. musculus, which interbreeds with M. musculus only in the laboratory to produce fertile F_1 females and sterile F_1 males.

DNA isolation

High molecular weight genomic DNA was extracted from livers of adult M. musculus male mice of the Swiss Ha/ICR strain and from M. caroli and M. spretus mice as described previously (16). Sperm nuclear DNA was extracted from sperm isolated from the vas deferentia and epididymes of dult mice as described (17).

Isolation of mouse repetitive DNA sequences

The isolation of mouse repetitive DNA (mrDNA) sequences is described in Results. Hybroxyapatite (HAP) chromatography (18) was performed as

described by Campo and Bishop (19).

The S_1 nuclease treatment was performed in 0.2 mM EDTA, 0.2 M NaCL, 30 mM Na acetate, 0.6 mM $ZnSO_4$, 0.053 N acetic acid, pH 4.5 at a final DNA concentration of 40 µg/ml. S_1 nuclease (Sigma) was used at 50 units/ml for 30 minutes at 37°C. This was the same treatment routinely used to degrade more than 98% of TCA precipitable single stranded DNA into TCA soluble form, except that the reaction was performed in 0.2 M NaCl instead of the usual0.1 M NaCl to attempt to stabilize mismatched repetitive sequence duplexes.

Cloning of mrDNA sequences

One microgram of mrDNA Isolated as described in the text was tailed with oligo dC residues in a 100 µl reaction containing 0.2 mM [α-^{32}-P]-dCTP, 0.1 M Na cacodylate pH 6.9, 1 mM $CaCl_2$, 0.1 mM dithiothreitol and terminal deoxynucleotidyl transferase (BRL) for 20 minutes at 15°C (20). An average of 10 nucleotides was added per 3' end of the mrDNA sample. Since the mrDNA was small, it was decided to isolate a fraction of mrDNA enriched in larger molecules. The tailed mrDNA was phenol extracted and loaded on a continuous 5-30% sucrose gradient (in 1 M NaCl, 0.1% sarkosyl, 10 mM Tris, pH 7.5 and 1 mM EDTA) and centrifuged at 10°C at 31,000 rpm for 16.5 hours in a Beckman SW41 rotor. The tailed mrDNA (assayed by incorporated ^{32}P) produced a single sharp peak in the gradient. Several fractions from the heavy side of the peak, constituting approximately 10% of the incorporated radoactivity, were pooled. The tailed mrDNA was annealed to pBR322 which had previously been linearized at the PstI site (21) and tailed with oligo dG residues. Competent cells of *E. coli* strain X1776 were transformed with the annealed mixture at a ratio of approximately 1 mrDNA molecule/viable cell. The applicable National Institutes of Health guidelines for research involving recombinant DNA Molecules were followed in all procedures. Full details of these procedures may be found in (22).

Purified plasmid DNA was extracted from selected clones by CsCl density gradient centrifugation following chloramphenicol amplification of plasmids (23).

Restriction endonuclease digestions

Total liver DNA from adult *M. musculus*, *M. caroli* and *M. spretus* mice was digested with AluI, EcoRII, or HaeIII according to conditions recommended by the suppliers (Boehringer Mannheim). Usually, 1 µg of DNA was digested overnight at 37°C with 2-3 units of enzyme. Restricted DNA

from M. musculus, M. caroli and M. spretus was subjected to electrophoresis in 0.8% agarose gels using a 36 mM Tris, 30 mM $NaH_2PO_4 \cdot H_2O$, 1 mM EDTA, pH 7.4 buffer. HindIII restricted lamba DNA provided appropriate size markers.

Nick translation of DNA

Recombinant plasmid DNA and total genomic mouse DNA were labeled in vitro by nick translation (24). [α-^{32}P] labeled nucleotides (400 Ci/mol, Amersham) were used to generate probes with specific activities of 0.5-2 X 10^8 dpm/µg.

Filter bound nucleic acid hybridizaton

Colony filter hybridization experiments were performed by a slight modification of the original procedure of Grunstein and Hogness (25) as described (16). Transfer of DNA from agarose gels to nitrocellulose filters was performed according to Southern (26). Denatured plasmid DNA in solution was bound to 0.8 cm diameter nitrocellulose filters as described (27). Hybridisation conditions and washes were carried out as described previously (16).

$C_o t$ analysis

Radiolabeled probes to insert sequences were prepared by first nick-translating genomic DNA to high specific activity with [α-^{32}P]-dTTP and then hybridizing to filter bound plasmid DNA. After extensive washes, the probes were eluted from the filters in 0.5x SSC, precipitated with EtOH and redissolved in water. The Tm of the probe was determined during elution and the size of the probes was determined by agarose gel electrophoresis after methylmercuric hydroxide denaturation (28).

Driver DNA was prepared from high molecular weight DNA by sonication in a Branson sonifier with a macroprobe attachment. Sonication was performed at the maximum power output of the machine for 8 bursts of 15 seconds at a DNA concentration of 1 mg/ml with 45 second coolings on ice between bursts. Annealing reactions were performed at a low stringency in 1 M NaCl, 20 mM hepes buffer pH 7.0, and 1 mM EDTA at 60°C. At various times samples were taken for S1 nuclease analysis as described (27).

Nucleotide sequence determination of pMR150

pMR150 was digested with either NciI or AvaII and the insert-containing fragment was isolated by preparative gel electrophoresis. The AvaII fragment was cleaved with SfaNI, and the NciI fragment with HhaI. The Sfa NI-cleaved fragment was treated with bacterial alkaline phosphatase prior to 5'-end-labelling with γ-^{32}P ATP in the presence of polynucleotide

kinase (29) and secondarily digested with NciI. The HhaI-digested fragment was 3'-end labeled with α-^{32}P-cordycepin 5'-triphosphate (New England Nuclear) in the presence of terminal deoxynucleotidyl transferase and secondarily cleaved with AvaII. In addition, the NciI fragment was 5'-end-labeled and secondarily digested with AvaII. The nucleotide sequence was determined for both complementary strands of the insert using the chemical degradation procedure of Maxam and Gilbert (30).

In situ hybridization

In situ hybridizations were performed by methods described previously (16).

RESULTS

Construction of a mouse repetitive DNA library

The cloning of repetitive sequences is usually composed of two steps. The first is the isolation of DNA enriched for repetitive sequences and the second is the insertion of these sequences into a bacterial cloning vector and subsequent propagation of the hybrid molecules in a bacterial host. The scheme employed in this case is shown in Figure 1.

High molecular weight DNA, approximately 100 kilobase pairs (kbp) in length, was extracted from the livers of male mice of the Swiss Ha/ICR strain of M.musculus. If any reassociation of the isolated single strands of this DNA were attempted, large tangled networks of DNA would have formed because of the presence of numerous highly repeated DNA sequences on molecules of this length. Therefore, in order to reduce the DNA to a practical size for reassociation, the high molecular weight DNA was partially digested with the restriction enzyme HaeIII to generate DNA with an average single stranded size of approximately 5 kbp. This size was chosen in an attempt to clone long fragments of some large repeated sequences, such as transposable elements. This 5 kbp DNA was deproteinized by phenol, ethanol precipitated, and dissolved in water.

The DNA solution was heat denatured at 100°C, cooled to 70°C, made 0.18 M in buffered NaCl, and incubated at 70°C to a C_ot of 0.03 mole seconds liter^{-1}. This is approximately the C_ot 1/2 of a sequence repeated 50,000 times per haploid genome. This reaction allowed much of foldback and satellite DNA to renature to form duplex DNA molecules. This DNA mixture was passed through an hydroxyapatite column to fractionate it into a double stranded pool (20%) which contained reassociated satellite and foldback DNA, and a single stranded pool (80%) which contained single copy

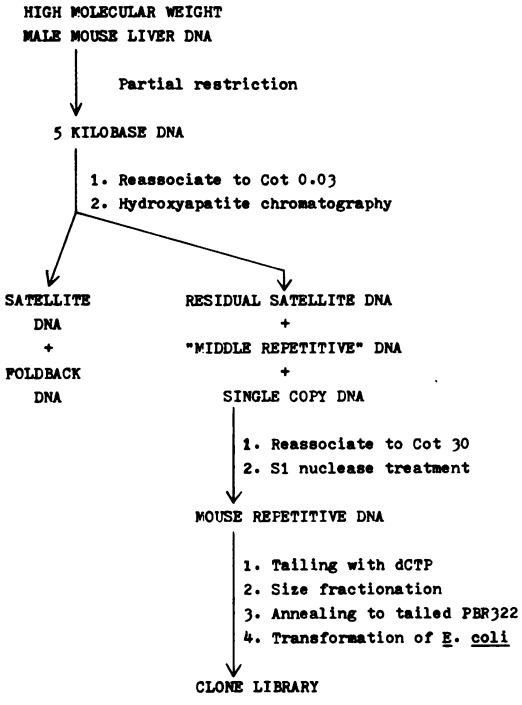

Figure 1. Outline of the scheme used for the construction of the M. musculus repetitive DNA library.

and middle repetitive DNA sequences. A major advantage of this scheme is that the mouse major satellite, repeated approximately 1 million times, is 95% reassociated at the end of this reaction. Therefore, only 50,000 copies of satellite sequence per genome equivalent of DNA will still be present in the single stranded pool. Thus, the concentration of satellite and the most highly repeated sequences relative to other sequences has been substantially reduced.

The single stranded DNA fraction was desalted, concentrated by ethanol precipitation, and redissolved in water. After thermal denaturation, the DNA was cooled to 70°C and made 1.0 M in buffered NaCl, and incubated to a C_ot of 30 mole seconds liter^{-1}. This C_ot is sufficient to allow approximately 50% renaturation of sequences repeated as little as 20-50 times per genome. The DNA was then treated with the single strand specific nuclease, S_1. This nuclease will degrade the single stranded single copy DNA , unrenatured repetitive DNA, and flanking sequences attached to the ends of short repetitive duplexes. It will also degrade any single

stranded DNA present within renatured molecules that contain mismatched segments. After nuclease digestion, the DNA was deproteinized and separated from S_1 digestion products by SP50-chelex chromatography.

At this stage, 4% of the original DNA reassociated to C_0t 30 mole seconds $liter^{-1}$ was recovered. An aliquot of this mouse repetitive DNA (mrDNA) was electrophoresed through a 5% polyacrylamide gel using HaeIII digested phage PM2 DNA as size markers (31). The mrDNA formed a broad smear with an average molecular weight of 100 bp (22). The mrDNA was tailed and a fraction enriched in larger molecules was isolated (see Materials and Methods). This mrDNA was cloned in pBR322 by the procedure of Villa-Komaroff et al. (21). The recombinant molecules were used to transform competent cells of E. coli strain X1776. 307 tetracycline resistant colonies were obtained by this procedure, of which approximately 125 contained mrDNA inserts, as judged by ampicillin sensitivity. Inserts contained in these recombinant plasmids were subsequently shown to have an average size of 150 bp (22).

Initial characterization of the library by colony filter hybridization

The most direct approach to identifying clones containing repetitive mouse DNA sequences is to use the colony filter hybridization (CFH) technique developed by Grunstein and Hogness (25). If nick-translated, mouse genomic DNA is used as a probe in such an assay, clones containing highly repeated mouse DNA inserts would be expected to give a positive signal. Also, the signal intensity of a particular clone might be expected to be a rough estimate of its repetition frequency in the mouse genome.

Figure 2 shows the results of several CFH experiments in which different probes were hybridized to replicate filters containing the mrDNA clones. Using male liver DNA as a probe (Figure 2A), approximately 50 of the 125 clones gave positive signals of varying intensity. Preliminary interpretations were that the colonies which produced the strong signals had inserts which corresponded to the highly repeated families, such as satellite and B_1 sequences (32), whereas the colonies which produced weak or no signals corresponded to families of middle and low repetition frequency, respectively.

The mrDNA library was constructed from liver DNA; one possibility was that some of the clones corresponded to families of relatively low repetition frequency in the germ line that had been amplified during somatic develop-ment. A test for amplification of sequences during somatic development can be performed using CFH under conditions where signal

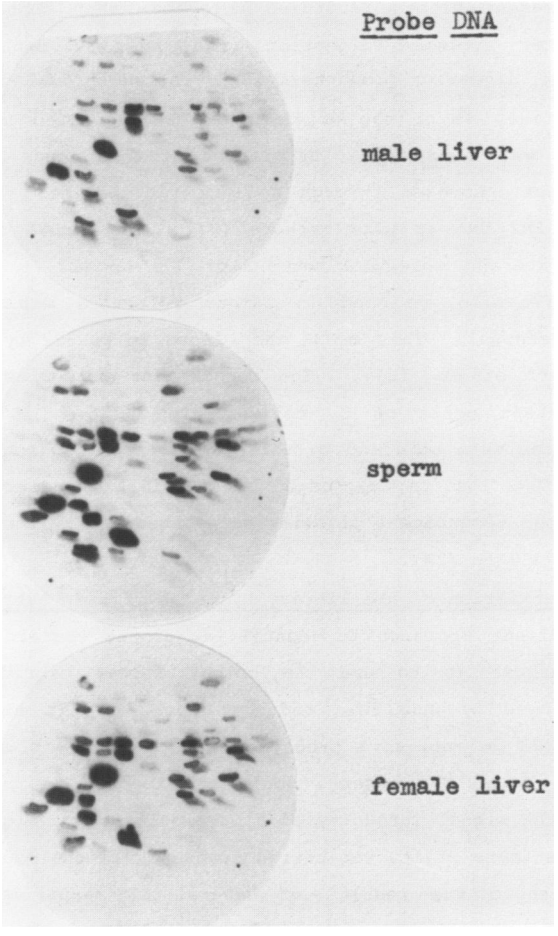

Figure 2. Grunstein-Hogness colony filter hybridizations using radiolabeled M. musculus A) male liver DNA, B) sperm DNA, and C) female liver DNA as probes. Autoradiographic exposure conditions were identical for all 3 filters.

intensity is proportional to repetition frequency. For this reason, sperm DNA was nick-translated and hybridized to the library (Figure 2B). Comparison of the autoradiographic signals obtained after such a hybridization of the sperm DNA with those previously observed with the liver DNA probe (Figures 2A & 2B) suggests that none of the more repeated families from the liver arose by somatic amplification.

The CFH assay was similarly used to determine whether any of the most repeated sequences were restricted to the Y chromosome. Female liver DNA

was nick-translated and hybridized to the filters (Figure 2C). All the clones that gave visible autoradiographic signals with male DNA (Figure 2A) gave signals of similar intensity when female DNA was used as probe, demonstrating that none of the highly repeated families in the mrDNA library are restricted to the Y chromosome. The few small differences between the 3 filters were probably due to varying colony size and loss of DNA from the filters during hybridization and washing.

Detection of satellite DNA clones by colony filter hybridization

Previous work demonstrated that satellite sequences are very divergent between mouse species (33,34). Clones containing satellite or highly divergent DNA inserts should be detected by probing the mrDNA library with M. caroli and M. spretus DNA. Colonies that gave positive signals with labeled M. musculus DNA, but reduced or no signals with M. caroli and/or M. spretus DNA might contain M. musculus satellite inserts. In preliminary experiments, a number of clones gave different signals when probed with DNA from different Mus species. These clones were restreaked on nitrocellulose filters and again probed with M.. musculus, M. spretus and M. caroli DNA. The results of this experiment are shown in Figure 3. Six clones (pMR 124, 150, 196, 238, 263, 286) showed strong autoradiographic signals when hybridized with radiolabeled M. musculus DNA. Five of these six clones showed reduced signals when hybridized with radiolabeled M. spretus and M. caroli DNA. One clone (pMR 150) gave an unexpected result; it produced a slightly stronger signal with M. spretus DNA than with M. musculus DNA, and hybridized poorly, if at all, with M. caroli DNA. Other clones gave more or less equal signal intensities with the 3 probes.

Southern blot analysis of clones pMR 196 and pMR 150

A Southern blot analysis of clones that exhibited differences between the Mus species was performed to determine whether these clones contained M. musculus satellite DNA inserts. M. musculus satellite DNA, isolated from CsCl density gradients, can be cut with specific restriction endonucleases to produce two types of restriction patterns, namely a type A and a type B pattern. A type A pattern (35) is characterized by all satellite sequences being cut into a regular series of low molecular weight fragments based on a monomer unit of 234 bp. A type B (36) pattern is produced when only a minor portion of total satellite sequences is cut into a series of integral multimers of 234 bp. M. musculus, M. spretus and M. caroli DNA's were restricted with enzymes known to generate either a type A (EcoRII) or type B (AluI) restriction pattern when hybridized with M.

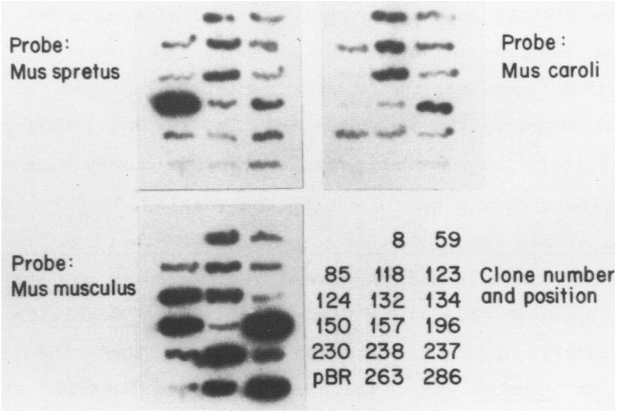

Figure 3. Species comparison of M. musculus repetitive DNA clones. Total liver DNA from M. musculus, M. caroli and M. spretus was nick-translated with (α-^{32}P)-dTTP and hybridized to filter bound plasmids as described in Materials and Methods. Approximately 3 x 10^7 dpm of each probe was added to each filter. The exposure time was 17 hours.

musculus satellite DNA (35-37). Clones of the first class (pMR 124,196,238, 263, 286) were suspected of being M. musculus satellite because a strong autoradiographic signal was produced by CFH only after hybridization to radiolabeled M. musculus DNA. A representative clone of the first class of sequences (pMR 196) was chosen to probe the restricted Mus DNA's (Figure 4). The results indicate that this clone identifies type A (lane 3) and type B (lane 4) restriction patterns, characteristic of M. musculus satellite DNA, when hybridized to M. musculus DNA. The restriction pattern seen with M. spretus (lanes 5,6) is similar to that with M. musculus, while M. caroli (lanes 1,2) exhibits numerous restriction site differences. The intensity of the autoradiographic signals of M. caroli and M. spretus are reduced compared to M. musculus (lanes 1,2,5,6 vs. lanes 3,4). After a five-fold longer exposure, M. caroli and M. spretus do not produce an autoradiographic signal equivalent in intensity to that of M. musculus, indicating that this sequence is less repeated or has diverged in M. caroli and M. spretus. These results are consistent with previous findings that the sequences of the major satellite have diverged considerably between M. musculus and M. caroli (33,34). By contrast, these sequences have been conserved between M. musculus and M.

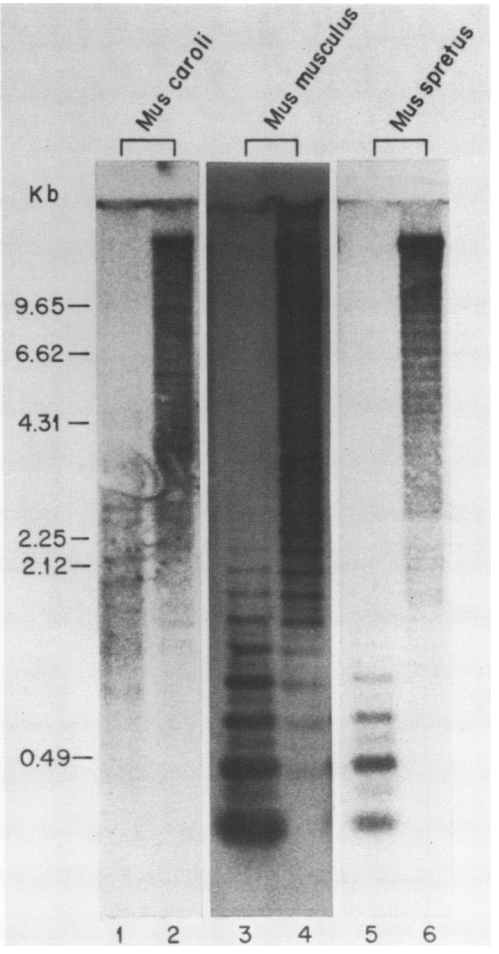

Figure 4. Southern blot analysis of M. caroli, M. musculus and M. spretus DNA probed with radiolabeled clone pMR 196 (1.5 x 10^7 dpm applied; Sp. Act. ~0.5 x 10^8 dpm/µg). Each lane contains 2 µg of restricted DNA. Lanes 1) 3) 5) are EcoRII restricted M. caroli, M. musculus and M. spretus DNA respectively. Lanes 2) 4) 6) are AluI restricted M. caroli, M. musculus and M. spretus DNA respectively. Lanes 3) 4) were exposed for 1 day and lanes 1) 2) 5) 6) were exposed 5 days to help visualize restriction patterns.

spretus, but are greatly reduced in M. spretus (38). Clones pMR 124, 238, 263, and 286 contain sequences homologous to the insert in pMR 196 as demonstrated both by CFH, using pMR 196 as probe, and by Southern blot analysis to M. musculus DNA (data not shown).

Striking differences were observed between Mus species in the

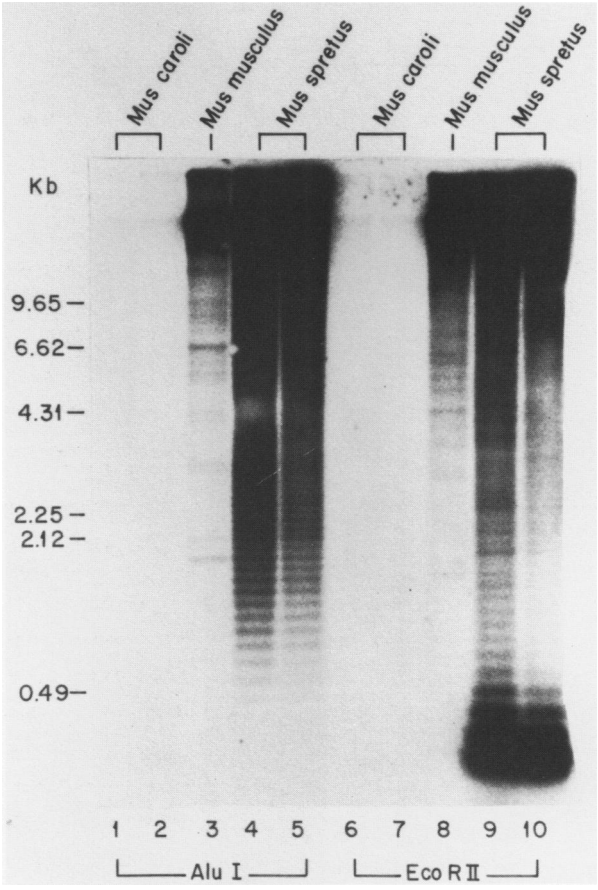

Figure 5. Southern blot analysis of M. caroli, M. musculus and M. spretus DNA probed with radiolabeled clone pMR 150 (2.3 x 10^7 dpm applied; Sp. Act. ~0.9 x 10^8 dpm/µg). Lanes 1) 2) are AluI restricted M. caroli DNA (2 µg). Each lane represents DNA from a different M. caroli mouse. Lane 3) is AluI restricted M. musculus DNA (2µg). Lanes 4) 5) are AluI restricted M. spretus DNA (2 µg and 1 µg respectively). These lanes represent DNA from two different M. spretus mice. Lanes 6) 7) are EcoRII restricted M. caroli DNA (4 µg). The lanes represent two different M. caroli mice. Lane 8) is EcoRII restricted M. musculus DNA (4 µg). Lanes 9) 10) are EcoRII restricted M. spretus DNA (4 µg respectively). These lanes represent DNA from two different M. spretus mice. Autoradiographic exposure time was 5 days.

restriction patterns identified by clone pMR 150 (Figure 5). First, M. caroli contains few, if any, sequences which are homologous to the insert of pMR 150 (lanes 1,2,6,7). Second, the intensity of the autoradiographic signals is greater with M. spretus DNA (lanes 4,5,9,10) than with M. musculus DNA (lanes 3,8) indicating that pMR 150 is more repeated in the M.

spretus genome. Both these results are consistent with the CFH (Figure 3). In addition, the restriction pattern of M. spretus DNA with either AluI or EcoRII reveals a series of multimers based on a monomer length of 100-140 bp. This restriction pattern is distinctly different from the 234 bp repeat revealed by the M. musculus major satellite when either EcoRII or AluI are used (Figure 4), and indicates a tandem arrangement of at least a subset of these sequences in the M. spretus genome. This restriction pattern is similar to a type B satellite pattern in which only a fraction of the total sequences are cut by a single restriction enzyme. The pMR 150 sequence is at least 5-fold more repeated in M. spretus than M. musculus, based on varying autoradiographic exposure times to produce signals of equivalent intensity (data not shown). These results indicate that the pMR 150 insert identifies a highly repeated sequence of M. spretus.

Partial Restriction Enzyme Analysis Indicates that the pMR 150 Sequence is In Tandem Arrays in Genomic DNA

In order to determine whether the pMR 150 sequence is in large tandem arrays in the genome, as suspected, mouse DNA was digested for various lengths of time with the enzyme MspI to obtain a series of partial and complete digestion products. Following gel electrophoresis and Southern transfer the DNA was hybridized with ^{32}P labelled pMR 150. Most of the pMR 150 material in the complete digest is detected as a series of bands up to a size of about 800 nucleotides (Figure 6). These bands increase in size by increments of approximately 130 nucleotides. Faint bands above this size range can also be detected. Analysis of the partial digestion products, particularly after one minute and 5 minutes digestion, reveals a clear series of bands increasing by increments of 130 nucleotides up to a size range of about 4 kb. At these early digestion times there is much more DNA in the large multimeric units than in the oligomeric units seen mainly in the complete digestion products. This data suggests that the pMR 150 sequence is organised in large tandem arrays of the 130 nucleotide repeat.

Repetition frequency of the pMR 150 satellite family

To determine the approximate repetition frequency of the M. musculus minor satellite, a denatured probe homologous to pMR 150 was annealed in solution to a vast excess of genomic M. musculus DNA as described in Materials and Methods. For comparison, a probe complementary to the major satellite clone pMR 196 was also annealed to M. musculus DNA (Figure 7). The major satellite probe, as expected, annealed with a $C_o t$ 1/2 of 0.00032

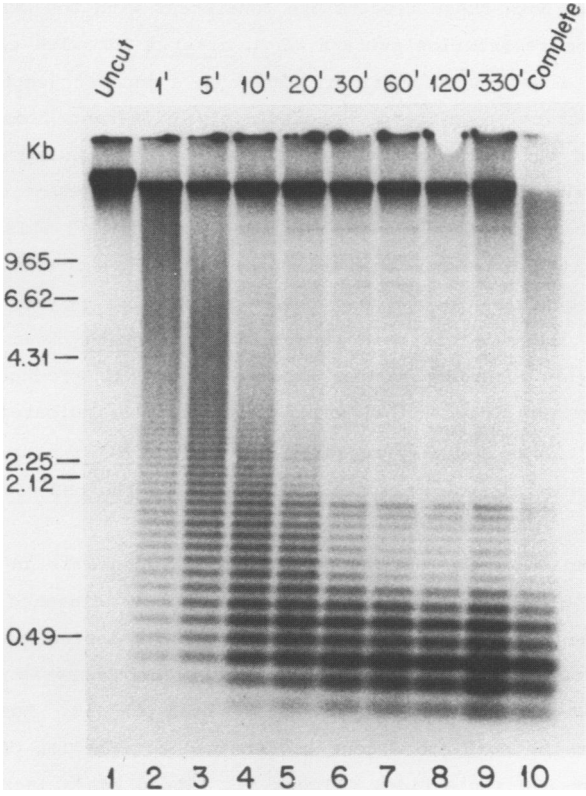

Figure 6 Partial restriction enzyme analysis of M. musculus DNA reveals the tandem nature of the pMR 150 satellite sequence.
M. musculus DNA was digested with $MspI$. At the times indicated 1 µg aliquots were removed from the reaction, electrophoresed on an agarose gel, transferred to nitrocellulose and hybridized to radiolabeled pMR 150.

mole seconds liter^{-1}, which is 700,000 times faster than single copy DNA. The $C_o t$ 1/2 of the pMR 150 probe suggests a repetition frequency of at least 50,000 times per haploid genome in M. musculus.

Sequence Analysis of the pMR 150 Insert Reveals a Region of Homology with the Major Mus musculus Satellite

Figure 8 shows the sequence of the pMR 150 insert determined as described in materials and methods. The insert is 94 nucleotides long and so does not represent a complete copy of the 130 nucleotide repeat. However a number of conclusions can be drawn from the data. Firstly, like the major satellite, the pMR 150 sequence contains 66% AT residues. Secondly, there is a stretch of 23 out of 28 nucleotides in the pMR 150

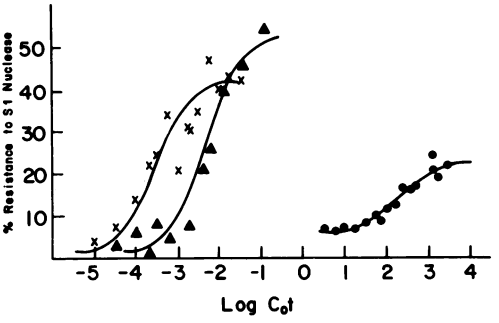

Figure 7 C_0t analysis to determine approximate repetition frequency of the minor M. musculus satellite. Probes corresponding to pMR 196 (X), pMR 150 (Δ) and mouse single copy DNA (o) were annealed in the presence of a vast excess of genomic mouse DNA as described in Materials and Methods.

sequence which is homologous to a region in the Mus musculus major satellite (37). This stretch in the major satellite represents one out of 8 of the major repeating subunits. The remaining 2/3 of the pMR 150 sequence shares no obvious homology with the major satellite. However, Horz and Altenburger (37) postulated that the prototype major satellite, 27 nucleotides in length, contains the sequence TGAAAAA. This stretch is present at positions 3-9 in the pMR 150 sequence. Also the sequence CAATGA occurs twice within 16 nucleotides in the pMR 150 sequence. The sequence AATGA occurs 4 times within the major satellite. There is no obvious internal periodicity within the pMR 150 sequence, unlike the case for the major satellite. Also the pMR 150 sequence contains no significant region of homology with the highly repeated dispersed mouse sequences B1 (32), B2 (39), and R (40).

M. musculus major satellite
sequence nucleotides 205-231.

```
5'-Gn A C T G A A A A A C A C A T T C G T T G G A A A C G G G A T T T

  G T A G A A C A G T G T A T A T C A A T G A G T T A C A A T G A

    G A A A C A T G G A A A A T G A T G A A A A C C A C A C T Cn-3'
      * * * *   * *   * * * * * * *   * * * *     * * * * * *
      A A A C G T G A A A A A T G A   G A A A T G C A C A C T
```

Figure 8 Nucleotide Sequence of the Cloned Plasmid pMR150. The first 3 lines show the sequence of the pMR150 insert. The fourth line shows the region of homology with the major satellite taken from Horz and Altenburger (54).

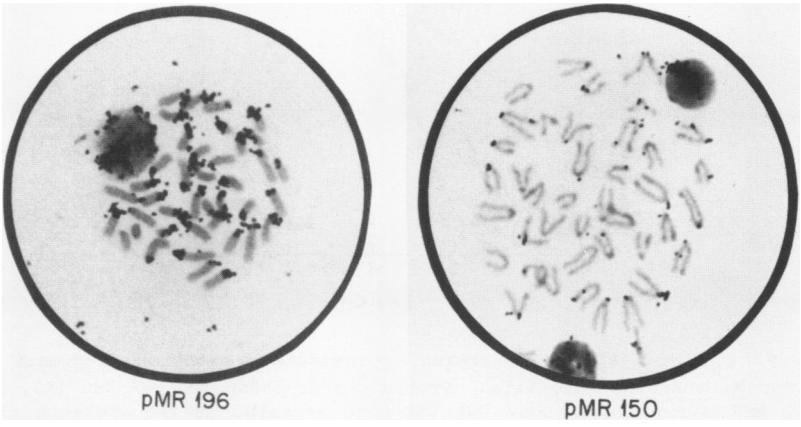

Figure 9 In situ hybridizations to bone marrow preparations from M. musculus male mice probed with ^{3}H-labeled pMR 196 and pMR 150. Both slides were coated with Kodak NTB-2 emulsion and developed after an 8 day (pMR 196) or 60 day (pMR 150) exposure.

In Situ Hybridization to Metaphase Chromosomes Indicates that the pMR 150 Satellite is Located Centromerically on at least Half of the Mouse Chromosomes

Figure 9 shows the results of in situ hybridization of the pMR 150 minor satellite and pMR 196 major satellite probes to metaphase chromosome preparations of Mus musculus. As expected the major satellite hybridized efficiently to centromeres of all chromosomes except the Y chromosome (41.42). The minor satellite produced a much weaker signal. However it is clear that on 16 out of 40 chromosomes there are grains located at the centromere. In some cases there are several centromeric grains on a single chromosome. From analysis of several metaphases we conclude that at least 20 out of 40 chromosomes have centromerically located pMR 150 sequences. Following hybridization to ^{3}H labelled PBR322 of the same specific activity and exposed to emulsion for the same period there are only two to three grain on average per metaphase plate (data not shown). Consistently for the pMR 150 sequence 20-30% of the total grains are observed on chromosome arms. This appears to be real hybridization as fewer grains are observed on chromosome arms when the major satellite probes are used even though the total number of grains is in great excess when the major satellite probe is used. Further studies using more sensitive techniques will be required to give a complete understanding of the chromosomal distribution of this minor satellite.

DISCUSSION

We have described the construction of a small library of recombinant plasmids which contain inserts corresponding to families of repetitive DNA in the mouse genome. An important advantage of this construction was theinitial reduction, in percentage, of satellite and the most highly repeated sequences from the main fraction of repetitive DNA. This resulted in the majority of clones possessing middle repetitive DNA inserts. Also, the S_1 treatment prior to cloning removed any single copy sequences which flanked the repetitive duplexes. The approach taken should ensure that each insert represents just one repetitive DNA family. Hence, these clones will be useful to study the structure and expression of individual repetitive DNA families.

In this initial characterization of the library, we have identified clones which are homologous to satellite families. The differences in autoradiographic signal intensities seen when the major satellite clone (pMR 196) was hybridized to *M. musculus* and *M. caroli* DNA (Figure 4) can be used to distinguish *M. musculus* chromosomes and interphase nuclei from those of *M. caroli* in *in situ* hybridizations (16). *In situ* hybridization of clone pMR 196 to histological sections of *M. musculus* ↔ *M. caroli* chimeras has provided a useful tool for distinguishing parental cell types in these chimeras (43). This type of cell marker system is ubiquitous and can be used to follow cell lineages during mouse embryogenesis. Thus, highly repeated satellite sequences which have diverged between species can provide useful tools to study embryogenesis as well as to determine the nature of these repetitive elements in mammalian genomes.

The newly described minor satellite shares properties with the major satellite but is obviously different. As shown in Figure 8 the 2 satellites share a region of homology and are both AT rich. Also both are located at the centromeric region. However, the majority of the minor satellite DNA sequence is different from the major satellite and it's repeating unit is 130 nucleotides compared with 234 nucleotides for the major satellite. Also they give different patterns when cut with different restriction enzymes. For example in *M. musculus* EcoRII sites are found in quite a high proportion of major satellite repeats but infrequently in minor satellite repeats (Figure 4). MspI sites however are often found in minor satellite members but infrequently within major satellite repeats (Figure 6 and Sanford and Hastie unpublished observations). Finally it appears that the minor satellite sequences may be more highly represented

on chromosome arms than the major satellite (Figure 9).

This library will also be useful for defining the properties of a number of dispersed repetitive sequence families in the mouse genome. We have characterized the structure and expression of the five most highly repeated dispersed mouse families which are each represented by several members in our library (15). Clones corresponding to a random sample of sequences of lower repetition frequency are also present in the library and we intend to study the properties of a number of these clones.

ACKNOWLEDGEMENTS

We wish to thank Mary Bodis, Fred Gaskin, Linda Gremke, Pauline Labrozzi and Sharon Sokoloswki for excellent technical assistance and Dr A.M. Buchberg for valuable scientific discussions. We thank Drs. W. A. Held, R. G. Hughes, Jr., and T. B. Shows for the generous use of laboratory equipment. We thank Cynthia Bell, Marcia Held and Katie Rae for typing the manuscript. This work was supported by grants from N.I.H. to Drs. V. M. Chapman (GM24125), K.W. Gross (GM19521, GM30248), and N. D. Hastie (GM24771).

*To whom correspondence should be addressed

[2]Current address: Department of Radiation Biology and Biophysics, University of Rochester School of Medicine and Dentistry, Rochester, NY, USA

[3]Current address: Carnegie Institution of Washington, Department of Embryology, 115 West University Parkway, Baltimore, MD 21210, USA

[4]Current address: MRC Clinical and Population Cytogenetics Unit, Western General Hospital, Crewe Road, Edinburgh EH4 2XU, UK

REFERENCES

1. Lewin, B. (1974) In Gene Expression-2, John Wiley & Sons Ltd., Great Britain.
2. Brutlag, D. L. (1980) Ann. Rev. Genet. 14, 121-144.
3. Davidson, E. H., Galau, G. A., Angerer, R. C. & Britten, R. J. (1975) Chromosoma 51, 253-259.
4. Singer, M. F. (1982) Cell 28, 433-434.
5. Singer, M. F. (1982) Intl. Rev. of Cytology 76, 67-112.
6. Scheller, R. H., Thomas, T. L., Lee, A. S., Klein, W. H., Niles, W. D., Britten, R. J. & Davidson, E. H. (1977) Science 196, 197-200.
7. Klein, W. H., Thomas, T. L., Lai, C., Scheller, R. H., Britten, R. J. & Davidson, E. H. (1978) Cell 14, 889-900.
8. Costantini, F. D., Scheller, R. H., Britten, R. J. & Davidson, E. H. (1978) Cell 15, 173-187.
9. Scheller, R. H., Costantini, F. D., Kozlowski, M. R., Britten, R. J. &

Davidson, E. H. (1978) Cell 15, 189-203.
10. Moore, G. P., Scheller, R. H., Davidson, E. H. & Britten, R. J. (1978) Cell 15, 649-660.
11. Moore, G. P., Costantini, F. D., Posakony, J. W., Britten, R. J. & Davidson, E. H. (1980) Science 208, 1046-1048.
12. Anderson, D. M., Scheller, , R. H., Posakony, J. W., McAllister, L. B., Trabert, S. G., Beall, C., Britten, R. J. & Davidson, E. H. (1981). J. Mol. Biol. 145, 5-28.
13. Scheller, R. H., Anderson, D. M., Posakony, J. W., McAllister, L. B., Britten, R. J. & Davidson, E. H. (1981) J. Mol. Biol. 149, 15-39.
14. Posakony, J. W., Scheller, R. H., Anderson, D. M., Britten, R. J. & Davidson, E. H. (1981) J. Mol. Biol. 149, 41-67.
15. Bennett, K. L., Hill, R. E., Pietras, D. F., Woodworth-Gutai, M., Kane-Haas, C. & Hastie, N. D. (1983) manuscript submitted.
16. Siracusa, L. D., Chapman, V. M., Bennett, K. L., Hastie, N. D., Pietras, D. F. & Rossant, J. (1983) J. Embryol. exp. Morph. 73, 163-178.
17. Shiurba, R. & Nandi, S. (1979) Proc. Natl. Acad. Sci. U.S.A. 76, 3947-3951.
18. Miyazawa, Y. & Thomas, C. A., Jr. (1965) J. Mol. Biol 11, 223-237.
19. Campo, M. S. & Bishop, J. O. (1974) J. Mol. Biol. 90, 649-663.
20. Roychoudhury, R., Jay, E. & Wu, R. (1976) Nuc. Acids Res. 3, 101-116.
21. Villa-Komaroff, L., Efstradiatis, A., Broome, S., Lomedico, P., Tizard, R., Naber, S. P., Chick, W. L. & Gilbert, W. (1978) Proc. Natl. Acad. Sci. U.S.A. 75, 3727-3731.
22. Pietras, D. F. (1981) Ph.D. Theis, State University of New York at Buffalo.
23. Clewell, D. B. & Helinski, D. R (1969) Proc. Natl. Acad. Sci. U.S.A. 62, 1159-1166.
24. Rigby, P., Dieckman, M., Rhodes, C. & Berg, P. (1977) J. Mol. Biol. 113, 237-251.
25. Grunstein, M. G. & Hogness, D. S. (1975) Proc. Natl. Acad. Sci. U.S.A. 72, 3961-3965.
26. Southern, E. M. (1975) J. Mol. Biol 98, 503-517.
27. Barth, R. K., Gross, K. W., Gremke, L. C. & Hastie, N. D. (1981) Proc. Natl. Acad. Sci. U.S.A. 79, 500-504.
28. Bailey, J. M. & Davidson, N. (1976) Analyt. Biochem. 70, 75-85.
29. Woodworth-Gutai, M. (1981) Virology 109, 344-352.
30. Maxam, A. M. and Gilbert, W. (1980) Meth. Enzymol. 65, 499-560.
31. Kovacic, R. T. & Van Holde, K. E. (1977) Biochemistry 16, 1490-1498.
32. Kramerov, D. A., Grigoryan, A. A., Ryskov, A. P. & Georgiev, G. P. (1979) Nuc. Acids Res. 6, 697-713.
33. Sutton, W. D., McCallum, M. (1972) J. Mol. Biol. 71, 633-656.
34. Rice, N. R. & Straus, N. A. (1973) Proc. Natl. Acad. Sci. U.S.A. 70, 3546-3550.
35. Southern, E. M. (1975) J. Mol. Biol. 94, 51-69.
36. Horz, W. & Zachau, H. G. (1977) Eur. J. Biochem. 73, 383-392.
37. Horz, W. & Altenburger, W. (1981) Nuc. Acids Res. 9, 683-696.
38. Brown, S. D. M. & Dover, G. A. (1980) Nature 285, 47-49.
39. Krayev, A. S., Markusheva, T. V., Kramerov, D. A., Ryskov, A. P., Skryabin, K. G., Bayev, A. A. and Georgiev, G. P. (1982) Nucleic Acids Res. 10, 7461-7475.
40. Gebhard, W., Meitinger, T., Hochtl, J. and Zachau, H. G. (1982) J. Mol. Biol. 157, 453-471.
41. Jones, K. W. (1970) Nature 225, 912-915.
42. Pardue, M. L. & Gall, J. G. (1970) Science 168, 1356-1358.
43. Rossant, J., Vijh, M., Siracusa, L. D. & Chapman, V. M. (1983) J. Embyryol. exp. Morph. 73, 179-191.