Towards construction of a high resolution map of the mouse genome using PCR-analysed microsatellites

Jennifer M.Love, Andrew M.Knight[1], Marcia A.McAleer and John A.Todd
Nuffield Department of Surgery, John Radcliffe Hospital, Headington, Oxford OX3 9DU and
[1]Transplantation Biology Section, Clinical Research Centre, Watford Road, Harrow, Middlesex HA1 3UJ, UK

Received April 26, 1990; Revised and Accepted May 31, 1990

ABSTRACT

Fifty sequences from the mouse genome database containing simple sequence repeats or microsatellites have been analysed for size variation using the polymerase chain reaction and gel electrophoresis. 88% of the sequences, most of which contain the dinucleotide repeat, CA/GT, showed size variations between different inbred strains of mice and the wild mouse, *Mus spretus*. 62% of sequences had 3 or more alleles. GA/CT and AT/TA -containing sequences were also variable. About half of these size variants were detectable by agarose gel electrophoresis. This simple approach is extremely useful in linkage and genome mapping studies and will facilitate construction of high resolution maps of both the mouse and human genomes.

INTRODUCTION

The construction of a high resolution map of the mouse and human genomes with markers at every 1cM will require the isolation of a large number of polymorphic DNA markers. Most DNA polymorphisms or variations have been detected using restriction enzymes, which detect sequence variation at restriction enzyme sites. These, by their nature, are rare and their detection can require the use of 20–30 different enzymes. In addition, blotting and hybridization protocols are time consuming.

As tools for genetic mapping, markers with more than two alleles are considerably more useful than standard, diallelic restriction fragment length polymorphisms (RFLP)[1]. Variable numbers of tandemly repeated (VNTR) DNA polymorphisms or minisatellites are extremely polymorphic and are widely distributed throughout human[2,3] and mouse[4,5] genomes. They can be analysed by conventional RFLP analysis and also by using the polymerase chain reaction (PCR)[6–8]. Because the repeat unit can be up to 50bp and the number of repeats large, the utility of PCR as a way of analysing VNTR length may be limited[6]. More recently, several reports have described length polymorphisms of DNA containing dinucleotide repeats, in particular $(CA)_n$ and $(GA)_n$ in human[9–11] and whale[9] genomes, referred to as microsatellites[10]. The number of dinucleotide repeats appears to be in the range of about 15–40, and therefore the length of DNA to be amplified by PCR is well within the limits of the Taq-Polymerase. $(CA)_n$ repeat-containing sequences are much more common than VNTR sequences. In the mouse and human genomes there are at least 5×10^4 sequences that contain $(CA)_n$ repeats[12]. In the human genome, they are apparently randomly distributed, and some of the loci that have been analysed are highly polymorphic with 6 or more alleles [9,10,13].

The aim of the present study is to analyse the feasibility of using microsatellites to study the mouse genome. The EMBL and Genbank databases were screened for $(CA)_{10}$, $(GT)_{10}$, $(GA)_{10}$ and $(CT)_{10}$ sequences. Fifty microsatellites were analysed by PCR. Of those analysed, 78% (39/50) are variable between inbred strains of mice and 92% (37/40) are variable between *M. spretus* and inbred strains. These results indicate that analysis of microsatellites will be a useful method in the construction of both human and mouse genetic maps.

MATERIALS and METHODS

DNA samples

DNA was prepared from liver using the sodium perchlorate method[14]. DNAs from DBA/2J, C57BL/6J, AKR/J, BALB/cBy, C57BL/10J, SWR/J, SJL/J, C3H/HeJ and CBA/J and from the 26 recombinant inbred (RI) strains of the BXD (C57BL/6JXDBA/2J) panel were purchased from the Jackson Laboratories[15]. DNA from *M. spretus* (abbreviated to SPE in Table 1) was a gift from Dr S. Brown. Livers from NOD (nonobese diabetic), NON (nonobese non diabetic), B6.PL and B10.H-2^{nod} (abbreviated to B10/W) were from Drs L.Wicker and L.Peterson (Merck, Sharp and Dohme, USA). B10.H-2^{nod} (B10/W) is an *H-2* congenic strain in which the NOD *H-2* has been transferred onto C57BL/10J.

Primers

Oligonucleotide primers, mostly 21 nucleotides long, were synthesized using an Applied Biosystems 380B DNA synthesizer. Deprotected primers were dried, dissolved in TE buffer (10mM Tris, 0.1mM EDTA, pH 7.5), extracted with phenol/chloroform (1:1), then ethanol-precipitated from TE buffer, once with ammonium acetate and twice with sodium acetate.

Polymerase chain reaction

PCR primers were designed to unique sequences flanking the repeat (Table1). The efficiency and specificity of each set of primers (a forward and a reverse primer) were tested by first optimising the Mg^{2+} concentration by titrating from 1–5mM in the standard reaction buffer, recommended for Taq-Polymerase (Amplitaq) by Perkin-Elmer/Cetus[16]. Gelatin was not added to this buffer. Reaction volumes were 25μl and samples were overlaid with light mineral oil (Sigma). 25–32 cycles of the PCR were carried out under standard conditions using a Perkin-Elmer/Cetus Thermocycler or Lep Prem III: 1min at 94°C, 1min at 55°C and 30s at 72°C, followed by a final incubation at 72°C for 10min. If gel analysis of the PCR products did not give a single band on a 4% agarose gel (3% NuSieve agarose/1% agarose) of approximately the same size as the expected product (taken from the database sequence), then the PCR was repeated either at higher annealing temperature (60°C or 65°C) or with an altered number of cycles (either increased to 45 cycles or decreased to 25 cycles). If a satisfactory profile was not obtained after these PCRs, a second forward and/or reverse primer was designed and tested for specificity by Mg^{2+} titration with one of the original primers. Most primer pairs worked well without the need to synthesize additional primers. In some cases, altering the amount of genomic DNA and/or the concentration of primers helped obtain a single major band after gel electrophoresis.

Assay of size variation between alleles

To date, microsatellites have been analysed by electrophoresis of radiolabelled PCR product on denaturing acrylamide sequencing gels[9–11]. The microsatellites identified in this report have been analysed, without radiolabelling during the PCR, by electrophoresis through standard, horizontal, non-denaturing agarose gels (Uniscience 7cm mini-gel apparatus). The gels consisted of 3% NuSieve agarose (FMC Corporation)/1% agarose (BRL) in TBE[17] containing 0.5μg/ml ethidium bromide. If the PCR products from different strains showed no detectable size variations, they were analysed further by electrophoresis through a vertical, 16cm or 20cm long, 0.75mm thick non-denaturing acrylamide gel in TBE buffer (using the Bio-Rad Protean II apparatus) and bands detected by ethidium bromide staining, or, if the amount of PCR product is low, by silver staining. PCR products of 75–320bp were resolved on 6–10% acrylamide gels. To test the resolving power of a vertical, 10cm long, 0.75mm thick, non-denaturing acrylamide gel, primers were designed to the *Plau* sequence (sequence 29.MMUPAA) that would produce PCR products that were 2, 4, 6 and 8bp smaller than the product size produced by the primers shown for sequence 29 in Table 1. The original two primers given in Table 1 are called 1-A (forward) and 2-A (reverse). The sequences of the additional primers are: forward primer1-B, 5′-CTGGCTAGGAATAAACAGAAA-3′ (with reverse primer 2-A, in Table 1, gives a product 2bp smaller than standard product size); reverse primer 2-B, 5′-GGAATTCATGTTCAGG-ATAAA-3′ (with primer 1-A gives a product 2bp smaller); reverse primer 2-C, 5′-AATTCATGTTCAGGATAAACA-3′ (with primer 1-A gives a product 4bp smaller); reverse primer 2-D, 5′-TTCATGTTCAGGATAAACAGG-3′ (with primer 1-A gives a product 6bp smaller); and a combination of primers 1-B and 2-D gives a product 8bp smaller than the standard product. Electrophoresis of 1–4μl of the 25μl PCR is sufficient to detect length differences of 2bp or more using this system.

Linkage analysis in RI strains

Strain distribution patterns (SDP) of microsatellite alleles were analysed in the 26 strains of the BXD RI panel[15]. Values for the recombination frequency (r) were calculated using the formula: $r = R/4-6R$, where R = number of discordant strains/number of total strains. Upper and lower 95% confidence limits for r were taken from Silver[18]. For estimating map distances between X-linked genes from RI data, the following equation was used, $r = 3R/(8-12R)$ (B.A. Taylor, personal communcation). The analysis was facilitated using the RIS (version 1.4) programme from Dr K. Manly (Roswell Park Cancer Institute, Elm and Carlton Streets, Buffalo, New York 14263). Generous help from Dr B. Taylor (Jackson Laboratory, Bar Harbor, Maine 04609) is also gratefully acknowledged.

RESULTS

Variability of mouse microsatellites

Table 1 gives the names of the sequences analysed according to the databases, the chromosomal location if known[19], the forward and reverse primer sequences, the expected product size, the sequence of the repeat and the relative sizes of alleles.

Table 2 gives the annealing temperatures and Mg^{2+} concentrations established by titration of each set of primers. Mg^{2+} concentrations and annealing temperatures are dependent on several factors that influence PCR including the type of PCR machine, the enzyme supplier and other components of the reaction buffer.

44/50 (88%) sequences were variable in size between any of the strains tested. 10/50 (20%) primer pairs did not produce a single, reproducible band on agarose using *M. spretus* DNA. *M. spretus* displays considerable sequence divergence from inbred laboratory strains of *Mus musculus domesticus* and therefore base mismatching at the sites of primer annealing may account for the failure to observe distinct products. Consistent with the latter suggestion, a lower annealing temperature did produce specific PCR products using *M. spretus* DNA (see below). Of the 40 sequences that were analysed successfully using *M. spretus* DNA, 37 (92%) had different sizes from any one of the inbred strain alleles and 25/40 (62%) were unique to *M. spretus*.

Between the inbred strains of mice tested including NOD and NON, 39/50 (78%) microsatellites were variable in size. Between two specific inbred strains, B10/W (B10.H-2^{nod}) and NOD, 20/45 (44%) sequences tested (excluding the four chromosome 17 MHC-linked sequences) were variable. Between the strains DBA/2J and B6/J, which are the parental strains of the BXD RI strains, 24/41 (58%) were variable in size and between B6/J and *M. spretus* 24/34 (70%) were variable. Microsatellite variation between B6/J and *M. spretus* may be higher than 70% because the 10 microsatellites that we failed to detect under the normal stringency conditions of PCR (annealing temperature of 55°C) using *M. spretus* DNA might also be variable. Nine of these microsatellites (Table 1: sequences 2, 7, 13, 15, 17, 18, 31, 37, 38 and 44) were tested again at an annealing temperature of 50°C at the Mg^{2+} concentrations given in Table 2 for 32 cycles. Microsatellites 2 and 38 remained unscoreable (even at an annealing temperature of 45°C). By agarose gel electrophoresis, sequences 7, 17 and 18 appeared to have B6/J and *M. spretus* alleles of equal size. It is noted, however, that size variation of less than 10bp is not reliably detected by agarose gel electrophoresis and these PCR products were not analysed

Table 1. Simple repeat containing DNA sequences or microsatellites in the mouse genome and allelic size variations analysed by PCR. Sequence names are according to the EMBL and Genbank databases. The chromosomal locations and distance from the centromere in cM are taken from the April 1990 edition of the 'Locus Map of the Mouse' from the Jackson Laboratory. The map locations of the following sequences/loci were either confirmed or obtained in this study: 4.MMGLN1/*Glns*, 5.MMHOX23R/*Hox-2*, 7.MLC1A/*Myla*, 9.MMINT2/*Int-2*, 28.MMHSP68/*Hsp68*, 29.MMUPAA/*Plau*, 30.MMCCE/*DXNds3*, 35.MMGSU1/*D6Nds6* and 38.MMBAND3I/*Empb3*. Details are given in the text and Table 3. Primer sequences from the public nucleic acid databases, EMBL and Genbank that were used in the PCR are given with the forward primer listed first and the reverse primer second. Both primers are shown in the 5′ to 3′ orientation. The PCR product size, which corresponds to the published DNA sequence of a particular mouse strain (and therefore only serves as a guideline for the expected size of the PCR product) and the repeat unit of the microsatellites are taken from the nucleic acid databases. The first size variants or allele listed has the highest molecular weight. '=' denotes that the alleles are identical in size on a 15−20cm long acrylamide gel (details given in Materials and Methods). '>' denotes that one allele is larger than another. 'ND' denotes not determined. Mouse strain names are explained in Materials and Methods.

Sequence	Locus	Chromosome (map location, cM)	Primer sequences	PCR product size (bp)	Repeat unit	Size variation
1.MMIGVH16	*Igh-V*	12 (73)	ACATGGTAATTTATGGGCAA CTGGATACCTGCAATAGTAGA	148	$(GT)_{28}$	B6/J=NOD=B6.PL=B10/W>NON=SPE>DBA/2J
2.MMIGVH28	*Igh-V*	12(73)	GTACAACTAATTGCATTCAT ATTACTAAGTCATGGATTTCA	212	$(CA)_{16}(TA)_4$	DBA/2J>B10/W=NOD=B6/J=NON
3.MMGMCSFG	*Csfgm*	11(30)	CTGTGCAACAGACTAAGCCT CTGTAACACAATAACCAGGCA	127	$(GT)_{14}$	DBA/2J=NOD=B10/W=NON>B6/J=B6.PL>SPE
4.MMGLN1	*Glns*	11(10)	AGCTTTGGAGACAACAATTAGATC TGTTCATCAGCTGAGGAATGGATG	181	$(GT)_{20}$	SPE>B6/J=B6.PL=B10/W>NOD=DBA/2J
5.MMHOX23R	*Hox-2*	11(54)	CCTTGCATTCTGAGGCTGAAGGAC TCAGAAGTCTTGCGCTGCATC	218	$(GT)_{24}$	SPE>B6.PL=B10/W=NON=NOD>DBA/2J
6.MMMOPC	*Cyp1a1*	9(28)	TGAATCTTACTGCAGCCTTCT AGTACCCGGGAAGCTCAATGTGGT	174	$(GT)_{12}AT(GT)_4$ $GCT(GT)_6$	SPE=CBA=B6.PL=NOD=NON=B6/J=DBA/2J=B10/W
7.MLC1A	*Myla*	11(62)	ACTAGTCCTACCGGTCTTCCA TGTCTGTTGCTTACTATGTGC	205	$(GT)_{16}$	B10/W=B6/J=B6.PL>NOD>DBA/2J
8.MMLBPA	*Ap2*	3(4)	TCCATAGCATTCATGCGTGCA GTCTGTTGCTTACTATGTGC	146	$(GT)_{20}$	NON>CBA>B10/W=B6.PL=NOD=SPE=B6/J>DBA/2J
9.MMINT2	*Int-2*	7 (74)	GTGACAATACATTCCTGCTGT CTCAGATCTTATCTCTAGCAC	161	$(GT)_{23}$	B10/W=NOD=B6/J=NON=B6.PL>SPE>DBA/2J
10.MMMETII	*Mt-2*	8 (35)	CATGCAGAAGCATGCATTGGTCAC AAGCTTACGGTTTAATCC	121	$(TG)_{24}$	NOD>B10/W=NON=B6/J=DBA/2J>SPE
11.MMPLP7A	*Plp*	X(56)	TAATATAACAGATAACCAACCATTC CATTTTGTAAGATGAGTTTCTA	120	$(CA)_{13}$	SPE>CBA=NOD=NON=B6.PL=B10/W=B6/J=DBA/2J
12.MMNCAM1A	*Ncam*	9(28)	CAGCGGCTTCACCAGAGCATC GGAATGTGTTTGTGTGTGCGT	117	$(CA)_{18}$	B6/J=SPE=B6.PL=NON=BIO/W>NOD=DBA/2J>CBA
13.MMMYCE12	*Myc*	15(18)	CGTCACTGATAGTAGGGAGTA TCAGCGTGCTGTACTTCCAAG	107	$(CA)_{20}$	B6.PL=B6/J=B10/W=DBA/2J=NON>NOD
14.MMNFIL	*Nfh*	ND	GCAATTAATCACTGCAGTCCATTA ATTCTTTTAGCCAGGGTCGCAT	197	$(GA)_4(TA)_{15}$	Balb/c>AKR=NOD=NON>B6/J=B6.PL=DBA/2J=B10/W=C3H=SPE
15.MMAC1A	*Ly-40*	ND	CAACAAGCAGGTCTAGATGGT GTGAGCCACACACAGAGCTTGCT	215	$(GT)_{10}(GA)_{10}$ $A_2(GA)_6$	NOD=B6.PL=B10/W=B6/J=NON
16.MMIGMUD3	*Igh*	12(65)	AGACTTATTGTACCCCACATGTTG TATCTTCCAATCCTAGTTAGGC	196	$(CT)_{30}(CA)_{30}(GA)_4$	NON>SPE>DBA/2J>B10/W=B6/J=B6.PL=NOD
17.MMMYOGG1	*DONds1*	ND	CTGAGTAAGATTCCTGGGTCT TGTGTGTGCATGTACATGTAC	160	$(TC)_{23}$	NOD=NON=B10/W=B6.PL=DBA/2J=B6/J
18.MMKALL	*Kal*	7	ACAGTCATTAGTCATTCAAAC TGGAGTACTTAGGTTGTGCTC	177	$(CT)_9T(TC)_{27}$	CBA=NOD=NON=B6.PL=B10/W=B6/J=DBA/2J
19.MMNGFG2	*Ngfg*	7(21)	CTCCACATGTGTATGTGTATG ATGGAGGCCGAAGAAAGAATC	147	$(TC)_{26}$	SPE>B6.PL=B10/W>NOD=NON
20.MMTHYS3	*DONds2*	ND	CTCTTATTCCTGTTCTACTCA ATTCTTTAGCATTTGTGGATC	88	$(CT)_{15}$	SPE=CBA=NOD=NON=B6.PL=B10/W=B6/J=DBA/2J
21.MMMHCQ4A	*Qa-4*	17(19)	CCTGGAGGAATATCAATAGTG ATACAGAGAAACCCTATCTCAA	214	$(TTC)_{31}(CT)_{32}$	SPE>B6/J=B6.PL>NON>DBA/2J>NOD=B10/W
22.MMTNFAB	*Tnfb*	17(19)	TTCCTGTGGCGGCCTTATCAG AGACAATGGGTAACAGAGGCA	135	$(TC)_{28}C_2(TC)_{12}$ $TT(CT)_5$	B6/J=B6.PL>B10/W=NOD=NON=DBA/2J>SPE
23.MMTNFAB	*Tnfa*	17(19)	GTTTCAGTTCTCAGGGTCCTA CAGGATTCTGTGGCAATCTGG	102	$(CA)_{20}$	B10/W=NOD=NON=SPE>DBA/2J=B10/J
24.MMBCL2	*Bcl-2*	1(41)	CATTATCAATGATGTACCATG GCAGTAAATAGCTGATTCGAC	132	$(CA)_{23}$	B6/J=B6.PL=B10/W=SPE>NOD=NON=DBA/2J
25.MMCD46	*Ly-4*	6(56)	AGGAGAGGATTAACTCTTGAA CATGCATGTGTGCAACATGCG	123	$(CA)_{11}$	B6/J=DBA/2J=NOD=NON=CBA=B6.PL=B10/W>SPE
26.MMREPGT4	*Hprt*	X(23)	TGACAACTTCTGTCCTCAACA ATGCCGTCCTTTATCTAGAAC	97	$(CA)_{14}$	SPE>NOD=NON>B6.PL=B10/W=CBA=DBA/2J=B6/J

27.MMMTCA4	*Tcra*	14(18)	GTCTTTAGTGGTCCTCACATA TTGCTCCTTCCTGTAAATAAG	129	$(TC)_{30}$	NON>CBA>NOD=DBA/2J>B6.PL>SPE=B10/W
28.MMHSP68C	*Hsp68*	17(19)	GTAATTGCGTTGACTGTTAAAT AGTGCTGCTCCCAACATTACT	96	$(TA)_{14}$	SPE>NOD=B10/W>B6.PL=B6/J>NON=DBA/2J
29.MMUPAA	*Plau*	14	TGCTGGCTAGGAATAAACAGA AGGGAATTCATGTTCAGGATA	188	$(GA)_{29}$	DBA/2J>SPE=NON=NOD>B6.PL=B10/W=B6/J
30.MMCCE	*DXNds3*	X(30)	ATGCTTGGCCAGTGTACATAG TCCGGAAAGCAGCCATTGGAGA	111	$(TG)_{29}$	NOD>B10/W=B6/J=B6.PL=SPE=NON>DBA/2J
31.MMDRF3	*D0Nds4*	ND	ATCGTAATCAGATTCCCTCTTCTG GTCATTACATGTATTCTCTGTGTT	121	$(TG)_{20}$	DBA/2J=B10/W=B6.PL>NOD=NON
32.MMRSL1R	*D0Nds5*	ND	CAAGCTGGAAGTTCATTAGGA AACAAGCAGGTATTGTCAACT	153	complex $(GA)_n$	NOD=B10/W=B6.PL=NON=DBA/2J>SPE
33.MMCYO3	*Cyp1a2*	9(28)	CAGCTTTTAAGATTCTGCTGA CCAAGCAAGCACTAGCATAAG	152	complex $(GA)_nCA$	SPE>NOD=B6.PL=B6/J=DBA/2J=B10/W
34.MMCYO3	*Cyp1a2*	9(28)	AGTTTTAGGCTAGTATAGGTT ACTGGAACCTTAGAGCATGAG	198	$(CAAG)_{11}$	SPE>B10/W=B6.PL=B6/J=NON>NOD=DBA/2J
35.MMGSU1	*D6Nds6*	6(32)	GAAGAAGAGAACTAATTCCTC AAGGAAGGTAAATTGTTTTGCTAT	201	$(GT)_4GC(GT)_6$	SJL>NOD>B6.PL=B6/J=DBA2=SPE>AKR>C3H
36.MMBPS2	*Mbp*	18(57)	CAGTACAGCCAGGACACAGAA ATGGCTGACCAACTCTCTAGC	144	$(CA)_{17}$	SPE>NOD=B6.PL=B10/W=CBA=B6/J=DBA/2J=NON
37.MMMUPBS6	*Mup-1*	4(29)	GATGATTAGTAAACCTTAATA TTAGCTATAAGATGGCACTGT	182	$(GT)_5CT(GT)_2$ $TT(GT)_{14}$	B6/J>B10/W>NOD
38.MMBAND3I	*Empb3*	11(65)	TCCTTATTCTGTTGATTGGCAG CTATAGAGAAATCCTGTCTTG	128	$(GT)_{30}$	DBA/2J>NOD=B10/W=B6/J=B6.PL=NON
39.MMGFAPD	*Gfap*	11(64)	AACTGTTCAAAGCCATTTCG CTATGGACTCACAGCCAGGCT	159	$(TG)_6(TC)_9(AC)_8$	SPE>DBA/2J>B10/W=NOD=B6.PL=B6/J=NON
40.MMGFAPD	*Gfap*	11(64)	TGAATTCTAGGACCAGCCAAGGCT ACCTCTAAGATCCTGTGCGAGGCT	277	$(GA)_{21}A(GAA)_8$ $(GA)_2GAAA(GT)_{15}$	DBA/2J>SPE>NOD>B6/J=NON=B6.PL=B10/W
41.MMGFAPD	*Gfap*	11(64)	CAAGCTCTGCCCTTCTGAGTG TTGGCCTTCTTTGGTGCTTTGC	154	$(TC)_{23}$	DBA/2J>SPE>B10/W=NOD=B6.PL=B6/J
42.MMIL4G12	*Il-4*	11(29)	GTCTGCTGTGGCATATTCTG GGCATTTCTCATTCAGATTC	117	$(CA)_{26}$	SPE=DBA/2J=NOD=B10/W>B6.PL=B6/J=NON
43.MMHOX31R	*Hox-3*	15	TTCCTGCTCCCACCTTCTGAG GAATCATCTTCTATATCTTCAGG	166	$(CA)_{19}$	SPE>B10/W=B6.PL=B6/J=NON>NOD=DBA/2J
44.MMGPD	Gdc-1	15(53)	ACACTGGAGAGGAAAACCATG GGTGTGTGCCTTCCACACCAC	160	$(CA)_{13}A(CA)_7$ $A(CA)_{14}$	NOD>B10/W=B6/J=B6.PL=NON=DBA/2J
45.MUSLYT3A1	*Ly-3*	6(32)	GATGCTGTATAGGATGAAGATGA ACCAGGAACGTAAGTGAGCT	147	$(CA)_{27}$	SPE=B10/W=B6.PL=B6/J=NON>DBA/2J>NOD
46.M13548	*Hbb*	7(49)	AAGAGTGTTCTACATGAGAG AGTTCTATACCCACATATCAC	128	$(CA)_{13}$	B10/W=NOD=SPE=B6/J=NON=DBA/2J
47.MMPRPMPB	*Prp*	6(60)	GGAAAAATAATAACAGCTTTCA TCAATTTTCTATCTCTATGTA	203	$(TA)_{15}$	SPE>NOD>B10/W
48. MMIL5G	*Il-5*	11(29)	CCTTTCTGAAAGTATTAAGAGT ACAACCATCTGCATATCCAGC	288	$(AT)_{17}(GT)_{13}(AT)_5$ $(GTAT)_8AT(GT)_{16}$	B6.PL=B6/J=DBA/2J>NOD=B10/W>NON>SPE
49. MMIL1BG	*Il-1b*	2(47)	CCAAGCTTCCTTGTGCAAGTA AAGCCCAAAGTCCATCAGTGG	257	complex $(TC)_n$	DBA/2J=B6.PL=NOD=B10/W=B6/J>NON>SPE
50. MUSACHRA	*Acrg*	1(17)	ACCGTTCACAGCTGACCTAGT GGGACACAGATGTACTAAGCT	112	$(CA)_{12}$	B6/J>NON=DBA/2J=NOD=B10/W>SPE

by acrylamide gel electrophoresis. The B6/J alleles of the microsatellites 13, 15, 37 and 44 were larger than the *M. spretus* alleles as detected by agarose gel electrophoresis (data not shown). Therefore, a total of 28/41 (68%) microsatellites tested on B6/J and *M. spretus* DNA showed variation.

Over the limited range of strains tested, many of the sequences have more than two alleles : 6/50 (12%) have one allele, 13/50 (26%) have 2 alleles, 23/50 (46%) have 3 alleles, 5/50 (10%) have 4 alleles and 3/50 (6%) have 5 alleles (Table 1). It is noted that the number of alleles detected depends on the number of different strains tested. Hence these values are likely to be underestimates because relatively few strains have been analysed.

28/50 (56%) sequences contain only CA/GT repeats, 11/50 (22%) contain only GA/CT repeats, 2/50 (4%) contain only AT/TA repeats (sequences 28 and 47), 1/50 has a CA/GT-AT/TA combination (sequence 2), 1/50 has a GA/CT-AT/TA combination (sequence 14), 4/50 have a CA/GT-GA/CT combination (sequences 15, 16, 33 and 39), 1/50 has $(CAAG)_{11}$ (sequence 34) and 2/50 have repeats with three core types (sequences 40 and 48).

Analysis of polymerase chain reaction products

Most of the 44 variable sequences analysed had at least one allele that could be resolved on agarose mini-gels. Of the 20 sequences that are variable between NOD and B10/W, 9 are agarose resolvable (sequences 13, 14, 29, 31, 34, 37, 43, 45 and 47). An example of an agarose resolvable size variant is shown in Fig. 1. Sequence 29 (*Plau*) is variable between strains C57BL/6J and DBA/2J and this variation was tested on the RI strain mapping panel, BXD. It was estimated, by using the sequence 29 primers

Table 2. PCR conditions (optimal Mg^{2+} concentration and annealing temperature).

Sequence	Temperature(°C)	Mg^{2+}(mM)	Sequence	Temperature(°C)	Mg^{2+}(mM)
1	55	3	26	55	4
2	55	3	27	55	2
3	55	2	28	55	1
4	55	1	29	55	2
5	55	1	30	55	1
6	55	1	31	60	1
7	55	1	32	55	2
8	55	2	33	55	1
9	55	2	34	55	1
10	55	3	35	55	1
11	55	2	36	55	1
12	55	1	37	55	3
13	55	1	38	55	1
14	55	1	39	55	1
15	55	1	40	55	1
16	55	1	41	55	1
17	55	1	42	55	1
18	55	3	43	55	1
19	60	1	44	55	1
20	55	3	45	55	2
21	55	1	46	55	2
22	55	1	47	55	4
23	60	1	48	55	2
24	60	2	49	55	1
25	55	2	50	55	4

described in Materials and Methods, that size differences of 10bp and greater can be clearly resolved on 7cm horizontal agarose gels (data not shown).

As also reported by others[9,10,20], we observed more than one product in the PCR after acrylamide gel electrophoresis and silver staining (Fig. 2). The most prominent additional product is 2bp smaller than the major product band on non-denaturing acrylamide gels. Additional products are not resolved by agarose gel electrophoresis and ethidium bromide staining[20]. Extra bands may be due to incomplete extension by the polymerase[20] and/or terminal transferase activity of the enzyme. We have been unable to eliminate these extra bands (for example, by treatment of the PCR product with mung bean nuclease or Klenow enzyme, data not shown), although the use of fewer cycles of PCR decreases the intensities of minor bands relative to the major band (data not shown). Multiple bands do not usually cause problems in scoring mice as homozygotes or heterozygotes at a particular locus because the major band is usually the most intense after silver staining, even when the difference between two alleles is only one dinucleotide repeat (2bp on a non-denaturing acrylamide gel).

Genetic mapping with microsatellites

To prove that the PCR bands on gels actually correspond to the locus to which the primers had been designed and also to assess the stability of these length variations over a large number of generations of breeding, six chromosomally mapped loci (Table 1)[19], *Hox*-2 (sequence 5; chromosome 11), *Gfap* (sequence 39; chromosome 11), *Myla* (sequence 7; chromosome 11), *Int*-2 (sequence 9; chromosome 7), *Ngfg* (sequence 19; chromosome 7) and *Plau* (sequence 29; chromosome 14) were analysed by PCR analysis of DNA from the BXD recombinant inbred (RI) strain panel (26 strains)[15]. The results are shown in Table 3 and

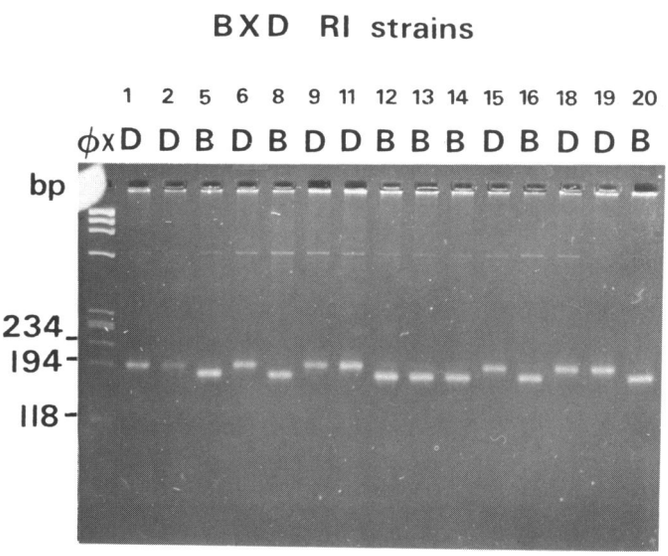

Figure 1. *Agarose gel electrophoresis of the Plau (sequence 29) microsatellite from the BXD RI strains 1–20 DNAs.* PCR reaction conditions[16] are given in Materials and Methods and the primer sequences are shown in Table 1. 30% of the PCR reaction in 1XTBE/bromophenol blue/sucrose buffer was loaded on a 4% (3% NuSieve/1% BRL agarose) horizontal, 7cm, agarose gel in TBE running buffer. Only the PCR products from the DNAs from the first 15 BXD strain DNAs are shown. The upper allele is from DBA/2J DNA and the lower allele is from C57BL/6J. DNA markers (first lane) are ϕx174 digested with *Hae*III (BRL).

are consistent (with one exception, see later) with published data[15,19,21–26].

Two loci, *Empb3* (sequence 38) and *Glns* (sequence 4), which have not been assigned to a mouse chromosome were mapped

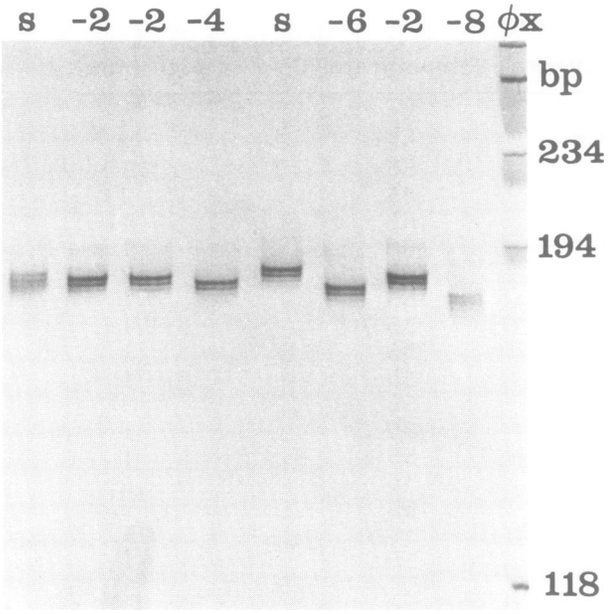

Figure 2: *Size differences of 2bp or more are resolvable on acrylamide gels.* Sequence 29 (*Plau*) was subjected to PCR using a set of primers that generates products, 4, 8, 12 and 16bp smaller than the primers shown in Table 1. The sequences of these primers are given in Materials and Methods. 10–30% of the PCRs were loaded in TBE/bromophenol blue/xylene cyanol/sucrose buffer onto an 8% non-denaturing, vertical, TBE acrylamide gel (0.75mm thick, 10cm long; Tall Mighty Small, Hoeffer). After electrophoresis, DNA was detected by silver staining. PCR products are: 's' (standard size), '−2' (2bp smaller than standard), '−2', '−4', 's', '−6', '−2' and '−8'. DNA markers (last lane) are φx174 digested with *Hae*III (BRL).

to chromosome 11 (Table 3). The location of *Empb3*, which encodes the erythrocyte surface protein band 3 on mouse chromosome 11, is supported by the presence of the homologous gene, EPB3 on human chromosome 17 because mouse chromosome 11 and human chromosome 17 share large regions of synteny[27].

The order of the chromosome 11 genes identified by microsatellite variation is consistent with previously published reports[15,21], with the exception of the *Gfap* locus. By comparison of BXD SDPs, *Gfap* appears to be much closer to *Myla* (3.5cM; Table 3) than shown in the Jackson Laboratory Mouse Locus Map, which places them 16cM apart.

26/26 RI strain DNAs tested for the *Int-2* microsatellite matched the pattern for the published *Int*-2 locus on chromosome 7 (Ref. 25). 25/25 RI strain DNAs tested for *Ngfg* matched the published pattern for this locus on chromosome 7 (Ref. 15). In addition, every PCR product that has been tested by DNA blotting and hybridization with radiolabelled probe containing the appropriate dinucleotide repeat sequence has shown positive hybridization (data not shown). Given the well established high specificity of PCR under optimal temperature and Mg^{2+} concentrations[16] and other data from human PCR-analysed dinucleotide repeat polymorphisms[9–11] the results indicate that the major product is derived from the target sequence.

The observation that the size of the microsatellite alleles in BXD RI strains remains constant indicates that these variants are extemely stable and are not subject to frequent mutation as has been reported for certain VNTR polymorphisms[4,5]. The stability of microsatellite variants increases the utility of these sequences in genetic mapping and also in DNA fingerprinting applications.

		BXD strain																											
	Locus	1	2	5	6	8	9	11	12	13	14	15	16	18	19	20	21	22	23	24	25	27	28	29	30	31	32	r(x100)	95% limits (x100) of r
(a)	*Pmv-22*	B	B X	B	D	B	D	B	B	D	D	B X	B	D	B	D	B	B	B X	D	B	D	D	D	B X	D	D	5.00	1.17-18.28
	Glns	B	D	B	D	B	D X	B	B	D	D	D	B	D	B X	D	B	B	D X	D	B	D	D	D	D	D	D	3.49	0.64-13.76
	Hba	B	D	B	D	B	B	B	B	D	D	D	B	D	D	D	B	B	B	D	B	D	D	D	D	B	D		
	Tca-3	B	D	D	B	B	D	B	D	B	D	D	D	B	D	D	D	D	B	B	B	B	B	D	B	B	D		
	Mpmv-4	B	D	D	B X	B X	D	B	D X	B	D	D	D	B	D	D	D	D	B	B X	B	B X	B	D	B X	B	D	8.82	2.59-31.61
	Hox-2	B X	D	D	D	D	D	B	B	B	D X	D X	D X	B X	D	D	D	D	B	D	B	D X	B	D	D	B	D	8.82	2.59-31.61
	Myla	D	D	D	D	D	D	B	B	B	B	B X	B	D	D	D	D	D	B	D	B	B	B	D X	D	B	D	2.17	0.24-10.08
	Empb3	D	D	D	D	D	D	B	B	B	B X	D	B	D	D	D	D	D	B	D	B	B	B	B	D	B	D	1.02	0.03-6.96
	Gfap	D	D	D	D	D	D	B X	B X	B X	D	D	B X	D	D	D	D	D	B	D	B	B	B	B	D	B	D	5.00	1.17-18.28
	Mpmv-8	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	B	D	B	B	B	B	D	B	D		
(b)	*Int-2*	D	D	D	B	B	D	B	D	D	B	D	B	D	B	D	B	B	D	B	D	D	D	D	B	D	D		
(c)	*Ngfg*	B	D	D	D	B	B	B	D	B	D	B	B	B	B	B	B	D	B	B	D	B	D	B	B	-	D		
(d)	*Plau*	D	D	B	D	B	D	D	B	B	B	D	B X	D	D	B	D	D	B	B X	B	B	B	D	B	D	D	2.38	0.26-11.34
	Odc-9	D	D	B	D	-	D	D	B	B	-	D	D	D	D	B	D	D	B	D	B	B	B	D	B	D	D		
(e)	*Hsp68*	D	B	D	D	B	D	D	D	B	B	B	D	D	B	B	D	D	B	D	D	D	D	B	D	D	D		
(f)	*DXNds3*	B	B	B	D	B	D	B	B	B	B	-	B	B	D X	B X	B	B	B	B	B	B	B X	B	B	B	D X	8.82	
	Rsvp	B	B	B	D	-	D	B	B	B	B X	B	B	-	B	D	B	B	B	B	B	B	D	B	B	B	B	1.67	
	Cf-8	B	B	B	D	-	D	B	B	B	D	B	B	-	B	D	B	B	B	B	B	B	D	B	B	B	B		

Table 3. Strain distribution patterns in BXD RI strains of (a) the chromosome 11 loci *Pmv-22*, *Hba*, *Mpmv-4*, *Hox-2*, *Mpmv-8*, *Gfap* and *Myla* with *Empb3* and *Glns* (b) *Int-2* on chromosome 7 (c) *Ngfg* on chromosome 7 (d) the chromosome 14 locus *Odc-9* with *Plau* (e) *Hsp68* on chromosome 17 (f) chromosome X loci *Cf-8* and *Rsvp* with sequence 30.MCCE, *DXNds3* 'X' denotes a crossover event between loci. Loci on chromosomes 11, 14 and X are shown in order from centromere (listed first) to telomere. r = recombination frequency[18].

Two other sequences of unknown chromosomal location were mapped using the BXD RI panel (Table 3). Sequence 28 (Table 1), heat shock protein 68 (ref. 29), designated *hsp68*, was mapped to *H-2S* in the Major Histocompatibility Complex (MHC) on chromosome 17 (SDPs from BXD RI strains were identical for *hsp68* and *H-2S*)[15]. This result is consistent with the location of a heat shock protein gene in the human MHC[30]. Sequence 30.MCCE, designated *DXNds3*, which is a DNA segment that contains an enhancer of retroviral gene expression[31] had a 19/23 match with the X chromosome locus, *Rsvp*, which is also linked to the locus *Cf-8*[32]. The assignment of *DXNds3* to chromosome X was further supported by the presence of only the B10/W allele in DNA from a (NODXB10/W) F1 male animal (where the Y chromosome is inherited from the NOD male parent). Sequence 26, *Hprt* is located on the X chromosome[19], and also shows only the B10/W microsatellite allele in DNA from the (NODXB10/W) F1 male (data not shown).

Sequence 35.MMGSU1[33], designated *D6Nds6* was mapped to chromosome 6 because it co-segregated with the chromosome 6 locus, *Ly-3* (sequence 45; which co-segregates with the immunoglobulin locus, *Igκ*[19]) in a NODX(NODXB6.PL) backcross (unpublished data). This is consistent with the previous observation that both the MMGSU1 (*D6Nds6*) and C_κ sequences are deleted in a λ light chain-producing tumour, which suggested that these two sequences may be linked physically[33]

DISCUSSION

The high level of variation of microsatellites between different strains of mice indicates that this approach will be a useful contribution to the construction of a high resolution map of the mouse genome. The approach is efficient and easy to use, particularly since access to a probe only requires the sequences of the primers. The abundance, stability and apparent random distribution of microsatellites in the mouse genome indicate that construction of a 1cM map should be feasible. Given that there are about 10^5 copies of $(CA)_n$ microsatellites (where it is assumed that $n > 10$) in the mouse genome[12] it can be calculated that, if 50% are variant between, for example, *M. spretus* and B6/J, then a marker should be present every 30kb.

On the basis of the relatively small amount of data available, the length of the repeat appears to be related to the number of alleles of the sequence. Calculation of the average number of repeats in a single block for each category of sequences with 1, 2, 3, 4 and 5 alleles showed a weak correlation between the number of alleles and the number of repeats. For 1, 2, 3, 4 and 5 allele sequences, respectively, the average number of dinucleotide repeats was 17, 18, 21, 20 and 23. Currently, we are cloning and sequencing microsatellites at random from the mouse genome. Based on the present results, the longer repetitive sequences have been selected by isolation of M13 clones that hybridize strongly to dinucleotide probes in order to isolate more variable microsatellites (Cornall, R., Aitman, T., Hearne, C., JAT, unpublished).

Generally, the size difference between alleles is in the order of 4−40bp (data not shown). Some of the *M. spretus* alleles are, however, considerably different in size from the inbred strain alleles. For example, the *M. spretus* allele of the chromosome 7 *Ngfg* gene is about 275bp in length compared to alleles of inbred strains which are about 150bp in length (data not shown). This *M. spretus* DNA-derived PCR band has been shown in a (*M. spretus*Xinbred strain) backcross to be linked to other RFLP markers of chromosome 7 and has allowed confirmation of the chromosome 7 location of *Ngfg* (Cavana, J.S., Greenfield, A.J., Rosalki, J., Steel, K., and Brown, S.D.M., unpublished).

GA/CT and AT/TA microsatellites are also variable in size between strains. A GA/TC sequence from whale DNA[11] and an AT-rich sequence from the human interleukin-6 gene[34] have been shown previously to be polymorphic. We have also found that stretches of A and T bases are variable between different inbred mouse strains (Aitman, T.J., McAleer, M., Hearne, C., JAT, unpublished). Some of these A- or T-rich sequences are adjacent to SINE repetitive DNA sequences and represent another abundant source of polymorphism[35]. Similarly, poly-A sequences adjacent to human Alu sequences have recently been shown to be variable in length[36].

One disadvantage of dinucleotide repeat microsatellites is that size variation between alleles can be small and, with extra PCR products, scoring can be difficult. A strategy to alleviate this problem is to isolate microsatellites that have 3-, 4- or 5-nucleotide repeats[37]. Microsatellite 34 (Table 1) is a 4-nucleotide repeat array that is easily resolved by agarose gel electrophoresis (data not shown). We have characterised several others in mouse and human which are also resolved by agarose gel electrophoresis and which are variable or polymorphic[37] (unpublished).

There are over 100 microsatellites in the mouse nucleic acid databases. The first 50 that we have studied show that PCR-analysed size variants are common, stable, simple to analyse and useful in genetic mapping. It is a simple procedure to isolate microsatellites from cosmid and phage clones of mouse (and human) DNA, obtain flanking sequences and design primers for PCR. With the 18 mapped markers that are polymorphic between NOD and B10/W, it is estimated that 39% of the genome would be covered in a NODX(NODXB10/W) backcross. Microsatellite variants in combination with standard restriction fragment length variants[38] will give over 50% coverage of the genome. These markers will therefore be useful in the search for non-MHC genes that influence susceptibility to insulin-dependent diabetes in NOD. Further, in Table 1 there are 33 microsatellites that are variable between *M. spretus* and NOD, giving an estimated 52% coverage of the genome in a NODX(NODX*M. spretus*) backcross. This set of markers and others we have characterised (data not shown) will contribute to a PCR-analysed genetic map of the mouse genome.

ACKNOWLEDGEMENTS

Support from the Medical Research Council as part of the UK Human Genome Mapping Project, the Wellcome Trust, the British Diabetes Association and the Juvenile Diabetes Foundation is gratefully acknowledged. Drs S. Brown, J. Peters, B. Taylor, E. Boyd, E. Simpson, other colleagues and one of the referees are thanked for critical review.

REFERENCES

1. Risch,N. (1990) Am. J. Hum. Genet., **46**,242−253.
2. Jeffreys,A.J., Wilson,V. and Thein,S.L. (1985) Nature, **314**,67−73.
3. Nakamura,Y., Leppert,M., O'Connell,P., Wolff,R., Culver,M., Martin,C., Fujimoto,E., Hoff,M., Kumlin,E. and White,R. (1987) Science, **235**,1616−1622.

4. Jeffreys,A.J., Wilson,V., Kelly,R., Taylor,B.A. and Bulfield,G. (1987) Nucleic Acids Res., **15**,2823–2836.
5. Julier,C., de Gouyon,B., Georges,M., Guenet,J.-L., Avner,P. and Lathrop,M. (1990) Proc. Natl. Acad. Sci. USA, in press
6. Jeffreys,A.J., Wilson,V., Neumann,R. and Keyte,J. (1988) Nucleic Acids Res., **16**,10953–10971.
7. Horn,G.T., Richards,B. and Klinger,K.W. (1989) Nucleic Acids Res., **17**,2140.
8. Boerwinkle,E., Xiong,W., Fourest,E. and Chan,L. (1989) Proc. Natl. Acad. Sci. USA, **86**,212–216.
9. Weber,J.L. and May,P.E. (1989) Am. J. Hum. Genet., **44**,388–396.
10. Litt,M. and Luty,J.A. (1989) Am. J. Hum. Genet., **44**,397–401.
11. Tautz,D. (1989) Nucleic Acids Res., **17**, 6463–6471.
12. Hamada,H., Petrino,M.G., Kakunaga,T. (1982) Proc. Natl. Acad. Sci. USA, **79**,6465–6469.
13. Litt,M., Sharma,V. and Luty,J.A. (1989) Cytogenet. Cell Genet., Human Gene Mapping **10**,A2053.
14. Graham,D. E. (1978) Anal. Biochem., **85**,609–613.
15. Taylor,B.A. (1989) In Lyon,M.F. and Searle,A.G. (eds.), Genetic Variants and Strains of the Laboratory Mouse. Oxford University Press, Oxford, second ed., pp. 773–796.
16. Saiki,R.K., Gelfand,D.H., Stoffel,S., Scharf,S.J., Higushi,R., Horn,G.T., Mullis,K.B. and Erlich,H.A. (1988) Science, **239**,487–491.
17. Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor
18. Silver,J. (1985) J. Hered., **76**,436–440.
19. Hillyard,A.L., Doolittle,D.P., Davisson,M.T. and Roderick,T.H. (1990) Locus Map of Mouse with Comparative Map points of Human on Mouse. The Jackson Laboratory, Bar Harbor, ME
20. Olsen,D.B. and Eckstein,F. (1989) Nucleic Acids Res. **17**, 9613–9620
21. Buchberg,A.M., Brownell,E., Nagata,S., Jenkins,N.A. and Copeland,N.G. (1989) Genetics, **122**,153–161.
22. Frankel,W.N., Stoye,J.P., Taylor,B.A. and Coffin,J.M. (1989) J. Virol., **63**, 3810–3821.
23. Frankel,W.N., Stoye,J.P., Taylor,B.A. and Coffin,J.M. (1990) Genetics, **124**,221–236
24. Wilson,S.D., Billings,P.R., D'Eustachio,P., Fournier,R.E.K., Geissler,E., Lalley,P.A., Burd,P.R., Houseman,D.E., Taylor,B.A. and Dorf,M.E. (1990) J. Exp. Med. **171**,1301–1314.
25. Silver,J. and Buckler,C.E. (1986) J. Virol., **60**,1156–1158.
26. Ceci,J.D., Kingsley,D.M., Silan,C.M., Copeland,N.G. and Jenkins,N.A. (1990) Genomics, **6**,673–678.
27. Solomon,E. and Barker,D.F. (1989) Cytogenet. Cell Genet., **51**,319–337.
28. Silver,J. and Buckler,C.E. (1986) Proc. Natl. Acad. Sci. USA, **83**,1423–1427.
29. Lowe, D.G. and Moran, L.A. (1986) J. Biol. Chem., **261**,2102–2112.
30. Sargent,C.A., Dunham,I. and Campbell,R.D. (1989) EMBO J, **8**,2305–2312.
31. Taketo,M. and Tanaka,M. (1987) Proc. Natl. Acad. Sci. USA, **84**, 3748–3752.
32. Mullins,L.J., Grant,S.G., Stephenson,D.A. and Chapman,V.M. (1988) Genomics, 3,187–194.
33. Hochtl,J. and Zachau,H.G. (1983) Nature, **302**,260–263.
34. Bowcock,A.M., Ray, A., Erlich,H.A. and Sehgal,B. (1989) Nucleic Acids Res., **17**,6855–6864.
35. Hastie,N.D. (1989) In Lyon, M.F. and Searle, A.G. (eds.), Genetic Variants and Strains of the Laboratory Mouse. Oxford University Press, Oxford, second ed., pp. 559–573.
36. Economou,E.P., Bergin,A.W., Warren,A.C. and Antonarakis,S.E. (1990) Proc. Natl. Acad. Sci. USA, **87**,2951–2954.
37. Todd,J.A., McAleer,M.A., Aitman,T.J., Hearne,C., Cornall,R., Love,J.M. (1990) In McPherson,M.C., Quirke,P. and Taylor,G. (eds.), PCR: A Practical Approach. Oxford University Press, Oxford, in press
38. Knight,A.M. and Dyson,P.J. (1990) Mol. Cellular Probes, in press