# Mutation at a hypervariable mouse minisatellite

Thesis submitted for the degree of
Doctor of Philosophy
at the University of Leicester

by

Mark Gibbs BSc. (Nottingham)
Department of Genetics
University of Leicester

May 1992

UMI Number: U548818

UMI

Dissertation Publishing

ProQuest

# Mutation at a hypervariable mouse minisatellite

Mark Gibbs, Department of Genetics, University of Leicester.

## Abstract

'Minisatellites' are a class of tandemly repeated sequences ubiquitous to eukaryotic genomes. A subset of minisatellites have been found to be highly variable in tandem repeat copy number. This variability makes such loci highly informative markers for genetic mapping, establishing family relationships in humans and other animals, and individual identification from forensic samples. One highly unstable mouse minisatellite locus, Ms6-hm, has been previously identified in mouse DNA fingerprints by cross-hybridization with the human minisatellite probe 33.6. Ms6-hm showed a high germline mutation rate to new length alleles (2.5% per gamete) causing multi-allelism and heterozygosity even within inbred strains. Mice mosaic for cells carrying a non-parental allele in somatic tissue are also seen (2.8% of mice). At reduced stringency Ms6-hm detects other loci, one of which, Hm-2, also appeared to be highly unstable This work describes the characterization of this locus. Hm-2 has been cloned, localized to chromosome 9 using recombinant inbred strains, and sequenced. The repeat sequence, $(GGCA)_n$, is short and the largest Hm-2 alleles can have over 5000 tandem repeats. Like Ms6-hm, Hm-2 shows a high germline mutation rate (3.6% per gamete). The incidence of mosaicism at Hm-2 (20.4% of adult mice) is much greater than at Ms6-hm, however, making Hm-2 an ideal locus for further study of somatic mutation events. Analysis of Hm-2 in embryonic and extra-embryonic tissues has shown that many mutation events occur early in development to produce in some cases divergence in Hm-2 genotype between the embryo and trophoblast. In others, mosaicism was shared between the embryo and trophoblast suggesting that these mutations arose very early in development before the fifth cell division following fertilization. The degree of mosaicism in 60 mosaics has been calculated and comparison of this data with computer simulations suggests that the somatic mutation rate is not constant throughout development but rather that mutations are biased towards the first two cell divisions after fertilization.

To Elizabeth and Tony

# Contents

## Chapter 1  Introduction

## Chapter 2  Materials and Methods

## Chapter 3  Cloning of *Hm-2*: a second highly unstable mouse minisatellite

## Chapter 4  Localization of *Hm-2*

# Chapter 5 Rates of Germline and Somatic Mutation at *Hm-2*

# Chapter 6 Somatic Mutation during Early development

# Chapter 7 Evidence for an early developmental 'window' of somatic mutation at *Hm-2*

# Chapter 8   Mutation processes at *Hm-2*

# Chapter 9   Discussion

# Abbreviations

| | |
|---|---|
| ATP | adenosine 5'-triphosphate |
| BCIG | 5-bromo-4-chloro-3-indoyl-β-D-galactopyranoside |
| bp, kb, Mbp | base pair, kilo~, mega~ |
| BSA | bovine serum albumin |
| BXD | C57BL/6J x DBA/2J (RI) strains |
| BXH | C57BL/6J x C3H/HeJ (RI) strains |
| Ci | Curie |
| CIP | calf intestinal phosphatase |
| cM | centiMorgan |
| dATP | 2'-deoxyadenosine 5'-triphosphate |
| dCTP | 2'-deoxycytosine 5'-triphosphate |
| dGTP | 2'-deoxyguanosine 5'-triphosphate |
| dTTP | 2'-deoxythymidine 5'-triphosphate |
| dNTP | 2'-deoxyribonucleoside 5'-triphosphate |
| $dH_2O$ | distilled water |
| ddATP | 2'3'-dideoxyadenosine 5'-triphosphate |
| ddCTP | 2'3'-dideoxycytosine 5'-triphosphate |
| ddGTP | 2'3'-dideoxyguanosine 5'-triphosphate |
| ddTTP | 2'3'-dideoxythymidine 5'-triphosphate |
| ddNTP | 2'3'-dideoxyribonucleoside 5'-triphosphate |
| DNA | deoxyribonucleic acid |
| dpm | disintegrations per minute |
| DTT | dithiothreitol |
| EDTA | ethylenediaminetetra-acetic acid |
| g, mg, μg, ng | grams, milli~, micro~, nano~ |
| HEPES | N-(2-hydroxyethyl)piparizine-N'-2(ethanesulphonic acid) |
| hr | hour |
| IPTG | isopropyl-β-D-galactopyranoside |
| l, ml, μl | litre, milli~, micro~ |
| LINE | long interspersed repeated element |
| LUA | Luria agar |
| LUB | Luria broth |
| M, mM, μM | molar, milli~, micro~ |
| MOPS | 3(N-morpholino)propanesulphonic acid |
| mtDNA | mitochondrial DNA |
| OD | optical density |

| | |
|---|---|
| o/n | overnight |
| PBS | phophate buffered saline |
| p.c. | *post coitum* |
| PCR | polymerase chain reaction |
| PEG | polyethylene glycol |
| PIC | polymorphism information content |
| PVP | polyvinyl pyrrolidone |
| RFLP | restriction fragment length polymorphism |
| RI | recombinant inbred |
| RNA | ribonucleic acid |
| rpm | revolutions per minute |
| r.t. | room temperature |
| SDP | strain distribution pattern |
| SDS | sodium dodecyl (lauryl) sulphate |
| SINE | short interspersed repeated element |
| SSC | saline sodium citrate |
| SWXL | SWR/J x C57/L (RI) strains |
| TAE | Tris-acetate EDTA |
| TBE | Tris-borate EDTA |
| TEMED | N,N,N'N'-tetramethyl-ethylenediamine |
| Tris | Tris-(hydroxymethyl)-methylamine[2-amino-(2-hydroymethyl)-propan-1,3-diol] |
| u | unit |
| UV | ultra violet |
| vol | volume |
| VNTR | variable number of tandem repeats |
| w/v, v/v | weight to volume, volume to volume |

# Acknowledgements

First and foremost, I would like to thank Alec whose enthusiastic and organized approach to science has always been an example and an encouragement throughout my time in his lab.

I would like to thank all members of G19, past and present for making my time within the lab entertaining and enjoyable, and for providing helpful advice, assistance, and chocolate when necessary: thanks to Andy, Annette, Darren, David, Esther, Ian, Ila, John, Keiji, Maureen, Max, Moira, Nicola, Rita, Robert, Sarah, Vicky, Yuri, and Zilla. In particular I would like to thank Robert, Andy, John, Nicola, and Darren for illuminating discussions.

Outside of G19 I have also been fortunate to be part of an extremely friendly and social department, for which I think much is owed to Terry Lymn , thank you everybody.

I would also like to thank all those who made my time in Leicester enjoyable, Cathryn, Louise, Lee, Richard, Darren, Karl, Gordon and rest of the Cornwall crowd for excursions near and far, barbecues and beers, Paul and Es for Sunday afternoon bashes and weekends in exotic sports halls, and finally, but far from least, I would like to thank Cathryn for encouragement, friendship and hugs.

# Publications

Some of the contents of this thesis have been previously published:

KELLY, R., BULFIELD, G., COLLICK, A., GIBBS, M., & JEFFREYS, A.J. (1989).
Characterization of a highly unstable mouse minisatellite locus:
evidence for somatic mutation during early development.
*Genomics* **5**: 844-856

KELLY, R., GIBBS, M., COLLICK, A., & JEFFREYS, A.J. (1991).
Spontaneous mutation at the hypervariable mouse minisatellite locus
*Ms6-hm*: flanking DNA sequence and early somatic mutation events.
*Proc. R. Soc. Lond.* B **245**: 235-245

JEFFREYS, A.J., ROYLE, N.J., PATEL, I., ARMOUR, J.A.L., MACLEOD, A.,
COLLICK, A., GRAY, I.C., NEUMANN, R., GIBBS, M., CROSIER, M.,
HILL, M., SIGNER, E. & MONCKTON, D.G. (1991).
*In,* "DNA Fingerprinting: Approaches and Applications" (Burke, T.,
Dolf, G., Jeffreys, A.J., and Wolff, R.Eds) pp1-19, Birkhauser Verlag, Bern.

GIBBS, M., KELLY, R., & JEFFREYS, A.J. (1991).
Evidence for somatic mutation during early development at highly
unstable mouse minisatellite loci.
*Genet. Res. Camb.* **58** 78

# 1
## Introduction

## 1.1 Repetitive DNA

The existence of repetitive DNA sequences in eukaryotic genomes has been known for over 20 years (Britten and Kohne, 1968), but even now, the organization and function of much of this DNA, which is estimated as 20-30% of the human genome (Schmid & Jelinek, 1982), but can be up to 50% of some genomes (such as the kangaroo rat *Dipodomys ordii*), remains unknown. The repetitive fraction can be divided into two classes; tandem repeats and dispersed repeats. The former category includes satellite, minisatellite, and microsatellite DNA; the latter category, as the name suggests, includes sequence families that are not tandemly repetitive, and are dispersed throughout the genome. Examples of these sequences are the SINEs (short interspersed repetitive elements), such as the B1 element in mice (Krayev *et al.*, 1982), which is related to the human Alu element; LINEs (long interspersed repetitive elements), such as the L1 element in humans and the related L1md sequences in mice (reviewed by Hastie, 1989), and also other elements related to retroviruses, and processed pseudogenes. This study is concerned with the tandem repetitive portion of the genome, and in particular, with minisatellite sequences. Of principal interest is the variation found in tandemly repetitive DNA sequences, in particular the extent of variability in the tandem repeat copy number at a given locus. Any polymorphic system, such as blood groups, protein electrophoretic variants, or restriction fragment length polymorphisms (RFLPs), is of use for genetic analysis, i.e., for the construction of genetic maps to find markers linked to traits of interest. Linked markers can be used for genetic tests even when the affected gene is unknown. The power of disease linked markers was demonstrated as early as 1978 by Kan and Dozy who described the predictive diagnosis of sickle cell anaemia using an RFLP located near the β globin gene. The informativeness of polymorphic loci is limited, however, when the number, or population distribution of different distinguishable allelic states is low; a system with with one very common allele and many rare variants is likely to be uninformative, despite the

variation. The majority of RFLPs, and almost all polymorphic protein markers have only a limited number of alleles (with the notable exception of the major histocompability loci, HLA and H-2 in mice). In contrast, those tandemly repetitive loci which have a large number of alleles within a given population make an exceptionally informative RFLP system for genetic linkage analysis, and have become vital for the construction of genetic maps (eg. Donis-Keller *et al.*, 1987) and for the positional cloning approach to identifying single gene defects (reviewed by Collins, 1992).

This introduction will review the organization of, and the extent of variability within tandemly repeated DNA, with especial reference to the biology of minisatellites. The possible mechanisms by which variation in tandem repeat copy number is generated will be discussed in chapters 8 and 9.

## 1.2    Organization of tandemly repetitive DNA

### 1.2.1    The archetypal tandem repeat: 'Satellite DNA'

The ability of CsCl density gradient centrifugation to separate distinct components of the genome was first observed using mouse DNA (Kit, 1961), when it was seen that in addition to a major band which represented DNA with an average G-C content of 42%, a further band (8% of the DNA) separated which reflected DNA with an average G-C content of only 30%; this fraction was termed the 'satellite' band. The DNA contained in this fraction was subsequently found to be highly repetitive and analagous to the highly repetitive sequence identified by reassociation kinetics. The term 'satellite' is now applied to any large tract of tandem repetitive sequence associated with heterochromatin, and includes blocks of repeats that do not have an atypical G-C content.

By far the most abundant satellite sequence in humans is the alphoid-satellite, consisting of a basic 171bp tandemly reiterated sequence, which may form higher-order tandem repeats. Alpha satellite sequences are found at the centromere of every human chromosome (Rosenberg *et al.*, 1978), each centromere array consisting of up to several thousand copies of individual higher-order repeats, giving a locus of 250kb to 4000kb (reviewed by Willard, 1990), and it has been suggested that the sequences might play a role in chromosome pairing at meiosis. The presence of a higher order structure to the satellite blocks giving a chromosome specific α-satellite on the metacentric chromosomes (Willard & Waye, 1987), suggests that the sequences are

propagated by some sort of crossover mechanism which amplifies a novel repeat unit type.

In mice, both the major 'γ' satellite, which has a 234bp monomer (Horz and Altenburger, 1981) organized in arrays of from 240kb to over 2000kb (Vissel and Choo, 1989), and the sequence related 'minor' satellite (120bp monomer unit present at 50,000 copies per genome, or 5% of the total (Wong and Rattner, 1988)) are found at the centromeres. There is evidence that on some of the acrocentric mouse chromosomes there is little, if any, 'typical' genomic DNA separating the proximal telomere repeats from the centromeric minor satellite (Kipling et al., 1991). In contrast to human metacentric chromosomes, both mouse acrocentric and human acrocentric chromosomes have homologous γ and α satellites respectively, believed to result from frequent recombinational exchange between non-homologous chromosomes. This may account for the increase in Robertsonian translocations in the mouse (Vissel and Choo, 1989).

### 1.2.2 The continuum of repetitive DNA

As more of the genome was characterized and sequenced, other, smaller tandemly repeated sequences were subsequently described, such as microsatellites and minisatellites. In some respects these are also dispersed repeats since they are found throughout the genome. However, they are distinguishable by their tandem repetitive nature, and the unique origin of each locus; there is no evidence to suggest a common origin for groups of similar loci, unlike, for example, both Alu and B1 SINEs, which show homology to the 5' and 3' ends of the 7SL RNA gene (Ullu and Tschudi, 1984) and are believed to be derived from a 7SL RNA pseudogene which lost the central region and was subsequently duplicated and transposed, via retroposition, throughout the genome.

Within eukaryotic genomes exists a continuum of repeat unit lengths with individual repeats ranging from a single nucleotide to many kilobases, and similar variation is found for the average copy number of repeats per array. This continuum has been somewhat arbitrarily broken down into distinct classes of tandem repeated sequences, with names following on from 'satellite'; i.e. 'midisatellites', 'minisatellites' (often called VNTR loci, for variable number of tandem repeat, although this description includes variable micro- and midisatellites), and 'microsatellites' (or short tandem repeats, STR), summarized in table 1.1. (Although many factors, such as the size of the repeat unit, overlap between groups.) It seems that the average length of array is the

most common definition of the different groups of tandem repeats, and perhaps the logical extension of this is to distinguish the groups by the technology required to resolve individual alleles, i.e. microsatellites are resolved by polyacrylamide gel electrophoresis of PCR amplified alleles (although high percentage agarose gels can resolve larger alleles), minisatellites by conventional agarose gel electrophoresis, and midisatellites by pulse-field gel electrophoresis. Each of these categories will be described, with the inclusion of tandemly repeated telomere sequences. Minisatellites will be described out of the logical order, since the bulk of the further introduction refers solely to these sequences.

**Table 1.1**     **Characteristics of tandemly repeated sequences**

|  | Satellite | Midisatellite | Minisatellite | Microsatellite |
|---|---|---|---|---|
| Repeat unit length (nucleotides) | large 100+ | medium - large 9-40+ | small-medium 5-50 | Small 1-4 |
| typical array length | 200-5000+kb | 10-500kb | 1-20kb | 0.02-0.3kb |

### 1.2.3 Midisatellites

Within the hierarchy of tandem repeats these sequences fill the gap between the extremely large satellite arrays and the much smaller minisatellites. Relatively few midisatellite sequences have been reported, perhaps reflecting the greater difficulties in cloning and characterizing larger loci. Two that have been described are a sequence near the telomere of human chromosome 1, with a tandem repeat of 40bp in an array of 250 to 500 kb (Nakamura et al., 1987b) and a sequence in the psuedoautosomal region of the X and Y chromosomes with a 61bp tandem repeat in arrays of 10 to 50kb (Page, 1987). Both showed length polymorphisms and internal repeat sequence polymorphisms. Two further midisatellites have been isolated from human DNA by I. Grey (pers. comm.) with array lengths of 50-200kb, but a relatively short repeat sequence (consensus AGCCAAGCC). Both loci showed high frequencies of diverged repeats and were highly polymorphic in internal structure, and although these loci await further characterization, it is believed that the rate of mutation to new length alleles by gain or loss of repeats at these loci would be high.

### 1.2.4 Microsatellite DNA

Microsatellites are usually defined as short (<1kb), simple tandem repeated (STR) loci, the length of the repeat is typically ≤6bp (Epplen *et al.*, 1991), although the distinction between various 'satellites', minisatellite and microsatellite on the basis of repeat length is highly arbitrary; the minisatellite *Ms6-hm*, has a 5bp repeat and alleles of up to 16kb in length (Kelly *et al.*, 1989), moreover, many tracts of satellite DNA, particularly in arthropods, consist of simple tandem repeats (John and Miklos, 1988).

The simple tandem repeats most characterized are the dinucleotide repeats, typically $(CA)_{10-30}$ (Litt and Luty, 1989; Weber and May, 1989; Tautz, 1989), although other short di- tri- and tetranucleotide repeat arrays have also been shown to be variable (Tautz, 1989). The small size of microsatellite alleles makes them particularly amenable to amplification by the polymerase chain reaction (PCR, Saiki *et al.*, 1988), and analysis of amplified microsatellites has shown that alleles show variation in tandem repeat copy number. Typically, the greater the average number of repeats per allele at a given locus, the greater the number of different allele lengths (Weber, 1990), although the variation at such small loci is, not surprisingly, lower than for a typical minisatellite, with fewer length states, and allele size frequency distributions dominated by one or a few alleles. The few new mutations that have been seen at microsatellite loci are consistent with a mutation rate at these regions being equivalent to minisatellites of equal variability (Kwiatkowski *et al.*, 1992).

Microsatellites are highly dispersed in human (Luty *et al.*, 1990) and other higher eukaryotic genomes, but are not found in bacteria (Gross and Garard, 1986). Estimates of the copy number in the human genome have ranged from 32,000 to 130,000 (Miesfeld *et al.*, 1981; Hamada and Kakunaga, 1982; Hamada *et al.*, 1982; Sun *et al.*, 1984). The lower variability and heterozygosity of microsatellites (Weber, 1990) makes them less informative as genetic markers, nevertheless, this is more than counterbalanced by their incidence and dispersion throughout the genome, such that it has been conservatively estimated that there are sufficient CA/GT microsatellite markers with a Polymorphism Information Content (PIC) of ≥ 0.5 such that genetic maps with an average resolution of approximately 0.3cM are theoretically possible (Weber 1990). The importance of microsatellite markers is already becoming apparent, microsatellite maps are being determined (for example, Hazan *et al.*, 1992), and microsatellite markers linked to disease genes

have been isolated and can be used for predictive diagnosis (for example Haemophilia A, Lalloz *et al.*, 1991, and polycystic kidney disease I, Harris *et al.*, 1991) or for isolation of the disease gene (reviewed by Collins 1992). The larger tri- and tetra-nucleotide repeats are also amenable for use as markers (Edwards *et al.*, 1991), and for individual identification (Edwards *et al.*, 1992).

The use of microsatellite markers for individual identification is limited by the lower variabilities in comparison to minisatellites, although by increasing the number of loci typed the statistical power may be improved such that a positive match may be declared. If loci are chosen carefully, two or more different loci may be amplified simultaneously by 'multiplex' PCR (Chamberlain *et al.*, 1988), expediting the analysis. The small size of microsatellite alleles is their forte, however, when analysing highly degraded DNA such as from forensic specimens. Smaller microsatellite alleles may be faithfully amplified in degraded DNA, and polymorphic microsatellites have been used to identify the skeletal remains of a murder victim, despite the DNA sample extracted from the bone being degraded to an average size of <300 nucleotides (Hagelberg *et al.*, 1991a). DNA has been shown to be recoverable from extremely ancient tissues, such as Egyptian mummies (Pääbo *et al.*, 1985), and mtDNA sequences have been amplified using PCR from 5000 yr old bones (Hagelberg *et al.*, 1989, 1991b ), and 7000 yr old brain tissue (Pääbo *et al.*, 1988). The 'record age' for recovery of ancient DNA is the amplification of chloroplast DNA from a 17-20 million year-old magnolia leaf (Golenberg *et al.*, 1990). The potential to amplify polymorphisms, including nuclear repetitive DNA sequences, from ancient DNA, has important implications for archaeological and anthropological studies.

## 1.2.5 Telomere repeats

Repetitive sequences form part of the DNA-protein complex found at the end of eukaryotic chromosomes that enables complete replication of the linear DNA molecule, mediated by the telomerase enzyme, a ribonucleoprotein reverse transcriptase (reviewed by Blackburn, 1991). Telomeres in higher eukaryotes are highly conserved and composed of heterogeneous arrays of simple tandemly repeated sequences, such as the $(TTAGGG)_n$ repeat which makes up the bulk of human and mouse telomeres, (Kipling and Cooke, 1990). Although similar in sequence, the telomeres of mice are larger (30-150kb) than human telomeres (Allshire *et al.*, 1989), and

have been shown to be extremely variable in length, displaying high levels of mutation to new length arrays (Starling *et al.*, 1990).

Blocks of interstitial telomere repeats are also found (Wells *et al.*, 1990) and other repetitive sequences, such as minisatellites, have been found adjacent to and within the actual telomeric sequences (Royle *et al.*, 1992). These are not believed to have a function in maintaining chromosome stability.

### 1.2.6 Tandemly repeated coding DNA

The sequences discussed thus far do not code for any useful product within the cell, although some, such as telomere sequences may serve a vital role. Other tandemly repetition is found within coding regions of the genome, either as internal tandem repetition, such as found in the collagen genes (reviewed by Vuorio and de Crombrugge, 1990), or as tandem repetition of entire genes to form gene clusters. These clusters may either be of identical genes, for example, the tandem repetition of the eukaryotic rRNA gene clusters (Sollner-Webb and Tower, 1986), or of related genes, such as the globin gene families (Efstradiadis *et al.*, 1980). In these cases tandem repetition has conferred a beneficial function, such as meeting the demand for rRNA molecules, or allowing regulated expression of proteins with similar and related functions (such as the developmental switching of globins), or is a result of structural necessity in the case of collagens. Further variation in the structure of these regions is therefore constrained by selection.

## 1.3    Minisatellites

### 1.3.1 History

The first highly polymorphic region in humans to be described was identified by Wyman and White in 1980 (and shown subsequently to be a tandemly repetitive sequence; Wyman *et al.*, 1986a). Other hypervariable regions were fortuitously identified flanking the human α globin genes (Higgs *et al.*, 1981), the Ha-*ras* gene (Capon *et al.*, 1983) and at the human insulin gene locus (Bell *et al.*, 1982), at which it was first shown, by sequence analysis, that the molecular basis for the polymorphism was change in the copy number of short tandemly repeated sequences, or 'minisatellites'.

The major breakthrough in the study of such sequences came with the discovery that repetitive sequences could be used to detect other hypervariable

minisatellite regions that had similar repeat sequences. Jeffreys *et al.*, in 1985 showed that a set of minisatellites that cross-hybridized to a tandem repeat found near the myoglobin locus, shared, with minor variations, a common 15bp 'core' sequence that bore similarity to the $\chi$ recombination signal of *E. coli*. Two of these loci, 33.6 and 33.15, containing repeats of this core sequence were shown to detect very efficiently other polymorphic loci in human DNA when used as hybridization probes at low stringency (Jeffreys *et al.*, 1985b). The number of loci that cross-hybridized was estimated to be about 1000, although the majority of loci were small and of low variability. However, the largest fragments (4-20kb) detected by Southern blot hybridization of genomic DNA (cleaved with a restriction enzyme that had no target site within the minisatellite repeats) were sufficiently polymorphic such that the pattern of cross-hybridizing bands produced an individual specific 'DNA fingerprint' (Jeffreys *et al.*, 1985b).

## 1.3.2 Applications of multi-locus DNA fingerprinting

The 'DNA fingerprints' produced by the multi-locus core probes were individual specific, with the exception of monozygotic twins, and individual bands showed mendelian inheritance, with the exception of new length mutant alleles. The obvious and immediate application of DNA fingerprints was in the identification of family relationships to resolve paternity disputes (Jeffreys *et al.*, 1985c). The two standard probes, 33.6 and 33.15 (Jeffreys *et al.*, 1985a) detect almost completely different sets of variable loci to produce independent DNA fingerprints, doubling the informativeness when establishing family relationships. Thus, in contrast to conventional genetic markers which have the power only to exclude relationship, the probes have sufficient power of discrimination to positively determine relationship with a low probability of error (Jeffreys *et al.*, 1985c, 1991a)

The high degree of individual specificity has also enabled identification of forensic samples (Gill *et al.*, 1985), and for diverse genetic problems such as following tumour progression (Thein et al., 1987), monitoring bone-marrow transplants (Thein *et al.*, 1986), and for identification of cell lines (Fay and Tobler, 1991).

Human multi-locus 'core' probes, 33.6 and 33.15, have been shown to cross-hybridize to a variety of variable loci in mammals (Jeffreys and Morton, 1987, Jeffreys *et al.*, 1987a), birds (Burke and Bruford, 1987, Wetton *et al.*, 1987) and even plants (Dallas, 1988, Rogstad *et al.*, 1988). In addition to these 'core'

probes other sequences have been found which detect hypervariable loci in humans and other animals. These include a sequence derived from the bacteriophage M13 (Vassart *et al.*, 1987), the α-globin 3'HVR (Fowler *et al.*, 1988), and various synthetic oligonucleotides comprising simple repeat motifs (Ali *et al.*, 1986, Vergnaud *et al.*, 1991, Epplen *et al.*, 1991), and an extremely broad range of animal and plant species are currently being studied using DNA fingerprinting probes (Amos, 1991), typically for the establishment of family relationships in wild populations, and maximising outbreeding in captive animals.

### 1.3.3  Cloning of minisatellites

Several approaches have been successful in isolating minisatellite loci, Aside from the first accidental isolation of polymorphic sequences (see section 1.2.1), the most successful strategies have used hybridization screening with multi-locus DNA fingerprint probes to isolate clones of individual loci, either by cloning selected alleles from a DNA fingerprint pattern (Wong *et al.*, 1986, 1987), or by screening cosmid libraries (Nakamura *et al.*, 1987a, 1988b, Armour *et al.*, 1990, Hanotte *et al.*, 1991). The majority of such clones isolated can then be used as locus specific probes in high stringency hybridizations (such that the probe will only anneal to its complementary sequence), or as multi-locus probes at low stringency conditions (which allow a degree of mismatch between the cross-hybridizing sequences so that a range of different loci similar in sequence may be detected). In the latter case novel loci may be detected within the set of cross-hybridizing loci, which if cloned and used as multi-locus probes may in turn detect further novel loci.

Tandem repeats are not confined to non-coding DNA, some coding sequences are also repetitive, for example the *per* gene repeat of *Drosophila*, and collagen genes. In general, however, it has been found that variable minisatellite loci are non-coding. The notable exception is the hypervariable MUC1 locus which encodes the human polymorphic epithelial mucin (PEM) (Swallow, *et al.*, 1987, Lancaster *et al.*, 1990).

### 1.3.4  Applications of single locus minisatellite probes

Single locus probes each detect a single minisatellite locus, and give a DNA 'profile' with either two bands per heterozygote (assuming both alleles are separated by agarose gel electrophoresis), or one band per homozygote.

Multi-locus probes require at least 500ng of genomic DNA for a 'DNA fingerprint' to be detected, which limits the use of such probes for forensic work. The use of single locus probes, which require as little as 10ng of genomic DNA, overcomes to some extent this sensitivity limitation, (Wong *et al.*, 1987), and a battery of single locus probes chosen for their high informativeness (the 6 loci cloned by Wong *et al.*, 1987 had heterozygosities from 90-99%) maintains the discriminatory power of a 'DNA fingerprint', although the power to distinguish close relatives is more limited due to the small number of loci typed. It is possible to further increase sensitivity to enable the typing of subnanogram quantities of DNA by amplifying minisatellites using the polymerase chain reaction (PCR, Saiki *et al.*, 1988). PCR primers complementary to the unique sequence DNA which flanks minisatellite repeats can be used to amplify alleles of up to 10kb to the point where they are detectable by Southern blot hybridization. With larger cycle numbers, alleles of up to 6kb can be directly visualized on ethidium stained agarose gels (Jeffreys *et al.*, 1988b), although at this stage the alleles are prone to collapse producing complex artifactual patterns.

Unlike DNA fingerprints, which give phenotypic information, the single locus pattern gives a genotype - information which can be used in linkage analysis. Informative genetic markers are vital for the construction of genetic maps (Donis-Keller *et al.*, 1987) and for the 'reverse genetics' or 'positional cloning' approach to identifying single gene defects. Since minisatellite length variations are insertion/deletion polymorphisms they are detectable by a wide variety of restriction enzymes, unlike polymorphisms that result from point mutations creating or destroying the restriction site for a single enzyme. This, together with the high polymorphic information content (PIC) which reflects the number of distinguishable alleles and their population frequency, makes minisatellite loci excellent markers for the investigation of human and animal disease. For example, the minisatellite cMS620 (D15S86, Armour *et al.*, 1990) has been used as a marker to study Angelmans syndrome (Malcolm *et al.*, 1991). Minisatellite loci have been identified on all chromosomes, but are not randomly distributed in humans; human minisatellites are most frequently found towards the ends of chromosomes (Royle, *et al.*, 1988; Vergnaud *et al.*, 1991), which limits their usefulness as markers. In contrast, in mice (Jeffreys *et al.*, 1987a) and cattle (Georges *et al.*, 1990) there appears to be no significant terminal clustering. Thus minisatellites are still a key element in genetic maps of domestic animals; minisatellite (and particularly microsatellite) markers are being isolated with

the aim of finding linkage to disease genes and economically important traits in livestock (for example, the European pig gene mapping project (PiGMaP), Archibald *et al.*, 1991).

## 1.3.5 Mutation at minisatellites

Minisatellite loci show heterogeneous variability; heterozygosities range from 0 to almost 100%. In the absence of selection, this variability reflects the mutation rate to new length alleles, which has been measured for some human minisatellites by pedigree analysis (Jeffreys *et al.*, 1988a, Armour *et al.*, 1989a). It was found that high heterozygosities were associated with high germline mutation rates, in agreement with the random-drift hypothesis for selectively neutral mutations (Kimura, 1983). The male and female germcells follow different developmental pathways (there are about 17 times more pre-meiotic divisions in the generation of a sperm (Vogel & Rathenberg, 1975)), and in a sample of 79 mutant bands detected at the human loci MS1, MS8, MS31 or p$\lambda$g3 there was a significant but small excess (64%) of paternally derived mutations (Jeffreys *et al.*, 1991a). At some other characterized loci this bias is extreme; the minisatellite CEB1 has a mutation rate of 15% in the male germline, but less than 1% in the female germline (Vergnaud *et al.*, 1991), and paternal bias has been observed at the murine minisatellite *Ms6-hm* (section 1.4.3.). The circumstances which produce these differing sexual biases await further investigation.

Gray and Jeffreys (1991) have suggested that the most unstable loci may arise very rapidly; the locus MS32 which has a heterozygosity of 97% in humans (Wong *et al.*, 1987) was found to have only 2 repeats in chimpanzees and apes, suggesting that the region has become hypervariable since the evolutionary divergence of man and the other primates. Similarly, computer simulations (Gray and Jeffreys 1991) have predicted that the most unstable loci are highly transient; starting with a system of two repeats which undergo random gain and loss of repeats through mutation it was shown that in a minority of cases (1 in 250) a high copy number, high heterozygosity locus was generated within a few thousand generations; equivalent to the situation observed at MS32. The simulation also predicted the eventual 'decay' of hypervariable loci back to a ground state of one repeat unit. In contrast, it has been demonstrated that a low variability human minisatellite does share common features in man and apes (Armour *et al.*, 1992), therefore stable minisatellites may 'survive' for millions of years.

Thus although multi-locus minisatellite probes detect a similar variety of variable loci in other animals (section 1.3.2) these are not necessarily exactly homologous groups of loci, even between closely related species. Some single-locus probes can detect a variable single-locus pattern in different species (Hanotte *et al.*, 1991, 1992), however, a consequence of the short lived nature of hypervariable minisatellites is that a hypervariable single locus probe from one species may detect only a monomorphic region in another species (Gray and Jeffreys 1991), thus for mapping purposes it may be necessary to isolate single locus probes for each different species.

## 1.4    Minisatellites in mice

### 1.4.1    *Detection by 'core' probes*

Human minisatellite probes cross-hybridize to mouse DNA and detect variable loci giving a DNA fingerprint similar in complexity to that in humans, although the DNA fingerprints of inbred mice strains differ relatively little (Jeffreys *et al.*, 1987a). In contrast to human minisatellite loci, murine minisatellites do not appear to be biased towards telomeric or sub-telomeric regions (Jeffreys *et al.*, 1987a, Julier *et al.*, 1990).

### 1.4.2    *Applications of polymorphic murine minisatellite loci*

Just as in humans, minisatellites are an abundant source of polymorphic loci for linkage studies in inbred mouse strains, an important experimental animal for many lines of research, including localization of mouse models of human genetic diseases (reviewed by Todd, 1992). For example, linkage of minisatellite markers to diabetes susceptibility genes in NOD mice (Todd *et al.*, 1991). This is aided by the apparent non-clustering of murine minisatellites.

In their own right, variable mini- and microsatellites giving individual or strain identification are of use as a means of genetic quality control to detect contamination of inbred lines (Russell and Deeny, 1992), and a characterized system of polymorphic markers may be of practical use for tracing cellular lineages and cell fate in aggregation chimeras.

There is much left to discover of the mechanisms by which minisatellites mutate and evolve; murine minisatellites provide an accessible system in which to study the genetic behaviour of individual loci, including

the processes involved in the generation and maintenance of variability at minisatellites in general.

### 1.4.3 Ms6-hm: *A highly variable mouse minisatellite.*

One minisatellite loci, called *Ms6-hm*, detected in murine DNA by the human probe 33.6 was seen to have extreme germline instability, resulting from a high mutation rate to new length alleles (Jeffreys, *et al.*, 1987a).

The *Ms6-hm* locus was cloned, sequenced, and mapped to an interstitial location on chromosome 4 (Kelly *et al.*, 1989). The repeat sequence $(GGGCA)_n$ had expanded from within a member of the MT SINE family (Mouse Transcript, Heinlein *et al.*, 1986), and the locus was further flanked by two other diverged MT elements and a B2 element (Kelly *et al.*, 1991). In turn, the *Ms6-hm* probe cross-hybridizes to other loci at low stringency, detecting a novel set of loci, one of which appeared to show similar features of high germline and somatic mutation rate. *Ms6-hm* was highly variable, even within inbred strains, and highly unstable; the germline mutation rate to new length alleles was measured directly in a large pedigree at 2.5% per gamete. These mutations showed a strong paternal bias, in contrast to most characterized human loci (section 1.3.5).

Most interestingly, a proportion of mice showed evidence for mosaicism for cells containing either the parental, or a common new length allele, resulting from a somatic mutation event. The degree of mosaicism found (10-60% of cell were mutant) and the uniform dispersion of mutant cells throughout the mosaic animals suggested that the mutation must have occurred during early development; in some mice the mosaicism extended to the germline, producing three-way segregation of the two parental and the mutant alleles.

### 1.5    Objectives of this work

Mice provide an accessible system in which to study the genetic behaviour of individual loci, with the advantages of genetically defined inbred strains and mouse breeding systems. By the isolation and comparison of different loci it may prove possible to determine which factors affect variability, and to discover more about the mechanisms that generate new length alleles. Ideally, for the latter purpose, a system which shows large

amounts of variability would be chosen. One such extremely variable locus (*Ms6-hm*) has already been characterized (Kelly *et al.*, 1989, 1991). This minisatellite was intriguing in that it showed a somatic mutation rate far in excess of any other characterized locus. An initial objective of this work was to clone and characterize a second hypervariable mouse minisatellite, and to further investigate the phenomenon of somatic mutation.

# 2
# Methods and Materials

## 2.1   Materials

### 2.1.1 Chemical Reagents

All chemical reagents, unless otherwise stated below, were supplied by Fisons, Loughborough. Antibiotics, BSA, HEPES, IPTG, polyethylene glycol (PEG), ficoll 400, TEMED, dithiothreitol (DTT), thymine, spermidine trichloride, salmon sperm DNA and agarose were supplied by the Sigma Biochemical Company, Poole. SeaKem™ HGT agarose and NuSieve™ agarose were supplied by ICN Biochemicals Ltd, High Wycombe. Ammonium persulphate was from Bio-Rad, Watford. BCIG was obtained from Anglian Biotechnology, Colchester. Acrylamide and urea were supplied by Serva, Heidelberg. N,N'-methylene-bisacrylamide was supplied by Uniscience, Cambridge. Deoxyribonucleotides and dideoxynucleotides were from Pharmacia, Milton Keynes. Radiochemicals were supplied by Amersham International plc, Little Chalfont. All reagents used were of analytical grade.

### 2.1.2 Oligonucleotides

Hexadeoxyribonucleotides for random oligonucleotide priming (section 2.4.1) were supplied by Pharmacia. Oligonucleotides for PCR amplification and the M13 sequencing 17-mer were synthesized in the Department of Biochemistry, University of Leicester on an Applied Biosystems 380B DNA synthesizer, using reagents supplied by Cruachem.

### 2.1.3 Enzymes

Restriction endonucleases were supplied by Gibco-BRL, Paisley, New England Biolabs (via CP Laboratories, Bishop's Stortford) and the Boehringer Corporation, Lewes. T4 ligase and REact™ buffers obtained from Gibco-BRL. Pharmacia supplied T4 polynucleotide kinase, T7 DNA polymerase and the Klenow fragment of DNA polymerase I. Calf intestinal alkaline phosphatase

(CIP) was obtained from the Boehringer Corporation. DNA polymerase from *Thermophilus aquaticus* was from Amersham International. Gigapack Plus™ *in vitro* packaging kit supplied by Stratagene Cloning Systems, CA, USA.

## 2.1.4 Molecular weight markers

λ DNA digested with *Hind*III and ΦX174 DNA digested with *Hae*III were obtained from Gibco-BRL.

## 2.1.5 Media

All media were obtained from Oxoid, Basingstoke, with the exception of yeast extract and tryptone, which were supplied by Difco, East Molesley.

LUB (Luria broth) consists of 10g bacto-tryptone, 5g bacto yeast extract and 5g NaCl per litre of dH$_2$O.

## 2.1.6 Bacterial strains

Two strains of *Escherischia coli* were used:

NM522    Δ(*lac-pro*AB), *thi⁻, sup*E, *hsd*R17(Rk⁻,Mk⁺), [F' *proAB*, lacI$^q$ZΔM15].

DH5a    F-, *end*A1, *hsd*R17(RK⁻,Mk⁺), *sup*E44, *thi*1, λ-, *rec*A1, *gyr*A96, *rel*A1, (*arg*F-*lacZ*YA) U169, Φ80d*lacI*$^q$, ZΔM15.

## 2.1.7   Cloning vectors

Charomid vectors (Saito and Stark, 1986) were provided by the Japanese Cancer Research Resources bank, Tokyo. M13 sequencing vectors M13mp18 and M13mp19 (Yanisch-Perron *et al.*, 1985)   replicative forms supplied by Gibco-BRL.   pUC18 and pUC19 (Yanisch-Perron *et al.*, 1985) were used for subcloning.

## 2.1.8   Mice and mouse DNA

C57BL/6J and DBA2/J inbred mice were received from Dr G. Bulfield (AFRC Institute of Animal Physiology and Genetics Research, Edinburgh), and were supplied to Edinburgh directly from The Jackson Laboratory, (Bar Harbor, ME, USA).   A, AKR, BALB/c, C3H/He, CBA, DBA2/J and SWR inbred mice

were obtained from Bantam and Kingman Ltd (Hull, England); the DBA/2J mice had originally been obtained from the Jackson Laboratory via Searle Ltd. C57BL/6J mice and CBA mice were also purchased from Harlan Olac, Bicester, England.

BXH and BXD RI series (and progenitor strain) DNAs were provided by Dr G. Bulfield (Edinburgh) and Dr. B. Taylor (The Jackson Laboratory).

## Methods

### 2.2 Mice

### 2.2.1 Housing

Mice were housed in the biomedical services of Leicester University. The mouse colony was maintained under a 12hr light/12hr dark cycle.

### 2.2.2 Identification

The existing Leicester pedigree colony had been marked by toe clipping following the numbering system of Allen et al., 1987. No more than two toes from each of two paws were clipped from any animal: from 0-200 the hundreds and tens were marked on the forepaws and the units on the right hindpaw. from 200-400 the units (multiples of ten omitted) were marked on the left hindpaw. The majority of mice used for the production of day 10 embryos were unmarked physically since once plugs were detected they were housed individually in separate cages. Some stud males were marked by an ear punch against the possibility of accidental mixing. Both toe clipping and ear punching were done under anaesthetic (halothane gas).

### 2.2.3 Procedures

#### 2.2.3.1.    Isolation of a length of tail

The mice were anaesthetised with halothane and 1-2 cm of tail severed with a scalpel. The tails were cleaned with sterile gauze and cauterized to prevent bleeding. Toe clipping or ear punching for identification was carried out in the same operation.

2.2.3.2.    Isolation of Day 10 embryos

Females were checked for vaginal plugs (day 1 *post coitum*, p.c.) on successive mornings after mating were set up. On day 10 p.c. gravid females were killed and the uterus dissected out into a petri dish containing ice cold PBS (phosphate buffered saline). Embryos were dissected from the uterus wall into drops of PBS using watch-makers forceps, and then further separated by pulling the decidua apart and gently removing the embryo. The trophoblast enclosing the embryo proper was then split and the embryo proper with associated visceral yolk sac exposed. These were then nipped away from the trophoblast into a further droplet of PBS and finally the embryo and the yolk sac were detached from each other. No effort was made to separate the yolk sac further into endoderm and mesoderm layers since the quantity of tissue was small. The trophoblast tissue was further scrutinized for adhering maternal cells before all three tissues were transferred to individual eppendorf tubes.

## 2.2.4  Preparation of mouse DNA

### 2.2.4.1    Tail DNA

Tails were finely chopped with a scalpel and suspended in 0.5ml 1xSE buffer (150mM NaCl, 100mM EDTA pH8.0) in a 1.5 ml eppendorf tube. To this was added 50$\mu$l of 10% SDS and 5$\mu$l of proteinase K (20mg/ml). The contents were mixed by inversion and incubated at 50$^\circ$C for 3-5hrs with occasional mixing. DNA was separated from protein and bone by two phenol/chloroform extractions (see section 2.3.1), and one chloroform extraction. DNA was precipitated from the aqueous layer by the addition of isopropanol, pelleted, rinsed with 80% ethanol and suspended in 100$\mu$l of distiled water.

### 2.2.4.2    Tissue DNA

Small scale tissue preparations were carried out as for tail except that the initial breaking up of tissue was done with a small hand homogeniser which reduced the time necessary for complete proteinase K digestion to 1-2hrs.

Dissected embryo trophoblast and yolk sac were broken up by pipetting through a 200$\mu$l tip and suspended in 200$\mu$l of TES (0.5% SDS, 10mM TrisCl, 100mM EDTA, pH8.0). 2$\mu$l of proteinase K (20mg/ml) were mixed into the suspension which was incubated at 50$^\circ$C for 3hrs. The digested tissues were

extracted once with phenol/chloroform and once with chloroform, precipitated with ethanol and rinsed in 80% ethanol. Pellets were air dried and suspended in 30μl dH$_2$O or 20μl dH$_2$O (yolk sac DNA).

Large scale tissue preparations followed a protocol modified from Maniatis *et al.*, (1982). The volume of tissue was estimated, and 10 volumes of 1xSE added. The tissue was then homogenised and transferred to a clean conical flask. 1/10 vol 10% SDS was added to lyse the cells, and the solution emulsified with 0.5 vol of phenol/chloroform by gentle swirling of the flask. This mixture was decanted to centrifuge tubes and spun at 10,000rpm for 5 minutes at room temperature to separate the aqueous and organic layers. The upper aqueous layer was removed using wide mouthed pipettes into a glass beaker, and the remaining layer back-extracted with dH$_2$O and pooled with the first. DNA was precipitated from this mixture by ethanol precipitation, rinsed in 80% ethanol, and resuspended in half the original volume of 0.1xTNE (1xTNE is 10mM TrisCL, 100mM NaCl, 1mM EDTA pH8.0). Pancreatic RNase to 20μg/ml was added to the resuspended DNA and the beaker incubated for 30 minutes at r.t. To remove the RNase, SDS was added to 1%, TNE to 1x, and the DNA re-extracted with phenol/chloroform. The DNA was than cleaned by ethanol precipitation with a final rinse in 80% ethanol. The pellet was allowed to dissolve in a small quantity of dH$_2$O in an eppendorf o/n at 4°C.

## 2.3    General Methods for DNA Handling

### 2.3.1    Phenol/chloroform extraction

Phenol/chloroform and chloroform was used to purify DNA from proteins during the initial extraction process and as an additional purification step following some enzymatic treatments. An equal volume of phenol/chloroform (Phenol:Chloroform:Isoamyl alcohol in the ratio 25:24:1. Phenol with hydroxyquinone added to 0.1%, equilibriated with TrisCl pH 8.0) was added to the nucleic acid sample and the mixture emulsified. The mixture was centrifuged to separate the phases and the upper aqueous layer transferred to another tube leaving behind the proteinacious interface. Further phenol/chloroform extractions were performed until the interface was clean and the aqueous layer clear. An extraction with chloroform was used to remove traces of phenol, then the DNA was ethanol precipitated. In situations where high molecular weight DNA was needed the emulsifying step used a

slowly rotating wheel to avoid shearing the DNA and large mouthed pipettes were used to transfer the DNA from one tube to another.

## 2.3.2 Ethanol precipitation

In order to concentrate, de-salt or recover DNA following manipulation, ethanol precipitation was used. 1/10 volume of 2M sodium acetate (pH 7) and 2 to 2.5 volumes of ethanol or 1 volume of isopropanol were added to the DNA solution. The solution was then mixed and chilled to precipitate the DNA, which was then pelleted by centrifugation in a bench top microfuge for 10 minutes at high speed (12,000rpm). The pellet was then rinsed in 70-80% ethanol, dried, and dissolved in the required amount of distiled water.

## 2.3.3 Estimation of DNA concentration

DNA concentration was assayed by measurement of UV absorbence at a wavelength of 260nm in a Cecil Instruments CE235 spectrophotometer, given that 0.02 OD units represents 1mg/ml of DNA (or 0.05 units for 1mg/ml of oligonucleotides). Smaller amounts of DNA, such as individual restriction digests, were estimated by visual comparison of aliquots of the samples after agarose gel electrophoresis and ethidium staining. In later experiments such estimates were confirmed using a DNA fluorimiter (TKO 100, Hoefer Scientific Instruments, San Francisco, USA).

## 2.3.4 Restriction endonuclease digests

Restriction digests were performed in the appropriate REact buffer (Gibco-BRL) at the temperature recommended by the suppliers. Incubations were typically for 1-5 hours with 1unit of enzyme per μg of DNA, or since the majority of enzymes commonly used maintained their activity over a longer time, complete digestion was achieved using <0.5u in an overnight incubation.

## 2.3.5 Agarose gel electrophoresis

Agarose gels in the concentration range 0.5 - 3% were used, depending on the size of DNA fragments to be resolved. 0.5 - 2% gels were cast from Sigma agarose or SeaKem HGT (FMC ICN Biomedicals). Gels for resolving

fragments of around 100bp used 3% NuSieve agarose (ICN Biomedicals). Gels were run in TAE (40mM tris-acetate, 20mM Na acetate, 0.2mM EDTA, pH8.3), ELFO buffer (40mM tris-acetate, 1mM Na acetate), or for gels requiring good resolution, TBE (0.089M tris-borate, 0.089M boric acid, 2mM EDTA). Ethidium bromide was added to all buffers to a concentration of 0.5mg/ml.

Gel length varied from 10-40cm, depending on the separation required. Loading mix was added to DNA samples prior to loading; 5x loading mix comprised 12.5% ficoll 400, 0.1% bromophenol blue in 5x TAE. DNA samples were run alongside markers of known molecular weight, either λ DNA cut with HindIII (range 23-2kb) or ΦX174 DNA cut with HaeIII (range 1.3kb to 70bp). Current applied to the gel was adjusted to suit a particular run, monitoring of the run could be achieved by following the migration of the bromophenol blue, or by observation of DNA under UV light from a hand held UV wand (Chromato-vue 57, UV Products Inc.). For photography, DNA was visualized by UV fluorescence on a transilluminator (Chomato-vue C-63, UV Products Inc., California), and photographed with a Polaroid MP-4 camera using Kodak negative film (T-max Professional 4052). Films were processed with Kodak LX24 developer, FX40 fixer and HX40 hardener.

## 2.3.6  Preparative gel electrophoresis

To selectively obtain and concentrate DNA fragments of a certain size prior to ligation, random oligonucleotide labelling etc., preparative gel electrophoresis (Yang et al., 1979) was used. The DNA samples were run in an agarose gel until adequate separation had been achieved. Dialysis membrane (Scientific Industries International Inc., Loughborough) was cut into sheets just wider than the gel slots and boiled in 10mM TrisCl pH 7.5, 1mM EDTA for 10 minutes. The gel was viewed using a hand-held UV source (Chromato-vue 57, UV Products Inc.) and an incision made with a scalpel immediately ahead of the band or smear to be purified. The dialysis membrane was inserted into this cut, and the DNA run onto the membrane at 10-15 volts/cm. An extra membrane was inserted behind the selected DNA to prevent the migration of unwanted higher molecular weight DNA when necessary. When all of the DNA was loaded onto the membrane (as judged by eye), the incision was extended across the gel to loosen the DNA side of the membrane from the agarose. With the current on, the membrane was rapidly removed to a 1.5ml microfuge tube using forceps. The DNA was collected by spinning the tube with the corner of the membrane held by the lid of the tube, and concentrated

by ethanol precipitation. This method allowed the recovery of small amounts of DNA with good yield. A modified version of this method was used in later experiments: the DNA was run as described previously, but then the block containing the desired size DNA was excised and placed into a slightly longer, precut slot into which the membrane had been lain. The membrane was then folded over the top of the agarose block and the DNA run onto the membrane as before.

### 2.3.7 Southern blotting (Southern, 1975).

The positions of the DNA markers were visualized by UV fluorescence and marked by notching the the sides of the gel to be blotted. Gels were shaken gently in 0.25M HCl for 2 x 7 minutes (depurination), 0.5M NaOH, 1M NaCl for 2 x 15 minutes (denaturation) and 0.5M TrisCl pH7.5, 3M NaCl for 2 x 15 minutes (neutralization), then placed on a wick of 3MM blotting paper (Whatman International Ltd., Basingstoke) previously soaked in transfer buffer over a reservoir of buffer. A pre-wetted membrane was than placed over the gel, followed by one or two sheets of wetted 3MM and a stack of blotting towels (either Kleenex paper towels or, in later experiments, 'Quick-draw'™ blotting paper (Sigma)). Weight was applied to the top of the stack via a glass plate to ensure even pressure onto the surface of the gel. Blots were left from 5 to 16 hours, longer times being used for the transfer of large fragments.

Membranes used were either Hybond-N™, (Amersham), or Immobilon-N™, (Millipore Corporation, Bedford, MA). For Hybond-N transfer solution used was 20xSSC (1x SSC is 0.15M NaCl, 15mM tri-sodium citrate pH 7), membrane prewetted in 3xSSC. After transfer the membrane was rinsed in 3xSSC and air dried. DNA was fixed using a UV transilluminator. For Immobilon-N, 10xSSC was used for transfer. Immobilon-N was prewetted in 100% ethanol, distiled water and equilibriated in 10xSSC for 15-30 minutes before use. Following transfer the membrane was immersed in 6xSSC for 5 minutes then air dried to bind DNA to the membrane.

## 2.4 DNA hybridization

### 2.4.1 Random oligonucleotide labelling of DNA fragments (Feinberg & Vogelstein, 1984)

5-10ng of gel-purified DNA were labelled as follows. The DNA was boiled for 3 minutes then cooled to 37°C, before adding to 1.2μl of 10mg/ml BSA (enzyme grade, Pharmacia) and 6μl of oligo-labelling buffer (OLB). OLB consisted of a mixture of 3 solutions: solution A: 1.25M TrisCl (pH 8), 125mM MgCl$_2$, 0.18% v/v 2-mercaptoethanol, 0.5mM dATP, 0.5mM dGTP, 0.5 mM dTTP; Solution B: 2M HEPES (pH 6.6) and solution C: hexadeoxyribonucleotides (Pharmacia) suspended in 3mM TrisCl, 0.2mM EDTA (pH 7.0) at 90 OD units/ml. Solutions A B and C were mixed in the ratio 2:5:3 respectively to give OLB. Following addition of OLB, 1-5μl of [α$^{32}$P]dCTP (1000Ci/mmol, Amersham) and 2.5 units of the Klenow fragment of DNA polymerase I (Pharmacia) were added. The reaction mixture was incubated for 1 hour at 37°C or overnight at room temperature. After labelling, 70μl of 'stop solution' (20mM NaCl, 20mM TrisCl pH 7.5, 2mM EDTA, 0.25% SDS) was added to the reaction, followed by 100mg of high molecular weight herring sperm DNA which acted as a carrier. The probe was recovered by ethanol precipitation, washed in 80% ethanol to remove unincorporated [α$^{32}$P]dCTP, and redissolved in 500μl distiled water. Probes were boiled for 3 minutes immediately prior to use.

### 2.4.2 Hybridization

Hybridization reactions were in either Denhardt's solution (0.2% ficoll 400, 0.2% polyvinyl pyrrolidone - PVP, 0.2% BSA in 3xSSC; Denhardt, 1966) for colony screening following plate lifts onto nitrocellulose filters or phosphate/SDS solution (0.5M sodium phosphate, pH 7.2, 1mM EDTA, 7% SDS; Church & Gilbert, 1984) for nylon membranes following Southern transfer. All hybridizations were performed in sealed perspex chambers in a shaking water bath.

2.4.2.1     Hybridizations in Denhardt's solution (modified by Jeffreys *et al.*, 1980).

Filters were pre-hybridized for 30 minutes in complete filter hybridization mix (CFHM: 1x Denhardt's solution, 0.1% SDS). Filters were then transferred to a minimal volume of CFHM + 6% polyethylene glycol (PEG) and the boiled probe and hybridized overnight at 65$^o$C with gentle shaking. Following hybridization, filters were washed at low stringency as described in section 2.4.3.

2.4.2.2     Phosphate/SDS hybridization

Filters were pre-hybridized in 0.5M sodium phosphate, pH 7.2, 1mM EDTA, 7% SDS for at least 15 minutes, and then hybridized in a fresh aliquot of the same solution with the boiled probe. Hybridizations to genomic DNA were at 65$^o$C with gentle shaking overnight.

2.4.3  Post-hybridization washing

Following hybridization, filters were washed at a stringency to suit the requirements of the experiment. These were usually either low stringency (1x SSC, 0.1% SDS, at 65$^o$C, or high stringency (0.1x SSC, 0.01% SDS, at 65$^o$C). Occasionally an intermediate stringency (0.5x SSC, 0.05-0.01% SDS, at 65$^o$C),was used when identification of a specific locus was required, but with the inclusion of other cross hybridizing loci to act as internal controls of DNA concentrations and 'gel wobble'. Washes were performed until the level of radioactivity stabilized, as monitored by a hand-held Geiger counter (Series 9000 mini-monitor, Mini Instruments Ltd., Essex). Later experiments used Immobilon-N (Millipore), which was washed at very high stringency, (0.05xSSC, 0.01%SDS at 68$^o$C).

2.4.4  Autoradiography

Filters were placed in autoradiographic cassettes separated from a sheet of Fuji RX100 X-ray film by a sheet of aluminium foil or clingfilm. Exposures were at room temperature or -80$^o$C with an intensifying screen for 1 hour to 14 days depending on the strength of the signal.

## 2.4.5   Densitometric Scanning

Densitometric scanning was carried out with an Ultrascan XL Laser Densitometer (LKB), with preflashed film which was exposed without intensifying screens.

## 2.4.5   Stripping filters for rehybridization

Nylon Hybond-N filters required for rehybridization with a different probe were stripped by immersing in boiling 0.1% SDS, which was then allowed to cool to room temperature on a shaking platform (manufacturers recommendations).

## 2.5   DNA sequencing

### 2.5.1   M13 subcloning

#### 2.5.1.1      Preparation of M13 vectors

M13mp18 or 19 vectors were prepared by linearization with an appropriate unique restriction enzyme present in the polylinker sequence. Alkaline phosphatase treatment was used to prevent self ligation of vectors cut with a single enzyme. Approximately (1 μl of 1/10 diluted) 0.01 units of calf intestinal alkaline phosphatase per picomole of 5' ends of DNA were added directly to the reaction mix and incubation continued at 37$^{\circ}$C for 30 minutes. The CIP was then deactivated by heating to 68$^{\circ}$C for 10 minutes (modified from Maniatis *et al.*, 1982). The efficiency of this treatment was  tested in a self ligation reaction: vectors which gave neither blue colonies or white colonies were ideal.  Vectors with two different cohesive ends were generated by two separate digestions.  The first digestion was set up as usual, and an aliquot checked for complete digestion to linear form.  The DNA was then cleaned and reconcentrated by ethanol precipitation before further digestion at the conditions required for the second enzyme.  Complete digestion could be confirmed by checking an aliquot of each double digest against the single digest; for the double digests the size of the large M13 linear fragment is reduced and the small fragment of excised polylinker can be seen.  To prevent religation of this polylinker fragment in subsequent ligation reactions the digested vector is excised from a preparative gel (see section 2.3.6).

### 2.5.1.2        Ligations

Ligation of DNA fragments into cloning vectors (M13mp18 or M13mp19) was performed as follows: a 2:1 molar ratio of vector to insert was mixed at low concentration ($\leq$30mg/ml) in the presence of 1mM ATP, 1x ligase buffer (50mM TrisCl pH 7.5, 10mM $MgCl_2$, 10mM DTT), 4mM spermidine and 0.1 Weiss units of T4 ligase per ml of reaction mix. Ligation reactions containing DNA molecules with 'sticky ends' were incubated at 15$^O$C overnight, whereas blunt-ended ligations were left at 4$^O$C overnight or longer.

### 2.5.2   Introduction of M13 into *E. coli*  by electroporation (Dower *et al.*, 1988)

### 2.5.2.1        Generation of competent cells

An overnight culture of *E. coli* was diluted 1/100 into fresh LUB. The fresh culture was incubated until cells were dividing logarithmically ($A_{600}$ of between 0.5 and 0.7, depending on the strain used).

Cells were harvested by centrifugation at 4,000rpm for 15mins at 4$^O$C, then resuspended in the original volume of ice cold 10% glycerol.

Cells were spun at 4000rpm for 15 minutes at 4$^O$C, and suspended in 1/2 volume of 10% glycerol.  After a further spin as previous, the cells were suspended in 1/50th volume of 10% glycerol, and after the final centrifugation (either as previous, or when volumes were small, in a chilled microcentrifuge at low speed for <2mins) the cells were resuspended in 1/500th volume of 10% glycerol.  This protocol gave 40µl of concentrated and washed cells for every 20mls of starting culture.

### 2.5.2.2        Transformation and plating

40µl of cells plus 1-2µl ligated DNA were subjected to an exponential pulse of 1.5kV (with a capacitance of 25mF and a parallel resistance of 940W) in a cuvette with a 2mm electrode gap, in a Bio-Rad Gene Pulser (The time constant was approximately 20 milliseconds). Immediately the pulse was delivered 1ml of SOC (2% w/v Difco bacto tryptone, 0.5% w/v yeast extract, 10mM NaCl, 10mM $MgCl_2$, 10mM $MgSO_4$, 20mM glucose; Hanahan, 1983) was added to the cuvette and the contents transferred to a test tube and shaken at 37$^O$C for 10 minutes.

300 µls plating mix (25µl 50mg/ml BCIG, 25µl 50mg/ml IPTG, 250µl plating cells) and 3mls of top agar (LUB with 7g/l agar) were added to each test tube before spreading onto LUA or BLA plates. Plates were incubated at 37°C overnight.

N.b. The electroporation of bacteria with plasmid or cosmid vectors was performed exactly as for M13 to the point at which SOC is added to the pulsed cuvette. After transformation the cells were incubated at 37°C for 30 minutes before plating onto LUA plates containing selective agents. (For charomids these are 25mg/ml ampicillin, 12.5mg/ml tetracycline, 25mg/ml BCIG, 25mg/ml IPTG.) Plates were incubated at 37°C overnight.

### 2.5.3   Library screening

Plate lifts were onto either nitrocellulose filters (Hybond-C, Amersham) by the method of Benton & Davis, 1977, or nylon filters (Hybond-N) by the method of Bulawela *et al.*, 1989. A filter was placed on the surface of each plate using sterile forceps. Holes were punched around the edge of the filter with a sterile needle for purposes of orientation. After 5 minutes nitrocellulose filters were carefully lifted off the plate and floated 'DNA side' down in 1.5M NaCl, 0.1M NaOH for 1 minute. Filters were then submerged in 2xSSC, 0.2M TrisCl (pH 7.5) for 1 minute, blotted dry on 3MM paper and then baked at 80°C for 3-4 hours prior to hybridization. Nylon membranes were transferred from the plate to a piece of 3MM moistened in 2xSSC, 5% SDS for 2 minutes, then the tray of filters was microwaved for 2.5 minutes at 650W setting in a standard microwave oven with a rotating turntable to lyse the cells and fix the DNA to the filter. Filters were then rinsed in 5xSSC 0.1%SDS to remove debris before transferring to a prehybridization solution (see section 2.4).

Positive plaques were picked from the library plates using a sterile pasteur pipette and deposited in 1ml of phage buffer (6mM tris-HCl pH 7.2, 10mM MgSO4, 0.005% gelatin).

### 2.5.4   Preparation of single-stranded DNA from M13 phage

0.5ml of a culture of NM522 which had been grown at 37°C overnight in LUB was diluted 1 in 100 into fresh LUB. 1.5ml of the diluted NM522 was mixed with 0.1ml of phage suspension in a large capped test-tube and shaken vigorously for 5-6 hours at 37°C. The culture was transferred to a microfuge tube and spun for 10 minutes at 15,000 rpm. 1ml of supernatant was

transferred to a fresh tube, and the phage precipitated by the addition of 200 µl of PEG 6000, 2.5M NaCl followed by incubation at 4ºC for 15 minutes. Phage particles were pelleted by centrifugation at 15,000 rpm for 10 minutes. Following removal of all traces of supernatant, the phage particles were resuspended in 100 µl of 1.1M sodium acetate (pH 7.0). The suspension was emulsified with an equal volume of phenol/chloroform (phenol: chloroform: isoamyl alcohol and 8-hydroxy-quinolene in the ratio 100:100:4:1, saturated with tris-HCl pH 7.5), and spun at 15,000 rpm for 5 minutes to separate the protein component of the phage particles. The upper aqueous layer was removed, and to maximise recovery, the remaining phenol layer was re-extracted with a second aliquot of 1.1M sodium acetate which was pooled with the first aqueous phase. The M13 DNA was then ethanol precipitated, rinsed with 80% ethanol, dissolved in 15µl of $dH_2O$, and stored at -20ºC.

## 2.5.5 Sequencing reactions

Single-stranded M13 DNAs were sequenced by the chain termination method of Sanger and Coulson (1977) using either the USB 'Sequenase' protocol with T7 DNA polymerase (Tabor & Richardson, 1987). or the Pharmacia T7 polymerase protocol.

### 2.5.5.1 Sequencing with T7

1-2µg of DNA was mixed with 2ng of the appropriate primer (M13 universal 17-mer) in sequencing buffer (40mM tris-HCl pH 7.5, 20mM $MgCl_2$, 50mM NaCl) in a total volume of 10µl. Template and primer were annealed by heating to 65ºC for 2 minutes in a water bath and gradually cooling to room temperature over a period of 0.5 to 1hr. To the annealed template primer, the following were added: 1µl 0.1M DTT, 2µl labelling mix (1.5µM dCTP, 1.5µM dGTP, 1.5µM dTTP), 0.5µl [α-$^{35}$S]dATP (10µCi/ml) and 2 units of T7 DNA polymerase (Pharmacia). This mixture was incubated for 5-10 minutes at room temperature, and then divided equally into four 2.5µl aliquots of termination mixes:

ddT termination mix: 80µM dATP, 80µM dCTP, 80µM dGTP, 80µM dTTP, 8µM ddTTP, 50mM NaCl.
ddC termination mix: 80µM dATP, 80µM dCTP, 80µM dGTP, 80µM dTTP, 8µM ddCTP, 50mM NaCl.

ddG termination mix: 80μM dATP, 80μM dCTP, 80μM dGTP, 80μM dTTP, 8μM ddGTP, 50mM NaCl.

ddA termination mix: 80μM dATP, 80μM dCTP, 80μM dGTP, 80μM dTTP, 8μM ddATP, 50mM NaCl.

The reaction temperature was increased to 37°C and incubation continued for 10mins. Reactions were stopped by adding 4μl of 'stop solution' (95% formamide, 20mM EDTA, 0.05% bromophenol blue, 0.05% xylene cyanol). Completed reactions were stored at -20°C prior to electrophoresis.

## 2.5.6   Polyacrylamide gel electrophoresis

Reactions were separated by electrophoresis through 0.4mm thick, 6% polyacrylamide gels containing 8M urea with a buffer gradient of 0.5-2.5x TBE. Samples were denatured by boiling for 3 minutes or heating to 80°C for 10 minutes prior to loading into the gel, which had been pre-warmed by running at 2000 volts for 15min; 2.5μl of each sample was loaded per well. Gels were run at 1500-2000 volts initially, until the plates had warmed to 50-60°C, after which the voltage was reduced to ~1200 volts. Gels were run until the bromophenol blue dye had migrated the length of the gel (~2.5 hrs). To resolve sequence >200bp from the primer, extension gels with a buffer gradient of 0.5-1.5x TBE were used with a longer running time (6-8 hrs). When the run was complete, gels were fixed in 10% methanol, 10% acetic acid for 10 minutes, transferred to a sheet of 3MM blotting paper, and dried under vacuum at 80°C for 1-2 hours in a Drygel Sr SE1160 (Hoefer Scientfic Instruments, San Francisco). X-ray film was placed directly in contact with the dried gel, exposures of 16 to 48 hrs were typical.

## 2.6   Computing

DNA sequences were analysed using a VAX 8650 mainframe computer operating on VMS 5.4-2, using  the Genetics Computer Group Sequence Analysis Software Package version 6.2 programs developed at the University of Wisconsin (Devereux *et al.*, 1984).  The EMBL DNA sequence database was scanned using the FASTN and FASTA  programs (Lipman & Pearson 1985).

## 2.7    Polymerase Chain Reaction (Saiki *et al.*, 1988).

### 2.7.1    General Precautions

Since the polymerase chain reaction will produce a product from very small amounts of template, precautions were taken to ensure that reagents and materials used for PCR were kept free of contaminating DNA. Thus PCR reactions were prepared away from the normal working area, using fresh reagents, clean pipettes and eppendorf tubes taken directly from the manufacturers packaging. All PCR reactions were performed in conjunction with zero DNA controls, which consistently gave no product.

### 2.7.2    General PCR protocol

PCR amplifications from genomic DNA were generally performed in a 10µl reaction mix comprising 45mM tris-HCl (pH 8.8), 11mM $(NH_4)_2SO_4$, 4.5mM $MgCl_2$, 6.7mM 2-mercaptoethanol, 4.4µM EDTA (pH8.0), 1mM dATP, 1mM dCTP, 1mM dGTP, 1mM dTTP, 113µg/ml BSA, 1µM of each primer and 1 unit of *Taq* polymerase. Reaction mixtures were overlaid with paraffin oil in a 0.5ml microfuge tube and amplifications carried out in a Perkin-Elmer Cetus DNA thermal cycler (Perkin-Elmer Cetus, Connecticut, USA). Amplimer sequences and parameters for the different primers are given in table 2.1. After amplification, PCR products were analysed by gel electrophoresis and ethidium staining, or by Southern hybridization.

**Table 2.1    PCR parameters**

| Region | Primer sequences (5'→3') | Denaturing | | Annealing | | Extension | | No. Cycles | Chapter |
|---|---|---|---|---|---|---|---|---|---|
| | | °C | mins | °C | mins | °C | mins | | |
| *Hm-2* CA microsatellite | R  TGGCCACACACTTAGCTTTC<br>L  CCCGCACACCTGCTGTGGTG | 96 | 1'20" | 64 | 1'00" | 70 | 1'00" | 27-30 | 4 |
| *Hm-2* minisatellite | A  CATGTCCACCTGTCTCTGAC<br>B  GGTGTAGCCTGATTAGAGGA<br>C  GATGACTGTCAGAGCAGGGA<br>D  CCCTCTGCTTTGTGCTTGTG | 96 | 1'00" | 65 | 1'00" | 70 | 10'00" | 18-20 | 5 |
| *Hm-2* flanking DNA | DR  CACAAGCACAAAGCAGAGGG<br>B  GGTGTAGCCTGATTAGAGGA | 96 | 1'20" | 58 | 1'00" | 70 | 4'00" | 25 | 3 |

# 3
## Cloning of *Hm-2*:
## a second highly unstable mouse minisatellite

## 3.1 Introduction

*Ms6-hm* has been shown to be unusual when compared to other characterized minisatellite loci. An initial objective of this work was to show that the properties of high germline and somatic mutation rates seen in *Ms6-hm* were not unique to this locus by cloning and characterizing another hypervariable locus.

It had been observed previously that when murine DNA was probed with pMm3-1, derived from the minisatellite *Ms6-hm* (see chapter 1), under low stringency conditions, a novel set of hybridizing loci were seen, giving a murine 'DNA fingerprint' which was variable both between strains and, to a lesser extent, within strains. In particular, one of the loci seen under these conditions, characterized by large (~20kb) fragments in C57BL/6J mice, showed evidence of instability in families. This locus, named *Hm-2*, for 'hypervariable minisatellite-2' was selected for cloning and characterization. If *Hm-2* did show similar features of high germline and somatic mutation rates, then it might be possible to search for common factors, such as sequence, or chromosomal localization, which determines this variability.

## 3.2 Identification of *Hm-2*

To clone and further characterize the locus, it was first necessary to define a method whereby *Hm-2* alleles could be distinguished amongst alleles of other loci which also hybridize to pMm3-1 at low stringency.

Assuming that each minisatellite locus is embedded in different DNA sequence, each separate minisatellite locus will have a different distribution of restriction enzyme recognition sites in the immediate flanking DNA; cleavage by restriction enzymes will therefore produce fragments for a given locus that vary in size, depending on the distance the target sites are from the minisatellite, and a flanking DNA 'signature', specific to that locus will be

produced. The sizes of these fragments relative to each other will remain constant regardless of the number of minisatellite repeats within the fragment, and so this signature is independent of allele length.

To identify alleles of *Hm-2*, various mouse DNAs were digested with one of four different restriction enzymes, *Hin*fI, *Alu*I, *Sau*3AI, and *Hae*III, to produce a characteristic flanking DNA 'signature' for the large C57BL/6 alleles. (Figure 3.1). This pattern is clearly recognizable in the other mice which have inherited C57BL/6 alleles. The pattern from four enzymes is sufficient to distinguish *Hm-2* from all other loci, and shows the presence of both mutant and mosaic alleles at *Hm-2*.

For example, $F_1$ 34 in figure 3.1 lacked the expected 19kb paternal allele seen in its litter mate, but had an additional band at about 12kb, which matched the expected pattern of sizes for *Hm-2* alleles, which showed that the 12kb band was the result of a germline mutation at *Hm-2* which reduced the parental 19kb allele to 12kb. This reduction in size was from a loss of minisatellite repeats only, with no effect on the DNA flanking the minisatellite.

C57BL/6J $F_0$ 7 (figure 3.1), which was heterozygous with *Hae*III allele sizes of 23 and 20kb, was seen to have a third band at 18.5kb which followed the signature of the others, this suggested that the mouse had three different *Hm-2* alleles; the alternative explanation was to postulate a separate cross-hybridizing fragment which had exactly the same pattern of length changes with the enzymes as *Hm-2*. The presence of three different *Hm-2* alleles is evidence of mosaicism; one of the parental alleles has mutated to a smaller size by loss of repeats, but this has happened subsequent to fertilization, so that cells with both parental type alleles are present in addition to mutant cells with one parental and one mutated allele. On close examination of a long exposure it was seen that $F_1$ 37 (figure 3.1) also has a faint band which shows the characteristic pattern of *Hm-2*, so this mouse is also a mosaic with one C57BL/6J allele, one DBA allele which as yet cannot be seen, and one mutant allele.

The signature cannot be clearly detected in the DBA mouse, indicating that *Hm-2* alleles in this strain either have different flanking signatures, or are much smaller than the C57BL/6J *Hm-2* alleles and were run off the bottom of the gel. It is also possible that the pattern is present, but masked by other cross-hybridizing fragments.

A restriction map of *Hm-2* was generated by digestion of $F_0$ DNA with various different combinations of four base pair recognition enzymes, and

**Figure 3.1    Genomic DNA probed with Mm3-1 at low stringency: detection of**
*Hm*-2 **by flanking DNA signature.**

DNA was analysed from different mice from a C57BL/6 x DBA pedigree bred by R.Kelly. From left; $F_0$ 7 - C57BL/6 founder, $F_0$ 13, - DBA founder; $F_1$ 34 and $F_1$ 37,- two C57BL/6 x DBA $F_1$ mice from the same family; $F_2$ 86,- C57BL/6 x DBA $F_2$ mouse. The final lanes (**BXD 12**) are digests of DNA from a mouse from BXD Recombinant Inbred strain 12.   DNA from each mouse was digested with four separate enzymes; *Hinf*I (**H**), *Alu*I (**A**), *Sau*3AI (**S**), and *Hae*III (**III**), and Southern blot hybridized with Mm3-1 at low stringency.

A characteristic 'signature' (marked by □ ) for the large C57BL/6J *Hm*-2 alleles in $F_0$ 7 is recognizable in C57BL/6 x DBA $F_1$ 37 and in the $F_2$ and BXD samples.   $F_1$ 34 in figure 3.1 lacks the expected 19kb paternal allele seen in its litter mate, but has an additional band at about 12kb, which matches the expected pattern of sizes.

C57BL/6J $F_0$ 7, which is heterozygous with *Hae* III allele sizes of 23 and 20kb, is seen to have a third band at 18.5kb which follows the signature of the others, suggesting that this mouse is mosaic.   On close examination of a long exposure it was seen that $F_1$ 37 also has a faint band which shows the characteristic pattern of *Hm-2*, so this mouse is also a mosaic with one C57BL/6J allele, one DBA allele which as yet cannot be seen, and one mutant allele.
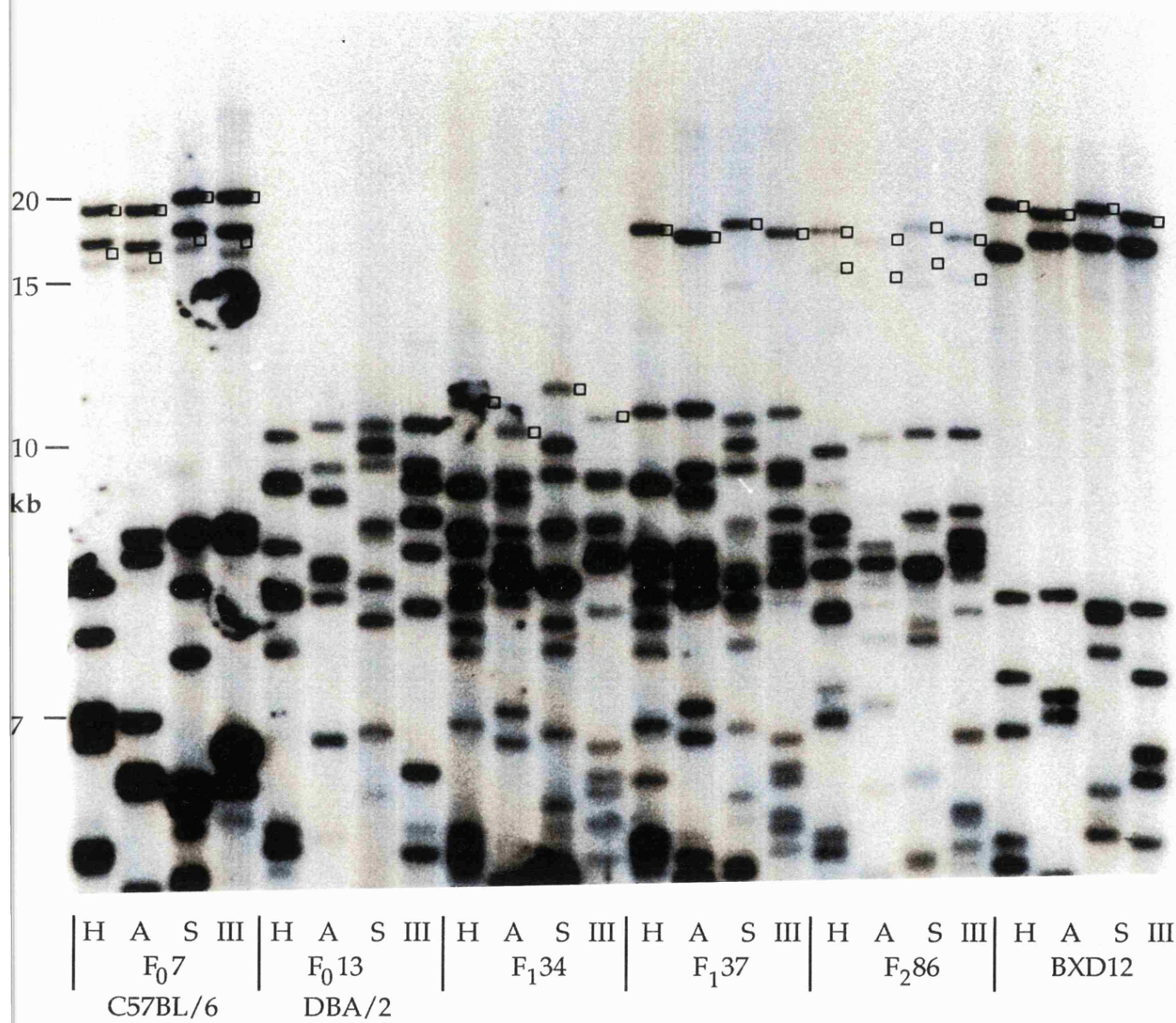
**Figure 3.1   Genomic DNA probed with Mm3-1 at low stringency: detection of *Hm-2* by flanking DNA signature**



| H  A  S  III | H  A  S  III | H  A  S  III | H  A  S  III | H  A  S  III | H  A  S  III |
|:---:|:---:|:---:|:---:|:---:|:---:|
| $F_0 7$ | $F_0 13$ | $F_1 34$ | $F_1 37$ | $F_2 86$ | BXD12 |
| C57BL/6 | DBA/2 | | | | |

with enzymes which would possibly be useful in later cloning. Figure 3.2 shows the initial map of nearest restriction sites generated for a C57BL/6 allele. This could only be an approximate map due to the size of the fragments generated, despite the use of a low percentage gel to maximize resolution.

## 3.3 Cloning a *Sau*3AI fragment from *Hm-2*

To further investigate *Hm-2* it was necessary to clone the locus both for sequence analysis and to allow the identification of smaller alleles which are obscured by other bands when the pMm3-1 probe is used at low stringency.

The starting point for cloning is to enrich for the sequence of interest by taking a small size range of digested DNA containing the desired locus. A library is made from this DNA and clones containing the fragment of interest isolated. In this case, the fragment to be cloned had to be one of the large (20kb) C57BL/6 alleles, since smaller alleles could not be identified. This was an advantage in that the enrichment for *Hm-2* by size selection would be large, but was a disadvantage because it is difficult to clone large DNA fragments, especially repetitive DNA such as minisatellites which collapse during propagation in bacterial hosts destroying viability of lambda clones to be packaged (Wong *et al.*, 1986). An initial approach, therefore, was to dispense with cloning the whole fragment, and to instead generate an M13 library from small fragments of the size fractionated DNA, and screen with pMm3-1 to detect *Hm-2* repeats. Positive clones could be used as locus specific probes for *Hm-2*, and sequencing of the minisatellite repeat would require no further subcloning. Such sequence, however, would be unlikely to include any flanking DNA.

DNA from C57BL/6J founder 1 was prepared from liver, brain, and kidney. 150μg of this DNA was digested with *Sau*3AI and size fractioned on an agarose gel. Five size fractions were collected by electroelution onto dialysis membrane and aliquots of these run on a second gel with *Sau*3AI digested $F_0$ DNA for comparison. (see figure 3.3). It was evident from this gel that the *Hm-2* allele was contained in the first fraction, but that this fraction also contained a second cross-hybridizing fragment which would generate a false positive if cloned. Therefore a second round of size selection was carried out, which produced a size fraction containing only one cross-hybridizing band corresponding to the size of the *Hm-2* allele in the control lane (not shown). This fraction (150ng) represents a 1000x purification assuming complete recovery at each step. The DNA was sonicated and fragments from between

# Figure 3.2 Restriction maps of Hm-2



(PstI, HincII & SphI cut within the flanking Sau3AI and HinfI sites)

SIII A H          A S III H

Genomic restriction map of C57BL/6 no.1 Hm-2 allele

Restriction map of insert from p9.2 charomid clone

SIII AH    A    S

Sp P

1kb

Restriction map of 9.2 insert from pUC18 subclone, and location of nearest restriction sites to minisatellite from M13 sequencing of 9.2 insert to scale

S IIIAHAS

Sp PHc T

Minisatellite repeats

Diagram shows positions of restriction sites nearest to the minisatellite repeat block for a Hm-2 allele and clones of the same Hm-2 allele. **S** - Sau3AI, **III** - HaeIII, **A** - AluI, **H** - HinfI, **Sp** - SphI, **P** - PstI, **Hc** - HincII, **T** - TaqI

**Figure 3.3    Enrichment for *Hm-2* by size fractionation**
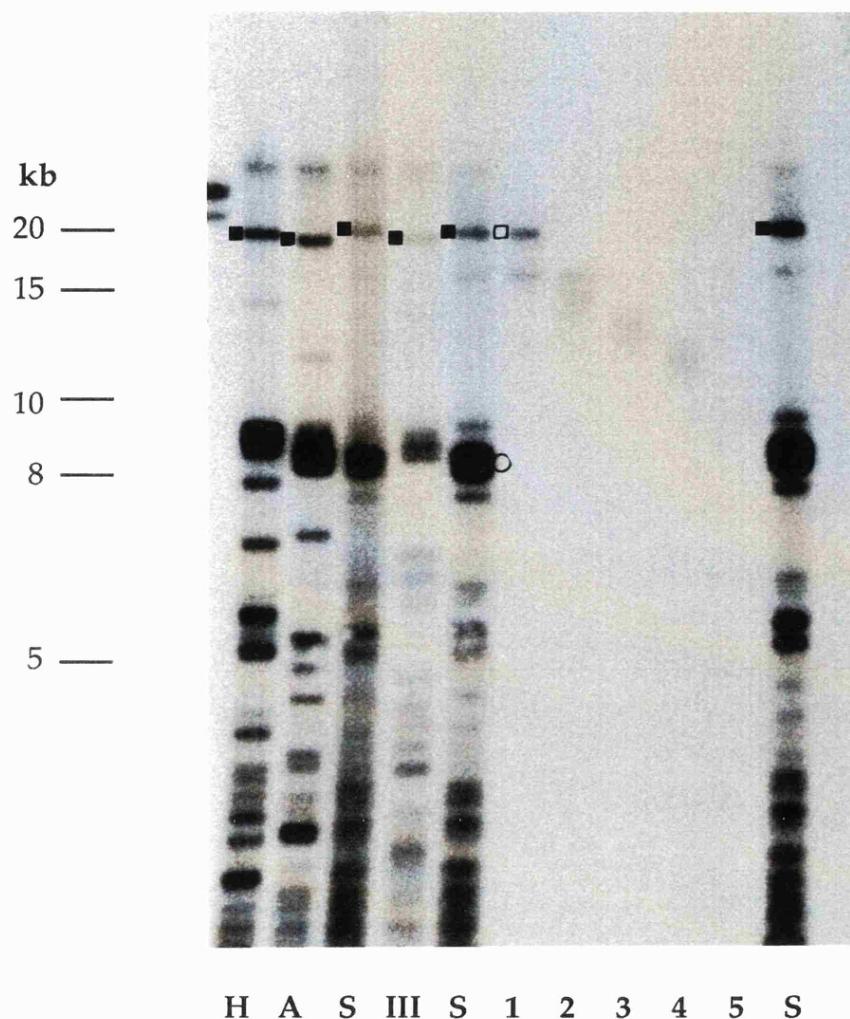


Size fractions in the range 10kb to 22kb were collected from *Sau*3AI digested genomic C57BL/6 DNA.  Aliquots of these fractions (lanes 1 to 5) and aliquots of the starting *Sau*3AI digest (S, flanking lanes 1 to 5) were separated on an agarose gel and Southern blot hybridized with Mm3-1 at low stringency.

To identify alleles of *Hm-2*, DNA from the same C57BL/6 mouse was digested with the enzymes *Hinf*I (H), *Alu*I (A), *Sau*3AI (S), and *Hae*III (III) to identify *Hm-2* by the flanking DNA signature.

*Hm-2* alleles marked with ■ , *Hm-2* allele in size fraction with □

The strongly hybridizing locus (marked ○ ) is *Ms6-hm*.

*Hm-2* is contained in fraction 1, but a smaller, non-*Hm-2* band is also present.

200-600 and 600+ bp collected. These were end-repaired and ligated into the *SmaI* site of CIP treated M13mp19 vector (Yanisch-Perron *et al.*, 1985) as described in chapter 2. Unfortunately the resulting library of plaques was too small to warrant screening despite the level of enrichment for *Hm-2*.

The second approach utilised cosmid vectors to overcome the problems of minisatellite collapse in λ vectors. These are vectors which contain λ cos sites and can therefore be efficiently packaged in vitro, but are then propagated as plasmids in transformed *E. coli* so collapse of the insert does not make the clone unviable. The cosmids used were Charomid vectors (Saito and Stark, 1986), which, unlike typical cosmids which require large inserts (typically 30-45kb), have a variable length spacer region allowing the size of the vector to be tailored to the size of the insert for efficient packaging.

Charomid vector 9-28 (Saito and Stark, 1986) was prepared from host *E. coli* DH1 cells by alkaline extraction (Birnboim and Doly, 1979). Charomid DNA was purified by linearizing with *Bam*HI, and two rounds of preparative gel electrophoresis with recovery of the 28kb fragments. Mouse DNA, prepared by size fractionation as previously, was ligated into the *Bam*HI site of the Charomid 9-28 polylinker in a ratio of 2:1 for 4 days at 15°C. Ligations packaged in vitro using Gigapack plus packaging extract (Stratagene), and absorbed onto *E.coli* DH5α. A library of an estimated 100,000 colonies was produced, which even assuming a 25% rate of non-recombinant colonies (Saito and Stark, 1986) gave a very high probability that *Hm-2* was represented. Colonies were lifted onto nitrocellulose membranes (Hybond-C, Amersham), and screened by hybridization with pMm3-1 at low stringency as described in chapter 2. Twelve putative positive clones were picked from this first round screen, replated at low density and rescreened with pMm3-1 with duplicate lifts. Duplicated positives were picked from five plates (1,4,9,10,11). Recombinant Charomid DNA was extracted from 4ml cultures, digested with *Sau*3AI to excise inserts, and electrophoresed on dual density 0.6/1.2% gels. Inserts could only be seen on an ethidium stained gel for clones 9.1 and 9.2. These gels were Southern blotted and hybridized with pMm3-1 to detect any inserts under 1kb which would be masked by fragments of vector. Autoradiographs showed cross-hybridizing insert bands for the positive clones from plates 4(not shown),9,10 and 11 (figure 3.4). A faint band was seen after a long exposure from the positives from plate 1. Insert sizes ranged from 300bp for clone 4.1, to 4kb for 9.1 and 9.2 which were the most strongly hybridizing clones. Because the cloned fragments were all less that 4kb and such small pieces of DNA would be unlikely to pass through two rounds of selection for fragments greater than

**Figure 3.4    Analysis of positive clones from charomid library by hybridization**

Cosmid DNA from positive colonies was digested with Sau3A1 to excise the insert DNA, electrophoresed through an agarose gel, and Southern blot hybridized with Mm3-1 at low stringency.  Hybridizing inserts are seen in all lanes.  Note the semi-ladder of fragments in 9.1 and 9.2, typical of a collapsing minisatellite.

Figure 3.4  Analysis of positive clones from charomid library by hybridization
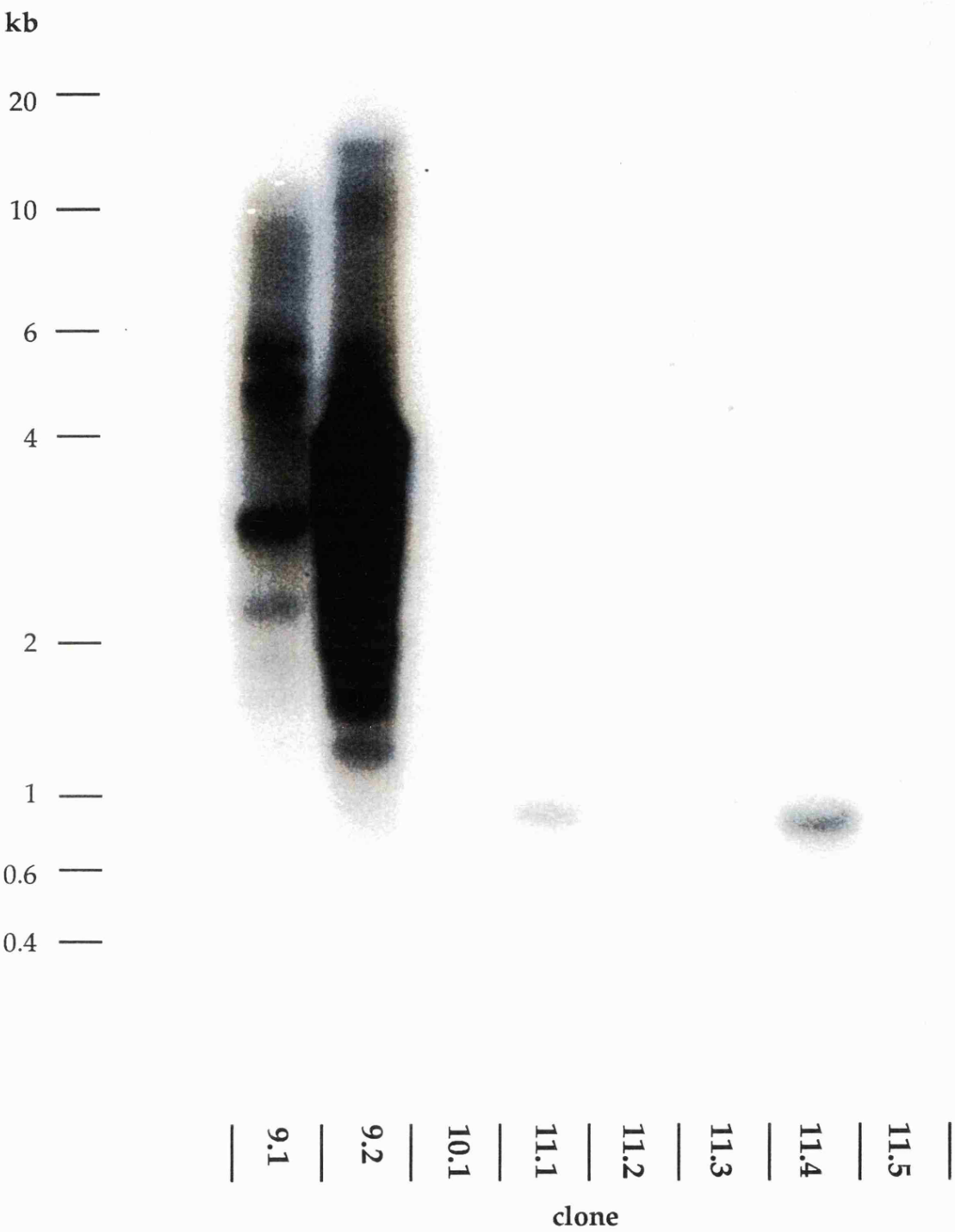
19kb, or be successfully packaged given the size restriction imposed by the Charomid vector (the 28kb vector will package successfully with inserts from between 10 and 24kb), it is apparent that inserts from all the clones have reduced in size after the cosmid entered the host *E.coli*. Collapse by loss of repeat units of cloned minisatellites has been described previously (Wyman *et al.*, 1986b; Wong *et al.*, 1986; Kelly *et al.*, 1989). Several distinct bands are visible in the 9.1 and 9.2 clones showing that repeat units are still being lost as the cosmid is replicated within the host cells (Figure 3.4). The two distinct sizes of inserts seen in positives from plate 11 could be due to one insert collapsing down to two stable forms or because two positives were picked at the same time from the first round screening.

Whole charomid DNA from six of the positives was labelled with $^{32}$P and used to probe whole mouse genomic DNA at high stringency. The two positives from plate 9 hybridized strongly to DNA with the correct *Hinf*I, *Alu*I, *Sau*3AI signature for *Hm-2*, and these bands were at a position matching the expected size of a *Hm-2* allele in the C57BL/6 mouse (figure 3.5). Other faintly hybridizing bands were seen at the base of the autoradiograph, these presumably represent other loci which are sufficiently similar to *Hm-2* to cross-hybridize under the conditions used. By isolating the *Sau*3AI insert it was shown that the insert and not the charomid DNA was responsible for the hybridization. (Subsequent hybridizations for *Hm-2* were with the purified insert DNA.)

In view of the tight size fractionation of the input DNA, plus the cross-hybridization of the clone to *Hm-2* alleles, clones 9.1 and 9.2 were considered clones of *Hm-2*.

The other putative positives tested in this manner did not produce any hybridization signal, and presumably represent extremely collapsed clones. As the objective of cloning an allele of *Hm-2* had been achieved, however, these clones were not investigated further.

Restriction mapping of the p9.2 clone showed that the region of the genomic map between the two *Sau*3AI sites agreed with the restriction map of the p9.2 insert, which further confirmed that p9.2 was a clone of *Hm-2*. The collapsed fragment was used to obtain a finer resolution map of this region (Figure 3.2). The *Sau*3AI insert from charomid clone p9.2 was subcloned into pUC18, prior to further manipulation of the insert DNA for sequencing. During subcloning the insert collapsed further to a size of 1.5kb. All deletions occurred between the *Hinf*I and *Alu*I sites, showing that only one minisatellite was present in the original clone.

**Figure 3.5  Analysis of genomic DNA fragments which cross-hybridize to the 9.2 clone**
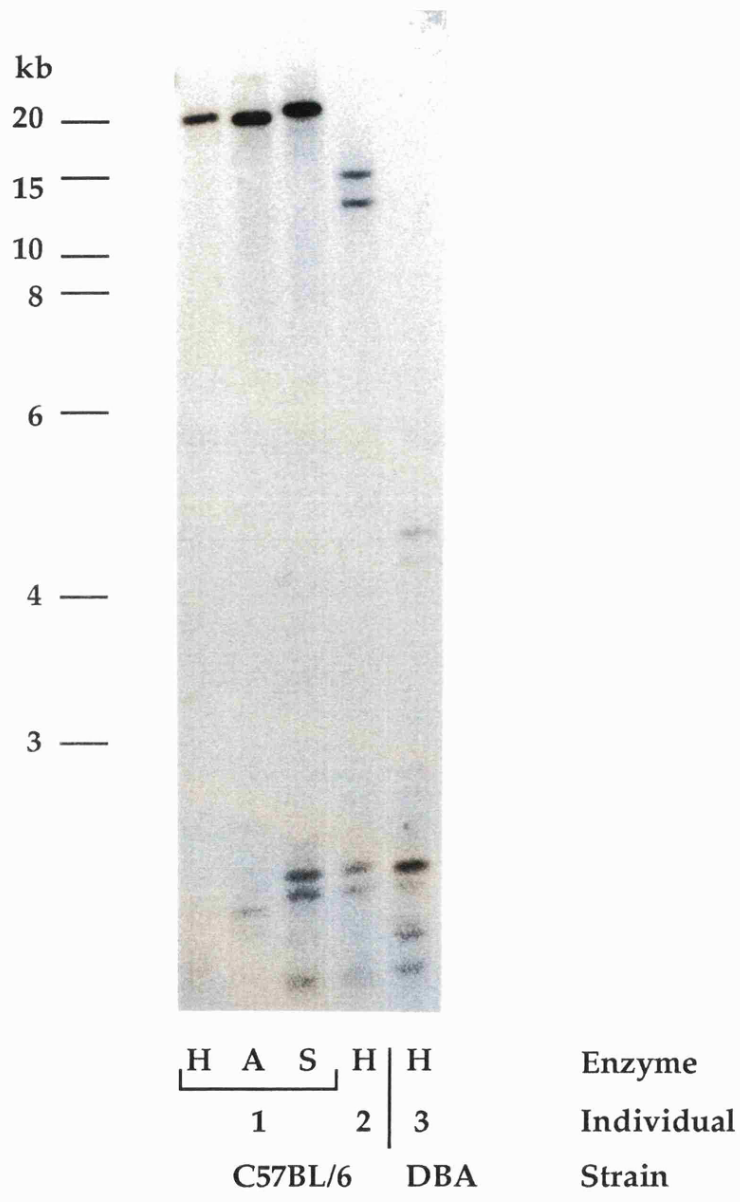


The entire p9.2 clone was labelled and used to probe digested genomic DNA at high stringency.

Lanes 1-3; C57BL/6 DNA from individual 1 digested with *Hinf*I (H), *Alu*I (A), *Sau*3AI (S). Lane 4; C57BL/6 DNA from individual 2 digested with *Hinf*I. Lane 5; DBA DNA digested with *Hinf*I.

The signature expected for an *Hm-2* allele is seen in lanes 1-3, showing that p9.2 is specific for *Hm-2*. The second C57BL/6 mouse is heterozygous.

## 3.4    p9.2 as a locus specific probe

As noted previously, p9.2 is an imperfect probe since it is not entirely specific to *Hm-2*, even under high stringency conditions. Experiments in which the stringency was increased by raising the washing temperature and concentration of mouse competitor DNA, and also decreasing the salt concentration made little, if any, difference to the intensity of DBA *Hm-2* alleles relative to the cross-hybridizing loci. These other loci do not seem to be variable  and have different flanking sequence to *Hm-2*, but hybridize to p9.2 equally efficiently as *Hm-2* alleles.

Another strategy to make a more specific probe was to use probes containing only sequences flanking the minisatellite. Such flanking probes would be single copy DNA probes and only the specific sequence should cross-hybridize. The 1.1kb *HinfI-Sau3AI* fragment flanking the minisatellite (figure 3.2) was isolated to create a more specific flanking probe, but this failed to generate any hybridization signal

When p9.2 was used as a probe at low stringency, a complex set of cross-hybridizing bands were detected. The set of loci detected were similar to those detected by the pMm3-1 probe at low stringency; the majority of differences between the DNA fingerprints created by the two probes were from differences in signal intensity from the same bands rather than the presence of novel bands. Within the small number of families screened at low stringency, none of the novel cross-hybridizing loci detected by p9.2 had the appearance of unstable loci, with additional bands or mutations. This suggests that the possibility for further probe walking to detect new loci is limited.

## 3.5    Sequencing

An initial strategy was to construct a M13 library of sonicated fragments of the p9.2 insert, from which a complete sequence would be built up from overlapping subclones. This strategy, however was unsuccessful because the library was too small, although the sequence of the ends of the 9.2 insert was obtained by cloning the the entire *Sau3AI* insert into the *Bam*HI site of M13.

To obtain complete sequence of the p9.2 clone a strategy of subcloning specific fragments of the entire insert was adopted. The  insert had been previously restriction mapped (Figure 3.2) with a selection of enzymes, in particular those with target sites in the M13mp18/19 polylinker, and was seen

to contain *Sph*I, *Pst*I and *Hinc*II restriction sites (the location of the *Hinc*II site was initially unknown, but was later mapped from the *Pst* sequence). These sites were in convenient locations to enable construction of a bank of subclones spanning almost the entire insert, in portions small enough to be accurately sequenced. The only sparse region was close to the minisatellite which could be covered by cloning the *Alu*I fragment into *Sma*I site (figure 3.6).

The p9.2 insert was digested with the appropriate enzyme and both fragments isolated by preparative gel electrophoresis. M13 vectors were prepared by double digestion of the circular form followed by preparative gel electrophoresis to remove the small fragment of excised polylinker to prevent recircularization. The orientation of the M13 polylinker was chosen so that when using the standard 17mer primer, sequence would be generated from the novel site, rather than from the *Bam*HI site from which sequence had already been obtained. Ligations, electroporation of phage, screening and preparation of ssDNAs were as previously described. Sequencing was by the dideoxynucleotide chain termination method (Sanger and Coulson, 1977) using T7 polymerase and the M13 17mer primer, or the M13 20mer primer to read sequences from the very beginning of the subclone. Fifty percent of the sequence was read from two separate subclones, which usually read opposite strands. Only in one case did overlapping sequences not match, when a 12bp mismatch was found. This was caused by a 2bp and 1bp change in sequence flanking an unchanged central region of 9bp. One of these sequences was rejected since it contained a *Taq*I site which was not present in the restriction map, and was therefore a sub-cloning or sequencing artifact.

## 3.6    Organization of *Hm-2*

The complete sequence of the 9.2 insert is 1415bp, excluding the minisatellite, and is shown in figure 3.7. This size corresponds to the bottom of the smear of collapsing minisatellite seen when cloning and subcloning the p9.2 insert, (see section 3.3). The restriction sites closest to the minisatellite present in this sequence tally with the restriction map of the p9.2 insert and the map generated from *Hm-2* alleles (figure 3.2).

# Figure 3.6 M13 subclones of p9.2 generated for sequence analysis



The section of p9.2 in each subclone is shown by

The extent of sequence read from each subclone is shown by

# Figure 3.7 Sequence of p9.2 clone
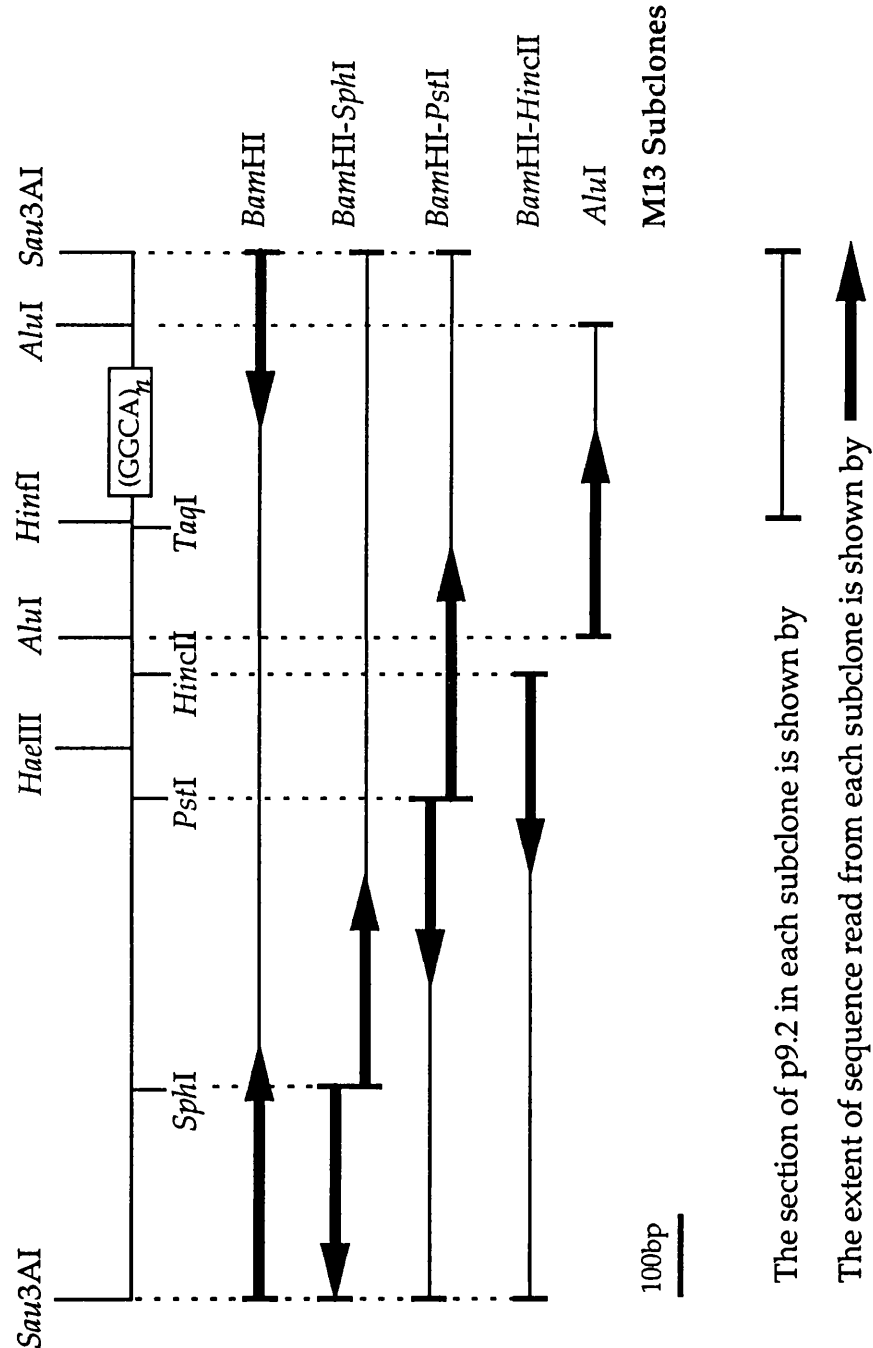
```
   1   gatcaaatga  agttagataa  taacataaag  ctgtgtgatg  gctggcgaga
  51   tgcttggggt  aaaggcaatt  aatgcctaaa  tctaatccct  gagacccaca
 101   cagtagaagg  taagaactga  ctcccatgga  ctcaactccc  gcacacctgc
 151   tgtggtgcac  acgaaCACAC  ACACACAC    ACACACAC    ACACACAaat
 201   gagtaaatgt  tacaaagatg  cacgaagaaa  ggaagatttg  tgaaagctaa
 251   gtgtgtggcc  aacagtgaaa  gccagtgttg  tgattttcat  ttctagctac
 301   tttaaatttc  ctatgattaa  aaggcatgcc  tgcctcctca  cacccagcag
 351   gcaggctgct  gttaaagggg  aacagaaagc  agctggaggc  ctgagggtgc
 401   ggaggtgaaa  ctggaaccct  ggtgtcctgc  tgatgagaac  agagagtggc
 451   atttgctcaa  aaagttaaat  tagcacaaga  gctaaatgtt  cctactctaa
 501   gtagagactg  aaaagaaatg  agttagaaca  ttgagtgctc  ctgtgcactg
 551   cggcattact  caccatacga  ctaggtggca  gcacctcagg  ccagcaacca
 601   acagagggat  taaaatatac  atttagcATA  TATATATgCA  CACACACACA
 651   CACACACACA  CAGAGAGAGA  GAGGGGGGGG  GAGAGGGAGA  GAGGGAGAGA
 701   GAGAGAgtgg  aatataattt  ggctcactag  aattgtcatc  tttccatcca
 751   ttgatgcacc  tgtctgtcta  cactgtgtct  acactggacc  tgcagcttgg
 801   ctctgtatct  atacagctgg  gttcatttgt  gtctagcaac  ttgtagctct
 851   acaccgaggc  ctccaggtat  agacctctga  aCACACACAC  ACACACACAC
 901   ACACACAC    ACACACACAC  ACACAagcta  atatggataa  ttaaaattgg
 951   aatggaggaa  catatatttg  ttaactgcta  gaaattgaga  ctacttggca
1001   gtttgaagta  aagcaactgg  tgtagcttat  gagcttaatg  ttaatgcagt
1051   aagttataat  gcttggaaag  acaaacatgt  ccacctgtct  ctgacacaat
1101   gtactcagga  caacttaaga  tggtatctca  gttactgtta  ctgctgtgcg
1151   aataccatga  ccaaggcaat  ttgtaaaagg  agcggtgtat  tgggatttac
1201   agtttcgaag  tttagagtcc  atgactgtca  gagcagggaa  tgtgacGGCA
1247   GGCAGGCAGG CA>>>>>>>> >>GGCAGGCA GGCAGGCAga  gaacttaact
1259   taatttctgc  ttcacaagca  caaagcacag  gggggagagc  tgaacttgga
1309   aaggcatggg  cttttgaaac  ctcaaagcct  gcctctagtg  atacatctcc
1359   tctaatcagg  ctacaccctg  taatccttcc  caaacagttc  tgatacctgg
1409   ggaccaa
```

Simple tandem repeat sequences (microsatellites) are shown in capitals. Only eight of the 56 *Hm-2* minisatellite repeats (GGCA) are shown in underlined capitals.

The sequence of the minisatellite repeat is

5' GGCA 3'

3' CCGT 5'

A total of 56 GGCA repeats were seen in the sequence of the AluI subclone, and in the opposite orientation 24 TGCC repeats were seen reading from the Sau3AI site. No variations in the repeat sequence were observed.

The GGCA Hm-2 repeat sequence has obvious similarity to the Ms6-hm repeat (GGGCA), although the homology between the two aligned repeats is only 40% concentrated to exact 6bp matches every 20bp. This match, however, was sufficient to allow cross-hybridization between the Mm3-1 probe and largeHm-2 alleles.

```
..GGCAGGCAGGCAGGCAGGCA..
..||   |     |   ||||..
..GGGCAGGGCAGGGCAGGGCA..
```

1246bp of sequence 5' to the $(GGCA)_n$ and 168bp of sequence 5' to the $(TGCC)_n$ repeat was read from the 9.2 insert. This sequence (not including the minisatellite) is 55% A/T, 45% G/C. Peculiarities of sequence are the unusually high frequency of microsatellite repeats in the 1kb of flanking sequence. Three blocks of dinucleotide repeats were found 5' to the GGCA repeat: A $(CA)_{16}$; a $(CA)_{22}$; and a composite of three microsatellites with the structure $(AT)_5G(CA)_{12}(GA/G)_{22}$.

The EMBL data base was searched for homologies to the sequence of Hm-2. All matches found were to the GGCA repeat or to the $(CA)_{16}$ and $(CA)_{22}$ repeats. When all repetitive sequences were removed from the 9.2 sequence, no significant matches were found against the remaining sequence. (Best match was 64% match over 200bp out of the 4000bp sequenced around the Mck gene (Sternberg et al.,1988).)

## 3.7  A polymorphism in the DNA flanking Hm-2

A group of families of C57BL/6 x CBA mice which had been bred to investigate early somatic mosaicism (see chapters 5 and 6) contained mice with a flanking DNA polymorphism. DNA from C57BL/6 female 1 was digested with the enzymes HinfI, AluI, MboI, and HaeIII, (figure 3.8). With the enzymes AluI, MboI, and HaeIII, the signature appeared normal for the bands that were believed to be Hm-2. With HinfI, however, one band was 2kb larger

than expected, whereas the other allele appeared normal. The variant band was allelic to the other maternal Hm-2 band and was therefore an Hm-2 allele. Because the increase in size was not seen with all restriction enzymes the polymorphism was located in the DNA flanking Hm-2 and did not result from changes to the repeat copy number. C57BL/6 female 1 was mosaic; the mutant band shared the signature of the typical Hm-2 allele rather than the variant, and therefore comprised at least the 5' flanking DNA from the standard allele (figure 3.8).

Two other C57BL/6 females used to set up the C57BL/6 x CBA families had genotypes identical to C57BL/6 female 1. DNA from one of these mice (16) was PCR amplified using four combinations of primers, as described in section 5b.5 (figure 3.9). With primer combinations A+D or C+D the PCR products matched the expected sizes of the HaeIII digested genomic DNA. When DNA was amplified with A+B or C+B, however, one band did show a 2kb increase in size. This result suggested that the HinfI polymorphism was the result of the insertion of 2kb of DNA between the B and D primer sequence. To verify this, a primer was designed to amplify across this region, extending away from the minisatellite towards primer B. (DR CACAAGCACAAAGCACAGGG, the reverse complement of D, figure 3.9). Various mouse DNAs were amplified with primers DR and B. In all mice a 105bp band amplified, as expected from the sequence of a standard allele from the 9.2 clone. When DNA from mouse 16 was amplified, a second band of 2.2kb amplified, verifying the presence of the additional DNA. Note that since the ancestral state is unknown, the polymorphism may reflect a 2kb deletion event, which has become the most common state. The PCR amplified alleles were Southern blot hybridized to probe 9.2 (figure 3.8). The 105bp bands cross-hybridized weakly, but unexpectedly, the 2.3 kb band cross-hybridized strongly to the 9.2 probe, suggesting that the additional DNA contained sequences homologous to repeat motifs {$(CA)_n$, $(GA)_n$, $(GGCA)_n$} in the 9.2 clone.

**Figure 3.9    Location of PCR primers and insertion/deletion polymorphism**
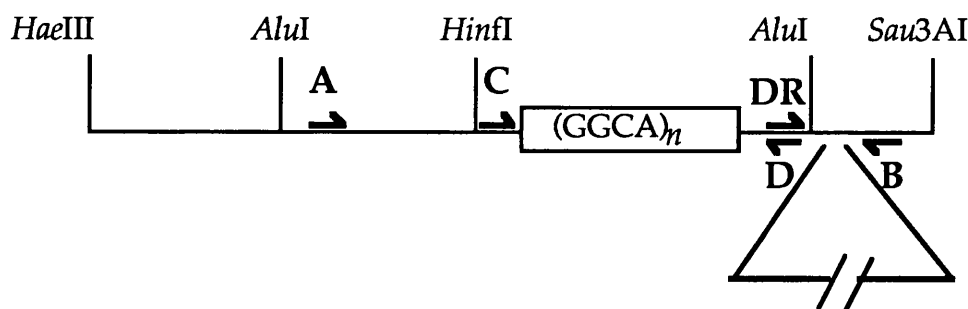
## Figure 3.8    A *Hinf*I flanking polymorphism

A    Tail DNA from C57BL/6 female 1 cut with the enzymes *Hinf*I (H), *Alu*I (A), *Mbo*I (M), and *Hae*III (III), Southern blot hybridized to $^{32}$P-labelled 9.2 at high stringency. *Hm-2* alleles labelled △

B    Schematic diagram of *Hm-2* alleles in C57BL/6 female 1. It is not known which alleles are maternal or paternal derived. One intense band and the fainter band share the same standard *Hm-2* signature. The other intense band has a signature with the enzymes *Alu*I, *Mbo*I, *Hae*III which is indistinguishable from the standard signature, but with the enzyme *Hinf*I has a size 2kb greater than expected. The fainter allele (labelled mutant) is presumed to have been derived at least in part from the allele labelled progenitor which has the same signature. It is possible, though unlikely, that the fainter band is the progenitor and the intense band the mutant.

C    Tail DNA (100ng) from C57BL/6 mice 17, 16 and 14 amplified with primers DR and B for 25 cycles, as described in chapter 2. Southern blot hybridized to $^{32}$P-labelled 9.2 at high stringency.

Figure 3.8    A *Hinf*1 polymorphism in the DNA flanking *Hm-2*
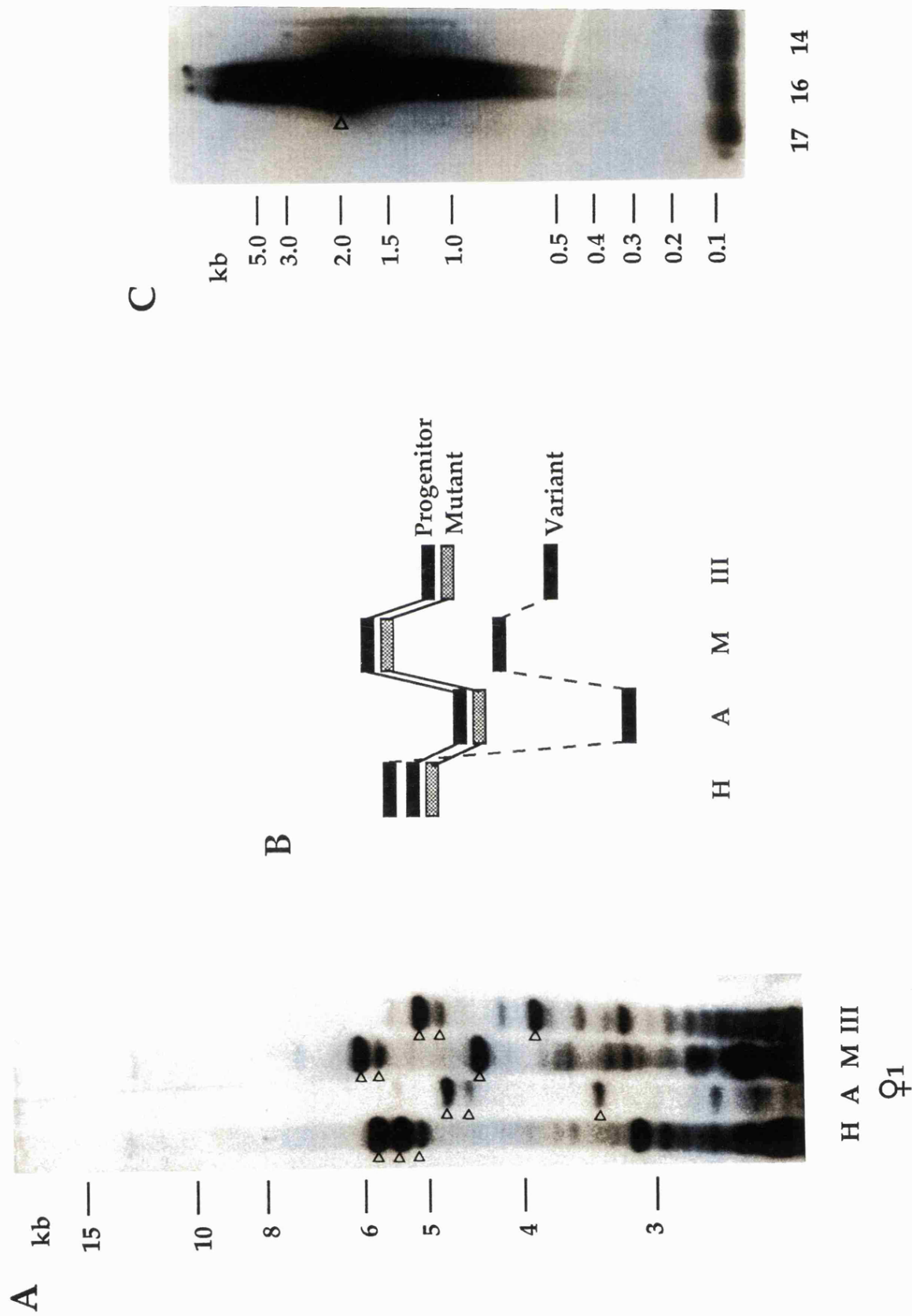
## 3.8 Summary

The unstable minisatellite locus *Hm-2* was isolated in clone p9.2 which could be used to detect alleles of *Hm-2*, although it was not a perfect locus specific probe. Sequence analysis of the p9.2 clone showed that the minisatellite repeat is GGCA, very similar to the GGGCA repeat of the other previously cloned hypervariable murine minisatellite *Ms6-hm* which was used to detect *Hm-2*.

# 4

## Localization of *Hm-2*

## 4.1 Introduction

The development of a locus-specific probe for the *Hm-2* locus allows its localization within the murine genome. The mouse is a organism for which detailed linkage maps are available, therefore it is easier to find linkage to other mapped loci. In addition, selective breeding over many years has enabled the use of techniques, such as analysis of recombinant inbred strains, which are not available for mapping in most other animals. Several linkage based methods are commonly used to map loci: classical linkage analysis using backcrosses in inbred strains, interspecific backcrosses between *Mus musculus* and *M. spretus* (Avner *et al.*, 1988, Copeland and Jenkins, 1991), and recombinant inbred strains (Taylor, 1978, 1989) can all allocate a locus to a particular linkage group and give a position on the chromosome if the linkage group has been localized cytogenetically. Alternatively, a chromosomal localization may be made more directly by *in situ* hybridization to metaphase chromosomes (Evans 1987), or by screening mouse/chinese hamster somatic cell hybrids which are segregating mouse chromosomes (Franke *et al.*, 1977).

This study took advantage of the increasing informativeness of recombinant inbred strains to localize *Hm-2*. Recombinant Inbred (RI) strains were developed in the 1970's (Bailey 1971, Taylor 1978). These strains have been derived from the cross of two unrelated inbred strains; the $F_1$ mice of the cross will be isogenic and heterozygous, the parental alleles will segregate in the $F_2$ generation and by inbreeding each line will become fixed for one or other of the parental alleles. On average half the lines will become fixed for one parental allele, and half for the other. For a characteristic determined by a single locus for which the two parental strains differ, a strain distribution pattern (SDP) of one or other parental allele across different RI strains will be seen, with one parent and half of the RI strains showing one form of the trait and the other parent and the rest of the RI strains showing the other. If this is

not seen it is evidence that the trait is not controlled by a single gene (Festing 1979).

Of particular interest in this case is the ability to use recombinant inbred strains to find linkage relationships of polymorphic loci. By using a RI strain for which the SDPs of a wide range of marker loci are known it is possible to assign linkage of another polymorphic locus by comparing the SDPs of the polymorphism, and the markers. Linked genes will have SDPs which are largely concordant, the closer the two loci are linked the greater will be the match between SDPs, and the polymorphic locus can be assigned a chomosomal location if the chromosomal position of the linked marker is known. There are currently about 20 sets of RI strains in existence although the number of generations of inbreeding is quite low in some (Taylor, 1989). The most informative series of RI strains for mapping purposes are those with a large number of strains and a large number of typed loci. On these criteria the most informative RI strains are the BXD series, from a cross between the C57BL/6J and DBA/2J strains of inbred mice (Taylor 1978). The informativeness of any set of recombinant inbred strains also relies on the locus under test having a detectable polymorphism between the parental strains.

## 4.2 There is no obvious SDP for *Hm-2* in BXD RI mice

From observations of allele lengths in inbred strains it was known that DBA/2 mice had allele sizes of about 4kb whereas C57BL/6 mice had allele sizes of around 19kb, therefore it should have been possible to score alleles in BXD strains to determine an SDP for *Hm-2*. A nylon membrane of *Alu*1 digested DNA from C57BL/6, DBA, and BXD RI mice prepared previously by R.Kelly was probed with $^{32}$P-labelled p9.2 (Figure 4.1). The sizes of *Hm-2* alleles in the BXD mice ranged from 1.2-21kb and showed no obvious relationship to the allele sizes in the parental C57BL/6 and DBA strains (19 and 4kb respectively). This presumably reflects high levels of mutation at this locus, which over the 60+ generations that the strains have been separated has resulted in the mutation of alleles both in the BXD RI strains and the parental strains, so that now no two alleles are the same length. Further evidence for a high mutation rate is shown by the frequency of heterozygosity: strains 1,8,15,18,19, and 20, plus both DBA mice and one C57BL/6 mouse were

**Figure 4.1** ***Hm-2* alleles in BXD Recombinant Inbred strains detected by hybridization of p9.2 to genomic DNA.**

Southern blot (prepared by R.Kelly) has approximately 3µg genomic DNA digested with *Alu*I from parental strains and each BXD strain (1-31). This filter was hybridized to p9.2 and washed at high stringency. A range of sizes of *Hm-2* alleles are seen, the majority between 4 and 21kb. For some strains there are no obvious *Hm-2* alleles; in these mice the alleles are small and obscured by other cross-hybridizing fragments. Strains 1 and 18 have an additional band ( △ ) and are mosaic.

B = C57BL/6, D = DBA parental strains

SDPs can be established for two of these cross-hybridizing loci. Locus 1 (arrowed) is seen as a 4.5kb band in DBA mice which is allelic to a 2.3 kb band in C57BL/6 mice. Locus 2 (asterisked) is seen as a 2kb band in C57BL/6 mice, no DBA counterpart to this allele is seen. SDPs for these loci are given below.

| | 1 | 2 | 5 | 6 | 8 | 9 | 11 | 12 | 13 | 14 | 15 | 16 | 18 | 19 | 20 | 21 | 22 | 24 | 25 | 27 | 28 | 29 | 30 | 31 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *locus1* | B | B | D | B | B | B | D | D | D | D | B | B | D | B | D | B | D | D | B | D | D | B | D | D |
| *locus2* | B | D | D | B | B | D | D | B | D | B | D | ? | D | D | B | B | D | B | D | D | B | B | D | B |

Figure 4.1 *Hm-2* alleles in BXD recombinant inbred strains detected by hybridization of 9.2 to genomic DNA

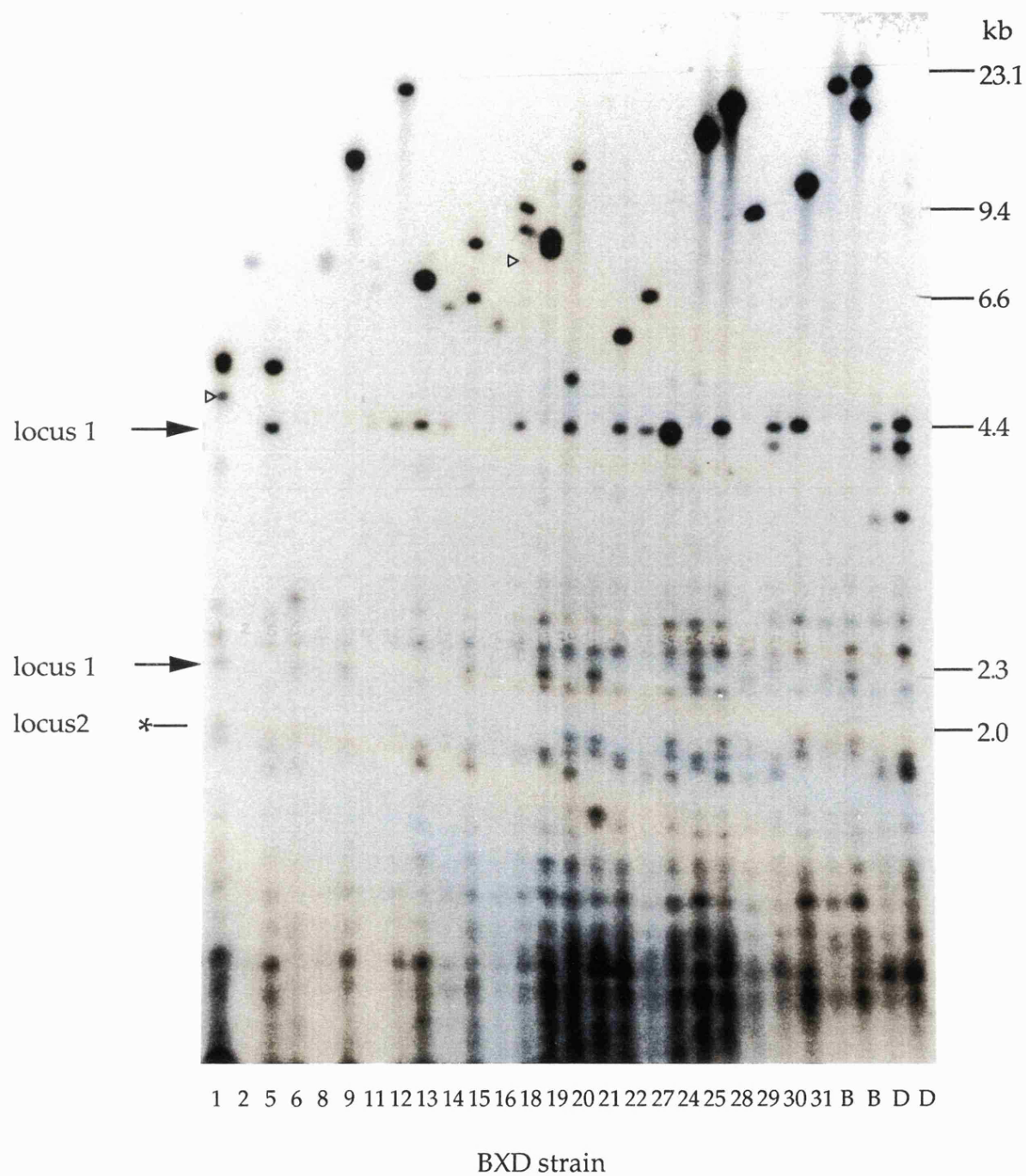**Figure 4.2  Size distribution of *Hm*-2 alleles in BXD mice**
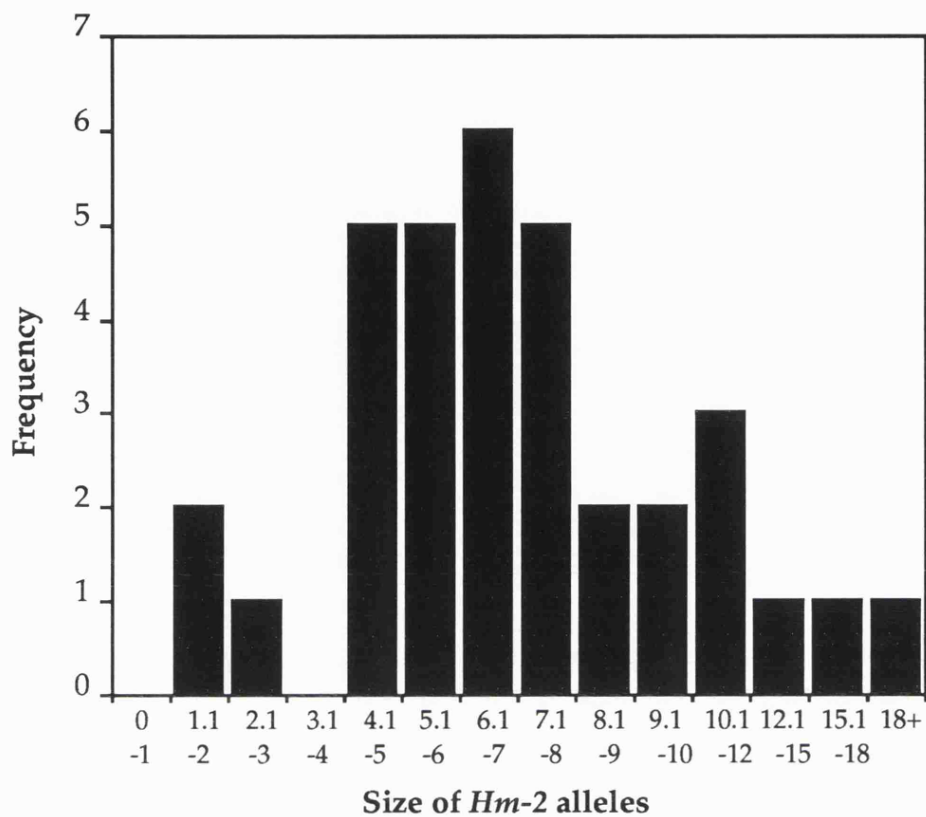


Size distribution of *Hm*-2 alleles in BXD mice (kb).  There is no bimodal size distribution, therefore it is not possible to group allele sizes into B type and D-type on the basis of size.

heterozygous at *Hm-2* in all or most of their cells. Strains 1 and 18 had an additional band and were mosaic.

It has been shown previously that the majority of length changes in minisatellites are small (Jeffreys *et al.*, 1988a). It might be expected, therefore, that alleles will only gradually mutate away from the original allele size, and would still be grouped into two distinguishable classes centred on the allele sizes of the original C57BL/6 and DBA mice used to set up the BXDs. This was, however, found not to be the case; as the histogram in figure 4.2 shows, there is a complete range of allele sizes with no evidence of a bimodal distribution.

SDPs could be read for two of the other cross-hybridizing variable DNA fragments. These are the 4.5kb band present in C57BL/6 and 13 of the BXDs which is allelic to a 2.3 kb band present in DBA mice and the other BXDs (locus 1); and a 2kb band which is present in C57BL/6 and either present or absent in the BXDs (locus 2, figure 4.1). The SDPs were compared to the database of known SDPs, but significant linkage to other loci could not be shown (B.Taylor, pers comm). This might be because the loci are in unmapped regions of the genome. It is also possible that the SDP of locus 2, which is scored from the presence or absence of a band, may be false if the locus is unstable and in some RI strains the allele is present, but has mutated to a different size.

## 4.3    A polymorphic CA microsatellite linked to *Hm-2*

The problem of distinguishing the two parental type alleles in the BXD RI strains on the basis of size would be common to all RI strains, therefore, an alternative means of finding the SDP for *Hm-2* needed to be found.

Sequence analysis of the p9.2 clone (chapter 3) revealed a microsatellite $(CA)_{16}$ about 1kb from the minisatellite. These small tandem repeats have been shown to be polymorphic (Weber and May, 1989, Litt and Luty, 1989), and if variable between C57BL/6 and DBA would provide an SDP for the linked *Hm-2* locus. PCR primers CAR (CCCGCACACCTGCTGTGGTG 5' to the CA strand) and CAL (TGGCCACACACTTAGCTTTC  5' to the TG strand) were designed to amplify the  microsatellite.

Amplification of C57BL/6 DNA produced a 125bp fragment (figure 4.3), as expected from the sequence of C57BL/6 $F_0$ no.1. Amplification of DBA DNA produced fragments of 131bp, an increase of 6bp, which was assumed to reflect a gain of three CA repeats.  C57BL/6 and DBA microsatellite alleles were

**Figure 4.3    Detection of polymorphism at the microsatellite linked to *Hm-2***

**A    Amplification from DNA from C57BL/6 and DBA mice.**

100ng DNA from C57BL/6 (BL/6), DBA (DBA) and C57BL/6 x DBA $F_2$ ($F_2$) mice was amplified using primers CAL and CAR (Denaturation 96°C 1'20", annealing 64°C 1'00", extension 70°C 1'00", for 30 cycles). Amplified products visualized by electrophoresis and ethidium staining on an agarose gel.

In C57BL/6 mice and in $F_2$ mice in which both *Hm-2* alleles are from C57BL/6 mice the size of product (125bp) was as predicted from the sequence of the p9.2 clone (chapter 3). In DBA mice and F2 mice with two DBA alleles a 6bp larger product amplifies. In heterozygotes a smear is seen between 125 and 131bp representing both the above allele sizes plus possibly the heteroduplex.

**B** = C57BL/6 allele, **D** = DBA allele.

**B    Amplification of CA microsatellite in other inbred strains.**

100ng of DNA from a pair of mice from six other strains was amplified as previously. No new alleles were detected; amplification products in five strains (A, AKR, C3H, CBA and BALB/c) matched the DBA product at 131bp. The other strain (SWR) matched the size of the C57BL/6 product of 125bp.

**Figure 4.3 Detection of polymorphism at the microsatellite linked to *Hm-2***

A



◄ 131bp
◄ 125bp

BL/6 DBA BL/6 DBA  D/D  D/B  B/B   0
  1   19   1   21    27   44   45  DNA

F₀                    F₂

B



- 131bp
- 125bp

BL/6 DBA   A    AKR   C3H   CBA  BALB/c  SWR   0
                                              DNA
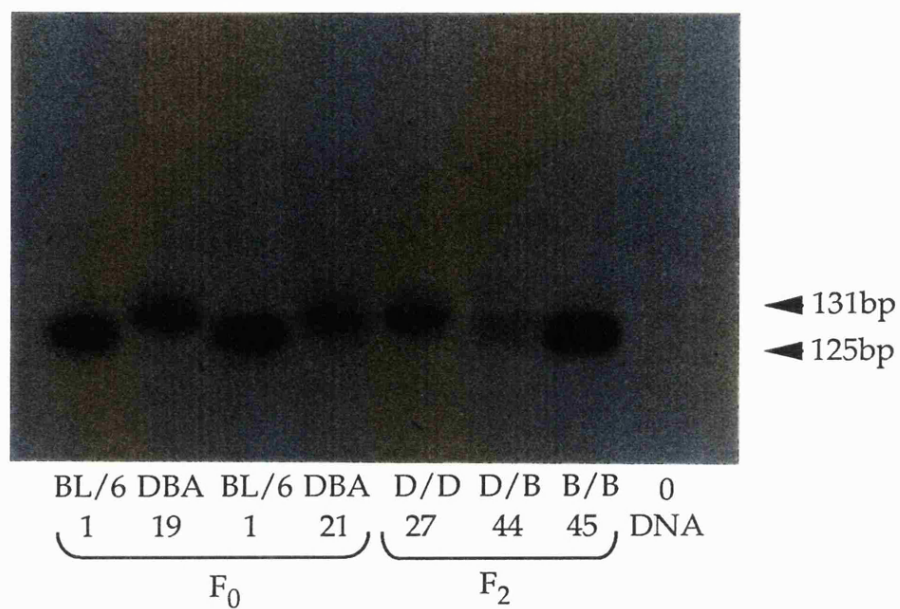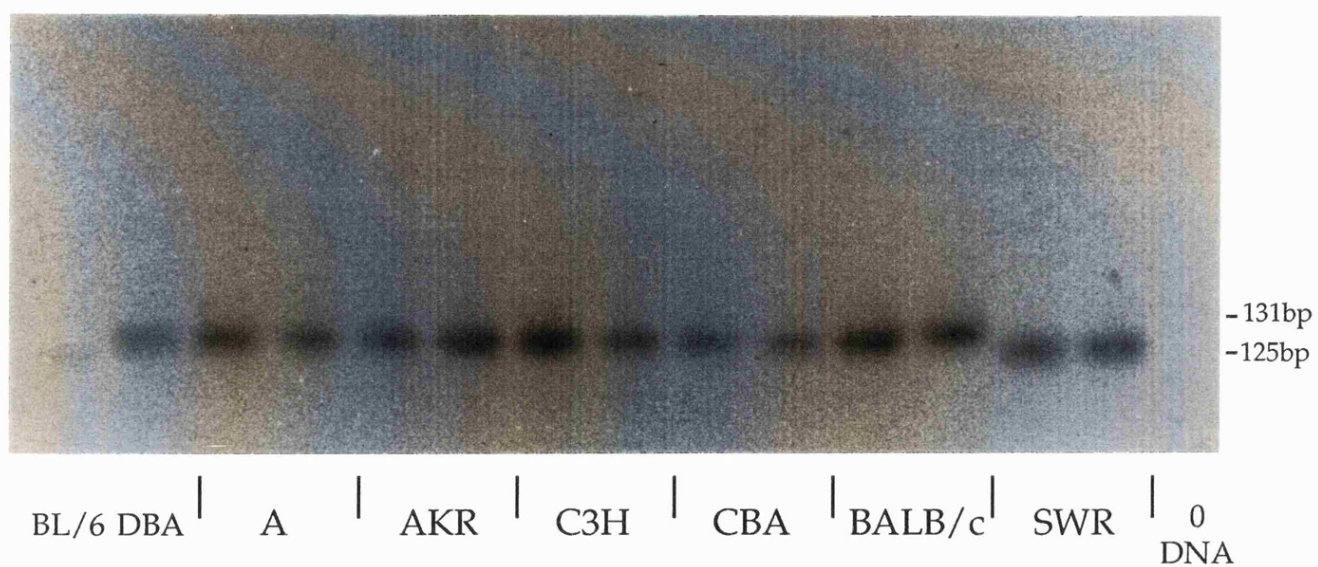
consistently inherited with C57BL/6 and DBA *Hm-2* alleles respectively. DNA from heterozygous $F_1$ or $F_2$ mice produced a smear between 125 and 131 bp representing the two parental microsatellite alleles and the heteroduplex. (figure 4.3). Therefore it was possible, albeit indirectly, to type *Hm-2* across the BXD RI series.

Variability at this microsatellite was also checked across other inbred mouse strains (figure 4.3). Within the resolution limits of agarose gel electrophoresis, the other strains had no new length alleles: Strains A, AKR, C3H, CBA, BALB/c had the 131bp DBA allele, and the SWR mice had the 125bp C57BL/6 allele. This extended the RI strains that would be informative to include the BXH RI series (from C57BL/6 x C3H/HeJ, 12 strains , Taylor 1989). Other series that have the potential to be typed at this microsatellite but for which DNAs were not held are the AXB, BXA and CXB series.

## 4.4 A microsatellite based Strain Distribution Pattern for *Hm-2*

PCR amplification of the microsatellite from BXD DNA showed in all cases but one, a single band which corresponded to either the DBA or the C57BL/6 type allele. The exception was BXD 6 in which the 131bp DBA allele amplified, and a second allele at 137bp was seen (figure 4.4A). Because of the presence of the DBA type allele and the complete absence of the C57BL/6 allele this strain was scored as 'D type'. The other allele presumably arose by a recent mutation event in strain BXD 6, giving a maximum limit for the mutation rate at the microsatellite of 0.002 per gamete. The complete strain distribution pattern is given in figure 4.4B.

The 12 BXH RI strains were also typed (figure 4.4C) to produce the SDP shown in figure 4.4D. Although only 3 out of the 12 strains are B type this is not a significant deviation from the 6/12 expected.

## 4.5 Chromosomal Localization of *Hm-2*

Comparison of the microsatellite SDP to the database of typed loci was carried out by B. Taylor. Strong evidence of linkage was found to the marker *Cck* (Friedman *et al.,* 1989) on chromosome 9, with 1 recombinant out of 26 BXD strains. The BXH data were consistent with a placement on chromosome 9 with 2 from 12 strains discordant with *Fv-2* (Jenkins *et al* , 1982). SDPs are shown in figure 4.5.

**Figure 4.4    Analysis of microsatellite alleles in Recombinant Inbred Strains**

A    Microsatellite alleles in BXD RI strain DNA.  100ng aliquots of genomic DNA were PCR amplified and typed as described in figure 4.3.  The mouse from BXD strain 6 has an additional band above the DBA sized allele.

B    Microsatellite SDP derived from BXD strains.  D; 131bp, DBA allele amplified.  **B**; 125bp,  C57BL/6 allele amplified.

C    Microsatellite alleles (positive image) in BXH RI strain DNA.  100ng aliquots of genomic DNA from BXH RI strains and the parental C57BL/6 and C3H strains were PCR amplified and typed as described in figure 4.3.  Two mice from strain 19 were amplified, both show the same size of amplification product.

D    Microsatellite SDP derived from BXH strains.  H; 131bp, C3H type band amplified.  **B**; 125bp,  C57BL/6 type band amplified.

# Figure 4.4  Analysis of CA microsatellite in recombinant inbred strains

## A    PCR amplification of CA microsatellite in BXD strains
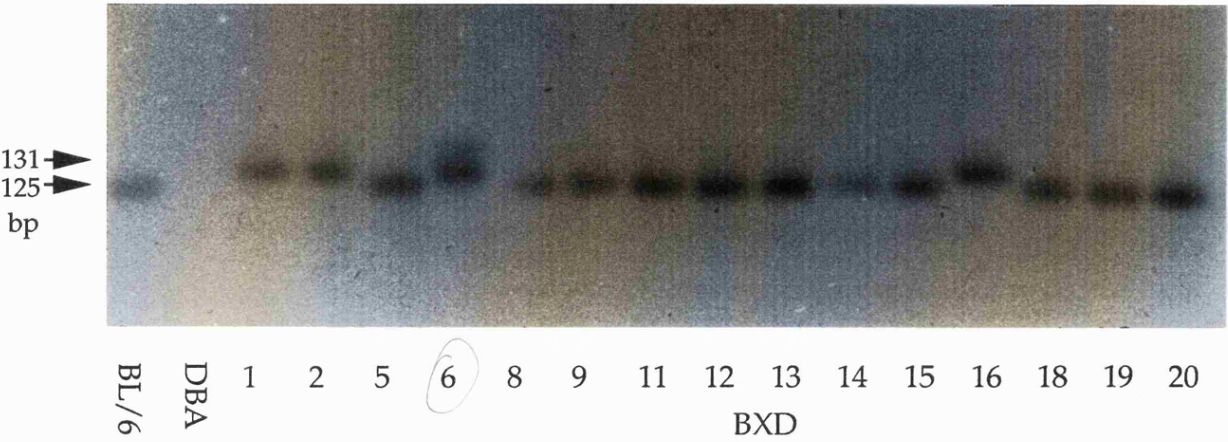


## B    SDP for CA microsatellite in BXD RI strains

| BXD | 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 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| CA (*Hm-2*) | D | D | B | D | B | B | B | B | B | B | B | D | B | B | B | D | D | D | D | B | B | B | B | D | B | D |

## C    PCR amplification of CA microsatellite in BXH strains
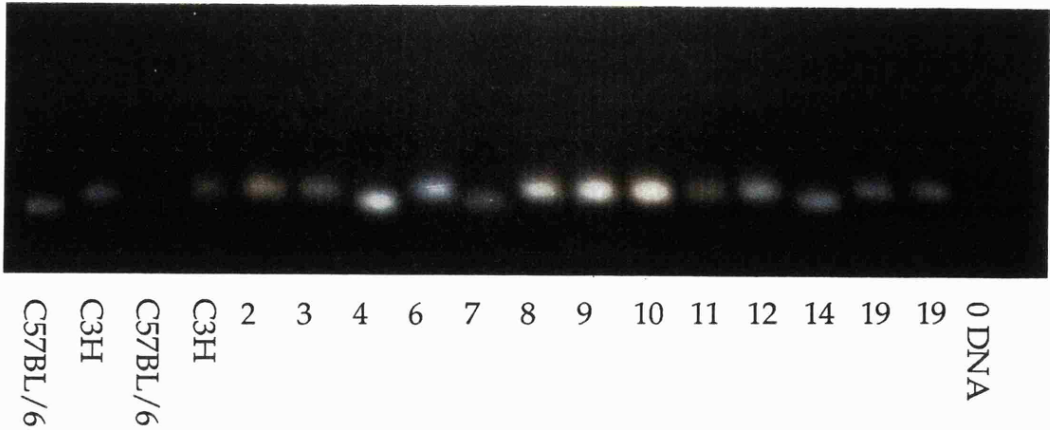


## D    SDP for CA microsatellite in BXH RI strains

| BXH | 2 | 3 | 4 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 14 | 19 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| CA (*Hm-2*) | H | H | B | H | B | H | H | H | H | H | B | H |

**Figure 4.5**      **Comparison of SDPs from loci on chromosome 9**

**A**

**BXD Strain**

|        | 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 |
|--------|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| *Fv-2*  | D | B | B | B | D | B | B | B | B | B | B | B | D | B | D | B | D | D | B | D | D | B | B | B | ? | ? |
| *Bgl-s* | B | D | B | B | D | B | B | B | B | B | B | B | D | B | D | D | D | B | B | D | B | B | B | B | D | ? |
| *Cck*   | B | D | B | D | B | B | B | B | B | B | B | D | B | B | B | D | D | D | D | B | B | B | B | D | B | D |
| *Hm-2*  | D | D | B | D | B | B | B | B | B | B | B | D | B | B | B | D | D | D | D | B | B | B | B | D | B | D |
| *Ms-TF* | D | D | D | B | B | B | B | B | B | B | B | D | B | B | B | D | D | ? | D | B | B | B | B | D | B | ? |

2   1   1   2   1   -   -   -   -   -   -   -   1   1   -   1   1   -   2   1   1   1   -   -   1   1   -

**B**

**BXH**

|       | 2 | 3 | 4 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 14 | 19 |
|-------|---|---|---|---|---|---|---|----|----|----|----|----|
| *Hm-2* | H | H | B | H | B | H | H | H | H | H | B | H |
| *Fv-2* | H | H | B | H | B | H | H | B | B | H | B | H |

**A**   Comparison of published BXD SDPs of loci on the distal region of chromosome 9 with the SDPs of *Hm-2* and Ms-TF. The loci *Fv-2:Bgl-s* and *Cck* are listed in the order in which they are mapped from centromere to telomere, *Hm-2* and Ms-TF are included in the positions which minimises the number of crossovers. Numbers below the table indicate the number of crossovers between *Fv-2* and Ms-TF in each strain. The order *Fv-2:Bgl-s*:Ms-TF:*Hm-2:Cck* would require a triple crossover in strain 1.

    **B**   Comparison of BXH SDPs of *Hm-2* and *Fv-2*.

Standard formulae are available for the calculation of recombination fractions between loci based on the analysis of RI strains (Taylor, 1978). For RI strains constructed by brother-sister matings, the relationship between the recombination fraction, r, and the probability, R , that a RI strain will carry a recombinant pair of alleles is given by

$$R = \frac{4r}{(1+6r)}$$  (Haldane and Waddington, 1931)

r is the map distance between the two loci (1 map unit = 1cM = 1% recombination), and an estimate of this can be obtained from the rearrangement of the formula above;

$$\hat{r} = \frac{R}{(4-6R)}$$

where $\hat{r}$ is the estimate of the true recombination fraction, based on the estimated value of R from the observed proportion of strains carrying recombinant combinations of alleles at the two loci. From this formula the estimate for the recombination distance between Hm-2 and Cck is 1.0 cM, shown in figure 4.6. 95% confidence intervals, which are calculated assuming that values of R follow a binomial distribution, are taken from standard tables (Silver, 1985). It can be seen that the confidence interval is high despite almost complete concordance in SDPs.

The SDP for Hm-2 also showed linkage to the locus Ms-TF, which was detected by cross-hybidization to a human telomere repeat probe (N.J. Royle, and R. Kelly, pers. comm.). Ms-TF was seen as a 4.0 kb band in AluI digested DBA DNA when probed with $(TTAGGG)_n$ at low stringency, and was present in 8/24 of the BXDs. No allelic C57BL/6 allele could be identified therefore, similar to locus 2, it is possible that this SDP is erroneous. Ms-TF is an interstitial telomere related sequence, which are often localized at proterminal sites (Wells et al., 1990). This would agree with the placement of both loci near to Cck, which is localized to the distal region of chromosome 9.

The SDP (Figure 4.5) for Ms-TF showed 2/24 discordances with Hm-2, and 3/24 with Cck, giving the order of the loci as Cck_Hm-2_Ms-TF, the alternative order, Cck_Ms-TF_Hm-2 required double cross-overs in two BXD strains. To orientate this grouping with respect to the centromere it was necessary to compare the SDPs of the two 'outside' markers, Cck and Ms-TF, to

**Figure 4.6    Estimation of Genetic Distances between loci**

|  |  | Discordancies | Distance (cM) |
|---|---|---|---|
| *Hm-2* | *Fv-2* | 11/24 | 36.7 |
| *Hm-2* | *Bgl-s* | 11/25 | 32.4 |
| *Hm-2* | *Cck* | 1/26 | 1.0 (0.0-7.0) |
| *Hm-2* | Ms-TF | 2/24 | 2.4 (0.3-11.3) |
| Ms-TF | *Fv-2* | 11/23 | 42.3 |
| Ms-TF | *Bgl-s* | 10/24 | 27.8 |
| Ms-TF | *Cck* | 3/24 | 3.9 (0.7-15.7) |
| *Cck* | *Fv-2* | 12/24 | 50.0 |
| *Cck* | *Bgl-s* | 10/25 | 25.0 |
| *Bgl-s* | *Fv-2* | 5/24 | 7.6 (2.0-28.7) |
| *Hm-2* | *Fv-2* | 2/12 (BXH) | 5.6 (0.5-44.2) |

Genetic distance in cM between pairs of loci calculated from the formula $\hat{r}=\dfrac{R}{(4-6R)}$ , where R is the proportion of recombinant strains.   95% confidence intervals are shown in parenthesis when appropriate for pairs in which linkage is definitely established

other markers distal and proximal of *Cck*. Figure 4.6 shows the number of recombinants between SDPs of these and two other markers *Fv-2* and *Bgl-s* (Meisler, 1976), both of which have been typed across the BXDs. *Fv-2* is mapped 13cM and *Bgl-s* 6cM from *Cck*, and the order *Fv-2_Bgl-s_Cck* is established (Hillyard *et al.*, 1991). The possible arrangements with respect to *Hm-2* are therefore:

> 1) (cen)_ *Fv-2_Bgl-s_* *Cck* _*Hm-2*_Ms-TF_(ter)
>
> 2) (cen)_ *Fv-2_Bgl-s*_Ms-TF_*Hm-2_Cck*_(ter)

From figure 4.6 it can be seen that Ms-TF and *Hm-2* mapped a greater distance from *Bgl-s* than *Cck*, so arrangement 1 seems likely. Arrangement 1 also requires fewer recombinations between all the markers: Arrangement 2 requires 19 recombinations between *Fv-2* and *Cck* one of which is a triple; arrangement 1 requires only 18 between *Fv-2* and Ms-TF.

The distance between the *Cck_Hm-2*_Ms-TF grouping and the loci *Fv-2* and *Bgl-s*, estimated from the SDPs across the BXD RI strains was very large; the confidence limits for the distances were such that non-linkage was not excluded. This is contrary to the established genetic distance between the *Cck* and *Bgl* loci on chromosome 9, on which *Bgl* is an 'anchor' locus and to which *Cck* has been localized by *Spretus* backcross and somatic cell hybrids (Friedman *et al* 1989). Moreover, the human homologues of *Cck* and *Bgl* are present on the same region of human chromosome 3 (3pter-3p21, Hillyard *et al* 1991). The larger than expected map distance is an unfortunate coincidence caused by large numbers of recombinations in this region in the BXD strains, and is an effect of using RI strains which although show linkage, cannot, except when the number of recombinants is low and the number of strains very high, localize a marker with confidence to a small region (Silver, 1985).

## 4.6    Relationship of BXD microsatellite alleles to BXD *Hm-2* alleles

The sizes of *Hm-2* alleles associated with the C57BL/6 microsatelite allele were compared with the sizes of *Hm-2* alleles associated with the DBA type microsatellite allele (figure 4.7). It was seen that there was a bimodal distribution of *Hm-2* allele sizes; DBA type BXD mice had small *Hm-2* alleles (average 4.2kb), and C57BL/6 type BXD mice had large *Hm-2* alleles (average 9.4kb). This was the distribution expected from allele sizes in the parental

**Figure 4.7    Relationship of BXD microsatellite alleles to BXD *Hm-2* alleles**

**A**    Size distribution of C57BL/6 type *Hm-2* alleles and DBA type *Hm-2* alleles in BXD RI mice.

**B**    Southern blot of genomic BXD DNA as described in figure 4.1. *Hm-2* alleles are labelled ( □ ) if the mouse had the C57BL/6 microsatellite allele, or ( ● ) if the mouse had the DBA microsatellite allele.

Figure 4.7 Relationship between BXD *Hm-2* alleles and microsatellite alleles
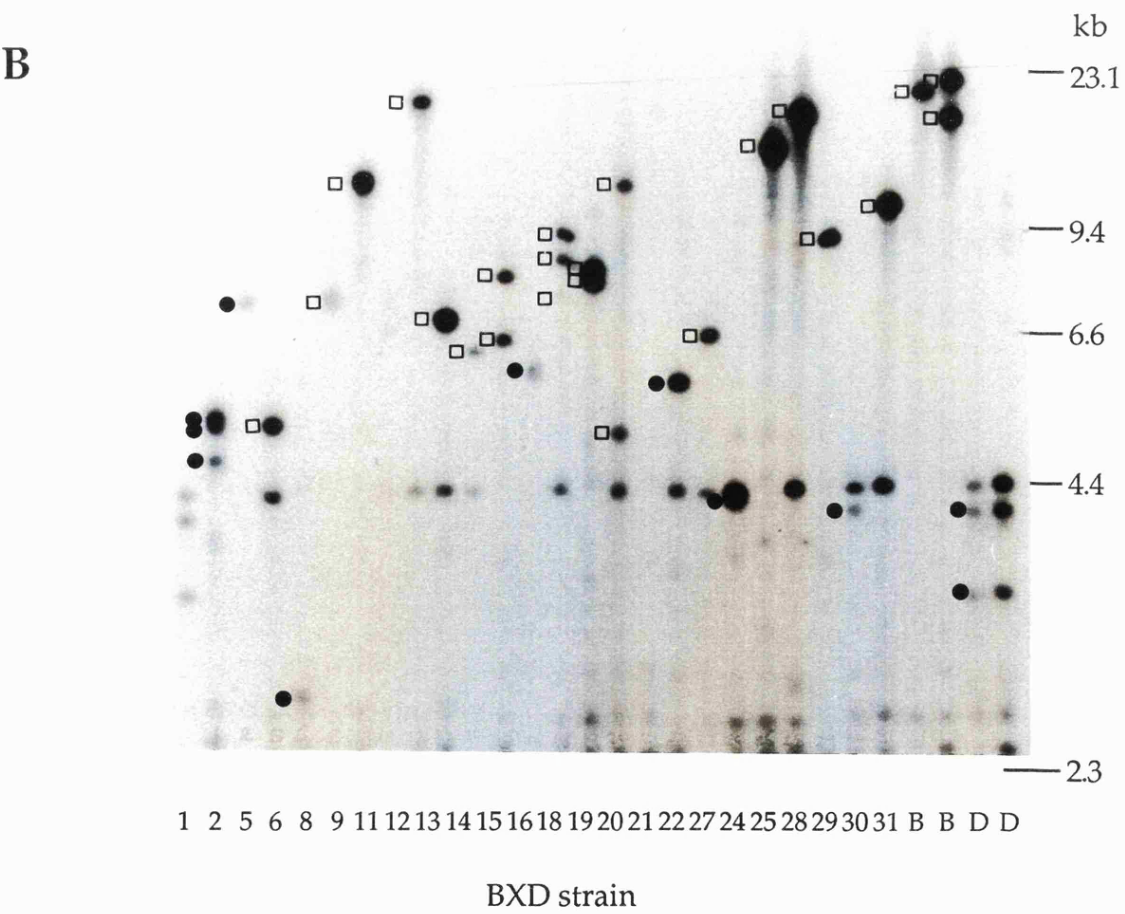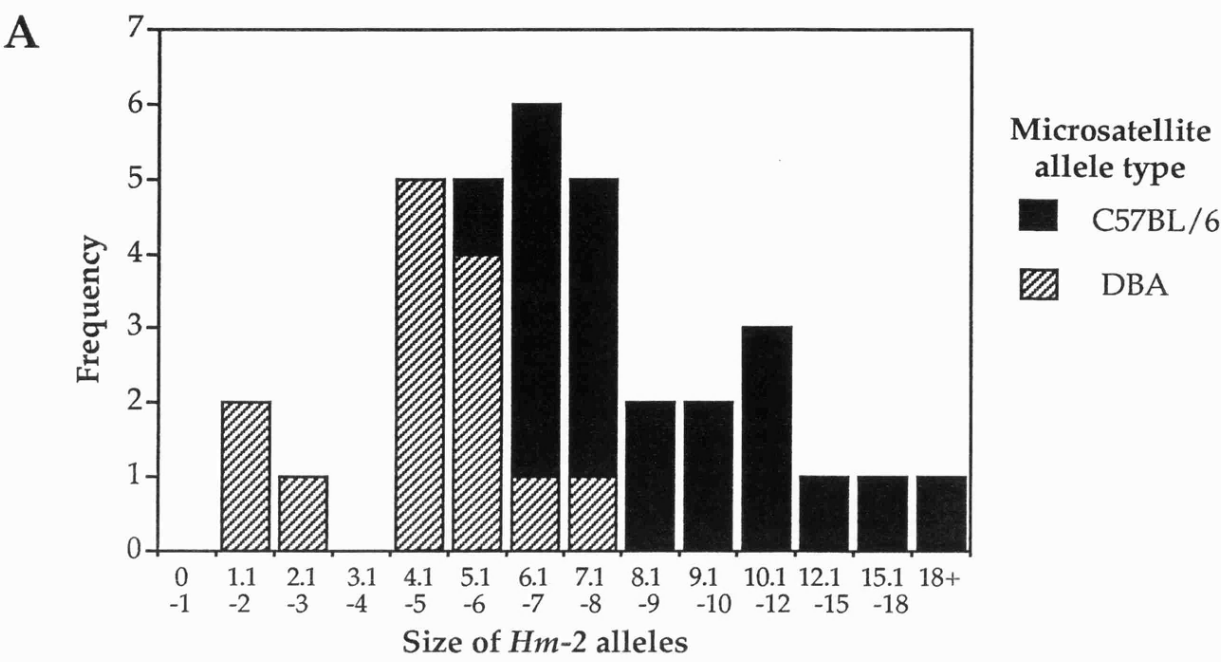
C57BL/6 and DBA strains, and validates the microsatellite SDP. The range of sizes, however, overlapped considerably: the largest DBA allele was 7.8kb, and the smallest C57BL/6 allele was 5.2kb, thus it would be impossible to construct an accurate SDP from allele sizes even with knowledge of the average sizes of C57BL/6 and DBA alleles.

## 4.7    Summary

*Hm-2* has been localized to the distal region of chromosome 9, near *Cck*. Although the ordering of the locus with respect its nearest neighbours will be accurate, the actual map distances obtained are unreliable. To pin-point the location of *Hm-2*, further genetic crosses need to be done.

# 5

# Rates of Germline and Somatic Mutation at *Hm-2*

## 5.1    General Introduction

The use of minisatellites as extremely variable and informative markers has relied on their variability in size, i.e. the number of tandem repeats in an allele, rather than any internal variation in the sequence of those repeats. The variation in minisatellite allele lengths is a direct result of mutations that change the repeat copy number of an allele. The possible mechanisms by which these changes arise will be discussed in chapter 8. In this chapter, the rates of occurrence of mutations that change allele length at *Hm-2* are evaluated.

Mutations can occur in either the germline or the soma. A germline mutation is seen in a mother-father-offspring trio, in which the parentage is beyond dispute, by the child inheriting a non-parental allele. Such a mutation has been spontaneously generated in the germline or gametes of the parent.

A new mutation in the soma will be inherited by all subsequent daughter cells. Its representation in the total cell population will depend on the number of descendant cells formed from this single mutant cell, which will in turn depend on the time the mutation first arose and the developmental fate of the mutant daughter cells. If the mutation occurs in a terminally differentiated cell then no further cells will be generated. A mutation in an adult stem cell, such as a haemopoietic stem cell, will continue to generate mutant cells for the life of the cell, leading to a small proportion of mutant cells in the blood. A mutation arising early in development will spread into a variety of different lineages and could possibly make up the majority of cells in a particular tissue. The organism in all these cases will therefore be mosaic for normal cells and mutant cells. However, unless the mosaicism extends to the germline, the mutation will not be inherited by any of the progeny, and therefore the new allele will not contribute to population variability. A mutation which occurs at a stage of development at which mutant daughter cells can still be allocated to both the

germline and soma will be mosaic both in the soma and germline. There is then the potential for mutant germ cells to be inherited and for the mutation to spread into the population. Unless, however, a mosaic individual transmits the mutant allele to offspring, it is not possible to determine whether the mutation affects the germline as well as the soma.

The processes of somatic and germline mutation, which may be identical or different, are linked together by their potentially overlapping effects on allelic variability in a population. Therefore, to distinguish these, but maintain the connection between them, this chapter has been subdivided into two sections. **5a Germline mutation** characterizes those mutations where a non-parental allele has been inherited instead of a parental allele. These are mutations which have arisen in, or were present in, the parental germcells or gametes. **5b Somatic mutation**, characterizes those mutations where two parental *Hm-2* alleles have been inherited, and in addition, a non-parental *Hm-2* allele is present. The mutant allele might also be present in the germcells. The detection of mosaicism assumes that the somatic mutant allele is sufficiently different in size from both parental alleles to be resolvable on an agarose gel, and that the dosage of the mutant allele is sufficient for a signal to be detected by Southern blot hybridization of total genomic DNA.

# 5a    Germline mutation at *Hm-2*

## 5a.1    Estimation of mutation rate from variability in BXD recombinant inbred strains

A population's heterozygosity in terms of allelic length variability is determined by the mutation rate to new length alleles at the locus and the effective population size. Thus a small population requires a high mutation rate to maintain variability at a given locus. Assuming an infinite allele model, heterozygosity (H) and mutation rate ($\mu$) are related by the equation

$$(1) \qquad H = \frac{4N_e\mu}{1+4N_e\mu}$$

where $N_e$ is the effective population size (Kimura and Crow, 1964). $N_e$ is related to the level of inbreeding ($\Delta F$) such that $N_e = 1/2 \, \Delta F$ (Falconer, 1960). For full sib mating the rate of inbreeding settles down to a constant value of 0.191 after four generations, giving a value for $N_e$ of 2.6. Thus, within an inbred strain the mutation rate may be estimated from the rearranged formula (1) with $N_e = 2.6$:

$$(2) \qquad \mu = \frac{H}{10.4(1-H)}$$

H is estimated from the number of heterozygous mice seen in an inbred population. Only strains where full sib mating has been strictly maintained can be used to estimate $\mu$, since any departure would raise the value of $N_e$. For this reason the BXD RI strains, which have been bred by strict brother - sister mating, were used.

The wide variety of allele lengths seen in the BXD series is evidence of high levels of mutation. The rate of mutation at *Hm-2* has been sufficient to mutate the original allele sizes of the parental C57BL/6 and DBA mice to create a spread of allele sizes in the BXD strains ranging from 1.2 to 21.0kb. Of 26 mice

examined in the BXD series, 7 had two *Hm-2* alleles of equivalent intensity, and were therefore presumed heterozygous (chapter 4, figure 4.1). This gives a heterozygosity (H) of 0.27, and from formula (2) the estimate of $\mu$, the germline mutation rate to new length alleles is calculated to be 0.036 per gamete.

## 5a.2   Direct estimation of germline mutation rate from mouse pedigrees

The mutation rate at highly variable human minisatellite loci previously studied has been sufficiently high for the rate to be scored directly from human pedigrees (Jeffreys *et al.*, 1988a). Similarly the mutation rate at *Ms6-hm*, a highly variable murine minisatellite, has been measured by the analysis of C57BL/6 x DBA families (Kelly *et al* 1989). The same pedigree has been used here to estimate the germline mutation rate at *Hm-2*.

A pedigree of C57BL/6 x DBA $F_1$ and $F_2$ mice (bred by R. Kelly) was scored for germline mutations at *Hm-2*. Figure 5.1 shows an example of a germline mutant. When digested genomic DNA was run on a low percentage agarose gel to increase resolution, it was seen that $F_2$ 117 had a mutant *Hm-2* allele (both alleles had the signature expected for *Hm-2* alleles with the enzymes *Hinf*I, *Alu*I, *Mbo*I, and *Hae*III). The mutant allele was presumably derived from the father since the maternal allele appeared normal. Since small changes in allele length are believed to be most frequent, if it is assumed that the mutation involved only a single allele, then the progenitor allele would be the paternal allele closest in size to the mutant allele. For mouse $F_2$117, it was therefore assumed that the progenitor paternal allele was the father's larger (~19kb) allele rather than the small (3.5kb allele).

A total of 18 mutations to a non-parental size were scored in 196 mice. Of these, 14 were different in size and were dispersed amongst families in which the parentage had been confirmed at other loci, and were therefore assumed to be genuine, independent, germline mutations. The other four mutants are discussed in section 5a.4

A second group of families of C57BL/6 x CBA mice bred to investigate early somatic mosaicism (chapter 6) were also scored for germline mutations; 4 out of 61 day 10 embryos had germline mutations.

The estimation of the total germline mutation rate is 18 mutations seen in 257 mother/father/offspring trios, or 3.5% per gamete (95% limits 0.023-0.051), which agrees remarkably well with the rate estimated from heterozygosity in BXD RI mice. Since mutations involving a very small

**Figure 5.1** **Resolving germline mutations at *Hm-2* by electrophoresis on low percentage agarose gels**

A      Genomic Southern blot of one family of C57BL/6 x DBA $F_2$ mice. DNA digests with *Hinf*I. 0.5% agarose gel.

The parental *Hm-2* alleles are marked by (o). Note that $F_2$ 116 has one mutant allele (●) which is the result of a germline mutation. $F_2$ 121 and 123 have somatic mutations (figure 5.5)

B      Genomic DNA digested with the enzyme *Hinf*I from $F_1$ parents and mouse $F_2$ 117. A low percentage (0.5%) agarose gel has been used to increase the resolution of the larger fragments. The mutant paternal allele is labelled µ117

C      Genomic digests of mouse 117 with the enzymes *Hinf*I (H), *Alu*I (A), *Mbo*I (M), and *Hae*III (III).

# Figure 5.1 Germline mutation



kb

20 ——
15 ——
10 ——

7 ——

4 ——

A

60 41 | 116 **117** 118 119 120 **121** 122 **123** 124
♀ ♂
$F_1$          $F_2$



kb

25 —

20 — ♀→
   ♂→ ←μ117

15 —

B    60 41 117
      ♀ ♂
       $F_1$ $F_2$

kb

25 ——

20 ——

15 ——

10 ——

H A M III
$F_2$117
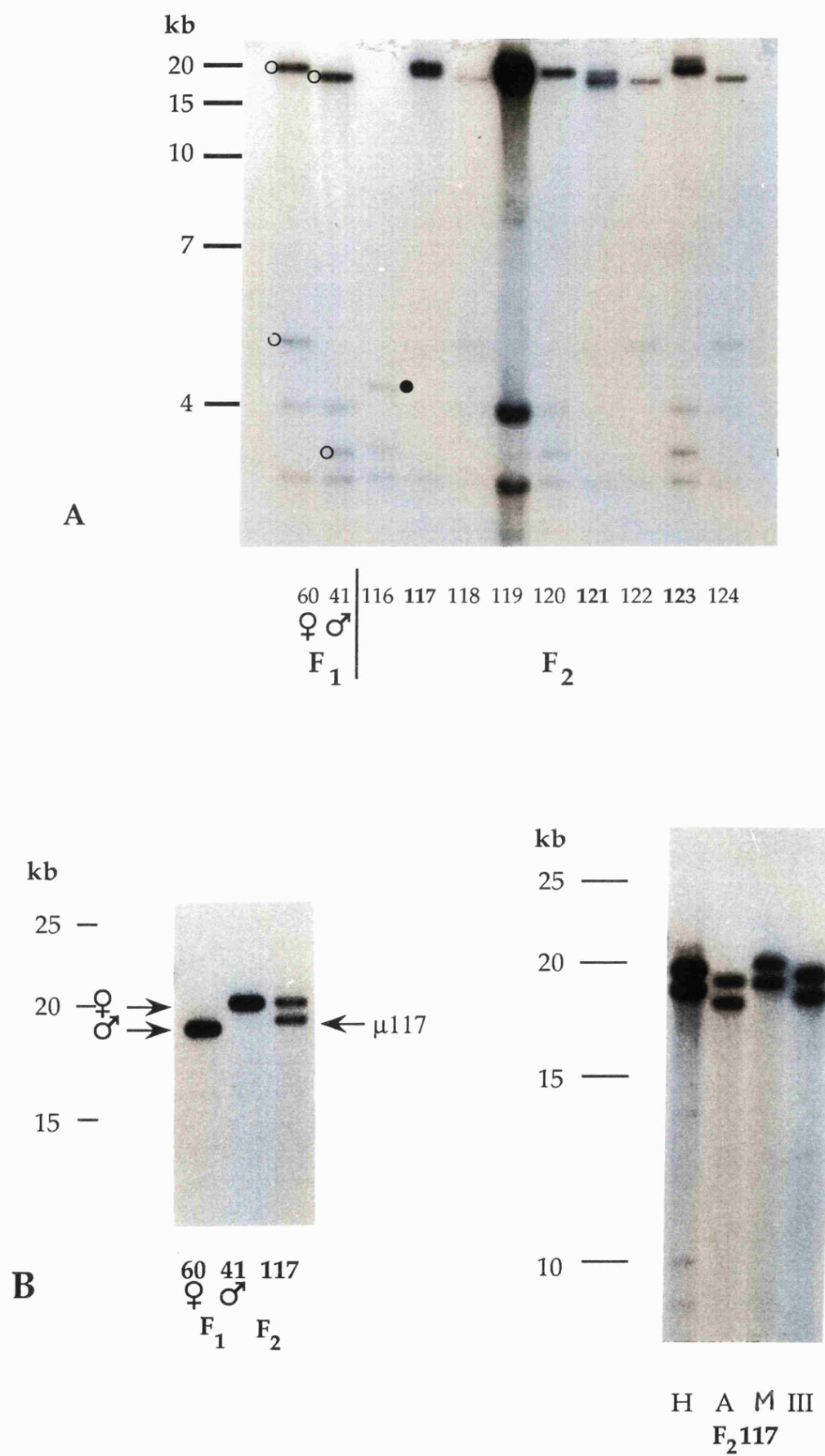
change in repeat copy number would not have been detected, this is therefore a minimum estimate of the incidence of gametes with non-parental alleles.

## 5a.3    Estimation of the contribution of somatic mosaicism

A somatic mutation that arises in early development may produce mosaicism in both the soma and germline, and could therefore be transmitted to offspring. In such cases the somatic mosaic is also a germline mutant. How often does a somatic mosaic produce an effect equivalent to a germline mutation, ie how often is a new mutant mosaic allele inherited?

For a mosaic mouse in which the dosage of the mutant allele is m (the proportion of cells with the mutant allele, estimated by densitometric scanning of autoradiographs), it is possible to calculate the germline transmission frequency (t), i.e. the proportion of progeny expected to receive the mutant allele assuming that the dosage in the germline is the same as the dosage in the soma, using the equation t=m/2.

Since several of the mutant C57BL/6 x DBA $F_1$ mice had been used to set up the $F_2$ matings, it was possible to determine whether the mutant mosaic allele was transmitted to the offspring, and to compare the rate of transmission to the rate expected from the somatic dosage of the mutant allele. The final column of table 5.2 gives the number of progeny that inherited the mutant allele for these mice. For 11 mosaic mice which had offspring, the mean dosage of the mosaic alleles was 0.152, giving an expected level of transmission of 0.152/2 = 0.076. In total, the 11 mice had 109 offspring, of which 8 had inherited the parental mutant allele, or an average transmission of 0.073, which agrees with the expected transmission ($\chi^2$=0.035, 0.9>p>0.8). Therefore, although the transmission of a mosaic allele from an individual mouse can differ significantly from expected (Kelly et al., 1991) (which presumably reflects chance sampling effects into a small pool of progenitor germcells), it is reasonable to use the average mosaicism over a large number of mosaic mice to estimate transmission frequency.

The mean dosage of mosaic alleles in 278 mice (mosaic and non-mosaic) was 0.048 (table 5.2), which gave a mean probability of transmission of a mosaic allele per mouse of 0.024.

## 5a.4 Further considerations

### 1) Complete replacement of parental allele by a somatic mutant allele

It is conceivable that some mutations that were thought to have arisen in the germline, are actually the result of somatic mutations which occurred in the early embryo. The population of mutant cells could have segregated by chance such that the embryo and hence the adult mouse was derived largely or entirely of mutant cells;- a "100%'mosaic". In the C57BL/6 x CBA families, however, in which DNA was analysed from both embryonic and extra-embryonic tissues of the offspring, germline mutations, which are seen as a complete replacement of one parental band in all tissues, can be distinguished from early somatic mutations which cannot replace parental type, non mutant cells, in all tissues. One embryo was seen which had a mosaic allele with a dosage of 100% in the embryo proper, whereas the extra-embryonic tissues still had the parental allele. Figure 5.2 shows embryo 15.1 which was a double mosaic. The larger mutant allele is derived from the paternal allele, the smaller mutant allele is derived from the maternal allele which is no longer visible by Southern blot hybridization of DNA from the embryo proper. (It is not possible to determine whether this mutant allele is present at low dosage in the yolk sac and trophoblast since the mutant allele comigrates with a faintly cross-hybridizing band.) If the extra-embryonic tissues had not been typed, this mouse would have been scored as a germline mutant. It is highly likely, although not certain, that this mouse was mutant in the germline as well as the soma, and would have transmited the mutant allele to its offspring.

### 2) The effect of germline mosaicism

A mutation that arises in the generation of the germline prior to the formation of gametes will result in germline mosaicism. It might then be possible for more than one sperm or oocyte to have identical mutant alleles, and thus the same mutant allele could be inherited by two or more offspring.

Only one possible instance of germline mosaicism in the absence of detectable somatic mosaicism was found in one family of $F_1$ mice (figure 5.3). Offspring should have all inherited the large (20kb) allele from the C57BL/6 father and a small (5.2kb) allele from the DBA mother. Mice 54, 57, 59, and 60, however, all had DBA alleles which were slightly bigger than the maternal allele, confirmed by the relative position of the band at 4.4kb (from a different

**Figure 5.2  A high dosage mosaic**
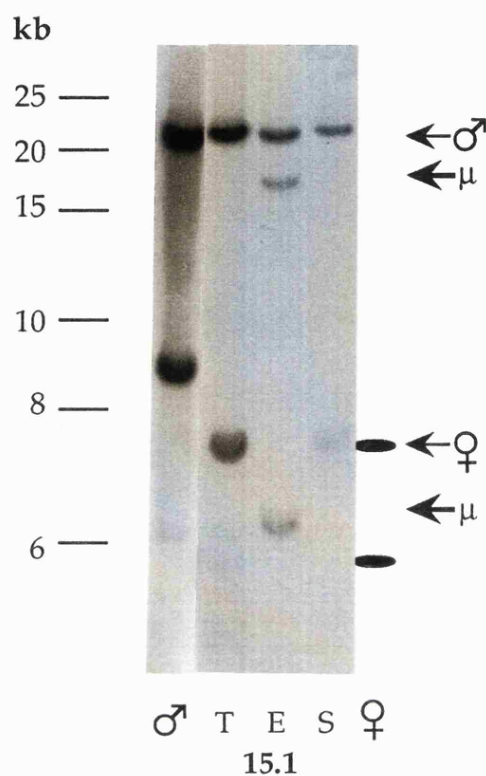


Lane 1, father,  Lanes 2-4, embryo 15.1.
T - trophoblast, E - embryo proper, S - yolk sac.
Southern blot hybridized to $^{32}$P-labelled 9.2 at high stringency.
Lane 5 shows the position of the maternal alleles

**Figure 5.3    Possible transmission of common mutant allele: evidence for germline mosaicism.**

DNA from a family of C57BL/6 x DBA mice (prepared by R. Kelly), digested with HinfI, and Southern blot hybridized with 9.2 at high stringency.

The $F_1$ family (**34-60**) should have all inherited the large (20kb) allele from the C57BL/6 father (**1**) and the small (5.2kb) allele from the DBA mother (**10**). (Both alleles marked ■)

Mice 54, 57, 59, and 60 have DBA alleles (□) which are slightly bigger than the maternal allele, confirmed by the relative position of the band from another cross-hybridizing locus at 4.4kb.

Note mice 34 and 36 are germline mutants and mice 37, 55 and 58 are mosaic for mutations affecting the paternal C57BL/6 allele (△).

**Figure 5.3 Possible transmission of a common mutant allele: evidence for germline mosaicism**
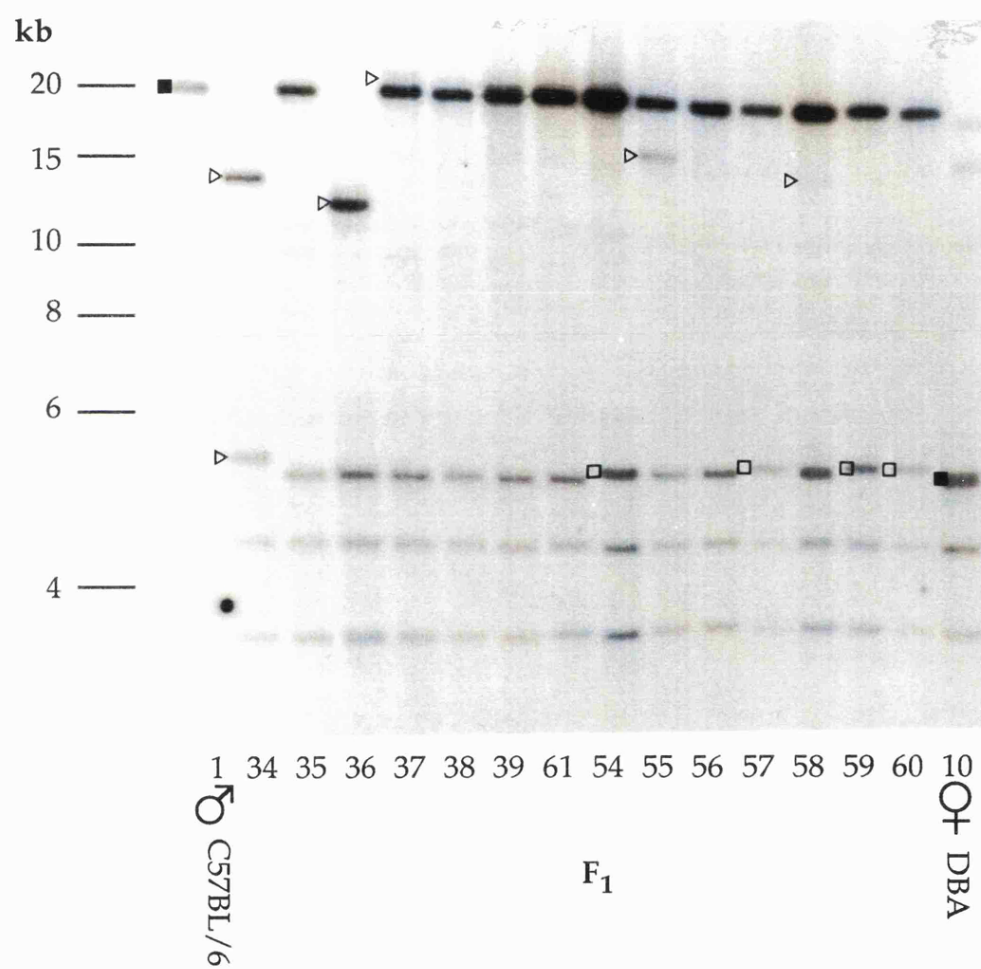
**Table 5.1**     Characteristics of germline mutation events at *Hm-2*

| Mouse | Sex | Progenitor parent | Progenitor allele repeats | | Change in No of repeats | % change |
|---|---|---|---|---|---|---|
| F₁ 34 | m | Male | 4690 | B | -1815 | -38.7 |
| F₁ 34 | m | Female | 1050 | D | +35 | +3.3 |
| F₁ 36 | m | Male | 4690 | B | -2115 | -45.1 |
| F₁ 77 | ? | Male | 4875 | B | -675 | -13.8 |
| F₁ 78 | ? | Male | 4875 | B | +525 | +10.8 |
| F₂ 13 | m | ? | 4325 | B | -2175 | -50.3 |
| F₂ 43 | m | Female | ✳ 550 | D | +15 | +2.7 |
|  |  |  | 5050 | B | -4485 | -88.8 |
| F₂ 47 | m | Male | ✳ 4750 | B | +400 | +8.4 |
|  |  |  | 1050 | D | +4100 | +390.5 |
| F₂ 70 | f | Female | ✳ 4625 | B | +225 | +4.9 |
|  |  |  | 1050 | D | +3800 | +361.9 |
| F₂ 74 | f | Female | ✳ 1050 | D | -100 | -9.5 |
|  |  |  | 4625 | B | -3675 | -79.5 |
| F₂ 95 | m | Female | ✳ 625 | D | +15 | +2.4 |
|  |  |  | 5050 | B | 4410 | -87.3 |
| F₂ 116 | f | Female | ✳ 1050 | D | -150 | -14.3 |
|  |  |  | 4690 | B | -3790 | -80.8 |
| F₂ 117 | f | Male | ✳ 4325 | B | +125 | +2.9 |
|  |  |  | 625 | D | +3825 | +612.0 |
| F₂ 130 | f | Female | ✳ 1050 | D | -100 | -9.5 |
|  |  |  | 4690 | B | -3740 | -79.7 |
| (F₁) 2.7 | ? | Male | ◉ 1350 | C | -525 | -36.2 |
|  |  |  | 3725 | C | -2900 | -77.9 |
| (F₁) 4.2 | ? | Male | ✳ 1400 | C | -350 | -25.0 |
|  |  |  | 3350 | C | -2300 | -68.7 |
| (F₁) 4.5 | ? | Male | ✳ 3350 | C | +450 | +13.4 |
|  |  |  | 1400 | C | +2400 | +171.4 |
| (F₁) 15.6 | ? | Female | ✳ 1250 | B | -475 | -38.0 |
|  |  |  | 1625 | B | -850 | -52.3 |

The folowing mice are possibly also germline mutants but not independent

| $F_1$ 54 | f | Female | ✳ 1050 | D | +15 | +1.5 |
|----------|---|--------|--------|---|-----|------|
|          |   |        | 4690   | B | -3625 | -77.3 |
| $F_1$ 57 | f | Female | ✳ 1050 | D | +15 | +1.5 |
|          |   |        | 4690   | B | -3625 | -77.3 |
| $F_1$ 59 | m | Female | ✳ 1050 | D | +15 | +1.5 |
|          |   |        | 4690   | B | -3625 | -77.3 |
| $F_1$ 60 | f | Female | ✳ 1050 | D | +15 | +1.5 |
|          |   |        | 4690   | B | -3625 | -77.3 |

**Table 5.1    Characteristics of germline mutation events at *Hm-2***

The 18 germline mutations are summarized showing the strain and sex of the affected mouse, and the direction and size of each mutation based on the size of the progenitor allele. The size of the progenitor parental allele is determined when the parent transmitting the allele is homozygous. When the parent is heterozygous, the parental allele closest in size to the mutant allele is taken to be the progenitor (shown by ✳ ) The allele marked ☉ was inherited from a mouse with one normal allele and one allele that had a polymorphic *Hinf*I site, and is the true progenitor allele, given that only one allele was involved in the mutation event.

*Mice*

> $F_1$    C57BL/6J x DBA $F_1$
> $F_2$    C57BL/6J x DBA $F_2$
> $(F_1)$  C57BL/6J x CBA F1

*Sex*

> m - male          f - female    ? - unknown

*Progenitor allele repeats*

Size of progenitor parent's alleles in number of GGCA repeats.

*Strain of progenitor allele*

> B    originally in C57BL/6J mouse
> D    originally in DBA mouse
> C    originally in CBA mouse

**Figure 5.4 Characteristics of germline mutation events at** *Hm-2*
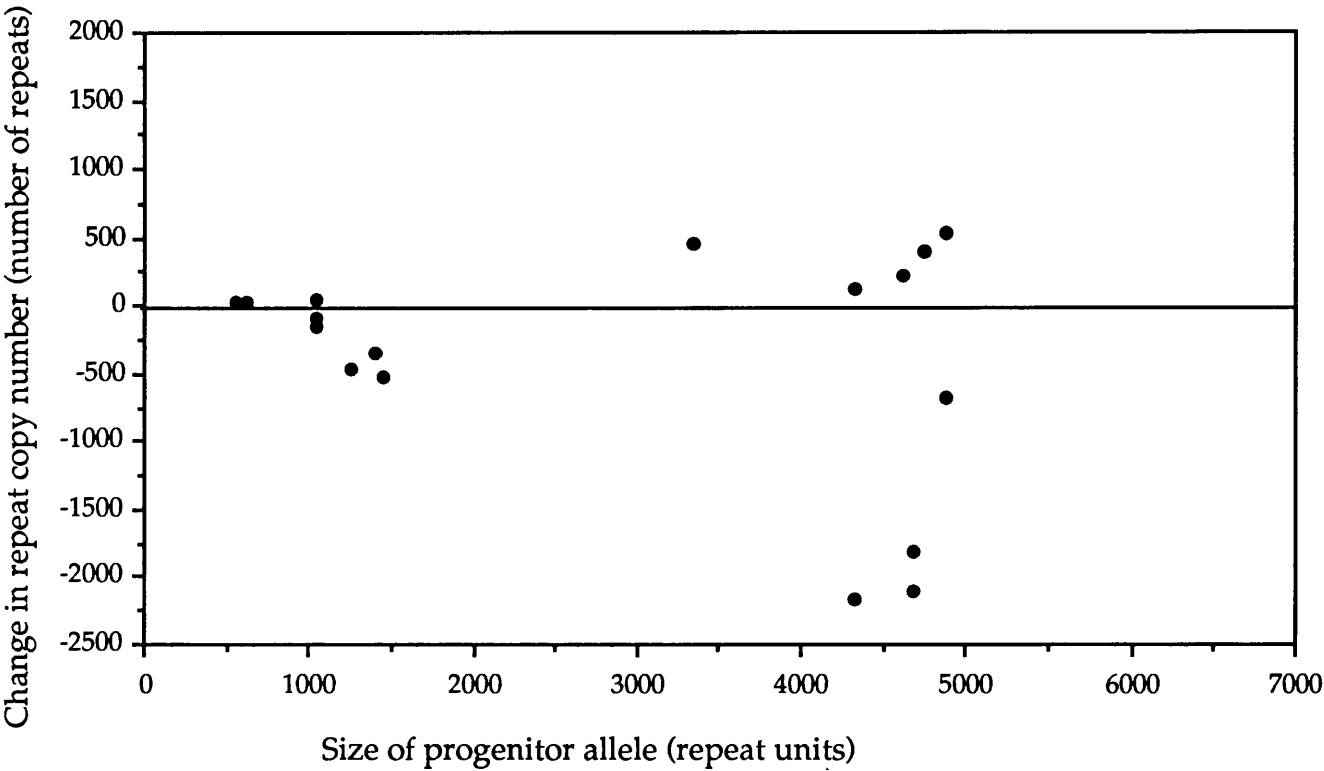


Relationship between the number of repeat units gained or lost in mutant *Hm-2* alleles and the number of repeat units in the progenitor allele. If the parent was heterozygous the progenitor allele was taken to be the parental allele closest in size to the mutant allele (marked with an asterisk in table 5.1)

cross-hybridizing locus), although the small differences in sizes are such that it is possible that the DBA female was heterozygous for these two alleles. Assuming these alleles are non-parental, they might either be the result of four germline mutations which, by coincidence, had the same gain in repeat copy number, or the result of DBA mouse number 10 being a germline mosaic, though not a detectable somatic mosaic. These mice were not included in the germline mutation totals, because of the uncertainty surrounding the real genotype of the mother

## 5a.5  Characteristics of germline mutations

Table 5.1 lists the germline mutants scored at *Hm-2* in adult and embryonic mice. The parental origin of the mutant allele could be identified, in all cases bar one in which the mouse had parents with identical heterozygous genotypes, by the loss of either the maternal or paternal allele. When the parent was homozygous, the size of the progenitor allele was also known. When the parent was heterozygous, such as in the $F_1$ generation, it was assumed that the allele closer in size to the new mutant allele was the progenitor allele, since it has been seen previously that small changes in allele length are more frequent (Jeffreys *et al.*, 1988a).

The mutations did not show any bias towards the gain or loss of repeats (10 lost, 8 gained), and were equally likely to involve an allele inherited from the mother as an allele inherited from the father (8 maternal, 9 paternal). The change in repeat copy number varied from 15 repeats to 2175 repeats, with small changes being more common. The largest changes in repeat copy number occurred at the largest alleles (figure 5.4), though when the change in repeat copy number is expressed as a percentage of the progenitor allele size the difference was not so apparent. The largest changes in size were all decreases in the number of repeat, mean change = -370 repeats, or -12.9% of the allele. Larger decreases than increases in allele size were also observed at the locus *Ms6-hm* (Kelly *et al.*, 1991).

## 5a.6  Summary

The frequency with which non-parental alleles arise in offspring gives a mutation rate of 3.6% per gamete at the *Hm-2* locus. Mutations are sporadic

and show no inclination towards the creation of larger or smaller alleles. It is possible that a proportion of germline mutants are a result of mosaicism in the parental germcells. There is also an input of new length mutant alleles into the population from alleles seen as somatic mosaicism, that were also present in the germline of the mosaic mouse. The probability of inheriting such an allele, averaged over all mice, has been estimated as 2.4% per mouse.

# 5b Somatic mutation at *Hm-2*

## 5b.1 Introduction

DNA from a mosaic tissue will contain both the parental alleles and non-parental mutant allele(s). In extreme cases in which a large proportion of cells are mutant, the mutant allele will be visible by hybridization in addition to both the parental alleles. This is an extremely rare occurrence for human minisatellite loci so far studied; only one case of human minisatellite mosaicism has been observed in over 6000 profiles (A.J.Jeffreys, pers comm).

In mice such 'three banded' individuals have been observed for the *Ms6-hm* (Kelly *et al.*, 1989) and *Hm-2* loci. These mice have a large proportion of mutant cells which is an indication that the mutation may have occurred during very early development. This will be considered in chapters 6 & 7. Here the rate of somatic mutation creating mice which are demonstrably mosaic by Southern blot hybridization of genomic DNA is evaluated, and the nature of the somatic mutations are determined.

## 5b.2 Evidence for somatic mosaicism at *Hm-2*

Somatic mosaicism was first observed at the locus *Ms6-hm* whilst screening for germline mutations (Kelly *et al.*, 1989). It was immediately apparent that at least one of the other loci detected by pMm3-1 at low stringency, *Hm-2*, also showed mosaicism, seen by three bands which have the same signature pattern on an autoradiograph (figure 3.1). The isolation of a *Hm-2* specific probe enabled the screening for the presence of third alleles in the same C57BL/6 x DBA and C57BL/6 x CBA families used previously. The C57BL/6 x CBA families consisting of mother-father-embryo trios were equivalent to mother-father-adult offspring trios, assuming there is no selection against mutant cells during development, so these were also scored.

Additional alleles, seen on Southern blots of tail DNA from the C57BL/6 x DBA families, were shown to be genuine *Hm-2* alleles rather than

alleles from a cross-hybridizing locus by their flanking DNA signature on gels (chapter 3). Figure 5.5 shows examples of two mosaics; mouse 121 had two large parental alleles (4325 and 4750 repeats) and a smaller non-parental allele (4100 repeats). The flanking DNA signature of both parental and non-parental alleles were the same and matched the expected flanking DNA signature for *Hm-2* alleles, showing that the mouse was mosaic. The upper maternal allele was reduced in intensity when compared to the paternal allele, suggesting that the mutant allele was derived from the maternal allele. Mouse 123 inherited a small 625 repeat paternal allele (see figure 5.5A), and a large maternal allele (4750 repeats), and also had a non-parental allele of 5150 repeats. The relative intensities of maternal and paternal alleles were not comparable over such a large difference in size, but it was assumed that the maternal allele was the progenitor since it was closer in size to the mutant allele.

All potential mutant alleles that were tested by the four enzyme signature test were shown to be genuine *Hm-2* alleles. This suggests the specificity of the 9.2 probe was sufficient such that even though other background loci were detected, *Hm-2* was the only unstable locus detected, and any additional bands were highly likely to be from *Hm-2*. Therefore, although mutants in the C57BL/6 x CBA pedigree were only confirmed by at most one extra restriction enzyme, they could be reliably scored as genuine *Hm-2* somatic mutations.

### 5b.3    Estimation of the somatic mutation rate at *Hm-2*

Table 5.2 lists the somatic mutations scored at *Hm-2* in adult and embryonic mice. These are alleles which could be resolved from the parental alleles, and had a dosage sufficient to produce a signal on an autoradiograph.

The total number of mosaic alleles seen in 279 mice was 62, giving a mean of 62/279 = 0.222 mosaic bands per mouse. Assuming that the number of mosaic alleles per mouse follows a Poisson distribution with a mean of 0.222, then the expected frequency of mice with 0, 1, 2, 3 additional mutant bands can be estimated (table 5.3)

**Figure 5.5     Demonstration of mosaicism - Mice with three *Hm-2* alleles**

A     Genomic Southern blot (prepared by R. Kelly) of one family of C57BL/6 x DBA $F_2$ mice.  Mice 121 and 123 have an additional non-parental allele (●).

B     $F_1$ parents and $F_2$ progeny mice with increased resolution on a low percentage agarose gel. The mutant alleles are arrowed.

C     Genomic digests of mice 123 and 121  with the enzymes *Hinf*I (H), *Alu*I (A), *Mbo*I (M),& *Hae*III (III).

A, B and C hybridizations were with $^{32}$P-labelled 9.2 at high stringency.

Schematic of *Hm-2* alleles of mouse 121.  The intensities of the alleles can be used to determine the progenitor allele and the dosage of the mutant allele.

Bands
seen in
mouse
121     dosage

♀  �  0.55
♂  ▬  1.00
mutant  ▓▓▓  0.45
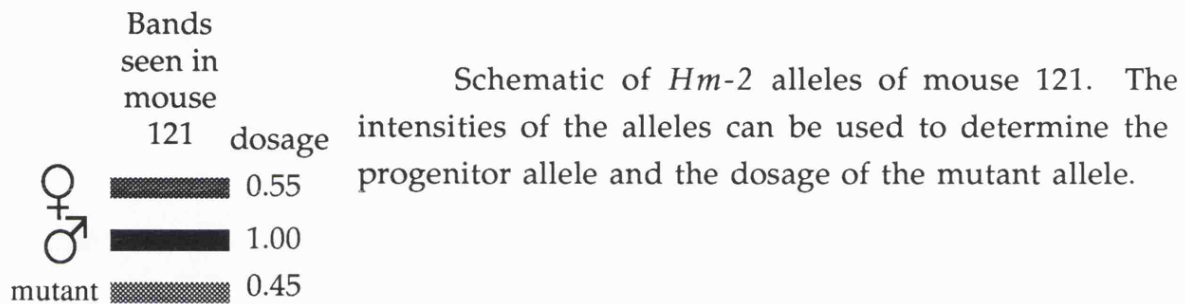
# Figure 5.5   Somatic Mosaicism



**A**

kb
20 —
15 —
10 —
7 —
4 —

| 60 | 41 | 116 | **117** | 118 | 119 | 120 | **121** | 122 | **123** | 124 |

♀ ♂

F₁   F₂



**B**

kb
25 —
20 —
15 —

← μ123
← ♀
← ♂
← μ121

| 60 | 41 | 117 | 121 | 123 |

♀ ♂

F₁   F₂



**C**

kb
25 —
20 —
15 —
10 —

| H | A | M | III | H | A | M | III |

F₂ 121   F₂ 123
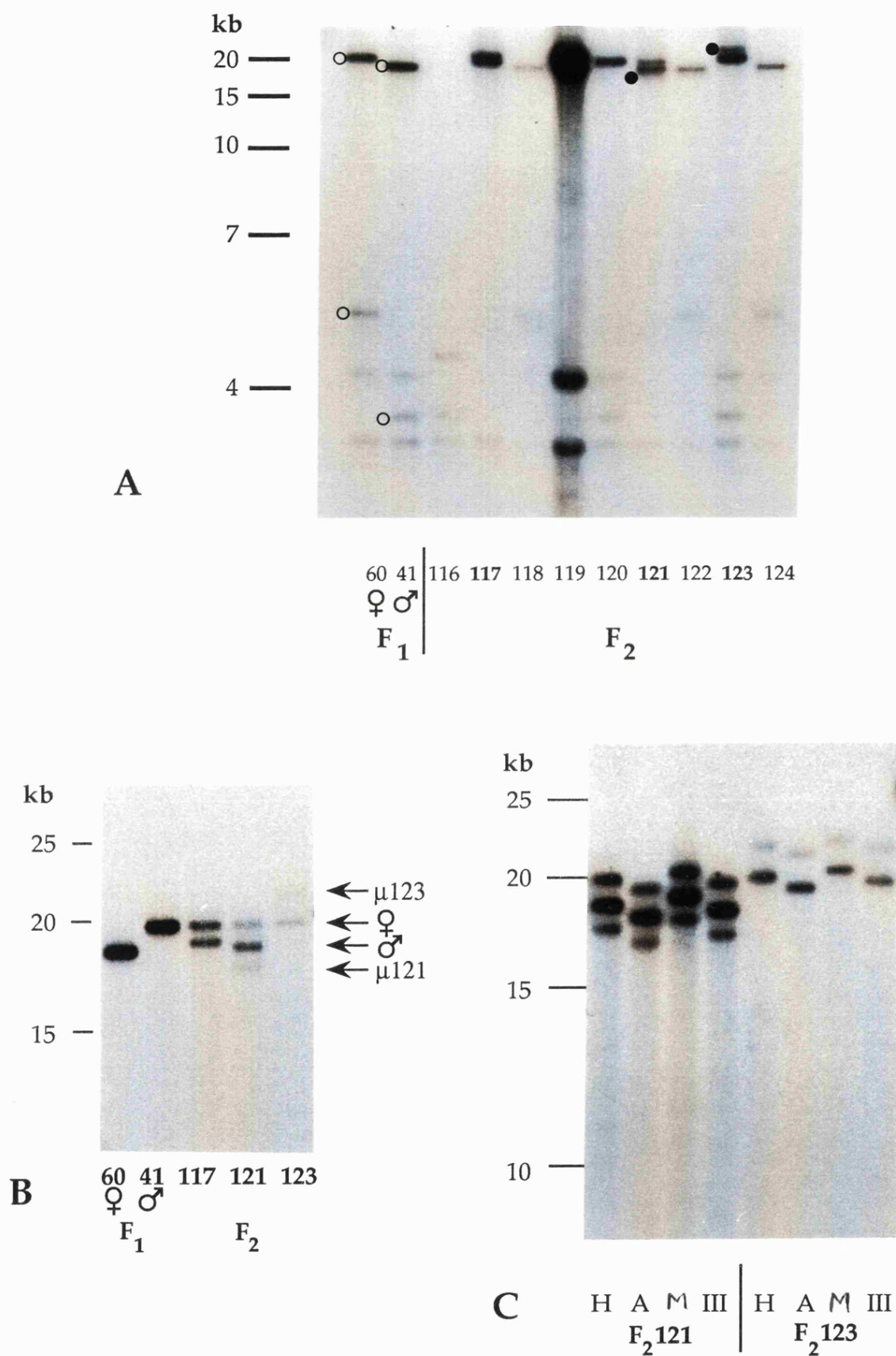
**Table 5.2**     Mice Mosaic at *Hm-2.*

| Mouse | Sex | Parental allele sizes | | | | Change in No of repeats | % change | Mosaic Allele Dosage | Inherited |
|---|---|---|---|---|---|---|---|---|---|
| BL/6J 1 | m | 4750 | | B | - | -300 | -6.3 | 0.04 | n.p. |
| | | 4625 | | B | - | -175 | -3.4 | | |
| BL/6J 3 | m | 4750 | | B | - | +875 | +18.4 | 0.51 | n.p. |
| | | 4750 | | B | - | +875 | +18.4 | | |
| BL/6J 4 | m | 4875 | | B | - | -800 | -16.4 | 0.45 | n.p. |
| | | 4875 | | B | - | -800 | -16.4 | | |
| BL/6J 7 | m | 5625 | | B | - | -1125 | -20.0 | 0.24 | n.p. |
| | | 4875 | | B | - | -375 | -7.7 | | |
| F₁ 33 | m | 625 | | D | P | +3425 | +548.0 | | |
| | | 4325 | * | B | M | -275 | -6.4 | 0.02 | 0/7 |
| F₁ 37 | f | 4750 | * | B | P | +600 | +12.6 | 0.03 | 0/22 |
| | | 1050 | | D | M | +4300 | +409.5 | | |
| F₁ 39 | m | 4625 | * | B | P | +175 | +3.8 | 0.28 | 2/7 |
| | | 1050 | | D | M | +3750 | +357.1 | | |
| F₁ 40 | f | 625 | | D | P | +2775 | +444.0 | | |
| | | 4325 | * | B | M | -925 | -21.4 | 0.07 | 0/13 |
| F₁ 47 | m | 625 | | D | P | +3775 | +604.4 | | |
| | | 4750 | * | B | M | -350 | -7.4 | 0.24 | 0/7 |
| F₁ 50 | m | 625 | | D | P | +3975 | +636.0 | | |
| | | 4750 | * | B | M | -150 | -3.2 | 0.04 | 0/8 |
| F₁ 51 | m | 625 | | D | P | +4625 | +740.0 | | |
| | | 5050 | * | B | M | +200 | +4.0 | 0.32 | n.p. |
| F₁ 52 | f | 550 | | D | P | +6100 | +1109.1 | | |
| | | 4750 | * | B | M | +1900 | +40.0 | 0.26 | 5/14 |
| F₁ 55 | f | 4750 | * | B | P | -975 | -20.5 | 0.19 | 0/8 |
| | | 1050 | | D | M | +2725 | +259.5 | | |
| F₁ 58 | m | 4625 | * | B | P | -1300 | -28.1 | 0.03 | 1/7 |
| | | 1050 | | D | M | +2275 | +216.7 | | |
| F₁ 62 | m | 4875 | * | B | P | -375 | -7.7 | 0.34 | 0/8 |
| | | 1050 | | D | M | +3450 | +328.6 | | |
| F₁ 75 | f | 4875 | * | B | P | +200 | +4.1 | 0.18 | n.p. |
| | | 1050 | | D | M | +4050 | +385.7 | | |
| F₁ 80 | f | 4875 | | B | P | -3675 | +75.4 | | |
| | | 1050 | * | D | M | +150 | +14.3 | 0.25 | n.p. |
| F₁ 83 | m | 550 | | D | P | +5200 | +945.5 | | |
| | | 5050 | * | B | M | +700 | +13.9 | 0.05 | n.p. |
| BL/6 1 | f | 1325 | ☉ | B | - | -75 | -5.7 | 0.17 | 0/8 |
| | | 875 | | B | - | +375 | +42.9 | | |
| CBA 10 | m | 4800 | | C | - | -1450 | -30.2 | <0.05 | 0/24 |
| | | 2000 | | C | - | +1350 | +67.5 | | |

| Mouse | | Sex | Parental allele sizes | | | | Change in No of repeats | % change | Mosaic Allele Dosage |
|---|---|---|---|---|---|---|---|---|---|
| F₂ | 2 | m | 4325 | ? | B | P | -700 | -16.2 | 0.14 |
| | | | 4325 | ? | B | M | -700 | | |
| F₂ | 7 | f | 4325 | | B | P | -1075 | -24.9 | |
| | | | 3450 | ✳ | B | M | -200 | -5.8 | 0.07 |
| F₂ | 17 | m | 4325 | ? | B | P | +1875 | +43.4 | 0.07 |
| | | | 4325 | ? | B | M | +1875 | | |
| | | | 4325 | ? | B | P | +550 | +12.7 | 0.24 |
| | | | 4325 | ? | B | M | +550 | | |
| F₂ | 28 | m | 1050 | | D | P | +3450 | +328.6 | |
| | | | 4750 | ✳ | B | M | -250 | -5.3 | 0.04 |
| F₂ | 31 | m | 625 | | D | P | +4025 | +644.0 | |
| | | | 4750 | ✳ | B | M | -100 | -2.1 | 0.29 |
| F₂ | 37 | f | 625 | | D | P | +2975 | +476.0 | |
| | | | 4750 | ✳ | B | M | -1150 | -24.2 | 0.20 |
| F₂ | 41 | m | 5050 | ◉ | B | P | -650 | -12.9 | 0.29 |
| | | | 4625 | | B | M | -225 | -4.9 | |
| F₂ | 58 | m | 4875 | ✳ | B | P | +1100 | +22.7 | 0.13 |
| | | | 625 | | D | M | +5350 | +856.0 | |
| F₂ | 63 | f | 4875 | | B | P | -100 | -2.1 | |
| | | | 4325 | ◉ | B | M | +450 | +10.4 | 0.93 |
| F₂ | 76 | m | 2575 | ✳ | B | P | -200 | -7.8 | 0.24 |
| | | | 550 | | D | M | +1775 | +322.7 | |
| F₂ | 82 | f | 1050 | | D | P | +5075 | +483.3 | |
| | | | 6650 | ✳ | B | M | -525 | -7.9 | 0.17 |
| F₂ | 86 | m | 625 | | D | P | +3425 | +548.0 | |
| | | | 4625 | ✳ | B | M | -575 | -12.4 | 0.26 |
| | | | 625 | | D | P | +4600 | +548.0 | |
| | | | 4625 | ✳ | B | M | +650 | +14.1 | 0.02 |
| F₂ | 92 | f | 625 | | D | P | +800 | +128.0 | |
| | | | 1050 | ✳ | D | M | +375 | +35.7 | 0.07 |
| F₂ | 106 | m | 550 | | D | P | +4100 | +745.5 | |
| | | | 4750 | ✳ | B | M | -400 | -8.4 | 0.22 |
| F₂ | 107 | f | 550 | | D | P | +3850 | +700.0 | |
| | | | 4750 | ✳ | B | M | -350 | -7.4 | 0.06 |
| F₂ | 121 | f | 4325 | | B | P | -225 | -5.2 | |
| | | | 4750 | ◉ | B | M | -650 | -13.7 | 0.45 |
| F₂ | 123 | f | 625 | | D | P | +4525 | +724.0 | |
| | | | 4750 | ✳ | B | M | +400 | +8.4 | 0.26 |
| F₂ | 137 | m | 2875 | ✳ | B | P | +150 | +5.2 | 0.33 |
| | | | 625 | | D | M | +2400 | +384.0 | |

| Mouse | Sex | Parental allele sizes | | | | Change in No of repeats | % change | Mosaic Allele Dosage |
|---|---|---|---|---|---|---|---|---|
| F₂  145 | m | 1050 | | D | P | +2725 | +259.5 | |
| | | 3450 | ✳ | B | M | +325 | +9.4 | 0.02 |
| | | 1050 | | D | P | +2150 | +204.8 | |
| | | 3450 | ✳ | B | M | -250 | -7.2 | 0.63 |
| F₂  179 | ? | 625 | | D | P | +3625 | +580.0 | |
| | | 4750 | ✳ | B | M | -500 | -10.5 | 0.09 |
| F₂  181 | ? | 4325 | ✳ | B | P | -325 | -7.5 | 0.04 |
| | | 1050 | | D | M | +2950 | +280.0 | |
| F₁  1.1 | ? | 4225 | ✳ | C | P | -1050 | -24.9 | 0.06 |
| | | 1325 | | B | M | +1850 | +139.6 | |
| F₁  2.3 | ? | 3725 | | C | P | -2625 | -70.5 | |
| | | 1200 | ✳ | B | M | -100 | -8.3 | 0.06 |
| F₁  2.6 | ? | 3725 | ? | C | P | -575 | -15.4 | 0.04 |
| | | 2275 | ? | B | M | +875 | +38.5 | |
| F₁  4.1 | ? | 1400 | | C | P | -325 | -23.2 | |
| | | 1025 | ✳ | B | M | +50 | +4.9 | 0.57 |
| F₁  4.3 | ? | 3350 | ✳ | C | P | +750 | +22.4 | 0.09 |
| | | 1075 | | B | M | +3025 | +281.4 | |
| F₁  4.5 | ? | 3800 | ✳ | C | P | +200 | +5.3 | 0.16 |
| | | 975 | | B | M | +3025 | +310.3 | |
| | | 3800 | ✳ | C | P | -1000 | -26.3 | 0.02 |
| | | 975 | | B | M | +1825 | +187.2 | |
| F₁  4.7 | ? | 3350 | ✳ | C | P | -500 | -14.9 | 0.19 |
| | | 975 | | B | M | +1875 | +192.3 | |
| F₁  12.4 | ? | 4800 | ✳ | C | P | -575 | -12.0 | 0.15 |
| | | 1300 | | B | M | +2850 | +219.2 | |
| F₁  14.5 | ? | 2000 | ✳ | C | P | -150 | -7.5 | 0.38 |
| | | 1250 | | B | M | +600 | +48.0 | |
| F₁  14.6 | ? | 4800 | ✳ | C | P | +700 | +14.6 | 0.21 |
| | | 925 | | B | M | +4625 | +500.0 | |
| F₁  15.1 | ? | 5600 | ✳ | C | P | -1375 | -24.6 | 0.33 |
| | | 1625 | | B | M | +2600 | +160.0 | |
| | | 5600 | | C | P | -4200 | -75.0 | |
| | | 1625 | ✪ | B | M | -225 | -13.8 | 1.00 |
| F₁  16.1 | ? | 2000 | | C | P | -800 | -40.0 | |
| | | 1250 | ✳ | B | M | -50 | -4.0 | 0.34 |
| F₁  16.4 | ? | 2000 | ✳ | C | P | -175 | -8.8 | 0.07 |
| | | 925 | | B | M | +900 | +97.3 | |
| F₁  16.5 | ? | 2000 | ✳ | C | P | +200 | +10.0 | 0.11 |
| | | 925 | | B | M | +1275 | +137.8 | |
| F₁  18.1 | ? | 4800 | ✳ | C | P | -925 | -19.3 | 0.32 |
| | | 1450 | | B | M | +2425 | +167.2 | |
| F₁  18.4 | ? | 2000 | ✪ | C | P | -250 | -12.5 | 0.21 |
| | | 1450 | | B | M | +300 | +20.7 | |

# Table 5.2    Somatic Mosaics

Table details all mosaic mice found in the C57BL/6J x DBA2J and C57BL/6J x CBA pedigrees.

*Mouse*

   BL/6J=C57BL/6J founder
   F₁=C57BL/6J x DBA/2J F₁ or C57BL/6J x CBA F₁
   F₂=C57BL/6J x DBA/2J F₂

*Sex*

   m=Male,  f=Female,  ?=unknown

*Parental allele size*

   Allele sizes are given in number of GGCA repeats

*Parental allele strain and sex origin.*

   B=Allele originally in C57BL/6J mouse

   D=Allele originally in DBA/2J mouse

   C=Allele originally in CBA mouse

   P=Paternally inherited allele    M=Maternally inherited allele

   The allele which is believed to have mutated, based on on the smallest change in allele length, is marked with a ✳  Alleles which can be seen to have mutated by a loss in intensity (figure 5.5) are marked with a ◎.

*Change in number of repeats*

   Difference in size between the parental alleles and the mutant measured in numbers of repeats.   These differences are expressed as a percentage of the parental allele size in the following column.

*Mosaic Allele Dosage* (m,  the proportion of cells containing the mutant allele).

   For heterozygous mice the non-progenitor allele has a dosage of 1, and the progenitor and mutant allele have a combined dosage of 1. m is related to the intensities of the progenitor (A) and non-progenitor (B) parental alleles by the following equations, which correct the intensity of

the parental allele to the equivalent intensity for an allele the same size as the mutant allele:

$$m = \frac{i_m}{I_A(S_m/S_A)+i_m} = \frac{i_m}{I_B(S_m/S_B)}$$

For homozygous mice the parental size allele is a combination of progenitor and non-progenitor alleles, which together with the mutant allele have a total dosage of 2. Thus m is calculated from the equation:

$$m = \frac{i_m}{\frac{1}{2}(I(S_m/S)+i_m)}$$

I =   intensity of parental allele          $i_m$ = intensity of mutant allele

S =   size of parental allele          $S_m$ = size of mutant allele (in repeats)

*Inheritance*

The fraction of offspring that have inherited the mutation is given. n.p. = no progeny available.

**Table 5.3    Distribution of mosaic alleles in mice**

| No. of mosaic alleles | Observed no. of mice | Expected no of mice |
|:---:|:---:|:---:|
| 0 | 222 | 223.4 |
| 1 | 52 | 49.6 |
| 2 | 5 | 5.5 |
| 3 | 0 | 0.4 |
| Total | 279 | 278.9 |

The observed and expected values agree closely. ($\chi^2$ 2df = 0.57, 0.8>p > 0.7), suggesting that mutations occur independently. The total incidence of mosaicism per mouse is 57/279 = 20.4% of mice. (95% confidence limits 15.4 - 24.2%, calculated assuming that the Poisson distribution tends to normality)

## 5b.4    Characteristics of Somatic mutations

The mutations involved both gain and loss of repeats, although there were significantly more losses than gains (22 gained, 39 lost, 1 unknown, $p(\leq 22$ gains) = 0.020) Changes involving large numbers of repeats were restricted to larger alleles, although the change in number of repeats as a percentage of the length of the progenitor was more even across all sizes of alleles. Figure 5.6 shows progenitor allele size plotted with change in repeat copy number.

The progenitor allele can be definitively determined in only a minority of cases, either from flanking DNA polymorphism (see section 3.7), or by reduction in intensity of the progenitor parental allele corresponding to its replacement in a large proportion of cells by the mutant allele (figure 5.5). In the majority of other cases, the mutant allele was close in size to one parental allele only, and this was assumed to be the progenitor allele.

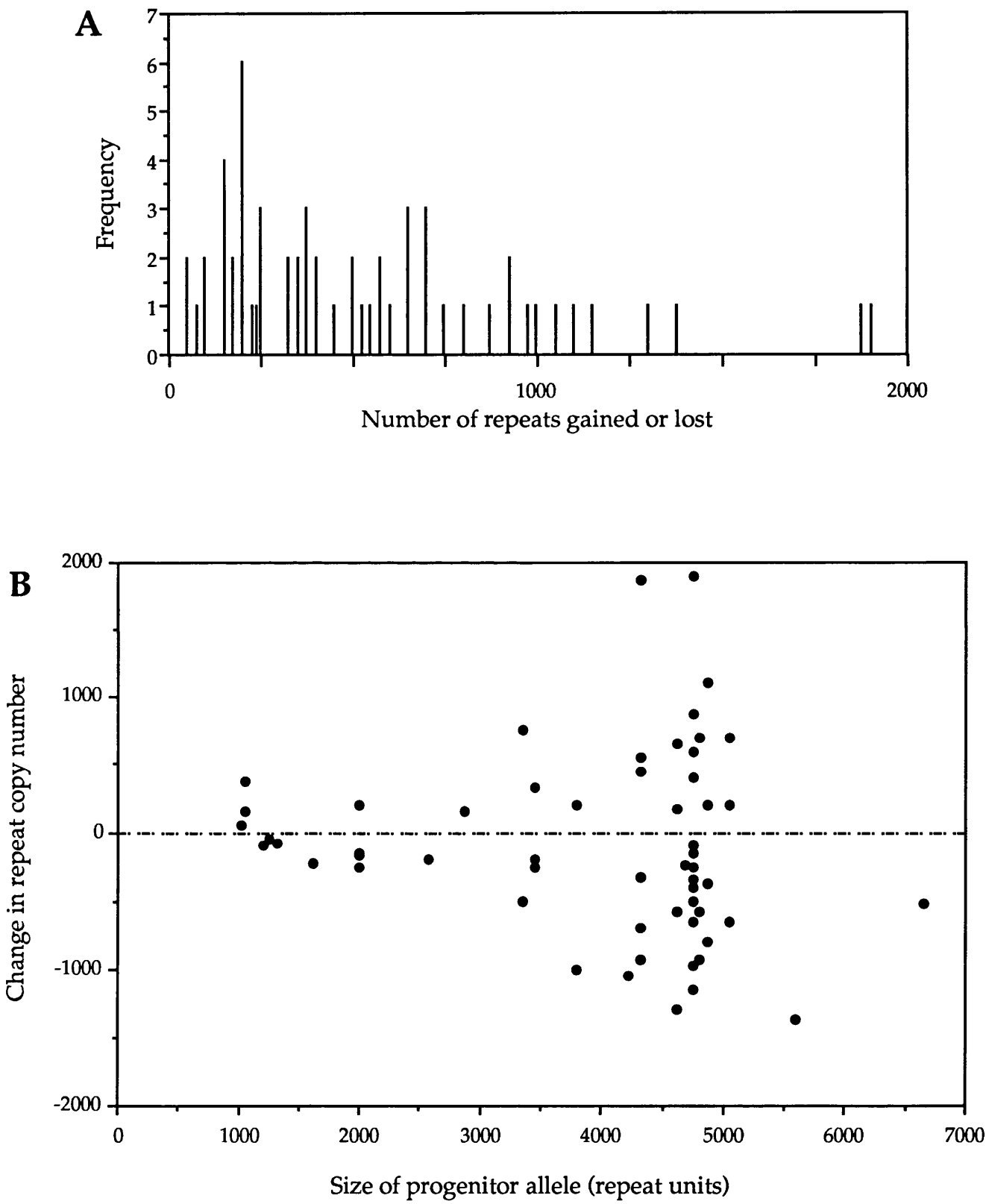There were significantly higher numbers of somatic mutations observed in $F_1$ mice when compared with $F_2$ mice. 32 mutations arose in 220 $F_1$ alleles, and 24 in 294 $F_2$ alleles, p<0.05. This was also the case at $Ms6\text{-}hm$ where all somatic mutations (7) in the DBA x C57BL/6 pedigree were in the $F_1$ or founder generations, whereas 221 $F_2$ mice had no detectable somatic mutations (p<0.025, Kelly 1990).

# Figure 5.6    Characteristics of somatic mutation events at *Hm-2*

A       Diagram summarizing the number of repeat units gained or lost during 62 somatic mutation events at *Hm-2*. The progenitor allele was taken to be the parental allele closest in size to the mutant allele (marked with a ✻ in table 5.2), or the allele which showed a decrease in intensity (marked with a ◉ in table 5.2)

B       Relationship between the number of repeat units gained or lost in mutant *Hm-2* alleles and the number of repeat units in the progenitor allele. Progenitor allele determined as above.

**Figure 5.6 Characteristics of somatic mutation events at *Hm-2***



A

Frequency (y-axis, 0 to 7)

Number of repeats gained or lost (x-axis, 0 to 2000)



B

Change in repeat copy number (y-axis, -2000 to 2000)

Size of progenitor allele (repeat units) (x-axis, 0 to 7000)

The degree of mosaicism was estimated by densitometric scanning of autoradiographs to determine the dosage of the mutant allele relative to the parental alleles (table 5.2), assuming that the signal intensity is proportional to the number of GGCA repeats.   Dosages ranging from 2% to 100% were scored in adult mice and embryos, although dosages of less than 5% were not expected to be detected except on best quality Southern blots
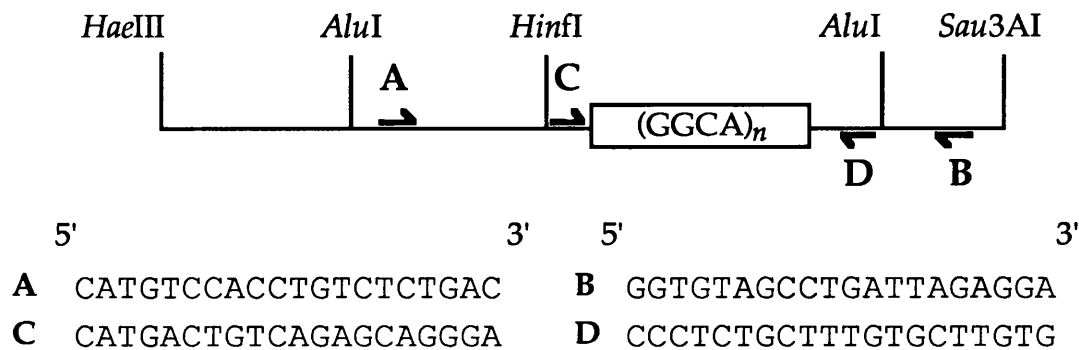
Several of the mutant C57BL/6 x DBA $F_1$ mice had been used to set up the $F_2$ matings, so it was possible to determine whether the mutant allele was transmitted to the offspring.  The final column of table 5.2 gives the number of progeny that have inherited the mutant allele.  It was seen that from 12 mosaic mice in which offspring could be scored, only 3 mice passed on the mutant allele to any of the progeny.  It is possible to calculate the germline transmission frequency (t), i.e. the proportion of progeny expected to receive the mutant allele assuming that the dosage in the germline is the same as the dosage in the soma, using the equation $t = \frac{m}{2}$ , where m is the proportion of cells with the mutant allele. Any significant deviation from the expected levels would suggest that the degree of mosaicism in the germline is different to the degree of mosaicism in the soma. In all cases the number of progeny inheriting or failing to inherit the mutant allele did not differ significantly from expected levels, although in the majority of cases the numbers of progeny observed were small.

## 5b.5    Application of PCR to the detection of mutant *Hm-2* alleles.

The Polymerase Chain Reaction can be used to amplify specific stretches of DNA, including minisatellites.   Thus PCR is a potential method of overcoming the problems of identifying *Hm-2* mosaic alleles against a background of cross-hybridizing loci.  Most minisatellites do not amplify well; due to the tandem repetitive nature of the sequence, collapse to heterogeneous smaller products occurs by misaligned annealing of incomplete PCR products as the reaction progresses.  Previously it has been shown that it is possible to amplify minisatellite alleles of up to 6kb to the point at which they become visible on an ethidium stained gel, and alleles of 10kb to a level at which they are detectable by cross-hybridization to ^{32}P-labelled DNA probes (Jeffreys *et al.*, 1988b).  The potential use of PCR for amplifying *Hm-2* alleles is therefore limited, since the majority of mice typed to date have alleles which are too

large to be amplified using current technology. To test the potential for PCR amplification, primers were designed flanking *Hm-2* (figure 5.7), and used to amplify genomic DNA from various mice.

**Figure 5.7    Location of primers for PCR amplification of *Hm-2***



| | 5' | | 3' | 5' | | 3' |
|---|---|---|---|---|---|---|
| **A** | CATGTCCACCTGTCTCTGAC | | | **B** | GGTGTAGCCTGATTAGAGGA | |
| **C** | CATGACTGTCAGAGCAGGGA | | | **D** | CCCTCTGCTTTGTGCTTGTG | |

Initial results were promising: alleles were amplified to a point at which they could be detected by hybridization to ^{32}P-labelled 9.2 probe (figure 5.8). The reactions collapsed a few cycles after this point, therefore it was not possible to amplify alleles to the point at which they could be visualized by ethidium staining.

## 5b.6    PCR amplification of *Hm-2* in BXD RI mice.

The BXD recombinant inbred series have been described in chapter 4. They provided a good specimen population with a range of known allele sizes determined by Southern blot hybridization of restriction digested genomic DNA.

Primers C and D were used to amplify *Hm-2* alleles in one mouse from each BXD strain (alleles ranged from 1.2 to 21.0kb, figure 4.1). Alleles of up to 12.8 kb were amplified from 100ng DNA to a level detectable by hybridization with the 9.2 probe (figure 5.9). Out of 26 DNAs, one from each of the 26 strains, 22 had alleles small enough to be amplified, and in all cases bands thought to be *Hm-2* from the Southern blot hybridization of restriction digested genomic DNA were amplified. Two mice, 1 and 18, previously thought to have three alleles from genomic Southern blot analysis, showed all three alleles amplified, proving mosaicism (data not shown).   A further four mice

**Figure 5.8** **Amplification of *Hm-2* alleles by the polymerase chain reaction (PCR)**
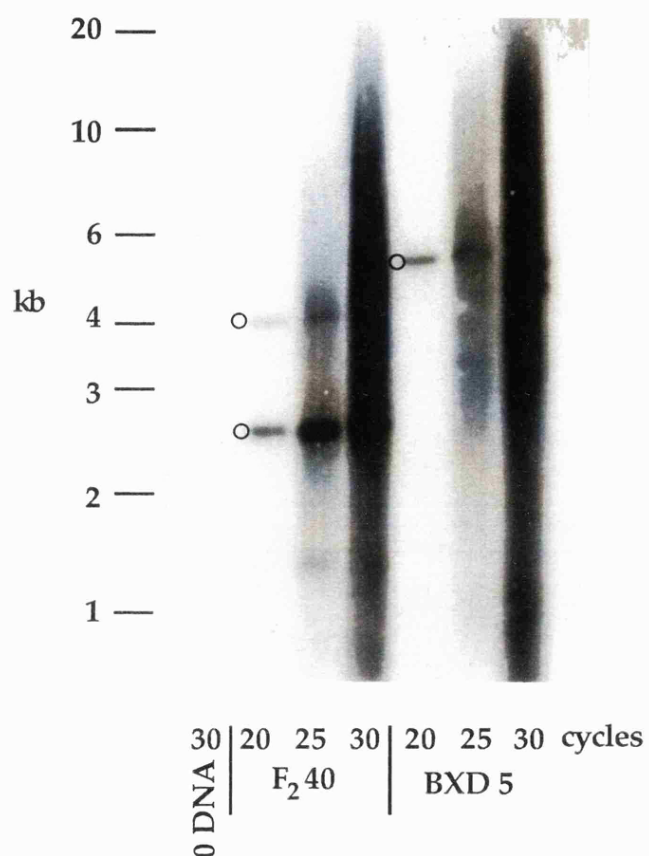


DNA (100ng) from C57BL/6 x DBA $F_2$ 40 and BXD 5 amplified using primers C and D in a 20μl reaction. 6μl aliquots were removed after 20, 25, and 30 cycles. Aliquots were electrophoresed through a 0.7% agarose gel and Southern blot hybridized with ^{32}P-labelled 9.2 at high stringency. Bands marked ○ match the size of *Hm-2* alleles seen on Southern blots of genomic DNA.

**Figure 5.9    PCR amplification of *Hm*-2 alleles from BXD genomic DNA**

DNA (100ng) from mice from BXD strains 32, 5, 8, 19, 31, and 9 were amplified using primers C and D in a 10µl reaction for 23 cycles and Southern blot hybridized with $^{32}$P-labelled 9.2 at high stringency.

Two exposures of the same filter are shown; lanes 32, 5, 31 and 19 are a 2 hr exposure, lanes 19, 31, 31 and 9 are a 16hr exposure. Mouse 19 is present in both exposures for comparison.  Sizes are of amplified bands .

Note that three bands amplified from BXD 32, suggesting that this mouse was mosaic.

**Figure 5.9 PCR amplification of *Hm*-2 alleles from genomic BXD DNA**



8.2
7.9
7.6
**kb**

5.2
4.8
4.5

1.8

12.8
10.2 **kb**
8.2
7.9

32   5   8   19 | 19   31   31   9   BXD strain

Short exposure | Long exposure
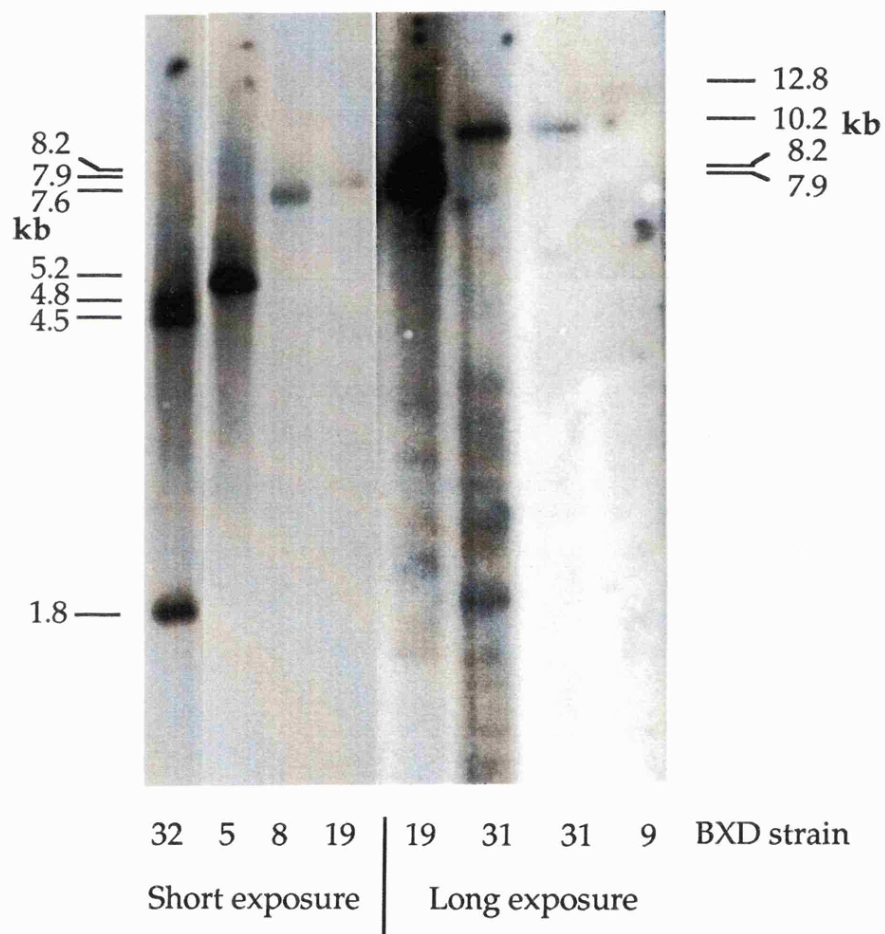
**Figure 5.10   Characterization of additional bands amplified from BXD mice.**

A      DNA (100ng) amplified using primers C and D in a 20μl reaction. 19 cycles, extension time reduced to 2 minutes.

An additional band (●) amplified in mice 6, 21, 30, and 32.

B      5μl of reaction products above denatured with formamide (ss) or left native (ds), electrophoresed through a 1% agarose gel and Southern blot hybridized with $^{32}$P-labelled 9.2 at high stringency.  Note that large fragments (o) have not transferred as well as in A.

The extra bands in A (●) from mice 30 and 32 denatured to two single stranded bands which have different intensities.

**Figure 5.10**
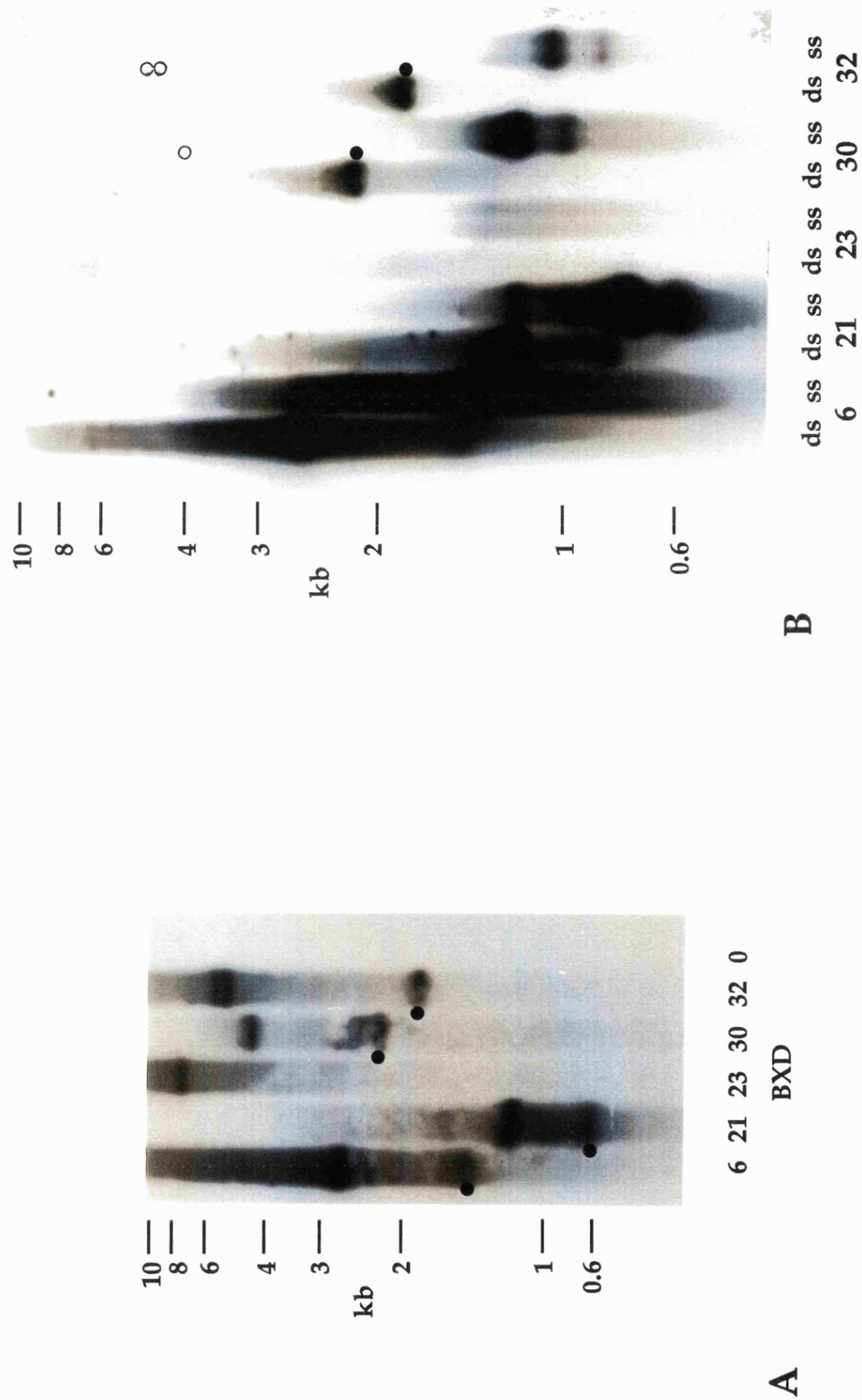
(6,21,30,32) had an additional band following amplification (figure 5.10A). These extra bands might also have been somatic mutations, or alternatively, they could have been PCR artifacts created by collapse of the minisatellite during amplification, or single stranded DNA products from incomplete reactions. To test whether additional bands were single stranded, reaction products were denatured, by the addition of formamide loading buffer and heating to 85°C for 5 minutes, before loading adjacent to undenatured PCR products. In two cases (mice 6 and 21) the sizes of extra bands seen in the double stranded lanes matched the size of one of the single strands and did not break down into different single stranded products themselves (figure 5.10B). These were therefore single stranded products, not somatic mutations. In the other two mice, 30 and 32, the additional bands did split into two single stranded products, and were therefore genuine double stranded products. These bands were amplified by the external primers (A & B) in addition to the standard primers, and repeat reactions continued to produce additional bands of the same size, suggesting that they were genuine alleles rather than collapsed products. It is possible, although unlikely, that these bands result from contamination of the original DNA samples, so these remain only candidate mosaics.

Four cases of somatic mosaicism were seen among this group of 26 BXD mice. Three of these were in mice with DBA type *Hm-2* alleles (determined from the CA repeat polymorphism, see chapter 4), and the mutations must have arisen from alleles smaller than 2000 repeats. The other mutant was scored as C57BL/6 type and had alleles of <2500 repeats. This was in contrast to the situation in the C57BL/6 x DBA pedigree where the mutant alleles were predominantly derived from large C57BL/6 alleles. It is possible, therefore, that a substantial number of mutations have been missed when scoring smaller DBA alleles in the C57BL/6 x DBA families. PCR amplification of a representative sample of mice with small alleles would confirm this.

## 5b.7    Analysis of size distributions of germline and somatic mutant alleles

Does the probability of mutation depend upon the size of the progenitor allele? Figures 5.11 and 5.12 show the size distributions of mutant alleles compared to the size distribution of the  alleles scored for mutation. An even distribution of germline mutations was seen, although the number of mutants

**Figure 5.11 Size distribution of alleles that have undergone germline mutation.**

A    Size distribution of progenitor *Hm-2* alleles which underwent germline mutation events.  Progenitor alleles determined as in figure 5.2.

B    Size distribution of *Hm-2* alleles sampled for germline mutation in C57BL/6,  DBA, CBA, C57BL/6 x DBA, and C57BL/6 x CBA inbred mice.

Figure 5.11

**A** Number of germline mutants at *Hm-2*

**B** Number of alleles sampled for germline mutation at *Hm-2*

Size of allele in repeat units

DBA
BL/6
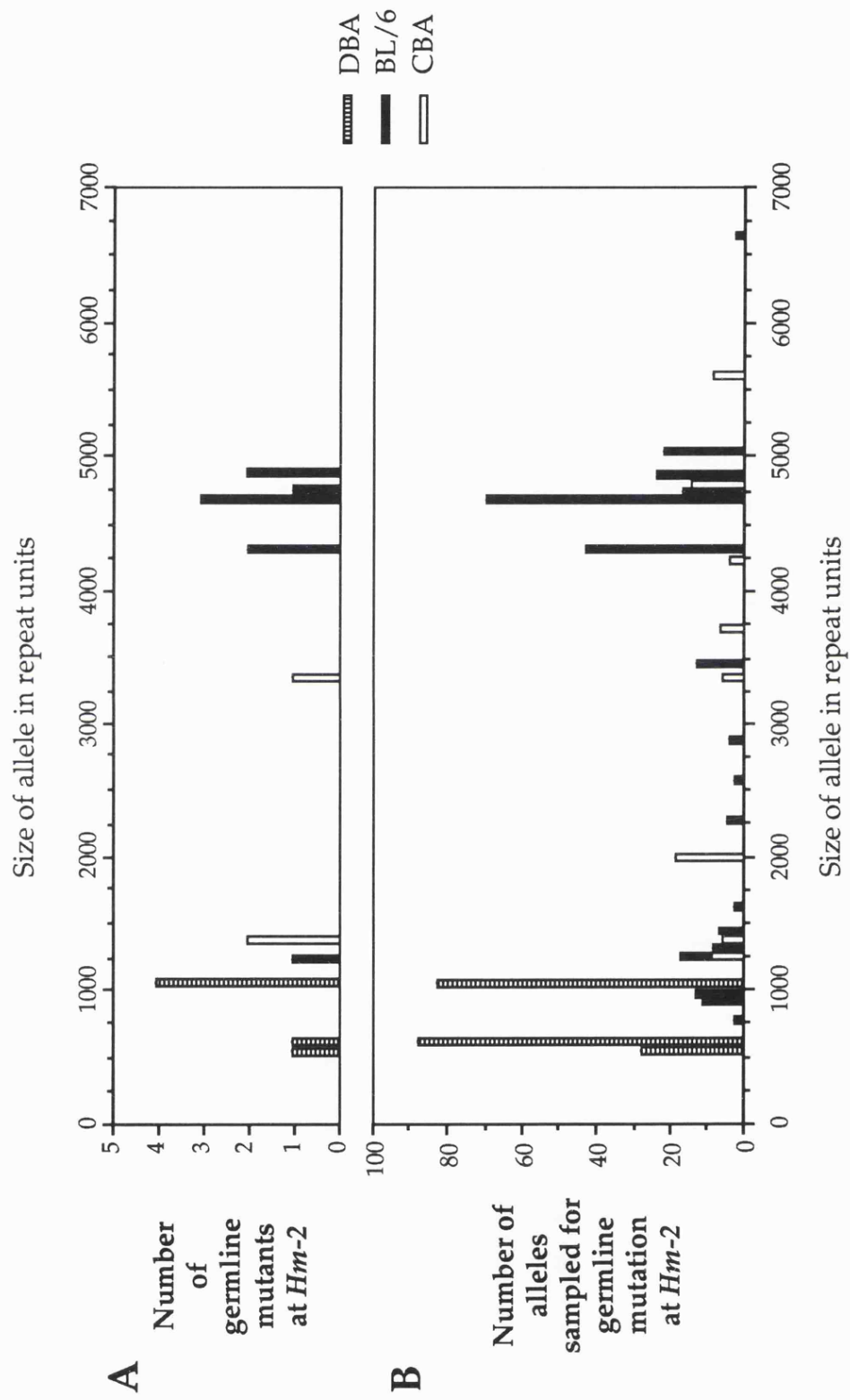CBA

**Figure 5.12  Size distribution of alleles that have undergone somatic mutation.**

A      Size distribution of progenitor *Hm-2* alleles which underwent somatic mutation events.  Progenitor alleles determined as in figure 5.6.

B      Size distribution of *Hm-2* alleles sampled for somatic mutation in C57BL/6,  DBA, CBA, C57BL/6 x DBA, and C57BL/6 x CBA inbred mice.

Figure 5.12
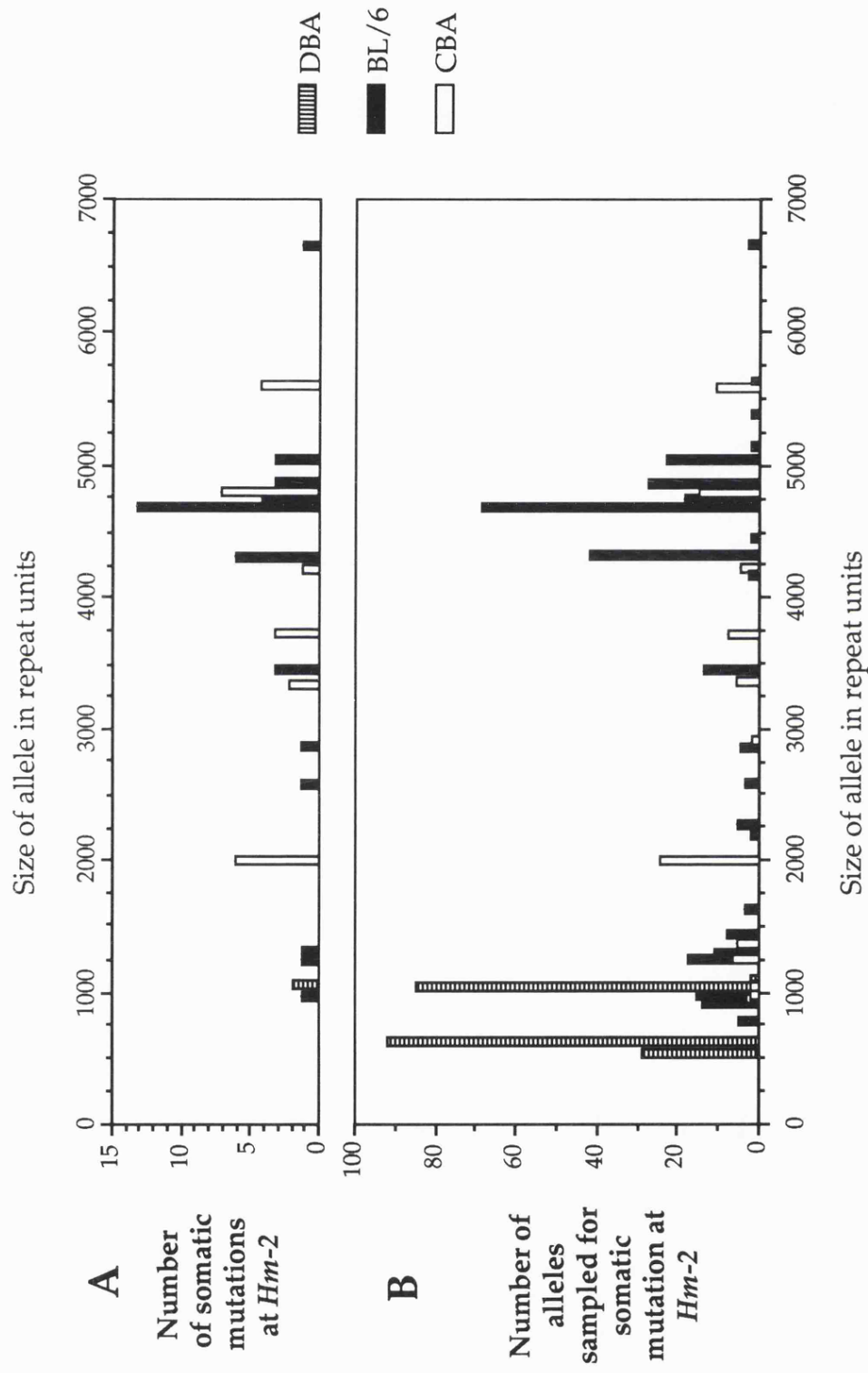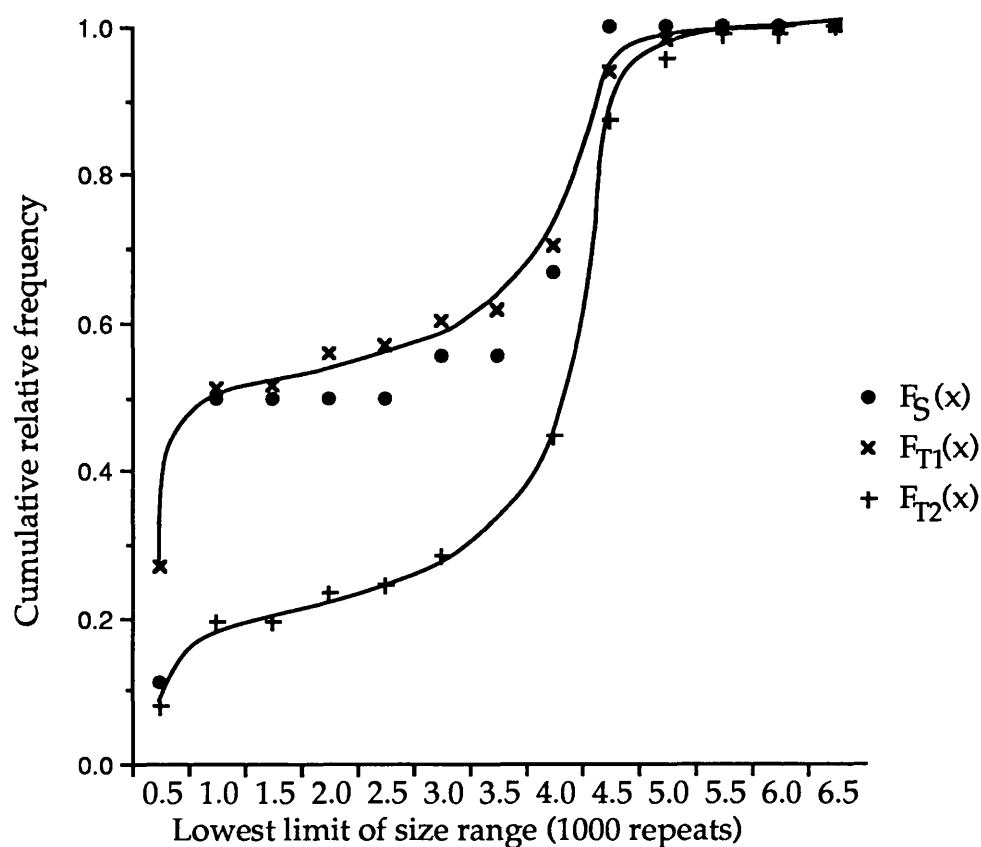


Size of allele in repeat units

DBA
BL/6
CBA

A

Number of somatic mutations at *Hm-2*

B

Number of alleles sampled for somatic mutation at *Hm-2*

# Figure 5.13 Analysis of size distribution of germline mutant alleles

## Cumulative frequency distribution of germline mutant alleles



Legend:
- $\bullet$ $F_S(x)$
- $\times$ $F_{T1}(x)$
- $+$ $F_{T2}(x)$

## Table 5.4 Calculation of Kolmogorov statistic

| x | $F_S(x)$ | $F_{T1}(x)$ | $F_{T2}(x)$ | $F_S(x)-$ $F_{T1}(x)$ | $F_S(x)-$ $F_{T2}(x)$ |
|---|---|---|---|---|---|
| 500-999 | 0.111 | 0.272 | 0.078 | **0.161** | 0.033 |
| 1000-1499 | 0.500 | 0.514 | 0.194 | 0.014 | **0.306** |
| 1500-1999 | 0.500 | 0.518 | 0.196 | 0.018 | 0.304 |
| 2000-2499 | 0.500 | 0.561 | 0.234 | 0.061 | 0.266 |
| 2500-2999 | 0.500 | 0.571 | 0.244 | 0.071 | 0.256 |
| 3000-3499 | 0.556 | 0.604 | 0.286 | 0.049 | 0.270 |
| 3500-3999 | 0.556 | 0.616 | 0.303 | 0.061 | 0.253 |
| 4000-4499 | 0.667 | 0.705 | 0.447 | 0.038 | 0.220 |
| 4500-4999 | 1.000 | 0.939 | 0.872 | 0.061 | 0.128 |
| 5000-5499 | 1.000 | 0.980 | 0.955 | 0.020 | 0.045 |
| 5500-5999 | 1.000 | 0.996 | 0.990 | 0.004 | 0.010 |
| 6000-6499 | 1.000 | 0.996 | 0.990 | 0.004 | 0.010 |
| 6500-6999 | 1.000 | 1.000 | 1.000 | 0.000 | 0.000 |

**Figure 5.14  Analysis of size distribution of somatic mutant alleles**

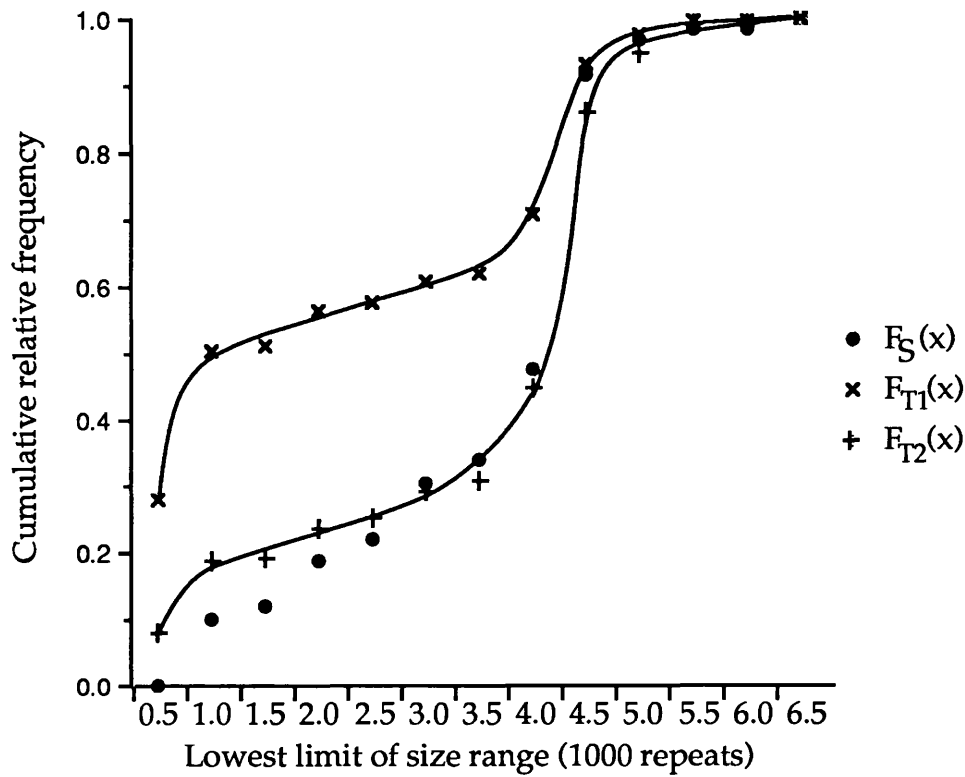Cumulative frequency distribution of somatic mutant alleles



**Table 5.5  Calculation of Kolmogorov statistic**

| x | $F_S(x)$ | $F_{T1}(x)$ | $F_{T2}(x)$ | $F_S(x)$-$F_{T1}(x)$ | $F_S(x)$-$F_{T2}(x)$ |
|---|---|---|---|---|---|
| 500-999 | 0.000 | 0.281 | 0.081 | 0.281 | 0.081 |
| 1000-1499 | 0.102 | 0.505 | 0.188 | **0.404** | **0.086** |
| 1500-1999 | 0.119 | 0.511 | 0.192 | 0.392 | 0.073 |
| 2000-2499 | 0.186 | 0.563 | 0.236 | 0.376 | 0.050 |
| 2500-2999 | 0.220 | 0.577 | 0.252 | 0.357 | 0.031 |
| 3000-3499 | 0.305 | 0.609 | 0.292 | 0.304 | 0.013 |
| 3500-3999 | 0.339 | 0.622 | 0.310 | 0.283 | 0.029 |
| 4000-4499 | 0.475 | 0.706 | 0.447 | 0.232 | 0.028 |
| 4500-4999 | 0.915 | 0.934 | 0.861 | 0.018 | 0.054 |
| 5000-5499 | 0.966 | 0.977 | 0.947 | 0.011 | 0.019 |
| 5500-5999 | 0.983 | 0.996 | 0.991 | 0.013 | 0.008 |
| 6000-6499 | 0.983 | 0.996 | 0.991 | 0.013 | 0.008 |
| 6500-6999 | 1.000 | 1.000 | 1.000 | 0.000 | 0.000 |

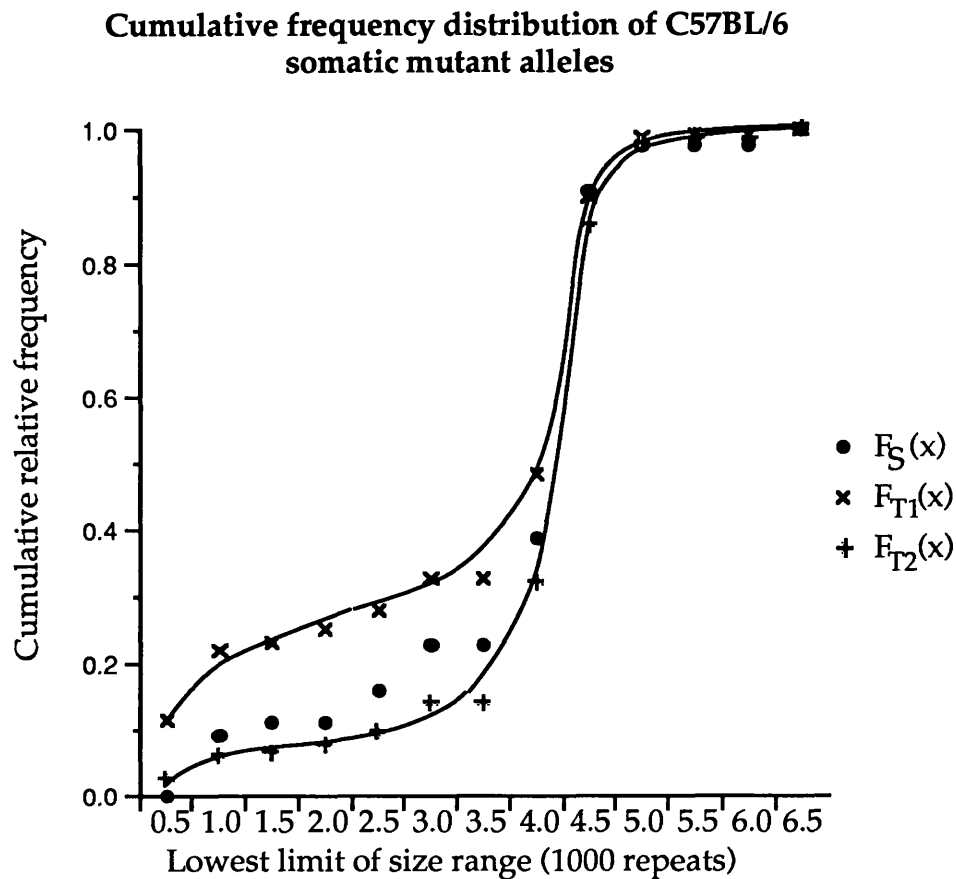**Figure 5.15   Analysis of size distribution of C57BL/6 somatic mutant alleles**

Cumulative frequency distribution of C57BL/6
somatic mutant alleles



**Table 5.5       Calculation of Kolmogorov statistic**

| x | $F_S(x)$ | $F_{T1}(x)$ | $F_{T2}(x)$ | $F_S(x)-$ $F_{T1}(x)$ | $F_S(x)-$ $F_{T2}(x)$ |
|---|---|---|---|---|---|
| 500-999 | 0.000 | 0.118 | 0.024 | 0.118 | 0.024 |
| 1000-1499 | 0.091 | 0.221 | 0.059 | 0.130 | 0.032 |
| 1500-1999 | 0.114 | 0.232 | 0.064 | 0.118 | 0.050 |
| 2000-2499 | 0.114 | 0.254 | 0.077 | **0.140** | 0.036 |
| 2500-2999 | 0.159 | 0.279 | 0.097 | 0.120 | 0.062 |
| 3000-3499 | 0.227 | 0.327 | 0.139 | 0.100 | **0.089** |
| 3500-3999 | 0.227 | 0.327 | 0.139 | 0.100 | **0.089** |
| 4000-4499 | 0.386 | 0.485 | 0.321 | 0.099 | 0.066 |
| 4500-4999 | 0.909 | 0.901 | 0.855 | 0.008 | 0.054 |
| 5000-5499 | 0.977 | 0.989 | 0.981 | 0.012 | 0.004 |
| 5500-5999 | 0.977 | 0.993 | 0.987 | 0.015 | 0.009 |
| 6000-6499 | 0.977 | 0.993 | 0.987 | 0.015 | 0.009 |
| 6500-6999 | 1.000 | 1.000 | 1.000 | 0.000 | 0.000 |

was small, whereas the somatic mutations appeared to cluster with large alleles.

To investigate possible size related effects on the mutation rate, the Kolmogorov/Smirnov goodness-of-fit test was used to compare the size distribution of mutant alleles to two theoretical distributions, which assumed the hypotheses:

1)The probability of mutation is independent of allele length.

2)The probability of mutation is proportional to allele length

The allele data was binned into 13 size categories, each containing alleles within a 500 repeat size range. The cumulative frequency distribution of the mutant alleles, $F_S(x)$, was compared to the theoretical cumulative frequency distributions $F_{T1}(x)$ and $F_{T2}(x)$, which were calculated from the distribution of scored alleles. Figure 5.13 shows the cumulative frequency distributions graphically, and tabulates the differences between the mutant and theoretical cumulative frequency distributions. The largest difference between the germline mutant and theoretical distributions gives the Kolmogorov test statistic, D, which is compared to standard tables to determine whether the sample distribution differs significantly from the theoretical distribution.

Comparison of the distribution of germline mutations to $F_{T1}(x)$ and $F_{T2}(x)$ produced the values $D_1$ (size independent mutation) of 0.161, and $D_2$ (mutation proportional to size) = 0.306. Neither values are significant at the 5% level, although $D_2$ is of borderline significance, $0.1 > p > 0.05$. Thus no correlation between size and the probability of germline mutation could be shown.

Similar calculations were carried out for the size distributions of alleles that underwent somatic mutations. Figure 5.14 shows the the distributions as described previously. The values of $D_1$ and $D_2$ in this case were 0.404 and 0.086 respectively. The $D_1$ value of 0.404 is significant ($p < 0.01$), and shows that the distribution of somatic mutant alleles does not agree with the distribution expected if mutation rate is size independent. Possible explanations for this difference are:

1) Somatic mutation rate is size dependant

2) The majority of small alleles were originally in DBA mice, and the majority of large alleles in C57BL/6 mice. Therefore there may be some factor,

other than allele length, specific to alleles from a particular strain that affects the probability of mutation.

3) Small mutant alleles are not reliably scored.

Since C57BL/6 mice had the widest range of allele sizes, the Kolmogorov test was carried out on the distribution of C57BL/6 mutations alone, as a single strain comparison. The cumulative frequency distribution of C57BL/6 mutations calculated (figure 5.15) did not differ significantly from either theoretical distribution.

Thus, the cumulative evidence against size dependant effects on the probability of mutation is:

1) Germline mutations, which are reliably scored even at small sizes, do fit well a size independent distribution, and the rate of mutation is not significantly different between DBA and C57BL/6 mice.

2) PCR does detect candidate mosaic alleles which are both small, and derived from small DBA alleles, contradicting the pattern seen on Southern blots.

3) The distribution of C57BL/6 mutations does fit a size independent distribution.

Thus a range of evidence has emerged which suggests that the low numbers of mutations at small alleles does not reflect any biological propensity, but rather is a by-product of the method of scoring mutations: somatic mutations which generate small alleles produce faint bands on Southern blots which are obscured against the background of cross-hybridizing loci seen in the low size range when hybridizing to 9.2.

## 5b.8    Summary

It has been estimated that 20% of mice are mosaic for cells with mutant and parental alleles at *Hm-2*. This estimate does not include mosaicism which is not discernible by Southern blot hybridization of genomic DNA, and is likely to be an underestimate, since both small changes in allele sizes, and mutant alleles of under 3kb are not reliably identified.

Alleles of up to 12kb can be amplified by PCR, and when mice are mosaic all three alleles amplify. It should be possible to utilize PCR to amplify both low level mosaics if they are present, and very small mutant alleles. This would provide further information on the estimated rate of somatic mutation.

Although fewer mutations were detected in small DBA alleles, there is no clear evidence for or against the idea that mutation rate increases with allele length.

# 6

# Somatic mutation during early development

## 6.1 Introduction

### 6.1.1 Evidence that suggests somatic mutation events occur in early development

This chapter investigates the time during development at which somatic mutation events occur. Certain features of mosaic alleles suggest that they are the products of somatic mutation events which take place early in development;

**a) Equal dosage of the non-parental allele in all tissues.**

Dissection of four mice mosaic at *Ms6-hm* has shown that the dosage of the mutant allele is indistinguishable in all adult tissues (Kelly *et al.*, 1991). This experiment has been repeated for two mice mosaic for *Hm-2* alleles with identical results (data not shown). This suggests that at both of these loci, the somatic mutation has preceded the allocation of the different tissue lineages.

**b) The mutant allele can be present in the germline.**

Either of the parental alleles or the non-parental allele can be inherited by the offspring of a mouse mosaic at *Ms6-hm* (Kelly *et al.*, 1989). Three mice mosaic at *Hm-2* also showed three-way segregation of parental and mutant alleles (table 5.2), showing that mosaicism is possible not only in the soma but also the germline, which has been determined by day-7 p.c. (McLaren 1991).

**c) The high dosage of mutant alleles.**

For a somatic mutation to be detectable in the DNA of an adult mouse, about 5% of the cells must contain the mutant allele (Kelly *et al.*, 1989). Such a mutation must, therefore, occur early enough for descendants of a single mutant cell to contribute substantially to adult tissues. Assuming that there has been no selective clonal proliferation of mutant cells, the mutation must have occurred when the total number of cells was small, i.e., at the first few cell divisions post fertilization.

Figure 6.1 Schematic representations of cleavage planes of 8-cell stage
polarized blastomeres and phenotypes of daughter cells produced.
(after Sutherland *et al.*, 1990)

PERICLINAL                                    ANTICLINAL

OBLIQUE

or

Polar cell                    Apolar cell
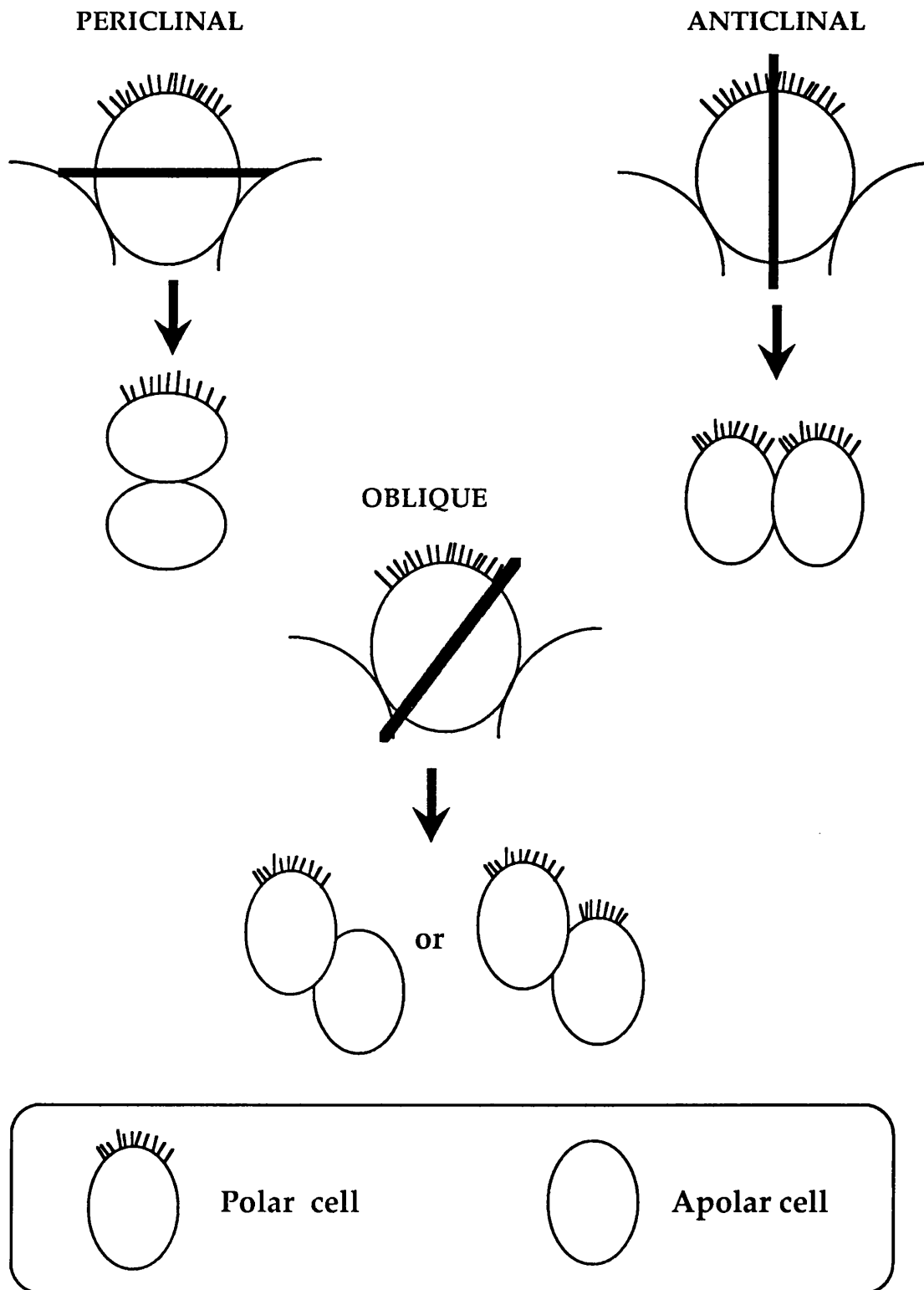
The trophoblast cells which surround the blastocoel cavity cease division and become polyploid. In contrast, the cells that remain in close proximity to the ICM (the polar trophoblast) continue to proliferate rapidly, eventually forming the chorion and placenta (reviewed by Gardner, 1983). Thus, the extra-embryonic tissues are derived from a subset of the trophoblast cells present at the 64-cell stage.

## 6.2    Analysis of extra-embryonic tissues.

The primary objective was to compare trophoblast and embryo proper DNA, derived from the same embryo, to find post-conception mutations common to the embryo and the extra-embryonic trophoblast, or to prove that such mutations did not occur. Any shared mutations would have to have occurred before the Trophoblast and ICM differentiated, at the 32-cell stage (figure 6.2).

Matings were therefore set up between CBA and C57BL/6J mice, chosen because previous mice typed had large (around 10kb) and very large (14-23kb) *Hm-2* alleles respectively. Alleles from both strains would, therefore, migrate to the upper regions of an agarose gel following electrophoresis, where the problem of background cross-hybridization of other loci would be minimized, whilst still being distinguishable in $F_1$ mice as coming from the CBA parent or the C57BL/6 parent. Moreover, it has been noticed previously that large alleles seem to have a higher probability of mutation (Jeffreys *et al.*, 1988a); if true, fewer animals would have to be analysed to find mutations.

Initial matings between C57BL/6 females and CBA males produced from 8 to 12 embryos (mean=9.5), whereas the reverse matings of C57BL/6 males and CBA females produced between 1 and 7 embryos only (mean=4.2). Therefore, subsequent matings were all between C57BL/6 females and CBA males.

Trophoblasts (T), embryos (E) and yolk sacs (S) from day-10 p.c. (post coitum, hereafter 'day-10') embryos were collected and DNA isolated from each tissue as described in chapter 2. Day-10 embryos were chosen as the stage at which it was still possible to separate cleanly the trophoblast from the maternal tissue, and the amount of tissue recovered was large enough to yield sufficient DNA for one or two Southern blots. (Once development proceeds past day-10 the trophoblast and maternal tissue become increasingly intermingled and difficult to separate cleanly.) DNAs were analysed by restriction digestion with the enzyme *Hae*III. If the quantity of DNA was sufficient, *Hinf*I digests were

also analysed. The amount of yolk sac DNA obtained was highly variable and never sufficient for more than one digest. Mosaic bands were confirmed as being mutant Hm-2 alleles by their cross-hybridization to the 9.2 probe, their absence from either parent, and when two separate enzymes were used, by having an expected size change with the second enzyme.

It was found that the C57BL/6 mice used in the matings had a different range of allele sizes to those observed previously. CBA alleles ranged from 5 to 23kb (1250-5600 rpts), and C57BL/6 alleles ranged from 3 to 9.7 kb (775-2275 rpts), rather than the the expected 14-23kb. This was not a problem, because few embryos inherited small C57BL/6 type alleles which migrated to the region of the gel where the background of other cross-hybridizing loci precluded the identification of shared TE or TES mosaicism. Mutations confined to a single tissue could still be seen in those cases, since in a non-mutant individual, all lanes are identical, and any extra bands in one lane are easily seen.

Some samples of trophoblast DNA were contaminated with maternal DNA from the placenta, seen by detection of the non-inherited maternal allele. This problem was eliminated in later experiments by a more rigourous separation of trophoblast from maternal tissue. Because the size of both maternal alleles was known, maternal contamination was not scored as a false trophoblast mutation. It is possible that some genuine mutations could exist in the trophoblast only, which by coincidence, are the same size as the non-inherited maternal allele; such mutants would not be scorable.

## 6.3 Evidence that somatic mutation events occur in early development

Each triplet (TES) or part triplet was scored for the presence of mosaic alleles at Hm-2. Table 6.1 shows the data from 7 litters.

Mutations were seen that were present in:

1) only the Trophoblast;

2) only the Embryo proper;

3) common to both the Embryo and the Yolk sac but not present in the Trophoblast;

4) common to Trophoblast, Embryo and Yolk sac.

Examples are shown in Figure 6.3.

Mutations that were common to both the trophoblast and to the embryo proper must have arisen before 32 cell stage prior to the determination of those lineages (Figure 6.2). Such mutation events must have occurred prior to the fifth cell division. Since the total number of daughter cells which could

**Table 6.1** **Characteristics of embryonic and extra-embryonic *Hm-2* mosaics**

| No. | Tissues | Parental allele sizes | | Change in repeats | % change in repeats | | Dosage of mutant allele |
|---|---|---|---|---|---|---|---|
| 1.1 | Te- | 4225 | * | -1625 | -38.5 | T | 0.06 |
|  |  | 1325 |  | +1275 | +96.2 |  |  |
|  | tE- | 4225 |  | -1050 | -24.9 | E | 0.06 |
|  |  | 1325 |  | +1850 | +139.6 |  |  |
| 1.2 | Tes | 1250 |  | +900 | +72.0 | T | 0.11- |
|  |  | 1325 |  | +825 | +62.3 |  | 0.07[1] |
| 1.8 | t-S | 4225 |  | -2975 | -70.4 |  |  |
|  |  | 1325 | * | -75 | -5.7 | S | 0.20 |
| 2.3 | tE- | 3725 |  | -2625 | -70.5 |  |  |
|  |  | 1200 | * | -100 | -8.3 | E | 0.06 |
| 2.4 | Te- | 3725 |  | -975 | -26.2 | T | <0.05[1] |
|  |  | 2275 |  | +475 | +20.9 |  |  |
| 2.6 | tE- | 3725 |  | -575 | -15.4 | E | 0.04 |
|  |  | 2275 |  | +875 | +38.5 |  |  |
| 2.9 | Te- | 3725 |  | -1750 | -47.0 |  |  |
|  |  | 2275 | * | -300 | -13.2 | T | 0.06 |
| 4.1 | TE- | 1400 |  | -325 | -23.2 |  |  |
|  |  | 1025 | * | +50 | +4.9 | T | 0.66 |
|  |  |  |  |  |  | E | 0.57 |
| 4.3 | tE- | 3350 | * | +750 | +22.4 | E | 0.09 |
|  |  | 1075 |  | +3025 | +281.4 |  |  |
| 4.4 | Te- | 3350 | * | +1125 | +33.6 | T | 0.14 |
|  |  | 975 |  | +3500 | +359.0 |  |  |
| 4.5 | Te- | 3800 | * | -200 | -5.3 | T | 0.21 |
|  |  | 975 |  | +2625 | +269.2 |  |  |
|  | tE- | 3800 | * | +200 | +5.3 | E | 0.16 |
|  |  | 975 |  | +3025 | +310.3 |  |  |
|  | tE- | 3800 | * | -1000 | -26.3 | E | 0.02 |
|  |  | 975 |  | +1825 | +187.2 |  |  |
| 4.6 | Te- | 3350 | * | -350 | -10.4 | T | 0.07 |
|  |  | 975 |  | +2025 | +207.7 |  |  |
| 4.7 | tE- | 3350 | * | -500 | -14.9 | E | 0.19 |
|  |  | 975 |  | +1875 | +192.3 |  |  |
| 12.3 | Tes | 4800 | * | +325 | +6.8 | T | 0.27 |
|  |  | 925 |  | +4125 | +445.9 |  |  |

**Table 6.1    Characteristics of embryonic and extra-embryonic *Hm-2* mosaics**

Characteristics of all mosaics seen in seven litters of CBA x C57BL/6J mice.  Each different mutant band is listed separately, although those from the same 'embryo' are listed together.

*Tissues*

The tissues in which mutation band was present; T, trophoblast; E, embryo proper; S, yolk sac. Bold capitals; tissue was mosaic. normal type, tissue non-mosaic;  '-' Tissue not scored. eg 'Te-', Mutation in trophoblast which was not in the embryo proper; yolk sac not scored.

*Parental allele sizes*

Size of inherited parental alleles given as number of repeats.  First allele listed is from the CBA father, second from the C57BL/6 mother.

The allele which is believed to have mutated, based on on the smallest change in allele length, is marked with a ✳  Alleles which can be seen to have mutated by a loss in intensity (see table 5.2 legend) are marked with a ◑.
[1]exact dosage depends on which allele was progenitor

*Change in number of repeats*

Difference in size between the parental alleles and the mutant measured in numbers of repeats.  These differences are expressed as a percentage of the parental allele size in the following column.

*Mosaic Allele Dosage*

The proportion of cells containing the mutant allele.  Calculated as described in the legend to table 5.2

**Figure 6.3    Analysis of _Hm-2_ alleles in embryonic and extra-embryonic DNA**

DNA from trophoblast (T), embryo (E) and yolksac (S) was digested with _Hae_III, and Southern blot hybridized with $^{32}$P-labelled 9.2 at high stringency.

**A    Mutation common to T, E and S**

Mutant band ($\mu$) is present in all three tissues. All other bands are due to cross-hybridization of other loci and were present in either or both of the parents (not shown)

**B    Mutation in Trophoblast only**

**C    Mutation in embryo and yolk sac**

Mutant band ($\mu$) is present in E and S at different intensities, but is undetectable in Trophoblast DNA.

Figure 6.3    Analysis of *Hm-2* alleles in embryonic and extra-embryonic tissues
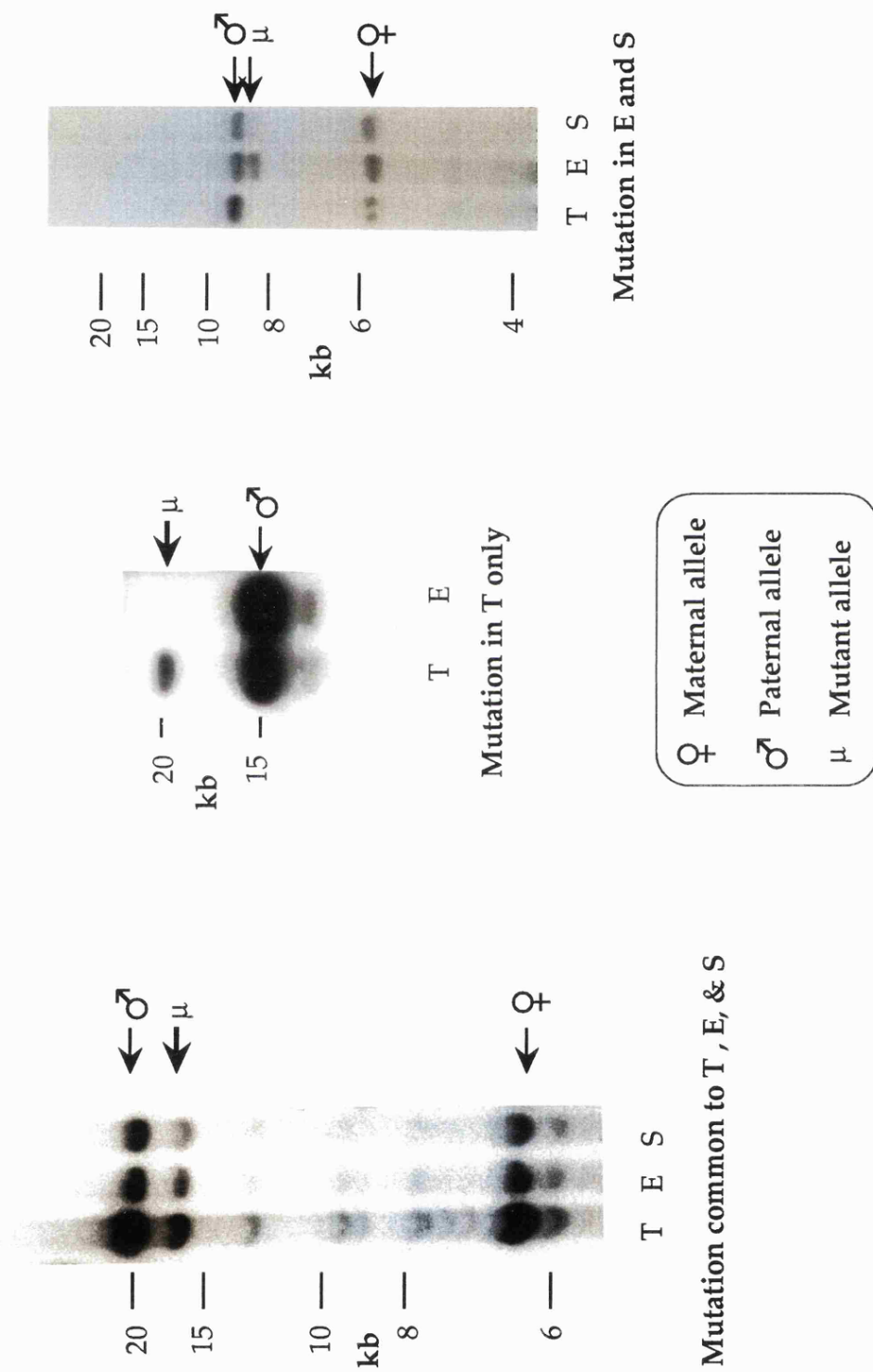
inherit a mutant allele and still pass daughter cells to both the trophoblast and the embryo is known, the number of shared mutations may be used to estimate the mutation rate during the first 4 cell divisions. This estimate will include neither those mutations which did occur before the fifth cell division, but the mutant cells then segregated into only the ICM or only the trophoblast; or those where the mutant cells in the trophoblast were not at the proliferating polar trophoblast and failed to contribute sufficient daughter cells to the day-10 trophoblast for a band to be seen.

Total number of trophoblast and embryonic pairs = 46

Number of pairs with mutant allele shared by trophoblast and embryo = 5

The proportion of all embryos in which the mutation occurred before the fifth cell division was 5/46 = 0.109 (95% limits 0.019-0.198)

The number of cells which could have mutated and still contributed cells to both the trophoblast and embryo is the number of cells up to and including the 16 cells produced by the 4th cell division. Mutations which occur at the 5th cell division will enter one of the 32 cells and be in either the trophoblast or ICM but not both. Thus this number is 2+4+8+16 = 30 cells.

If a total of 30 daughter cells could mutate and still contribute descendants to both the T and the E , the minimum estimate for the mutation rate per daughter cell per cell division is 0.109/30 = 0.004.

## 6.4 Distribution of mosaicism in trophoblast, embryo and yolk sac triplets.

To compare the rates of mosaicism in all seven categories (T, E, S, TE, ES, TS, and TES) the data was reduced to only those 24 triplets (table 6.2) for which all three tissues could be scored for the presence/absence of mutations.

**Table 6.2    Distribution of Hm-2 mosaicism in 24 TES triplets**

| Mutant Tissues | none | T | E | S | TE | ES | TS | TES |
|---|---|---|---|---|---|---|---|---|
| No of triplets (total 24) | 8 | 8 | 3 | 0 | 0 | 2 | 0 | 3 |
| Proportion of embryos | 0.33 | 0.33 | 0.13 | 0.00 | 0.00 | 0.08 | 0.00 | 0.13 |

The frequency of mosaicism in day 10 Embryos (E+ES+TES 8/24 = 33%, 95%limits 0.14-0.52) is directly comparable to the frequency seen in adult mice (27% seen in $F_1$ C57BL/6 x DBA mice (chapter 5b)). In contrast, the total rate of

mosaicism including mutations in the extra-embryonic tissues was extraordinarily high (16/24 or 67%, 95% limits 0.48-0.86).

There were no mutations shared by the Trophoblast and the Yolk sac that were not also in the Embryo proper, although there were mutations shared by the Embryo and Yolk sac (figure 6.4). This would be expected from the pattern of determination of the lineages. The ICM and Trophoblast split first, followed by the ICM splitting into the primitive ectoderm, which develops into the embryo proper, and the primitive endoderm, which develops into the yolk sac. A mutation in the trophoblast and yolk sac must have been present in the ICM and would also be in the embryo unless all mutant cells in the ICM segregated into the yolk sac lineage.



Figure 6.4    Distribution of mutations in the trophoblast, embryo and yolk sac.
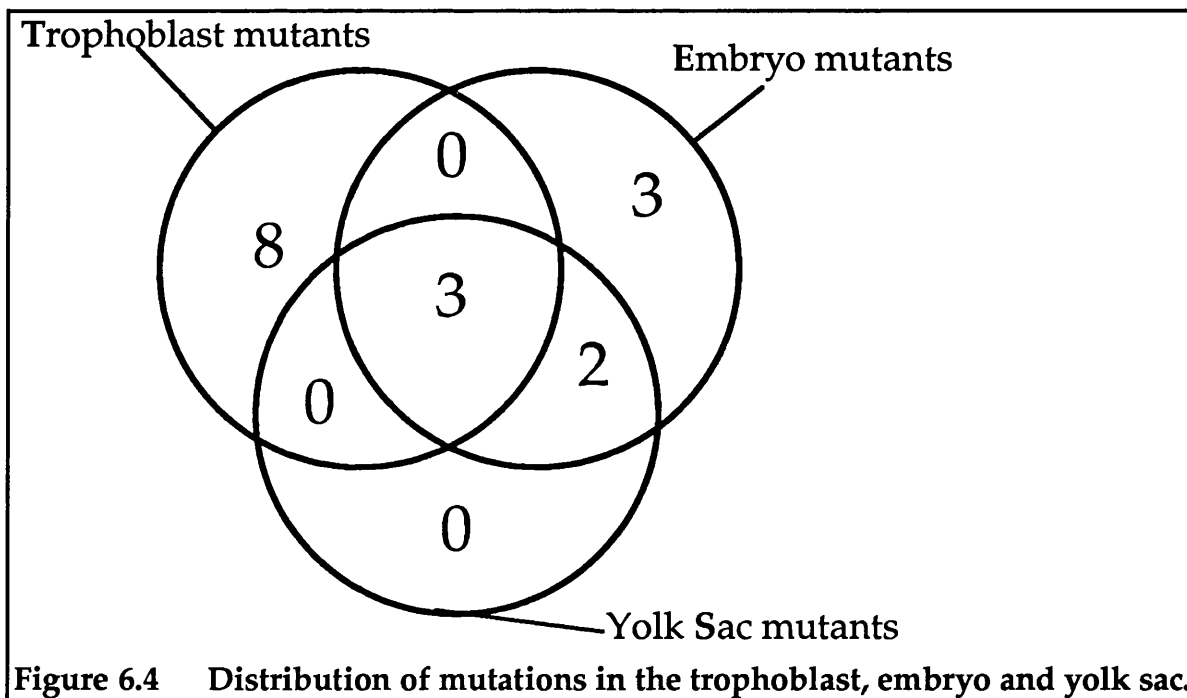
## 6.5    Summary

Analysis of Hm-2 in embryonic and extra-embryonic tissues has shown that many mutation events occur early in development, to produce in some cases divergence in Hm-2 genotype between the embryonic and extra-embryonic tissues, and in others, mosaicism common to the embryo and trophoblast, suggesting that mutations can occur at the earliest stages of development, prior to the fifth cell division following fertilization. A minimal estimate for the proportion of $F_1$ mice with such mutations was 11%.

This figure was used to find a minimal estimate for the average somatic mutation rate over the first 4 divisions of 0.004 per cell per division.

The proportion of embryonic mosaics is similar to the proportion of adult mosaics. If mosaicism in the extra-embryonic tissues is included, the total proportion of mice that are mosaic at *Hm-2* increases to 67%. The absence of mutations shared by the Trophoblast and yolk sac but not the embryo fits the predictions of early developmental pathways.

# 7

## Evidence for an early developmental 'window' of somatic mutation at *Hm-2*

## 7.1 Introduction

It has been shown that somatic mutations occur at a high frequency in the first few cell divisions following fertilization (Chapter 6). A question remaining is whether this process continues throughout life at the same rate, or whether high levels of mutation are preferentially confined to a "window" of instability in early development.

If there is a constant mutation rate per cell per cell division, then as development proceeds, the frequency of mutant cells will increase with each successive cell generation, although mutations which occur later in development will be inherited by fewer daughter cells than those which occur very early in development and will therefore make up a lower proportion of the total number of cells. After a certain time, therefore, mutant alleles will fail to contribute sufficiently to a DNA sample to be detected by Southern blot hybridization. It is possible that the mosaicism described thus far (chapter 5 and 6) may be just 'the tip of the iceberg', and in reality, mice are mosaic for numerous other low dosage mutant alleles. Thus what has been shown in chapter 6 is not that all somatic mutations occur in early development, but that those somatic mutations that could be detected by Southern blot analysis of genomic DNA occurred early in development. This chapter will attempt to show that in fact the somatic mutation rate does decrease with time.

Two observations already point to a non-constant mutation rate:

### a)   No tissue specific mutations have been seen

If mutations happen later in development, some might occur early in a developmentally separate lineage, to give tissue specific mosaicism. No cases of tissue specific mosaicism have been observed in a total of 6 mosaic animals that have been dissected (see section 6.2). However, this assumes that the number of cells initially allocated to each lineage is small enough for descendants of a new mutant to represent a detectable proportion of the total, or, if cells for a lineage are allocated from a large pool of cells, that there is the

possibility of chance sampling events to take mutant cells which are a small proportion of the pool, but a large proportion of the sample.

### b) There are no mutations seen only in the yolk sac

The only mutations seen in the yolk sac are those that are shared by the embryo, or by both the embryo and the trophoblast (see chapter 6, figure 6.4). This suggests that by the time cells are allocated to the extra-embryonic ectoderm mutations are rare; the yolk sac may or may not share in the mosaicism of the ICM, but yolk sac cells rarely, if ever, undergo new mutations which would then be specific to the yolk sac.

## 7.2 Relationship between time at which somatic mutation arises, and dosage of the mosaic allele

A normal, heterozygous individual has two alleles, each with a dosage of 1. A mosaic individual with one mutant and two parental alleles will have one parental allele (the non-progenitor allele) with a dosage of 1, whereas the other parental allele (the progenitor allele) and the mutant allele will have a combined dosage of 1. A mutation that occurs at the first cell division producing one mutant and one normal daughter cell will result in an embryo in which half the cells will have the mutant allele, and the dosage of the mutant allele will be 0.5. Mutations that occur later in development will have smaller dosages as the proportion of mutant cells decreases.

The intensities of alleles seen on an autoradiograph reflect the dosage of the mutant and parental alleles. The dosage (m, the proportion of affected cells) can be calculated by comparison of the intensity of the mutant allele relative to the parental alleles (see table 5.2, somatic mutations, and table 6.1 analysis of embryonic mutants). These dosages can then be used to gain information about the cell division at which the mutation occurred. Thus it should be possible to estimate the numbers of mutation events expected at each cell division, and compare this value with the observed frequencies of mutations of particular dosages.

If the probability of mutation is equal per cell per cell division, it would be expected that a higher incidence of low dosage mutants would be observed. This is because the number of cells which have a chance of mutating increases with each successive division, but each mutation affects a decreasing proportion of the total cell number.

To use the dosage of a mutant allele to estimate the time at which a mutation arose, it is necessary to calculate the proportion of mutant cells in the

entire embryo, including extra-embryonic tissues, and to assume that the proportion of mutant and non mutant cells has remained constant throughout development.

There are problems, however, in using the dosage in the entire embryo as an estimate of the cell division at which the mutation occurred. Dosage in the trophoblast is affected by the stochastic proliferation of the trophoblast, which will randomly increase or decrease the proportion of mutant cells depending on how many of the mutant cells form part of the polar trophectoderm (see section 6.1.2). Thus the dosage of a mutant allele in a day 10 trophoblast does not reflect accurately the proportion of mutant cells in the starting population of trophoblast cells, and moreover, the dosage of the entire embryo cannot be used to estimate time of mutation, since a major component will be the extra-embryonic tissues. Therefore, only the sections of the embryo that have relatively equal rates of cell division, i.e. the embryo proper and the yolk sac, can be used to extrapolate back from the dosage in a day 10 embryo to the dosage at mutation, which reflects the number of cells in the organism when the mutation occurs. Assuming that the proportion of mutant cells in the embryo proper remains constant, the dosage seen in an adult mouse can be considered to be equivalent to a day 10 embryo. (It has been shown (see section 6.1.1a) that not only are mutant alleles present in all adult tissues, but moreover, that the dosage is approximately equal in all tissues. This result suggests that the mutant cells are divided equally between the different tissue lineages and contribute equally to adult tissues.)

## 7.3    Classification of observed mutant alleles by dosage

The dosages of mosaic alleles in 60 mosaic embryos and adult mice were calculated by densitometric scanning (see table 5.2). The proportion of mutant cells ranged from 2% to 100%.

To relate the observed dosages to the developmental time at which they were generated it was necessary to calculate the range of possible dosages in the ICM from mutations at each cell division. The number of cells that enter the ICM from mutations which occur in the first five cell divisions is variable. For example, if one cell is mutant at the 4-cell stage, 8 cells will be mutant at the 32-cell stage. Of these 8, from 0 to 6 will have been allocated to the ICM. Therefore, although the dosage in total is 8/32 = 25%, the dosage seen in the ICM may be between 0 and 60%. The range of possible dosages in the ICM for mutations at each cell division were calculated by determining the minimum and maximum number of mutant cells that could be allocated to a mutant

ICM (which contains between 10-16 cells at the 32 cell stage (Fleming, 1987)), as shown in table 7.1.1.

**Table 7.1.1 Calculation of possible dosages of mutant cells in ICM at 32 (64) cell stage from mutations at each cell division**

| | Cell division at which mutation occurs | | | | | |
|---|---|---|---|---|---|---|
| | 1st | 2nd | 3rd | 4th | 5th | 6th |
| Number of mutant cells in whole embryo | 16 | 8 | 4 | 2 | 1 | (1) |
| min and max no. of mutant cells in ICM | 1-12 | 1-6 | 1-4 | 1-2 | 1 | (1) |
| max and min no. of cells in ICM | 10-16 | 10-16 | 10-16 | 10-16 | 10-16 | (20-32) |
| dosage range | 0.06 - 1.00 | 0.06 - 0.60 | 0.06 - 0.40 | 0.06 - 0.20 | 0.06 - 0.10 | 0.03 - 0.05 |

Since the range of dosages for each division overlap, it is not possible to say that a mutation with a dosage of 'X%' occurred at the 'Yth' cell division. Neither was it possible to prove that all mutant cells had entered the ICM by screening trophoblasts to see if mutant cells were present, because mutant cells may have entered the trophoblast but failed to contribute enough cells to the day 10 trophoblast to be detected by Southern hybridization. The categories used to bin the data, (table 7.1.2) are not, therefore, precisely related to any single cell division; the dosage levels for each bin were within the possible range of dosages for mutations at each cell division, but would not contain all mutant alleles generated at any single cell division.

**Table 7.1.2 Dosage range in each bin**

| Bin | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| range of dosages in each bin | >0.60 | 0.40 - 0.60 | 0.20 - 0.40 | 0.10 - 0.20 | 0.05 - 0.10 | 0.03 - 0.05 |

The dosages of observed mutant alleles were binned into the categories above and the numbers in each bin plotted. The histogram (figure 7.1) showed

a peak in bin 3, containing mutant alleles with dosages of between 0.2 and 0.4. Fewer mutant alleles were scored with lower dosages; there was a steady decline in the numbers of mutations in bins 4 to 6. This reduction was not due to poor scoring of faint mosaics, since the decline began in bins 4 and 5, which contained mutants with dosages of 5 to 20% - easily detectable on Southern blots, although dosages of less than 5% might have been missed.

## 7.4    Comparison of the frequency distribution of the observed dosages to a frequency distribution of dosages created by a computer model using a constant mutation rate per cell per cell division.

To compare the frequency distribution of the observed mutant allele dosages (figure 7.1) with the distribution expected from a constant mutation rate, a computer program was written to simulate somatic mutation over the first 6 cell divisions following fertilization. Mutations were created at random in the dividing array of cells (although each cell represented a single allele, rather than a real diploid cell). At division 4 and 5 cells were allocated to the ICM, as described in section 6.1.2. The model takes into account anticlinal and periclinal divisions of polar cells (figure 6.1), and assumes that descendants of ICM cells remain in the ICM. The model generates embryos with the ratios of ICM to trophoblast cells observed by Fleming (1987), with the probabilities of different ratios taken from Fleming, 1987, figure 2 and 6. At the 64 cell stage the dosage of the mutant allele in the ICM was calculated and binned as previously described (table 7.1.2). The program was written in BBC BASIC on a Apple Macintosh, and is given in Appendix A.

Ten thousand alleles were simulated with a constant mutation rate per cell per cell division of 0.004, the minimal estimate of mutation rate calculated from the numbers of shared TE mutations (chapter 6). The histogram (figure 7.1) showed that the numbers of mutants increased in successive bins. This did not correspond with the observed data in which frequencies declined after bin 3. Comparison of the observed and expected data (table 7.2) using the $\chi^2$ goodness-of-fit test showed significant deviation ($\chi^2$ with 5df = 103.7, p<0.0001), and thus the hypothesis of constant mutation rate was rejected.

**Figure 7.1 Histograms of observed and simulated dosage data**

A  Observed data.  Error bars show 95% confidence limits

B  Simulation of 10000 alleles with constant mutation rate per cell of 0.004
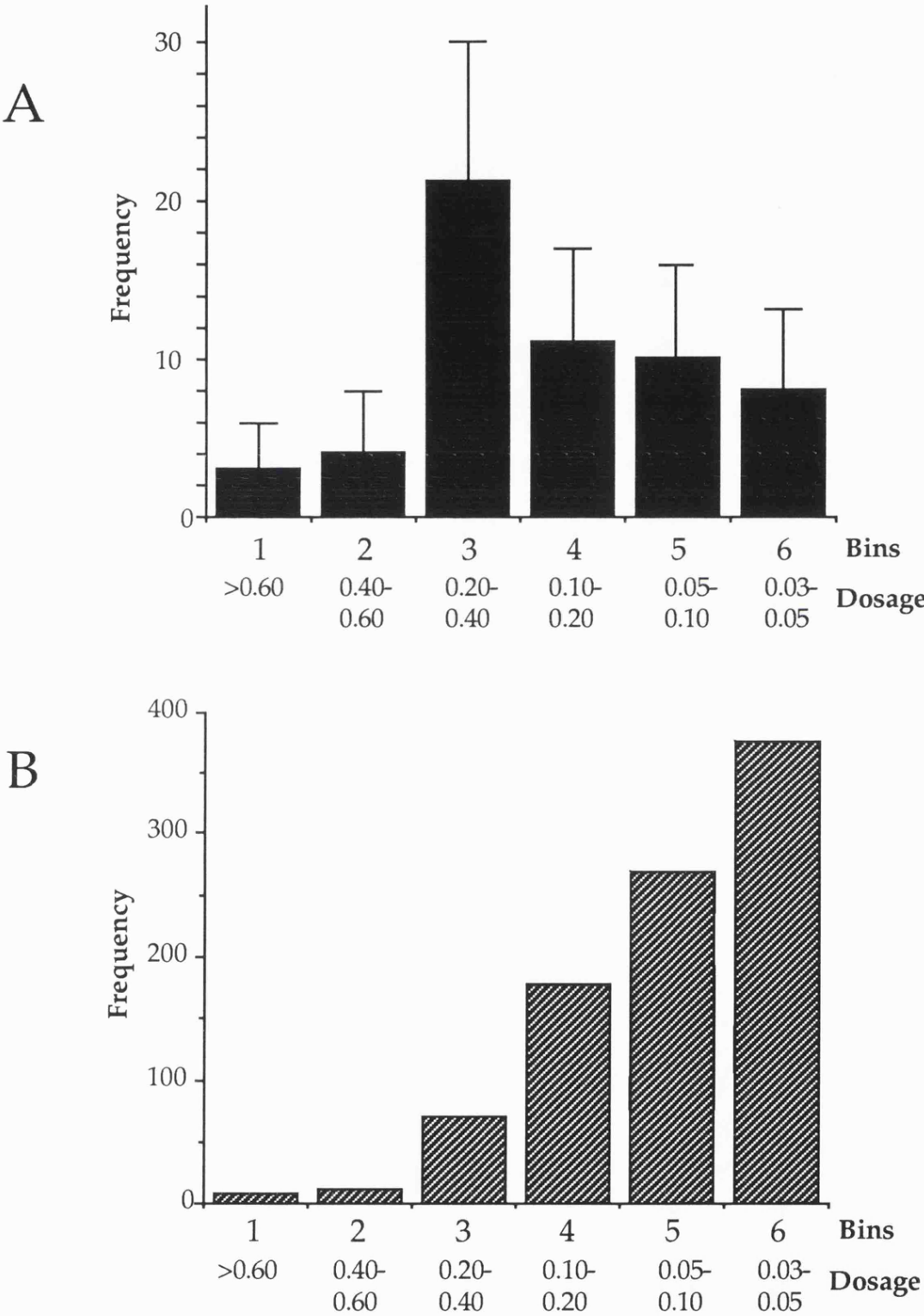
**Table 7.2 Analysis of observed and simulated dosages**

| Dosage | Observed frequency | Simulated (Expected) frequency |
|---|---|---|
| >0.60 | 3 | 0.5 |
| 0.40 -0.60 | 4 | 0.7 |
| 0.20 -0.40 | 21 | 4.4 |
| 0.10 -0.20 | 11 | 11.1 |
| 0.05 -0.10 | 10 | 16.8 |
| 0.03 -0.05 | 8 | 23.5 |
| total | 57 | 57.0 |

## 7.5 Relationship between cell division in which mutations arose and their final allocation to bins

The relatively low frequency of low dosage mutants suggests a decline in the mutation rate at later cell divisions. To quantify this, the distribution of mutations from each cell division into each bin was calculated by analysing the final distribution into bins of over 1000 simulated mutations generated at each of the first 5 cell divisions. Table 7.3 shows the final destination by proportion into each bin of mutations from each cell division.

**Table 7.3     Bin distribution of mutations from each cell division**

| Mutations generated at division; | 1 | 2 | Bin 3 | 4 | 5 | 6 | Not in ICM |
|---|---|---|---|---|---|---|---|
| 1st | 0.33 | 0.30 | 0.30 | 0.05 | 0.01 | 0.00 | 0.00 |
| 2nd | | 0.07 | 0.60 | 0.17 | 0.08 | 0.00 | 0.07 |
| 3rd | | | 0.10 | 0.59 | 0.13 | 0.00 | 0.19 |
| 4th | | | | 0.35 | 0.14 | 0.00 | 0.51 |
| 5th | | | | | 0.42 | 0.00 | 0.58 |
| 6th | | | | | | 0.42 | 0.58 |

These calculations showed that the observed peak in bin 3, would actually contain very few mutations that arose at the third cell division, but

rather, that the major portion of this peak would be from mutations arising at
the second and the first cell divisions.

## 7.6    Estimation of mutation rates over the first 6 cell divisions

The mosaicism distributions in table 7.3 suggests an approach to
determine the mutation rates at each cell division using the observed numbers
of mutations at each dosage level.  For example, the distribution of mosaics
arising at division 1 suggests that only 1/3 of 1st division mutants will appear
in bin 1.  Since three mutants were seen in bin 1 it can be surmised that 9
mutants arose at division 1.  More generally, the relationship between p1 to p6
and b1 to b6, where p1 to p6 are the true numbers of mutations that occurred at
each division, and b1 to b6 are the numbers of mutants per bin, gives the
simultaneous equations:

$$0.33p1 = b1 \ (3)$$

$$0.30p1 + 0.07p2 = b2 \ (4)$$

$$0.30p1 + 0.60p2 + 0.10p3 = b3 \ (21)$$

$$0.05p1 + 0.17p2 + 0.59p3 + 0.35p4 = b4 \ (11)$$

$$0.01p1 + 0.08p2 + 0.13p3 + 0.14p4 + 0.42p5 = b5 \ (10)$$

$$0.42p6 = b6 \ (8)$$

(the numbers in brackets are the observed values of  b1 to b5)

A second program was written (Appendix B), which used all possible
values of p1 to p5 to calculate the expected number of mutants in each bin,
using these equations.  For each combination of p1 to p5, the ($\chi^2$) deviation of
the expected bin totals from the observed bin totals was calculated, and the
values giving the lowest deviation identified.  The values of p1 to p5 which
best fit the observed totals are given in table 7.4.  The figures generated are an
estimation of the number of mutations at each cell division in the sample
investigated, which was a total of 279 mice.  Of these, therefore, 8.4 mutated at
the first cell division, which gave a frequency of mutant alleles at division 1 of
8.4/279 = 0.03.  There are a total of two cells produced at this division, therefore
the probability of mutation per daughter cell is 0.03/2 = 0.015.  The figures
produced for probability of mutation at each cell division are given in table 7.4,
and shown in figure 7.2.

**Table 7.4     Estimates of mutation rate per daughter cell**

| Division | Number of mutations occurring at cell division | Frequency of mutation per division (n/279) | Number of cells | Probability of mutation per daughter cell |
|---|---|---|---|---|
| 1 | 8.4 | 0.0301 | 2 | $1.5 \times 10^{-2}$ |
| 2 | 28.8 | 0.1032 | 4 | $2.5 \times 10^{-2}$ |
| 3 | 9.5 | 0.0341 | 8 | $4.3 \times 10^{-3}$ |
| 4 | 0.2 | 0.0007 | 16 | $4.5 \times 10^{-5}$ |
| 5 | 15.1 | 0.0541 | 32 | $1.7 \times 10^{-3}$ |
| 6 | 19.0 | 0.0681 | 64 | $1.1 \times 10^{-3}$ |

**Figure 7.2     Mutation rate per cell division**
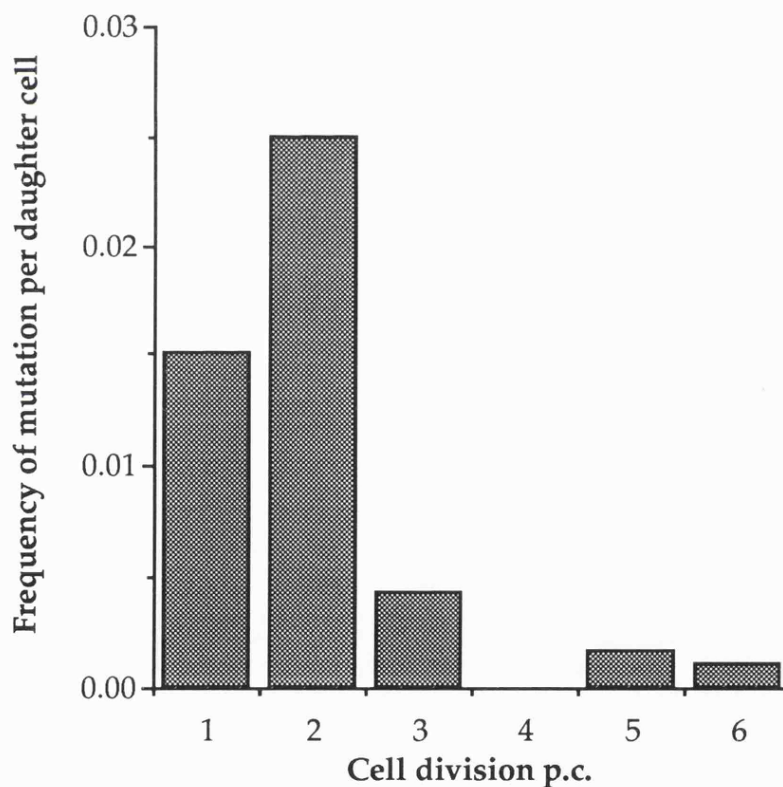


These results strongly suggest that the mutation rate reaches a peak at division 2, then declines rapidly to much lower levels (see figure 7.2).

When the mutation rates shown in table 7.4 were used in the computer model, and 25000 alleles simulated as previous, the histogram produced closely resembled the histogram produced from the observed data (Figure 7.3).

Comparison of the observed and simulated data using the $\chi^2$ test showed that the observed sample did not differ significantly from the simulated distribution ($\chi^2$ with 5df = 3.609, p>0.50).

**Figure 7.3   Comparison of actual data, with data simulated using mutation rates in table 7.5**



Observed data from 279 embryos     Simulation of 12500 embryos
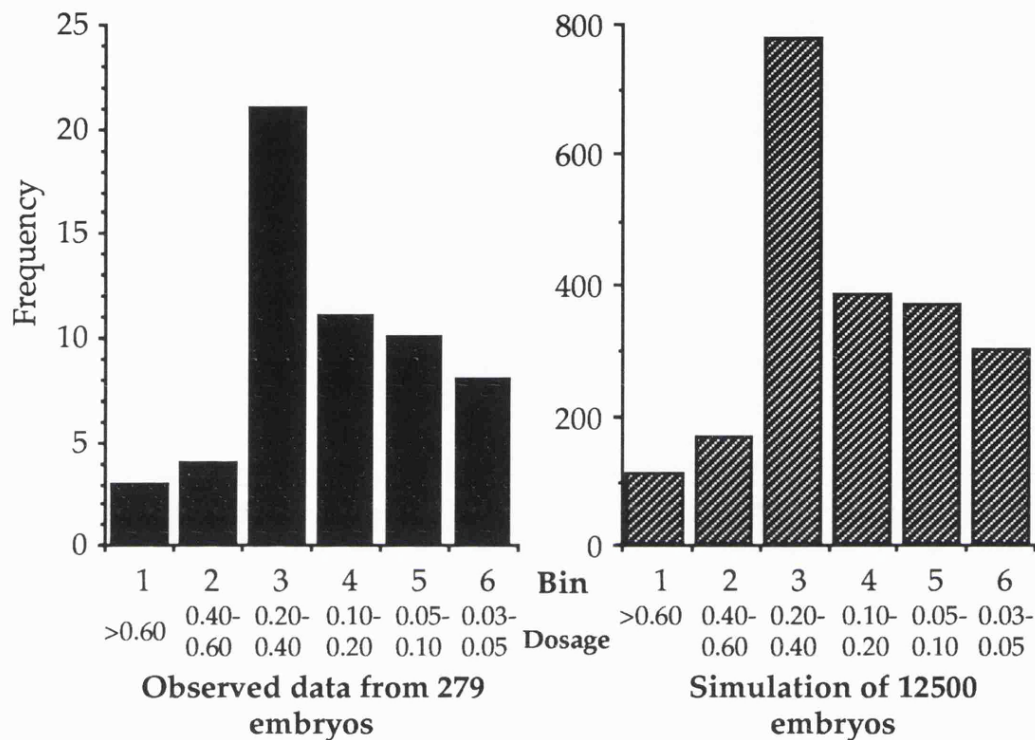
## 7.7   Summary

The observed dosages of mutant alleles did not fit a model in which the mutation rate is constant; there was a decrease in the incidence of low dosage mosaics from the expected frequency under a constant mutation rate. This decrease could not be attributed to non-detection of low level mosaics, since the decrease in frequency began in a range of dosages that could be readily scored.

The results suggest that there is a period of high frequency mutation ($\sim 10^{-2}$ per cell) which extends through the 1st and 2nd cell division, the rate thereafter rapidly declines, to $\sim 10^{-3}$ per cell at divisions 4-6. It can be assumed that this rate either remains the same or decreases further because other evidence suggests that somatic mutations in later development occur very infrequently.

# 8

## Mutation processes at *Hm-2*

## 8.1 Introduction

Several models have been proposed for the primary mechanism by which changes in repeat copy number in minisatellite alleles are generated. These are inter-allelic unequal recombination, i.e. recombination between homologous chromosomes, and intra-allelic processes, either unequal recombination between two sister chromatids (unequal sister chromatid exchange, hereafter SCE), or an internal DNA slippage process during replication (Jeffreys *et al.*, 1985b).

Evidence has accumulated to suggest that inter-allelic recombination with exchange of flanking markers is not a major mechanism for the generation of mutations (Wolff *et al.*, 1988; Kelly *et al.*, 1991), although recently analysis of the internal structure of minisatellite alleles has shown that some inter-allelic processes are occurring (Jeffreys *et al.*, 1991). It has also been suggested that gene conversion following an inter-allelic cross-overs could create recombinant alleles without exchange of flanking markers (Wolff *et al.*, 1990, 1991)

## 8.2 Distinguishing the products of unequal exchange and slippage

A major obstacle to investigating how a particular mutant allele arose is that processes that affect only a single allele (SCE and slippage) cannot be distinguished by their effect on the flanking DNA. The two mechanisms do differ in that a slippage mechanism affects only one molecule and only a single new length allele is created, whereas a mutation event that involves unequal exchange between two DNA molecules, whether between sister chromatids or homologous chromosomes, will produce two mutant alleles, which will differ in size reciprocally from the parental alleles, the gain in repeat units from one allele will be matched by the loss of repeat units in the other (figure 8.1). Germline mutations generated by these processes will produce a pair of reciprocal gametes in sperm, but the chances of both gametes successfully

**Figure 8.1** **Segregation of reciprocal products into daughter cells following recombination and unequal exchange.**
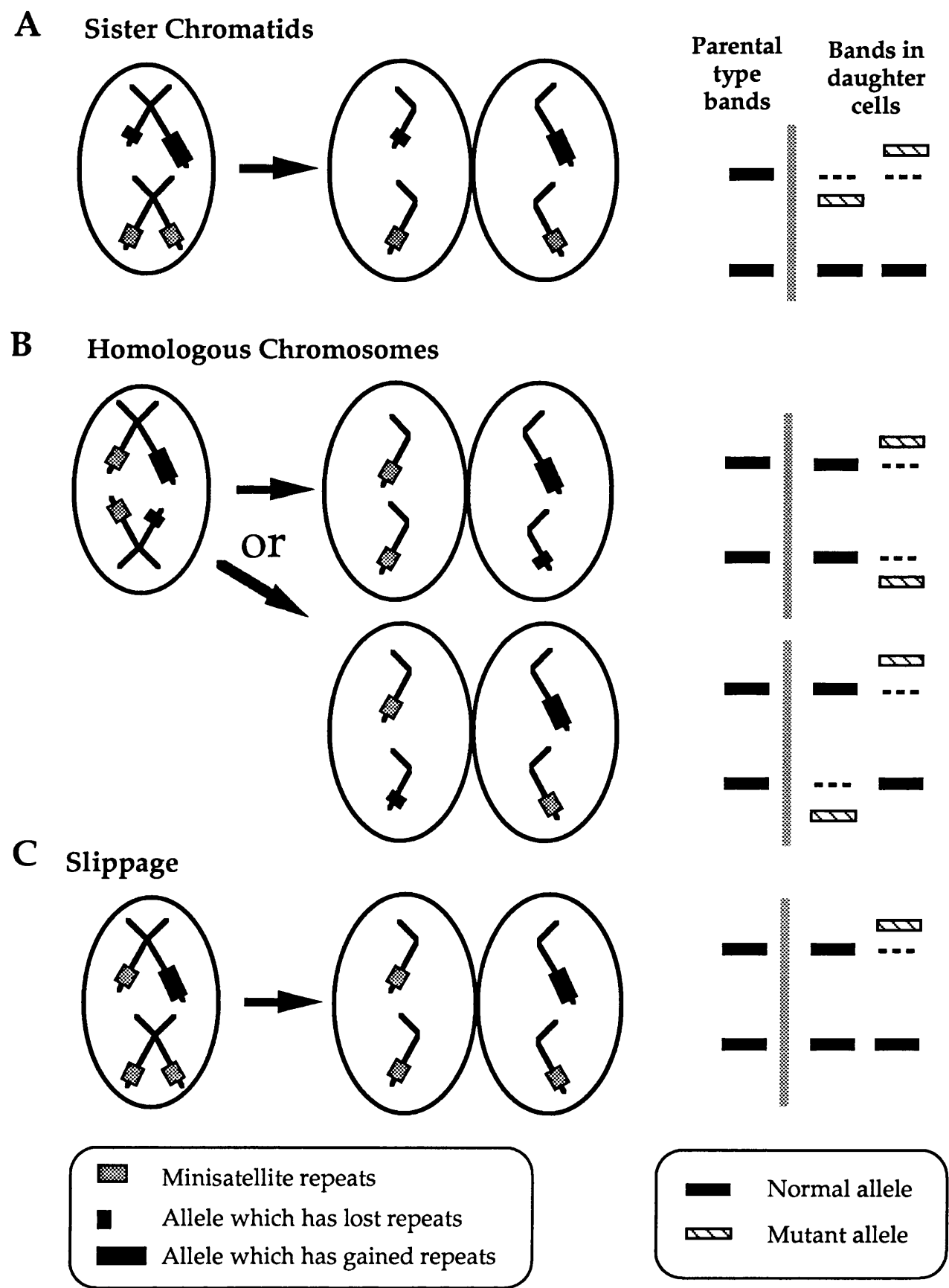
Schematic diagrams of segregation of minisatellite alleles during mitosis. One pair of homologous chromosomes containing a minisatellite locus are shown. Parental alleles are shown by shaded boxes, mutant length minisatellite alleles are shown by filled boxes.

A     The result of sister chromatid exchange. At anaphase the two new length alleles (shown as open bars) are partitioned into different daughter cells.

B     The result of an inter-allelic recombination. The pair of reciprocal products can segregate into either the same daughter cell or different daughter cells.

C     The single mutant allele generated by a slippage event may segregate into either daughter cell.

# Figure 8.1 Segregation of reciprocal products into daughter cells following recombination and unequal exchange or slippage

## A Sister Chromatids



## B Homologous Chromosomes



## C Slippage



Minisatellite repeats

Allele which has lost repeats

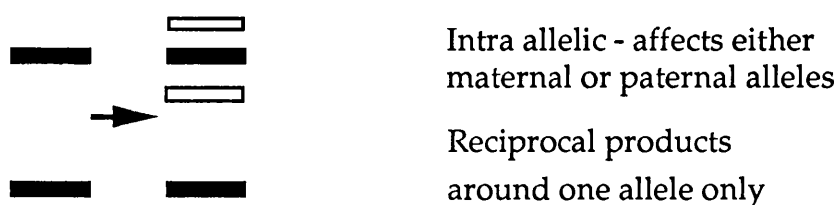Allele which has gained repeats

Normal allele

Mutant allele

**Figure 8.2 The expected patterns of reciprocal products on Southern blots from different kinds of mutation events.**
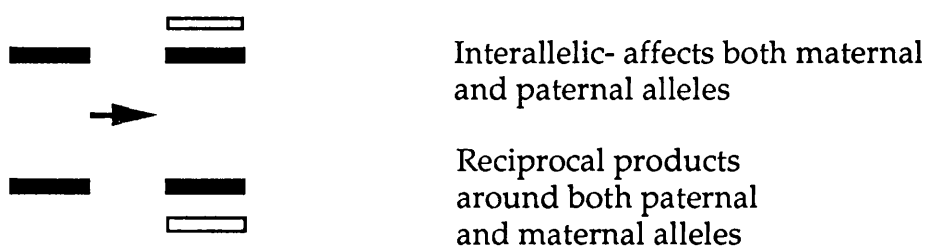
Assuming that a tissue is mosaic for both reciprocal products, the patterns below would be seen
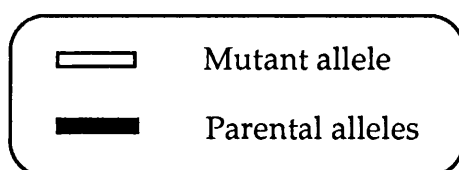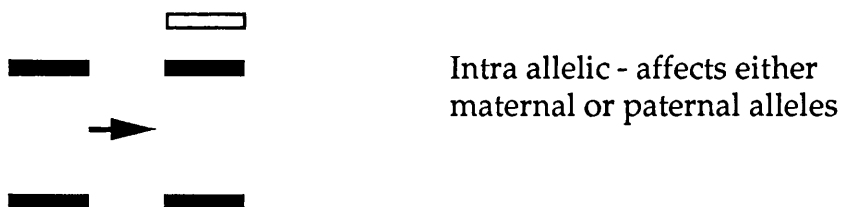
# Recombination mechanisms

## Sister chromatid exchange

Intra allelic - affects either maternal or paternal alleles

Reciprocal products around one allele only

## Unequal exchange between homologous chromosomes

Interallelic- affects both maternal and paternal alleles

Reciprocal products around both paternal and maternal alleles

# Non-recombination mechanisms (Slippage)

Intra allelic - affects either maternal or paternal alleles

Mutant allele

Parental alleles

Figure 8.3 Cell divisions up to the blastocyst stage showing one possible result of a mutation created by an unequal sister chromatid exchange at the 2nd cell division

A

+ x rpts     - x rpts

Cells forming the
Trophoblast lineage

Cells forming the
ICM lineage

12/20 Normal cells
6/20 Mutant (gained repeats)
2/20 Mutant (lost repeats)

4/12 Normal
2/12 Mutant (gained repeats)
6/12 Mutant (lost repeats)
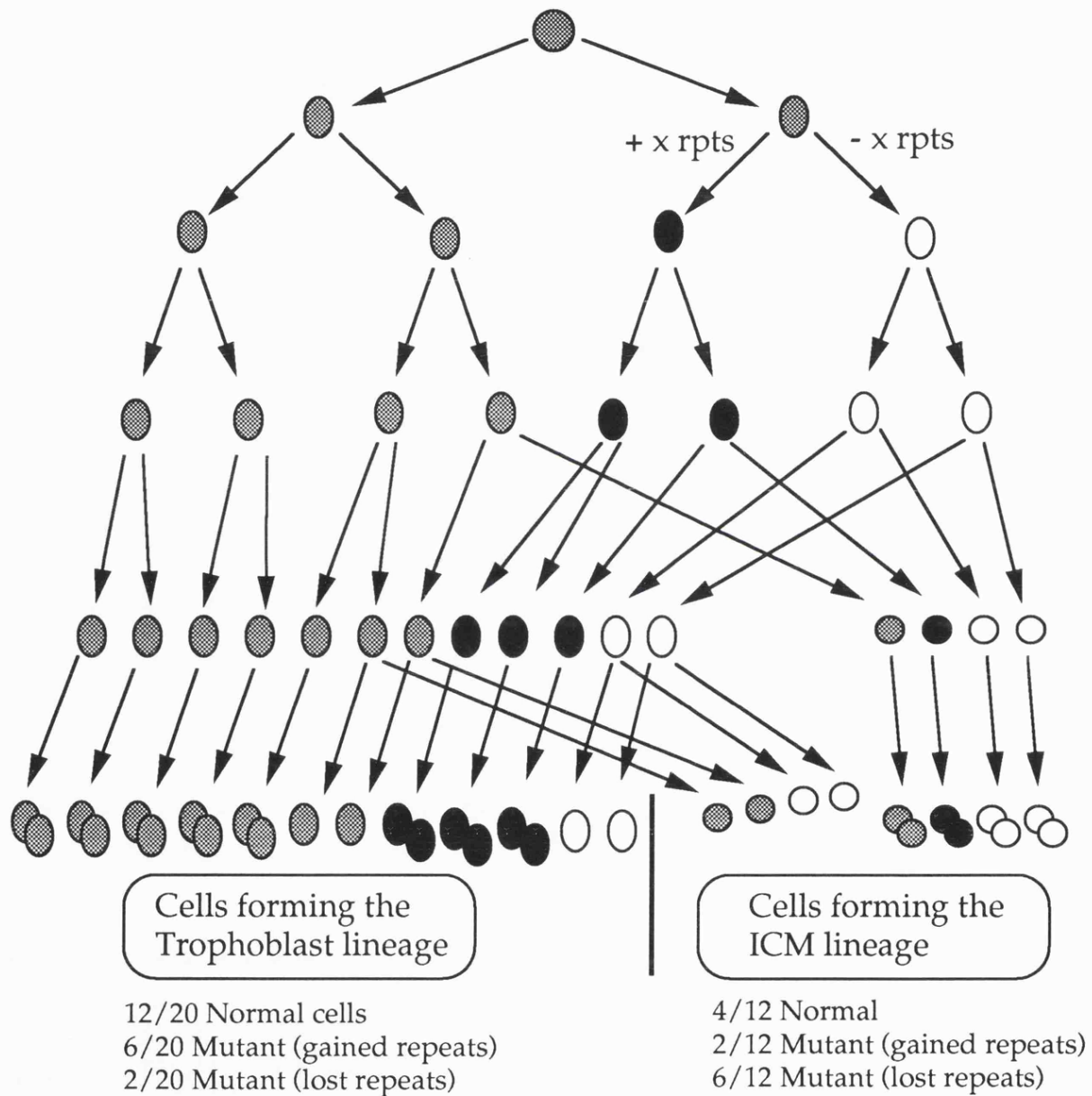
Normal cell          (two parental type alleles)

Mutant cell          (one parental, one mutant allele)

Mutant cell          (one parental, one mutant allele)

**Figure 8.3**

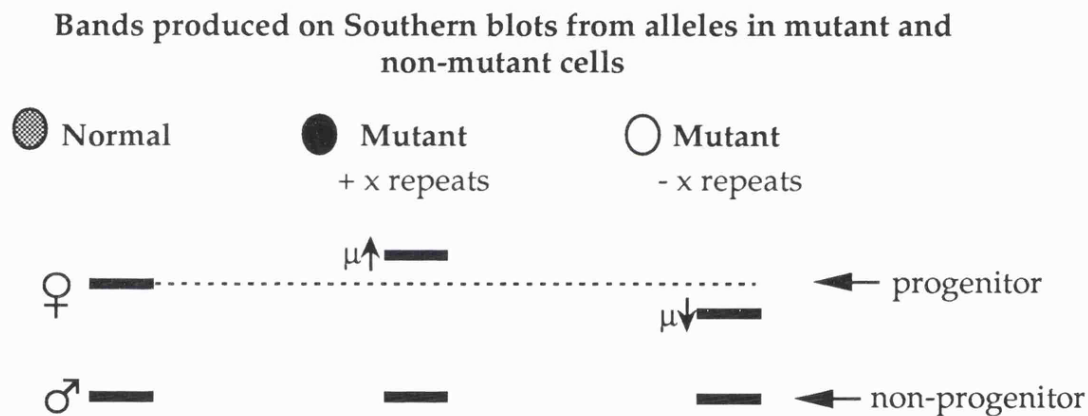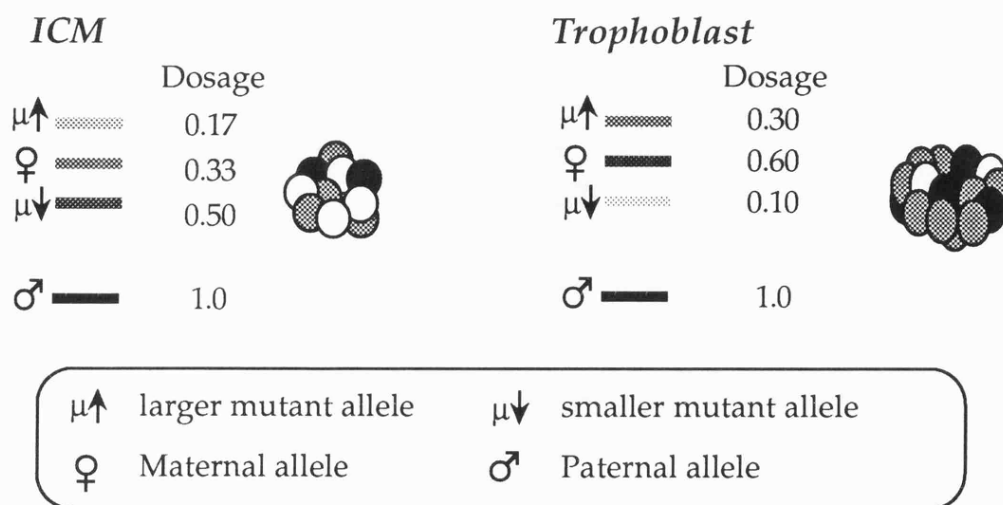**A** Schematic of one hypothetical distribution of mutant cells following a mutation at the 2nd cell division which creates two reciprocal products.

The daughter cells can be distributed disproportionately between the ICM and trophoblast lineages, producing unequal dosages of the mutant alleles in adult or embryonic mice. Mosaic tissues have all alleles in varying dosages, producing differing intensities on an autoradiograph. Shown below (**B**) are the dosages produced in the embryo for the example given. Note that although the progenitor parental allele will have a reduced dosage compared to a non-mosaic tissue, the non-mutant parental allele will still have a dosage of 1 in all tissues.

**Bands produced on Southern blots from alleles in mutant and non-mutant cells**

⊚ Normal    ● Mutant    ○ Mutant
                + x repeats      - x repeats

μ↑ ▬

♀ ▬ - - - - - - - - - - - - - - - - - - - - - - - - - - ◄— progenitor
                                    μ↓ ▬

♂ ▬              ▬              ▬ ◄— non-progenitor

**B  Band pattern seen with single locus probe**

*ICM*

|  |  | Dosage |
|---|---|---|
| μ↑ | ▬ | 0.17 |
| ♀ | ▬ | 0.33 |
| μ↓ | ▬ | 0.50 |
| ♂ | ▬ | 1.0 |

*Trophoblast*

|  |  | Dosage |
|---|---|---|
| μ↑ | ▬ | 0.30 |
| ♀ | ▬ | 0.60 |
| μ↓ | ▬ | 0.10 |
| ♂ | ▬ | 1.0 |

μ↑  larger mutant allele     μ↓  smaller mutant allele

♀  Maternal allele           ♂  Paternal allele

fertilizing eggs is negligible and, therefore, only one of the two reciprocal products will be detected. When the reciprocal mutation happens in a somatic cell, however, the mutant alleles will segregate into the two daughter cells, and both of these cells can continue to replicate and therefore both products of a somatic mutation may be seen in mosaic tissue. Thus if somatic mutations are created by recombination, it might prove possible to identify them by the presence of two mutant alleles, reciprocally positioned around either or both of the parental alleles. (figure 8.2).

Some assumptions, however, must be made:

1) Every cell must continue to divide at a similar rate in the adult or embryonic tissues so that descendants of both daughter cells are represented, not necessarily at equal dosage, but sufficiently for both alleles to be seen by Southern blot hybridization. It is likely that the reciprocal products would not be at equal dosages: it has already been shown that mutations happen predominantly during the first few cell divisions (chapters 6 & 7), and it would be unlikely for the descendants of both mutant daughter cells to segregate equally into the embryo and trophoblast (figure 8.3).

2) The unequal recombination must be the only event. Any alteration to the reciprocal sizes, such as by replication slippage, will mask the evidence of recombination.

## 8.3    Analysis of adult, embryonic, and extra-embryonic double mosaics

The C57BL/6 x DBA and C57BL/6 x CBA families were analysed as described in chapters 5 and 6. Eight mice or embryos were found to be double mosaics, and one embryo/extra-embryonic tissue pair was a triple mosaic (table 8.1). In four cases the lengths of a pair of mutant alleles, within sizing error, were reciprocal with respect to one of the parental alleles consistent with unequal sister chromatid exchange. Figure 8.4 shows examples of double mutants. $F_2 17$, and 18.5 have two mosaic bands which are not reciprocal and are presumed to be from independent mutations. 18.5 and $F_2 86$ have a pair of mutant alleles which might have arisen by unequal sister-chromatid exchange.

## Figure 8.4    *Hm-2* double mosaics

Southern blot hybridization of genomic DNA cut with *Hae*III to $^{32}$P-labelled 9.2. Parental alleles are labelled ● , mutant alleles are labelled o.

T = trophoblast,    E = Embryo,   S = yolk sac

**A**    $F_2$86 has inherited a very small (625 repeat) paternal allele (not shown), and a large (4625 repeat) allele from the mother.  In addition, two other bands (labelled μ)  are seen which are reciprocally located around the maternal allele. Note the father, $F_1$33 is also mosaic with a faint band of approximately the same size as the smaller mutant allele, but this has not been inherited.

**B**    18.4 has two non-parental alleles flanking the paternal allele. The sizes of the two mutant alleles are reciprocal. The relative intensities of the paternal and smaller mutant allele suggest that the paternal allele is the progenitor allele. The dosage of the smaller mutant allele decreases in the embryo and yolk sac lanes and is matched by a corresponding increase in dosage of the paternal allele.  The larger mutant allele is presumed to have arisen from the paternal allele which is closest in size.

**C, D**        Both mice have two mutant alleles (labelled μ), but these are not reciprocally located around the parental alleles.

# Figure 8.4 *Hm-2* double mosaics
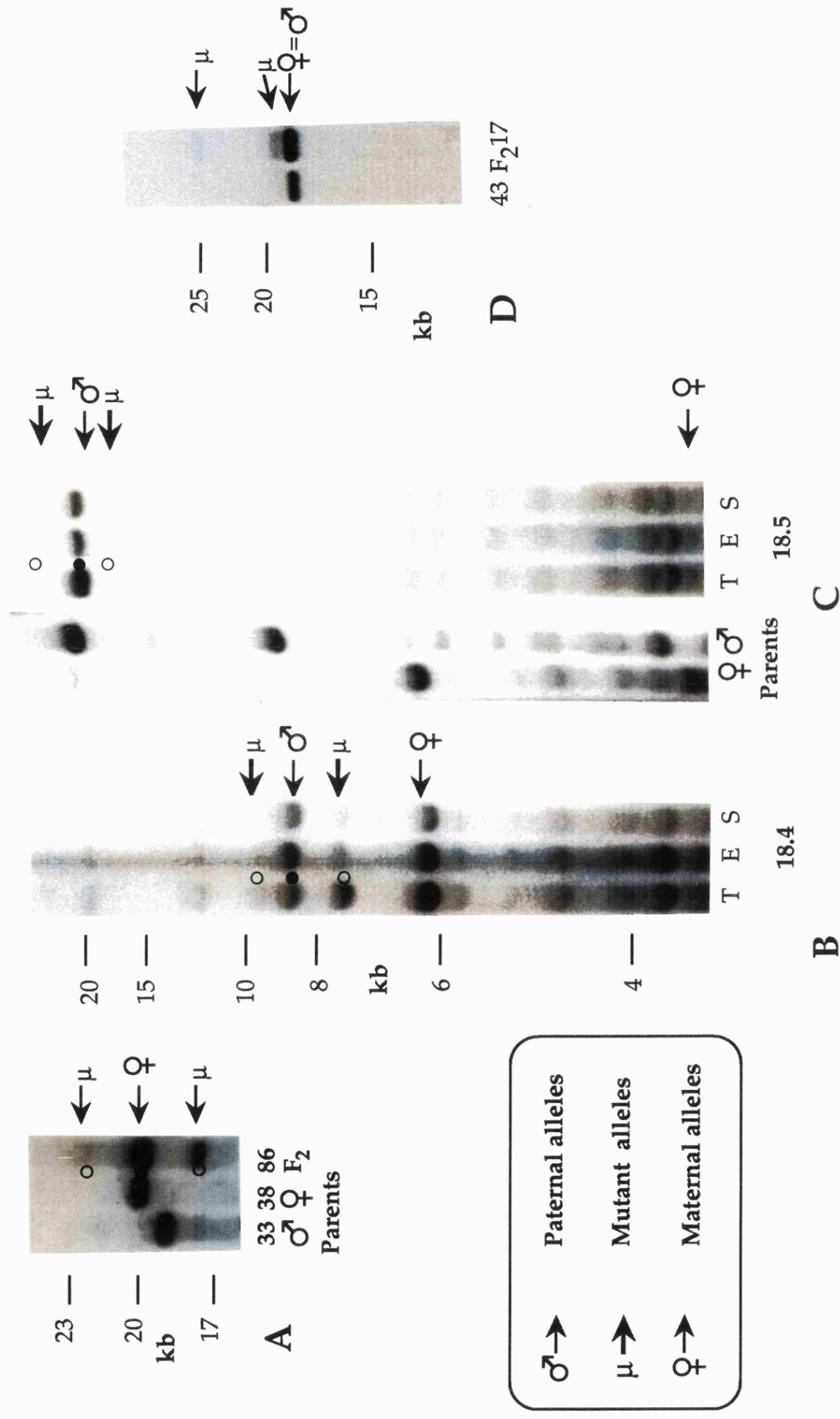
**Table 8.1    Double and triple mosaics at *Hm-2***

| Mouse | alleles affected<br>p - paternal<br>m - maternal | reciprocal<br>sizes? | |
|---|---|---|---|
| 1.1 | ? | no | Double mutant |
| 15.1 | p and m | no | Double mutant |
| 16.5 | both p | no | Double mutant |
| 18.4 | both p | **Yes** | Possible SCE |
| 18.5 | both p | no | Double mutant |
| 4.5 (Triple) | all p | **Yes** and no | Possible SCE and double mutant |
| F₂ 17 | ? | no | Double mutant |
| F₂ 86 | both m | **Yes** | Possible SCE |
| F₂ 145 | both m | **Yes** | Possible SCE |

## 8.4    Conclusions

### 8.4.1    No evidence for inter-allelic recombination.

No double mosaics were seen with reciprocal mutations around both parental alleles (figure 8.2). Moreover, if it is assumed that it is equally likely for the two reciprocal products of inter-allelic recombination to segregate into the same daughter cell as different daughter cells (figure 8.1), it would be expected that if homologous recombination is the major mutational mechanism then at least one half of all adult mosaics would have mutant cells containing both reciprocal products, and thereby show reciprocal mosaicism, with both mutant alleles at equal intensity. Mutants in which the two reciprocal products segregated into different daughter cells could still be seen in adult mice if descendants of both cells enter the embryo. No mosaic mice, either embryos and extra-embryonic tissues or adult mice, had reciprocal mutations of the type predicted by inter-allelic recombination. If inter-allelic events have occurred they have not, therefore, produced a reciprocal gain and loss of repeats.

### 8.4.2  Evidence for unequal SCE

Four mice were seen with mutant alleles that matched the predictions of unequal SCE. The reciprocal alleles formed by SCE must segregate into different daughter cells. Thus, for an adult mouse to have both products of an unequal SCE, both of the daughter cells must have contributed descendants to the embryonic lineage. Of the four double mosaics seen that could be the result of unequal SCE, two were detected in adult DNA, one had one allele in the the embryo and one in the trophoblast, and the other had both alleles in the trophoblast.

### 8.4.3  The majority of mosaics have only a single mutant allele

The majority of mutations are seen as one mutant allele per embryo, or rarely, have two mutant alleles which are unrelated. It has already been shown (chapter 5) that the distribution of mice with 0, 1 or 2 alleles do fit a Poisson distribution, suggesting that mutations occur independently and create only a single mutant allele.

Together these last two observations could suggest that either;

1) SCE does happen, but the major cause of mutation is a single molecule mechanism, such as replication slippage.

2) SCE never happens, the reciprocal sizes are coincidental and the major cause of mutation is replication slippage.

3) SCE is a major mechanism, but in most cases one reciprocal fails to be detected by Southern blot hybridization, due to, for example, trophoblast proliferation. This possibility could be shown by PCR amplification of the rarer reciprocal product.

### 8.5  Influence of reciprocal products on estimates of mutation rate

If eight mutants previously identified were derived from just four unequal SCE events, i.e. they have two mutant alleles resulting from a single mutation event, how are the estimates of mutation rates made in chapter 5b, and the conclusions drawn in chapters 6 and 7 about the timing of mutation affected?

*Chapter 5b    Estimation of somatic mutation rate*

In practise the numbers that are altered are small enough so that all the estimates made so far are not significantly altered.  The proportion of mosaic mice remains the same, although naturally, the proportion of mice in which there have been two independent mutations decreases. Two of the candidate SCE were detected in trophoblast DNA and were not included in the general mutation rate.  If the remaining two pairs of mutant bands were the result of unequal SCE, then the proportion of  alleles that have undergone mutation reduces from 62 out of 558 alleles to 60/558 alleles (11.1% to 10.8%).

*Chapter 6    Mutations shared by the trophoblast and embryo occurred prior to the fifth cell division.*

The numbers of mutations shared between trophoblast and embryo does not alter, since none of the SCE mutants have both mutant alleles scored as shared.  (If *either* of the reciprocal alleles are shared then it is still valid to deduce that the mutation must have arisen before the fifth cell division.  The only potential error is in scoring a pair of reciprocal products, *both* of which are shared by the trophoblast and embryo, as two shared mutations, rather than a single mutation.)

*Chapter 7    The dosages of mosaic alleles can be used to estimate the time of mutation*

Three of the possible SCE mutants have one intense and one faint allele (figure 8.4, table 5.2, table 6.1).  The true time at which the mutation occurred is best estimated from the most intense non-parental allele. For example, in fig 8.3 the hypothetical ICM has mutant alleles with dosages of 0.17 and 0.50.  The former would erroneously be classified as a mutation event at the 4th cell division (see chapter 7), whereas the latter would be correctly classified as mutation that arose at the 2nd cell division.  Therefore, if these mutants are SCEs, they would best be classified by ignoring the least intense allele and basing the classification on the higher dosage allele.  Thus, if the fainter alleles were removed from the data for these four cases, the effect would be to tighten the clustering of mutations to the first two cell divisions, maintaining, if not increasing, the estimation of the peak period for mutations to this time.

## 8.6 Summary

It should be possible to see the effects of somatic recombination between minisatellites by the appearance of two reciprocal mutant alleles, the products of unequal exchange. A few mutants have been seen that match the pattern expected to be produced by an intra-allelic unequal exchange, but not the pattern expected from inter-allelic cross-overs. Unequal sister chromatid exchange may occur, but it is not the major mechanism by which length changes are generated in somatic cells at *Hm-2*.

# 9

# Discussion

## 9.1 Introduction

In this chapter the principal results will be discussed and the possible reasons for the existence of hypervariable DNA considered further.

## 9.2 Cloning of *Hm-2*

### 9.2.1 Cloning of hypervariable arrays

A major impediment to the cloning of minisatellites has been the tendency for tandemly repeated sequences to collapse within bacterial hosts, making size constrained vectors such as lambda inviable; indeed the locus *Ms6-hm* was successfully isolated in λ only due to the co-cloning of a second non-repetitive insert which maintained the size of the vector despite the collapse of the repeat array (Kelly *et al.*, 1989). The charomid technique used to clone *Hm-2* circumvents this problem whilst maintaining the efficiency of *in vitro* packaging. The system has been very successfully developed as a methodical approach for isolating human minisatellites, using an ordered array of charomid clones which can be screened systematically with different multi-locus probes (Armour *et al.*, 1990), an approach which has also proved successful in birds (Hanotte *et al.*, 1991).

### 9.2.2 Probe walking

The minisatellite *Ms6-hm* was first detected by cross-hybridization to the human probe 33.6 (Jeffreys *et al.*, 1987a). In turn, the Mm3-1 probe (specific to *Ms6-hm*) detects a number of novel loci, including *Hm-2* (Kelly *et al.*, 1989). Thus it might be expected that further loci would be detected by cross-hybridization to *Hm-2*. When the *Hm-2* specific probe 9.2 was used as a multi-locus probe in mice, however, the DNA 'fingerprint' pattern detected was very similar to the pattern detected with Mm3-1. Furthermore, the majority of the 'novel' bands seen in the *Hm-2* fingerprint are detected by the human probe 33.15. Thus in this case the potential for further probe walking is limited.

## 9.2.3 Sequences surrounding Hm-2

There are at least three instances where minisatellites have been shown to have expanded from within dispersed repeat elements (Armour *et al.*, 1989b; Kelly *et al.*, 1989 (section 1.4.3.); Armour *et al.*, 1992) and minisatellites are generally associated with dispersed and other tandemly repeated elements (Armour *et al.*, 1989b). Although the sequence immediately flanking Hm-2 shows no homology to any dispersed repeats, the presence of three simple tandem repeat regions (figure 9.1) within 1kb of the Hm-2 repeat is unexpected considering the general non-clustering of microsatellites in humans (Luty *et al.*, 1990). One of these is polymorphic (section 4.3), and since the longest stretch of perfect repeats in microsatellites is the best indication of their potential variability (Weber, 1990) the perfect $A(CA)_{22}$ found within 200bp of the Hm-2 repeats (figures 3.7 and 9.1) is a good candidate for a further polymorphic marker (Weber, 1990). The final repetitive region, although the largest, is composed of several different repeats and is, therefore, unlikely to be variable within inbred mouse strains. For the study of mutation processes (see section 9.4) at Hm-2 it would be useful to have a polymorphic marker on the other side of the Hm-2 repeat array (see figure 9.1). The polymorphic insertion detected by the enzyme HinfI in some C57BL/6 mice (section 3.6) provides such a marker, and further experiments could determine whether mutations at Hm-2 are associated with exchange of these flanking markers.

**Figure 9.1    Location of flanking markers**



$A(CA)_{16}$

$(AT)_5 G(CA)_{12} (AG)_6 (G)_8 (AG)_3 G(GA)_3$

$A(CA)_{22}$

$(GGCA)_n$

2kb Insertion
polymorphism
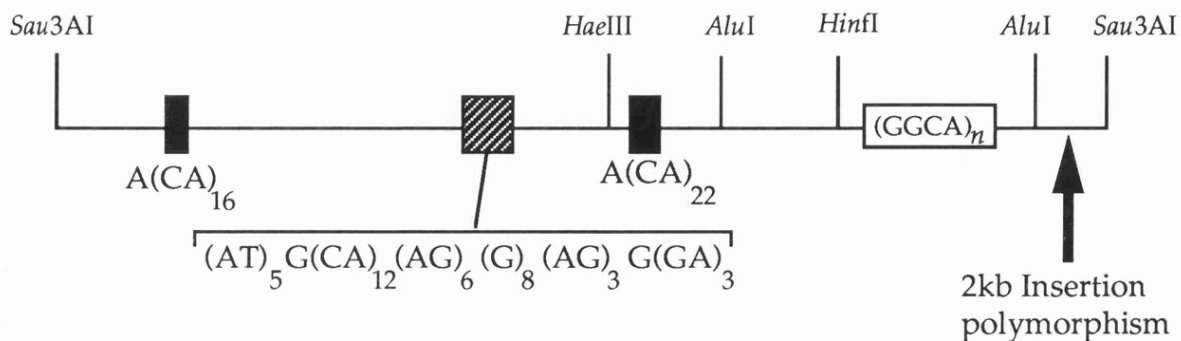
## 9.3     Mutational mechanisms

A variety of mechanisms have been proposed to generate new length alleles in tandem repeated DNA; meiotic or mitotic unequal exchange, inter or intra-allelic gene conversion, sister-chromatid exchange, and slippage and reannealing at replication. Intra-molecular recombination or 'looping out' is a further possibility but can be ruled out as the only major mechanism since only deletions would be generated.

### 9.3.1     Inter-molecular mechanisms

Mechanisms based on recombination, either between arrays on homologous chromosomes, or between sister-chromatids (Jeffreys *et al.*, 1985a), have been the most obvious candidates for the principal generators of new length alleles at minisatellites. This was based on the observation of mutant alleles which had undergone a large change in repeat copy number, unlikely, therefore to have resulted from slippage. All studies to date, however, have failed to demonstrate any exchange of flanking markers in mutant alleles (Wolff *et al.*, 1989; Kelly *et al.*, 1991, Vergnaud *et al.*, 1991). Although it is possible that the mechanisms of germline and somatic mutation are dissimilar, no reciprocal products were found in 15 mutant alleles in human tumours (Armour *et al.*, 1989a) and the present study also failed to find evidence of reciprocal exchanges between the maternal and paternal alleles (chapter 8) supporting the view that unequal exchange between homologous chromosomes is not a principal mechanism, although it is has not been ruled out that some alternative recombination mechanism, such as conversion, that maintains the phase of flanking markers is involved (Wolff *et al.*, 1991).

Evidence for or against the role of unequal sister chromatid exchange is harder to obtain, although such exchanges are believed to be the mechanism by which repeat unit fidelity is maintained in satellite DNA (Smith, 1976). This mechanism is further implied by the identification on human chromosomes of higher-order chromosome specific tandem repetitive structures likely to be due to propagation of novel (mutant) repeat unit types by crossovers (Willard and Waye, 1987).

Analysis of somatic mutations (chapter 8) found some candidate intra-allelic exchange reciprocal products, and furthermore, the lack of variation in repeat sequence found at *Hm-2* and *Ms6-hm* agrees with the predictions by

computer simulations that high rates of crossing over would generate arrays of near identical short tandem repeats (Stephan, 1989).

Analysis of the internal structure of minisatellite alleles by mapping the distribution of variant repeats has enabled more detailed analysis of the products of minisatellite mutations. Studies of the internal structure of large deletion mutants at the human minisatellite MS32 (D1S8) using restriction enzymes and PCR based assays revealed all mutants to be the result of intra-allelic events with no evidence for inter-allelic recombination (Jeffreys et al., 1990). The more recent MVR-PCR technique for analysing internal variation (Minisatellite Variant Repeat PCR, Jeffreys et al., 1991c) which may be used to characterize mutant alleles which have increased in length or undergone very small length changes has, however, provided firm evidence for the involvement of inter-allelic recombination and/or conversion in mutation events at the human loci MS32 and MS205 (Jeffreys et al., 1991c, J. Armour, pers. comm). Since small length change mutations are more common, it is possible that a majority of mutation events involve such events; the failure to detect intermolecular events previously reflects the lower resolving power of the assays used previously. Hm-2 (and Ms6-hm) lack internal variation in repeat sequence and cannot be analysed by this method, although other human loci are being investigated (J. Armour, D. Neil and A.J.Jeffreys pers. comm.).

### 9.3.2 'Slippage' mechanisms

It has been suggested that slippage events occur when duplex DNA strands mispair within regions of short tandem repeats during DNA repair or replication, generating short duplications and deletions (Levinson and Gutman, 1987a). Slippage is believed to be the principal mechanism by which mutation occurs at microsatellites, although there is no direct evidence for this mechanism in vivo. Evidence to support slippage is seen in the typically small changes in copy number in variant microsatellite alleles, and conversely, no evidence of unequal-crossing over with exchange of flanking markers has been seen at microsatellites (Morral et al., 1991; Kwiatkowski et al., 1992), although this does not exclude sister-chromatid exchange or conversion events.

The lower variability seen at microsatellites with imperfect repeats, in which the ability to anneal out of alignment but in register would be impaired, has been suggested to be further evidence for slippage mechanisms (Weber

1990). Finally slippage has been demonstrated to increase lengths of di- and trinucleotide repeat arrays *in vitro* (Schlötterer and Tautz, 1992).

At minisatellites, the evidence is more against recombination than for slippage. It might be presumed, however, that as repeat units get larger the potential for in-register strand slippage would be reduced. Thus perhaps slippage is only particularly effective when repeat sequences are small, a criterion which *Hm-2* fulfills. If a slippage mechanism is responsible for generating mutations at microsatellites it might be argued that *Hm-2* would undergo the same processes.

## 9.4    Factors affecting variability and mutation rate at minisatellite loci

Tandem repeats have been suggested to be the natural 'ground state' for non-selected DNA (Levinson and Gutman 1987a). Beyond the obvious tandem repetitive nature of a large portion of eukaryotic genomes, more complex regions of DNA have been found which are composed of cryptic repeats (Tautz *et al.*, 1986). Gray and Jeffreys (1991) have suggested that minisatellite loci are transient, the most variable loci are believed to exist as high copy number regions for a short evolutionary time, whereas less variable and more stable regions may persist longer (section 1.3.5). A question of fundamental importance is what are the mechanisms which initiate and maintain variability, and what are the factors that actually determine the variability and mutation rate to new length alleles at given points in the genome?

The *Hm-2* minisatellite has a short G/C rich repeat in very long tandem arrays; all these features are candidate factors promoting mutation.

### 9.4.1    Sequence

The sequence of the repeat unit itself has been proposed to be a determining factor. The majority of hypervariable minisatellites isolated from higher eukaryotes to date have a G/C rich repeat sequence. This implies conservation, and a possible functional role for these sequences. The resemblance of the original 'core' sequence (GGAG<u>GTGGGC</u>A<u>GG</u>A<u>RG</u> derived from the repeat sequences of minisatellites detected by 33.6 and 33.15) to the chi signal of *E. coli* which promotes homologous recombination (similarities underlined above) prompted suggestions that minisatellites may promote recombination, and new length alleles could result from unequal

recombination between chromosomes during meiosis or mitosis (Jeffreys *et al.*, 1985a). Other consensus minisatellite repeats have been proposed, GNNGTGGG (Nakamura *et al.*, 1987a), GGCAGG, (Mitani *et al.*, 1990) which resemble both $\chi$ and the 'Jeffreys core' and are G-rich, although the prevalence of G/C rich variable minisatellites may in part reflect a bias imposed by identifying new loci by cross-hybridization to G/C rich multi-locus probes. Minisatellite core probes have been shown to preferentially hybridize to chiasmata in human bivalents (Chandley & Mitchell, 1988) and the distribution of minisatellites to predominantly the proterminal regions of human chromosomes (Royle *et al.*, 1988, Armour *et al.*, 1990)) matches the distribution of recombination nodules associated with synaptonemal complexes believed to be the visible cytological representation of meiotic recombination events (Solari, 1980), further suggesting a possible role in promoting chromosome pairing and recombination.

Further evidence for the association of tandem repeats with recombination hotspots comes from the hotspots found at the mouse major histocompatability complex (reviewed by Steinmetz *et al.*, 1987), two of which have been characterized and found to be close to tetra-nucleotide repeat arrays; $(CAGG)_{7-9}$ at the hotspot within the E$\beta$ gene (Steinmetz *et al.*, 1986), and $(CAGA)_{4-6}$ in the hotspot between the A$\beta$3 and A$\beta$2 genes (Uematsu *et al.*, 1986). Interestingly the former is the same repeat motif as *Hm-2*, although unlike *Hm-2*, variability at the E$\beta$ (CAGG) array is minimal.

The presence of minisatellites on plasmids has been shown to enhance recombination at a distance; Wahls *et al.*, 1990 found an elevated level of homologous recombination in plasmids containing tandem repeats based on the core sequence when introduced into EJ bladder carcinoma cells.

Allied to this evidence is the presence in eukaryotic cells of several proteins that bind to minisatellite repeats (Collick and Jeffreys, 1990; Collick *et al.*, 1991; Wahls *et al.*, 1991). Although these are not yet fully characterized, their presence suggests that interactions might be occurring at these sequences. It is difficult to reconcile a minisatellite function, however, given that minisatellites are not evolutionary conserved (Gray and Jeffreys 1991) although the processes acting upon minisatellites may merely be a by-product of some other cellular function. The relationship of the core sequence to the *E. coli* $\chi$ sequence, therefore, is probably coincidental (Although it has been suggested that the introduction of clones containing high numbers of copies of sequences similar to a genomic regulatory signal sequence may explain the

generally refractory nature of minisatellites to cloning in *E. coli* (Armour, 1990.))

Despite the association of tandem repeat arrays with recombination hotspots, no evidence suggests that the link is a functional one: at the murine MHC hotspots meiotic recombination events characterised at the DNA level were seen not to change the length of the repeat arrays, and other hotspots have been described at which tandem repeats play no part. For example, a 1.9kb region upstream of the β-globin gene, containing a $(CA)_n$ microsatellite, significantly enhanced recombination when transfected into yeast cells, but continued to enhance recombination after deletion of the microsatellite (Treco *et al.*, 1985). Furthermore, many monomorphic minisatellites have chi-like motifs or 'core' repeats, variable AT rich minisatellites have been isolated (Stoker *et al.*, 1985, Knott *et al.*, 1986) and variable DNA fingerprints can be detected using short synthetic AT rich sequences (Epplen *et al.*, 1991), suggesting that a G/C rich motif is not the final arbiter of variability, although the incidence of G/C rich minisatellites in the human genome suggests that a 'core' like repeat may be a promoter of activity.

### 9.4.2 Position of repeat sequences in the genome

Recent analysis of the internal structure of MS32 alleles (Jeffreys *et al.*, 1991c) has provided evidence that mutation events at MS32 are polarized to one end of the array. All the mutations within MS32 alleles studied to date have been at same end of the array (subsequently termed the 5' end), suggesting a bias of mutation. Further evidence for polarity of mutation is given by the internal structure of normal MS32 alleles which typically have very variable 5' ends but 3' ends that may be grouped into similar related haplotypes (Jeffreys *et al.*, 1990, 1991c, Tamaki *et al.*, manuscript in preparation). The variability of the 5' end extends into the flanking DNA, with a greater than average number of polymorphic sites (Monckton *et al.*, manuscript in preparation). The available evidence therefore points to the influence of a region flanking the minisatellite that elevates the mutation rate, but this effect gradually diminishes with distance, which would suggest that there would be no correlation between array length and mutation rate. Thus, given that tandem repeats are not themselves promoters of recombination, it is still possible that some or all repetitive regions are associated with flanking hotspots which generate variability. Studies involving transgenic mice harbouring MS32 repeat arrays are underway, and will hopefully give an

insight into the effect of surrounding sequence on minisatellite mutation rates (A.J. Jeffreys, A. Collick, and M. Allen, personal communication).

### 9.4.3   Size of alleles: does the length of an allele determine mutation rate?

In general, the most variable loci have alleles which lie between 4 and 20 kb; the majority of variation seen on a DNA fingerprint is seen amongst the largest fragments, which suggests that variability is a property of large alleles. Thus it might be expected that at a given locus the mutation rate would increase in proportion to allele size, but as yet, no firm evidence has been produced to support this theory.   There was no significant evidence for size bias for mutations at *Hm-2*, and on the contrary, there are clear examples of loci which do not follow this tendency.   Firstly, minisatellites which have similar allele size distributions may have different heterozygosities, which correlate with higher and lower mutation rates (Eg MS8 and MS31, Wong *et al.*, 1987).   Secondly, there are loci which are small, yet are highly variable, eg YNZ22 (Wolff *et al.*, 1988) and MS228B (Armour, 1990).   Finally, there are large arrays of tandem repeats which nevertheless appear to be monomorphic within a population; indeed the invariant locus that cross-hybridizes to the 9.2 probe at high stringency (section 3.4) must bear a strong sequence similarity to *Hm-2*, and moreover, the size of this cross-hybridizing fragment is comparable to the size of smaller *Hm-2* alleles.   This suggests that neither sequence similarity or array size are sufficient to generate hypervariability.

In contrast, at the smaller microsatellite loci (which may have an allele size distribution under 30bp and yet still be variable) there is a correlation between increasing copy number and variability.

There is a strong correlation between the number of perfect repeats and variability at microsatellites (Weber 1990), and the rate of slippage events has been shown to increase with increasing length of blocks of repeats for $(CA)_n$ sequences in M13 (Levinson and Gutman, 1987b) and for (CTCTT)n sequences in *Escherichia coli* (Murphy *et al.*, 1989).   This association may also hold true for minisatellites, especially those with smaller repeat lengths; in general the loci with large arrays of short repeats show extreme variability.   For example, MS1 (the most variable human locus characterized) *Ms6-hm*, and *Hm-2*  have small repeats, (9, 5 and 4bp) and are therefore quite similar to microsatellites. Recently polymorphic trinucleotide repeat sequences have been discovered within the regions responsible for fragile X-linked mental retardation ($(CGG)_n$), myotonic dystrophy (DM, $(CTG)_n$), (reviewed by Richards and Sutherland,

1992) and X-linked spinal and bulbar muscular atrophy (Kennedy's disease $(CAG)_n$, La Spada *et al.*, 1991). Intriguingly, all the evidence points to an increase in copy number of the repeat sequences being responsible for the DNA instability in these disorders, although the molecular basis for this remains unknown. In normal individuals the fragile-X CGG repeats have a copy number between 6 and 54; carrier males, which have a much increased probability of further mutation and array expansion have between 52 and 200 repeats, affected individuals have over 200 repeats (Fu *et al.*, 1991). There is evidence that meiotic recombination is not the cause of the change in copy number, since flanking markers are not exchanged. At the fragile X and myotonic dystrophy repeats, the copy number of the repeat has a dramatic affect on the likelihood of further expansion of the array and progression to a fully affected phenotype in following generations. Within an affected individual, the germline instability of carrier males extends to somatic instability, with multiple alleles or a heterogeneous smear of bands detected in affected individuals (Fu *et al.*, 1991). On the basis of copy number, therefore, it would not be surprising that *Hm-2* with a tandem array of 500 to 5000 repeats would be massively unstable, and in many respects the variability seen at *Hm-2* mirrors high copy number DM and fragile X repeats; a large repeat array undergoes frequent mutation both in the germline and soma.

The sequences of both the DM and Kennedy's disease repeats are essentially the same, although transcription is from opposite strands. Interestingly, this sequence (..GCA..) bears the same relationship to *Hm-2* (.GGCA.) as *Hm-2* does to *Ms6-hm* (.GGGCA.). Although there was no significant correlation between array length and mutation rate at *Hm-2* (chapter 5), this might reflect a threshold that is below the size of all of the alleles analysed, but is otherwise similar to the situation in the fragile X and DM repeat sequences. In the latter cases, selection maintains the array size at low copy number, whereas maybe *Hm-2* is an unstable sequence that has escaped the constraints of selection.

Thus there is evidence that location near a recombinational hotspot may be the major factor at some loci, or else chance expansion of tandem repeats past a threshold level promotes further mutation. The midisatellites which are yet to be fully characterized may be examples of minisatellites that have passed this threshold.

## 9.5 Mutation and variability at *Hm-2*

### 9.5.1 Germline mutation

The germline mutation rate to new length alleles at *Hm-2* has been estimated as 3.6% per gamete (chapter 5a). This puts the locus amongst the most unstable minisatellites yet characterized, despite the degree to which this rate is underestimated due to the large allele sizes and short repeat sequence preventing detection of small length changes. Like other minisatellite loci this high mutation rate was reflected in high allelic length variability with heterozygous individuals found even within an inbred strain. Variation was restricted to some extent, as shown by the clustering of BL/6 and DBA allele sizes in BXD RI mice (figure 4.7). Thus although the frequency of mutation events was great enough such that all alleles in the BXD mice were of different sizes, the amount of length change accompanying each mutation event was small, and therefore allele lengths tend only to depart gradually from their original length. The difference in typical allele sizes in different populations of C57BL/6 and CBA mice (section 6.2) presumably reflects the effects of this gradual drift coupled with regular population bottlenecks in commercial colonies.

### 9.5.2 Applications of a highly variable marker

The extreme variability seen at *Hm-2* would suggest that the locus would be an ideal polymorphic marker (see chapter 1), and *Hm-2* and *Ms6-hm* complemented each other in that they were controls against DNA mixing, and gave proof of parentage in the pedigrees used in this study. However, the extreme instability of *Hm-2* would detract from some applications, such as quality control of inbred strains, since a strain specific pattern would soon vary even in the absence of contamination.

The fate of mutant cells resulting from an early somatic mutation will be reflected in the distribution of mosaicism in adult tissues. Thus somatic mutant alleles might provide informative and innocuous lineage markers. The most important features of lineage markers are heritability, ubiquitous distribution, and *in situ* detection (reviewed by Rossant, 1987). Minisatellite markers fulfil the first two criteria and would be developmentally neutral. The uses would be limited, however, because the majority of naturally

occurring new mutations occur at the first few cell divisions and would be less likely to provide information about later developmental pathways.

### 9.5.3 Somatic mutation at minisatellites

In complete contrast to the overwhelming majority of other characterized minisatellites, both *Hm-2* and *Ms6-hm* undergo a significant level of mutation in somatic cells (Kelly *et al.*, 1991, chapter 5b). Mitotic events have been shown to occur at other (human) loci by PCR amplification of DNA from sperm and blood (Jeffreys *et al.*, 1990), and analysis of tumours (Armour *et al.*, 1989a). The rates observed (a maximum of $4\times10^{-6}$ per mitosis in gastrointestinal tumours (Armour *et al.*, 1989)), however, were very low in comparison to *Hm-2* and *Ms6-hm* and in general, other hypervariable minisatellites appear to be somatically stable. Furthermore, the available evidence strongly suggests that the somatic mutations at *Hm-2* and *Ms6-hm* are predominantly confined to the first few cell divisions post fertilization (chapter 7), whereas there has only been one example from over 6000 human single locus profiles of a human minisatellite somatic mutation with the same characteristics of high dosage (detectable by conventional Southern blotting techniques) and presence in different tissue layers (detected in blood and buccal cells) similar to the mutations seen at *Hm-2* and *Ms6-hm* which might, therefore, have resulted from an early somatic mutation (A.J. Jeffreys and I. Patel, personal communication).

### 9.5.4 What is the basis for the early somatic mutation events at Ms6-hm and Hm-2

There are two central questions; firstly, why do these particular sequences have a high somatic mutation rate when other minisatellites do not? Secondly, why do these mutations occur predominantly within a certain developmental period?

The majority of somatic mutations at *Hm-2* occur during the first few cell divisions (chapter 7), and therefore some process, specific to or concentrated at this time is raising the mutation rate at these minisatellite arrays. The 1 to 2-cell stage is a period marked by a number of important developmental events, notably the activation of transcription by the embryonic genome and the inactivation or degradation of much of the maternal mRNA (reviewed by Telford *et al.*, 1990). In summary, RNA

synthesis is initiated at the 1-2 cell stage which is reflected in changes in protein patterns within the same period, although protein synthesis through the first cleavage is supported almost exclusively by maternal mRNA. Many transcripts found in 2-cell embryos are not present in the oocyte and appear to represent a set of embryo specific genes (Taylor and Pikó, 1987). Furthermore, the turnover and degradation of these early embryonic transcripts is variable, suggesting that some proteins are transiently expressed. With this in mind, a plausible cause of elevated mutation might be the transient expression of factor that promotes instability, the levels of expression of this factor would closely follow the pattern of mutation shown in figure 7.2., and could also account for sequence specificity by binding only to (G)GGCA repeats. However, at least one invariant GGGCA array is present in the mouse genome (B10, Kelly 1990), although this had a length of only 4 repeats. If this region is truly stable and not simply a 'frozen accident' this suggests that any binding activity is size dependant. This is consistent with the observation that at least one minisatellite binding protein requires a minimum of two 12 nucleotide G-rich repeats for binding, and the efficiency increases if a substrate with 5 repeats is used (Collick and Jeffreys, 1991). A remaining question is what role such a protein may play within the cell.

An alternative explanation for somatic mutations is that merely due to their extreme length these sequences generate early somatic instability. The remarkably high rates of mutation seen at fragile X CGG repeats have been suggested to arise from difficulties for DNA replication through regions of short repeats; once the number of repeats has passed a threshold level (~50 repeats) replication pauses within the repeats, with subsequent resolution of the free ends by an error-prone mechanism (Fu et al., 1991). This agrees with the association of free ends with slippage in vitro; Schlötterer and Tautz (1992) proposed that the polymerase enzyme leaves the synthesized strand and slippage occurs before a new polymerase complex is established. The single stranded 'bulge' created may then be the target of repair mechanisms, which either recreate the original copy number or elongate the other strand.

What could account for the increased error rate during the first few cell divisions? One possibility is that some factor required for replication, such as the availability of nucleotides, is at a limiting concentration in the early embryo. Thus the polymerase complex may stall or incorporate incorrect bases creating mismatches when long stretches of repeats are encountered, particularly in regions that are last to be replicated. When metabolic inhibitors are added to block de novo purine synthesis 55% of embryos are blocked at the

1-cell stage, and the remainder at the 2-cell stage (Alexiou and Leece, 1992). This suggests that the store of maternal nucleotides in the fertilized egg is sufficient only for at most a single cell division, and thus towards the end of the first replication, the concentrations of nucleotides are reaching a minimum, and only begin to rise when fresh synthesis begins at the 2-cell stage. Perhaps, therefore, the determining factor is the distance a suitable 'error promoting' repeat lies from an origin of replication.

A related question is "why have early somatic mutations not been observed more often at human loci?" If a transiently expressed minisatellite binding protein is responsible, this may simply be absent in human cells, or since the activation of the human embryonic genome occurs later than in mice, at the 4 to 8 cell stage, patterns of expression may be different. Alternatively large arrays of (G)GGCA repeats may be absent in humans, although there has been no specific search for such sequences in the human genome and currently all human minisatellites isolated have repeat sequences ≥8bp, and therefore copy numbers much lower than Hm-2 and Ms6-hm. Neither Mm3-1 or 9.2 produce good fingerprints from human DNA, despite the presence of core related minisatellites, and as further circumstantial evidence, when the 9.2 GGCA probe was used to probe a human genomic library the two most strongly cross-hybridizing clones contained variable midisatellites with GGAA based repeat sequences (Ian Gray, pers comm). Extreme instability has also been observed in fragments detected by a GGCA probe in chickens (J.T. Epplen, pers comm), although the nature of this variable region is unknown.

It is not yet clear to what extent minisatellite repeat sequences are species specific; perhaps the amplification to large arrays is random with respect to the starting repeats, thus large (G)GGCA minisatellites and the associated early somatic instability may be found only within a subset of species. Characterization of further large minisatellite arrays would help to determine the factors responsible for somatic instability.

### 9.5.5. What is the basis for the different incidence of somatic mutation in $F_1$ and $F_2$ mice at Hm-2 and Ms6-hm?

There was a significantly higher number of somatic mutations at Hm-2 observed in $F_1$ mice when compared to $F_2$ mice (section 5b.4). This was also the case at Ms6-hm (Kelly, 1990). An increase in mutation in $F_1$ hybrids has some similarities to the hybrid dysgenesis seen in some Drosophila crosses,

which has been shown to be the result of activation of P element transposition. Dysgenesis occurs when a male of a P element bearing strain is crossed with a female which lacks P elements (M strain). The offspring of such a cross have high levels of mutation and chromosome rearrangements which usually result in sterility, whereas matings with P strain females produce viable offspring which have a P cytotype. P elements code for a transposase which is transcribed in germcells. A truncated version of this protein is produced in somatic cells and it is believed that this acts as a repressor of transposition. Thus dysgenesis only occurs when P elements are introduced into maternal germcells in which the repressor is not present in the cytoplasm (Snyder and Doolittle, 1988).

Somatic mutations at *Hm-2*, however, are equally likely to occur in male or female mice and to have come from the the maternal or paternal allele. Also the differences in mutation rate are not large and furthermore mutations were seen in the pure bred founder strains.

Another possibility is that there is a strain specific effect on mutation rate. No mutations were seen in the DBA founder mice although the number of founder strain mice analysed was small and the different incidence of mutation was only of marginal significance (5 mutations in 34 alleles in C57Bl/6 mice, 0/24 in DBA mice and 1/16 in CBA mice, $0.1>p>0.05$).

Engler *et al.*, (1991) describe a strain specific modifier responsible for methylation of a transgene in C57BL/6 mice independently of the transgene's location. This methylation was gradually lost as the genes were bred into a DBA background, but could be returned by crossing an unmethylated transgene into a methylating strain. Methylation has been proposed to interfere with site specific recombinase activity (Engler *et al.*, 1991); in mice with methylated and unmethylated copies of the transgene described above (which contained V(D)J joining target sequences) the unmethylated copies could be rearranged whereas the methylated copies were not.

Analysis of a methylation sensitive *Hpa*II site immediately flanking *Ms6-hm* demonstrated some evidence for tissue and strain specific differential methylation (Kelly 1990), with increased methylation in C57BL/6 mice particularly in brain DNA. Similarly, a *Hpa*II site located 3' to the *Hm-2* GGCA repeats (outside the 9.2 clone) in tail DNA was analysed. Although both C57BL/6 and DBA DNA showed partial methylation, a greater proportion of the DNA was cut by *Hpa*II in DBA mice than in C57BL/6 mice suggesting slight differential methylation (data not shown). This is the opposite relationship between methylation and increased mutation to that described by Engler *et al*,

although obviously the processes of site specific recombination and minisatellite mutation are different.

Assuming that the difference in mutation rate between the parental C57BL/6 and DBA strains is real it is possible to theoretically explain the different mutation rates in $F_1$ and $F_2$ mice by the mixture and dilution of factors determining high and low level mutation rate. Thus in C57BL/6 mice this factor is at a high level, but low in DBA mice (Although the variation within strains may mean the distributions overlap). In the $F_1$ hybrids all mice are intermediate, and the mutation rate is higher than in the DBA mice but lower than in the C57BL/6 mice. In the $F_2$ the variance is increased which could account for the decrease in mutation rate in $F_2$ mice if the association between the mutation rate and the level of 'mutation promoting factor' is non-linear.

Possible factors which may be influencing the incidence of somatic mutation are the level of nucleotides in the early embryo or the expression of a minisatellite recombinase (see section 9.3.4). Both of these could be candidates for explaining differential mutation rates in $F_1$ and $F_2$ mice if, for example, nucleotide stores are strain specific or strain specific methylation promotes or represses recombinase activity.

This phenomenon requires further investigation for any conclusions to be drawn. Particularly, the mutation rate within different inbred strains needs further quantifying, and the change in mutation rate in hybrid strains and hybrids backcrossed to parental strains requires investigation.

## 9.6    Final Conclusions

The isolation of a second hypervariable locus, *Hm-2*, confirmed the existence of the phenomena observed at *Ms6-hm* and is a step towards finding similarities between such loci which may reveal the biological mechanisms involved in the generation of variability. It must be remembered that only a subset of minisatellite loci have been investigated, and that these are typically the most active and variable. It is counter-intuitive, however, that a single mechanism affecting defined sequences could account for the sheer variety of different variable and non-variable tandemly repeated sequences within the genome, but rather, that several different factors promote variability, these may not be mutually exclusive, and chance may play a very large part in the initial expansion of repeat arrays.

The extraordinary incidence of somatic mosaicism seen at *Ms6-hm* and *Hm-2* demonstrates that not all minisatellites behave the same way, and suggests that further study of minisatellites will produce other surprises.

# Appendix A: Computer simulation of early somatic mutations

240-360 allocate 2-7 cells to the ICM at the 16 cell stage. The probabilities of each are taken from Fleming 1987. Only one of the two daughter cells may be allocated to simulate a periclinal or oblique division (Sutherland *et al.*, 1990)

370-420 increase size of array to 32, calculates further mutations and determines any further allocation of cells to the ICM.

430-467 increases array to 64, checks for mutation, then bins mutations according to dosage in the ICM.

480-500 used for verifying program.

505-END prints results.

## Subroutines

520-560 prints entire array of cells; used for verifying program (routine never called).

570-670 calculates dosage in the trophoblast and ICM. M is a count of the number of mutation events per run (each 'embryo'). If this is >1 then the dosages will be affected, line 640 flags this and the mutation is not included in the totals. Lines 650 and 660 are used to verify program.

680-920 determines mutations at each cell stage.

680-760 doubles array size, each previous cell is given two identical daughter cells.

770 randomly determines mutation. 'tim' records which cell division mutation occurs. Nb. model assumes a mutation is inherited by only one of the two daughter cells.

800-840 Mutation allocated to either daughter cell.

850 used for verifying program

910-920 change the probability of mutation at each cell division.

930-1060 Allocates further cells to ICM to bring the total at the 32-cell stage to between 10 and 16. The probabilities of each are taken from Fleming 1987.

1070-1130  Chooses a random cell until a non-ICM cell is found, then changes to an ICM cell.

2000-21000 Adds to bin totals.  D1 to D6 contain the estimated time of mutation from the dosage in the ICM (D0 counts the mutations which failed to enter the ICM).  The actual cell division at which the mutation was generated is stored in REALT(1-6).

**Program**

```
10 CLEAR
15 D0=0:D1=0:D2=0:D3=0:D4=0:D5=0:D6=0:FAIL=0
20 DIM a(64):    REM ARRAY OF CELLS
30 INPUT "MUTATION RATE"; U
35 REALU=U:      REM REQUIRED IF U CHANGES DURING RUN
40 INPUT "NO OF RUNS"; NR
77 DIM REALT(6):  REM COUNTS NO OF MUTATIONS AT EACH DIV
80 FOR NO=1 TO NR
90 REM RESET VARIABLES AT THE START OF EACH NEW RUN
RANDOMIZE TIMER
95 U=REALU:      REM RETURNS U TO ORIGINAL VALUE FOR NEW RUN
100 J=0:         REM HOLDS CURRENT CELL DIVISION (1-6)
110 ICM=0:       REM NUMBER OF CELLS IN ICM
120 N=0:M=0:     REM COUNT
130 FOR CL=1 TO 64
140 a(CL)=0
150 NEXT CL
160 TR=0:        REM NUMBER OF TROPHOBLAST CELLS
170 IM=0:        REM NUMBER OF MUTANT ICM CELLS
180 TM=0:        REM NUMBER OF MUTANT TROPHOBLAST CELLS
190 T=1:         REM TOTAL CELL NUMBER
200 REM FIRST FOUR DIVS
210 FOR J=1 TO 4
220 GOSUB 680:REM CREATES ARRAY AND CALCULATES MUTATIONS
230 NEXT J
240 REM DETERMINES NUMBER OF CELLS INTO ICM (FLEMING)
250 K=RND(1)
260 IF K<.24 THEN ICM=7: GOTO 320
270 IF K<.44 THEN ICM=6: GOTO 320
280 IF K<.65 THEN ICM=5: GOTO 320
290 IF K<.87 THEN ICM=4: GOTO 320
300 IF K<.96 THEN ICM=3: GOTO 320
310 IF K<.101 THEN ICM=2: GOTO 320
320
330 TR=16-ICM
```

```
340 FOR L=1 TO (2*ICM) STEP 2
350 a(L)=a(L)+10
360 NEXT L
370 REM 32 CELL STAGE
380 ICM=ICM*2
390 TR=TR*2
400 J=5
410 GOSUB 680:REM CALC MUTATIONS
420 GOSUB 930:REM ALLOCATE ANY FURTHER CELLS TO ICM
430 REM 64 CELL STAGE (IF WANT ONLY 32 GOTO 440)
432 ICM=ICM*2
433 TR=TR*2
435 J=6
437 GOSUB 680:REM CALCULATE MUTATIONS
440 GOSUB 570:REM CALCULATE DOSAGE
450 ICMR=(IM/ICM)
460 TRMR=(TM/TR)
467 IF M>=1 THEN GOSUB 2000:REM BIN
470 NEXT NO :REM *END OF ONE RUN*
480 REM FOR I=1 TO (NO-1)
490 REM PRINT "ICM ";ICMR(I);" TR ";TRMR(I);WARN$(I)
500 REM NEXT I
505 CLS
507 PRINT "ESTIMATED CELL DIVS: 1=";D1;"2=";D2;"3=";D3
PRINT "4=";D4;"5=";D5;"6=";D6;"0=";D0 "FAILS=";FAIL
508 PRINT "REAL CELLDIVS";
509 FOR Z=1 TO 6 :PRINT Z;"=";REALT(Z);:NEXT Z
INPUT a
END

520 FOR P=1 TO T
530 PRINT a(P);
540 NEXT P
550 PRINT "NO MUTS="; M
560 RETURN

570 REM CALCULATE T AND ICM MUT DOSAGE
580 REM BUT DOESN'T DISTINGUISH BETWEEN
590 REM DIFFERENT MUTATION EVENTS
600 FOR G=1 TO T
610 IF a(G) >0 AND a(G) <10  THEN TM=TM+1
620 IF a(G) >10 THEN IM=IM+1
630 NEXT G
640 IF M>1 THEN WARN$="!!!!"
650 REM PRINT "ICM MUT ";IM;"/";ICM;" = ";IM/ICM
660 REM PRINT "TROP MUT ";TM;"/";TR;" = ";TM/TR
670 RETURN
```

```
680 REM CALCULATES NEXT MUTATION STAGE
690 N=2^(J-1)
700 FOR I = N TO 1 STEP-1
710 X=RND(1)
730  a=(2*I)-1
740  B=2*I
750  a(B)=a(I)
760  a(a)=a(I)
770 IF X<U THEN
780  M=M+1
785 tim=J:REM IF MORE THAN ONE MUT WILL BE REJECTED LATER ANYWAY
790  Y=RND(1)
800  IF Y<=.5 THEN
810   a(a)=a(a)+1
820  ELSE
830   a(B)=a(B)+1
840  END IF
850 REM  PRINT A,A(A),B,A(B)
860 ELSE
870 END IF
880 NEXT I
890 T=T*2
910  IF J=1 THEN U=.026
915  IF J=2 THEN U=.0042
917  IF J=3 THEN U=.000044
918  IF J=4 THEN U=.0017
919  IF J=5 THEN U=.00106
920 RETURN

930 REM CHANCE OF FURTHER ALLOCATIONS
940 XX=RND(1)
950 IF ICM<10 THEN
960 REM ADD ONE TO ICM UNTIL =10 (MINIMUM)
970 GOSUB 1070
980 GOTO 950
990 ELSE
1000 IF XX>.08 AND ICM<11 THEN GOSUB 1070: GOTO 1010
1010 IF XX>.267 AND ICM<12 THEN GOSUB 1070: GOTO 1020
1020 IF XX>.454 AND ICM<13 THEN GOSUB 1070: GOTO 1030
1030 IF XX>.621 AND ICM<14 THEN GOSUB 1070: GOTO 1040
1040 IF XX>.808 AND ICM<15 THEN GOSUB 1070: GOTO 1050
1050 IF XX>.975 AND ICM<16 THEN GOSUB 1070: GOTO 1060
END IF
1060 RETURN
```

```
1070 REM ADD ONE RANDOM CELL TO ICM
1080 cell=INT(32*RND+1)
1090 IF a(cell)>=10 THEN GOTO 1080
1100 a(cell)=a(cell)+10
1110 ICM=ICM+1
1120 TR=TR-1
1130 RETURN

2000 REM CALCULATE BINS
2005 IF WARN$="!!!!" THEN WARN$ = ""  :FAIL=FAIL+1:GOTO 21000
2008 IF ICMR<.03 THEN D0=D0+1:EST=0:GOTO 2100
2010 IF ICMR>=.03 AND ICMR<=.05 THEN D6=D6+1:EST=6:GOTO 2100
2020 IF ICMR>.05 AND ICMR<=.1 THEN D5=D5+1:EST=5:GOTO 2100
2030 IF ICMR>.1 AND ICMR<=.2 THEN D4=D4+1:EST=4:GOTO 2100
2040 IF ICMR>.2 AND ICMR<=.4 THEN D3=D3+1:EST=3:GOTO 2100
2050 IF ICMR>.4 AND ICMR<=.6 THEN D2=D2+1:EST=2:GOTO 2100
2060 D1=D1+1:EST=1
2100 PRINT NO,EST,B
2110 REALT(tim)=REALT(tim)+1
tim =0 :ICMR = 0
21000 RETURN
```

## Appendix B: Iterative 'best fit' program

This program was used to calculate the mutation rates per division which best fitted the frequency distribution of the observed mutant allele dosages. The program was written and compiled using QuickBASIC (Microsoft).

**b1** to **b6** are the observed numbers of mutation per bin (table 7.

**p1** to **p6** are the true numbers of mutations per division

**exb1** to **exb6** contain the expected number of mutations per bin with each different value of **p1** to **p6**. These are used to calculate the individual $\chi^2$ terms which are stored in **chi1** to **chi6**.

**fp1** to **fp6** store the current 'best' set of **p1** to **p6**, based on the lowest calulated $\chi^2$ which is stored in **chilow**.

The spread of values of **p1** etc tested are determined by the FOR... NEXT.. STEP loops. To avoid excessive computing time these are altered to 'home in' with increasing accuracy on the values which best fit the observed data. The values shown were those used to arrive at the final figures for **p1** to **p6**.

The program returns the values of **p1** to **p6** which best fit the observed totals (**b1** to **b6**) and the value of $\chi^2$

For further details see section 7.6.

```
b1=3
b2=4
b3=21
b4=11
b5=10
chilow=10000
FOR p1=8.7 TO 8.2 STEP -.1
   PRINT p1
   exb1=.33*p1
   chi1=(b1-exb1)^2/exb1
   FOR p2=29.5 TO 28.5 STEP -.1
      exb2=.3*p1+.07*p2
      chi2=(b2-exb2)^2/exb2
      FOR p3=9.5 TO 7.5 STEP -.1
         exb3=.3*p1+.6*p2+.1*p3
         chi3=(b3-exb3)^2/exb3
         FOR p4=3 TO .1 STEP -.1
            exb4=.05*p1+.17*p2+.59*p3+.35*p4
            chi4=(b4-exb4)^2/exb4
            FOR p5=15.5 TO 14.5 STEP -.1
               exb5=.01*p1+.08*p2+.13*p3+.14*p4+.42*p5
               chi5=(b5-exb5)^2/exb5
               chi=chi1+chi2+chi3+chi4+chi5
               IF chi<chilow THEN fp1=p1:fp2=p2:fp3=p3:fp4=p4:fp5=p5
               IF chi<chilow THEN chilow=chi
            NEXT p5
         NEXT p4
      NEXT p3
   NEXT p2
NEXT p1
PRINT "p1=";fp1
PRINT "p2=";fp2
PRINT "p3=";fp3
PRINT "p4=";fp4
PRINT "p5=";fp5
PRINT "chi=";chilow
INPUT a
END
```

# Literature Cited

ALEXIOU, M., AND H.J. LEESE. (1992)
Purine utilisation, de novo synthesis and degradation in mouse preimplantaion embryos. *Development* **114**: 185-192.

ALI, S., C.R. MULLER AND J.T. EPPLEN. (1986)
DNA fingerprinting by oligonucleotide probes specific for simple repeats. *Hum. Genet.* **74**: 239-243.

ALLEN, N.D., S.C. BARTON, M.A.H. SURANI AND R. WOLF. (1987)
Production of transgenic mice. *In* "Mammalian development, a practical approach." IRL press, Oxford.

ALLSHIRE, R.C., G.R. GOSDEN, S.H. CROSS, G. CRANSTON, D. ROUT, N. SUGAWARA, J.W. SZOSTACK, P.A. FANTES AND N.D. HASTIE. (1989)
Telomeric repeat from *T.thermophila* cross-hybridizes with human telomeres. *Nature* **332**: 656-659.

AMOS, B. (1991)
Fauna and flora fingerprinting factfinder *Fingerprinting News* **3**: 15-16.

ARCHIBALD, A., C.S. HALEY, L. ANDERSSON, A.A. BOSMA, W. DAVIES, M. FREDHOLM, H. GELDERMANN, J. GELLIN, M. GROENEN, I. GUSTAVSSON, L. OLLIVIER, E.M. TUCKER AND A. VAN DE WEGHE. (1991)
PiGMaP: A European initiative to map the porcine genome.
*Animal Genet.* **22**: suppl. 82-83.

ARMOUR, J.A.L., I. PATEL, S.L. THEIN, M.F. FEY AND A.J. JEFFREYS. (1989a)
Analysis of somatic mutations at minisatellite loci in tumours and cell lines. *Genomics* **4**: 328-334.

ARMOUR, J.A.L., Z. WONG, V. WILSON, N.J. ROYLE AND A.J. JEFFREYS. (1989b)
Sequences flanking the repeat arrays of human minisatellites: association with tandem and dispersed repeat elements.
*Nucl. Acids Res* **17**: 4925-4935.

ARMOUR, J.A.L., S. POVEY, S. JEREMIAH AND A.J. JEFFREYS. (1990)
Systematic cloning of human minisatellites from ordered array charomid libraries. *Genomics* **8**: 501-512.

ARMOUR, J.A.L. "The isolation and characterization of human minisatellites." PhD, University of Leicester, 1990.

ARMOUR, J.A.L., M. CROSIER AND A. J. JEFFREYS. (1992)
Human minisatellite alleles detectable only after PCR amplification
*Genomics* **12**: 116-124.

AVNER, P., L. AMAR, L. DANDOLO AND G.L. GUENET. (1988)
Genetic analysis of the mouse using interspecific crosses
*Trends Genet.* **4**: 18-23.

BAILEY, D.W. (1971)
Recombinant inbred strains. An aid to finding identity linkage and
function of histocompatability and other genes.
*Transplantation* **11**: 325-327.

BELL, G.I., M.J. SELBY AND W.I. RUTTER. (1982)
The highly polymorphic region near the human insulin gene is
composed of simple tandemly repeated sequences. *Nature* **295**: 31-35.

BENTON, W.D. AND R.W. DAVIS. (1977)
Screening λgt recombinant clones by hybridization to single plaques in
situ *Science* **196**: 180-182.

BIRNBOIM, H.C. AND J. DOLY. (1979)
A rapid alkaline procedure for screening recombinant plasmid DNA.
*Nucl. Acids Res.* **7**: 1513-1523.

BLACKBURN, E.H. (1991)
Structure and function of telomeres. *Nature* **350**: 569-573.

BRITTEN, R.J. AND D.E. KOHNE. (1968)
Repeated sequences in DNA. Hundreds of thousands of copies of DNA
sequences have been incorporated into the genomes of higher
organisms. *Science* **161**: 529-540.

BULAWELA, L., A. FORSTER, T. BOEHM AND T.H. RABBITTS. (1989)
A rapid procedure for colony screening using nylon filters
*Nucleic Acids Res.* **17**: 452.

BURKE, T. AND M.D. BRUFORD. (1987)
DNA fingerprinting in birds. *Nature* **327**: 149-152.

CAPON, D., E. CHEN, A.D. LEVINSON, P.H. SEEBERG AND D.V. GOEDELL. (1983)
Complete nucleotide sequence of the T24 human bladder carcinoma
oncogene and its normal homologue. *Nature* **302**: 33-37.

CHAMBERLAIN, J.S., R.A. GIBBS, J.E. RANIER, P.-N. NGUYEN AND C.T. CASKEY.
(1988)    Deletion screening of the Duchenne muscular dystrophy
locus via multiplex DNA amplification.
*Nucleic Acids Res.* **16**: 11141-11156.

CHANDLEY, A.C. AND A.R. MITCHELL. (1988)
Hypervariable minisatellite regions are sites for crossing-over at meiosis in man. *Cytogenet. Cell. Genet.* **48**: 152-155.

CHURCH, G.M. AND W. GILBERT. (1984)
Genomic sequencing. *Proc. Natl. Acad. Sci. USA* **81**: 1991-1995.

COLLICK, A. AND A.J. JEFFREYS. (1990)
Detection of a novel minisatellite-specific DNA-binding protein. *Nucl. Acids Res* **18**: 2256-2266.

COLLICK, A., M. G. DUNN AND A. J. JEFFREYS. (1991)
Minisatellite binding protein Msbp-1 is a sequence-specific single-stranded DNA-binding protein *Nucleic Acids Res* **19**: 6399-6404.

COLLINS, F.S. (1992)
Positional cloning: Let's not call it reverse anymore. *Nature Genet.* **1**: 3-6.

COPELAND, N.G. AND N.A. JENKINS. (1991)
Development and applications of a molecular genetic linkage map of the mouse genome. *Trends Genet.* **7**: 113-118.

CRUZ, Y.P. AND R.A. PEDERSEN. (1985)
Cell Fate in the Polar Trophectoderm of Mouse Blastocysts as Studied by Microinjection of Cell Lineage Tracers
*Developmental Biology* **112**: 73-83.

DALLAS, J.F. (1988)
Detection of DNA 'fingerprints' of cultivated rice by hybridization with a human minisatellite DNA probe.
*Proc. Natl. Acad. Sci. USA* **85**: 6831-6835.

DENHARDT, D. (1966)
A membrane filter technique for the detection of complementary DNA. *Biochem. Biophys. Acta* **23**: 641-646.

DEVEREUX, J., P. HAEBERLI AND O. SMITHIES. (1984)
A comprehensive set of sequence analysis programs for the VAX *Nucl. Acids Res.* **12**: 387-395.

DONIS-KELLER, H., P. GREEN, C. HELMS, S. CARTINHOUR, B. WEIFFENBACH, K. STEPHENS, T.P. KEITH, D.W. BOWDEN, D.R. SMITH, E.S. LANDER, D. BOTSTEIN, G. AKOTS, K.S. REDIKER, T. GRAVIUS, V.A. BROWN, M.B. RISING, C. PARKER, J.A. POWERS, D.E. WATT, E.K. KAUFMAN, A. BRIKER, R. PHIPPS, H. MULLER-KHALE, T.R. FULTON, S. NG, J.W. SCHUMM, J.C. BRAMAN, R.G. KNOWLTON, D.F. BARKER, S.M. CROOKS, S.E. LINCOLN, M.J. DALY AND J. ABRAHAMSON. (1987)
A genetic linkage map of the human genome. *Cell* **51**: 319-377.

DOWER, W.J., J.F. MILLER AND C.W. RAGSDALE. (1988)
High efficiency transformation of *E. coli* by high voltage electroporation.
*Nucleic Acids Res.* **16**: 6127-6145.

EDWARDS, A., A. CIVITELLO, H.A. HAMMOND AND C.T. CASKEY. (1991)
DNA typing and genetic mapping with trimeric and tetrameric repeats.
*Am. J. Hum. Genet.* **49**: 746-756.

EDWARDS, A., H.A. HAMMOND, L. JIN, C.T. CASKEY AND
R. CHAKRABORTY. (1992)
Genetic variation at five trimeric and tetrameric tandem repeat loci in
four human population groups. *Genomics* **12**: 241-253.

EFSTRADIADIS, A., J.W. POSAKONY, T. MANIATIS, R.M. LAWN, C. O'CONNELL,
R.A. SPRITZ, J.K. DERIEL, B.G. FORGET, S.M. WEISSMAN, J.L. SLIGHTOM,
A.E. BLECHL, O. SMITHIES, F.E. BAVALLE, C.C. SHOULDERS AND
N.J. PROUDFOOT. (1980)
The structure and evolution of the human β-globin gene family
*Cell* **21**: 653-658.

ENGLER, P., D. HAASCH, C.A. PINKERT, L. DOGLIO, M. GLYMOUR, R. BRINSTER
AND U. STORB. (1991)
A strain-specific modifier on mouse chromosome 4 controls the
methylation of independant transgene loci. *Cell* **65**: 939-947.

EPPLEN, J.T., H. AMMER, C. EPPLEN, C. KAMMERBAUER, R. MITREITER,
L. ROEWER, W. SCHWAIGER, V. STEIMLE, H. ZISCHLER, E. ALBERT,
A. ANDREAS, B. BEYERMANN, W. MEYER, J. BUITKAMP, I. NANDA,
M. SCHMID, P. NURNBERG, S.D.J. PENA, H. POCHE, W. SPRECHER,
M. SCHARTL, K. WEISING AND A. YASSOURIDIS. (1991)
Oligonucleotide fingerprinting using simple repeat motifs: A
convenient way to detect hypervariability for multiple purposes. *In*
"DNA fingerprinting: approaches and applications.", ed. T. T. Burke,
G. Dolf, A.J. Jeffreys, and R. Wolff. :Birkhauser Verlag, Bern.

EVANS, E.P. (1987)
Karyotyping and sexing of gametes, embryos, and foetuses and *in
situ* hybridization to chromosomes. *In* "Mammalian development, a
practical approach": IRL press, Oxford.

FALCONER, D.S. Introduction to quantitative genetics. Essex: Longman, (1960).

FAY, M.F. AND A. TOBLER. (1991)
Assessment of DNA fingerprinting as a method for validating the
identity of cancer cell lines maintained in long term culture.
*Nucleic Acids Res.* **19**: 3464.

FEINBERG, A.P. AND B. VOGELSTEIN. (1984)
A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* **137**: 266-267.

FESTING, M.F.W. (1979)
Inbred strains in biomedical research. Macmillan Press, London .

FLEMING, T.P. (1987)
A Quantitative Analysis of Cell Allocation to Trophectoderm and Inner Cell Mass in the Mouse Blastocyst *Developmental Biology* **119**: 520-531.

FOWLER, S.J., P. GILL, D.J. WERRETT AND D.R. HIGGS. (1988)
Individual specific DNA fingerprints from a hypervariable region probe: alpha globin 3' HVR. *Hum. Genet.* **79**: 142-146.

FRANKE, U., P.A. LALLEY, W. MOSS, I. IVY AND J.D. MINNA. (1977)
Gene mapping in *Mus musculus* by interspecific cell hybridization: assignment of the genes for tripeptidase-1 to chromosome 10, dipeptidase-2 to chromosome 18, acid phosphatase-1 to chromosome 12, and adenylate kinase-1 to chromosome 2.
*Cytogenet. Cell. Genet.* **19**: 57-84.

FRIEDMAN, J.M., B.S. SCHNEIDER, D.E. BARTON AND U. FRANCKE. (1989)
Level of Expression and Chromosome Mapping of the Mouse Cholecystokinin gene: Implications for Murine Models of Genetic Obesity. *Genomics* **5**: 463-469.

FU, Y-H., D.P.A. KUHL, A. PIZZUTI, M. PIERETTI, J.S. SUTCLIFFE, S. RICHARDS, A.J.M.H. VERKERK, J.J.A. HOLDEN, R.G. FENWICK, S.T. WARREN, B.A. OOSTRA, D.L. NELSON AND C.T. CASKEY. (1991)
Variation of the CGG repeat at the Fragile X site results in genetic instability: resolution of the Sherman paradox. *Cell* **67**: 1047-1058.

GARDNER, R.L. (1983)
Origin and differentiation of extra-embryonic tissues in the mouse. *Int. Rev. Exp. Pathol.* **24**: 63-133.

GEORGES, M., M. LATHROP, P. HILBERT, A. MARCOTTE, A. SCHWERS, S. SWILLENS, G. VASSART AND R. HANSET. (1990)
On the use of DNA fingerprints for linkage studies in cattle. *Genomics* **6**: 461-474.

GILL, P., A.J. JEFFREYS AND D.J. WERRETT. (1985)
Forensic application of DNA 'fingerprints'. *Nature* **318**: 577-579.

GOLENBERG, E.M., D.E. GIANNASI, M.T. CLEGG, C.J. SMILEY, M. DURBIN, D. HENDERSON AND G. ZURAWSKI. (1990)
Chloroplast DNA sequence from a Miocine *Magnolia* species. *Nature* **344**: 656-658.

GRAY, I.C. AND A.J. JEFFREYS. (1991)
Evolutionary transience of hypervariable minisatellites in man and the primates. *Proc. R. Soc. Lond. B.* **243**: 241-253.

GROSS, D.S. AND W.T. GARRARD. (1986)
The ubiquitous potential Z-forming sequence of eukaryotes (dT-dGn.(dC-dA)n is not detectable in the genomes of eubacteria, archaebacteria or mitochondria. *Mol. Cell. Biol.* **6**: 3010-3013.

HAGELBERG, E., B. SYKES AND R. HEDGES. (1989)
Ancient bone DNA amplified. *Nature* **342**: 485.

HAGELBERG, E., I. C. GRAY AND A. J. JEFFREYS. (1991a)
Identification of the skeletal remains of a murder victim by DNA analysis. *Nature* **352**: 427-429.

HAGELBERG, E., L.S. BELL, T. ALLEN, A. BOYDE, S.J. JONES AND J.B. CLEGG. (1991b)
Analysis of ancient bone DNA: techniques and applications. *Phil. Trans. R. Soc. Lond. B* **333**: 399-407.

HALDANE, J.B.S AND C.H. WADDINGTON. (1931)
Inbreeding and Linkage *Genetics* **16**: 357-374.

HAMADA, H. AND T. KAKUNAGA. (1982)
Potential Z-DNA forming sequences are highly dispersed in the human genome. *Nature* **298**: 396-398.

HAMADA, H., M.G. PETRINO AND T. KAKUNAGA. (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.

HAMADA, H., M.G. PETRINO, T. KAKUNAGA, M. SEIDMAN AND STOLLAR B.D. (1984)
Characterization of genomic Poly(dT-dG).Poly(dC-dA) sequences: structure, organization and conformation. *Mol. Cell. Biol.* **4**: 2610-2612.

HANAHAN, D. (1983)
Studies on transformation of *Esherishia Coli* with plasmids *J. Mol. Biol.* **166**: 557-580.

HANOTTE, O., T. BURKE, J.A.L. ARMOUR AND A.J. JEFFREYS. (1991)
Hypervariable minisatellite DNA sequences in the Indian peafowl, *Pavo cristatus. Genomics* **9**: 587-597.

HANOTTE, O., E. CAIRNS, T. ROBSON, M. DOUBLE, AND T. BURKE. (1992)
Cross species hybridization of a single-locus minisatellite probe in passerine birds. *Molecular Ecology* **1** (in press)

HARRIS, P.C., S. THOMAS, P.J. RATCLIFFE, M.H. BREUNING, E. COTO AND
C. LOPEZ-LARREA. (1991)
Rapid genetic analysis of families with polycystic kidney disease I by
means of a microsatellite marker. *Lancet* **338**: 1484-1487.

HASTIE, N. D. (1989)
Highly repeated DNA families in the genome of *Mus musculus*.
*In* "Genetic variants and strains of the laboratory mouse", ed. M.F Lyon
and A.G. Searle. 2nd ed. Oxford University Press, Oxford.

HAZAN, J., C. DUBAY, M.-P. PANKOWIAK, N. BECUWE
AND J. WEISSENBACH. (1992)
A genetic linkage map of human chromosome 20 composed entirely of
microsatellite markers. *Genomics* **12**: 183-189.

HIGGS, D.R., S.E.Y. GOODBOURN, J.S. WAINSCOAT, J.B. CLEGG AND
D.J. WEATHERALL. (1981)
Highly variable regions of DNA flank the human alpha globin genes.
*Nucl. Acids Res.* **9**: 4213-4224.

HILLYARD, A.L., D.P. DOOLITTLE, M.T. DAVISSON AND T.H. RODERICH. (1991).
Locus map of the Mouse with comparative map points of human on
mouse. The Jackson Laboratory,

HORZ, W. AND W. ALTENBURGER. (1981)
Nucleotide sequence of mouse satellite DNA.
*Nucleic Acids Res.* **9**: 2989-2998.

JEFFREYS, A.J., V. WILSON, D. WOOD, J.P. SIMONS, R.M. KAY AND
J.G. WILLIAMS. (1980)
Linkage of adult $\alpha$ and $\beta$-globin genes in *X. laevis* and gene duplication
by tetraploidisation. *Cell* **21**: 555-564.

JEFFREYS, A.J., V. WILSON AND S.L. THEIN. (1985a)
Hypervariable 'minisatellite' regions in human DNA.
*Nature* **314**: 67-73.

JEFFREYS, A.J., V. WILSON AND S.L. THEIN. (1985b)
Individual-specific 'fingerprints' of human DNA. *Nature* **316**: 76-79.

JEFFREYS, A.J., J.F.Y. BROOKFIELD AND R. SEMEONOFF. (1985c)
Positive identification of an immigration test case using human DNA
fingerprints. *Nature* **317**: 818-819.

JEFFREYS, A.J. AND D.B. MORTON. (1987)
DNA fingerprints of dogs and cats. *Anim. Genet.* **18**: 1-15.

JEFFREYS, A.J., WILSON, V., KELLY, R., TAYLOR, B.A., BULFIELD, G. (1987a)
Mouse DNA 'fingerprints': analysis of Chromosome localization and
germ-line stability of hypervariable loci in recombinant inbred strains.
*Nucleic Acids Research* **15**: 2823-2836.

JEFFREYS, A.J., J. HILLEL, N. HARTLEY, G. BULFIELD, D.B. MORTON, V. WILSON,
Z WONG AND S. HARRIS. (1987b)
The implications of hypervariable DNA regions for animal
identification: hypervariable DNA and genetic fingerprints.
*Anim. Genet.* **18**: 141-142.

JEFFREYS, A.J., N.J. ROYLE, V. WILSON AND Z. WONG. (1988a)
Spontaneous mutation rates to new length alleles at tandem-repetitive
hypervariable loci in human DNA. *Nature* **332**: 278-281.

JEFFREYS, A.J., V. WILSON, R. NEUMANN AND J. KEYTE. (1988b)
Amplification of human minisatellites by the polymerase chain
reaction: towards DNA fingerprinting of single cells
*Nucleic Acids Research* **16**: 10953-10971.

JEFFREYS, A.J., R NEUMANN AND V WILSON. (1990)
Repeat Unit Sequence Variation in Minisatellites: A Novel Source of
DNA Polymorphism for Studying Variation and Mutation by Single
Molecule Analysis. *Cell* **60**: 473-485.

JEFFREYS, A.J., M. TURNER AND P. DEBENHAM. (1991a)
The efficiency of multilocus DNA fingerprint probes for
individualization and establishment of family relationships,
determined from extensive casework. *Am. J. Hum. Genet.* **48**: 824-840.

JEFFREYS, A. J., A. MACLEOD, K. TAMAKI, D. L. NEIL
AND D. G. MONCKTON. (1991c)
Minisatellite repeat coding as a digital approach to DNA typing
*Nature* **354**: 204-209.

JENKINS, N.A., N.G. COPELAND, B.A. TAYLOR, H.G. BEDIGAN
AND B.K. LEE. (1982)
Ecotropic murine leukaemia virus DNA content of normal and
lymphomatous tissues of BXH-2 recombinant inbred mice.
*Journal of Virology* **42**: 379-388.

JOHN, B. AND G.L.G. MIKLOS.
The eukaryote genome in development and evolution.
Allen & Unwin, England (1988).

JOHNSON, M.H. AND B. MARO. (1986)
Time and space in the mouse early embryo: a cell biological approach to cell diversification. *In* "Experimental approaches to mammalian embryonic development.", ed. Rossant and Pedersen.
Cambridge University Press, New York .

JULIER, C., B. DE GOUYON, M. GEORGES, J-L. GUENET, Y. NAKAMURA, P. AVNER AND G.M. LATHROP. (1990)
Minisatellite linkage maps in the mouse by cross-hybridization with human probes containing tandem repeats.
*Proc. Nat. Acad. Sci. USA* **87**: 4585-4589.

KAN, Y.W. AND A.M. DOZY. (1978)
Polymorphism of DNA sequence adjacent to human β-globin structural gene: relationship to sickle mutation.
*Proc. Natl. Acad. Sci. USA* **75**: 5631-5635.

KELLY, R., BULFIELD,G., COLLICK,A., GIBBS,M. AND JEFFREYS, A.J. (1989)
Characterization of a Highly Unstable Mouse Minisatellite Locus: Evidence for Somatic Mutation during Early Development
*Genomics* **5**: 844-856.

KELLY, R.G. "Hypervariable minisatellites in mouse DNA."
PhD, University of Leicester, (1990).

KELLY, R., M. GIBBS, A. COLLICK AND A.J. JEFFREYS. (1991)
Spontaneous mutation at the hypervariable mouse minisatellite locus *Ms6-hm*: flanking DNA sequence and analysis of early and somatic mutation events. *Proc. R. Soc. Lond. B.* **245**: 235-245.

KIMURA, M. AND J.F. CROW. (1964)
The number of alleles that can be contained in a finite population
*Genetics* **49**: 725-738.

KIMURA, M. (1983)
The neutral theory of molecular evolution.
Cambridge University Press, Cambridge .

KIPLING, D. AND H.J. COOKE. (1990)
Hypervariable ultra-long telomeres in mice. *Nature* **347**: 400-402.

KIPLING, D., H.E. ACKFORD, B.A. TAYLOR AND H.J. COOKE. (1991)
Mouse minor satellite DNA genetically maps to the centromere and is physically linked to the proximal telomere. *Genomics* **11**: 235-241.

KIT, S. (1961)
Equilibrium sedimentations in density gradients of DNA preparations from animal tissues. *J. Mol Biol.* **3**: 711-716.

KNOTT, T., S. WALLIS, R. PEASE, L POWELL AND J. SCOTT. (1986)
A hypervariable region 3' to the human apolipoprotein B gene.
*Nucl. Acids Res.* **14**: 9215.

KRAYEV, A.S., T.V. MARKUSHEVA, D.A. KRAMEROV, A.P. RYSKOV,
K.G. SKRYABIN, A.A. BAYEV AND G.P. GEORGIEV. (1982)
Ubiquitous transposon-like repeats B1 and B2 of the mouse genome: B2
sequencing. *Nucl. Acids Res.* **18**: 1121-1127.

KWIATKOWSKI, D.J., E.P. HENSKE, K. WEIMER, L. OZELIUS, J.F. GUSELLA AND
J. HAINES. (1992)
Construction of a GT polymorphism map of human 9q.
*Genomics* **12**: 229-240.

LA SPADA, A.R., E.M. WILSON, D.B. LUBAHN, A.E. HARDING AND
K.H. FISCHBECK. (1991)
Androgen receptor gene mutations in X-linked spinal and bulbar
muscular atrophy. *Nature* **352**: 77-79.

LALLOZ, M.R.A., J.H. MCVEY, J.K. PATTINSON AND E.G.D. TUDDENHAM. (1991)
Haemophilia A diagnosis by the analysis of a hypervariable dinucleotide
repeat within the factor VIII gene. *Lancet* **338**: 207-211.

LANCASTER, C. A., N. PEAT, T. DUHIG, D. WILSON, J. PAPADIMITRIOU, J. TAYLOR
AND S. J. GENDLER. (1990)
Structure and expression of the human polymorphic epithelial mucin
gene: An expressed VNTR unit
*Biochem Biophys Res Commun* **173**: 1019-1029.

LEVINSON, G. AND G.A. GUTMAN. (1987a)
Slipped-strand mispairing: a major mechanism for DNA sequence
evolution. *Mol. Biol. Evol.* **4**: 203-221.

LEVINSON, G. AND G.A. GUTMAN. (1987b)
High frequencies of short frameshifts in poly-CA/TG tandem repeats
borne by bacteriophage M13 in *Escherichia coli* K-12.
*Nucleic Acids Res.* **15**: 5223-5228.

LIPMAN, D.J. AND W.R. PEARSON. (1985)
Rapid and sensitive protein similarity searches. *Science* **227**: 1435-1441.

LITT, M. AND J.A. LUTY. (1989)
A hypervariable microsatellite revealed by in vitro amplification of a
dinucleotide repeat within the cardiac muscle actin gene
*Am. J. Hum. Genet.* **44**: 397-401.

LOVE, J.M., A.M. KNIGHT, M.A. MCALEER AND J.A. TODD. (1990)
Towards construction of a high resolution map of the mouse genome
using PCR-analysed microsatellites *Nucl. Acids Res.* **18**: 4123-4130.

LUTY, J.A., Z. GUO, H.F. WILLARD, D.H. LEDBETTER AND M. LITT. (1990)
Five polymorphic VNTRs on the human X chromosome.
*Am. J. Hum. Genet.* **46**: 776-783.

MALCOLM, S., J. CLAYTON-SMITH, M. NICHOLS, S. ROBB, T. WEBB,
J.A.L. ARMOUR, A.J. JEFFREYS AND M.E. PEMBREY. (1991)
Uniparental paternal disomy in Angelman's syndrome.
*Lancet* **337**: 694-697.

MANIATIS, T., E.F. FRITSCH AND J. SAMBROOK. (1982)
Molecular cloning: A laboratory manual. 1st ed.
Cold Spring Harbor Laboratory Press, N.Y.

MCLAREN, A. Conference communication at 2nd Biennial Mammalian
Developmental Genetics Workshop (1991). Jackson Laboratory,
Bar Harbor, Maine.

MEISLER, M.H. (1976)
Effects of the *Bgs* locus on mouse β galctosidase
*Biochemical Genetics* **14**: 921-932.

MIESFELD, R., M. KRYSTAL AND N. ARNHEIM. (1981)
A member of a new repeated sequence family which is conserved
throughout eukaryotic evolution is found between the human δ and β-
globin genes. *Nucl. Acids Res.* **9**: 5931-5947.

MITANI, K., Y. TAKAHASHI AND R. KOMINAMI. (1990)
A GGCAGG motif in minisatellites affecting their germline instability.
*J. Biol. Chem.* **265**: 15203-15210.

MORRAL, N., V. NUNES, T. CASALS AND X. ESTIVILL. (1991)
CA/GT microsatellite alleles within the cystic fibrosis transmembrane
conductance regulator (CFTR) gene are not generated by unequal
crossingover. *Genomics* **10**: 692-698.

MURPHY, G.L., T.D. CONNELL, D.S. BARRITT, M. KOOMEY
AND J.G. CANNON. (1989)
Phase variation of gonococcal protein II: Regulation of gene expression
by slipped-strand mispairing of a repetitive DNA sequence.
*Cell* **56**: 539-547.

NAKAMURA, Y., M. LEPPERT, P. O'CONNELL, R. WOLFF, T. HOLM, M. CULVER,
C. MARTIN, E. FUJIMOTO, M. HOFF, E. KUMLIN AND R. WHITE. (1987a)
Variable number of tandem repeat (VNTR) markers for human gene
mapping. *Science* **235**: 1616-1622.

NAKAMURA, Y., C. JULIER, R. WOLFF, T. HOLM, P. O'CONNELL, M. LEPPERT
AND R. WHITE. (1987b)
Characterization of a human 'midisatellite' sequence.
*Nucl. Acids Res.* **15**: 2537-2547.

PAGE, D.C., K. BEIKER, L.E. BROWN, S. HINTON, M. LEPPERT, J.-M. LALONEL,
M. LATHROP, M. NYSTROM-LAHTI, A. DE LA CHAPELLE AND R. WHITE.
(1987)    Linkage, physical mapping and DNA sequence analysis of
pseudoautosomal loci on the human X and Y chromosomes.
*Genomics* **1**: 243-256.

PÄÄBO, S. (1985)
Molecular cloning of ancient Egyptian mummy DNA.
*Nature* **340**: 465-467.

PÄÄBO, S., J.A. GIFFORD AND WILSON A.C. (1988)
Mitochondrial DNA sequences from a 7000-year old brain.
*Nucl. Acids Res.* **16**: 9775-9787.

RICHARDS, R.I. AND G.R. SUTHERLAND. (1992)
Heritable unstable DNA sequences *Nature Genet.* **1**: 7-9.

ROGSTAD, S.H., J.C. PATTON II AND B.A. SCHAAL. (1988)
A human minisatellite probe reveals RFLPs among individuals of two
angiosperms. *Nucleic Acids Res.* **18**: 1081.

ROSENBERG, H., M. SINGER AND M. ROSENBERG. (1978)
Highly reiterated sequences of simiansimiansimiansimiansimian.
*Science* 200: 394-402.

ROSSANT, J. (1987)
Cell lineage analysis in mammalian embryogenesis
*Current topics in developmental biology* **23**: 115-146.

ROYLE, N.J., R.E. CLARKSON, Z. WONG AND A.J. JEFFREYS. (1987)
Clustering of hypervariable minisatellites in the proterminal regions of
human autosomes. *Genomics* 3: 352-360.

ROYLE, N.J., M.C. HILL AND A.J. JEFFREYS. (1992)
Isolation of telomere junction fragments by anchored polymerase chain
reaction. *Proc. R. Soc. Lond. B* **247**:57-61.

RUSSELL, R. AND A. DEENY. (1992)
Genetic minitoring using DNA fingerprinting.
*Mouse Genome* **90**: 79-80.

SAIKI, R.K., D.H. GELFAND, S. STOFFEL, S.J. SCHARF, R. HIGUCHI, G.T. HORN, K.B. MULIS AND H.A. ERLICH. (1988)
Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239**: 487-491.

SAITO, I AND G.R. STARK. (1986)
Charomid vectors for the efficient cloning and mapping of large or small restriction fragments. *Proc. Natl. Acad. Sci. USA* **83**: 8664-8668.

SAMBROOK, J., E.F. FRITSCH AND T. MANIATIS. (1989)
Molecular cloning: A laboratory manual. 2nd ed.
Cold Spring Harbor Laboratory Press, N.Y.

SANGER, F AND A.R. COULSON. (1977)
DNA Sequencing with chain terminating inhibitors.
*Proc. Natl. Acad. Sci. USA* **74**: 5463-5467.

SCHLÖTTERER, C. AND D. TAUTZ. (1992)
Slippage synthesis of simple sequence DNA
*Nucleic Acids Res.* **20**: 211-215.

SCHMID, C.W. AND W.R. JELINEK. (1982)
The Alu family of dispersed repetitive sequences.
*Science* **216**: 1065-1070.

SILVER, J. (1985)
Confidence limits for estimates of gene linkage based on analysis of recombinant inbred strains. *The Journal of Heredity* **76**: 436-440.

SMITH, G.P. (1976)
Evolution of repeated sequences by unequal crossing-over.
*Science* **191**: 528-535.

SNYDER, M. AND W.F. DOOLITTLE (1988)
P elements in Drosophila: selection at many levels.
*Trends Genet.* **4**:147-149

SOLARI, A.J. (1980)
Synaptonemal complexes and associated structures in microspread human spermatocytes. *Chromosoma* **81**: 307-314.

SOLLNER-WEBB, B. AND J. TOWER. (1986)
Transcription of eukaryotic ribosomal RNA genes.
*Ann. Rev. Biochem.* **55**: 801-830.

SOUTHERN, E.M. (1975)
Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* **98**: 503-517.

STARLING, J.A., J. MAULE, N.D. HASTIE AND R.C. ALLSHIRE. (1990)
Extensive telomere repeat arrays in mice are hypervariable
*Nucl. Acids Res.* **18**: 6881-6888.

STEINMETZ, M., D. STEPHAN AND K. FISCHER-LINDAHL. (1986)
Gene organization and recombinational hotspots in the murine major
histocompatability complex. *Cell* **44**: 895-904.

STEINMETZ, M., Y UEMATSU AND K. FISCHER-LINDAHL. (1987)
Hotspots of homologous recombination in mammalian genomes.
*Trends Genet.* **3**: 7-10.

STEPHAN, W. (1989)
Tandem-repetitive noncoding DNA: forms and forces.
*Mol. Biol. Evol.* **6**: 198-212.

STERNBERG, E.A., G. SPIZZ, W.M. PERRY, D. VIZARD, T. WEIL
AND E.N. OLSON. (1988)
Identification of upstream and intragenic regulatory elements that
confer cell-type-restricted and differentiation-specific expression on the
musclecreatine kinase gene. *Mol. Cell. Biol.* **8**: 2896-2909.

STOKER, N.G., K. CHEAH, J. GRIFFIN, F. POPE, AND E. SOLOMON. (1985)
A highly polymorphic region 3' to the human type II collagen gene.
*Nucl. Acids Res.* **13**: 4613-4622.

SUN, L., K.E. PAULSON, C.W. SCHMID, L. KADYK AND L. LEINWAND. (1984)
Non-alu family interspersed repeats in human DNA and their
transcriptional activity. *Nucl. Acids Res.* **12**: 2699-2690.

SUTHERLAND, A.E., T.P. SPEED AND P.G. CALARCO. (1990)
Inner Cell Allocation in the Mouse Morula: The Role of Oriented
Division during Fourth Cleavage. *Developmental Biology* **137**: 13-25.

SWALLOW, D.M., S. GENDLER, B. GRIFFITHS, G. CORNEY,
J. TAYLOR-PAPADIMITRIOU AND M.E. BRAMWELL. (1987)
The human tumour-associated epithelial mucins are encoded by an
expressed hypervariable gene locus PUM. *Nature* **328**: 82-84.

TABOR, S. AND C.C. RICHARDSON. (1987)
DNA sequence analysis with a modified bacteriophage T7 DNA
polymerase. *Proc. Natl. Acad. Sci. USA* **84**: 4767-4771.

TAUTZ, D., M. TRICK AND G.A. DOVER. (1986)
Cryptic simplicity in DNA is a major source of genetic variation.
*Nature* **322**: 652-656.

TAUTZ, D. (1989)
Hypervariability of simple sequences as a general source of polymorphic DNA markers. *Nucl. Acids Res.* **17**: 6463-6471.

TAYLOR, B. A. (1978)
Recombinant inbred strains: use in gene mapping. *In* "Origins of Inbred Mice", ed. Herbert C. Morse. pp423-438. Academic Press, New York. .

TAYLOR, B.A. (1989)
Recombinant inbred strains. *In* "Genetic variants and strains of the laboratory mouse", ed. M.F Lyon and A.G. Searle. 2nd ed. Oxford University Press, Oxford.

TAYLOR, K. D. AND L. PIKÓ. (1987)
Patterns of mRNA prevalence and expression of B1 and B2 transcipts in early mouse embryos. *Development* **101**: 877-892.

TELFORD, N.A., A.J. WATSON AND G.A. SCHULTZ. (1990)
Transistion from maternal to embryonic control in early mammalian development: a comparison of several species.
*Mol. Reprod. Devel.* **26**: 90-100.

THEIN, S.-L., A. J. JEFFREYS AND H. A. BLACKLOCK. (1986)
Identification of post-transplant cell population by DNA fingerprint analysis. *Lancet* **ii** 37.

THEIN, S.L., A.J. JEFFREYS, H.C. GOOI, F. COTTER, J. FLINT, N. O'CONNOR AND J.S. WAINSCOAT. (1987)
Detection of somatic changes in human cancer DNA by DNA fingerprint analysis. *Br. J. Cancer* **55**: 353-356.

TODD, J.A., T.J. AITMAN, R.J. CORNALL, S. GHOSH, J.R.S. HALL, C.M. HEARNE, A.M. KNIGHT, J.M. LOVE, M.A. MCALEER, J-B. PRINS, N. RODRIGUES, M. LATHROP, A. PRESSEY, N.H. DELARATO, L.B. PETERSON AND L.S WICKER. (1991)
Genetic analysis of autoimmune type 1 diabetes mellitus in mice. *Nature* **351**: 542-547.

TODD, J.A. (1992)
The benefits of using rodents to map and analyse human disease genes. *Gnome News* **9**:

TRECO, D., B. THOMAS AND N. ARNHEIM. (1985)
Recombination hotspots in the human β-globin gene cluster: meiotic recombination of human DNA fragments in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **5**: 2629-2638.

UEMATSU, Y., H. KIEFER, R. SCHULZE, K. FISCHER-LINDAHL
AND M. STEINMETZ. (1986)
Molecular characterization of a meiotic recombinational hotspot
enhancing homologous equal crossing-over. *EMBO. J.* **5**: 2123-2129.

ULLU, E. AND C. TSCHUDI. (1984)
Alu sequences are processed 7SL RNA genes. *Nature* **312**: 171-172.

VASSART, G., M. GEORGES, R. MONSIEUR, H. BROCAS, A.S. LEQUARRE AND
D. CHRISTOPHE. (1987)
A sequence in M13 phage detects hypervariable minisatellites in human
and animal DNA. *Science* **235**: 683-684.

VERGNAUD, G., D. MARIAT, F. APIOU, A. AURIAS, M. LATHROP AND
V. LAUTHIER. (1991)
The use of synthetic tandem repeats to isolate new VNTR loci: cloning
of a human hypermutable sequence. *Genomics* **11**: 135-144.

VISSEL, B. AND H.K. CHOO. (1989)
Mouse major ($\gamma$) satellite DNA is highly conserved and organized into
extremely long tandem arrays: implications for recombination between
non-homologous chromosomes. *Genomics* **5**: 407-414.

VOGEL, F. AND R. RATHENBERG. (1975)
Spontaneous mutation in man. *Adv. Hum. Genet.* **5**: 223-318.

VUORIO, E. AND B. DE CROMBRUGGE. (1990)
The family of collagen genes. *Ann. Rev. Biochem.* **59**: 837-872.

WAHLS, W.P., L.J. WALLACE AND P.D. MOORE. (1990)
Hypervariable minisatellite DNA is a hotspot for homologous
recombination in human cells. *Cell* **60**: 95-103.

WAHLS, W.P., G. SWENSON AND P.D. MOORE. (1991)
Two hypervariable minisatellite DNA binding proteins.
*Nucl. Acids Res.* **19**: 3269-3274.

WEBER, J.L. AND P.E. MAY. (1989)
Abundant class of human DNA polymorphisms which can be typed
using the Polymerase Chain Reaction *Am. J. Hum. Genet.* **44**: 388-396.

WEBER, J.L. (1990)
Informantiveness of (dC-dA)n.(dG-dT)n polymorphisms.
*Genomics* **7**: 524-530.

WELLS, R.A., G.G. GERMINO, S. KRISHNA, V.J. BUCKLE AND
S.T. REEDERS. (1990)
Telomere related sequences at interstitial sites in the human genome.
*Genomics* **8**: 699-704.

WETTON, J.H., R.E. CARTER, D.T. PARKIN AND D.T. WALTERS. (1987)
Demographic study of a wild house sparrow population by DNA
fingerprinting. *Nature* **327**: 147-149.

WILLARD, H.F. AND J.S. WAYE. (1987)
Hierarchical order in chromosome-specific alpha satellite DNA.
*Trends Genet.* **3**: 192-198.

WILLARD, H.F. (1990)
Centromeres of mammalian chromosomes. *Trends Genet.* **6**: 410-416.

WOLFF, R., Y. NAKAMURA AND R. WHITE. (1988)
Molecular characterization of a spontaneously generated new allele at a
VNTR locus: no exchange of flanking sequence. *Genomics* **3**: 347-351.

WOLFF, R.K., R. PLAETKE, A.J. JEFFREYS AND R. WHITE. (1989)
Unequal crossing-over between homologous chromosomes is not the
major mechanism involved in generation of new alleles at VNTR loci.
*Genomics* **5**: 382-384.

WOLFF, R., Y. NAKAMURA, S. ODELBERG, R SHIANG AND R. WHITE. (1991)
Generation of variability at VNTR loci in human DNA. *In* "DNA
fingerprinting: approaches and applications.", ed. T. T. Burke, G. Dolf,
A.J. Jeffreys, and R. Wolff. pp20-38. Birkhauser Verlag, Bern

WONG, Z., V. WILSON, A.J. JEFFREYS AND S.L. THEIN. (1986)
Cloning a selected fragment from a human DNA 'fingerprint': isolation
of an extremely polymorphic minisatellite.
*Nucl. Acids Res.* **14**: 4605-4616.

WONG, Z., V. WILSON, I. PATEL, S. POVEY AND A.J. JEFFREYS. (1987)
Characterization of a panel of highly variable minisatellites cloned from
human DNA. *Ann. Hum. Genet.* **51**: 269-288.

WONG, A.K.C. AND J.B. RATTNER. (1988)
Sequence organisation and cytological localisation of the minor satellite
of mouse. *Nucl. Acids Res.* **16**: 11645-11661.

WYMAN, A. AND R. WHITE. (1980)
A highly polymorphic locus in human DNA.
*Proc. Natl. Acad. Sci. USA* **77**: 6754-6758.

WYMAN, A.R., J. MULHOLLAND AND D. BOTSTEIN. (1986a)
Oligonucleotide repeats involved in the highly polymorphic locus
D14S1. *Am. J. Hum. Genet.* **39**: A226.

WYMAN, A.R., L.B. WOLFE AND D. BOTSTEIN. (1986b)
Propagation of some human DNA sequences in bacteriophage λ requires mutant *Escherischia coli* hosts. *Proc. Natl. Acad. Sci. USA* **77**: 2880-2884.

YANG, R. C-A., J. LIS AND R. WU. (1979)
Elution of DNA from agarose gels after electrophoresis. *Meth. Enzymol.* **68**: 176-182.

YANISCH-PERRON, C., J. VIERA AND J. MESSING. (1985)
Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33**: 103-119.