HYPERVARIABLE MINISATELLITES IN MOUSE DNA

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

by

Robert George Kelly BA (Trinity College, Dublin)
Department of Genetics
University of Leicester

February 1990

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

I would primarily like to thank Professor Alec Jeffreys for his continual help, encouragement, and interest in this work during my time in his laboratory, for all I have learnt during my stay, and for instilling in me his enthusiasm for molecular biology.

I would also like to thank all the members of Alec's labotatory for their support throughout this work; in particular Vicky Wilson and Ila Patel for maintaining order, Nicola Royle, John Armour, and Ian Gray for illuminating discussions, and Andy Collick and Mark Gibbs for assistance murine (and illuminating discussions). Outside G19, I gratefully acknowledge the help and advice of Drs Baljinder Mankoo, Justin Thackeray, and Ross Hawkins. Thanks also to the technical staff of the Genetics Department for material assistance, and Dr David Morton and his staff at Biomedical services for maintaining our mouse colony in excellent order. Outside Leicester, I am grateful to Dr Grahame Bulfield (AFRC, Roslin, Edinburgh) and his laboratory for many supplies of mice and mouse DNAs, Dr Peter Goodfellow (ICRF, London) for kindly supplying cell line DNAs, and Dr Ben Taylor (The Jackson Laboratory, Maine, USA) for supplying mouse DNAs and kindly analysing data from recombinant inbred strains of mice at many stages during this study.

Outside science, I am indebted to many people for making Leicester such an interesting place; BJ, Justin, Ross, Serge, Steve, Gary, Bren and Vidya among others, also the Phoenix, the Magazine players, and the cartographers of the Leicestershire 1:25,000 series. And many more. I would especially like to thank my parents for all their support, and encouraging the taking of welcome draughts of Cumbrian air. And far from least, for inspiration at a distance, E.

I acknowledge the receipt of a British Council scholarship from October 1986 - December 1989.

ABBREVIATIONS

A (dATP)	2'-deoxyadenosine 5'-triphosphate
BCIG	5-bromo-4-chloro-3-indoyl-β-D-galactopyranoside
bp	base pair
BSA	bovine serum albumin
BXD	C57BL/6J x DBA/2J (RI) strains
BXH	C57BL/6J x C3H/HeJ (RI) strains
C (dCTP)	2'-deoxycytidine 5'-triphosphate
Ci	Curie
cM	centiMorgan
ddNTP	2',3'-dideoxynucleoside 5'-triphosphate
dH_2O	distilled water
DMSO	dimethyl sulphoxide
DNA	deoxyribonucleic acid
dNTP	2'-deoxynucleoside 5'-triphosphate
dpm	disintegrations per minute
dr	direct repeat
G (dGTP)	2'-deoxyguanosine 5'-triphosphate
HEPES	N-2-hydroxyethylpiperazine-N'-2-ethanesulphonic acid
hr	hour
IPTG	isopropyl-β-D-galactopyranoside
kb	kilobase pair
LINE	long interspersed repeated element
LTR	long terminal repeat
min	minute
Mkb	megakilobase pair (10^9 bp)
MOPS	3-(N-morpholino)propanesulphonic acid
MYA	million years ago
N	A, C, T or G
nt	nucleotide
O.D.	optical density
O/N	overnight
PEG	polyethylene glycol
R	purine (A or G)
RI	recombinant inbred
RNA	ribonucleic acid
RPII	RNA polymerase II
RPIII	RNA polymerase III
RT	room temperature
SDS	sodium dodecyl sulphate
SINE	short interspersed repeated element
SSC	saline sodium citrate
SWXL	SWR/J x C57/L (RI) strains
T (dTTP)	2'-deoxythymidine 5'-triphosphate
TEMED	N,N,N',N'-tetramethylethylenediamine
Tris	2-amino-2(hydroxymethyl)propane-1,3-diol
Y	pyrimidine (C or T)

Some of this work has been published :

Jeffreys, A.J., Wilson, V., Kelly, R., Taylor, B.A., and Bulfield, G. 1987. Mouse DNA 'fingerprints' : analysis of chromosome localisation and germ-line stability of hypervariable loci in recombinant inbred strains. *Nucleic Acids Res.* **15**:2823-2836.

Kelly, R., Bulfield, G., Collick, A., Gibbs, M., and Jeffreys, A.J. 1989. Characterisation of a highly unstable mouse minisatellite locus ; evidence for somatic mutation during early development. *Genomics* **5**:844-857.

To my parents

I. INTRODUCTION

1.1 Mammalian genome organisation

The eukaryotic genome is a dynamic structure composed of unique and repetitive DNA sequences. It is important to investigate the organisation of these components of the genome in order to understand the contemporary structure of eukaryotic chromosomes, the processes by which this structure has evolved, and the genomic aberrations associated with the many genetic disorders of man.

DNA reassociation studies have shown that the repetitive fraction of the genome is represented in highly and moderately repeated components, the former frequently associated with heterochromatin (satellite DNA), and the latter often interspersed with single copy sequences in the genome. Subsequent molecular studies have uncovered a complex spectrum of repeated DNA sequences, such that some of the dispersed sequences are as highly repeated as the heterochromatic sequences. This study will review the complex organisation of both satellite and dispersed repeated sequences (principally in mammals), and then concentrate on a component of the moderately repetitive dispersed complement, tandemly repeated minisatellite sequences, examining the genomic organisation, hypervariability, and applications of minisatellite loci in man and mouse.

1.1.1 Highly repeated satellite DNA

Extraordinary variation in the chromosomal DNA content of related organisms (the C-paradox) has been observed for many years (Mirsky and Ris, 1951); this suggests that there is an additional, variable, component of the genome beyond that which is necessary to encode the functions required for the development and structure of an organism. Highly repeated satellite DNA may constitute such a component, and comprises an extremely heterogeneous class of DNA sequences (see Singer, 1982a).

Many satellite DNA sequences have a different guanine and cytosine content to the remainder of the genome and were initially observed as a distinct 'satellite' peak on isopycnic centrifugation of genomic DNA (Sueoka, 1961, Kit, 1961). Other (cryptic) satellite sequences are masked in the main genomic band. Satellite DNA is organised into long tandemly repeated arrays of a unit sequence, and comprises from 5% (human) to over 50% (kangaroo rat, *Dipodomys ordii*) of genomic DNA. The repeat units vary from

simple sequence to complex units with internal and higher-order repeat structures. Cytological studies have shown that satellite DNA sequences are predominantly associated with centromeric heterochromatin (Pardue and Gall, 1970).

The organisation of the major satellite in the mouse (constituting approximately 10% of the genome) was elucidated by Southern (1975a) using restriction endonucleases which cleave most repeat units (such as *Eco*RII). The monomer has been shown by sequencing to be 234bp (Horz and Altenburger, 1981). The observation of fractional multiples of 234bp with *Eco*RII suggests that the satellite has evolved by amplification and divergence from smaller ancestral repeat units, perhaps through processes of unequal exchange (Smith, 1976) and slippage (Levinson and Gutman, 1987, see section 1.3). The mouse major satellite is found in uninterrupted arrays from 240kb to over 2000kb in length (Vissel and Choo, 1989) within the centromeric heterochromatin of all chromosomes except the Y (Pardue and Gall, 1970). The sequence related minor satellite of the mouse is also centromerically located (Pietras *et al.*, 1983).

The main primate satellite DNA, or α-satellite, is located at the centromeres of all human chromosomes (Rosenberg *et al.*, 1978), where it is organised as distinct chromosome-specific subsets of diverse arrays of a 171bp monomer (see Willard and Waye, 1987). Individual subsets are characterised by long range periodicities defining units of amplification which have been analysed from several different chromosomes. The length of these centromeric arrays is highly variable (Waye *et al.*, 1987, Tyler-Smith and Brown, 1987), and further polymorphism within arrays provides evidence for unequal recombination events between higher-order repeats (Waye and Willard, 1986). This variation suggests that satellite DNA is a rapidly evolving component of the genome, consistent with the divergent organisation of satellite DNA sequences within the genus *Mus* (see Hastie, 1989).

The diversity of eukaryotic satellite DNAs obscures any functional role such sequences may have. Conceivably centromeric repeated DNA may be involved in either homologue interactions or interactions between chromosomes and nuclear structures. The heterogeneity and rapid evolution of centromeric satellite arrays argues against any sequence-specific role. It has been proposed that much of the repetitive component of the genome may be 'junk' DNA, diverging and amplifying under little or no selection through processes of unequal exchange and slippage (Ohno, 1970). While this may explain the redundancy and divergence of satellite DNA it is not true in every case : a centromeric satellite repeat of *Drosophila melanogaster* appears to be functional as the *responder* element of *segregation distorter*, causing dysfunction of the satellite bearing sperm, although no mechanism for this association has yet been elucidated (Wu *et al.*,

1988). Thus, while the importance of centromeric arrays remains functionally unassessed, a role for these sequences cannot be ruled out.

1.1.2 Telomeric repeats

Tandem arrays of heterogeneous length are found at the telomeres of eukaryotic chromosomes (Szostak and Blackburn, 1982), and may be structurally compared to the satellite component of the genome. In contrast to centromeric arrays, telomeric tandem repeat units are short (generally <10bp), and are of a similar sequence in many eukaryotes (Allshire *et al.*, 1988). In man and mouse the principal repeat unit sequence is TTAGGG (Moyzis *et al.*, 1988, see Hastie and Allshire, 1989) which diverges among proximal repeats (Allshire *et al.*, 1989); this sequence constitutes the major satellite of the guinea pig (*Cavia porcella*) (Southern, 1970) and the kangaroo rat (Fry and Salser, 1977). In addition, interstitial blocks of TTAGGG constitute a highly variable fraction of dispersed repetitive DNA sequences in man. The most distal telomeric sequences vary in length from 10-15kb in man (see Hastie and Allshire, 1989), and are of unknown size in mice. The evolution and detailed comparative organisation of telomeric repeat arrays remains to be elucidated.

In contrast to centromeric satellite DNA, telomeric repeats are critical determinants of chromosomal integrity (see Hastie and Allshire, 1989). In association with a ribonucleoprotein telomerase activity (Greider and Blackburn, 1989, Morin, 1989, Shippen-Lentz and Blackburn, 1990), telomeric repeats are responsible for maintaining chromosome length at replication and preventing inter-chromosomal fusions. Furthermore, human telomeres have been shown to be functional in yeast cells (Cross *et al.*, 1989, Brown, 1989).

1.1.3 Dispersed repeated DNA

Studies of genomic organisation using DNA renaturation techniques, hydroxylapatite binding methods, and DNA hyperchromicity led to the observation that much of the human genome consists of single copy DNA interspersed with repetitive sequences, many of which are present as inverted repeats (Schmid and Deininger, 1975). Isolation of total human repetitive DNA, after reannealing and S1 nuclease treatment, revealed that most of the interspersed repetitive sequences in man belonged to a single family (see Schmid and Jelinek, 1982). Subsequent molecular approaches have discovered most of the highly repeated sequence families in the human and mouse genomes (Sun *et al.*, 1984, Bennet *et al.*, 1984). These include short and long interspersed repetitive elements (SINES and LINES, Singer, 1982b), which may be either

mobile in the genome (retroposons, see Rogers, 1985), or independently amplified tandem repeats of common unit sequences (Tautz and Renz, 1984). In addition to the most repeated families of mammalian genomes there are many less abundant interspersed repeated sequences, including a large number of retroposon and processed pseudogene families (see Rogers, 1985), clustered and dispersed multi-gene families (D'Eustachio and Ruddle, 1983), endogenous retroviral and retroviral-like elements (see Finnegan, 1989) and cryptic and low copy number tandemly repeated sequences (Tautz *et al.*, 1986) - illustrating the complex and dynamic organisation of mammalian euchromatin.

1.1.3.a Short interspersed retroposons

The archetypal short interspersed retroposon is the Alu sequence (see Schmid and Jelinek, 1982). The human genome contains up to 900,000 members of this family. Alu sequences are 300bp long, and composed of two imperfectly repeated units separated by an (A)-rich sequence. Alu sequences are the predominant SINE in all primate genomes, and are related to SINE families of many mammalian species (see Rogers, 1985).

The structure of Alu elements suggests that they are dispersed in the genome by replication through an RNA intermediate (see Sharp, 1983). Most Alu members which have been isolated are flanked by direct repeats of 7-20 nucleotides. These are hallmarks of integrated sequences, resulting from staggered nicks at the site of insertion. A run of deoxyadenosine residues typically precedes the 3' direct repeat, in an equivalent position to the poly(A) tail of an RNA transcipt. The retroposition intermediate is likely to be an RNA polymerase III transcript, as Alu elements contain an internal RPIII promoter. There is no evidence for specific *in vivo* transcription of Alu sequences by RPIII, although many are co-transcribed by flanking promoters. It is thought instead that almost all Alu sequences may be pseudogenes generated as by-products of RPIII transcription of one or a small number of source genes (see Deininger and Daniels, 1986, Britten *et al*, 1988). Source gene transcription must occur in the germline or early embryo to allow the inheritance of newly integrated elements. Changes in the source gene over evolution are reflected in the discontinuous sequence classes of Alu sequences in the contemporary genome (see Britten *et al.*, 1989).

Alu units are very similar in sequence to the 7SL RNA component of the signal recognition particle (Ullu and Tschudi, 1984). It is possible that the source gene encodes an as yet unidentified 7SL-related RNA component of a ribonucleoprotein particle. The conservation of such a progenitor source gene in different species would explain the presence of Alu-related SINE families in a wide range of species.

In the mouse there are two principal SINE families, B1 (Krayev *et al.*, 1980) and B2 (Krayev *et al.*, 1982). B1 elements are homologous to a single Alu unit, and are about 130bp in length. B1 sequences are the most abundant mouse SINE (130,000-180,000 copies) constituting about 1% of the genome. B2 elements are approximately 190bp long and are repeated 80,000-100,000 times in the mouse genome. B2 elements share the RPIII promoter boxes and other retroposon features with Alu-type sequences, but are otherwise unrelated. B2 sequences are rodent-specific, and individual elements are 8% divergent from the consensus sequence (whereas Alu-type retroposons are 13% divergent, Rogers, 1985). Rogers (1985) suggested that B2 transcripts might form a tRNA-like secondary structure; Sakamoto and Okada (1985) proposed that the B2 source gene may have in part evolved from a lysine tRNA gene. It is likely that all SINE families have small RNA gene progenitors which have by chance evolved into prolific source genes.

Other less abundant SINE families exist in the human and mouse genomes. MT (Mouse Transcript) elements are a rodent-specific SINE family which were first detected in mouse DNA, and are highly represented in cerebellar mRNA (Heinlein *et al.*, 1986, Bastien and Bourgaux, 1987). Although MT elements are poorly characterised, Heinlein *et al.* (1986) estimate that there are 40-90,000 MT sequences (each of 400bp) in the mouse genome. Other interspersed repeated families may also be selectively associated with brain transcripts (e.g., Anzai *et al.*, 1986); however, the discovery that the 'identifier' sequence, initially found in rat brain transcripts (Sutcliffe *et al.*, 1982), is in fact found with similar abundance in the RNA of other tissues (Owens *et al.*, 1985), advises caution in assigning significance to such observations. It is likely that many more low copy number retroposon families remain undiscovered (Sun *et al.*, 1984).

1.1.3.b Long interspersed retroposons

The predominant LINEs of mammalian genomes are all related to the L1 family (see Singer and Skowronski, 1985, and Rogers, 1986). The *Kpn*I elements of primates (Shafit-Zagardo *et al.*, 1982) and *Bam*HI elements of mice (L1Md, Voliva *et al.*, 1984) are 6-7kb long, although most copies are randomly truncated 3' fragments. Thus the 5' sequences are repeated less frequently in the genome (about 10,000 copies in the mouse) than the 3' sequences (100,000 copies). As retroposons, L1 elements are flanked by direct repeats and have a poly(A) tail at the 3' end. L1 elements may be dispersed in the genome via an RPII transcript; a central region has 60% homology between mouse and human LINES. Loeb *et al.* (1986) isolated a full length (6.85kb) L1Md element and found that this central region contained two overlapping open reading frames, one of which showed similarity to viral reverse transcriptase domains. This has important

implications for the amplification of LINES, as full length elements may encode all the functions neccessary for their own transposition. These workers also noted an array of tandem repeats at the 5' end which may carry internal promoters and be important in the maintainance of 5' sequences.

1.1.3.c Significance of retroposons

Many roles have been proposed for middle repetitive DNA, including involvement in gene expression (Britten and Davidson, 1969), transcript processing (Tchurikov *et al.*, 1982), and replication (Krayev *et al.*, 1982). It may rather be that many retroposon families are non-functional, and are dispersed as 'selfish' (i.e., sequence dependent, Orgel and Crick, 1980, Doolittle and Sapienza, 1980), or 'ignorant' (i.e., sequence independent, Dover, 1980), components of the genome. It is hard to reconcile the variety of retroposon families in different species with conserved function. There is no evidence for activity of individual Alu or L1 elements; Alu elements are eroded by mutation at an equal rate to non-coding flanking sequences (see Deininger and Daniels, 1986, Britten *et al.*, 1988), and the majority of L1 elements are defective through truncation. In some cases (e.g., Alu) it is likely that most or all of the repeats are ignorant by-products of a functionally important gene (Britten *et al.*, 1989); alternatively (e.g., LINES) interspersed repeat sequences may be the successful and partial amplification products of a selfish gene (see Rogers, 1986).

Apart from any direct functional significance, retroposons are undoubtedly very important in shaping the contemporary structure and continual rearrangement of the mammalian genome. In man Alu-Alu recombination events are associated with both deletions and duplications within the LDL receptor gene causing familial hypercholesterolaemia (Lehrman *et al.*, 1986, 1987), and with deletions within the α-globin cluster causing haemoglobinopathies (Nichols *et al.*, 1987). The human growth hormone and chorionic somatommamotropin gene cluster may have evolved from a single ancestral gene by recombination between flanking Alu elements (Barsh *et al.*, 1983). Similarly, recombination between flanking L1 elements may be responsible for duplication of the γ-globin gene cluster in old world monkeys (Maeda and Smithies, 1986). These examples illustrate the extent of such duplication and deletion events in the genome, and the importance of this variation as both a positive and negative influence in genome evolution.

Continuing integration is a further source of retroposon-derived variation in contemporary genomes. Independent *de novo* L1 insertions are responsible for two cases of haemophilia A in man (Antonarakis *et al.*, 1988). In mice, both B1 (King *et al.*, 1986)

and B2 (Kominami *et al.*, 1983a, Kress *et al.*, 1984) elements have been shown to be polymorphic among contemporary alleles, in one case possibly influencing the expression of an MHC gene at the site of integration (Kress *et al.*, 1984). Rogers (1985) has noted that the insertion sites of retroposons are often (A)-rich, and that retroposons frequently insert into pre-existing retroposon tails. Thus clusters of retroposons arise in the genome (see Rogers, 1985), with implications for the evolution of such sequences (for example, novel combinations of retroposons may be generated). Furthermore, retroposon clusters are often associated with simple tandem repeat sequences (Gebhard and Zachau, 1983, Kominami *et al.*, 1983b), suggesting that particular regions of the genome may be prone to the accumulation of dispersed repetitive elements. Retroposon integration events must occur either in the germline or during early development to be heritable. Interestingly the expression of both B2 elements (Murphy *et al.*, 1983, Bennet *et al.*, 1984, Vasseur *et al.*, 1985), and L1 repeats (Skowronski and Singer, 1985), has been observed in cells of the early embryo (see 1.1.3.e).

As the long range organisation of the mammalian genome emerges, it is becoming apparent that different classes of retroposons may be dispersed differently in the genome (see Bickmore and Sumner, 1989). Korenberg and Rykowski (1988) used high resolution *in situ* hybridisation and solid state quantitative imaging to show that Alu and L1 elements tend to be inversely distributed in the human genome. Whereas Alu sequences appear to predominate within reverse geimsa staining chromosomal bands, L1 sequences tend to predominate in geimsa positive bands. Reverse bands are where CpG islands are found (see Bird, 1987), and contain most, if not all, genes (see Bickmore *et al.*, 1989). The significance of such sequence partitioning for the process of retroposition and for the organisation and evolution of mammalian chromosomes remains unclear.

1.1.3.d Retroviral-like elements

The most complex class of mobile repeated sequences in the mammalian genome are the retroviral-like elements. These elements are bounded by long terminal direct repeats (LTRs) and contain open reading frames encoding functions required for their transposition, including reverse transcriptase. Retroviral-like elements are thought to transpose via an RNA intermediate (see Finnegan, 1985); on integration these elements are flanked by short direct repeats (4-6bp) of target-site sequence. In addition to the endogenous retroviral elements found in the mouse genome (e.g., Jenkins *et al.*, 1981), several murine retrovirus-like elements have been described. These are found as proviral-like genomes (which may in some cases have internal deletions) or as solo LTRs, resulting from precise excision of the internal sequences by recombination between

flanking LTRs. The major murine retroviral-like elements are MuRRS (7.5kb, 100 copies (LTR, 1000 copies), Schmidt *et al.*, 1985), VL30 (5.5kb, Hodgson *et al.*, 1990, >1100 copies, Rotman *et al.*, 1984), ETn-like (>6kb, 1000-3000 copies, Brulet *et al.*, 1983), and IAP (3.5-7kb, 1000 copies, Leuders and Kuff, 1977). In man a transposon-like element has been identified (THE family, Paulson *et al.*, 1985); the associated 350bp LTR (O-LTR) is also found as a solitary repeat (Sun *et al.*, 1984).

There is good evidence for the mobility of these elements. An increasingly large number of mutations in the mouse are known to be due to retroviral and retroviral-like elements. For example, an endogenous provirus at the *dilute* locus on chromosome 9 is responsible for the light hair pigmentation characteristic of *dd* mice (Jenkins *et al.*, 1981, Copeland *et al.*, 1983). Experimental infection of the germline and early embryo with murine leukaemia viruses generates insertional mutations of value in developmental studies (e.g., Jaenisch *et al.*, 1983). The arrangement of LTR-IS sequences in the genome is polymorphic (Wirth *et al.*, 1984), implying past activity of MuRRs elements. In several cases the integration of an IAP element has been associated with altered gene expression (Kuff *et al.*, 1982, Rechavi *et al.*, 1982, Burt *et al.*, 1984). As with retroposons it is not known when these elements transpose, however IAP and ETn-like transcripts are found in early embryos, where the former are associated with non-infectious virus-like particles (Piko *et al.*, 1984, Brulet *et al.*, 1983). The early developmental expression of IAP elements is regulated autonomously, rather than by flanking sequences (Howe and Overton, 1986).

1.1.3.e Processed pseudogene families

Mammalian genomes contain many families of non-functional gene sequences. Such pseudogenes may be found within gene clusters, presumably arising through duplication and divergence associated with loss of function (Proudfoot *et al.*, 1982), or dispersed in the genome as 'processed' pseudogenes (see Vanin, 1984). Processed pseudogenes lack introns and generally possess the structural features of integrated retroposons, including 9-14bp direct repeats (Vanin, 1984). The most abundant families are those of small nuclear RNA pseudogenes (100-1000 copies/family, see Rogers, 1985). Other families include ribosomal protein pseudogenes (Wagner and Perry, 1985), and pseudogenes from a variety of structural genes (generally 'housekeeping' genes) with a typical copy number of 1-5 (see Rogers, 1985). Most processed pseudogenes are non-functional as the promoter is lost during transposition (in the case of RPII transcripts), however occasionally flanking sequences may give rise to novel patterns of expression (McCarrey and Thomas, 1987). Thus while the majority of processed pseudogenes

appear to be 'ignorant' by-products of transcription, they may still contribute to evolution within the genome.

Processed pseudogenes, and both short and long interspersed retroposons therefore appear to be dispersed by similar mechanisms (see Rogers, 1985). It is interesting that only in mammals have retroposons become the predominant dispersed repetitive sequences (other species tend to have low copy number SINE families); similarly processed pseudogenes are rarely found outside the class. The expression of L1 and B2 sequences in the early embryo, and the observation that only in mammals does the embryonic genome become transcriptionally active at the two cell stage (see Davidson, 1969), suggests that the cleavage stage embryo could be where new retroposons are made (Rogers, 1986). Furthermore, the parallel phylogenetic distribution of endogenous retroviruses (see Wagner, 1986), and the expression of retrovirus-like elements in the early embryo, suggests that these elements (which are also thought to be dispersed through a RNA intermediate) might mediate retroposition in the mammalian genome, perhaps providing a source of reverse transcriptase activity.

1.1.3.f Dispersed tandem repeats

Not all interspersed repeats in the genome are mobile. Dispersed tandem repeats of simple sequences, which have presumably arisen independently at different loci, are ubiquitious components of eukaryotic genomes (Tautz and Renz, 1984). These 'microsatellite' sequences are commonly di- or tri-nucleotide repeats (e.g., $(CA)_n$ (Hamada *et al.*, 1984) and $(AT)_n$ (Greaves and Patient, 1985)). In the mammalian genome there may be as many as 500,000 $(CA)_n$ arrays each of 20-60bp (Hamada *et al.*, 1984).

The initial isolation of $(CA)_n$ and other simple sequence repeats adjacent to genes and retroposons, and their frequent co-transcription by flanking promoters, invoked many theories as to the functional importance of such sequences (see Rogers, 1983). There are precedents for an involvement of simple repeats in illegitimate recombination (Stringer, 1982, Hasson *et al.*, 1984), gene conversion (Slightom *et al.*, 1980), and deletion and duplication events (Hellman *et al.*, 1988). Such sequences can assume unusual DNA conformations *in vitro* (Neidle, 1983, Wells, 1986), and Gilmour *et al.* (1989) found that S1 sensitive sites upstream of several *Drosophila* genes identify regions of alternating C and T residues which bind nuclear proteins and may mediate an effect on transcription. More generally such sequences are likely to be functionless, as illustrated by Tautz and Renz (1984) who showed that simple repeats were lost from the nucleus of the protozoan *Stylonchia* during chromosome diminution.

Simple quadruplet repeats, including $(GATA)_n$ and $(GACA)_n$ sequences have been described in vertebrates and invertebrates (see Epplen, 1988). GATA repeats were originally isolated as a satellite DNA fraction in DNA from the banded krait (*Bungarus fasciatus*, Singh *et al.*, 1980). The chromosomal distribution of GATA repeats in a range of eukaryotic species prompted claims that these sequences were conserved due to their role in sex-determination (Singh *et al.*, 1984). In the mouse large GATA and GACA arrays are present on the Y chromosome, and shorter arrays are dispersed over other mouse chromosomes (Schafer *et al.*, 1986. More detailed investigation of the organisation of $(GATA)_n$ repeats and flanking DNA sequences in different species led to the more prosaic conclusion that such arrays have arisen independently (Levinson *et al.*, 1985). Furthermore, other species apparently lack large GAYA repeat arrays (see John and Miklos, 1988).

Short simple sequence arrays are thought to have expanded through slippage mechanisms (Levinson and Gutman, 1987, see section 1.3.3). As such they have been shown to be hypervariable; several studies have used the polymerase chain reaction to reveal sometimes extensive repeat copy number polymorphism within these arrays (Weber and May, 1989, Litt and Luy, 1989, Tautz, 1989). This variability, coupled with the abundance of simple repeat loci in the genome, makes them extremely promising genetic markers.

1.2 Hypervariable minisatellites

Eukaryotic genomes appear to harbour a continuum of dispersed tandem repeat types with respect to unit length and sequence. Arrays based on a spectrum of more complex repeat units may continualy be expanding and contracting at many loci through processes of slippage and unequal exchange. Such loci comprise the hypervariable minisatellite component of the genone.

1.2.1 Historical

The first multiallelic locus identified in human DNA was discovered on screening a random DNA library for long singe copy sequences (Wyman and White, 1980). More than 80 alleles have since been resolved at this locus (D14S1) in the range 3.7-26kb (Balazs *et al.*, 1986), and the polymorphism arises due to a variable number of $^{T}_{C}GG$ repeats (Wyman *et al.*, 1986).

Other hypervariable regions (HVRs) in man have been discovered by chance near cloned genes, including HVRs upstream of the insulin gene (Bell *et al.*, 1982), within the α-globin gene cluster (Proudfoot *et d.*, 1982, Jarman *et al.*, 1986), downstream of the c-

Ha-ras oncogene (Capon *et al.*, 1983, 5' to J_H in the human immunoglobulin heavy chain gene (Silva *et al.*, 1987), and within the human factor VII gene (O'Hara and Grant, 1988). While these loci all contain G-rich repeat units, other minisatellites have been discovered which have A-rich repeat units, including HVRs 3' to the human Type II collagen and apolipoprotein B genes (Stoker *et al.*, 1985, Knott *et al.*, 1986), and within the pseudoautosomal region of the human sex chromosomes (Simmler *et al.*, 1987). An HVR has also been reported within a coding sequence, responsible for a correspondingly polymorphic epithelial mucin-type glycoprotein (Swallow *et al.*, 1987). In addition to short tandem arrays, considerably larger loci have been described; for example, a 250-500kb 'midisatellite' array has been described on chromosome 1, apparently consisting entirely of 40bp repeat units (Nakamura *et al.*, 1987a). Many HVRs have also been cloned by screening genomic libraries for single-copy sequences (Knowlton *et al.*, 1986), and by hybridisation with oligomeric probes (Nakamura *et al.*, 1987b).

It is apparent from these studies that are a large number of minisatellites in the human genome. The molecular basis for variation at each locus is allelic polymorphism in the number of oligonucleotide repeat units. Thus HVRs (or VNTR loci, variable number of tandem repeats (Nakamura *et al.*, 1987)) define many highly polymorphic DNA markers throughout the genome whichare of value for the genetic analysis of man.

1.2.2 The importance of informative genetic markers in man

Investigation of the linear and functional organisation of the human genome through linkage analysis is dependent on the availability of polymorphic markers. These are used to distinguish and follow the segregation of specific regions of homologous chromosomes. Classical genetic markers in man include the ABO blood groups (Landsteiner, 1900), and other cell marker systems (e.g., histocompatibility antigens); further genetic polymorphism has been detected at the protein level, through the electrophoretic separation of variant aleles of enzymes and structural proteins (Harris and Hopkinson, 1972). In the last ten years, however, analysis of variation in the primary structure of DNA has revolutionised human genetic analysis.

Jeffreys (1979) demonstrated that nucleotide sequence variations occurred on average once in every hundred base pairs. Nucleotide changes which destroy or create endonuclease target sites can be detected as restriction fragment length polymorphisms (RFLPs) using unique sequence probes. Kan and Dozy (1978) showed that a polymorphic *Hpa*I site linked to the β-globin locus could be used to indirectly diagnose sickle cell anaemia. Botstein *et al.* (1980) argued that it would be possible not only to use such polymorphic sites as diagnostic markers for inherited disorders, but also to construct

linkage maps of entire human chromosomes which would enable the localisation of all disease loci associated with segregating genetic disorders (McKusick, 1986).

Since then RFLPs have contributed enormously to human molecular genetics, and enabled the characterisation of many otherwise intractable inherited disorders of man, including Duchenne muscular dystrophy (Davies *et al.*, 1983, Monaco *et al.*, 1986), retinoblastoma (Cavenee *et al.*, 1983, Friend *et al.*, 1986), cystic fibrosis (Tsui *et al.*, 1985, Rommens *et al.*, 1989), and autosomal retinitis pigmentosa (McWilliam *et al.*, 1989, Dryja *et al.*, 1990). Linkage has been found between RFLPs and many other monogenetic disorders (such as Huntington's Chorea (Gusella *et al.*, 1983)), and possibly to major genes associated with complex polygenic disorders, such as schizophrenia (Sherrington *et al.*, 1988). In these cases more tightly linked markers can be developed as an approach towards the isolation and characterisation of the disease locus. RFLPs have enabled the establishment of detailed linkage maps of human chromosomes using large pedigrees (Donis-Keller *et al.*, 1987); such maps will be invaluable aids in the elucidation of the structure and organisation of the human genome. Furthermore, studies of the genetic distribution of RFLPs in different populations can reveal how contemporary genome organisation has evolved (Higgs *et al.*, 1986).

Despite their continuing success most RFLPs are not ideal linkage markers. The usefulness of a genetic marker depends on the likelihood that a particular family will be informative for a locus with a given heterozygosity, and is related to the frequency of alleles at that locus (Botstein *et al.*, 1980). For a diallelic RFLP heterozygosity cannot exceed 0.5. Furthermore, the overall nucleotide variation in the human genome is low (Jeffreys, 1979), and extensive screening may be required to find an RFLP for a given probe. This can be overcome to some extent by using enzymes which recognise the mutable CpG doublet in their target site (Litt and White, 1985), and by detecting nucleotide variation using other techniques, such as denaturing gradient gels (Myers *et al.*, 1985), and mismatch sensitive endonucleases (Gibbs and Caskey, 1987). However the ideal highly informative marker must have many alleles and be heterozygous in a large proportion of individuals. Hypervariable tandem repeat loci, including dinucleotide arrays and minisatellites, constitute such a class of DNA marker, and have contributed significantly to both the development of human genetic linkage maps (Donis-Keller *et al.*, 1987) and the localisation of genetic disorders in man (e.g., adult polycystic kidney disease, Reeders *et al.*, 1986).

1.2.3 Human DNA fingerprints

There are many related subsets of loci within the spectrum of minisatellite sequences in man. Under low stringency hybridisation conditions it is possible to detect distinct 'families' of sequence-related loci in genomic DNA. Hypervariable DNA fingerprints were first detected in human DNA by Jeffreys *et al.* (1985a).

1.2.3.a Multilocus probes

A monomorphic G-rich tandem repeat region lies within the first intron of the human myoglobin gene (Weller *et al.*, 1984). Using a hybridisation probe based on this 33bp repeat Jeffreys *et al.* (1985a) screened a human genomic library, and isolated eight cross-hybridising loci. These were all minisatellites composed of 16-64bp repeat units repeated from 3 to 29 times. Comparison of the repeat units from each locus revealed that they all contained a 10-15bp G-rich 'core' sequence, GGAGGTGGGCAGGARG. Two of these clones, 33.6 and 33.15, cross-hybridised at low stringency to many minisatellite fragments in human genomic DNA, generating highly variable and individual specific patterns, or DNA fingerprints. These cross-hybridising fragments were shown to segregate in a Mendelian fashion and were found to be somatically stable (Jeffreys *et al.*, 1985b).

Minisatellite probes 33.6 and 33.15 detect non-overlapping sets of loci in human DNA. DNA fingerprints are also generated by 'multi-locus' probes which appear to be unrelated to the G-rich 'core' sequence. These include the α-globin 3'-HVR (Jarman *et al.*, 1986), probes based on the telomeric repeat TTAGGG (see Hastie and Allshire, 1989), an oligonucleotide repeat within M13 (Vassart *et al.*, 1987), a mouse clone which cross-hybridises to a tandem repeat sequence from the *Drosophila per* locus (Georges *et al.*, 1987), and several simple sequence tandem repeat probes (Ali *et al.*, 1986, Epplen, 1988). While the minisatellite loci detected in these systems may overlap to some extent, Vergnaud (1989) has shown that a variety of random sequence oligonucleotide repeat probes can, with varying efficiency, detect distinct sets of minisatellites in human DNA at low stringency, illustrating the unprecedented abundance and sequence range of tandem repeat loci in the human genome.

1.2.3.b DNA fingerprints : applications

Jeffreys *et al.* (1985b) demonstrated that probes 33.6 and 33.15 each detect about 15 resolvable bands within the size range 4-20kb in an individual's DNA, and that the mean probability (s) of a fragment being present in the DNA fingerprint of a second

unrelated individual is approximately 0.2. As these fragments assort independently the probability of all bands being shared between two individuals is vanishingly small, even if they are related (thus elevating the value of s), with the exception of identical twins.

This individual specificity can be applied to problems of human identification. The advantage of DNA fingerprinting over conventional methods of identity testing is that it enables positive identification, rather than relying on exclusion (Jeffreys *et al.*, 1985b). DNA fingerprinting is therefore an extremely valuable technique in the resolution of paternity and immigration disputes (Jeffreys *et al.*, 1985c), the matching of biological samples to suspects in forensic work (Gill *et al.*, 1985, Wambaugh, 1989), and a variety of medical analyses, including the detection of chromosomal changes in tumour cells (Thein *et al.*, 1987), the monitoring of bone-marrow transplants (Thein *et al.*, 1986), and the determination of twin zygosity (Hill and Jeffreys, 1985).

Jeffreys *et al.* (1986) demonstrated how DNA fingerprints may be used to study the segregation of multiple HVR loci in human pedigrees. They showed that most of the variable fragments within a DNA fingerprint are recombinationally separable, and therefore identify dispersed minisatellite loci. Using probes 33.6 and 33.15, up to 34 loci could be examined simultaneously from a pool of approximately 60 cross-hybridising loci. Although this is a powerful method for detecting linkage within a single large family, a different fraction of loci will be scored in different families, necessitating the cloning of any fragment cosegregating with a disease trait in order to follow up the analysis.

Human minisatellite probes 33.6 and 33.15 cross-hybridise to a wide range of animal DNAs, and in many cases generate DNA fingerprints as complex and variable as those of man. These include mice (see section 1.4), cats, dogs, and birds, among a wide range of vertebrate species (Jeffreys *et al.*, 1987, Jeffreys and Morton, 1987, Burke and Bruford, 1987, Georges *et al.*, 1988). The potential applications of DNA fingerprinting in animal breeding are numerous, and include pedigree (as well as strain and line) determination, the verification of samples for artificial insemination, and the search for linkage to economically important loci (Soller and Beckmann, 1983). DNA fingerprinting, among other molecular approaches, is revolutionising analysis of the genetic structure of wild populations (see Burke, 1989), and could be of value in maximising the outbreeding of endangered species in zoo colonies. Probes 33.6 and 33.15 also generate DNA fingerprints in plant species (Dallas, 1988).

1.2.4 Characterisation of human minisatellites

In order to investigate the processes of length change associated with hypervariability at HVRs, it was neccessary to isolate such loci from within the DNA

fingerprint and analyse them individually. Accordingly Wong *et al.* (1986, 1987) cloned and characterised a panel of human minisatellites. By selectively cloning the largest fragments detected by 'core' probes 33.6 and 33.15, Wong and colleagues isolated 6 loci which are among the most variable yet described in the human genome. Sequence analysis revealed that these clones contained tandem repeat sequences composed of units ranging from 9 to 45bp; variant repeat units were found dispersed through these arrays. As locus-specific hybridisation probes these cloned minisatellites detected loci with heterozygosities ranging from 90 to 99%.

These minisatellites are therefore highly informative genetic markers, and may be used (either singly or together) for individual identification. The cumulative probability that two unrelated individuals share alleles at all 6 loci is comparable with the probability that they would have identical DNA fingerprints detected by one multilocus probe, but involves the interpretation of a much less complex pattern (Wong *et al.*, 1987). The tandem repetitive nature of these clones makes them extremely sensitive hybridisation probes, and therefore of value in forensic science where DNA samples may be very small (Wong *et al.*, 1987). This sensitivity of detection can be greatly enhanced using the polymerase chain reaction, such that alleles may be amplified to detectable levels from single cells (Jeffreys *et al.*, 1988b).

These 6 loci were localised in the genome using somatic cell hybrid panels (Wong *et al.*, 1987), and *in situ* hybridisation (Royle *et al.*, 1988), and were found to be dispersed over 4 autosomes, showing preferential, though not exclusive, localisation to terminal G-bands. Indeed the pattern of HVR distribution which is emerging from the work of several groups is that minisatellites tend to be found towards the ends of human chromosomes (Donis-Keller *et al.*, 1987, Royle *et al.*, 1988). While this sub-telomeric distribution makes minisatellites ideal markers to investigate chromosomal rearrangements in tumours (Thein *et al.*, 1987), and is an intriguing observation with regard to the origin of HVRs, the possibility of saturating the human genetic linkage map with these highly polymorphic markers becomes questionable.

1.3 Evolution of minisatellite loci

Minisatellite loci are of heterogeneous variability; a spectrum of heterozygosity is observed among the individual loci which have been examined. Minisatellite variability arises due to germline mutation to new length alleles, involving gain or loss of an integral number of repeat units. High heterozygosities have been shown to be associated with high rates of germline mutation, in agreement with the random drift hypothesis for selectively neutral mutations (Jeffreys *et al.*, 1988a, Kimura, 1983). Both the repeat unit

sequence and DNA sequences flanking the tandem repeat region are likely to have an important influence on mutation rate at any one locus. The mechanisms driving the origin and continued evolution of minisatellites are unknown, and may be multiple. Several lines of evidence suggest that hypervariable minisatellites may be associated with high levels of recombination in the human genome.

1.3.1 Unequal crossing-over

It has been argued that DNA which is not selected for coding functions will exhibit tandem repetitive patterns through unequal exchange processes dependent on sequence similarity (Smith, 1976, Stephan, 1989). Different rates of unequal recombination, relative to the base substitution rate, will generate distinct array types. Stephan (1989) demonstrated theoretically that when the recombination rate (and therefore the rate of unequal recombination) is low, long periodicities with extensive heterogeneity and higher order structures will arise as a result of unequal crossing-over and replication slippage (see section 1.3.3). Such a process might explain the structure of the highly repeated satellite DNA arrays which lie in meiotic-recombination deficient heterochromatin.

In contrast, high recombination rates will produce short periodicities with nearly identical tandem repeats, similar to the structure of HVR loci. While both processes will tend to homogenise repeat units across an array (cross-over fixation), at any one time an array is likely to be composed of several variant repeat units in a transition state (Smith, 1976). Variant repeat units are often diffused acrosss several non-adjacent repeats within minisatellite arrays (Jeffreys *et al.*, 1985a). Such a distribution is more consistent with amplification through unequal exchange than through slippage alone, which would tend to cause lateral diffusion into adjacent repeat units (see section 1.3.3). Furthermore, although most mutation events observed at minisatellites are small, large length changes also occur (Jeffreys *et al.*, 1988a), which are unlikely to be the product of slippage events. However it is important to note that there are a variety of possible recombination based events which might be involved, including meiotic, mitotic, and sister chromatid exchange events (Jeffreys *et al.*, 1985a), and associated non-reciprocal exchange mechanisms such as gene conversion (Kourilsky, 1986).

The sub-telomeric bias in minisatellite distribution observed by Royle *et al.* (1988) is similar to the distribution of chiasmata in human metaphase chromosomes (Laurie and Hulten, 1985). Chiasmata are thought to be cytologically visible consequences of crossing-over events (Janssens, 1909). Low stringency *in situ* hybridisation of minisatellite probe 33.15 to human meiotic metaphase I chromosomes showed that autoradiographic grains clustered over chiasmata (Chandley and Mitchell, 1988). This

observation supports an association between core-containing minisatellites and meiotic recombination. Furthermore, Jeffreys *et al.* (1988a) reported that the mutation rate at human minisatellites was similar in maternal and paternal germlines, even though there are many more mitoses in spermatogenesis than oogenesis. Mutation events appear therefore to be confined to a common stage of gametogenesis, possibly meiosis.

The rate of unequal exchange required to maintain the observed level of variability at core-related minisatellites in human populations has been estimated to be approximately 10^{-4} per kilobase of minisatellite (Jeffreys *et al.*, 1985a). As this is higher than the estimated mean rate of meiotic recombination in man (about 10^{-5} per kilobase, Botstein *et al.*, 1980), core-containing minisatellites may be meiotic recombination hotspots; alternatively mutation processes at these loci may be limited to sister chromatid exchange events.

1.3.2 Hotspots of recombination

Recombination rates are not constant throughout the genome. The comparison of physical and genetic distances within linkage groups has revealed that meiotic recombination between genetic markers may be enhanced or suppressed (see Steinmetz *et al.*, 1987). Several hotspots of enhanced meiotic recombination have been reported in eukaryotic genomes. Two of the best characterised meiotic hotspots are found in the major histocompatibility complex on mouse chromosome 17 (Kobori *et al.*, 1986, Uematsu *et al.*, 1986). These hotspots are MHC haplotype dependent, and are active even if present on only one homologous chromosome (see Steinmetz *et al.*, 1987). In both cases molecular analysis has revealed short repeated sequencs with some similarity to the myoglobin-related core sequence close to the cross-over points. The Eβ hotspot contains 7-9 tandem repeats of the sequence (AGGC), and the Aβ3/Aβ2 hotspot contains 4-6 (CAGA) repeats. However, there is no direct evidence that these repeats contribute to the enhanced recombination rate. Both hotspots are further associated with the presence of a dispersed repetitive element of the MT family (see Discussion).

The pseudoautosomal region of the sex chromosomes is associated with a high rate of meiotic exchange in males (Cooke *et al.*, 1985, Simmler *et al.*, 1985, Rouyer *et al.*, 1986). Obligatory pseudoautosmal recombination may be necessary to maintain pairing between the X and Y chromosomes (Burgoyne, 1986). In man this region contains many hypervariable sequences of which two have been found to be A-rich minisatellites associated with a low level of variability (Simmler *et al.*, 1987). Internally repetitive hypervariable sequences have also been described flanking a retroviral insertion in the pseudoautosomal region of the mouse (Harbers *et al.*, 1986).

HVRs are not associated with every hotspot : Chakravarti *et al.* (1984) identified a region with a 30 fold enhanced recombination rate upstream of the β-globin gene in man. This region engages in enhanced reciprocal exchange in a yeast system compared to neighbouring sequences. Although this region contains a (CA) dinucleotide repeat array, the repeat can be deleted without diminishing this effect (Treco *et al.*, 1985). Endogenous retroviral-like LTR-IS elements in the mouse genome recombine at a high frequency with exogenous retroviral LTRs (Schmidt *et al.*, 1984). Edelmann *et al.* (1989) have studied this activity *in vitro* and demonstrated that deletion of a 37bp region strongly suppressed the recombinational activity of the LTR, and that this region interacted with at least two nuclear proteins. A related sequence is found in the Aβ hotspot of the mouse MHC. Advances in long-range physical mapping of mammalian chromosomes permit the comparison of genetic and physical distances over very large intervals. Kingsmore *et al.* (1989) examined a 3Mbp (0.7cM) interval of mouse chromosome 1, and found that crossovers were not uniformly distributed over this region; these authors suggested that there may be a relationship between meiotic recombination frequency and the presence of CpG islands.

Recombination systems are more clearly understood in prokaryotic systems. The cross-over hotspot instigator (chi) sequence, GCTGGTGG, is thought to mediate homologous recombination in *E.coli* (and certain lambda mutants) by interacting with the *rec*BC gene product to unwind and nick the duplex DNA molecule. The *rec*A gene product then promotes assimilation of the invading single strand into the homologous duplex DNA molecule to produce a D-loop and a Holliday junction (Smith, 1983). Jeffreys *et al.* (1985a) noted that the myoglobin-related core sequence bears a resemblance to chi (underlined) :

core GGAGGTGGGCAGGARG

It was proposed that core-related minisatellites might identify a similar signal in man which would recognise and interact with eukaryotic recombinases. Consistant with this, the most variable minisatellite loci which have been described are those related to the core sequence (Wong *et al.*, 1987).

Several other core-like sequences are associated with recombinationally active DNA sequences. The switch regions which mediate somatic rearrangement of the immunoglobulin heavy chain genes in man are composed of simple tandem repeats of the sequences GAGCT and GGGGT (Nikaido *et al.*, 1981). Two hotspots of recombination have been identified outside the switch sequences, and are associated with a multigene deletion (Keyeux *et al.*, 1989). Oligonucleotides related to the minisatellite core are present in both sequences; however the significance of this observation is unclear. In yeast a sequence which confers high postmeiotic segregation frequency to heterozygous

deletions, and which may be involved in recombination, is similar to the myoglobin-related core sequence (White *et al.*, 1988). The Foldback family of transposable elements in *Drosophila* are both mobile in the genome and variable in sequence (Truett *et al.*, 1981); the inverted repeats of these elements consist of tandem repeats of a core-like sequence which vary in copy number. These examples provide indirect, although highly circumstantial, support for an association between core-related sequences and recombination events.

1.3.3 Slippage events

Spontaneous mutational 'hotspots' were discovered in prokaryotic systems many years ago (Benzer, 1961). Farabaugh *et al.* (1978) analysed the molecular basis of a hotspot in the *lac*I gene of *E.coli* and showed that this was due to a high frequency of deletion and addition of the unit CTGG at a tandem array of three such repeats. While they found that an array of four repeat units reverted at a very high frequency, arrays of two repeat units were extremely stable. As the reversion rate remained high in a *rec*A host this appeared to be the result of recombination independent slippage events.

Slippage events occur when duplex DNA strands mispair within regions of short tandem repeats, generating duplications and deletions on DNA replication (Levinson and Gutman, 1987). In contrast to unequal exchange, slippage events involve a single DNA duplex, and the size of the length change is biased to the shortest length compatible with mispairing. However, the frequency of both processes increases with the size of the tandem array. Levinson and Gutman (1988) demonstrated high rates of slipped strand mispairing within (CA) dinucleotide arrays in both *rec*A$^+$ and *rec*A$^-$ *E.coli* hosts. These authors (Levinson and Gutman, 1987) argue that slipped strand mispairing may be a ubiquitous force in the evolution of the eukaryotic genome, expanding short, simple tandem repeats into longer arrays. Such arrays may then be predisposed to more rapid expansion by further slippage and unequal crossing-over events; the net result is dependent on the relative rates of these processes (Stephan, 1989). Thus a combination of evolutionary processes may be envisaged to explain the diversity of dispersed tandem repeat loci found in the genome.

1.3.4 Significance of minisatellite evolution

Tandem repeated simple sequences may therefore be the natural ground state of any non-selected DNA sequence (Smith, 1976, Levinson and Gutman, 1987). This argument is supported by the abundance and diversity of tandem repeat loci dispersed in the genome. Additional regions of 'cryptic simplicity', containing scrambled repeat units,

have been found to underlie many more complex sequences (Tautz *et al.*, 1986). The expansion and contraction of tandem repeat arrays is therefore likely to have contributed significantly to the evolution of contemporary genomes.

Unequal recombination events and slippage driven mutations, by which simple tandem arrays evolve, are likely to occur at a higher rate than point mutations. These processes may therefore be important in the rapid evolution of coding sequences. Tandem CAG repeats may contribute to evolutionary divergence in developmentally important genes of *Drosophila*, such as *hunchback*, a gap segmentation gene, (Treier *et al.*, 1989), *Notch* (Wharton *et al.*, 1985), and *engrailed* (Kassis *et al.*, 1986). The human involucrin gene has evolved very recently through successive amplification and duplication events, and may have arisen from a simple CAG tandem array (Eckert and Green, 1986). Similarly, the ancestral mammalian proline-rich protein gene may have evolved through amplification and divergence of a 42bp GC-rich repeat unit (Ann and Carlson, 1985). Oligonucleotide repeats have even been proposed to be primal ordered molecules from which all complex genetic organisation has evolved (Ohno, 1984).

1.4 Hypervariable minisatellite loci in mice

Minisatellite loci related to the myoglobin core sequence are detected in the mouse genome by probes 33.6 and 33.15; these mouse DNA fingerprints are as complex and variable as those of man (Jeffreys *et al.*, 1987). The strength of classical mouse genetics can therefore be applied to the analysis of murine minisatellites, and to the comparison of minisatellite organisation in mouse and man. In order to appreciate the advantages afforded by such studies it is neccessary to follow the development of inbred and recombinant inbred strains of mice.

1.4.1 Inbred strains of mice

1.4.1.a Historical

Darwin (1859) proposed that evolutionary processes act on phenotypic variation in natural populations. It was not until the rediscovery of Mendel's (1866) observations on the particulate nature of inheritance that the genetical basis of natural variation was understood (Correns, 1900, De Vries, 1900, von Tschermak, 1900). At the same time Johannsen (1903) observed that without variation, within pure (or inbred) lines, selection was ineffective. It became clear that it would be necessary to work with such genetically defined populations in order to study the relative contribution of genotypic and environmental effects to phenotype. The breeding of inbred strains of mice, including

most laboratory strains commonly used today, dates back to this period (see Morse, 1978).

There were many advantages to using the mouse as a model mammalian system. Small body size, resistance to infection, large litter size, and relatively rapid generation time made mice particularly suitable for laboratory studies. Furthermore, a pool of coat colour and behavioural mutations were available through stocks bred by mouse fanciers. Mice provide highly informative model systems for many human disorders, and it is through a comparative analysis of human and mouse genetics that their value becomes evident.

There are many inbred strains of mice available today (Festing, 1979) including recombinant inbred (see below) and congenic inbred strains, which differ from other strains at a single locus. Many of the established inbred strains are derived from the same original mouse stocks, and on the basis of mitochondrial DNA RFLPs are thought to have descended from a common female ancestor (Ferris *et al.*, 1982). The genomes of inbred mouse strains are known to have genetic contributions from the two major taxonomic groups of wild mice, *Mus musculus domesticus* (Western Europe and USA), and *Mus musculus musculus* (Eastern Europe, Russia and China) (Bishop *et al.*, 1985).

Inbred strains of mice are used in every area of mammalian biology. The earliest studies with inbred mice, carried out by the innovative pioneers of mouse genetics (including C.Little, L.Strong, L.Loeb and H.Bagg, see Morse, 1978), involved investigations into the genetics of cancer susceptibility. Much of the subsequent development of inbred mice strains has been coordinated through the Jackson Laboratory, Maine, founded in 1929. Among the many contributions of inbred mice to mammalian biology are advances in cellular and molecular immunology (including the discovery of antigenic variation and the murine histocompatibility complexes), radiation biology, drug and carcinogen testing, and disease models for human disorders (see Morse, 1978, and Festing, 1979). The potential contribution of inbred mouse strains to studies of genomic organisation and evolution is now being realised through recent advances in molecular biology.

1.4.1.b Inbreeding

The most efficient inbreeding strategy in mammals is full-sib mating. After 20 generations the coefficient of inbreeding, which is the probability that two alleles at a locus are identical by descent (Wright, 1934, Malecot, 1948) becomes 0.986 (see Falconer, 1960). Theoretically within an inbred strain all individuals are genetically identical (isogenic), and each individual is homozygous at every locus. However in

practice there will invariably be low levels of heterozygosity, persisting due to incomplete inbreeding, and arising through strain contamination and new mutation. New mutation is responsible for the divergence of inbred strain sublines which are bred at different locations.

The high level of homozygosity within an inbred strain is associated with a decrease in vigour, termed inbreeding depression (see Falconer, 1960). This is a compound effect resulting from homozygosity for recessive alleles at many loci; however after the initial generations of full-sib mating the fitness of surviving inbred strains tends to stabilize. The converse of inbreeding depression is hybrid vigour, or heterosis (see Falconer, 1960). F_1 hybrid mice resulting from the cross of two inbred strains are isogenic, and each individual is heterozygous at every locus. Such mice are often used in research where more robust animals and larger litters are required.

1.4.1.c Recombinant inbred strains of mice

As a knowledge of mouse linkage groups expanded, and many new polymorphic genetic markers were identified, the need for more efficient and detailed mapping systems emerged. Recombinant inbred (RI) strains, developed by Bailey (1971) and Taylor (1978), have proved to be an important advance in mouse mapping technology.

Recombinant inbred strains are derived from the cross of two unrelated but each highly inbred progenitor strains. The heterozygous F_1 mice are full-sib mated, the progenitor alleles segregating in the F_2 generation. F_2 pairs are then crossed and thereafter maintained independently under a strict regimen of full-sib inbreeding to generate a number of RI strains. Each strain will be fixed at every genetic locus for the allele of one or other progenitor strain. On average half of the RI strains in a panel should become fixed for the allele contributed by one parent, and half for the allele contributed from the other. Thus for any locus at which the progenitor strains differ a clear strain distribution pattern (SDP) for the progenitor alleles should emerge across the panel of RI strains.

The primary application of RI strains is in linkage analysis. Genetic markers which are unlinked will be found equally in parental and recombinant phases, whereas linked genetic markers will be found in the same combinations in which they entered the cross. The closer two loci are linked, the more their SDPs will match. Taylor and Meier (1976) first demonstrated this application of RIs by mapping the adrenal lipid depletion gene to chromosome 1. The power of such a system is that data accumulates as more loci are mapped. However, as SDPs can only be derived for markers which are polymorphic between the progenitor strains, RIs cannot be used to directly map new

mutations. RIs may also be used to investigate the number of genes responsible for particular traits. While a clear SDP of one or other progenitor allele is evidence for a single gene mode of inheritance, complete phenotypic gradation between RI and progenitor strains is evidence for a polygenic mode of inheritance. Furthermore, because RIs represent a single segregation which has been amplified over many generations, RI panels can be used to estimate the rate of strain divergence due to new mutation.

There are currently over 15 established RI panels (see Taylor, 1989). The most informative panels contain a large number of RI strains which increases the likelihood of finding significant linkage between two markers, and are the products of a cross between two genetically distinct progenitor strains, thus increasing the number of polymorphic markers which may be analysed. The BXD series, composed of 26 RI strains bred for 63-88 generations are derived from C57BL/6J and DBA/2J mice (Taylor, 1989). The BXD panel have been typed for at least 200 hundred genetic markers; the cumulative genetic distance covered makes it highly likely that linkage will be detected for any new polymorphism analysed.

1.4.2 Mouse DNA fingerprints

Inbred strains are homozygous at most loci and therefore have less complex DNA fingerprints than wild mice (Jeffreys *et al.*, 1987). Within an inbred strain the DNA fingerprints of individual mice are very similar, with only minor variation above that which may be due to Y-specific minisatellite loci. Such strain-specific patterns should prove useful in strain identification, and in monitoring the extent and rate of subline divergence.

It is possible to genetically dissect the DNA fingerprints of inbred mouse strains by studying the segregation of the component DNA fragments in RI strains. Such studies can investigate whether the same loci are scored in the DNA fingerprints of different strains (allelism), how many loci are represented within the DNA fingerprint of a particular strain (linkage), and how frequently minisatellite loci mutate to new length alleles (germline stability).

C57BL/6J and DBA/2J mice each have distinct DNA fingerprint patterns detected by human minisatellite probes 33.6 and 33.15. Jeffreys *et al.* (1987) used the BXD RI panel to compare the DNA fingerprints of these two strains. The SDPs for 15 B (C57BL/6J) and 13 D (DBA/2J) specific fragments were determined. These identified 13 loci, 5 detected by 33.6 (*Ms6-1* to *Ms6-5*), and 8 detected by 33.15 (*Ms15-1* to *Ms15-8*). Of these loci 10 had alleles detectable in only one of the two progenitor strains. This is comparable with the fraction of scorable heterozygous loci represented by a single allele

in individual human DNA fingerprints (Jeffreys *et al.*, 1986). While most of these loci were represented by a single segregating fragment, one locus, *Ms15-1*, was more complex, and consisted of 10 cosegregating fragments in the B allele, and 2 in the D allele. This locus maps to chromosome 4 and appears to be a midisatellite, with a size range of 19 to >90kb, comparable to that identified in man by Nakamura *et al.* (1987a).

The SDPs for 8 of these loci showed significant linkage to the SDPs of previously mapped mouse genetic markers. The remaining loci presumably lie in regions of the genome for which no BXD polymorphisms have been typed. Alternatively, new mutation during the breeding of the RIs has masked the correct SDPs of these loci. The loci which were mapped were dispersed over 5 mouse autosomes. Two loci mapped to each of chromosomes 4, 5, and 14, and one to chromosomes 6 and 7. Of the syntenic pairs, only the loci on chromosome 14 showed significant linkage to each other, and were estimated to be 5 centimorgans apart. In contrast to observations in humans (Royle *et al.*, 1988) these loci were not preferentially distributed near the ends of chromosomes. This lack of association with highly repeated satellite sequences suggests that minisatellites will provide valuable dispersed markers for linkage analysis in the mouse.

Elliott (1986) used a probe containing simple-copy sequence to detect multiple strain-specific minisatellite fragments in mouse DNA. The segregation of some of the fragments was followed in the BXD and AKXL RI strains. These fragments represented recombinationally seperable loci, which were dispersed throughout the mouse genome. These loci are distinct from those detected by probes 33.6 and 33.15 (Jeffreys *et al.*, 1987). Other minisatellite loci have been identified in mouse DNA. Kominami *et al.* (1987) used a probe based on the myoglobin-related core sequence to isolate a clone from a BALB/c mouse genomic library which contained a minisatellite with a 14bp repeat unit. This fragment in turn cross-hybridised to other minisatellite loci in the mouse genome. Using a tandem repeat sequence from the *Drosophila per* locus Shin *et al.* (1985) isolated a mouse clone which cross-hybridised to many variable fragments in human DNA (Georges *et al.*, 1987).

Many short tandem repeat loci which are likely to represent HVRs have been identified in the mouse genome close to isolated genes and repetitive elements. In some cases these have been shown to be variable, such as the quadruplet repeats present at the Aβ and Eβ MHC hotspots (Uematsu *et al.*, 1986). However in most cases it is not known whether such repeats are polymorphic. The extent to which short tandem repeat sequences are variable in contemporary genomes may now be directly analysed through the polymerase chain reaction (Saiki *et al.*, 1988).

1.5 Aims of project

The development of single locus probes has yielded much information on the structure, distribution, and mutation rate of minisatellite loci in man (Wong *et al.*, 1987, Royle *et al.*, 1988, Jeffreys *et al.*, 1988). Such loci provide the most polymorphic and informative genetic markers identified in the human genome. Using a similar approach to isolate and analyse HVR loci in the mouse it should be possible to compare the organisation of minisatellite loci in the two species. In addition, the advantages of genetically controlled inbred strains and mouse breeding systems will allow the genetical behaviour of individual loci to be directly studied.

The most variable minisatellites are those associated with the highest mutation rates (Jeffreys *et al.*, 1988a). As in man, minisatellite loci in mice detected by probes 33.6 and 33.15 show a wide range of germline stability (Jeffreys *et al.*, 1987). By isolating individual highly unstable loci it will be possible to assess the relative contributions of repeat unit sequence and DNA context to hypervariability, and to design experiments to directly examine the mechanisms involved in new-length allele generation. If minisatellites are hotspots of meiotic recombination, for example, are mutation events accompanied by the exchange of polymorphic markers flanking the minisatellite? It might be anticipated that loci with extremely high germline mutation rates would be associated with some level of somatic mutation; if so how do germline and somatic events compare mechanistically?

Jeffreys *et al.* (1987) observed that one locus in particular was extremely unstable in the BXD RI strains. The objectives of this research were to isolate and characterise this hypermutable locus and thus provide a model with which to study processes of allelic change at hypervariable loci.

II. MATERIALS AND METHODS

2.1 STOCKS AND REAGENTS

2.1.1 Mice (*Mus musculus*) and mouse DNA

C57BL/6J and DBA/2J 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, Maine, USA. C57BL/6J mice also purchased from the Radiobiology Unit (Harwell, England). A, AKR, BALB/c, C3H/He, CBA, DBA/2J and SWR inbred mice were purchased from Bantin and Kingman Ltd. (Hull, England); the DBA/2J mice had originally been obtained from the Jackson Laboratory via Searle Ltd. DBA/2J, 129, AKR, BALB/c, SWR and SJL inbred mice were purchased from Harlan Olac Ltd. (Bicester, England); the DBA/2J mice originally came from the Jackson Laboratory via the MRC Laboratory Animal Centre (Carshalton, England).

C57BL/6J mice and animals from the BXD RI series were bred by Dr. G. Bulfield in Edinburgh, and supplied as DNAs. BXH (and progenitor strains) and Peruvian feral mice livers were also obtained from Dr G. Bulfield, as were DNAs from European and Japanese feral mice. SWXL (and progenitor strains) and additional BXD RI strain DNAs were received from Dr. B. Taylor (The Jackson Laboratory, Maine, USA). F9 and derived cell line DNAs were provided by Dr. P. Goodfellow (ICRF, London, England).

2.1.2 Chemicals, enzymes and antibiotics

Restriction endonucleases, 10x REact buffers, M13mp18 and M13mp19 RF DNAs, λ*Hin*dIII and ΦX174*Hae*III DNAs, DH5 and DH5α competent cells, and T4 DNA ligase were obtained from Gibco-BRL plc., Paisley, Scotland.

NaAmpicillin, bovine serum albumin, dithiothreietol, Ficoll 400, HEPES (N-2-hydroxyethylpiperazine-N'-2-ethanesulphonic acid), IPTG (isopropyl-β-D-galacto-pyranoside), PEG (polyethylene glycol) 6000, salmon sperm DNA (Na salt), spermidine trichloride, TEMED (N,N,N',N'-tetramethylethylenediamine), ribonuclease A, proteinase K, lysozyme and thymine were supplied by Sigma Chemical Co., Poole, England. MOPS (3-(morpholino)propanesulphonic acid), Tris (2-amino-2(hydroxy methyl) propane-1,3-diol) and dimethyldichlorosilane were obtained from BDH, Poole, England. Formamide, urea,

phenol and polyvinylpyrrolidine were supplied by Fisons, Loughborough, England. Acrylamide was obtained from Serva, Heidelberg, West Germany; N,N'-methylenebisacrylamide from Uniscience, Cambridge, England; SeaPlaque agarose (FMC) and SeaKem HGT agarose (FMC) from ICN Biomedicals Ltd., High Wycombe, England; ammonium persulphate from Bio-Rad Laboratories, Watford, England. BCIG (5-Bromo-4-chloro-3-indoyl-β-D-galactopyranoside) was obtained from Anglian Biotechnology, Colchester, England; Biolabs T4 DNA ligase from CP Laboratories, Bishop's Stortford, England; deoxyribonucleotides, dideoxyribonucleotides, hexadeoxyribonucleotides, DNA polymerase I (Klenow fragment) and T7 DNA polymerase from Pharmacia, Milton Keynes, England. Calf intestinal phosphatase was supplied by Boehringer Corporation (London) plc., Lewes, England.

Bacterial media was supplied by Oxoid Ltd., Basingstoke, England, Becton Dickinson Ltd., Oxford, England (BBL trypticase), and Difco Ltd., East Molesley, England (Difco tryptone, Difco agar and Bacto yeast extract).

Radionucleotides, Hybond-N and -C filters and Taq DNA polymerase were supplied by Amersham International plc., Little Chalfont, England; Gigapack Plus *In vitro* packaging kit by Stratagene Cloning Systems, California, USA; nitrocellulose filters by Schleicher and Schuell, Dassel, West Germany.

Oligonucleotides were synthesised by Mr. J. Keyte, Biochemistry Department, University of Leicester, on an Applied Biosystems 380B DNA synthesiser using reagents supplied by Cruachem.

All other chemicals were of analytical grade.

2.1.3 Media

Luria broth (LUB) contained 10g Difco bacto tryptone, 5g Difco bacto yeast extract and 5g NaCl per litre dH_2O. Luria agar plates (LB) were prepared by solidifying LUB with 15g Difco bacto-agar per litre. NaAmpicillin was added to 50μg/ml (from a stock of 25mg/ml in 50% ethanol) to select for cells harbouring β-lactamase encoding plasmids.

S.O.C. contained 2% (w/v) bactotryptone, 0.5% yeast extract, 10mM NaCl, 10mM $MgCl_2$, 10mM $MgSO_4$, and 20mM glucose. Filter sterilised $MgCl_2$, $MgSO_4$, and glucose were added to the other components which had been dissolved in water and autoclaved.

Phage were plated in soft overlay agar (BTL, containing 10g BBL trypticase, 5g NaCl, 2.5g $MgSO_4.7H_2O$ and 6g Sterilin or Difco agar per litre dH_2O) onto a BLA base (as BTL except 15g agar/litre).

E.coli JM101 was maintained on glucose supplemented minimal medium plates (42mM Na_2HPO_4, 22mM KH_2PO_4, 18mM NH_4Cl, 8mM NaCl, 22mM glucose, 0.1mM $CaCl_2$, 1mM $MgSO_4$, 3μM thiamine HCl, 0.17mM proline and 15g Davis agar per litre).

2.1.4 Bacterial strains

The following strains of *Escherischia coli* were used :

DH5	F$^-$, *end*A1, *hsd*R17(Rk$^-$,Mk$^+$), *sup*E44, *thi*1, λ$^-$, *rec*A1, *gyr*A96, *rel*A1 (Hanahan, 1983).
DH5α	F$^-$, *end*A1, *hsd*R17(Rk$^-$,Mk$^+$), *sup*E44, *thi*1, λ$^-$, *rec*A1, *gyr*A96, *rel*A1, (*arg*F-*lac*ZYA) U169, Φ80d*lac*I^q,ZΔM15 (BRL Focus, 1986).
WL95	803, *sup*E, *hsd*Rk, *hsd*Mk, *ton*A, *trp*R, *met*B (Loenen and Brammar, 1980).
ED8910	*sup*E44, *sup*F58, *rec*B21, *rec*C22, *hsd*S, *met*B, *lac*Y1, *gal K2*, *gal*T22 (Loenen and Brammar, 1980).
JM101	Δ(*lac-pro*), *sup*E44, *thi*1, F'*tra*D36, *pro*AB, Φ80d*lac*I^q, *lac*ZΔM15 (Messing, 1981)
NM522	Δ(*lac-pro*AB), *thi*$^-$, *sup*E, *hsd*R17(Rk$^-$,Mk$^+$), [F'*pro*AB, *lac*I^q,ZΔM15]
NM554	*rec*A13, *mcr*A$^-$, *mcr*B$^-$, *hsd*R, Δ(*lac*)X74 (Raleigh *et al.*, 1988)

2.1.5 Cloning vectors

λL47.1 (Loenen and Brammar, 1980) and charomid 9-32 (Saito and Stark, 1986) were used to construct genomic libraries. Plasmid pUC13 (Vieira and Messing, 1982) was used for subcloning. M13mp18 and M13mp19 (Yanis-Perron *et al.*, 1985) were used to subclone DNA fragments for DNA sequencing.

2.2 GENERAL METHODS FOR DNA HANDLING

2.2.1 Estimation of DNA concentration

DNA yield was measured by absorbance of ultraviolet light at 260nm on a Cecil Instruments CE 235 spectrophotometer (a reading of 0.02 O.D. units (0.05 units for oligonucleotides) corresponding to a nucleic acid concentration of 1μg/ml). Alternatively, samples were visually compared with known concentrations of phage DNA on agarose gels.

2.2.2 Restriction endonuclease digestions

Reactions containing DNA at concentrations of 0.05-0.5mg/ml were incubated at 37°C for 0.5hr (plasmid) to 4hr (genomic) in the presence of the appropriate REact buffer and restriction endonuclease (in excess of 1 unit/0.5μg DNA). Spermidine trichloride was added to a final concentration of 4mM to suppress non-specific cleavage (Pingoud, 1985).

2.2.3 Polynucleotide kinase treatment

As described in Maniatis *et al.* (1982), DNA was incubated at 37° for 15 minutes in the presence of 1x kinase buffer (50mM Tris-HCl (pH 7.6), 10mM $MgCl_2$, 5mM dithiothreitol), 1mM ATP, and T4 polynucleotide kinase (>10 units/pmole 5' end DNA). Further enzyme and ATP were added and incubation continued for 30 minutes. Kinase was inactivated by heating samples to 65° for 5 minutes.

2.2.4 Alkaline phosphatase treatment

To remove the 5' phosphate group from DNA fragments, calf intestinal phosphatase (0.2U/μg DNA) was added after restiction endonuclease digestion, and incubation at 37° continued for 30 minutes. The reaction was stopped by adding EDTA to 20mM and SDS to 0.5%, and the phosphatase inactivated at 65° for 10 minutes. DNA was recovered with two phenol/chloroform extractions.

2.2.5 End-repair of recessed 3'ends

DNA (1-10μg) was incubated in 1x ligase mix (see section 2.2.8), 10mM Tris-HCl (pH 8.0), 10mM $MgCl_2$, 4mM spermidine trichloride, 25μM dNTPs with 1 unit of Klenow polymerase at RT for 30 minutes. Polymerase was removed by phenol/chloroform extraction followed by ethanol precipitation.

2.2.6 Phenol/chloroform extraction

DNA solutions were emulsified with an equal volume of phenol/chloroform (phenol, chloroform, isoamyl alcohol and 8-hydroxyquinoline prepared in the ratio 100:100:4:0.1 (w:v:v:w), saturated with Tris-HCl (pH 7.5) (stored at 4°)). After centrifugation the upper aqueous layer was carefully removed and transferred to another tube. To maximise recovery 0.5 volume of dH_2O was added to the phenol layer and re-emulsified. After centrifugation the aqueous phases were pooled.

2.2.7 Ethanol precipitation

To concentrate, or recover DNA after manipulation, one tenth volume of sodium acetate (pH 5.6) and 2.5 volumes of 100% ethanol were added. The solution was mixed and placed at -80° (or in a dry-ice/ethanol bath) for at least 5 minutes (longer for low DNA concentrations) and centrifuged for 5 or more minutes. To maximise yield the solution was re-chilled and re-centrifuged (with a 180° rotation of the tube in the microcentifuge). The ethanol was removed and the pellet rinsed in at least the original volume of 80% ethanol by centifugation. The ethanol was removed as before and the pellet dried under a vacuum and dissolved in the required volume of dH_2O. DNA solutions were generally stored at -20°.

2.2.8 Ligations

Reactions were generally carried out in 10µl volumes with less than 100 ng DNA in the presence of 1x ligase buffer (50mM Tris-HCl (pH7.5), 10mM $MgCl_2$, 10mM dithiothreitol), 1mM ATP, 4mM spermidine trichloride and 0.5U T4 DNA ligase. Cohesive and blunt-end ligations were incubated at 4° overnight. λ*Hin*dIII (or ΦX174*Hae*III) DNA fragments were usually ligated in parallel control reactions and analysed on a mini-gel to monitor efficiency of ligation. The ligase was inactivated by heating to 65° for 10 minutes.

2.3 MICE

2.3.1 Housing

Mice were housed in the Biomedical Services of Leicester University. The mouse room was maintained under a 12hr dark/12hr light cycle.

2.3.2 Identification

Toe-clipping was used to identify mice in the colony. The numbering system adopted was basically that of Allen *et al.* (1987). No more than two toes off each of two paws were clipped from one 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 ommited) were marked on the left hindpaw. Toes were clipped under anaesthetic (halothane).

2.3.3 Removal of a length of tail

The mice were anaesthetised under halothane, and 1-2cm of tail cut off with a scalpel blade. The tails were cleaned with a sterile gauze and cauterised to prevent bleeding. Toes were clipped in the same operation.

2.3.4 Isolation of day 8 embryos

Females were checked for vaginal plugs (day 1) on successive mornings after matings were set up. On day 8 pregnant mice were sacrificed and the uterus dissected out into a Petri-dish containing ice-cold PBS (phosphate-buffered saline). Embryos were carefully dissected out of the decidua in drops of PBS under a binocular microscope as described by Beddington (1987).

2.3.5 Preparation of mouse DNA

2.3.5.a Tail DNA

A protocol modified from Allen *et al.* (1987) was followed. Using a scalpel 1-2cm of tail was cut into small pieces and resuspended in 0.5ml 1x SE (150mM NaCl, 100mM EDTA (pH 8.0)) in a 1.5ml eppendorf microcentifuge tube. 50µl of 10% SDS and 5µl proteinase K (20mg/ml) were added, the suspension mixed by inversion and incubated at 50° for 3-5hr (with occasional mixing). Protein and bone were removed by phenol extraction : 0.5ml phenol/chloroform was added and mixed until the solution was thoroughly emulsified and the sample was then centifuged for 5 minutes. The viscous upper layer was removed using a cut-off 1ml disposable tip. Care was taken to avoid all contact with the organic/aqueous interphase. Chloroform (0.5ml) was added, emulsified, and the aqueous layer removed after centrifugation as before. An equal volume of isopropanol was added, and a small clot of DNA precipitated on inversion. The DNA was rinsed O/N in 80% ethanol (0.5ml) and resuspended in 100µl dH_2O after removing all traces of ethanol. This gave approximately 100µg high quality DNA with very little variation in yield between samples.

2.3.5.b Tissue DNA

Small-scale tissue preparations from dissected mice were carried out as for tails, except that incubations were for 1-2hr at 50°. Embryo DNA was similarly prepared in 50µl volumes with a 1hr incubation at 50°. Large-scale tissue preparations followed a

protocol modified from Maniatis *et al.* (1982). All glassware involved in DNA preparation was carefully rinsed twice in dH_2O before use. The volume of tissue was estimated, and 10 volumes of 1x SE added; the tissue was then hand homogenised in a glass tube and transferred to a conical flask. 1/10 vol 10% SDS was added and the flask gently swirled to assist lysis. The viscous solution was emulsified with 0.5 vol phenol/chloroform by further gentle mixing. After decanting to centrifuge tubes the mixture was spun at 10,000rpm for 5 minutes (RT). The aqueous layer was removed using a broken-off siliconised Pasteur pipette and the organic layer re-extracted with dH_2O. The aqueous layers were pooled in a beaker and ethanol precipitated : a large stringy clot of DNA appeared on swirling. This was washed in 80% ethanol and resuspended in 0.5x original volume of 0.1x TNE (1x TNE is 10mM Tris-HCl, 100mM NaCl, 1mM EDTA (pH8.0)). Pancreatic RNase A was added to 20µg/ml and the beaker incubated at RT for 30 minutes (stock RNase : 20mg/ml in 0.15M NaCl, heated to 80° for ten minutes (stored at -20°)). SDS was added to 1%, TNE to 1x, and the mixture emulsified with phenol/chloroform. After centifugation and re-extraction, the DNA was ethanol precipitated and re-dissolved at least three times before a final 80% ethanol rinse. The DNA was transferred to an eppendorf tube and dissolved in water at 4° O/N.

2.3.5.c Removing polysaccharides from liver DNA

To remove polysaccharides from high molecular weight liver DNA, the following additional step was incorporated. DNA was dissolved in a small volume of 0.1x TNE to which an equal volume of 2.5M potassium phosphate (pH 8.0), and an equal volume of 2-methoxyethanol were added. After mixing and centrifugation at 10,000rpm for 2 minutes the viscous upper layer was removed. The lower layer was re-extracted with 1 vol dH_2O, 1 vol K phosphate, 1 vol 2-methoxyethanol. The pooled aqueous phases were ethanol precipitated. Ethanol precipitations were repeated until the K phosphate was removed.

2.4 AGAROSE GEL ELECTROPHORESIS

2.4.1 Pouring gels

The concentration of agarose used depended on the resolution required and varied within the range 0.5% (>10kb) to 2% (<500bp). Electrophoresis buffers used were either ELFO (40mM Tris.acetate, 1mM EDTA, 0.5µg/ml ethidium bromide, pH 7.7) or TAE (40mM Tris.acetate, 20mM sodium acetate, 0.2mM EDTA, 0.5µg/ml ethidium bromide, pH 8.3). Agarose was dissolved in buffer in a microwave oven and allowed to

cool to 40^{o} before pouring into a mould (usually a glass plate of the required size, bordered with tape, with a suitable comb to form the wells in position). Plate size depended on the resolution required; routinely 10x10cm plates were used, for genomic analysis plates of 20x20cm or 20x30cm (for resolution of large fragments) were used. Well sizes ranged from 0.2cm to 14cm; generally 0.5cm wells were used to analyse up to 5μg DNA.

2.4.2 Electrophoresis

Gels were immersed in tanks containing buffer (either ELFO or TAE). Loading dye was added to the samples (5x mix contained 0.2M Tris.acetate, 0.1M sodium acetate, 1mM EDTA (pH 8.3), 12.5% ficoll 400 and 0.1% bromophenol blue) to prevent diffusion and monitor migration during electrophoresis. Ethidium bromide (0.5μl of 5mg/ml stock) was added to digested genomic DNA (generally 5μg of DNA in 20μl) to enhance even migration within each track. Size markers were loaded in parallel with the samples; usually λ DNA digested with *Hin*dIII (23-0.1kb range) and/or ΦX174 digested with *Hae*III (1.3-0.07kb range). A current was applied to the gel and the voltage adjusted to suit a particular run.

2.4.3 Gel analysis

The intercalation of ethidium bromide (which fluoresces under ultraviolet light and is present in the running buffer) into DNA allows the visualisation of fragments separated in the gel (Sharp *et al.*, 1973). Gels were monitored during a run using a hand-held long wave UV lamp, and after a run on a bench-top UV transilluminator (short wave UV; Chromato-Vue C-63 UV Products Inc., San Gabriel, California, USA). Gels were photographed using Kodak negative film (T-Max Professional 4052). The film was processed using Kodak LX24 developer, FX40 fixer and HX40 hardener (Kodak, Hemel Hempstead, England).

2.4.4 Southern blotting (Southern, 1975b)

Gels to be blotted were photographed and the size markers lightly notched with a blade under UV light. Gels were then shaken gently (on a Gerhardt shaker) in 0.25M HCl for 7 minutes (depurination - to enhance the transfer of large molecules) and the solution aspirated off. This was repeated once more. Gels were then shaken twice in alkali (0.5M NaOH, 1M NaCl) for 15 minutes (denaturation) and twice in neutralising solution (0.5M Tris-HCl (pH7.5), 3M NaCl) for 15 minutes.

Gels were then transferred onto a wick of 3MM (Whatman) paper soaked in 20x SSC (1x SSC is 0.15M NaCl, 15mM trisodium citrate pH 7.0) on a glass plate over a reservoir of 20x SSC. A filter cut to cover the area of interest was then carefully placed on the gel (avoiding all air bubbles). Filters were pre-soaked in 3x SSC (Hybond-N, Amersham), or water and then 10x SSC (nitrocellulose, Schleicher and Schuell). A similarly cut and pre-wet sheet of 3MM paper was placed on the filter (again, avoiding all air bubbles). A stack of paper towels were placed on top of the 3MM paper to draw solution upwards. A glass plate and a 500g weight were placed on top to maintain tight juxtaposition between the gel and the filter. The towels were changed regularly during the first hour of blotting to avoid patchy transfer. Blots were left for at least 2hr (genomic blots usually overnight).

After blotting the position of the size markers was marked on the filter with a biro, the filter rinsed in 3x SSC, and blotted dry in 3MM paper. DNA transferred to nylon filters was fixed by a 2 minute exposure to short wave UV light on the transilluminator (through a layer of Saran-Wrap (Dow Inc., USA)). Nitrocellulose filters were baked for 2-4hr in an 80^{o} oven. Filters were then cut into strips ready for hybridisation.

2.4.5 Preparative gels

DNA was collected from preparative gels by electrophoresis onto dialysis membrane (Yang *et al.*, 1979). There were generally two scales of operation : large scale preparative gels to fractionate genomic DNA for library construction (input >50μg), and small scale preparative gels to collect defined DNA fragments for subcloning (input <5μg).

2.4.5.a Large scale preparative gels

A 2% agarose base-plate was poured into a 20x20cm mould (100ml). A 14cm slot former was clamped above this and 0.8% agarose in 1/3x ELFO was poured in (at 4^{o}) to make a deep gel (>1cm). Samples (which had been monitored on a mini-gel for complete digestiom) were loaded in a 1.8ml volume with 200μl 5x loading mix and 50μl 5mg/ml ethidium bromide between two λ*Hin*dIII marker lanes. Gels were run in the dark-room in 1/3x ELFO at 20-30V overnight. Dialysis membrane, cut into as many single layer 15cmx1.5cm sheets as required, was boiled for 5 minutes in 10mM Tris-HCl (pH 7.5), 1mM EDTA and rinsed in dH_2O. The current was switched off, the gel viewed with the long-wave UV wand, and cut at the required positions to 0.5cm either side of the main track. UV exposure was minimised to reduce the formation of thymine dimers. Using two pairs of millipore forceps a sheet of dialysis membrane was inserted vertically into each slot, so that it rested on the glass plate, and the concave face of each strip faced the

wells. Current was then applied at 200V until each fraction was fully loaded (this could be monitored both by fluorescence of the main digest and the migration of the markers at either side). With the current still on, the cuts were extended to the edges of the gel, the two halves of the gel gently prised apart to open up a 1-2mm gap and free the membrane, which was grasped by two forceps and transferred in a single movement to a Sterilin pot. The transfer was as rapid as possible - out of the current only surface tension holds the DNA to the membrane. The membrane was trapped under the lid of the pot and centrifuged for 1 minute. The DNA was transferred to an eppendorf tube and the membrane rinsed (twice) in 200µl dH_2O and re-centrifuged. The pooled DNA was centrifuged for 2 minutes to remove agarose solids, any pellet rinsed, and the DNA finally redissolved in dH_2O after ethanol precipitation.

2.4.5.b Small scale preparative gels

Generally less than 5µg DNA was loaded into a 7mm slot. As above, gels were run in the dark-room, and viewed with a UV wand before cuts were made. Single sheets of dialysis membrane (treated as above), just wider than the track, and with a lug to aid removal, were inserted into each slot, and the DNA was run onto the membranes at 150V. An extra membrane was generally inserted as a trap above the fragment to be collected. Again, each cut was extended to the edges of the gel before the membrane was transferred to an eppendorf tube. The membrane was trapped under the lid and the tube centifuged for 30 seconds, the membrane rinsed with 30µl dH_2O, and the tube re-centrifuged. The DNA was cleaned by phenol/chloroform extraction, followed by ethanol precipitation, and resuspended in the required volume of dH_2O.

2.5 DNA HYBRIDISATION

2.5.1 Preparation and labelling of probe DNA

Two labelling procedures were used to incorporate α^{32}P-dCTP imto DNA fragments for use as hybridisation probes. Random oligonucleotide priming (Feinberg and Vogelstein, 1984) was used in most labelling reactions, while M13 primer extension (Jeffreys *et al.*, 1985) was used to label probes to very high specific activities for DNA fingerprinting (probes 33.6 and 33.15).

2.5.1.a Oligo-labelling DNA fragments

Fragments were isolated either from a preparative gel as described above (2.4.5), or from 0.5% low gelling temperature agarose gels (SeaPlaque LGT). In the latter case the gel was viewed under the UV wand and gel containing at least 100ng of the required fragment excised. The volume of gel was estimated in a pre-weighed eppendorf tube and two volumes of dH_2O were added. Probes were stored at -20°.

Before labelling, probes were placed at 65° for 5 minutes. 5-10ng of DNA was diluted to 30µl with dH_2O, and was denatured at 100° for 3 minutes. After cooling at RT, 1.2µl 10mg/ml BSA (enzyme grade, Pharmacia) and 6µl OLB were added. The constituents of OLB are Solution A : 1.25M Tris-HCl (pH 8.0), 125mM $MgCl_2$, 0.18% v/v β-mercaptoethanol, 0.5mM dATP, 0.5mM dGTP and 0.5mM dTTP; Solution B : 2M HEPES, (pH 6.6 with NaOH); and Solution C : Hexadeoxyribonucleotides (Pharmacia) suspended in 3mM Tris-HCl, 0.2mM EDTA (pH7.0) at 90 O.D. units/ml. These three are mixed in the ratio 2:5:3 (vol). 2µl of α^{32}P-dCTP (Amersham, 3000Ci/mmol, 10µCi/µl) and 0.5µl Klenow fragment of DNA polymerase I (2.5U) were added and the solution mixed by pipetting. Labellings were incubated for >5hr at RT or 1hr at 37°.

Probes were recovered after labelling to remove unincorporated α^{32}P-dCTP from the solution. 70µl 'stop mix' (20mM NaCl, 20mM Tris-HCl (pH7.5), 2mM EDTA and 0.25% SDS) was added and the mixture transferred to a disposable siliconised glass tube. 100µg of high molecular weight carrier herring sperm (or human placental) DNA was added. After ethanol precipitation the DNA was rinsed in 80% ethanol and resuspended in 500µl dH_2O. A specific activity >10^9 dpm/µg was routinely obtained.

2.5.1.b M13 Primer extension

0.4µg of the appropriate single-stranded recombinant M13 DNA (see section 2.8) was annealed with 4ng of M13 17mer sequencing primer in 1x TM buffer (10mM MgCl2, 10mM Tris-HCl (pH 8.0)) at 60° for 30 minutes. 10µl AGT mix (0.125mM dATP, 0.125mM dGTP, 0.125mM dTTP, 2.5mM Tris-HCl (pH 8.0), 0.25 mM EDTA), 6µl TE buffer (10mM Tris-HCl (pH8.0), 1mM EDTA)), 3µl α^{32}P-dCTP, and 5U Klenow polymerase were added and primer extension carried out at 37°. After 15 minutes 2.5ml 0.5M dCTP (cold chase) was added and incubation continued for a further 15 minutes. The DNA was digested with a restriction endonuclease to liberate the newly synthesized strand (*Eco*RI for 33.6 or *Bam*HI for 33.15) and then denatured by the addition of 1/10 vol 1.5M NaOH, 0.1M EDTA. The labelled single stranded fragment extended from the primer was recovered by electophoresis through a 1.2% low gelling temperature agarose gel.

2.5.2 Hybridisation

Pre-hybridisation and hybridisation were carried out in perspex hybridisation chambers (5x20cm, 20x20cm and 10x10cm (for round filters)) which were incubated at 65° in a shaking water-bath. Two different protocols were followed :

2.5.2.a Phosphate/SDS hybridisations (Church and Gilbert, 1984).

This procedure was generally used for nylon filters with both oligo-labelled polycore and locus specific minisatellite probes and unique sequence probes. Filters were pre-hybridised in 0.5M Na_2HPO_4 (pH 7.2) (1M stock : 128g Na_2HPO_4 and 6.7ml 88% H_3PO_4 to 1 litre), 1M EDTA, 7% SDS for at least 10 minutes at 65°. Filters were transfered to a second chamber containing the same solution (pre-warmed to 65°) plus boiled radioactive probe and hybridised overnight at 65°. Filters of recombinant DNA or PCR products were hybridised for 2-3hr. If required, single stranded competitor DNA was added with the probe, such as vector and *E.coli* DNA (approximately 5µg/ml) for probing filters of recombinant DNA.

2.5.2.b Denhardt hybridisations (Denhardt, 1966, Jeffreys *et al.*, 1980).

This procedure was used for nitrocellulose filters either from Southern blots (for DNA fingerprints) or plaque and colony lifts. Filters were rinsed at 65° in 1x SSC for 5 minutes and transferred to 1x Denhardt's solution (0.2% Ficoll 400, 0.2% polyvinylpyrrolidine, 0.2% BSA in 3x SSC) for 30 minutes (Southern blot filters only). Filters were then placed in 1x Denhardt's, 0.1% SDS (CFHM, pre-warmed and de-gased under a vacuum) for 30 minutes. Filters were then transferred to de-gased CFHM + 6% polyethylene glycol 6000 for 15 minutes before being placed in CFHM + 6% PEG + boiled radioactive probe at 65° overnight. Filters from phage and plasmid subclone lifts were hybridised for 2-3hr in the presence of single stranded *E.coli* and vector competitor DNA (approximately 5µg/ml). The optimum amount of competitor in Denhardt hybridisations with probe 33.6 was found to be 0.5µg/ml single stranded salmon sperm DNA.

2.5.3 Post-hybridisation washes

Filters were generally washed to either low (1x SSC, 0.1% SDS, 65°) or high (0.1x SSC, 0.01% SDS, 65°) stringency. After hybridisation filters were transfered to the appropriate washing solution (pre-heated to 60°) and washed at 65° with several changes of solution until no more radioactivity came off the filters (or an included blank filter was

cold). Filters were then rinsed in 3x SSC at RT and dried in 3MM paper. If the filters had been cut before hybridisation they were reassembled before autoradiography. Filters to be re-hybridised were stipped in 0.4M NaOH at 65° for 5 minutes (with changes of solution), neutralised in 0.1x SSC, 0.1% SDS, 0.2M Tris-HCl (pH7.5) at 65° for 30 minutes, and rinsed in dH_2O.

2.5.4 Autoradiography

Filters were exposed to Fuji RX 100 X-ray film in autoradigraphic cassettes using intensifier screens (Kyokko LFII, Kasei Optonix Ltd, Japan) at -80°. Library plate filters were orientated using ^{35}S radioactive ink. Exposure times ranged from 30 minutes to 10 days. Film was developed as described in section 2.4.3. For densitometric scanning film was preflashed before exposure and autoradiographs were scanned using an Ultrascan laser densitometer (LKB).

2.6 LAMBDA AND COSMID CLONING

2.6.1 Preparation of lambda arms

λL47.1 (Loenen and Brammar, 1980) arms were prepared by A.J.Jeffreys. λL47.1 DNA was prepared by the method of Blattner *et al.* (1977) and completely digested with *Bam*HI to release the 'stuffer' (internal) fragment of the phage. The arms (23.6 and 10.4kb) were recovered by electroelution onto dialysis membrane, and their cohesive ends annealed by incubation at 68° for three minutes in the presence of 0.1M Tris-HCl (pH7.5) and 10mM $MgCl_2$. Ligations were performed as described in section 2.2.8 with a three-fold excess of arms to insert DNA. The efficiency of ligation was monitored on a mini-gel, after heating input and ligated samples to 68° for three minutes in the presence of 66mM Tris-HCl (pH7.5) and 10mM $MgCl_2$. The ligated DNA was ethanol precipitated and redissolved in a small volume of packaging buffer A (20mM Tris-HCl (pH 7.5), 3mM $MgCl_2$, 0.05% (v/v) β-mercaptoethanol, 1mM EDTA (pH 8.0)).

2.6.2 *In vitro* packaging and infection of *E.coli*

A commercial packaging kit, Gigapack Plus (Stratagene), was used to maximise library size. Care was taken to use extracts immediately after thawing. Ligated DNA (in buffer A) was added to 10μl of freeze/thaw extract. To this 15μl of sonicated extract was added, and, after gentle mixing, the contents were incubated at RT for 2hr. 500μl of λ bufffer (6mM Tris-HCl (pH7.2), 10mM $MgSO_4$, 0.005% gelatin) and a drop of chloroform

(to stabilize the phage particles) were added to the packaged DNA (phage stored at 4^{o}). *E.coli* cells of the required genotype (L95 or NM554) were grown to mid-logarithmic phase in LUB supplemented with maltose (0.2%) and $MgSO_4$ (10mM). For library plates, 150µl of packaged DNA was used to infect 150µl of *E.coli* cells. A series of contols and dilutions to estimate the recombinant phage titre were also set up. Infections were carried out in glass tubes at RT for 15 minutes with occasional shaking. 3ml of BTL (section 2.1.3, with 0.2% maltose, 10mM $MgSO_4$, maintained at 50^{o} in a heating block) was then added and the mixture poured on to BLA plates (section 2.1.3), avoiding air bubbles, and incubated at 37^{o} overnight when set.

2.6.3 Library screening

2.6.3.a Lifts (Benton and Davis, 1977).

Nitrocellulose filters (Schleicher and Scheull or Amersham) were placed on each library plate using two sterile Millipore forceps. Each filter was numbered and orientated using a sterile needle, and the unique pattern of holes marked on the back of the plate with a red-hot needle. After 5 minutes the filter was carefully lifted off the plate and floated phage side downwards on 1.5M NaCl, 0.1M NaOH for one minute in a shallow tray. The filter was then submerged in 2x SSC, 0.2M Tris-HCl (pH 7.5) for one minute, and blotted dry on 3MM paper. Filters were then baked at 80^{o} prior to hybridisation (2.5.2.b). Library plates were stored at 4^{o} for up to three months.

2.6.3.b Second round screening

Positive plaques were picked from the library plates using a sterile Pasteur pipette into 500µl λ buffer and 20µl chloroform. 100µl of this stock and a series of 10 fold dilutions were mixed with 100µl of log phase cells of the required *E.coli* strain (ED8910 or NM554). Phage were plated out as before. Plates of suitable phage density were lifted and hybridised. If necessary, a third round of screening was performed. Single, well separated, positive plaques were picked into 500µl λ buffer + $CHCl_3$ and stored at 4^{o} (permanent stocks).

2.6.4 Lambda DNA mini-preps

A scaled-down protocol of that of Blattner *et al.* (1977) was used.

2.6.4.a Purification of phage particles

5, 20 and 100µl of phage stock were used to infect 100µl of *E.coli* cells grown to late log phase in 10ml glass tubes. After 15 minutes at RT, 8ml of LUB (with 10mM $MgSO_4$ and 20µg/ml thymine) was added, and the tubes incubated at 37° overnight on an angled rack. The following morning mini-lysates were clearer and contained more bacterial debris than uninfected control cultures. The overnight cultures of each phage were pooled into 30ml Corex centrifuge tubes (acid washed) and spun at 9,000rpm for 5 minutes. The supernatant was decanted into another Corex tube (100µl removed to add to the permanent stock at 4°) and 6.7ml 10% PEG 6000 in 2.5M NaCl was added. After mixing by inversion the tube was left on ice for 15 minutes to precipitate phage. The sample was centrifuged at 9,000rpm for 10 minutes, the phage pellet resuspended in 0.5ml λ buffer, and transferred to an eppendorf. After centrifugation to remove contaminants, and rinsing of any pellet in 0.5ml λ buffer, phage was again precipitated on ice with 0.3ml PEG/NaCl. The pellet was resuspended in 250µl λ buffer.

2.6.4.b Isolation of phage DNA

Phage particles were lysed by the addition of 50µl 10% SDS. After three chloroform extractions, two ethanol precipitations and an 80% ethanol wash, the pellet was resuspended in 25µl 1x REact 4 with 20µg/ml RNase and incubated at 37° for 30 minutes. SDS was added to 1%, and after another phenol extraction, ethanol precipitation and rinse, the pellet was finally suspended in 20µl dH_2O. Although yields were low with this procedure (50-250ng), enough DNA was generally recovered for analysis on a mini-gel and for purification of the insert for subcloning.

2.6.5 Cosmid cloning

Charomid vector 9-32 (Saito and Stark, 1986) was prepared as described in section 2.7.10. DNA was digested with *Bam*HI and linear charomid purified from preparative gels. Ligations were performed as described in section 2.2.8. Ligated DNA was ethanol precipitated and resuspended in packaging buffer A (section 2.6.1). Gigapack plus packaging extracts were used as for lambda cloning (section 2.6.2). After infection of *E.coli*, 1ml of LUB supplemented with glucose (20mM) was added, and the tubes incubated for 40 minutes at 37° to allow β-lactamase expression. The cells were pelleted briefly and resuspended in 120µl Luria broth. 100µl (with parallel controls and a dilution series to determine the titre of infected cells) was plated out onto Luria agar plates (supplemented with ampicillin) and grown at 37° overnight. Library plates were

screened as described in section 2.7.2.

2.7 PLASMID SUBCLONING

2.7.1 Transformation

Competent *E.coli* JM101 cells were prepared and transformed as described in section 2.8.10. Alternatively, commercially prepared competent *E.coli* DH5 and DH5α (BRL) were used, which routinely gave efficiencies of >1 x 10^7 colonies/μg. 1-10ng of ligated DNA (in 1μl) was added to 20μl of freshly thawed cells and mixed by gentle pipetting. The cells were left on ice for 30 minutes, heat-shocked at 42° for 45 seconds, and placed on ice for 2 minutes. 90μl of SOC (section 2.1.3), or LUB supplemented with 20mM glucose, was added, and the tube shaken at 37° for 1hr to express β-lactamase. 25μg BCIG (in dimethylformamide) and 25μg IPTG (in H_2O) were added in transformations of *lacZ* carrying plasmids into *E.coli* DH5α. Non-recombinant plasmids are able to complement the chromosomal *lacZ* deletion, and thus produce β-galactosidase in the presence of inducer (IPTG), giving rise to a blue colour on cleavage of the chromogenic substrate BCIG. Recombinant plasmids, which contain an insert interrupting *lacZ*, will give rise to white colonies. Cells were then plated onto Luria agar (+ ampicillin) plates (with parallel controls and a dilution series) and incubated overnight at 37°.

2.7.2 Colony screening

Colonies were transferred to either nitrocellulose or nylon filters. After approximately 10hr growth, when the colonies were just visible as pin-pricks, filters were placed on the plates with sterile Millipores and orientated (as described in 2.6.3.a). The filters were carefully lifted off and placed colony side up on a fresh Luria agar (+ ampicillin) plate. If required, a second lift could be made at this stage. Plates were incubated for a further 10hr. Library plates were stored at 4° (for up to 3 months). Subsequent treatment of the filters depended on the type of filter used.

2.7.2.a Nitrocellulose lifts (Grunstein and Hogness, 1975).

Filters were placed colony side up on blotting paper saturated with 0.5M NaOH in a shallow tray for 5 minutes, to lyse cells. Filters were then transferred to blotting paper soaked in denaturing solution (1.5M NaCl, 0.1M NaOH) for 10 minutes, and dried again. Filters were next placed on blotting paper soaked in neutralising solution (1.5M NaCl,

0.5M Tris-HCl (pH 7.4)) for 4 minutes and blotted dry. Filters were then immersed in neutralising solution, scrubbed to remove bacterial debris, blotted dry on 3MM paper, and baked at 80° for 2-4hr.

2.7.2.b Hybond-N (nylon) lifts (Buluwela *et al.*, 1989).

Filters were placed colony sided up on 3MM paper soaked in 2x SSC, 5% SDS in a shallow tray for 2 minutes. The tray and filters were then placed in a microwave oven (with a rotating turntable) for 2.5 minutes at full setting (650W) to lyse the cells and fix the DNA to the filter in a single step. Filters were then immersed in 5x SSC, 0.1% SDS and lightly scrubbed to remove bacterial debris. Filters were rinsed and blotted dry prior to hybridisation.

2.7.2.c Storing plasmid clones

Single positive colonies were picked from library plates with a sterile loop after alignment with the autoradiograph, and grown overnight in Luria broth (+ ampicillin). An aliquot from the overnight bacterial culture was streaked out onto Luria agar (+ ampicillin) plates and incubated overnight at 37° to produce single colonies (plates then stored at 4° for up to 3 months). 0.5ml of the overnight culture was mixed with 0.5ml sterile 40% (v/v) glycerol/Luria broth in a half dram vial and stored at -80° as a permanent stock.

2.7.3 Plasmid DNA preparation

A modified alkaline lysis method was followed (Ish-Horowicz and Burke, 1981).

2.7.3.a Small-scale preparation

3ml of Luria broth (+ ampicillin) was inoculated with a colony picked from a fresh selective plate and shaken overnight at 37° in a 10ml culture tube. 1.5ml of the culture was transferred to an eppendorf tube and centrifuged for 1 minute to pellet the cells. This step was repeated once. Cells were resuspended in 100µl lysis solution (25mM Tris-HCl (pH 8.0), 10mM EDTA, 50mM glucose, 1mg/ml lysozyme (added fresh)) and left on ice for 10min. 200µl of 0.2N NaOH, 1% SDS was added, the solution mixed and left on ice for 5min. 150µl 3M KAc was then added, the tube mixed by inversion, and left on ice for a further 10min before centrifugation to pellet the cellular debris and chromosomal DNA (white precipitate). The supernatant was carefully removed and filtered through a cut-off 1ml disposable tip containing polyallomer wool into a fresh eppendorf tube to

remove all traces of precipitate. Nucleic acid was ethanol precipitated twice, washed in 80% ethanol, and redissolved in 40µl dH_2O. RNase A treatment (20µg/ml) was incorporated in restriction endonuclease digestions at 37°. This protocol generally yielded 1-4µg plasmid DNA.

2.7.3.b Large-scale plasmid DNA preparation

This was a scaled up version of the above procedure. Overnight cultures were used to seed up to 100ml Luria broth (+ ampicillin) in a conical flask which was shaken overnight at 37°. The culture was centrifuged at 7,000rpm for 10min (RT). As in (a) three solutions were added (volumes for 50ml culture : 3ml, 6ml, 4.5ml respectively) with incubations on ice. After centrifugation at 10,000rpm for 10min (4°) the supernatant was filtered through polyallomer wool into a 30ml Corex tube. Isopropanol (0.5 vol) was added, the tube mixed, and left at RT for 10min before centrifugation at 10,000rpm for 10min (4°). After an 80% ethanol wash the pellet was redissolved in dH_2O (1.5ml for 50ml culture). NH_4Ac was added to 3.75M, the tube left at RT for 10min to precipitate contaminating proteins. Alternatively DNA was cleaned by a phenol/chloroform extraction. DNA was then precipitated with ethanol and redissolved in dH_2O (500µl for a 50ml culture).

2.8 M13 DNA SEQUENCING

2.8.1 Subcloning into M13

DNA fragments to be sequenced, and appropriate M13mp18 or M13mp19 vector DNA fragments, were isolated from preparative gels (section 2.4.5.b) and ligated as described in section 2.2.8. Alternatively, 'random' DNA fragments were 'shotgun' cloned (Sanger *et al.*, 1980) into M13mp18 by one of two procedures :

2.8.1.a Sonication

5-10µg of the DNA fragment of interest was purified and self-ligated overnight (to avoid over-representation of fragment ends among the clones). DNA was then sheared in a sonicating bath (Kerry Ultrasonics Ltd.). Sonications were carried out in 15 second bursts. Successful sonication generated a mist of fine droplets on the inside of the tube, which were collected by centrifugation between bursts. The degree of shearing was monitored on a mini-gel, and sonication continued until a size range of 100-1200bp was obtained. Two size fractions (from 600-1000bp and 1000-1200bp) were selected on a

preparative gel and end-repaired (section 2.2.5). Samples were extracted with phenol/chloroform and ethanol precipitated before ligation into an M13 vector, linearised with *Sma*I and treated with Calf intestinal phosphatase.

2.8.1.b Partial digestion

Aliquots of DNA fragment were digested for 10 minutes with a concentration series of the required restriction enzyme designed to give a range of partial digestion products (e.g., 200ng fragment would be incubated with 5, 1, 0.5, 0.1, 0.05, and 0.01 units of enzyme). A sample from each was analysed on a mini-gel and partially digested aliquots pooled, size fractionated on a preparative gel, and ligated into a suitably prepared M13 vector. Combinations of *Alu*I partials into *Sma*I cut vectors, and *Sau*3AI partials into *Bam*HI cut vectors were generally used to give an informative series of clones.

2.8.2 Transformation of *E.coli*

2.8.2.a Competent cells

A 3ml culture of *E.coli* JM101 or NM522 was grown O/N in LUB at 37°. 50ml LUB was seeded with 0.5ml of overnight bacterial culture and grown at 37° to mid-log phase (A_{600} 0.45 O.D. units). 1.5ml cells were transferred to an eppendorf tube and centrifuged briefly. The cell pellet was gently resuspended in 0.5ml MR (10mM MOPS (pH7.0), 10mM RbCl) and centifuged for 30 seconds. Cells were then gently resuspended in 0.5ml MRC (100mM MOPS (pH6.5), 10mM RbCl, 50mM $CaCl_2$) and left on ice for 30 minutes. After a 30 second centifugation the cells were finally resuspended in 150µl MRC and held on ice.

2.8.2.b Transformation

5µl of ligated DNA (containing 5-20ng vector) and 3µl DMSO were added to 150µl aliquots of cells and gently mixed by pipetting. DNA was adsorbed onto the cells on ice for 1hr. Cells were heat shocked at 55° for 35 seconds, cooled on ice for one minute, and then held at RT. 200µl log-phase cells, 25µg BCIG (in dimethylformamaide), and 25µg IPTG (in dH_2O) were added to the cells, the mixture transferred to small glass tubes containing 3ml BTL (held at 50° in a heating block), and poured onto BLA plates, avoiding air bubbles. On drying, plates were incubated at 37° overnight.

2.8.3 Screening M13 plaques

Nitrocellulose lifts were made as for λ libraries (Benton and Davis, 1977, described in section 2.6.3.a). Positive plaques were picked with a sterile Pasteur pipette into 500µl λ buffer and stored at 4°.

2.8.4 M13 DNA preparation

An overnight culture of JM101 or NM522 was diluted 1:100 into fresh LUB. 1.5ml of this dilution was infected with 0.1ml of phage suspension and incubated on an angled rack for 5-6hr at 37°. The culture was transferred to an eppendorf tube and centrifuged for 2x 5 minutes (with a 180° twist) to pellet cells. Double-stranded replicative form DNA was prepared from the pellet (as described in section 2.7.4.a) to confirm the structure of the subclone and purify inserts for oligolabelling. Single-stranded DNA sequencing template was prepared from the supernatant, 1ml was transferred to a fresh tube, 300µl 10% PEG 6000, 2.5M NaCl added, and phage particles precipitated by incubation at 4° for 15 minutes. After a 10 minute centrifugation the small white phage pellet (absent from uninfected controls) was drained of all PEG/NaCl with a drawn-out Pasteur pipette and resuspended in 100µl 1.1M NaAcetate (pH7.0). DNA was prepared by a vigorous phenol/chloroform extraction, followed by an ethanol precipitation and 80% ethanol wash. The DNA pellet was vacuum dried, dissolved in 25µl dH_2O, and a 2µl aliquot analysed on a mini-gel.

2.8.5 Sequencing reactions

The dideoxynucleotide chain termination procedure of Sanger *et al.* (1977) was followed using either the Klenow fragment of *E.coli* DNA polymerase or T7 DNA polymerase. Reactions were carried out in either eppendorf tubes or micro-titre trays (for large numbers of clones). M13 sequencing primers used were either the universal 17mer or a 15mer (TGCAGCACTGACCCT) further away from the polylinker, which enabled sequence to be read across the cloning sites.

2.8.5.a. Klenow reactions

Single stranded M13 recombinant DNAs were incubated at 60° for 10min, then placed at RT. 5µl of each clone was mixed with 5µl of primer mix (150ng/ml primer, 20mM Tris-HCl (pH 8.0), 20mM $MgCl_2$) and annealed at 60° for 1hr. 2µl DNA was then added to each of 4 tubes containing 2µl of each nucleotide mix in 10mM Tris.HCl (pH 8.0), 0.1mM EDTA :

T mix	125μM dCTP, 125μM dGTP, 6.25μM dTTP, 62μM ddTTP
C mix	125μM dTTP, 125μM dGTP, 6.25μM dCTP, 40μM ddCTP
G mix	125μM dTTP, 125μM dCTP, 6.25μM dGTP, 80μM ddGTP
A mix	125μM dTTP, 125μM dCTP, 125μM dGTP, 12μM ddATP.

2μl Klenow mix (containing 0.5U Klenow polymerase, 0.7μl α^{35}S-dATP (10μCi/μl) in dH_2O; prepared on ice immediately before use) was added to each tube, the contents gently mixed and incubated at 37° for 20 minutes. 2μl of sequence chase mix (10mM Tris-HCl (pH8.0), 0.1mM EDTA, 0.25mM dNTPs) was then added, and tubes incubated at 37° for a further 20 minutes. 4μl of formamide dye mix (stock : 10ml deionised formamide, 10mg xylene cyanol FF, 10mg bromophenol blue, 0.2ml 0.5M EDTA (pH 8.0), stored at -20°) was added to terminate the reactions. Reactions could be stored at -20° for up to one week.

2.8.5.b T7 polymerase reactions

To anneal primer and template DNAs, 7μl of each clone DNA was mixed with 1ng primer and 2μl sequenase buffer (40mM Tris-HCl (pH 7.5), 20mM $MgCl_2$, 50mM NaCl), placed at 65° for 2 minutes and cooled to RT over 30 minutes. To each clone was added 1μl 0.1 dithiothreitol, 0.5μl α^{35}S-dATP (10μCi/μl), 2μl labelling mix (1.5μM dTTP, 1.5μM dCTP, 1.5μM dGTP), and 2U T7 DNA polymerase. 5x labelling mix was used to read sequences beyond 300 nucleotides. Reactions were gently mixed and incubated at 37° for 5-10 minutes. 3.5μl of each reaction was then added to 2.5ml of each ddNTP mix and incubated at 37° for 10-20 minutes (chain termination). Dideoxyribonucleotide termination mixes were 80μM for each of the four dNTPs and 8μM for the specific ddNTP in each mix in 50mM NaCl. 4μl formamide mix was then added to each reaction.

2.8.6 Sequencing gels

2.8.6.a. Preparation of the gel

One large (330x420x3.5mm) and one small (330x395x3.5mm) glass plate were thoroughly washed with detergent, rinsed with dH_2O, and wiped with ethanol to ensure both plates were dust-free. The smaller plate was coated with dimethyldichlorosilane and rinsed with dH_2O after 10 minutes. Cleaned and greased (Apiezon AP101, May and Baker Ltd.) 0.4mm spacers (Gibco, BRL) were placed between the plates, the sides and base taped, and the sides clamped with bulldog clips. If not for immediate use the open

end was covered with cling-film.

0.5x and 2.5x TBE acrylamide stocks were used for buffer gradient gels (1x TBE is 100mM Tris base, 80mM boric acid, 3mM EDTA). Both 0.5x and 2.5x stocks were 6% acrylamide (19:1 acrylamide:bisacrylamide) and 8M urea. The 2.5x stock was supplemented with 5% (w/v) sucrose and 100µg/ml bromophenol blue. Stocks were polymerised by the addition of APS (ammonium persulphate, made fresh) and TEMED. 0.42ml of 10% (w/v) APS, 28µl TEMED were added to 90ml of 0.5x stock; and 0.175ml APS, 12µl TEMED to 20ml 2.5x stock. 8ml 0.5x stock and 12ml 2.5x stock were drawn into a 20ml syringe and poured carefully into the mould, avoiding air bubbles. For extension gels (with an altered buffer gradient designed to resolve sequence beyond 250-300nt from the primer) 14ml 0.5x stock and 6ml 2.5x stock were used. For both types of gel this was followed by 60ml 0.5x stock, poured carefully down either side to create an even gradient. The blank edge of the comb was placed 4mm into the gel to form a slot and the top of the mould clamped.

2.8.6.b Electrophoresis

The comb and the tape at the base of the mould were removed and the gel placed in a vertical electrophoresis tank. The well slot was flushed out with 0.5x TBE, with which the upper compartment was filled, and after any leaks were sealed with grease, the lower compartment filled with 2.5x buffer (1.5x for extension gels). The shark's tooth edge of the comb was inserted until the teeth just penetrated the gel surface to create 3mm wells. The gel was pre-run at 2000V to warm up the plates. Samples were either boiled (with caps off) or placed at 80° for 3 minutes, and 2.5µl loaded per slot (clones ordered T,C,G,A). The voltage was reduced to 1200V and adjusted to maintain the plates at appoximately 55° until the bromophenol blue dye front had run off the bottom of the gel (about 3hr). Extension gels were run for at least 2hr after the xylene cyanol front had run off (for 6-10hr).

After the run, the gel was removed from the tank and the siliconised plate lifted off. The gel was fixed in 10% methanol, 10% acetic acid (v/v) for 15 minutes, drained and transferred onto a sheet of 3MM paper. Gels were dried in a vacuum gel dryer (Bio Rad) for 1-2hr at 80° and autoradiographed with Fuji RX 100 X-ray film for 1-10 days.

2.9 POLYMERASE CHAIN REACTION

Enzymatic amplification of minisatellite alleles was carried out using the polymerase chain reaction (Saiki *et al.*, 1988). For all PCR experiments fresh solutions in dH_2O were used and extra care was taken to avoid any reagent or sample contamination.

2.9.1 Preparation of primers and buffer

Flanking 5' and 3' synthetic 24mer oligonucleotides were recovered by ethanol precipitation in dH_2O and diluted to 10μM (79.2μg/ml). PCR reaction buffer was prepared as a 10x stock : 0.5M Tris-HCl (pH 8.8), 123mM ammonium sulphate, 50mM $MgCl_2$, 74mM β-mercaptoethanol, 50 μM EDTA (pH 8.0), 11.1mM dATP, 11.1mM dCTP, 11.1mM dGTP, 11.1mM dTTP, 1.25mg/ml BSA. Primers and reaction mix were stored at -20°.

2.9.2 Amplifications

Amplifications were carried out on either a Perkin Elmer DNA thermal cycler (Perkin Elmer Cetus, Connecticut, USA), or a Techne programmable dri-block (Techne, Cambridge, England). Reactions were in 7-100μl volumes (in 0.3ml eppendorf tubes) as described by Jeffreys *et al.* (1988b). Input genomic DNA varied in concentration from 500ng to 20pg; for very low concentrations (<1μg/ml) DNA was diluted in 5mM Tris-HCl (pH7.5) containing 0.1μM primers as carrier. For reactions, primers were added to a concentration of 1μM, and PCR buffer to 0.9x; Taq polymerase (Anglian Biolabs) was then added (1.5 units for 7 and 10μl reactions). The solution was mixed by pipetting, and overlaid with a drop of paraffin oil to prevent evaporation; after a brief centrifugation tubes were placed in the heating block. Typical cycle parameters were 96°, 1.2min (strand separation); 60°, 1min (primer annealing); 70°, 5min (primer extension). Cycle number varied from 14-25 depending on the experiment. Amplifications were followed by an additional annealing and extension step to remove single stranded product. Final reaction products were stored at 6° or -20°.

2.9.3 Analysis of PCR products

Reactions were generally analysed by electrophoresis and Southern blot hybridisation (section 2.4). Paraffin oil was either removed from the sample before electrophoresis by two ether extractions, or samples were directly collected and loaded from under the oil. DNA hybridisations were carried out as described in section 2.5.2.a.

2.9.4 Oligonucleotide amplification using PCR

Synthetic oligonucleotides were used to create repetitive probes. Complementary overlapping oligonucleotides, containing up to four copies of the repeat of interest, were recovered by ethanol precipitation. The oligonucleotides (2μg each) were annealed at 60°

for 30min, and kinased as described in section 2.2.3. The annealed phosphorylated oligonucleotides were then self-ligated at 4° overnight to provide a minimal substrate for amplification (2.2.8). Amplification was carried out as described above, with extensions at 70° reduced to 2min. DNA was run out on low gelling temperature agarose and a fraction of high molecular weight product excised for oligo-labelling (2.5.1.a).

2.10 COMPUTING

DNA sequences were analysed on a VAX 8650 mainframe computer using programs developed at the University of Wisconsin (Devereux *et al.*, 1984). EMBL and GENBANK DNA sequence databases were scanned using the FASTN programme (Lipman and Pearson, 1985). Statistical programs were written by A.J.Jeffreys on a BBC micro-computer.

2.11 CONTAINMENT

All experiments described in this thesis were carried out with reference to the Genetic Manipulation Advisory Group guidlines on safety and containment conditions. Handling of radioisotopes followed University rules in accordance with Government guidelines. Regulated procedures on mice were carried out under a personal licence (PIL 40/01384) granted by the Home Office, under the Animals (Scientific Procedures) Act of 1986.

III. *MS6-HM* : IDENTIFICATION, CLONING AND STRUCTURE

3.1 Introduction

Many minisatellite loci in the mouse genome cross-hybridise to human DNA fingerprinting probes 33.6 and 33.15. The resulting patterns can be genetically dissected using panels of recombinant inbred (RI) strains to investigate linkage, allelism and germline stability of the minisatellite fragments resolved in these DNA fingerprints. DNA fingerprint analysis of the BXD RI strains defined 13 variable loci in the progenitor strains (C57BL/6J and DBA/2J), eight of which were found to be dispersed over five mouse autosomes (Jeffreys *et al.*, 1987). Of these loci only three were scored in both progenitor strains, and the majority were represented by a single fragment present in only one progenitor, which segregated to approximately half of the BXD RI strains. In contrast to the germline stability of these minisatellites, one locus, termed *Ms6-hm*, was observed to be highly unstable.

3.2 Identification of a highly unstable minisatellite

A 7kb *Hin*fI fragment detected by probe 33.6 in C57BL/6J DNA was absent from 25 different BXD RI strains. Instead, many of the BXD strains carried DNA fragments of corresponding intensity which varied in length from 5-13kb. These fragments were seen in neither progenitor nor other RI strain DNAs. It was proposed that these fragments defined a highly unstable locus at which new length alleles had arisen by mutation during the breeding of the BXD RI strains (Jeffreys *et al.*, 1987).

Restriction endonucleases cleaving outside the minisatellite were used to investigate allelism. Alleles at one locus will show a characteristic 'signature' which is independent of allele length and determined by the location of the target sites for these enzymes in the flanking DNA. The same relative sizes for different fragment lengths were observed with endonucleases *Hin*fI, *Alu*I, and *Sau*3AI (Fig.3.1). This result strongly suggested that the variable fragments detected by probe 33.6 were derived from a single hypermutable minisatellite locus, termed *Ms6-hm*. The large number of alleles at *Ms6-hm* in different DNA fingerprints precluded both the identification of a DBA/2J allele and the establishment of a strain distribution pattern (SDP) across the RI strains.

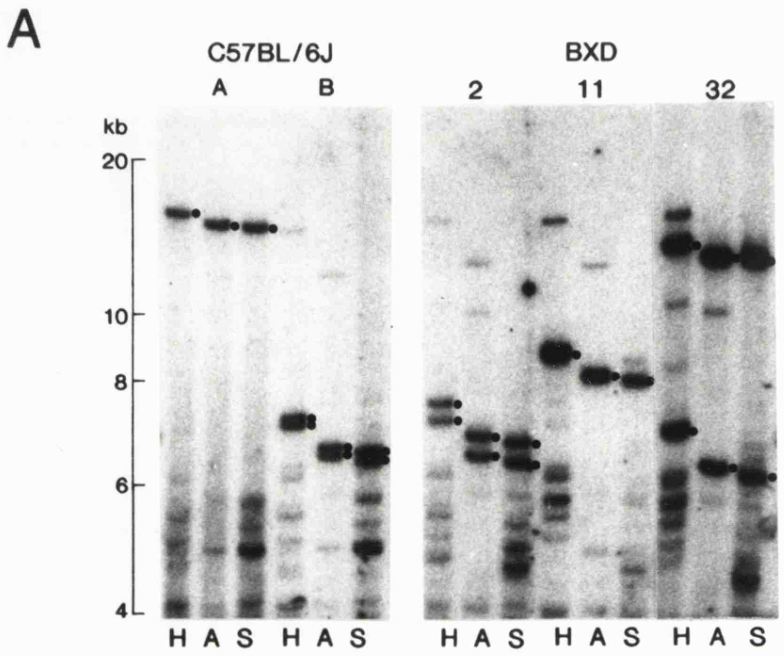
Multi-allelism at *Ms6-hm* was also observed among C57BL/6J mice. While most *Hin*fI alleles were in the size-range 6-8kb, two mice with very large (16kb) alleles were

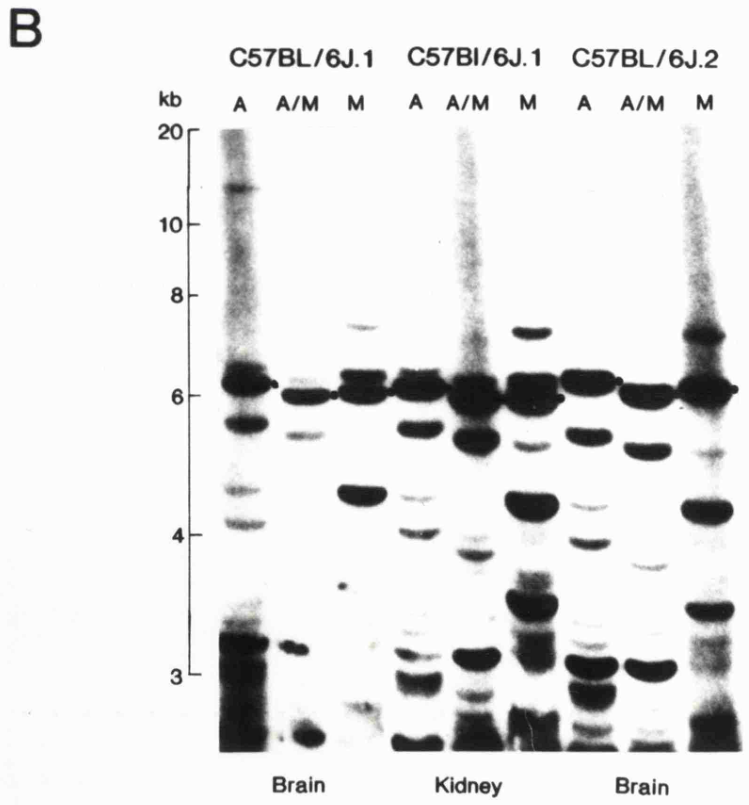
Figure 3.1

Identification of a highly unstable minisatellite in mouse DNA.

A. Characteristic 'signature' patterns of alleles derived from *Ms6-hm* in inbred mouse DNA digested with *Hin*fI (H), *Alu*I (A), and *Sau*3AI (S), and detected by human minisatellite probe 33.6 at low stringency. Alleles at *Ms6-hm* are indicated by dots. Note that C57BL/6J B, BXD 2 and BXD 32 each carry two detectable *Ms6-hm* fragments and are presumably heterozygous at this locus. C57BL/6J A has a large (16kb *Hin*fI) allele at *Ms6-hm*.

B. DNA from C57BL/6J mice digested with *Alu*I (A) and *Mbo*I (M, an isoschizomer of *Sau*3AI) and hybridised with probe 33.6 at low stringency. Alleles at *Ms6-hm* (marked •) appear to be somatically stable in brain and kidney DNA from C57BL/6J 1. Note that the double digest produces a fragment smaller than either single digest, suggesting that *Alu*I trims a *Sau*3AI fragment carrying *Ms6-hm* by 50-100bp.


A
C57BL/6J
A
B
BXD
2
11
32
kb
20
10
8
6
4
H A S H A S
H A S H A S H A S


B
C57BL/6J.1
C57Bl/6J.1
C57BL/6J.2
kb
A A/M M A A/M M A A/M M
20
10
8
6
4
3
Brain
Kidney
Brain

scored. Furthermore, several inbred individuals were found to be heterozygous at this locus (for example, BXD 2 and BXD 32, Fig.3.1), consistent with a high germline mutation rate at *Ms6-hm*. The 7.5kb size difference between the two alleles of BXD 32 indicates that large mutation events can occur at *Ms6-hm*, as these segregating alleles must have arisen from one ancestral allele in a recent generation. In order to investigate the structure, chromosomal location, and processes of mutation at this hypervariable minisatellite, experiments were initiated to clone *Ms6-hm*.

3.3 Cloning *Ms6-hm*

A C57BL/6J mouse whose DNA fingerprint contained a 7kb *Sau*3AI allele of *Ms6-hm* detected by probe 33.6 was identified. This cross-hybridising DNA fragment was purified away from the remainder of the DNA fingerprint by two rounds of preparative gel electrophoresis. 600μg of liver DNA was digested with *Sau*3AI and split into two aliquots for broad size fractionation. DNA fragments from 4-9kb were collected and subjected to a second round of tight size fractionation, resulting in two series of 6 subfractions in the 4-9kb range. These subfractions were electrophoresed and hybridised with probe 33.6 to identify those which contained fragments derived from *Ms6-hm* (Fig.3.2). The degree of purification was calculated by estimating the yield of total DNA in the positive subfractions and comparing positive hybridisation signals to an unfractionated input of known concentration using scanning densitometry. Pooling the positive fractions, a purification of 800x was achieved (yield 30ng DNA).

The enriched fraction was then ligated into annealed *Bam*HI digested λL47.1 arms, packaged *in vitro*, and plated onto *E.coli* L95 (*rec*$^+$). The resulting library of 4500 plaques was screened by hybridisation with probe 33.6. A 7kb fragment purified 800x can be expected to yield approximately 8 positive clones in a library this size (haploid genome genome size of mouse is 3Mkb). 28 positive plaques, of varying hybridisation intensity were picked on the first round of screening. Of these, 19 clones were successfully replated on *E.coli* ED8910 (*rec*BC). DNA from these phage was prepared and analysed on mini-gels and by Southern blot hybridisation with probe 33.6 and total mouse genomic DNA (Fig.3.2).

*Sau*3AI, which recognises the sequence GATC, is expected to cleave every 256bp in random sequence DNA. Thus a library of large *Sau*3AI-resistant fragments would be expected to be enriched for non-random sequences; several of the recombinant inserts isolated from the library described here contained sequences which are highly repeated in the mouse genome (i.e., hybridised strongly to radioactively labelled total mouse genomic DNA). Many of the phage contained contained two recombinant inserts,

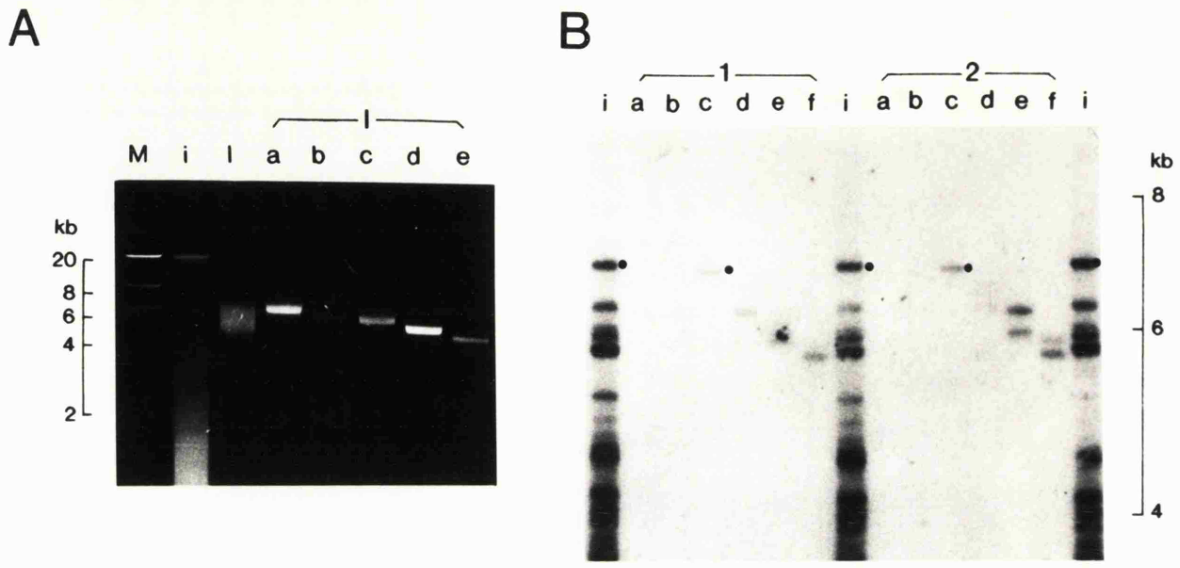
Figure 3.2

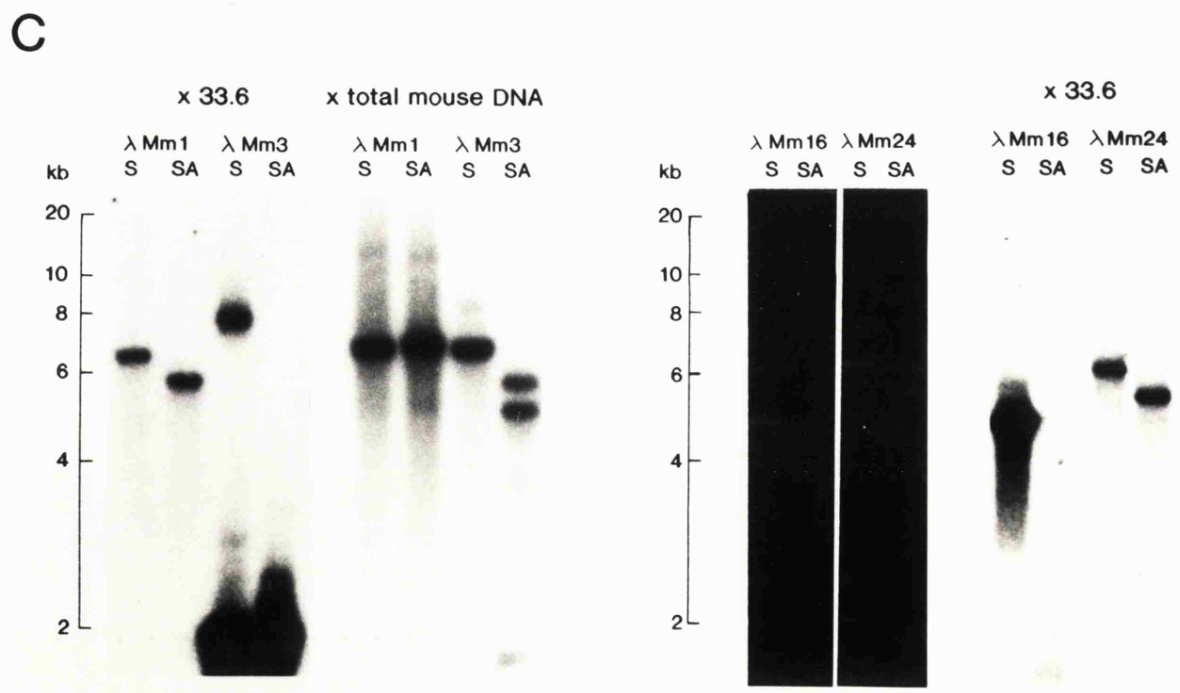
Cloning a *Sau*3AI fragment derived from *Ms6-hm*.

A. Ethidium bromide stained agarose gel of C57BL/6J DNA digested with *Sau*3AI (i) and aliquots of size-fractionated *Sau*3AI digested DNA samples isolated by preparative gel electrophoresis. (I) is a broad size-fraction from 4.5-7kb, and I_a-I_e are tight subfractions from within this size-range prepared by a second round of fractionation.

B. Aliquots of DNA from two duplicate series of size-fractions prepared from *Sau*3AI digested C57BL/6J DNA (1_a-1_f and 2_a-2_f) and 5μg *Sau*3AI digested input DNA (i), hybridised with probe 33.6 at low stringency. Fragments derived from *Ms6-hm* (marked •) are contained in subfractions 1_c and 2_c.

C. Analysis of recombinant phage DNA digested with *Sau*3AI (S) and *Sau*3AI with *Alu*I (SA), and hybridised with human minisatellite probe 33.6 or total mouse genomic DNA. Both λMm1 and λMm3 contain two inserts; λMm1 contains a 6.2kb *Sau*3AI minisatellite insert which cross-hybridises with probe 33.6 and is trimmed to 5.6kb by *Alu*I, and a second insert of 6.5kb which is not trimmed by *Alu*I. This fragment hybridises to total mouse genomic DNA and is therefore highly repeated in the mouse genome. λMm3 contains a 6.5kb *Sau*3AI insert which is highly repeated in the mouse genome, and a second insert of 2kb which hybridises strongly to probe 33.6. λMm16 contains a 5kb *Sau*3AI insert which hybridises to probe 33.6 and disappears with *Alu*I. λMm24 contains a second isolate of the 6.2kb minisatellite fragment present in λMm1.


A
M i l a b c d e
l
kb
20
8
6
4
2
B
i a b c d e f i a b c d e f i
1
2
kb
8
6
4


C
x 33.6
x total mouse DNA
λ Mm1 λ Mm3 λ Mm1 λ Mm3
S SA S SA S SA S SA
kb
20
10
8
6
4
2
λ Mm16 λ Mm24
S SA S SA
x 33.6
λ Mm16 λ Mm24
S SA S SA

presumably as the result of a high insert to vector ratio in the ligation. Inserts derived from *Ms6-hm* were expected to be trimmed by *Alu*I by less than 0.1kb (Fig.3.1). Initially this was not observed in any of the recombinants. Instead many phage contained a 6.2kb *Sau*3AI fragment which hybridised with probe 33.6 and was trimmed to 5.6kb with *Alu*I. On subcloning, this fragment was found to account for 15 of the positive clones. All had stable inserts derived from a minisatellite locus termed *Mm-1*, which is only weakly detected in the 33.6 DNA fingerprint of C57BL/6J mice, and was co-purified with the *Ms6-hm* allele during fractionation. One isolate of *Mm1* carried a second insert which, on hybridisation, detected the 250bp ladder typical of the mouse major satellite in *Alu*I digested genomic DNA (Southern, 1975a). DNA fragments from a second minisatellite locus which cross-hybridised to probe 33.6, *Mm16*, were also co-purified with the *Ms6-hm* allele during fractionation, accounting for 3 of the positive plaques. *Mm-1* and *Mm-16* are described in Chapter 8.

Another class of positive clones had inserts of <7kb which hybridised with probe 33.6. One clone in particular, λMm3, had two *Sau*3AI inserts, one of 2kb which hybridised with 33.6, and a second of 6.5kb which did not hybridise with the minisatellite probe. Since it is unlikely that this 2kb insert could have survived two rounds of size-selection during preparative gel electrophoresis, it must therefore have collapsed from a larger fragment during cloning, the recombinant remaining viable (large enough to be repackaged) due to the presence of the second insert. The 2kb insert of λMm3 was subcloned into pUC13 to generate pMm3, and hybridised to mouse genomic DNA where it detected *Ms6-hm* in the midst of other weakly cross-hybridising fragments (Fig.3.4).

3.4 Instability in *E.coli*

Genetic instability of exogenous tandemly repeated DNA sequences in *E.coli* has been documented previouly for clones of *Drosophila* satellite DNA (Brutlag *et al.*, 1977), mouse rDNA (Arnheim and Keuhn, 1979), and minisatellites (Wong *et al.*, 1986). Rearrangements, predominantly deletions, occur even in recombination deficient strains of *E.coli* (Brutlag *et al.*, 1977, Wyman *et al.*, 1985).

The deletion event from 7 to 2kb in λMm3 may have occurred during initial propagation in *rec*$^+$ *E.coli*. On subcloning into pUC13, and propagation in *E.coli* DH5α (*rec*A), the 2kb fragment generated a heterogenous set of inserts from 2 to 0.4kb. This was observed on mini-gels as a smear below the DNA fragment carrying the minisatellite (Fig.3.3). Restriction endonuclease analysis of two subclones revealed major *Sau*3AI insert fragments of 1.45kb (pMm3-I) and 1.35kb (pMm3-II) which were trimmed by

Figure 3.3

Instability of Mm3 in *E.coli.*

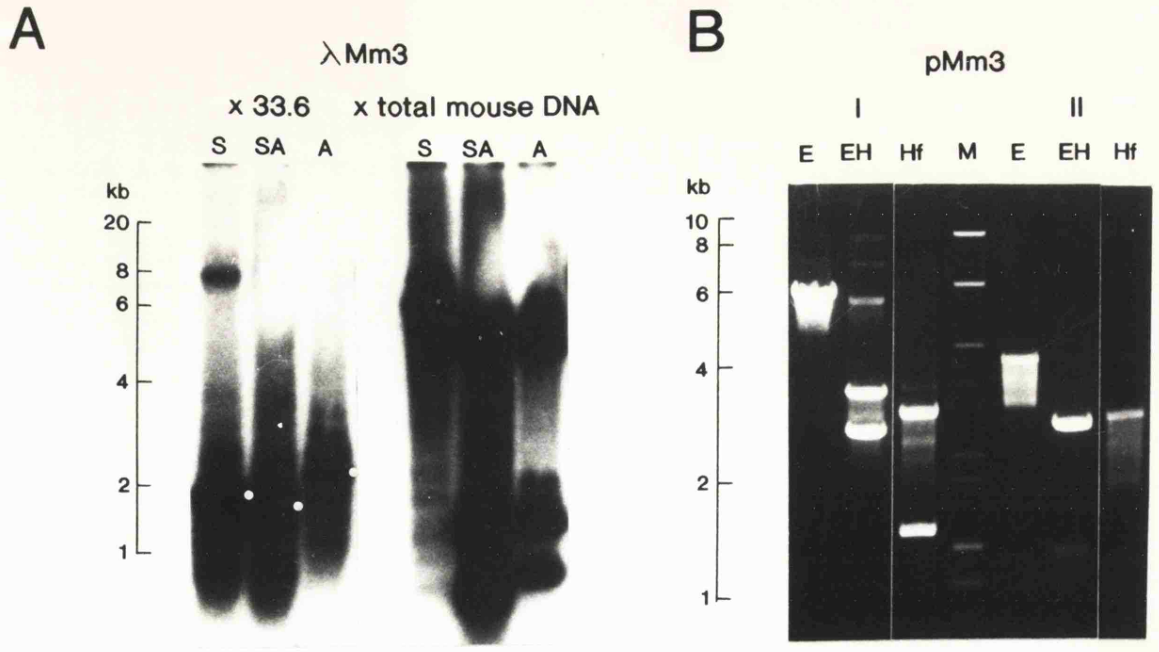
A. λMm3 DNA digested with *Sau*3AI (S) and *Alu*I (A), and hybridised with probe 33.6 or total mouse DNA. λMm3 contains two distinct inserts, a minisatellite fragment and a highly repetitive DNA fragment. The minisatellite fragment (marked by dots) appears to be heterogeneous in length.

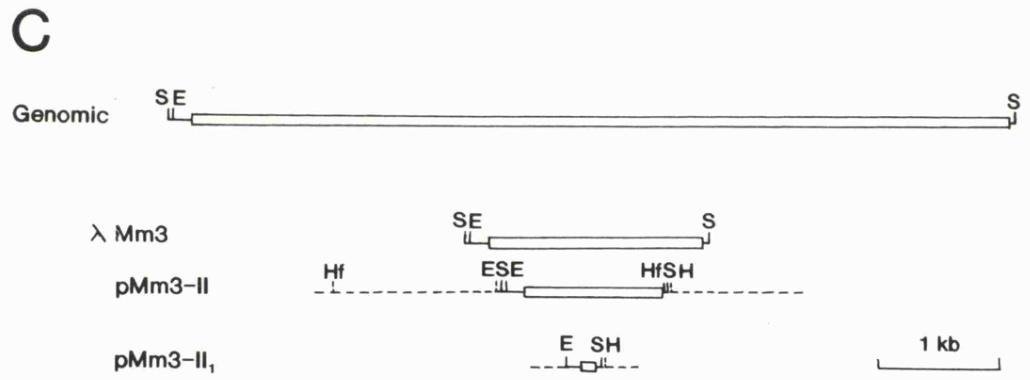
B. DNA prepared from two plasmid subclones containing minisatellite inserts derived from λMm3 digested with *Eco*RI (E), *Hin*dIII (H) and *Hin*fI (Hf), and electrophoresed on an ethidium bromide stained agarose gel. pMm3-I also contains a 1.5kb λ DNA *Sau*3AI fragment. Note the smear below the minisatellite containing fragments. M, λ*Hin*dIII/ΦX174*Hae*III DNA marker fragments.

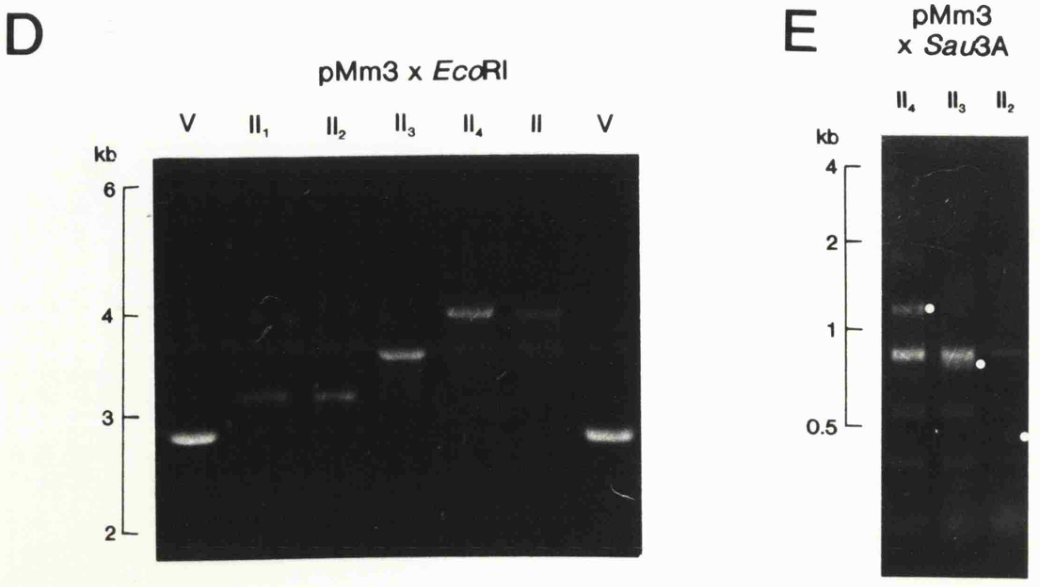
C. Structure of *Ms6-hm* and clone Mm3, determined by restriction endonuclease mapping, illustrating the loss of repeat units on cloning and the derivation of the stable subclone pMm3-II_1. Tandem repeated minisatellite sequence is denoted by an open box, and pUC13 vector sequences by a broken line. S, *Sau*3AI; E, *Eco*RI; Hf, *Hin*fI; H, *Hin*dIII.

D. Deletion derivatives of pMm3-II digested with *Eco*RI and electrophoresed on an ethidium bromide stained agarose gel. V, linearised pUC13 vector; pMm3-II_1 and -II_2, linearised subclones with stable 334bp inserts which hybridised only weakly with probe 33.6. pMm3-II_3 and -II_4, linearised unstable subclones which hybridised intermediately and strongly (respectively) with probe 33.6. Note that in these larger derived subclones, and in the original subclone pMm3-II, the smear below the linearised plasmid stops abruptly at a size corresponding to the insert of subclones pMm3-II_1 and -II_2.

E. Subclones pMm3-II_4, -II_3, and -II_2 digested with *Sau*3AI and analysed on an agarose gel. The minisatellite containing insert fragments are marked by dots. These fragments also contain 120bp of vector sequences, since one flanking *Sau*3AI site was lost on recircularisation (see C).


A
λMm3
x 33.6 x total mouse DNA
S SA A S SA A
kb
20
8
6
4
2
1
B
pMm3
I II
E EH Hf M E EH Hf
kb
10
8
6
4
2
1


C
Genomic
S E
S
λ Mm3
S E
S
pMm3-II
Hf
E S E
Hf S H
pMm3-II1
E S H
1 kb


D
pMm3 x EcoRI
V II1 II2 II3 II4 II V
kb
6
4
3
2
E
pMm3
x Sau3A
II4 II3 II2
kb
4
2
1
0.5

approximately 50bp with *Alu*I. This suggested that the genomic organisation of the flanking DNA was maintained and that rearrangements were confined to the minisatellite repeat region.

In Fig.3.3 it can be seen that the smear below the major insert fragment stops abruptly, corresponding to a *Sau*3AI insert of 400bp. It was proposed that these smallest insert fragments represented flanking DNA with little or no minisatellite. To test this hypothesis, pMm3-II DNA was digested with *Eco*RI to linearise the plasmid, and run on a low-gelling temperature agarose gel. A second *Eco*RI site in the insert resulted in the generation of an additional 60bp fragment. The lowest part of the smear below the linearised plasmid was excised, religated at low concentration to recircularise the plasmid (with the loss of the 60bp fragment), and transformed into *E.coli* DH5α, followed by colony hybridisation with probe 33.6. A range of hybridisation intensities was observed, consistent with heterogeneous insert sizes. DNA was prepared from representative strong, medium and weak cross-hybridising colonies. In Fig.3.3 it can be seen that the intensity of hybridisation is proportional to insert size; two colonies hybridising only weakly to probe 33.6 harbour plasmids with minimal 340bp inserts (pMm3-II_1 and pMm3-II_2).

The stable insert from pMm3-II_1 was force-cloned into M13mp18 and M13mp19. DNA sequence analysis revealed a residual minisatellite consisting of 19 repeats of the unit GGGCA. Significantly, this sequence is similar to the bacterial recombinogenic signal chi (GCTGGTGG, Smith, 1983), and a repeat (CTGG) associated with an insertion/deletion hotspot in the *lac*I gene of *E.coli* (Farabaugh *et al.*, 1978). All 19 GGGCA repeats present in the pMm3-II_1 were identical. DNA sequence analysis of larger plasmid inserts from pMm3-II revealed many more perfect GGGCA repeats. This is consistent with repeat unit loss being responsible for the observed size heterogeneity. Furthermore, the flanking DNA sequence revealed that *Ms6-hm* is contained within a dispersed repeat of the MT (mouse transcript) family (see section 3.8); colinearity with the consensus MT sequence either side of the minisatellite provided additional evidence that the flanking sequences had not rearranged during cloning. Collapse of the genomic allele in *E.coli* therefore appears to be the result of sequential deletions within the minisatellite, causing the loss of almost all of the approximately 1340 repeat units present in the original allele.

3.5 *Ms6-hm*: variability in inbred strains

The stable insert from pMm3-II_1 (hereafter Mm3-1) was excised on an *Eco*RI-*Hin*dIII fragment and used as a hybridisation probe. Under conditions of high stringency

Figure 3.4

Variability at *Ms6-hm* in inbred strains

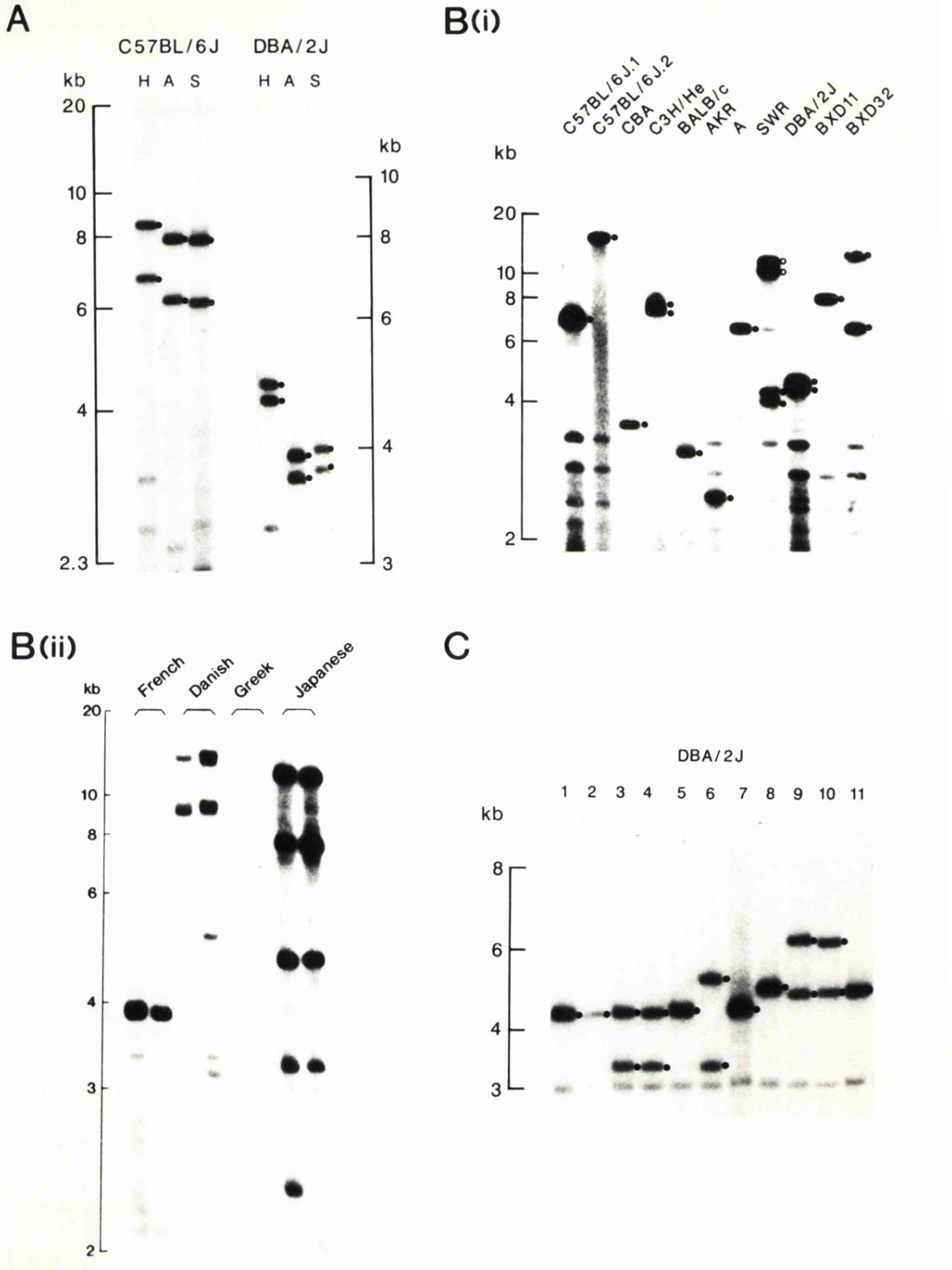
A. *Ms6-hm* alleles detected in C57BL/6J and DBA/2J DNA by hybridisation with Mm3-1 at high stringency. Note that additional weakly cross-hybridising fragments are seen even under conditions of high stringency. Alleles at *Ms6-hm* (marked •) can be identified by the characteristic signature seen in *Hin*fI (H), *Alu*I (A), and *Sau*3AI (S) digests. Both mice are heterozygous at *Ms6-hm*. The DBA/2J signature is distinct from the C57BL/6J signature due to an additional *Alu*I site (see Fig.3.6).

B.

i. Variability at *Ms6-hm* in inbred strains of mice. High stringency hybridisation of Mm3-1 to different inbred strain DNAs digested with *Hin*fI reveals a wide range of allele sizes from 2.5kb (AKR) to 16kb (C57BL/6J 2). Alleles at *Ms6-hm* are marked (•). A second large locus (o) is seen in SWR mice (see Chapter 7).

ii. Variability at *Ms6-hm* in wild mice. High stringency hybridisation of Mm3-1 to different mouse DNAs digested with *Alu*I. The pairs of wild mice were trapped from the same local populations. Note that 4-5 large DNA fragments hybridise strongly to Mm3-1 in DNA from Japanese wild mice.

C. Variability at *Ms6-hm* within inbred strains; high stringency hybridisation of Mm3-1 to 11 DBA/2J mice. Alleles at *Ms6-hm* are marked (•). Mice 1 and 2 were obtained from the same source, mice 3-7 from a second source, and mice 8-11 from a third. At least 6 alleles at *Ms6-hm* are resolved within a range of 3.5-6kb.


A
C57BL/6J
DBA/2J
kb
H A S
H A S
20
10
8
6
4
2.3
kb
10
8
6
4
3
B(i)
C57BL/6J.1
C57BL/6J.2
CBA
C3H/He
BALB/c
AKR
A
SWR
DBA/2J
BXD11
BXD32
kb
20
10
8
6
4
2
B(ii)
kb
French
Danish
Greek
Japanese
20
10
8
6
4
3
2
C
DBA/2J
1 2 3 4 5 6 7 8 9 10 11
kb
8
6
4
3

(0.1x SSC, 65°) this insert detected *Ms6-hm* almost exclusively in mouse genomic DNA, confirming that Mm3 was derived from this hypermutable locus (Fig.3.4). In addition, other weakly cross-hybridising fragments were detected at high stringency. Alleles at *Ms6-hm* can be unambiguously distinguished from these fragments by the locus-specific signature obtained with *Hin*fI, *Alu*I and *Sau*3AI (Fig.3.4, see section 3.2). The cross-hybridising fragments constitute a DNA fingerprint under conditions of low stringency (1x SSC, 65°), generating a pattern distinct from that seen with either 33.6 or 33.15 (see Chapter 7). This cross-hybridisation is due to the tandem repeat unit of *Ms6-hm*, rather than the flanking sequences, since an indistinguishable DNA fingerprint is generated by a synthetic $(GGGCA)_n$ probe. Chapter 7 examines the genetic and molecular properties of this novel DNA fingerprint.

A wide size range of alleles at *Ms6-hm* across different inbred strains was observed (Fig.3.4), from 2 to 16kb (*Alu*I alleles). Each strain tends to have a specific size range, indicating that the majority of mutation events at *Ms6-hm* involve small length changes. AKR mice have the smallest alleles of the strains investigated (200-500 repeat units), and C57BL/6J the largest (>1000 repeat units). In some mice additional variable fragments were detected at high stringency, for example the large locus seen in SWR mice, and the strongly hybridising bands present in Japanese wild mice (Fig.3.4). These may represent independent amplifications of the sequence GGGCA at different genomic locations and are examined in more detail in Chapter 7.

There is also much variation within inbred strains at *Ms6-hm*, as illustrated in Fig.3.4. At least six different alleles can be resolved within the size range 3.5-6kb among 11 DBA/2J mice obtained from three different sources, and 5 of these mice are heterozygous at *Ms6-hm*. These observations are consistent with a high rate of germline mutation at *Ms6-hm*.

3.6 Genomic organisation of *Ms6-hm*

Mm3-1 was used to construct a genomic map of nearest neighbour restriction endonuclease target sites around the minisatellite. Emphasis was on a comparative map between C57BL/6J and DBA/2J strains in order to detect polymorphisms in the flanking DNA. These would be used as stable genetic markers to obtain an SDP across the BXD recombinant inbred strains in order to map *Ms6-hm* in the mouse genome, and also to investigate the mechanism of mutation at *Ms6-hm*. Furthermore, a detailed map would facilitate the isolation of a larger DNA fragment derived from this locus.

Examples of gels used to construct such a map are shown in Figs.3.5 and 3.6. A C57BL/6J mouse with a small *Ms6-hm* allele was chosen to facilitate accurate mapping.

Figure 3.5

A genomic restriction map around *Ms6-hm*

A. Two examples of gels used to generate a genomic restriction map of nearest neighbour endonuclease target sites around *Ms6-hm* in C57BL/6J and DBA/2J DNA using probe Mm3-1. The C57BL/6J mouse has three alleles at *Ms6-hm* of 15, 12.5 and 8kb (with *Hin*dIII, see Chapter 6); the DBA/2J mouse is heterozygous for two *Ms6-hm* alleles of approximately 8kb (with *Hin*dIII). Fragments derived from a second locus cross-hybridising to Mm3-1, which is common to both strains, are marked (•). H, *Hin*dIII; E, *Eco*RI; B, *Bam*HI; P, *Pst*I.

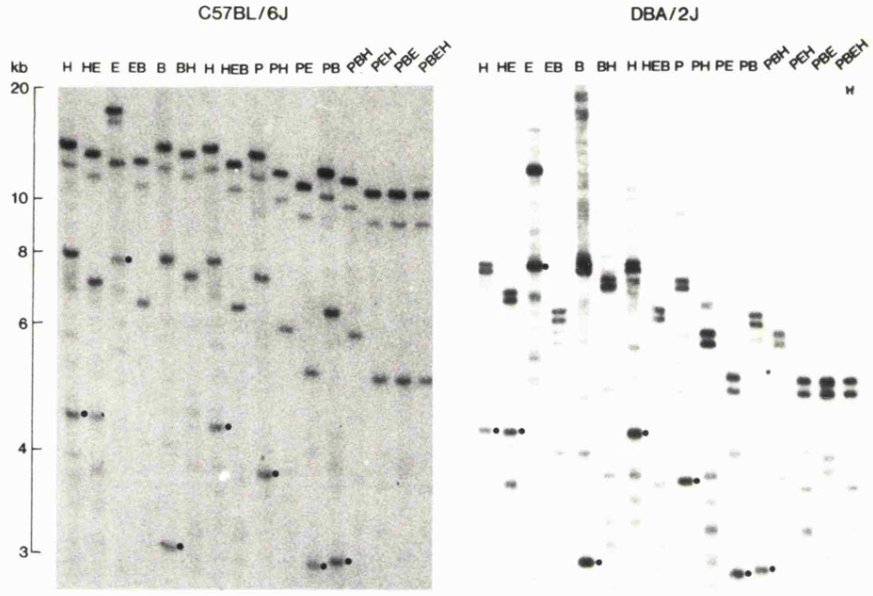
B. Map of nearest neighbour restriction endonuclease sites around *Ms6-hm* in C57BL/6J and DBA/2J DNA. This map is based on gels such as those in A and Fig 3.6. Tandem repeated minisatellite sequence is denoted by an open box.

(i). 6bp target site endonucleases, illustrating a 2kb polymorphism 5' to the minisatellite, detected with *Stu*I (S), and *Pvu*II (P) by Mm3-1, and with *Eco*RI (E) by a *Bam*HI-*Eco*RI flanking probe (see text). P, *Pst*I; B, B*am*HI; H, *Hin*dIII; N, *Nco*I; Sm, *Sma*I; EV, *Eco*RVI.

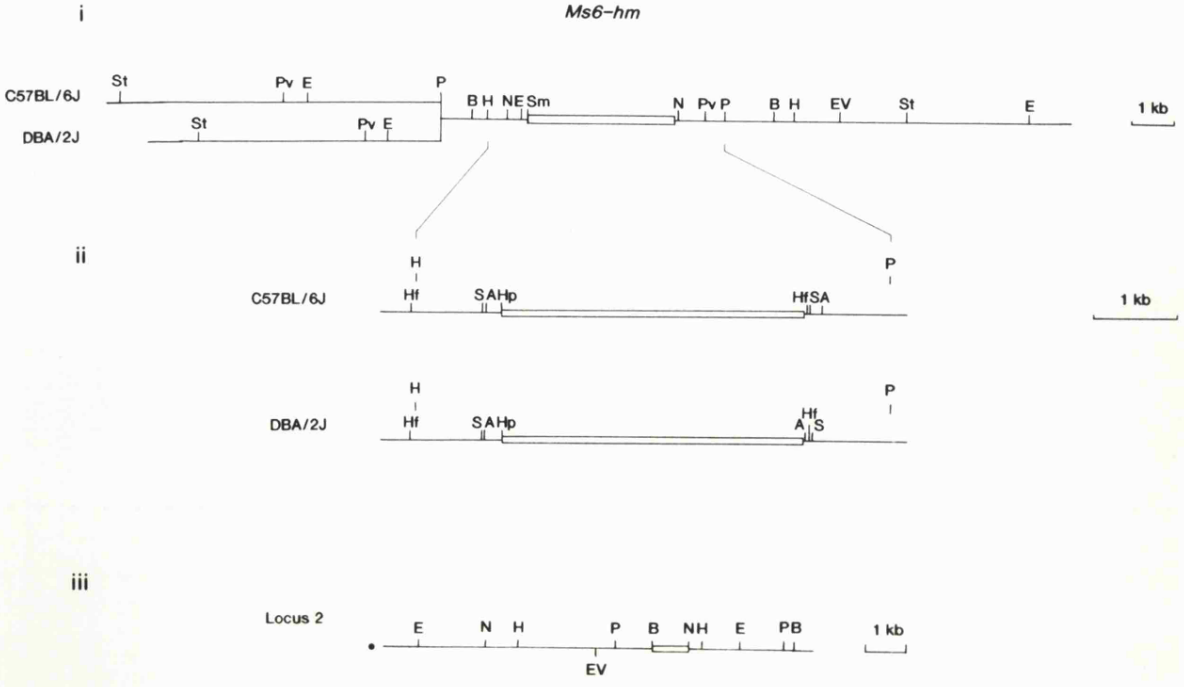
(ii). 4bp target site endonucleases; Hf, *Hin*fI; S, *Sau*3AI; A, *Alu*I; Hp, *Hpa*II. Note the additional *Alu*I site present 3' to the minisatellite in the DBA/2J allele.

(iii). Restriction endonuclease map of a cross-hybridising minisatellite locus (marked (•) in A). A polymorphic *Eco*RV site is indicated below the line.

A


C57BL/6J
DBA/2J
kb
H HE E EB B BH H HEB P PH PE PB PBH PEH PBE PBEH
20
10
8
6
4
3

B


i
Ms6-hm
C57BL/6J
DBA/2J
St
Pv E
P
B H NE Sm
N Pv P
B H
EV
St
E
1 kb
ii
H
P
Hf
S A Hp
Hf S A
C57BL/6J
DBA/2J
1 kb
iii
Locus 2
E N H P B NH E PB
EV
1 kb

As restriction endonucleases with 6bp recognition sites cut less frequently in DNA than those with 4bp target sites, minisatellites will be isolated on fragments containing more flanking DNA; this complicates mapping by bringing strongly cross-hybridising loci, which are normally isolated on fragments too small to be resolved, up into the size range of *Ms6-hm* alleles. One cross-hybridising locus in particular, shared by C57BL/6J and DBA/2J strains, was mapped from these gels, and an *Eco*RV recognition site polymorphism was even detected at this locus. The compound restriction map around *Ms6-hm* (and this cross-hybridising locus) is illustrated in Fig.3.5.

Two polymorphisms linked to *Ms6-hm* were revealed :

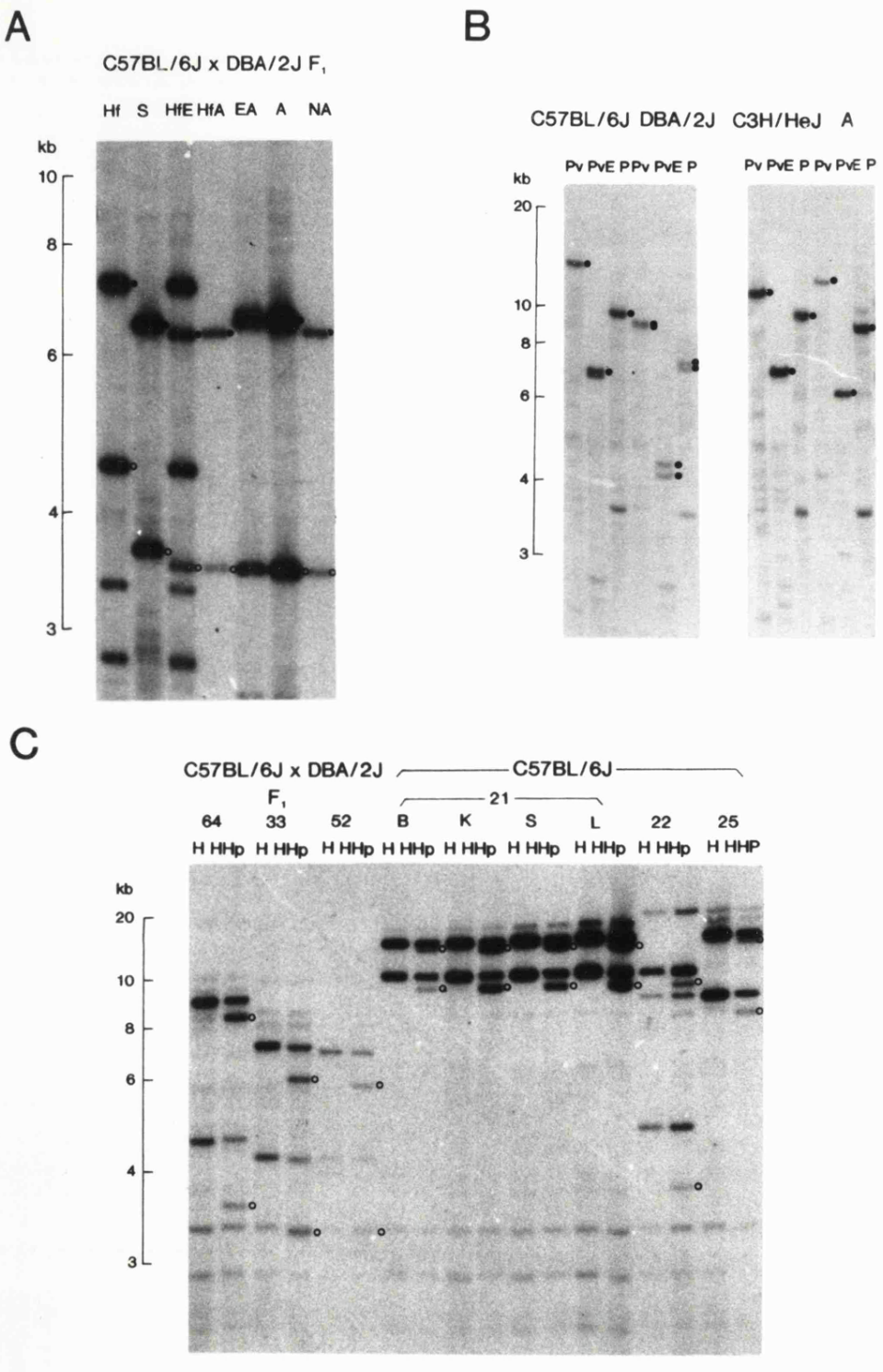
a. A polymorphic *Alu*I site which was present in DBA/2J mice and absent from C57BL/6J mice. The resulting RFLP was used to map *Ms6-hm* in the BXD RI series (Chapter 4). The additional site lies close to the relative 3' end of the minisatellite; DBA/2J *Ms6-hm* alleles therefore have a different *Hin*fI, *Alu*I, *Sau*3AI signature to C57BL/6J alleles (Fig.3.4). Signatures of DNA from other inbred mouse strains revealed that the polymorphic *Alu*I site was also present in SWR *Ms6-hm* alleles, and absent from C3H and BALB/c alleles.

b. An insertion/deletion event 2-3kb 5' to the minisatellite. This was detected as a multi-enzyme polymorphism by *Stu*I and *Pvu*II, and later by *Eco*RI (using a flanking probe, see section 3.7). An insertion of approximately 2kb was seen in C57BL/6J DNA. This insertion was shared by A mice, and absent from all other strains investigated.

The methylation patterns of allelic sites on homologous chromosomes has been shown to vary in an allele-specific fashion in different tissues using sequences adjacent to minisatellite loci in man (Silva and White, 1988). The methylation status of a *Hpa*II site at the 5' boundary of *Ms6-hm* was investigated in C57BL/6J and C57BL/6J x DBA/2J F_1 mice. Although no difference was observed between maternal and paternal homologues, the site appeared to be partially methylated in both alleles (Fig.3.6). This was observed for DNA from different individual animals, and alleles at *Ms6-hm* appeared to be overmethylated in brain DNA by comparison with liver and kidney DNA. It is unclear whether this represents any significant *in vivo* difference.

3.7 Isolation of additional sequence flanking *Ms6-hm*

As clone pMm3 contains only 300bp of DNA sequence flanking the minisatellite, experiments were initiated to isolate a clone carrying more flanking DNA. This would allow further elucidation of the primary structure of the locus and provide a minisatellite-free hybridisation probe for *Ms6-hm*. However two specific problems were anticipated. Any size fraction of DNA from a 6bp target endonuclease digestion (which would be

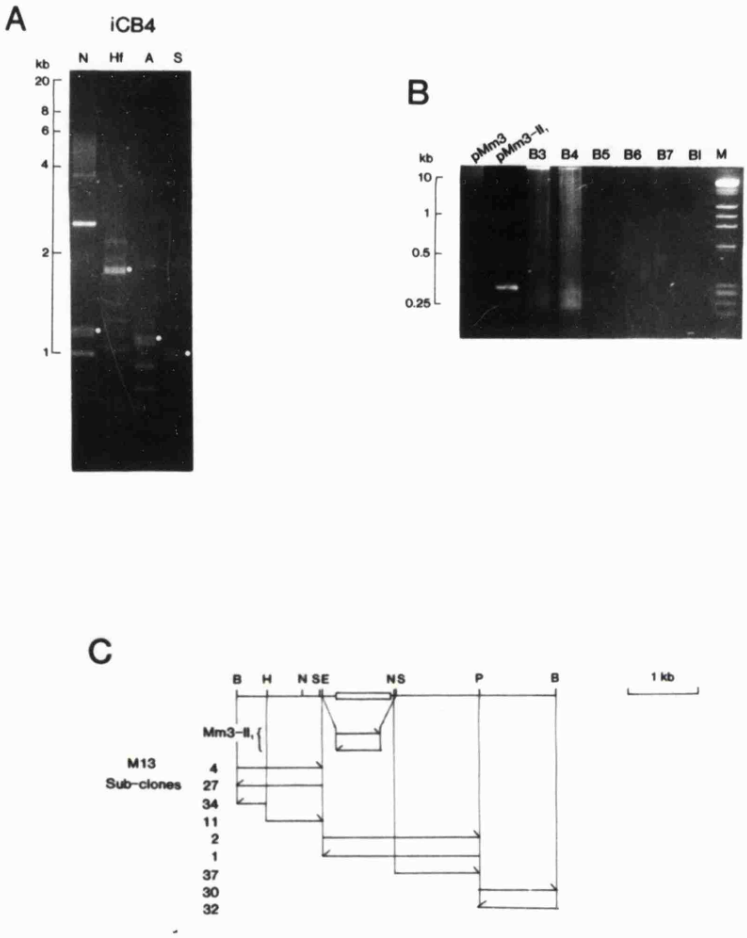

A
C57BL/6J x DBA/2J F1
Hf S HfE HfA EA A NA
kb
10
8
6
4
3
B
C57BL/6J DBA/2J C3H/HeJ A
Pv PvE P Pv PvE P
Pv PvE P Pv PvE P
kb
20
10
8
6
4
3
C
C57BL/6J x DBA/2J
F1
C57BL/6J
21
64 33 52 B K S L 22 25
H HHp H HHp H HHp H HHp H HHp H HHp H HHp H HHp H HHP
kb
20
10
8
6
4
3

ne- cessary to carry more flanking DNA) would be rich in cross-hybridising minisatellite DNA fragments. Screening a library of such inserts with a probe (Mm3-1) containing a mouse dispersed repetitive element and a simple tandem repeat sequence (which is likely to be present in low copy number at many loci) would be expected to yield a high number of false positive clones. By selecting recombinants positive with both Mm3-1 and 33.6 the number of false positives should be reduced. Secondly, if the recombinant insert collapsed, the phage would be too small to package and therefore unable to re-infect *E.coli* cells, and under-represented or absent from the library.

These points were borne out when a λL47.1 library constructed from a tight size fraction around a 15kb *Bam*HI allele of *Ms6-hm* (also digested with *Stu*I to further increase purification, see Fig.3.5) was screened with Mm3-1. Analysis of 12 strong positive clones from a library of 25,000 plaques showed that none of these cross-hybridised with probe 33.6 at low stringency. 4 clones were further investigated, 3 of which were all from the same locus, *B10* (described in Chapter 8), which contains only 4 perfect GGGCA repeats.

An alternative approach was to use a cosmid vector. One advantage of cosmid cloning is that the recombinant phage is only packaged *in vitro* (before any possible collapse) and subsequently replicates inside the cell as a plasmid, where insert collapse is irrelevant. The charomid series of cosmid vectors (Saito and Stark, 1986) have the additional advantage of enhancing selection for a particular size fraction. Charomid 9-36 (capable of accepting inserts of 10-15kb) was therefore digested with *Bam*HI and ligated to a highly purified DNA fraction containing a 15kb *Bam*HI allele of *Ms6-hm* from a C57BL/6J mouse. After packaging, *E.coli* NM554 (which is an *mcr*$^-$ strain and permits the replication of highly methylated exogenous DNA) was infected with the recombinant charomids, yielding a library of 30,000 colonies.

5 clones which hybridised strongly to Mm3-1 were picked, and all 5 hybridised to some extent with probe 33.6. One of these (CB4) was further characterised by restriction endonuclease mapping and Southern blot hybridisation with probe Mm3-1 and total mouse genomic DNA to build up the map illustrated in Fig.3.7. Inserts derived from *Ms6-hm* were expected to be trimmed by 1.2kb with *Eco*RI (see Fig.3.5). CB4 had a 4.4kb *Bam*HI insert which was cleaved to generate two fragments of 3.2 and 1.2kb by *Eco*RI, and the expected *Hin*fI, *Alu*I, and *Sau*3AI signature was observed. As this fragment was derived from a 15kb size fraction it had therefore collapsed on cloning; furthermore, DNA isolated from CB4 was actually heterogeneous for two similarly sized inserts. These data strongly suggested that CB4 was derived from *Ms6-hm*. This was confirmed by amplification of the CB4 insert by the polymerase chain reaction using primers based on the sequence of pMm3 (Fig.3.7). It is interesting to note that CB4 did not collapse with


A
iCB4
N Hf A S
kb
20
8
6
4
2
1
B
pMm3
pMm3-II,
B3 B4 B5 B6 B7 BI M
kb
10
1
0.5
0.25
C
B H N SE NS P B
1 kb
Mm3-II,
M13
Sub-clones
4
27
34
11
2
1
37
30
32
D
C57BL/6J
B BH H HE E BE
kb
20
10
6
4
3
2
1
E
AKR BALB/c CBA C3H SWR C57BL/6J DBA/2J
M F M F M F M F M F M F M F
kb
10
8
6
4
3

the finality of pMm3. Possible reasons for this include the additional flanking DNA which may partially stabilize the insert in *E.coli*, the different host strain, or the different vector.

The cloning of CB4 allowed the development of a minisatellite-free flanking hybridisation probe for *Ms6-hm*. A 1.2kb *Bam*HI to *Eco*RI fragment 5' to the minisatellite was isolated. This was used to confirm the genomic organisation of *Ms6-hm* (Fig.3.7). The expected fragments were observed with *Bam*HI and *Hin*dIII; a background smear is explained by the presence of a dispersed repetitive sequence in the probe (as discussed in section 3.8). A 5.4kb *Eco*RI fragment (from the *Eco*RI site immediately flanking the minisatellite to the next 5' *Eco*RI site) was detected. This was cleaved to a 1.2kb fragment with *Bam*HI. However, an anomalous 2kb fragment was also detected in the *Bam*HI/*Eco*RI double digest (Fig.3.7). This is unlikely to be a partial digestion product as previous hybridisation of this blot with another probe (from *B10*) showed no evidence of partial digest products. Whether sequences other than the dispersed repeat on the 1.2kb *Bam*HI-*Eco*RI fragment are present at another site in the genome is unclear. There is no evidence for additional cross-hybridising DNA fragments in either *Bam*HI or *Eco*RI single digests.

The flanking probe was also used to confirm the strain distribution of the 5' flanking polymorphism, as previous analysis of large *Pvu*II and *Stu*I fragments would not accurately reveal small differences in allele length between strains. Hybridisation of this probe to genomic DNA digested with the endonuclease *Eco*RI revealed either a 5.4kb (C57BL/6J) or 3.4kb (other strains) fragment (Fig.3.7). Since three different restriction endonucleases detect the same 2kb size difference, the polymorphism cannot be due to simple point mutation, but is likely instead to be due to a 2kb insertion/deletion event between 2 and 3.4kb 5' to *Ms6-hm*.

3.8 Primary structure of *Ms6-hm*

Using the restriction endonuclease map described above, fragments from CB4 were force-cloned into M13 vectors for DNA sequencing, to generate the series of subclones illustrated in Fig.3.7. The results from these templates, and data from the original *Sau*3AI clone, are presented in Fig.3.8. The entire flanking sequence from the 5' *Bam*HI site to the 3' *Pst*I site (2.6kb) was determined (only 19 minisatellite repeat units are included in Fig.3.8). Two sections of the *Pst*I to 3' *Bam*HI 1.2kb fragment were sequenced, covering an (A)-rich region and a sequence highly repeated in the mouse genome (see Fig.3.8 and below). Most of this sequence was determined on one strand only, with 10-50bp overlaps between neighbouring clones.

Although the entire minisatellite array of CB4 was not sequenced, no variation

Figure 3.8

Primary structure of *Ms6-hm*

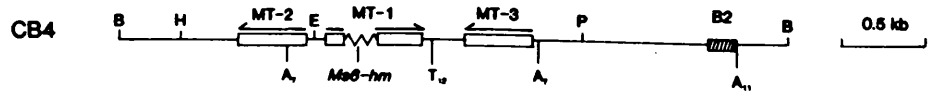
A. Structure of CB4. The position of minisatellite, dispersed repetitive, and poly(A) sequences within the clone are indicated. The arrows above the three MT elements indicate the direction of the strand similar to the consensus sequence of Bastien and Bourgaux, (1987). B, *Bam*HI; H, *Hin*dIII; E, *Eco*RI; P, *Pst*I.

B. The DNA sequence of *Ms6-hm* as determined from the CB4 and Mm3 M13 subclones illustrated in Fig.3.7. Only 19 GGGCA repeats (underlined by arrows) are included in the sequence, starting at nt 1432 (only 19 repeats remain in the stable subclone pMm3-II_1). The approximate positions of the MT elements in the sequence are MT-1, 1300-1775; MT-2, 780-1295 (reverse and complement); MT-3, 2000-2393 (reverse and complement); regions showing similarity to the consensus MT sequence of Bastien and Bourgaux (1987) are underlined.

The (A)-rich region starts at approximately position 2880, and the B2 repeat starts 200bp from the 3' end of CB4 (see text). A gap of about 200bp is indicated (...), starting at nt 3208; note that DNA sequence 3' to this gap is numbered from the 3' *Bam*HI site.

As the CB4 M13 subclones were constructed by a force-cloning strategy it is possible that sites selected for cloning may be clustered in the original sequence. This does not apply to the *Eco*RI and *Nco*I sites used, as these sites were sequenced on other clones (see Fig.3.7). The *Hin*dIII site 5' to the minisatellite could be a cluster of sites, however examination of an extention gel of subclone 4 suggests that this is a single site. Thus only the *Pst*I site 3' to the minisatellite could be a cluster of sites.

A



B

CB4

```
   1 GGATCCCTGG GCTATGTAGT CTATGGTTCT TGTCAGCTAA GCAGTGTTGA GTATGGATTG CATCTCATGG GGTGAGGGTT TAGTTAAGCT AGATATTGGT
 101 TGGTTACTCC CACAAATTTT GTGCCACCAT TGCATTTATG TATCTTACAG GCAGAACATC ATTGTAAATC AAAGATGTTG TGACTGGATT GGTGCTTATG
 201 TTTCTCCTTT GGTAGTGTGA AGAGAACCTT CCTGTACCAA AGACACTAGC ACATATGCGT GAAGGTTCTA TGTAGGTTCC AGCTTGACCT TTCTATGTTC
 301 GATGAGTTGT ATATGTGTTG TCTTCAGCAA TAGGGCTCTG TCTCATTTTG TGGAGAAGAA AGTATAATCT TGCAACAGCC TGGGTTTGAG GATTCCCATA
 401 GGACCTCTTT GGACAATAAC TCAATTAGAT GTAACCCAGT CTTGGTCCTG GAAGCTTCAT TCGGTGACAA AATATGGCTA GTTGGGGCTC TGTCTTCCCC
 501 AATTATTTGG TGGTTTCATT TATATCACCT TGGTATGTGT GCATATTTTA GAAACCATCT ACTATATTGT GTTTCTATAC TATCCCTCCG ATGGCCCTTA
 601 ATTTTAGCTG TCTTTTCCCA TATTCCCTCC CTCTTCCCCA TCTCCACTTG ATCCTCTCAT TCCAGCCATA CCCCCATCCA TCCATAACTA TCTATTCTAC
 701 CTCCCTTTCC TAGGGAAAGC TATCTGACAT GCCTTCCCTG ATACACAGTC CCTTTATCTA AACTCTGTGG TTCTGTGAAT TGTAGTTTGG TTGTTATGCT
 801 TCATCGTCTT ATGTCATTTT GACACAAGCC TCAATTAAGA AATGCCTCCA TAGGATCGGC TGTAGTCAGT CCATTGTCAG GTATTTTCTT ATTTAGTAAT
 901 TGTAGCCCCC TACAAGGTCA TTTGTGTGGG TGGTGCATCC CTGGCCATAT GGTCCTGGGT TCTATAAGAA AGCAGACTGA ACAAGCCATG GGGAGCAAGC
1001 CAGTAAGCAG CACTCTCCAT GGCTTCTATA TCAGCTCCTG CCTCCAGGTT TCTGCCCTGC TTGAGTTCCT CCCTGACTTC TTTGATAATG AACAGTGATA
1101 TGGAAGTGTA AGCCGAATTT TAAAAAAATC TAACCATTGC ATACACATTT TCTACATTTG CCAATGCTTT AGTTCAAAAG ACTAACTGGC AGATCCCAAA
1201 TGCTCAGTTA CACTTGCAAA CTTCTACAGA TAGAGAAGTT TAGCTTGCCT GGAATTCTTC CAAATGATGA CAGTGATGTA AACCAGACTA TATGGCTAAC
1301 TGTTTTAGTT AGAGTTTCTA GTTGCTGTGA CCAACACCAT GACCAAAAAG CAAGTTGGGG AGGAAAGGAT TTATTTGACT TACACTTCCA TATAACTGTT
1401 CATCATCAAA AGAAATCAGG ACAGAAACCC GGGGGCAGGG CAGGGCAGGG CAGGGCAGGG CAGGGCAGGG CAGGGCAGGG CAGGGCAGGG CAGGGCAGGG
1501 CAGGGCAGGG CAGGGCAGGG CAGGGCAGGG CTGATGTAGC GTCACTGAGG AGTCCTGCTT CCTACTTTGC TTCCCATGGG TGGATCAGCC TGCTTTCTTA
1601 TAGATCCCAA AACCAGGGAT ACCACCGCCT ACAATGGCCT GGACCTTCCT CCATTGATCG CTCATTGAGA AAATGCCTTA GAACTGACTC TCATAGAGGC
1701 ATTTCCTCAA CTGAGCCTCC TTCCTCTCTG ATGACTCTAG CTTTTGTCAA GATGACATAA AATCAACCAG TACACTGACC TACCTGGGAT TGTTGTAACT
1801 GTTGCCATAT TGTGTATTTT TTTTTTTTAA AGAAAGAAGT AGAACATATT AGAATTGAAG GGGAGAACGG AAGCTGGTCA AAGCTGTTTC TGTTATACAA
1901 TAAAAACTGT TAGGTATCTG GTAATAGACC ATATTAATGT CTACATTTTT ATTTAATTAC TAAATATATT TAAGTTCAAG AAATTAAGAG AAGCACAGTT
2001 CTCTGTGTGA GTAACTAGGA ATGAGGGATT GTGTCTCATC AAGAGGAAAG AGTAATGTGC TGTCAAATGG TGTATTAATT TCCATTAGGT AGAAGTTGAT
2101 CATTGACCNG AAGGANNGGG TTGTTAGTGT TTTTGTCAAT TGAGACACCT GGGAGAGCAA CTCATTAAGG ATTGCTCAGC TGCCTATGGA AGCNNNTGCG
2201 CATGAGCCAG GAAGTGAGCA GCTCTTCTGA AAATCTTCTG AAATTTCTGC TTCTGTCATT CCTGCTTGGG TTTCTGCTCT CACTTCCCTC AGCAATGAAC
2301 TATGACCTGA AGTGTGAGTC GATAAGCCCT CCCTCTCCTA AGCTGTTTTG GTCATGATGT TTACCACAGC AACAGAAAGC AAACAAGGAC AAATACTTTC
2401 TTCAACTACT AAAAAAATCA ACAGAGGGAT GATGCTAAGA GTATCTGCAC AAGGAGATTT GGGTGTTTTG AACATGCGTT ACAAGTCAGG CTTTCACAAG
2501 AAGAGAACTG AGATTACGTG GAAACTTTGA ATTGGAAGAA GAAACATAAA ATATGCTGTT GGAGACCAGG GCCATAGCTT ACCAGCTAGA AAGATCCTGA
2601 TAGGTATACA GGCCAATATG GACAGAATAC AACCAGGTTG TATTCTGTAT TCTAAAGGCT AGGTTATTAC CTGACAAAAT TGTGTATTTT GTGCCAAGAA
2701 GACCAATTCT ATCTTCTCAC TGCAGGGTTC TACACTTATG TAATAACTAC TTAGCTTGCA GAATGTGCGA GTCAAGTAAC TTCCAGTATC CCTTCTGCAA
2801 GTAAGTAAGT GACTAGTAAC AATCACGTTT TCTAATAGGG TCTCAAGGAG AAAACAAACA CTCTAATAGG GCGTGGCAAG GCAGAACAAC AACAAGAAGA
2901 AGAAGAGGAG TAGGAGGCAG AAAGAAAAAA CAGACAAACA ACAAAAACAA AACAAAACAA AAAGCCCTCA GAACATAGAG AGGTAAGAAC CTTAAAATCT
3001 GCTTTAGACT CAAGGGCACT TACTGACCAC AGCAATAATG CTCTCTTCTC TCCTCTTCTT TCCCCGTGGA AAAGAAGGCT CTTCTACATT TTTTTTGTGG
3101 GCTAGCAACT CATCATATCT TGTACCCTAA ATAGATTCCA GTTATCTGTT AACTAATGTC ATTTGGTTGC TGCTAATACT TAGCATTACT TACAGAGACC
3201 AGCAATTG ...

        ... AGACCTTCTT CTCTCTGTAT TAATTCCTGC TGTGACAATG AATGTCCTAA AATACATTTA AGACCATGTA AATTCATTGT TCTGGAGCTC  451
    ACACGTACAA AAAATAGATC TTACTGGGGC TGGTGAGATG GCTCAGTGGT AAGAGCACCG ACTGCTCTTC TGAAGGTCCG GAGGTCAATC CCAGCAACAC  351
    ATGGTGGCTC ACAACCATCC GTAATGNNAT CTGACTCCCT CTTCTGGAGT GTCTGAAGAC AGCTACAGTG TACTTACATA AAATAAATAA ATAAATAAAT  251
    AAATAAATAA ATCTGGTTTG TTTAAAATAA AAAAAAAAAT AGATCTTACT TTGCTGCAAA TCCCTGCATA GACAGCACAC ATTCCTTTTG GAGACTCTGG  151
    GAATAGTTAT TTTTTATATT TTAGAGTCTA TACATGCTCT TCTGTCCGCC CCACCCCACC TTTCCTACAT TTTAAAACCT AGCTATATTG GGCTAAATCT   51
    GCTTCATATT ATTCTCTCTG ATTGTCTTCC TGTCCCACGT CTAAGGATCC
```

was observed from the 5bp repeat unit GGGCA in over 60 repeat units sequenced (from subclone 2). GGGCA forms the central part of the core sequence common to many hypervariable human minisatellite loci (GGAGGTGGGCAGGARG, Jeffreys *et al.*, 1985). The 36bp repeat unit of probe 33.6 (essentially composed of 3 core-like subunits) contains 3 copies of the permutation AGGGC (underlined below), accounting for the cross-hybridisation between the two loci at low stringency :

33.6 AGGAAGGGCTGGAGGAGGGCTGGAGGAGGGCTCCGG

GGGCA resembles the repeat units of a number of other tandemly repeated loci, presumably due to the independent expansion of similar sequences. Of note among these is a GGCA repeat found at the Eβ hotspot in the mouse major histocompatibility complex (Kobori *et al.*, 1986).

The EMBL and Genbank DNA sequence data bases were searched for entries related to sequences flanking *Ms6-hm*. This search revealed that there are three dispersed repetitive elements of the MT (mouse transcript) family associated with *Ms6-hm* (MT-1, MT-2, and MT-3). MT elements are 400bp long rodent-specific dispersed repetitive sequences which are represented 40-90,000 times in the mouse genome (Heinlein *et al.*, 1986; Bastien and Bourgaux, 1987). The central element (MT-1, from nt 1301-1775, Fig.3.8) shows approximately 75% similarity to the consensus MT sequence of Bastien and Bourgaux (1987) throughout its length, and of the three elements associated with *Ms6-hm* is most similar to the consensus. The minisatellite lies within MT-1, starting 135bp into the consensus sequence; at this point in the consensus there are two copies of the minisatellite repeat unit (GCAGG) separated by 7bp. Beyond the minisatellite MT-1 continues from the second GCAGG repeat. This suggests that *Ms6-hm* has expanded from within MT-1.

The flanking MT elements, MT-2 and MT-3, are orientated in the opposite direction to MT-1. MT-2 (the 5' MT element, from approximately nt 770-1200) shows about 60% similarity to the consensus sequence (Fig.3.9). This similarity increases to 70% if the first 55bp are discounted, a stretch of T_7A_4 preceding the start of strong similarity. The 3' element (MT-3, from approximately nt 2000-2395) shows much less similarity to the consenus. However this is unambiguously an MT element; within the first 135bp the similarity is about 70%. After 135bp there is a 10bp direct repeat (TTTCAGAAGA) beyond which there is no detectable similarity to the MT consensus sequence; this is precisely the point at which the minisatellite has expanded from within MT-1. Further evidence that this represents a particularly labile region within MT sequences is addressed in the Discussion. Pairwise comparison of each MT sequence associated with *Ms6-hm* to the others did not reveal any similarity greater than 65% over

the entire best fit consensus regions. All three are identical at 20/31bp between nt 76 and 105 of the consensus sequence (65%). These data are illustrated in Fig.3.9.

Bastien and Bourgaux (1987) suggest that MT elements have the structural features of retroposons. They found 8-10bp direct repeats with 1-3 mismatches flanking three genomic MT elements at a variable distance 5' (0-5bp) and 3' (9-35bp) to the element. Of the MT sequences associated with *Ms6-hm*, only for MT-1 can short direct repeats (dr) be found. The sequence TAACTGTT overlaps with the first four bp of MT-1, and is found perfectly repeated 20bp 3' to the element. The significance of this is unclear because the integration mechanism proposed for retroposons requires that the dr be of target site sequence only (see Rogers, 1985). None of the three elements at this locus have an obvious poly(A) 3' sequence typical of retroposons. These features are in contrast to the long direct repeats and poly(A) sequences associated with other retroposon sequences in the mouse genome.

A B2 dispersed repetitive element was found 212bp from the 3' *Bam*HI site of CB4, and shows 94% identity to the consensus sequence of Krayev *et al.* (1982) (Fig.3.10). The B2 element is flanked by 11bp perfect direct repeats, and the 3' dr is preceded by an 11bp poly(A) tail. The 3' AT rich block which precedes the poly(A) tail is extended, and contains 8 tandem copies of the unit (AAAT). Thus two adjacent dispersed repetitive sequences contain an amplified simple repeat.

Other sequences of interest within CB4 include an (A)-rich region starting approximately 200bp 3' to the *PstI* site. About 17 repeats of variations on ($A^GC(A)_n$) are found. The sequence $(T)_{12}$ is found 40bp 3' to MT-1; as discussed above such sequences ($(A)_n$) may be associated with the 3' end of retroposons. In fact an 8bp sequence preceding this is repeated 850nt further 3'. The significance, if any, of this, and of several other oligonucleotide direct repeats within CB4, is unclear. Two short sequence elements (AAATGCTCA and GATGTAAA) separated by 70bp lie between MT-1 and MT-2 (Fig.3.10); these are of interest because they are found a similar distance apart adjacent to another mouse minisatellite with the repeat sequence GGGCT which was detected by cross-hybridisation to the terminal repeats of the Herpes Simplex Virus (Gomez-Marquez *et al.*, 1985). Again, the significance of this observation is questionable.

3.9 Summary

The unstable mouse minisatellite *Ms6-hm* was identified as a highly variable locus within the 33.6 DNA fingerprint of C57BL/6J mice. On cloning, Ms6-hm was found to show extensive size variation between inbred strains of mice, and multiallelism and heterozygosity within inbred strains, consistent with a high germline mutation rate.

Figure 3.10

DNA sequence of the B2 dispersed repetitive element within CB4

DNA sequence of the B2 element from CB4 (top line), compared to the B2 consensus sequence of Krayev *et al.*, 1982. The sequences are aligned with gaps (-) to maximise similarity; nucleotides which differ from the consensus are in bold type. The 3' direct repeat starts 200bp from the 3' end of CB4 (see Fig.3.8, note that the numbering in this figure is different to Fig.3.8). The 11bp direct repeats are at nt 7-17 and nt 236-246 (underlined), and the 3' direct repeat is preceded by an 11bp poly(dA) tail. There is a 6bp overlap between the poly (dA) tail and the direct repeat, which, if included, increases the dr to 17bp. Note that the sequence AAAT is repeated 8 times at the 3' end of the element.

```
      1                                                  50
CB4   AAAAAATAGA TCTTACTGGG GCTGGTGAGA TGGCTCAGTG GT-AAGAGCA CC-GACTGCT CTTCTGAAGG TCCGGAGGTC AAT-CCCAGC AAC-ACATGG
cons                    GGG GCTGGAGAGA TGGCTCAGTG GTTAAGAGCA CCTGACTGCT CTTCCGAAGG TCC-GAGTTC AATTCCCAGC AACCACATGG

      101                                                150
CB4   TGGCTCACAA CCATCCGTAA TGNNATCTGA CTCCCTCTTC TGGAGTGTCT GAAGACAGCT ACAGTGTACT TACATAAAAT AAATAAATAA ATAAATAAAT
cons  TGGCTCACAA CCATCCGTAA TGAGATCTGA TGCCCTCTTC TGGAGTGTCT GAAGACAGCT ACAGTGTACT TACATATAAT AAATAAATAA AA

      201                                                250
CB4   AAATAAATCT GGTTTGTTTA AAATAAAAAA AAAAATAGAT CTTACTTTGC
```

Alleles derived from this locus were also unstable in both *rec*$^+$ and *rec*$^-$ strains of *E.coli*, resulting in the deletion of most of the GGGCA minisatellite repeat units. DNA sequence analysis revealed that the minisatellite has evolved by amplification from within a member of the MT family of dispersed repetitive elements. Furthermore, the minisatellite is flanked by two additional diverged MT elements, and a B2 element is found in the 3' flanking sequences. This clustering of mobile repetitive elements, combined with the hypervariability of the minisatellite, suggests that *Ms6-hm* lies in a highly unstable region of the mouse genome.

IV. CHROMOSOMAL LOCALISATION OF *MS6-HM*

4.1 Introduction

The development of a specific DNA hybridisation probe for *Ms6-hm* allows the allocation of this locus to a mouse linkage group. A variety of techniques are used to map loci in the mouse genome, from classical linkage analysis, using intraspecific and interspecific backcrosses, and recombinant inbred strains, through screening mouse-chinese hamster somatic cell hybrids containing whole or partial mouse chromosomes (Franke *et al.*, 1977), to physical localisation by hybridisation to metaphase chromosomes (Evans, 1987). The approaches used in this study were linkage analysis using the BXD RI strains, and direct breeding to confirm the localisation.

4.2 BXD recombinant inbred strain analysis

Under high stringency hybridisation conditions (0.1x SSC, 65°) Mm3-1 detected a wide range of allele lengths across the BXD RI strains (Fig.4.1). On closer analysis the RI strains fall into two classes on the basis of allele length; this bimodal allele distribution centres on both C57BL/6J and DBA/2J progenitor allele sizes. This is consistent with the strain-specific allele size ranges observed in Chapter 3. On the basis of this bimodal size distribution a provisional SDP was derived for *Ms6-hm* : all strains with alleles of more than 1000 repeat units were proposed to be inherited from the C57BL/6J progenitor strain, and those with smaller alleles from the DBA/2J progenitor strain (Fig.4.1).

The polymorphic *Alu*I site 3' to the minisatellite (see Chapter 3) was used to identify the progenitor allele for which each BXD RI strain was fixed. The RI DNAs were typed with endonucleases *Alu*I and *Sau*3AI (Fig.4.1). *Alu*I alleles which are trimmed by *Sau*3AI are derived from the C57BL/6J progenitor, whereas the opposite is true of DBA/2J derived alleles. The resulting *Alu*I polymorphism-based SDP matched the provisional size-based SDP with one exception (BXD 31). This result suggests that most length changes at *Ms6-hm* which accumulate during the breeding of the BXD RI strains are small, such that alleles at this locus in a particular RI strain are contained within a size range defined by the progenitor allele for which that strain is fixed. Furthermore, this suggests that the strain-specific allele size ranges noted among different inbred strains are not defined by the different genetic backgrounds of each strain (which are

Figure 4.1

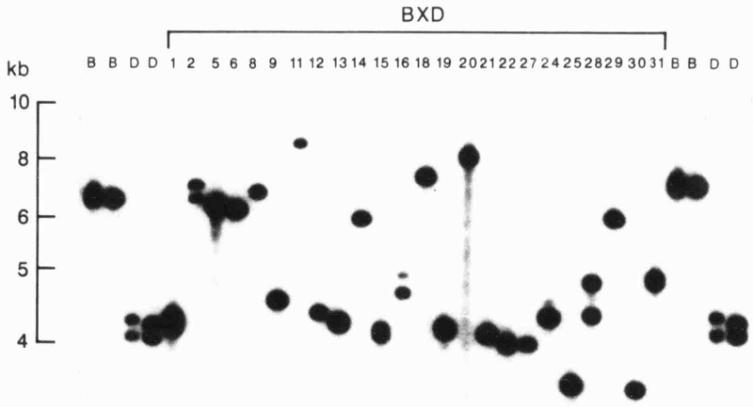
Identification of *Ms6-hm* alleles in BXD recombinant inbred strains

A. BXD RI and progenitor strain DNAs digested with *Alu*I and hybridised with probe Mm3-1 at high stringency. There is a wide range of allele sizes (from 3-15kb) across the BXD RI strains; this distribution is bimodal, the larger alleles (>5.5kb) centering on the C57BL/6J (B) progenitor allele size, and the smaller alleles (<5.0kb) centering on the DBA/2J (D) progenitor allele size. Note that the DBA/2J progenitors, and BXD RI strains 2 and 28, are heterozygous at this locus, and that BXD 16 is mosaic for cells carrying a somatic mutant allele at *Ms6-hm* (see Chapter 6).

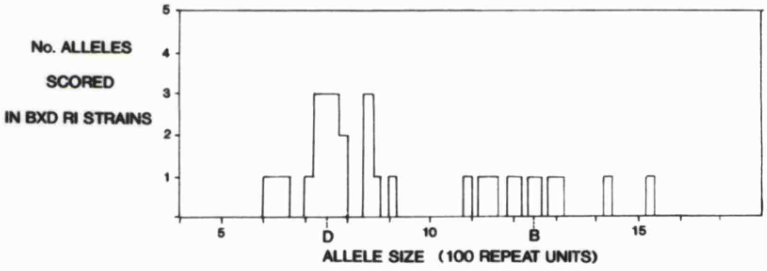
B. *Ms6-hm* allele sizes scored across the BXD RI strains. The number of repeat units per allele (N) is related to *Alu*I DNA fragment length (Lbp) by N = (L-400)/5 for C57BL/6J alleles, and by N = (L-200)/5 for DBA/2J alleles. The allele sizes are grouped in 20 repeat unit intervals. Note the bimodal distribution of allele size, centering on the allele size of each progenitor strain (B and D).

C. DNA from parental and selected BXD RI strains digested with *Alu*I (A) and *Sau*3AI (S) and hybridised with Mm3-1. The additional *Alu*I site flanking *Ms6-hm* in DBA/2J mice provides a polymorphic marker which gives a strain distribution pattern across the RI strains for this locus. The larger alleles (such as BXD 8) show the C57BL/6J pattern, and the smaller alleles (such as BXD 25) show the DBA/2J pattern. BXD 16 is a somatic mosaic at *Ms6-hm* (putative non-parental allele marked •).

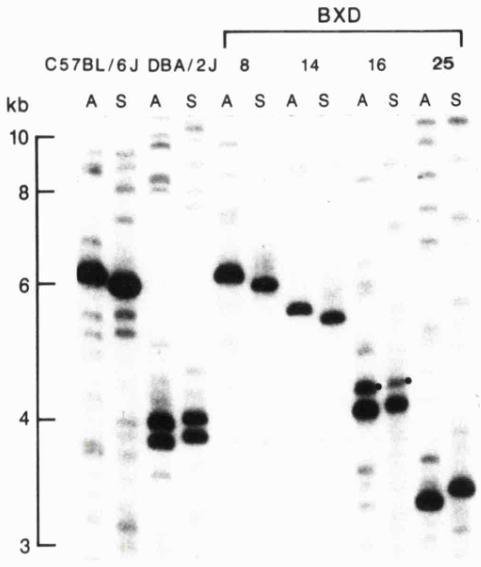
A


BXD
B B D D 1 2 5 6 8 9 11 12 13 14 15 16 18 19 20 21 22 27 24 25 28 29 30 31 B B D D
kb
10
8
6
5
4

B


No. ALLELES
SCORED
IN BXD RI STRAINS
5
4
3
2
1
5
D
10
B
15
ALLELE SIZE (100 REPEAT UNITS)

C


BXD
C57BL/6J DBA/2J 8 14 16 25
kb
A S A S A S A S A S A S
10
8
6
4
3

heterogeneous among the RI strains), but rather by selection for a common allele length imposed by the process of inbreeding.

It is interesting to note the wider range of allele size among C57BL/6J derived alleles. While DBA/2J alleles vary within a 500 repeat unit range, C57BL/6J alleles vary from 1000-2500 repeat units (and alleles of 3000 repeat units have been observed (Fig.3.1)). This spread may be due to either more frequent length changes, or mutation events involving larger length changes, within C57BL/6J alleles. Either or both of these possibilities could be explained by the larger number of repeat units in C57BL/6J, relative to DBA/2J, alleles. The implications of a relationship between the mutation process at *Ms6-hm* and allele length will be discussed in the concluding chapter.

4.3 Strain distribution pattern of *Ms6-hm*

Analysis of the SDP of *Ms6-hm*, based on the polymorphic *Alu*I site, showed significant linkage between *Ms6-hm* and *Ifa*, the murine interferon-α gene cluster (Nadeau *et al.*, 1986) (2/26 recombinants, Fig.4.2). *Ifa* maps distal to the brown coat colour gene (*b*), interstitially on mouse chromosome 4 (De Maeyer and Dandoy, 1987), and 5/26 RI strains are recombinant between *Ms6-hm* and *b*. Two minisatellite loci, *Ms15-1* and *Ms6-2*, map distal to these loci on chromosome 4 (Jeffreys *et al.*, 1987); 7/25 RI strains are recombinant between *Ms6-hm* and *Ms15-1*, and 12/24 are recombinant between *Ms6-hm* and *Ms6-2*.

Taylor (1978) has shown how SDP data can be quantitated for linkage analysis. The probability, R, that an RI strain will carry a non-parental combination of alleles at two loci can be expressed as a function of the recombination frequency, r, in a single meiosis. For brother-sister matings:

$$R = 4r/(1+6r)$$

(Haldane and Waddington, 1931). Thus r, which is also the map distance between the two loci in centiMorgans, can be estimated from R, the ratio of recombinant strains relative to the total number of strains :

$$r = R/(4-6R)$$

for R lying in the range $0 < R < 0.5$. Values of r, and associated confidence limits (Silver, 1985), can then be estimated for each pair of loci, to construct the linkage map illustrated in Fig.4.2. The order cen - *b* - *Ifa* - *Ms6-hm* - *Ms15-1* minimises the number of double crossovers in the RI strains.

Figure 4.2

Strain distribution pattern and linkage analysis of *Ms6-hm*

A. SDP across 26 BXD RI strains obtained for *Ms6-hm* from an *Alu*I site flanking the minisatellite which is polymorphic between C57BL/6J and DBA/2J progenitor strains. An SDP is also obtained from the bimodal distribution of allele sizes across the RIs (expressed in 100 repeat units). Large alleles (>1000 repeat units (underlined)) center on the C57BL/6J progenitor allele size, and smaller alleles (<1000 repeat units) on the DBA/2J progenitor allele size. These two SDPs match in all but one strain, BXD 31.

B. Comparison of the *Alu*I polymorphism-based SDP for *MS6-hm* and the SDPs of linked loci. Computer analysis revealed two recombinant strains out of 26 with *Ifa* (Nadeau *et al.*, 1986). *Ms6-hm* and *Ifa* lie between *b* (Taylor and Shen, 1977) and *Ms15-1* (Jeffreys *et al.*, 1987) on chromosome 4.

C. Linkage map around *Ms6-hm* on chromosome 4. Map distances in cM were calculated from BXD RI data according to standard methods (Taylor, 1978). The 95% confidence limits were taken from a published table (Silver, 1985). The order is in agreement with the minimal number of cross-overs in the RI strains.

A

BXD RI

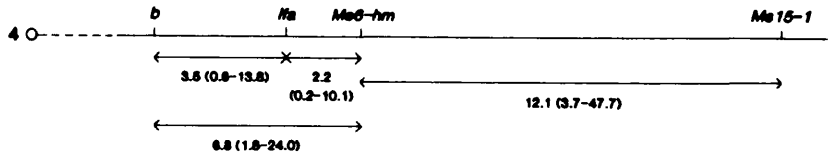
	B	D	1	2	5	6	8	9	11	12	13	14	15	16	18	19	20	21	22	23	24	25	27	28	29	30	31	32
+/- AluI site	-	+	+	-	-	-	-	+	-	+	+	-	+	+	-	+	-	+	+	+	+	+	+	+	-	+	-	-
Allele size (100 RU)	12.4	7.4	7.8 7.6	12.8 12.0	11.8	11.4	12.2	8.4	15.2	7.8	7.6	11.2	7.4 7.2	9.0 8.6	13.0	7.4	14.2	7.2	7.0	7.4	7.6	6.2	6.4	8.4 7.6	10.8	6.0	8.4	25.4 12.4

B

BXD RI

	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
b	D	B	B	B	B	D	B	B	D	B	D	B	B	B	B	D	B	B	D	D	D	D	B	D	B	B
Ifa	D	B	B	B	B	D	B	D	D	B	D	B	B	D	B	D	D	B	D	D	D	D	B	D	B	B
Ms6-hm	D	B	B	B	B	D	B	D	D	B	D	D	B	D	B	D	D	D	D	D	D	D	B	D	B	B
Ms15-1	D	B	D	B	B	D	B	D	B	D	D	D	D	D	B	D	D	-	B	D	D	D	D	D	B	D

C



4.4 Linkage of *Ms6-hm* to *b* and *Ms15-1*

The placement of *Ms6-hm* on mouse chromosome 4 was confirmed by direct progeny analysis of C57BL/6J x DBA/2J F_2 mice, from crosses designed primarily for mutational analysis (see Chapter 5). F_1 mice, heterozygous for C57BL/6J and DBA/2J *Ms6-hm* alleles, were crossed with F_1 heterozygotes carrying a different pair of *Ms6-hm* alleles, such that all four alleles at this locus could be distinguished in the progeny. F_2 mice were also scored for coat colour and typed with minisatellite probe 33.15.

DBA mice are homozygous for dilute, brown, and non-agouti alleles at the *d*, *b*, and *a* loci respectively (Little and Tyzzer, 1915). C57BL/6J mice are also non-agouti, but are wild-type at the *d* and *b* loci. Thus whereas all F_1 mice are black (genotype +/*d* +/*b*), alleles at these two loci assort independently in the F_1 germline to generate four different coat colour phenotypes among the F_2 progeny. These are non-dilute black, dilute black (grey), non-dilute brown, and dilute brown (original DBA colour), and will be found in the ratio 9:3:3:1. Both brown and dilute brown mice are homozygous *b*/*b* and can be scored for the co-inheritance of DBA/2J alleles at *Ms6-hm* and *Ms15-1*.

Similarly all F_1 mice are heterozygous for the C57BL/6J-specific linked haplotype at *Ms15-1* (see Chapter 8). One quarter of F_2 mice will inherit two DBA/2J alleles at this locus, and thus the C57BL/6J-specific haplotype will be absent from their DNA fingerprint. Such mice can be scored for the inheritance of two DBA/2J alleles at *Ms6-hm*, and two *b* alleles at the brown locus. Due to the complexity of the 33.15 DNA fingerprint it was found difficult to ascertain whether F_2 mice carrying the C57BL/6J-specific haplotype were heterozygous or homozygous, so for these data only homozygous negative mice were scored.

Examples of two families, and data for 18 such families, can be seen in Fig.4.3. Between 5 and 16 offspring were typed for each F_1 x F_1 cross. The map distance in centiMorgans between a pair of markers, r, may be calculated from the number of recombinant gametes over the total number of gametes scored :

$$r = n/2N$$

where n is the number of F_2 mice which are recombinant for these markers among the informative F_2 mice scored (N). For *b* and *Ms15-1*, only F_2 mice homozygous for *b* and *Ms15-1*^{D} respectively were informative. Two estimates of r for each pair of markers can be calculated; for example the genetic distance between *b* and *Ms6-hm* can be estimated either from the number of *b*/*b* mice which have not inherited both DBA/2J alleles at *Ms6-hm*, or from the number of mice with two DBA/2J alleles at *Ms6-hm* which are not *b*/*b* (however in this case mice with two recombinant chromosomes will not be detected, resulting in an underestimate of n when r is large). From these data *Ms6-hm* maps 1.3-

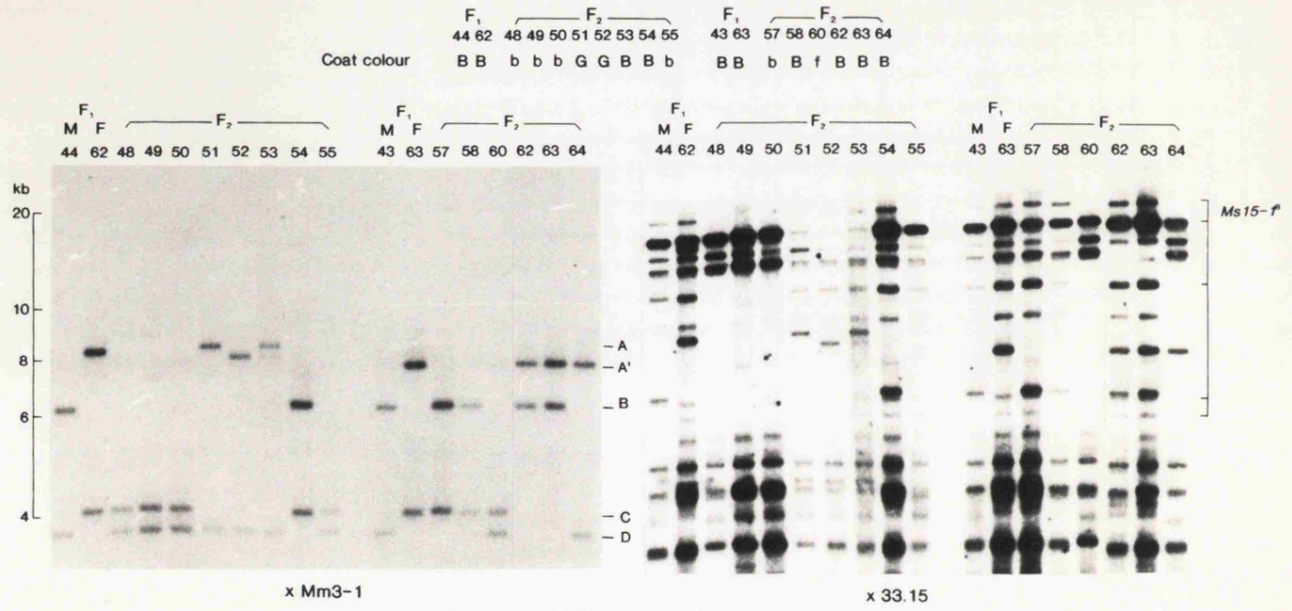
Figure 4.3

Segregation of alleles at *Ms6-hm*, *b*, and *Ms15-1* in C57BL/6J x DBA/2J pedigrees

A. Two C57BL/6J x DBA/2J F_2 families (A5 and A6), showing coat colour and tail DNA digested with *Hin*fI and hybridised with either probe Mm3-1 at high stringency, or probe 33.15 at low stringency. F_2 mice 48, 49, 50, 55 and 60 have all inherited two DBA/2J derived alleles at *Ms6-hm* (alleles C and D). These mice also have a brown coat colour (i.e., are *b*/*b*), and are homozygous for the DBA/2J allele at *Ms15-1* (i.e., lack the C57BL/6J-specific linked haplotype at this locus). F_2 mouse 57 is brown yet has not inherited two DBA/2J alleles at either *Ms6-hm* or *Ms15-1*; this mouse is therefore recombinant between *b* and the two minisatellites. Note that F_2 mouse 52 has inherited a new length mutant allele from F_1 mouse 62 (see Chapter 5). B, black; G, grey; b, brown; f, fawn (DBA). A, A', and B, C57BL/6J derived alleles at *Ms6-hm*; C and D, DBA/2J derived alleles at *Ms6-hm* (see Chapter 5).

B. Summary of linkage data for *Ms6-hm*, *b*, and *Ms15-1* from C57BL/6J x DBA/2J pedigree analysis, showing the number of F_2 mice homozygous for two DBA/2J alleles at each locus which were also scored at the other two loci. For example, 44 brown mice (*bb*) were scored at *Ms6-hm*, 42 of which were homozygous for DBA/2J alleles at both loci (i.e., parental), and 2 of which were recombinant. Two estimates of the map distance (r) in cM may be calculated for each pair of markers (see text). These map distances are in agreement with those calculated from the BXD RI strain distribution patterns, and with the map illustrated in Fig.4.2. *bb*, homozygous brown; CD, two DBA/2J alleles at *Ms6-hm*; $15^{D}15^{D}$, homozygous for the DBA/2J allele at *Ms15-1*.

A



B

Summary of linkage data for *Ms6-hm* from C57BL/6J x DBA/2J F_2 mice

Genotype	No. mice also scored at		No. mice of non-recomb. genotype		No. recomb.	r	95% confidence limits
bb	*Ms6-hm*	44	*bb* CD	42	2	0.023	0.004-0.059
	Ms15-1	43	*bb* 15^D15^D	34	9	0.105	0.055-0.175
CD	*b*	45	CD *bb*	42	3	0.033	0.009-0.083
	Ms15-1	44	CD 15^D15^D	36	8	0.091	0.046-0.160
15^D15^D	*b*	46	15^D15^D *bb*	34	12	0.130	0.080-0.020
	Ms6-hm	43	15^D15^D CD	36	7	0.081	0.038-0.147

2.3cM from *b* and 8.1-9.1cM fom *Ms15-1*, and *Ms15-1* maps 10.5-13.0cM from *b*. These distances are consistent with the map produced using the BXD RI data (see Fig.4.2), and unequivocally confirm the assignment of *Ms6-hm* close to *Ifa* on chromosome 4. While *Ifa* is syntenic with human chromosome 9p (Nadeau *et al.*, 1986), both the origin of *Ms6-hm* within a rodent-specific dispersed repetitive element, and the rapid evolution of allele length at *Ms6-hm*, make it improbable that an orthologous minisatellite locus exists in man (see Chapter 7).

4.5 Application of a highly informative marker on chromosome 4

The construction of detailed genetic linkage maps is critically dependent on the availability of informative DNA markers (Botstein *et al.*, 1980). The discovery of highly polymorphic DNA markers in man (Wyman and White, 1980, Goodbourn *et al.*, 1983, Jeffreys *et al.*, 1985a) has greatly facilitated both the construction of linkage maps of whole human chromomsomes (Donis-Keller *et al.*, 1987) and the localisation of disease loci segregating in affected pedigrees (for example, Reeders *et al.*, 1985, see Introduction). Similarly, hypervariable minisatellite loci in mice provide valuable markers for linkage analysis.

The diabetes locus (*db*), one of two complementary loci which cause heritable obesity in mice (Coleman and Hummel, 1967, 1969), maps close to *Ifa* on mouse chromosome 4. The homozygous *db* phenotype is thought to reflect the loss of a receptor for a circulating appetite suppressor. By a combination of genetical and molecular approaches it is hoped to isolate the *db* gene and characterise the biochemical basis for this mutation (for example, see Friedman *et al.*, 1989).

Two mouse stocks in which *db* and misty (*m*, a recessive coat colour dilution gene, which maps proximal to *db* on chromosome 4 (Wooley, 1945)) are segregating in repulsion were analysed to investigate whether the minisatellite loci *Ms6-hm* or *Ms15-1* were cosegregating with these markers. The stocks differ in regard to the strain background (C57BL/6J and C57BL/KsJ). Carrier mice are bred to propagate both recessive alleles, ensuring a predictable supply of sterile *db*/*db* mice. Occasionally a crossover in the *db*-*m* interval disturbs this mating system such that one or other homozygous offspring is not produced. In these cases the mice are further bred to test whether there has been a crossover, and if so, which mouse carries the recombinant genotype. Characterisation of these recombinant genotypes will help formally define the linkage map interval between the two markers. As the *db* and *m* mutations arose on different backgrounds (C57BL/KsJ and DBA respectively) these loci should be flanked by strain-specific polymorphisms which continue to segregate in these stocks. It was hoped

Figure 4.4

Mapping the *db* locus

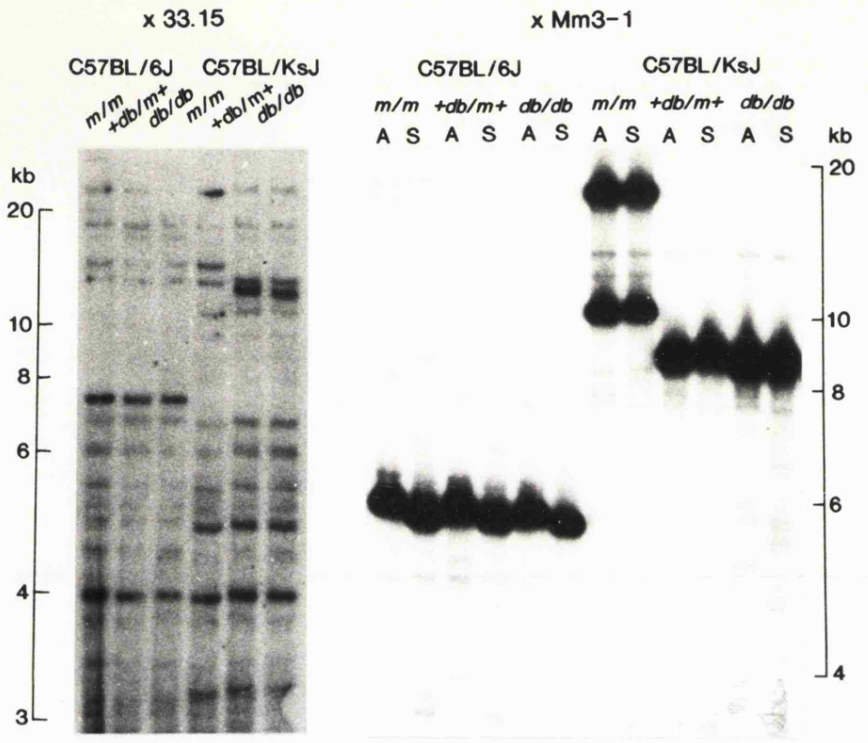
A. DNA from two mouse stocks in which *db* and *m* are segregating in repulsion, either digested with *Hin*fI and hybridised with probe 33.15, or digested with *Alu*I (A) and *Sau*3AI (S) and hybridised with Mm3-1. The two minisatellite loci do not appear to be segregating with *db* or *m* in either stock. Note that C57BL/KsJ mice have a DBA/2J-type signature at *Ms6-hm*.

B. Linkage maps of the region of chromosome 4 around *db* and *Ms6-hm*.

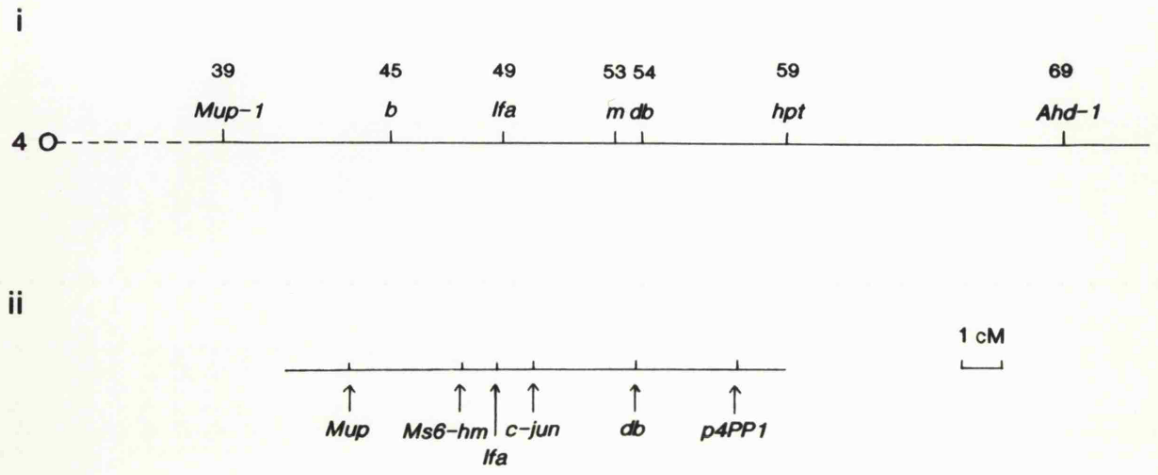
(i) This map is adapted from that of Davisson and Roderick (1989), and is based on data from classical genetic crosses and recombinant inbred strains; recombination distances are given as distance in cM from the centromere.

(ii) This map is based on genetic crosses with *db* mice using a variety of polymorphic markers in this region, including *Ms6-hm* (J.Friedman, personal communication).

A


x 33.15
C57BL/6J
C57BL/KsJ
m/m
+db/m+
db/db
kb
20
10
8
6
4
3
x Mm3-1
C57BL/6J
C57BL/KsJ
m/m
+db/m+
db/db
A S A S A S A S A S A S
kb
20
10
8
6
4

B


i
39
45
49
53 54
59
69
Mup-1
b
lfa
m db
hpt
Ahd-1
4
ii
1 cM
Mup
Ms6-hm
lfa
c-jun
db
p4PP1

that either *Ms6-hm* or *Ms15-1* might detect such a polymorphism, and could then be used to screen recombinants, and help define the position of *db*.

The *db* and *m* genotypes in both stocks carried the C57BL haplotype at *Ms15-1* (Fig.4.4). Similarly, *Ms6-hm* was not segregating for the *Alu*I polymorphism within either stock; however while both C57BL/6J genotypes lacked the polymorphic *Alu*I site linked to *Ms6-hm*, this site was present in the C57BL/KsJ genotypes. Mice from the C57BL/KsJ stocks also had large and highly variable alleles at *Ms6-hm* (9-20kb, Fig.4.4). These results suggested that neither locus would be informative for further analysis of the *db-m* recombinants.

In another study *Ms6-hm* has proved a useful marker in orientating the search for *db* relative to *Ifa* and other loci in this region, including *c-jun* and *Mup-1* (Fig.4.4). The results of linkage analysis using a *Mus spretus* x C57BL/6J interspecific backcross place *db* 5cM distal to *Ms6-hm*, and *Ms6-hm* 1cM proximal to *Ifa*, in contrast to the order consistent with the minimum number of double crossovers relative to *b* in the BXD RIs (J.Friedman, personal communication). Thus while *Ms6-hm* and *Ifa* are tightly linked, the orientation of these two loci relative to the centromere is unclear.

4.6 Summary

Linkage analysis has localised *Ms6-hm* near the brown coat colour gene (*b*) on chromosome 4, where the minisatellite locus in turn provides a highly polymorphic marker for further linkage studies. The hypervariability of *Ms6-hm* makes this one of the most informative DNA markers in the mouse genome. However, the instability of this locus in the BXD RIs suggests that *Ms6-hm* may be too unstable for linkage analysis in extended pedigrees. The congruence of the allele size-based and linked polymorphism-based SDPs for *Ms6-hm* across the BXD RIs suggests that most length changes at this locus are small. Correspondingly, the high degree of intrastrain variation at this locus tends to be contained within a strain-specific size range.

V. GERMLINE MUTATION AT *MS6-HM*

5.1 Introduction

The spontaneous mutation rate to new-length alleles at the most variable human minisatellite loci is sufficiently high to be directly measurable in human pedigrees (Jeffreys *et al.*, 1988a). Such mutation events, which may involve processes of unequal exchange or strand slippage, are responsible for maintaining variability at minisatellite loci (see Introduction). The mechanisms involved in allele length change may be indirectly studied by analysing the rate at which non-parental alleles arise at such extremely unstable loci, and by comparing new-length and progenitor parental alleles. This chapter describes the characteristics of germline mutation events at *Ms6-hm*, and directly assesses the association between such events and unequal meiotic exchange.

5.2 Indirect estimation of mutation rate from heterozygosity levels in inbred strains

Under the random drift hypothesis, heterozygosity in the number of repeat units at a tandemly repeated array within a population is determined by both the mutation rate to new-length alleles at that locus and the effective size of the population. This hypothesis assumes that the length of the tandem array is under no selective constraints. The level of heterozygosity at *Ms6-hm* within an inbred strain therefore provides an indirect estimate of the mutation rate at this locus. Heterozygosity (H) and mutation rate (μ) for neutral alleles are related by the equation

$$H = 4N_e\mu/(1 + 4N_e\mu),$$

where N_e is the effective population size (Kimura and Crow, 1964). N_e is related to the rate of inbreeding ΔF such that $N_e = 1/2\Delta F$ (see 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 mouse strain the mutation rate may be estimated from the equation

$$\mu = H/10.4(1 - H).$$

Many suppliers may propagate inbred mouse strains by mating stud males with seveal females, under the assumption that the mice are already completely inbred. This departure from full-sib mating elevates the effective population size of the inbred strain; recombinant inbred strains, however, are bred under a regimen of strict brother-sister mating, and may be used to estimate heterozygosity. Of 26 mice examined in the BXD RI

series (see Chapter 4), 5 individuals, each from one strain, are heterozygous at *Ms6-hm* (H = 0.19). This gives a germline mutation rate to new length alleles at *Ms6-hm* of 0.023 per gamete. This rate is comparable to that estimated from the BXH RI series (see Chapter 7) where 3 individuals from 12 strains tested were heterozygous, giving μ = 0.032 per gamete.

5.3 Germline mutation in mouse pedigrees

The germline mutation rate at *Ms6-hm* has been directly measured by scoring for new-length alleles in mouse pedigrees. A six-generation C57BL/6J pedigree was bred by successive rounds of strict brother-sister mating (G.Bulfield, Edinburgh) and liver DNA scored for germline mutation at Ms6-hm. 8 new-length mutations were observed among 71 informative mice (i.e., mother/father/offspring trios). In the fourth generation (Fig.5.1) two mutation events, involving changes of approximately 2 and 1.75kb (or 400 and 350 repeat units) generated heterozygotes 61 and 63 respectively. These new mutant alleles were stably transmitted to the next generation, confirming that the germline of these mice also carries the mutant allele, to produce several mice which departed completely from the original allele size. In Fig.5.1 it can be seen that another mutation event caused an additional increase of approximately 150 repeat units in mouse 72.

In order to further investigate mutation events at *Ms6-hm* a large number of C57BL/6J and DBA/2J F_1 and F_2 mice were scored for new-length alleles at this locus. These strains were chosen for several reasons; the fine structure of Ms6-hm had been analysed in detail in these two strains (see Chapter 3), both strains have distinct yet easily scorable allele size ranges, and these strains carry genetic markers making such a cross useful for linkage analysis (Chapter 4). In addition, there is an advantage to scoring F_2 mice in that the hybrid vigour of the F_1 females makes them particularly suited to produce large numbers of offspring. By scoring tail DNA for new-length alleles it is possible to design the structure of the pedigree to study the subsequent inheritance of mutant alleles.

C57BL/6J and DBA/2J 'founder' mice were obtained from several different sources in order to find mice of each strain which were homozygous for different alleles. The allele size ranges of these mice were consistent with those observed in the BXD RI analysis (Chapter 4). C57BL/6J mice with alleles AA' and BB were crossed with DBA/2J mice with alleles CC and DD respectively (see Fig.5.2). Thus two sets of F_1 heterozygotes were obtained carrying alleles A and C or B and D. F_1 heterozygotes were then crossed with F_1 mice carrying a different pair of alleles such that all four alleles at *Ms6-hm* could

Figure 5.1

New-length mutation at *Ms6-hm* in mouse pedigrees

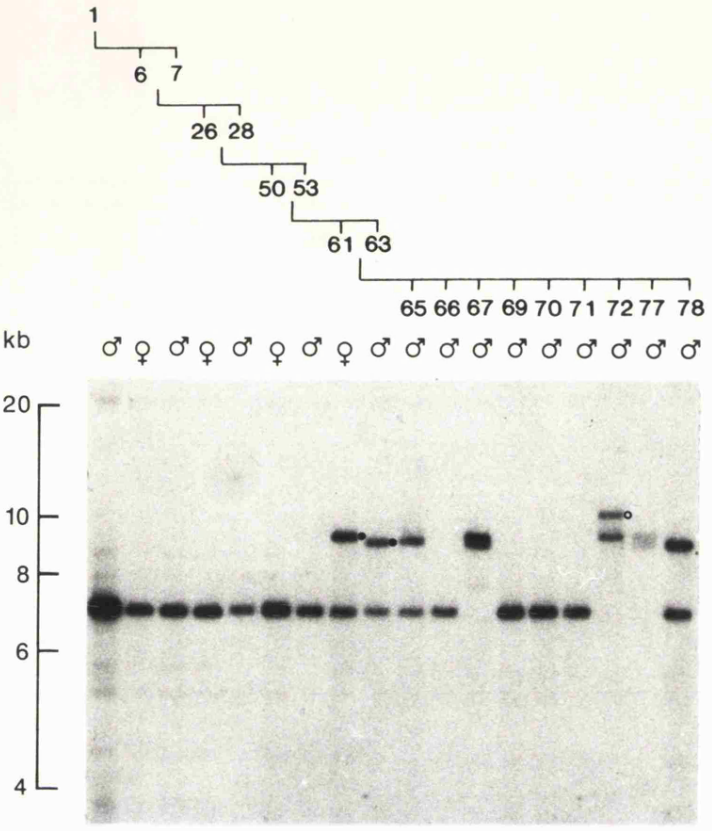
A. *Ms6-hm* alleles in a six-generation C57BL/6J pedigree. Mice were bred by successive rounds of brother-sister mating. Liver DNA was digested with *Hin*fI and hybridised with Mm3-1 at high stringency. DNA from the original female was unavailable. Mice 61 and 63 have new-length mutant alleles at *Ms6-hm* (•) absent from either maternal (50) or paternal (53) DNAs. These alleles are transmitted to the next generation, one subsequently mutating further in mouse 72 (⊙). As this mouse has inherited a normal maternal allele, the mutant allele is of paternal origin.

B. Germline mutation at *Ms6-hm* in C57BL/6J x DBA/2J pedigrees.

i. Breeding strategy. C57BL/6J mice with alleles AA' (which differ by approximately 50 repeat units) and BB were mated with DBA/2J mice with alleles CC and DD respectively, to generate two sets of F_1 heterozygotes, carrying alleles A and C or B and D. F_1 AC heterozygotes were then crossed with F_1 BD heterozygotes to enable all four alleles at *Ms6-hm* to be distinguished among the progeny. 25 F_2 families were analysed. Approximate allele length in repeat units is given; the number of 5bp repeat units per allele is related to the *Hin*fI DNA fragment length (L bp) by N = (L - 1070)/5.

ii. *Ms6-hm* alleles in a three-generation pedigree. Tail DNA was digested with *Hin*fI and hybridised with Mm3-1 at high stringency. F_2 mouse 52 has a new-length allele (•) not seen in parental DNAs (44 and 62). In this mouse the non-mutant allele is inherited from the mother and the mutant allele is therefore of paternal origin. The presumed mutation event is a small (0.3kb) length change from the larger paternal allele in the father's germline. MM (8) and FF (2), mother's mother and father's father, C57BL/6J; MF (21) and FM (11), mother's father and father's mother, DBA/2J; M (44), mother, and F (62), father, F_1 hybrid mice; 48-54, F_2 hybrid mice.

A



B i

C57BL/6J DBA/2J C57BL/6J DBA/2J

$\frac{A}{A'} \times \frac{C}{C}$ $\frac{B}{B} \times \frac{D}{D}$

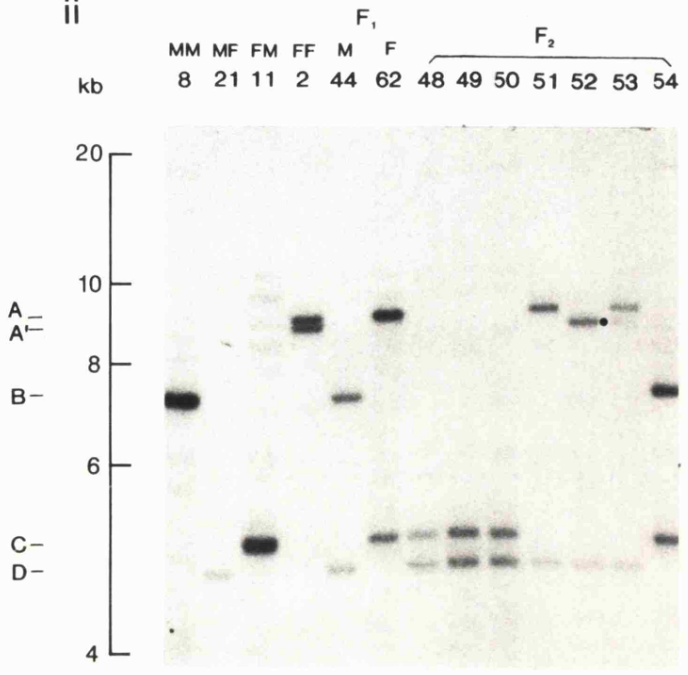
F_1 $\frac{A}{C}$ $\frac{A'}{C}$ x $\frac{B}{D}$

F_2 $\frac{A}{B}$ $\frac{A}{D}$ $\frac{C}{B}$ $\frac{C}{D}$

Repeat units

A	A'	B	C	D
1640	1590	1240	750	690

ii



be distinguished among the progeny. In such fully informative pedigrees the origin of any mutant allele in an F_2 mouse could be clearly determined. 76 F_1 and 221 F_2 mice were scored in this way, yielding a total of 11 germline muations.

An example of such a family is illustrated in Fig.5.1. F_2 mouse 72 has a new-length mutation at *Ms6-hm* that is paternal in origin, as the non-mutant allele in this mouse comes from the mother. If the progenitor allele in this case is assumed to be the larger paternal allele (closest in size to the mutant), then the mutation event involves the loss of about 300bp (60 repeat units). Each mutant allele has been characterised in this way to generate information on the rate and other features of spontaneous germline mutation at *Ms6-hm*.

5.4 Germline mutation rate

A total of 19 new mutant alleles at *Ms6-hm* have been observed in 373 mother/father/offspring trios (Fig.5.3), giving a germline mutation rate of 0.025 per gamete (95% confidence limits 0.017-0.037). This rate is in agreement with the indirect estimate based on heterozygosity levels in inbred strains and is similar to that of the most unstable human minisatellite loci so far described (Jeffreys *et al.*, 1988a). As mutation events involving very small length changes (particularly at large alleles) will not be resolved due to the short repeat unit of *Ms6-hm*, the observed germline mutation rate is likely to underestimate the true mutation rate.

The progenitor parental allele could be accurately determined only when the parent transmitting the mutant allele was homozygous at *Ms6-hm*. Alternatively, if this parent was heterozgous at *Ms6-hm*, the mutant allele was assumed to be derived from the closest parental allele. This assumption was previously made for mutant alleles derived from heterozygous parents at the human minisatellite locus *λMS1* (Jeffreys *et al.*, 1988a); markers flanking this minisatellite were found in the parental combinations predicted on the basis of the closest parental allele in 11/11 new-length alleles (A.Jeffreys, Leicester, personal communication). In the case of C57BL/6J x DBA/2J F_2 mice with non-parental alleles at *Ms6-hm* the mutant allele was shown by the signature test to inherit the flanking *Alu*I polymorphism of the predicted progenitor allele. However, if a process of unequal meiotic recombination is responsible for germline mutation events at *Ms6-hm* neither parental allele can be described as the sole progenitor of a mutant allele.

The 19 mutant alleles are described in Table 5.1 and Fig.5.2. Mutation events at *Ms6-hm* involve either gain or loss of repeat units, each with approximately equal frequency (11 and 8, respectively). The majority of mutation events at *Ms6-hm* involve

Table 5.1

Summary of germline mutation data at *Ms6-hm*

Table summarising the data from nineteen mice carrying germline mutant alleles at *Ms6-hm*, showing the sex of each mouse (m, male; f, female), and the direction and size of each mutation event. The progenitor parental allele could be accurately determined only when the parent transmitting the mutant allele was homozygous at *Ms6-hm* (see text). In cases where this parent is heterozygous, the parental allele closest in size to the mutant allele (which is most likely to be the progenitor allele) is indicated by an asterisk. For F_2 mice 8, 52 and 126 this assignment has been shown to be correct using flanking polymorphisms (see text).

The parental origin of the progenitor allele could not be determined in cases where the parents were both either homozygous (for example, C57BL/6J 16) or heterozygous (for example, F_2 8) for indistinguishable alleles. The parents of F_2 mice 8 and 17 both carried the same sets of alleles (B and D, see Fig.5.1). For the relationship between the number of repeat units per allele and the *Hin*fI DNA fragment length see the legend to Fig.5.1. F_1, C57BL/6J x DBA/2J F_1; F_2, C57BL/6J x DBA/2J F_2; P, paternal; -, could not be determined.

Germline mutation at *Ms6-hm*

	Mouse	Sex	No. repeat units in progenitor allele (origin)	Repeat unit change	% change
C57BL/6J	16	m	1240 (-)	+ 100	8.1
	32	f	1240 (-)	- 20	1.6
	36	f	1240 (-)	+ 30	2.4
	46	m	1240 (-)	+ 20	1.6
	49	f	1240 (-)	+ 100	8.1
	61	f	1240 (-)	+ 400	32.2
	63	m	1240 (-)	+ 350	28.2
	72	m	* 1590 (P)	+ 150	9.4
			1240 (P)	+ 500	40.3
F_1	36	m	* 1640 (P)	+ 110	6.7
			1590 (P)	+ 160	10.0
	50	m	690 (P)	- 80	11.7
	55	f	* 1640 (P)	+ 350	21.3
			1590 (P)	+ 400	25.2
	63	m	* 1590 (P)	- 100	6.3
			1640 (P)	- 150	9.2
	72	m	690 (P)	- 20	2.9
	218	f	690 (P)	+ 20	2.9
F_2	8	f	* 1240 (-)	- 250	20.2
			690 (-)	+ 300	43.5
	17	m	* 690 (-)	+ 10	1.4
			1240 (-)	- 540	43.5
	52	f	* 1640 (P)	- 60	3.6
			750 (P)	+ 830	110.7
	126	m	* 1240 (P)	- 250	20.2
			690 (P)	+ 300	43.5
	242	-	1240 (P)	- 570	46.0
			* 750 (P)	- 80	10.7

small length changes (<200 repeat units); the mean change in allele repeat copy number is 10%. However, large length changes can also occur at *Ms6-hm*, as inbred individuals heterozygous for very different sized alleles are found (for example, BXD 32, Fig.3.1). These characteristics are consistent with those of germline mutation events observed at human minisatellite loci (Jeffreys *et al.*, 1988a), and with the hypothesis that random genetic drift accounts for the size distribution of minisatellite alleles. Six of these mice were mated, and in all cases the non-parental allele was stably transmitted to the next generation.

Mutations at *Ms6-hm* arise sporadically, with no significant clustering in families; the 19 mutant alleles were distributed over 15 out of 35 families. This is the expected number for a random distribution of mutant offspring; the mean incidence of mutant offspring per family is 19/35 or 0.543, and the chance that a family would have no mutant offspring is given by the Poisson distribution as $e^{-0.543}$ or 0.58. Thus the number of families out of 35 with no mutant offspring is 35x0.58, or 20.3 (assuming that all families are of equal size). Furthermore, there were no common non-parental alleles among the mutant offspring in the four families which had two mutant offspring, indicating that there is no significant level of germline mosaicism for mutant alleles at *Ms6-hm*.

There is no clear association between parental allele size and the rate or number of repeat units gained or lost in germline mutation events at *Ms6-hm* (Fig.5.3). However a wider range of length changes is observed among large alleles (>1000 repeat units). Fewer mutant alleles have been scored which are derived from small alleles (<1000 repeat units, 4/19), and therefore it is not possible to say whether large length changes never occur for small alleles, or rather have just not been observed in this study. Alleles greater than 1000 repeat units are C57BL/6J derived, and those less than 1000 repeat units are DBA/2J derived (above, and Chapter 4). Therefore the observation that large length changes appear to be confined to large alleles may be explained by a strain difference between C57BL/6J and DBA/2J alleles, rather than simply on the basis of parental allele size. The observation that less mutant alleles are scored among DBA/2J alleles than C57BL/6J alleles is not significant (0.5>p>0.25) because a larger number of C57BL/6J alleles were scored for mutation.

It is not possible to determine the parental origin of all 19 mutant alleles (for example, both parents of the mutant offspring may be share alleles at *Ms6-hm*). However in every case in which the parental origin of the mutant allele could be determined (10 out of 19 cases), the new-length allele was paternal in origin (the probability of this occuring by chance is $1/2^9$, or 0.002). This contrasts with the observation that the

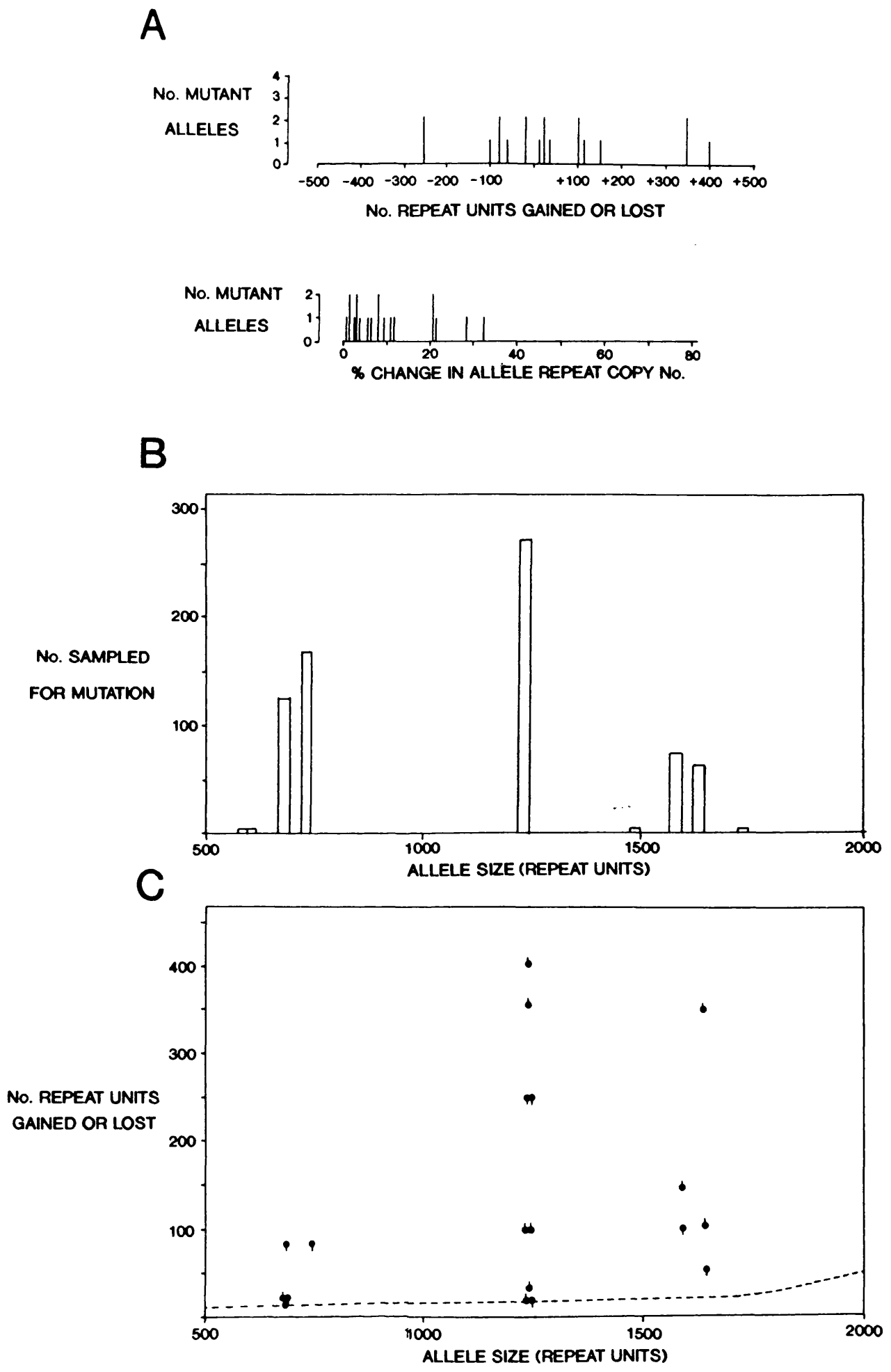
Figure 5.2

Characteristics of germline mutation events at *Ms6-hm*

A. Diagrams summarising the number of repeat units gained and lost, and the percentage change in allele length, during 19 germline mutation events at *Ms6-hm*. The progenitor allele is taken to be the closest parental allele (marked by an asterisk in Table 5.1) in cases where the parent transmitting the mutant allele is heterozygous. The mean change in allele repeat copy number is 10%.

B. Size distribution of parental *Ms6-hm* alleles sampled for mutation in C57BL/6J and C57BL/6J x DBA/2J pedigrees. For the relationship between the number of repeat units per allele and the *Hin*fI DNA fragment length see the legend to Fig.5.1. The size distribution is given in 125bp intervals, though each size interval contains multiple resolvable alleles.

C. Relationship between the number of repeat units gained or lost in mutant *Ms6-hm* alleles and the number of repeat units in the parental allele. As in A, the closest parental allele (marked by an asterisk in Table 5.1) is taken to be the progenitor allele in cases where the parent transmitting the mutant allele is heterozygous. ♦, mutant allele with gain of repeats; ♦, mutant allele with loss of repeats. The dotted line represents the approximate resolution limit of the gel; any mutations below this line are not resolvable from the parental allele.


A
No. MUTANT ALLELES
4
3
2
1
0
-500 -400 -300 -200 -100 +100 +200 +300 +400 +500
No. REPEAT UNITS GAINED OR LOST
No. MUTANT ALLELES
2
1
0
0 20 40 60 80
% CHANGE IN ALLELE REPEAT COPY No.
B
No. SAMPLED FOR MUTATION
300
200
100
500 1000 1500 2000
ALLELE SIZE (REPEAT UNITS)
C
No. REPEAT UNITS GAINED OR LOST
400
300
200
100
500 1000 1500 2000
ALLELE SIZE (REPEAT UNITS)

mutation rate at human minisatellite loci is indistinguishable in sperm and oocytes (Jeffreys *et al.*, 1988a).

5.5 Mutation processes at *Ms6-hm*

The mechanisms responsible for mutation at minisatellite loci remain unknown. Two processes which are likely to be involved are unequal exchange (either between homologues or sister chromatids) and strand slippage (see Introduction). By analysing parental and new-length mutant alleles at *Ms6-hm* it is possible to test predictions associated with different models of minisatellite mutation.

One model suggests that new-length alleles are generated by unequal exchange between homologous chromosomes at meiosis or mitosis. This predicts that markers flanking the minisatellite will be recombinant following the mutation event. Wolff *et al.* (1988) examined one mutant allele at a human minisatellite which differed by a single repeat unit from the parental allele, and showed by DNA sequence analysis that flanking markers were not exchanged. Subsequently Wolff *et al.* (1989) characterised 12 new-length alleles at the hypervariable human minisatellite λ*MS1*, and found that mutation events were not obviously associated with the exchange of markers 6.8 and 3.9cM either side of the minisatellite. The two polymorphisms between C57BL/6J and DBA/2J mice immediately flanking *Ms6-hm* (see Chapter 3), provide the opportunity of investigating whether flanking markers are exchanged during mutation events at this highly unstable locus.

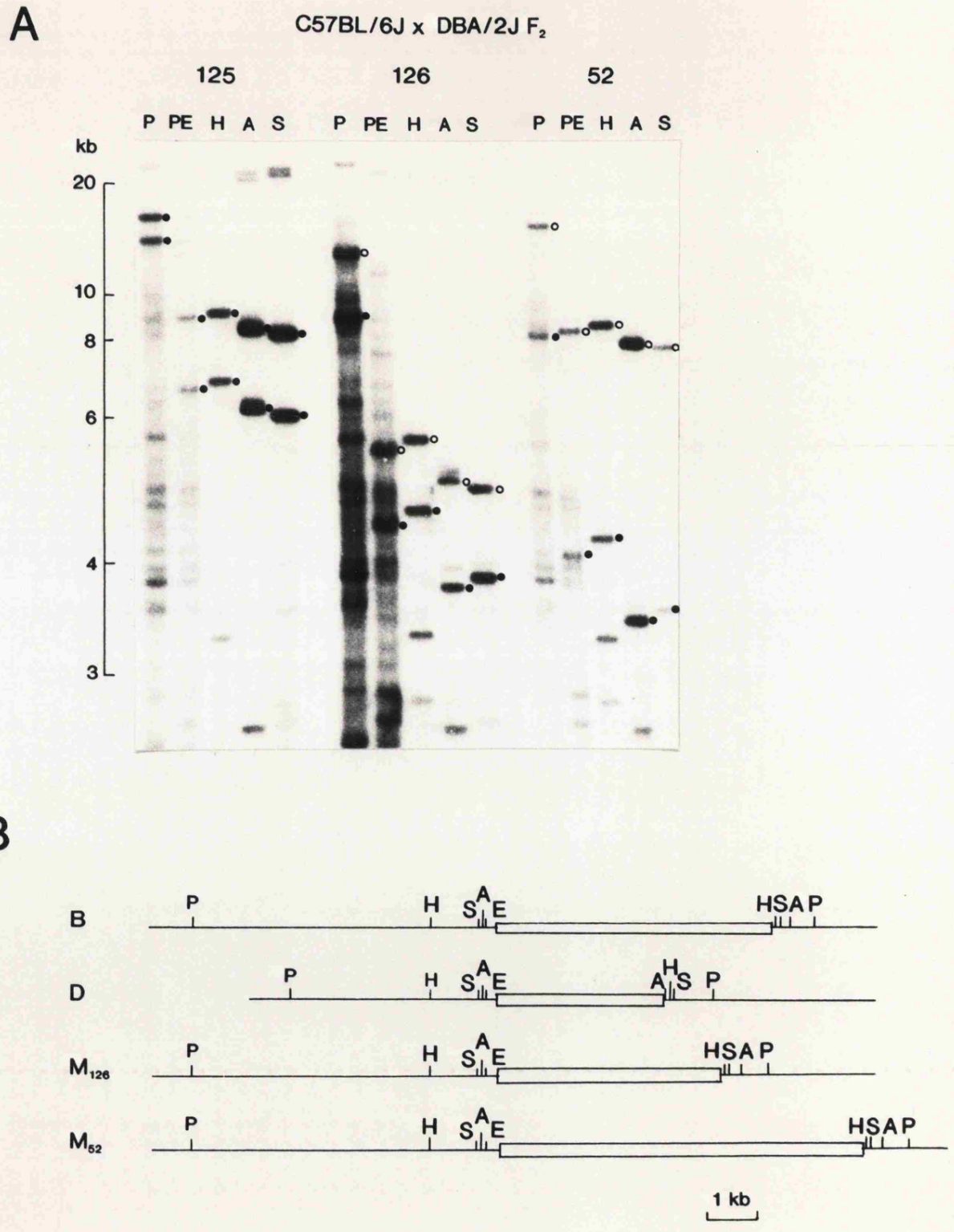
New-length alleles scored in C57BL/6J x DBA/2J F_2 mice have presumably arisen in the germline of F_1 mice, where C57BL/6J and DBA/2J chromosomes pair during meiosis. Thus it is possible to investigate whether the F_2 mutant alleles are parental or recombinant for the flanking polymorphisms between C57BL/6J and DBA/2J alleles. Using the restriction endonucleases *Pvu*II, *Pvu*II and *Eco*RI, *Hin*fI, *Alu*I, and *Sau*3AI, three F_2 mutant alleles have been shown to be parental for the 5' insertion/deletion event and the +/- *Alu*I polymorphism (F_2 mice 126, 52 (Fig.5.4), and 8). All three mutant alleles are derived from C57BL/6J parental alleles. This result rules out inter-allelic unequal exchange at meiosis or mitosis associated with the exchange of flanking markers, as a possible model for all mutation events at *Ms6-hm*. Furthermore, these results indicate that no more than 63% of mutation events at *Ms6-hm* could be associated with unequal exchange of flanking markers ($p>0.95$).

Figure 5.3

Flanking markers and new-length mutant alleles at *Ms6-hm*

A. DNA from three C57BL/6J x DBA/2J F_2 mice digested with *Pvu*II (P), *Pvu*II and *Eco*RI (PE), *Hin*fI (H), *Alu*I (A) and *Sau*3AI (S), and hybridised with Mm3-1 at high stringency. F_2 mouse 125 has two C57BL/6J derived parental alleles at *Ms6-hm* (marked •); F_2 mice 126 and 52 each have one new-length mutant allele (marked o) and one DBA/2J derived parental allele (•) at *Ms6-hm*. Compare the *Pvu*II and *Pvu*II/*Eco*RI digests, and the *Hin*fI, *Alu*I, and *Sau*3A signature of the different alleles.

B. Restriction maps of four *Ms6-hm* alleles from the blot illustrated in A (compare with the map in Fig.3.5). Tandem repeated minisatellite sequence is denoted by an open box. The C57BL/6J allele (B) is the lower allele of mouse 125, and the DBA/2J allele (D) is the lower allele of mouse 126. Note that the new-length alleles in mice 126 (M_{126}) and 52 (M_{52}) are non-recombinant for both the 5' insertion/deletion event and the 3' polymorphic *Alu*I site, and that both mutant alleles are C57BL/6J derived.


A
C57BL/6J x DBA/2J F2
125
126
52
P PE H A S P PE H A S P PE H A S
kb
20
10
8
6
4
3
B
B
D
M126
M52
P
H
S
A
E
HSAP
1 kb

5.6 Summary

Multiallelism and heterozygosity at *Ms6-hm* within inbred strains of mice result from a high germline mutation rate to new-length alleles (2.5% per gamete). Mutation events at *Ms6-hm* are sporadic, and show no substantial bias towards allele expansion or contraction, although larger alleles may be associated with larger size changes. Mutation events at *Ms6-hm* preferentially occur in the male germline, which may reflect the large number of cell divisions during spermatogenesis compared with oogenesis. This suggests that mitotic events may be involved in the generation of new-length alleles at *Ms6-hm*. Furthermore, mutation events at this locus are not accompanied by the exchange of flanking markers in three mutant alleles analysed.

VI. SOMATIC MUTATION AT *MS6-HM*

6.1 Introduction

Hypervariable minisatellite loci detected by probes 33.6 and 33.15 in human DNA appear to be somatically stable (Jeffreys *et al.*, 1985b). In addition, all new-length allele mutations observed at five of the most unstable human minisatellite loci in 344 offspring are thought to be germline in origin (Jeffreys *et al.*, 1988a). In common with germline mutation events at *Ms6-hm* (Chapter 5), the mutant offspring are clonal, with no trace of the original parental allele in addition to the new mutant allele. Mutant alleles were never shared by more than one offspring in a sibship, and the new mutant alleles are transmitted faithfully from parent to offspring. In contrast to these observations, mice have been found which are mosaic for cells carrying a common non-parental allele at *Ms6-hm* in somatic tissue, and in some cases also in the germline. This suggests that additional mutation events occur at *Ms6-hm* in early mouse development which precede the separation of the somatic and germ cell lineages.

6.2 Evidence for somatic and germline mosaicism at *Ms6-hm*

While screening C57BL/6J and C57BL/6J x DBA/2J pedigrees for germline mutation events at *Ms6-hm*, a small number of mice were found to have a third, non-parental, allele at this locus in somatic tissue. The *Hin*fI, *Alu*I, and *Sau*3AI locus-specific signature was used to confirm that the non-parental DNA fragment in each mosaic mouse was an allele of *Ms6-hm*, rather than a new-length fragment derived from a cross-hybridising locus (Fig.6.1). Thus mouse C57BL/6J x DBA/2J F_1 52 has two parental alleles and one non-parental allele at *Ms6-hm* in adult tail DNA. The flanking signature of the new-length mutant allele shows that it is derived from the DBA/2J (paternal) allele (Fig.6.1). The tail of this mouse must therefore be a mosaic of two types of cells, those which contain the two parental alleles, and those which contain the maternal allele and a common new-length mutant allele.

The *Ms6-hm* alleles of mosaic mouse 52 underwent non-Mendelian three way segregation into her offspring (Fig.6.1). Thus the germline of this mouse must be mosaic for cells containing the same new-length allele at *Ms6-hm* which was found in tail DNA, and each germ cell must exclusively contain either one of the two parental alleles or the new length mutant allele at this locus. Mouse C57BL/6J 22 similarly transmitted any

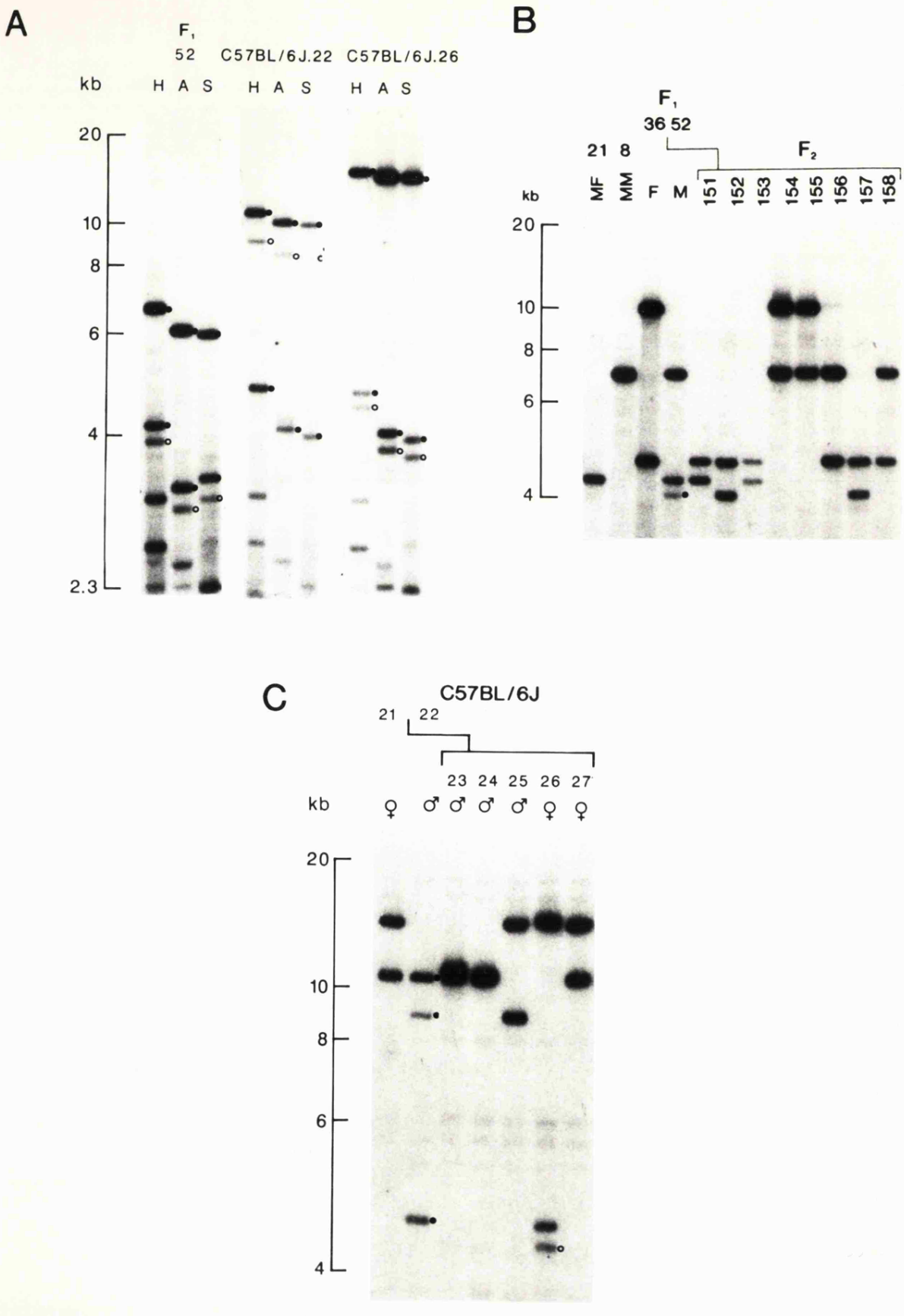
Figure 6.1

Somatic and germline mosaicism at *Ms6-hm*

A. Examples of mice with an additional allele at *Ms6-hm*, showing the characteristic signature for *Ms6-hm* with restriction endonucleases *Hin*fI (H), *Alu*I (A), and *Sau*3AI (S). Parental alleles (•) and non-parental alleles (o) are marked. Mouse 52 is a C57BL/6J x DBA/2J F_1; the DBA/2J allele (lower) has mutated somatically at an early stage of development. In mice C57BL/6J 22 and 26, the third allele has presumably arisen from the closest parental allele.

B. Tail DNA from a C57BL/6J x DBA/2J pedigree digested with *Hin*fI and hybridised at high stringency with Mm3-1. C57BL/6J x DBA/2J F_1 mouse 52 has a somatic mutant allele (•) derived from her paternal allele. The germline of this mouse is also mosaic; either of the parental alleles or the non-parental allele can be transmitted to her progeny. This suggests that the somatic mutation event preceded the separation of the germline and soma. MF (21), DBA/2J father of 52; MM (8), C57BL/6J mother of 52; 36 (F, father) and 52 (M, mother), C57BL/6J x DBA/2J F_1 mice; 151 - 158, F_2 mice.

C. Brain DNA from a C57BL/6J pedigree digested with *Hin*fI and hybridised at high stringency with Mm3-1. C57BL/6J 22 has three alleles at *Ms6-hm* (•). Each of these can be transmitted to the next generation, showing that the germline of this mouse is also mosaic. A subsequent somatic mutation event generates another mosaic mouse (C57BL/6J 26) with a non-parental allele (o). This mutant allele is presumably derived from the lower (paternal) parental allele.


A
F1
52
C57BL/6J.22
C57BL/6J.26
kb
H A S H A S H A S
20
10
8
6
4
2.3
B
F1
36 52
21 8
F2
MF MM F M 151 152 153 154 155 156 157 158
kb
20
10
8
6
4
C
C57BL/6J
21
22
23 24 25 26 27
kb
♀ ♂ ♂ ♂ ♂ ♀ ♀
20
10
8
6
4

one of three alleles to the next generation, one allele subsequently mutating further in one of the offspring to generate another mosaic mouse (C57BL/6J 26, Fig.6.1).

6.3 Characteristics of somatic mutation events

The degree of mosaicism in somatic mosaic mice may be estimated by densitometric scanning of allele intensities. The relative intensity of any minisatellite band on an autoradiograph is proportional to allele size (the number of repeat units) and dosage; whereas the allele dosage is 2 for a homozygous fragment, and 1 for a heterozygous fragment, the dosage of a mosaic fragment must lie in the range 0-1. The dosage of a somatic mutant allele can be estimated by comparing the intensity of the new-length fragment in DNA from a mosaic mouse with the same amount of DNA from a heterozygous offspring to which the mutant allele has been transmitted. Alternatively, dosage may be estimated by comparing the intensities of the parental and non-parental alleles in DNA from a mosaic mouse and correcting for allele length (see legend to Table 6.1).

Of 471 mice typed, 13 were demonstrably somatic mosaics at *Ms6-hm*, giving a frequency of 0.028 (Table 6.1). The dosage of the mutant allele in these mice ranged from 0.08 to 0.6 indicating that 8-60% of cells in the mosaic tissue typed contain the new mutant allele rather than the progenitor allele. Four out of the five mosaic mice which were mated transmitted the new-length allele to the next generation. As with germline mutation events at *Ms6-hm*, somatic mutation events tend to involve small length changes (<200 repeat units); the mean change in allele repeat copy number is 13% (Fig.6.2). Although 10 of the 13 mutation events involve a loss of repeat units, this is not a significant bias ($p=0.09$, two-tailed binomial); all somatic mutation events involving length changes of more than 40 repeat units, however, are deletion events (Fig.6.2). Two of these events involve the loss of >500 repeat units, whereas the largest germline length change observed at this locus involves the gain of 400 repeat units.

Somatic mutation events appear to be sporadic; the 7 mosaic mice which were scored in pedigrees were distributed over 5 out of 35 families (for a random distibution 7 mosaic mice would be expected to occur in 6 families out of 35 (see Chapter 5)). However no somatic mosaic mice were observed among 221 C57BL/6J x DBA/2J F_2 mice scored; this represents a significant deviation from the frequency of mosaicism among C57BL/6J and C57BL/6J x DBA/2J F_1 mice ($0.025>p>0.01$), and suggests that somatic mutation events may be influenced by additional unknown factors.

The parental origin of the progenitor allele of 6 somatic mutant alleles could be determined either by the signature test (in the 4 F_1 mosaic mice), or by the closest

Table 6.1 Summary of somatic mutation data at *Ms6-hm*

This table summarises the data from thirteen mice carrying somatic mutant alleles at *Ms6-hm*, showing the sex of each mouse (m, male; f, female), the direction and size of the mutation events, the dosage of the mosaic allele in DNA from adult tissue (t, tail; l, liver; b, brain), and the germline transmission ratio for those mosaics which were bred from, expressed as the number of offspring inheriting the mutant allele over the total number of offspring, and as the transmission frequency. Those mice which were dissected are marked (D). In the case of heterozygous mosaic mice both parental alleles are included.

The first 7 somatic mosaic mice were identified during analysis of C57BL/6J and C57BL/6J x DBA/2J pedigrees, and the parental origin of the non-mutant alleles in these mice is known. For heterozygous mosaic mice where it has been possible to determine the progenitor parental allele, this is indicated by an asterisk; in the case of C57BL/6J x DBA/2J F_1 mice this was determined by the *Hin*fI, *Alu*I, *Sau*3AI signature test, and in other cases where the mutant allele is clearly closer to one parental allele by the smaller length change. In C57BL/6J 67 and 52 the two parental alleles are very close in size, and therefore the parental origin of the mutant allele is uncertain. SJL F_2.1 was identified among the progeny of a cross between a C57BL/6J x DBA/2J F_2 mouse and an SWR x SJL F_1 mouse (see Chapter 7); in this mouse the allele marked by an asterisk is C57BL/6J derived.

The degree of mosaicism (m) in adult tissue was estimated by densitometric scanning of autoradiographs to determine the dosage of the mutant allele relative to each of the parental alleles. For homozygous mosaic mice m is related to fragment intensity by the equation $m = i_m/(I(s_m/2S))$ where i_m is the intensity of the mutant allele; I is the intensity of the parental fragments; s_m is the size of the mutant allele (in repeat units); and S is the size of the parental alleles. For heterozygous mosaic mice m is related to the intensities of the progenitor (A) and the non-progenitor (B) parental alleles by the equations $m = i_m/(I_A(s_m/S_A) + i_m) = i_m/i_B(s_m/S_B)$. For mice with equal somatic and germline mosaicism, the relationship between the degree of mosaicism in somatic tissue (m) and the germline transmission frequency (t) is $t = m/2$. BL/6J, C57BL/6J; F_1, C57BL/6J x DBA/2J F_1; P, paternal; M, maternal; -, could not be determined.

Somatic mosaicism at *Ms6-hm*

Mouse	Sex	No. repeat units in parental allele (origin)	Repeat unit change	% change	Mosaic allele dosage (tissue)	Germline transmission ratio (frq)	
BL/6J 67	m	* 1590 (P)	- 260	16.4	0.10 l	-	
		1640 (M)	- 310	18.9			
BL/6J 52	m	* 1640 (M)	+ 20	1.2	0.15 l	-	
		1740 (P)	- 80	4.6			
F_1 37	f	* 1590 (P)	- 360	22.6	0.40 t	1/39	D
		750 (M)	+ 480	64.0		(0.02)	
F_1 52	f	1240 (M)	- 590	47.6	0.45 t	5/16	D
		* 690 (P)	- 40	5.8		(0.30)	
F_1 213	m	* 1590 (P)	- 40	2.5	0.08 t	0/37	
		750 (M)	+ 800	106.7			
F_1 207	f	* 750 (M)	- 160	21.3	0.15 t	4/28	
		1640 (P)	- 1050	64.0		(0.15)	
BL/6J 26	f	3190 (M)	- 2460	77.1	0.45 b	-	D
		* 790 (P)	- 60	7.6			
BL/6J 22	m	* 2240 (-)	- 550	24.6	0.20 b	1/5	D
		790 (-)	+ 900	114.0		(0.20)	
BXD 16	m	860 (-)	+ 40	4.6	0.20 l	-	
BXH 2	m	1290 (-)	- 160	12.4	0.30 l	-	
BXH 12	m	1230 (-)	- 600	48.8	0.60 l	-	
BL/6J A	m	1240 (-)	+ 20	1.6	0.15 b	-	
SJL F_2.1	m	2790 (M)	- 1570	56.3	0.60 t	-	
		* 1240 (P)	- 20	1.6			

Figure 6.2

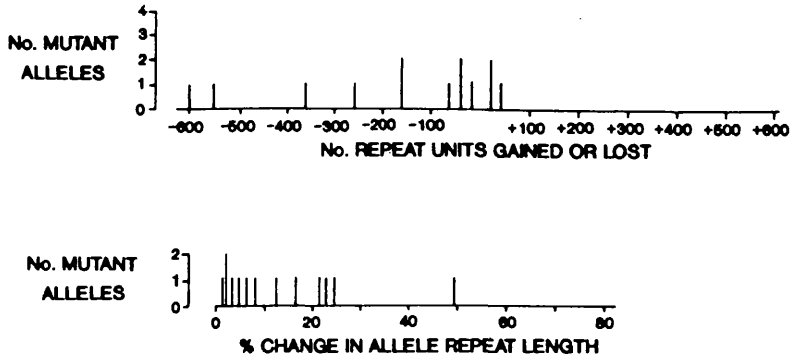
Characteristics of somatic mutation events at *Ms6-hm*

A. Diagrams summarising the number of repeat units gained and lost, and the percentage change in allele length, during somatic mutation events at *Ms6-hm*. The progenitor allele in heterozygous mosaic mice was determined either by the *Hin*fI, *Alu*I, *Sau*3AI signature, or is taken to be the closest parental allele (marked by an asterisk in Table 6.1). The mean change in allele repeat copy number is 13%.

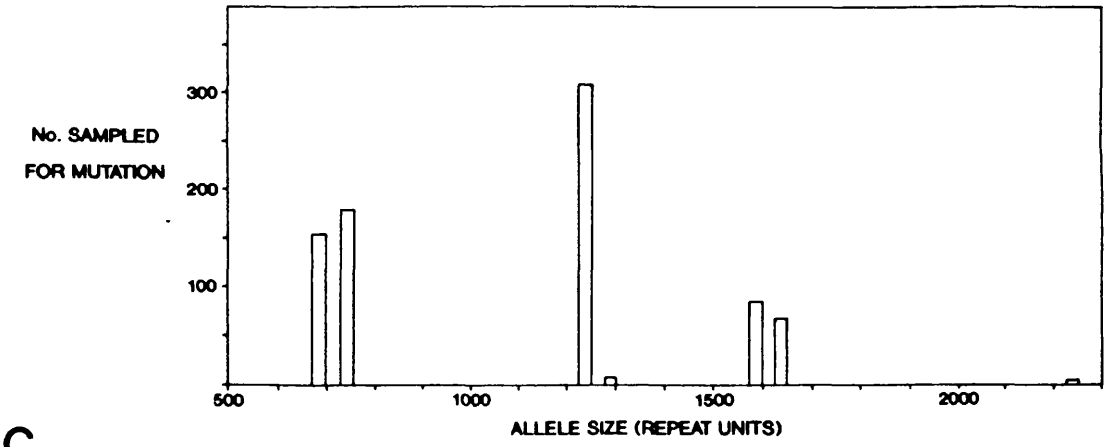
B. Size distribution of *Ms6-hm* alleles sampled for somatic mutation in C57BL/6J, DBA/2J, C57BL/6J x DBA/2J, C3H, BXD and BXH inbred mice. For the relationship between the number of repeat units and the *Hin*fI DNA fragment length see the legend to Fig.5.1. The size distibution is given in 125bp intervals, though each size interval contains multiple resolvable alleles.

C. Relationship between the number of repeat units gained or lost in somatic mutant *Ms6-hm* alleles and the number of repeat units in the progenitor allele. The progenitor allele for heterozygous mosaic mice was determined either by signature, or is taken to be the closest parental allele (marked by an asterisk in Table 6.1). ↓, mutant allele with gain of repeats; ↑, mutant allele with loss of repeats. The dotted line represents the approximate resolution limit of the gel; any mutations below this line are not resolvable from the parental allele.

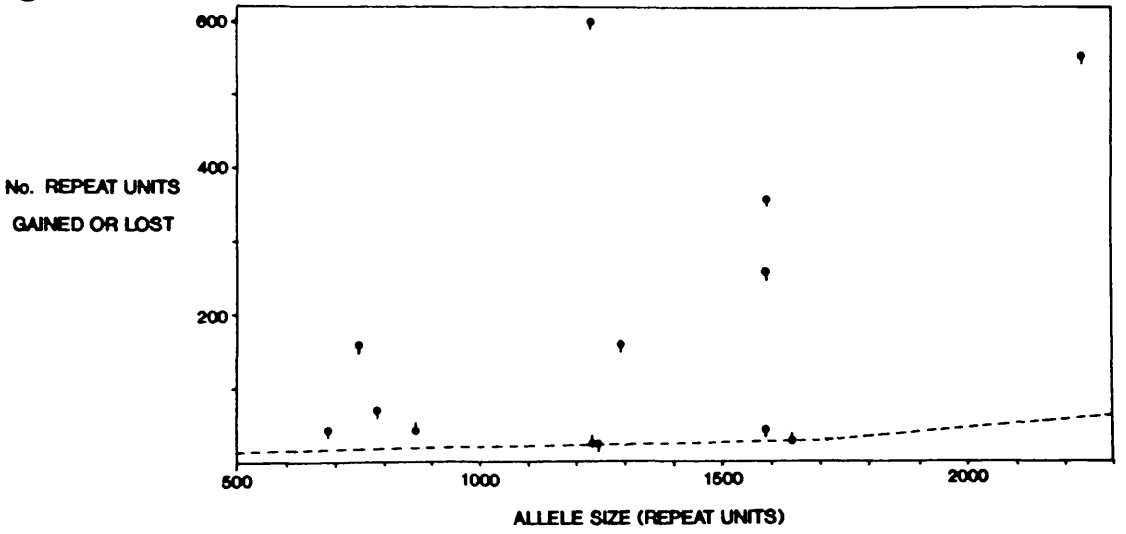
A


No. MUTANT ALLELES
4
3
2
1
0
-600 -500 -400 -300 -200 -100 +100 +200 +300 +400 +500 +600
No. REPEAT UNITS GAINED OR LOST
No. MUTANT ALLELES
2
1
0
0 20 40 60 80
% CHANGE IN ALLELE REPEAT LENGTH

B


No. SAMPLED FOR MUTATION
300
200
100
500 1000 1500 2000
ALLELE SIZE (REPEAT UNITS)

C


No. REPEAT UNITS GAINED OR LOST
600
400
200
500 1000 1500 2000
ALLELE SIZE (REPEAT UNITS)

parental allele in cases where the two parental alleles were different in size (e.g., C57BL/6J 26, Fig.6.1). Furthermore, if the level of mosaicism is high, the progenitor parental allele will be visibly under-represented in mosaic mice. In 5 of these 6 mosaic mice the new-length allele is derived from the paternally inherited allele, which does not represent a significant bias (p=0.22, two-tailed binomial).

Somatic length change events at *Ms6-hm* are summarised in Fig.6.2 as a function of all alleles scored for mutation. In common with germline mutations at this locus, there is no clear association between parental allele size and mutation rate, although no large length changes (>160 repeat units) are observed for alleles of <1000 repeat units (DBA/2J alleles, Fig.6.2). It is evident that there is a wider range of length change events associated with larger alleles; however no simple relationship emerges between parental allele size and either the rate or length change associated with somatic (or germline) mutation events at *Ms6-hm*.

6.4 Somatic mutation events at *Ms6-hm* occur in early development

Mice which are mosaic for cells carrying non-parental alleles at *Ms6-hm* in both the soma and the germline suggest that somatic mutation events at this locus precede the separation of these lineages, and therefore occur in early development. In order to investigate the distribution of somatic mutant alleles at *Ms6-hm* in different adult tissues, four mosaic mice were dissected (C57BL/6J mice 22, 26, and C57BL/6J x DBA/2J F_1 mice 37, 52). Of these mice three transmitted the non-parental allele to the next generation, and one (C57BL/6J 26) was not tested. In each case the dosage of the mutant allele was compared in a range of different somatic tissues (including brain, lung, kidney, liver, uterus, tail, and muscle and bone from the limbs, vertebral and sternal regions, Fig.6.3). In all four mice the progenitor and new-length mutant alleles were found to be present in indistinguishable ratios in all tissues tested. The different somatic tissues of each of these mice must therefore contain an equal proportion of cells carrying the non-parental allele; in addition, the somatic mutation event in each of these mice must have preceded the allocation of somatic lineages.

The dosage of the non-parental allele in the germline is not always equivalent to that in the soma. C57BL/6J x DBA/2J F_1 mouse 37, which is 40% mosaic for a non-parental allele in adult tail DNA, transmitted the mutant allele to only one of 39 offspring (Table 6.1). As this transmission frequency is significantly lower than that expected for 40% mosaicism (p=0.002), the degree of mosaicism in the germline of this mouse is likely to be much lower than that in somatic tissues. C57BL/6J x DBA/2J F_1 mouse 213 (8% mosaic in adult tail DNA) did not transmit the mutant allele to any of 37 offspring,

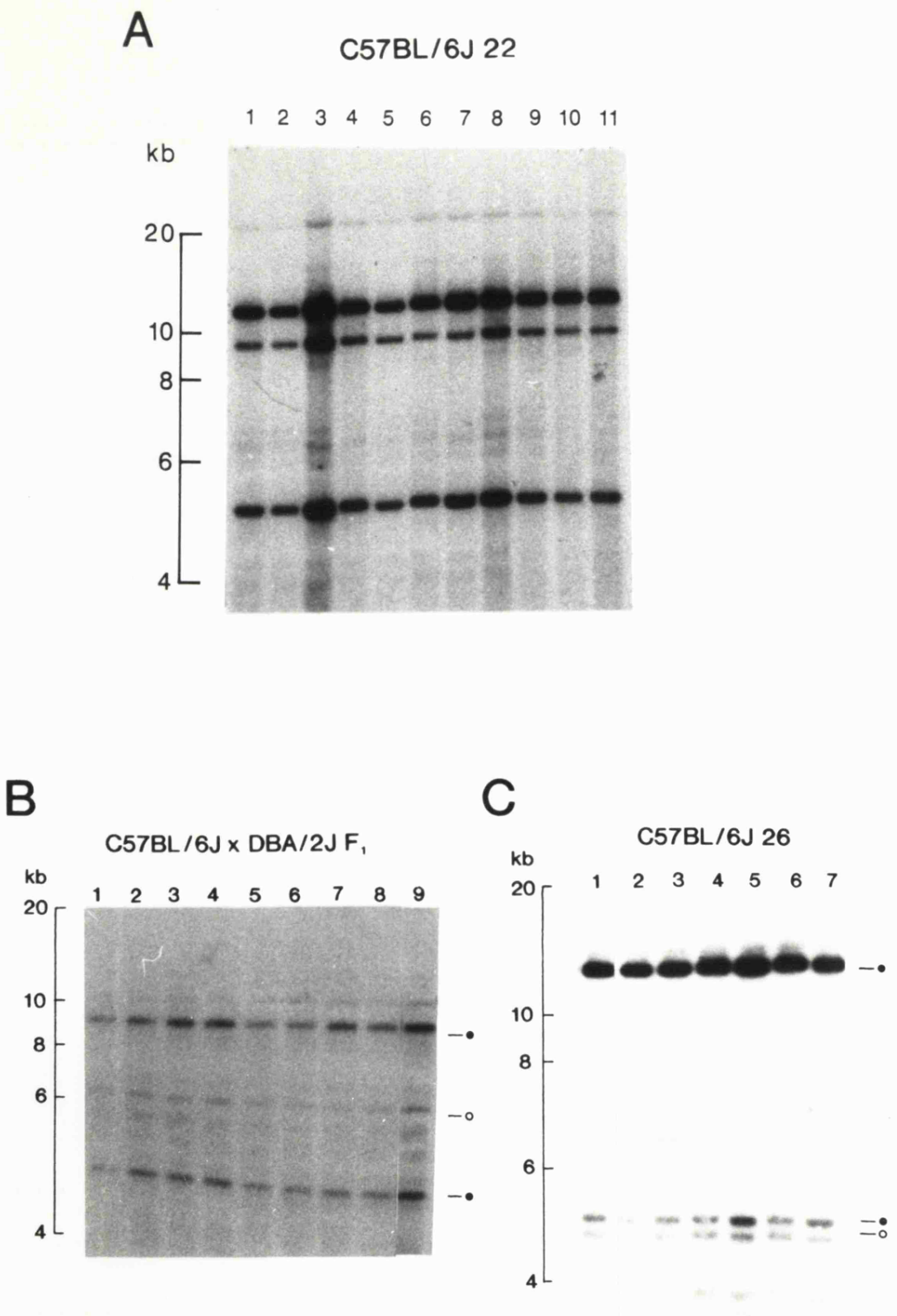
Figure 6.3

Distribution of somatic mutant alleles of *Ms6-hm* in adult tissues

A. DNA prepared from different regions of mouse C57BL/6J 22 digested with *Hin*fI and hybridised at high stringency with Mm3-1. 1, lower jaw; 2, brain; 3, midvertebral region; 4, sternal region; 5, liver; 6, left kidney; 7, right kidney; 8-11, tail sections. The dosage of the mutant allele is approximately equivalent in all 11 DNA preparations.

B. DNA prepared from different regions of C57BL/6J x DBA/2J F_1 mouse 37 digested with *Hin*fI and hybridised at high stringency with Mm3-1. Parental (•) and non-parental (○) alleles at *Ms6-hm* are marked. 1, left lung; 2, right lung; 3, sternal region; 4, front (l) leg muscle; 5, front (r) leg muscle; 6, heart; 7, liver; 8, uterus; 9, tail.

C. DNA prepared from different regions of C57BL/6J 26 digested with *Hin*fI and hybridised at high stringency with Mm3-1. Parental (•) and non-parental (○) alleles at *Ms6-hm* are marked. 1, liver; 2, kidney; 3, lower jaw; 4, midvertebral region; 5, sternal region; 6 and 7, tail sections.


A
C57BL/6J 22
1 2 3 4 5 6 7 8 9 10 11
kb
20
10
8
6
4
B
C57BL/6J x DBA/2J F1
kb
1 2 3 4 5 6 7 8 9
20
10
8
6
4
C
C57BL/6J 26
kb
1 2 3 4 5 6 7
20
10
8
6
4

however this is not a significant deviation from the expected transmission frequency (p=0.22).

6.5 Somatic mutation events *in vitro*

Cell lines derived from embryonic carcinomas (EC cell lines) provide an *in vitro* model for differentiation in the early mouse embryo (Hogan *et al.*, 1983). The F9 EC cell line is derived from a teratocarcinoma isolated from a 129/sv mouse (Bernstine *et al.*, 1973). F9 cells can be induced by retinoic acid to differentiate into cells similar to the extraembryonic endoderm of the parietal and visceral yolk sacks which surround the developing mouse embryo (Hogan *et al.*, 1983). Since somatic mutation events at *Ms6-hm* appear to be selectively restricted to the early embryo, the stability of this locus was studied in F9 and related cell lines.

F9 and PCC4 cell lines were isolated from different explants of the same *in vivo* transferred tumour. The cell lines PCC4.IRA, PCC4.Neo, MCP-6, and MP2H4 were derived from PCC4 cells as illustrated in Fig.6.4; the number of generations between cloning and DNA preparation in these cell lines was unknown. DNA from F9 cells, and PCC4 derived cell lines, was hybridised with Mm3-1 (Fig.6.4). This analysis was complicated by the presence of a second locus in the DNA of 129 mice which cross-hybridises strongly to Mm3-1 at high stringency. Using the three enzyme locus-specific signature test it was shown that alleles at *Ms6-hm* are indistinguishable and therefore stable in the 5 DNA samples screened. This may reflect both the small *Ms6-hm* allele size in 129 mice (325-375 repeat units) and the limited sample of cell lines.

The second locus, in contrast, was highly unstable (Fig.6.4). Both alleles differ between F9 and PCC4.IRA cells (perhaps the result of *in vivo* mutation events), and one allele further mutated *in vitro* in MCP-6 EC cells. *Hin*fI alleles at this locus (*129-LL*) lie in the range 3-5kb, and are recombinationally seperable from *Ms6-hm* alleles. *129-LL* is further described in Chapter 7.

6.6 Application of the Polymerase Chain Reaction to the analysis of somatic mutation events at *Ms6-hm* in early development

Early somatic mutation events at *Ms6-hm* precede the allocation of somatic and germ cell lineages; however the precise time in development at which such events occur remains unknown. It is possible that these events are confined to the first few cell divisions of the fertilised egg. By screening for non-parental alleles at *Ms6-hm* in products of the morula additional to the embryo, such as the trophectoderm and

Figure 6.4

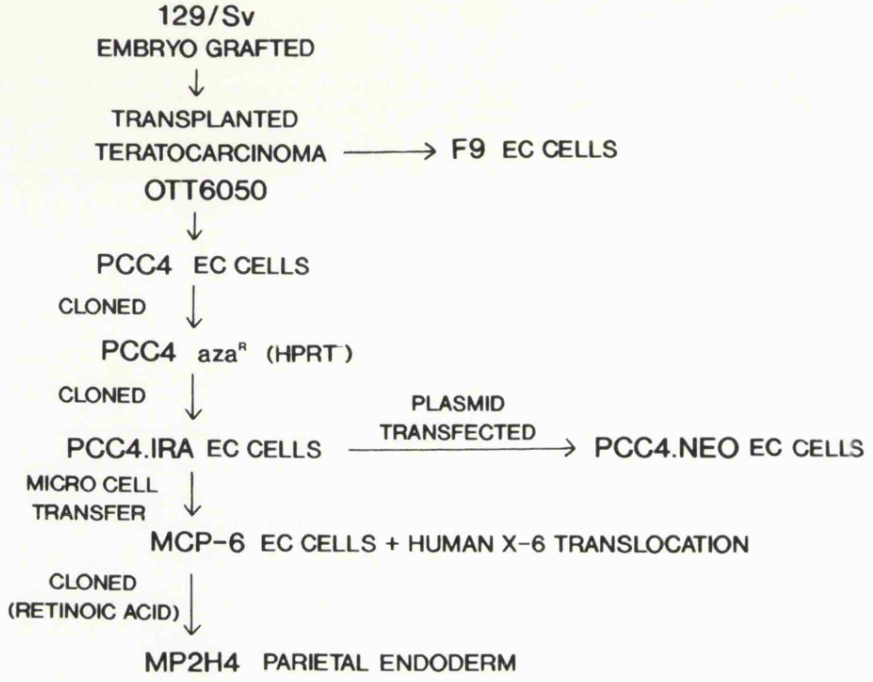
Somatic mutation in embryonic carcinoma cell lines

A. History of F9, PCC4 and PCC4 derived embryo carcinoma (EC) cell lines. The F9 and PCC4 EC cell lines were established in culture from different explants of the *in vivo* transferred tumour OTT6050, itself derived from a 6-day male embryo transplanted into the testis (see Hogan *et al.*, 1983). All the derived cell lines are EC lines except for MP2H4 which is a differentiated (parietal endoderm) clone from MCP-6. HPRT, hypoxanthinephosphoribosyltransferase; aza^R, 8-azaguanine resistant.

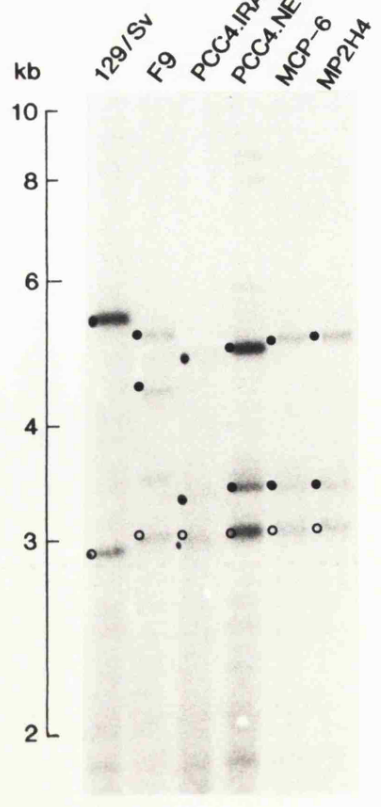
B. DNA from F9, PCC4 and PCC4 derived cell lines digested with *Hin*fI and hybridised at high stringency with Mm3-1. Two strongly hybridising DNA fragments are seen in 129/Sv DNA; the smaller fragment (o) is derived from *Ms6-hm*, and the larger fragment (•) is derived from a second locus, *129-LL*. As this DNA is not from the original 129/Sv mouse from which the F9 and PCC4 EC cells were derived, and alleles at *Ms6-hm* and *129-LL* are polymorphic among 129 mice, fragments derived from these loci differ beteen lane 1 and the cell line DNAs. While *Ms6-hm* alleles are stable in the different cell lines, fragments derived from *129-LL* are extremely unstable.

C. *Hin*fI (H), *Alu*I (A), and *Sau*3AI digests of 129/Sv, F9 and PCC4.Neo DNA hybridised at high stringency with Mm3-1. 129 mice have a DBA/2J-type signature at *Ms6-hm* (i.e., the polymorphic *Alu*I site adjacent to the minisatellite is present). The signature of DNA fragments derived from *129-LL* (•) is distinct from that of alleles at *Ms6-hm* (o), allowing each locus to be distinguished in the different cell lines DNAs.

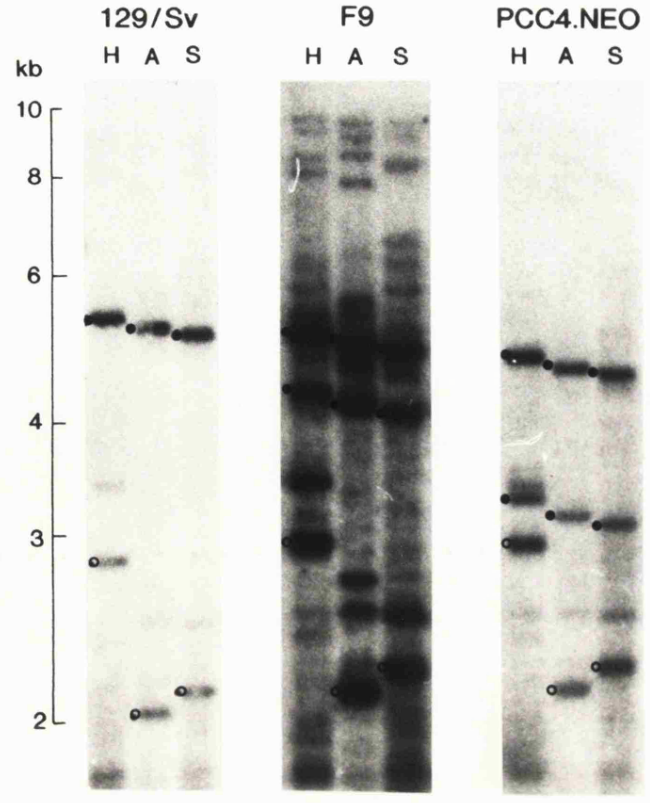
A


129/Sv
EMBRYO GRAFTED
TRANSPLANTED
TERATOCARCINOMA
F9 EC CELLS
OTT6050
PCC4 EC CELLS
CLONED
PCC4 aza^{R} (HPRT^{-})
CLONED
PLASMID
TRANSFECTED
PCC4.IRA EC CELLS
PCC4.NEO EC CELLS
MICRO CELL
TRANSFER
MCP-6 EC CELLS + HUMAN X-6 TRANSLOCATION
CLONED
(RETINOIC ACID)
MP2H4 PARIETAL ENDODERM

B


kb
129/Sv
F9
PCC4.IRA
PCC4.NEO
MCP-6
MP2H4
10
8
6
4
3
2

C


kb
129/Sv
F9
PCC4.NEO
H A S
H A S
H A S
10
8
6
4
3
2

extraembryonic tissues, it should be possible to define more precisely the developmental window within which somatic length changes at this locus may take place. The polymerase chain reaction (PCR) enables the amplification of DNA fragments from very small quantities of DNA (down to a single cell, Saiki *et al.*, 1988), and may be used to examine alleles at *Ms6-hm* in minute DNA samples from different zygotic tissues.

Tandem repeated minisatellite sequences may be successfully amplified by PCR (Jeffreys *et al.*, 1988b). PCR amplification, however, is efficient only for small alleles (<5kb). Two primers flanking *Ms6-hm* were synthesised. Each was 24 nucleotides long; one ended 5bp 5' to MT-1, the MT element within which *Ms6-hm* is embedded, and the other lay within MT-1, at a sequence divergent among MT elements (Fig.6.5). Alleles at *Ms6-hm* were amplified using these primers with denaturation at 96° for 1.2 minutes due to the high GC content (80%) of the tandem repeat sequence. Amplified alleles were detected by hybridisation with a $(GGGCA)_n$ probe to reduce interference from spurious amplification products. This probe was synthesised by annealing and ligating two overlapping oligonucleotides of complementary sequence which each contained 4 repeat units of 5 nucleotides. In addition, by specifically amplifying alleles from *Ms6-hm*, cross-hybridising loci (which would normally be detected in restriction endonuclease digests of genomic DNA) will not complicate the analysis.

Initial experiments using a pool of different inbred mouse strain DNAs demonstrated the range of *Ms6-hm* allele sizes which may be amplified by PCR (Fig.6.5). AKR (1.5kb), BALB/c (2.2kb), CBA (2.7kb), and DBA/2J (3.2 and 3.7kb) alleles could be amplified, but C57BL/6J alleles (6 and 9kb) were too large to be detected even after 25 cycles. Amplifications of *Ms6-hm* were not highly efficient; this might be improved if both primers lay outside the MT element. Alternatively, GC rich sequences may not denature efficiently even at 96°, causing reduced amplification of alleles at this locus.

AKR and BALB/c mice containing small (1-2.5kb) and more efficiently amplified *Ms6-hm* alleles were chosen for further experiments. Tail DNA was amplified from an AKR family. Amplified alleles were just visible on an ethidium stained gel after 24 cycles (input 100ng genomic DNA), and hybridisation confirmed that these represented alleles at *Ms6-hm* (Fig.6.5). Preliminary experiments were carried out to investigate whether this approach would be suitable to examine early zygotic tissues. As the embryo is derived from a subset of blastomeres, a higher frequency of mosaicism might be expected if DNA isolated from the whole blastocyst was screened. 2-3 month old BALB/c females and AKR males were mated, and the females checked for vaginal plugs. 7 days after fertilisation the implanted blastocysts were removed from pregnant females for DNA preparation. Blastocyst and parental tail DNA was amplified for 20-25 cycles, and Southern blot hybridised with a synthetic $(GGGCA)_n$ probe.

Figure 6.5

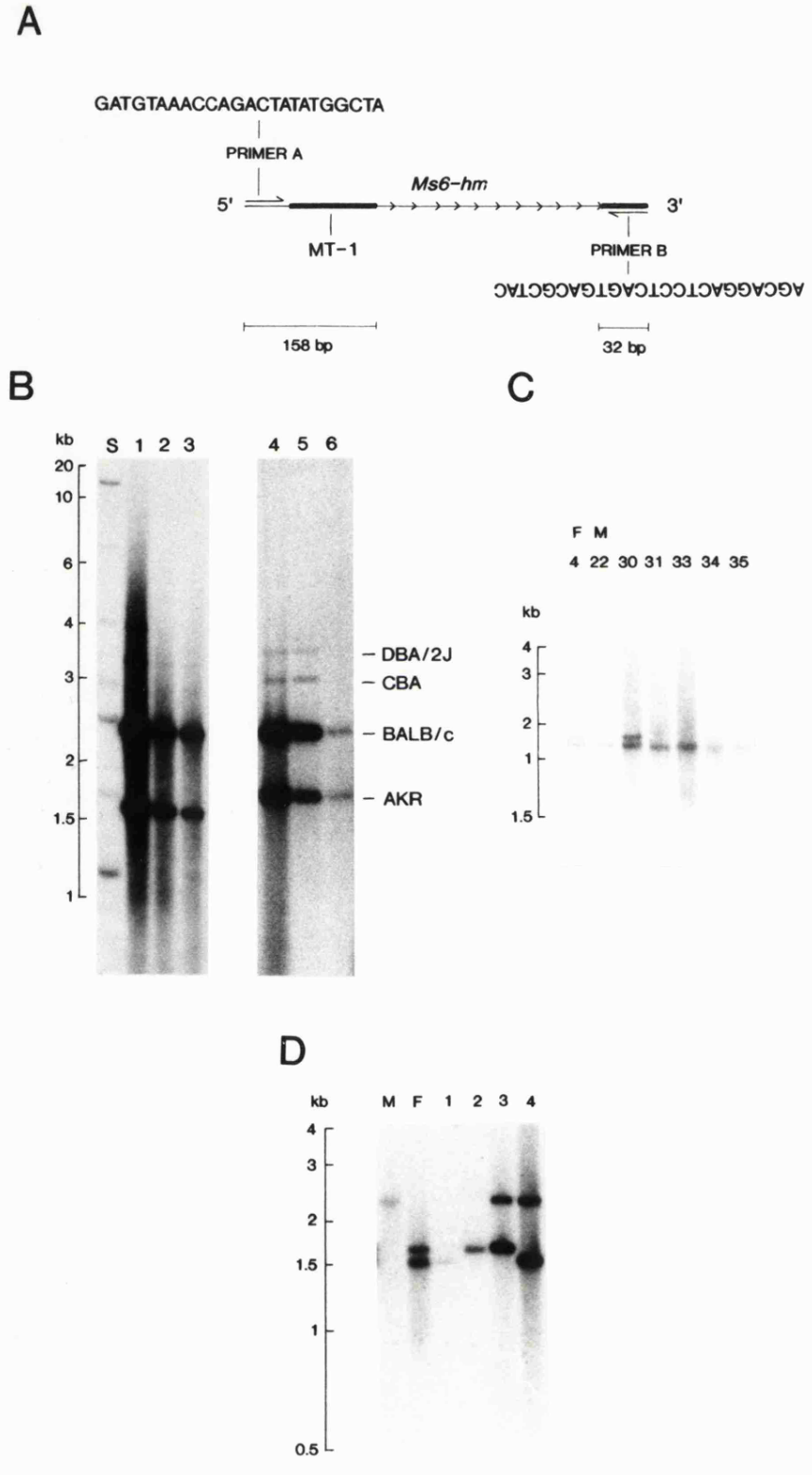
Amplification of *Ms6-hm* alleles using the polymerase chain reaction

A. The DNA sequence and position of the 5' and 3' flanking 24mer oligonucleotide primers used to amplify alleles at *Ms6-hm* by PCR, showing the minisatellite repeat units (arrows) and position of MT-1 (filled box).

B. Amplification of *Ms6-hm* alleles by PCR. 0.5-50ng aliquots of DNA from 5 individual inbred mice, each from a different strain (AKR, BALB/c, CBA, DBA/2J and C57BL/6J), together containing 7 different *Ms6-hm* alleles ranging from 1.5 to 9.0kb were pooled and amplified for 15 cycles in 10μl reactions containing 1 unit of Taq polymerase plus the flanking primers illustrated in A. PCR products were electrophoresed in a 1% agarose gel and Southern blot hybridised with a synthetic $(GGGCA)_n$ probe. S, 2μg of each mouse DNA digested with *Sau*3AI; *Sau*3AI sites flanking *Ms6-hm* are positioned such that each *Sau*3AI allele is 0.1kb longer than its corresponding PCR product. 1 and 2, amplifications of 50ng aliquots of each mouse DNA with polymerase extention time at 70° for 10min, primer annealing at 60° for 1min, and denaturation at 95° for 1.5min (1) and 2min (2). 3, as (1), except primer annealing was extended to 2min. 4, 5, and 6, 50, 5 and 0.5ng of each DNA pooled and amplified for 15 cycles with denaturation at 96° for 1.2min, primer annealing at 60° for 1min, and polymerase extention at 70° for 5min.

C. Amplification of *Ms6-hm* alleles in an AKR pedigree by PCR. 0.1μg of each DNA was amplified for 24 cycles (as in B.1, above), reaction products electrophoresed on a 1.2% agarose gel, and Southern blot hybridised with a synthetic $(GGGCA)_n$ probe. These fragments were just visible on the agarose gel. Autoradiography was for 30min.

D. Amplification of alleles at *Ms6-hm* in AKR x BALB/c 7 day blastocysts. 5ng of parental tail DNA and 1/10 DNA prepared from each isolated blastocyst (1-5ng) was amplified for 20 cycles as in B.4-6 above. M, BALB/c mother; F, AKR father; 1-4, 7 day blastocysts.


A
GATGTAAACCAGACTATATGGCTA
PRIMER A
Ms6-hm
5'
3'
MT-1
PRIMER B
AGCAGGACTCCTCAGTGACGCTAC
158 bp
32 bp
B
kb
S 1 2 3
4 5 6
20
10
6
4
3
2
1.5
1
DBA/2J
CBA
BALB/c
AKR
C
F M
4 22 30 31 33 34 35
kb
4
3
2
1
1.5
D
kb
M F 1 2 3 4
4
3
2
1.5
1
0.5

The result from one such experiment is illustrated in Fig.6.5; parental alleles, and a subset of blastocyst alleles, were clearly amplified to a detectable level. However it would be necessary to achieve more efficient amplification to ensure that all *Ms6-hm* alleles in each sample are successfully amplified, and to reduce the background hybridisation smear, before carrying out a larger analysis using this approach. It should then be possible to compare *Ms6-hm* alleles in trophoblast, embryo, and yolk sack DNA from the same blastocyst. However as only 2.8% of mice have a non-parental allele at *Ms6-hm* (and it remains to be seen whether this frequency may be enhanced by screening the whole blastocyst rather than just the embryo) a large number of experiments may need to be carried out to observe even a few mosaic blastocysts. It is also possible that the rate of somatic mutation for large alleles (e.g., C57BL/6J and DBA/2J) is higher than that for small alleles (e.g., AKR and BALB/c), thus detracting from the advantages of using PCR.

6.7 Summary

Mutation events at *Ms6-hm* are not confined to the germline but can also occur during early mouse development. Such somatic mutation events result in mice which are mosaic for cells carrying a common non-parental allele at *Ms6-hm*. These mutation events are similar in rate and average length change to germline events at *Ms6-hm*. It should be possible to study somatic mutation events in cell culture, and in post-implantation embryos, using the polymerase chain reaction; at present, however, the significance and precise timing of somatic mutation events at *Ms6-hm* remain unknown. As yet no similar early somatic mutation events have been observed at any human minisatellite loci, despite screening of DNA from thousands of individuals with several locus-specific probes.

VII. MM3-1 : A MULTI-LOCUS PROBE

7.1 Introduction

DNA fingerprinting probes 33.6 and 33.15 are each derived from a single human minisatellite locus (Jeffreys *et al.*, 1985a). Both loci contain a common G-rich 'core' sequence embedded within each repeat unit; probe 33.15 is comprised of 29 repeats of a 16bp variant of the core sequence, and the 36bp repeat unit of probe 33.6 contains three core-like subunits and is repeated 18 times (Jeffreys *et al.*, 1985a). The difference in sequence and repeat length of these probes results in their detecting different patterns of hypervariable minisatellite fragments in human DNA at low stringency. While essentially any tandemly repeated oligonucleotide probe may, to a varying extent, detect multiple loci in human DNA (Vassart *et al.*, 1987, Vergnaud, 1989), the most informative and variable DNA fingerprints are those generated by G-rich probes related to the core sequence (GGAGGTGGGCAGGARG, Jeffreys *et al.*, 1985a). Correspondingly, the GGGCA repeat unit of *Ms6-hm* cross-hybridises to many highly variable minisatellite loci in the mouse genome to generate a novel and highly individual specific mouse DNA fingerprint.

7.2 A novel DNA fingerprint in mice

Subclone Mm3-1, derived from *Ms6-hm*, contains 19 GGGCA repeat units and cross-hybridises weakly to other variable DNA fragments in the mouse genome, under conditions of high stringency (filters washed in 0.1x SSC, 65°). This requires that the *Hin*fI, *Alu*I, and *Sau*3AI locus-specific signature test be used to confirm whether novel DNA fragments observed in pedigree analysis are alleles at *Ms6-hm* or are derived from cross-hybridising loci (see Chapter 3). However Mm3-1 may also be used as a multi-locus probe; under low-stringency washing conditions (1x SSC, 65°), Mm3-1 detects multiple variable loci in mouse genomic DNA to produce a novel DNA fingerprint (Fig.7.1).

The DNA fragments detected by Mm3-1 at low stringency segregate within pedigrees in a Mendelian fashion, and assort independently at meiosis suggesting that they are derived from recombinationally separable minisatellite loci (see section 7.3); several of these fragments are also highly unstable (see sections 7.4 and 7.5). The Mm3-1 DNA fingerprint is distinct from those generated by probes 33.6 and 33.15, and is therefore essentially comprised of a novel subset of mouse minisatellite loci. However, the loci detected by the three sequence-related probes overlap to some extent under the

Figure 7.1

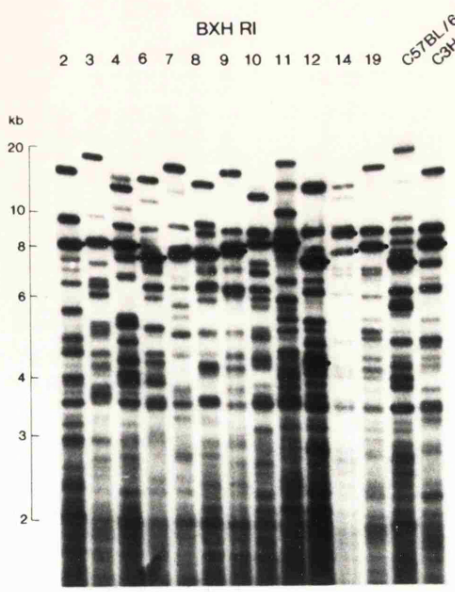
Segregation analysis of minisatellite fragments in BXH RI strains

A. Liver DNA from BXH RI and progenitor strains (C57BL/6J and C3H/HeJ) digested with *Hin*fI and hybridised with Mm3-1 at low stringency (1x SSC, 65°). Under these conditions Mm3-1 detects multiple variable loci to produce a novel DNA fingerprint. The pattern of cross-hybridising fragments is distinct from that detected by probes 33.6 or 33.15 (compare with B, below). Note that most of the DNA fragments larger than 4kb are transmitted to only one or a few RI strains, and that many new-length fragments are present in only one RI strain, such that no strain distribution patterns can be determined for these fragments. Alleles at *Ms6-hm* are marked (•).

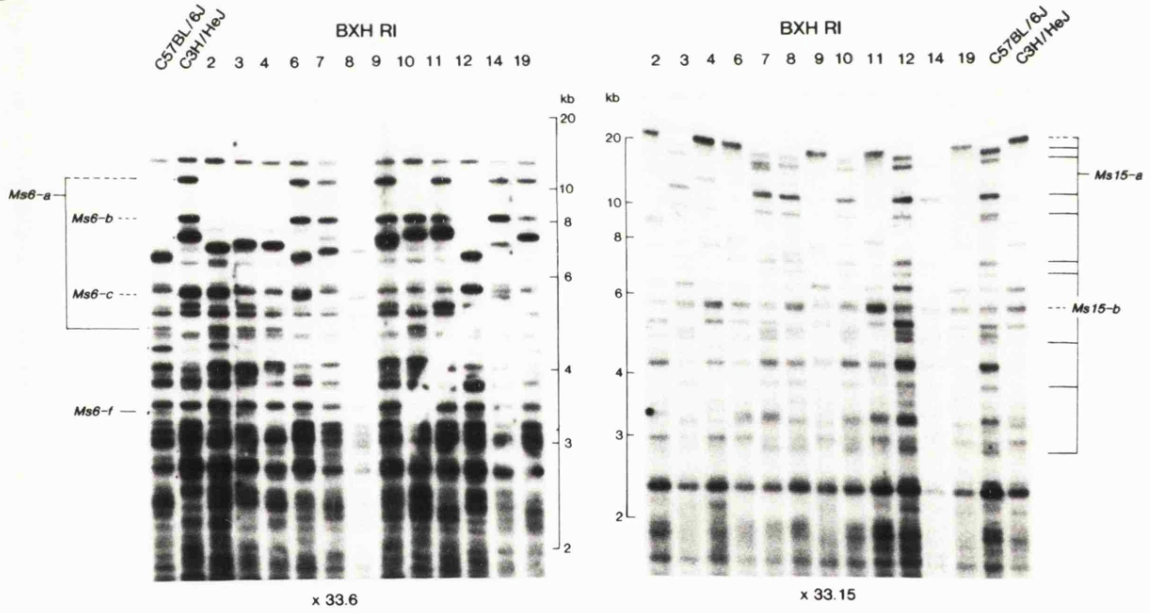
B. Liver DNA from BXH RI and progenitor strains digested with *Hin*fI and hybridised with 33.6 or 33.15 at low stringency (1x SSC, 65°). Note that the majority of fragments detected by these probes are stably transmitted to approximately half the RI strains. Strain distribution patterns were obtained for the 16 progenitor fragments indicated; —, fragment present in C57BL/6J strain DNA; ---, fragment present in C3H/HeJ strain DNA. Additional fragments were either poorly resolved, shared by the progenitor strains, or absent from most or all RI strains, and were not scored. The 16 scored fragments define 6 different loci termed *Ms6-a*, *Ms6-b*, *Ms6-c*, *Ms6-f*, *Ms15-a*, and *Ms15-b*. Allelic and linked fragments are joined by vertical lines. Locus *Ms15-a* is also termed *Ms15-1* and is described in Chapter 8.

C. Table showing the strain distribution patterns of minisatellite loci in the BXH RI strains, corresponding to the cross-hybridising variable fragments shown in B, and linkage between these loci and other genetic markers typed in BXH strains. Allelic fragments were detected only for *Ms6-a* and *Ms15-a*; each other locus was resolved in only one of the two progenitor strains. The estimated map distance between linked loci was calculated according to standard methods (Taylor, 1978), and 95% confidence limits were taken from a published table (Silver, 1985). References for the linked markers; (1) D'Eustachio, 1984; (2) B.Taylor, The Jackson Laboratory, personal communication; (3) Lusis *et al.*, 1987; (4) Mishkin *et al.*, 1976; (5) Jenkins *et al.*, 1982.

A



B



C

STRAIN DISTRIBUTION PATTERNS OF MINISATELLITE LOCI IN THE BXH RI STRAINS

Locus	Scored in	2	3	4	6	7	8	9	10	11	12	14	19
Ms6-a	BH	B	B	B	H	H	B	H	B	H	B	H	H
Ms6-b	H	B	B	B	H	H	B	H	H	H	B	H	H
Ms6-c	H	H	B	B	H	B	H	B	B	B	B	H	B
Ms6-f	B	B	B	H	B	H	H	B	B	H	B	B	H
Ms15-a	BH	H	B	H	H	B	B	H	B	H	B	B	H
Ms15-b	H	B	H	B	B	B	B	H	B	B	H	B	H

LINKAGE TO OTHER GENETIC MARKERS

Locus	Chromosome	Linked marker (ref)	Recombinants /total	Map distance (cM) 95% confidence limits
Ms6-a	12	*D12Nyu-2* (1)	0/12	0.0 (0.0-11.0)
	4	*Lck* (2)	1/12	2.4 (0.1-22.8)
Ms6-b	12	*Apob* (3)	0/12	0.0 (0.0-11.0)
Ms6-f	16	*Pmv-35* (2)	0/12	0.0 (0.0-11.0)
	11	*Es-3* (4)	1/12	2.4 (0.1-22.8)
Ms15-a	4	*Pmv-23* (2)	1/12	2.4 (0.1-22.8)
Ms15-b	8	*Es-1* (5)	1/12	2.4 (0.1-22.8)

hybridisation conditions used in this study, i.e., in the absence of competitor DNA (apart from high molecular weight human or herring sperm carrier DNA used to recover labelled probe, at approximately 1.5μg/ml in overnight hybridisations). Very approximately 10-25% of loci detected by Mm3-1 in *Hin*fI digests of mouse DNA are also detected by probe 33.6, and approximately 25-40% by 33.15 (Fig.7.1). These estimates are based on the number of bands larger than 4kb which are detected by both probes within the DNA fingerprints of individual mice, and are likely to overestimate the number of loci held in common due to chance co-migration of fragments detected by different probes. One locus in particular, *Ms6-hm*, is detected by all three probes under these hybridisation conditions.

The 12 BXH recombinant inbred strains are derived from C57BL/6J and C3H/HeJ progenitors, and have been strictly brother-sister mated for 71-86 generations (Taylor, 1989). The DNA fingerprints of these RI strains, and the two progenitor strains, detected by probes 33.6, 33.15 and Mm3-1 at low stringency, are illustrated in Fig.7.1. The majority of loci detected by probes 33.6 and 33.15 in the progenitor DNAs are stably transmitted to approximately half the RI strains, and the strain distribution patterns of the most clearly resolved fragments (and putative linkage assignments) are illustrated in Fig.7.1. In contrast, most of the DNA fragments larger than 4kb detected by probe Mm3-1 in C57BL/6J and C3H/HeJ DNA are transmitted to only one or a few RI strains. Many new-length fragments are present in only one RI strain, such that no SDPs can be assigned to these DNA fragments. This result suggests that many of the cross-hybridising fragments are associated with a high rate of germline mutation to new-length alleles during the breeding of the BXH RI strains.

The low stringency hybridisation of Mm3-1 to the BXD RI panel is illustrated in Fig.7.2. The Mm3-1 DNA fingerprints of both the BXH and BXD RI panels are as variable as those of a series of unrelated inbred strains. In contrast, probes 33.6 and 33.15 (and almost all other conventional polymorphisms tested) detect, in the same RI panels, with few exceptions, loci which are stably inherited from the progenitor strains (Jeffreys *et al.*, 1987, Taylor, 1989). Thus Mm3-1 detects a more unstable subset of minisatellite loci in the mouse genome than either probe 33.6 or 33.15. The same DNA fingerprints are detected by a synthetic $(GGGCA)_n$ repeat probe, suggesting that the cross-hybridising loci are detected solely by the GGGCA tandem repeat component of Mm3-1 (Fig.7.2).

7.3 Segregation analysis of multiple DNA markers in mouse pedigrees

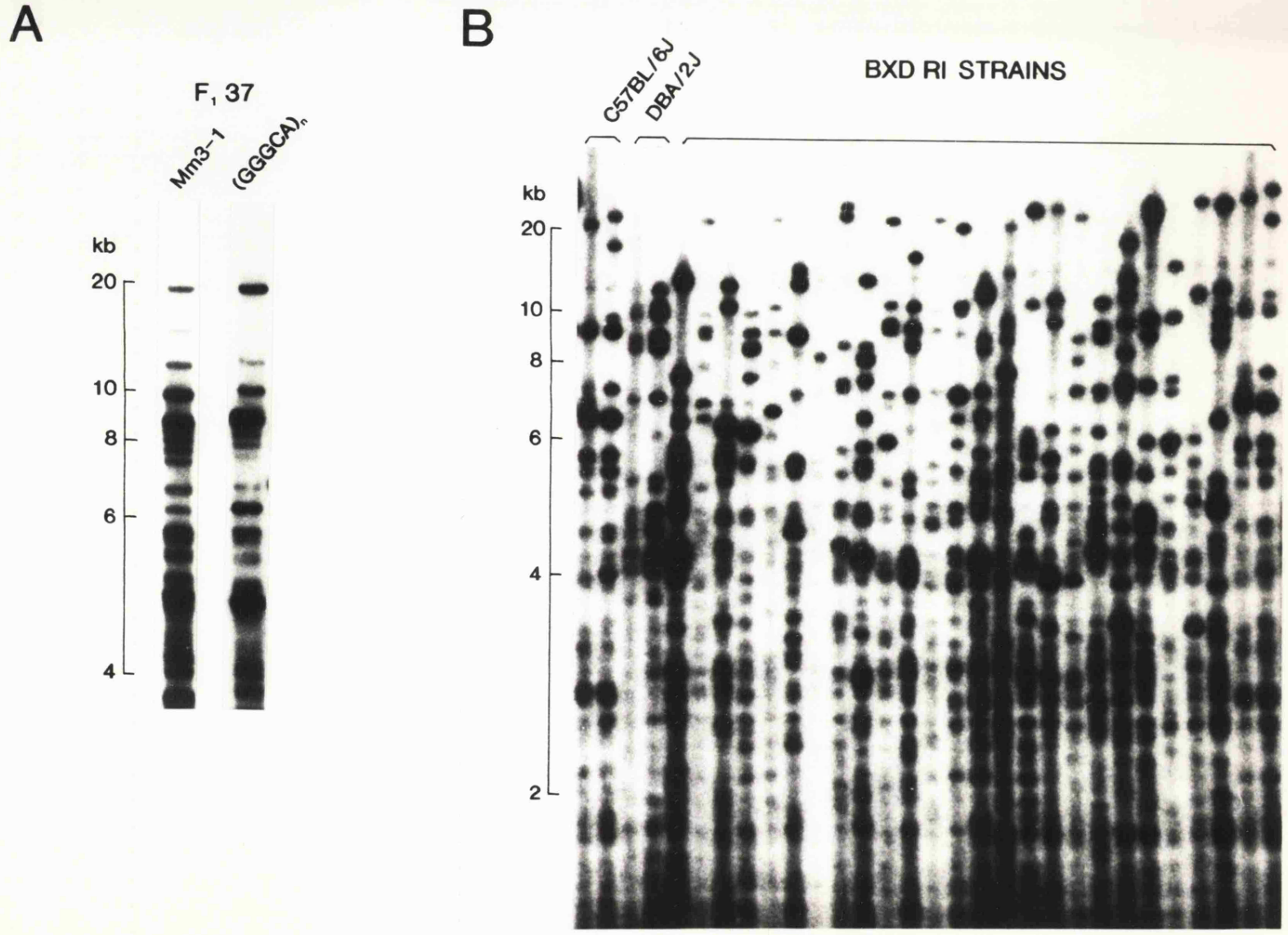
Inbred mouse pedigrees, designed for mutational and linkage analyses at *Ms6-hm* (see Chapter 5), were used to study the segregation of individual minisatellite fragments

Figure 7.2

$(GGGCA)_n$ DNA fingerprints

A. Comparison of DNA fingerprints generated by probes Mm3-1 and $(GGGCA)_n$. DNA from C57BL/6J x DBA/2J F_1 mouse 37 digested with *Hin*fI and hybridised with Mm3-1 or $(GGGCA)_n$. Differences in hybridisation intensity of particular DNA fragments are routinely observed in different hybridisations with either individual probe; for example, the faintly hybridising second largest fragment detected by Mm3-1 is detected to a varying extent by both probes, although not clearly visible in the $(GGGCA)_n$ track illustrated. Apart from such fluctuations in hybridisation intensity, these probes detect an indistinguishable set of minisatellite loci in the mouse genome.

B. Liver DNA from the BXD RI and progenitor strains (C57BL/6J and DBA/2J) digested with *Hin*fI and hybridised with Mm3-1 at low stringeny (1x SSC, 65°). Note that most of the DNA fragments larger than 4kb are transmitted to only one or a few RI strains, and that many new-length fragments are present in only one RI strain, such that no strain distribution patterns can be determined for these fragments.


A
F1 37
Mm3-1
(GGGCA)n
kb
20
10
8
6
4
B
C57BL/6J
DBA/2J
BXD RI STRAINS
kb
20
10
8
6
4
2

detected by probe Mm3-1 at low stringency. Any individual inbred mouse is likely to be homozygous at most minisatellite loci, including even the highly unstable loci detected by probe Mm3-1. Thus every C57BL/6J x DBA/2J F_1 mouse has a complex *Hin*fI DNA fingerprint composed of almost all the bands present in each parent (Fig.7.3). A minority of the more unstable loci (such as *Ms6-hm*) may be heterozygous in the founder inbred mice and will segregate to approximately half the F_1 mice. However, as each F_1 mouse is heterozygous for a C57BL/6J and a DBA/2J allele at every locus, the F_1 x F_1 crosses provide informative segregation data for any DNA fragments not shared by the two individual F_1 mice mated (Fig.7.3).

The segregation of paternal and maternal bands was analysed in two of the larger F_2 families. Only those DNA fragments larger than 6kb were clearly resolved and were used in this analysis (8-12 bands per F_1 parent). Of these DNA fragments, 3-6 were specific to each F_1 parent (the remainder were shared). Paternal and maternal specific fragments were transmitted to approximately half the offspring, consistent with 1:1 segregation (see Fig.7.3). The number of F_2 mice receiving each DNA fragment followed the expected binomial distribution, in which the proportion of parental fragments that are transmitted to precisely r of n F_2 mice is ${}^nC_r/2^n$ (Fig.7.3). This observation is consistent with the Mendelian inheritance of these minisatellite loci.

By pairwise comparison of the segregation patterns for the large paternal and maternal specific DNA fragments in these families it is possible to identify allelic pairs of fragments, as well as fragments showing tight linkage in coupling (see Jeffreys *et al.*, 1986). No evidence for allelism among either paternal or maternal fragments larger than 6kb was found. Presumably the fragments which are allelic to those scored are among the many unresolved fragments smaller than 6kb. This is the case for *Hin*fI alleles at *Ms6-hm*; C57BL/6J alleles (of 7-8kb) are scored in this analysis, while DBA/2J alleles (4-5kb) are not.

The observed pairwise segregation distribution was consistent with that expected if the loci were unlinked, in which case the number of pairwise comparisons giving r concordant offspring (out of n offspring) for two fragments is given by the binomial distribution ${}^nC_r/2^n.[L(L-1)/2]$, (where L is the number of paternal (or maternal) specific fragments scored, Jeffreys *et al.*, 1986). The number of loci contributing to the entire Mm3-1 DNA fingerprint, including unresolved and therefore unscored fragments, is related to the number of fragments, and the number of linked and allelic pairs of fragments, which are scored. This cannot be estimated from the present data, because no alleles were detected among the large fragments scored. However it is likely that there are at least 30 contributing loci, which is the number of heterozygous loci estimated by similar pedigree analysis to be detected in human DNA by probe 33.6 or 33.15 (Jeffreys

Figure 7.3

Segregation analysis of minisatellite fragments detected by Mm3-1 in mouse pedigrees

A. Tail DNA from C57BL/6J (father), DBA/2J (mother) and F_1 mice digested with *Hin*fI and hybridised with Mm3-1 at low stringency (1x SSC, 65°). The 10 F_1 mice have complex DNA fingerprints composed of most of the fragments present in each parent, confirming that the parents are homozygous at most cross-hybridising minisatellite loci. Mouse 34 has a new-length mutant fragment (•) derived from the largest paternal fragment.

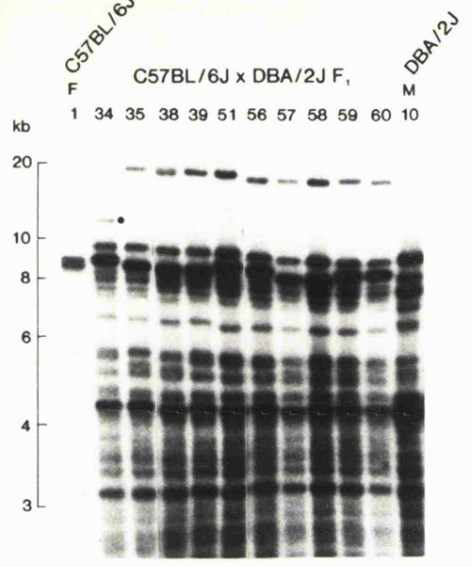
B. Tail DNA from a three generation C57BL/6J x DBA/2J pedigree digested with *Hin*fI and hybridised with Mm3-1 at low stringency (1x SSC, 65°). F_1 mice 44 and 62 are heterozygous at most cross-hybridising loci and the segregation of paternal and maternal fragments can be followed in their offspring. Those DNA fragments larger than 6kb which were analysed in the F_2 mice are indicated; fragments smaller than 6kb are poorly resolved. DNA fragments marked (•) are paternal-specific, and those marked (o) maternal-specific; other fragments were shared by the F_1 parents.

Note that the largest paternal-specific fragment is not transmitted to any of 8 F_2 mice; this fragment is absent from the DNA of the father's parents (2 and 11) and has been shown by signature to be a somatic mutant allele at *Hm-2* (derived from the largest cross-hybridising fragment shared by 44 and 62, see Fig.7.6). 8 (MM, mother's mother), C57BL/6J; 21 (MF, mother's father), DBA/2J; 11 (FM, father's mother), DBA/2J; 2 (FF father's father), C57BL/6J; 44 (M, mother), 62 (F, father), C57BL/6J x DBA/2J F_1 mice; 48-55, F_2 mice.

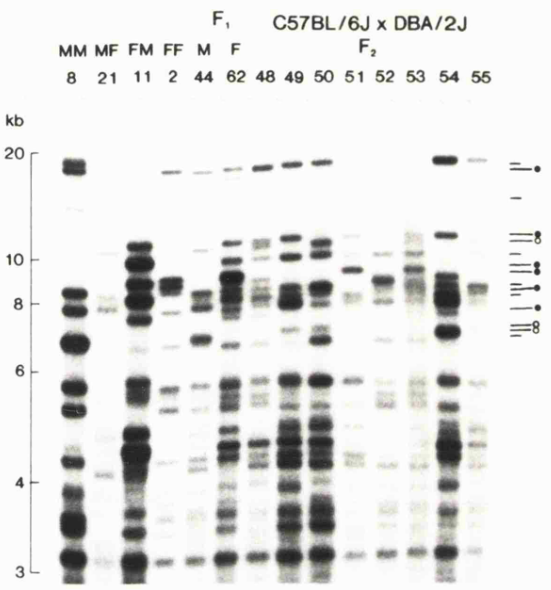
C. Segregation of minisatellite fragments in two C57BL/6J x DBA/2J F_2 pedigrees, showing the number of paternal and maternal loci scored, the transmission frequency of paternal and maternal fragments, and linkage between pairs of fragments. The number of fragments detected by Mm3-1, out of those scored as specific to each parent, which were transmitted to precisely r of n offspring is compared with the expected number given by the binomial distribution ${}^nC_r/2^n$, assuming 50% transmission. The mean transmission frequencies (+/- SEM) are also given. The paternal fragment in pedigree A5 which is transmitted to none of 8 offspring is a somatic mutant allele at *Hm-2* (and is therefore expected to be transmitted to less than 50% of the offspring).

Linkage between pairs of fragments AB was investigated by scoring the number of offspring which were concordant for AB (either AB or --), using all possible pairwise comparisons of paternal or maternal fragments (giving L(L-1)/2 pairwise comparisons for L fragments analysed in each parent). Pairs of fragments falling into zero- or all-offspring classes represent alleles or tighly linked pairs, respectively. The observed distribution is compared with that expected if all L fragments are derived from unlinked loci, in which case the number of pairwise comparisons giving r (AB or --) offspring is given by the binomial distribution $({}^nC_r/2^n.[L(L-1)/2]$. The fragments analysed in pedigree A5 are those illustrated in B above.

A



B



C

C57BL/6J x DBA/2J F_2 Family B9

	Maternal	Paternal
No. fragments scored	12	8
No. parental specific fragments	6	3

	MATERNAL				PATERNAL			
	single fragment		pair (AB or --)		single fragment		pair (AB or --)	
Transmission to No. offspring	Obs.	Exp.	Obs.	Exp.	Obs.	Exp.	Obs.	Exp.
0	0	0.006	0	0.015	0	0.003	0	0.003
1	0	0.06	0	0.15	1	0.03	0	0.03
2	0	0.3	0	0.7	0	0.1	0	0.1
3	1	0.7	1	1.8	0	0.4	1	0.4
4	1	1.2	4	3.1	0	0.6	0	0.6
5	2	1.5	6	3.7	1	0.7	1	0.7
6	1	1.2	2	3.1	1	0.6	1	0.6
7	0	0.7	2	1.8	0	0.4	0	0.4
8	1	0.3	0	0.7	0	0.1	0	0.1
9	0	0.06	0	0.15	0	0.03	0	0.03
10	0	0.006	0	0.015	0	0.003	0	0.003
Transmission frequency	51.7% (+/- 6.4%)				40.0% (+/- 12.5%)			

C57BL/6J x DBA/2J F_2 family A5

	Maternal	Paternal
No. fragments scored	9	12
No. parental specific fragments	3	6

	MATERNAL				PATERNAL			
	single fragment		pair (AB or --)		single fragment		pair (AB or --)	
Transmission to No. offspring	Obs.	Exp.	Obs.	Exp.	Obs.	Exp.	Obs.	Exp.
0	0	0.01	0	0.01	1	0.02	0	0.04
1	1	0.1	0	0.1	0	0.15	0	0.3
2	0	0.3	0	0.3	1	0.5	1	1.1
3	0	0.7	1	0.7	0	1.1	2	2.2
4	2	0.8	1	0.8	2	1.4	3	2.8
5	0	0.7	1	0.7	1	1.1	4	2.2
6	0	0.3	0	0.3	0	0.5	3	1.1
7	0	0.1	0	0.1	1	0.15	2	0.3
8	0	0.01	0	0.01	0	0.015	0	0.03
Transmission frequency	37.5% (+/- 10.2%)				55.0% (+/- 9.1%)			

et al., 1986). Furthermore, as these loci assort independently, they are likely to be substantially distributed in the mouse genome, providing highly informative DNA markers for linkage analysis.

7.4 Intrastrain variation within the Mm3-1 DNA fingerprint

The BXD and BXH RI analysis (see section 7.2) suggests that Mm3-1 detects a more unstable subset of minisatellite loci in mouse DNA than probes 33.6 or 33.15. As a result, highly individual-specific DNA patterns can be obtained with Mm3-1, even for mice from within an inbred strain (Fig.7.4).

*Hin*fI DNA fingerprints of different inbred strains are illustrated in Fig.7.4. DNA fragments larger than 4kb show a high level of variation between inbred strains, comparable with probe 33.6 and 33.15 DNA fingerprints (Jeffreys *et al.*, 1987). However, a higher level of intrastrain variation is observed within Mm3-1 DNA fingerprints than within 33.6 or 33.15 DNA fingerprints, such that 6 DBA/2J mice obtained from the same source could all be distinguished on the basis of their Mm3-1 DNA fingerprint (Fig.7.4). There is a range of variability among the cross-hybridising fragments, from those, like *Ms6-hm*, which are extremely unstable, to others, generally <4kb in length, which are shared between different inbred strains.

Mm3-1 therefore detects unprecedented levels of variation among mice which are generally considered to be isogenic. This variation presumably arises due to high germline mutation rates at the cross-hybridising loci. The ability to distinguish individual mice which are otherwise almost entirely inbred may be useful in many areas of research where inbred mouse strains are used, including the monitoring of genetic divergence in isolated stocks of the same strain.

7.5 New-length mutation events within the Mm3-1 DNA fingerprint

Instability within the Mm3-1 DNA fingerprint has been directly observed. In the C57BL/6J x DBA/2J pedigree analysis (see section 7.3), new-length fragments have been scored among both F_1 and F_2 mice which are absent from either parental DNAs (Fig.7.5). Under high stringency washing conditions (0.1x SSC, 65°) these novel fragments disappear, and therefore cannot be somatic mutant alleles at *Ms6-hm*. In several cases the mutant allele in an F_1 mouse was stably transmitted to the next generation (Fig.7.5). In most instances the origin of the new-length fragment cannot be determined due to the large number of cross-hybridising loci.

Mutant alleles at one locus, however, could be clearly observed. This minisatellite, termed *Hm-2* (second hypermutable locus), is defined by 15-20kb *Hin*fI

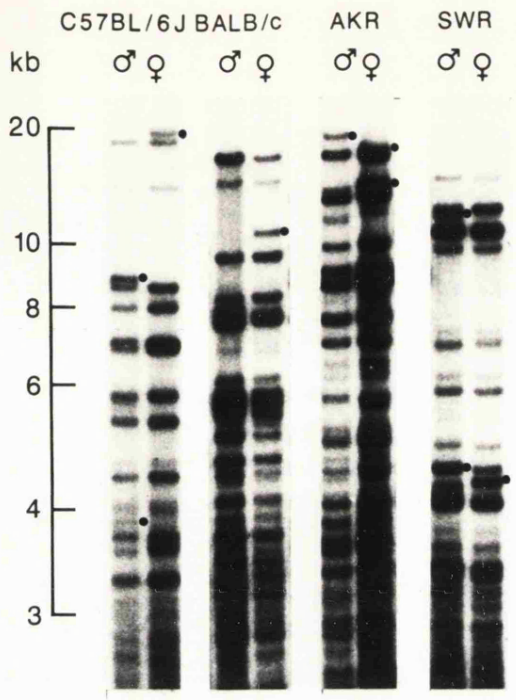
Figure 7.4

Intrastrain variation within the Mm3-1 DNA fingerprint

A. DNA from four pairs of mice from different inbred strains digested with *Hin*fI and hybridised with Mm3-1 at low stringency (1x SSC, 65°). Fragments varying among individuals from one strain are marked (•).

B. Comparison of high (0.1x SSC, 65°) and low (1x SSC, 65°) stringency hybridisation with Mm3-1. DBA/2J tail DNA was digested with *Hin*fI. The six DBA/2J mice were obtained from the same source. Under conditions of high stringency (0.1x SSC, 65°) *Ms6-hm* alleles are primarily detected (•); Mm3-1 hybridised to the same filter at low stringency (1x SSC, 65°) detects multiple fragments, many of which are shared, but of which at least three (o) are variable, allowing the six inbred mice to be individually distinguished.

A


C57BL/6J
BALB/c
AKR
SWR
kb
♂ ♀ ♂ ♀ ♂ ♀ ♂ ♀
20
10
8
6
4
3

B

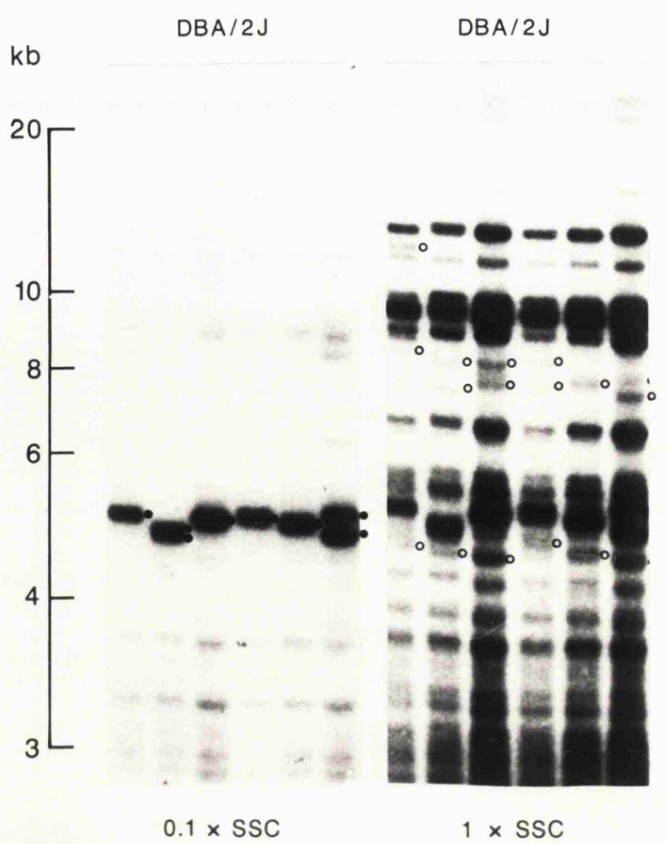
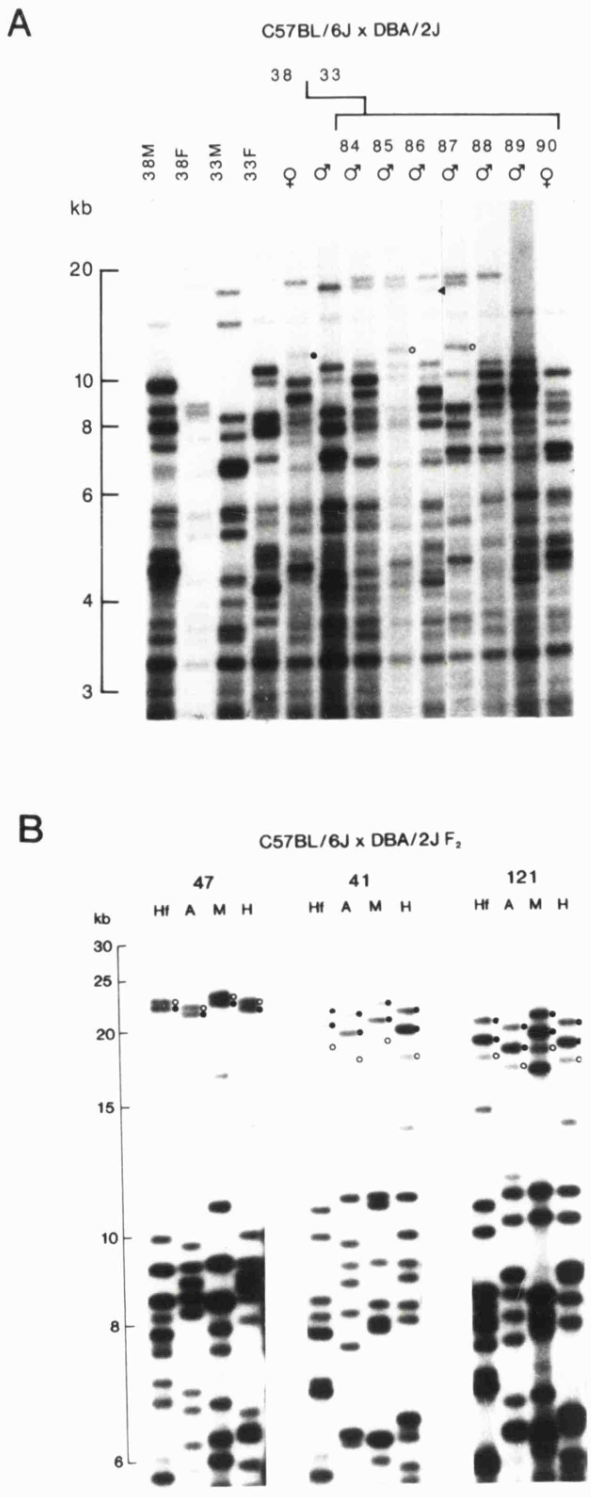

DBA/2J
DBA/2J
kb
20
10
8
6
4
3
0.1 × SSC
1 × SSC

Figure 7.5

New-length mutation events within the Mm3-1 DNA fingerprint

A. *Hin*fI DNA fingerprints of a C57BL/6J x DBA/2J mouse pedigree detected by probe Mm3-1 at low stringency (1x SSC, 65°). A new mutant fragment (•) seen in F_1 female 38 is absent from her maternal (38M) or paternal (38F) DNAs. This fragment is transmitted to two of her progeny (o); as this band is not seen at high stringency it cannot be a somatic mutant allele at *Ms6-hm*. A second mutational event is seen in F_2 mouse 86 (◁). 38M, 33F, DBA/2J; 38F, 33M, C57BL/6J; 38, 33, C57BL/6J x DBA/2J F_1 mice; 84-90, F_2 mice.

B. Signature analysis of three C57BL/6J x DBA/2J F_2 mice carrying non-parental alleles at *Hm-2* using *Hin*fI (Hf), *Alu*I (A), *Mbo*I (M), and *Hae*III (H). Parental (•) and non-parental (o) alleles at *Hm-2* are marked. Mouse 47 has a germline mutant allele at *Hm-2* while mice 41 and 121 have three alleles at this locus and are therefore somatic mosaics at *Hm-2*. Note that in both mosaics the progenitor parental allele, which hybridises to Mm3-1 with reduced intensity, is likely to be that furthest in length from the mutant allele. Gel and hybridisation by M.Gibbs, Leicester.


A
C57BL/6J x DBA/2J
38
33
84 85 86 87 88 89 90
38M
38F
33M
33F
♀
♂ ♂ ♂ ♂ ♂ ♂ ♀
kb
20
10
8
6
4
3
B
C57BL/6J x DBA/2J F2
47
41
121
Hf A M H
Hf A M H
Hf A M H
kb
30
25
20
15
10
8
6

alleles in C57BL/6J mice which are large enough to be easily scored in the F_1 and F_2 pedigrees (Fig.7.5). A four enzyme locus-specific signature test was used to distinguish alleles at this locus (*Hin*fI, *Alu*I, *Sau*3AI, and *Hae*III, Fig.7.5). Mice carrying three alleles at *Hm-2* were also identified (Fig.7.5); somatic mosaicism at *Hm-2* was confirmed by the signature test, and, for certain F_1 mosaics, by segregation analysis.

Hm-2 has been cloned into a Charomid vector from the DNA of a C57BL/6J mouse (M.Gibbs, Leicester, unpublished data). This clone (CMm9), like those derived from *Ms6-hm*, underwent a large deletion on cloning, such that a 20kb genomic allele resulted in a 4kb insert. CMm9 specifically detects *Hm-2* in mouse genomic DNA at high stringency (0.1x SSC, 65°), in addition to more stable fragments of similar hybridisation intensity (one of 3.2kb common to C57BL/6J and DBA/2J mice, and one DBA/2J-specific fragment of 4.0kb, Fig.7.6). *Hm-2* DBA/2J *Hin*fI alleles lie in a similar size range to *Ms6-hm* DBA/2J alleles (3-6kb, Fig.7.7). CMm9 also detects smaller, monomorphic, fragments similar to those detected by Mm3-1 at high stringency. At low stringency (1x SSC, 65°) CMm9 weakly cross-hybridises to a set of loci indistinguishable from those detected by Mm3-1 under the same conditions.

DNA sequence analysis has revealed that *Hm-2* contains the 4bp repeat unit GGCA, contained within both the human core sequence and the repeat unit sequence of *Ms6-hm*. The similarity of this sequence to the GGGCA repeat unit of *Ms6-hm* accounts for the low stringency cross-hybridisation between these loci. The repeat sequence of *Hm-2* is identical to that of the tandem array at the Eβ MHC meiotic recombination hotspot (see Introduction).

CMm9 detects a continuous size distriubution of *Hm-2* alleles across the BXD RI strains such that no allele size-based SDP can be assigned to *Hm-2*. The stable DBA/2J-specific cross-hybridising DNA fragment of 4.0kb can be assigned an SDP across the BXD RI series, but this does not show tight linkage to any other loci typed across the panel (B.Taylor, The Jackson Laboratory, personal communication). Thus at present the chromosomal localisation of *Hm-2* is unknown; however, from the segregation of alleles at this locus in C57BL/6J x DBA/2J pedigrees, *Hm-2* is known to be recombinationally separable from *Ms6-hm*.

Screening C57BL/6J x DBA/2J pedigrees with CMm9 revealed 14 germline mutations among 206 F_1 and F_2 mice (M.Gibbs, Leicester, personal communication); thus the germline mutation rate at this locus is 0.034 per gamete, similar to that of *Ms6-hm*. 18 of the mice scored appeared to be mosaic for non-parental alleles at *Hm-2*, suggesting that as many as 8.8% of mice show detectable levels of somatic mosaicism at *Hm-2*. The majority of the mutation events at *Hm-2* have involved large (C57BL/6J) alleles; since distinguishing small length changes at these alleles is difficult, improved

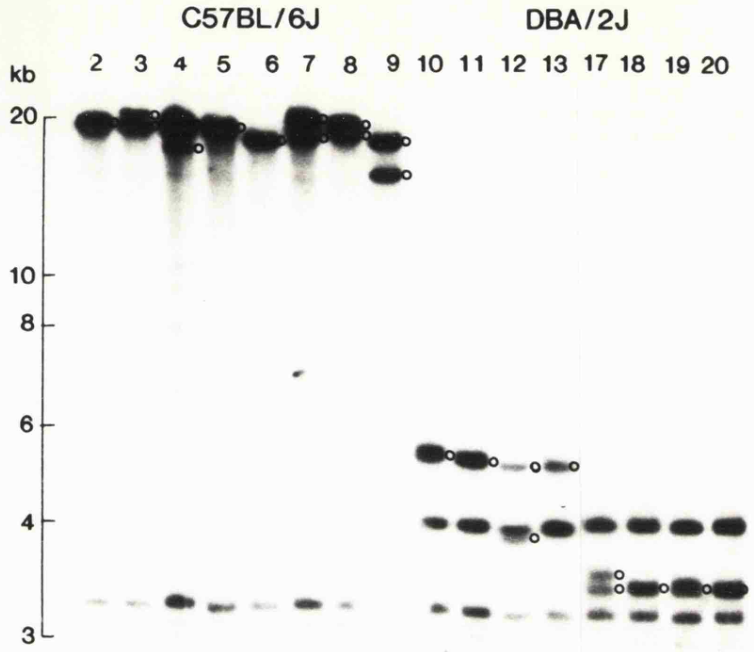
Figure 7.6

Identification of *Hm-2* alleles in C57BL/6J and DBA/2J DNA

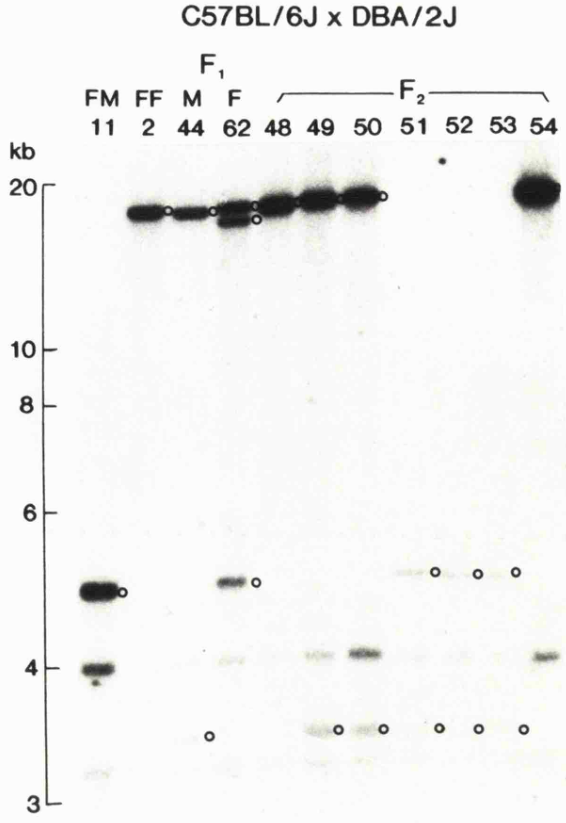
A. Tail DNA from 8 C57BL/6J and 8 DBA/2J mice digested with *Hin*fI and hybridised with CMm9 at high stringency (0.1x SSC, 65°). Alleles at *Hm-2* are marked (o). CMm9 detects additional stable fragments in mouse DNA; a 3.2kb fragment common to C57BL/6J and DBA/2J mice, and a 4.0kb DBA/2J-specific fragment. Mice 3 and 4 are likely to be somatic mosaics at *Hm-2*. Hybridisation by M.Gibbs, Leicester.

B. Tail DNA from a C57BL/6J x DBA/2J pedigree digested with *Hin*fI and hybridised with CMm9 at high stringency (0.1x SSC, 65°). Alleles at *Hm-2* are marked (o). Compare this blot with the low stringency hybridisation of probe Mm3-1 to the same mouse pedigree (A5, Fig.7.3). Hybridisation by M.Gibbs, Leicester.

A


C57BL/6J
DBA/2J
kb
2 3 4 5 6 7 8 9 10 11 12 13 17 18 19 20
20
10
8
6
4
3

B


C57BL/6J x DBA/2J
F1
F2
FM FF M F
11 2 44 62 48 49 50 51 52 53 54
kb
20
10
8
6
4
3

resolution may reveal even higher levels of germline mutation and mosaicism. No mice with two non-parental alleles at *Hm-2* have yet been observed.

7.6 A double mosaic mouse

As with *Ms6-hm*, mosaicism at *Hm-2* is likely to be due to somatic mutation events confined to early mouse development; firstly, there is no evidence that there is an equal probability of a somatic mutation event occurring at *Hm-2* at every mitotic cell division throughout development (which would result in large numbers of less intensely hybridising non-parental bands). Secondly, mice are found which are mosaic for the same non-parental *Hm-2* allele in both adult tail DNA and the germline, suggesting that the mutation event precedes the separation of the somatic and germ cell lineages (see Chapter 5).

It is expected, at a low frequency (2.8% x 8.8% of mice, or approximately 1 in every 400 mice), to find mice which are mosaic for a non-parental allele at both *Ms6-hm* and *Hm-2*. By examining the segregation of all 6 alleles in the offspring of such a double mosaic mouse the relative timing of the two somatic mutation events can be studied. C57BL/6J x DBA/2J F_1 52 (female) was such a double mosaic mouse (in adult tail DNA); at Ms6-hm a non-parental allele of 650 repeat units was likely to be derived by a 40 repeat unit deletion from the paternally inherited DBA/2J allele (tested by signature, see Chapter 6). At Hm-2 a 25kb non-parental allele was likely to be derived from a 20kb maternally inherited C57BL/6J allele (the closest parental allele, Fig.7.7).

The segregation of parental and non-parental alleles at *Ms6-hm* and *Hm-2* was analysed in the 16 offspring of mouse 52 (Fig.7.7). As non-parental alleles from both loci segregated into the F_2 mice, the germline of mouse 52 was also mosaic for the same non-parental alleles at *Ms6-hm* and *Hm-2* which were detected in adult tail DNA. Among these 16 F_2 mice, 2 inherited the non-parental allele at both loci. Thus both somatic mutation events must have occurred in the same mitotic lineage to give rise to a gamete carrying both non-parental alleles. No progeny were scored in either of the two classes which would permit the order of the two events to be elucidated; i.e., offspring inheriting the mutant allele at one locus and the progenitor parental allele at the other (see Fig.7.7). However, an approximately equal ratio of offspring were found which had inherited either two non-parental alleles, the non-parental allele at *Ms6-hm* and the stable parental allele at the *Hm-2*, or the stable parental allele at *Ms6-hm* and the non-parental allele at *Hm-2*. 2, 3 and 4 offspring (out of 16) were found in these three classes, respectively (see Fig.7.7). Assuming that the germline founder cells are represented equally among the progeny of this mouse, either of the last two classes would be expected to be 2^n times

Figure 7.7

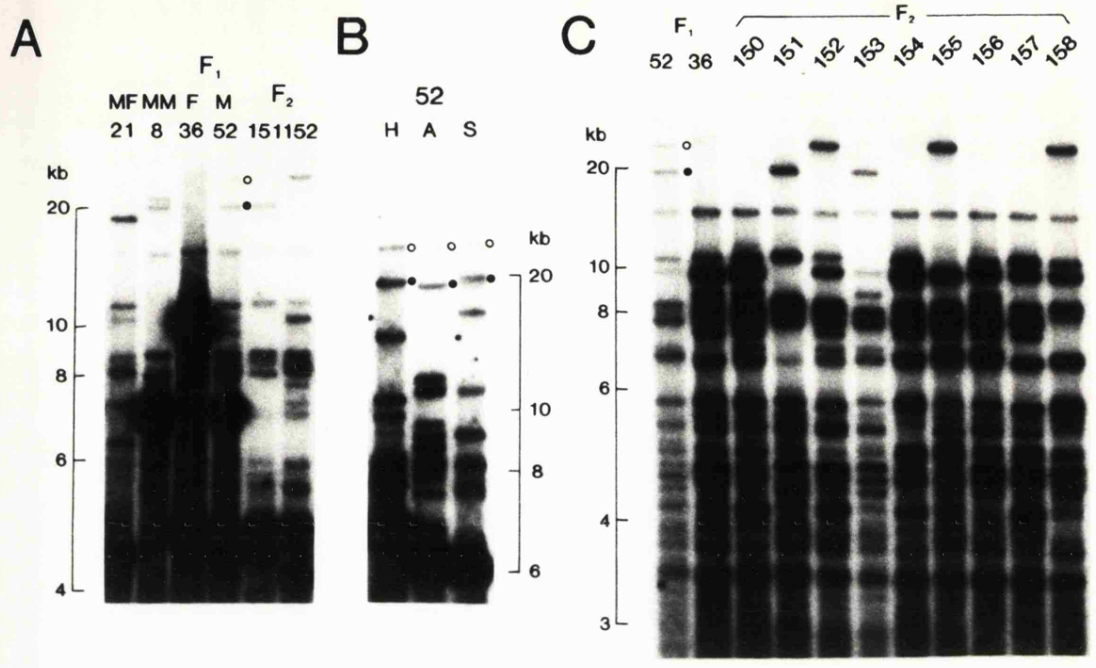
C57BL/6J x DBA/2J F_2 52 : a double mosaic mouse

A. Low stringency hybridisation of Mm3-1 to mouse tail DNA digested with *Hin*fI. F_1 mouse 52 has inherited a maternally derived allele at *Hm-2* (•), and carries a non-parental fragment derived from this locus (marked o). Somatic and germline mosaicism at *Ms6-hm* in mouse 52 are illustrated in Fig.6.1. 21, father of 52, DBA/2J; 8, mother of 52, C57BL/6J; 36, 52 F_1 mice, 151,152, F_2 mice.

B. Signature analysis of mouse 52 using *Hin*fI (H), *Alu*I (A), and *Sau*3AI (S), confirming that the new-length mutant fragment is derived from *Hm-2*. Alleles at *Hm-2* are marked; (•), parental, (o), non-parental.

C. Low stringency hybridisation of Mm3-1 to tail DNA from F_1 mice 52 and 36 and nine F_2 progeny, illustrating the transmission of parental (•, to 151 and 153) and non-parental (o, to 152, 155 and 158) alleles at *Hm-2* into the progeny of mouse 52.

D. Diagram and table to illustrate the segregation of alleles at *Ms6-hm* and *Hm-2* into the progeny of mouse 52, showing the four possible germline precursor cells (depending on the order of the two somatic mutation events), and all possible classes of gametes resulting from these cells, with the number of progeny (out of 16) scored for each class (see text). As no offspring are scored carrying alleles (M+) or (Dm), the order of the two mutation events cannot be determined. Alleles at *Ms6-hm* are B, non-progenitor C57BL/6J parental allele, D, progenitor DBA/2J parental allele (shown by signature), M, non-parental somatic mutant allele. Alleles at *Hm-2* are +, progenitor C57BL/6J parental allele (closest parental allele, •, in A, B, and C above), -, non-progenitor DBA/2J parental allele (not scored, but detected on high stringency hybridisation with CMm9, data not shown), m, non-parental somatic mutant allele. (o, in A, B, and C above).



D

SEGREGATION OF ALLELES AT *Ms6-hm* AND *Hm-2* IN THE PROGENY OF MOSAIC MOUSE C57BL/6J x DBA/2J 52

		Non-parental			
		first event			second event
Mouse 52	Parental	*Ms6-hm*		*Hm-2*	
Ms6-hm	BD	BM	or	BD	BM
Hm-2	+-	+-		m-	m-
Gametes	B B D D	B B M M		B B D D	B B M M
	+ - + -	+ - + -		m - m -	m - m -

Gametes of mouse 52

Ms6-hm	B	B	B	D	D	D	M	M	M
Hm-2	+	-	m	+	-	m	+	-	m
No. scored	2	3	4	2	0	0	0	3	2

more abundant than the other classes for every n cell divisions separating the two mutation events (precisely which class would be more abundant depending on which mutation event occurred first).

This result suggests that either the two mutation events occurred at the same cell division (the mutant alleles at both loci segregating into the same daughter cell), or the two mutation events occurred in closely following cell divisions within the same mitotic lineage (see Fig.7.7). The number of cells from which the germline is allocated must be at least two (as all three alleles in a mosaic mouse can contribute to the gametes, see Chapter 5 and Fig.7.7), however the precise number of germline progenitor cells, the time of germline allocation, and the relative degree to which germ cells contribute to the gametes after allocation are largely unknown. The actual stage of early development at which the somatic mutation events occurred in mouse 52 is therefore unclear; it remains formally possible, however, that both mutation events occurred as early as the first zygotic cell division.

7.7 Human DNA fingerprints obtained with Mm3-1

Core-related minisatellite sequences are found in the genomes of many species; human minisatellite probes 33.6 and 33.15 cross-hybridise to mouse DNA (Jeffreys *et al.*, 1987), and similarly Mm3-1 cross-hybridises to human DNA (Fig.7.8). The patterns of minisatellite fragments in *Hin*fI digests of human DNA detected by Mm3-1 are different to those detected by probe 33.6 (Fig.7.8). Furthermore, a synthetic GGGCA repeat probe detects a different set of minisatellite loci to either probe 33.6 or 33.15 in a human genomic DNA library of 4-8kb *Sau*3AI fragments (J.Armour, Leicester, personal communication).

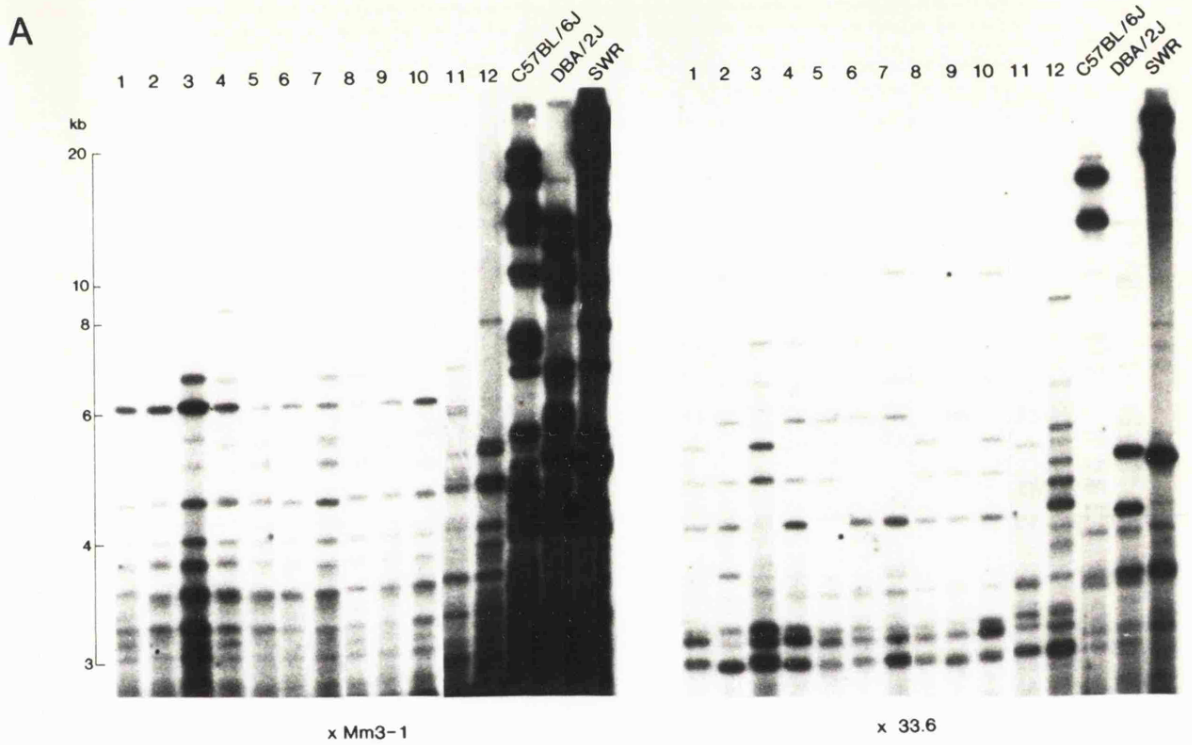
While the DNA fingerprints generated in mice by human probes 33.6 and 33.15 are as complex and variable as those of man, the human DNA fingerprints obtained with mouse probe Mm3-1 are less complex than those of mice, although the larger fragments are highly variable between unrelated individuals (Fig.7.8). The majority of *Hin*fI fragments cross-hybridising to Mm3-1 in human DNA are smaller than 6.5kb, and hybridise less intensely than fragments detected in an equivalent amount of mouse DNA (which range in size up to 20kb). Complex Mm3-1 DNA fingerprints are also observed in wild mice and rat (*Rattus norvegicus*) DNAs (Fig.7.8). The cross-hybridising loci detected by probe Mm3-1 in mouse DNA may therefore represent a relatively rodent-specific subset of minisatellite sequences.

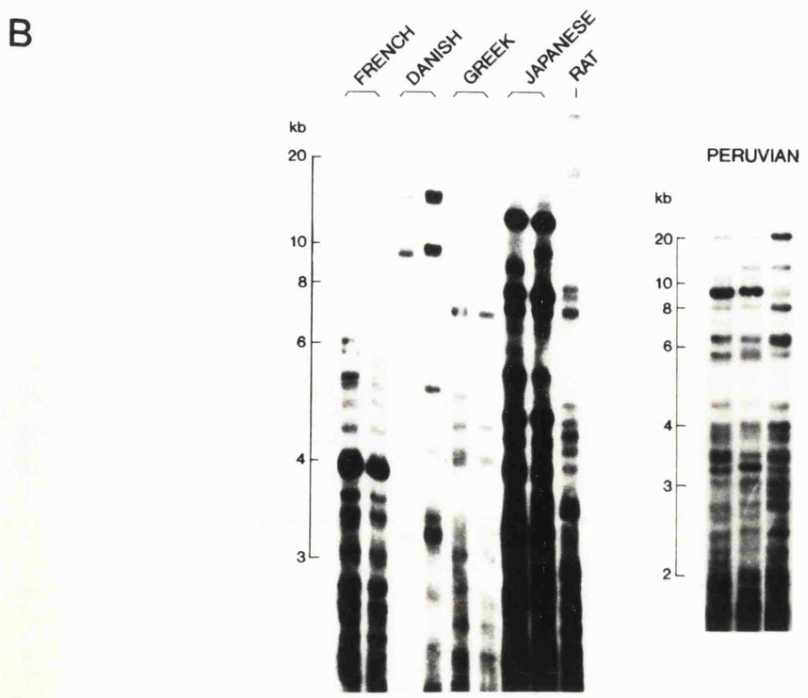
Figure 7.8

Comparison of human and mouse DNA fingerprints obtained using probes Mm3-1 and 33.6 at low stringency

A. 3μg of human and mouse DNA digested with *Hin*fI and Southern blot hybridised with either probe Mm3-1 or probe 33.6 at low stringency (1x SSC, 65°). Individuals 1-10 are sibs from CEPH family 14.13, and individuals 11 and 12 are unrelated. Note that Mm3-1 detects a different set of minisatellite loci in human DNA to probe 33.6. Compare the intensity of the mouse and human DNA fingerprints in each hybridisation.

B. DNA fingerprints of wild mice and a rat generated by probe Mm3-1. All digests are *Hin*fI. The pairs of wild mice were trapped from the same local populations; the three Peruvian mice are sibs. Note that the rat DNA fingerprint contains cross-hybridising fragments ranging in size up to at least 20kb.


A
1 2 3 4 5 6 7 8 9 10 11 12 C57BL/6J DBA/2J SWR
kb
20
10
8
6
4
3
x Mm3-1
1 2 3 4 5 6 7 8 9 10 11 12 C57BL/6J DBA/2J SWR
x 33.6


B
FRENCH
DANISH
GREEK
JAPANESE
RAT
kb
20
10
8
6
4
3
PERUVIAN
kb
20
10
8
6
4
3
2

7.8 Other loci detected by Mm3-1 at high stringency in inbred strains of mice

In individual DNAs from a minority of inbred mouse strains a larger number of minisatellite fragments hybridise strongly to probe Mm3-1 at high stringency (0.1x SSC, 65°) than the expected one, two, or three allelic fragments at *Ms6-hm*. In SWR and SJL mice very large DNA fragments of 10-20kb are observed (see Fig.3.4), whereas strain 129 mice have 2-4 DNA fragments in the 2-5kb range (see Fig.6.4). Using the *Hin*fI, *Alu*I, and *Sau*3AI locus-specific signature test it has been shown for these strains that DNA fragments derived from other minisatellite loci are present in addition to *Ms6-hm* alleles (see Figs.6.4 and 7.9). To explain the high stringency hybridisation signal these loci presumably contain perfect GGGCA repeats, and are thus distinct from those minisatellites which cross-hybridise with Mm3-1 at low stringency (such as *Hm-2*).

7.8.a A second locus in SWR mice

SWR mice are an albino strain of Swiss origin, and have have been inbred since 1929 (see Festing, 1979). Taylor (1982) showed that SWR mice, out of 27 inbred strains, had the largest number of rare alleles at 16 polymorphic loci; thus SWR mice are thought to be genetically distinct from other inbred mouse strains. *Ms6-hm* alleles in SWR mice have a DBA/2J-type signature (i.e., the polymorphic *Alu*I site flanking the minisatellite is present, see Chapter 3, section 4), and are found in a 3-6kb size range. The additional large minisatellite fragments observed in SWR DNA are derived from a single 'large locus' (*SWR-LL*), which has a different locus-specific signature to *Ms6-hm* and is extremely variable within a 10-20kb size range among SWR mice (Fig.7.9). Furthermore, mice have been found with more than two alleles at this locus, implying that *SWR-LL* is also somatically unstable (Fig.7.9). It is possible that *SWR-LL* is the result of a sequence duplication, including *Ms6-hm*, on chromosome 4 in SWR mice, in which case *SWR-LL* would be expected to be closely linked to *Ms6-hm*. Alternatively *SWR-LL* may have arisen through amplification of a GGGCA array at a second site in the mouse genome. To distinguish between these two possibilities, experiments were designed to investigate whether *SWR-LL* is linked to *Ms6-hm*.

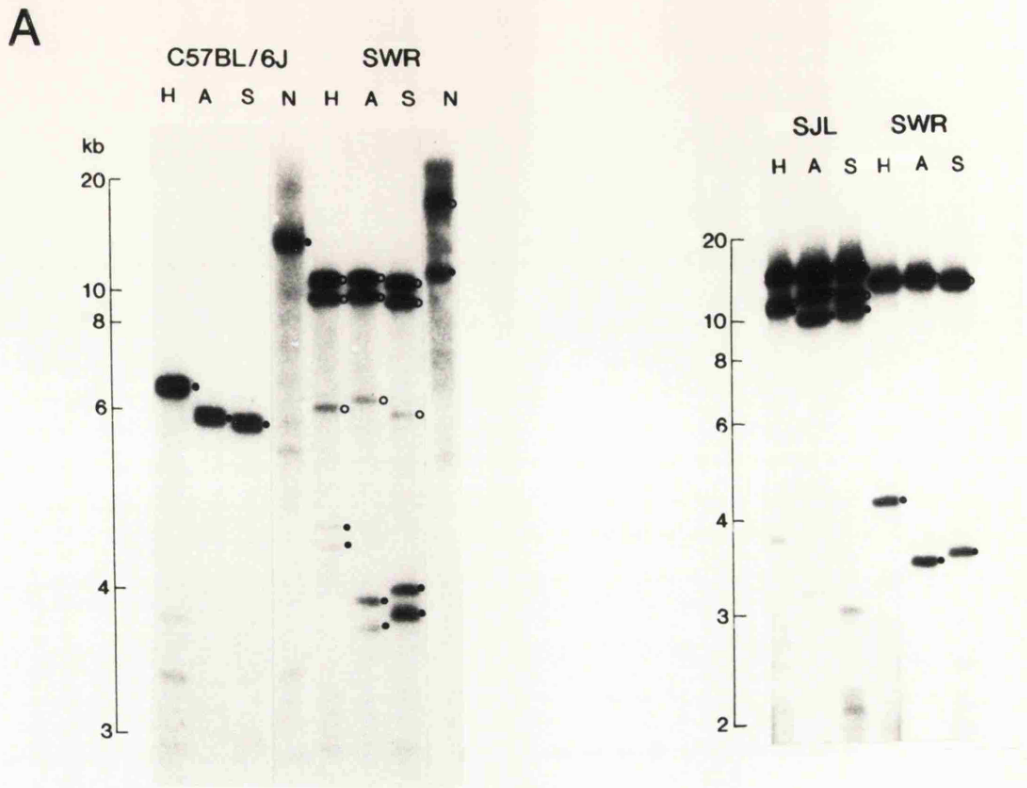
Fig.7.9 illustrates that the two loci in SWR DNA can be resolved on fragments at least 20kb in length using 6bp target site restriction endonucleases, and are thus unlikely to be the result of a local duplication involving only the minisatellite and a small amount of flanking DNA. Furthermore, at least two point mutations would be required to interconvert the locus-specific signatures of *Ms6-hm* and *SWR-LL*. However, *SWR-LL* might still be tightly linked to *Ms6-hm*, and two approaches were used to test for linkage between these loci.

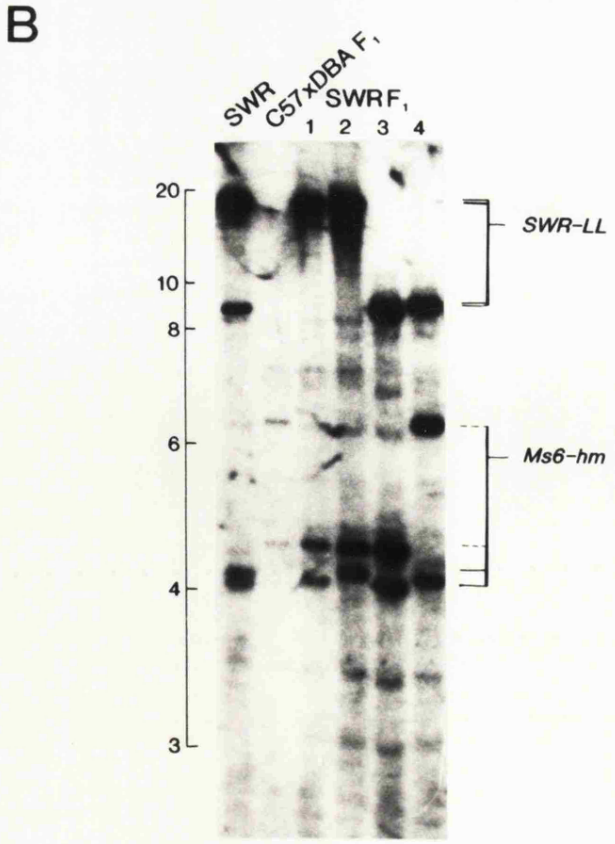
Figure 7.9

Other loci detected by Mm3-1 in mouse DNA at high stringency

A. Signature analysis of C57BL/6J, SWR and SJL DNA digested with *Hin*fI (H), *Alu*I (A), *Sau*3AI (S), and *Nci*I (N), and hybridised with Mm3-1 at high stringency (0.1x SSC, 65°). Alleles at *Ms6-hm* are marked (•); SWR and SJL mice have a DBA/2J-type signature at this locus. Alleles at *SWR-LL* are marked (o), and have a distinct signature to *Ms6-hm* alleles; note that the first SWR mouse has three alleles at *SWR-LL*, and is therefore likely to be mosaic for a non-parental allele at this locus. Note also that *Ms6-hm* and *SWR-LL* are resolved on separate fragments even with the six bp target site endonuclease *Nci*I. Alleles at both *Ms6-hm* and *SWR-LL* lie in a 10-20kb size range in SJL DNA.

B. Tail DNA from the parents and four progeny of an SWR x (C57BL/6J x DBA/2J F_1) cross digested with *Hin*fI and hybridised with Mm3-1 at high stringency (0.1x SSC, 65°). Alleles at *Ms6-hm* and *SWR-LL* are indicated; —, SWR derived *Ms6-hm* alleles; = , SWR derived *SWR-LL* alleles; ---, C57BL/6J x DBA/2J derived *Ms6-hm* alleles. Alleles at each locus are joined by vertical lines. The SWR father is heterozygous at both *Ms6-hm* and *SWR-LL*; in the 4 offspring each *SWR-LL* allele is transmitted with either *Ms6-hm* allele, thus demonstrating that the two loci are recombinationally separable.


A
C57BL/6J
SWR
H A S N H A S N
kb
20
10
8
6
4
3
SJL
SWR
H A S H A S
20
10
8
6
4
3
2


B
SWR
C57xDBA F1
SWR F1
1 2 3 4
20
10
8
6
4
3
SWR-LL
Ms6-hm

The first approach involved a direct breeding experiment. A male SWR mouse, which was heterozygous at both *SWR-LL* and *Ms6-hm*, was mated with a C57BL/6J x DBA/2J F_1 female carrying *Ms6-hm* alleles of a size distinct from all 4 SWR fragments (Fig.7.9). The segregation of the SWR fragments was analysed in 4 offspring. All four allele combinations were observed in the progeny, demonstrating that the two loci are recombinationally separable ($r>0.1$, $p>0.95$).

SWR-LL and *Ms6-hm* were also demonstrated to be unlinked using the SWXL recombinant inbred series. There are 7 SWXL RI strains which are derived from SWR/J and C57L/J progenitor strains, and have been inbred for 56-84 generations (Taylor, 1989). C57L/J inbred mice (*a*, *b*, and *ln*) have a grey coat colour (due to the *leaden* mutation), and are derived from a brown subline of the original C57 cross which also produced the C57 black subline (see Festing, 1979). The 7 RI and progenitor strain DNAs were digested with *Hin*fI, *Alu*I, and *Sau*3AI (to distinguish C57L/J and SWR *Ms6-hm* alleles), and hybridised with Mm3-1 (Fig.7.10).

As observed in the BXD RI analysis (see Chapter 4, section 2), the size distribution of *Ms6-hm* alleles in SWXL RI strains is consistent with the signature of each strain. Thus three strains with *Ms6-hm* alleles larger than 1000 repeat units have the C57L/J-type signature, and three strains with smaller alleles have the SWR-type signature. SWXL-12 is an exception to this, and in this strain an *Ms6-hm* allele of approximately 1200 repeat units has an SWR type signature. Both *SWR-LL* and *Ms6-hm* show extreme instability across the SWXL RI strains; no two strains share the same *Ms6-hm* allele (three are heterozygous), and the single strain inheriting an SWR-derived allele is heterozygous at *SWR-LL*. At low stringency Mm3-1 cross-hybridises to other minisatellite fragments showing instability across the SWXL RI strains, and for which no SDPs can be obtained (Fig.7.10).

Ms6-hm and *SWR-LL* have distinct SDPs (based on signature), with 5 (out of 7) discordant RI strains, although only the SWR allele at *SWR-LL* could be scored. This result confirms that these two loci are recombinationally separable. The SDP of *Ms6-hm* differs from that of the brown locus (*b*) by a single strain (SWXL-4), and the SDP of *SWR-LL* is identical to that of the T-cell receptor β chain locus on chromosome 6, suggesting possible linkage to that chromosome (B.Taylor, The Jackson Laboratory, personal communication).

A fragment in C57L/J DNA (*Hin*fI alleles 1.5-3kb) shares both the signature and segregation pattern of *SWR-LL* (Fig.7.10). This fragment was also highly unstable across the RI panel, and is not seen in C57BL/6J DNA; this minisatellite may be derived from *SWR-LL*. It therefore appears that the sequence GGGCA has expanded independently at two different points in the mouse genome. Alleles at *SWR-LL* in C57L/J mice may

Figure 7.10

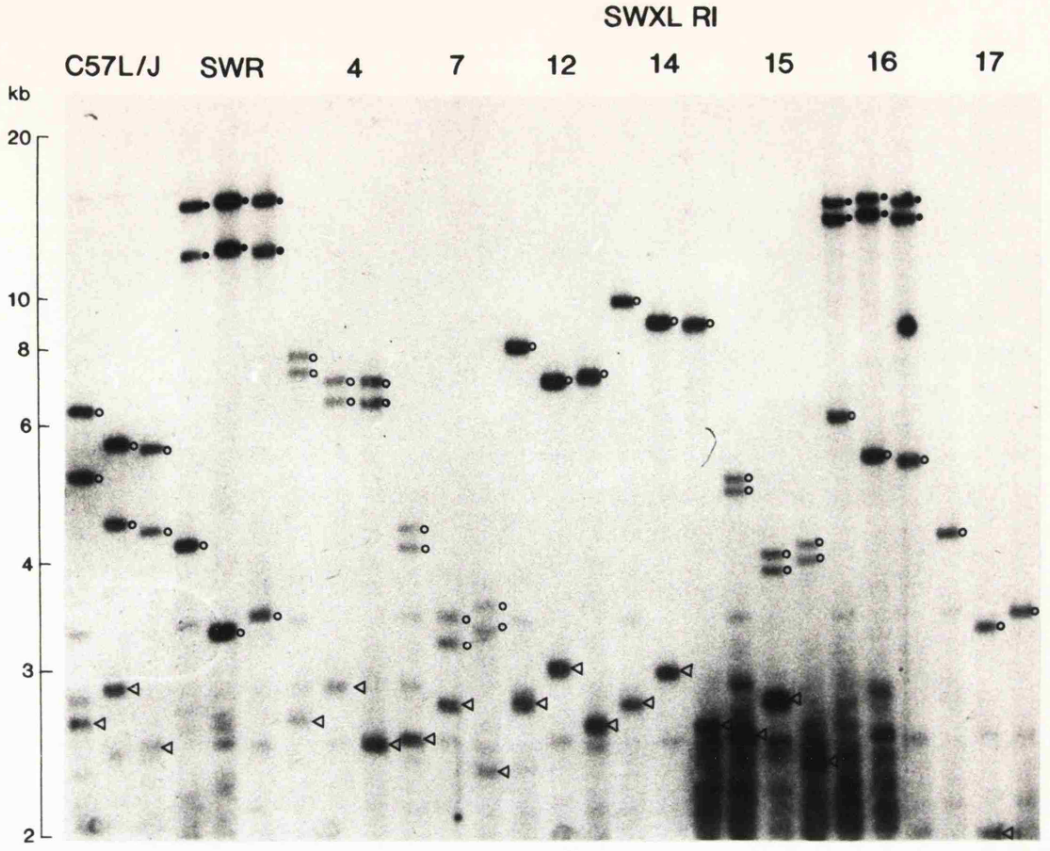
Segregation analysis of DNA fragments cross-hybridising to Mm3-1 in SWXL RI strains

A. SWXL RI and progenitor strain (C57L/J and SWR) DNAs digested with *Hin*fI, *Alu*I, and *Sau*3AI, and hybridised with Mm3-1 at high stringency (0.1x SSC, 65°). Alleles at *Ms6-hm* (o) and *SWR-LL* (•) are marked. The strain distribution patterns of these two loci are distinct. Note that the larger *Ms6-hm* alleles tend to have the C57-type signature (with the exception of SWXL 12). DNA fragments derived from a second locus in C57L/J DNA are marked (◁); the signature and SDP of this locus match those of *SWR-LL*. *Hin*fI and *Sau*3AI alleles at this locus in SWXL 17 are smaller than 2kb.

B. Strain distribution patterns of *Ms6-hm* (o above), *SWR-LL* (• above), and the lower C57L/J locus (◁ above) in SWXL RI strains.

C. SWXL RI and progenitor strain (C57L/J and SWR) DNAs digested with *Hin*fI and hybridised with Mm3-1 at low stringency (1x SSC, 65°). Note that most of the DNA fragments larger than 4kb are transmitted to only one or a few RI strains, and that many new-length fragments are present in only one RI strain, such that no strain distribution patterns can be determined for these fragments.

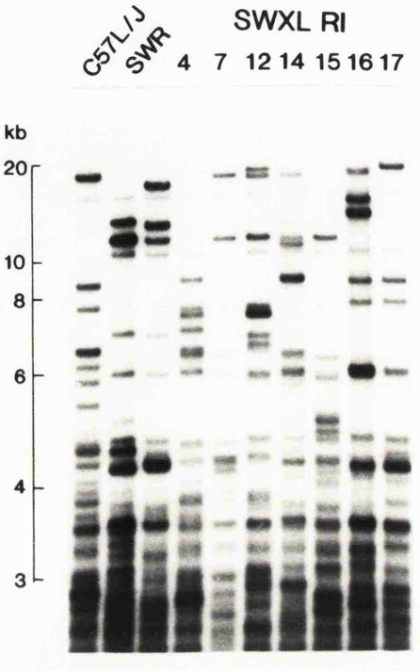
A



B

		SWXL						
		4	7	12	14	15	16	17
Ms6-hm ○	LS	L	S	S	L	S	L	S
SWR-LL ●	S	-	-	-	-	-	S	-
C57L/J ◁	L	L	L	L	L	L	-	L

C



The first approach involved a direct breeding experiment. A male SWR mouse, which was heterozygous at both *SWR-LL* and *Ms6-hm*, was mated with a C57BL/6J x DBA/2J F_1 female carrying *Ms6-hm* alleles of a size distinct from all 4 SWR fragments (Fig.7.9). The segregation of the SWR fragments was analysed in 4 offspring. All four allele combinations were observed in the progeny, demonstrating that the two loci are recombinationally separable ($r>0.1$, $p>0.95$).

SWR-LL and *Ms6-hm* were also demonstrated to be unlinked using the SWXL recombinant inbred series. There are 7 SWXL RI strains which are derived from SWR/J and C57L/J progenitor strains, and have been inbred for 56-84 generations (Taylor, 1989). C57L/J inbred mice (*a*, *b*, and *ln*) have a grey coat colour (due to the *leaden* mutation), and are derived from a brown subline of the original C57 cross which also produced the C57 black subline (see Festing, 1979). The 7 RI and progenitor strain DNAs were digested with *Hin*fI, *Alu*I, and *Sau*3AI (to distinguish C57L/J and SWR *Ms6-hm* alleles), and hybridised with Mm3-1 (Fig.7.10).

As observed in the BXD RI analysis (see Chapter 4, section 2), the size distribution of *Ms6-hm* alleles in SWXL RI strains is consistent with the signature of each strain. Thus three strains with *Ms6-hm* alleles larger than 1000 repeat units have the C57L/J-type signature, and three strains with smaller alleles have the SWR-type signature. SWXL-12 is an exception to this, and in this strain an *Ms6-hm* allele of approximately 1200 repeat units has an SWR type signature. Both *SWR-LL* and *Ms6-hm* show extreme instability across the SWXL RI strains; no two strains share the same *Ms6-hm* allele (three are heterozygous), and the single strain inheriting an SWR-derived allele is heterozygous at *SWR-LL*. At low stringency Mm3-1 cross-hybridises to other minisatellite fragments showing instability across the SWXL RI strains, and for which no SDPs can be obtained (Fig.7.10).

Ms6-hm and *SWR-LL* have distinct SDPs (based on signature), with 5 (out of 7) discordant RI strains, although only the SWR allele at *SWR-LL* could be scored. This result confirms that these two loci are recombinationally separable. The SDP of *Ms6-hm* differs from that of the brown locus (*b*) by a single strain (SWXL-4), and the SDP of *SWR-LL* is identical to that of the T-cell receptor β chain locus on chromosome 6, suggesting possible linkage to that chromosome (B.Taylor, The Jackson Laboratory, personal communication).

A fragment in C57L/J DNA (*Hin*fI alleles 1.5-3kb) shares both the signature and segregation pattern of *SWR-LL* (Fig.7.10). This fragment was also highly unstable across the RI panel, and is not seen in C57BL/6J DNA; this minisatellite may be derived from *SWR-LL*. It therefore appears that the sequence GGGCA has expanded independently at two different points in the mouse genome. Alleles at *SWR-LL* in C57L/J mice may

Figure 7.10

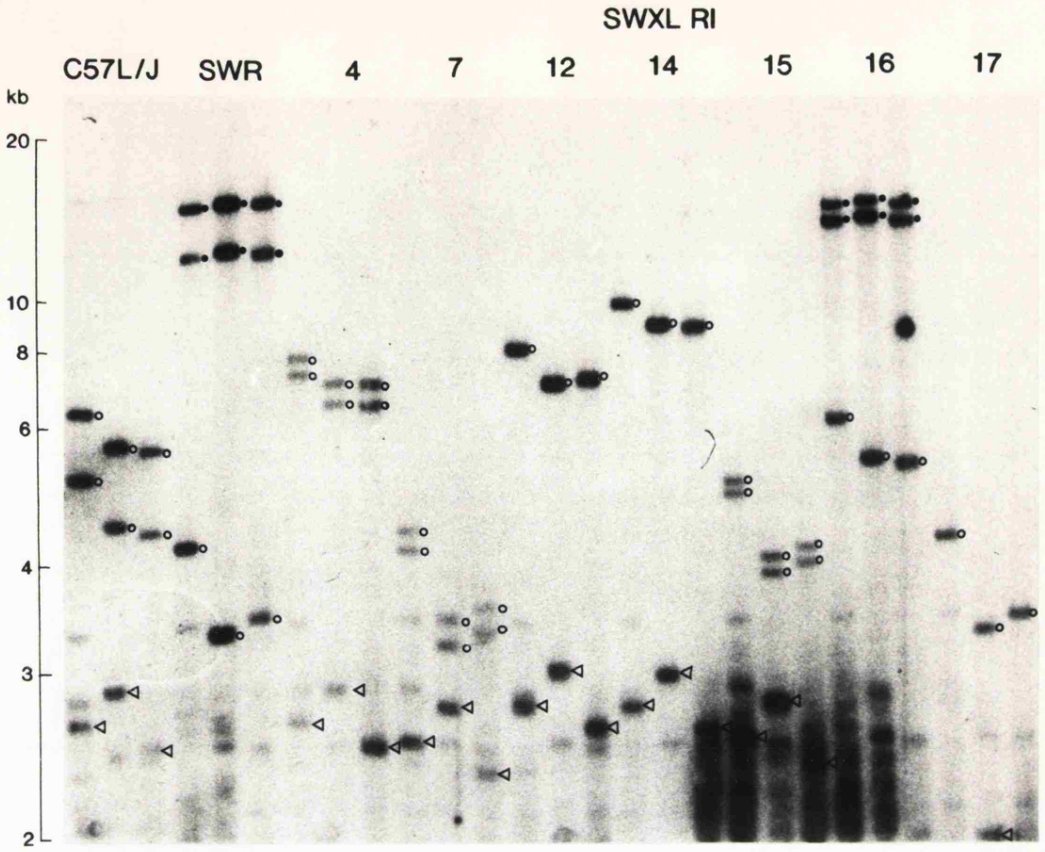
Segregation analysis of DNA fragments cross-hybridising to Mm3-1 in SWXL RI strains

A. SWXL RI and progenitor strain (C57L/J and SWR) DNAs digested with *Hin*fI, *Alu*I, and *Sau*3AI, and hybridised with Mm3-1 at high stringency (0.1x SSC, 65°). Alleles at *Ms6-hm* (o) and *SWR-LL* (•) are marked. The strain distribution patterns of these two loci are distinct. Note that the larger *Ms6-hm* alleles tend to have the C57-type signature (with the exception of SWXL 12). DNA fragments derived from a second locus in C57L/J DNA are marked (◁); the signature and SDP of this locus match those of *SWR-LL*. *Hin*fI and *Sau*3AI alleles at this locus in SWXL 17 are smaller than 2kb.

B. Strain distribution patterns of *Ms6-hm* (o above), *SWR-LL* (• above), and the lower C57L/J locus (◁ above) in SWXL RI strains.

C. SWXL RI and progenitor strain (C57L/J and SWR) DNAs digested with *Hin*fI and hybridised with Mm3-1 at low stringency (1x SSC, 65°). Note that most of the DNA fragments larger than 4kb are transmitted to only one or a few RI strains, and that many new-length fragments are present in only one RI strain, such that no strain distribution patterns can be determined for these fragments.

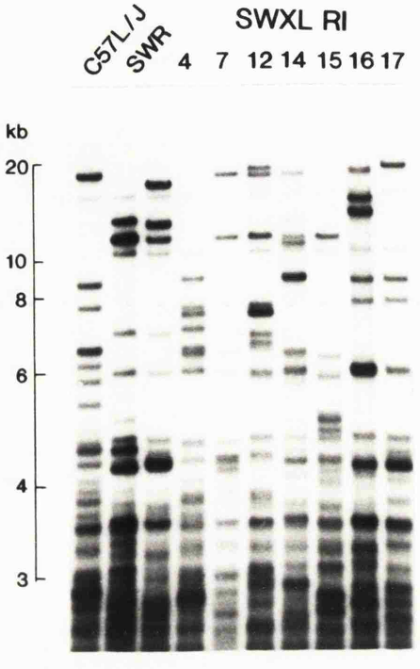
A



B

	SWXL							
	4	7	12	14	15	16	17	
Ms6-hm ○	LS	L	S	S	L	S	L	S
SWR-LL ●	S	-	-	-	-	-	S	-
C57L/J ◁	L	L	L	L	L	L	-	L

C



represent an intermediate stage between the small unresolved *SWR-LL* alleles of C57BL/6J mice (and other strains), and the very large *SWR-LL* alleles found in SWR mice.

7.8.b A second locus in SJL mice

SJL mice are also of Swiss origin, and have been inbred since 1955 (see Festing, 1979). 2-4 hypervariable *Hin*fI DNA fragments in the 10-20kb size range hybridise strongly with Mm3-1 in SJL DNA (Fig.7.9). The locus-specific *Hin*fI, *Alu*I, and *Sau*3AI signature test suggested that these fragments may be alleles of *Ms6-hm* and *SWR-LL*. In order to confirm whether these fragments are derived from *Ms6-hm* and *SWR-LL*, an SJL female with four large minisatellite fragments was crossed with an SWR male which was heterozygous at both *Ms6-hm* and *SWR-LL* (Fig.7.11). Each of four offspring had three fragments in the 8-20kb size range, and one or other of the SWR *Ms6-hm* alleles. Three of these large fragments were non-parental, presumably arising by mutation in the parental germline; of these, two were shared by two offspring, suggesting germline mosaicism (Fig.7.11).

An SWR x SJL F_1 female (A11.4) carrying 4 parental fragments was mated with a C57BL/6J x DBA/2J F_2 mouse which had two *Ms6-hm* alleles distinct in size from all 4 cross-hybridising DNA fragments in A11.4 (Fig.7.11). The segregation of the SWR x SJL F_1 fragments was examined in the progeny of this cross. Each of 8 offspring had three strongly hybridising DNA fragments; in each mouse two fragments were inherited from A11.4, and one was a paternal *Ms6-hm* allele. From A11.4, each mouse inherited either the largest DNA fragment or the SWR derived *SWR-LL* allele, and either the third largest DNA fragment or the SWR derived *Ms6-hm* allele (Fig.7.11). This segregation data suggests that the large 10-20kb DNA fragments observed in SJL mice are alleles at *Ms6-hm* and *SWR-LL*.

In one mouse (SJL F_2.1), a non-parental allele at *SWR-LL* was observed; this mouse is also mosaic for a new-length fragment derived from the paternal *Ms6-hm* allele (Fig.7.11).

7.8.c A second locus in 129 mice

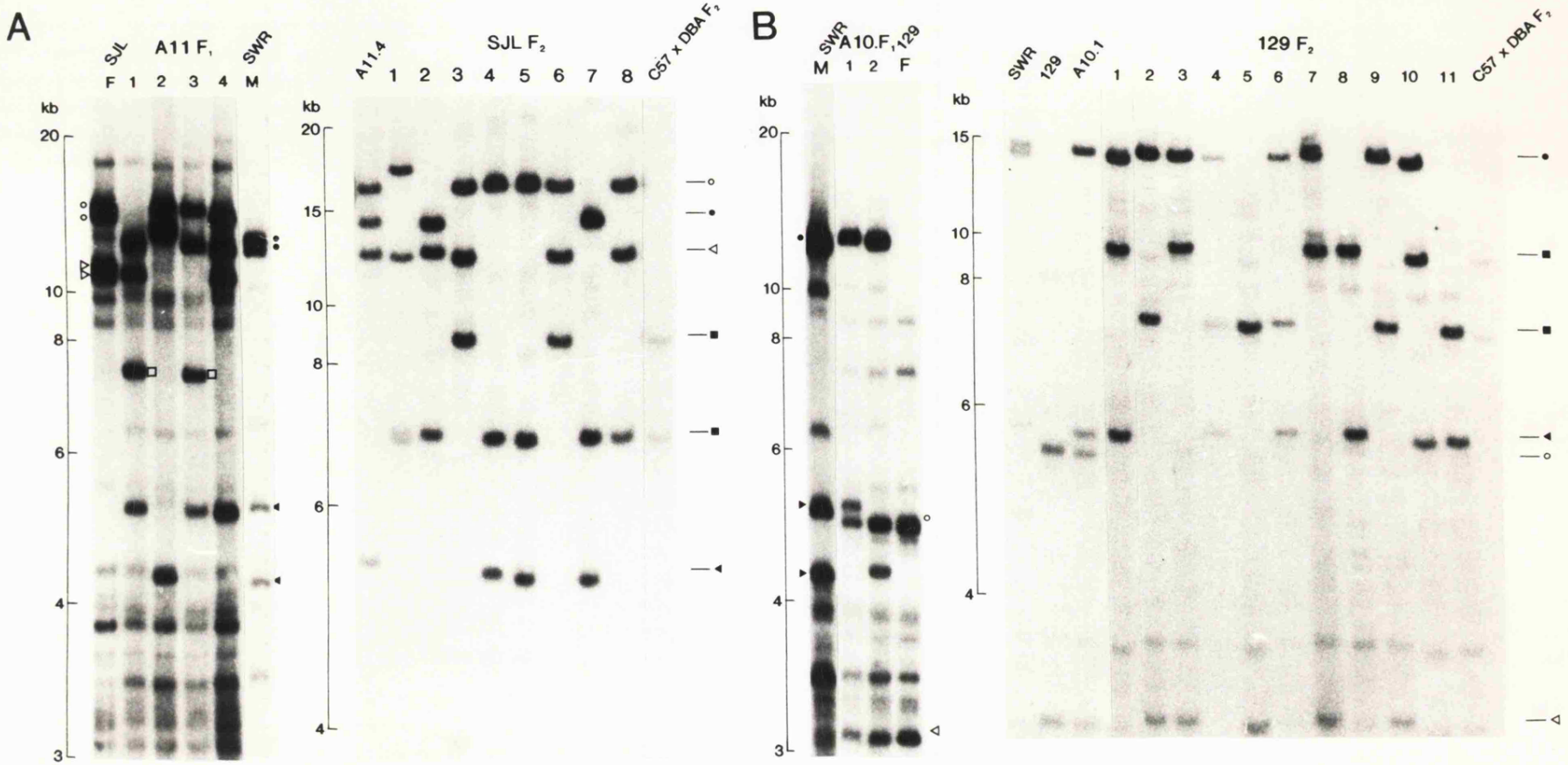
Fragments derived from two strongly hybridising hypervariable loci were detected by Mm3-1 in DNA from 129 mice while screening embryonic carcinoma cell lines for *in vitro* somatic mutation events (Chapter 6, section 5). The *Hin*fI, *Alu*I and *Sau*3AI signature of one of these loci matched that of *Ms6-hm* (DBA/2J-type). The signature of

Figure 7.11

Segregation analysis of loci cross-hybridising to Mm3-1 at high stringency in DNA from SWR, SJL and 129 inbred mice

A. Analysis of cross-hybridising loci in SJL DNA; tail DNA from SWR x SJL F_1 and F_2 mice digested with *Hin*fI and hybridised with Mm3-1 at high stringency (0.1x SSC, 65°). SWR derived alleles at *Ms6-hm* (◄) and *SWR-LL* (•), and SJL derived alleles at *Ms6-hm* (▷) and *SWR-LL* (○) are marked. Two non-parental alleles in the F_1 mice are marked (□). A11 F_1 mouse 4 (female) was crossed with a C57BL/6J x DBA/2J F_2 male, producing 8 F_2 mice. Paternally derived alleles at *Ms6-hm* are marked (■). The segregation of the SWR and SJL derived alleles at *Ms6-hm* and *SWR-LL* can be followed in the F_2 mice. SJL F_2 mouse 1 has a germline mutant allele at *SWR-LL* and a somatic mutant allele at *Ms6-hm* (which is presumably derived from the paternally inherited allele, closest parental allele).

B. Analysis of cross-hybridising loci in 129 DNA; tail DNA from 129 x SJL F_1 and F_2 mice digested with *Hin*fI and hybridised with Mm3-1 at high stringency (0.1x SSC, 65°). SWR derived alleles at *Ms6-hm* (◄) and *SWR-LL* (•) and 129 derived alleles at *Ms6-hm* (▷) and *129-LL* (○) are marked. A10 F_1 mouse 1 (female) was crossed with a C57BL/6J x DBA/2J F_2 male, producing 11 F_2 mice. Paternally derived alleles at *Ms6-hm* are marked (■). The segregation of the SWR derived alleles at *Ms6-hm* and *SWR-LL*, and the 129 derived alleles at *Ms6-hm* and *129-LL*, can be followed in the F_2 mice.


A
SJL
A11 F1
SWR
F 1 2 3 4 M
kb
20
10
8
6
4
3
A11.4
SJL F2
1 2 3 4 5 6 7 8
C57 x DBA F2
15
B
SWR
A10.F1,129
M 1 2 F
129 F2
129
A10.1
9 10 11

the second locus (termed *129-LL*) was distinct from that of either *Ms6-hm* or *SWR-LL*, and would require at least two point mutations to resemble the signature of either locus.

A mating was set up in order to investigate whether this locus was allelic to *SWR-LL*, or linked to either *Ms6-hm* or *SWR-LL*. A 129 female with two strongly hybridising fragments (i.e., homozygous at *Ms6-hm* and *129-LL*) was crossed with an SWR male which was heterozygous at *Ms6-hm* and *SWR-LL* (Fig.7.11). Both 129 fragments were transmitted to each of 2 offspring, which also each inherited an SWR derived allele at *Ms6-hm* and *SWR-LL*. One of these offspring, female A10.3, was mated with a C57BL/6J x DBA/2J F_2 mouse which had *Ms6-hm* alleles distinct in size from the 4 cross-hybridising DNA fragments of A10.3.

The segregation of the SWR x 129 F_1 DNA fragments was studied in the 11 progeny of this cross (Fig.7.13). Each mouse inherited a paternal *Ms6-hm* allele, and either a 129 or SWR derived maternal *Ms6-hm* allele, confirming the results of the signature test outlined above. The SWR derived *SWR-LL* allele was transmitted to 9 of the 11 offspring, and the fragment from the second 129 locus to none of 11. The explanation for the apparent segregation distortion of this *129-LL* allele in the F_2 mice is unclear. There are 2/11 discordancies between *SWR-LL* and *129-LL* suggesting that these two loci are not allelic, although they may be loosely linked in repulsion (r=0.2, 95% confidence limits 0.03-0.47). Alternatively it is possible that *SWR-LL* and *129-LL* are allelic, and that non-parental (and unresolved) alleles at *129-LL* were transmitted to the two mice not inheriting the SWR derived *SWR-LL* allele. *129-LL* and *Ms6-hm* segregate independently and are therefore not closely linked.

7.9 Summary

Under low stringency hybridisation conditions the GGGCA repeat of Mm3-1 cross-hybridises to many minisatellite loci in the mouse genome to generate a novel and highly individual specific DNA fingerprint. These minisatellite loci have been shown by pedigree analysis to be inherited in a Mendelian fashion, and to be dispersed in the mouse genome. Many of the larger loci detected by Mm3-1 are highly unstable : there is a high level of intrastrain variation within the DNA fingerprint, SDPs cannot be obtained for these loci using RI strains, and in some cases new-length mutation events have been directly observed. In addition, somatic mutant alleles arising in early mouse development have been observed at one of these loci, *Hm-2*.

Additional highly unstable loci are detected by Mm3-1 even under conditions of high stringency; such loci appear to be specific to particular inbred strains. These loci are unlinked to *Ms6-hm*, and may result from the expansion of GGGCA repeats at

different points in the mouse genome; such loci are also somatically unstable. The germline and somatic hypervariability of these cross-hybridising loci contrasts with the stability of loci detected by probes 33.6 and 33.15, suggesting that long minisatellite arrays related to the sequence GGGCA show preferential instability in the mouse genome. GGGCA-related loci are poorly represented in man, and may represent a relatively mouse-specific sequence-related subset of hypermutable minisatellites.

VIII. TANDEM REPETITIVE LOCI IN THE MOUSE GENOME

8.1 Introduction

Simple tandem repeat sequences are a ubiquitous component of mammalian genomes (see Chapter 1). During the course of this study several tandem repeated sequences in mouse DNA were characterised, in addition to *Ms6-hm* and the loci to which Mm3-1 cross-hybridises. The examples of tandem repeat loci described in this chapter illustrate the complexity of the simple sequence component of the mouse genome.

8.2 *Ms15-1*

Ms15-1 is one of the minisatellite loci detected in mouse genomic DNA digested with *Hin*fI by human probe 33.15 (Jeffreys *et al.*, 1987). 10 fragments detected in C57BL/6J DNA by probe 33.15 showed complete linkage across the BXD RI panel, being either all present or all absent in each BXD strain (Jeffreys *et al.*, 1987). These fragments are therefore likely to be all derived from a single minisatellite locus, *Ms15-1*, containing internal *Hin*fI cleavage sites. By summing the sizes of these fragments, Jeffreys *et al.* (1987) estimated that the C57BL/6J allele of *Ms15-1* must be at least 90kb in length. This locus is therefore comparable with the 250-500kb 'midisatellite' locus which has been described on human chromosome 1 (Nakamura *et al.*, 1987a).

The length and internal organisation of *Ms15-1* varies between inbred mouse strains. Two fragments in DBA/2J DNA cosegregate to those BXD strains which do not contain the C57BL/6J *Ms15-1* haplotype. These fragments define a minimum DBA/2J allele length of 19kb. The complementary SDPs of B and D alleles at *Ms15-1* allowed the assignment of this locus to mouse chromosome 4 (see Chapter 4). In C3H/HeJ DNA a large (approximately 30kb) fragment is detected by probe 33.15. Across the BXH RI panel, this large fragment segregated to those strains not containing the C57BL/6J *Ms15-1* haplotype (see Fig.7.1). Other inbred strains (SWR, BALB/c) contain large (>30kb) hybridising fragments detected by probe 33.15. In the SWXL RI strains the large SWR fragment segregates to those strains lacking the C57L/J haplotype of at least 8 cosegregating fragments (minimal allele length 70kb, Fig.8.1). In SWR DNA a 5.5kb fragment cosegregates with the large *Ms15-1* fragment.

C57BL/6J and AKR strains, which were originally inbred from separate mouse stocks, have similar DNA fingerprints with probe 33.15; this similarity is primarily due to

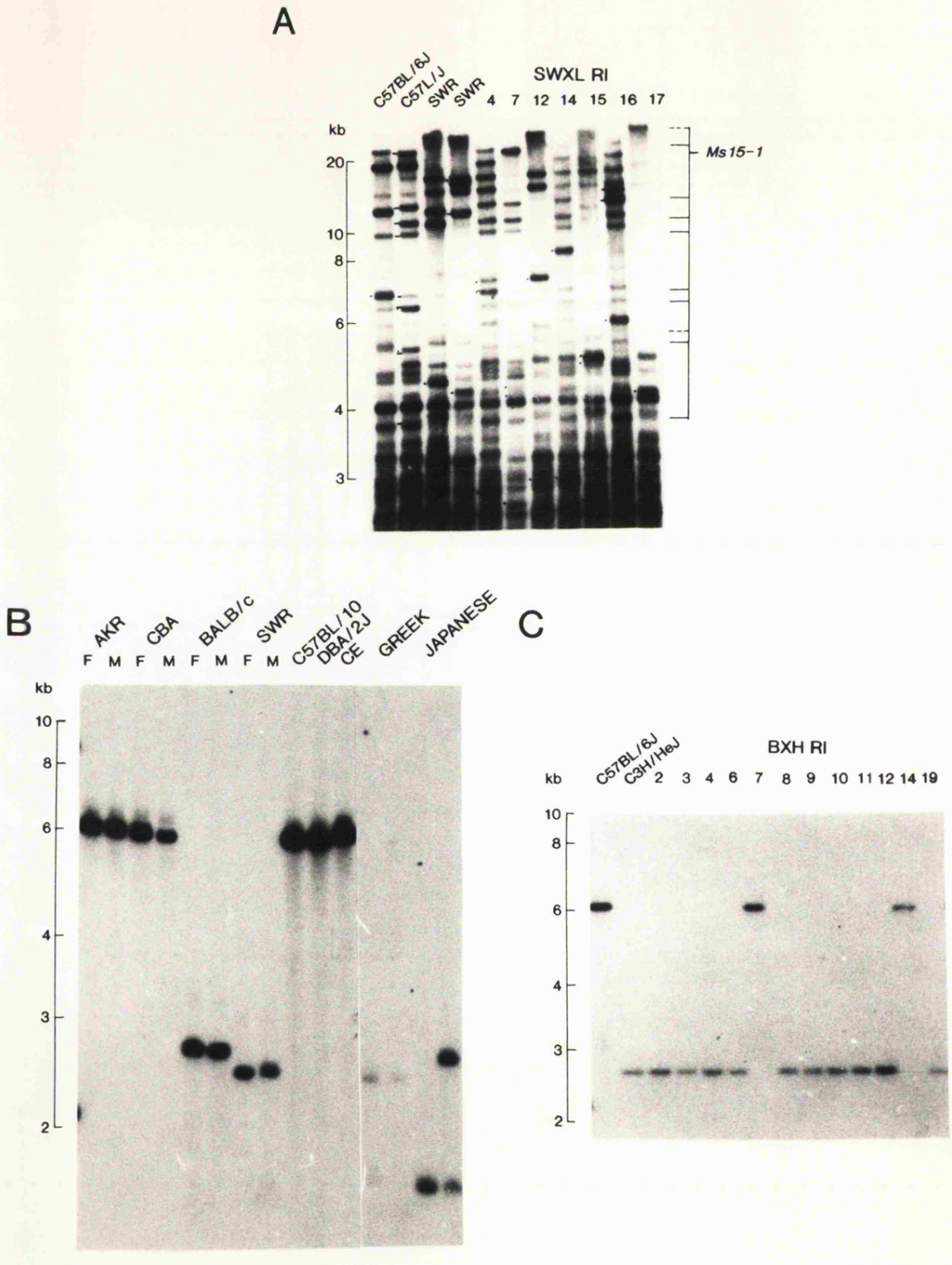
Figure 8.1

Ms15-1 and *Mm1*

A. SWXL and progenitor strain (C57L/J and SWR) and C57BL/6J DNAs digested with *Hin*fI and hybridised with probe 33.15 at low stringency (1x SSC, 65°). Fragments derived from *Ms15-a* are indicated; —, C57L/J derived fragments; ---, SWR derived fragments. The minimal allele length of the 8 cosegregating C57L/J fragments is 70kb.

B. DNA from inbred and wild mice digested with *Sau*3AI and hybridised with Mm1 at high stringency (0.1x SSC, 65°). Alleles at *Mm1* are stable within inbred strains, although variable among wild mice. F, female; M, male.

C. BXH RI and progenitor (C57BL/6J and C3H/He/J) strain liver DNAs digested with *Hin*fI and hybridised with Mm1, showing the segregation of *Mm1* alleles across the BXH RIs. Note that the progenitor strain alleles are stably inherited in each RI strain. The SDP of *Mm1* is identical to that of *Mod-2* on mouse chromosome 7 (B.Taylor, The Jackson Laboratory, personal communication).


A
C57BL/6J
C57L/J
SWR
SWR
SWXL RI
4 7 12 14 15 16 17
kb
20
10
8
6
4
3
Ms15-1
B
AKR
CBA
BALB/c
SWR
C57BL/10
DBA/2J
CE
GREEK
JAPANESE
F M F M F M F M
kb
10
8
6
4
3
2
C
C57BL/6J
C3H/HeJ
BXH RI
kb
2 3 4 6 7 8 9 10 11 12 14 19
10
8
6
4
3
2

the presence in AKR DNA of 8 of the 10 *Ms15-1*^{B} haplotype fragments. The conservation of this haplotype suggests that it may not be rare in mouse populations, and furthermore that *Ms15-1* is stable in the mouse genome. There is no evidence for the loss of any cosegregating *Ms15-1* fragments in any RI strains examined, further suggesting that *Ms15-1* shows substantial germline stability. In the absence of a locus-specific probe for *Ms15-1* the repeat unit sequence and detailed organisation of this presumed midisatellite locus remain unknown.

8.3 *Mm1*, *Mm6*, and *Mm16* : three mouse minisatellites

The clone Mm3-1, derived from *Ms6-hm*, was isolated by hybridisation to probe 33.6 from a C57BL/6J genomic library of size-selected *Sau*3AI DNA fragments cloned in λL47.1 (see Chapter 3, section 3). While screening this library, two clones derived from loci other than *Ms6-hm* were also isolated (*Mm1* and *Mm16*). A clone derived from a third locus (*Mm6*) was isolated by cross-hybridisation to probe 33.6 from a second library resulting from a control ligation of smaller size selected DNA fragments (5.5-6.0kb).

8.3.a Mm1

15 of the 28 plaques hybridising positively with 33.6 in the library from which λMm3 was isolated were derived from *Mm1*. Isolates of *Mm1* contained a stable 6.2kb *Sau*3AI insert which was trimmed to 5.6kb by *Alu*I (see Chapter 3, section 3). The *Sau*3AI insert from λMm1 was subcloned in pUC13; two independent subclones had similar 6.2kb plasmid inserts. When used as a hybridisation probe, the insert from one of these subclones, pMm1$_{I}$, detected a single locus in mouse genomic DNA at high stringency (0.1x SSC, 65°, Fig.8.1). The *Hin*fI, *Alu*I, and *Sau*3AI signature of this locus in C57BL/6J DNA was distinct from that of *Ms6-hm*.

Mm1 is variable among wild mice, with *Sau*3AI alleles in a 1.5-3kb size range. Every individual mouse examined within a particular inbred strain was homozygous for a strain-specific *Mm1* allele; there is therefore no evidence for variation at this locus within inbred strains. AKR, CBA, C57BL/6J, C57BL/10, DBA/2J and CE mice share a 6.2kb *Sau*3AI allele at *Mm1*. Therefore no SDP could be obtained for *Mm1* across the BXD RI strains on the basis of allele length. C3H/HeJ, BALB/c, A, and SWR mice have alleles between 2.5 and 3kb in length. An SDP for *Mm1* was obtained from the BXH RI strains (Fig.8.1), and revealed 0/12 discordancies with *Mod-2*, a gene encoding a mitochondrial malic enzyme, on mouse chromosome 7 (B. Taylor, The Jackson Laboratory, personal communication; the 95% confidence limits for the distance between *Mm1* and *Mod-2* are 0 and 11cM). The progenitor strain *Mm1* alleles were stably inherited in the BXH RI

Figure 8.2

DNA sequence of *Mm1*

The DNA sequence of three M13 subclones containing random fragments from within a 6.2kb *Sau*3AI fragment derived from *Mm1*. Clone 1 contains flanking and minisatellite DNA sequence, while clones 2 and 3 contain only repeat units from the tandem array. The 40bp minisatellite repeat unit consensus sequence is also shown; deviations from this sequence in the M13 minisatellite subclones are either underlined or indicated by gaps (-). The 11bp sequence within this consensus which precisely matches the core-like sequence of probe 33.6 is also underlined. The consensus repeat sequence is compared to that of probe 33.6 and the human minisatellite 'core' sequence.

Mm1

Clone 1

AGTCTGGTCCAGACGAGCAGCAGCCTTATCGCCAACACCCCCTGGAGTGAGGAAGGCAAACACAGGAC

----------GGGCTGGAGGACTGACCTAGTGC-

-TACAGTCTGTCCCATGGGCTGG

Clone 2

AGTGTT

ATACAGTCTGTCCCATGGGCTGGAGGACTGACCTAGTGTT

ATACAGTCTGTCCCATGGGCTGGAGGACTGACCTAGTGTT

ATACAGTCTGTCCCATGGGCTGGAGGACTGACCTAGTGC-

-TACAGTCTATCCCATGGGCTGGAGGACTGACCTAGTGTT

ATACAGACTGTCCCATGGGCTGGAGGACTGACCTAGTGTT

ATACAGTCTGTCCCATGGGCTGGAGGACTGACCTAGTGTT

ATACAG

Clone 3

CCTAGTGTT

ATACAGTCTGTCCCATGGGCTGGAGGACTGACCTAGTGTT

ATACAGTCTGTCCCATGGGCTGGAGGACTGACCTAGTGTT

ATACAGTCTGTCCCATGGGCTGGAGGACTGACCTAGTGTT

ATACAGTCTGTCCCATGGGCTGGAGGACTGACCTAGTGC-

-TACAGTCTGTCCCATGGGCTGGAGGACTGACCTAGTGC-

-TACAGTCTGTCCCATGGGCTGGAGGACTGACCTAGTGC-

-TACAGTCTGTCCCATGGGCTGGAGGACTGACCTAGTGTT

ATACAG

Consensus ATACAGTCTGTCCCATGGGCTGGAGGACTGACCTAGTGTT

33.6 AGGAAGGGCTGGAGGAGGGCTGGAGGAGGCTCCGG

core GGAGGTGGGCAGGARG

strains, suggesting that this is a relatively stable locus in the mouse genome; the observation that *Mm1* alleles are held in common by different inbred strains supports this conclusion.

The 6.2kb insert from pMm1$_1$ was sonicated, and subcloned into M13mp19 for DNA sequence analysis (Fig.8.2). Three clones were sequenced, two hybridising strongly with probe 33.6 (containing only tandem repeats), and one hybridising weakly with probe 33.6 (containing predominantly flanking DNA sequence). *Mm1* has a 40bp repeat unit, containing 11bp which precisely match the core-like sequence of 33.6 (of which each repeat unit is a trimer, see Fig.8.2). 13 complete and 6 partial repeat units were sequenced (Fig.8.2). 5 of these were variant repeat units; 4 contained a TTA replaced by a C, of which 3 were adjacent, the fourth also containing a single nucleotide change. The other variant repeat unit contained a single nucleotide change. The DNA sequence of these repeat units and 70bp of flanking DNA is illustrated in Fig.8.2. The flanking sequence shows no significant similarity to other DNA sequences in the EMBL DNA sequence database.

8.3.b Mm6

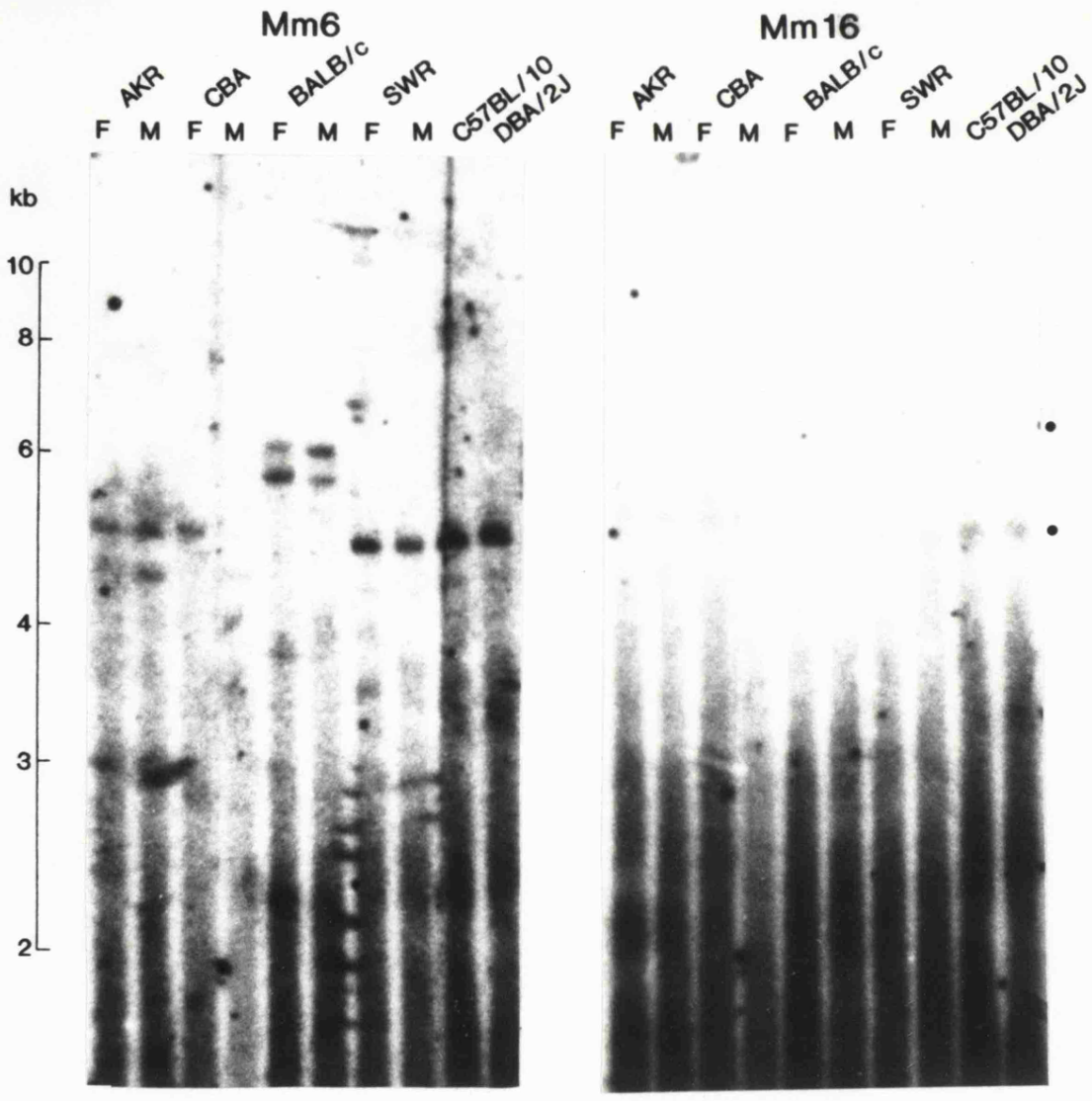
Two plaques from the library of 5.5-6.0kb DNA fragments hybridised to probe 33.6, both of which were stable isolates of *Mm6*. The 5.5kb insert of λMm6 was subcloned in pUC13 (to generate pMm6), and used as a hybridisation probe at high stringency (0.1x SSC, 65°) against mouse genomic DNA. This probe detected a 5.5kb fragment in C57BL/6J DNA digested with *Sau*3AI, but no cross-hybridising fragment was seen in *Alu*I digested DNA, suggesting that most, or all, *Mm6* repeat units contain a cleavage site for *Alu*I.

DNA from one male and one female mouse from each of several inbred strains was digested with *Sau*3AI and hybridised with the insert from pMm6 ; alleles were observed in a 4.5-6.5kb size range (Fig.8.3). A background smear indicated that the insert of pMm6 contained a sequence which was highly repeated in the mouse genome. C57BL/6J and DBA/2J mice shared a 5.5kb DNA fragment and therefore no SDP could be obtained for this locus from the BXD RI strains. This appears to be a stable minisatellite allele. In other mouse strains (BALB/c and AKR) the insert from pMm6 detected two strongly hybridising fragments. It is not clear in these cases whether both mice are heterozygous at *Mm6* or whether there is a second sequence-related locus in the genomes of these strains.

Figure 8.3

Mm6 and *Mm16*

DNA from selected inbred mouse strains digested with *Sau*3AI and hybridised with either Mm6 or Mm16. While alleles at *Mm6* are variable between inbred strains, Mm16 detects two monomorphic DNA fragments common to all strains tested (marked •). The hybridisation smear visble on each blot suggests that both probes contain DNA sequences which are highly repeated in the mouse genome. F, female; M, male.


Mm6
Mm16
AKR
CBA
BALB/c
SWR
C57BL/10
DBA/2J
F M F M F M F M
kb
10
8
6
4
3
2

8.3.c Mm16

At least three of the positive plaques in the library from which λMm3 was isolated were derived from a third locus, *Mm16*. Two isolates, λMm16 and λMm22 contained differently sized inserts (5.0kb and 5.7kb respectively). These inserts were subcloned into pUC13, and used as hybridisation probes; both inserts detected the same fragments in genomic DNA. In C57BL/6J DNA digested with *Sau*3AI two fragments of 5.8kb and 4.2kb were detected; in addition, a background hybridisation smear was observed, suggesting that the probe contains sequences which are highly repeated in the mouse genome (Fig.8.4). No DNA fragments were detected in *Alu*I or *Hin*fI digests. These fragments were shared by all inbred strains investigated, with no sign of variation or additional bands.

Neither *Mm16* nor *Mm6* were localised in the genome, and the minisatellite repeat sequence of these loci remains unknown.

8.4 *B10* : a locus rich in simple tandem repeat sequences

A λL47.1 library of large (15kb) size-selected *Bam*HI DNA fragments from a C57BL/6J mouse homozygous for a 15kb *Bam*HI *Ms6-hm* allele, was screened with Mm3-1 in an attempt to isolate more flanking DNA from *Ms6-hm* (see Chapter 3, section 7). 12 plaques which hybridised strongly with Mm3-1 were picked, replated and hybridised with probe 33.6 at low stringency, since isolates of *Ms6-hm* should hybridise strongly with both Mm3-1 and 33.6. However, none of these 12 clones were positive with 33.6; 4 of these, those which hybridised most strongly with Mm3-1, were further analysed by restriction mapping, and 3 were shown to be derived from the same locus (*B10*). In order to further characterise the cross-hybridising sequence in these 3 clones, a 3.6kb *Bam*HI-*Eco*RI fragment which cross-hybridised with Mm3-1, and contained presumptive minisatellite sequence, was purified from one of the multiple isolates (λB10) and subcloned into pUC13 to generate pB10.

A 1.6kb *Sau*3AI fragment, which cross-hybridised with Mm3-1, was isolated from the insert of pB10 and used as a hybridisation probe against mouse genomic DNA. Under low stringency washing conditions (1x SSC, 65°) this probe detected multiple fragments in *Hin*fI digested DNA (Fig.8.4). There were two components to the B10 DNA fingerprint pattern. First, a faint DNA fingerprint indistinguishable from that detected by Mm3-1 was observed in DNA from male and female mice. In DNA from male mice this was superimposed on a complex pattern of cross-hybridising fragments ranging in length up to >30kb. The male-specific fragments were highly variable between different inbred strains, showing much less variation within inbred strains; two BXD RI males had an

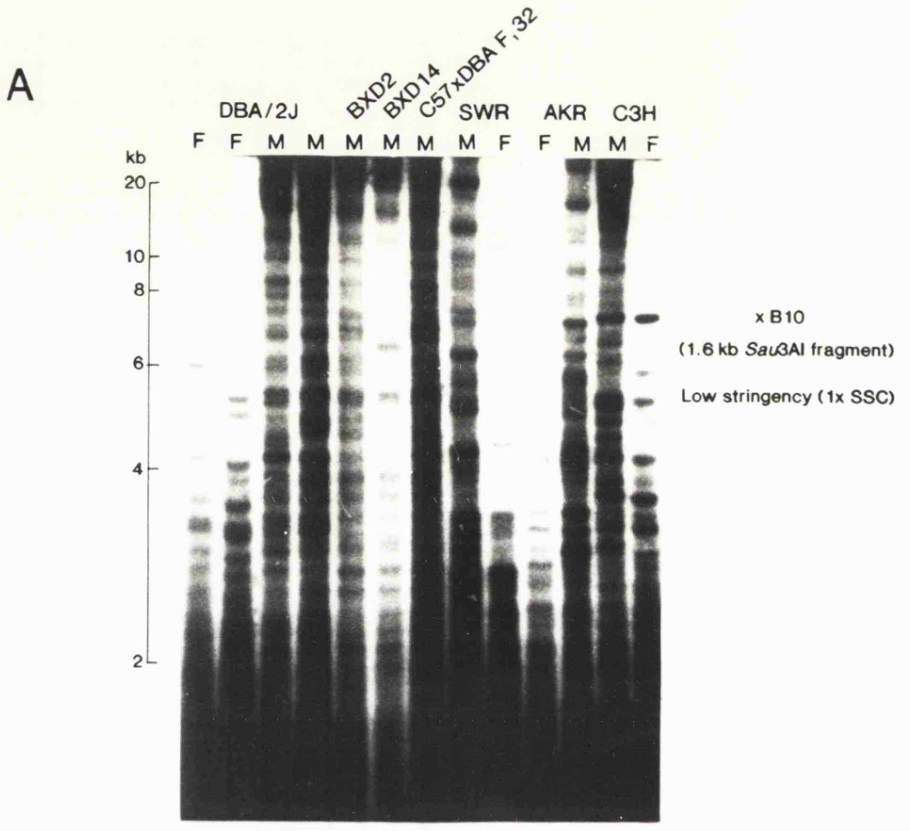
Figure 8.4

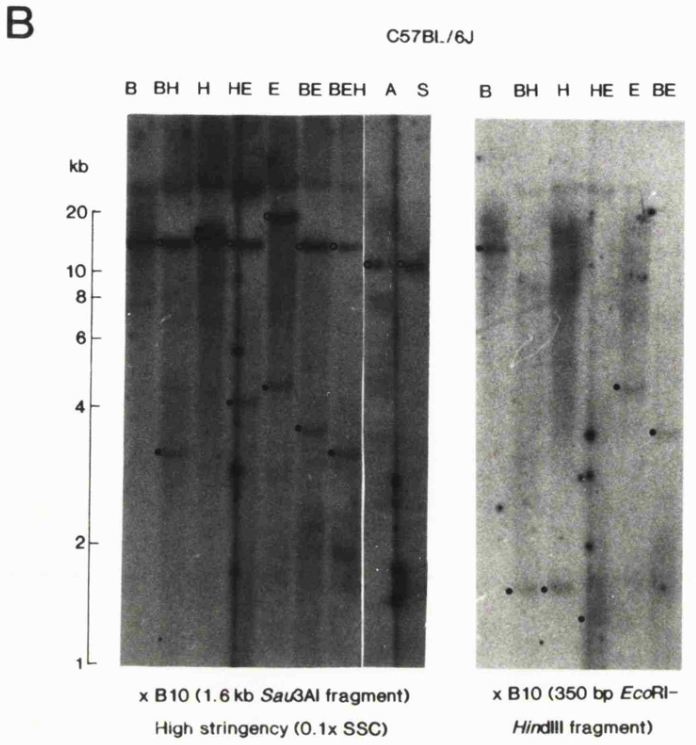
B10

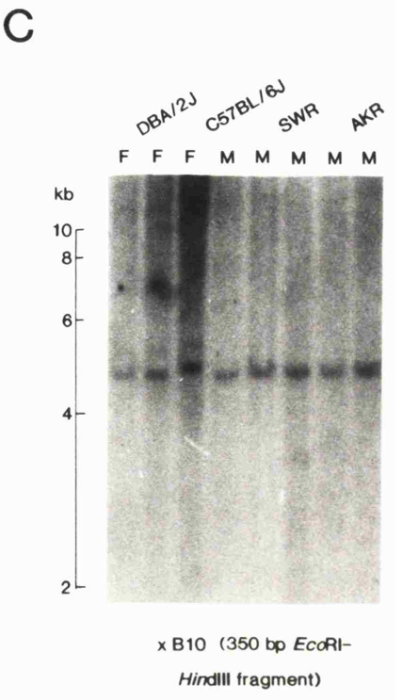
A. DNA from mice of several inbred strains digested with *Hin*fI and hybridised at low stringency (1x SSC, 65°) with a 1.6kb *Sau*3AI fragment derived from *B10*. This probe contains three simple repetitive elements; a GGGCA array which generates the DNA fingerprint most clearly observed in female (F) mice; a GATA array which generates the complex DNA fingerprint specific to male mice (M); and a GAAAAA array. Note that the cross-hybridising DNA fragments of BXD RI and DBA/2J male mice are similar.

B. C57BL/6J (male) brain DNA digested with a variety of restriction endonucleases and hybridised at high stringency (0.1x SSC, 65°) with either a 1.6kb *Sau*3AI probe (used in A) or a 350bp *Eco*RI-*Hin*dIII probe derived from *B10*. Genomic fragments derived from *B10* (•) and cross-hybridising fragments at *Ms6-hm* (o) are marked. Note that in this mouse 15kb *Bam*HI hybridising fragments are detected at both *Ms6-hm* and *B10*. The *Eco*RI-*Hin*dIII probe detects fragments at *B10* only, although in digests with *Hin*dIII the fragments observed are different to those detected by the 1.6kb *Sau*3AI probe (see Fig.8.5).

C. DNA from inbred mice of four different strains digested with *Eco*RI and hybridised with an *Eco*RI-*Hin*dIII fragment from *B10*. This probe detects a single monomorphic DNA fragment common to male (M) and female (F) mice of each strain.


A
DBA/2J
BXD2
BXD14
C57xDBA F,32
SWR
AKR
C3H
F F M M M M M M F F M M F
kb
20
10
8
6
4
2
x B10
(1.6 kb Sau3AI fragment)
Low stringency (1x SSC)


B
C57BL/6J
B BH H HE E BE BEH A S
B BH H HE E BE
kb
20
10
8
6
4
2
1
x B10 (1.6 kb Sau3AI fragment)
High stringency (0.1x SSC)
x B10 (350 bp EcoRI-
HindIII fragment)


C
DBA/2J
C57BL/6J
SWR
AKR
F F F M M M M M
kb
10
8
6
4
2
x B10 (350 bp EcoRI-
HindIII fragment)

almost identical pattern to a DBA/2J male, suggesting that the majority of male-specific fragments detected by this probe show substantial germline stability. The striking sex-difference in hybridisation pattern obtained with pB10 suggests that the 1.6kb *Sau*3AI fragment contains sequences which cross-hybridise to satellite sequences on the mouse Y chromosome, and resembles the pattern obtained in mouse DNA using a $(GATA)_n$ oligonucleotide probe (Shafer *et al.*, 1986).

Under high stringency washing conditions (0.1x SSC, 65°) the DNA fingerprint disappeared, and the only fragments detected by pB10 in *Hin*fI, *Alu*I or *Sau*3AI digested genomic DNA were alleles at *Ms6-hm* (in addition to a faint hybridisation smear at the exclusion point and throughout the DNA track of male mice). This result suggests that *Ms6-hm* is the only large sequence-related locus detected by pB10 under these hybridisation conditions which is not cleaved by these three enzymes. In *Bam*HI, *Hin*dIII and *Eco*RI restriction digests a second locus, *B10* itself, was detected (Fig.8.4). The restriction map of *B10* in C57BL/6J genomic DNA (from the mouse whose DNA was used to construct the library) precisely matched that of the insert from λB10. In this mouse *Bam*HI alleles at both *Ms6-hm* and *B10* are 15kb in length, explaining the presence of fragments derived from *B10* in the size-selected library. Both *Hin*fI and *Alu*I cleave the 1.6kb *Sau*3AI DNA fragment from pB10 to fragments less than 400bp in length; *B10* alleles are therefore not resolved in a *Hin*fI Mm3-1 DNA fingerprint.

The genomic library from which λB10 was isolated was constructed from the DNA of a male mouse; *B10* itself might therefore be a Y linked locus. A 350bp *Eco*RI-*Hin*dIII fragment from outside the 1.6kb *Sau*3AI fragment of pB10 was used as a single-locus hybridisation probe for *B10* (Fig.8.4). This probe detected the same 4.5kb *Eco*RI fragment in DNA from male and female mice of several inbred strains; a hybridisation smear was also detected, in DNA from both sexes, suggesting that the probe contains sequences which are highly repeated in the mouse genome (Fig.8.4). *B10* is therefore autosomal or pseudoautosomal in location, and, as this *Eco*RI fragment includes the tandem repeat region, is not itself highly variable.

The 1.6kb *Sau*3AI fragment of pB10, and several internal *Alu*I partial digest fragments, were subcloned into M13mp19 to examine whether the sequences responsible for the GATA-type and Mm3-1-type DNA fingerprints were components of the same tandem array. The DNA sequence of this region is illustrated in Fig.8.5. Three major tandem arrays were found:

a. 4 perfect GGGCA tandem repeat units in a 40-50bp GC rich region (80%). This accounts for the strong cross-hybridisation to Mm3-1, and is presumably not long enough to hybridise to probe 33.6 (which hybridises only very weakly to the 19 GGGCA repeats of $pMm3\text{-}II_1$ (see Chapter 3)).

Figure 8.5

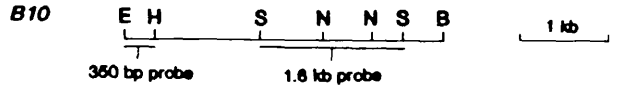
Structure and DNA sequence of *B10*

A. Structure of the 3.6kb *Bam*HI-*Eco*RI fragment derived from *B10* (insert of pB10), showing the *Eco*RI-*Hin*dIII and *Sau*3AI fragments used as probes in the experiments illustrated in Fig.8.4. E, *Eco*RI; H, *Hin*dIII; S, *Sau*3AI; N, *Nco*I; B, *Bam*HI.

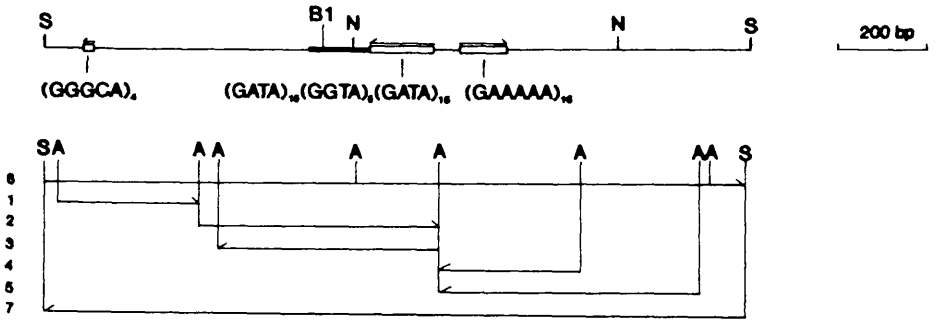
B. Structure of the 1.6kb *Sau*3AI fragment derived from *B10*, showing the position of the three tandem repetitive arrays (denoted by open boxes) and the diverged B1 element (closed box). The M13 sub-clones constructed for DNA sequence analysis of this fragment are also illustrated; these contained either the entire *Sau*3AI fragment, or internal partial *Alu*I fragments, cloned into M13mp19 linearised with either *Bam*HI, or *Sma*I, respectively. A, *Alu*I; N, *Nco*I; S, *Sau*3AI.

B. DNA sequence of the 1.6kb *Sau*3AI fragment determined from the sub-clones illustrated above. DNA sequences immediately adjacent to the *Sau*3AI sites were not obtained. The three tandem repeat arrays are underlined; nucleotides showing identity to the B1 repeat sequence of Kalb *et al.* (1983) are in bold type (5' end at position 544). The precise position of the direct repeats flanking the B1 element is unclear; however, 6/7 of the nucleotides immediately preceding the B1 element are found repeated at nt 860 and nt 1110, suggesting that the GATA tandem array (nt 695-834), and possibly also the $(GAAAAA)_n$ array (nt 901-996), have expanded from within the 3' tail of the B1 element.

A



B



C

B10 : Sau3AI Fragment

```
   1  ATAGACATCA CCAAACCCAG CATTCCATTA CTCTCTATGT GTGAAGCTGA GAGATTCTAA NNNNCCCGNC CCGCCCTGCC GTGCCCTGCC CTGCCCTGCC
 101  CCGCCCTGTC CCCAAACTGT ACAGTAACAA GGAGGAGGCA AGAGATGATG ATGGTTTTGA GCAAATAAAC TATGGATGGT TGATGGAGAA TGTAGAAATT
 201  TACGAAGACA TTTAGGAGGT TTGCTTAAGG GATGATTCTT CCTACCTTCA TTCATCAGTA AAATATTTGT ATGTTTTCAA CACAGTGTGG GGTTGCATTT
 301  GAATTAAGGT TGTTTAAAGT GGGGCTTCCA ATGAAGCTGA AAAAAGAGGG ATTCTAGGTA GGGAGGGAAT CACAATGGTC TAGCTTAAAA GATTCTAAGA
 401  CTTAGATAGT ATATAGTAAT TTAAAAATGA ACCTCAACAA TGTATTCAGC AAATTATTTA AAGTTGAATT ATGAATACCT GTTATGTGAC ACCTACCTGC
 501  CTGCCTTAAT AAATATCTGG ACAAACAGGA AGGGAATTTA GAACCAGGAA TGGGGACACA CCTTTAATCT CAGCATTCTG GGGGCAGAGG CAGGTATGAC
 601  CCTGTGAATC CAAGCCAGTC TGGCCGACAT AAAGAATTAT AGATAATCCA TGGCTATCCT CTAGAGCTCT GTCTAAAAAG AAAAGGCTAA ATAGTATCTA
 701  TCTATCTATC TATCTATCTA TCTATCTATC TATCTATCTA TCTATCTATC TATCTACCTA CCTACCTACC TACCTATCTA TCTATCTATC TATCTATCTA
 801  TCTATCTATC TATCTATCTA TCTATCTATC TATCTAGATA AAACAAAAAT TTAAGTAACA TTTAAATGTT ATAACAATTG TTCAAGCTCC ACTTTAAAAA
 901  GAAAAAGAAA AAGAAAAAGA AAAGAAAAA GAAAAAGAAA AAGAAAAAGA AAAGAAAAA GAAAAAGAAA AAAAAAAAGA AAAGAAAAA GAAAAATGCT
1001  GTCATCTGGA AACTTTAGTC CAGAACAAAG AAGGTTATTC TATTCCTTCC TCTCTCCAGA TGACTGACTT CATTCCTACT CGTTAAAAGA TGACTCTCCC
1101  ACCCATATTC ACTTCCACCC ACAACCATAT ATTGTTTAAC AATCCTAGAC TTGTTCTGTG GTAATATATA GAGTCAATAA CAAAGCTAAC TGTAGTTACC
1201  ATGGCTGAGT GCACTTAGCA CGGCAGTATT TAATATGTAG TAAGTATTCA AAGACAGAAC TGACTCATTA CAGTGCAAGA AATATTTTCC CTTATAACAT
1301  CCAAATAACA GTCAAAAATT ATTTCAAATG TTAAAGGTAA AATTTGAACC AATCCAGTTC TCTGTGTATC ACTCCTACTC TTTTTTCTGA AATCATTGTC
1401  TTCCCTTAAC TCTGAATGGC TTATCTCATA TTTCATTAAA GTTTTATGGC AGGTTTTCCT CTCAGCTTTA TGGGATTTTT TAAATATATA ATAAGAGTTG
1501  TAAAAGGAAG CTAGTGACAT ATATACTTAA ATTCATACTG TTTATAAAAT ACAAAATAAA AAT
```

b. Approximately 540bp from the GGGCA array is a simple quadruplet tandem repeat array. This consists (on the same strand as the GGGCA array) of 15 GATA repeat units either side of 5 variant GGTA repeat units. This sequence accounts for the male-specific hybridisation signal. The GATA array is in turn flanked on either side by an extremely T rich region (24/35 adjacent nucleotides).

c. A third, unexpected, simple sequence array was found 59 nucleotides beyond the GATA repeat, composed of 16 TTTTTC tandem repeat units (on the same strand as the GATA and GGGCA arrays). One repeat unit had a C to T nucleotide difference. The 59bp between the GATA and TTTTTC arrays are 80% AT rich.

The non-tandemly repeated sequences within this clone were used to search the EMBL and GenBank DNA sequence databases. A diverged B1 element lies between the GGGCA and GATA arrays (Fig.8.5). This element shows approximately 65% similarity to the consensus sequence of Kalb *et al.* (1983) over the entire 130bp, and approximately 80% over the first 50bp. The (A)-rich tail starts 20nt before the GATA array, suggesting that this array (TATC) has expanded from within the 3' sequences of the B1 element. There are no obvious long direct repeats flanking this element, and it is therefore unclear whether the GAAAAA array has also expanded from within the 3' tail. Short repeats of sequences immediately preceding the element are found either side of the GAAAAA array (6/7nt, at positions 860 and 1110 in Fig.8.5). This complex collection of repetitive elements suggests that *B10*, like *Ms6-hm*, is prone to the accumulation of both simple and retroposon dispersed repeat sequences (see Discussion).

The autosomal localisation of *B10* remains unknown. It is possible that the tandem arrays within this sequence exhibit microheterogeneity, which could be exploited using the polymerase chain reaction to provide strain-specific markers. This might in turn allow the assignment of an SDP to this locus.

8.5 TTAGGG-related sequences in mouse DNA

The tandem repeat sequence TTAGGG is found at the telomeres of human chromosomes (see Allshire *et al.*, 1988, Richards and Ausubel, 1988, Moyzis *et al.*, 1988). Probes based on this sequence also detect interstitial loci which are highly variable in human DNA. A TTAGGG repeat probe was synthesised from two complementary oligonucleotides by annealing, ligation, and amplification using the polymerase chain reaction (N.Royle, Leicester). This probe was hybridised to *Alu*I digested DNA from the BXD RI strains, in order to investigate the organisation of TTAGGG-related loci in the mouse genome (Fig.8.6).

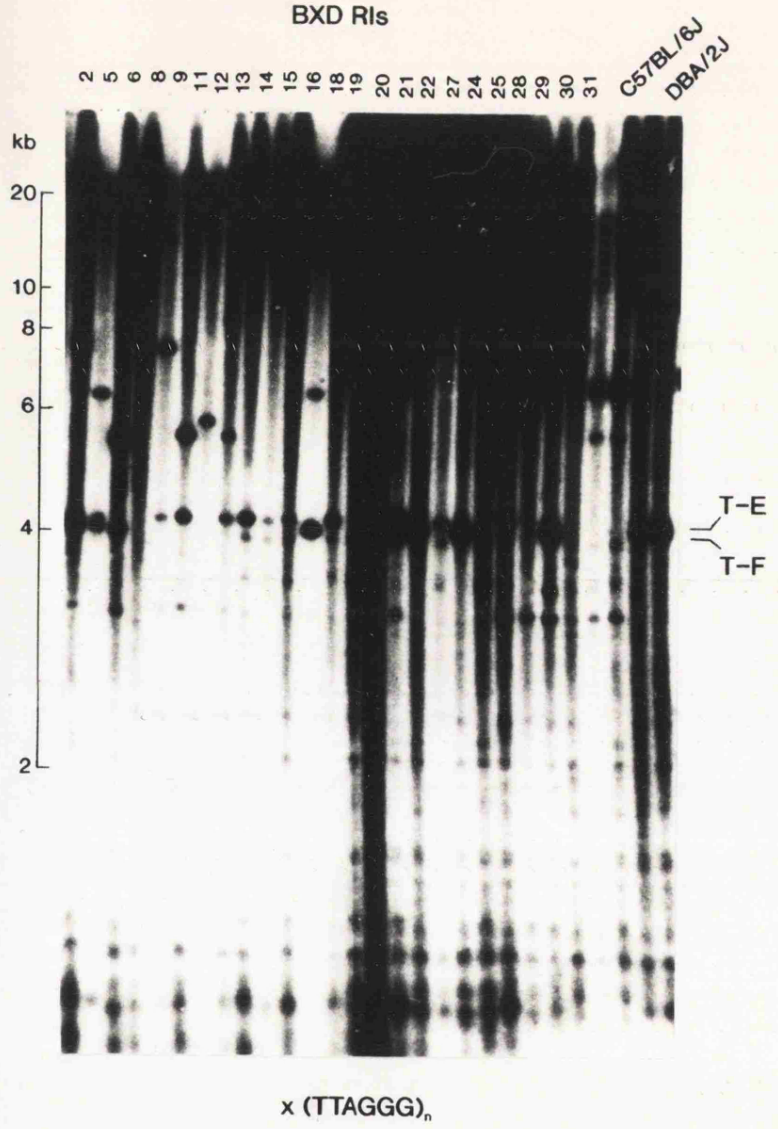
Figure 8.6

Segregation analysis of TTAGGG-related loci in the BXD RI strains

A. BXD RI and progenitor strain (C57BL/6J and DBA/2J) DNAs digested with *AluI* and hybridised with a synthetic $(TTAGGG)_n$ probe. The very large (>15kb) cross-hybridising DNA fragments may correspond to mouse telomeric sequences, while the discrete fragments (<10kb) may represent internal TTAGGG-related tandem arrays. While some of the smaller fragments are found in only one or a few RI strains such that they cannot be assigned an SDP, others are more stably inherited, in particular fragments T-E and T-F, scored in the DBA/2J progenitor strain, for which SDPs have been obtained. Hybridisation by N.Royle, Leicester.

B. Strain distribution patterns of TTAGGG-related DNA fragments T-E and T-F from A, above. The C57BL/6J allelic fragments could not be resolved for these loci.

A



B

Locus		BXD RI																							
		1	2	5	6	8	9	11	12	13	14	15	16	18	19	20	21	22	24	25	27	28	29	30	31
T-E	(D)	D	D	B	B	D	D	B	D	D	D	D	B	D	B	B	D	B	B	B	D	D	B	D	B
T-F	(D)	D	D	D	B	B	B	B	B	B	B	B	D	B	B	B	D	D	D	B	B	B	B	D	B

The telomeric arrays at the ends of human chromosomes are heterogeneous in length, resulting from the addition of different numbers of terminal repeats in different cells, and are detected as a smear in human DNA (Cooke *et al.*, 1985). Very large (>30kb) strongly hybridising fragments common to C57BL/6J and DBA/2J progenitor strains segregated to all the RI strains. These fragments may be heterogeneous, and are assumed to represent the ends of mouse chromosomes. As these DNA fragments are larger than the 10-15kb telomeric smears detected in *Sau*3AI or *Hin*fI digested human DNA (Hastie and Allshire, 1989), it appears that mice may have longer telomeric arrays than man. Addtional 10-20kb fragments are detected by TTAGGG in some, but not all, BXD RI strains.

Many fragments smaller than 10kb are detected in *Alu*I digested mouse DNA by the TTAGGG probe (Fig.8.6). These are discrete fragments, and are thought to represent interstitial TTAGGG-related arrays. TTAGGG-related segments have also been isolated on *Sau*3AI inserts in a mouse genomic library (M.Festing, personal communication). The proximity of TTAGGG-related sequences to the natural end of a chromosome can be investigated by examining the sensitivity of the corresponding fragments to *Bal*31 exonuclease digestion (Cooke *et al.*, 1985). Studies have shown that discrete TTAGGG-related DNA fragments of less than 10kb are insensitive to *Bal*31 digestion, and are thus not located at chromosome termini (see Hastie and Allshire, 1989). The majority of these fragments observed in the BXD RI strains are hypervariable, and found in only one or a small number of strains. It is therefore impossible to obtain SDPs for these fragments on the basis of allele length.

Other, more stable, TTAGGG-related DNA fragments may be assigned SDPs; for two of these fragments linkage to other genetic markers has been detected (B.Taylor, The Jackson Laboratory, personal communication). One of these has 3/24 discordancies with the *Fgr* proto-oncogene on the distal part of chromosome 4, and also with a sequence on the proximal part of chromosome 3. The other stable fragment for which an SDP has been found also shows 3/24 discordancies with two different loci, one near the middle of chromosome 5, and another (*Cck*) on the distal end of chromosome 9. While the correct chromosomal assignments for these two loci remain unclear, it seems likely that interstitial TTAGGG-related arrays are dispersed in the mouse genome.

8.6 Summary

The loci described in this chapter illustrate the extent to which core-related minisatellite loci are only the 'tip of the iceberg' of tandem repeated sequences in the mouse genome (Dover, 1989). The structure and germline stability of these loci may be

compared : *Ms6-hm*, which consists of a homogeneous array of 5bp tandem repeat units is extremely unstable, whereas *Mm1*, which consists of a heterogeneous array based on a 40bp repeat unit, is relatively stable. Such comparisons allow insights into the processes associated with minisatellite evolution (see Discussion).

The DNA sequence of *B10* demonstrates that not all GGGCA tandem repeat loci in the mouse genome are equally hypervariable; variant *B10* alleles, of length differences exceeding 20 repeat units, would have been observed across different inbred strains. It is possible that there are only a small number of large extremely unstable minisatellite loci in the mouse genome; these include TTAGGG-related arrays which may evolve similarly to minisatellite loci (through slippage and unequal exchange events), or alternatively (and perhaps additionally) through the internalisation of telomeric repeats through chromosome fusions (see Hastie and Allshire, 1989). More detailed analysis of the molecular organisation of mouse chromosomes is required in order to understand the mechanisms by which the variety of tandem repeat arrays in the mouse genome have arisen, and by which these highly unstable sequences are continuing to evolve.

IX. DISCUSSION

9.1

Characterisation of human and mouse minisatellite loci has provided insights into the structure of unstable sequences and the processes associated with instability in the mammalian genome. The hypervariable mouse minisatellite *Ms6-hm* has been cloned from the DNA of a C57BL/6J mouse, allowing analysis of the primary structure and new-length mutation events at this locus, and the application of *Ms6-hm* as both a lineage marker for early mouse development and a probe to detect other unstable sequences in the mouse genome.

9.2 Sequences associated with *Ms6-hm*

Analysis of the DNA sequence flanking *Ms6-hm* revealed that the minisatellite lies within a member of the MT family of short interspersed repetitive elements (Heinlein *et al.*, 1986). This MT element (MT-1) is flanked by two additional MT elements in reverse orientation, MT-2 (100bp 5') and MT-3 (200bp 3'). A fourth dispersed repetitive sequence, a B2 element, lies 2.2kb 3' to MT-1, in the same orientation.

MT elements are a newly described high copy number rodent-specific SINE family which are repeated 40-90,000 times in the mouse genome (see Introduction, Heinlein *et al.*, 1986, Bastien and Bourgaux, 1987). The relatively recent characterisation of this repeat family suggests that MT elements may have an unusual distribution in the genome. In addition to the seven elements which were cloned by these groups to establish MT consensus sequences only 12 MT elements are currently in the EMBL DNA sequence database (Release No. 19). Bastien and Bourgaux (1987) demonstrated by heteroduplex mapping that MT elements are 400bp in length, and identified two sequences in the MT consensus sequence which displayed weak similarity with the A (8/11bp) and the B block (11/12bp) of the RNA polymerase III promoter (Galli *et al.*, 1981), in addition to a sequence similar to the replication origin of SV40 (11/12bp, 150bp into the consensus sequence, Fig.9.1). These authors reported that three elements, isolated by sequence similarity to an MT element in the mouse Ins sequence, terminate with a short 3' sequence rich in (A) residues, and are flanked by short direct repeats.

The GGGCA repeat unit of *Ms6-hm* appears to have expanded from within MT-1, presumably through processes of slipped strand mispairing and unequal exchange (see Introduction). The minisatellite starts approximately 135bp into the element at a point where there are two copies of the tandem repeat unit GCAGG, separated by 7bp, in the MT consensus sequence (Bastien and Bourgaux, 1987). MT-2 and MT-3 are less similar to the consensus sequence than MT-1; while similarity at MT-2 starts 55bp into the consensus sequence, similarity at MT-3 ceases after 135bp, precisely the point at which the minisatellite has expanded from MT-1. Two of the MT elements sequenced by Bastien and Bourgaux (1987) have small (7 and 9bp) non-repetitive insertions between the two GCAGG repeats at this point; furthermore, at a shortened MT element upstream of the cytochrome P450 oxidoreductase gene of the rat, strong similarity to the MT consensus sequence starts in this region (Fig.9.1, Gonzalez and Kasper, 1983). These data suggest that this region may be particulary labile within MT elements.

Recombination between dispersed repeats is associated with deletions and duplications in the mammalian genome (see Introduction). The organisation of the MT elements around *Ms6-hm* suggests that MT-2 and MT-3 may once have been a single ancestral MT element which was disrupted by a nomadic sequence carrying MT-1 (or into which MT-1 subsequently integrated). MT-1 shows greater similarity to the MT consensus sequence than either MT-2 or MT-3, and therefore might be argued to have integrated more recently. However, within the 80bp region of the consensus MT sequence to which both MT-2 and MT-3 show similarity there is no greater similarity between these two elements than between either of these and any other MT repeat (from dotplot analysis). Furthermore, there is no evidence for any short sequence 5' of MT-3 which is directly repeated 3' of MT-2.

Bastien and Bourgaux (1987) argued that MT elements have the structural properties of retroposons. Rogers (1985) has analysed the sequences generated upon retroposon insertion in 109 cases including B1, B2, Alu, LINE, and pseudogene retroposons; the majority of (A)-rich tails are of the strucure A_n, $(NA_x)_y$, or A_n-$(NA_x)_y$ (N being frequently C), with a minority of (A) tails in each group being irregular. The average length of the terminal repeats in each retroposon group was 12-13bp. These features, common to all major retroposon groups, contrast with the characteristics of sequences flanking MT elements. The three '(A)-rich tails' identified by Bastien and Bourgaux (1987) are highly irregular and contain only 3/8, 7/23 and 15/33 (A) residues. Furthermore, none of the MT elements associated with *Ms6-hm* have detectable (A)-rich tails, nor do other MT elements in the EMBL database.

Direct repeats flanking MT elements are also non-typical; two of those reported by Bastien and Bourgaux (1987) have 3/8 mismatches, and two do not directly abut the 5'

end of the element. 8bp, overlapping with the first 4bp of MT-1, are repeated perfectly 20bp 3' to the element; however, the direct repeats flanking retroposons should be strictly of target sequence, and the 3' tail (of which 3nt out of 20 are (A)) is highly irregular (Rogers, 1985). It is possible that MT-1 is truncated by 4bp at its 5' end; MT-2 and MT-3 have no detectable direct repeats at either the expected or observed boundaries of similarity with the MT consensus sequence. None of the other MT elements in the EMBL database for which the 5' and 3' ends have been completely sequenced have longer or more perfect direct repeats than those reported by Bastien and Bourgaux (1987).

MT elements are therefore not a typical retroposon family. This is clearly illustrated by contrasting the characteristics of the MT elements associated with *Ms6-hm* with the long poly(A) tail and direct repeats of the downstream B2 element. Dotplot analysis comparing 8 MT elements against each other (at a stringency of 11/17 nucleotides) revealed that in no case does sequence similarity extend beyond the 400bp consensus established by Bastien and Bourgaux (1987). There was no evidence for retroposon-like (A)-rich tails at either end of any of these sequences. In addition, many of these MT elements were truncated, or showed patchy similarity to the MT consensus sequence; the fact that MT elements have not been found in libraries of highly repetitive DNA sequences further suggests that individual elements may not be closely related. The most simple explanation for these results is that those MT elements which have been analysed to date were active shortly after the divergence of rodents (65 MYA), and have since been eroded by point mutation, masking the hallmarks of retroposition. It is likely, by analogy with Alu sequences in primates (Britten *et al.*, 1988), that there may be more than one MT consensus sequence corresponding to different source genes active during rodent evolution; perhaps there is even a precise class of MT elements currently active in the mouse genome. Alternatively, characterisation of individual MT elements may reveal that they are not retroposons, and uncover novel mechanisms for their dispersal in the genome.

MT elements are often associated with unstable regions of the genome. The first of six examples discussed here was also the first MT sequence to be isolated, and lies within a region of somatic instability, the mouse Ins sequence. Ins is a 1.6kb cellular sequence which is precisely excised at a high frequency with viral DNA from a polyoma virus transformed mouse cell line (Fig.9.1, Bourgaux *et al.*, 1982). In this particular cell line there are two complete polyoma genomes joined head to tail in tandem at a single integrated site (Chartrand *et al.*, 1981). On induction of viral replication a homogeneous population of recombinant molecules is produced, each molecule consisting of a complete polyoma genome (5.3kb) with a 182bp direct repeat and 1.6kb of flanking cellular sequences (Ins). This chimaeric molecule (RmI) shares an identical arrangement of

Figure 9.1

MT elements in mouse DNA

A. Consensus MT sequence of Bastien and Bourgaux (1987), determined from the Ins MT sequence and three MT elements isolated by cross-hybridisation to the Ins MT sequence. This consensus is 97% similar to the trunctated MT consensus of Heinlein *et al.* (1986). The first two underlined sequences (nt 2-12, nt 56-67) display weak similarity with, respectively, the A block (8/11nt), and the B block (7/12nt) of the RNA polymerase III promoter (Galli *et al.*, 1981). Two GCAGG sequences at the point from which *Ms6-hm* has expanded from MT-1 are also underlined (nt 132-136, nt 144-148). A sequence showing similarity to the SV40 origin of replication is boxed (11/12nt, nt 154-165).

B. Structure of seven loci associated with unstable DNA sequences, and the upstream region of a rat P-450 gene, which contain MT sequences (see text). Note that the structure of PRP M14 is illustrated to half the scale of the other loci.

Key - dispersed repeats : MT, Mouse Transcript element (arrow indicates direction of strand similar to the consensus sequence); B1, B2, mouse B1 or B2 element; IAP, intracisternal A-particle (showing LTRs); LTR at Aβ hotpsot, 36bp region similar to a portion of the LTR of endogenous MuRRS elements (Schmidt *et al.*, 1985); L1Md, mouse LINE; simple repeat regions are also indicated, with the repeat unit sequence in brackets. Polyoma virus integrated genome (Py DNA) is linked to the Ins sequence; broken line, cellular α-site; continuous line, cellular β-site (see text). Minimal regions containing meiotic cross-over points within the MHC recombination hotspots are indicated by broken arrows. Gene structure for SAA2 and PRP M14 : I-IV, exons; 5', 3', boundaries of homology between the SAA2 and SAA1 genes (and of the putative gene conversion region); A, B, PRP M14 3' polyadenylation sites.

References - *Ms6-hm*, present work; Ins, Sylla *et al.*, 1984a, Bourgaux-Ramoisy *et al.*, 1986, Allard *et al.*, 1988; MHC Aβ and Eβ meiotic recombination hotspots, Uematsu *et al.*, 1986, Kobori *et al.*, 1986, Shiroishi, T., Hanzawa, N., Sagai, T., Steinmetz, M., and Moriwaki, K., personal communication; Renin-2 gene (3' IAP), Burt *et al.*, 1984; serum amyloid A2 gene, Lowell *et al.*, 1986; proline rich gene M14, Ann *et al.*, 1988; rat NADPH-cytochrome P-450 oxidoreductase gene, Gonzalez and Kasper, 1983.

C. Dotplots comparing MT elements at the loci illustrated in B to the MT consensus of Bastien and Bourgaux (1987). The sequences are compared in windows of 17 nucleotides, of which 11 need to match to place a dot on the graph. Arrows indicate the point at which the GGGCA array has expanded from MT-1; the direct repeat within MT-3 at which similarity to the MT consensus ends; the start of strong similarity to the MT consensus within the rat P450 MT element; the border of the region of sequence similarity with SAA1 within the SAA2 MT element, and the point within the *Ren-2* MT element at which the IAP genome has integrated.

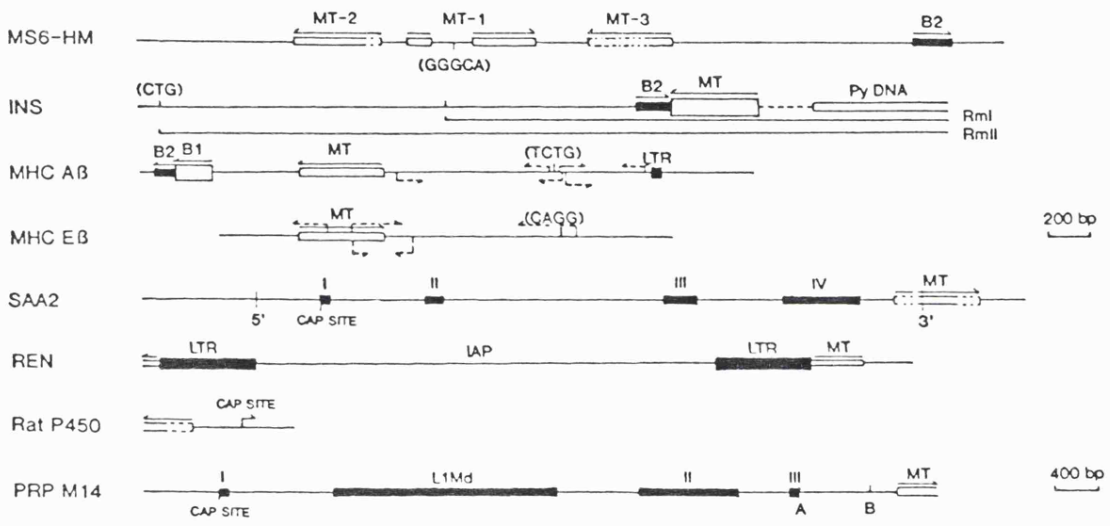
The sequences included are : MT-1, CB4, nt 1280-1790 (only 5 GGGCA repeats included to maximise the alignment); MT-2, CB4, nt 760-1200 (reverse and complement); MT-3, CB4, nt 1980-2410 (reverse and complement); INS, EMBL: MMINS01, nt 240-680 (Sylla *et al.*, 1984a); Rat P-450, EMBL: RNCYC450, nt 1-250 (reverse and complement, Gonzalez and Kasper, 1983); MHC Aβ, EMBL: MMAB322, nt 684-1124 (reverse and complement, Uematsu *et al.*, 1986); MHC Eβ, EMBL: MMMHIEBF, nt 1026-1465 (Kobori *et al.*, 1986); SAA2, EMBL: MMSAA2B, nt 3600-4020 (Lowell *et al.*, 1986); PRP M14, New EMBL: J03891, nt 6920-7306 (Ann *et al.*, 1988); REN, Genbank: IAPMSV, nt 1-97, 3106-3303 (reverse and complement, deleting IAP genome to maximise alignment, Burt *et al.*, 1984).

A

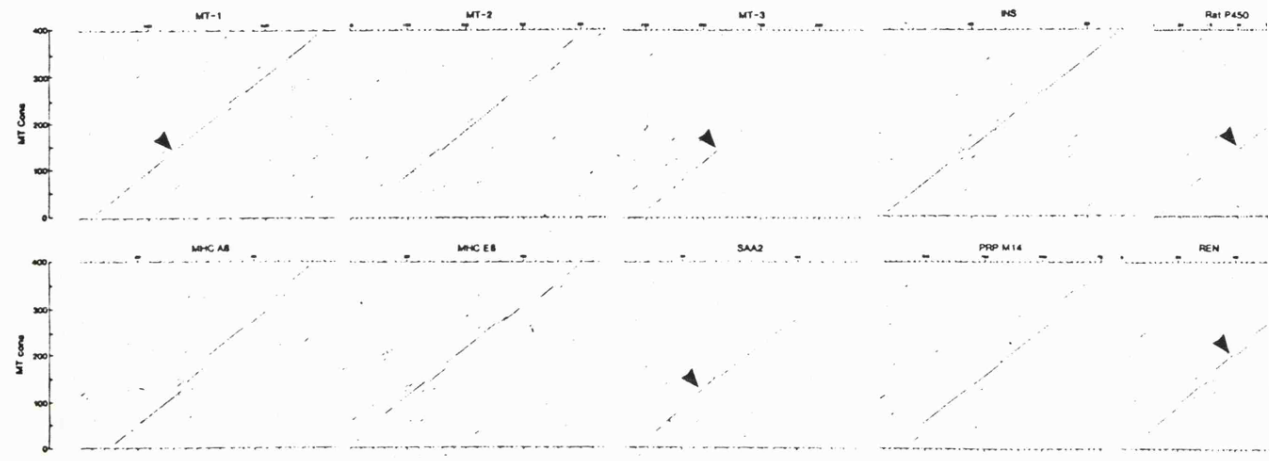
MT CONSENSUS SEQUENCE

```
  1  CTGTCTTAGT CAGGGTTTCT ATTCCTGCAC AAACATCATG ACCAAGAAGC AAGTTGGGGA GGAAAGGGTT TATTCAGCTT ACACTTCCAC ACTGCTGTTC
101  ATCACCAAAG GAAGTCAGGA CTGGAACTCA AGCAGGTCAG AAAGCAGGAG CTCATGCAGA GGCCATGGAG GGATGTTGCT TACTGGCTTG CTTCCCTTGG
201  CTTGCTCAGC CTGCTTTCTT ATGGAACCCA AGACTACCAG CCCAGGGATG GCACCACCCA CAATGGGCCC TTCCCCCTTG ATCACTAATT GAGAAAATGC
301  CTTACAGTTG GATCTCATGG AGGCATTTCC TCAAGGGAGG TTCCTTTCTC TGTGATAACT CCAGCTTGTG TCAAGTTGAC ACACAAAACC AGCCAGTACA
```

B



C



sequences on either side of one viral-cellular join with the insertion site in uninduced cells. The Ins sequence is thought to mediate excision of the Ins-polyoma chimaeric molecule through a site-specific recombinational process (Sylla *et al.*, 1984a).

Wallenberg *et al.* (1984) transfected RmI into rat cells and found that, under conditions of viral replication, RmI integrated non-randomly into the genome, implying that there was a specific interaction between Ins and cellular sequences during integration as well as excision. RmI itself is too large to be packaged into viral capsids; however, this molecule, with the structural properties of a bacterial cointegrate, recombines to produce unit sized polyoma genomes which may be packaged to form infectious particles (Piche and Bourgaux, 1987). This recombination event occurs with $1/20^{th}$ the efficiency of the initial precise excision of RmI.

Isolation of a second chimaeric molecule, RmII, containing 3kb of flanking DNA (including Ins), allowed analysis of the highly efficient recombination event which generates RmI (Sylla *et al.*, 1984b); patches of in-register homology were found on either side of the crossover site, and the two sequences which join to create RmI were found to contain a 12-14bp inverted repeat. Subsequent cloning of cellular sequences allowed analysis of the alternative recombination event which generated RmII (Bourgaux-Ramoisy *et al.*, 1986). This crossover occurs within a simple sequence array ($(CTG)_{33}$), at a point where a G to A transition generates a 5bp match with the viral sequence. Oligonucleotides either side of this array may be responsible for aligning the viral and cellular sequences prior to recombination. By comparing cellular sequences from the occupied and unoccupied integration site, Allard *et al.* (1988) found that the original integration event was associated with a deletion (of unknown extent), such that the Ins sequence is composed of DNA segments from two distinct sites; the 210bp adjacent to the genomic viral-cellular join (or α-site) were juxtaposed to another cellular sequence at the time of integration. This second cellular sequence (or β-site) is polymorphic for a 1.8kb insertion/deletion event between different mouse strain DNAs.

The Ins MT element appears to be full length, and lies within the sequence derived from the β-site; the 5' end of the element is 36bp from the cellular-cellular junction (Fig.9.1, Sylla *et al.*, 1984a). At the 3' end a truncated B2 element (5' 100bp) lies in the opposite direction, directly following an inverted repeat of the last 15bp of the MT element (12/15bp). This is the only example (out of 15 MT elements with known flanking sequences) of the putative insertion of an MT element, although there is no evidence for continuation of the B2 repeat 5' to the MT sequence. The best direct repeat flanking the Ins MT element is a 5/5bp repeat immediately preceding and following the element. There is no evidence for a poly(A) tail.

A second MT element associated with genomic instability is found 3' to both the *Ren-1* and *Ren-2* genes of DBA/2J mice (Fig.9.1). The *Ren-2* gene is thought to have arisen from the *Ren-1* gene through a tandem duplication event. A retroviral intracisternal A-particle (IAP) genome has integrated into this element at the *Ren-2* locus, 205bp into the MT consensus sequence (Burt *et al.*, 1984). The IAP genome is flanked by 6bp direct repeats, and the MT sequence continues beyond the second direct repeat. This MT element shows poor similarity to the consensus sequence.

Thirdly, an MT element is found in the 3' sequences of the serum amyloid A 2 (SAA2) gene of the mouse (Fig.9.1). A region of 3215bp shows 96% homology between the SAA1 and SAA2 genes, including intervening, upstream and downstream sequences and is flanked by non-homologous sequences (Lowell *et al.*, 1986). These authors argue that this extensive homology is the result of gene conversion events. A B1 element is found 73bp upstream of the 5' boundary of homology in the SAA1 gene. The MT element lies 3' to the SAA2 gene; the boundary of homology with the SAA1 gene and the start of strong similarity to the MT consensus sequence are both found approximately 125bp into the MT element (close to the labile site discussed above). Only the central third of this MT element shows strong similarity to the consensus sequence of Bastien and Bourgaux (1987), or any other (of 7) MT elements by dotplot analysis.

MT elements are also found at the Aβ3/Aβ2 and Eβ hotspots of meiotic recombination in the mouse major histocompatibility complex (see Introduction; Shiroishi, T., Hanzawa, N., Sagai, T., Steinmetz, M., and Moriwaki, K., personal communication). The molecular organisation of these two hotspots is remarkably similar; at each locus the tetrameric tandem repeat sequences ($(CAGA)_{4-6}$ at the Aβ, or $(CAGG)_{7-9}$ at the Eβ hotspot) lie approximately 1kb beyond the distal end of the sequence related to the papovavirus replication origin in each MT element (Fig.9.1). The $(CAGR)_n$ strand at each hotspot is orientated in the same direction as the MT consensus sequence. In addition, approximately 400bp beyond the tetrameric array, Uematsu *et al.* (1986) identified a 36bp sequence (containing a breakpoint in one recombinant strain) which shows 72% similarity to a portion of the LTR sequence associated with MuRRS (retroviral-related) elements (Fig.9.1). Edelmann *et al.* (1989) have demonstrated that this region of the LTR interacts with at least two nuclear proteins *in vitro*, and is likely to be responsible for the high frequency of recombination between MuRRS LTRs in the mouse genome. At the Aβ hotspot truncated B1 and B2 elements are found 500bp from the MT sequence (at the opposite side to the CAGA array, Uematsu *et al.*, 1986). Both the Aβ and Eβ recombination hotpsots are MHC haplotype-dependent; as the MT-tetranucleotide repeat configuration is shared by all sequenced MHC haplotypes at both loci it is possible that distant haplotype-specific enhancer-like elements act on this

structure to facilitate recombination. The Eβ MT element is shortened at the 5' end, and both the Aβ and Eβ MT elements show patchy homology to the MT consensus sequence. The physical arrangement of the two sequence elements at these meiotic hotspots differs from the arrangement at *Ms6-hm*; however, the orientation of MT-1 and the CAG strand in the tandem array is similar. For minisatellite alleles greater than 1kb in length, the distance between the two repeats will then be the same as that at the MHC hotspots.

Finally, an MT element is found in the 3' sequences of a mouse proline-rich salivary protein (PRP) gene (M14, Fig.9.1). The proline-rich proteins of the mouse are encoded by a multigene family clustered on chromosome 6 (Azen *et al.*, 1989). Ann *et al.* (1988) cloned two tandemly arrayed PRP genes (MP2 and M14) on a contiguous block of 77kb, and showed by DNA sequence comparison that these genes were likely to have arisen by duplication of a common ancestral gene. Much of the PRP coding sequence is composed of tandem repeats of a 42bp GC-rich repeat unit (from which the entire ancestral PRP gene may have originated, Ann and Carlson, 1985). MP2 and MP14 differ in the number of 42bp tandem repeats in exon 2; in addition an L1Md element has integrated into the first intron of M14, close to two simple sequence repeat arrays. This is therefore a rapidly evolving region of the mouse genome. The start of similarity to the MT consensus sequence lies approximately 360bp 3' to the second polyadenylation singal noted by Ann *et al.* (1988); this element is truncted at both the 3' and 5' ends. However, it is not clear from the published sequence data whether there is also an MT element in the 3' sequences of the MP2 PRP gene.

The correlation between MT sequences and genomic instability is provocative, however it is unclear whether MT elements are a cause or a consequence of instability. Again the question arises, what is the definition of an MT element? What is the significance of their fissiparous nature? Does the association between MT elements and sequence instability explain the divergence between individual MT elements, or reflect the mechanism by which MT elements are mobile in the genome? It is conceivable that these elements may mobilise surrounding sequences (such as the Ins sequence) or promote local recombination events. The answers to these questions must await more detailed characterisation of individual members of this sequence family.

Evidence from analysis of the DNA sequences flanking human minisatellites suggests that there is also an association between tandem and dispersed repetitive elements in man (Armour *et al.*, 1989). In less than 6.5kb of DNA sequence flanking 6 minisatellite loci these authors found 6 dispersed repetitive elements (4 Alu elements, 1 L1 element, and a novel human SINE), suggesting that there is an excess of dispersed repeats near minisatellites. In addition, one of the human minisatellites, λ*MS32*, is flanked by retroviral LTR sequences; the alignment between the flanking DNA sequence

and the retroviral LTR extends to the boundary of the minisatellite and resumes beyond the repeat array. This suggests that λ*MS32* may have expanded from within a member of this LTR family, much as *Ms6-hm* has expanded from within MT-1. Using the polymerase chain reaction, the orthologous locus in old world primates has been shown to be monomorphic, containing 2.5 repeat units corresponding to the 3' end of the human minisatellite array; this presumably represents the organisation of the ancestral LTR from which λ*MS32* expanded (I.Gray and A.Jeffreys, manuscript in preparation).

There are other examples of tandem repeat arrays within dispersed repetitive elements; a $(AAAT)_8$ array is found within the B2 element adjacent to *Ms6-hm*. This repeat lies in a 3' (A)-rich region preceding the 3' tail which is known to be heterogeneous in length among B2 elements (Krayev *et al.*, 1982, Kramerov *et al.*, 1985). Using PCR it would be possible to determine whether this array is polymorphic between different mice at any one B2 element, and in particular at the B2 element linked to *Ms6-hm*. Similarly, a GATA (and possibly also a GAAAAA) array at *B10* may have expanded from within the 3' sequences of a diverged B1 repetitive element; the polymerase chain reaction may also reveal microheterogeneity at this locus. Brown and Piechaczyk (1982), studying sequence variation within the mouse LINE family, found one element which contained four 5bp tandem repeats of the sequence CAAAA; this particular element was highly diverged from other L1Md elements throughout its length. MT-1, in contrast, is relatively similar (75%) to the consensus MT sequence (by comparison with other MT elements). Within the third intron of a mouse transplantation antigen pseudogene in the MHC a B1 element was found which contained the hexanucleotide repeat GAGGCA amplified 17-fold (Steinmetz *et al.*, 1981, Rogers, 1985). The extent of variation at this locus and the L1Md CAAAA array is unknown.

In conjunction with the clustering of dispersed repetitive elements (see Introduction and Rogers, 1985) it therefore appears that certain regions of the mammalian genome are prone to the accumulation of dispersed and tandem repetitive elements. Furthermore, characterisation of several linked pairs of minisatellites in man, and the observation that HVR loci tend to be found towards the ends of human chromosomes, suggests that minisatellites may also cluster in the genome (Royle *et al.*, 1988, Armour *et al.*, 1989). It is likely that these clusters define 'junk' regions of mammalian chromosomes in which repeated DNA sequences may accumulate in the absence of selective constraints; the regional chromatin structure at these loci may somehow promote this accumulation. Alternatively, sequence elements within such clusters may be involved in functional interactions; it is provocative that three highly unstable loci (the Aβ hotspot, the Ins sequence, and *Ms6-hm*) all contain an MT element, a B2 element, and a simple sequence component. These three middle repetitive elements

may somehow interact to promote instability at these loci. However, the extent to which flanking sequences have actively contributed to the evolution of the minisatellite *Ms6-hm* remains unclear.

Long hypervariable GGGCA arrays have also evolved at other sites in the genome of certain mouse strains. These are known to be recombinationally separable from *Ms6-hm* and therefore are unlikely to have arisen through local duplication events. Instead such loci may have evolved through the independent expansion of GGGCA repeat units by slippage and unequal recombination events, although it is conceivable that these minisatellites arose through the mobility of retroposed copies of the expanded MT-1 element from *Ms6-hm*.

9.3. Localisation of *Ms6-hm*

Ms6-hm has been localised to mouse chromosome 4. *Ifa*, to which *Ms6-hm* is tightly linked, is syntenic with chromosome 9p in man (Nadeau *et al.*, 1986). Loci distal to, and inclusive of, *c-jun* on mouse chromosome 4, however, are syntenic with human chromosome 1p; the breakpoint between these syntenic groups therefore lies between *Ifa* and *c-jun* which are less than 1cM apart (Friedman *et al.*, 1989). In man, the sites of ancestral chromosomal fusions may be represented by some contemporary fragile sites, regions of breakage and recombination in human chromosomes (see Hastie and Allshire, 1989). An interstitial telomeric-like repeat array was mapped by *in situ* hybridisation to a site on human chromosome 2 to which two fragile sites map, and which is thought to be the site of fusion of two ancestral acrocentric chromosomes (Allshire *et al.*, 1988). If *Ms6-hm* lies distal to *Ifa* it is possible that this minisatellite may be associated with the synteny breakpoint which has been mapped to this region. Similarly, the highly variable interstitial loci detected by the TTAGGG repeat probe in BXD RI DNA may identify sites of chromosomal fusion during the evolution of the contemporary mouse karyotype. Characterisation of the sequences present at chromosomal breakpoints will facilitate an understanding of the evolution of mammalian chromosomes.

9.4 Mutation processes at *Ms6-hm*

The germline mutation rate to new length alleles at *Ms6-hm* has been estimated to be 2.5% per gamete. It is this extremely high mutation rate which accounts for the extensive variation and high levels of heterozygosity at *Ms6-hm* which are found even within inbred strains of mice. However most germline mutation events at this locus are small, and inbreeding has a strong homogenising effect on allele size, such that alleles within any one strain tend to lie within a particular size range.

In contrast, other minisatellite loci show substantial stability within inbred strains. Alleles at *Mm1*, for example, may be shared by different inbred strains; this locus has a 40bp repeat unit which is heterogeneous in sequence within the minisatellite array. Stephan (1989) predicted, by computer simulations, that tandem arrays which undergo crossing-over (and therefore unequal exchange events) most frequently will have the shortest repeat units. This is consistent with the observation that *Ms6-hm* and *Hm-2* (with 5 and 4bp repeat units respectively) are the most variable minisatellite loci known in mice, and that the most variable locus known in man, λ*MS1*, has a relatively short 9bp repeat unit (Jeffreys *et al.*, 1988a). The repeat unit sequence at *Ms6-hm* appears to be extremely homogeneous within the minisatellite array. Higher-order structures of variant repeat units (as observed for *Mm1*) only arise when recombination rates are low relative to the base substitution rate (Stephan, 1989). Thus the GGGCA array at *Ms6-hm* is likely to be the product of extremely high rates of recombination and slippage at this locus.

The GGGCA repeat unit of *Ms6-hm* is the central part of the core sequence common to many human minisatellites (Jeffreys *et al.*, 1985a). The most unstable minisatellites identified in the human and mouse genomes have repeat units related to this core sequence (*Ms6-hm*, *Hm-2*, Wong *et al.*, 1987). Furthermore, those loci in the mouse genome detected by a GGGCA repeat probe at low stringency (and in particular all long (>2kb) minisatellite arrays of the repeat sequence GGGCA (*Ms6-hm*, *SWR-LL*, *SJL-LL*, and *129-LL*)) are even more variable than loci related to the human DNA fingerprinting probes, resulting in extreme intrastrain variation. As discussed in the Introduction, core-related sequences are also associated with other recombinationally active DNA sequences, including the meiotic hotspots found in the mouse MHC. Therefore, by virtue of their repeat unit sequence, core-containing HVRs may be actively more variable than other internally repeated sequences.

It is possible that the core sequence acts as a signal for recombinases or other cellular proteins which might enhance the variability associated with core-related tandem repeat arrays. A binding activity has been detected in extracts from murine somatic tissues and canine testes which specifically interacts with tandem repeated polycore arrays, and has been demonstrated by protein-blotting to be a 40kD protein (MSBP1, Collick and Jeffreys, 1990). The structure and biological function of this protein remain to be elucidated, and it is unknown whether additional proteins interact with minisatellite sequences in the germline.

Whether other tandem repeated sequences are as variable as core-related minisatellites remains to be seen. Interstitial telomeric arrays, for example, are extremely variable, and although unrelated to the core sequence are presumably 50% G-rich. Core-related minisatellites may therefore be hypervariable simply by virtue of being G-rich.

Only by assessing the variability at tandem repeat arrays with a wide spectrum of repeat unit sequences will the true relationship between variability and repeat unit sequence emerge.

Variability at any one minisatellite locus is likely to result from a combination of factors. The large number of short monomorphic arrays of core-related sequences which are detected in human and mouse DNA fingerprints suggest that array length is likely to contribute to instability (although the bulk of these cross-hybridising loci may mask a minority of highly variable loci). The GGGCA array at *B10*, for example, indicates that not all GGGCA arrays are as hypervariable as *Ms6-hm*. There are very few large GGGCA arrays in the mouse genome; perhaps only those loci which are longer than a certain threshold length are hypervariable. As outlined in the Introduction long arrays may evolve faster through additional mechanisms unavailable to short arrays (such as unequal exchange). Consistent with this idea is the observation that C57BL/6J *Ms6-hm* alleles (which generally contain more than 1000 repeat units) are found over a more extensive size range than DBA/2J alleles (generally 500-1000 repeat units in length). No significant difference in germline mutation rate, however, is detectable between C57BL/6J and DBA/2J alleles; the wider range of allele sizes could equally well be explained by an association between large alleles and large length change events. At *Hm-2* the majority of new-length mutant alleles which have been scored are derived from C57BL/6J alleles (which are extremely large), even though the ability to resolve small length changes is much greater for short DBA/2J alleles.

In the Introduction arguments suggesting that hypervariable minisatellite loci are hotspots of recombination were reviewed. It is debatable whether minisatellites are active instigators of recombination in the way that chi sequences in *E.coli* initiate homologous recombination, or are merely passive products of dynamic processes, including slippage and unequal recombination, which occur in the genome (see Jarman and Wells, 1989, Dover, 1989). Initial claims that two human minisatellites were associated with hotspots of meiotic recombination have been demonstrated to be untrue after more extensive haplotype analysis showed that markers flanking the HVRs are in linkage disequilibrium (at both the human insulin locus, Cox *et al.*, 1989, and the α-globin locus, Higgs *et al.*, 1986).

Elucidation of the mechanisms by which minisatellites change in length should allow a better understand of their role in the genome. Wolff *et al.* (1989) have demonstrated that unequal meiotic recombination associated with the exchange of flanking markers is extremely unlikely to be the primary mechanism by which repeat copy number changes at the human minisatellite locus λ*MS1*. Similarly 3/3 length change events at *Ms6-hm* are not accompanied by the exchange of flanking markers.

Conversely, DNA sequence analysis of 4 recombinant haplotypes at the Eβ hotspot (Kobori *et al.*, 1986), and 5 at the Aβ hotspot (Uematsu *et al.*, 1986), has shown that meiotic recombination events at the mouse MHC do not alter the number of tetrameric repeats at these loci.

These results are consistent with a variety of alternative models for the generation of new-length alleles at minisatellite loci, including several recombination based models - meiotic or mitotic gene conversion, and meiotic or mitotic sister chromatid exchange. Intramolecular 'loop-out' events may be ruled out as a primary mechanism, as these would only allow alleles to shorten in length. Replication based models are also possible, specifically slipped strand mispairing, although large length change events, and the distribution of variant repeat units across minisatellite arrays argue against this being the only mechanism. Other less orthodox models might involve the activity of a telomerase-type enzyme which could add or remove minisatellite repeat units at a nick within a tandem array, or transposition type micro-insertion/deletion events; however these models do not explain either the homogeneity of repeat unit sequence within minisatellite arrays, or the diversity of minisatellite sequences found in the genome.

The advent of the polymerase chain reaction has enabled the structural analysis of minisatellite alleles amplified directly from genomic DNA. The distribution of variant repeat units across a minisatellite array can now be mapped by end-labelling and partially digesting amplified minisatellite alleles with restriction endonucleases which cut either every repeat unit or every variant repeat unit, followed by electrophoresis and autoradiography (Jeffreys *et al.*, 1990). The human minisatellite locus λ*MS32* has been analysed by internal mapping, allowing any allele at this locus to be assigned a specific binary code. The GGGCA array of *Ms6-hm* lacks suitable restriction enzyme sites for this analysis.

Internal mapping has been used to analyse the population distribution of allele-types at the human minisatellite λ*MS32*, and, through amplification of single molecules, to compare the distribution of variant repeat units across mutant λ*MS32* alleles with the structure of parental alleles. Mutant λ*MS32* alleles amplified from single sperm DNA molecules are entirely derived from one or other parental allele (Jeffreys *et al.*, 1990). This result clearly shows that interallelic unequal exchange between homologues at meiosis or mitosis is rare, and favours mitotic models, such as sister chromatid exchange, for the primary mechanism of length change at λ*MS32*. Furthermore, there is a significant level of germline mosaicism for specific non-parental alleles in sperm DNA, which must result from pre-meiotic mutation events at this locus.

Distinct haplotypes of variant repeat unit distribution are observed at λ*MS32* alleles within human populations; the comparison of allele structure between different

populations will allow direct insights into the evolution of alleles at this locus. In relation to the evolution of minisatellites, it is interesting to note that a minisatellite with a 14bp repeat unit is found in the first intron of the embryonic α-globin gene of man (Proudfoot *et al.*, 1982), and, at the same position, in the orthologous gene of the goat (Wernke and Lingrel, 1986). The goat minisatellite also has a 14bp repeat unit, but a different (yet closely related) repeat unit sequence, suggesting that the goat and human arrays have become fixed for different minisatellite repeat units over evolution.

Mutation events at three human minisatellite loci occur at a similar rate in the paternal and maternal germlines, consistent with length changes arising primarily at a specific stage of gametogenesis (Jeffreys *et al.*, 1988a, and Introduction). In contrast, germline mutation events at *Ms6-hm* appear significantly to involve the paternally derived allele ($p>0.99$). This paternal bias may simply reflect the large number of cell divisions during spermatogenesis (compared with oogenesis), suggesting that mitotic events are involved in the generation of new-length mutant alleles at *Ms6-hm*.

9.5 Somatic mutation at *Ms6-hm*

Allelic length change events at human minisatellites are not restricted to the germline, but can also arise in other tissues. Such somatic mutant alleles have been detected in clonally derived tumour cells, and by PCR amplification of single molecules from blood DNA (Armour *et al.*, 1989, Jeffreys *et al.*, 1990). In the analysis of non-parental alleles from clonal tumour cells, Armour *et al.* (1989) found that most length changes, as in the germline, were of only a few repeat units, although large length changes were also observed. No reciprocal products were seen in 15 mutant tumours suggesting that unequal mitotic recombination between homologues was unlikely to be the primary mechanism of length change (one-quarter of unequal mitotic recombinations would be expected to partition both mutant products into the same daughter cell, and thus one third of mutant tumours should show a reciprocal product). In addition, mutant λ*MS32* alleles amplified from blood DNA appear, like mutant alleles in sperm DNA, to be entirely derived from one or other parental allele (Jeffreys *et al.*, 1990).

The incidence of mutations per allele at λ*MS1* in gastrointestinal tumours was comparable to the incidence of germline mutations at this locus, despite the large number of post-zygotic mitoses in gut epithelial lineages (>10,000 compared with 400 in the male and 24 in the female germline), suggesting that the mutation rate per mitosis at this locus is extremely low ($<10^{-5}$ per allele per cell division). Non-parental λ*MS32* alleles amplified from blood DNA are found at a correspondingly low frequency, and there is a very low level of mosaicism for somatic mutant alleles (Jeffreys *et al.*, 1990). In

polyclonal tissues, therefore, mutant alleles will be heterogeneous in size and will not be detectable on analysis of bulk tissue DNA. No evidence for significant mutational mosaicism at any of 6 human minisatellite loci has been observed by Southern blot hybridisation of bulk genomic DNA, despite screening more than 1000 individuals. Two non-parental alleles were observed at λ*MS32* in DNA prepared from individual lymphoblastoid cell lines, however these may be oligoclonal cell lines, mosaic for pre-existing low-level mutant alleles in blood cells; alternatively length-change events may have arisen during passage *in vitro* (Armour *et al.*, 1989).

In contrast to these observations, a significant proportion (3%) of mice show evidence of an early somatic mutation event at *Ms6-hm*, resulting in mosaicism for cells containing either the parental or a new-length allele in somatic and germline lineages. Somatic mutation events at *Ms6-hm* do not have an equal likelihood of occuring at every mitotic cell division; this would result in the detection of many less intense somatic mutant alleles on Southern blot hybridisation. In fact the proportion of cells containing the mutant allele varies between 8 and 60%, even though 5% 'mosaicism' could be detected in allele mixing experiments; thus mutation events at *Ms6-hm* which occur later in somatic lineages must, as at human minisatellite loci, be rare. Early somatic mutation events at *Ms6-hm*, in contrast, appear to be confined to a narrow developmental window during embryogenesis. Mice mosaic for the same non-parental allele in the soma and germline clearly demonstrate that the mutation events at *Ms6-hm* precede the separation of these lineages.

The earliest developmental decisions in mouse embryogenesis involve the allocation of cells to the trophectoderm and the inner cell mass (ICM), by 3 days postcoitum (pc). The trophectoderm differentiates into foetal placenta, while the ICM differentiates into primitive ectoderm and primitive endoderm by the late blastocyst stage (4.5 days pc, see Rossant, 1984). The primitive ectoderm gives rise to the somatic and germ cell lineages of the embryo. The embryo founder cells proliferate to form the egg cylinder, which initially consists of 500-600 cells which divide rapidly during primitive streak development (7 days pc), differentiating into the three definitive germ layers, the embryonic ectoderm, endoderm and mesoderm, and increasing at gastrulation to 12,000-15,000 cells (Monk and Harper, 1979, Snow, 1977). At this stage the somites appear and the body plan of the foetus is established (see Hogan *et al.*, 1986).

The lineage relationships of cells in early mouse development have been analysed using chimaeric embryos constructed by aggregating genetically distinct cleavage stage embryos (Mintz, 1974), or by injection of marker embryonic cells into recipient blastocysts (Gardner, 1968). Such studies, however, are likely to distort normal developmental interactions. Genotypic selection, or chimaeric drift, may result in the

selection of one parental genotype over the other during embryogenesis (Warner *et al.*, 1977). Although the abnormally large mosaic blastocysts are subject to size regulation preceding gastrulation, the extent to which normal cell-cell interactions are distorted is unknown (Lewis and Rossant, 1982). Mosaicism generated by female X-inactivation has also been used in lineage analysis, however this only gives information on developmental decisions subsequent to the time of X-inactivation, which is thought to occur between 4.5 and 5.5 days pc (McMahon *et al.*, 1983, Gardner *et al.*, 1985). These different approaches have shown that mammalian development involves the irreversible and heritable restriction of groups of cells to different lineages, but that between lineage allocation and the restriction of cell potential positional cues are very important, allowing flexibility of response to changing conditions (see Rossant, 1984).

Somatic mutation events at *Ms6-hm* generate genetically mosaic embryos without disturbing normal development. Allelic length change events at this locus within an early zygotic cell result at cell division in one daughter cell carrying two parental alleles, and the other one parental and one new-length mutant allele. The segregation of the mitotic progeny of the daughter blastomere carrying the non-parental allele is reflected in the mosaic distribution of the mutant allele in different tissues of the adult mouse. Thus somatic mutant alleles at *Ms6-hm* provide highly informative and innocuous cell markers with which to study the allocation of cells to distinct lineages during mouse development.

Dissection of four mosaic mice has shown that the dosage of non-parental alleles at *Ms6-hm* is indistinguishable in different adult tissues. There are several important conclusions about lineage relationships in the soma which follow from this result. Firstly, the somatic mutation events at *Ms6-hm* in these mice preceded the allocation of primitive ectoderm cells to specific somatic lineages. Secondly, there must be extensive cell mixing, and also a large number of primitive ectoderm cells, at the time of tissue allocation, such that an equivalent fraction of mosaic cells from the same undifferentiated pool are partitioned into each lineage. Thirdly, subsequent to the allocation of somatic lineages every founder cell must divide and undergo cell death at an approximately equal rate such that each adult tissue contains an indistinguishable fraction of mosaic cells. These conclusions may be oversimplified as the adult tissues analysed consist of more than one cell type. However, if each somatic lineage was derived from only a small number of cells, of which perhaps only 20% were mosaic, one could envisage that cells carrying mutant alleles would not contribute to particular lineages, or would contribute disproportionally. It seems likely, therefore, that there is a large pool of intermingling progenitor cells at the time of allocation of somatic lineages.

Similar conclusions come from the analysis of mosaic transgenic mice generated by the late integration of exogenous plasmid DNA or recombinant retroviral DNA into pre-

implantaion embryos (Wilkie *et al.*, 1986, Soriano and Jaenisch, 1986). These integrated sequences, like somatic mutant alleles at *Ms6-hm*, provide presumably neutral DNA markers to study the developmental history of cells in early mouse development. Both studies reported a uniform distribution of transgenic cells in all somatic tissues of mosaic mice, with occasional incomplete cell mingling (Soriano and Jaenisch, 1986). Soriano and Jaenisch (1986) estimate the time of genetic marking (foreign DNA integration) to be between the 4th and 8th zygotic cell division while Wilkie *et al.* (1986) estimate that integration occurs shortly after the first round of DNA replication. Early somatic mutation events at *Ms6-hm* must occur between the first zygotic cleavage and the allocation of somatic linaeges (presumably before day 6 pc).

Comparison of DNA prepared from placental and embryonic tissues revealed that mosaic transgenic cells were often found in one or other tissue, although the same mosaic trangenic cells were rarely found in both (Wilkie *et al.*, 1986, Soriano and Jaenisch, 1986). This is consistent either with integration occuring subsequent to blastocyst differentiation, or with an asymmetric allocation of transgenic cells into trophectoderm and ICM lineages. Similar comparative experiments would be difficult at *Ms6-hm* because of the maternal contribution to placental tissue, although this problem can be avoided by dissection of early post-implantation embryos. The low frequency of somatic mosaic mice at *Ms6-hm* suggests that a large number of embryos would need to be dissected to obtain significant results. However if somatic mutation events at *Ms6-hm* occur in the first few zygotic cell divisions, and since only a fraction ($^1/_4$-$^1/_8$th, Gardner, 1978) of blastomeres contribute to the embryonic lineages, the incidence of mosaicism within the whole blastocyst is likely to be higher than the incidence in the embryo. Now that a locus-specific probe for a second somatically unstable minisatellite locus (*Hm-2*) is available this analysis is more feasible (although ideally several other such loci would be scored), and should provide information on the timing of somatic mutation events at these loci relative to the earliest developmental decisions of embryogenesis.

The time at which the germline is allocated during early mouse development is unclear (see McLaren, 1984). Using X-inactivation mosaics, McMahon *et al.* (1983) showed that germ cells and somatic lineages were derived from a common precursor pool after X-inactivation, and suggested that up to the time of gastrulation the potential to develop as either somatic or germ cell depended on the position a cell occupied within the egg cylinder, and therefore that there is a large germline precursor pool (>100 cells). In contrast, studies of germline mosaicism in chimaeric embryos suggest that the pool of germline progenitor cells contains between 2 and 9 cells (Mintz, 1974). Gardner *et al.* (1985) cloned cells from pre-implantation embryos by injection into genetically distinct blastocysts, and demonstrated that at least 2 primitive ectoderm cells from 5th day

blastocysts (out of approximately 23 primitive ectoderm cells, Gardner, 1985) could contribute to the germline. These authors concluded that possibly all early primitive ectoderm cells may be able to contribute to the germline, although germline allocation within a normal embryo was not addressed.

Mice mosaic for a common non-parental allele at *Ms6-hm* in the germline and soma have been found. For three of these mice the level of mosaicism in the germline was similar to that in the soma. However in one mouse the dosage of the non-parental allele in the germline was significantly reduced compared to somatic tissue. The dosage of the mutant allele in this mouse was found to be indistinguishable in different adult somatic tissues; the germ cells of this mouse must therefore be allocated differently from the somatic lineages. In this case the germline must be allocated either before the somatic lineages, or from a small subset of the pool of primitive ectoderm cells giving rise to the soma. Similarly Wilkie *et al.* (1986) and Soriano and Jaenisch (1986) found that the fraction of transgenic cells was not always equivalent in somatic and germ cell lineages of mosaic mice. Both groups identified mosaic mice which transmitted the transgene to less than the expected number of offspring, and Soriano and Jaenisch (1986) found two mice in which a substantial part of the germline was composed of cells which did not detectably contribute to the somatic lineages. These authors argued that this was extremely unlikely to be due to retroviral infection of prospective germ cells, and concluded that cells destined to give rise to the germline were set aside early, and may be restricted from freely mingling with cells which give rise to somatic lineages.

Soriano and Jaenisch (1986) argued that the minimum contribution of one genotype ever observed in the somatic lineages of a mosaic mouse represents the progeny of one marked progenitor cell. These authors found no evidence for mosaic proviruses with dosages lower than 0.12, and concluded that $^1/_{0.12}$, or 8, cells may be the number of progenitor cells for the adult mouse. However, as Rossant (1986) points out, this is more strictly the number of possible progenitor cells present at the time of genetic marking. The estimated progenitor number for the somatic lineages is larger when calculated from X-inactivation mosaics, where mosaicism is established later (not less than 20-30 progenitor cells, Nesbitt, 1971). It would therefore be unwise to attempt to predict the number of progenitor cells for the somatic lineages from the range of mosaicism found in mosaics at *Ms6-hm* until the time at which somatic mutation events occur at this locus is known more accurately.

Early somatic length-change events at mouse minisatellite loci are not confined to *Ms6-hm*, but have been documented for one other locus *Hm-2*. One mouse, C57BL/6J x DBA/2J F_1 52, was found to be mosaic at both *Ms6-hm* and *Hm-2*. The segregation pattern of the non-parental alleles in the offspring of this mouse revealed that both

mutation events occurred in the same mitotic lineage. Furthermore, from the ratio of allelic combinations at these two loci among the progeny of this mouse, it appears that the mutation events occurred in closely following cell divisions. No offspring were scored which would determine the order in which the two events occurred, and therefore it is possible that both events occurred at the same cell division (the non-parental alleles at each locus segregating to the same daughter cell). This result suggests that the developmental window within which *Ms6-hm* and *Hm-2* are somatically unstable may be extremely narrow.

It is important to note that some mice scored as 'germline' mutants at *Ms6-hm* might instead have arisen from early somatic mosaic blastocysts, in which only cells containing the non-parental allele have contributed to the primitive ectoderm. Furthermore, somatic mutant alleles which are partitioned to the germline precursor cells, and are undetectable in the somatic tissues, would result in germline mosaicism in the adult mouse, the offspring of which would then be scored as 'germline' mutants. The distinction between 'somatic' and 'germline' mutation events at this locus is therefore ill-defined. It is unlikely, however, that all 'germline' mutation events at *Ms6-hm* occur in early development. An approximately equal rate of germline and somatic mutation is observed in adult mice. As there is no trace of parental alleles in 'germline' mutant mice, nor any cases of common new-length alleles among sibs, the rate of germline mutation appears to be too high to be purely the effect of random partitioning of mosaic cells within the primitive ectoderm. In addition, the paternal bias observed for germline mutation events at *Ms6-hm* is not observed for all somatic mutation events at this locus, although 5/6 events do involve the paternally inherited allele ($p=0.22$). At *Hm-2* neither germline nor somatic mutation events show a bias for the involvement of the paternal allele.

Germline and early developmental allelic length changes at *Ms6-hm* are therefore likely to be distinct events. While meiotic models for germline events cannot be ruled out, the early events are clearly mitotic, and, by the same arguments used for somatic mutant alleles in human tumours, are unlikely to arise through unequal recombination between homologues. Furthermore, the non-parental 'somatic' mutant allele in C57BL/6J x DBA/2J F_1 mouse 37 was non-recombinant for C57BL/6J markers either side of the minisatellite (data not shown). Thus unequal sister chromatid exchange, mitotic or sister chromatid gene conversion, replication slippage, and possibly intramolecular loop-out events (for the large length changes, which all involve deletions) remain the contending mechanisms for somatic mutation events at *Ms6-hm*. The only minisatellite loci at which early somatic mutation events are known to occur have repeat units of GGGCA (*Ms6-hm*, and possibly *129-LL* and *SWR-LL*) or GGCA (*Hm-2*). It is

conceivable that during very early development there is transient expression of an MSBP1-type protein, which is perhaps also expressed in the germline, and which recognises the GGCA motif and enhances minisatellite instability in a sequence dependent fashion.

Alternatively *Ms6-hm* might be somatically unstable in early development as a consequence of sequence elements in the surrounding DNA. Activation of the zygotic genome (at the two cell stage) may be associated with a period of genomic instability. Transcription of many dispersed repetitive elements (including B2 elements) is known to occur in early development (see Introduction), and this is likely to be when retroposition, including the generation of pseudogenes, occurs. Retroviral-like elements are also actively expressed at this time (see Introduction). This genomic activity must precede the separation of the soma and germline in order to ensure the heritability of novel integrants; perhaps mosaicism for novel integrants reduces their genetic load which might otherwise be deleterious, while allowing high levels of retroposon associated variation. This period of genomic activity coincides with the time at which somatic mutation events occur at *Ms6-hm*. It is unknown whether MT elements are specifically expressed in early development; transcriptional activity at *Ms6-hm* might be associated with minisatellite instability. The abundance of dispersed repeats around *Ms6-hm* suggests at least that this locus is exposed to nomadic elements at the time they are mobile in the genome. It is also unknown whether MT elements, or other dispersed repeat sequences, are associated with *Hm-2* (although a $(CA)_n$ array is linked to this minisatellite).

It will be extremely interesting to find out how representative *Ms6-hm* and *Hm-2* are of all minisatellite loci in the mouse genome. By cloning other individual loci, and experimenting with synthetic tandem repeated hybridisation probes it should be possible to define the extent of somatic instability in the genome, and identify the common features of somatically unstable loci. Finally, it will be important to find out whether early somatic events occur at any human minisatellite loci. The evidence to date suggests that such events do not occur in man; this could be due to different patterns of genome activity in man and mouse during early embryogenesis. Alternatively, the finding that $(GGGCA)_n$ is a poor DNA fingerprinting probe in man suggests that large GGGCA-related minisatellites, and perhaps also the early somatic instability associated with these loci, may be specific to the mouse genome.

9.6 Conclusions

The mammalian genome contains a complexity of tandem repeat loci with respect to repeat unit sequence and variability. The repeat unit GGGCA is found at one of the most variable loci yet identified in the mouse genome, *Ms6-hm*. This repeat sequence is likely to contribute significantly to the high mutation rate at this locus, as are three closely linked members of a dispersed repeat family associated with other unstable sequences in the mouse genome.

Mitotic processes are associated with mutation events at *Ms6-hm*; new-length alleles at this locus arise in the male germ line and in the early embryo. Early mutation events at this locus provide insights into the lineage relationships of cells during mouse embryogenesis, revealing that cells of the primitive ectoderm mix extensively prior to the allocation of somatic lineages, while germ cells are allocated differently.

How representative is *Ms6-hm* of the diverse spectrum of mouse minisatellite loci? At least one other locus with a related G-rich repeat sequence exhibiting a high level of germline and early somatic instability has been identified in the mouse genome. Further analysis of these two loci, in parallel with studies on human minisatellites, will provide a clearer understanding of the mechanisms by which tandem repeat arrays evolve, and define the role which G-rich minisatellites play in maintaining hypervariability in the mammalian genome.

Mr. Gall Junior brought his work to a close. He reassembled the piano case, put back his tools in their bag, and stood up.

The mice have returned, he said.

Samuel Beckett
Watt

BIBLIOGRAPHY

Ali, S., Muller, C.R., and Epplen, J.T. 1986. DNA fingerprinting by oligonucleotide probes specific for simple repeats. *Hum. Genet.* **74**:239-243.

Allard, D., Delbecchi, L., Bourgaux-Ramoisy, D., and Bourgaux, P. 1988. Major rearrangement of cellular DNA in the vicinity of integrated polyomavirus DNA. *Virology* **162**:128-136.

Allen, N.D., Barton, S.C., Surani, M.A.H., and Wolf, R. 1987. Production of transgenic mice. *In* "Mammalian development, a practical approach", IRL Press, Oxford, England.

Allshire, R.C., Gosden, J.R., Cross, S.H., Cranston, G., Rout, D., Sugawara, N., Szostak, J.W., Fantes, P.A., and Hastie, N.D. 1988. Telomeric repeat from *T. thermophila* cross-hybridises with human telomeres. *Nature* **332**:656-659.

Allshire, R.C., Dempster, M., and Hastie, N.D. 1989. Human telomeres contain at least three types of G-rich repeat distributed non-randomly. *Nucleic Acids Res.* **17**:4611-4627.

Ann, D.K., and Carlson, D.M. 1985. The structure and organisation of a proline-rich protein gene of a mouse multigene family. *J. Biol. Chem.* **260**:15863-15872.

Ann, D.K., Smith, M.K., and Carlson, D.M. 1988. Molecular evolution of the mouse proline-rich protein. *J. Biol. Chem.* **263**:10887-10893.

Antonarakis, S.E., and Kazazian, H.H. 1988. The molecular basis of haemophilia A in man. *Trends Genet.* **4**:233-237.

Anzai, K., Kobayashi, S., Kitamura, N., Kanai, Y., Nakajuma, H., Suehiro, Y., and Goto, S. 1986. Isolation and sequencing of a genomic clone for a mouse brain specific small RNA. *J. Neurochemistry* **47**:673-677.

Armour, J.A.L., Patel, I., Thein, S.L., Fey, M., and Jeffreys, A.J. 1989. Analysis of somatic mutations at human minisatellite loci in tumours and cell lines. *Genomics* **4**:328-334.

Armour, J.A.L., Wong, Z., Wilson, V., Royle, N.J., and Jeffreys, A.J. 1989. Sequences flanking the repeat arrays of human minisatellites : association with tandem and dispersed repeat elements. *Nucleic Acids Res.* **17**:4925-4935.

Arnheim, N., and Kuehn, M. 1979. The genetic behaviour of a cloned mouse ribosomal DNA segment mimics mouse ribosomal gene evolution. *J. Mol. Biol.* **134**:743-765.

Azen, E.A., Davisson, M.T, Cherry, M., and Taylor, B.A. 1989. *Prp* (proline-rich protein) genes linked to markers *Es-12* (esterase-12), *Ea-10* (erythrocyte alloantigen), and loci on distal mouse chromosome 12. *Genomics* **5**:415-422.

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

Balazs, I., Baird, M.L., Wexler, K., and Wyman, A.R. 1986. Characterisation of the polymorphic DNA fragments detected with a new probe derived from the D14S1 locus. *Am. J. Hum. Genet.* **39**:A229.

Barsh, G.S., Seeburg, P.H., and Gelinas, E. 1983. The human growth hormone gene family : structure and evolution of the chromosomal locus. *Nucleic Acids Res.* **11**:3939-3988.

Bastien, L., and Bourgaux, P. 1987. The MT family of mouse DNA is made of short interspersed repeated elements. *Gene* **57**:81-88.

Bedington, R. 1987. Isolation, culture and manipulation of post-implantation mouse embryos. *In* "Mammalian development, a practical approach", IRL Press, Oxford, England.

Bell, G.I., Selby, M.J., and Rutter, W.J. 1982. The highly polymorphic region near the human insulin gene is composed of simple tandemly repeated sequences. *Nature* **295**:31-35.

Bennet, K.L., Hill, R.E., Pietras, D.F., Woodworth-Gutai, M., Kane-Haas, C., Houston, J.M., Heath, J.K., and Hastie, N.D. 1984. Most highly repeated dispersed DNA families in the mouse genome. *Mol. Cell Biol.* **4**:1561-1571.

Benton, W.D., and Davis, R.W. 1977. Screening λgt recombinant clones by hybridisation to single plaques *in situ*. *Science* **196**:180-182.

Benzer, S. 1961. On the topography of the genetic fine structure. *Proc. Natl. Acad. Sci. USA* **47**:403-415.

Bernstine, E.G., Hooper, M.L., Grandchamp, S., and Ephrussi, B. 1973. Alkaline phosphatase activity in mouse teratoma. *Proc. Natl. Acad. Sci. USA* **70**:3899-3903.

Bickmore, W.A., and Sumner, A.T. 1989. Mammalian chromosome banding - an expression of genome organisation. *Trends Genet.* **5**:144-148.

Bickmore, W.A., Porteous, D.J., Christie, S., Seawright, A., Fletcher, J.M., Maule, J.C., Couillin, P., Junien, C., Hastie, N.D., and van Heyningen, V. 1989. CpG islands surround a DNA segment located between translocation breakpoints associated with genitourinary dysplasia and aniridia. *Genomics* **5**:685-693.

Bird, A.P. 1987. CpG islands as gene markers in the vertebrate nucleus. *Trends Genet.* **3**:342-347.

Bishop, C.E., Boursot, P., Baron, B., Bonhomme, F., and Hatat, D. 1985. Most classical *Mus musculus domesticus* laboratory mouse strains carry a *Mus musculus musculus* Y chromosome. *Nature* **315**:70-72.

Blattner, F.R., Williams, B.G., Blechl, A.E., Denniston-Thompson, K., Faber, H.E., Furlong, L., Grunwald, D.J., Kiefer, D.O., Moore, D.D., Schumm, J.W., Sheldon, E.L., and Smithies, O. 1977. Charon phages : safer derivatives of bacteriophage lambda for DNA cloning. *Science* **196**:161-169.

Botstein, D., White, R., Skolnick, M., and Davis, R.W. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphism. *Am. J. Hum. Genet.* **32**:314-331.

Bourgaux, P., Sylla, B.S., and Chatrand, P. 1982. Excision of polyoma virus DNA from that of a transformed mouse cell : identification of a hybrid molecule with direct and inverted repeat sequences at the viral and cellular junctions. *Virology* **122**:84-97.

Bourgaux-Ramoisy, D., Gendron, D., Chartrand, P., and Bourgaux, P. 1986. An excision event that may depend on patchy homology for site specificity. *Mol. Cell. Biol.* **6**:2727-2730.

Britten, R.J., Baron, W.R., Stout, D.B., and Davidson, E.H. 1988. Sources and evolution of human Alu repeated elements. *Proc. Natl. Acad. Sci. USA* **85**:4770-4774.

Britten, R.J., and Davidson, E.H. 1969. Gene regulation in higher cells : a theory. *Science* **165**:349-358.

Britten, R.J., Stout, D.B., and Davidson, E.H. 1989. The current source of human Alu retroposons is a conserved gene shared with old world monkeys. *Proc. Natl. Acad. Sci. USA* **86**:3718-3722.

Brown, S.D.M., and Piechaczyk, M. 1983. Insertion sequences and tandem repetitions as sources of variation in a dispersed repeat family. *J. Mol. Biol.* **165**:249-256.

Brown, W.R.A. 1989. Molecular cloning of human telomeres in yeast. *Nature* **338**:774-776.

Brulet, P., Kaghad,M., Xu, Y-S., Croissant, O., and Jacob, F. 1983. Early differential tissue expression of transposon-like repetitive DNA sequences of the mouse. *Proc. Natl. Acad. Sci. USA* **80**:5641-5645.

Brutlag, D., Fry, K., Nelson, T., and Hung, P. 1977. Synthesis of hybrid bacterial plasmids containing highly repeated satellite DNA. *Cell* **10**:509-519.

Bulawela, L., Forster, A., Boehm, T., and Rabbits, T.H. 1989. A rapid procedure for colony screening using nylon filters. *Nucleic Acids Res.* **17**:7630.

Burgoyne, P. 1986. Mammalian X and Y cross-over. *Nature* **319**:258-259.

Burke, T. 1989. DNA fingerprinting and other methods for the study of mating success. *Trends Ecol. Evol.* **4**:139-144.

Burke, T., and Bruford, M.W. 1987. DNA fingerprinting in birds. *Nature* **327**:149-152.

Burt, D.W., Reith, A.D., and Brammar, W.J. 1984. A retroviral provirus closely associated with the *Ren-2* gene of DBA/2J mice. *Nucleic Acids Res.* **12**:8579-8593.

Capon, D., Chen, E., Levinson, A.D., Seeburg, P.H., and Goedell, D.V. 1983. Complete nucleotide sequence of the T24 human bladder carcinoma oncogene and its normal homologue. *Nature* **302**:33-37.

Cavenee, W.K., Dryja, T.P., Phillips, R.A., Benedict, W.F., Godbout, R., Gallie, B.L., Murphree, A.L., Strong, L.C., and White, R.L. 1983. Expression of recessive alleles by chromosomal mechanisms in retinoblastoma. *Nature* **305**:779-784.

Chakravarti, A., Buetow, K.H., Antonarakis, S.E., Waber, P.G., Boehm, C.D., and Kazazian, H.H. 1984. Non-uniform recombination within the human β-globin gene cluster. *Am. J. Hum. Genet.* **36**:1239-1258.

Chandley, A.C., and Mitchell, A.R. 1988. Hypervariable minisatellite regions are sites for crossing-over at meiosis in man. *Cytog. Cell. Genet.* **48**:152-155.

Chartrand, P., Gusew-Chartrand, N., and Bourgaux, P. 1981. Integrated polyoma genomes in inducible permissive transformed cells. *J. Virol.* **39**:185-195.

Church, G.M., and Gilbert, W. 1984. Genomic sequencing. *Proc. Natl. Acad. Sci. USA* **81**:1991-1995.

Coleman, D.L, and Hummel, K.P. 1967. Studies on the mutation *diabetes* in the mouse. *Diabetologia* **3**:238-248.

Coleman, D.L., and Hummel, K.P. 1969. Effects of parabiosis of normal with genetically diabetic mice. *Am. J. Physiol.* **217**:1298-1304.

Collick, A., and Jeffreys, A.J. 1990. Detection of a novel minisatellite-specific DNA binding protein. *Nucleic Acids Res.* **18**:625-629.

Cooke, H.J., Brown, W.R.A., and Rappold, G.A. 1985. Hypervariable telomeric sequences from the human sex chromosomes are pseudoautosomal. *Nature* **317**:687-692.

Copeland, N.G., Hutchinson, K.W., and Jenkins, N. A. 1983. Excision of the DBA ecotropic provirus in Dilute coat colour revertants of mice occurs by homologous recombination involving the viral LTRs. *Cell* **33**:379-387.

Correns, C. 1900. G. Mendels Regel uber das Verhalten der Nachkommenschaft der Rassenbartarde. *Ber. dt. bot. Ges.* **18**:158-168.

Cox, N., Bell, G., and Xiang, X-S. 1989. Linkage disequilibrium in the human insulin/insulin-like growth factor II region of human chromosome 11. *Am. J. Hum. Genet.* **43**:495-501.

Cross, S.H., Allshire, R.C., McKay, S.J., McGill, N.I., and Cooke, H.J. 1989. Cloning of human telomeres by complementation in yeast. *Nature* **338**:774-776.

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

Darwin, C.R. 1859. The origin of species by means of natural selection. John Murray, London, England.

Davidson, E.H. 1969. Gene activity in early development. Academic Press, London, England.

Davies, K.E., Pearson, P.L., Harper, P.S., Murray, J.M., O'Brien, T., Sarfarazi, M., and Williamson, R. 1983. Linkage analysis of two cloned sequences flanking the Duchenne muscular dystrophy locus on the short arm of the human X chromosome. *Nucleic Acids Res.* **11**:2303-2312.

Davisson, M.T., and Roderick, T.H. 1989. Linkage Map. *In* "Genetic variants and strains of the laboratory mouse", Second edition, Lyon, M.F., and Searle, A.G., Eds., Oxford University Press, Oxford, England.

Deininger, P.L., and Daniels, G.R. 1986. The recent evolution of mammalian repetitive elements. *Trends Genet.* **2**:76-80.

De Maeyer, E., and Dandoy, F. 1987. Linkage analysis of the murine interferon alpha locus (*Ifa*) on chromosome 4. *J. Heredity* **78**:143-146.

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

D'Eustachio, P., and Ruddle, F.H. 1983. Somatic cell genetics and gene families. *Science* **220**:919-924.

D'Eustachio, P. 1984. A genetic map of mouse chromosome 12 composed of polymorphic DNA fragments. *J. Exp. Med.* **160**:827-838.

Devereux, J., Haeberli, P., and Smithies, O. 1984. A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.* **12**:387-395.

De Vries, H. 1900. Das spaltungsgesetz der Bastarde. *Ber. dt. Bot. Ges.* **18**:83-90.

Donis-Keller, H., Green, P., Helms, C., Cartinhour, S., Weiffenbach, B., Stephens, K., Keith, T.P., Bowden, D.W., Smith, D.R., Lander, E.S., Botstein, D., Akots, G., Rediker, K.S., Gravius, T., Brown, V.A., Rising, M.B., Parker, C., Powers, J.A., Watt, D.E., Kaufman, E.K., Briker, A., Phipps, P., Muller-Kahle, H., Fulton, T.R., Ng, S., Schumm, J.W., Braman, J.C., Knowlton, R.G., Barker, D.F., Crooks, S.M., Lincoln, S.E., Daly, M.J., and Abrahamson, J. 1987. A genetic linkage map of the human genome. *Cell* **51**:319-337.

Doolittle, W.F., and Sapienza, C. 1980. Selfish genes, the phenotype paradigm and genome evolution. *Nature* **284**:601-603.

Dover, G.A. 1980. Ignorant DNA. *Nature* **285**:618-620.

Dover, G.A. 1989. DNA fingerprints : victims or perpetrators? *Nature* **342**:347-348.

Dryja, T.P., McGee, T.L., Reichel, E., Hahn, L.B., Cowley, G.S., Yandell, D.W., Sandberg, M.A., and Berson, E.L. 1990. A point mutation of the rhodopsin gene in one form of retinitis pigmentosa. *Nature* **343**:364-366.

Eckert, R., and Green, H. 1986. Structure and evolution of the human involucrin gene. *Cell* **46**:583-589.

Edelmann, W., Kroger, B., Goller, M., and Horak, I. 1989. A recombinational hotspot in the LTR of a mouse retrotransposon identified in an *in vitro* system. *Cell* **57**:937-946.

Elliott, R.W. 1986. A mouse 'minisatellite'. *Mouse News Lett.* **74**:115-117.

Epplen, J.T. 1988. On simple repeated $GA^{T}_{C}A$ sequences in animal genomes : a critical reappraisal. *J. Hered.* **79**:409-417.

Evans, E.P. 1987. Karyotyping and sexing of gametes, embryos, and fetuses and *in situ* hybridisation to chromosomes. *In* "Mammalian development, a practical approach", IRL Press, Oxford, England.

Falconer, D.S. 1960. Introduction to quantitative genetics. Longman, Essex, England.

Farabaugh, P., Schmeissner, U., Hofer, M., and Miller, J. 1978. Genetic studies of the *lac* repressor. VII : on the molecular nature of spontaneous hotspots in the *lac*I gene. *J. Mol. Biol.* **126**:847-863.

Feinberg, A.P., and Vogelstein, B. 1984. A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* **137**:266-267.

Ferris, S.D., Sage, R.G., and Wilson, A.C. 1982. Evidence from mt DNA sequences that common laboratory strains of inbred mice are descended from a single female. *Nature* **295**:163-165.

Festing, M.F.W. 1979. Inbred strains in biomedical research. The Macmillan Press, London, England.

Finnegan, D.J. 1985. Transposable elements in eukaryotes. *Int. Rev. Cytol.* **93**:281-326.

Finnegan, D.J. 1989. Eukaryotic transposable elements and genome evolution. *Trends Genet.* **5**:103-107.

Franke, U., Lalley, P.A., Moss, W., Ivy, I., and Minna, J.D. 1977. Gene mapping in *Mus musculus* by interspecific cell hybridisation : assignment of the genes for tripeptidase-1 to chromosome 10, dipeptidase-2 to chromosome 18, acid posphatatse-1 to chromosome 12, and adenylate kinase-1 to chromosome 2. *Cytogenet. Cell. Genet.* **19**:57-84.

Friedman, J., Fletcher, C., Bahay, N., Noman, D., Leibel, R., Albright, K., Cram, S., and Heintz, N. 1989. Molecular mapping of mouse chromosomes 4 and 6 : use of a chromosome specific library constructed from a flow sorted Robertsonian chromosome. *Mouse News Lett.* **85**:84.

Friend, S.H., Bernards, R., Rogel, S., Weinbertg, R.A., Rapaport, J.M., Albert, D.M., and Dryja, T.P. 1986. A human DNA segment that predisposes to retinoblastoma and osteosarcoma. *Nature* **323**:643-646.

Fry, K., and Salser, W. 1977. Nucleotide sequences of HS-α satellite DNA from kangaroo rat *Dipodomys ordii* and characterisation of similar sequences in other rodents. *Cell* **12**:1069-1084.

Galli, G., Hofstetter, H., and Bernstiel, M.L. 1981. Two conserved sequence blocks within eukaryotic tRNA genes are major promoter elements. *Nature* **294**:626-631.

Gardner, R.L. 1968. Mouse chimeras obtained by the injection of cells into the blastocyst. *Nature* **220**:596-597.

Gardner, R.L. 1978. The relationship between cell lineage and differentiation in the early mouse embryo. *In* "Genetic mosaics and cell differentiation", Gehring, W.J., Ed., Springer-Verlag, Berlin, FR Germany.

Gardner, R.L. 1985. Regeneration of endoderm from primitive ectoderm in the mouse embryo : fact or artifact? *J. Embryol. Exp. Morph.* **88**:303-326.

Gardner, R., Lyon, M., Evans, E., and Burtenshaw, M. 1985. Clonal analysis of X-chromosome inactivation and the origin of the germline in the mouse embryo. *J. Embryol. Exp. Morph.* **88**:349-363.

Gebhard, W., and Zachau, H.G. 1983. Simple DNA sequences and dispersed repetitive elements in the vicinity of mouse immunoglobulin K light chain genes. *J. Mol. Biol.* **170**:567-573.

Georges, M., Cochaux, P., Lequarre, A-S., Young, M.W., and Vassart, G. 1987. DNA fingerprinting in man using a mouse probe related to part of the *Drosophila per* gene. *Nucleic Acids Res.* **15**:7193.

Georges, M., Lequarre, A-S., Castelli, M., Hanset, R., and Vassart, G. 1988. DNA fingerprinting in domestic animals using four different minisatellite probes. *Cytog. Cell. Genet.* **47**:127-131.

Gibbs, R., and Caskey, C.T. 1987. Identification and localisation of mutants at the Lesch-Nyhan locus by ribonuclease A cleavage. *Science* **236**:303-305.

Gill, P., Jeffreys, A.J., and Werrett, D. 1985. Forensic application of DNA 'fingerprints'. *Nature* **318**:577-579.

Gilmour, D.S., Thomas, G.H., and Elgin, S.C.R. 1989. *Drosophila* nuclear proteins bind to regions of alternating C and T residues in gene promoters. *Science* **245**:1487-1490.

Gomez-Marquez, J., Puga, A., and Notkins, A.L. 1985. Regions of the terminal repetitions of the herpes simplex virus type 1 genome. *J. Biol. Chem.* **260**:3490-3495.

Gonzalez, F.J., and Kasper, C.B. 1983. Cloning and characterisation of the rat NADPH-cytochrome P-450 oxidoreductase gene. *J. Biol. Chem.* **258**:1363-1368.

Goodbourn, S.E.Y., Higgs, D.R., Clegg, J.B., and Weatherall, D.J. 1984. Molecular basis of length polymorphism in the human ζ-globin gene complex. *Proc. natl. Acad. Sci. USA* **80**:5022-5026.

Greaves, D., and Patient, R. 1985. $(AT)_n$ is an interspersed repeat in the *Xenopus* genome. *EMBO J.* **4**:2617-2626.

Greider, C.W., and Blackburn, E.M. 1987. The telomere terminal transferase of *Tetrahymena* is a ribonucleoprotein enzyme with two kinds of primer specificity. *Cell* **51**:887-898.

Grunstein, M., and Hogness, D.S. 1975. Colony hybridisation : a method for the isolation of cloned DNAs that contain a specific gene. *Proc. Natl. Acad. Sci. USA* **72**:3961-3965.

Gusella, J.F., Wexler, N.S., Conneally, P.M., Naylor, S.L., Anderson, M.A., Tanzi, R.E., Watkins, P.C., Ottina, K., Wallace, M.R., Sakaguchi, A.Y., Young, A.B., Shoulson, I, Bonilla, E., and Martin, J.B. 1983. A polymorphic DNA marker genetically linked to Huntington's disease. *Nature* **306**:234-238.

Haldane, J.B.S., and Waddington, C.H. 1931. Inbreeding and linkage. *Genetics* **16**:357-374

Hamada, H., Petrino, M., Kakunaga, T., Seidman, M., and Stollar, B. 1984. Characteristics of genomic poly(dT-dG).poly(dCdA) sequences : structure, organisation and conformation. *Mol. Cell. Biol.* **4**:2610-2621.

Hanahan, D. 1983. Studies on transformation of Esherishia coli with plasmids. *J. Mol. Biol.* **166**:557-580.

Harbers, K., Soriano, P., Muller, U., and Jaenisch, R. 1986. High frequency of unequal recombination in pseudoautosomal region shown by proviral insertion in transgenic mice. *Nature* **324**:682-685.

Harris, H., and Hopkinson, D.A. 1972. Average heterozygosity per locus in man : an estimate based on the incidience of enzyme polymorphisms. *Ann. Hum. Genet.* **36**:9-19.

Hasson, J-F., Mougneau, E., Cuzin, F., and Yaniv, M. 1984. SV40 illegitimate recombination occurs near short direct repeats. *J. Mol. Biol.* **177**:53-68.

Hastie, N.D. 1989. Highly repeated DNA families in the genome of *Mus musculus*. *In* "Genetic strains and variants of the laboratory mouse", Second edition, Lyon, M.F., and Searle, A.G., Eds., Oxford University Press, Oxford, U.K.

Hastie, N.D., and Allshire, R.C. 1989. Human telomeres : fusion and interstitial sites. *Trends Genet.* **5**:326-331.

Heinlein, U.A.O., Lange-Sablitzky, R., Schaal, H., and Wille, W. 1986. Molecular characterisation of the MT-family of dispersed middle-repetitive elements. *Nucleic Acids Res.* **14**:6403-6416.

Hellman, L., Steen, M-L., Sundvall, M., and Petersson, U. 1988. A rapidly evolving region in the immunoglobulin heavy chain loci of rat and mouse : postulated role of $(dC\text{-}dA)_n.(dG\text{-}dT)_n$ sequences. *Gene* **68**:93-100.

Higgs, D., Wainscoat, J., Flint, J., Hill, A., Thein, S., Nicholls, R., Teal, H., Ayyub, H., Peto, T., Falusi, A., Jarman, A., Clegg, J., and Weatherall, D. 1986. Analysis of the human α-globin gene cluster reveals a highly informative genetic locus. *Proc. Natl. Acad. Sci. USA* **81**:5165-5169.

Hill, A.V.S., and Jeffreys, A.J. 1985. Use of minisatellite probes for determination of twin zygosity at birth. *Lancet* **ii**:1394-1395.

Hodgson, C.P., Fisk, R.Z., Arora, P., and Chotani, M. 1990. Nucleotide sequence of mouse virus-like (VL30) retrotransposon BVL-1. *Nucleic Acids Res.* **18**:673.

Hogan, B.L.M., Barlow, D.P., and Tilly, R. 1983. F9 teratocarcinoma cells as a model for the differentiation of parietal and visceral endoderm in the mouse embryo. *Cancer Surveys* **2**:115-140.

Hogan, B., Constantini, F., and Lacy, E. 1986. Manipulating the mouse embryo, a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, USA.

Horz, W., and Altenburger, W. 1981. Nucleotide sequence of mouse satellite DNA. *Nucleic Acids. Res.* **9**:683-696.

Howe, C.C., and Overton, G.C. 1986. Expression of the intracisternal A-particle is elevated during differentiation of embryonal carcinoma cells. *Mol. Cell. Biol.* **6**:150-157.

Ish-Horowicz, D., and Burke, J.F. 1981. Rapid and efficient cosmid cloning. *Nucleic Acids Res.* **9**:2989-2998.

Jaenisch, R., Harbers, K., Schnieke, A., Lohler, J., Chumakov, I., Jahner, D., Grotkopp, D., and Hoffman, E. 1983. Germline integration of Moloney Murine Leukaemia virus at the Mov13 locus leads to recessive lethal mutation and early embryonic death. *Cell* **32**:209-216.

Janssens, F.A. 1909. Spermatogenese dans les Batraciens. V. La theorie de la chiasmatypie. Nouvelles interpretation des cineses de maturation. *Cellule* **25**:387-411.

Jarman, A.P., Nicholls, R.D., Weatherall, D.J., Clegg, J.B., and Higgs, D.R. 1986. Molecular characterisation of a hypervariable region downstream of the human alpha-globin cluster. *EMBO J.* **5**:1857-1863.

Jarman, A.P., and Wells, R.A. 1989. Hypervariable minisatellites : recombinators or innocent bystanders? *Trends Genet.* **5**:367-371.

Jeffreys, A.J. 1979. DNA sequence variants in the $^{G}\gamma$-, $^{A}\gamma$-, δ- and β-globin genes of man. *Cell* **18**:1-10.

Jeffreys, A.J., Wilson, V., Wood, D., Simons, J.P., Kay, R.M., and Williams, J.G. 1980. Linkage of adult α- and β-globin genes in *X.laevis* and gene duplication by tetraploidisation. *Cell* **21**:555-564.

Jeffreys, A.J., Wilson, V., and Thein, S.L. 1985a. Hypervariable 'minisatellite' regions in human DNA. *Nature* **314**:67-73.

Jeffreys, A.J., Wilson, V., and Thein, S.L. 1985b. Individual-specific 'fingerprints' of human DNA. *Nature* **316**:76-79.

Jeffreys, A.J., Brookfield, J.F.Y., and Semeonoff, R. 1985c. Positive identification of an immigration test-case using human DNA fingerprints. *Nature* **317**:818-819.

Jeffreys, A.J., Wilson, V., Thein, S.L., Weatherall, D.J., and Ponder, B.A. 1986. DNA 'fingerprints' and segregation analysis of multiple markers in human pedigrees. *Am. J. Hum. Genet.* **38**:11-42.

Jeffreys, A.J., and Morton, D.B. 1987. DNA fingerprints of dogs and cats. *Animal Genet.* **18**:1-15.

Jeffreys, A.J., Wilson, V., Kelly, R., Taylor, B.A., and Bulfield, G. 1987. Mouse DNA 'fingerprints' : analysis of chromosome localisation and germ-line stability of hypervariable loci in recombinant inbred strains. *Nucleic Acids Res.* **15**:2823-2836.

Jeffreys, A.J., Royle N.J., Wilson, V., and Wong, Z. 1988a. Spontaneous mutation rates to new-length alleles at tandem-repetitive hypervariable loci in human DNA. *Nature* **332**:278-281.

Jeffreys, A.J., Wilson, V., Neumann, R., and Keyte, J. 1988b. Amplification of human minisatellites by the polymerase chain reaction : towards DNA fingerprinting of single cells. *Nucleic Acids Res.* **16**:10953-10971.

Jeffreys, A.J., Neumann, R., and Wilson, V. 1990. Repeat unit sequence variation in minisatellites : a novel source of DNA polymorphism for studying allelic variation and mutation by single molecule analysis. *Cell*, in press.

Jenkins, N.A., Copeland, N.G., Taylor, B.A., and Lee, B.K. 1981. Dilute (*d*) coat colour mutation of DBA/2J mice is associated with the site of integration of an ecotropic MuLV genome. *Nature* **293**:370-374.

Jenkins, N.A., Copeland, N.G., Taylor, B.A., Bedigan, H.G., and Lee, B.K. 1982. Ecotropic murine leukaemia virus DNA content of normal and lymphomatous tissues of BXH-2 recombinant inbred mice. *J. Virol.* **42**:379-388.

Johannsen, W. 1903. Uber erblichkeit in Populationen und in reinen Linien. Jena, Gustav Fischer.

John, B., and Miklos, G.L.G. 1988. The eukaryote genome in development and evolution. Allen and Unwin, London, England.

Kalb, V.F., Glasser, S., King, D., and Lingrel, J.B. 1983. A cluster of repetitive elements within a 700 base pair region in the mouse genome. *Nucleic Acids Res.* **11**:2177-2184.

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

Kassis, J.A., Poole, J., Wright, D.K., and O'Farrell, P. 1986. Sequence conservation in the protein coding and intron regions of the *engrailed* transcription unit. *EMBO J.* **5**:3583-3589.

Keyeux, G., Lefranc, G., and Lefranc, M-P. 1989. A multigene deletion in the human IGH constant region locus involves highly homologous hot spots of recombination. *Genomics* **5**:431-441.

Kimura, M., and Crow, J.F. 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, England.

King, D., Snider, L.D., and Lingrel, J.B. 1986. Polymorphism in an androgen-regulated mouse gene is the result of the insertion of a B1 repetitive element into the transcription unit. *Mol. Cell. Biol.* **6**:209-217.

Kingsmore, S.F., Watson, M.L., Howard, T.A., and Seldin, M.F. 1989. A 6000kb segment of chromosome 1 is conserved in human and mouse. *EMBO J.* **8**:4073-4080.

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

Knott, T., Wallis, S., Pease, R., Powell, L., and Scott, J. 1986. A hypervariable region 3' to the human apolipoprotein B gene. *Nucleic Acids Res.* **14**:9215.

Knowlton, R.G., Brown, V., Braman, J., Barker, D., Schumm, J., Murray, C., Takvorian, T., Ritz, J., and Donis-Keller, H. 1986. Use of highly polymorphic DNA probes for genotypic analysis following bone marrow transplantation. *Blood* **68**:378-385.

Kobori, J., Strauss, E., Minard, K., and Hood, L. 1986. Molecular analysis of the hotspot of recombination in the murine major histocompatibility complex. *Nature* **234**:175-179.

Kominami, R., Muramatsu, M., and Moriwaki, K. 1983a. A mouse type 2 Alu sequence (M2) is mobile in the genome. *Nature* **301**:87-88.

Kominami, R., Urano, Y., Mishima, Y., Muramatsu, M., Moriwaki, K., and Yoshikura, H. 1983b. Novel repetitive sequence families showing size and frequency polymorphism in the genomes of mice. *J. Mol. Biol.* **165**:209-228.

Kominami, R., Mitani, K., and Muramatsu, M. 1987. Nucleotide sequence of a mouse minisatellite DNA. *Nucleic Acids Res.* **16**:1197.

Korenberg, J., and Ryowski, M. 1988. Human genome organisation : Alu, LINES, and the molecular structure of metaphase chromosomes. *Cell* **53**:391-400.

Kourilsky, P. 1986. Molecular mechanisms for gene conversion in higher cells. *Trends Genet.* **2**:60-63.

Kramerov, D.A., Tillib, S.V., Lekakh, I.V., Ryskov, A.P., and Georgiev, A.P. 1985. Nucleotide sequence of small polyadenylated B2 RNA. *Nucleic Acids Res.* **13**:6423-6437.

Krayev, A.S., Kramerov, D.A., Skryabin, K.G., Ryskov, A.P., Bayev, A.A., and Georgiev, G.P. 1980. The nucleotide sequence of the ubiquitous repetitive DNA sequence complementary to the most abundant class of mouse fold-back RNA. *Nucleic Acids Res.* **8**:1201-1215.

Krayev, A.S., Markusheva, T.V., Kramerov, D.A., Ryskov, A.P., Skryabin, K.G., Bayev, A.A., and Georgiev, G.P. 1982. Ubiquitous transposon-like repeats B1 and B2 of the mouse genome : B2 sequencing. *Nucleic Acids Res.* **10**:7461-7475.

Kress, M., Barra, Y., Seidman, J.G., Khoury, G., and Jay, G. 1984. Functional insertion of an Alu type 2 (B2 SINE) repetitive sequence in murine class I genes. *Science* **226**:974-977.

Kuff, E.L., Feenstra, A., Leuders, K., Smith, L., Hawley, R., Hozumi, N., and Shulman, M. 1983. Intracisternal A-particle genes as movable elements in the mouse genome. *Proc. Natl. Acad. Sci. USA* **80**:1992-1996.

Landsteiner, K. 1900. Zur kenntnis der antifermentatinen, lytischen und agglutinierenden wirkungen des blutserums und der lymphe. *Zentralbl Bakteriol* **27**:357-362.

Laurie, D.A., and Hulten, M.A. 1985. Further studies on chiasma distribution and interference in the human male. *Ann. Hum. Genet.* **49**:203-214.

Lerhman, M.A., Russell, D.W., Goldstein, J.L., and Brown, M.S. 1986. Exon-Alu recombination deletes 5 kilobases from low density lipoprotein receptor gene, producing null phenotype in familial hypercholesterolemia. *Proc. Natl. Acad. Sci. USA* **83**:3679-3683.

Lerhman, M.A., Goldstein, J.L., Russell, D.W., and Brown, M.S. 1987. Duplication of seven exons in LDL receptor gene caused by Alu-Alu recombination in a subject with familial hypercholesterolemia. *Cell* **48**:827-835.

Leuders, K.K., and Kuff, E.L. 1977. Sequences associated with intracisternal A-particles are reiterated in the mouse genome. *Cell* **12**:963-972.

Levinson, G., Marsh, J.L., Epplen, J.T., and Gutman, G.A. 1985. Cross-hybridising snake satellite, *Drosophila*, and mouse sequences may have arisen independently. *Mol. Biol. Evol.* **2**:494-504.

Levinson, G., and Gutman, G.A. 1987. Slipped-strand mispairing : a major mechanism for DNA sequence evolution. *Mol. Biol. Evol.* **4**:203-221.

Levinson, G., and Gutman, G. 1988. High frequencies of short frameshifts in poly-CA/TG tandem repeats borne by phage M13 in *E.coli* K12. *Nucl Acids Res.* **15**:5323-5338.

Lewis, N.E., and Rossant, J. 1982. Mechanism of size-regulation in mouse embryo aggregtates. *J. Embryol. Exp. Morph.* **72**:169-181.

Lipman, D.J., and Pearson, W.R. 1985. Rapid and sensitive protein similarity searches. *Science* **227**:1435-1441.

Litt, M., and White, R. 1985. A highly polymorphic locus in human DNA revealed by cosmid derived probes. *Proc. Natl. Acad. Sci. USA* **82**:6206-6210.

Litt, M., and Luty, J.A. 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.

Little, C.C., and Tyzzer, E.E. 1915. *J. Med. Res.* **33**:393.

Loeb, D.D., Padgett, R.W., Hardies, S.C., Shehee, W.R., Comer, M., Edgell, M.H., and Hutchinson III, C.A. 1986. The sequence of a large L1Md element reveals a tandemly repeated 5' end and several features found in retroposons. *Mol. Cell. Biol.* **6**:168-182.

Loenen, W.A., and Brammar, W.J. 1980. A bacteriophage λ vector for cloning large DNA fragments made with several restriction enzymes. *Gene* **10**:249-259.

Lowell, C.A., Potter, D.A., Stearman, R.S., and Morrow, J.F. 1986. Structure of the murine serum amyloid A gene family. *J. Biol. Chem.* **261**:8442-8452.

Lusis, A.J., Taylor, B.A., Quon, D., Zollman, S., and LeBoeuf, R.C. 1987. Genetic factors controlling structure and expression of apolipoproteins A and B in mice. *J. Biol. Chem.* **262**:7594-7604.

Maeda, N., and Smithies, O. 1986. The evolution of multigene families : human haptoglobin genes. *Ann. Rev. Genet.* **20**:81-108.

Malecot, G. 1948. Les mathematics de l'heredite. Masson, Paris, France.

Maniatis, T., Fritsch, E.F., and Sambrook, J. 1982. Molecular cloning, a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, USA.

McCarrey, J.R., and Thomas, K. 1987. Human testis-specific PGK gene lacks introns and possesses characteristics of a processed gene. *Nature* **326**:501-505.

McKusick, V.A. 1986. Mendelian inheritance in man : catalogs of autosomal dominant, autosomal recessive, and X-linked phenotypes, 7th edition. John Hopkins University Press, Baltimore, USA.

McLaren, A. 1984. Germ cell lineages. *In* "Chimeras in developmental biology", Le Douarin, N., and McLaren, A., Eds., Academic Press, London, England.

McMahon, A., Fosten, M., and Monk, M. 1983. X-chromosome inactivation mosaicism in the three germ layers and the germ line of the mouse embryo. *J. Embryol. Exp. Morphol.* **74**:207-220.

McWilliam, P., Farrar, G.J., Kenna, P., Bradley, D.G., Humphries, M.M., Sharp, E.M., McConnell, D.J., Lawler, M., Sheils, D., Ryan, C., Stevens, K., Daiger, S.P., and Humphries, P. 1989. Autosomal dominant retinitis pigmentosa (ADRP) : localisation of an ADRP gene to the long arm of chromosome 3. *Genomics* **5**:619-622.

Mendel, G. 1866. Versuche uber Pflanzenhybriden. *Vrh. naturf. Ver. Brunn.* **4**:3-44.

Messing, J., Crea, R., and Seeburg, P.H. 1981. A system for shotgun DNA sequencing. *Nucleic Acids Res.* **9**:309-321.

Mintz, B. 1974. Gene control in mammalian development. *Ann. Rev. Genet.* **8**:411-470.

Mirsky, A.E., and Ris, H. 1951. The desoxyribonucleic acid content of animal cells and its evolutionary significance. *J. gen. Physiol.* **34**:451-462.

Mishkin, J.D., Taylor, B.A. and Mellman, W.J. 1976. *Glk* : a locus controlling galactokinase activity in the mouse. *Biochem. Genet.* **14**:635-640.

Monaco, A., Neve, R., Colletti-Feener, C., Bertelson, C., Kurnit, D., and Kunkel, L. 1986. Isolation of candidate cDNAs for portions of the Duchenne muscular dystrophy gene. *Nature* **323**:646-650.

Monk, M., and Harper, M. 1979. Sequential X-chromosome inactivation coupled with cellular differentiation in early mouse embryos. *Nature* **281**:311-313.

Morin, G.B. 1989. The human telomere terminal transferase enzyme is a ribonucleoprotein that synthesises TTAGGG repeats. *Cell* **59**:521-529.

Morse, H.C., Ed. 1978. Origins of inbred mice. Academic Press, New York, USA.

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

Murphy, D., Brickell, D.S., Latchman, D.S., Willison, K., and Rigby, P.W.J. 1983. Transcripts regulated during normal embryonic development and oncogenic transformation shares a repetitive element. *Cell* **35**:865-871.

Myers, R.M., Lumelsky, N., Lerman, L.S., and Maniatis, T. 1985. Detection of single base substitutions in total genomic DNA. *Nature* **313**:495-498.

Nadeau, J.H., Berger, F.G., Kelley, K.A., Pitha, P.M., Sidman, C.L., and Worrall, N. 1986. Rearragement of genes located on homologous chromosomal segments in mouse and man : the location of genes for alpha- and beta-interferon, alpha-1 acid glycoprotein-1 and -2, and aminolevulinate dehydratase on mouse chromosome 4. *Genetics* **104**:1239-1255.

Nakamura, Y., Julier, C., Wolff, R., Holm, T., Leppert, M., and White, R. 1987a. Characterisation of a human 'midisatellite' sequence. *Nucleic Acids Res.* **15**:2537-2547.

Nakamura, Y., Leppert, M., O'Connell, P., Wolff, R., Holm, T., Culver, M., Martin, C., Fujimoto, E., Hoff, M., Kumlin, E., and White, R. 1987b. Variable number of tandem repeat (VNTR) markers for human gene mapping. *Science* **235**:1616-1622.

Neidle, S. 1983. New twists to left-handed DNA. *Nature* **302**:574.

Nesbitt, M.N. 1971. X-chromosome inactivation mosaicism in the mouse. *Dev. Biol.* **26**:252-263.

Nicholls, R.D., Fischel-Ghodsian, N., and Higgs, D.R. 1987. Recombination at the human α-globin gene cluster : sequence features and topological constraints. *Cell* **49**:369-378.

Nikaido, T., Nakai, S., and Honjo, T. 1981. Switch region of immunoglobulin C_μ gene is composed of simple tandem repeats. *Nature* **292**:845-848.

O'Hara, P.J., and Grant, F.J. 1988. The human factor VII gene is polymorphic due to variation in repeat copy number in a minisatellite. *Gene* **66**:147-158.

Ohno, S. 1970. So much 'junk' DNA in our genome. *In* "Evolution of genetic systems", Smith, H.H., Ed., *Brookhaven Symposia in biology* **23**. Gordon and Breach, New York.

Ohno, S. 1984. Repeats of base oligomers as the primordial coding sequences of the primeval earth and their vestiges in modern genes. *J. Mol. Evol.* **20**:313-321.

Orgel, L.E., and Crick, F.H.C. 1980. Selfish DNA : the ultimate parasite. *Nature* **284**:604-607.

Owens, G., Nirupa, C., Hahn, W. 1985. Brain 'identifier sequence' is not restricted to brain : similar abundance in nuclear RNA of other organs. *Science* **229**:1263-1265.

Pardue, M.L., and Gall, J.G. 1970. Chromosomal localisation of mouse satellite DNA. *Science* **168**:1356-1358.

Paulson, K.E., Deka, N., Schmid, C.W., Misra, R., Schindler, C.W., Rush, M.G., Kadyk, L., and Leinwand, L. 1985. A transposon-like element in human DNA. *Nature* **316**:359-361.

Piche, A., and Bourgaux, P. 1987. Resolution of a polyoma virus-mouse hybrid replicon : release of genomic viral DNA. *J. Virol.* **61**:840-844.

Pietras, D.F., Bennet, K.L., Siracusa, L.D., Woodworth-Gutai, V.M., Chapman, V.M., Gross, K.W., Kane-Haas, C., and Hastie, N.D. 1983. Construction of a small *Mus musculus* repetitive DNA library : identification of a new satellite sequence in *Mus musculus*. *Nucleic Acids Res.* **11**:6965-6983.

Piko, L., Hammons, M.D., and Taylor, K.D. 1984. Amounts, synthesis, and some properties of intracisternal A-particle-related RNA in early mouse embryos. *Proc. Natl. Acad. Sci. USA* **81**:488-492.

Pingoud, A. 1985. Spermidine increases the accuracy of type II restriction endonucleases. *Eur. J. Biochem.* **147**:105-109.

Proudfoot, N.J., Gil, A., and Maniatis, T. 1982. The structure of the human zeta-globin gene and a closely linked, nearly identical pseudogene. *Cell* **31**:553-563.

Raleigh, E.A., Murray, N.E., Revel, H., Blumenthal, R.M., Westaway, D., Reith, A.D., Rigby, P.W.J., Elhai, J., and Hanahan, D. 1988. *Mcr*A and *Mcr*B restriction phenotypes of some *E.coli* strains and implications for gene cloning. *Nucleic Acids Res.* **16**:1563-1575.

Rechavi, G., Givol, D., and Canaazi, E. Activation of a cellular oncogene by DNA rearrangement : possible involvement of an IS-like element. *Nature* **300**:607-611.

Reeders, S.T., Breuning, M.H., Davies, K.E., Nicholls, R.D., Jarman, A.P., Higgs, D.R., Pearson, P.L., and Weatherall, D.J. 1985. A highly polymorphic DNA marker linked to adult polycystic kidney disease on chromosome 16. *Nature* **317**:542-544.

Rogers, J.H. 1983. CACA sequences - the ends or the means? *Nature* **305**:101-102.

Rogers, J.H. 1985. Origin and evolution of retroposons. *Int. Rev. Cytol.* **93**:187-279.

Rogers, J.H. 1986. The origin of retroposons. *Nature* **319**:725.

Rommens, J.M., Ianuzzi, M.C., Kerem, B-S., Drumm, M.L., Melmer, G., Dean, M., Rozmahel, R., Cole, J.L., Kennedy, D., Hidaka, N., Zsiga, M., Buchwald, M., Rordan, J.R., Tsui, L-C., and Collins, F.S. 1989. Identification of the cystic fibrosis gene : chromosome walking and jumping. *Science* **245**:1059-1065.

Rosenberg, H., Singer, M.F., and Rosenberg, M. 1978. Highly repeated sequence of simiansimiansimiansimiansimian. *Science* **200**:394-402.

Rossant, J. 1984. Somatic cell lineages in mammalian chimaeras. *In* "Chimaeras in developmental biology", LeDouarin, N. and McLaren, A., Eds., Academic Press, London, England.

Rossant, J. 1986. Retroviral mosaics : a new approach to cell lineage analysis in the mouse embryo. *Trends Genet.* **2**:302-303.

Rotman, G., Itin, A., and Keshet, E. 1984. 'Solo' long terminal repeats (LTR) of an endogenous retrovirus-like gene family (VL30) in the mouse genome. *Nucleic Acids Res.* **12**:2273-2282.

Rouyer, F., Simmler, M-C., Johnsson, C., Vergnaud, G., Cooke, H.J., and Weissenbach, J. 1986. A gradient of sex linkage in the pseudoautosomal region of the human sex chromosomes. *Nature* **319**:291-295.

Royle, N.J., Clarkson, R.C., Wong, Z., and Jeffreys, A.J. 1988. Clustering of hypervariable minisatellites in the proterminal regions of mammalian autosomes. *Genomics* **3**:352-360.

Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B., and Erlich, H.A. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239**:487-491.

Saito, I., and Stark, G.R. 1986. Charomid vectors for efficient cloning and mapping of large or small restriction fragments. *Proc. Natl. Acad. Sci. USA* **83**:8664-8668.

Sakamoto, K., and Okada, N. 1985. Rodent type 2 Alu family, rat identifier sequence, rabbit C-family, and bovine or goat 73bp repeat may have evolved from tRNA genes. *J. Mol. Evol.* **22**:134-140.

Sanger, F., Nicklen, S., and Coulson, A.R. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**:5463-5467.

Sanger, F., Coulson, A.R., Barrell, B.G., Smith, A.J.H., and Roe, B.A. 1980. Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. *J. Mol. Biol.* **143**:161-178.

Schafer, R., Ali, S., and Epplen, J.T. 1986. The organisation of the evolutionary conserved GATA/GACA repeats in the mouse genome. *Chromosoma* **93**:502-510.

Schmid, C.W., and Deininger, P.L. 1975. Sequence organisation of the human genome. *Cell* **6**:345-358.

Schmid, C.W., and Jelinek, W.R. 1982. The Alu family of dispersed repetitive sequences. *Science* **216**:1065-1070.

Schmidt, M., Glogger, K., Wirth, T., and Horak, I. 1984. Evidence that a major class of mouse endogenous long terminal repeats (LTRs) resulted from recombination between exogenous retroviral LTRs and similar LTR-like elements (LTR-IS). *Proc. Natl. Acad. Sci. USA* **81**:6696-6700.

Schmidt, M., Wirth, T., Kroger, B., and Horak, I. 1985. Structure and genomic organisation of a new family of murine retrovirus-related DNA sequences (MuRRS). *Nucleic Acids Res.* **13**:3461-3470.

Shafit-Zagardo, B., Maio, J. and Brown, F. 1982. *Kpn*I families of long interspersed repetitive DNAs in human and other primate genomes. *Nucleic Acids Res.* **10**:3175-3196.

Sharp, P.A., Sugden, B., and Sambrook, J. 1973. Detection of two restriction endonuclease activities in *Haemophilus influenzae* using analytical agarose-ethidium bromide electrophoresis. *Biochemsitry* **12**:3055-3063.

Sharp, P.A. 1983. Conversion of RNA to DNA in mammals : Alu-like elements and pseudogenes. *Nature* **301**:471-472.

Sherrington, R., Brynjolfsson, J., Petursson, H., Potter, M., Dudleston, K., Barraclough, B., Wasmuth, J., Dobbs, M., and Gurling, H. 1988. Localisation of a susceptibility locus for schizophrenia on chromosome 5. *Nature* **336**:164-167.

Shin, H-S., Bargiello, T.A., Clark, B.T., Jackson, F.R., and Young, M.W. 1985. An unusual coding sequence from a *Drosophila* clock gene is conserved in vertebrates. *Nature* **317**:445-448.

Shippen-Lentz, D., and Blackburn, E.H. 1990. Functional evidence for an RNA template in telomerase. *Science* **247**:546-552.

Silva, A., Johnson, J., and White, R. 1987. Characterisation of a highly polymorphic region 5' to J_H in the human immunoglobulin heavy chain. *Nucleic Acids Res.* **15**:3845-3857.

Silva, A.J., and White, R. 1988. Inheritance of allelic blueprints for methylation patterns. *Cell* **54**:145-152.

Silver, J. 1985. Confidence limits for estimates of gene linkage based on analysis of recombinant inbred strains. *J. Hered.* **76**:436-440.

Simmler, M-C., Rouyer, F., Vergnaud, G., Nystrom-Lahti, M., Ngo, K.Y., de La Chapelle, A., Weissenbach, J. 1985. Pseudoautosomal DNA sequences in the pairing region of the human sex chromosomes. *Nature* **317**:692-697.

Simmler, M-C., Johnsson, C., Petit, C., Rouyer, F., Vergnaud, G, and Weissenbach, J. 1987. Two highly polymorphic minisatellites from the pseudoautosomal region of the human sex chromosomes. *EMBO J.* **6**:963-969.

Singer, M.F. 1982a. Highly repeated sequences in mammalian genomes. *Int. Rev. Cytol.* **76**:67-112.

Singer, M.F. 1982b. Highly repeated SINEs and LINEs : short and long interspersed sequences in mammalian genomes. 1982. *Cell* **28**:433-434.

Singer, M.F., and Skowronski, J. 1985. Making sense out of LINES : Long interspersed repeat sequences in mammalian genomes. *Trends Bichem. Sci.* **10**:119-121.

Singh, L., Purdom, I.F., and Jones, K.W. 1980. Sex chromosome associated satellite DNA. Evolution and conservation. *Chromosoma* **79**:137-157.

Singh, L., Phillips, C., and Jones, K.W. 1984. The conserved nucleotide sequences of Bkm, which define *Sxr* in the mouse, are transcribed. *Cell* **36**:111-120.

Skowronski, J., and Singer, M. 1985. Expression of a cytoplasmic LINE-1 transcript is regulated in a human teratocarcinoma cell line. *Proc. Natl. Acad. Sci. USA* **82**:6050-6054.

Slightom, J., Blechl, A., and Smithies, O. 1980. Human fetal $^{G}\gamma$- and $^{A}\gamma$-globin genes : complete nucleotide sequences suggest that DNA can be exchanged between these duplicated genes. *Cell* **21**:627-638.

Smith, G.P. 1976. Evolution of repeated sequences by unequal crossing-over. *Science* **191**:528-535.

Smith, G.R. 1983. Chi hotspots of generalised recombination. *Cell* **34**:709-710.

Snow, M.H.L. 1977. Proliferative centres in embryonic development. *In* "Development in mammals", Volume 3, Johnson, M., Ed., Elsevier, Amsterdam, Holland.

Soller, M., and Beckmann, J.S. 1983. Genetic polymorphism in varietal identification and gametic improvement. *Theor. Appl. Genet.* **67**:25-33.

Soriano, P., and Jaenisch, R. 1986. Retroviruses as probes for mammalian development : allocation of cells to the somatic and germ cell lineages. *Cell* **46**:19-29.

Southern, E.M. 1970. Base sequence and evolution of guinea pig α-satellite DNA. *Nature* **227**:794-798.

Southern, E.M. 1975a. Long range periodicities in mouse satellite DNA. *J. Mol. Biol.* **94**:52-69.

Southern, E.M. 1975b. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* **98**:503-517.

Steinmetz, M., Moore, K.W., Frelinger, J.G., Sher, B.T., Shen, F-W., Boyse, E.A., and Hood, L. 1981. A pseudogene homologous to mouse transplantation antigens : transplantation antigens are encoded by eight exons that correlate with protein domains. *Cell* **25**:683-692.

Steinmetz, M. Uematsu, Y., and Lindahl, K.F. 1987. Hotspots of homologous recombination in the mammmalian genome. *Trends Genet.* **3**:7-10.

Stephan, W. 1989. Tandem-repetitive noncoding DNA : forms and forces. *Mol. Biol. Evol.* **6**:198-212.

Stoker, N.G., Cheah, K., Griffin, J., Pope, F., and Solomon, E. 1985. A highly polymorphic region 3' to the human Type II collagen gene. *Nucleic Acids Res.* **13**:4613-4622.

Stringer, J.R. 1982. DNA sequence homology and chromosomal deletion at a site of SV40 DNA integration. *Nature* **296**:363-366.

Sueoka, N. 1961. Variation and hetrogeneity of base composition of deoxyribonucleic acids : a compilation of old and new data. *J. Mol. Biol.* **3**:31-40.

Sun, L., Paulson, K., Schmid, C., Kadyk, L., and Leinwand, L. 1984. Non-Alu family interspersed repeats in human DNA and their transcriptional activity. *Nucleic Acids Res.* **12**:2669-2690.

Sutcliffe, J.G., Milner, R.J., Bloom, R.E., and Lerner, R.A. 1982. Common 82-nucleotide sequences unique to brain RNA. *Proc. Natl. Acad. Sci. USA* **79**:4942-4946.

Swallow, D.M., Gendler, S., Griffiths, B., Corney, G., Taylor-Papadimitriou, J., and Bramwell, M.E. 1987. The human tumour-associated epithelial mucins are encoded by an expressed hypervariable gene locus *PUM*. *Nature* **328**:82-84.

Sylla, B.S., Allard, D., Roy, G., Bourgaux-Ramoisy, D., and Bourgaux, P. 1984a. A mouse DNA sequence that mediates integration and excision of polyoma virus DNA. *Gene* **29**:343-350.

Sylla, B.S., Huberdeau, D., Bourgaux-Ramoisy, D., and Bourgaux, P. 1984b. Site-specific excision of integrated polyoma DNA. *Cell* **37**:661-667.

Szostak, J.W. and Blackburn, E.H. 1982. Cloning yeast telomeres on plasmid vectors. *Cell* **29**:245-255.

Tautz, D., and Renz, M. 1984. Simple sequences are ubiquitous repetitive components of eukaryotic genomes. *Nucleic Acids Res.* **12**:4127-4138.

Tautz, D., Trick, M., and Dover, G.A. 1986. Cryptic simpicity in DNA is a major source of genetic variation. *Nature* **322**:652-656.

Tautz, D. 1989. Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Res.* **17**:6463-6473.

Taylor, B.A., and Meier, H. 1976. Mapping the adrenal lipid depletion gene of the AKR/J mouse strain. *Genet. Res. Camb.* **26**:307-312.

Taylor, B.A., and Shen, F-W. 1977. Location of *lyb-2* on mose chromosome 4 : evidence from recombinant inbred strains. *Immunogenetics* **4**:587-599.

Taylor, B.A. 1978. Recombinant inbred strains : use in gene mapping. *In* "Origins of inbred mice", Morse, H.C., Ed., The Jackson Laboratory, Bar Harbor, Maine, USA.

Taylor, B.A. 1982. Genetic relationships between inbred strains of mice. *J. Heredity* **63**:83-86.

Taylor, B.A. 1989. Recombinant inbred strains. *In* "Genetic variants and strains of the laboratory mouse", Second edition, Lyon, M.F., and Searle, A.G., Eds., Oxford University Press, Oxford, England.

Tchurikov, N.A., Naumova, A.K., Zelentsova, E.S., and Georgiev, G.P. 1982. A cloned unique gene of *Drosophila melanogaster* contains a repetitive 3' exon whose sequence is present at the 3' ends of many different mRNAs. *Cell* **28**:365-373.

Thein, S.L., Jeffreys, A.J., and Blacklock, H.A. 1986. Identification of post-transplant cell population by DNA fingerprint analysis. *Lancet* **ii**:37

Thein, S.L., Jeffreys, A.J., Gooi, H.C., Cotter, F., Flint, J., O'Connor, N., and Wainscoat, J.S. 1987. Detection of somatic changes in human cancer DNA by DNA fingerprint analysis. *Br. J. Cancer* **55**:353-356.

Treco, D., Thomas, B., and Arnheim, N. 1985. Recombination hotspot in the human β-globin gene cluster : meiotic recombination of human DNA fragments in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **5**:2029-2038.

Treier, M., Pfeifle, C., and Tautz, D. 1989. Comparison of the gap segmentation gene *hunchback* between *Drosophila melanogaster* and *Drosophila virilis* reveals novel modes of evolutionary change. *EMBO J.* **8**:1517-1525.

Truett, M.A., Jones, R.S., and Potter, S.S. 1981. Unusual structure of the FB family of transposable elements in *Drosophila*. *Cell* **24**:753-763.

Tschermak, E. Von. 1900. Ube kunstliche Kreuzung bei *Pisum sativum*. *Ber. dt. bot. Ges.* **18**:232-239.

Tsui, L-C., Buchwald, M., Barker, D., Bramn, J.C., Knowlton, R., Schumm, J.W., Eiberg, H., Mohr, J., Kennedy, D., Plavsic, N., Zsiga, M., Markiewicz, D., Akots, G., Brown, V., Helms, C., Gravius, T., Parker, C., Rediker, K., and Donis-Keller, H. 1985. Cystic fibrosis locus defined by a genetically linked polymorphic DNA marker. *Science* **230**:1054-1057.

Tyler-Smith, C., and Brown W.R.A. 1987. Structure of the major block of alphoid satellite on the human Y-chromosome. *J. Mol. Biol.* **195**:457-470.

Uematsu, Y., Keifer, H., Schulze, R., Fischer-Lindahl, K., and Steinmetz, M. 1986. Molecular characterisation of a meiotic recombinational hotspot enhancing homologous equal crossing-over. *EMBO J.* **5**:2123-2129

Ullu, E., and Tschudi, C. 1984. Alu sequences are processed 7SL RNA genes. *Nature* **312**:171-172.

Vanin, E.F. 1984. Processed pseudogenes : characteristics and evolution. *Biochem. Biophys. Acta.* **782**:231-241.

Vassart, G., Georges, M., Monsieur, R., Brocas, H., Lequarre, A.S., Christophe, D. 1987. A sequence in M13 phage detects hypervariable minisatellites in human and animal DNA. *Science* **235**:683-684.

Vasseur, M., Condamine, H., and Duprey, P. 1985. RNAs containing B2 repeated sequences are transcribed in the early stages of mouse embryogenesis. *EMBO J.* **4**:1749-1753.

Vergnaud, G. 1989. Polymers of random short oligonucleotides detect polymorphic loci in the human genome. *Nucleic Acids Res.* **17**:7623-7630.

Vieira, J., and Messing, J. 1982. The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. *Gene* **19**:259-268.

Vissel, B., and Choo, H.K. 1989. Mouse major (γ) satellite DNA is highly conserved and organised into extremely long tandem arrays : implications for recombination between non-homologous chromosomes. *Genomics* **5**: 407-414.

Voliva, C., Martin, S., Hutchinson, C., and Edgell, M. 1984. Dispersal process associated with the L1 family of interspersed repetitive sequences. *J. Mol. Biol.* **178**:795-813.

Wagner, M, and Perry, R.P. 1985. Characterisation of the multigene family encoding the mouse S16 ribosomal protein : strategy for distinguishing an expressed gene from its processed pseudogene counterparts by an analysis of total genomic DNA. *Mol. Cell. Biol.* **5**:3560-3576.

Wagner, M. 1986. A consideration of the origin of the origin of processed pseudogenes. *Trends Genet.* **2**:134-137.

Wallenburg, J.C., Nepveu, A., and Chartrand, P. 1984. Random and nonrandom integration of a polyoma virus DNA molecule containing highly repetitive cellular sequences. *J. Virol.* **50**:678-683.

Wambaugh, J. 1989. The blooding. Bantam Press, London, England.

Warner, C.M., McIvor, J.L., and Stephens, T.J. 1977. Chimaeric drift in allophenic mice. *Transplantation* **24**:183-193.

Waye, J.S., and Willard, H.F. 1986. Structure, organisation, and sequence of alpha satellite DNA from human chromosome 17 : evidence for evolution by unequal crossing-over and an ancestral pentamer repeat shared with the human X chromosome. *Mol. Cell. Biol.* **6**:3156-3165.

Waye, J.S., Durfy, S.J., Pinkel, D., Kenwrick, S., Patterson, M., Davies, K.E., and Willard, H.F. 1987. Chromosome-specific alpha satellite DNA from human chromosome 1 : Hierarchical structure and genomic organisation of a polymorphic domain spanning several hundred kilobase pairs of centromeric DNA. *Genomics* **1**:43-51.

Weber, J.L., and May, P.E. 1989. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am. J. Hum. Genet.* **44**:388-396.

Weller, P., Jeffreys, A., Wilson, V., and Blanchetot, A. 1984. Organisation of the human myoglobin gene. *EMBO J.* **3**:439-446.

Wells, R.D. 1988. Unusual DNA structures. *J. Biol. Chem.* **263**:1095-1098.

Wernke, S., and Lingrel, J. 1986. Nucleotide sequence of the goat embryonic α-globin gene (γ) and linkage and evolutionary analysis of the complete α-globin cluster. *J. Mol. Biol.* **192**:457-471.

Wharton, K., Yedvobnick, B., Finnerty, V., and Artavanis-Tsakonas, S. 1985. Opa : a novel family of transcribed repeats shared by the *Notch* locus and other developmentally regulated loci in *Drosophila melanogaster*. *Cell* **40**:55-62.

White, J.H., DiMartino, J.F., Anderson, R.W., Lusnak, K., Hilbert, D., and Fogel, S. 1988. A DNA sequence conferring high post-meiotic segregation frequency to heterozygous deletions in *Saccharomyces cerevisiae* is related to sequences associated with eukaryotic recombination hotspots. *Mol. Cell. Biol.* **8**:1253-1258.

Wilkie, T.M., Brinster, R.L., and Palmiter, R.D. 1986. Germline and somatic mosaicism in transgenic mice. *Dev. Biol.* **118**:9-18.

Willard, H.F., and Waye, J.S. 1987. Hierarchical order in chromosome-specific alpha satellite DNA. *Trends Genet.* **3**:192-198.

Wirth, T., Schmidt, M., Baumruker, T., and Horak, I. 1984. Evidence for mobility of a new family of mouse middle repetitive elements (LTR-IS). *Nucleic Acids Res.* **12**:3603-3610.

Wolff, R.K., Nakamura, Y., and White, R. 1988. Molecular characterisation of a spontaneously generated new allele at a VNTR locus. *Genomics* **3**:347-351.

Wolff, R.K., Plaetke, R., Jeffreys, A.J., and White, R. 1989. Unequal crossingover between homologous chromosomes is not the major mechanism involved in the generation of new alleles at VNTR loci. *Genomics* **5**:382-384.

Wong, Z., Wilson, V., Jeffreys, A.J., and Thein, S.L. 1986. Cloning a selected fragment from a human DNA 'fingerprint' : isolation of an extremely polymorphic minisatellite. *Nucleic Acids Res.* **14**:4605-4616.

Wong, Z., Wilson, V., Patel, I., Povey, S., and Jeffreys, A.J. 1987. Characterisation of a panel of highly variable minisatellites cloned from human DNA. *Am. J. Hum. Genet.* **51**:269-288.

Wooley, G.W. 1945. Misty dilution in the mouse. *J. Hered.* **36**:269-270.

Wright, S. 1934. The results of crosses between inbred strains of guinea-pigs, differing in the number of digits. *Genetics* **19**:537-551.

Wu, C-I., Lyttle, T., Wu, M-L., and Lin, G-F. 1988. Association between a satellite DNA sequence and the *responder* of *segregation distorter* in *D.melanogaster*. *Cell* **54**:179-189.

Wyman, A., and White, R. 1980. A highly polymorphic locus in human DNA. *Proc. Natl. Acad. Sci. USA* **77**:6754-6758.

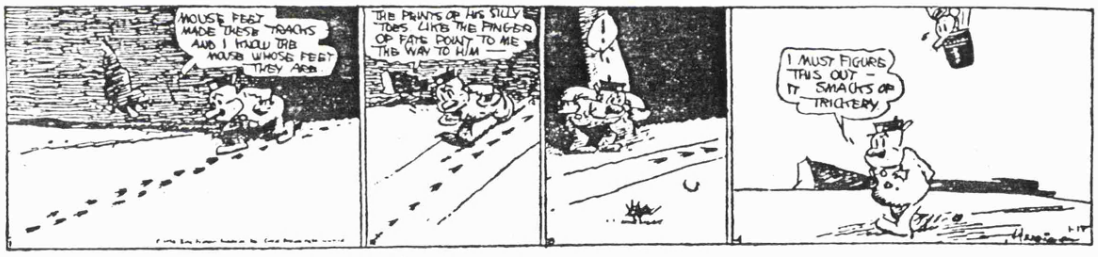
Wyman, A.R., Wolfe, L.B., and Botstein, D. 1985. Propagation of some human DNA sequences in bacteriophage λ vectors requires mutant *Esherishia coli* hosts. *Proc. Natl. Acad. Sci. USA* **82**:2880-2884.

Wyman, A.R., Mulholland, J. and Botstein, D. 1986. Oligonucleotide repeats involved in the highly polymorphic locus D14S1. *Am. J. Hum. Genet.* **39**:A226.

Yang, R.C-A., Lis, J., and Wu, R. 1979. Elution of DNA from agarose gels after electrophoresis. *Meth. Enzymol.* **68**:176-182.

Yanis-Perron, C., Vieira, J., and Messing, J. 1985. Improved M13 phage cloning vectors and host strains : nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33**:102-119.

Krazy Kat George Herriman



HYPERVARIABLE MINISATELLITES IN MOUSE DNA

Robert Kelly

Hypervariable minisatellite loci provide highly informative genetic markers in mammalian genomes. Hybridisation probes based on a G-rich core sequence simultaneously detect many minisatellite loci in human DNA, generating individual specific and highly variable DNA fingerprints with a wide range of applications. Human minisatellite probes cross-hybridise to mouse DNA, generating DNA fingerprints as complex and variable as those of man.

Inbred strains of mice have strain-specific DNA fingerprints which can be analysed using recombinant inbred strains. Analysis of the BXD recombinant inbred series using human minisatellite probes 33.6 and 33.15 revealed that the cross-hybridising loci show variation in germline stability; one locus in particular, *Ms6-hm*, detected by probe 33.6, exhibited multiallelism across the BXD strains, and heterozygosity among contemporary C57BL/6J inbred mice. *Ms6-hm* alleles were cloned from C57BL/6J DNA by cross-hybridisation to probe 33.6; on propagation in *E.coli* the majority of minisatellite repeat units were lost. DNA sequence analysis revealed that *Ms6-hm* consists of a homogeneous array of the repeat unit GGGCA which has evolved by amplification from within a member of the MT (mouse transcript) family of interspersed repetitive elements. *Ms6-hm* is flanked by two additional, diverged, MT elements, and there is further evidence that MT elements may be associated with other unstable regions of the mouse genome.

Multiallelism and heterozygosity at *Ms6-hm* (which maps near *brown* on chromosome 4) result from a high germline mutation rate to new length alleles (2.5% per gamete). Mice mosaic for cells carrying common non-parental *Ms6-hm* alleles in somatic tissue, and in some cases also in the germline, provide evidence for additional, early developmental, mutation events at this locus. Such somatic mutant *Ms6-hm* alleles provide innocuous and informative markers with which to analyse the lineage relationships of cells in early mouse development. In four mosaic mice the fraction of cells containing the non-parental allele has been shown to be indistinguishable in different adult tissues, suggesting that the mutation events preceded the allocation of the somatic lineages, and that the same pool of primitive ectoderm cells contributes equally to all somatic tissues. Under low-stringency hybridisation conditions the minisatellite repeat array of *Ms6-hm* cross-hybridises to other unstable minisatellite loci in the mouse genome to generate a novel and highly individual specific DNA fingerprint; at least one cross-hybridising locus is also somatically unstable during early mouse development.