THE ISOLATION AND CHARACTERISATION OF CANINE MINISATELLITE DNA SEQUENCES

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

by

Shirin Susan Joseph
Department of Biochemistry
University of Leicester

January 1994

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

THE ISOLATION AND CHARACTERIZATION OF CANINE MINISATELLITE DNA SEQUENCES

Shirin Susan Joseph

ABSTRACT

A Charomid ordered array library containing the 2-16kb size fraction of *Mbo*I - digested canine genomic DNA was screened with the multilocus probes, 33.6 and 33.15. Testing for polymorphism of the minisatellite loci in 48 resulting positive clones yielded seven polymorphic minisatellites with heterozygosities in the range 20-88%.

Mendelian inheritance was confirmed shown for two of the polymorphic minisatellites. DNA fingerprinting studies of the level of inter- and intra-breed variation did not show any significant difference between the two. Analysis of intra-breed variation in Bedlington Terriers using two polymorphic minisatellites as single-locus probes revealed a significant reduction in the number of alleles in comparison with an agglomerated population sample, consistent with the high level of inbreeding within this breed. Multi-locus canine minisatellite probe analysis of unrelated species showed that related repeat sequences are present in numerous species. Use of single-locus canine minisatellite probes to analyse related canids suggested that polymorphic canine minisatellites are likely to show transience in their variability and detection, whereas monomorphic minisatellites are stable and more readily detected in related canids. Use of c*Cfa*MP5, the most polymorphic canine minisatellite isolated to date, as a single-locus probe in paternity analysis demonstrates its applicability to forensic problems.

Flanking sequence and partial repeat sequence data obtained for the minisatellite in c*Cfa*MP5. The variable region in this minisatellite region is similar to many human minisatellites which show a distinct purine or pyrimidine strand bias. A mechanism whereby this minisatellite might have evolved is suggestedon the basis of the distribution and kinds of repeat units. An initial MVR-PCR analysis of *Cfa*MP5 has been carried out and, with future optimization, it should be possible to digitally type canine minisatellite alleles, thereby widening the scope of the analysis of canine minisatellite variation.

To my parents, Joseph and Joyce K. Varughese

ACKNOWLEDGEMENTS

I would like to thank Dr. Jeff Sampson for his great help, advice, support, encouragement and understanding over the years. I would also like to thank the numerous folk that have provided me with a variety of technical help, advice, encouragement, constructive criticism and kindness over the years: Sismindari Sujadi, Mike Bruford, Olivier Hanotte, John Armour, Cathryn Mellersh, Mark Gibbs, Prof. Robert Smith, Sheikh Oumar and Iris van Pijlen.

I am deeply grateful to Peter and Moira Chinnery for providing me with a 'home away from home' and priceless hospitality, care and kindness. I am also deeply grateful to Clive and Rachel Surrell for continuous support and help in all kinds of ways over the years. My thanks also go to Kirupa Sathasivam, Vanessa Baguley, Judith Chinnery, Ruth Orr, Jenny Andrews and Jill Silverthorne for their friendship.

This work would not have been possible without the continuous support and love of my parents, Joseph and Joyce K. Varughese. Finally, I would like to quote my favourite Psalm from the Bible :

"The Lord is my shepherd; I shall not want. He maketh me to lie down in green pastures: he leadeth me beside the still waters.

He restoreth my soul: he leadeth me in the paths of righteousness for his name's sake.

Yea, though I walk through the valley of the shadow of death, I will fear no evil: for thou art with me; thy rod and thy staff they comfort me.

Thou preparest a table before me in the presence of mine enemies: thou anointest my head with oil; my cup runneth over.

Surely goodness and mercy shall follow me all the days of my life: and I will dwell in the house of the Lord for ever."

ABBREVIATIONS

ATP - adenosine triphosphate
AT-rich - rich in adenosine and thymidine residues
bp - base pair
BSA - bovine serum albumin
°C - degrees centigrade
$CaCl_2$ - calcium chloride
cm - centimetre
cpm - counts per minute
CsCl - caesium chloride
$CsSO_4$ - caesium sulphate
dH_2O - distilled water
dATP - deoxy adenosine triphosphate
dCTP - deoxy cytosine triphosphate
dGTP - deoxy guanosine triphosphate
dTTP - deoxy thymidine triphosphate
ddATP - dideoxy adenosine triphosphate
ddCTP - dieoxy cytosine triphosphate
ddGTP - dideoxy guanosine triphosphate
ddTTP - dideoxy thymidine triphosphate
DNA - deoxyribonucleic acid
ds - double-stranded
DTT - dithiothreitol
EDTA - ethylenediaminetetraacetic acid
g - centrifugal force
GC-rich - rich in guanosine and cytosine residues
HCl - hydrochloric acid
HEPES - N-2-hydroxyethylpiperazine-N-ethanesulfonic acid
hrs - hours
hr - hour
IMS - industrial methylated spirit
IPTG - isopropylthio-β-D-galactoside
kb - kilobase
LMP - low melting point
λ - lambda bacteriophage
l - litre
mins - minutes
μg - microgram
mg - milligram
ml - millilitre
μl - microlitre
mm - millimetre
mM - millimolar
M - molar
mCi - milliCurie
$MgCl_2$ - magnesium chloride
$MgSO_4$ - magnesium sulphate

NaCl - sodium chloride
NaOH - sodium hydroxide
Na - sodium
nm - nanometre
ng - nanogram
MVR-PCR - minisatellite variant repeat mapping by polymerase chain reaction
OD_x - optical density at wavelength = x nanometres
OLB - oligonucleotide labelling buffer
PCR - polymerase chain reaction
PEG - polyethylene glycol (6000)
PIC - polymorphism information content
pmol - pico mole
K - potassium
/ - per
RNase - ribonuclease
RFLP - restriction fragment length polymorphism
rpm - revolutions per minute
2X - twice
SDS - sodium dodecyl sulphate
sec - second
ss - single-stranded
TCA - trichloroacetic acid
TBE - tris-borate-EDTA
TEMED - N,N,N',N' tetramethyl-ethylene-diamine
Tris.HCl - Tris[hydroxymethyl]aminomethane (brought to correct pH using hydrochloric acid)
TY - tryptone-yeast
u.v. - ultra violet
V - volts
x-gal - 5-bromo-4-chloro-3-indoyl-β-D-galactoside

CONTENTS

CHAPTER 3 - GENERATION AND ANALYSIS OF A CANINE GENOMIC LIBRARY CONTAINING CLONED MINISATELLITE SEQUENCES

FIGURES

TABLES

Chapter 3

Chapter 4

Chapter 5

5.1 PCR conditions used for MVR-PCR analysis of c*Cfa*MP5

CHAPTER 1

INTRODUCTION

1.1 Minisatellite Biology

1.1.1 The Main Features of Minisatellites

Minisatellite DNA sequences consist of arrays containing tandemly repeated units ranging in size from 2bp to 250bp (Brereton *et al.*, 1993; Paulsson *et al.*, 1992). An individual locus can vary between individuals in its repeat copy number (Jeffreys *et al.*, 1985a) and the level of variation shown by these loci can exceed heterozygosities of 99% (Brereton *et al.*, 1993). Many polymorphic minisatellites have been shown to be GC-rich (Jeffreys *et al.*, 1985a; Jacobson *et al.*, 1992; Gyllensten *et al.*, 1989) and yet numerous AT-rich polymorphic minisatellites have also been reported (Simmler *et al.*, 1987; Berg & Olaisen, 1993; Desmarais *et al.*, 1993). Several minisatellites have also been found to display strand asymmetry i.e. a purine- or pyrimidine-rich strand bias (Brereton *et al.*, 1993; Jacobson *et al.*, 1992). Furthermore, the level of variation at minisatellite loci has been shown to extend to the sequence of individual repeat units within an array (Jeffreys *et al.*, 1990, 1991a) (see Section 1.1.2). Royle *et al.*, (1988) have shown that human minisatellites are most frequently found towards the ends of chromosomes, but this is not the case for mice (Jeffreys *et al.*, 1987) and cattle (Georges *et al.*, 1990) where there appears to be no significant terminal clustering.

1.1.2 The History of Minisatellites

The first minisatellite was found quite fortuitously after screening a human genomic library of random genomic segments (Wyman and White, 1980; Wyman *et al.*, 1986). This was followed by the finding of other minisatellites flanking the human α-globin genes (Higgs *et al.*, 1986), the Ha-ras gene (Capon *et al.*, 1983) and at the human insulin gene (Bell *et al.*, 1982).

The finding of a short minisatellite consisting of a 33bp repeating monomer in an intron of the human myoglobin gene by Weller *et al.*, (1984), was followed by the detection of multiple polymorphic loci by a

pure repeat probe prepared from the myoglobin minisatellite (Jeffreys *et al* 1985a). The resulting individual-specific DNA banding pattern was termed a DNA 'fingerprint'. The generation and sequencing of λ clones containing human minisatellite regions showed that individual repeats shared a core region with the almost invariant sequence, GGGCAGGAXG. Thus, these minisatellites belonged to a subset of human minisatellites all of which showed similarities to the myoglobin minisatellite and were GC-rich.

The generation of a DNA fingerprint was a turning point in minisatellite biology and has led to their application in numerous fields of biology as well as for public use (see Sections 1.1.3 & 1.3). The cloning of individual human minisatellites however has been followed by the rapid isolation of minisatellites in a whole range of other species: Indian peafowl (Hanotte *et al.*, 1991), mice (Kelly *et al.*, 1989), chickens (Bruford *et al.*, 1993), whales (Van Pijlen, I. - manuscript submitted), geese (Rowe, G. - personal communication), Drosophila (Jacobson *et al.*, 1992) and salmon (Bentzen & Wright, 1992).

The demonstration of the use of PCR amplification of individual minisatellites in forensic analyses (Jeffreys *et al.*, 1988) was followed by its use in the detection of a novel source of variation at minisatellite loci (Jeffreys *et al.*, 1990) i.e. internal sequence variation of repeats. This technique (MVR mapping) involved the amplification of a minisatellite locus (clone MS32 at D1S8) using primers containing an *Eco*RI site, followed by partial restriction digestion of the amplified product to produce a continuous ladder of labelled DNA fragments from which the number of repeat units could be determined. MVR mapping was subsequently simplified further to generate a similar ladder (Jeffreys *et al.*, 1991a), only here the primers used were as follows: an anchor primer complementary to unique flanking sequence together with either of two primers each specific to one of the two variant repeat units at D1S8. Internal priming of the primers specific to the repeat units was prevented by attaching a 20-nucleotide 5' extension (TAG) to each variant repeat-specific primer as well as carrying out amplifications at low concentrations of either of the repeat-specific primers and high concentrations of the anchor and the TAG sequence itself. This results in products which vary in length depending on the relative positions of a specific variant repeat unit within the tandem array, enabling the positions of each variant repeat unit to be known.

1.1.3 The Major Applications of Minisatellites

The isolation and sequencing of a number of human minisatellite sequences led to the use of these sequences as multilocus probes to produce the well renown 'DNA fingerprint' (Jeffreys *et al.*, 1985a). As described above, the technique uses a single-copy minisatellite sequence as a DNA probe at low stringency enabling the detection of numerous related sequences which may or may not be homologous. The individual-specific nature of DNA fingerprints has enabled their application in a variety of fields : forensic analysis (Gill *et al.*, 1985), paternity analysis (Hermans *et al.*, 1991), population and ecological studies (Wetton *et al.*, 1987; Wetton *et al.*, 1992; Haig *et al.*, 1993), segregation analysis (Jeffreys *et al.*, 1986), immigration analysis (Jeffreys *et al.*, 1985), analysis of inbreeding (Kuhnlein *et al.*, 1990) and linkage analysis (White *et al.*, 1985a).

The use of minisatellite DNA as single-locus probes at high stringency has also paved the way for their application in cross-species hybridization analysis (Hanotte *et al.*, 1992a), mutational analyses at minisatellite loci (Kelly *et al.*, 1989, 1991; Gibbs *et al.*, 1993), forensic analysis (Jeffreys *et al.*, 1991b and references therein) and linkage analysis (Malcolm *et al.*, 1991). Furthermore, PCR amplification of individual minisatellite loci has enabled the typing of individual cells and co-amplification of several minisatellite loci results in the generation of DNA fingerprints from nanogram quantities of DNA (Jeffreys *et al.*, 1988). Finally, MVR-PCR, a technique which results in a ladder of DNA bands each corresponding to consecutive repeat units within a repeat array, has enabled the analysis of the internal structures of minisatellite alleles (Jeffreys *et al.*, 1991a) (see also Sections 1.1.2 and 1.1.6). The applications of this have been shown to be useful in forensic analysis and individual identification (Jeffreys *et al.*, 1991a) as well as differentiating between true and pseudo-homozygotes (Monckton & Jeffreys, 1991).

1.1.4 The Function(s) of Minisatellites

Many minisatellites, although found to be associated with genes, are not transcribed and therefore no overall specific *in vivo* function has been attributed to these loci: However, cases of minisatellites existing within coding regions include the hypervariable MUC1 locus, which

encodes a highly polymorphic mucin (Swallow *et al.*, 1987), the Balbiani ring 1 gene of *Chironomus tentans* (Paulsson *et al.*, 1992) and a gene coding for a breast-cancer-associated antigen (Tsarfaty *et al.*, 1990). In the latter two cases, the size of the repeat regions are conserved; the size of the repeat region in the Balbiani ring 1 gene varies by not more than 10% whereas that in the gene coding for a breast-cancer-associated antigen is completely conserved. Minisatellites have also been shown in a number of cases to bind nuclear proteins and binding factors thereby implying a possible role in gene regulation (Treppichio & Krontiris, 1992, 1993). Sequence similarities found by Jeffreys *et al.*, (1985a) between the core consensus sequence of a subset of human minisatellites and the *E.coli* generalised recombination signal (GCTGGTGG), led to speculation that the core sequence may play a role in the generation of such sequences. This has led to the search for DNA-binding proteins involved in the mutational processes within minisatellites, for which the core sequences might act as recognition sites. To date, four minisatellite-specific binding proteins have been detected (Collick & Jeffreys, 1990, Wahls *et al.*, 1991, Yamazaki *et al.*, 1992).

1.1.5 Proposed Mechanisms of the Generation of Variation at Minisatellite Loci

The mechanisms that are currently viewed to play probable roles in the origin and the generation of variation at minisatellite loci are:

(i) inter-allelic unequal exchange at meiosis and mitosis
(ii) sister chromatid exchange
(iii) gene conversion
(iv) DNA slippage during replication (Jeffreys *et al.*, 1990).

Initial studies of the internal structures of parent and mutant alleles by Wolff *et al.*, (1988) showed no exchange of flanking markers consistent with sister chromatid exchange and DNA slippage events being active. Further studies by Wolff *et al.*, (1989) of the flanking markers at the minisatellite locus, D1S7, has supported this initial finding indicating that unequal exchange requiring recombination between sequences flanking minisatellites is not the major mechanism whereby variation at minisatellite loci is generated. This was rapidly followed by the studies of

Jeffreys *et al.*, (1990) of the internal structures of minisatellite alleles using MVR-mapping which further supported the findings of Wolff *et al.*, (1989). However, a few cases of unequal exchange between paternal alleles have subsequently been detected (Jeffreys *et al.*, 1991a), indicating that such events do occur. Recently, Madsen *et al.*, (1993) have been able to show via a series of *in vitro* primer extension reactions that deletion events at a mitochondrial repeat domain can be mainly attributed to DNA replication errors, thereby further supporting preceding evidence (see above). In a similar vein, Strand *et al.*, (1993) have found that mutations in genes involved in DNA mismatch repair lead to 100- to 700-fold increases in the instability of simple repetitive DNA tracts. Therefore, current evidence seems to point towards the predominance of slippage and intra-allelic unequal exchange events over inter-allelic unequal exchanges in the generation of variation at minisatellite loci.

1.1.6 Mutations at Minisatellite Loci

Mutations at minisatellite loci have been studied in various mammals such as humans (Jeffreys *et al.*, 1988, 1990, 1991b; Armour *et al.*, 1989b), mice (Kelly *et al.*, 1989, 1991) and cattle (Georges *et al.*, 1990). Both germline and somatic events have been detected and the mutation rates generally increase with the level of heterozygosity seen at a given locus (Armour *et al.*, 1989b). The germline mutation rate at a human minisatellite locus has been estimated to be approximately ~ 0.0014/gamete (Jeffreys *et al.*, 1990), whereas the corresponding rate at two murine minisatellite loci is 0.025-0.035/gamete (germline) (Kelly *et al.*, 1991; Gibbs *et al.*, 1993). Somatic mutations at the mouse minisatellite loci, *Ms6-hm* and *Hm-2*, have been shown to be substantially confined to the first few cell divisions (Kelly *et al.*, 1989; Gibbs *et al.*, 1993). Whether similar biases will be found for other minisatellite loci remains to be seen. Mutation rates vary substantially from locus to locus, rising to 0.102/gamete for the most unstable locus, *Hm-2* (Gibbs *et al.*, 1993).

Mutations at minisatellite loci have been shown to be a result of a change in the length of the repeat array since mutant alleles are detected when restriction enzymes are used which only cleave outside the array. Most length changes at minisatellite loci involve only a few repeat units although very large changes have also been observed (Armour *et al.*, 1989b; Kelly *et al.*, 1991). The length change of the repeat array in the

majority of mutant alleles studied (78) has not been attributed to the exchange of flanking markers (Jeffreys *et al.*, 1990; Kelly *et al.*, 1991), although a few examples do exist (2) (Jeffreys *et al.*, 1991a).

For both humans (Jeffreys *et al.*, 1988) and mice (Kelly *et al.*, 1991) mutations arise sporadically and show no clustering within families. However, this seems not to be the case for cattle where there is evidence for 'bursts' in the germline mutation rate of individual animals (Georges *et al.*, 1990). Although Kelly *et al.*, (1991) have reported that in all cases where the parental origin of a mutant allele could be specified, these were all paternal in origin, no such bias has been seen in other studies by Jeffreys *et al.*, (1991) and Gibbs *et al.*, (1993). High mutation rates at minisatellite loci requires the measurement of mutation rates at these loci prior to their use in studies such as paternity analysis (where incorrect assignment of parentage could occur as a result). Finally, analysis of internal repeat sequence variation at the human minisatellite, MS32, by Jeffreys *et al.*, (1990, 1991a) has shown that most mutation events are clustered at one end of alleles, suggesting the presence of a localised mutational 'hot-spot'.

1.2 Other Types of Tandemly Repeated Arrays

There are four main types of tandemly repeated DNA sequences which have been identified in eukaryotes. They are (in descending order of size):

(i) satellites,
(ii) midisatellites
(iii) minisatellites
(iv) microsatellites.

Satellites were initially classed as sequences which could be separated from most of the eukaryotic genome using isopycnic centrifugation in gradients of CsCl and $CsSO_4$. However some arrays resist separation and remain instead with the main density fraction of genomic DNA (Singer, 1982). Repeat units vary from 2bp in the crab AT satellite to 1408bp in the calf satellite (John & Miklos, 1979). Characteristics of many satellites (though not all) include: (i) association with heterochromatin (ii) lack of measurable transcription (iii) replication late in S-phase, and (iv) underreplication in polytene chromosomes (Singer, 1982).

A polymorphic locus near the telomere of the short arm of chromosome 1 was termed as a midisatellite as it was found to be 'considerably larger' than minisatellite loci (a tandem repeat of 40bp in an array of 250-500kb) (Nakamura *et al.*, 1987). Page *et al.*, (1987) have also described a midisatellite in the pseuodoautosomal region of the X and Y chromosomes with a 61bp tandem repeat in arrays of 10-50kb.

Microsatellites are stretches of DNA which consist of only one, or a few tandemly repeated nucleotides, for example, poly (dA). poly (dT) to poly (dG - dT).poly (dC - dA). The repeat length is typically less than 6bp and the array length less than 1kb (Epplen *et al.*, 1991). They show variation in the copy number of repeat units between individuals although this variation is generally lower than that seen for hypervariable minisatellites (Holmes *et al.*, 1993). Due to their small size they can be amplified by PCR and visualised on polyacrylamide gels although larger alleles can also be separated by high percentage agarose gels. Unlike minisatellites they are dispersed throughout the human as well as other genomes (Luty *et al.*, 1990) although they are not found in bacterial genomes (Gross & Garard, 1986). However, they are similar to minisatellites in that they can act as single-locus markers, the locus specificity being determined by their flanking sequences (Weber, 1990).

1.3 Canine minisatellites and microsatellites

Jeffreys & Morton (1987) generated canine *Hinf* I DNA fingerprints using the human minisatellite probes, 33.6 and 33.15 (Jeffreys *et al.*, 1985a). This analysis showed that a complex pattern of canine DNA fragments could be produced. The probe 33.6 detected a pattern which was similar in complexity to its human counterpart, whereas the largest DNA fragments detected by 33.15 were fainter than those detected in human DNA fingerprints. Furthermore, canine DNA fingerprints detected by 33.6 were distinct from those generated using 33.15, although a number of loci were seen to cross-hybridize (23%). This is in contrast to human DNA fingerprints where few if any fragments cross-hybridize. The number of minisatellite fragments larger than 6kb detected in canine DNA fingerprints was less than those for humans and a number of large homozygous canine DNA fragments were observed which are seldom seen in humans. The comparison of the DNA fingerprints of two Whippets showed that intra-breed variability was not significantly less

than inter-breed variability. However, a subsequent review of DNA fingerprinting in animal populations by Hill (1987) has suggested that a more detailed analysis of the canine population would be likely to display substructuring. The overall level of band sharing is ~ 46% whereas that in human DNA fingerprints is ~ 22%, reflecting a lower variability for canine DNA fingerprints. As in humans, canine DNA fingerprints show germ-line stability, although in contrast to human DNA fingerprints, observed instances of allelism between DNA fragments reduce the total number of distinct and recombinationally separable hypervariable loci to 13. All segregating DNA fragments were seen to be autosomal i.e. none were specifically transmitted to sons or daughters. There was no evidence of clustering in the genome, indicating that canine minisatellites are likely to be dispersed over canine autosomes.

DNA fingerprinting has now been commercially used in the identification of Greyhounds used for racing (Jeff Sampson - personal communication) and paternity analysis (Georges *et al.*, 1989, Hermans *et al.*, 1991). It has also been used in a number of population and ecological studies in related canids i.e. wolves (Wayne *et al.*, 1991a) and foxes (Gilbert *et al.*, 1990; Wayne *et al.*, 1991b).

The isolation and characterization of canine microsatellites with a view to using them in linkage analysis is now underway. Holmes *et al.*, (1993) have reported the isolation and characterization of 10 canine microsatellite loci containing $[dC\text{-}dA]_n.[dG\text{-}dT]$ repeats. The microsatellites varied in length from 13 to 23 repeat units. The overall observed heterozygosities were lower than that seen for polymorphic human minisatellites. Of the ten microsatellites isolated, nine have PIC (polymorphism information content - see Botstein *et al.*, 1980) values which are sufficiently high for their use in linkage analysis. A comparison of intra-breed versus inter-breed variation using several of the microsatellites indicates that there is no significant difference. However, they do accede to the view that 'different pedigree dog populations will have varying degrees of inbreeding' thereby affecting the individual informativeness of a given microsatellite locus depending on the breed being analysed. Finally, they have also shown that microsatellites display Mendelian inheritance in a family of Irish Setters.

1.4 The Implications of Minisatellite Biology on Studies of the Canine Genome

It can be seen that those applications of human minisatellites described in Section 1.1.3 for various analyses are also applicable to other species, providing it is shown that similar repeat sequences exist in these species. The first way in which one might show this for dogs, is by the generation of a DNA fingerprint using numerous minisatellite probes that are currently available for research use e.g. the human minisatellite loci, 33.6 and 33.15 (Jeffreys *et al.*, 1985a). This has already been done, initially by Jeffreys & Morton (1987), and subsequently by various groups involved in canine paternity analyses (Georges *et al.*, 1989, Hermans *et al.*, 1991).

One major application of minisatellite biology to the canine genome that requires some attention is that of linkage analysis, due to its significant importance in studies of the segregation of genetically inherited diseases in dogs. The study of the cosegregation of a specific informative genetic marker with a locus of interest within pedigrees, has been termed 'linkage analysis.' Such studies have led to the localization of disease loci such as the cystic fibrosis gene (White *et al.*, 1985b). The generation of a linkage map of genetic markers for the canine genome therefore would be invaluable in future studies of the segregation of disease loci in pedigrees. As stated above, linkage analysis requires the use of informative genetic markers. By this, it is meant that these markers must display polymorphism. There are various DNA sequences which have been identified and shown to be polymorphic between individuals: RFLPs (restriction fragment length polymorphisms), minisatellites and microsatellites. The informativeness of RFLPs is limited in that at most they can only be informative in 50% of cases (due to the presence of absence of a restriction enzyme site) (Jeffreys *et al.*, 1985a). However, the variation shown by minisatellites and microsatellites is due to differences in the copy numbers of repeat units (see above - Sections 1.1.1 and 1.3). Such loci can show very high levels of polymorphism, although in practice the variation shown by microsatellites is generally lower than that for minisatellites (Amos & Pemberton, 1992). Thus, both minisatellites and microsatellites show potential for being highly useful genetic markers for linkage analysis.

1.5 Canine Diseases

Many canine diseases are known to be genetically inherited, but due to the paucity of genetic information regarding their mode of inheritance, they are not well characterized e.g. hip dysplasia and epilepsy. There are however several for which pedigree analysis has been done, e.g. achondroplasia, pituitary dwarfism, progressive retinal atrophy and elbow dysplasia. Many of the latter have been shown to follow an autosomal recessive mode of inheritance (Willis, 1989). Furthermore, a number of canine diseases have human equivalents e.g. hip dysplasia, elbow dysplasia and epilepsy.

The use of DNA markers in the localization of the genes controlling such diseases would be of great importance in selective breeding systems where dogs are tested soon after birth for a tendency towards a particular condition. It is therefore important to have access to disease pedigrees thereby enabling the use of a bank of polymorphic minisatellite DNA probes (or microsatellites) to study the cosegregation of a given marker with a disease locus. It should be noted that such linkage analysis will only be applicable to diseases where the mode of inheritance is controlled by a single gene. It is possible however, that analysis of some pedigrees will result in the linkage of a marker with a gene specific to the pedigree. In such cases one can only conclude that the condition is controlled by a polygenic mode of inheritance. Therefore, providing one is dealing with a disease which is controlled by a single gene, the detection of DNA markers linked to the gene can eventually lead to the genetic characterization of the condition, enabling early diagnosis and possibly even gene therapy. The implications of such testing are not of small significance commercially at least, where dog breeders can thereby produce lines which are no longer susceptible to such disease.

1.6 Aims in the Present Study; Outline of Strategies involved

For the reasons stated above, the aim of this project was to isolate and characterize polymorphic canine minisatellite sequences. The method used to isolate these loci is based on the work of Armour *et al.*, (1990b). Although individual minisatellites have been isolated by various methods such as cloning into: (i) λ (Wong *et al.*, 1986, 1987) and (ii) cosmid vectors (Nakamura *et al.*, 1987, 1988). Such procedures have specific

disadvantages, for example, cloning into λ allows only a subset of minisatellites to be cloned, whereas cloning into cosmid vectors requires the screening of large libraries and also does not select for larger and more variable minisatellites (Armour *et al.*, 1990b).

The strategies involved in the isolation of polymorphic minisatellites have been described subsequently in detail by Bruford *et al.*, (1992) and those specific to the isolation of canine polymorphic minisatellites are described in detail in Chapter 3. Therefore as an introduction, only an outline of these strategies will be presented here (see Figure 1.1 for an overview).

The generation of a DNA fingerprint for a given species provides the following information: (i) whether minisatellites exist within the genome and, if present, (ii) the size range within which the most variable minisatellite DNA fragments are found. Since the ultimate aim was to isolate the most polymorphic minisatellite loci, the generation of a DNA fingerprint indicates the size range of genomic DNA that should be isolated prior to cloning. The selected size fraction, after several rounds of purification, can then be cloned into a vector such as Charomid 9, of which there are numerous variants, depending on the number of 2kb spacer fragments present (see Figure 1.2) e.g. Charomid 9-36 has 15 spacer fragments which results in an overall size of 36kb.

The cloning of a specific size range of DNA fragments is achieved both through the initial stage of purification and through packaging *in vitro* of recombinant Charomid molecules after ligation. The principle behind this is that the packaging system of λ can only accept inserts ranging from 38-52kb. Thus, a 36kb-sized Charomid molecule will only be packaged *in vitro* if it has ligated to insert DNA in the 2-16kb size range. After packaging *in vitro* the resulting λ particles are used to infect NM554 bacteria (ampicillin sensitive and recombination deficient). The presence of a selectable marker in the Charomid vector (ampicillin resistant) enables subsequent isolation of only those bacterial colonies where infection has been successful. Prior to the storage of the resulting library on microtitre plates, it is advisable to estimate the number of recombinants within the library, thereby determining its worth. Providing a sufficient number of recombinants are known to be present, the establishment of the library in an ordered array i.e. on microtitre plates, means that it can be consecutively screened using multiple minisatellite probes e.g. 33.6, 33.15 (Jeffreys *et al.*, 1985a), 3'α-globin (Fowler *et al.*, 1988)

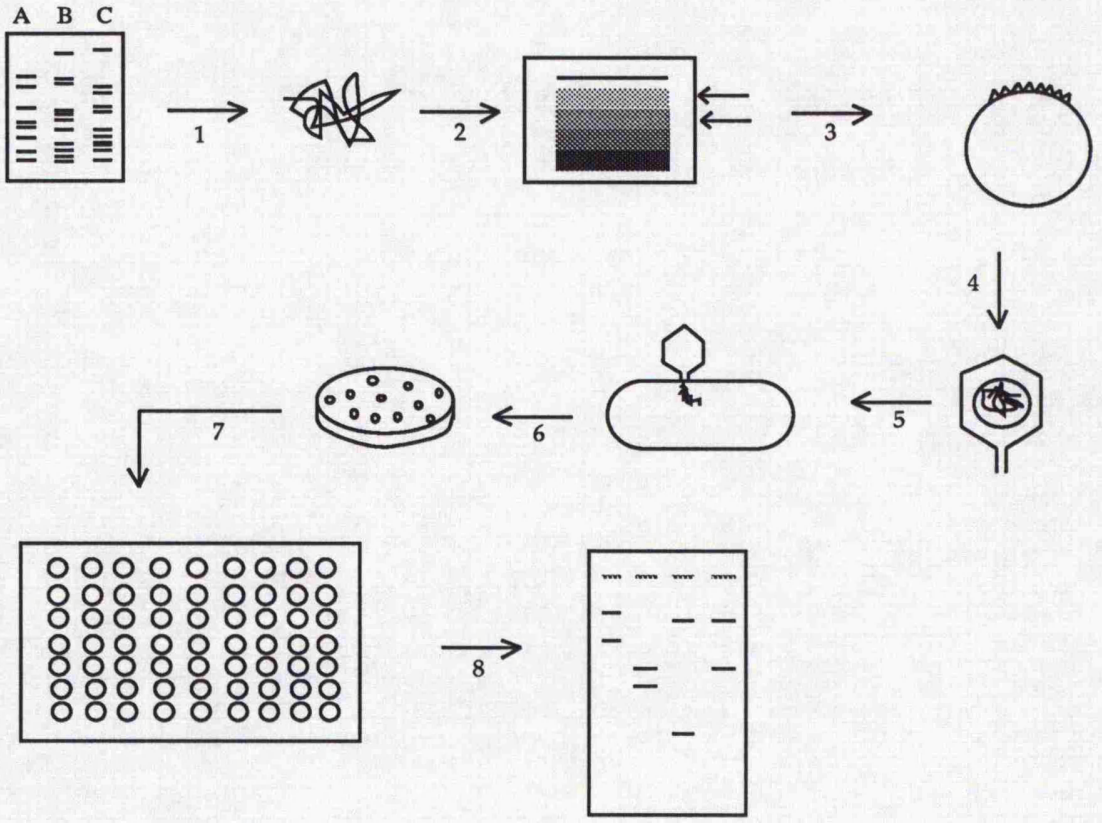
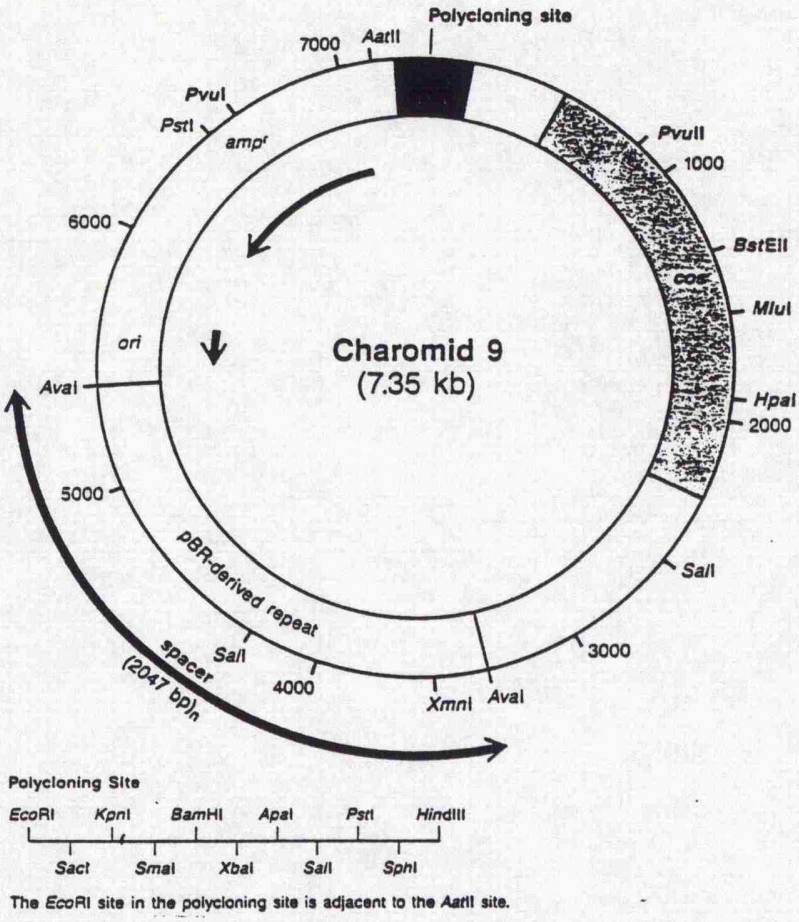


Figure 1.1: Generation of an Ordered Array Canine Genomic Library 1 - after determination of the genomic size range containing the most variable DNA fragments in a canine DNA fingerprint, purified genomic DNA pooled from 10-20 dogs, is digested to completion with *Mbo* I. 2 - The genomic size fraction is purified by preparative gel electrophoresis. 3 - The selected genomic DNA is ligated into the *Bam* HI site of the Charomid 9 vector. 4 - The ligation is packaged in vitro. 5 - The resulting λ particles are used to infect NM554 bacteria (ampicillin sensitive, reccombination deficient). 6 - Colonies containing the Charomid are selected for using ampicillin. 7 - after determining the number of recombinants in the library individual colonies are picked into the wells of microtitre plates. 8 - After screening the library consecutively with a number of minisatellite probes at low stringency, the DNA inserts from any resulting positive clones are used as single-locus probes at high stringency to probe a Southern blot containing canine genomic DNA from 4-6 individuals digested to completion. Individuals that show one positively hybridizing band are homozygous at the given locus (uninformative) whereas individuals showing two positively hybridizing bands, (representing two different sized alleles) are heterozygous and informative.



The *Eco*RI site in the polycloning site is adjacent to the *Aat*II site.

Figure 1.2 The Charomid 9 Vector

and M13 (Vassart *et al.*, 1987). Isolation of canine DNA inserts from positive clones can then be followed by their use as single-locus probes at high stringency to detect individual minisatellite loci in canine genomic DNA. Categorization of the presence or absence of polymorphism, depends on (i) the number of different alleles seen and (ii) the number of individuals within the population that are heterozygous.

CHAPTER 2

METHODS AND MATERIALS

2. General Molecular Biological Methods

2.1 Agarose Gel Electrophoresis

0.5 - 1.0% Seakem agarose gels in 1X TBE (89mM Tris.HCl, 89mM Boric acid, 2mM EDTA pH 8.0) were used unless stated otherwise, containing 0.5µg/ml ethidium bromide. Electrophoresis was normally carried out at 50 - 100V for 1-2hrs. 1/5vol of 6X loading buffer (0.25% bromophenol blue, 0.25% xylene cyanol, 15% Ficoll type 400 in dH_2O) was added to each DNA sample prior to loading. After the run, the DNA was visualised using u.v. light from a Fluo-link transilluminator. Photographs of gels were taken using a Land Polaroid MP-4 camera and Kodak T-max 100 film.

2.2 Digestion of DNA with Restriction Endonucleases

Genomic DNA : For single- and multi-locus DNA fingerprinting, 5.5µg of DNA were digested to completion with a four-base recognition restriction endonuclease (*Mbo*I, *Alu*I, *Hae*III or *Hinf*I)(Gibco-BRL), overnight, at 37°C, using 2 units of endonuclease/µg of DNA. The extent of digestion was checked by the electrophoresis of an aliquot (0.5µg) from each digest on a 1% agarose gel. The amount of endonuclease required to digest 1µg of DNA to completion was determined empirically (Hoelzel, 1992).

Plasmid or Bacteriophage DNA : These DNAs were digested with a given restriction endonuclease according to the manufacturer's instructions (Gibco BRL). Complete digestion was normally achieved using a maximum of 3hrs at 37°C.

2.3 Oligolabelling

Oligolabelling was based on standard methods as described by Feinberg & Vogelstein (1984). 10ng of a given double-stranded DNA molecule was dissolved in a final volume of 15µl in sterile dH_2O. This

was denatured by boiling for 10 mins, spun briefly and immediately placed at 37°C. 5µl of OLB (Solutions A, B & C mixed in the ratio 2:5:3 where Solution A : 1.25M Tris.HCl, 125mM $MgCl_2$, 0.18% v/v 2-mercaptoethanol, 0.5mM dCTP, 0.5mM dGTP, 0.5mM dTTP; Solution B : 2M HEPES pH 6.6; Solution C : Hexanucleotides 90 OD units/ml in 3mM Tris.HCl, 0.2mM EDTA pH 7.0), 2µl of BSA (10mg/ml), 2µl of ^{32}P-dATP (10mCi/ml) and 3 units of DNA polymerase I (Klenow fragment) (Gibco - BRL) were then added. This was mixed, spun briefly and placed at 37°C for 3hrs. After 3hrs, 70µl of oligo stop solution (20mM NaCl, 20mM Tris.HCl pH 7.5, 2mM EDTA, 0.25% SDS) was added and the ^{32}P - labelled probe (labelled to a specific activity of ~ 10^9 cpm/µg DNA) boiled for 10 mins prior to the addition of the probe to the hybridization solution.

2.4 Ethanol Precipitation of DNA

Plasmid DNA : 1/10 vol of 3M Na Acetate pH 5.2 and 3 vol of absolute ethanol was added, mixed by inversion and incubated at -70°C for 30mins - 1hr. If very small amounts of DNA were known to be present, incubation was carried out at -20°C for 16hrs. After chilling, the DNA precipitate was microcentrifuged at 13000g for 30mins. The supernatant was removed, the DNA pellet rinsed in 70% ethanol, spun for 10 mins, the supernatant removed again and the pellet vacuum dried. The DNA was then dissolved in an appropriate volume of $TE_{0.1}$ (10mM Tris.HCl pH 8.0, 0.1mM EDTA) or sterile dH_2O.

Genomic DNA : 3 vol of absolute ethanol were added, mixed by inversion, and the DNA strands collected into an eppendorf tube using the looped end of a Pasteur pipette. (If at this stage, no strands of DNA were seen, this usually indicated a low yield of DNA, and required incubation of the ethanol supernatant for 16hrs at -20°C prior to collection of the DNA. Collecting the DNA involved a 30 min centrifugation at 10000g). The DNA strands were spun down briefly and rinsed in 200 - 500µl of 70% ethanol. The DNA was spun for 10 mins at 13000g, the 70% ethanol removed, the pellet vacuum dried for 3-5 mins or air dried at room temperature and dissolved in an appropriate volume of $TE_{0.1}$.

2.5 Phenol Extraction

DNA was purified from protein by adding an equal volume of Tris-saturated phenol to the aqueous solution. This was mixed according to the size of a given DNA molecule; plasmid DNA solutions were vortexed briefly whereas genomic DNA solutions were mixed by slow inversion, or swirling in a conical flask. The resulting mixture was separated by microcentrifugation at 13000g for 5mins (eppendorf tubes) or centrifugation at 10000g for 10mins (corex tubes). The aqueous phase was removed and transferred to a fresh tube. (For genomic DNA, removal of the aqueous phase was achieved using a wide-mouthed 1ml Gilson tip to minimise shearing of the DNA). An equal volume of phenol/chloroform/iso-amyl alcohol (25:24:1) was added, mixed and separated as above and the aqueous phase transferred again. Finally, an equal volume of chloroform/iso-amyl alcohol (24:1) was added, mixed and separated as above and the aqueous phase transferred to a fresh tube. The resulting DNA was ethanol precipitated and dissolved in an appropriate volume of $TE_{0.1}$.

2.6 DNA Concentration Determination

(Sambrook *et al.*, 1990)
The concentration of DNA was estimated by : 1) the electrophoresis of an aliquot of the DNA together with λ DNA concentration standards (λ DNA diluted from 500 - 31.25ng) on a 1% agarose gel 2) calculated using an OD_{260} measurement on a spectrophotometer.
1 OD unit = 50μg/ml.

2.7 Preparation of Single and Double-stranded M13 Bacteriophage DNA

2.7.1 Preparation of Competent Cells

(Sambrook *et al.*, 1990)

5ml of 2X TY (16g Bactotryptone, 10g yeast extract, 5g NaCl / l) in a Sterilin tube was inoculated with JM101 bacteria (see appendix) and grown

at 37°C overnight with shaking at 260 rpm. 10ml of 2X TY was inoculated with 100µl of the fresh overnight culture of JM101, and grown for 3hrs at 37°C with shaking at 260 rpm until the OD_{550} was ~ 0.3. This culture was centrifuged at 4000g for 5 mins and the supernatant discarded. The pellet of cells was resuspended gently (kept always on ice during resuspension) in 5ml of ice-cold 50mM $CaCl_2$. The tube was placed on ice for 20 mins. The suspension was centrifuged at 4000g for 5 mins and the supernatant discarded. The cells were resuspended in 2.5ml of ice-cold 50mM $CaCl_2$, and kept on ice.

2.7.2 Transformation of Competent Cells with M13 DNA

10ng of ss or ds M13 DNA was added to 300µl of competent cells, gently mixed and left to stand on ice for 1hr. The cells were then placed at 42°C for 3 mins and then immediately placed on ice for 2-3 mins. Following this, 200µl of the fresh or overnight culture of JM101 was added and mixed. The competent cell mixture was transferred into a 15ml Sterilin tube (kept at 55°C) containing 40µl 20mg/ml x-gal, 4µl 200mg/ml IPTG and 3ml molten top agar (Top agar :10g Bactotryptone, 8g NaCl, 8g agar /l). This was mixed by inversion and poured into a Bottom agar plate (Bottom agar :10g Bactotryptone, 8g NaCl, 12g agar /l). The top agar was allowed to set for 5 mins, the plates inverted and transferred to a 37°C oven overnight.

2.7.3 Preparation of Single-stranded M13 DNA for Sequencing

A fresh JM101 overnight culture was diluted 1:100 in 2X TY. An individual plaque was stabbed with a sterile Pasteur pipette (taking up the cylinder of agar) and blown into 1.5ml of the diluted culture in a sterile 15ml tube. This was incubated at 37°C with shaking at 260 rpm for 5hrs. The culture was microcentrifuged for 5 mins at 13000g and the supernatant transferred to a fresh tube. The pellet containing the ds M13 DNA was kept and treated as described in Methods Section 2.7.4. 200µl of 2.5M NaCl/20% PEG 6000 was added to the supernatant, mixed by inversion and left to stand on ice for 1hr. This was spun at 13000g for 10 mins, the supernatant discarded and the contents of the tube spun again at 13000g very briefly. The remainder of the supernatant was removed and the pellet dissolved in 500µl of $TE_{0.1}$. An equal volume of 2.5M NaCl/20%

PEG 6000 was added and the tube left at room temperature for 30 mins. After 30 mins, the tube was spun at 13000g for 10mins and the resulting pellet dissolved in 500µl $TE_{0.1}$. This solution was extracted with phenol, phenol/chloroform/isoamyl-alcohol (2X), and chloroform/isoamyl-alcohol. Finally the DNA was ethanol precipitated, dissolved in sterile dH_2O and stored at -20°C.

2.7.4 Preparation of Double-stranded M13 DNA

The bacterial pellet was resuspended by vortexing it in 100µl lysis buffer (50mM glucose, 50mM Tris.HCl pH 8.0, 10mM EDTA) and placed on ice for 10 mins. 200µl freshly-prepared 0.2M NaOH/ 1% SDS was added, mixed by inversion (5X) and the tube placed on ice for 5 mins. This was followed by the addition of 150µl 3M K Acetate pH 5.2, followed by mixing by vortexing upside down and placing the tube on ice for 10 mins. The resulting mixture was spun at 13000g for 10 mins and the supernatant transferred to a fresh tube. The mixture was spun again as before and the supernatant transferred once more to a fresh tube. (This repetition served to ensure the maximum removal of contaminant bacterial debris). The supernatant was extracted with phenol, phenol/chloroform/isoamyl-alcohol and chloroform/isoamylalcohol. The DNA was ethanol precipitated and dissolved in $TE_{0.1}$.

2.8 Preparation of High Quality Double-stranded M13 DNA for Sequencing

(Sambrook *et al.*, 1990)
(Note : All large volumes stated in this method were divided into 30ml corex tubes for centrifugation steps).

A 200ml bacterial culture was spun for 10 mins at 10000g in a Sorvall RC-5B (Dupont Instruments), the cell pellet resuspended in 16ml of lysis buffer (see Methods Section 2.7.4) and placed on ice for 10 mins. 32ml of 0.2M NaOH/1% SDS was then added, mixed by inversion and the tube placed on ice for 5 mins. This was followed by the addition of 24ml of 3M K Acetate pH 5.2, mixing by inversion and the tube placed on ice for 10 mins. After 10 mins, the cell debris was spun down at 10000g for 10 mins. The supernatant was removed and kept. The nucleic acid was precipitated by the addition of 71ml of isopropanol with chilling at -70°C for 1hr

followed by centrifugation for 15 mins at 10000g. The resulting pellet was dissolved in 6ml of $TE_{0.1}$, followed by the addition of 6.5ml of 4.4M LiCl, mixed, and the solution left to stand for 2hrs at room temperature. The RNA pellet was spun down for 15 mins at 10000g, and the supernatant removed and kept. 10ml of absolute ethanol was added to the supernatant to precipitate the DNA. This was incubated at -70°C for 1hr and spun at 10000g for 30 mins. The resulting pellet was washed in 70% ethanol, vacuum dried and dissolved in $TE_{0.1}$. Any remaining RNA was digested with 100µg/ml RNase A at 37°C for 15 mins and at 70°C with the further addition of 20µl of 10% SDS. An equal volume of 1.6M NaCl/ 13% PEG was then added, mixed well and placed on ice for 1hr. The DNA was recovered by spinning at 12000g for 10 mins. The supernatant was completely removed by aspiration and the pellet dissolved in 500µl of $TE_{0.1}$. This was followed by a further PEG precipitation with an incubation time of 30 mins at room temperature. The DNA solution was extracted with phenol, phenol/chloroform/isoamyl-alcohol (2X) and chloroform. The DNA was ethanol precipitated and dissolved in an appropriate volume of $TE_{0.1}$ or dH_2O.

2.9 Autoradiography

Filters were exposed to Fuji Medical x-ray film for 16hrs - 2 weeks at -70°C with or without screens depending on the number of counts present after post-hybridisation washes. Kodak diagnostic film was only used for short exposures. Sequencing gels were exposed to Fuji Medical film at room temperature.

2.10 Southern Blotting

(Southern, 1975)

The agarose gel was placed in 0.25M HCl for 15 mins with shaking (2X). The acid was removed and the gel rinsed briefly in dH_2O. This was removed and the gel soaked in denaturing solution (0.5M NaOH, 1.5M NaCl) for 15 mins with shaking (2X). This was removed, the gel rinsed briefly in dH_2O and placed in neutralising solution (0.5M TRIS.HCl pH 7.5, 1.5M NaCl) for 15 mins with shaking (2X). The gel was then placed upside down on a 3MM Whatmann paper wick soaking in 20XSSC (3M NaCl, 0.3M tri-sodium citrate). A piece of Hybond N membrane cut to fit the size

of the gel (Nfp -Amersham) was placed on top of the gel ensuring that no air bubbles were present between the gel and membrane. This was followed by two pieces of 3MM Whatmann paper soaked in 3XSSC. Cling film was then used to cover each edge of the gel. A stack of paper towels 5cm thick was placed on top, followed by a glass plate and then a 500g weight. This was left for 16hrs to ensure complete transfer of DNA. The Hybond N membrane was dried at 65°C for 10 mins, and the DNA fixed to the membrane by u.v. transillumination on a Fluo-Link, TFL-35M transilluminator for ~ 1-2 mins (the exact time was used was determined by calibrating the transilluminator).

2.11 DNA Sequencing

(Tabor & Richardson, 1987)
DNA sequencing was done using the Sequenase Version 2.0 kit from USB.

2.11.1 Annealing Template and Primer

The components of the annealing reaction (for either ss or denatured ds DNA) were as follows: 0.5 pmol primer, 0.5 - 1.0 pmol DNA, 2µl of annealing buffer (200mM Tris.HCl pH 7.5, 100mM $MgCl_2$, 250mM NaCl) made up to a final volume of 10µl with sterile dH_2O. For ss DNA, after mixing and a brief spin, the primer/template mixture was placed at 65°C for 2mins and the mixture cooled slowly to room temperature over a period of about 30 mins. However, for denatured ds DNA the heating step was found to be unnecessary as the primers readily annealed within approximately 10 mins incubation at room temperature. (Double-stranded DNA was denatured prior to sequencing by incubating 8µg DNA in a final concentration of 0.2M NaOH for 10 mins at 37°C and ethanol precipitation.

2.11.2 Labelling Reaction

Prior to setting up the labelling reaction, 2.5µl of each termination mixture was placed into eppendorf tubes and pre-warmed at 37°C (Each mixture contains 80mM dATP, 80mM dTTP, 80mM dCTP, 80mM dGTP and 50mM NaCl. In addition the "G" mixture contains 8mM ddGTP; the "A" mix, 8mM ddATP; the "T" mix, 8mM ddTTP; and the "C" mix, 8mM ddCTP). Then the following was added to the annealed template/primer

mixture: 1µl 0.1M DTT (dithiothreitol), 2µl of labelling Nucleotide Mix (for use with radiolabelled dATP) (1.5mM dGTP, 1.5mM dCTP, 1.5mM dTTP), 5mCi [α-^{35}S] dATP (Amersham) and 3 units of Sequenase enzyme with the total volume being about 15µl. This was mixed thoroughly and incubated for
1-5 mins at room temperature.

2.11.3 Termination Reactions

When the labelling reaction was complete, 3.5µl of it was transferred to each termination mixture, incubated for 2-5 mins at 37°C and the reaction stopped by the addition of 4µl of stop solution (95% formamide, 20mM EDTA, 0.05% bromophenol blue, 0.05% xylene cyanol). This was stored at -20°C until required. Prior to loading on the gel the samples were heated to 80°C for 4 mins and 5µl loaded in each lane.

2.11.4 Preparation of the Sequencing Gel and Gel Electrophoresis of the DNA

(Sambrook *et al.*, 1990)

21g of ultrapure urea (USB) was dissolved in dH_2O by gentle heating on a heating magnetic stirrer (B & T Hotspin). One of two sequencing gel plates (20 x 40 cm) was siliconised (BDH Repelcote TM) on the inner side in a fume hood. Both were then washed with water and wiped with IMS. They were taped together (vinyl tape) with the inclusion of 0.35mm thick plastic card spacers (1cm x 37cm). The plates were clamped and placed at an angle of 30° from the horizontal. The dissolved urea was added to 5ml of 10X TBE (seq) (0.9M Tris.HCl, 0.89M boric acid, 25mM EDTA) and 7.5ml of 40% acrylamide (Accugel 40TM - National Diagnostics) and made up to a final volume of 50ml. This was mixed and filtered. Before pouring the gel, 300µl 10% ammonium persulphate and 50µl TEMED were added and mixed. The gel was allowed to set for 2 - 24hrs before running. Prior to loading the DNA samples, the gel was pre-run at 1200V for 30 mins. After the run, the gel was fixed in 10% TCA/10% methanol for 5 mins, transferred to 3MM Whatmann paper, covered with cling film and then dried in a BIORAD Model 583 gel dryer. The gel was exposed to Fuji Medical film without screens at room temperature.

2.12 DNA Extraction from Canine Blood, Blood Clots and Tissue

Blood

2 vol of 1X SSC was added to 0.5ml of blood in a 1.5ml eppendorf tube and the contents mixed thoroughly. This was spun for 30 mins at 10000g. The supernatant was removed and a further 2 vol of 1X SSC added, mixed and spun again.The loose pellet was then transferred to an eppendorf tube. The remainder of the procedure was as described below.

Blood Clot

0.5 mg of blood clot was minced using a sterile scalpel blade and placed in a 1.5ml eppendorf tube. The remainder of the procedure was as described below.

Tissue

0.1mg of tissue was minced using a sterile scalpel blade and placed in a 1.5ml eppendorf tube. The remainder of the procedure was as described below.

1ml of extraction buffer (150mM NaCl, 100mM EDTA pH 8.0, 1%SDS, 0.5mg/ml proteinase K - Gibco-BRL) was added to the tube and placed in a 55°C water bath for 3hrs or for 16hrs at 42°C on a rotatory platform. The mixture was extracted with phenol, phenol/chloroform/isoamyl-alcohol and chloroform/isoamyl-alcohol. The resulting solution was ethanol precipitated, dissolved in an appropriate volume of $TE_{0.1}$ and stored at -20°C.

2.13. Generation of Individual-specific Canine DNA Fingerprints

2.13.1 Digestion of Canine Genomic DNA

5.5µg of genomic DNA/individual sample was digested to completion with a four base recognition restriction endonuclease (see Methods Section 2.2).

2.13.2 Agarose Gel Electrophoresis of Digested DNA

5µg of the digested DNA was run on a (23.6 x 20 cm) 0.8% agarose gel (Sigma type I, low EEO) in 1X TBE (see Methods Section 2.1). λ Hind III or 1kb ladder molecular weight marker DNA (Gibco) was run alongside the digested DNAs. Electrophoresis was carried out at 35V over 48hrs and the running buffer was changed twice. The gel was Southern blotted (see Methods Section 2.10).

2.13.3 Prehybridization

The membrane was incubated in prehybridisation solution 1 (6XSSC, 5mM EDTA, 6% PEG 6000, 1% SDS, 0.25% Marvel milk) for 3hrs at 60°C in a shaking water bath.

2.13.4 Multi-locus Probe Synthesis

The recombinant 33.6 M13 mp8 and 33.15 M13mp19 bacteriophage DNAs (Jeffreys *et al* 1985a) were digested with *Eco* RI and *Hind* III. The 720 bp insert of 33.6 and the 592 bp insert of 33.15 containing variants of the core minisatellite sequence were isolated by running the digests into 0.8% LMP agarose and cutting out the bands of interest. Each minisatellite DNA insert was then melted and 10ng placed in an oligolabelling reaction.

2.13.5 Hybridization

The boiled probe was added to the prehybridization solution and hybridization carried out overnight with shaking at 60°C.

2.13.6 Post-hybridization Washes

The membrane was washed in 1X SSC/ 0.1% SDS for 15 mins (3X) or more until a plateau of counts was reached. It was then wrapped in cling film and exposed to x-ray film for 16hrs - 2 weeks.

2.14 Removal of Probe from Hybond N Membranes for Re-use

(Bruford *et al.*, 1992)

The membrane was incubated with shaking at 45°C for 15 mins in 0.4M NaOH. It was then transferred to 0.1XSSC / 0.1% SDS/ 0.2M Tris.HCl pH 7.5 for 30mins at 45°C. The wash in 0.1XSSC /0.1% SDS/ 0.2M Tris.HCl pH 7.5 was then continued until the counts, detected using a mini-monitor, (Mini- intruments, type 5.10 or series 900) were negligible. The membrane was exposed to x-ray film for a minimum of 16hrs with screens to check that stripping of the probe was complete.

2.15 Generation of the Canine Genomic DNA Library (SSJ1) in the Charomid 9-36

(The methods described in this section are based on those described by Bruford *et al.*, 1992).

2.15.1 Purification of Canine Genomic DNA Prior to Size-Fractionation

High molecular weight canine genomic DNA was isolated as described in Methods Section 2 from 10-20 breeds and pooled. This was further purified by extracting with phenol, phenol/chloroform/isoamylalcohol, and chloroform/isoamylalcohol several times as described in Methods Section 2.5. The DNA was then ethanol precipitated and the purity determined by :

1) electrophoresing an aliquot to check the purity of the DNA. (Any protein impurity remains in the well - this gives a preliminary estimate of the purity) and

2) determining the OD_{260}/OD_{280} of the DNA.

The ratio between the readings at 260nm and 280nm (OD_{260}/ OD_{280}) provides an estimate of the purity of the nucleic acid. A pure preparation of DNA has an OD_{260}/OD_{280} value of 1.8. If there is contamination with protein or phenol, this ratio will be significantly less than 1.8, and accurate quantitation of the amount of nucleic acid will not be possible.

After ensuring maximum removal of protein by the methods described above, the DNA was dialysed overnight in 10mM Tris.HCl pH8.0, 1mM EDTA pH 8.0 (this removes small organic molecules such as phenol and chloroform). After dialysis, the concentration and purity of the DNA was determined by carrying out electrophoresis of an aliquot together with a dilution series of λ DNA (Gibco) and by checking the OD_{260}/OD_{280}. Optical density measurements at 260nm and 280nm were recorded after phenol extraction and dialysis so as to determine the efficacy of each purification procedure.

2.15.2 Size-fractionation of Canine Genomic DNA using Preparative Gel Electrophoresis

2.15.2.1 Digestion of Canine Genomic DNA with *Mbo* I

One milligram of pooled, purified canine genomic DNA was digested to completion with *Mbo*I. Prior to digesting the whole sample, the amount of *Mbo*I required to digest 1μg of DNA was determined. 0.5units *Mbo*I/μg DNA was then used to digest 1mg of high molecular weight canine genomic DNA. The digest was incubated overnight at 37°C. An aliquot was then electrophoresed to ensure complete digestion before continuing.

2.15.2.2 Size-selection of the 2-16kb Genomic Fraction

75 ml of 2% agarose (Seakem) was poured into a taped 20 X 20 cm glass plate. A 2.5mm thick comb with a long central 14 cm slot and two small 5 mm slots on either side was clamped 2mm above the agarose base plate. 600ml of 0.8% (Sigma Type I, low EEO) agarose was poured on top of the agarose base. The set gel was placed in a gel tank with 1X TBE 2mm above the surface of the gel. (The gel tank used, allowed a voltage to be applied after removal of the lid). 3μg λ *Hind* III DNA in a final volume of 30μl and 6μl of loading buffer was loaded into the outermost 5mm slots. (The 5mm slots were 2/3 filled with loading buffer before loading the marker DNA). The digested DNA was made up to a final volume of 1.8ml with sterile dH_2O and 200μl of loading buffer added. This was loaded into the 14cm slot and run at 20-30V overnight. The migration of the markers was checked using a 254nm long wave u.v. wand (Mineralight

lamp,model UVG - 54, Ultra-Violet Products, USA). Pieces of dialysis tubing were cut to a single layer 1cm longer than the main slot and 0.5cm wider than the depth of the gel. The positions of the markers and the edges of the main DNA sample were marked by injecting the gel with 5µl of loading buffer. An incision using a scalpel blade was made straight across starting from one edge of the main DNA sample to the outer edge, avoiding the marker DNA. The position of the incision depended on the size fraction of genomic DNA chosen. A dialysis membrane was inserted into each incision using two blunt Millipore tweezers, such that the concave face of the membrane was facing the loading slot. The DNA was electrophoresed on to the dialysis membranes at 200V. When all of the size-selected fraction was loaded onto the membrane at the front (2kb), the voltage was reduced to 100V and the incision extended at both sides so that the gel was cut in half. The lower half of the gel was prised gently away from the upper half to open up a gap of 1-2mm such that the rear of the membrane was no longer held in place by the gel. With the current still on, the membrane was grasped with two blunt tweezers (Millipore) and transferred in one single movement (as quickly as possible - < 1sec) into a 30ml Sterilin tube (note that all of the above was done as quickly as possible to minimise risks of electrocution!). Part of the membrane was pinched by screwing down the lid and centrifuged for 1min at full bench speed in a MSE bench top centrifuge. The liquid was transferred to an eppendorf tube. Both sides of the membrane was moistened with 200µl of sterile dH_2O, bench centrifuged and the liquid pooled with the first sample. The membrane was checked under u.v. light to ensure complete recovery. The DNA was ethanol precipitated overnight at -20°C and the DNA pellet dissolved in an appropriate volume of $TE_{0.1}$. The concentration was estimated by the electrophoresis of a small aliquot with concentration standards.

2.16 Preparation of *Bam*HI - digested Charomid 9-36 DNA

Charomid 9-36 DNA in the bacterium DH5α (see appendix) supplied by the Japanese Cancer Research Resources Bank was grown overnight at 37°C for 16hrs only. This was done due to the following reason:
Charomid 9-36 contains 15 spacer fragments that consist of tandem repeats. Therefore, the number of generations of bacterial growth must be kept to a minimum during the propagation of the vector due to its tendency to lose

spacer fragments. Charomid DNA was isolated as described in Methods Section 2.23.1 and digested with *Bam*HI. After digestion, the Charomid 9-36 DNA was size-selected (see Methods Section 2.23.2) by running λ*Xho*I marker DNA alongside.

2.17 Ligation of the 2-16kb Canine Genomic DNA Size Fraction into the *Bam* HI Site of Charomid 9-36

A 2 : 1 molar ratio of insert : vector was used in the ligation. 200ng of the 2-16kb size fraction and 1.2µg *Bam* HI-digested Charomid 9-36 DNA (donated by Dr Olivier Hanotte, Department of Zoology, Leicester University; see Section 2.16 for preparation) was made up to a final volume of 10µl containing 1X ligase buffer (660mM Tris.HCl pH 7.6, 50mM $MgCl_2$, 50mM DTT), 1mM ATP, 4.5mM spermidine trichloride and 0.5 units DNA ligase. This was incubated at 14°C for 3 days. After 3 days, 0.5 units ligase, 0.5µl 10mM ATP (SIGMA) and 0.3µl 10X ligase buffer were added (to ensure ligation of any remaining unligated DNA) and the ligation mixture incubated at 14°C for a further 2 days. The ligation mixture was stored at -20°C.

2.18 Packaging *in vitro*

Packaging *in vitro* was carried out using the commmercial packaging extract Gigapack II Plus from Stratagene according to the protocol provided.

2.19 Infection and Titration of NM554 bacteria with Packaged Phage

NM554 bacteria (see appendix) were grown up overnight at 37°C by inoculating 2ml of 2X TY containing 20% maltose and 10mM $MgSO_4$ with one colony. A negative control containing 2ml of 2X TY was set up alongside. 4ml of fresh 2X TY was added to the 2ml overnight culture and grown for 3hrs at 37°C. Titration of packaged λ particles was then carried out by diluting 21µl of the packaged extract in 80µl of SM buffer followed by several 100 fold dilutions of this to a final dilution of 5×10^6 fold. 200µl of freshly grown NM554 bacteria was added to each tube and incubated at room temperature for 20 mins. 700µl of 2X TY was then added and the

tubes incubated for 1hr at 37°C. 200µl of the cultures from each tube was plated out onto 5 separate agar plates containing ampicillin (50µg/ml). The plates were incubated at 37°C overnight.

2.20 Determination of the Size of SSJ 1

The approximate titre of the total library suspension was determined from a plate where the colonies could be easily counted. The total number of clones in the total library suspension was calculated as follows:

No. of colonies in 1ml of 1st dilution = No. of colonies from 200µl of 1st dilution x 5

1ml of first dilution contains 20µl from a total of 500µl of library suspension.

No. of recombinants in 500µl of stock library suspension = No. of colonies from 1ml of 1st dilution x 25

2.21 Establishment of SSJ1 onto Microtitre Plates

(Bruford *et al.*, 1992)

120µl of 2X TY containing 15% glycerol and 50µg/ml ampicillin was added to each well of a microtitre plate. Individual colonies generated as described in Methods Section 2.19 were picked from plates using a P200 Gilson to suck them up into the tip and each well inoculated by pipetting the medium up and down 2-3 times. The lid was then secured with tape and the plate stored at -20°C .

2.22 Screening of SSJ1 with the Multi-locus Minisatellite Probes 33.6 and 33.15

2.22.1 Replication of SSJ1 onto Hybond N Memranes

(Bruford *et al.*, 1992)

1200 ml of Bottom agar containing ampicillin (50µg/ml) was poured into a large (38 x 33 cm), sterile, stainless steel tray and allowed to set in a sterile fume hood (Gelaire Flow Laboratories). Hybond N membranes were cut to fit the size of a microtitre plate and placed on the set agar, avoiding air bubbles between the agar and the membrane. The prongs of a 'hedgehog' microtitre replicating device (Microtiter-Dynatech) were sterilized by dipping in IMS and flaming. This when cool, was placed in the microtitre plate and then pressed onto a membrane. The 'hedgehog' was flamed between plates and allowed to cool. After pressing the colonies onto the mebranes, the tray was covered with aluminium foil and incubated in a 37°C oven overnight. The resultant bacterial colonies were lysed and the DNA fixed as follows: The membranes were placed on 3MM Whatmann filter paper, soaked previously in 2X SSC/ 5% SDS for 2 mins in a glass pyrex tray. The membranes were then heated at the maximum setting for 2.5 mins in a microwave oven (Hitachi Model MR 7300, 1.4kW) (Buluwela *et al.*, 1989). The membranes were covered in cling film until required.

2.22.2 Prehybridisation, Multi-locus Probe Synthesis, Hybridisation and Post-hybridisation washes

(Bruford *et al.*, 1992)

10 membranes containing fixed and lysed colonies were placed consecutively in 50ml of prehybridisation solution 1 (see Methods Section 2.13.3) in a 10x15 cm hybridisation chamber. This was incubated for 16hrs at 60°C prior to hybridisation. Before adding a given multi-locus probe, the prehybridisation solution 1 was removed and replaced with an equal amount of the same solution. The multi-locus probe as prepared in Methods Section 2.13.4 was added and hybridisation was carried out overnight at 60°C. The filters were washed in 1X SSC/ 0.1% SDS at 60°C until a plateau of counts using a mini-monitor was reached. The filters

were then wrapped in cling film and exposed to x-ray film as described in Methods Section 2.9.

2.23 Testing of Positive Clones for Polymorphism
(Bruford *et al.*, 1992)

2.23.1 Small-scale Purification and Isolation of Charomid DNA from Positive Clones

Microtitre plates from the library identified above as potential positive recombinants containing canine minisatellite sequences, were thawed and 10ml of 2X TY inoculated with 10µl from the required wells. This was grown overnight at 37°C with shaking at 260rpm. The cells were pelleted in a bench centrifuge for 5 mins at 4000g, the supernatants were removed and the pellets resuspended in 200µl of lysis buffer (see Methods Section 2.7.4). This was transferred to eppendorf tubes and placed on ice for 10 mins. 400µl of 0.2M NaOH/1% SDS was then added, mixed by inversion and left on ice for 5 mins. This was followed by 300µl of 3M K Acetate pH 5.2, mixing by vortexing the tubes upside down briefly and the tubes left to stand on ice for 10 mins. The solutions were microcentrifuged for 10 mins and the supernatants removed to fresh eppendorf tubes. This was repeated once more. 1ml of absolute ethanol was added and the tubes chilled at -70°C for 10 mins. They were microcentrifuged for 10 mins, the supernatants removed and the DNA pellets dissolved in 100µl of sterile dH_2O. 50µl of 7.5M ammonium acetate was then added and the tubes left at room temperature for 10 mins. They were microcentrifuged for 10 mins and the supernatants transferred to fresh tubes. 300µl of ice-cold absolute ethanol was added to the supernatants, mixed and the tubes chilled for 10 mins at -70°C. This was followed by microcentrifugation for 10 mins. The resulting DNA pellets were washed in 70% ethanol, vacuum dried and dissolved in 20µl dH_2O.

2.23..2 Purification of *Sau* 3AI-digested Canine DNA inserts from Positive Clones

Charomid DNA, isolated from positive clones, was digested with 20 units of *Sau*3AI for 3hrs at 37°C. No ribonuclease A was included as minute traces of DNAase activity destroys Charomid DNA very efficiently.

Aliquots were removed to check for complete digestion and the remainder of each sample was loaded into alternate wells of a 1% agarose gel. The gel was run for 1hr at 100V or until each *Sau*3AI insert was sufficiently separated from the Charomid backbone. The insert bands were recovered as follows: Each band was cut out and kept in an microfuge tube. A slot was then cut, the size of the insert band, out of another 1% agarose gel. A piece of dialysis membrane was cut to the width of the slot and 1.5 cm longer than the length of the slot. The insert band was placed on top of the dialysis membrane and fitted into the slot. The DNA in the agarose was run onto the membrane at 130V for 15-30mins. (Note:- all of the above manipulations were carried out as quickly as possible to minimise risks of electrocution!). The progress of the DNA was monitored using a long wave u.v. wand. When the membrane was fully loaded, the membrane was firmly grasped at one end (nearest to the DNA) and in one single movement (< 1sec) transferred into an eppendorf tube. This was microcentrifuged briefly and the solution transferred to a fresh eppendorf. The membrane was secured onto the lid and microcentrifuged again. The remainder of the solution was pooled with the first sample. This DNA was then ethanol precipitated overnight at -20°C and dissolved in 10µl $TE_{0.1}$. The concentration was determined and the volume increased to a suitable concentration (5ng/µl) for subsequent oligolabelling reactions.

2.23.3 Use of Canine DNA inserts as Single-locus Probes

10ng of gel-purified insert DNA was oligolabelled (see Methods Section 2.3) and used to probe Southern blots of canine genomic DNA digested to completion with *Mbo*I. The blots were incubated in 20ml of prehybidisation solution 2 (0.263M sodium hydrogen phosphate pH 7.2, 7% SDS, 1mM EDTA, 1% BSA (added fresh at < 65°C)) for 3hrs at 65°C in a Hybaid oven prior to the addition of the probe. Canine genomic DNA (10µg/ml prehybridisation solution 2) denatured by boiling for 20 mins and sheared by syringing through a 0.5 x 25 (Microlance) needle was added as competitor. After hybridisation of the probe at 65°C overnight, the blots were washed in 40mM sodium phosphate/ 1% SDS at 65°C for 10 mins, followed by 0.1X SSC/0.01% SDS at 65°C for 10 mins or until a plateau of counts was reached using a mini-monitor. The filters were wrapped in cling film and autoradiographed (see Methods Section 2.9) for 16hrs - 2 weeks.

2.24 Subcloning of Canine DNA Inserts into the pBluescript and M13 Bacteriphage mp18 and mp19 vectors

Gel purified restriction fragments generated from *Sau*3AI digests were cloned into the *Bam*HI site of pBluescript and restriction mapped. The bacterium XL1-Blue was used as a host for pBluescript. The ligation conditions used were identical to those described in Methods Section 2.17 except that incubation times were generally not more than 2-3 days. Transformations were carried out as described in Methods Sections 2.7.1 & 2.7.2 with the following modifications: Competent cells were heat shocked for 90 sec at 42°C. Following heat shocking, 1ml of 2X TY supplemented with 20mM glucose was added and the cells allowed to recover growth at 37°C. They were then plated out onto Bottom agar plates (~200µl/plate) containing x-gal (32µg/ml), IPTG (32µg/ml), ampicillin (50µg/ml) and tetracycline (12.5µg/ml). Smaller fragments containing minisatellite inserts were identified by probing Southern blots of digested insert DNAs with the multi-locus probe each clone originated from. The reduced inserts were then subcloned (ligation conditions once again being the same as those described above) into the M13 mp18 and mp19 bacteriophage vectors for simplified single and double-stranded M13 Sanger dideoxy chain termination sequencing.

APPENDIX

Bacterial Strains

All bacterial srains except for DH5α were purchased from Stratagene.

JM101 *supE, thi-1,* Δ (*lac-proAB*), [F' *traD36, proAB,* lacI ᑫZ Δ*M15*]

NM554 *recA13, araD139,* Δ(*ara-leu*)*7696,* Δ(*lac*)*l7A, galU, galK, hsdR, rpsL* (str^{r}), *mcrA,*$^{-}$ *mcrB-*$^{-}$

XL1-Blue *recA1, endA1, gyrA96, thi-1, hsdR17, supE44, relA1, lac,* [F' *proAB, lac* ᑫZ Δ*M15,* Tn*10* (tet^{r})].

DH5α *hsd* R17, *rec* A1, *mcr A,*$^{+}$ *mcr B.*$^{+}$ (Raleigh *et al.,* (1988)).

Solutions

Unless stated otherwise, compounds necessary for all solutions were purchased from the following: BDH, FISONS, USB, OXOID & SIGMA.

CHAPTER 3

GENERATION AND ANALYSIS OF A CANINE GENOMIC LIBRARY CONTAINING CLONED MINISATELLITE SEQUENCES

SUMMARY

A canine genomic library enriched for minisatellite sequences was generated in the vector, Charomid 9-36. 2940 colonies were screened with two multi-locus probes, 33.6 and 33.15. 93 positive clones were obtained as a result. Initial analysis of 32 positive clones indicated that a high proportion of clones contained multiple canine DNA inserts. This necessitated re-screening of DNA prepared from individual positive clones with the respective multi-locus probe to indicate which inserts contained repeat sequences. 48 individual positively-hybridizing inserts were isolated, and used at high stringency in the presence of canine, competitor DNA, to probe Southern blots of *Mbo*I-digested DNA from 4-6 unrelated canine individuals of different breeds. This analysis yielded a total of 28 variable canine minisatellite loci (the remainder were monomorphic in the animals tested). 15 of these require further analysis to determine the origin of the variation detected within the patterns obtained. Of the remaining 13 clones, which detect scorable single locus patterns, 6 detect loci that are dimorphic within the individuals tested and have heterozygosity values up to 50%. The remaining 7 clones detect distinct polymorphic loci with heterozygosity values of 20 - 88% with 5-35 alleles/locus.

INTRODUCTION

Existing canine breeds have been established by selective line breeding to give the breed-specific, phenotypic characteristics seen within the population. In a number of cases, this has resulted in the emergence of genetically-inherited diseases which have been shown to be a result of a single recessive mutation (Willis, 1989). The establishment of conclusive tests for the detection of carriers of particular mutations, would therefore be extremely useful to selectively breed unaffected individuals. The majority of canine genetically-inherited diseases can only be diagnosed phenotypically. However, an indirect approach may be applied to this problem by the isolation and utilization of localised canine DNA markers, to study the co-segregation of a given DNA marker with a disease locus within a pedigree.

Several sequences which share the property of highly polymorphic length have been discovered in the human genome within the past few years (Jeffreys *et al.*, 1991b). These hypervariable minisatellite regions consist of tandemly repetitive arrays of often, GC-rich, repeat units of approximately 16 - 70bp. Their variability lies in the ability of these arrays to undergo rearrangements via processes such as unequal sister chromatid exchange at mitosis and slippage during DNA replication (Jeffreys *et al.*, 1990). Such rearrangements result in allelic differences in the number of repeat units present at such loci and are reflected as length polymorphisms. Polymorphic minisatellite loci have a tendency towards high levels of heterozygosity and can serve as highly informative DNA markers for linkage analysis. The informative capacity of polymorphic minisatellite loci as DNA markers for linkage analysis, depends on their uniform distribution throughout a genome. Studies to date on the location of such sequences within the human genome have shown that they tend to be preferentially localised within the pro-terminal regions of human autosomes (Royle *et al.*, 1988). On the other hand, murine minisatellite sequences have been shown to be randomly dispersed over autosomes (Jeffreys *et al.*, 1987b).

The main aim of this study has been to apply the current methods of isolating polymorphic minisatellite loci to the canine genome. Ordered array Charomid libraries have been successfully used to isolate minisatellite loci in a number of species (Hanotte *et al.*, 1992b; Bruford *et al.*, 1992; Armour *et al.*, 1990b). This chapter therefore describes the

generation and analysis of a Charomid ordered array library containing cloned canine loci enriched for the isolation of variable minisatellite loci.

RESULTS

3.1 Preparation of Canine Genomic DNA

3.1.1 Collation of Canine Genomic DNA from Various Breeds

Canine genomic DNA of individuals from 10 breeds was isolated and equal amounts of each DNA pooled to give a total of 1mg. This increases the likelihood that canine minisatellite loci consisting of alleles varying greatly in size in different individuals are represented within the pooled sample.

Bruford *et al.*, (1992) suggest choosing individuals from the heterogametic sex to ensure equal representation of sex-linked loci. However, this was not possible at the time due to the paucity of canine DNA samples of known sex. The DNA isolated and pooled was from the following breeds: Labrador, West Highland White, Bernese Mountain dog, Poodle, Irish Setter, Blood Hound, Great Dane, Lhasa Apso, Pekinese and Border Terrier.

3.1.2 Purification of Pooled Canine Genomic DNA

The pooled canine genomic DNA was purified as described in Methods Section 2.15.1. It was necessary to ensure maximal purification so that 1mg of the pooled DNA could subsequently be digested using the minimum amount of *Mbo*I.

3.1.3 Selection of the Canine Genomic DNA Fraction for Cloning into Charomid 9-36

The aim of generating a canine genomic library in the Charomid vector 9-36 was to harvest as many polymorphic canine minisatellite loci as possible. Therefore, to enrich the library for the isolation of polymorphic minisatellite loci, a specific canine genomic size fraction was selected for subsequent cloning. The rationale behind this was as follows: Canine DNA fingerprints have been shown previously by Jeffreys &

Morton, (1987) to display the individual-specific characteristics now shown to be present in numerous other species. Therefore, the size fraction within a canine DNA fingerprint showing the greatest variation in minisatellite banding patterns, between individuals from different breeds, was speculated to be a suitable region of the canine genome from which polymorphic minisatellite loci might be isolated. The selection of this size fraction involved the generation of canine *Mbo*I DNA fingerprints of individuals from different breeds, using the multilocus probes, 33.15 and 33.6 (Figure 3.1). These indicated that the greatest variation in minisatellite DNA banding patterns and therefore the majority of polymorphic loci, was mainly contained within the genomic size range above 2kb. The variation within the genomic size fraction above 2kb was further assessed by determining the level of band sharing between the individuals tested. The band sharing coefficient was therefore calculated using the formula:

$$x = ((N_{ab}/N_a) + (N_{ab}/N_b))/2$$ (Bruford *et al.*, 1992)

where N_{ab} = number of bands of similar intensity and electrophoretic mobility in individuals a and b. N_a = total number of bands in a which could be scored, if present, in b and N_b = total number of bands in b which could be scored, if present, in a. Thus, the band sharing coefficient for the 33.15 DNA fingerprint was 0.31 and that for the 33.6 DNA fingerprint was 0.13. The overall level of band sharing seen therefore is 22% over the resolved part of the canine DNA fingerprints. This value is considerably lower than that determined by Jeffreys & Morton, (1987) (~46%) but is likely to be a reflection of sampling rather than a real difference.

The values for the levels of band sharing in the canine *Mbo*I DNA fingerprints were determined in a fewer number of individuals (Poodle, Springer Spaniel & Bernese Mountain Dog) than those studied by Jeffreys & Morton, (1987). The levels of band sharing seen between breeds is influenced by the inbreeding coefficients specific to each breed (Kuhnlein *et al* (1990)). Furthermore, Hanotte *et al.*, (1992b) have found that in gallinaceous birds the choice of restriction enzyme for a particular DNA fingerprint analysis influences the index of similarity (or band sharing value). They did, however, find that these two parameters showed concordance within species. Therefore, the use of a greater number of

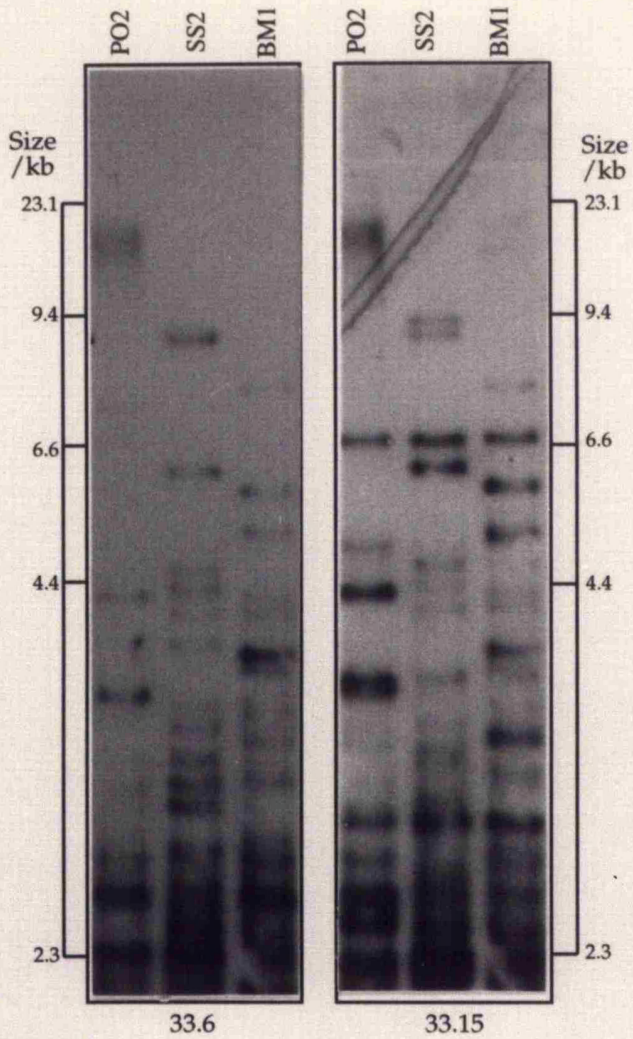


Figure 3.1 : 33.6 and 33.15 *Mbo*I DNA fingerprints of 3 breeds. PO2 - Poodle, SS2 - Springer Spaniel, BM1 - Bernese Mountain Dog.

individuals from a wider spectrum of breeds would probably give a more representative indication of the levels of band sharing in canine DNA fingerprints as a whole (see Chapter 4, Section 4.2).

Burke & Bruford, (1987) have shown that the multi-locus minisatellite probe, 33.15 detects a large smear within *Hinf*I - digested Japanese quail DNA, instead of the standard series of individual minisatellite DNA bands normally seen in DNA fingerprints. This has been attributed to the presence of an abundant class of tandem-repetitive satellite DNA containing occasional *Hinf*I restriction sites. It was therefore important to determine that the restriction enzyme used in the cloning process did not reveal a similar class of satellite sequences, which would hybridize to multilocus probes, completely obscuring any underlying minisatellite pattern within canine genomic DNA. Figure 3.1, therefore, also indicates that *Mbo*I does not produce a profile rich in satellite DNA classes within canine genomic DNA. This information is important as it is beneficial to minimise the cloning of satellite DNA. (The reasons for this will become clearer later on in this chapter). Consequently, the 2-16kb size fraction from *Mbo*I - digested canine genomic DNA was chosen for isolation and subsequent cloning into the vector, Charomid 9-36.

It is important to note that although the fraction of *Mbo*I - digested canine genomic DNA above 16kb also contained polymorphic minisatellite sequences, only the 2-16kb size fraction was isolated for the purposes of this study. The λ packaging system used to generate the library can only efficiently accept DNA molecules within the 38 - 52kb size range. Thus, a Charomid of 36kb such as Charomid 9-36 would only be efficiently packaged if it contained canine DNA inserts within the 2-16kb size range (see Introduction Section 1.6). It would have been possible to clone *Mbo*I fragments above 16kb if the Charomid vectors, 9-28 and 9-20 had been used to generate two further libraries containing canine genomic DNA within the 10-24kb and 18-32kb size ranges, respectively. However, this task would have been by no means a simple or an inexpensive one, as *Mbo*I digests of canine genomic DNA result in most of the DNA being below the 4kb size range, and very large quantities of starting DNA would have to be digested so as to obtain a suitable yield of DNA within the 10-32kb size range.

3.1.4 Digestion of Purified, Pooled Canine Genomic DNA

One milligram of pooled, purified canine genomic DNA was digested to completion with *Mbo*I as described in Methods Section 2.15.2.1: Prior to digesting the whole pooled sample, the minimum amount of *Mbo*I required to digest 1µg of the pooled DNA to completion was determined. A series of 1µg DNA samples were digested with varying amounts of *Mbo*I at 37°C for 16hrs. These were electrophoresed to check the extent of digestion. Figure 3.2 shows that the minimum amount of *Mbo*I required for complete digestion of 1µg of canine pooled DNA was 0.4units. However, so as to ensure complete digestion, 1mg of pooled canine DNA was digested using 0.5units of *Mbo*I /µg of DNA at 37°C for 16hrs. An aliquot was then checked for complete digestion before proceeding to the next stage. *Mbo*I is a restriction enzyme that cleaves frequently within genomic DNA but rarely within minisatellite sequences (Bruford *et al.*, 1992), and complete digestion of canine DNA results in most of the DNA being below the 2kb region. Thus, it was necessary to digest a large amount of canine genomic DNA to begin with so that a sufficient amount of DNA could be isolated from the region within 2-16kb.

The restriction enzyme *Mbo*I was chosen for the digestion of the canine genomic DNA prior to size selection for the following reasons:

(i) it is a four base recognition restriction enzyme (5' GATC 3') which cleaves eukaryotic genomic DNA frequently, generally leaving large fragments likely to be repetitive sequences such as minisatellite loci intact.

(ii) it is the only four base recognition restriction enzyme which is compatible with the *Bam* HI restriction site present in Charomid 9-36.

(iii) it does not reveal an abundant class of satellite sequences within a canine DNA fingerprint, thus indicating that a reduced proportion of such sequences will be present within the 2-16kb genomic fraction of *Mbo*I - digested canine DNA (see Results Section 3.1.3).

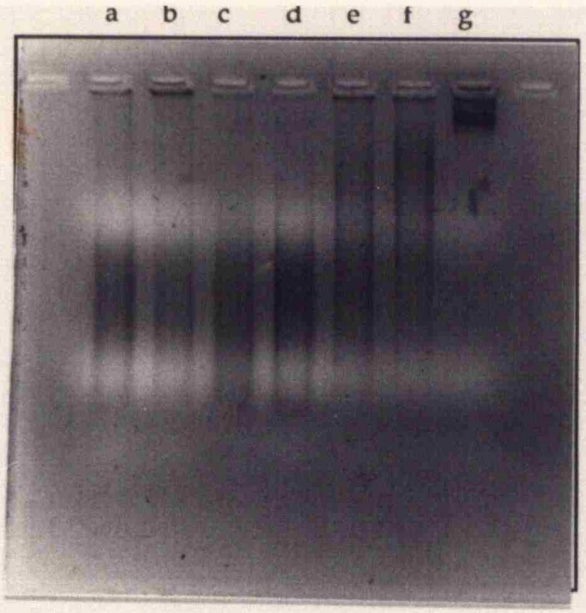


Figure 3.2: Determination of the minimum amount of *Mbo*I required to digest one microgram of purified, pooled canine genomic DNA to completion. A series of 1µg DNA samples were digested with varying amounts of *Mbo*I as described in Methods Section 2.2; a-f - 1, 0.75, 0.5, 0.4, 0.25, 0.1 units of *Mbo*I/µg DNA respectively. g - undigested DNA.

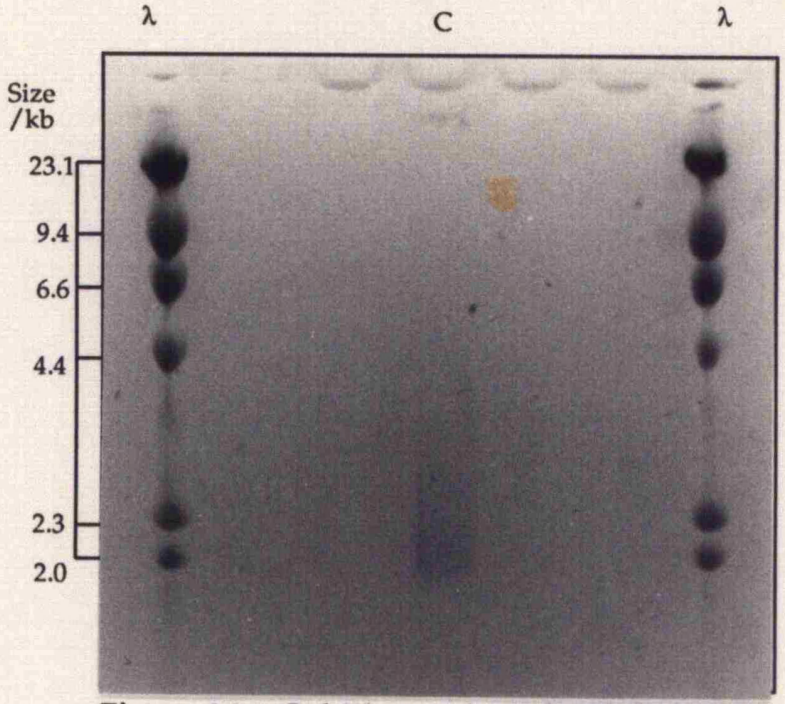


Figure 3.3 : Gel Electrophoresis of the 2-16kb canine genomic DNA fraction. After isolation, an aliquot of the size-selected DNA was electrophoresed alongside λ*Hind* III molecular weight marker DNA to determine the yield of DNA and to check the efficiency of the size-selection procedure. λ - λ *Hind* III molecular weight marker DNA, C - 2-16kb canine genomic fraction.

3.1.5 Isolation of the 2-16kb Canine Genomic DNA Fraction

The 2-16kb fraction of the *Mbo*I - digested canine genomic DNA was isolated as described in Methods Section 2.15.2.2. After isolation, an aliquot of the fractionated DNA was electrophoresed alongside λ *Hind* III marker DNA to check the concentration of the DNA and the efficiency of the size-selection procedure (Figure 3.3).

The yield of DNA from the 2-16kb size fraction was in total ~3μg (estimated by agarose gel electrophoresis of an aliquot alongside digested genomic DNA of a known concentration). Bruford *et al.*, (1992) state that it is necessary to repeat the size fractionation due to low molecular weight fragments aggregating during electrophoresis. However, due to several previous instances of repetitive, substantial loss of DNA after the second size fractionation, and since analysis of the DNA after the initial size fractionation suggested that there was minimal contamination of the 2-16kb size fraction with low molecular weight DNA fragments, a second size fractionation was not conducted in this study.

3.2 Ligation and Packaging *in vitro*

The 2-16kb size fraction was ligated into the *Bam*HI site of Charomid 9-36, as described in Methods Section 2.17. The 2:1 molar ratio of insert to vector DNA was calculated by considering the 2-16kb canine genomic DNA size fraction to mainly consist of fragments within the 3kb region: The greater the number of restriction enzyme sites for a given restriction enzyme within genomic DNA, the lower the proportion of larger fragments within a given amount of DNA. This is reflected in the complete digestion of canine genomic DNA using *Mbo*I, and visual assessment of the electrophoresed 2-16kb size fraction showed that most of the digested DNA present was within the 2-4kb region.

Thus, the calculation of a mean fragment size based on there being an equal representation of each fragment size present within the 2-16kb size fraction is not valid. Consequently, for the purposes of determining a 2:1 insert to vector molar ratio, 3kb was decided upon to be a suitable mean insert fragment size (see Figure 3.3). Electrophoresis of the control ligation, performed alongside the test ligation of vector and canine DNA, contained

1μg of λ *Hind*III fragments and showed a 47kb high molecular weight band as expected, indicating that the conditions used for the ligation were optimal (data not shown). 0.25μg of DNA from the test ligation was therefore packaged *in vitro* using the commercial packaging extract, Gigapack II Plus from Stratagene, as stated in Methods section 2.18.

3.3 Infection and Titration of NM554 Bacteria

The approximate titre of the library suspension was determined as described in Methods Section 2.19. The total number of clones in the library was estimated to be ~ 10^5 (see Methods Section 2.20). Gigapack II Plus packaging extracts are stated by Stratagene to have maximum packaging efficiencies of 1.0 x 10^9 pfu/μg of wild type λ DNA. The number of clones obtained within SSJ1 is 10^3 fold less than that this, i.e. ~ 10^5 clones were obtained from 0.25μg of DNA from the ligation, thus representing a packaging efficiency of 4.0 x 10^5 clones/μg of DNA from the ligation. The packaging efficiency obtained here is within the range generally obtained by Bruford *et al.*, (1992) i.e. 10^5 - 10^6 clones/μg DNA using identical protocols. The decrease generally seen in the efficiency of packaging recombinant Charomid molecules *in vitro* , compared with that quoted by Stratagene, could be due to reasons associated with possible, accumulative inefficiencies, inherent within the general protocol:

(i) Incomplete digestion of the Charomid 9-36 DNA would reduce the number of ligateable vector molecules.

(ii) Prolonged periods of incubation of canine genomic DNA with *Mbo*I would result in unligateable ends due to the possible presence of non-specific exonucleases/endonucleases .

3.4 Estimation of the Percentage of Recombinant Clones within the Canine Genomic Library SSJ1.

Estimation of the percentage of clones that contained DNA inserts within the library involved the isolation of Charomid DNA from 20 randomly picked colonies (Figure 3.4). The DNA inserts were liberated using *Sau*3AI (an isochizomer of *Mbo*I) as described in Methods Section 2.23.2.

It was necessary to use *Sau*3AI instead of *Mbo*I due to the *dam* modification system known to be present in NM554 (Pirrotta, 1976). The

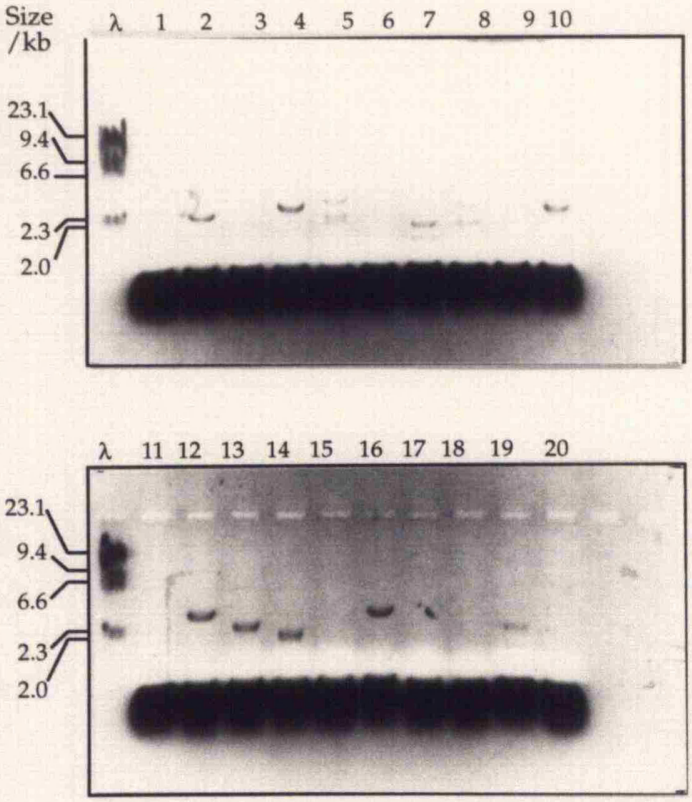


Figure 3.4 : Estimation of the percentage of recombinants within SSJ1.
Sau3AI - digested Charomid DNA from 20 randomly-picked colonies was electrophoresed to check for the presence of DNA inserts. 11/20 colonies were seen to contain DNA inserts indicating that 55% of the clones within SSJ1 were recombinant.

dam gene product is a specific methylase which transfers methyl groups from S-methyl-adenosylmethionine to the N^6 position of the adenine residue within the palindromic recognition sequence 5'GATC 3' to yield 5'GA^{m}TC 3'. The ability of certain restriction enzymes to cleave DNA is known to be inhibited by the presence of methylated residues within their recognition sites. *Mbo*I is unable to cleave DNA containing methylated adenine residues within its recognition site, whereas the presence of methylated cytosine residues within the recognition site of *Sau*3AI results in its inhibition. Thus, following the infection of NM554 bacteria with the packaged λ particles, any 5'GATC 3' sequences within the Charomid DNA would contain methylated adenine residues due to the *dam* methylase, and digestion of isolated Charomid DNA with *Mbo*I would be inhibited. Furthermore, *Sau*3AI could not have been used in the cloning process instead of *Mbo*I, because eukaryotic DNA is known to contain methylated cytosine residues (Lewin, 1990).

11/20 randomly-chosen clones were found to contain DNA inserts (Figure 3.4). This indicated that approximately 55% of the clones within the library contained DNA inserts (2.2×10^5 clones/μg DNA from ligation). Bruford *et al.*, (1992) state that a suitable library has 60-80% recombinants. The main aim in generating the canine genomic library as stated previously was to isolate as many polymorphic minisatellite loci as possible. Thus, it was important to determine at this stage whether or not the library generated would be suitable for futher analysis. This was done by considering the library size required, for subsequent screening, in order to find a specific minisatellite sequence with a certain level of probability, within the region of the genome represented in the 2-16kb size range. This can be estimated by using the following formula:

$$N = \frac{\ln (1 - P)}{\ln (1 - f)}$$ (see Sambrook *et al.*, 1990)

where P is the desired probability, f is the fractional proportion of the size selected genome in a single recombinant, and N is the required number of recombinants. Thus, to achieve a 99% probability of representing a specific locus in a library containing fragments averaging 3kb, from a size fraction representing 1% of the haploid genome, it would be necessary to screen 46,048 clones (see appendix). However, since only approximately 55% of the clones within the library contain inserts, the required number of clones to be screened would be increased to 83,725. Since the library

potentially contained ~10^5 clones containing inserts, this number was sufficient to ensure that a given locus within the 2-16kb size range would be represented at least once within the library.

3.4..1 A Few Inferences from the Liberation of Canine DNA Inserts from Twenty Randomly -Picked Colonies

Of the 11 clones containing inserts, 4 were found to contain more than one DNA band. This could be due to one of the following reasons:

(i) It is possible that multiple *Mbo*I - digested insert fragments religated to each other within the ligation, resulting in concatamers consisting of multiple, tandem inserts together with Charomid DNA.

(ii) On the other hand, a large minisatellite sequence could have undergone the loss of repeat unit sequences, during growth of the bacterium, NM554 (Kelly *et al.*, (1989) - see Section 3.5.2 for a more detailed explanation).

(iii) Another possibility, although less likely, is the presence of a large minisatellite sequence containing multiple internal *Mbo*I sites, which, when cleaved with *Sau*3AI would be reduced to a number of smaller fragments. It should be noted that this would only occur if the minisatellite in question had internal *Mbo*I sites where the cytosine residue was not methylated (see above).

It is important to note that NM554 is a *rec* A^- host and is therefore recombination deficient. However, this deficiency is not absolute i.e. the level of recombination is only diminished with respect to *rec A* strains and therefore repeat sequences that are prone to recombination will not necessarily always be stable.

10 out of the 11 clones contained DNA inserts within the size range, 2 - 3.5kb. This is indicative of the fact that, as explained previously in Section 3.2 of this chapter, most of the 2-16kb genomic size fraction consists of fragments within the lower end of the size range.

3.5 Screening of SSJ1 with the Multilocus Probes, 33.15 and 33.6

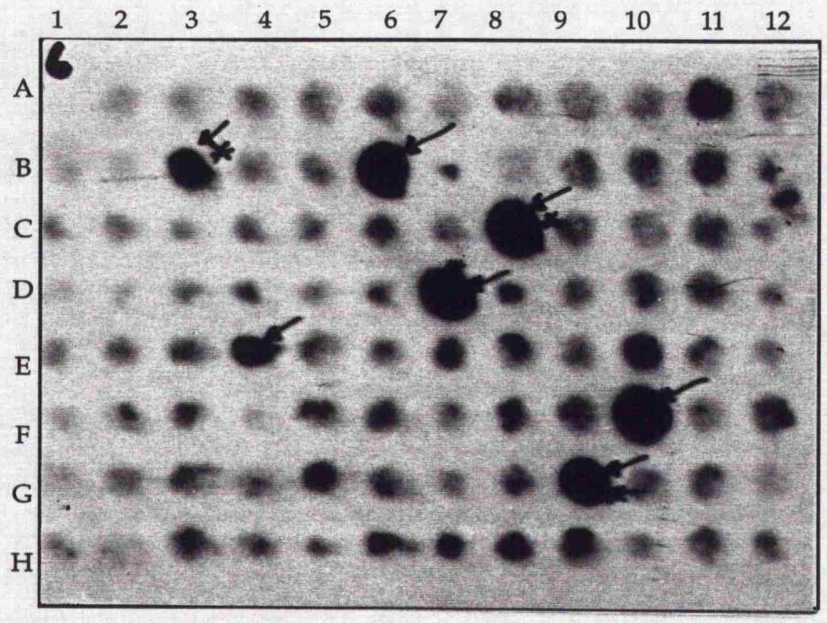
2940 NM554 colonies previously infected with recombinant phage were individually picked and transferred into microtitre plates to generate an ordered array of bacterial colonies (Armour *et al.*, 1990b). These were replica-plated onto Hybond N membranes and screened with the multilocus probes, 33.15 and 33.6 as described in Methods Section 2.22 (Figure 3.5). As a result, a sum total of 93 'strong to weakly hybridizing positive clones were detected. These can be represented as 5.8% of the total number of clones containing DNA inserts. It might therefore be suggested that the myoglobin - derived minisatellite sequences represented by 33.15 and 33.6, do not represent an abundant class of DNA within canine genomic DNA. This is further supported by the fact that 33.15 and 33.6 canine DNA fingerprints have been shown to contain a lower number of minisatellite loci than their human counterparts (Jeffreys & Morton, 1987).

3.5.1 Analysis of Initially Isolated Positive Clones Indicates a Larger Size Range of Inserts

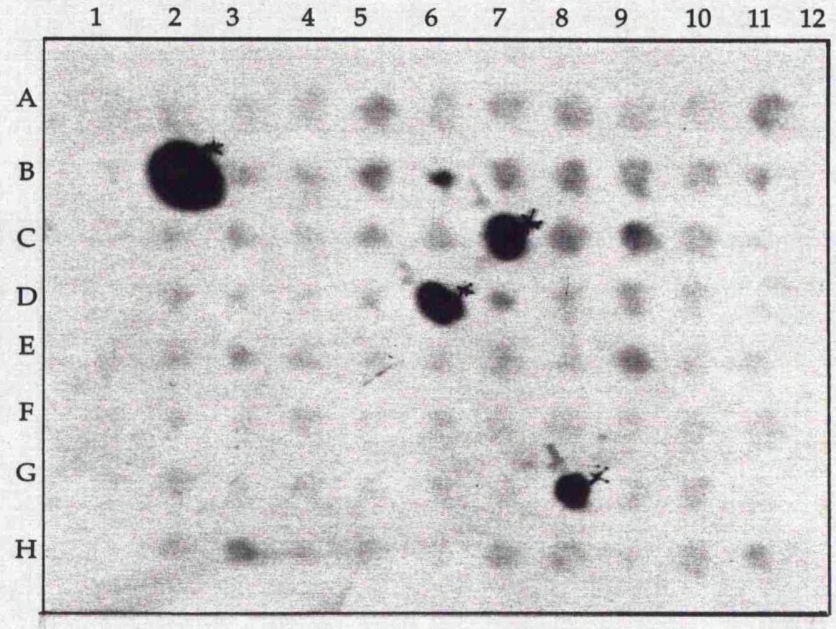
Charomid DNAs from each of the positive clones were isolated and digested with *Sau*3AI, so as to isolate each canine DNA insert for further analysis. Figure 3.6 shows a representative sample of the electrophoresis of *Sau*3AI-digested Charomid DNA from 32 positive clones, isolated initially using the multilocus probe, 33.15. All the positive clones can be seen to contain canine DNA inserts. However, from this analysis, it could be seen that the size range represented by the canine DNA inserts within these positive clones was considerably greater i.e. approximately 0.8 - 10kb. This however, still does not cover the whole size range spanned by the size fraction used for the cloning, and probably points to the paucity of fragments above 10kb within the 2-16kb genomic size fraction of *Mbo*I - digested canine DNA.

3.5.2 Detection of Fragments Smaller than 2kb within the 2-16kb Genomic Size Fraction

14/32 clones could be seen to contain multiple inserts among which there were DNA fragments of a lower size than the 2-16kb size fraction



33.6



33.15

Figure 3.5 : Positive clones obtained from screening with the multilocus probes 33.6 and 33.15.
All 4 positives obtained from the use of 33.15 to screen plate 6, cross-hybridize to 33.6. Filters were exposed to x-ray film with screens for 16hrs - 1 week.

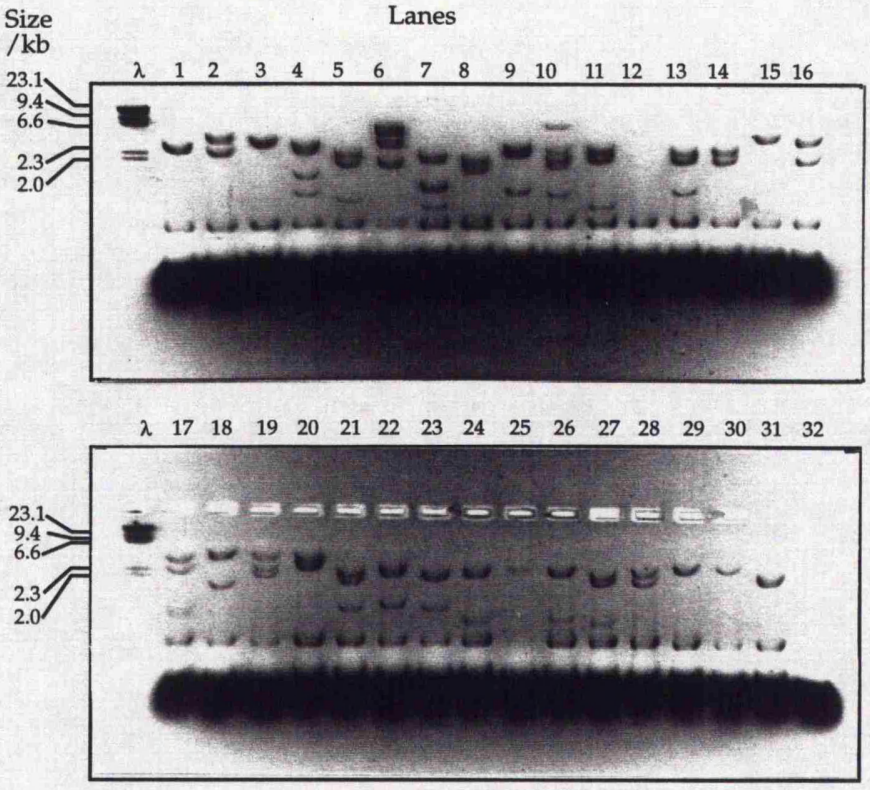


Figure 3.6 : *Sau*3AI-digestion of Charomid DNA from 32 positive clones obtained by probing SSJ1 with 33.15.

used for cloning in the library. Kelly *et al.*, (1989) have found that a 7kb allele of a highly unstable mouse minisatellite locus, *Ms6-hm*, collapsed into a 400bp plasmid insert on propagation in *E.Coli* due to the loss of minisatellite repeat units. Thus, it is possible that some of these lower sized inserts consist of collapsed regions of a larger minisatellite locus. However, it is more likely that residual contamination of the 2-16kb size fraction with fragments smaller than 2kb was present after the first size fractionation. As stated in Section 3.1.5, a second size fractionation was not carried out in this study due to reasons already stated.

3.5.3 Consequences of the Presence of Multiple Canine DNA inserts within Positive Clones

Seventy-two percent of the positive clones could be seen to contain multiple canine DNA inserts. As expected, all recombinant Charomid clones, represented a packaged Charomid size range of 38 - 52kb. None of the 32 clones contained solitary inserts of less than 2kb. The percentage of clones containing multiple canine DNA inserts is quite high and is probably due, in each case, to one of the reasons enumerated earlier in Section 3.4.1. A 0.5kb DNA band representing the Charomid backbone was common to all digests. This was a result of the presence of multiple *Sau*3AI restriction sites within Charomid 9-36.

The presence of multiple canine DNA inserts within 72% of the positive clones necessitated further probing of Southern blots of *Sau*3A-digested Charomid DNAs from all positive clones with the multilocus probes, 33.15 and 33.6. Figure 3.7 shows an example of such analysis; the gel shown in Figure 3.6 was probed with the multilocus probe, 33.15. This indicated which canine DNA bands among each set of multiple inserts contained a sequence similar to the 33.15 multilocus minisatellite. Positively hybridising bands were detected in 22/32 clones. Therefore, 10 /32 clones when probed again with 33.15 did not give a positive signal. This prompted the re-examination of these particular clones as to whether or not they were weakly hybridizing positives and therefore, originally, false positives. It was clear from all cases examined that each clone that did not hybridize after the second probing was originally a comparatively weakly-hybridizing positive clone.

Lanes 2,3 and 20 seem to contain two bands that hybridize but this is purely artifactual due to the way in which the gel was Southern

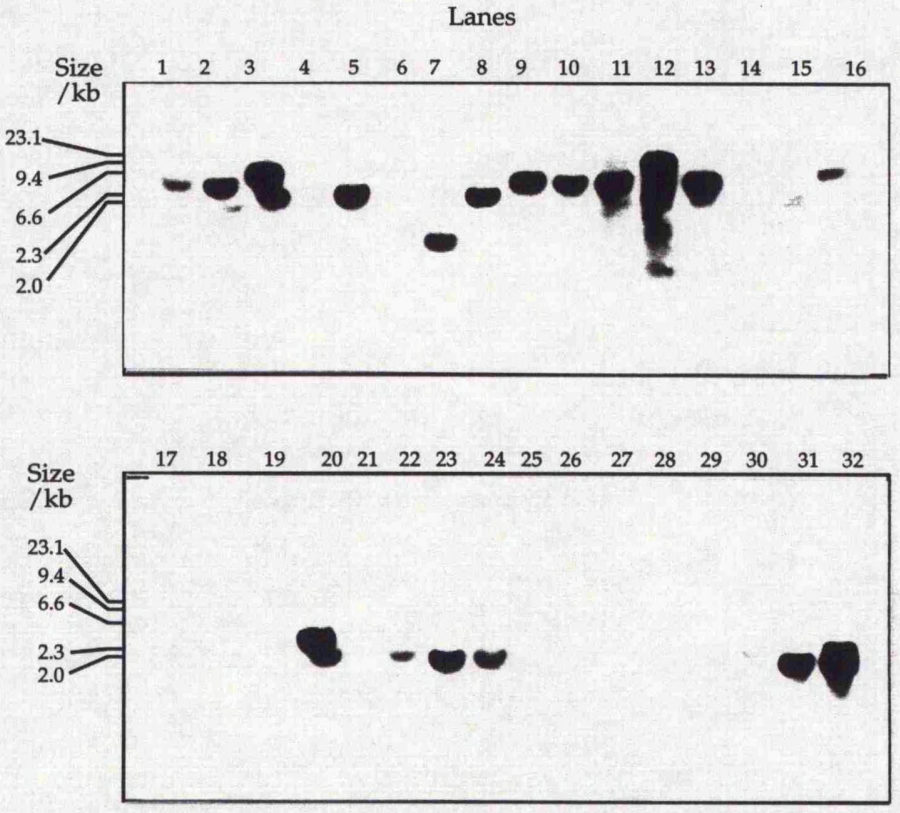


Figure 3.7 : Re-screening of 33.15 - derived positive clones with 33.15 to detect minisatellite sequences among multiple inserts.
Molecular weight markers are derived from λ *Hind* III DNA. The smear seen in Lane 12 indicates collapse of the minisatellite sequence during propagation in NM554; this particular clone was subsequently shown to be polymorphic and is described in Table 3.3 as c*Cfa*MP1.

transferred. However, lanes 12 and 32 do contain multiple bands that hybridize to the probe. In the case of the clone in lane 32, two faint canine DNA inserts can be seen on the respective agarose gel (see Figure 3.6). These two inserts could represent either:

(i) two canine DNA inserts containing minisatellite sequences from distinct loci.

(ii) a larger minisatellite insert sequence that has internal *Mbo*I sites.

(iii) the collapse of a larger canine minisatellite sequence.

The true nature of each insert can only be known by analysing each insert separately as single-locus probes. (In the case of this particular clone it was shown subsequently that each insert detected the same locus).

Lane 12, on the other hand, contains a major DNA band of approximately 9kb in size and a faint ladder of DNA bands directly below it. This seemed to indicate that this clone probably consisted of a canine minisatellite sequence that had collapsed during growth of the bacterium. Hypervariable minisatellite loci are known to be relatively unstable in the genome, due to frequent rearrangements via the speculated processes of unequal exchange and/or slippage during DNA replication. Two such clones, where re-probing of the respective *Sau*3AI - digested Charomid DNA molecules resulted in positively-hybridizing smears, were subsequently shown to detect polymorphic single-locus patterns (see Table 3.3). These were the only clones where insert fragments of a lower size than the 2kb were seen to hybridize after a second screening with the multi-locus probes, 33.6 and 33.15. This further supports the suggestion that the majority of insert fragments, that were smaller than 2kb, were cloned as a result of contamination (see Section 3.5.2).

3.5..4 Reduction of the Initial Number of Positive Clones

As stated previously, 31% of the positive clones shown in Figure 3.7 could not be seen to contain positively hybridising canine DNA inserts after a second screening with 33.15. These clones were all weakly hybridizing positives. Thus, re-screening of the 93 positives obtained initially, reduced the overall total to 48 'true' positives (i.e. 48.4% of the total number of positives obtained initially were eliminated from subsequent analysis). This reduction mainly resulted from the second screening with the multi-locus minisatellite probes, 33.15 and 33.6 but was also partly due to a number of clones which could not be grown up.

Clones which could not be grown up after the initial screening might be attributed to loss of ampicillin resistance by the bacteria. This could be due to either loss of the Charomid or a rearrangement or mutation within the Charomid such that ampicillin resistance was lost.

For there to be a 99% probability that any given sequence within the 2-16kb size fraction was represented, it would have been necessary to screen approximately 83,725 colonies. The 2940 colonies screened within this study therefore represent 3.5% of this total (see appendix to this chapter).

3.6 Testing of Positive Clones for Polymorphism

The canine DNA inserts from 48 positive clones were analysed as single - locus probes (see Methods Section 2.23). Figure 3.8 shows a representative sample of the various banding patterns displayed by a number of the single-locus probes obtained. A variety of banding patterns were detected. Apart from those clones that showed obviously mono- or polymorphic patterns, there were also clones that detected satellite, multi-band and dimorphic patterns. The results of this analysis are summarised in Tables 3.1 - 3.5. The nomenclature for each clone is derived from that suggested by Bruford *et al.*, (1992).

3.6.1 Clones Detecting Multi-band Patterns

Table 3.1 gives a summary of those clones detecting multi-band patterns. These consist of those that detect patterns that seem to correspond to:

(i) multiple loci
(ii) two loci or
(iii) a single locus containing an internal *Mbo*I restriction site.

There were a number that would probably serve as informative multi-locus probes (i.e. they detected patterns that correspond to characteristic DNA fingerprint patterns rather than single-locus patterns)(see Fig. 3.8c).

The informativeness of those that detect patterns within the above latter two categories can be determined by segregation analysis in pedigrees as well as the use of different restriction enzymes, so as to obtain the characteristic scorable single-locus pattern. In the case of (ii), segregation analysis would show independent assortment of specific DNA bands

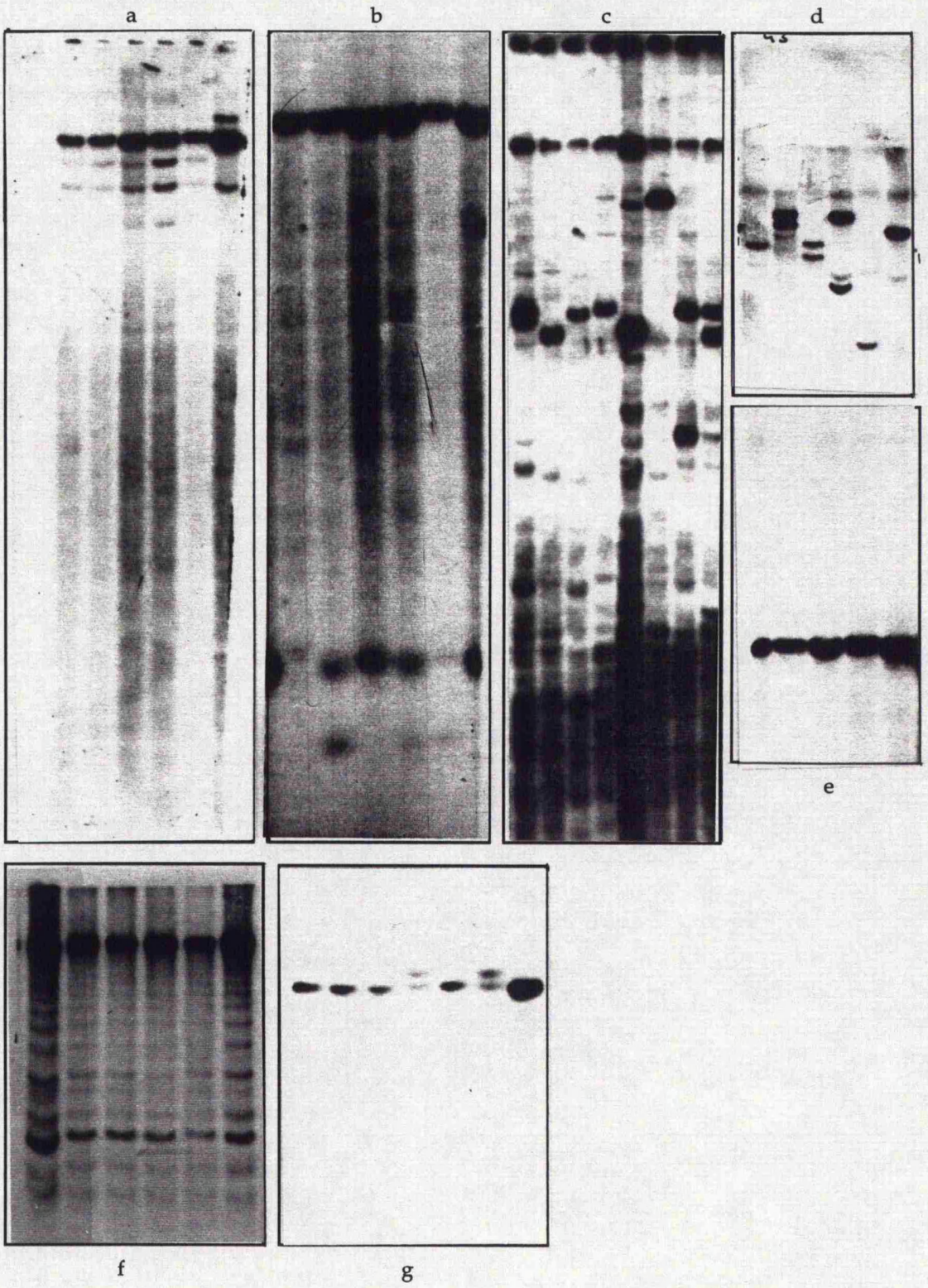


Figure 3.8 : Banding patterns seen during the testing of canine minisatellite loci for polymorphism. a, b & c - multiband patterns, d - polymorphic pattern, e - monomorphic pattern, f - satellite pattern, g - dimorphic pattern. Note that c consists of a DNA fingerprint-like pattern where many related minisatellite loci are detected. Such loci are likely to be useful as canine-specific multilocus DNA fingerprinting probes. Multi-band patterns a & b probably contain either multiple *Mbo*I sites or detect more than one locus.

Table 3.1 Summary of multiband pattern-detecting clones

Clone	Plate	Origin	Size of insert/kb	Band size range/kb
c*Cfa*MB1	1A10*	33.15	2.55	4.40 - 6.80
c*Cfa*MB2	3E5*	33.15	2.55	5.10 - 19.00
c*Cfa*MB3	3E6	both	2.30	1.55, 2.15, 2.30, 2.40, 2.75
c*Cfa*MB4	6C8*	both	3.50	3.45, 3.90, 7.20, 28.00
c*Cfa*MB5	6F10	33.6	2.80	3.90,4.00,4.30, 44.00
c*Cfa*MB6	6G9	both	2.50	1.80, 3.40, 3.70, 7.00, 21.00
c*Cfa*MR7	7H1	33.6	2.35	3.80 - 14.50
c*Cfa*MB8	9B4*	both	3.60	3.60 - 13.00
c*Cfa*MB9	9F10	both	2.85	1.75, 2.10, 3.80, 4.40
c*Cfa*MB10	10F10	both	2.35	2.50 - 4.50
c*Cfa*MB11	12F7	33.6	2.30	2.30, 2.40, 2.60
c*Cfa*MB12	19B9	33.15	2.35	5.10, 0.35, 0.40
c*Cfa*MB13	22G3	33.15	4.80	4.20 - 12.00
c*Cfa*MB14	29F10	33.6	3.20	3.05, 3.20, 3.30, 3.60,5.10
c*Cfa*MB15	33D10	33.15	2.80	3.80 - 46.00

* clones detecting a multilocus pattern. The nomenclature for each multi-band detecting clone is derived from c*Cfa*MBN where
c - clone, *Cfa - Canis familiaris* , MB - multi-band, N - Nth clone isolated.

Table 3.2 Clones detecting satellite patterns

Clones	Plate	Origin	Size of insert /kb
c*Cfa*SS1	1C11	33.15	2.55
c*Cfa*SS2	3A8	33.15	3.50
c*Cfa*SS3	5H6	33.6	3.50
c*Cfa*SS4	6B3	33.15	3.50
c*Cfa*SS5	21B4	33.6	3.10
c*Cfa*SS6	21C4	both	2.30
c*Cfa*SS7	23A4	33.15	2.10
c*Cfa*SS8	23H2	33.6	3.30

The nomenclature for each satellite-detecting clone is derived from c*Cfa*MBN where c - clone, *Cfa - Canis familiaris* , SS - satellite, N - Nth clone isolated.

corresponding to a given locus within a pedigree. Restriction analysis would confirm the presence of an internal restriction site by the use of other four base recognition site restriction enzymes to demonstrate the recovery of a diallelic locus in a heterozygous individual or an isoallelic locus in a homozygous individual. This analysis, however, has not been done within this study.

3.6.2 Clones Detecting Patterns Corresponding to Satellite Loci

As can be seen by Table 3.2, 8/48 positives yielded satellite patterns. Such patterns consisted of mainly monomorphic ladders of bands most of which were uninformative, and were attributed to satellite sequences (see Figure 3.8 (a)). There were some, however, that did contain one or two variable bands within the satellite ladder. However, since it was clear that the informativeness of these loci would generally be negligible, no further analysis was done. 83.3% of the total number of positive clones detected loci that were not of satellite origin. This reaffirms the results of the 33.6 and 33.15 *Mbo*I DNA fingerprint profiles obtained, as they were not found to be rich in satellite sequences (see Section 3.1.3 of this chapter).

3.6.3 Clones detecting Polymorphic Single-locus Patterns

Table 3.3 gives a summary of the characteristics of the 7 polymorphic clones obtained. These clones detect patterns that vary in the degree of informativeness. For example, the range in terms of the levels of heterozygosity is 20% - 88%. Furthermore, the number of alleles detected by each probe varies from 5 - 35. Figure 3.9 shows the heterozygosity analysis of c*Cfa*MP5 : DNA from 25 unrelated individuals from different breeds was digested with *Mbo*I, Southern blotted and probed with c*Cfa*MP5 at high stringency. c*Cfa*MP5 is the most informative clone isolated thus far and could possibly be considered a canine counterpart of the hypervariable MS1 locus found by Jeffreys *et al.*, (1988). It detects 35 alleles within the size range 1.75 - 11.00kb and has a heterozygosity of 88% within unrelated individuals from different breeds. Studies using polymorphic single-locus probes against individuals from particular breeds are described in Chapter 4.

Table 3.3 Clones detecting polymorphic single-locus patterns

Clone	Plate	Origin	Size of insert /kb	No.of alleles	Allele size range /kb	Hetero-zygosity
c*Cfa*MP1	6D7$	both	10.50	18	3.00 - 9.20	59%
c*Cfa*MP2	9C1	both	2.85	5	3.60 - 6.60	20%
c*Cfa*MP3	10H1	both	2.00	7	3.70 - 6.00	56%
c*Cfa*MP4	12B9	33.6	2.90	8	2.70 - 3.30	35%
c*Cfa*MP5	19F3$	both	2.60	35	1.75 - 11.00	88%
c*Cfa*MP6	10D7	33.6	2.10	5	2.8, 2.85, 3.1, 3.2, 3.9	33%
c*Cfa*MP7	22G12	33.6	2.85	6	3.90 - 7.20	67%

The nomenclature for each polymorphic clone is derived from c*Cfa*MPN where c - clone, *Cfa* - Canis familiaris, MP - minisatellite (polymorphic), N - Nth clone isolated. $ - clones the inserts of which were seen to collapse after propagation in NM554. Heterozygosities were determined by testing 10-20 canine breeds for polymorphism,

Table 3.4 Clones detecting dimorphic single-locus patterns

Clone	Plate	Origin	Size of Insert /kb	Allele size range/kb	Hetero-zygosity
c*Cfa*MD1	5H8	both	3.50	2.05,2.20	50%
c*Cfa*MD2	6E4	33.6	1.75	2.35, 2.50	20%
c*Cfa*MD3	7B9	33.6	3.30	1.55, 1.80	45%
c*Cfa*MD4	8A2	both	5.20	4.60 & 6.20	10%
c*Cfa*MD5	10C4	33.15	3.25	2.70, 2.85	50%
c*Cfa*MD6	21H6	both	5.70	2.60,2.70	29%

The nomenclature for each dimorphic clone is derived from c*Cfa*MDN where c - clone, *Cfa* - Canis familiaris, MD - minisatellite (dimorphic),

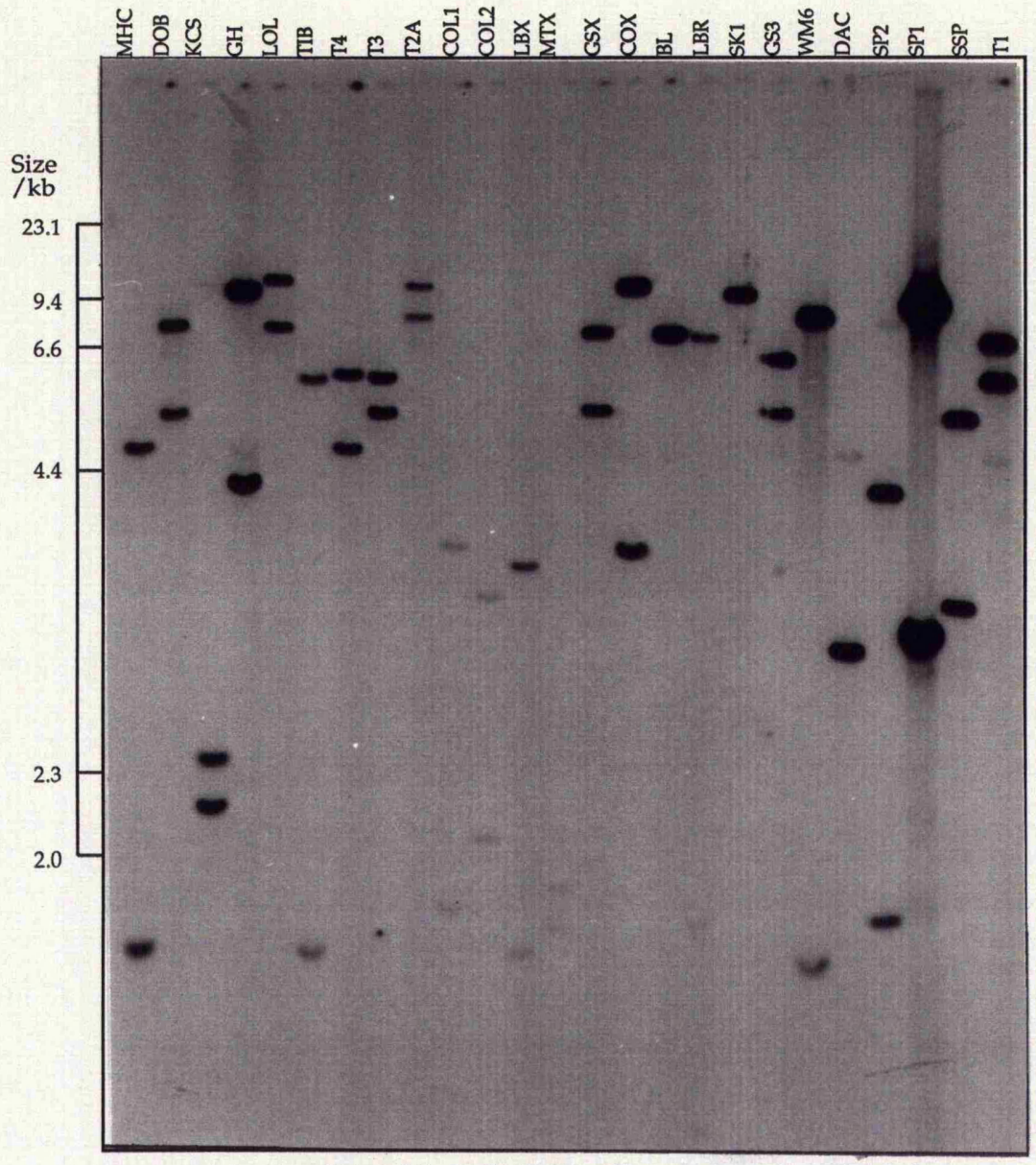


Figure 3.9 : Southern blot of *Mbo*I - digested canine DNA from 25 unrelated individuals from various breeds probed with c*Cfa*MP5 at high stringency.

MHC - collie, DOB - Doberman, KCS - King Charles Spaniel, GH - Grey Hound, LOL - Bedlington Terrier, TIB - Tibetan Terrier, T1-T4 & MTX - Terrier crosses, COL1 & COL2 - Collie crosses, LBX - Labrador cross, GSX - German Shepherd cross, COX - Collie cross, BL - Black Labrador, LBR - Labrador, SK1 - Schipperke, GS3 - German Shepherd, WM6 - Irish Setter, DAC - Dachshund, SP1 & SP2 - Spaniels, SSP - Springer Spaniel

3.6.4 Clones Detecting Dimorphic Loci

The clones that detect only two allelic forms of a locus have been defined as dimorphic. To confirm that these clones are true dimorphic VNTRs as opposed to RFLPs, further restriction analysis would be required. The principles behind this would be as follows: If the locus in question is a true VNTR, using other restriction enzymes should display the same morphism. This is because the regions flanking the repeat array will be homologous, and therefore any difference between alleles, due to the size of the repeat array, will only serve to shift the unique, flanking regions with respect to each other. Thus, the single-locus patterns seen due to each restriction enzyme should display the same dimorphism differing only in the presence of a shift in the sizes of the alleles, which will depend on the positions of each flanking restriction enzyme (i.e. the size difference between alleles will be similar)(see Figure 3.10 a). A RFLP is defined as a sequence that displays polymorphism due to the presence or absence of a specific restriction enzyme site. Therefore, the use of several other restriction enzymes to confirm the presence of such a locus within the DNA of a given individual would result in patterns which could be:

(i) monomorphic, due to the absence of any difference in the positioning of restriction sites on the corresponding homologous chromosome or

(ii) dimorphic but with allele sizes which do not correspond to similar differences in size between alleles (see Figure 3.10 b). Thus, it cannot be ruled out that a number of the clones categorised as dimorphics could, in fact, be RFLPs.

It should be noted that the levels of heterozygosity stated in Table 3.4, with regard to each dimorphic single-locus probe are determined, in most cases, from only 6 unrelated individuals. Therefore, it is possible, were a greater number of individuals studied in each case, that in reality the heterozygosity values could vary significantly. It is probable, however, that any changes involved would generally lower these values due to the level of inbreeding within the canid species. This is supported by the fact that one of the dimorphic clones (c*Cfa*MD3) when used to screen 20 unrelated individuals from different breeds has a characteristically low level of heterozygosity (see Table 3.4).

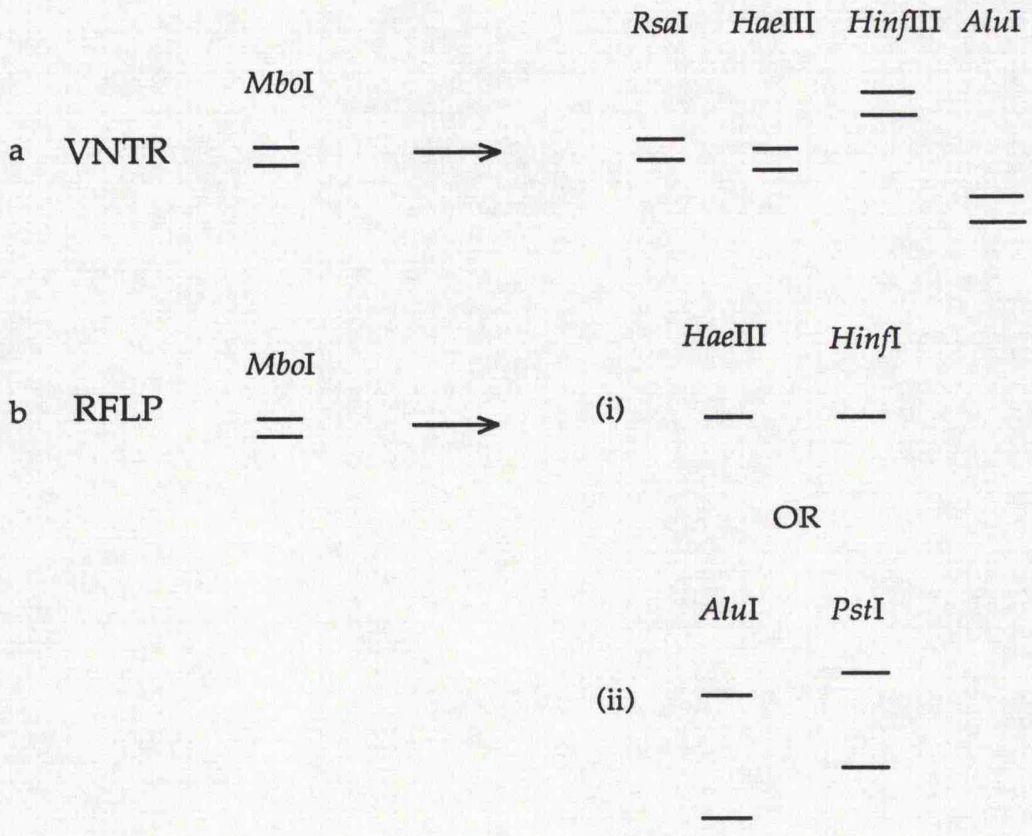


Figure 3.10 : Restriction analysis of dimorphic loci to determine the origin of the variation seen. a - Restriction analysis of a VNTR sequence. Digestion of a VNTR with multiple restriction enzymes that flank the VNTR would reveal diallelic patterns in an individual heterozygous for the locus, due to the homology present between the flanking regions. These patterns would be similar in that the difference in size between alleles would be identical.
b - Restriction analysis of a RFLP sequence. Digestion of a RFLP with multiple restriction enzymes, other than the restriction enzyme (*Mbo*I) used to display the dimorphism, would either :
(i) no longer display this variation due to the absence of a restriction site polymorphism for the given restriction enzyme or
(ii) display a variation due to a different restriction site polymorphism which would no longer show the original size difference between alleles.

3.6.5 Clones Detecting Monomorphic Single-locus Patterns

Table 3.5 groups together those clones that detect monomorphic loci. The patterns detected by such clones consist of a single-hybridizing fragment which showed no variation in size between the unrelated individuals tested. Three of the clones are identical in terms of the size of the locus detected. It is possible that these clones represent repeat isolates of the same original locus. This would be confirmed by the demonstration of cross-hybridization of the various loci at high stringency and the identical nature of restriction maps at the given loci. It should be noted that a number of monomorphic clones detect loci which are larger in size than the insert used to detect them. This is likely to be due to the collapse of the repeat array within the canine DNA fragments cloned, during propagation within NM554.

3.6.6 Comparison of Positive Clones Obtained with each Multi-locus Probe

Table 3.6 groups the clones in terms of the numbers obtained within each stated category and with respect to the probe that they originated from. It can be seen that a total of 17 clones cross-hybridize to both probes, 16 hybridize only to 33.6 and a further 15 hybridize only to 33.15. Jeffreys & Morton (1987) in the analysis of canine DNA fingerprints using 33.6 and 33.15, have found that only a limited number of canine loci cross-hybridize to both multi-locus probes. However, the somewhat high proportion of cross-hybridizing clones obtained in this study, seems to indicate that a greater proportion of canine minisatellite loci than was previously thought, cross-hybridize to both multi-locus probes. Both probes detect comparable numbers of clones in total. However, it is notable that 33.15 seems to detect a slightly greater number of clones within categories that are uninformative for the purposes of this project, and furthermore all the polymorphic clones that it does detect also cross-hybridize to 33.6. This could be an indication that 33.15 when used as a multi-locus probe to screen canine genomic libraries detects canine counterparts that generally tend to be uninformative. This feature has also been seen in the analysis of a minke whale library where 33.6 has been seen to detect more polymorphic loci than 33.15 (personal communication - Iris A. van Pijlen). Furthermore, Hanotte *et al.*, (1992a) have found that

Table 3.5 Clones detecting monomorphic single-locus patterns

Clone	Plate	Origin	Size of insert /kb	Locus size /kb
c*Cfa*MM1	5G2	33.15	2.30	3.30*
c*Cfa*MM2	6B6	33.6	3.50	3.40
c*Cfa*MM3	8A5	both	2.35	2.80
c*Cfa*MM4	9A3	both	3.60	3.60
c*Cfa*MM5	9A7	33.6	2.70	2.80
c*Cfa*MM6	9C2	both	2.55	3.30*
c*Cfa*MM7	10E5	33.15	2.45	4.00*
c*Cfa*MM8	20A8	33.15	1.90	2.80
c*Cfa*MM9	20H6	33.15	1.50	8.60*
c*Cfa*MM10	21C6	33.6	2.60	6.30*
c*Cfa*MM11	22A8	33.6	2.20	3.10
c*Cfa*MM12	23A11	33.15	2.50	3.50*

The nomenclature was derived from c*Cfa*MMN where : c - clone, *Cfa* - Canis familiaris, MM - monomorphic minisatellite, N - Nth clone isolated. * - the loci detected by these clones are larger than the inserts cloned.

Table 3.6 Summary of the Number and Types of Clones detected by the Multi-locus Probes 33.15 and 33.6

Morphism	33.15	33.6	both	Total
monomorphic	5	4	3	12
multi-band	5	4	6	15
satellite	4	3	1	8
polymorphic	-	3	4	7
dimorphic	1	2	3	6
total	15	16	17	48

33.15 cross-hybridizes strongly with abundant sequences present in chicken and Japanese quail DNA cut with different restriction enzymes, and suggest that the detection of satellite and satellite-like sequences might be more common than previously thought. However, it would be necessary to screen and analyse a much larger number of clones than those analysed here before this feature could be considered to be significant.

3.6.7 Yields of Informative and Uninformative Canine Minisatellite Loci

Table 3.7 gives a summary of the number and percentage of clones present within each category, in terms of both the total number of clones screened from the library, and also in terms of the total number of positive clones analysed. This indicates that only 1.63% (as opposed to the initial 5.8% - see Section 3.5) of the total number of clones screened have yielded loci which show similarity to the consensus 'core' minisatellite sequences within the muti-locus probes, 33.15 and 33.6. This percentage has been found to be similar to that obtained in a Charomid library of cloned chicken fragments enriched for minisatellite sequences (Bruford, M. - thesis - 1992). The decrease in the yield of 33.6 and 33.15-derived canine minisatellite positive clones has been attributed to a number of factors (see Section 3.5).

Table 3.7 Summary of the Analysis of the Positive Clones Obtained within each Morphism

Morphism	Number	C/%	P/%
monomorphic	12	0.41	25.00
multi-band	15	0.51	31.25
satellite	8	0.27	16.67
polymorphic	7	0.24	14.58
dimorphic	6	0.20	12.50

C - percentage of clones obtained in terms of the total number of clones screened; P - percentage of clones obtained in terms of the total number of positive clones analysed.

Approximately 42% of the positive clones analysed were uninformative. Of the remaining 58% (28), approximately half detected multiband patterns not easily scorable without further restriction or segregation analysis. A further 21% (6) of the remaining 28 loci were dimorphic and are likely to be limited in their informativeness. Such

probes would mainly be limited to use in segregation analyses where both parents are heterozygous with respect to the given locus. Thus, only 25% (7) of the total number of potentially informative positive clones were found to detect distinct, polymorphic and easily scorable single-locus patterns.

Therefore, considering the results obtained so far, the total number of positive clones that might be expected from the screening, using the probes 33.6 and 33.15, so that any given DNA fragment within the 2-16kb size fraction can be represented once i.e. screening 13,403 colonies, would be 424. However, this total might be expected to be reduced by about 48% due to reasons already stated to approximately 220 positive clones. Of these, 92 might be uninformative, leaving 128 potentially informative clones of which 32 might be expected to be distinct, polymorphic, easily-scorable single-locus clones.

DISCUSSION

A canine genomic library (SSJ1) has been generated in the vector, Charomid 9-36. 2940 colonies have been screened with the multi-locus minisatellite probes 33.6 and 33.15, from which 48 positive clones were obtained. These clones were screened for polymorphism by using them as probes at high stringency with the presence of canine competitor DNA to minimise the interaction of the repeat sequence in the probe with other repeat sequences, and thereby maximise the detection of the specific locus that had been cloned. Within the five categories of morphisms seen (multi-band, satellite, polymorphic, dimorphic and monomorphic) two were seen to be uninformative i.e. satellite and monomorphic patterns (Tables 3.2 & 3.5). Among the informative patterns, the determination of the origin of each multi-band pattern detected by 15 of the clones (Table 3.1) required further analysis of each probe via restriction and/or segregation analysis. This analysis has not been done within this study.

The other two categories of morphisms represented single loci that were either di- or multi-allelic. The dimorphic clones were limited in their degree of informativeness, i.e. their heterozygosity levels were not more than 50%. However, the clones categorized as polymorphic have heterozygosity levels up to 88%. Screening the library with 33.6 and 33.15 therefore resulted in the detection of easily-scorable, polymorphic loci by

25% of the informative clones isolated and 14.6% of the total number of positive clones isolated.

The main aim in generating a canine genomic library in Charomid 9-36 was to isolate as many polymorphic loci as possible with a view to ultimately using them as DNA markers in linkage analysis. The isolation of 28 informative clones within this study is therefore a step towards achieving this aim. However, it is clear that it will be necessary to analyse approximately half of these (15 clones detecting multi-band patterns) further, to determine their true nature so that accurate inferences can be made from their use in subsequent studies. Furthermore, the six dimorphic clones isolated can only detect two allelic states between individuals from different breeds, and are therefore likely to yield limited information with respect to most canine breeds that are known to be line bred.

Although other researchers do not find it uncommon to obtain clones containing multiple DNA inserts, it is clear that the proportion of positive clones containing multiple DNA inserts is quite high in SSJ1 (72%). This was probably due to an underestimation of the concentration of the size-selected DNA resulting in an excess of insert DNA with respect to the vector in the ligation. Furthermore, it is possible that the second size fractionation might have reduced the presence of contaminating small DNA within the 2-16kb size fraction, thereby also reducing the presence of multiple inserts within clones.

Table 3.8 compares the main characteristics of SSJ1 with those of a similar library generated from chicken DNA (Bruford - thesis (1992)) and also to those of a Charomid library generated from human DNA (Armour - thesis (1990)). Both the size fractions chosen and the relative sizes of the canine and chicken libraries are comparable. Furthermore, both the domestic dog and chicken species are known to consist of highly line bred populations.

The number of positive clones tested for polymorphism vary from the number of positive clones actually isolated for the canine and chicken libraries due to reasons already described in Section 3.5.3 of this chapter. Thus, for the chicken library, the initial number of positive clones have been reduced by 34% whereas those for the canine library have been reduced by 48%. Considering that only 1/5 of the percentage of the genome represented by the 2-16kb size fraction has been screened and a 48% reduction occurred in the initial number of positives, the yield of

Table 3.8 Comparison of the Main Characteristics of SSJ1 with Other Charomid Libraries

	Human	Chicken	Dog
Size Fraction Used/kb	4.0-9.0	2.0-7.0	2.0-16.0
Library Size	3×10^5	1.15×10^5	2.25×10^5
Colonies Screened	3123 (3)	4275 (4)	2940 (0.2)
No.of Positives Obtained	137	127	93
No.of Positives Screened	137	84	48
Number of polymorphic clones	39 (28%)	35 (42%)	7 (15%)
Number of dimorphic clones	3 (2%)	2 (2%)	6 (13%)
Number of monomorphic clones	28 (20%)	17 (20%)	12 (25%)
Number of satellite clones	12 (9%)	25 (30%)	8 (17%)
Number of multi-band clones	4 (3%)	5 (6%)	15 (31%)
Mean Hetero-zygosity/%	71%	76.4	43.2
Median Hetero-zygosity	77%	87.0	45.0

The numbers in brackets below each value of the number of colonies screened represent the genome equivalents corresponding to each library. The data for the human library have been taken from the thesis of J.A. Armour (1990) and those for the chicken library are from M. Bruford *et al.*, (1993) (in press). In the case of the human library the total percentage of clones screened for polymorphism is less than 100% due to the presence of repeat isolates.

polymorphic loci obtained from SSJ1 is not so low as might be initially thought. Therefore, were sufficient colonies (46,048) screened from SSJ1 to comprise ~ 5X the genome represented by the 2-16kb size fraction, the subsequent yield of polymorphic clones might be expected to be approximately 32. This is not significantly different from the corresponding number of polymorphic clones isolated from a 4 genome equivalent in the chicken library. However, it should be noted that reduction of observed yields of polymorphic clones could result from the repeated isolation of the same minisatellite locus.

The 13 clones which detect easily-scorable informative, single loci (i.e. both dimorphic and polymorphic clones) from SSJ1 have a mean heterozygosity of 43.2% and a median heterozygosity of 45%. (The value of a mean heterozygosity will be more affected by extremely low or high values, and therefore for nonsymmetrical distributions the median is considered to be a better measure for statistical description). Both the canine mean and median heterozygosities are considerably lower than those obtained for the chicken library represented. Calculation of the heterozygosities for the various polymorphic clones from the canine and chicken libraries involved the testing of each clone against unrelated line bred individuals.

Inbreeding has the effect of increasing the level of homozygosity within the genome (Willis, 1989).This might therefore be expected to have a concomittant reductive effect on the overall level of heterozygosity of minisatellite loci within the genome. Thus, it could be postulated that the isolation of minisatellite loci from the canine genome, would be likely to veer in the direction of large yields of monomorphic minisatellite loci among which any polymorphic loci isolated would have generally low overall levels of heterozygosity. Examination of the overall heterozygosites obtained for the various canine informative loci isolated in this study, certainly seems to support this postulate. However, since chickens are also known to be a line bred species, the considerably higher levels of heterozygosity observed is not consistent with the above postulate. Furthermore, since the percentages of monomorphic loci obtained with respect to the total number of positives screened, is 20-25% in chickens, dogs and humans, this would indicate that the level of inbreeding within a given population, does not influence the yield of monomorphic loci from Charomid libraries. Therefore, it is clear that until further analysis is done so that at least a full genome equivalent has

been screened from SSJ1 with the two multi-locus minisatellite probes, 33.6 and 33.15 (Jeffreys *et al.*, 1985a) as well as other multi-locus probes, little can be concluded about the overall levels of heterozygosity of minisatellite sequences within the canine genome.

It should be sufficient to screen SSJ1 with another 4 multi-locus probes in order to obtain a sizeable number of canine minisatellite informative loci: Armour *et al.*, (1992a) have compared the use of NTR (naturally occurring tandem repeat) probes with STR (synthetic tandem repeat) probes to screen a human ordered array charomid library to isolate hypervariable minisatellite loci. This study found that although a sizeable number of polymorphic loci (41) were detected by the use of a total of 10 NTR and STR probes, most of the loci would have been detected using a smaller selection of the probes. They infer therefore that for those whose initial priority is to isolate a large number of informative loci quickly, that little is gained from the use of more than 4 to 6 multi-locus probes.

Table 3.8 shows that approximately 31% of the total positives screened for polymorphism detect multi-band patterns. It is possible that further segregation analysis for such clones where multiple bands represent multiple loci, or restriction analysis for those clones where multiple bands represent internal *Mbo*I restriction sites, will cause some of these clones to be categorized together with the present polymorphic clones. However, judging from the numerous loci (i.e. more than two) detected by some of these clones at high stringency, such loci might be useful as canine-specific, multi-locus probes (see Figure 3.8 c).

The 2-16kb genomic size fraction has been seen to contain the most variable loci within 33.6 and 33.15 DNA fingerprints. However, it is clear from the analysis of the positive clones analysed that most of the DNA isolated from such a size-selected genomic fraction consists mainly of DNA fragments within the lower size range. The lower regions of canine DNA fingerprints have been seen to consist mainly of invariant loci (Jeffreys & Morton, 1987). Consequently, the proportions of less informative loci are likely to be higher in lower genomic size fractions. This might have had subsequent additive consequences on the yield of polymorphic loci with lower levels of heterozygosity. Thus, it would probably be useful to generate and analyse together with SSJ1, another library containing cloned canine genomic DNA fragments within a higher size range, e.g. 6 - 16kb.

Section 3.5.2 describes the re-screening of clones containing multiple inserts with the respective multi-locus probes. Armour *et al.*, (1992a) have compared the sequences of human minisatellite loci detected to the sequences of the probes whereby they were isolated. These comparisons have shown that 'a relatively poor match over a short region is sufficient to produce a positive hybridization signal in library screening,' and therefore, this has been thought to account for the substantial overlap seen between sets of loci detected by some multi-locus probes. It is possible therefore, that in the process of such elimination of insert fragments, that minisatellite sequences which hybridize to 33.6 and 33.15 relatively weakly, but possibly represent loci that are detected by other multi-locus probes, have been eliminated from subsequent analysis.

J. Armour (thesis - 1990) has screened an ordered array Charomid library containing cloned human DNA fragments with 6 multi-locus probes, and having obtained an initial number of polymorphic clones, used one of these to re-screen the library. This resulted in the isolation of 48 further positive clones, 32 of which were unique to the single-locus probe used. It is likely, therefore, that were a similar strategy applied to SSJ1, a further source of positive clones might be obtained, a significant proportion of which might be expected to be unique to the canine genome.

Inferences

(i) Although the overall level of heterozygosity of the polymorphic clones isolated to date is relatively low, this does not necessarily reflect the overall *in vivo* state of canine polymorphic minisatellite loci.

(ii) Comparison of the yields of monomorphic loci isolated within SSJ1 with those from libraries generated from chicken and human genomic DNA, indicates that the levels of inbreeding within a species does not necessarily reflect on the relative yields of monomorphic versus polymorphic loci.

(iii) Future analysis of the library generated might include the following:

  (a) The screening of a further 13,403 colonies so that a one genome equivalent might be analysed.

  (b) Screening of a one genome equivalent using a further 4 multi-locus probes to isolate a sizeable number of informative, canine minisatellite loci.

(c) The use of a hypervariable canine minisatellite sequence (e.g. c*Cfa*MP5) to re-screen the library, which might result in the isolation of a further subset of minisatellite sequences which are unique to the minisatellite sequence used.

(iv) The generation of another library containing fragments from the 6-16kb size range might facilitate the isolation of polymorphic loci with higher levels of heterozygosity.

APPENDIX

FOR THE CANINE GENOME :

N (required number of recombinants $= \frac{\ln (1 - P)}{\ln (1 - f)}$

$$= \frac{\ln (1 - 0.99)}{\ln(1 - 0.0001)}$$

$$= \frac{-4.605}{-0.0001}$$

= 46,048 colonies need to be screened to have a 99% probability that any given fragment in the 2-16kb fraction is represented within the library, where :

P (desired probability) = 0.99

f (fractional proportion of the size-selected genome in a single recombinant) = $(3 \times 10^3) \div 0.01(3 \times 10^9) = 0.0001$.

If only 55% of the colonies are recombinant, $\Rightarrow$ 46,048 $\equiv$ 25, 326 recombinant colonies.

Therefore, 46,048 x $\frac{100}{55}$ = 83,725 colonies need to be screened in total from SSJ1 to ensure a 99% probability that any given fragment is represented at least once.

Total length of DNA represented by 46,048 colonies each containing a 3kb DNA insert = 46,048 x 3 = 138,144 = 1.4×10^5 kb = 1.4×10^8 bp.

2-16 kb size fraction represents ~ 1% of total genome = 3×10^7 bp.

Since, $\frac{(1.4 \times 10^8)}{(3 \times 10^7)} = 4.7$.

Therefore, the 2-16kb genomic size fraction would be represented ~ 5X after screening 46,048 colonies.

A single representation of the 2-16kb size fraction would therefore require the screening of 9797 colonies.

But only 55% of 9797 colonies from the library are recombinant, therefore:

9797 colonies actually represent 5389 colonies with inserts, and a further 4408 recombinant colonies are required.

Therefore,
No. of colonies required for a single representation of the 2-16kb size fraction $= \frac{(4408)}{(5389)} \times 9797 + 5389$

$= 8014 + 5389 = 13{,}403.$

Therefore, 2940 colonies represent ~1/5 of the total number of colonies required for a single representation of any given DNA fragment with a 99% probability within the 2-16kb size fraction.

CHAPTER 4

APPLICATIONS OF CANINE MINISATELLITE LOCI

SUMMARY

A number of canine minisatellite loci were used as multi- and single-locus probes for various analyses : Mendelian inheritance was demonstrated for the two polymorphic canine minisatellite loci tested. In both cases, both parents were heterozygous at these loci although this Irish Setter breed is known to be highly inbred. Thus, despite the presence of high levels of inbreeding within the canine population, analysis of specific pedigrees should be possible, providing a bank of variable minisatellite loci are available.

The use of the two polymorphic loci, c*Cfa*MP1 and c*Cfa*MP5 (see Table 3.3, Chapter 3) to analyse 18 Bedlington Terriers demonstrated a reduction in the level of variation compared to that seen in the population as a whole. This was not surprising since high levels of inbreeding are generally known to be present within pedigree populations of domestic dogs.

Analysis of species related to the domestic dog using the most variable of the canine minisatellite loci (c*Cfa*MP5) detected equivalent loci (all of which were variable) in only 3/35 wolves and coyotes at high stringency. The lack of detection in 32 of the related canids might be due to the presence of null alleles resulting in the inability to detect these loci under normal high stringency conditions. Detection of the c*Cfa*MP5 equivalent in 3/35 canids suggests the maintenance of variability at this locus in a proportion of the canid population since the advent of domestication in the Canidae family. The detection of similar invariant loci of identical size in wolves and coyotes by canine monomorphic minisatellite loci confirms their stability in these sub-families. The use of c*Cfa*MP5 as a multilocus probe to screen unrelated species has shown that related minisatellite loci exist in a wide variety of species.

The resolution of a paternity dispute has been achieved via the use of c*Cfa*MP5 as a single-locus probe. This latter analysis also identified the presence of offspring for whom the parental combination had been wrongly assigned.

Finally, inter - breed and intra - breed analysis using DNA fingerprint analysis has shown that intra - breed variability is two fold less than that between breeds. Although this difference in variability was shown to be statistically insignificant, it would be surprising if a more detailed analysis did not reveal the level of substructuring currently known to be present within the canine population.

INTRODUCTION

The isolation of individual minisatellite loci in man and subsequently in a wide spectrum of species of which some are birds (Bruford & Burke, 1994), fish (Bentzen *et al.*, 1993), and insects (Jacobson *et al.*, 1992), has demonstrated their use in a wide variety of analyses. In man, birds, pigs and the domestic dog, these loci have been isolated in the main to serve as DNA marker loci in genomic linkage maps. The isolation of polymorphic minisatellite loci in an inbred species such as the domestic dog although successful, has been slow, due to the repeated isolation of numerous uninformative minisatellite loci.

However, as Chapter 3 describes, 7 polymorphic canine minisatellite loci have been isolated and characterized. The use of such loci in genetic analyses are manifold. This chapter simply serves to illustrate a number of these uses for the canine genome as well as related and unrelated species. This was done by :

(i) demonstrating Mendelian inheritance of these loci in pedigrees,
(ii) showing limited variation of these loci in specific breeds,
(iii) detecting related loci in other canids and
(iv) both resolving a paternity dispute and detecting the presence of offspring for which the parental combination has been incorrectly assigned.

Finally, an inter - and intra - breed analysis has been done of canine DNA fingerprints following on from the initial work of Jeffreys & Morton (1987). Although this analysis cannot be considered comprehensive, it suggests that the degree to which intra - breed variation coincides with that between different breeds depends on the inbreeding coefficient of a specific breed.

RESULTS / DISCUSSION

4.1 Mendelian Inheritance of Two Polymorphic Canine Minisatellite Loci

The canine minisatellite clones, c*Cfa*MP4 and c*Cfa*MP5, were used as single-locus probes at high stringency (Methods Section 2.23.3) against Southern blots of *Mbo*I - digested canine genomic DNA from a family of Irish Setters. This Irish Setter family is known to have members that are affected by PRA (progressive retinal atrophy) (Millichamp *et al* 1988). Both clones detected informative single-locus patterns showing simple Mendelian inheritance of the respective minisatellite loci within the family. c*Cfa*MP4 detected two alleles of 3.35kb and 3.9kb (Figure 4.1). Both the parents, RFS (sire) and HON (dam) were heterozygotes. 6/9 offspring were homozygous for the 3.35kb allele, whereas the remaining three had identical genotypes to the parents. Only two of the possible three genotypes were observed among the offspring. However, it is likely that were a larger number of meioses examined for the same two parents at this locus, the third genotype (i.e. homozygosity for the 3.9kb allele) would eventually manifest itself. Parental origins of the various allele combinations seen at this locus could not be determined from this analysis due to the identical genotypes of both parents. Co-segregation could not be seen between PRA and either of the two alleles at this locus.

c*Cfa*MP5 detected three alleles of 8.0kb, 2.2kb and 1.75kb within the Irish Setter family (Figure 4.2). The sire was heterozygous for alleles of 2.2kb and 1.75kb, whereas the dam was heterozygous for alleles of 8.0kb and 1.75kb. The genotypes of the offspring displayed typical allele combinations as might be expected from an independent assortment of parental alleles. The parental origin of the various allele combinations seen in the offspring could be determined in all cases except for those individuals that were homozygous at this locus for the 1.75kb allele. None of the three alleles present at c*Cfa*MP5 could be seen to co-segregate with the PRA locus in this family.

4.2 DNA Fingerprint Analysis of Inter - and Intra - breed Variation

Jeffreys & Morton (1987) have studied the levels of band sharing within the DNA fingerprints of cats and dogs using the multilocus

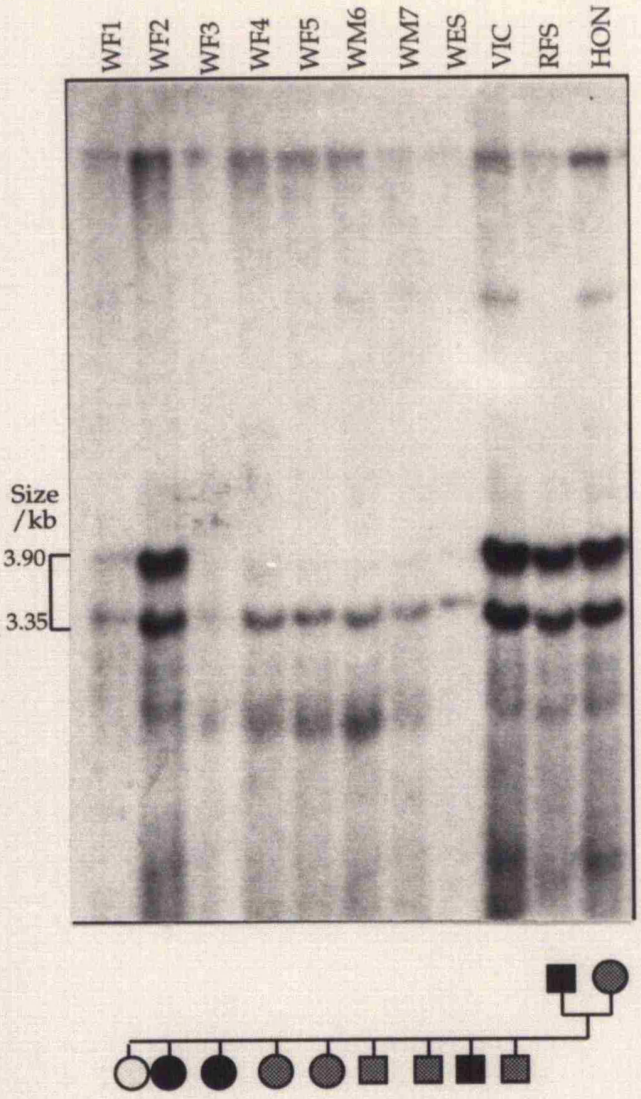


Figure 4.1 : Simple Mendelian inheritance of the canine minisatellite locus, c*Cfa*MP4 in an Irish Setter family.
WF1 - WF5 - female offspring, WM6, WM7,
WES & VIC - male offspring, RFS - sire, HON - dam. PRA (progressive retinal atrophy) is known to segregate within this family. c*Cfa*MP4 does not segregate with the PRA locus.

- female carrier
- male carrier
- unknown genotype
- affected female
- affected male

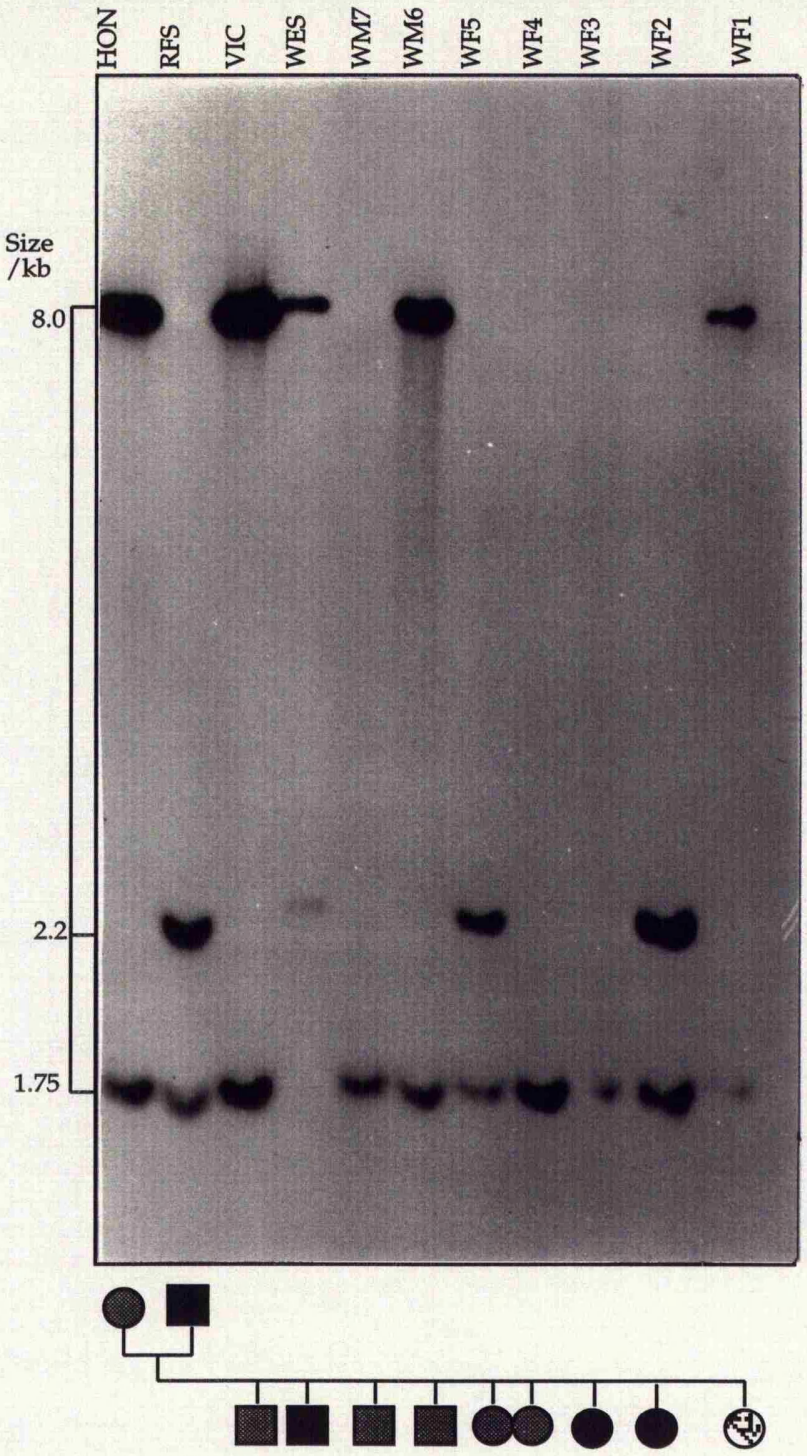


Figure 4.2 : Simple Mendelian inheritance of the canine minisatellite locus, c*Cfa*MP5 in an Irish Setter family. HON - dam, RFS - sire, WF1 - WF5 - female offspring, WM6, WM7, WES & VIC - male offspring. PRA (progressive retinal atrophy) is known to be segregating in this family. c*Cfa*MP5 does not segregate with the PRA locus.

- female carrier
- male carrier
- unknown genotype
- affected female
- affected male

minisatellite probes, 33.6 and 33.15. They were able to show that the level of band sharing between different dog breeds is similar to that between different shorthair domestic cats. This level of band sharing was seen to be considerably higher than that seen in humans, suggesting a lower level of variability in cats and dogs. Their study included the DNA fingerprints of two Whippets which were not seen to be significantly more similar to each other than to other dog breeds. They therefore suggested that intra - breed variability is unlikely to be substantially less than that seen between breeds. However, a subsequent review of the application of DNA fingerprinting to animal and bird populations has stated, that 'it would be surprising if a more thorough study revealed no breed structure.' (Hill, 1987).

The number of loci identified by the multilocus probes 33.6 and 33.15, in dogs has been estimated to be ~ 13/ individual (see Introduction, Section 1.3). Although instances of allelism and linkage have been found in a sibship of 10 'greagles' (offspring of a greyhound and a beagle), a more detailed analysis of the segregation of pairs of parental loci revealed no evidence for clustering in the genome suggesting that these loci must be dispersed at least to a certain extent over canine autosomes (Jeffreys & Morton 1987). This would suggest that most of the bands scored per individual dog represent individual loci thereby probably having minimal effects on the estimations of intra- and inter-band variability obtained in the following analysis. Such information is important since evidence for allelism among DNA fragments in DNA fingerprints of the parrot, Amazona ventralis, has been found by Brock & White (1991) where as few as 2-5 loci were identified by the multilocus probe, 33.15.

*Alu*I-DNA fingerprints of pairs of individuals from various breeds, using the multilocus minisatellite probes, 33.6 and 33.15, were generated (Figures 4.3 & 4.4). Due to the faintness of the bands within the obtained 33.15 *Alu*I - DNA fingerprint (seen also by Jeffreys & Morton - 1987), the following analysis was only done on the pattern generated by the 33.6 *Alu*I-DNA fingerprint of individuals from six breeds that displayed bands of sufficient intensity : Tables 4.1 and 4.2 show the levels of inter- and intra-breed band sharing seen in the 33.6 *Alu*I-DNA fingerprint. (The value varies from zero when there are no bands in common to one when the two tracks are identical). Band analyses were restricted to the larger, more clearly resolved fragment sizes i.e. above 2kb. Furthermore, band sharing coefficients were determined only from comparisons of adjacent

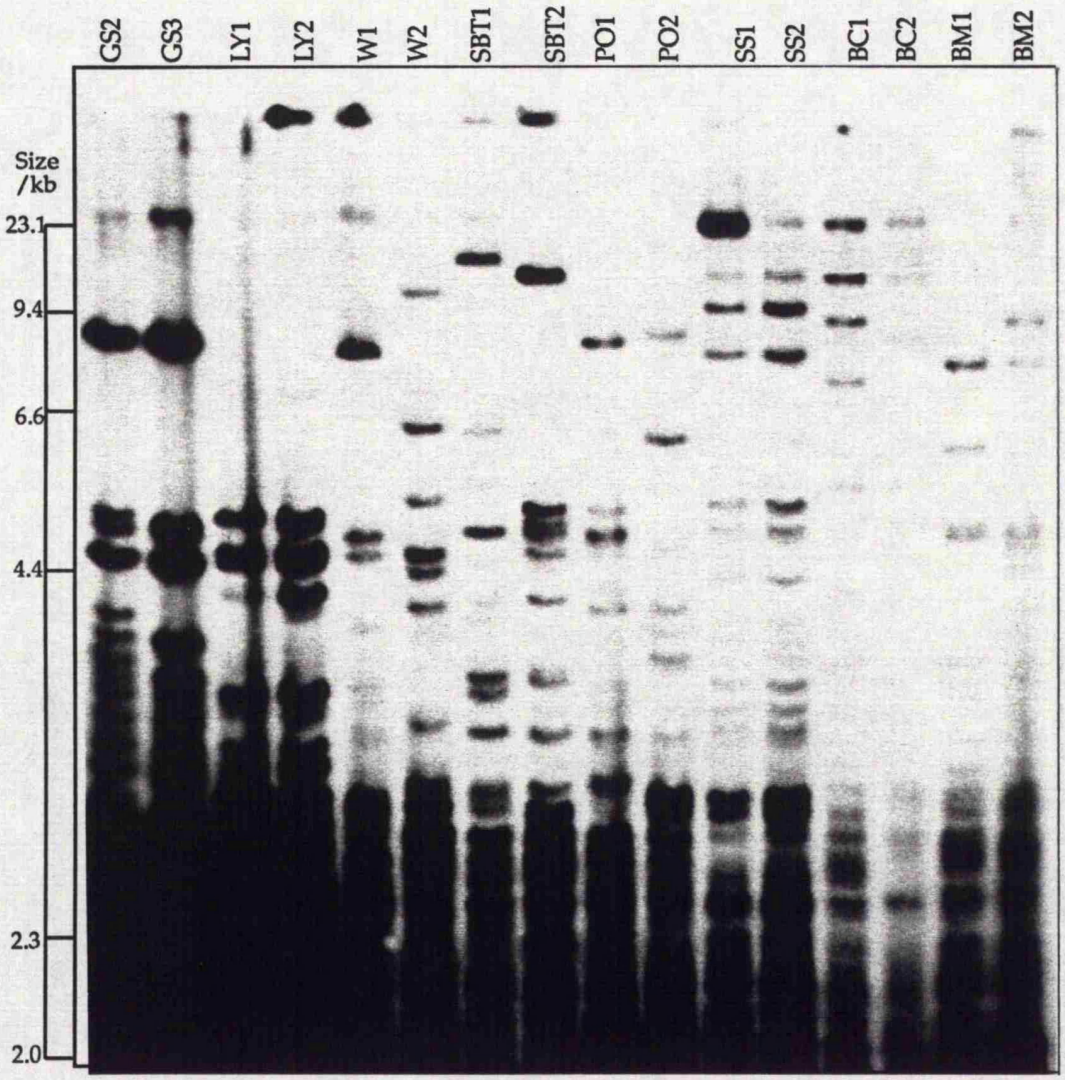


Figure 4.3 : An *Alu* I DNA fingerprint of 8 pairs of individuals from various canine breeds using 33.6.
Breeds analysed were as follows: GS2 & GS3 - German Shepherds, LY1 & LY2 - Labradors (LY2 is a female), W1 & W2 - West Highland Whites, SBT1 & SBT2 - Staffordshire Bull Terriers, PO1 & PO2 - Poodles, SS1 & SS2 - Springer Spaniels, BC1 & BC2 - Border Collies, BM1 & BM2 - Bernese Mountain Dogs. The size range of minisatellite DNA bands detected is indicated by λ *Hind* III molecular weight markers. All breeds consisted of unrelated individuals of unknown sex. The autoradiogram shown was exposed to x-ray film with screens for 48hrs.

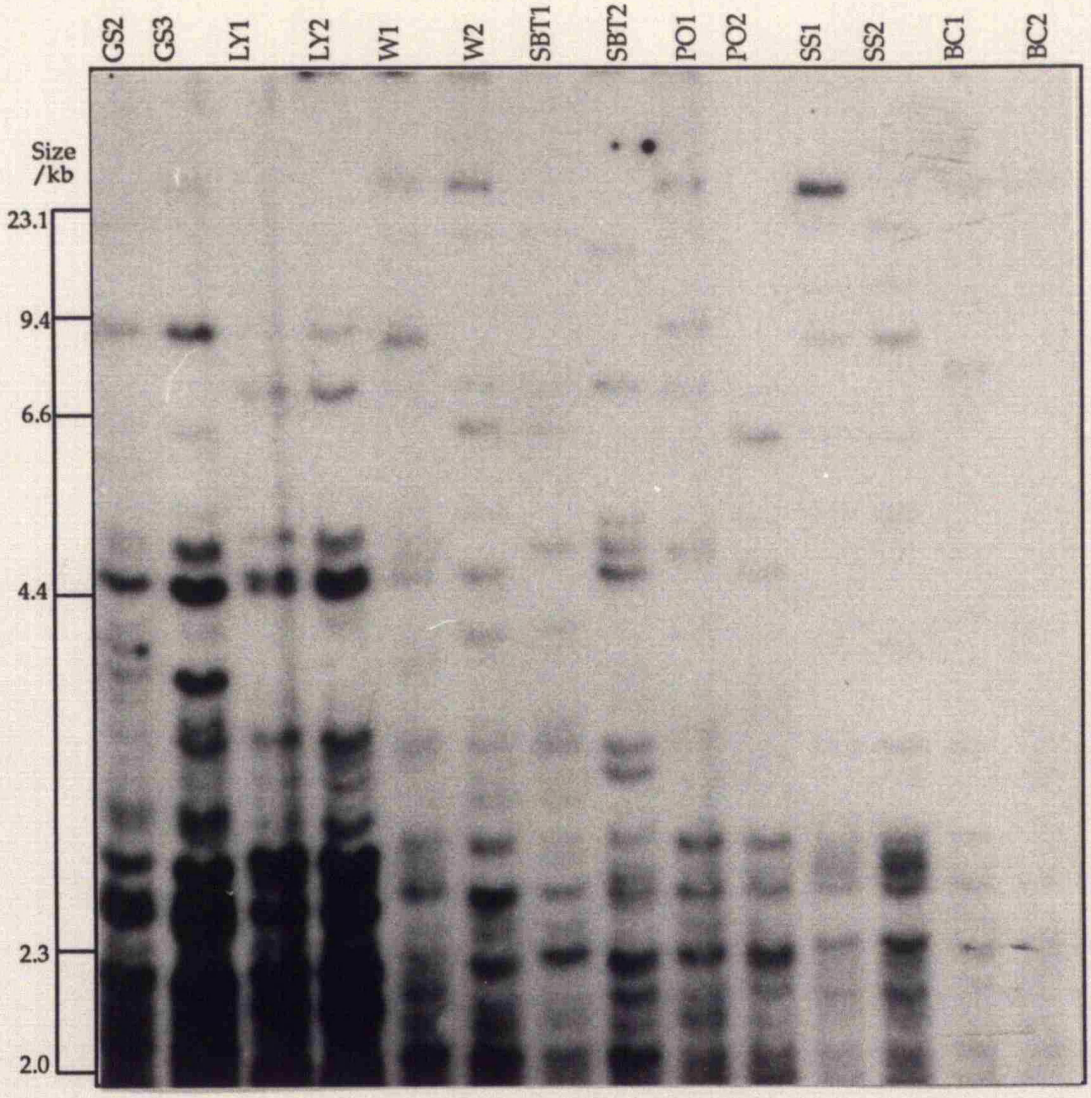


Figure 4.4 : An *Alu* I DNA fingerprint of 7 pairs of individuals from various canine breeds using 33.15.
Breeds analysed are as follows: GS2 & GS3 - German Shepherd, LY1 & LY2 - Labradors (LY2 is a female), W1 & W2 - West Highland Whites, SBT1 & SBT2 - Staffordshire Bull Terriers, PO1 & PO2 - Poodles, SS1 & SS2 - Springer Spaniels, BC1 & BC2 - Border Collies. The size range of minisatellite DNA bands detected is indicated by λ *Hind* III molecular weight markers. All breeds were unrelated individuals of unknown sex. The autoradiogram shown was exposed to x-ray film for 1 week.

Table 4.1 Inter - breed variation within *Alu*I - DNA fingerprints of individuals from 6 pedigree breeds.

	Bands scored	No. of unique bands	Bands shared	Band sharing coefficient
GS3	12	17	2	0.19
LY1	9			
LY2	11	18	1	0.10
W1	9			
W2	7	16	2	0.20
SBT1	13			
SBT2	10	13	3	0.32
PO1	9			
PO2	7	19	2	0.2
SS1	16			

The mean band sharing coefficient is 0.17. GS3 - German Shepherd, LY1& LY2 - Labradors, W1 &W2 - West Highland Whites, SBT1 & SBT2 - Staffordshire Bull Terriers, PO1 & PO2 - Poodles, SS1 - Springer Spaniel.

Table 4.2 Intra - breed variation within *Alu*I - DNA fingerprints of individuals from 6 pedigree breeds.

	Bands scored	No. of unique bands	Bands shared	Band sharing coefficient
GS2	14	8	9	0.69
GS3	12			
LY1	9	4	8	0.80
LY2	11			
W1	9	14	1	0.13
W2	7			
SBT1	13	21	1	0.09
SBT2	10			
PO1	9	12	2	0.25
PO2	7			
SS1	16	3	15	0.91
SS2	17			

The mean band sharing coefficient is 0.48. GS2 & GS3 - German Shepherd, LY1& LY2 - Labradors, W1 &W2 - West Highland Whites, SBT1 & SBT2 - Staffordshire Bull Terriers, PO1 & PO2 - Poodles, SS1 & SS2 - Springer Spaniels.

fingerprints. First of all, it can be seen that there is a approximately two fold difference between the mean intra- and inter-breed band sharing coefficients. Furthermore, the mean inter-breed band sharing coefficient obtained by Jeffreys & Morton (1987) (0.46) and Georges *et al.*, (1988) (0.45) is approximately two fold higher than that obtained here (0.20).

Hanotte *et al.*, (1992b) have found that the choice of restriction enzyme for DNA fingerprint analyses in gallinaceous birds affects the index of similarity (or the band sharing coefficient), although there is strong concordance between these two parameters within species. The results obtained here indicate that this is not the case in the domestic dog i.e. the choice of restriction enzyme does seem to affect the band sharing value within *Canis familiaris* . The studies of Jeffreys & Morton (1987) and Georges *et al.*, (1988) used the restriction enzymes *Hinf* I and *Hae* III respectively.

The two fold difference seen between the mean intra- and inter-breed band sharing coefficients seemed to indicate that there is a considerable difference in the degree of band sharing, between individuals from the same breed to that between individuals from different breeds. Certainly, German Shepherds, Labradors and Springer Spaniels all display relatively high levels of band sharing. High genetic similarity as indicated by band sharing coefficients might be due to unusually low minisatellite mutation rates, strong selection or recent common ancestory among dogs. The latter of three explanations is likely to be the case in dogs. On the other hand, Poodles, West Highland Whites and particularly Staffordshire Bull Terriers have considerably lower levels of band sharing similar to those obtained when comparing individuals from different breeds (see Table 4.2). The higher levels of band sharing among German Shepherds, Labradors and Springer Spaniels are thus likely to be indicative of a greater degree of genetic homogeneity within these breeds compared with the others.

The significance of the two fold difference was tested using a Mann-Whitney test and found to be non-significant ($U = 4 > 3$ when $n_1 = 6$ and $n_2 = 5$). These results therefore support the findings of Jeffreys & Morton (1987) (see above) indicating that intra-breed variability is not significantly less than that between individuals from different breeds. However the general canine population consists of subpopulations (breeds) among which the degree of genetic similarity probably depends on the evolutionary ancestory of individual breeds. The present analysis simply

shows that a study of the levels of inter- and intra-breed variability will probably require much larger samples for band sharing analyses to detect a significant level of substructuring within the canine population. It can be seen however that the range of band sharing values between individuals from the same breed can vary considerably from breed to breed. The level of intra - breed band sharing probably reflects the level of inbreeding within a particular breed. It was not possible to calculate inbreeding coefficients for the various breeds analysed here due to the paucity of information regarding their respective pedigrees. However, it would have been interesting to correlate this information with the degree of band sharing seen within these DNA fingerprints.

Such correlations have been done by Kuhnlein *et al.*, (1990) in defined strains of chickens where inbreeding coefficients are known per strain. They found that the mean band sharing increases with the level of inbreeding, although the relationship between band sharing and the inbreeding coefficient is nonlinear and has to be fitted by a higher order approximation. However, providing that the frequency of a given band is known within the population, the average band frequency can be determined. They were able to determine the average band frequencies in seven strains of chickens and showed that this increased approximately linearly with the level of inbreeding. Thus, to correlate the level of inbreeding with the levels of band sharing seen within canine breeds it would be necessary to :

(i) analyse numerous individuals from various breeds to determine the average band frequencies within the population corresponding to a particular breed, and

(ii) to analyse breeds with known pedigrees so that inbreeding coefficients can be determined accurately.

Inbreeding has been traditionally thought to have undesirable effects on the viability of a species. However, such beliefs are not always true in nature as is shown by the high genetic similarity of different colonies of the eusocial naked mole-rat. Reeve *et al.*, (1990) found that the success of the colonies lies in the use of inbreeding to secure the future of existing genes and results in each member carrying the same highly adapted and desirable combination of genes. The only problem with this is that as soon as there is an adverse change in the environment or if disease strikes, 'all inmates of the colony are affected equally' (Young, 1990). Animals other than dogs that are believed to have undergone

extensive inbreeding are koalas, due to isolation from other breeding groups (Taylor *et al* 1991; Cocciolone & Timms, 1992), mice (Mannen *et al.*, 1993), captive cheetahs (O'Brien *et al.*,1985) and beavers (Ellegren *et al.*, 1993).

4.3 Analysis of the Variability of Two Canine Polymorphic Minisatellite Loci Within and Between Canine Breeds

The two canine minisatellite clones, c*Cfa*MP1 and c*Cfa*MP5, were used as single-locus probes (see Methods Section 2.23.3) against *Mbo*I - digested canine DNA from various breeds and their heterozygosities determined to be 59% and 88%, respectively (see Chapter 3, Table 3.3). They were then used in turn to study individuals from a Bedlington Terrier breed which is known to be highly inbred. In a number of cases these individuals are the offspring of incestuous matings (data not shown).

Inbreeding results in an increase in the number of loci that exist in a homozygous state within the genome. The consequence of this on minisatellite loci depends on the inherent variability of the given locus. If a minisatellite locus is highly variable, such as the human minisatellite, MS1, it is less vulnerable to the effects of inbreeding and very similar allele frequency distributions are found even among radically different ethnic groups (Jeffreys *et al.*, 1991b).

Figure 4.5 shows the result of using c*Cfa*MP5 as a single-locus probe to analyse 18 individuals from the Bedlington Terrier breed and 4 other cross breeds. There are four alleles at this locus in Bedlington Terriers, of 9.4kb, 8.2kb, 7.6kb and 6.8kb. 13/18 individuals share the 9.4kb allele. On the other hand, the four Terrier cross-breed dogs have a number of different-sized alleles to those in the Bedlington Terrier breed, indicating the greater level of variation present at this locus in the total population as a whole (see Table 3.3, Chapter 3). c*Cfa*MP5 detects at least 35 different alleles among 25 unrelated individuals from different breeds (see Chapter 3, Figure 3.9). Yet this variation is reduced to only four alleles in Bedlington Terriers. This suggests that Bedlington Terriers could have lost variability subsequent to domestication from the founder population of grey wolves (Wayne, 1993).

It should be noted however that the number of alleles detected depends on the size of the population analysed and the resolving power of

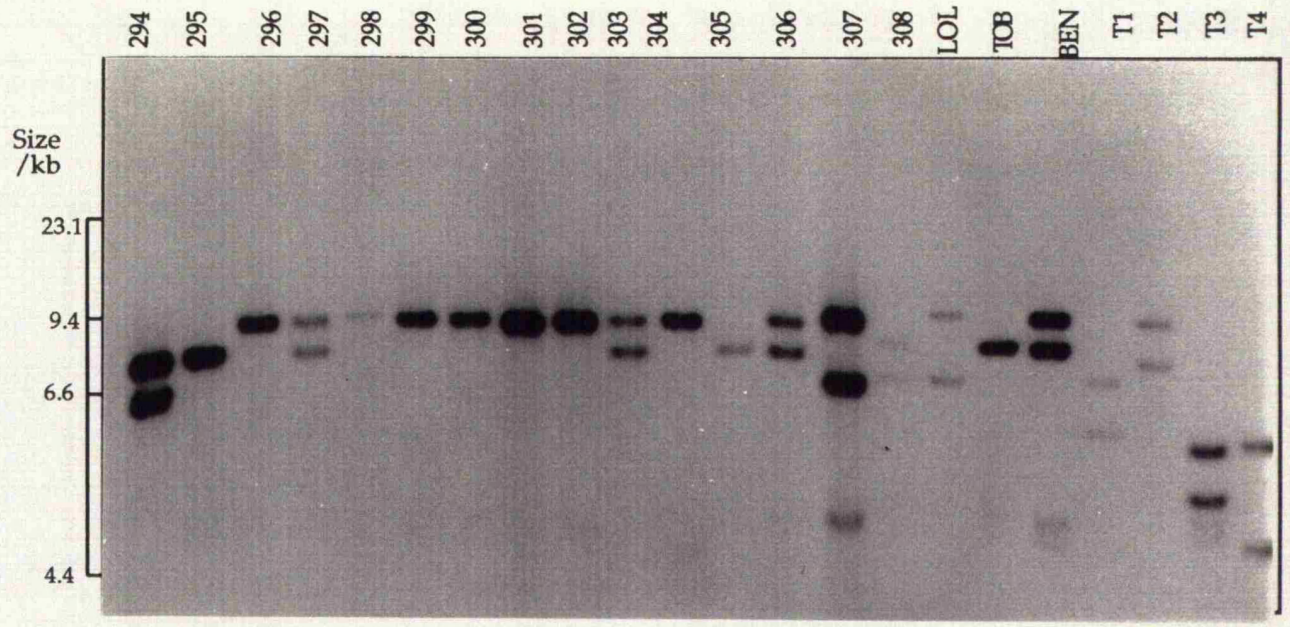


Figure 4.5 : Southern blot of *Mbo*I - digested canine DNA from 18 Bedlington Terriers and 4 Terrier crosses probed with c*Cfa*MP5 at high stringency. Individuals 294 - 308, LOL, TOB & BEN are Bedlington Terriers. T1 - T4 are Terrier crosses. λ *Hind* III molecular weight markers are shown. Four alleles are seen within the Bedlington Terrier breed, of 9.4kb, 8.2kb, 7.6kb and 6.8kb. The heterozygosity of c*Cfa*MP5 within this breed is therefore ~ 44%. 7 alleles which are not present in the group of Bedlington terriers can be seen to be present within the four Terrier cross dogs analysed.

the electrophoretic system. In the construction of allele frequency tables, alleles that differ only by a few repeat units may not be distinguished from each other. Furthermore, alleles of extreme size (low molecular weight) will not be detected due to lack of hybridization thereby increasing the apparent homozygosity and affecting the frequency estimates accordingly.

The heterozygosity seen at c*Cfa*MP5 in this breed approximates to ~ 44%, a further reflection of the marked reduction in the level of variation at this locus for this breed. A major factor in this reduction is present in the fact that a number of these individuals are directly related to each other and all are related to each other indirectly (data not shown). 7/18 individuals analysed here are homozygous for the 9.4kb allele and 3/18 individuals are homozygous for the 7.6kb allele. Similar losses in heterozygosity as measured by allozyme electrophoresis between individuals of geographically separated wolf packs have been found by Wayne *et al.*, (1991a). However, they have found that a significant proportion of the variability found among mainland wolves is retained in isolated island wolves - the amount lost being consistent with that expected given complete isolation and small effective population size. In a similar vein, Polynesians have been shown to display a lower level of heterozygosity than Melanesians, the reduced genetic diversity being attributed to genetic bottlenecks (Martinson *et al.*, 1993).

Figure 4.6 shows the results of probing the same filter as that in Figure 4.5 with c*Cfa*MP1 at high stringency. As in the case of c*Cfa*MP5, a limited number of alleles (5) are seen in the Bedlington Terrier breed compared to those seen in the population as a whole (18). A Kolmogorov-Smirnov two-tailed test of the null hypothesis that the total allele distributions are identical between population and breed for the minisatellites c*Cfa*MP1 and c*Cfa*MP5 could be rejected (c*Cfa*MP1: $D_{mn} = 0.49$, $p < 0.01$, c*Cfa*MP5: $D_{mn} = 0.42$, $p < 0.01$). The sizes of the alleles seen in Bedlington Terriers for c*Cfa*MP1 are 6.2kb, 4.9kb, 4.8kb, 4.6kb and 4.2kb. The level of heterozygosity seen is 33%, a reduction of 26% from that seen in the population as a whole. 11/18 of the Bedlington terriers were homozygotes at c*Cfa*MP1. Of these, 9 were homozygous for the 4.8kb allele, one was homozygous for the 4.9kb allele and one was homozygous for the 4.6kb allele. For both c*Cfa*MP1 and c*Cfa*MP5, it would be invalid to test for Hardy-Weinberg equilibrium in both the breed and the population samples due to the way in which the sampling was done; nothing is known about the generations as to which the various individuals come

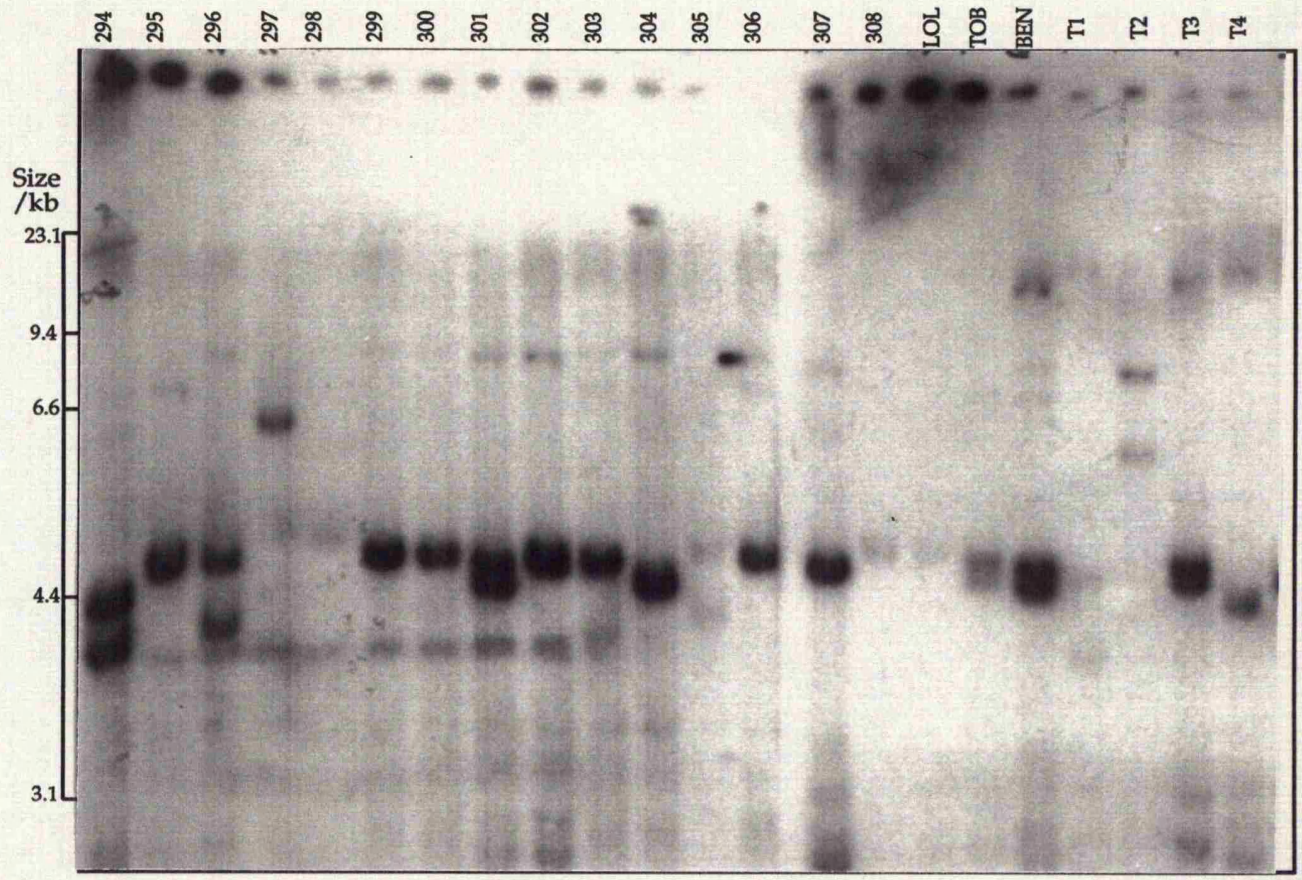


Figure 4.6 : Southern blot of *Mbo*I - digested canine DNA from 18 Bedlington Terriers and 4 Terrier crosses probed with c*Cfa*MP1 at high stringency. Individuals 294 - 308, LOL, TOB & BEN are Bedlington Terriers. T1 - T4 are Terrier crosses. λ - λ *Hind*III molecular weight markers are shown. Five alleles are seen within the Bedlington Terrier breed of 6.2kb, 4.9kb, 4.8kb, 4.7kb, 4.6kb and 4.2kb. The heterozygosity of c*Cfa*MP1 within this breed is therefore approximately 33%. At least five of the alleles present in the Terrier cross individuals are not present in the Bedlington Terriers.

from within the population as a whole and in the breed sample the individuals come from different generations (data not shown).

The reductions in the number of alleles from population to breed can be seen visually in Figure 4.7 which shows the allele frequency distributions obtained for c*Cfa*MP1 and c*Cfa*MP5 both in Bedlington Terriers and the population. Alleles have been detected for the size ranges of 3.3 - 11.5kb for c*Cfa*MP1 (mean allele frequency, q = 0.04) and 1.75 - 9.7kb for c*Cfa*MP5 (q = 0.03) in the population. The corresponding mean allele frequencies in c*Cfa*MP1 and c*Cfa*MP5 for Bedlington Terriers are q = 0.20 and q = 0.25, respectively. For c*Cfa*MP1, the 4.8kb allele in Bedlington Terriers corresponds to the second largest peak in the population sample, whereas in the case of c*Cfa*MP5 the 9.4kb allele corresponds to one of two peaks both of which are equal in size. c*Cfa*MP1 shows a bimodal allele frequency distribution. How might these modes be generated? Two possible explanations are:

1) multiple modes may indicate some form of genetic heterogeneity within the population
2) current distributions of alleles reflect their evolutionary antiquity which assumes that the modal classes reflect alleles that are older than others (see below).

Studies of VNTR allele frequencies in ethnic human populations have shown that alleles shared among populations are often the most frequent allele within each population (Chapman *et al.*, 1986; Deka *et al.*, 1991; Flint *et al.*, 1989). This suggested that the most common alleles are the most ancient and therefore, ancestral (Deka *et al.*, 1991). Such a circumstance may also apply to canine alleles which show a greater frequency both within the general population and individuals breeds. Furthermore the presence of multiple modes in the population allele frequency distributions supports the currently held view that *Canis familiaris* consists of a heterogeneous mixture of numerous subpopulations.

The genotypes of individuals 297 and 298 are quite different from those of the other Bedlington Terriers for c*Cfa*MP1 i.e. 297 is heterozygous at this locus with two alleles of 6.2kb and 4.9kb, whereas 298 is homozygous for the 4.9kb allele, neither of which are present in any of the other Bedlington Terriers. This anomaly can be explained by examination of the pedigree of these two individuals which shows that the degree of relatedness between individuals 297 and 298 and the rest of the individuals in this Bedlington pedigree is relatively small (data not

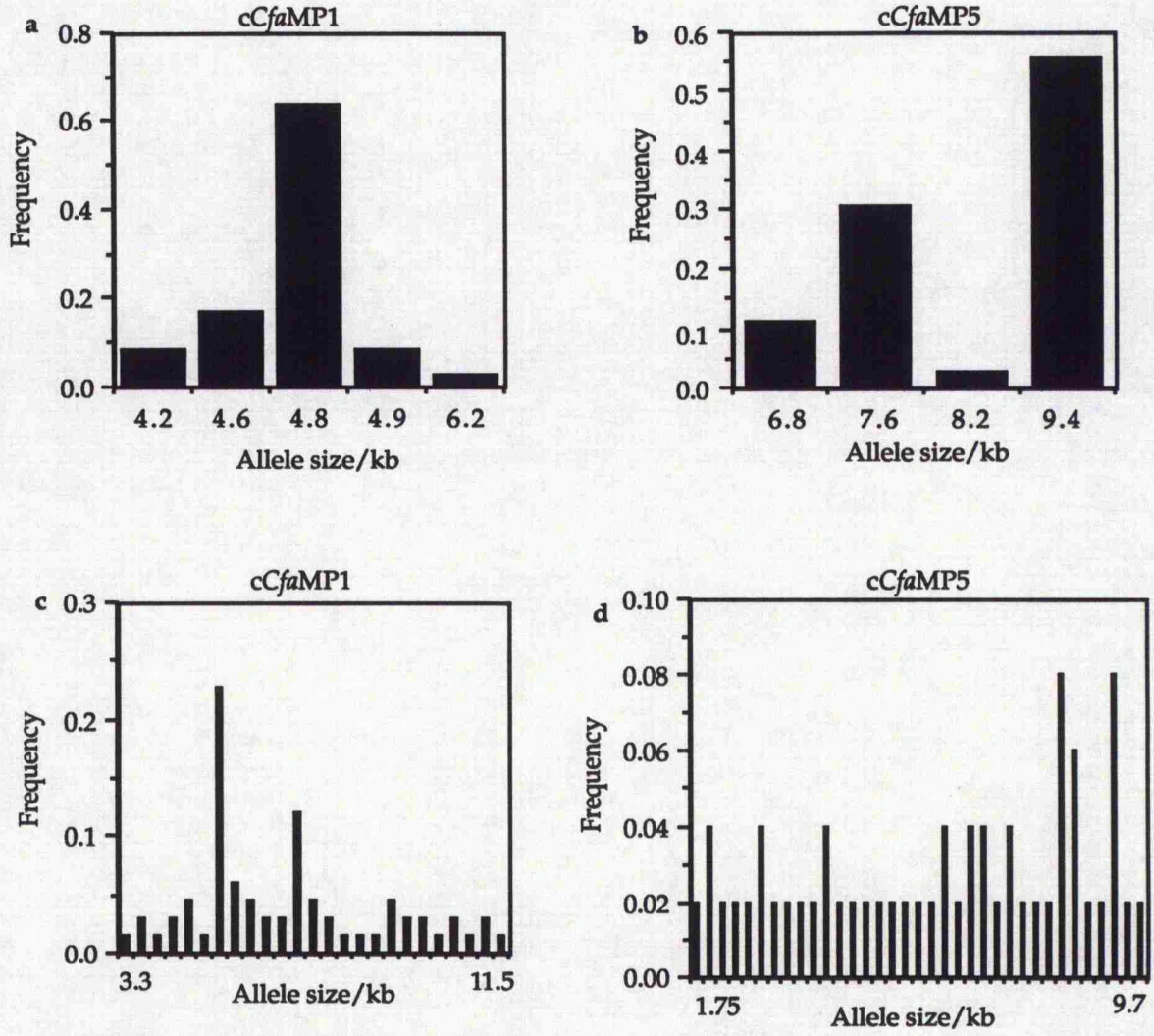


Figure 4.7 : Comparison of c*Cfa*MP1 and c*Cfa*MP5 allele frequency distributions in 18 Bedlington Terriers and a population sample consisiting of 25 individuals for c*Cfa*MP5 and 33 individuals for c*Cfa*MP1. a & b - allele frequency distributions in Bedlington Terriers; **c & d** - allele frequency distributions in the population sample.

shown). It is interesting however that this distance in the degree of relatedness, is not reflected by c*Cfa*MP5, which shows a greater level of heterozygosity than c*Cfa*MP1 both in the Bedlington Terrier population as well as the canine population at large (see Figure 4.5). This is a reflection of the vulnerability of c*Cfa*MP1 to minor genetic divergence effects within an otherwise inbred pedigree. It also indicates that different polymorphic single-locus loci are likely to vary in the degree of information that they reveal about various pedigrees, and that the less variable minisatellite loci might be more useful for studies of relatedness or divergence within pedigrees.

It can be seen from both Figures 4.5 and 4.6 that both c*Cfa*MP1 and c*Cfa*MP5 detect each other, albeit more weakly in the case of c*Cfa*MP5, in their single-locus profiles. This suggests that a considerable degree of similarity exists between the repeat arrays within each locus. This is reflected by the fact that both minisatellite loci cross-hybridize to the multilocus probes, 33.6 and 33.15 (see Chapter 3, Table 3.3) both of which belong to a subset of human minisatellite loci that show similarity in their core consensus sequences (Jeffreys *et al.*, 1985a).

Vergnaud *et al.*, (1991) have investigated the effect on the hybridization profile of small sequence changes in synthetic polymers. One of the changed polymers (16C18) when hybridized under the conditions used for the original polymer (16C2), did not detect the predominant locus identified by 16C2, but cross-hybridized with a new predominant locus that appeared as a weak band with 16C2. The characterization of c*Cfa*MP1 and c*Cfa*MP5 by obtaining sequence information about their respective repeat arrays would demonstrate the similarities present between them. This information has only been obtained for c*Cfa*MP5 (see Chapter 5).

Certain alleles are present in Bedlington Terriers but not in the general population, but these are generally those alleles that are at low frequencies within Bedlington Terriers. This is most likely to be due to sampling effects and the production of *de novo* mutations in each of the populations after their foundation. For c*Cfa*MP1, two alleles of 4.5kb and 4.8kb show the highest individual frequencies within the population sample; the 4.5kb allele is shared by 4 individuals from different pedigree breeds, 4 cross-bred individuals and 2 individuals from the same breed, whereas the 4.8kb allele is shared by 5 individuals from different pedigree breeds. For c*Cfa*MP5, two alleles of 9.4kb and 6.9kb show the highest allele

frequencies; the 9.4kb allele is shared by 3 individuals from different pedigree breeds whereas the 6.9kb allele is shared by 3 individuals, two of which come from the same breed and one is a crossbred individual. This indicates that alleles of polymorphic canine minisatellite loci are likely to be shared between unrelated individuals from the same breed as well as individuals from phenotypically diverse breeds such as Dachshunds, German Shepherds, West Highland Whites, Staffordshire Bull Terriers and Labradors. Thus, the likelihood of detecting a breed-specific minisatellite allele which has arisen due to substructuring effects will be small due to the relatively recent domestication of wolves (Wayne, 1993).

c*Cfa*MP1 was used to obtain a single-locus profile of this locus in the Schipperke breed. Figure 4.8 demonstrates the homozygosity present in this breed as a direct result of inbreeding (S1 -S6). Although alleles of varying sizes are detected in the other breeds shown, only a single form of this locus (8.8kb) is present within the genomes of the various Schipperkes analysed. All dogs within this breed can be traced back to two males and five female ancestoral dogs known to be subsequently line bred from as long ago as 1886 (personal communication - J. Sampson). The level of inbreeding in this particular breed is therefore very high, and c*Cfa*MP1 as a polymorphic minisatellite locus displaying a lower level of heterozygosity within the population than c*Cfa*MP5, yields a monomorphic single-locus pattern in Schipperkes indicating the vulnerability of less variable loci to inbreeding effects (Jeffreys *et al* 1991b). However, c*Cfa*MP5 has also been used as a single-locus probe to analyse Schipperkes (SK5) in separate analyses (data not shown) and it has been shown that a degree of heterozygosity (although unknown to what extent) can be detected within this breed. Thus, the above analyses indicate that although a probe which is polymorphic for numerous breeds, might detect an uninformative single-locus profile in a given breed, the use of a bank of highly polymorphic canine minisatellite probes should eventually yield an informative pattern for the purposes of specific analyses.

4.4 The Use of Canine Minisatellite Loci to Detect Related Loci Within the Genomes of Unrelated and Related Species

The clone c*Cfa*MP5 was used as a multi-locus probe (see Methods Section 2.13) against a Southern blot of *Mbo* I - digested genomic DNA from various species as indicated in Figure 4.9. This shows that at low

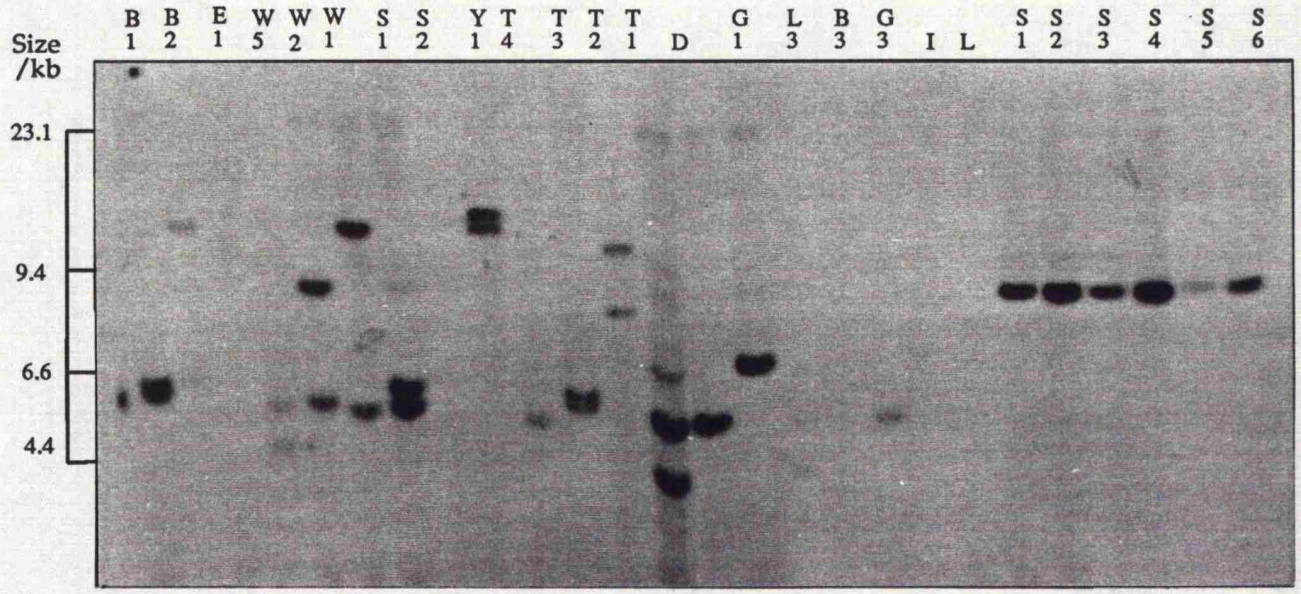


Figure 4.8 : Southern blot of *Mbo*I-digested canine DNA from a selection of individuals from various breeds probed with c*Cfa*MP1 at high stringency. Breeds analysed were as follows: B1 & B2 - Bernese Mountain, E1 - English Bull Terrier, W5, W2 & W1 - West Highland White, S1 & S2 - Staffordshire Bull Terrier, Y1 - Yorkshire Terrier, T1, T2, T3 & T4 - Terrier crosses (mongrels), D - Dachshund, G - Great Dane, L3 - Labrador, B3 - Border Collie, G3 - German Shepherd, I - Irish Setter, L - Lhasa Apso, S1 - S6 - Schipperke. λ *Hind* III DNA molecular weight markers are shown. cCfaMP1 does not detect equivalent loci in a number of cases i.e. English Bull Terrier, Staffordshire Bull Terrier 2, Border Collie 3, Irish Setter and Lhasa Apso. This is more likely to be due to the lack of sufficient DNA present in these samples rather than the absence of the locus within the genomes of these individuals. It is notable that whereas c*Cfa*MP1 is informative i.e. detects 5 different alleles in 3 West Highland White dogs, it is completely uninformative for Schipperke dogs. This is probably a reflection of the inbreeding coefficients of the two breeds.

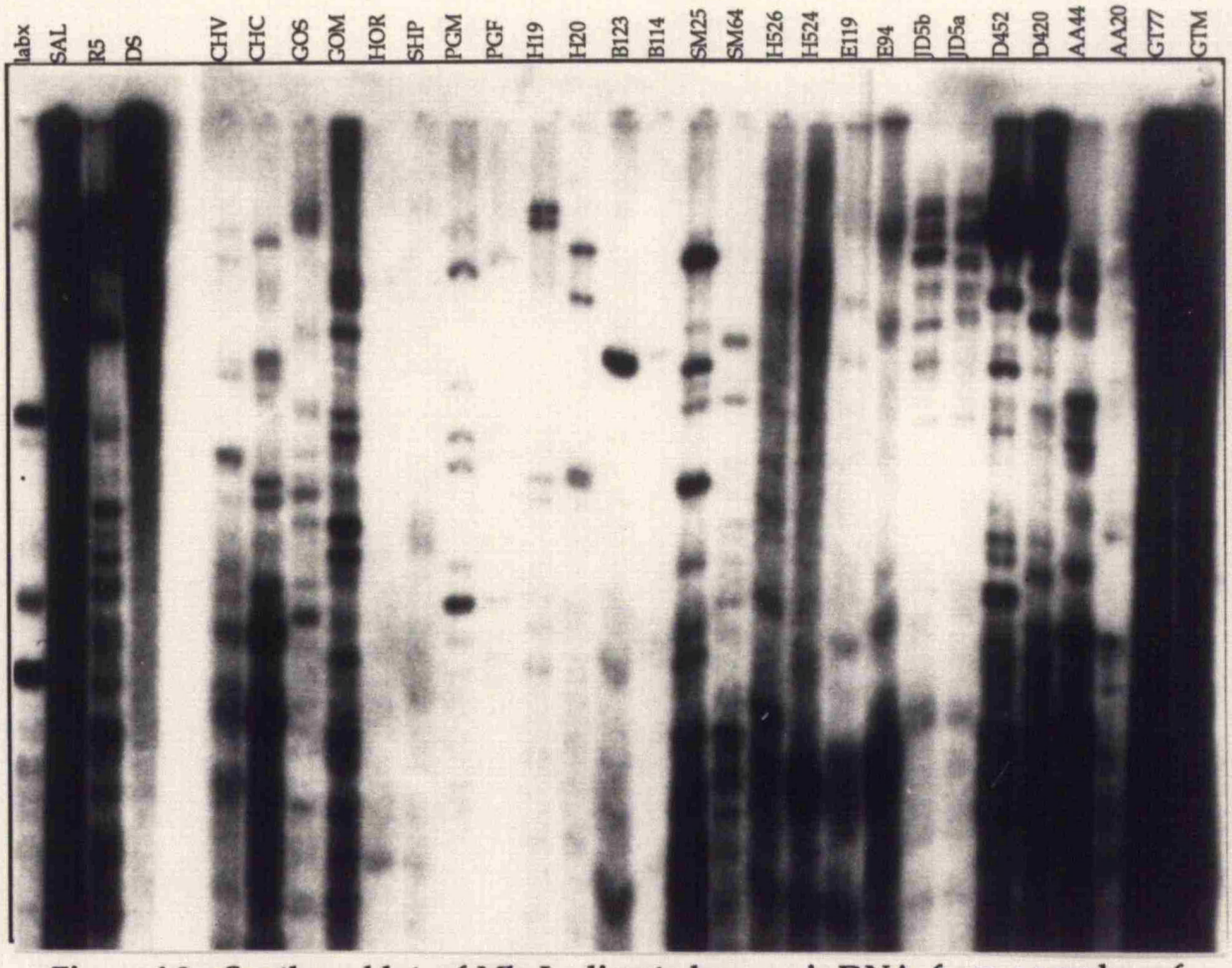


Figure 4.9 : Southern blot of *Mbo*I - digested genomic DNA from a number of species probed at low stringency with c*Cfa*MP5.

Labx - Labrador cross, SAL - salmon, R5 - Rutilus alburnoides (fish), DS - Drosophila, CHV & CHC - female chimpanzees, GOS - female gorilla, GOM - male gorilla, HOR - horse, SHP - sheep, PGM - male pig, PGF - female pig, H19 & H20 - hyenas, B123 & B114 - badgers, SM25 & SM64 - Sand martins, H526 & H524 - House sparrows, E119 & E94 - Eastern Bluebirds, JD5a & b - Jackdaws, D452 & D420 - Dunnocks, AA44 & AA20 - Alpine Accentors, GT77 & GTM - Great tits.

stringency this minisatellite sequence detects DNA fingerprint patterns in most of the species analysed. The exceptions to this are salmon, Drosophila, house sparrow and great tits whose genomes seem to contain an extremely abundant class of repeat sequences which are probably similar to those seen by Burke & Bruford (1987) in the Japanese quail.

It is somewhat surprising that c*Cfa*MP5 does not seem to detect any strongly-hybridizing minisatellite DNA fragments in either the horse or the sheep genome, considering that c*Cfa*MP5 is able to detect related sequences in phylogenetically unrelated species such as Drosophila. However, it is possible that this result might be due to some experimental defect such as degradation of the DNA prior to electrophoresis and perhaps future repetition of this analysis will serve to clarify this anomaly.

Interestingly, the two badgers show only one major strongly - hybridising minisatellite DNA band per individual within the resolvable region of the DNA fingerprint. Furthermore, the DNA fingerprint detected by c*Cfa*MP5 in the labrador cross breed is dominated by the presence of the two alleles detected by this probe at high stringency, among which there are a number of weakly-hybridising canine minisatellite DNA loci.

The detection of the DNA fingerprint patterns at low stringency with c*Cfa*MP5 in the various species analysed indicates that multiple related minisatellite sequences exist in the genomes of quite a large spectrum of species. This result is not altogether surprising since minisatellite single-locus probes isolated from humans (Wolff *et al.*, 1991), birds (Hanotte *et al.*, 1991) and fish (Bentzen & Wright, 1992), have been shown to cross-hybridize to numerous other species at high stringency. The major minisatellite DNA band detected in badgers might represent a homologous locus to c*Cfa*MP5, since badgers belong to the Mustelidae family which is thought to be linked phylogenetically to the Canidae family, which encompasses the domestic dog (Macdonald, 1992). Although the DNA sequence structure of c*Cfa*MP5 has been determined in the domestic dog, it would be necessary to determine the sequence of the equivalent locus in badgers, before any conclusions could be drawn. The detection of two strongly - hybridizing bands corresponding to the locus, detected by this probe at high stringency in the Labrador, seems to indicate that the unique DNA sequence flanking the repeat array in this genomic clone is involved in the degree of specificity shown by this probe even under non-stringent conditions.

c*Cfa*MP5 is not taxon-specific in its recognition of polymorphic loci i.e. it detects multiple variable loci in unrelated species. This result as well as the guanine-rich nature of c*Cfa*MP5, suggests that it may function like L17 (Gyllensten *et al.*, 1989) a bird minisatellite which as a result of a cryptic similarity to minisatellite repeats in unrelated loci functions as a multilocus probe. This cross-hybridization of c*Cfa*MP5 to unrelated species at low stringency indicates that it can be used to study genetic relationships among individuals in a wide range of species thus complementing human minisatellite sequences. In contrast, Bentzen & Wright, (1992) have suggested that the hybridization of a fish minisatellite, Ssal-5', to multiple loci only in salmonids may reflect a common relationship by descent for these loci, through transposition and/or duplication of the minisatellite and associated flanking sequence.

Hanotte *et al.*, (1992a) have shown that it is possible to detect specific, variable loci in species (Passeridae family) related to the house sparrow by a hypervariable house sparrow minisatellite single-locus probe. c*Cfa*MP5, which is the most variable canine minisatellite clone isolated to date (heterozygosity - 88%), was used as a single-locus probe at high stringency to analyse *Mbo*I - digested DNAs from related species such as wolves, coyotes and foxes, as well other less related mammals such as hyenas, pigs, badgers and humans. Figure 4.10 shows the results of this analysis : Several DNAs from various breeds of dog were included as positive controls. It can be seen that although this polymorphic locus detects homologous loci in the various domestic dogs analysed as would be expected, it does not detect an equivalent in any of the related canids, nor in any of the unrelated species analysed. This was somewhat surprising since wolves, coyotes and foxes belong to the same phylogenetic family, (the Canidae) as the domestic dog and grey wolves are generally thought to be the ancestors of the present day domestic dog (Wayne, 1993).

c*Cfa*MP5 was therefore used again as a single-locus probe at high stringency to screen 35 wolves and coyotes from various regions of the North American continent and Alaska (DNA samples were donated by M. Roy, UCLA) together with some other vertebrates (fish and newt). In all but three related canids (individuals 195, W29 & W73), no equivalent minisatellite loci were detected (data not shown). Furthermore the intensity of the signal obtained in the Southern blot analysis for these few wolves and coyotes is considerably lower than that seen with dogs. On the other hand, a high molecular weight smear was detected in both the fish

and newt genomes. Hanotte *et al.*, (1992a) have found that at high stringency a house sparrow minisatellite sequence detected a complex high molecular weight pattern in the acorn woodpecker, and have suggested that this represents an abundant repeated sequence in this species. Therefore, the smears detected by c*Cfa*MP5 in fish (salmon) and newt (*Triturus sp*) at high stringency might also represent abundant related repeated sequences in these two species.

The inability of c*Cfa*MP5 to detect similar loci in other wolves and coyotes can perhaps be explained by considering the observations of Gray & Jeffreys, (1991). They amplified the primate equivalents of two human hypervariable minisatellite loci, MS1 and MS32, and found that the ability of hypervariable minisatellite sequences to detect highly informative loci by cross-species hybridization is unpredictable due to the evolutionary transience of these sequences. Furthermore, they were not able to detect MS32 equivalents in prosimian (lemur) DNA by cross-hybridization and attempts to amplify this locus in New World monkeys or lemurs were completely unsuccessful. It is possible therefore that the informativeness of highly polymorphic minisatellite sequences isolated from the canine genome, when used as single-locus probes could be limited in related species for the following reason: Polymorphic minisatellite loci show variation within a given population due to the ability of these sequences to undergo mutations, thereby maintaining their transient nature over an evolutionary period of time. This transiency could be reflected in the inability of c*Cfa*MP5 to detect a canid equivalent in many of the related canids used in this study at high stringency. Gray & Jeffreys (1991) suggest in conclusion from their results, that although multilocus probes can detect multiple variable loci at low stringency in a wide range of species, the patterns generated by such probes do not necessarily represent homologous groups of hypervariable loci in different species, even if these species are closely related. Therefore, given that c*Cfa*MP5 detects a single major strongly-hybridizing band in the DNA fingerprints of the two badgers studied at low stringency, it is possible that the lack of sufficient homology between this locus and c*Cfa*MP5, results in a lack of signal at high stringency (see Figure 4.10). The fact that a canid equivalent of c*Cfa*MP5 is however detected in only a few canid individuals at high stringency confirms that this locus is present in wolves and coyotes but in several cases exists as either a much smaller locus (a null allele?) (Armour *et al.*, 1992b) or is completely absent. The resolution of the above question

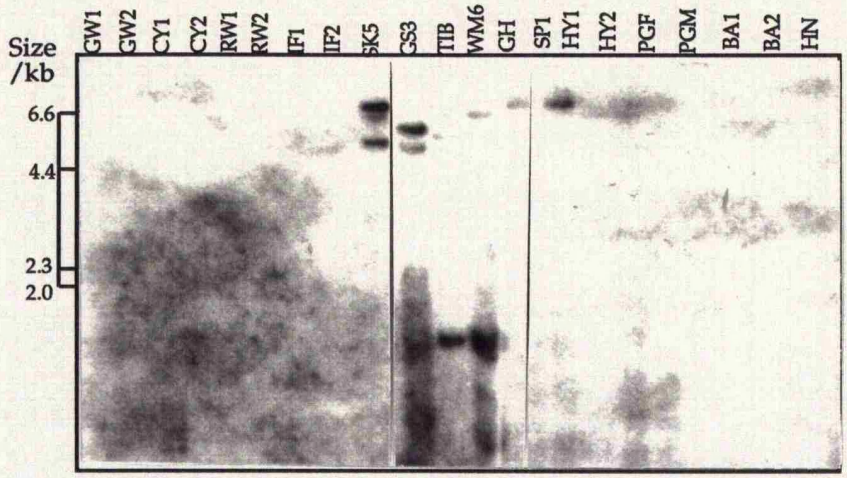


Figure 4.10 : Autoradiogram of *Mbo*I - digested genomic DNA from various species probed at high stringency with c*Cfa*MP5.
GW1 & GW2 - Grey wolves, CY1 & CY2 - Coyotes, RW1 & RW2 - Red Wolves, IF1 & IF2 - Irish Foxes, SK5 - Schipperke, GS3 - German Shepherd, TIB - Tibetan Terrier, WM6 - Irish Setter, GH - Grey Hound, SP1 - Spaniel, HY1 & HY2 - hyenas, PGF - female pig, PGM - male pig, BA1 & BA2 - badgers, HN - human.

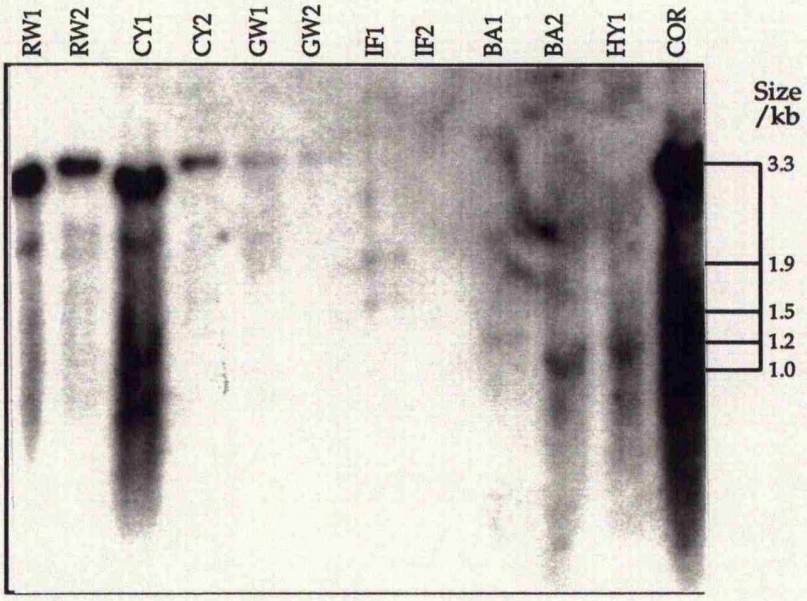


Figure 4.11 : Southern blot of *Mbo*I - digested genomic DNA from various canids probed with c*Cfa*MM3 at high stringency. The minisatellite loci detected by this probe are of identical size in wolves, coyotes and the domestic dog, and reflect the degree of genetic homogeneity between these canids. See Figure 4.10 legend for details of RW1 to HY1. COR - Corgie.

requires amplification of this locus using primers flanking the repeat array within wolves and coyotes which has yet to be done. The weak signal obtained for three of the individuals studied could be due to the existence of a repeat array which although a similar size to that found in dogs could have a limited number of repeat units that are homologous to those present in the dog.

The use of a monomorphic canine minisatellite clone c*Cfa*MM3 to screen DNA of individuals from related species showed that although this probe detected a monomorphic equivalent in wolves and coyotes, it was no longer monomorphic with respect to the other less related species: As can be seen in Figure 4.11, the locus c*Cfa*MM3 detects a canid equivalent in most of the related canids. There seems to be a weak interaction corresponding to approximately 1kb in size in the hyena but this requires repetition of this analysis for sufficient confidence in its validity. c*Cfa*MM3 does not seem to detect a canid equivalent in one of the Irish foxes (IF2) analysed. This is probably due to a lack of sufficient DNA in this track. Interestingly, c*Cfa*MM3 detects an informative locus in the other Irish fox i.e. it is heterozygous at this locus with two alleles of 1.9kb and 1.5kb. The locus detected by this probe in the wolves and coyotes is identical in size to that detected in the domestic dog. On the other hand, the loci detected by this probe in badgers consists of two alleles of 1.2kb and 1.0kb, for which one badger is homozygous for the 1.2kb allele (BA1) and the other homozygous for the 1.0kb allele (BA2). Why does this monomorphic canine minisatellite detect informative loci in badgers and foxes? It is possible that even at high stringency the loci detected in these species are similar but not homologous to that in dogs, wolves and coyotes, so that degrees of polymorphism may be detected in less related species by c*Cfa*MM3. These findings contrast with those of Gray and Jeffreys (1991) who found that PCR amplification of MS1, a human minisatellite, shows extreme repeat-copy-number variability in man compared with low copy number and minimal variability in great apes. Testing the above hypothesis that c*Cfa*MM3 detects similar, but not homologous loci in badgers and foxes, might involve similar amplification of the locus in these species and subsequent sequencing of the region.

The above results support the phylogenetic grouping of domestic dogs with wolves and coyotes in the family Canidae. It is probable therefore that at least with respect to monomorphic minisatellite sequences the three sub-families have a great deal in common. Such

similarity is likely to be due to the relative stability that monomorphic minisatellite sequences show at least in size, if not in homology, between highly related canids. It should be noted that this similarity ends when less related species (e.g. the fox and the badger) are probed with such sequences. Therefore, although monomorphic minisatellite loci are generally uninformative for most purposes within *Canis familiaris* , such loci are likely to prove useful for inter-species population studies, e.g. studies of the degree of genetic homogeneity between various canids. It follows then that the use of polymorphic minisatellite loci with lower levels of variation might be able to differentiate between various populations of wolves and also be informative with respect to the lineages of specific kinds of wolves. Such information would be highly useful for the purposes of conservation issues such as the preservation of hybrid species such as the red wolf (see Wayne, 1993).

4.5 Paternity Analysis in *Canis familiaris*

DNA profiling with the variable minisatellite clone c*Cfa*MP5 was used to resolve a paternity dispute in an Irish Greyhound family. (Identification of parentage is important in domestic species where large sums of money are paid for inseminations or where animals are being progeny tested). The case involved a litter of 10 pups (A - J), a dam (M) and two putative sires (F1 & F2) of which F2 was known to be the father of F1. c*Cfa*MP5 was used to probe a Southern blot of *Hinf* I - digested genomic DNAs from the family at high stringency, the result of which is shown in Figure 4.12. This analysis shows that there are four alleles present at this locus of 10.8kb, 8.0kb, 6.5kb and 1.3kb. The lack of the 6.5kb allele in one of the putative sires (F1) and the presence of this allele in five of the offspring supports the exclusion of this individual as the sire of individuals B,D,F,G &H. F2 shares an allele with F1, which is also present in five of the offspring (A, C, E, I & J) and so from this particular analysis, it is not possible to exclude F2 from having sired pups A, C, E, I & J. However, subsequent microsatellite analysis of this litter has shown that F2 is the true sire of the litter (C.Mellersh - personal communication).

There is a notable anomaly in this litter in that although pup B has a 6.5kb allele consistent with it being inherited as a paternal allele from F2, its other allele (10.8kb) (presumably its maternal allele), is not present in the dam (M) from whom all the other maternal alleles clearly originate.

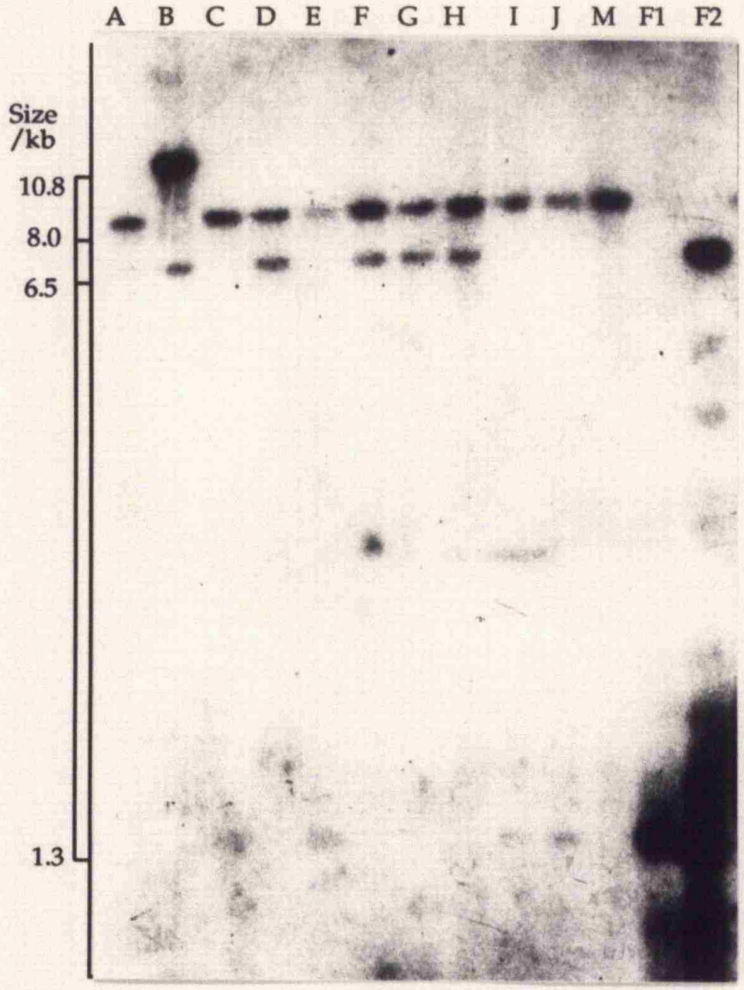


Figure 4.12 : Paternity analysis in a Greyhound pedigree. Canine genomic DNA was digested with *Hinf*I to completion, Southern blotted and probed with c*Cfa*MP5 at high stringency. Individuals A - J are the offspring of dam M and sire F2. Individual F1, who is also the son of individual F2, has been excluded from having sired the pups due to the lack of a band at 6.5kb in his single-locus profile.This band which is present in five of the offspring has been inherited from individual F2. Note that individual B has a non-maternal allele of 10.8kb.

There are two possible explanations for this:

(i) the 10.8kb allele could have arisen by somatic mutation of the maternal allele to result in an allelic size change at this locus

(ii) the 10.8kb allele might originate from another dam (unusual but entirely possible) indicating the presence of an impostor among the offspring of dam (M).

Amplification of a microsatellite locus in this family has subsequently also shown that the dam (M) is unlikely to be the actual mother of pup B, and that there has been either infiltration of a dam from another kennel or that pup B has been assigned to the wrong mother (unpubl. data - C. Mellersh, 1993). It has been proposed that VNTR-containing sequences are frequent sites for *de novo* mutations. This is based on the assumption that the large number of individual variations in the number of tandem repeats could arise by rearrangements in the germ cells. The possibility of occurrence of recombinant alleles is important for the interpretation of paternity test results since they may lead to the exclusion of the alleged father or mother. Therefore it is essential that cases of exclusion be confirmed by more than one locus.

Armour *et al.*, (1992) have shown that a proportion of alleles at two human minisatellite loci is undetected by Southern blot analysis. In each case the missing allele can be identified after PCR amplification and correspond to repeat arrays which are too short to be detected by hybridization. Although a more detailed analysis might show evidence for the existence of null alleles for c*Cfa*MP5 and c*Cfa*MP4, none has been found thus far. However, for c*Cfa*MP1, it is possible that null alleles exist, as no hybridization signal was obtained for a Schipperke dog analysed (data not shown). However, this would have to be confirmed by segregation studies.

CONCLUSIONS

The applications of minisatellite sequences as multilocus and single-locus probes are many e.g. paternity analysis (Jeffreys *et al.*, 1985b, Georges *et al.*, 1989, Hermans *et al.*, 1991), population studies (Gilbert *et al.*, 1990), linkage analysis (White *et al.*, 1985a & b), assessment of inbreeding (Kuhnlein *et al.*, 1990) and pedigree analysis (Jeffreys *et al.*, 1986). The isolation of the limited number of canine polymorphic minisatellite loci

in this study has made the first step forward into the use of such sequences in analyses of the canine genome. It is clear that a much greater number of polymorphic canine minisatellite loci will have to be isolated and characterized, before a suitable canine linkage map of DNA markers spanning most of the canine genome can be generated. However, the isolation and characterization of the 7 polymorphic canine minisatellite loci in this study has enabled the use of a few of these loci as single-locus probes, to :

(i) show Mendelian inheritance of these loci in canine pedigrees,
(ii) compare the levels of genetic variation in a canine subpopulation sample to that of an agglomerated population sample using two polymorphic canine minisatellite loci,
(iii) study the similarity of unrelated and highly related species to dogs,
(iv) resolve paternity disputes.

The attempt to show Mendelian inheritance of these polymorphic minisatellite loci in the Irish Setter family, emphasises that the degree of polymorphism shown by these loci within the population as a whole, has to be re-defined with respect to each breed analysed. This was because a number of the variable probes when used as single-locus probes against the Irish Setter family were uninformative. Thus, it is likely that it will be necessary to characterize any polymorphic canine minisatellite loci isolated with respect to a large panel of different breeds, so that a particular group of such loci that are likely to be informative within a given breed, is first used to test the breed for the purposes of specific analyses. It is clear that the more variable a locus is with respect to the canine population at large, the greater the chances of it being informative for a given breed. However, such loci might not be able to detect levels of relatedness within a breed, the effects of which are more likely to be displayed readily by less variable minisatellite loci.

The results of the *Alu*I DNA fingerprint analysis support the view that inter - breed variation in canine DNA fingerprints is not significantly greater than intra - breed variation. This does not however, preclude the fact that a more detailed analysis may eventually show substructuring of the canine population.

Comparison of the genetic variation of two polymorphic minisatellite loci within an agglomerated population sample to that

within a breed shows that there is a general reduction in the genetic variation seen from population to breed. For example, the heterozygosity levels seen for c*Cfa*MP5 in the population is 88%, compared to 44% in Bedlington Terriers. Furthermore the number of alleles is reduced by a quarter from population to breed. This is not unexpected since it is well known that the present canine population consists of numerous subpopulations (breeds), each having arisen due to a mixture of inbreeding, assortative mating and selection. It is not inconceivable however that the loci used in this study are more vulnerable to processes such as those stated above, due to their lower rates of mutation being less effective against the forces of genetic drift in breeds. Therefore, it is possible that the isolation and characterization of highly polymorphic minisatellite loci in dogs (heterozygosities > 99%), like those detected in mice (Kelly *et al.*, 1991) and humans (Gray & Jeffreys, 1991) will eventually enable similar levels of inter- and intra-breed variation to be obtained.

The studies using canine minisatellite loci both as single-locus and multilocus probes to screen individuals from related and unrelated species provide information on several fronts:

(i) Loci similar to canine minisatellite sequences are present in both related and unrelated species, although these might not be homologous to their canine counterparts in all cases.

(ii) The c*Cfa*MP5 family of repeats shows varying degrees of conservation between unrelated species.

(iii) The canid equivalents of polymorphic minisatellite sequences from the domestic dog are not necessarily detectable by single-locus probing in all members of highly related canids, such as wolves and coyotes. This could be due to the reasons enumerated above i.e. a smaller locus representing 'null' alleles or although less likely, the total lack of this particular locus in some wolves and coyotes.

(iv) The detection of the presumptive homologues of c*Cfa*MM1 in wolves and coyotes supports the phylogenetic grouping of wolves, coyotes and domestic dogs into one family, the Canidae, and serves to further confirm the view that the present domestic dog is probably the descendant of the grey wolf (see Wayne, 1993).

In summary, the applications of minisatellite biology to the canine genome are in league with those already in existence to the human genome. In the long term, the addition of such sequences to current

canine linkage analysis will serve to fill any gaps left by microsatellite DNA markers, the easier isolation and characterization of which, has meant that these are at present the preferred route to the generation of a canine linkage map of DNA markers.

CHAPTER 5

DNA SEQUENCE ANALYSIS OF THE CANINE MINISATELLITE SEQUENCE IN *cCfa*MP5

SUMMARY

A canine minisatellite sequence (c*Cfa*MP5) has been isolated from a canine genomic library, subcloned into the M13 bacteriophage mp18 and mp19 vectors, and sequenced to obtain flanking and repeat sequence. This minisatellite is similar to many other already isolated minisatellites in that it is GC-rich, shows strand asymmetry and shows similarity to the *E.coli* recombination signal, GCTGGTGG. Sequence comparisons with those in the Genbank and EMBL databases using GCG software did not detect any significant matches. The repeat array contains repeat units of 11bp of which there are at least four variants. These variants show 73-90% homogeneity, any heterogeneity between repeat variants resulting from point mutation events, only in four bases near the 3' end of the repeat unit. Although complete flanking sequence has been obtained, only partial repeat sequence has been obtained, 441bp at one end and 144bp at the other. Sequencing of the repeat region proved difficult due to the high purine/pyrimidine strand bias. The repeat array is cleaved by the restriction enzyme, *Alu* I the recognition site of which is present in only one of the repeat variants. Only one side of the margin of the repeat array shows divergence from the middle region in that a truncated version of one of the variant repeats exists at the 5'end. A mechanism whereby this minisatellite might have evolved has been postulated on the basis of the distribution and kinds of variant repeats. This envisages initial duplication events of the motif (GGA)n, followed by point mutations within a relatively stable 11bp sequence. Variants of this could have subsequently undergone duplication as well as possibly intra-allelic unequal exchange events to create the intermingled distribution of variant repeats seen within the array.

INTRODUCTION

The rapid detection, isolation and characterization of numerous minisatellites in a wide range of species has succeeded the initial discovery by Wyman & White, (1980) of a polymorphic human DNA locus. This has led to their application in various scientific and practical analyses: paternity analysis (Jeffreys *et al.*, 1985b, Georges *et al.*, 1988, Hermans *et al.*, 1991), population studies (Gilbert *et al.*, 1990), linkage analysis (White *et al.*, 1985a), assessment of inbreeding (Kuhnlein *et al.*, 1990), pedigree analysis (Jeffreys *et al.*, 1986) and forensic analysis (Gill *et al.*, 1985).

The examination of a subset of human minisatellite sequences by Jeffreys *et al.*, (1985a) resulted in the detection of a core consensus sequence which was shown to be shared by these minisatellites. The similarity that this core sequence bears to the *E.Coli* generalised recombination signal led to speculation that it might also play a similar role to promote unequal exchange at minisatellite loci. This speculation was supported by several other lines of evidence, a few of which are as follows: Minisatellites show preferential localization to the proterminal regions of human autosomes, as do chiasmata (Royle *et al.*, 1988); *in situ* hybridization of the 33.15 sequence to meiotic chromosomes showed that it is localized at chiasmata (Chandley & Mitchell, 1988); a core-related sequence is present at or near the Eβ MHC meiotic recombination hotspot in the mouse (Steinmetz *et al.*, 1986); paternal and maternal mutations arise with similar frequencies in human pedigrees, suggesting that mutation events are largely restricted to one step of gametogenesis, possibly meiosis (Jeffreys *et al.*, 1988). However, recent studies into the internal structures of minisatellites have indicated that although inter-allelic unequal exchanges do occur at minisatellites, the majority of mutational events can only be attributed to intra-allelic unequal exchange events or slippage events (Jeffreys *et al.*, 1990,1991a).

The isolation and characterization of minisatellite sequences is crucial to the understanding of the mechanisms underlying length variation at minisatellites and their origin. The characterization of the canine minisatellite sequence in c*Cfa*MP5 serves to add to the current' lake of knowledge' concerning minisatellites. This chapter therefore describes the isolation and sequence analysis of this minisatellite, correlating its similarities and disimilarities to the characteristics of known minisatellite

sequences in a wide range of species, as well as speculating on the mechanisms by which it originated.

RESULTS

5.1 Subcloning of the 2.5kb Canine DNA Fragment in c*Cfa*MP5 into pBluescript DNA

Recombinant Charomid DNA was isolated from c*Cfa*MP5 as described in Methods Section 2.23.1 and digested with *Sau* 3AI to liberate any canine DNA inserts. Agarose gel electrophoresis of the digested products showed that c*Cfa*MP5 contained two insert DNA fragments of 2.5kb and 2.8kb. Previous Southern blot analysis showed that the 2.5kb insert contained the repeat array (see Chapter 3, Results Section 3.5.3). Therefore, the 2.5kb fragment was isolated by preparative gel electrophoresis (Methods, Section 2.23.2) and the DNA concentration determined. A ligation was set up containing a 4:1 molar ratio of pBluescript to 2.5kb canine insert DNA (see Methods Section 2.24).

The ligation mixture was then used to transform XL1-Blue bacteria (Methods Section 2.24) and any resulting colonies were picked and analysed to check for inserts of the correct size: Initially 18 clones were digested with *Pst* I which has one restriction site within the multiple cloning site of pBluescript DNA. Single DNA bands of approximately 2.96kb corresponding to the size of pBluescript were seen after gel electrophoresis of the digested products for 15 of the clones (Figure 5.1a). However, 3 clones were seen to contain inserts, two of which lacked a *Pst* I site within the insert. Clone 4 was seen to contain two DNA bands, one corresponding to the size of unrecombinant pBluescript DNA and the other of ~ 2.3kb. Since the canine DNA insert from the Charomid was known to be ~ 2.5kb, this seemed to indicate that a mutation had probably occurred in clone 4 during propagation in XL1Blue bacteria, resulting in the introduction of a *Pst* I site other than the one in the multiple cloning site and a reduction in the size of the 2.5kb insert. Subsequent *Sau* 3AI - digestion of the clones 3, 4 and 11 showed that all contained inserts of approximately the same size and also corresponded to the size of the 2.5kb canine DNA insert in c*Cfa*MP5 (Figure 5.1b). The presence of a repeat array was confirmed by probing the gel in Figure 5.1b with the multilocus minisatellite probe, 33.6 (Figure 5.1c).

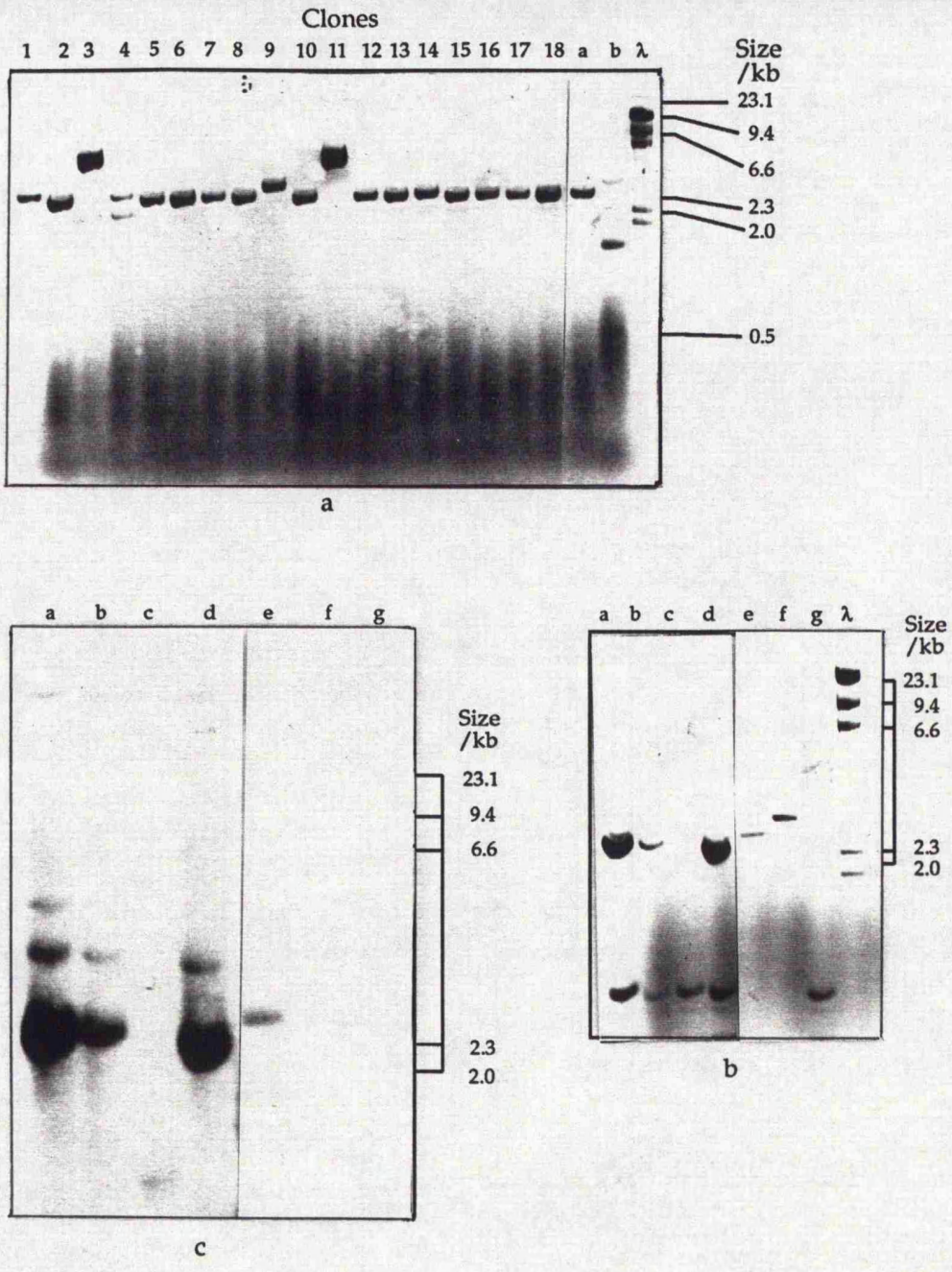


Figure 5.1 : Analysis of XL1-Blue Transformants for the presence of the 2.5kb canine DNA insert from c*Cfa*MP5. Figure 5.1a shows the digestion of KS+ pBluescript DNA from recombinant transformants with *Pst*I. From this analysis, three clones were seen to potentially contain inserts of the correct size (clones 3,4 & 11), and one seemed to contain an insert of a somewhat diminished size (clone 9). a - pBluescript DNA cleaved with *Pst*I, b - uncut pBluescript DNA, λ - λ*Hind*III DNA. These four clones were further analysed by *Sau*3AI-digestion of the vector DNA molecules to liberate each insert if present. This analysis is shown in Figure 5.1b. As can be seen, only clones 3, 4 & 11 contained correctly sized 2.5kb insert fragments. Lanes : a - clone 3, b - clone 4, c - clone 9, d - clone 11, e - 2.5kb *Sau*3AI DNA insert 19F32, f - 2.8kb *Sau* 3AI DNA insert 19F31, g - pBluescript DNA cleaved with *Sau*3AI, λ - λ*Hind*III DNA. Figure 5.1c shows the results of probing of the gel in Figure 5.1b with 33.6. As can be seen, clones 3, 4 & 11 contain inserts that hybridize to 33.6 indicating that these clones probably contain repeat sequences that are related to the consensus sequence in 33.6. The 2.5kb *Sau* 3AI DNA insert from Charomid 9-36 hybridizes positively but the 2.8kb insert does not, indicating that the correct insert has been sub-cloned.

5.2 Restriction Mapping of the 2.5kb canine DNA fragment in pBluescript

Since clone 11 lacked the *Pst* I sites, it was used for subsequent restriction mapping experiments. Clone 11 was therefore digested with a bank of restriction enzymes which cleaved only once within the pBluescript backbone to determine the smallest fragment containing the whole repeat array, which could be isolated and re-subcloned for simplified dideoxynucleotide sequencing. Of the number of enzymes used, only *Sac* I could be seen to reduce the cloned fragment in size. Figure 5.2 shows that digestion of clones 11 with *Sac* I results in two fragments of approximately 3.6kb and 2.1kb. This gel was therefore Southern blotted and the resulting filter probed with 33.6 to determine whether the whole repeat array was encompassed within the 2.1kb *Sac* I fragment in the two clones (Figure 5.3). This analysis showed that the 2.1kb *Bam* HI/*Sac* I fragment contained the complete repeat array.

5.3 Subcloning of the 2.1kb canine DNA fragment from clone 11 into the M13 mp18 and mp19 Bacteriophage Vectors for Dideoxynucleotide Sequencing

The 2.5kb *Sau* 3AI fragment from clone 11 was digested with *Sac* I and two fragments of ~2.1kb and ~0.4kb isolated by preparative gel electrophoresis for subsequent ligation into M13 mp18 and mp19 bacteriophage DNAs double-digested with *Bam* HI and *Sac* I (Figure 5.4). After transformation of the above ligation into JM101, the number of transformants were very low (1 or 2 white plaques only). However, single-stranded M13 bacteriophage DNA was isolated from both mp18 and mp19 bacteriophage transformants (Methods Section 2.7.3). Sequencing of a M13 mp18 bacteriophage clone showed the correct cloning site to be present (*Bam* HI), and the repeat array was apparent after 232bp of unique flanking DNA sequence (Figure 5.5). However, single-stranded dideoxynucleotide sequencing of a M13 mp19 bacteriophage clone showed that the cloning site was *Sac* I. This was probably due to a cloning artefact i.e. the digestion of the 2.5kb *Sau* 3AI/*Sac* I DNA fragment with *Sac* I had resulted in a population of fragments of 2.0-2.1kb which were indistinguishable under the electrophoresis conditions used. Therefore, the 2.0kb DNA fragment which was present in the mp19 vector was not present in the equivalent

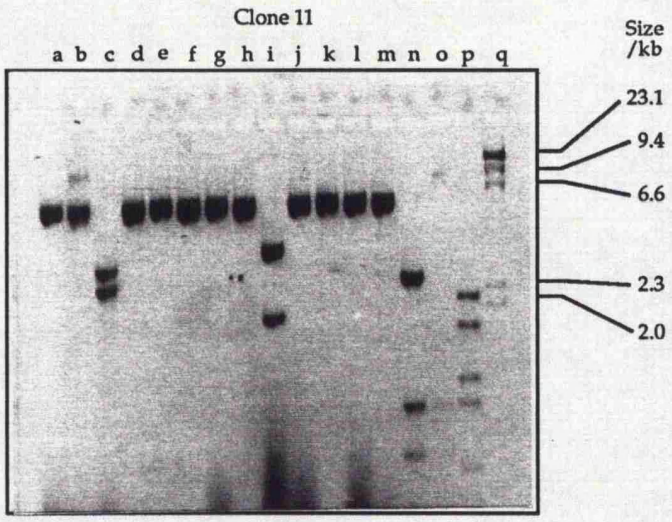


Figure 5.2 : Restriction analysis of pBluescript DNA from the recombinant clone 11 containing the 2.5kb minisatellite insert from c*Cfa*MP5.
Restriction enzymes used were as follows:
a - *Acc*I, b - *Apa*I, c - *Bam*HI, d - *Cla*I, e - *Eco*RI, f - *Hind*III, g - *Kpn*I,
h - *Pst*I, i - *Sac*I, j - *Sca*I, k - *Sma*I,
l - *Xba*I, m - *Xho*I, n - *Sau*3AI,
o - unrecombinant pBluescript DNA cleaved with *Sau*3AI, p - 2kb ladder,
q - λ *Hind*III DNA.

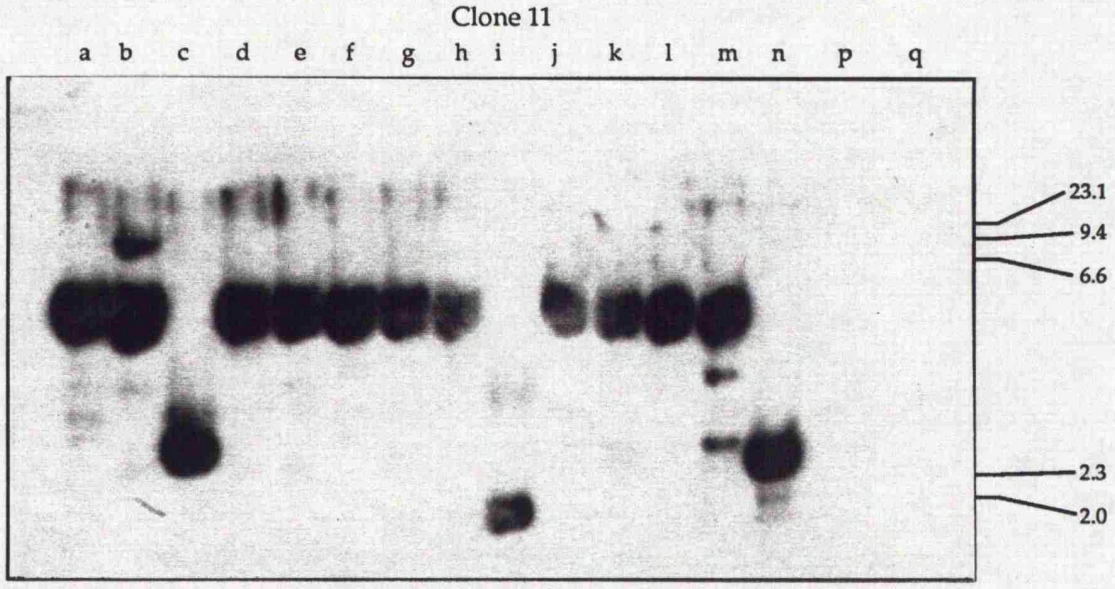


Figure 5.3 : Southern blot of gel in Figure 5.2 probed with 33.6 to detect the position of the minisatellite sequence within the clone 11. Restriction enzymes used were as follows: a - *Acc*I, b - *Bam*HI, d - *Cla*I, e - *Eco*RI, f - *Hind*III, g - *Kpn*I, h - *Pst*I, i - *Sac*I, j - *Sca*I, k - *Sma*I, l - *Xba*I, m - *Xho*I, n - *Sau*3AI, o - unrecombinant pBluescript DNA cleaved with *Sau*3AI, p - 2kb ladder, q - λ *Hind*III DNA.

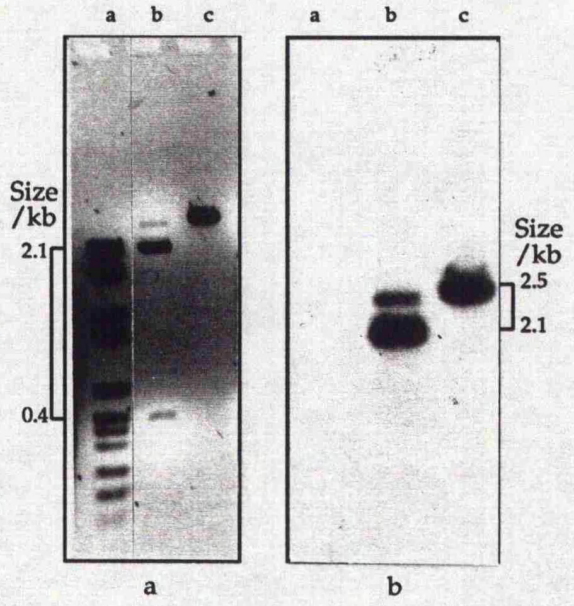


Figure 5.4 : Analysis of the minisatellite sequence in c*Cfa*MP5 by digestion with *Sac*I and probing with 33.6. The 2.5kb canine DNA insert from Charomid 9-36 was isolated by preparative gel electrophoresis, digested with *Sac*I and probed with 33.6 to determine the position of the repeat array. Figure 5.4a shows that digestion with *Sac*I reduces the 2.5kb insert to fragments of 2.1kb and 0.4kb. a - 2kb ladder, b - 2.5kb *Sau*3AI insert from c*Cfa*MP5 digested with *Sac*I. c - undigested 2.5kb *Sau*3AI insert from c*Cfa*MP5. Probing with 33.6 at high stringency in the presence of canine competitor DNA shows that the 2.1kb fragment contains the minisatellite sequence (Figure 5.4b). This fragment was therefore used for subsequent cloning into the M13 mp18 and mp19 bacteriophage vectors.

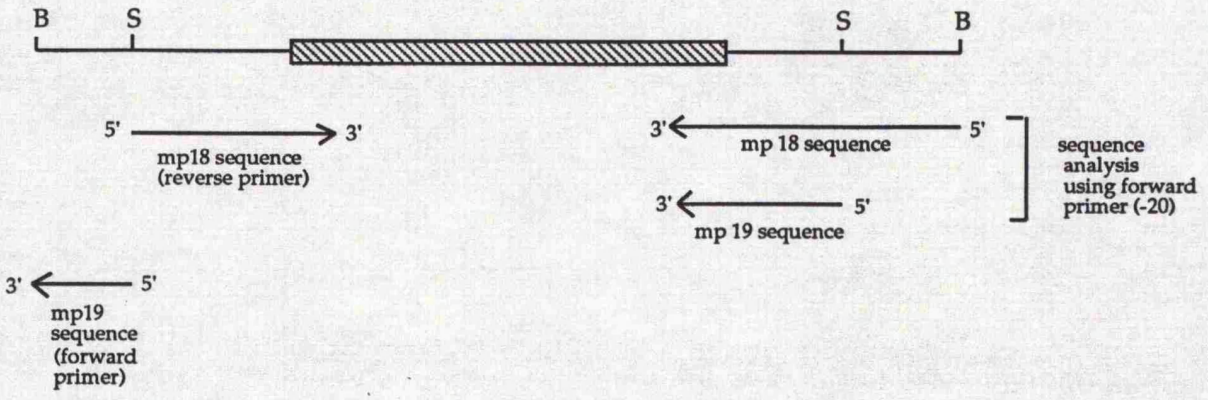


Figure 5.5 : Sequence analysis of the repeat region in c*Cfa*MP5.

mp18 bacteriophage clone. The mp19 bacteriophage clone isolated had clearly resulted due to probable incomplete digestion of the M13 vector with *Bam* HI and *Sac* I enabling a *Sac* I/ *Sac* I fragment to be cloned. Initially, the two mp18 and mp19 bacteriophage clones were sequenced using the 15 base M13 forward primer to obtain sequence information at the 5' end of the minisatellite array. To obtain sequence information from the other end, double-stranded M13 mp18 bacteriophage DNA was isolated (Methods Section 2.8) and sequenced using the 17 base reverse primer of M13. In order to get more flanking DNA sequence at the 3'end of the minisatellite, the *Sac* I/*Sau* 3AI fragment of ~0.4kb was also subcloned into the M13 mp18 and mp19 bacteriophage vectors and sequenced completely from either end.

The results of the sequence information is shown in Figure 5.6. Figure 5.6A presents the sequences flanking and partially within the minisatellite array in clone c*Cfa*MP5. For simplicity, the flanking sequence is given in its entirety, but the minisatellite sequence is represented by the appropriate order of the repeat motifs using the nomenclature presented in Figure 5.6B. A few higher order repeats can be seen and the order of the various repeat variants is not preserved i.e. there is considerable intermingling of the different repeat units. The canine minisatellite contained four variant repeat units, each of 11 bp (Figure 5.6B). These were conserved in all but three positions. Differences between repeat units were due to single-base substitutions at these positions. A base substitution (G→ C) in variant repeat a results in an *Alu* I site in variant repeat 3. The consensus sequence of the various repeats seen in c*Cfa*MP5 is therefore GGAGGARRGST where R is a purine and S is a either a cytosine or a guanine. The start of the repeat array contains a truncated version of the 11bp repeat which aligns with 8bp of the variant repeat 1 at the 3' end. Figure 5.6C shows the alignment of the various variant repeats with the consensus sequences of 33.6 and 33.15 and the chi sequence. There are multiple regions within the array which show similarity to the *E.coli* chi sequence, (GCTGGTGG) a generalised recombination signal. It has not yet proved possible to determine the sequence of the entire minisatellite repeat array. Sequencing of the repeat array was difficult due to premature loss of signal intensity. This is probably due to the mainly purine or pyrimidine strand bias of the repeat array causing localised depletion of nucleotides. Similar difficulties were experienced by Brereton *et al.*, (1993) in obtaining the complete sequence of a polypurine repeat

Figure 5.6 : Summary of DNA Sequence Analysis of the Canine Minisatellite Locus in c*Cfa*MP5. A - DNA sequence of the 5' and 3' flanking regions together with the sequence of variant repeats known to be present either end of the minisatellite array. Each variant repeat has been designated by the nomenclature used in 5.6B. Dots in the middle region depict the inner region of the minisatellite which has not been sequenced. Higher order repetitions are shown by arrows. B - sequences of the four variant repeat units (1-4). Variant 3 contains an *Alu* I recognition site (5' -AG↓CT-3'). C - alignment of the consensus sequences of 33.6, 33.15 and chi to the various repeats in c*Cfa*MP5. The three bases at the 3'end of variant repeat 3 show similarity together with the next five bases of the 5'end of any of the variants shown in B with the chi sequence. D - organisation of the DNA fragment in clone c*Cfa*MP5; the black boxes depict the regions of the repeat array which have been sequenced. Both the 5' and 3' flanking regions have been sequenced completely. The length of the repeat array was estimated from the size of the *Sau* 3AI fragment on an agarose gel.

FIGURE 5.6

A

```
   1 GGATCCAGCT GGAGTTACAG CACAGCGCTG CTTCCCTTGT NNCGGCAGAC
  51 AGCCAGCCCC CCACCGCGGC GTCGGCCTCC CTTCCCGCTG CCCTGTCCTG
 101 CTGCCTGGCA CTCACCGCCG GGGCAGCGGA GTCAGAGCTC GGCAGCCTTC
 151 CCACAGCCAC GTGTCACCAC AACAGTTTTG CTGATTTTTG CACCAGTGCG
 201 TTCGCCGATG AAGCTGGGGC AGAGAGGAGA GGT
```

```
  1122113113113112111133133232312122113l1..2121122142111  MINISATELLITE
                                                           REPEAT
```

```
1801 TGGNCTTCCA GTTGCCCNNC CTTCCNGGCC AAGGGAACCA AGGNACNNNT
1851 TACCGGGCCC CGAATTAGAG GGAAAAGTCG TCGTTGGNCC CGATGACTAT
1901 CCCTGCTTAA ACAACACGGG TCCAGCAGTC GANCGACGCT CCCGACTCTC
1951 AGGAGCAACG TTCAGAGANA TCAGGGCCCT TCAGACAGGG GCTCCCTAGG
2001 CGAGGGCCTG CCCAGAGCAC GCAAGGCCAC GACCAAGTTT CCAGGAGTTC
2051 CATTCAGAGC TCAGAAGAGG CGCTACGCGT GCACTAGGTA CATCCCTTTG
2101 ATTGCAGCAG CTGAAGAAAG TGTTTAGAGA AAAACTGTTT AAAGGTGATT
2151 GACCCTCGGC AGGAACAGCT GCGTGAAGAG CTGAGGCATT GAGGGCCGGT
2201 TCCTGGTCGT CTTAACAAAG CAAATGACTC CCAACTGTTT CTCTTTGCTC
2251 TTGAGTTTCC AGACAGGACG GCCGCTCCTC AGTAGTGGTT TCTCTCTTAA
2301 GTCCACACCA AACGTTAAGT CACTAAAGAA TTGCCCTCTA AGAATGGTGT
2351 GCACGGAGCA GGGGGACGTG AGGATNTCTT CGGGAAGCTG AGGAGGGGCT
2401 CGTCTAGGAC GCTGCAGCGA ATCCACCGGT ACAAGTGATA GTGATAAATA
2451 GATAGGGCCA TAAGATTATT GGCCATTGAT AGACTTCATG AATAGGATCC
```

B

Variant repeat

```
1.          GGAGGAGAGGT
2.          GGAGGAGGGGT
3.          GGAGGAGAGCT
4.          GGAGGAAGGGT
```

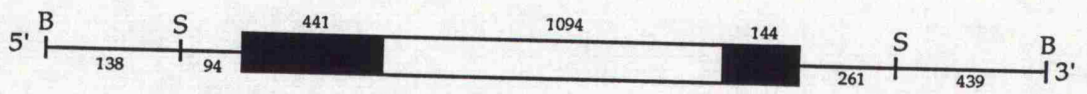
C

```
          4. GGAGGAAGGGT          4. GGAGGAAGGGT
             ||||||||||              ||||||||||
33.6    TGGAGGAAGGGCTGGAGGAGGCTCCGGAGGAAGGGC
                       |||||||
                    1. GGAGGAGAGGT
                    2. GGAGGAGGGGT
                    3. GGAGGAGAGCT
                    4. GGAGGAAGGGT

33.15  GAGGTGGGCAGGTGGAGAGGTGGGCAGGTGGA
                     || ||||||||
                  1. GGAGGAGAGGT
                  2. GGAGGAGGGGT
                  3. GGAGGAGAGCT
                  4. GGAGGAAGGGT

chi sequence                    GCTGGTGG
                                |||||  ||
                          GGAGGAGAGCTGGAGG
                             3       x
```

D



array at the D8S210 locus on the human chromosome 8. However, the regions of the minisatellite array the sequence for which is known are depicted in Figure 5.6D. Flanking sequences of 221bp and 700bp have been obtained. Only 441bp of repeat sequence on one end and 144bp on the opposite end of the array has been obtained. The minisatellite is flanked by non-repeated DNA, the 5' region of which is considerably GC-rich. Comparisons of the entire ~2.5kb region with all other sequences in the Genbank and EMBL databases did not detect any significant homology, although the repeat region showed varying degrees of similarity to a number of repeat regions in the genome of herpes simplex virus and also to a purine/pyrimidine rich region of the second intron of the human azurocidin gene.

DISCUSSION

There have been numerous reports of the sequences of minisatellite loci from a wide range of species; insects (Paulsson *et al.*, 1992; Jacobson *et al.*, 1992), mice (Kominami *et al.*,1988; Mitani *et al.*,1990; Gibbs *et al.*, 1993), rats (Mori *et al.*, 1992), humans (Jeffreys *et al.*, 1990), fish (Bentzen & Wright, 1993), birds (Gyllensten *et al.*, 1989) and primates (Gray & Jeffreys, 1991). These display varying degrees of polymorphism from the most variable MS1, a human minisatellite (heterozygosity > 99%) (Wong *et al.*, 1987; Jeffreys *et al.*, 1988) to the minimally variable equivalent of MS1 in primates (Gray & Jeffreys, 1991). They have a tendency to be GC-rich (Jacobson *et al.*, 1992) although AT-rich minisatellites have also been reported (Simmler *et al.*, 1987; Berg & Olaisen, 1993; Desmarais *et al.*, 1993). The repeating units within minisatellites vary from 2bp at the D8S210 locus (Brereton *et al.*, 1993) to 250bp in the minisatellite in the Balbiani ring 1 gene of *Chironomus tentans* (Paulsson *et al.*, 1992). Individual repeat units vary in sequence within repeat arrays (Jeffreys *et al.*, 1990, 1991a), from highly divergent repeats in *Drosophila* (Jacobson *et al.*, 1992) to repeat units in the human minisatellite MS32 that are virtually identical (Wong *et al.*, 1987, Jeffreys *et al.*, 1990). Furthermore, the distribution of variant repeat units within alleles differs from a well preserved ordered structure in the COL2A1 VNTR (Berg & Olaisen, 1993), to conserved margins with internal variation in alleles of the D8S210 locus (Brereton *et al.*, 1993), to diverged margins in MS32 (Wong *et al.*, 1987, Jeffreys *et al.*, 1990) and the DXYS17 HVR (Simmler *et al.*, 1987). Finally, a another common feature of minisatellites is their tendency to show strand assymetry i.e. either a purine-rich or pyrimidine rich strand bias (Coggins & O'Prey, 1989; Brereton *et al.*, 1993). It is clear therefore that the features of polymorphism, sequence content, size, variation in repeat heterogeneity/homogeneity and the distribution of variant repeats within minisatellite alleles are highly diverse among minisatellites.

A specific function has not yet been attributed to minisatellite sequences although the detection of core sequences within minisatellites that bear similarity (Jeffreys *et al.*, 1985a), and in some cases, complete identity (Jacobson *et al.*, 1992), with the *E.Coli* generalised recombination signal (Smith, 1983) has led researchers to speculate that they may be involved in promoting homologous recombination. Studies into the

ability of such sequences to promote homologous recombination involve those of Wahls *et al.*, (1990) where a synthetically derived consensus minisatellite sequence was shown to stimulate homologous recombination between two nonreplicating plasmids introduced into human somatic cells. Furthermore, attempts to find DNA-binding proteins that are involved the putative recombinogenic activity of these minisatellites, has resulted in the detection of four different minisatellite-specific DNA binding proteins (Collick & Jeffreys, 1990, Wahls *et al.*, 1991, Yamazaki *et al.*, 1992). Subsequent reports by Trepicchio & Krontiris (1992, 1993) have shown that minisatellites associated with or within genes can bind nuclear proteins and in doing so, at least in the case of that within the D_H-J_H interval of the immunoglobulin heavy chain gene, are capable of suppressing enhancer-mediated activation of the adenovirus major late promoter. Finally, three reports of the presence of minisatellite repeat arrays in coding genes suggests that a significant proportion of minisatellites exist as specific, functional sequences within the genome (Swallow *et al.*, 1987; Tsarfaty *et al.*, 1990; Paulsson *et al.*, 1992).

Studies into the mechanisms of the maintenance of length variation at these sequences have shown that these involve mainly intra-allelic unequal exchange or replication slippage events (Armour *et al.*, 1989b; Mitani *et al.*, 1990; Jeffreys *et al.*, 1990; Schlotterer & Tautz, 1992; Berg & Olaisen, 1993; Desmarais *et al.*, 1993) although instances of inter-allelic unequal exchange events, though less frequent, have also been detected (Jeffreys *et al.*, 1991a). Furthermore, Brereton *et al.*, (1993) have found that polymorphism at the D8S210 locus extends to the loss of single copy flanking sequence probably having arisen due to some mechanism other than those operating at repeat arrays.

How does the canine minisatellite sequence in c*Cfa*MP5 fit into this minisatellite 'arena'? The size of the repeating units in c*Cfa*MP5 (11bp) is certainly within the range described above and is both identical in size and similar in sequence to those found within the human hypervariable locus, D8S210 by Brereton *et al.*, (1993). The number of variant repeat units detected thus far are four although more may be actually present within the internal region of the array. This results in a degree of homogeneity between repeat units which varies from 73-90%. Tandem repeat arrays that undergo high rates of mutational length change through unequal crossing-over or slipped strand mispairing events has been suggested to consist of homogeneous arrays of short repeat units (Stephan, 1989). Two

hypervariable mouse minisatellites have been identified which have very short repeat units among which no variant repeat units have been detected (4bp for *Hm-2* - Gibbs *et al.*, 1993; 5bp for *Ms6-hm* - Kelly *et al.*, 1991). This has led to the suggestion that the related short repeat units of the two arrays may contribute to their similar patterns of instability. However, the most hypervariable human minisatellite isolated to date is MS1 which has a repeat unit size of only 9bp (Wong *et al.*, 1987; Jeffreys *et al.*, 1988) but in which the repeat units are not completely homogeneous. Furthermore, Brereton *et al.*, 1993 have reported the detection of a polypurine hypervariable locus, D8S210, (heterozygosity > 99%) where repeat units have been shown to vary in size from 2bp to 11bp. Mitani *et al.*, (1990) compared two polymorphic mouse minisatellites, Pc-1 and Pc-2, to a monomorphic locus, mo-1, and showed that they differed in repeat unit lengths. However, the analysis of a monomorphic probe by Zischler *et al.*, (1992) showed that the reason for its uniformity was in its repeat structure; this was short, inhomogeneous and consists of four different motifs encompassing only about 50bp. Thus, the evidence does not indicate one way or another as to the roles of short repeat unit length and degrees of homogeneity on the hypervariability of minisatellites.

There is a considerable purine/pyrimidine strand bias within the c*Cfa*MP5 repeat array, again similar to human minisatellites but not as extreme as that found at the D8S210 locus. Minisatellites that show strand asymmetry have also been shown to form triplex and knot-like tertiary structures: Coggins *et al.* (1992) have shown that linearized plasmids containing the 33.6 and 33.15 minisatellite repeat arrays (Jeffreys *et al.*, 1985a), when denatured and reannealed, form homoduplexes containing knot-like tertiary structures. They detected metastable interhelical associations between such homoduplexes and have suggested that these could promote recombination within the minisatellite array or associated flanking sequences. Brereton *et al.* (1993) have shown that the polypurine region of the D8S210 locus can adopt triplex conformations in vitro. They have suggested that the introduction of single-stranded regions *in vivo* by such conformations, could initiate recombination mechanisms and lead to the high rate of length mutation seen at D8S210. It is therefore possible that the purine/pyrimidine strand bias in c*Cfa*MP5 will render it susceptible to similar conformations, which could play a role in the generation of variation seen for this minisatellite.

Flanking sequence comparisons with existing sequences in the Genbank and EMBL databases were unsuccessful in detecting significant matches, indicating that the isolation and characterization of this canine minisatellite sequence is the first of its kind. There have been reports of the association of human and mouse minisatellites with repeat elements (Armour *et al.*, 1989a, 1992; Kelly *et al.*, 1989; Hendriks *et al.*, 1992). Furthermore, minisatellites have also been shown to be associated with microsatellite sequences (Iwasaki *et al.*, 1992; Bentzen & Wright, 1992). Although no such association was found in this study, it is not inconceivable that further analyses will reveal similar associations in other eukaryote genomes. Comparisons of the repeat region with those in the database detected some similarity to numerous repeat regions within the genome of Herpes simplex virus and also to a purine/pyrimidine rich region of the second intron of the human azurocidin gene. This is neither surprising nor significant since numerous repeat arrays have now been isolated containing purine/pyrimidine strand biases.

The only restriction enzyme found to cleave repeatedly within the minisatellite in c*Cfa*MP5 is *Alu* I. *Alu* I sites are scattered throughout the repeat array and exist only in repeat type 3. This is analogous to the human minisatellite MS32 in which all repeat units are virtually identical, other than a T→ G transversion that occurs in approximately two-thirds of the repeat units (Wong *et al.*, 1987; Jeffreys *et al.*, 1988). Furthermore, the junction of each 3 type repeat variant with any other repeat variant contains a sequence which is identical, excepting one base mismatch, to a GGCAGG motif. Mitani *et al.*, (1990) have proposed that this motif contributes to the germline instability of mouse minisatellites due to the observation that mutations in this sequence in result in the detection of monomorphisms. It is not inconceivable that a similar motif may be involved in contributing to the germline instability of canine minisatellites. Future analyses of further polymorphic canine minisatellite sequences will serve to resolve this speculation.

Figure 5.6b displays the similarities present between the repeat units in c*Cfa*MP5 and the chi sequence. These lie within the junction of every repeat type 3 and any other and are identical to the chi sequence except for a single base mismatch. The detection of sequences similar or identical to the chi sequence has, as described above, led to the suggestion that minisatellites may be involved in recombinogenic activity. However, the fact that several hypervariable minisatellite loci have been identified

which bear no resemblance to the chi sequence (Simmler *et al.*, 1987; Berg & Olaisen, 1993; Desmarais *et al.*, 1993) is inconsistent with this suggestion. Furthermore, population studies on the α-globin complex and the insulin minisatellite do not show that these are specific 'hot spots' of recombination (Higgs *et al.*, 1986; Cox *et al.*, 1989). Although there is a wealth of evidence for a link between recombination rate and minisatellite abundance, it remains to be seen whether this is simply ' a passive reflection of DNA behaviour and long-range genomic structure' or whether minisatellites really do have a role to play in recombination (Jarman & Wells, 1989 and references therein).

The similarity that the repeat units in c*Cfa*MP5 show to the consensus sequences of the human minisatellites, 33.6 and 33.15, is unsurprising as these were used as multilocus probes in the isolation of c*Cfa*MP5. However, it does not always follow that the use of specific repeat sequences in the isolation of minisatellite loci results in loci which show sequence similarity to the probe used for their isolation. Armour *et al.*, (1992a) have found that sequence comparisons with probe(s) used to detect human minisatellite loci suggest that a relatively poor match is sufficient for the positive detection of tandem repeats in a clone. This is further supported by the evidence of Zischler *et al.*, (1992) who found that the screening of human genomic libraries with $(CAC)_5$ yielded two clones, HZ32 and HZ41, neither of which contained a perfect $(CAC)_5/(GTG)_5$ target sequence.

Both Zischler *et al.*, (1992) and Brereton *et al.*, (1993) have suggested mechanisms whereby the repeat arrays in the minisatellites they have isolated (the clone HZ42 and the D8S210 locus, respectively) may have been generated. They have both envisaged single base mutation as the initial event after which subsequent events consist of larger scale intra-allelic events such as *en bloc* duplication. In c*Cfa*MP5, all repeats contain a conserved region of 6bp (GGAGGA) at the 5' end, which appears to have arisen from a duplication of the $(GGA)_n$ simple motif. One could envision the 11bp repeat arising through the duplication of this simple motif, followed by a series of point mutations resulting in the more divergent 3' end of the repeat. This event might have been followed by several *en bloc* duplications to constitute the reiterating 11bp repeat unit, which was subsequently changed slightly by point mutations. Such a scenario is very similar to that proposed by Zischler *et al.*, (1992) for HZ42. However, it should be noted that the above speculation as to the mechanism of

generation of c*Cfa*MP5 only holds true if the clone sequenced has not been rearranged *in vitro*.

The presence of higher order repeats in c*Cfa*MP5 suggests that events within this minisatellite subsequent to the occurrence of slight changes due to point mutations (see above) were larger duplications of certain groups of repeats. Finally, it should be noted that the association of minisatellites with dispersed repeat elements or microsatellites (see above), has given support to the view that minisatellites (as longer repeats) probably arose in these cases, as a consequence of mutational events creating new, longer motifs from tandemly arranged shorter ones (Levinson & Gutman, 1987).

The 5' margin of the minisatellite array in c*Cfa*MP5 shows divergence from the consensus sequence. This is unlike many other minisatellites where both margins tend to be either converged or diverged with respect to the consensus repeat sequence. Although totally unlike in sequence composition, c*Cfa*MP5 is, however, similar to the AT-rich DXYS15 HVR which also displays this unilateral divergence. The presence of multiple copies of each variant repeat diffused along the repeat array in c*Cfa*MP5 is in keeping with the suggestion that these are intermediates of crossover fixation; at ultravariable loci variant repeats are not usually present in single copies but in multiple copies as expected for intermediates of crossover fixation (Smith, 1976).

The nature of the repeat sequence in c*Cfa*MP5 is both similar to that of the D8S210 locus (Brereton *et al.*, 1993) and to that of the DXYS82 locus (Armour *et al.*, 1992b) where the bulk of the array consists of octanucleotide repeats (GGAGGGAA) which may be reduced to a twofold repeat of (GGAR). However, the latter locus has been shown to be a poor hybridization probe requiring prolonged exposure to detect longer alleles. This inefficient hybridization has been attributed to the simple nature of the repeat unit sequence causing relatively high levels of background signal by cross-hybridizing to other purine-rich regions of the genome. In contrast to the cMS630 minisatellite at the DXYS82 locus, c*Cfa*MP5 functions as a highly efficient single locus probe producing strong hybridization signals with very low background signals. Yet this efficiency is not shown by another canine minisatellite clone, c*Cfa*MP1, which when used as a single-locus probe detects less prominent hybridizing fragments together with the main locus (see Chapter 3, Results Section 3.6.3). This would suggest that specific sequence differences or degrees of cryptic

sequence similarity between various minisatellites determine their ability to function as single- or multi-locus probes.

As previously stated in the Introduction, Section 1.1.2, MVR mapping by PCR (MVR-PCR) is a technique which enables the determination of internal sequence variation within minisatellites. The first minisatellite to be mapped in this way has been the human minisatellite, MS32, at the locus D1S8 (Jeffreys *et al.*, 1991a). Subsequent to this, two further human minisatellite loci, MS31A at D7S21 (Neil & Jeffreys, 1993) and MS205 at D16S309 (Armour *et al.*, 1993) have also been mapped. The successful application of MVR-PCR relies on the availability of a locus which has (i) allelic variability, (ii) suitable internal variation and (iii) a constant repeat unit length. MS32 suits these requirements in that it consists of a 29 bp repeat unit which shows internal variation due to the presence or absence of a *Hae* III restriction site and has a heterozygosity of 97.5% based on allele length (Wong *et al.*, 1987).

MVR mapping of MS32 alleles (Jeffreys *et al.*, 1990, 1991a) has shown that different alleles can have related haplotypes with most of the inter-allelic differences being clustered towards one end of the map. These findings suggest that MVR mapping will be highly useful in population studies where the degree of divergence of specific populations is of interest. Regarding dogs, it is not inconceivable that individuals of the same breed may have related haplotypes. However, since MVR-PCR diploid codes are based on a single locus, siblings have a one in four chance of having the same diploid code, and therefore, an essential requirement of any population study would be the use of unrelated individuals. Admittedly, in the canine world, this is not an easy task, as dog breeders in this country often use the same 'champion' sire or immediate relatives for generating a pedigree population for 'show' purposes. In such circumstances, it would be advisable to collect unrelated canine DNA samples of a given breed from geographically isolated populations.

The current population of domestic dogs has been proposed to be the progeny of the domestication of the grey wolf (Wayne, 1993). If this is so, comparisons of MVR haplotypes of these related canids might show the degree to which the two have diverged since the domestication of wolves. Studies of the internal structures of the human MS32 and MS1 minisatellite allele equivalents in primates led Gray & Jeffreys, (1991) to conclude that the variation shown by minisatellites displays evolutionary

transience i.e. the primate equivalents of these loci were smaller than the human equivalents and showed little or no variation between different primate individuals. Although the period since domestication has only been several thousands of years (Scott, 1968) canids may have diverged sufficiently at variable loci such as minisatellites, to show differences in their MVR haplotypes.

Examination of c*Cfa*MP5 single-locus profiles of canine individuals from both different and identical breeds has shown that alleles can be shared between individuals. In the case of identical breeds, shared alleles are probably more often due to the effects of inbreeding rather than population substructuring. However, in the case of different breeds, these shared alleles (determined by Southern blot analysis), might represent alleles which although fortuitously identical in size, are different in their internal sequence structure. MVR mapping provides a means of distinguishing between identically sized alleles and would therefore show the true nature of shared alleles.

Although c*Cfa*MP5 fits the requirements of constant repeat unit size and allelic variation for the application of MVR-PCR in dogs, it falls 'short of the mark' due to the type of internal variation present between repeat units. As described in Chapter 5, Section 5.3, the repeat units in c*Cfa*MP5 contain a region of similarity at their 5' end, consisting of the sequence, GGAGGA. MVR-PCR requires the use of primers specific to each variant repeat unit. The presence of internal similarity between repeat units, however, would cause non-specific priming within the repeat array resulting in uninterpretable MVR-PCR patterns. Therefore the application of MVR-PCR to c*Cfa*MP5 has required the synthesis of degenerate primers and has been carried out by R.Regan (personal communication).

Initially primers were made to amplify the c*Cfa*MP5 locus in various dogs. Figure 5.7 shows the various primers used initially in PCR amplification of c*Cfa*MP5 in various dogs (E & F) and also the various primers used in the MVR-PCR analysis (A-D). Two variant repeat-specific primers were used: variant repeat primer A is specific to repeat variants 1, 2 or 4, whereas variant repeat primer B is specific to variant repeat 3 only. Table 5.1 shows the composition of and the conditions used for MVR-PCR analysis of c*Cfa*MP5. Although preliminary MVR-PCR analysis using the primers depicted in Figure 5.7 showed ladders of DNA fragments characteristic of each repeat variant 3 with respect to the others (data not shown), it requires further optimization. However, future MVR-PCR

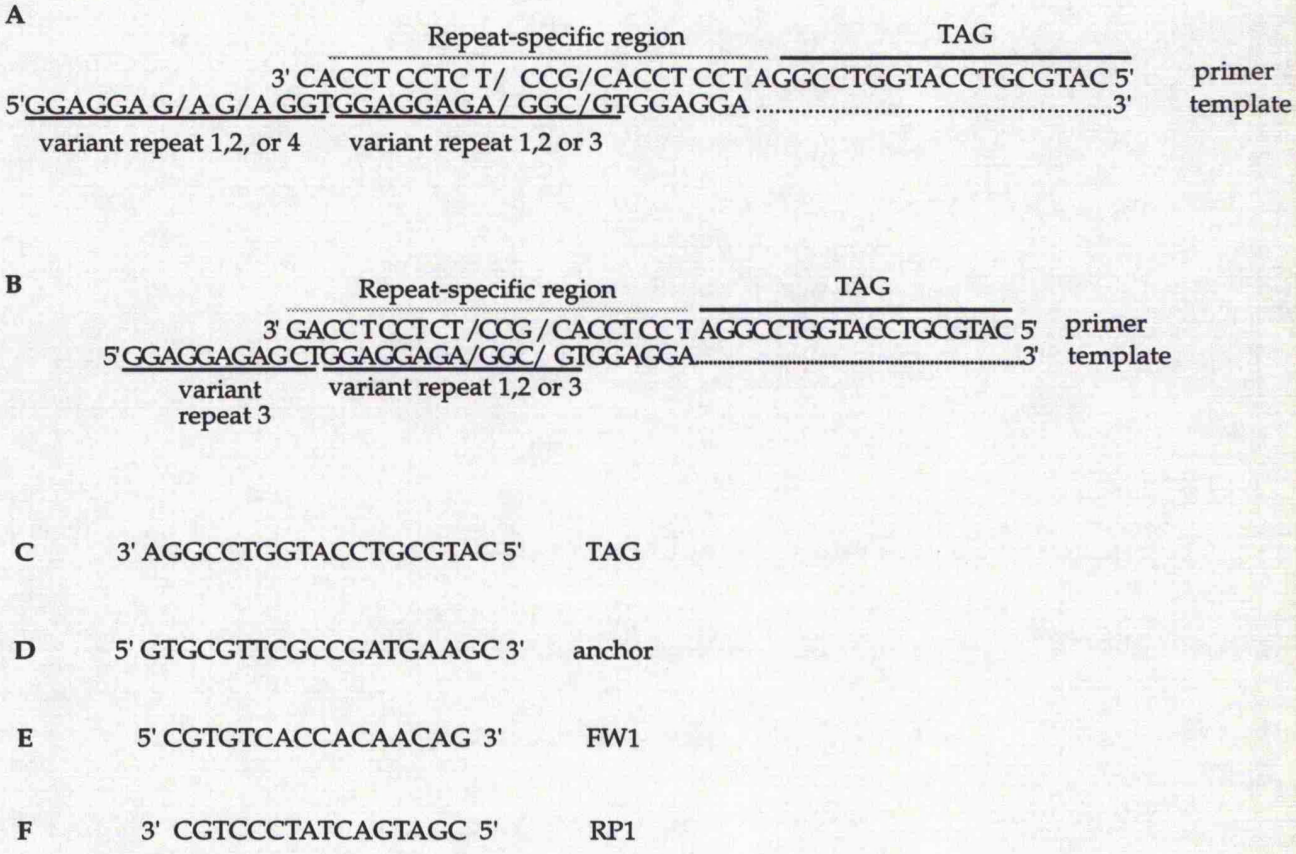


Figure 5.7 : Primers used in PCR amplification of and MVR-PCR analysis of c*Cfa*MP5. A - variant repeat primer specific to variant repeats 1, 2 or 4 which is adjacent to any of variant repeats 1, 2, or 3. **B** - variant repeat primer specific to variant repeat 3 which is adjacent to any of variant repeats 1, 2, or 3. **C** - TAG primer. **D** - anchor primer. **E** & **F** - primers used for PCR amplification of c*Cfa*MP5 in various dogs.

PCR Conditions						
Primers	**Concentration /μM**	**Denaturation**	**Annealing**	**Extension**	**No. of cycles**	**Taq polymerase**
anchor	1	1) 90°C for 1.3 mins	65°C for 1 min	70°C for 10 mins	18	1 unit in 10μl final volume
TAG	1	2)	65°C for 1 min	70°C for 10 mins	1	
A	0.07					
B	0.07					
PCR Buffer : 45mM Tris.HCl (pH 8.8), 11mM $(NH_4)_2 SO_4$, 4.5mM $MgCl_2$, 6.7mM 2-mercaptoethanol, 4.4μM EDTA (pH 8.0), 1mM each of dATP, dCTP, dTTP & dGTP, 113μg/ml BSA						

Table 5.1: PCR conditions used for MVR-PCR analysis of c*Cfa*MP5.

analysis of c*Cfa*MP5 promises a greater understanding of internal repeat sequence variation at this locus.

Conclusions

(i) The features of polymorphism, sequence content, size, variation in repeat heterogeneity/homogeneity and the distribution of variant repeats within minisatellite alleles are highly diverse among minisatellites.

(ii) Although it has been shown that there is a link between recombination rate and minisatellite abundance, it remains to be seen if an *in vivo* role for minisatellites, in maintaining length variation at these loci actually exists. It is known that minisatellites can bind nuclear proteins thereby suppressing the trancriptional activation of bacterial promoters. Furthermore, a number of minisatellites exist within coding regions of the genome thereby contributing to specific functions.

(iii) The main processes whereby length variation is generated at polymorphic minisatellite loci are currently thought to be intra-allelic unequal exchanges and DNA slippage during replication, although inter-allelic unequal exchange events have also been shown to occur.

(iv) Examination of various polymorphic minisatellite sequences indicates that a short repeat unit length is consistent with hypervariability although the role of the degree of homogeneity between repeat units within minisatellites is still unclear.

(v) The high purine/pyrimidine content of the repeat array in c*Cfa*MP5 suggests that it may display triplex conformations similar to those shown by other purine/pyrimidine-rich minisatellites, which could play a role in the generation of the variation seen for this minisatellite.

(vi) Any similarity shown between the repeat array in c*Cfa*MP5 and sequences such as the chi sequence, 33.6, 33.15, repeats within the DR2 region of Herpes simple virus or in the azurocidin gene, can be attributed to their common GC-rich nature and/or their high purine/pyrimidine strand asymmetry.

(vii) Finally, the sequence similarity shown by c*Cfa*MP5 and the repeat array at the locus DXYS82 contrasts sharply with their individual efficiencies as single-locus probes, indicating that the degree to which a minisatellite shows cryptic similarity to other repeats in the genome, plays an important role in determining the background signal intensity in their single-locus DNA profiles.

CHAPTER 6

DISCUSSION

The screening of 2940 colonies from the canine genomic library, SSJ1, has enabled the isolation of 7 polymorphic canine minisatellite loci which show heterozygosities of 20-88% and display more than two allelic forms within the individuals studied. There are a number of factors which have to be considered when comparing this yield to those of similar libraries generated in humans and chickens:

(i) The number of recombinants in SSJ1 was 0.3 fold less than in the chicken and human libraries. This means that a larger number of colonies from SSJ1 must be screened in order to achieve a one genome equivalent.

(ii) Although the size fractions used in the three libraries are comparable, the presence or absence of polymorphic minisatellite loci in the respective size fractions depends on the particular species.

(iii) The overall level of inbreeding in the human population as compared to that in the chicken and canine populations can be considered to be less, since line breeding does not generally occur in humans.

(iv) The composition of the canine 2-16kb genomic size fraction mostly consists of DNA fragments within the 2-4kb size range. This might explain the higher yield of less informative loci from SSJ1 compared to the human and chicken libraries, since the 2-4kb size range of canine *Mbo* I DNA fingerprints contains a higher proportion of invariant minisatellite loci.

Thus, the yield of polymorphic canine minisatellite loci is not significantly low considering the above mentioned factors. However, it is clear that the overall heterozygosity of the polymorphic clones isolated is less than their chicken and human counterparts. This could be a feature of continuous inbreeding in dogs over the last hundred or so years, or equally likely a feature of the size fraction chosen for the generation of the library. The generation and analysis of a library containing a larger genomic size fraction might resolve this question.

Is it worth screening SSJ1 with a further 4 multilocus probes (see Chapter 3)? Considering the lower number of recombinants present and the average size of the inserts cloned (2-4kb), it would seem to be a better option to screen another library containing a larger size fraction to obtain more polymorphic minisatellites with higher heterozygosities. However,

since the general shift towards the use of microsatellites as DNA markers for genome mapping rather than minisatellites due to their ubiquitous nature, their wide distribution and ease of isolation and subsequent analysis (Amos & Pemberton, 1992), it is unlikely that the present canine genomic library will be used to isolate more minisatellites. It will however be useful to screen the library to locate large inserts containing specific microsatellites so that linkage groups of microsatellites may be generated.

The minisatellites generated thus far can however be used to aid in filling gaps in the canine linkage map of microsatellite DNA markers and to aid in possible future paternity or forensic analyses where confirmation is required from several loci of the origins of a specific DNA sample. They will also be useful as single-locus probes in inter- and intra-breed analyses, population studies such as those including comparisons of related canids and some might be useful as canine-specific multi-locus probes in DNA fingerprinting analyses.

Future DNA fingerprinting studies using both the Jeffreys' minisatellite multilocus probes and the canine-specific ones generated in this study, might indeed show differences in the levels of band sharing between canine individuals from the same breed and those from different breeds. One might expect this difference to display itself in higher levels of band sharing between individuals from the same breed and relatively lower levels among individuals from different breeds. The careful collection of canine DNA from geographically isolated populations is necessary for such analyses since interpretation of data would be plagued by any degree of relatedness between individuals from the same breed.

The main difference between human and canine DNA fingerprints has been a lower number of DNA fragments in the latter and in the case of 33.15 the bands generated are of a lower intensity as well (Jeffreys & Morton, 1987). The use of the canine minisatellite polymorphic loci generated here as multilocus probes in DNA fingerprint analysis, should increase further the band intensity and the number of variable loci detected, thereby increasing the number of minisatellite probes available for research use.

The studies of inter- versus intra-breed variation using the minisatellites, c*Cfa*MP1 and c*Cfa*MP5, as single-locus probes, has shown that allelic variation is not constant throughout the canine population. Due to a combination of population substructuring, inbreeding effects and the inherent variability of each minisatellite, allelic variation is reduced

within breeds. It has yet to be seen whether this reduction in allelic variation is present among geographically isolated individuals from a specific breed. Furthermore, it is not possible to exclude the fact that the future use of a canine minisatellite equivalent to the highly variable human MS1 minisatellite (Gray & Jeffreys, 1991) as a single-locus probe, will no longer reveal a reduction in allelic variation. Such a circumstance might arise due to the high mutability of a given locus thereby counteracting the effects of genetic drift and preventing any particular alleles from attaining high frequencies within the population.

It was somewhat surprising that Southern blot analysis of a total of 39 different wolves did not yield readily detectable minisatellite loci in most cases. This is unlikely to be due to experimental error, since two separate analyses yielded similar results. It is however possible that in a similar manner such as the comparison of MS32 and MS1 human minisatellite equivalents in primates (Gray & Jeffreys, 1991), the wolf equivalent of c*Cfa*MP5 is no longer of a similar size as the canine one and requires PCR amplification for its detection. Such future analysis would be relatively simple to do since flanking primers to the c*Cfa*MP5 minisatellite have already been generated for MVR-PCR analysis of the locus.

The detection of c*Cfa*MM1 monomorphic equivalents of identical size in wolves suggests that this locus has remained stable since the domestication of wolves. What is it about certain minisatellite loci that enables them to remain stable through numerous generations? Sequencing of the flanking regions of minisatellites in mice and humans has revealed their association with dispersed repeat elements. Furthermore, MVR-PCR analysis of minisatellite alleles (Jeffreys *et al.*, 1991a; Armour *et al.*, 1993) has shown that there is a polarity of variation suggesting the involvement of cis-acting sequence elements that control *de novo* mutation in allele structure and direct its positioning within loci. It remains possible therefore that monomorphic minisatellite loci have lost such regulatory elements thus preventing any further mutation.

The demonstration of Mendelian inheritance within the Irish Setter family indicates the individuality of each minisatellite locus isolated. Athough only two of the seven isolated polymorphic minisatellites (c*Cfa*MP4 and c*Cfa*MP5) have been shown here to display Mendelian inheritance, c*Cfa*MP1 has also been used as a single-locus probe to screen the family (data not shown). The latter however was uninformative for the Irish Setter family used. This suggests that the level

of variation displayed by a given locus, although important (c*Cfa*MP1-heterozygosity = 59%; c*Cfa*MP4 - heterozygosity = 35%; c*Cfa*MP5 - heterozygosity = 88%) does not necessarily guarantee that it will be informative for a given pedigree, i.e. an individual who is heterozygous at one locus may be homozygous and so uninformative at another locus. Therefore, the use of polymorphic canine minisatellites to screen pedigrees requires the availability of a bank of such probes for testing, to allow for some which might be uninformative.

Paternity analysis using single-locus profiles requires a knowledge of the mutation rate and allele frequencies at a given minisatellite locus so that one can specify the probability of a chance match. However, in the case studied in Chapter 4, sufficient background information was known prior to analysis for the case to be solved without knowledge of mutation rates. Unfortunately, at present we do not have access to a sufficient number of families for studies into mutation rates.

The sequencing of the first canine minisatellite, c*Cfa*MP5, has shown that like many other mammalian minisatellites, it shares a constant repeat unit size (11bp), displays internal repeat sequence variation and is GC-rich with numerous internal sites showing similarity to the *E Coli.* chi recombination signal. Comparison of the complete sequence with those in the EMBL/Genbank databases showed no significant similarities, the only cryptically similar sequences being an intronic repeat region of the azurocidin gene and repeats within the genome of Herpes simplex virus. The isolation and characterization of c*Cfa*MP5 therefore adds to the current 'lake' of knowledge of the nature and features of minisatellites in general. However, it also provides a beginning for future studies into the evolution and origins of canine minisatellite alleles, enabling inter-breed comparisons to be made via MVR-PCR analysis. At present the MVR-PCR analysis of c*Cfa*MP5 alleles has yet to be optimized, but this is unlikely to be an unsurmountable problem.

The informativeness of MVR-PCR analysis on c*Cfa*MP5 alleles is unlikely to be great for related canids, since high stringency Southern blot analysis of wolves, coyotes and foxes has shown that c*Cfa*MP5 equivalents either no longer exist in these canids or are present as greatly reduced forms. Therefore, considering that it was possible to detect the monomorphic minisatellite equivalent, c*Cfa*MM1, in the above canids (see above) suggests that future sequencing of less variable minisatellites

might provide suitable loci for MVR-PCR analysis, which might then be used in related canid analysis.

Southern blot analysis using c*Cfa*MP5 as a single-locus probe of individuals from numerous breeds, has shown that alleles are shared between pedigree breed dogs and cross-breeds. However, whether this is due to a fortuitous sharing of identically-sized alleles which in actual fact are nonidentical in their internal repeat sequence structures, will only be known after MVR-PCR analysis of the respective alleles. Similarly, canine alleles that are rendered monomorphic according to Southern blot analysis may in fact show greater variation when their MVR-PCR internal repeat maps are examined. Such a study might have a considerable bearing on the conclusions drawn from the present analysis of the canine genomic library, SSJ1.

One of the main reasons for which this study has been conducted has been the necessity to generate markers which might prove useful in the diagnosis of affected individuals and carriers of genetically inherited disease. Although the family used to demonstrate Mendelian inheritance (Chapter 4), is known to have the disorder, progressive retinal atrophy (PRA), no linkage to the disease of either of the two minisatellite loci studied has been detected. This does not preclude the future detection of linkage were other loci analysed, although such efforts would be unnecessary since this locus has already been detected via the 'candidate gene' approach. However, there are many other canine genetically-inherited diseases for which linked DNA markers have yet to be found. The minisatellites isolated in this study provide a small fraction of the markers that will eventually be required to ensure that any given disease locus will be readily located.

Botstein *et al.*, (1980) have suggested the need for DNA markers to be spaced every 20 cM apart. For the canine genome this would require the isolation of approximately 160 DNA markers (haploid canine genome size - 3.3×10^9 bp) for the establishment of a set of well-spaced, highly polymorphic DNA markers covering the entire canine genome. This requires the isolation of markers which are known to be abundant and randomly distributed within the genome. Although mouse minisatellites have been shown to be randomly distributed within the genome this is not the case with human minisatellites, which have been found to be preferentially localised within the proterminal regions of human autosomes (Royle *et al.*, 1988). Jeffreys & Morton (1987) have found no

evidence for clustering of minisatellites within the canine genome from studies into the segregation of parental loci, and suggest that canine minisatellites must be dispersed within the genome at least to a certain extent. Whether this is definitely the case can only be confirmed by localising existing DNA markers on canine autosomes.

Finally, how important is genome mapping in dogs? Regarding it from the disease point of view, the presence of several diseases in dogs which are also present in man (see Introduction, Section 1.5), mean that the dog could be regarded as yet another animal model for subsequent analysis of these diseases and gene therapy. However, the commercial aspect of genome mapping in dogs is of considerable importance as the line breeding of dogs for sale to the general public as well as for 'show' purposes can provide real financial gain for dog breeders.

Concluding remarks:

(i) The generation of a canine genomic library to isolate polymorphic canine minisatellites has shown that they are of a lower variability to most already isolated. This could be due to both the method of the generation of the library as well as feature of inbreeding among dogs in general.
(ii) Application of the isolated minisatellites to numerous genetic analyses shows their versatility.
(iii) Sequencing of the most polymorphic minisatellite, c*Cfa*MP5, isolated to date shows its similarity to other reported GC-rich minisatellites.
(iv) Initial MVR-PCR analysis of c*Cfa*MP5 suggests its potential as a beginning to digital typing in dogs providing a method of comparing c*Cfa*MP5 alleles in different breeds and related canids.

The analysis of minisatellites in dogs therefore has not only opened the way for the future development of a canine linkage map of DNA markers but has enabled their use in various genetic analyses. Future analysis of canine minisatellite alleles will provide further information on the mutational processes whereby the variation seen at these loci is generated.

REFERENCES

Amos, B. & Pemberton, J. (1992). DNA fingerprinting in non-human populations. *Current Opinion in Genetics and Development* 2: 857-860.

Armour, J.A.L., Wong, Z., Wilson, V., Royle, V. & Jeffreys, A.J. (1989a). Sequences flanking the repeat arrays of human minisatellites : association with tandem and dispersed repeat elements. Nucleic Acids Research **17** (13): 4925-4835.

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

Armour, J.A.L. (1990a). 'The Isolation and Characterization of Human Minisatellite Loci.' *PhD thesis*, University of Leicester.

Armour, J.A.L., Povey, S., Jeremiah, S. & Jeffreys, A.J. (1990b). Systematic cloning of human minisatellites from ordered array Charomid libraries. *Genomics* **8**: 501-512.

Armour, J.A.L., Vergnaud, G., Crosier, M. & Jeffreys, A.J. (1992a). 'Isolation of human minisatellite loci detected by synthetic tandem repeat probes: direct comparison with cloned DNA fingerprinting probes.' *Human Molecular Genetics* Vol **1**, No.5, 319-323.

Armour, J.A.L., Crosier, M. & Jeffreys, A.J. (1992b). Human minisatellite alleles detectable only after PCR amplification. *Genomics* **12**: 116-124.

Armour, J.A.L., Harris, P.C. & Jeffreys, A.J. (1993). Allelic diversity at minisatellite MS205 (D16S309): evidence for polarized variability. *Human Molecular Genetics* **2** (8) : 1137-1145.

Bell, G.I., Selby, M.J. & Rutter, W.I. (1982). The region near the human insulin gene is composed if simple tandemly repeated sequences. *Nature* **295**: 31-35.

Bentzen,P. & Wright, J. (1992). Nucleotide sequence and evolutionary conservation of a minisatellite variable number of tandem repeat cloned from the Atlantic salmon, *Salmo salar* . *Genome* **36**: 271-277.

Berg, E.S. & Olaisen, B. (1993). Characterization of the COL2A1 VNTR polymorphism. *Genomics* **16**: 350-354.

Botstein, D., White, R.L., Skolnick, M., & Davis, R.W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics* **32**: 314-331.

Brereton, H.M., Firgaira, F.A. & Turner, D.R. (1993). Origins of polymorphism at a polypurine hypervariable locus. *Nucleic Acids Research* 21 (11): 2563-2569.

Brock, M.K. & White, B.N. (1991). Multifragment alleles in DNA fingerprints of the parrot, *Amazona ventralis*. Journal of Heredity 82: 209-212.

Burke, T. & Bruford, M.W. (1987). DNA fingerprinting in birds. *Nature (London)* 327: 149-152.

Bruford, M.W. (1992). 'Hypervariable Markers in the Chicken Genome.' *PhD thesis*, University of Leicester.

Bruford, M.W., Hanotte, O., Brookfield, J.F.Y. & Burke, T. (1992). 'Single-locus and multi-locus DNA fingerprinting'. *Molecular Genetic Analysis of Populations. A Practical Approach.* Ch 225-269 IRL Press Oxford.

Bruford, M.W., Hanotte, O. & Burke, T. (1994). Minisatellite DNA markers in the chicken genome. *Animal Genetics* (in press).

Buluwela, L. Forster, A. Boehm, T. & Rabbits, T.H. (1989). A rapid procedure for colony screening using nylon filters. NAR 17: 452.

Burke, T. & Bruford, M.W. (1987). 'DNA fingerprinting in birds.' *Nature*, 327: 149-151.

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

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

Chapman, B.S., Vincent, K.A. & Wilson, A.C. (1986). Persistence of rapid generation of DNA length polymorphism at the zeta-globin locus of humans. *Genetics* 112: 79-92.

Cocciolone, R.A. & Timms, P. (1992). DNA profiling of Queensland Koalas reveals sufficient variability for individual identification and parentage determination. *Wildlife Research* 19: 279-287.

Coggins, L.W. & O'Prey, M. (1989). DNA tertiary structures formed in vitro by misaligned hybridization of multiple tandem repeat sequences. *Nucleic Acids Research* 17: 7417-7426.

Coggins, L.W., O'Prey, M. & Akhter, S. (1992). Intra-helical pseudoknots and interhelical associations mediated by mispaired human minisatellite DNA sequences *in vitro* . *Gene* **121**: 279-285.

Collick, A. & Jeffreys, A.J. (1990). Detection of a novel minisatellite-specific DNA -binding protein. *Nucleic Acids Research* **18** (3): 625-629.

Cox, N.J., Bell, G.I., & Xiang, K. (1989). Linkage disequilibrium in the human insulin/insulin-like growth factor II region of human chromosome II. *American Journal of Human Genetics* **43** : 495-501.

Deka, R., Chakraborty, R. & Ferrell, R.E. (1991). A population genetic study of six VNTR loci in three ethnically defined populations. *Genomics* **11**: 83-92.

Desmarais, E., Vigneron, S., Buresi, C., Cambien, F., Cambou, J.P. & Roizes, G. (1993). Variant mapping of the Apo(B) AT rich minisatellite. Dependence on nucleotide sequence of the copy number variations. Instability of the non-canonical alleles. *Nucleic Acids Research* **21** (9): 2179-2184.

Ellegren, H., Hartman, G., Johansson, M. & Andersson, L. (1993). Major histocompatibility complex monomorphism and low levels of DNA fingerprinting variability in a reintroduced and rapidly expanding population of beavers. *Proceedings of the National Academy of Science USA.* **90** : 8150-8153.

Epplen, J.T., Ammer, H., Epplen, C., Kammerbauer, C., Mitreiter, R., Roewer, L., Schwaiger, W., Steimle, V., Zischler, H., Albert, E., Andreas, A., Beyermann, B., Meyer, W., Buitkamp, J., Nanda, I., Scmid, M., Nurnberg, P., Pena, S.D.J., Poche, H., Sprecher, W., Schartl, M., Weising, K. & Yassouridis, A. (1991). Oligonucleotide fingerprinting using simple repeat motifs: A convenient way to detect hypervariability for multiple purposes. In "DNA fingerprinting: approaches and applications." ed. T. Burke, G.Dolf, A.J. Jeffreys & R. Wolff. Birkhauser Verlag, Bern.

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

Flint, J., Boyce, A.J., Martinson, J.J. & Clegg, J.B. (1989). Population bottlenecks in Polynesia revealed by minisatellites. *Human Genetics* **83** : 257-263.

Fowler, S.J., Gill, P., Werrett, D.J. & Higgs, D.R. (1988). Individual specific DNA fingerprints from a hypervariable region probe: Alpha-globin 3'HVR. *Human Genetics* **79** : 142-146.

Georges, M., Lequarre, A.-S., Castelli, M., Hanset, R. & Vassart, G. (1988) DNA fingerprinting in domestic animals using four different minisatellite probes. *Cytogenetics and Cell Genetics* **47**: 127-131.

Georges, M., Hilbert, P., Lequarre, A., Leclerc, V., Hanset, R. & Vassart, G. (1989). Use of DNA bar codes to resolve a canine paternity dispute. *JAVMA* **193** (9) : 3-6.

Georges, M., Lathrop, M., Hilbert, P., Marcotte, A., Schwers, A., Swillens, S., Vassart, G., & Hanset, R. (1990). On the use of DNA fingerprints for linkage studies in cattle. *Genomics* **6** : 461-474.

Gibbs, M., Collick, A., Kelly, R.G. & Jeffreys, A.J. (1993). A tetranucleotide repeat mouse minisatellite displaying substantial somatic instabiltiy during early preimplantation development. *Genomics* **17**: 121-128.

Gilbert, D.A., Lehman, N., O'Brien, S.J. & Wayne, R.K. (1990). Genetic fingerprinting reflects population differentiation in the California Channel Island fox. *Nature* **344**: 764-766.

Gill, P., Jeffreys, A.J. & Werret, D.J. (1985). Forensic applications of DNA fingerprints. *Nature* **318**: 577-579.

Gray, I.C. & Jeffreys, A.J. (1991). Evolutionary transience of hypervariable minisatellites in man and the primates. *Proceedings of the Royal Society of London B* **243** : 241-253.

Gross, D.S. & Garrard, W.T. (1986). The ubiquitous potential Z-forming sequence of eukaryotes (dT-dG)n. (dC-dA)n is not detectable in the genomes of eubacteria, archaebacteria or mitochondria. *Molecular and Cellular Biology* **6** : 3010-3013.

Gyllensten, U.B., Jakobsson, S., Temrin, H. & Wilson, A.C. (1989). Nucleotide sequence and genomic organization of bird minisatellites. *Nucleic Acid Research* **17**: 2203-2215.

Haig,S.M., Belthoff, J.R. & Allen, D.H. (1993). Examination of population structure in red-cockaded woodpeckers using DNA profiles. *Evolution* **47** (1) : 185-194.

Hanotte, O., Burke, T., Armour, J.A.L., & Jeffreys, A.J. (1991). Cloning , characterization and evolution of Indian Peafowl *Pavo cristatus* minisatellite loci. In: Burke, T., Dolff, G., Jeffreys, A.J. & Wolff, R. (eds) *DNA Fingerprinting : Approaches and Applications*, Birkhauser Verlag, Basel, pp 193-216.

Hanotte, O., Cairns, E., Robson, T., Double, M.C. & Burke, T. (1992a). Cross-species hybridization of a single-locus minisatellite probe in passerine birds. *Molecular ecology* 1: 127-130.

Hannotte, O., Bruford, M.W., & Burke, T. (1992b). Multilocus DNA fingerprints in gallinaeceous birds: general approach and problems. *Heredity* **68** : 481-494.

Hendriks, R.W., Hinds, H., Chen, Z. & Craig, I.W. (1992). The hypervariable DX255 locus contains a LINE-1 repetitive element with a CpG island that is extensively methylated only on the active X chromosome. *Genomics* **14** : 598-603.

Hermans, I.F., Atkinson, J., Hamilton, J.F. & Chambers, G.K. (1991). Three cases of disputed paternity in dogs resolved by the use of DNA fingerprinting. *New Zealand Veterinary Journal* **39** : 61-64.

Higgs, D.R., Wainscoat, J.S., Flint, J., Hill, A.V.S., Thein, S.L., Nicholls, R.D., Teal, H., Ayyub, H., Peto, T.E.A., Falusi, A.G., Jarman, A.P., Clegg, J.B. & Weatherall, D.J. (1986). Analysis of the human a-globin gene cluster reveals a highly informative genetic locus. *PNAS* **83** : 5165-5169.

Hill, W.G. (1987). DNA fingerprints applied to animal and bird populations. *Nature* **327**: 98-99.

Hoelzel, A.R. (ed) (1992) Molecular Genetic Analysis of Populations. A Practical Approach. pp 225-269 IRL Press Oxford.

Holmes, N.G., Mellersh, C.S., Humphreys, S.J., Binns, M.M., Holliman, A., Curtis, R., Sampson, J. (1993). Isolation and characterization of microsatellites from the canine genome. *Animal Genetics* **24** : 289-292.

Iwasaki, H., Stewart, P.W., Dilley, W.G., Holt, M.S., Steinbrueck, T.D., Wells, S.A. & Donis-keller, H. (1992). A minisatellite and a microsatellite polymorphism within 1.5kb at the human muscle glycogen phosphorylase (PYGM) locus can be amplified by PCR and have combined informativeness of PIC 0.95. *Genomics* **13** : 7-15.

Jacobson, J.W., Guo, W. & Hughes, C.R. (1992). A *Drosophila* minisatellite contains multiple chi sequences. *Insect Biochemistry & Molecular Biology* 22 (8) : 785-792.

Jarman, A.P. & Wells, R.A. (1989). Hypervariable minisatellites: recombinators or innocent bystanders? *Trend in Genetics* 5 (11): 367-371.

Jeffreys, A.J., Wilson, V. & Thein, S. L. (1985a). Hypervariable 'minisatellite' regions in human DNA. *Nature* **314** : 67-73.

Jeffreys, A.J., Brookfield, J.F.Y. & Semeonoff, R. (1985b). 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. & Ponder, B.A.J. (1986). DNA fingerprints and segregation analysis of multiple markers in human pedigrees. *American Journal of Human Genetics* **39** : 11-24.

Jeffreys, A.J. & Morton, D.B. (1987). DNA fingerprints of dogs and cats. *Animal Genetics* **18** : 1-15.

Jeffreys, A.J., Wilson, V., Kelly, R., Taylor, B.A., Bulfield, G. (1987). Mouse DNA 'fingerprints' : analysis of chromosome localization and germ-line stability of hypervariable loci in recombinant inbred strains. *Nucleic Acids Research* Vol **15** (7) 2823-2836.

Jeffreys,A.J., Royle, N.J., Wilson, V. & Wong, Z. (1988). Spontaneous mutation rates to new length alleles at tandem-repetitive hypervariable loci in human DNA. *Nature* **332** : 278-281.

Jeffreys, A.J., Neumann, R. & Wilson, V. (1990). Repeat unit sequence variation in minisatellites: a novel source of DNA polymorphism for studying variation and mutation by single molecule analysis. *Cell* **60** : 473-485.

Jeffreys, A.J., Macleod, A., Tamaki, K., Neil, D.L. & Monckton, D. (1991a). Minisatellite repeat coding as a digital approach to DNA typing. *Nature* **354** : 204-209.

Jeffreys, A.J., Royle, N.J., Patel, I., Armour, J.A.L., Macleod, A., Collick, A., Gray, I.C., Neumann, M., Gibbs, M., Crosier, M., Hill, M., Signer, E. & Monckton, D. (1991b).'Principles and Recent Advances in Human DNA Fingerprinting.' In: Burke, T., Dolff, G., Jeffreys, A.J. & Wolff, R. (eds) *DNA Fingerprinting : Approaches and Applications*, Birkhauser Verlag, Basel, pp 1-19.

John, B. & Miklos, G.L.G. (1979). Functional aspects of satellite DNA and heterochromatin. *International Review of Cytology* **58** : 1-114.

Kelly, R., Bulfield, G., Collick, A., Gibbs, M. & Jeffreys, A.J. (1989). Chracterization of a highly unstable mouse minisatellite locus: evidence for somatic mutation during early development. *Genomics* **5** : 844-856.

Kelly, R., Gibbs, M., Collick, A. & Jeffreys, A.J. (1991). Spontaneous mutation at the hypervariable mouse minisatellite locus *Ms6 - hm*: flanking DNA sequence and analysis of germline and early somatic mutation events. *Proceedings of the Royal Society of London* B. **245** : 235-245.

Kominami, R., Mitani, K. & Muramatsu, M. (1988). Nucleotide sequence of a mouse minisatellite DNA. *Nucleic Acids Research* **16** (3): 1197.

Kuhnlein, U., Zadworny, D., Dawe, Y., Fairfull, R.W. and Gavora, J.S. (1990). Assessment of inbreeding by DNA fingerprinting: development of a calibration curve using defined strains of chickens. *Genetics* **125** : 161-165.

Levinson, G. & Gutman, G.A. (1987). Slipped-strand mispairing : a major mechanism for DNA sequence evolution. Molecular Biology and *Evolution* **4** (3): 203-221.

Lewin, B. (1990). 'The Nature of Active Chromatin.' *Genes IV* Ch 22 : 436-439.

Luty, J.A., Guo, Z., Willard, H.F., Ledbetter, D.H. & Litt, M. (1990). Five polymorphic VNTRs on the human X cromosome. *American Journal of Human Genetics* **46** : 776-783.

Macdonald, D. (1992). The Velvet Claw. Publ. BBC Books, Woodlands, 80, Wood Lane, London, W12 OTT.

Madsen, C.S., Ghivizzani, S.C. & Hauswirth, W.W. (1993). In vivo and in vitro evidence for slipped mispairing in mammalian mitochondria. *Proceedings of the National Academy of Sciences USA.* **90** : 7671-7675.

Malcolm, S., Clayton-Smith, J., Nichols, M., Robb, S., Webb, T., Armour, J.A.L., Jeffreys, A.J., Pembrey, M.E. (1991). Uniparental paternal disomy in Angelman's syndrome. *The Lancet* **331**: 694-697.

Mannen, H., Tsuji, S., Goto, N. & Fukuta, K. (1993). Identification of inbred strains of mice and genetic relationships between strains as assessed by DNA fingerprinting. *Exp. Anim .* **42** (2) : 169-173.

Martinson, J.J., Harding, R.M., Philippon, G., Flye Sainte-Marie, F., Roux, J., Boyce, A.J. & Clegg, J.B. (1993). Demographic reductions and genetic bottlenecks in humans: minisatellite allele distributions in Oceania. *Human Genetics* **9** : 445-450.

Millichamp, N.J., Curtis, R. & Barnett, K.C. (1988). Progressive retinal atrophy in Tibetan Terriers. *JAVMA* **192** (6) : 769-776.

Mitani, K., Takahashi, Y. & Kominami, R. (1990). A GGCAGG motif in minisatellites affecting their germline instability. *Journal of Biological Chemistry* **265** (25): 15203-15210.

Monckton, D.G. & Jefferys, A.J. (1991). Minisatellite 'isoallele' discrimination in pseudohomozygotes by single molecule PCR and variant repeat mapping. *Genomics* **11**: 465-467.

Mori, M., Ishzaki, T., Serikawa, T. & Yamada, J. (1992). Instability of the minisatellite sequence in the first intron of the rat renin gene and localization of the gene to chromosome 13q13 between FH and PEPC loci. *Journal of Heredity* 83 : 204-207.

Nakamura, Y., Leppert, M., O'Connell, P., Wolff, R., Holm, T., Culver, M., Martin, C., Fujimoto, E., Hoff, M., Kumlin, E. & White, R. (1987). Variable number of tandem repeat (VNTR) markers for human gene mapping. *Science* 235 : 1616-1622.

Nakamura, Y., Lathrop, M., O'Connell, P., Leppert, M., Barker, D., Wright, E., Skolnick, M., Kondoleon,S., Litt, M., Lalouel,J. M. and White, R. (1988). A mapped set of DNA markers for human chromosome 17. *Genomics* 2 : 302-309.

Neil, D.L. & Jeffreys, A.J. (1993) Digital typing at a second hypervariable locus by minisatellite variant mapping. *Human Molecular Genetics* 2 (8) : 1129-1135.

O'Brien, S.J.,Roelke, M.E., Marker, L., Newman, A., Winkler, C., Mettzer, D., Colly, L., Evermamann, J.F., Bush, M. & Wildt, D.E. (1985). Genetic basis for species vulnerability in the cheetah. *Science* 227: 1428-1434.

Page, D.C., Beiker, K., Brown, L.E., Hinton, S., Leppert, M., Lalonel, J.-M., Lathrop, M., Nystrom-Lahti, M., De La Chapelle, A. & White, R. (1987). Linkage, physical mapping and DNA sequence analysis of pseudoautosomal loci on the human X and Y chromosomes. *Genomics* 1: 243-256.

Paulsson, G., Hoog, C., Bernholm, K., Weislander, L. (1992). Balbiani ring gene 1 in *Chironomus tentans* : sequence organization and dynamics of a coding minisatellite. *Journal of Molecular Biology* 225 : 349-361.

Pirotta, V. (1976) Two restriction endonucleases from*Bacillus globiggi*.. *Nucleic Acids Research* 3 : 1747-1759.

Raleigh, E.A., Murray, N.E., Revel, H., Blumenthal, R.M., Westaway, D., Reith, A.D., Rigby, P.W.J., Elhai, J. & Hanahan, D. (1988). Mcr A and Mcr B restriction phenotypes of some E.Coli strains and implications for gene cloning. Nucleic Acids Research 16 : 1563-1575.

Reeve,H.K., Westneat, D.F., Noon, W.A., Sherman, P.W. & Aquadro, C.F. (1990). DNA fingerprinting reveals high levels of inbreeding in colonies of the eusocial naked mole-rat. *Proceedings of the National Academy of Sciences, USA* . 87 : 2496-2500.

Royle, N.J., Clarkson, R.E., Wong, Z. & Jeffreys, A.J. (1988). 'Clustering of hypervariable minisatellites in the proterminal regions of human autosomes. *Genomics* **3** : 854-859.

Sambrook, J., Fritsch, E.F., & Maniatis, T. (1990). *Molecular Cloning: a laboratory manual* (2nd edn). Cold Spring Harbor Laboratory Press. Cold Spring Harbor, New York.

Schlotterer, C. & Tautz, D. (1992). Slippage synthesis of simple sequence DNA. *Nucleic Acids Research* **20** (2) : 211-215.

Scott, J.P. (1968). Evolution and domestication of the dog. *Evolutionary Biology* **2** : 243-275.

Simmler, M.C., Johnsson, C., Petit, C., Rouyer, F., Vergnaud, G. & Weissenbach, J. (1987). Two highly polymorphic minisatellites from the pseudoautosomal region of the human sex chromosomes. *The EMBO Journal* **6** (4): 963-969.

Singer, M.F. (1982). Highly repeated sequences in mammalian genomes. *International Reviews in Cytology* **76** : 67-112.

Smith, G.P. (1976). Evolution of repeated sequences by unequal crossing-over. *Science, Wash* . **191**: 528-535.

Smith, G.R. (1983). Chi hotspots of generalized recombination. *Cell* **34** : 709-710.

Southern, E. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98** : 503.

Steinmetz, M., Stephan, D. & Lindahl, K. (1986). Gene organization and recombinational hotspots in the murine major histocompatibility complex. *Cell* **44** : 895-904.

Stephan, W. (1989). Tandem-repetitive noncoding DNA: forms and forces. *Molecular Biology and Evolution* **6** (2): 198-212.

Strand, M., Prolla, T.A., Liskay, R.M. & Petes, T.D. (1993). Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. *Nature* **365** : 274-276.

Swallow, D.M., Gendler, S., Griffith, B., Corney, G., Taylor-Papadimitriou, J. & Bramwell, M.E. (1987). The human tumour-associated epithelium mucins are coded by an expressed hypervariable gene locus PUM. *Nature* **328** : 82-84.

Tabor, S. & Richardson, C.C. (1987). DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. PNAS **84** : 4767-4771.

Taylor, A.C., Marshall Graves, J.A., Murray, N.D. & Sherwin, W.B. (1991). Conservation genetics of the Koala (*Phascolarctos cinereus*) II. Limited variability in minisatellite DNA sequences. *Biochemical Genetics* **29** (7/8) : 355-363.

Trepicchio, W.L. & Krontiris, T.G. (1992). Members of the *rel* /NF-*x* B family of trancriptional regulatory proteins bind the *HRAS1* minisatellite DNA sequence. *Nucleic Acids Research* **20** (10): 2427-2434.

Trepicchio, W.L. & Krontiris, T.G. (1993). *IGH* minisatellite suppression of USF-binding-site- and Eμ-mediated transcriptional activation of the adenovirus major late promoter. *Nucleic Acids Research* **21** (4): 977-985.

Tsarfaty, I., Hareuveni, M., Horev, J., Zaretsky, J., Weiss, M., Jeltsch, J.M., Garnier, J.M., Lathe, R., Keydar, I. & Wreschner, D.H. (1990). Isolation and characterization of an expressed hypervariable gene coding for a breast-cancer-associated antigen. *Gene* **93** : 313-318.

Van Pijlen, I., Amos, B. & Burke, T. (1994) Patterns of genetic variability at individual minisatellite loci in Minke whales (*Balaenoptera Acutorostrapa*) populations in different oceanic systems. Molecular Biology & Evolution (manuscript submitted).

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.

Vergnaud, G., Mariat, D., Zoroastro, M. & Lauthier, V. (1991). Detection of single and multiple polymorphic loci by synthetic tandem repeats of short oligonucleotides. *Electrophoresis* **12** : 134-140.

Wahls, W.P., Wallace, L.J. & Moore, P.D. (1990). Hypervariable minisatellite DNA is a hotspot for homologous recombination in human cells. *Cell* **60** : 95-103.

Wahls, W.P., Swenson, G. & Moore, P.D. (1991). Two hypervariable minisatellite DNA binding proteins. *Nucleic Acids Research* **19** (12): 3269-3274.

Wayne, R.K., Lehman, N., Girman, D., Gogan, P.J.P., Gilbert, D.A., Hansen, K., Peterson, R.O., Seal, U.S., Eisenhawer, A., Mech, L.D., Krumenaker, R.J. (1991a). Conservation genetics of the endangered Isle Royale gray wolf. *Conservation Biology* **5** (1) : 41-51.

Wayne, R.K., George, S.B., Gilbert, D., Collins, P.W., Kovach, S.D., Girman, D. & Lehman, N. (1991b). A morphological and genetic study of the island fox, *Urocyon Littoralis*. *Evolution* **45** (8): 1848-1868.

Wayne, R.K. (1993). Molecular evolution of the dog family. *Trends in Genetics* **9** (6) : 218-224.

Weber, J.L. (1990). Informativeness of $(dT\text{-}dG)_n$. (dC-dA)n polymorhisms. *Genomics* **7**: 524-530.

Weller, P., Jeffreys A.J., Wilson, V. Blanchetot, A. (1984). Organisation of the human myoglobin gene.*EMBO J.* **3** : 439-446.

Wetton, J.H., Carter, R.E., Parkin, D.T. & Walters, D. (1987). Demographic study of a wild house sparrow population by DNA fingerprinting. *Nature* **327**: 147-149.

Wetton, J.H., Parkin, D.T. & Carter, R.E. (1992). The use of genetic markers for parentage analysis in Passer domesticus (House Sparrows). *Heredity* **69** : 243-254.

White, R., Leppert, M., Bishop, D.T., Barker, D., Berkowitz, J., Brown, C., Callahan, P., Holm, T. & Jerominski, L. (1985a). Construction of linkage maps with DNA markers for human chromosomes. *Nature* **313** : 101-105.

White, R., Woodward, S., Leppert, M., O'Connell, P., Hoff, M., Herbst, J., Lalouel, J-M., Dean, M. & Woude, G.V. (1985b). A closely linked genetic marker for cystic fibrosis. *Nature* **318** : 382.

Willis, M.B. (1989). Genetics of the dog. Published by H.F. & G. Witherby Ltd. London.

Wolff, R.K., Nakamura, Y. & White, R. (1988). Molecular characterisation of a spontaneously generated new allele at a VNTR locus : no exchange of flanking sequence. *Genomics* **3** : 347-351.

Wolff, R.K., Plaetke, R., Jeffreys, A.J. & 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.

Wolff, R., Nakamura, Y., Odelberg, S., Shiang, R. & White, R. (1991). Generation of variability at VNTR loci in human DNA. In: Burke, T., Dolf, G., Jeffreys, A.J. & Wolff, R. (eds) *DNA fingerprinting: approaches and applications* , pp. 20-38. Birkhauser Verlag, Basel.

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. *Nucl. 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. *Annals of Human Genetics* **51** : 269-288.

Wyman, A.R. & White, R. (1980). A highly polymorphic locus in human DNA. *Proceedings of the National Academy of Sciences USA.* **77** (11): 6754-6758.

Wyman, A.R., Mulholland, J. & Botstein, D. (1986). Oligonucleotide repeats involved in the highly polymorphic locus D14S1. *American Journal of Human Genetics* **39** : A226.

Yamazaki, H., Nomoto, S., Mishima, Y. & Kominami, R. (1992). A 35-kDa protein binding to a cytosine-rich strand of hypervariable minisatellite DNA. *Journal of Biological Chemistry* **267** (17): 12311-12316.

Young, S. (1990). Naked mole rats keep it in the family. *New Scientist* , 12 May.

Zischler, H., Kammerbauer, C., Studer, R., Grzeschik, K.-H. & Epplen, J.T. (1992). Dissecting $(CAC)_5/(GTG)_5$ multilocus fingerprints from amn into individual locus-specific, hypervariable components. *Genomics* **13** : 983-990.