USE OF AN OVINE BACTERIAL ARTIFICIAL CHROMOSOME LIBRARY FOR THE STUDY OF BOVIDAE GENOMES

by

Clare Alexandra Gill

December, 1999

A thesis submitted in partial fulfilment
of the requirements of the award of
PhD in the Department of Animal Science
at the University of Adelaide

CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

'	minute
"	second
A	adenine
BAC	bacterial artificial chromosome
bp	base pair
5-BrdU	5-bromodeoxyuridine
BSA	bovine serum albumin
BTA	*Bos taurus*
C	cytosine
°C	degree Celsius
cDNA	complementary deoxyribonucleic acid
cfu	colony forming units
CHEF	contour-clamped homogeneous electric field
CHI	*Capra hircus*
CIAP	calf intestinal alkaline phosphatase
cm	centimetre
cM	centiMorgan
dATP	2'-deoxyadenosine 5'-triphosphate
dCTP	2'-deoxycytosine 5'-triphosphate
dGTP	2'-deoxyguanosine 5'-triphosphate
dH_20	distilled water
DMSO	dimethyl sulphoxide
DNA	deoxyribonucleic acid
dTTP	2'-deoxythymidine 5'-triphosphate
EDTA	ethylenediamine tetra acetic acid

FCS	foetal calf serum
FISH	fluorescence *in situ* hybridisation
FITC	fluorescein isothiocyanate
g	gram
G	guanine
HAT	hypoxanthine, aminopterin, thymidine
HBSS	Hank's balanced sodium salts
HPRT	hypoxanthine phosphoribosyl transferase
HSA	*Homo sapiens*
h	hour
IPTG	Isopropyl-ß-D-thiogalactopyranoside
ISH	*in situ* hybridisation
KAP	keratin associated protein
kb	kilobase
kDa	kiloDaltons
KIF	keratin intermediate filaments
KRT	keratin
L	litre
LB	Luria broth
LINE	long interspersed nuclear element
M	molar
mA	milliampere
ml	millilitre
mM	millimolar
mRNA	messenger ribonucleic acid
Mb	megabase
nm	nanometres

ng	nanograms
OAR	*Ovis aries*
OD	optical density
PAC	P1 derived artificial chromosome
PBS	phosphate buffer saline
PCR	polymerase chain reaction
PHA	phytohaemaglutinin
pmol	picomoles
PMSF	phenylmethyl sulfonyl fluoride
PPD11	ρ-phenylenediamine dihydrochloride, pH 11
PRINS	oligonucleotide primed *in situ* hybridisation
RH	radiation hybrids
RLGS	restriction landmark genome scanning
rpm	revolutions per minute
RNA	ribonucleic acid
S-phase	stationary phase
SCH	somatic cell hybrids
SDS	sodium dodecyl sulphate
SINE	short interspersed nuclear element
SNP	single nucleotide polymorphism
SPRINT	super-fast primed *in situ* targeting
SSC	sodium chloride sodium citrate
SSM	slipped-stranded mispairing
STC	sequence-tagged connectors
T	thymine
TAE	tris acetate ethylenediaminetetra-acetic acid
TBE	tris borate ethylenediaminetetra-acetic acid

TE	tris ethylenediaminetetra-acetic acid
TK	thymidine kinase
µg	microgram
µl	microlitre
µM	micromolar
U	restriction endonuclease unit
UTR	untranslated region
UV	ultraviolet
V	volts
v/v	volume for volume
w/v	weight for volume
X-Gal	5-bromo-4-chloro-3-indolyl-ß-D-galactopyranoside
YAC	Yeast artificial chromosome

ABSTRACT

Incorporating genes for wool quality traits into breeding programmes would benefit producers, but to target these genes, their location in the genome must be known. The aim of this project was to develop tools to assist in the construction of the ovine genome map so economic trait loci in sheep can be identified. Methods leading to a rapid improvement in the status of the ovine physical map were of particular interest.

In an attempt to utilise short PCR products or oligonucleotides for physical mapping, super-fast primed *in situ* targeting (SPRINT) was developed. To demonstrate the efficacy of SPRINT for detecting repeats, human satellite III and ovine satellite I core sequences were used to prime *in situ* incorporation of fluorescein into a newly synthesised strand. SPRINT, however, was ineffective at detecting single copy sequences, and consequently, an ovine bacterial artificial chromosome (BAC) library was generated as a source of large-insert probes for fluorescence *in situ* hybridisation (FISH).

A library of 59,904 *HindIII* BAC clones with an average insert size of 103kb representing two genome equivalents was constructed from the DNA of a purebred Suffolk ram. The ovine BAC library was characterised by PCR-based screening of pooled DNA, and BAC clones for 69 microsatellites and genes were isolated. Fifty loci were anchors for the ovine genetic map and 19 were comparative mapping loci. At least one BAC clone was physically assigned by FISH to every sheep chromosome, except OAR18, and no chimaeric clones were detected by chromosomal *in situ* hybridisation. These anchored clones were used to confirm chromosome identification in dual localisation experiments with previously unmapped loci. Additionally, ovine BAC clones were simultaneously mapped to sheep, cattle and goat metaphase chromosomes to improve the knowledge of genome organisation and rearrangements that have occurred in the evolution of these three closely related members of the Bovidae family.

To demonstrate the utility of BAC clones for investigating genome structure and function, a BAC clone containing a keratin associated protein gene (KAP1.1) that mapped to 11q3.1 was characterised by long-range restriction mapping and shot-gun sequencing. Sequences from the ovine KAP BAC clone were compared to GENBANK sequences from a homologous human BAC clone. There was strong support for the localisation of an as yet unidentified type I keratin gene, as well as the high sulphur KAP1.1, KAP1.3, KAP1.4, KAP2.3 and KAP3.4 genes within the ~100kb BAC clone. A striking feature of both the ovine and human sequences was the localised, high concentration of repetitive DNA in this region. It is speculated that since KAP genes lack introns, they may require the accumulated repetitive DNA in the intergenic regions for proper regulation or function.

The ovine BAC library is available for collaborative research and will be a valuable resource for comparative mapping, positional cloning, and studying gene function and regulation.

DECLARATION

I certify that this thesis does not contain material which has been accepted for the award of any degree or diploma; and that to the best of my knowledge and belief it does not contain any material previously published or written by another person except where due reference is made in the text. I give consent to this copy of my thesis, when deposited in the University Library, being available for loan and photocopying.

C.A. Gill

9/11/00

DEDICATION

To Jeff, for his loving support, motivation and patience.

To my family, for all their encouragement.

ACKNOWLEDGMENTS

I would sincerely like to thank my supervisor, Prof. Cindy Bottema, for her guidance and enthusiastic support throughout my PhD and for her editorial comments on this thesis. I thank my co-supervisor, Dr. Barry Powell for his helpful comments on the keratin study.

I am indebted to Dr. Scott Davis, Dr. Jerry Taylor and Dr. Noelle Cockett for their contributions to this project, which was also supported by the International Wool Secretariat. I thank Li Cai, Song and Xiaodong Wu, Ryan Harbinger, Jeff Gill and Brenden Holland for their help in constructing the BAC library and everyone in the lab at Texas A&M who made my stay so enjoyable.

Thanks also to Dr. Dan Gallagher and Jan Cook for helping me to initiate and maintain the fibroblast cell lines; Tracy Shay and Helena Kojevnikoff for helping to extract the BAC DNA pools; Lillian Allen, Annette Pollington, and Bozena Kruk for their assistance with library screening and FISH; Dr. Graham Webb and Judy Cheong for their help with photography. I also thank the other staff and students in the Dept. of Animal Science who have helped make the past few years very enjoyable and productive. Finally, many thanks to my friends and family for all their support and for helping to keep me sane!

CHAPTER 1

LITERATURE REVIEW

1.1 Introduction

Wool produced in Australia in 1996-97 had a gross value of $2.6 billion with 156 million sheep and lambs shorn. The yield of clean wool, free of impurities in the fleece (eg. grease, dirt and vegetable matter), that is used by the textile industry is about 65% of the shorn fleece (Australian Bureau of Statistics). Over the past decade the gross value and clean wool yield have stabilised, so methods that help to improve the quality and quantity of clean wool need to be adopted by producers to increase profitability. By incorporating genes for economically important traits (eg. fibre diameter, fibre strength, clean fleece weight) into breeding programmes, producers will be able to make genetic gains more rapidly than by traditional methods. However, before producers can target these genes, the regions of the genome where they are located must first be identified.

1.2 Objectives

The aim of this project was to develop a set of tools to assist in the construction of a dense genome map for sheep so that the regions of the genome containing genes for economically important traits could be identified, and subsequently isolated and characterised. In particular, the objectives were:

1) To develop novel molecular biological techniques to physically localise genes, polymorphic microsatellite markers, and other repetitive DNA elements from the ovine genome that would be useful for genome mapping. Specifically,

i) to improve the sensitivity, resolution, speed and cost-effectiveness of *in situ* hybridisation methodologies for sheep, and

ii) to construct an ovine bacterial artificial chromosome library.

2) To use these newly developed tools to investigate the organisation of the ovine genome, including the structural and functional roles of genes and interspersed repetitive

elements in the genome. In particular, the organisation of the keratin associated protein gene family and its associated repetitive elements was investigated as this gene family may be involved in wool quality traits.

1.3 Literature Review

1.3.1 Genome mapping

A genome map is a combination of the genetic and physical maps for a species, which gives the location of genes and their genetic neighbours. Such a map also provides information on the evolution of genomes, gene structure and functional aspects of gene location (Goodfellow, 1993). The ultimate goal of ovine genome mapping, from a production perspective, is to isolate and characterise economically important genes (eg. genes that affect growth rate, body composition, disease resistance and wool characteristics). In order to achieve this, a dense map of polymorphic markers and physically localised genes must be developed (Beckman and Soller, 1983; Crawford *et al.,* 1995; de Gortari *et al.,* 1998). This could be accomplished by using comparative mapping information from other closely related livestock species (de Gortari et al., 1997; Schibler et al., 1998b) and 'map-rich' species like human and mouse (REFS) as a source of genes and new markers to generate *de novo* mapping information in sheep.

The human and mouse genomic maps are very well developed because of the vast resources allocated to these projects. Rather than independently duplicate the production of very dense genome maps that has been one focus of the human and mouse projects, livestock genome mappers can target their gene searches based on the information that has already been generated in these 'map-rich' species. The human and mouse maps consist of over 10,000 and 7000 microsatellites, respectively, with an average spacing between markers of about 0.2cM or 400kb (Dib *et al.,* 1996, Dietrich *et al.,* 1996). The relatively small spacing between markers is now adequate to isolate genes by shot-gun sequencing strategies (Claverie, 1994; Roach, 1995). By 1996,

~16000 genes had been mapped in humans and 4000 genes had been mapped in mouse (Dib *et al.*, 1996; Schuler *et al.*, 1996; Dietrich *et al.*, 1996). More than 30,000 genes have now been mapped in humans (Deloukas *et al.*, 1998; Weissenbach, 1998; Strausberg *et al.*, 1999) and this number continues to increase rapidly. Full-length sequences for only about 6000 genes are in the Genbank database (Strausberg *et al.*, 1999). The remaining assigments were made using expressed sequence tags (ESTs) that have been clustered into "Unigenes" based on sequence similarity. There are now numerous Internet sites to access the enormous amount of mapping and sequence data being generated by the human genome project which can be exploited for comparative mapping in 'map-poor' species (Table 1.1).

Livestock gene mapping projects have generally adopted the strategies developed by the human genome project. Initially, most mapping in sheep was performed using somatic cell hybrids (Tucker *et al.*, 1981; Saidi-Mehtar *et al.*, 1991; Burkin *et al.*, 1998). The advent of PCR technology (Saiki *et al.*, 1988) and development of microsatellite markers (Dietrich *et al.*, 1992; Serikawa *et al.*, 1992; Weissenbach *et al.*, 1992) was the main impetus for the construction of genetic maps (Crawford *et al.*, 1995; de Gortari *et al.*, 1998). Fluorescence *in situ* hybridisation (Bauman *et al.*, 1980) has superceded radioactive *in situ* hybridisation (Pardue and Gall, 1969) for the physical localisation of genes, and fine mapping of candidate regions has utilised a variety of large-insert libraries such as cosmids (Buchanan *et al.*, 1993) and YAC clones (Broom and Hill, 1994). Bacterial artificial chromosomes (BAC) are now the library of choice for gene mapping (Kim *et al.*, 1996) and an ovine BAC library will be another tool for gene isolation in sheep and other livestock species. It seems likely that the next tools to be adopted by livestock researchers will be genotyping with microarrays of single nucleotide polymorphisms (SNP-chips; Ramsay, 1998; Sapolsky *et al.*, 1999) and physical mapping by PCR-based screening of radiation hybrid (RH) panels, although an RH

panel is not yet available for sheep (McCarthy *et al.*, 1997; Deloukas *et al.*, 1998; Womack *et al.*, 1997; Yerle *et al.*, 1998).

1.3.2 Ovine Genome Maps

Beckman and Soller (1983) proposed that a genetic map consisting of 200 markers spaced at regular intervals over the entire mammalian genome would be needed to give about a 90% chance that any one of the markers would segregate with a gene of economic interest. The size of the ovine genome was estimated to be a minimum of 2770cM based on the number of chiasmata per cell (Chapman and Bruere, 1977). The first generation ovine genome map covered 2070cM and consisted of 246 DNA polymorphisms with an average spacing between the 174 framework markers of 14.4cM (Crawford *et al.*, 1995). Markers comprised 86 anonymous microsatellites derived from the sheep genome, 126 microsatellites from cattle, 1 microsatellite from deer and 33 polymorphic markers associated with known genes. The physical location of 65 of these markers was determined by *in situ* hybridisation. There was poor coverage on chromosomes 5, 7, 11 and 13 with linkage groups spanning <50cM. Although the general location of several economically important genes (eg. Booroola fecundity gene) was determined from this genome map, the spacing between markers (several million bases) was too large to efficiently isolate genes by a positional candidate cloning approach. Once critical regions of the genome were identified, fine mapping was undertaken to place more informative markers within the regions and thereby reduce the size of the critical regions.

Many of the markers on the ovine genetic map were assigned using 'comparative' gene mapping strategies that rely on the conservation of unique sequences between species. Microsatellites are flanked by unique sequences and it has been found that 40-80% of the polymorphic microsatellites from cattle can be used in sheep, goats and other Bovidae (Georges *et al.*, 1991; Vaiman *et al.*, 1994).

Table 1.1: Genomic Mapping Databases on the World Wide Web

	Web Site	Web Address
Human	Human genome database	http://www.gdb.org
	Whitehead/MIT Center for Genome Research	http://www-genome.wi.mit.edu/
	Généthon	http://www.genethon.fr/genethon_en.html
	Human Transcript Map	http://www.ncbi.nlm.nih.gov/science96/
	Human Genome Project	http://www.ornl.gov/TechResources/Human_Genome/project/hgp.html
	The Institute for Genomic Research	http://www.tigr.org/
	Human Genome Organisation (HUGO)	http://hugo.gdb.org/
	Stanford Human Genome Centre:	http://shgc.stanford.edu/
Murine	Mouse genome informatics	http://www.informatics.jax.org/
	MRC HGU Mouse Atlas	http://glengoyne.hgu.mrc.ac.uk/
Bovine	US Bovine gene mapping	http://bos.cvm.tamu.edu/bovgbase.html
	Bovine ArkDB	http://www.ri.bbsrc.ac.uk/cgi-bin/arkdb/browsers/browser.sh?species=cattle
	BovMap – INRA	http://locus.jouy.inra.fr/cgi-bin/bovmap/intro.pl
	Animal Genome Database in Japan	http://ws4.niai.affrc.go.jp/jgbase.html
	MARC	http://sol.marc.usda.gov/
	Cattle genome database (CSIRO, Australia)	http://spinal.tag.csiro.au/
Ovine	Sheep ArkDB	http://www.ri.bbsrc.ac.uk/cgi-bin/arkdb/browsers/browser.sh?species=sheep
	AgResearch SheepMap Database	http://dirk.invermay.cri.nz/
	CAB sheep map	http://rubens.unimelb.edu.au/~jillm/pages/isag_chr.htm
Caprine	GoatMap	http://locus.jouy.inra.fr/bovmap/goatmap
Comparative	Bovine and Human on mouse comparative maps	http://bos.cvm.tamu.edu/htmls/HBM.html
	Human and Mouse on bovine comparative maps	http://bos.cvm.tamu.edu/htmls/BHM.html
	INRA comparative homology database	http://locus.jouy.inra.fr/cgi-bin/lgbc/mapping/common/taxonomy.pl
	Homologous unigenes between human and mouse	http://www.tree.caltech.edu/Hs-Mm.html
Other	Online mendelian inheritance in animals	http://probe.nalusda.gov:8300/animal/omia.html
	Agricultural Genome Information System (AGIS)	http://probe.nalusda.gov:8000/
	National Centre for Biotechnology Information	http://www.ncbi.nlm.nih.gov/
	Genomic and Genetic Resources on the WWW	http://www.nhgri.nih.gov/Data/
	Livestock Animal Genome Databases	http://probe.nalusda.gov:8300/animal/index.html
	Roslin Institute	http://www.ri.bbsrc.ac.uk/homepage.html
	Online cytogenetics of animals	http://www.angis.su.oz.au/Databases/BIRX/ocoa/
	Mendelian inheritance in sheep (COGNOSAG)	http://www.angis.su.oz.au/Databases/BIRX/mis/

The human and mouse maps are very extensive and many of the markers on the bovine map were derived by using sequences from these maps (Womack, 1991), which have in turn been used on the sheep map.

The second generation ovine genetic linkage map consists of 519 markers and 70% of the markers are common to the USDA MARC bovine map (Kappes *et al.*, 1997). The average spacing between adjacent linked markers is ~6.4cM (de Gortari *et al.*, 1998), but the physical map is still less well developed. There are 222 genes and 207 markers on the ovine physical map (ArkDB, 6/8/99).

1.3.3 Tools for genome mapping

1.3.3.1 Reference flocks for genetic linkage analysis

Physical mapping techniques localise genes to specific chromosomal regions, but the genetic association of genes can only be measured by the frequency with which they are co-inherited. A genetic map is constructed by testing the inheritance of genetic markers within families. An essential requirement of genetic linkage studies is families in which the segregation of loci can be demonstrated unequivocally. The number of families required to construct a genetic map depends on the density of the map, the degree of informativeness of the families due to allelic segregation, and the family size (Crawford *et al.*, 1995; Montgomery and Crawford, 1997; Broad *et al.*, 1998).

International ovine 'reference mapping families' have been established in New Zealand and Australia as a common resource for researchers involved in genome mapping (Broad and Hill, 1994). The five breeds that contributed to the pedigrees (Texel, Coopworth, Perendale, Romney and Merino) were used to try to maximise the heterozygosity of the F_1 generation, and thus maximise the informativeness of markers within the population. Large full-sib families (7 to 17 individuals) were established using two rounds of multiple ovulation and embryo transfer (Crawford *et al.*, 1995).

Nine three-generation, full-sib pedigrees with a total of 98 progeny were used for segregation analysis to generate the most recent genetic linkage maps (Crawford *et al.*, 1995; de Gortari *et al.*, 1998).

Genetic data collected from the International Mapping Flock was merged with data from the USDA reference population to produce the second-generation linkage map (de Gortari *et al.*, 1998). The USDA reference population of 247 backcross progeny was produced by mating four F1 rams (two Suffolk x Romanov, a Rambouillet x Romanov and a Roman x Rambouillet) to 44 Romanov ewes (de Gortari *et al.*, 1997).

1.3.3.2 Genetic linkage analysis

Genetic linkage is the co-segregation of pairs of genes or genetic markers due to their proximity on the same chromosome. During meiosis, frequency of recombination is proportional to the distance between two loci on the chromosome and rarely occurs between loci that are close together. If the parental genotypes are known, then the distance between two loci can be estimated from the genotypes of their offspring, because they indicate the proportion of recombination events that occurred during meiosis (Hetzel, 1991; Broad and Hill, 1994). The unit of distance in a genetic linkage map is centiMorgans (cM), where 1cM corresponds to a recombination frequency of 1%.

Linkage is exploited when analysing the effect of one or more markers on a phenotype in animals from a pedigree. Genetic linkage analysis is carried out using a panel of polymorphic markers, selected from a genetic map on the basis of their distribution. This is followed by some form of interval mapping analysis in which recombination fractions and the size of gene effects are estimated. Linkage between markers and economically important traits will enable the introduction of these traits into elite germlines by marker assisted selection or by genetic engineering. In turn, this will lead to improvements in animal health and productivity (O'Brien *et al.*, 1988).

1.3.3.3 Microsatellites

Microsatellites are ideal genetic markers because they are highly polymorphic and easy to analyse by the polymerase chain reaction (PCR). Microsatellites are tandem arrays of a 2-6bp core sequence that are embedded in DNA which usually is unique sequence. Microsatellites are ubiquitous in eukaryotes (Valdes *et al.,* 1993) and are abundant throughout the genome in coding and non-coding regions, occurring approximately once every 10kb (Tautz, 1989). Microsatellites are common in introns, 3' untranslated regions and 5' untranslated regions, but exonic microsatellites are rare (Ellegren, 1993). Dinucleotide repeats are the most abundant microsatellites and nearly all the markers on the ovine genetic map are $(GT)_n$ repeats (de Gortari *et al.,* 1998).

In general, the inheritance of microsatellite alleles is stable and Mendelian. However, several human diseases have been described that are due to an expansion in the number of trinucleotide repeat units from one generation to the next (Richards and Sutherland, 1992). Examples include Fragile X caused by expansion of a CCG repeat and Huntington's disease which is due to expansion of a CAG repeat (Mandel, 1993).

There is considerable variation in the repeat motifs present in different species. For instance, arrays of (TAA) are common in plants, but are rare in mammals (Tautz *et al.,* 1986). There is also variation in the repeat motifs that are present in coding regions. Dinucleotides and tetranucleotides are generally found in non-coding regions, but 57% of GC rich trinucleotides are found in coding regions (Wang *et al.,* 1994).

1.3.3.3.1 Mechanisms for generating microsatellite length variation

Microsatellite length variation in eukaryotes is biased towards expansion. Slipped-stranded mispairing (SSM) during replication has been invoked as the mechanism for the evolution of microsatellites and for the generation of length variability. SSM is an intra-helical event involving two strands of a single DNA duplex. Local denaturation and displacement of a strand of the DNA duplex causes mispairing of

complementary bases within a microsatellite (Levinson and Gutman, 1987). Polypyrimidine or polypurine DNA tracts, common in microsatellites, are susceptible to local denaturation and would allow SSM (Murphy *et al.,* 1989). Replication or repair gives insertion or deletion of repeat units. The most common change should be a single repeat loss or gain (Tautz, 1992; Strand *et al.,* 1993).

SSM, as described above, cannot be the mechanism for the massive expansions that have been observed in human diseases caused by trinucleotide repeat instabilities (eg. Fragile X). In these diseases, the microsatellite in the offspring is often larger than the sum of the repeat alleles inherited from the parents (Richards and Sutherland, 1994). Mitotic cross-over is not the source of the expansion because linkage disequilibrium is observed, meaning that alleles on either side of the repeat are always co-inherited with the disorder from the transmitting parent (Oostra and Verkerk, 1992; The Huntington's disease collaborative research group, 1993).

Richards and Sutherland (1994) proposed a new model for SSM that explains both small variations in microsatellite length and the generation of large expansions (Figure 1.1). Anchored slippage events with single-strand breaks cause loop formations. When the Okazaki fragment is anchored at the 5' end, repair mechanisms add bases. The increased length means that the chance of another single-strand break is also increased. When two single-strand breaks occur, the Okazaki fragment is no longer anchored and can slide along the sequence, adding many bases.

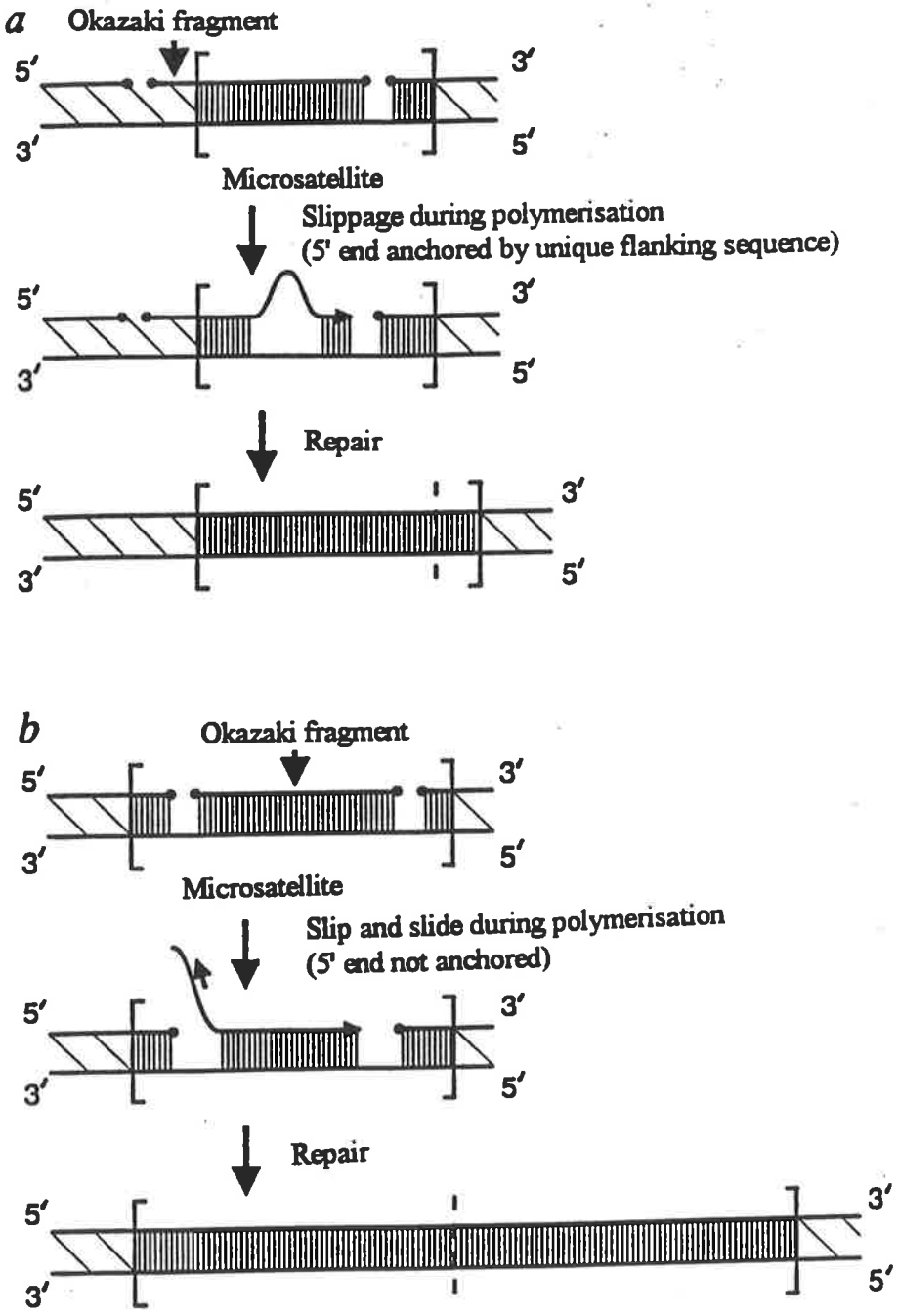


Figure 1.1: Models for microsatellite variability generated by slipped-stranded mispairing. (a) When the microsatellite is <80 repeats, only one single-strand breaks is likely to occur within the repeat during replication. Slippage during polymerisation can result in the addition of a few copies of the repeat. (b) For microsatellites >80 repeats, it is possible that two single-strand breaks within the repeat occur during replication. The strand between these breaks is not anchored by unique sequence and is free to slide during polymerisation, enabling the addition of many more repeat units than were present in the original sequence. (Adapted from Richards and Sutherland, 1994).

1.3.3.4 Physical mapping strategies

1.3.3.4.1 Chromosome identification

Physical mapping is the localisation of sequences to specific chromosomal regions by physical methods, such as the detection of probe sequences hybridised to metaphase chromosomes. Consequently, the unambiguous identification of chromosomes is essential for physical mapping.

A 'karyotype' refers to the number and appearance of the chromosomes of a species. For instance, the human karyotype consists of 17 pairs of 'metacentric' autosomes of various sizes, 5 pairs of 'acrocentric' autosomes and 2 metacentric sex chromosomes, whilst the bovine karyotype is 29 pairs of similarly sized acrocentric autosomes and 2 metacentric sex chromosomes. Metacentric means that the chromatid arms are joined by a centromere that is positioned centrally and acrocentric means that the centromere is positioned towards one end. The ovine karyotype consists of 27 pairs of chromosomes including three pairs of metacentric autosomes. The remaining autosomes and X chromosome form a gradually descending size-series of acrocentrics. The Y chromosome is a small metacentric (Broad and Hill, 1994).

A comparison of the bovid karyotypes has revealed extensive chromosomal homology. Chromosome 1, the largest of the sheep metacentrics is the equivalent of cattle chromosomes 1 and 3 fused at the centromeres. Chromosome 2 is the homologue of cattle chromosomes 2 and 8 centrically fused and chromosome 3 is the equivalent of cattle 5 fused to cattle 11. This homology extends to all of the other chromosomes except cattle chromosomes 8 and X (ISCNDA, 1990).

The unambiguous identification of sheep and cattle chromosomes is difficult because the chromosomes cannot be identified on the basis of morphology alone, since they are mainly acrocentric and decrease gradually in size. Metaphase chromosomes can be identified by the banding patterns that are generated by different stains, due to the differential uptake of the stain by various domains along the chromosome arms.

Structural banding methods are based on the intrinsic nature of the chromosomes and are applied to fixed chromosomal preparations, while dynamic banding techniques are based on a modification of the chromosomal DNA induced by the incorporation of a thymidine analogue during cell culture (Broad *et al.,* 1998). Common structural banding techniques are Q-banding produced by quinacrine mustard (Casperson *et al.,* 1968); G-banding, produced by trypsin digestion and Giemsa dye (Wang and Federoff, 1972); R-banding produced by treatment with hot salt solutions and Giemsa dye (Sehested, 1974), and C-banding produced by multiple treatments in acidic, basic and hot salt solutions followed by Giemsa staining (Pardue and Gall, 1970). Dynamic methods include replication banding produced by exposing the cells to 5-BrdU in late S phase and staining with Hoescht (Latt *et al.,* 1976). Simultaneous R- or G-banding of propidium iodide stained chromosomes can also be induced using an alkaline (pH 11) p-phenylenediamine antifade solution (Lemieux *et al.,* 1992).

Unless sequential banding is performed, these banding techniques are unable to distinguish several of the ovine autosomes (specifically 21, 24, 25 and 26). A combination of Q-banding, G-banding and silver staining of the nucleolar organiser regions on prometaphase and early metaphase chromosomes (for increased resolution) from sheep with Robertsonian translocations (centromeric fusions) has resolved these ambiguities (Ansari *et al.,* 1993). Banding techniques are important tools in physical gene mapping because (in combination with *in situ* methods) they help to confirm the order of genes along the chromosomes (Burkholder, 1993).

1.3.3.4.2 Somatic cell hybrids

Somatic cell hybrids (SCH) have been used extensively as a tool for assigning genes and markers to chromosomes. They have been particularly useful for mapping in livestock species where large backcross families for genetic linkage analysis are expensive to generate.

SCH contain chromosomes from two different species (eg. sheep x hamster; (Burkin *et al.*, 1998)) and are formed by virus or polyethylene glycol (PEG) mediated cell fusion (Harris and Watkins, 1965; Pontecorvo, 1975). SCH can be used for gene mapping because although the hybrids will initially contain a complete complement of chromosomes from both species, in the next few divisions, chromosomes from one species will be lost (Abbott and Povey, 1995). A gene can be assigned to a particular chromosome by examining a panel of hybrid cell lines that have retained different combinations of chromosomes from the species of interest and by testing for the presence of absence of the gene or a gene product. Isozyme analysis, Southern blotting and PCR are commonly used to make these assignments.

Hamster tumour cells are often used as 'acceptor' cells in fusions because the chromosomes are rarely lost in culture. Blood lymphocytes or cultured fibroblasts from the species of interest can be used as 'donor' cells. To select for hybrids, a mutation is introduced into the rodent parent that can be complemented by the normal chromosomes of the species of interest. The chromosome (or partial chromosome) that carries the functional gene will always be retained. Hamster cells with mutations in hypoxanthine phosphoribosyl transferase (HPRT) or thymidine kinase (TK) are often used (Abbott and Povey, 1995).

HPRT is utilised in an alternative pathway for the synthesis of nucleotides from exogenous hypoxanthine supplied to the cell. $HPRT^{-ve}$ mutants can be selected by adding 8-azaguanine or 6-thioguanine to the culture medium. Only those cells that do not incorporate these base analogues, that are competitive inhibitors of hypoxanthine, into their DNA can survive. TK enables cells to incorporate thymidine from the culture medium into nucleotides, avoiding the *de novo* pathway. TK^{-ve} mutants can be selected by adding the analogue 5-bromodeoxyuridine to the medium (Abbott and Povey, 1995).

Hybrid cells are selected by culture in a medium that allows them to out-grow the parental cells. In general, only the fastest growing parent is inhibited (ie. the hamster

cells). For example, ouabain, an ATPase inhibitor, is frequently added to the media at concentrations that do not affect the fused cells but selectively kills donor parent cells. HAT (hypoxanthine, aminopterin, thymidine) selection is also commonly used (Szybalski *et al.,* 1962; Littlefield, 1964). Aminopterin inhibits folic acid reductase and blocks *de novo* synthesis of nucleotides (Figure 1.2). Instead, cells must use the alternative pathway of nucleotide synthesis that requires incorporation of hypoxanthine and thymidine. Consequently, normal cells can grow but $HPRT^{-ve}$ or TK^{-ve} cells cannot. Hybrids selected in HAT medium will always retain the chromosomes equivalent to human chromosome 17 or X that carry the genes for TK or HPRT, respectively.

To generate panels of hybrids that retain different combinations of chromosomes for gene mapping, auxotrophic hamster mutants with different nutritional requirements, or temperature sensitive mutants have been used. Alternatively, dominant acting bacterial markers (eg. antimetabolite resistance) can be introduced by DNA transfection, electroporation or retroviral infection (Athwal *et al.,* 1985; Warburton *et al.,* 1990; Kurdi-Hairdar *et al.,* 1993). Incorporation of the bacterial marker is random, but the chromosome carrying it will be retained.

Sheep x rodent somatic cell hybrids (Cianfriglia, 1979; Saidi-Mahter *et al.,* 1981; Tucker *et al.,* 1981) have been used extensively to establish syntenic groups, but the chromosomes retained in these original hybrids were not characterised. Cytogenetic characterisation is important because the chromosome content of hybrids is not always stable. Discordancy between markers in uncharacterised hybrids might be due to further chromosome loss and could lead to incorrect physical assignments.

Recently, a cytogenetically characterised sheep x hamster somatic cell hybrid panel was completed (Burkin *et al.,* 1998). Lymophocytes from sheep carrying normal and Robertsonian translocation chromosomes were fused with different Chinese hamster auxotrophs to selectively "capture" specific sheep chromosomes. Particular cell hybrids were treated with BrdU or grown in non-selective medium to promote further sheep

chromosome loss (Burkin *et al.*, 1993; Broad *et al.*, 1995; Broad *et al.*, 1996; Burkin *et al.*, 1997a; Burkin *et al.*, 1998). A large number of hybrids were cytogenetically characterised to assemble the panel of 30 hybrids. More than half of the hybrids in the panel contain single sheep chromosomes. These monochromosomal hybrids are a resource for positional cloning and for the development of radiation hybrids.

More than 65% of the assignments (153/234) on the ovine physical map (ArkDB 11/9/98) have been made using SCH. The complete sheep x hamster panel was used to assign 9 genes and to confirm the chromosomal assignments of 110 microsatellite markers (Burkin *et al.*, 1998).

The advantage of mapping with SCH is that, provided the gene (or gene product) from the two species can be distinguished, physical assignments do not require polymorphism. However, physical localisations made with hybrids are often limited to a whole chromosome assignment and do not give the order of genes along the chromosome. A subsequent localisation by *in situ* hybridisation is often performed (Ryan *et al.*, 1993).

SCH will continue to be useful for mapping short cDNA probes that would be difficult to assign by *in situ* hybridisation. However, *in situ* hybridisation methods are more efficient than SCH for localising genes or markers to discrete chromosomal bands provided large-insert clones are available.

1.3.3.4.3 Radiation hybrids

Goss and Harris (1975) initially described the production of radiation hybrids (RH), but the technique was not widely used until the advent of PCR-based screening methods (Cox *et al.*, 1990). Analysis of RH is a powerful statistical tool to produce maps of gene order and relative distance, regardless of allelic variation (Schläpfer *et al.*, 1997).

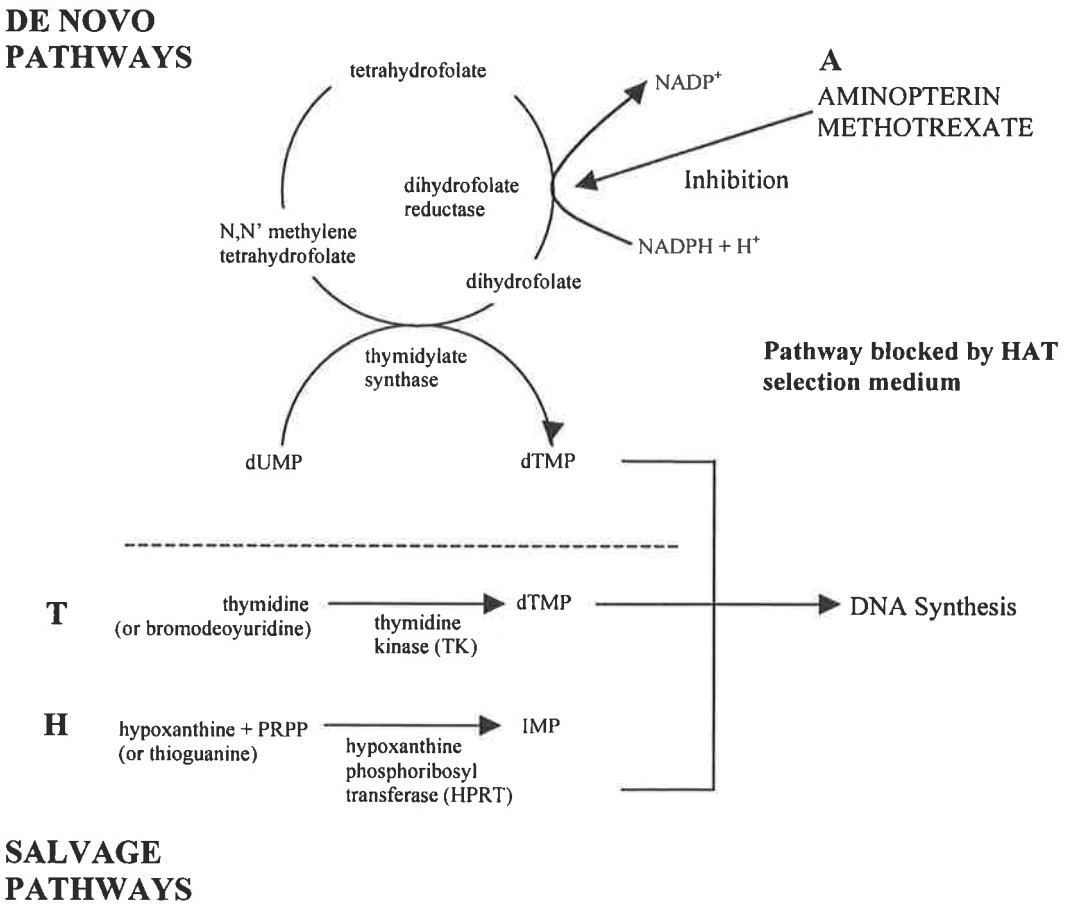


Figure 1.2: HAT (hypoxanthine, aminopterin, thymidine) selection for somatic cell hybrids. Aminopterin inhibits folic acid reductase and blocks *de novo* synthesis of nucleotides. To continue to divide, cells must use the alternative pathway of nucleotide synthesis that requires incorporation of hypoxanthine and thymidine. (From Abbott and Povey, 1995).

Radiation hybrids are made by subjecting a monochromosomal hybrid and a marker gene (eg. HPRT or TK) to lethal irradiation that fragments the chromosomes (Cox *et al.*, 1990). Alternatively, diploid cells can be irradiated as a source of donor DNA for 'whole-genome radiation hybrids' (Walter *et al.*, 1994). In either case, the fragments are fused to non-irradiated, HPRT or TK deficient hamster cells in HAT medium. The gene for the selective marker will always be retained, but the greater the distance between other markers, the more likely they will be contained in different fragments and consequently will segregate independently. Resolution of RH maps is dependent on the radiation dosage.

Human and mouse RH panels have enabled the construction of dense physical maps (Deloukas *et al.*, 1998; McCarthy *et al.*, 1997). These maps are a powerful resource for positional candidate cloning, comparative mapping and for establishing minimum-tiling paths of large-insert clones (eg. YACs and BACs) for large-scale sequencing (Mahairas *et al.*, 1999; Siegel *et al.*, 1999). Mapping 3' UTR regions of cDNAs on two RH panels produced the human RH map of 30,181 unique genes (Deloukas *et al.*, 1998). The GB4 3000-rad panel is suitable for long-range ordering of markers and consists of 93 hybrids each retaining ~32% of the human genome in random fragments of ~10Mb (Walter *et al.*, 1994). The G3 10000-rad panel is 83 hybrids that each retain ~15% of the human genome in ~4Mb fragments (Schuler *et al.*, 1996; Deloukas *et al.*, 1998). Using the G3 panel, it is possible to order markers that are ~1Mb apart with odds greater than 1000:1 and the average resolution of the G3 map is ~500kb. Recently, the TNG3 30,000-rad panel was generated by Research Genetics and can resolve the order of markers that are separated by just 100kb (Carey, 1997). Whilst the G3 panel (and probably the TNG3 panel) allows high-resolution mapping, the panel does not represent the complete genome and so it must be used in tandem with the G4 panel (McCarthy *et al.*, 1997).

The mouse RH panel was produced using a 3000-rad dose of radiation to create a panel with a high retention rate (27.6%) of donor DNA for optimal mapping. A preliminary map of 271 markers was produced to characterize this panel (McCarthy *et al.*, 1997).

Recently, a bovine RH panel and a porcine RH panel were completed (Womack *et al.*, 1997; Yerle *et al.*, 1998). The bovine 5000-rad panel consists of 101 hybrids and the retention rate is estimated to be about 30%, although it is not clear whether the bovine content of the hybrids has been cytogenetically tested. The 152 porcine hybrids produced by fusing lymphocytes or fibroblasts irradiated with 6000-7000-rad to hamster cells were cytogenetically characterized. After assessing the porcine content of each hybrid, 118 were selected to form a RH mapping panel. The retention rate within the mapping panel varied from 9-50% (average 29.3%; Hawken *et al.*, 1999).

The availability of these livestock RH panels will greatly accelerate the production of the physical maps for these economically important species. RH mapping enables type I and type II markers to be integrated into a single map (Hawken *et al.*, 1999) and allows comparative mapping information from 'map-rich' species to be efficiently utilised. For example, parellel RH mapping in humans and cattle for 24 orthologous genes was able to identify three chromosomal breakpoints and reveal internal structural rearrangements between HSA17 and BTA19 (Yang and Womack, 1998). Radiation hybrid maps for other bovine chromosomes are currently being developed (J.F. Taylor, Texas A&M University, pers. comm.). A RH map of porcine chromosome 15 demonstrated that the resolution was 18 times greater than the genetic linkage map (Yerle *et al.*, 1998). Nine hundred type I and type II markers have now been scored on the porcine RH panel to produce a first-generation RH map of 757 linked markers for pigs (Hawken *et al.*, 1999). To date, a sheep RH panel has not been produced, but with the completion of the SCH panel containing monochromosomal hybrids (Burkin *et al.*, 1998), it seems likely that one will soon be available.

1.3.3.4.4 *In situ* hybridisation

In the absence of radiation hybrids, *in situ* hybridisation methods are the best way to determine physical gene order and to localise markers to discrete chromosomal regions. *In situ* hybridisation (ISH) was first described by Pardue and Gall (1969) and is the molecular hybridisation of a radioactive DNA probe to chromosomal DNA. The 'sensitivity' of ISH is related to the length of the target sequence on the metaphase chromosomes and to the length of the probe able to hybridise to the target, and then to itself, so as to create concatenated matrices. ISH is most efficient with satellite DNA that is highly repeated but most difficult with single-copy sequences. The probe is labelled by incorporating radioactive nucleotides into the sequence by nick translation or by other DNA labelling methods. The radioactivity is detected by autoradiography and the sequence is localised to a specific chromosome region by examining the distribution of silver grains on metaphase spreads (Gosden and Lawson, 1994).

ISH has a number of disadvantages including low sensitivity, high background (ie. silver grains on other chromosomes or not associated with any chromosomes) and long exposure times for experimental results. It can take anywhere from days to months for sufficient autoradiographic signal to enable detection of the sequence. Also, only one radioactive probe can be detected at a time. The need for a more rapid and sensitive, non-isotopic physical mapping strategy was realised (Bauman *et al.*, 1980) and led to the development of fluorescence *in situ* hybridisation (FISH).

1.3.3.4.5 Fluorescence *in situ* hybridisation

Fluorescence *in situ* hybridisation (FISH) is the hybridisation of biotin-labelled or digoxygenin-labelled probes to metaphase chromosomes. Biotin is an endogenous component of several organs and is often present in active cells. Digoxygenin is as sensitive as biotin, but has no endogenous production (Warford and Lauder, 1991).

Fluorescein-conjugated avidin (for biotin) or anti-digoxygenin antibody is bound to the probe after hybridisation. Consequently, the site of probe hybridisation can be visualised by fluorescent microscopy.

The advantages of FISH compared with ISH are safety, high spatial resolution, sensitivity and the fact that different targets can be visualised simultaneously in the same sample by using probes with different coloured fluorescent labels (Shuxian *et al.*, 1993). Although physical mapping by SCH has been more commonly used in sheep than ISH, SCH assignments are generally limited to entire chromosomes. ISH and FISH have been used to localise 91 genes or markers to discrete bands on ovine metaphase chromosomes (ArkDB, 6/8/99). Continued efforts are needed to increase the number of genes (or comparative mapping markers) that have been localised to particular chromosomal bands because they enable mapping information to be extrapolated from 'map-rich' species. A dense physical map assembled either by ISH or RH mapping will greatly assist the search for economically important traits as it provides physical map locations for candidate genes.

Plasmid-derived sequences are generally insufficient for FISH mapping because the target is too small and the intensity of the fluorescent signal is dependent on probe length. Large-insert clones (eg. cosmids, BACs, PACs and YACs) are required for efficient physical mapping by FISH. Long hybridisation times (several hours) limit the number of markers that can be processed and the immunochemistry involved in the detection of biotin-labelled or digoxygenin labelled probes is laborious and expensive. Oligonucleotide primed *in situ* hybridisation (PRINS) is a recent advance in physical mapping which may overcome many of these problems.

1.3.3.4.6 Primed *in situ* hybridisation

PRINS is a rapid alternative method to FISH that is based on sequence-specific annealing of unlabelled oligonucleotides to chromosomes *in situ*. The oligonucleotide is a

primer for strand elongation catalysed by a DNA polymerase that uses target chromosomal DNA as the template (Koch *et al.*, 1989). The new strand can be detected indirectly by using biotin-labelled or digoxygenin-labelled nucleotides as a substrate for strand elongation. Alternatively, the new strand can be labelled directly with fluorescein-labelled nucleotides. Direct detection is faster, easier, produces less background signal and is as sensitive as the indirect method (Koch *et al.*, 1992; Terkelson *et al.*, 1993).

Since the priming oligonucleotide is not labelled, high concentrations can be used to allow fast hybridisation without the production of extensive background. Good signals can be obtained from repetitive sequences in less than an hour and unique sequences can be detected in less than three hours (Volpi and Baldini, 1993). Differences in signal strength are due only to differences in the number and organisation of target sequences because short oligonucleotide primers give rise to as much strand elongation as longer probes or primers (Hindkjær *et al.*, 1994).

Several probes can be detected simultaneously by repeating the PRINS reaction multiple times with different coloured nucleotides for each probe. The chromosomal DNA may only be completely denatured at the start of the first PRINS reaction, so the first probe used may give the strongest signal. Therefore, the probe with the smallest target should be used first to maximise the signal for that probe (Volpi and Baldini, 1993; Hindkjær *et al.*, 1994).

Cycles of denaturation, annealing and extension (as for PCR) can be used to amplify PRINS signal. Instead of a single strand being synthesised, multiple strands are produced by each subsequent cycle. This increases the amount of fluorescent label accumulating at the target site, and therefore, increases the visualisation of short targets. The new strands remain loosely associated with the target, and there is limited diffusion of product away from the chromosomes (Gosden and Hanratty, 1993).

The advantages of PRINS compared with FISH are the lower background, shorter analysis time and increased signal intensity (Shuxian *et al.*, 1993). PRINS should become the method of choice for physical mapping of anonymous markers and may enable plasmid-derived sequences to be physically mapped. The main disadvantage of PRINS is that the thermal kinetics of the reaction are not fully understood, so the procedure needs to be optimised for each new probe sequence used (Gosden and Lawson, 1994).

1.3.3.5 Large-insert libraries

A genomic library is a collection of clones that has a high probability of containing at least one copy of every sequence in the genome. The number of clones (N) required for a genomic library is a function of the desired probability of finding any given unique sequence within the library (P), the average insert size of the cloned DNA (I), and the genome size (GS) (Clarke and Carbon, 1976):

$$N = \frac{\ln(1-P)}{\ln(1-I/GS)}$$

$$P = 1 - (1-I/GS)^N$$

Large-insert DNA libraries are a resource for the construction of dense physical maps (Kotani *et al.*, 1997), for the development of minimum tiling paths for large-scale sequencing (Mahairas *et al.*, 1999; Siegel *et al.*, 1999) and for the identification and isolation of markers and genes (Davis *et al.*, 1999; Pirottin *et al.*, 1999). Vectors for cloning large fragments are just specialised plasmid or bacteriophage systems.

1.3.3.5.1 Cosmid libraries

Cosmids are modified plasmids (Figure 1.3a) with *cos* sequences for packaging DNA into bacteriophage lambda (Royal *et al.*, 1979). The cloning capacity of cosmids

is a function of the size of DNA that can be packaged into the head of lambda, and the vector size. Cosmid inserts generally range from 33-47kb, and cosmids were the first specialised cloning system that enabled the isolation of relatively large genes within a single insert. High-resolution physical mapping by FISH (eg. to determine gene order) became routine once cosmids with large inserts were available (Lichter *et al.,* 1990).

Cosmid libraries have been constructed for humans (Van Dilla and Deaven, 1990), mice (Chen *et al.,* 1982), rats (Hassett *et al.,* 1989), chickens (Buitkamp *et al.,* 1998), pigs (Haley *et al.,* 1987), cattle (*Clontech,* Palo Alto, CA) and sheep (Buchanan *et al.,* 1993). Clones isolated from livestock cosmid libraries have been used to investigate gene expression and regulation (Kojima *et al.,* 1997; Rijnkels *et al.,* 1998), for comparative mapping studies (Comincini *et al.,* 1997; Sonstegard *et al.,* 1998), and to integrate physical and genetic linkage maps (Toldo *et al.,* 1993; Ellegren *et al.,* 1994; Ferretti *et al.,* 1997; Robic *et al.,* 1997).

The frequency of repetitive sequences (eg. SINEs and microsatellites) in the sheep genome was estimated by analysing cosmid clones (Buchanan *et al.,* 1993). Porcine and bovine cosmid-derived $(CA)_n$ microsatellites were isolated and physically assigned by ISH or FISH to anchor and orientate linked markers on the physical maps (Ellegren *et al.,* 1994; Toldo *et al.,* 1993; Mezzelani *et al.,* 1995; Ferretti *et al.,* 1997). An iterative PCR-based screening method was utilised to efficiently isolate cosmids linked to specific chromosomal regions (Heaton *et al.,* 1997). Cosmids containing homologues of human genes can be isolated and physically assigned to identify regions of conserved synteny, establish boundaries of conservation and evaluate gene order within livestock maps (Sonstegard *et al.,* 1997).

Cosmids now have limited use for large-scale physical mapping projects because of a number of technical problems associated with them. Libraries constructed in cosmid vectors frequently contain clones that do not carry inserts of foreign DNA. Cosmids are often chimaeric, carrying two or more DNA fragments that are not

contiguous in the genome. Some cloned DNA fragments are unstable with loss or rearrangement of cloned segments due to recombination between repetitive elements (Poustka *et al.,* 1984). Chimaerism and instability are problematic when chromosome-walking experiments are undertaken (Yokobata *et al.,* 1991). Another problem is differential growth of clones leading to over- or under-representation of particular cloned sequences in a DNA library. For example, in *Clontech's* bovine cosmid library, a cloning artefact similar to 1.715 bovine satellite DNA is found in 60% of all the clones (Vaiman *et al.,* 1993).

Improvements have been made to cosmid vectors to increase insert stability and to enable transfection and selection in mammalian cells (Poustka *et al.,* 1984; Kioussis *et al.,* 1987; Ishiura *et al.,* 1989). However, the relatively small insert size (compared with YACs, BACs, and PACs) means that 350,000 clones are required to achieve a 99% probability that a particular single-copy sequence will be represented in the library. Cloning systems that accept larger inserts were developed to reduce the number of clones required for good genome coverage and contig assembly.

1.3.3.5.2 Yeast artificial chromosomes

Burke, Carle and Olson (1987) first constructed yeast artificial chromosomes (YAC) that are able to maintain inserts from 500-1000kb. YAC clones have been used to construct large-scale, 'sequence-ready' physical maps by walking along regions of mammalian chromosomes (Orti *et al.,* 1997; Wong *et al.,* 1997) and to examine the structure and function of large genes (Pearce *et al.,* 1993; McCormick *et al.,* 1995). YAC clones are constructed by isolating large genomic restriction fragments and ligating vector arms to their ends. All the sequences required for the constructs to function as chromosomes in yeast are on the vector arms.

The YAC vector (eg. pYAC4; Figure 1.3b) is propagated as a circular plasmid in *E. coli*. It contains a unique restriction endonuclease site (eg. *EcoRI*) in the SUP4 gene,

as well as ARSI (for initiation of DNA replication) and CEN4 (yeast centromere) elements required for stable single-copy propagation of the artificial chromosome. TRPI, HIS3 and URA3 are selectable markers. TEL sequences derived from *Tetrahymena* telomeres function as telomeres in yeast. To clone an insert, the vector is digested with *BamHI* (which cuts adjacent to the TEL sequences) and with *EcoRI*. YAC clones are introduced into the host yeast strain by spheroblast transformation and plated onto selective medium. Several YAC vectors are now available with improved selectable markers or genes allowing for selection of mammalian cells transfected with YAC clones (Riley *et al.,* 1992; Markie *et al.,* 1993; Tucker *et al.,* 1997).

Although their very large-insert size makes YAC clones a powerful resource for physical mapping, the YAC cloning system has a number of disadvantages. Yeast cell densities in an overnight culture are significantly lower than *E. coli* meaning the yield of DNA is low. Only 2-8% of the DNA recovered is YAC DNA and it is difficult to separate from yeast DNA. Manipulation of YAC DNA can be problematic if intact molecules are needed and YAC DNA may need to be prepared in agarose plugs to limit shearing (Moir and Smith, 1998).

About 1% of YAC clones are unstable giving rise to deletion derivatives, while 10% of YAC clones carry two independent YACs because of co-transformation. The most serious problem is that 40-60% of YAC clones are chimaeric. It is thought that most chimeras are caused by recombination within the yeast cell between repetitive elements on co-transformed YACs, rather than from co-ligation events (Green *et al.,* 1991; Larionov *et al.,*1994). Use of recombination-deficient yeast hosts has been shown to result in fewer chimaeric clones (Ling *et al.,* 1993; Haldi *et al.,* 1994) although the cloning efficiency is lower in these hosts. Bacterial artificial chromosomes (BACs) and P1-derived artificial chromosomes (PACs) are recent innovations that should avoid the problems associated with YACs.

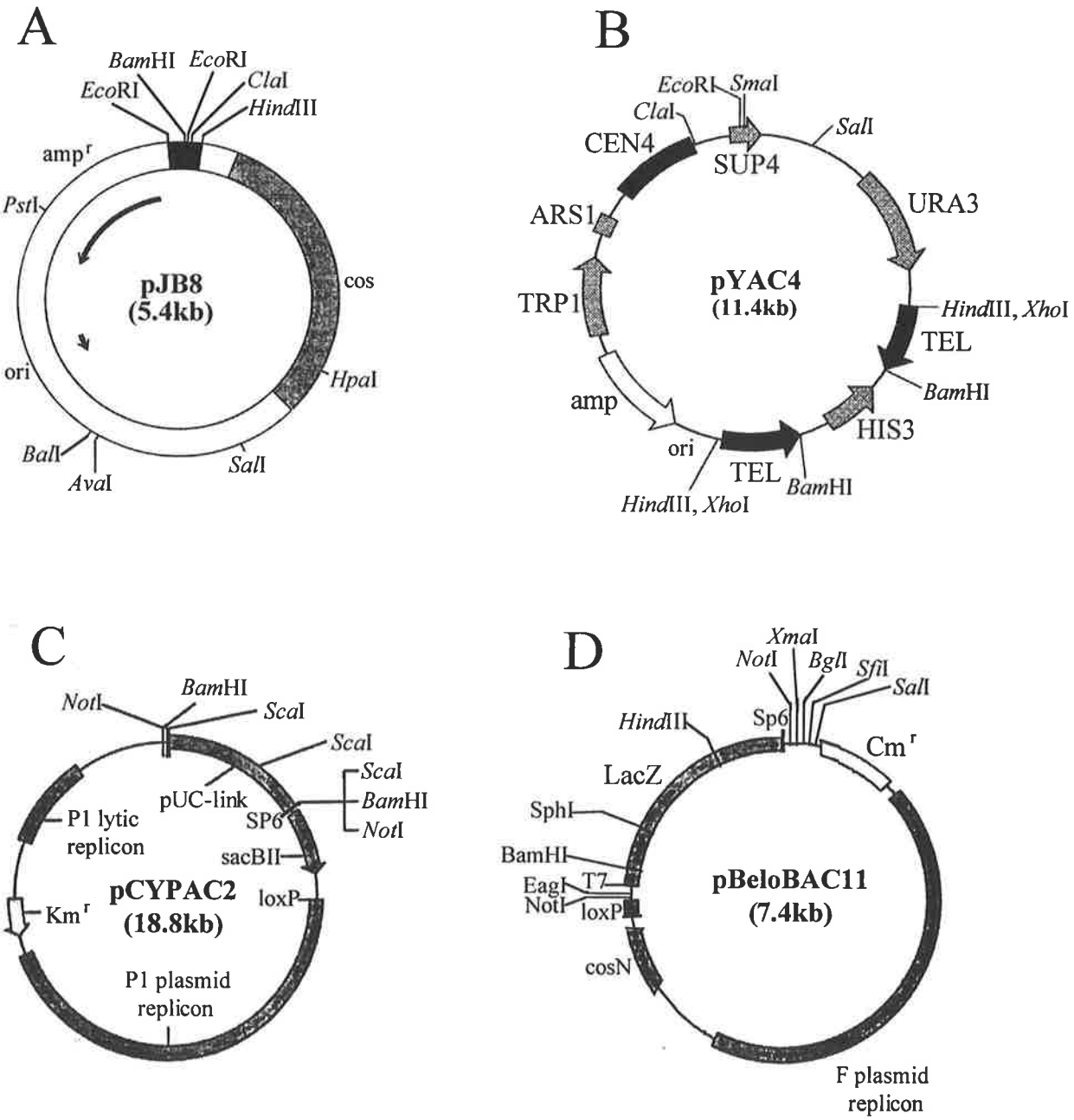


Figure 1.3: Examples of large-insert cloning vectors. A) Cosmid vector, pJB8. B) YAC vector, pYAC4. C) PAC vector, pCYPAC2. D) BAC vector, pBeloBAC11. (Adapted from Moir and Smith, 1998).

1.3.3.5.3 BAC and PAC libraries

Bacterial artificial chromosomes (BACs; Shizuya *et al.,* 1992) and P1-derived artificial chromosomes (PACs; Ioannou *et al.,* 1994) have a cloning capacity of 100-300kb and high cloning efficiency (10^4-10^6 transformants per μg DNA). BAC vectors (eg. pBeloBAC11; Figure 1.3c) are based on the *E. coli* F factor and PAC vectors (eg. pCYPAC2; Figure 1.3d) are based on the P1 phage replicon. pBeloBAC11 has all the essential genes from the F factor (*oriS*, *repE*, *parA* and *parB*) which maintain the vector at one or two copies per *E. coli* genome, as well as the chloramphenicol resistance gene and a polycloning site in the *lacZ* gene for blue-white colour selection of recombinant clones (Wang *et al.,* 1997). The pCYPAC vectors are derived from the P1 vector, pAd10SacBII, and have both a plasmid and P1 lytic replicon, as well as the kanamycin resistance gene and a polycloning site in the *sacB* gene for selection of recombinants. The *sacB* gene converts saccharose to levan, which is toxic to *E. coli*. A pUC19 linker in *sacB* silences the gene during vector preparation, and must be completely excised prior to ligation of insert into the *sacB* polycloning site, otherwise a high percentage of clones will be non-recombinants (Ioannou *et al.,* 1994).

Both PAC and BAC clones are stable with <3% of clones showing rearrangements after 100 generations (Shizuya *et al.,* 1992; Ioannou *et al.,* 1994; Woo *et al.,* 1994; Wang *et al.,* 1994; Cai *et al.,* 1995; Al-Bayati *et al.,* 1999). Maintenance and stability of large-inserts has been attributed to characteristics of the host, DH10β *E. coli*, which has a defective *recA* gene so homologous recombination is not possible (Wyman and Wertman, 1987; Hanahan *et al.,* 1991). By extrapolating results from FISH experiments, it has been estimated that BAC libraries contain <4% chimaeric clones (Shizuya *et al.,* 1992; Cai *et al.,* 1995; Schibler *et al.,* 1998a). Almost no chimerism has been detected in PAC libraries (Ashworth *et al.,* 1995; Hubert *et al.,* 1997; Matsumoto *et al.,* 1997).

BAC and PAC clones are now preferred for fine-scale physical mapping of candidate regions (Taylor *et al.,* 1998; Vaiman *et al.,* 1999a) and as templates for large-

scale sequencing (Cai *et al.*, 1998) because of the low frequency of rearrangement and the relative ease of DNA purification. BAC and PAC DNA may be separated from the endogenous *E. coli* chromosomal DNA by standard alkaline lysis procedures similar to those used for plasmid and cosmid preparations (Sinnett *et al.*, 1998). A BAC or PAC library covering the ovine genome would be a key resource for identifying economically important genes by positional cloning and for investigating other aspects of genome structure and function.

1.3.4 Investigation of genome structure and function

Once a comprehensive genome map for a species is developed, the next challenge is to understand *in vivo* gene function and the mechanisms that underly gene expression including cell specificity, developmental regulation and interactions between genes. From an animal production perspective, the goal is to identify the function of genes that confer economically important traits and to characterise the mutations that affect expression of these genes.

Transgenesis provides a means for investigating gene function and helps elucidate the role of genome structure in gene expression. Initially, transgenic experiments had unpredictable results, often with the transgene failing to be expressed in the target tissue. Short plasmid or bacteriophage inserts containing a gene and promoter region were mainly used and transgene expression was shown to be related to the integration site (Linton *et al.*, 1993; Rosen, *et al.*, 1996). By using large-insert clones for transgenesis, it has recently been demonstrated that position-independent expression is related to the occurrence of dominant *cis*-regulatory elements such as enhancers, insulators and locus-control regions (Peterson *et al.*, 1993; Peterson *et al.*, 1995; Lien *et al.*, 1997; Nielsen *et al.*, 1997; Porcu *et al.*, 1997; Nielsen *et al.*, 1999; Stinnakre *et al.*, 1999). These regions may control several genes with related fuctions within a multigene locus and can be located more than 50kb away (Nielsen *et al.*, 1997; Nielsen *et al.*, 1999).

Given that only ~10% of mammalian genomes are coding DNA, it will be interesting to see whether the remaining "junk" DNA, like microsatellites, short interspersed nuclear elements (SINEs) and long interspersed nuclear elements (LINEs), has any modulating or regulatory role in gene expression. These repetitive elements are capable of forming unusual DNA structures, such as Z-DNA (Vogt, 1990) and cruciforms (Jelinek and Schmid, 1982; Hanke *et al.,* 1995). They have been implicated in translational and rotational positioning of histone octamers influencing nucleosome formation over neighbouring regions (Englander *et al.,* 1995), in interactions with DNA-binding proteins and other cellular factors that are important for appropriate gene expression or cell proliferation (Sakamoto *et al.,* 1991; Cox *et al.,* 1998), and in rearrangements and deletions leading to altered structure and function of key genes (Lehrman *et al.,* 1987; Purandare and Patel, 1997).

Transgenesis not only provides the opportunity to study the fundamentals of genome structure and function *in vivo*, but enables directed changes or improvements to be made to economically important traits in livestock (Wall *et al.,* 1997; Moffat, 1998; Muller and Brem, 1998). Traditionally, improvements in wool quality and clean wool yield have arisen through selection of natural mutations that alter the expression of wool genes. Mutations may change the properties, abundance or location of proteins in the fibre, or affect the expression of whole gene families. Transgenesis is being used to make directed changes in wool gene expression that alter the properties of wool and enable the relationship between the structure of the wool fibre and its performance as a textile fibre to be investigated (Powell *et al.,* 1994; Bawden *et al.,* 1998). The availability of a library of stable, large-insert clones will be valuable for sheep transgenesis experiments as a source of transgenes and their long-range regulatory elements.

1.4 Research Plan

The aim of this project was to develop a set of tools to improve the ovine genome map so that the regions of the genome containing genes for economically important traits in sheep could be identified.

The US Department of Agriculture distributes ovine microsatellite primers for genetic linkage analysis. The primers are designed in the unique DNA sequence that flanks ovine microsatellites. If these primers (or PCR products generated by them) could be directly localised on metaphase chromosomes, then a fully integrated genomic map could be rapidly developed. Novel methods for physically localising these oligonucleotide primers on metaphase chromosomes were investigated in an attempt to improve the ovine physical map. When it became apparent that a modified PRINS procedure was not suitable for the localisation of single-copy sequences, sources of large-insert probes for FISH were examined.

To generate a source of stable, non-chimaeric, large-insert clones for FISH, an ovine BAC library was constructed and characterised. Two BAC clones for every chromosome were isolated from the library to directly anchor and orient the ovine genetic maps. These anchored clones were also used to confirm chromosome identification in dual localisation experiments with previously unmapped genes or markers. Ovine BAC clones were simultaneously mapped to sheep, cattle and goat metaphase chromosomes to improve the comparative map between these three livestock species.

To demonstrate the utility of BAC clones for investigating genome structure and function, a BAC clone containing a keratin associated protein gene was isolated. A restriction map of the BAC for KAP1.1 was developed. The location of repetitive elements (microsatellites, SINEs) and other KAP genes within the BAC was investigated by Southern blot hybridisation.

The ovine BAC library is a powerful resource for physical mapping and for the identification and isolation of economically important genes. In the future, it will continue to be a source of probes for physical and comparative mapping. Efforts to expand and order the ovine BAC library would benefit the search for economically important genes.

CHAPTER 2

MATERIALS AND METHODS

MATERIALS AND METHODS

The recipes for all solutions are listed in Appendix I.

2.1 Tissue culture

Peripheral blood and fibroblast cell lines derived from skin were the two sources of cultured cells used to prepare metaphase chromosome spreads for physical mapping by *in situ* hybridisation. All tissue culture was performed in a laminar flow hood using aseptic techniques.

2.1.1 White blood cell lymphocyte culture

White blood cell lymphocytes were used as a source of metaphase chromosomes for super-fast primed *in situ* targeting experiments (section 2.2). Sheep and cattle blood were collected in lithium-heparin vacutainers (*Becton Dickinson*) from the jugular vein using an 18 gauge needle (*Terumo*), mixed thoroughly to prevent clotting and stored at 4°C. The culture conditions were optimised to obtain a high mitotic index and elongated ("stringy") metaphase chromosomes. Various methods for enriching lymphocytes (sections 2.1.1.1-2.1.1.3) were investigated and the media and supplements were varied to try to increase the mitotic index (section 2.1.1.4). The amount of colchicine used to break the spindle fibres was titrated since it affects chromosome contraction (section 2.1.1.5) and stringy chromosomes are desired for gene mapping. A high mitotic index and chromosome banding were produced by synchronising the cells at mid-S phase with 5-bromodeoxyuridine (section 2.1.1.6).

2.1.1.1 Lymphocyte enrichment using Ficoll-Hypaque

A 5ml aliquot of heparinised blood was mixed with 5ml Hank's Balanced Sodium Salts (HBSS, *Life Technologies*) and the diluted mixture was carefully overlayed on 5ml Ficoll-Hypaque (*ICN*) in a 15ml Falcon tube. The tubes were centrifuged at 600g for 40' at 20°C in the swing bucket rotor of a Sorvall RT 6000D. The portion of the

gradient containing the opaque band of white blood cell lymphocytes was carefully removed and washed in 9ml HBSS and then centrifuged at 200g for 20'. The pellet was resuspended in 10ml culture medium (section 2.1.1.4).

2.1.1.2 Lymphocyte enrichment using 0.85% ammonium chloride

A 5ml aliquot of heparinised blood was mixed with 45ml 0.85% ammonium chloride and incubated at 4°C for 2h. The mixture was centrifuged at 200g for 10'. The pellet was washed in 50ml 0.85% ammonium chloride and centrifuged again at 200g for 10'. The wash and centrifugation were repeated, and the pellet was resuspended in 10ml culture medium (section 2.1.1.4).

2.1.1.3 Lymphocyte enrichment by centrifugation

Whole blood was centrifuged at 350g for 10' and then 10ml culture medium (section 2.1.1.4) was inoculated with 0.5ml cells taken from the buffy coat at the plasma-blood interface.

2.1.1.4 Evaluation of media

A comparison was made between Ham's F10 medium (*CSL*) and RPMI 1640 (*CSL*) for sheep and cattle white blood cell lymphocyte culture (Lin *et al.*, 1976; Lin *et al.*, 1977). The combinations of lymphocyte treatment, medium and supplements that were used are listed in Table 2.1. In addition to foetal calf serum (*CSL*) and glutamine (*CSL*), each culture was supplemented with 1.8% phytohaemagglutinin-M (PHA-M, *Life Technologies*) which is a mitotic stimulant and 1% penicillin-streptomycin-fungizone solution (*CSL*). PHA-M primarily affects T-cells and although pokeweed mitogen can be used to stimulate B cells, its inclusion in the culture medium was not evaluated in these experiments.

Table 2.1: Treatments and media used to culture white blood cell lymphocytes from sheep and cattle.

Lymphocyte treatment	Medium	Supplements
Ficoll enriched	Ham's F10	20% FCS
Ficoll enriched	RPMI 1640	20% FCS
NH_4Cl enriched	Ham's F10	20% FCS
NH_4Cl enriched	RPMI 1640	20% FCS
Enriched by centrifugation	RPMI 1640	20% FCS
Enriched by centrifugation	RPMI 1640	15% FCS + 1% glutamine
Whole blood	Ham's F10	15% FCS + 1% glutamine
Whole blood	RPMI 1640	15% FCS + 1% glutamine
Whole blood	RPMI 1640	20% FCS

The culture tubes were incubated with loose lids on their sides at 37°C/5% CO_2 for 68.5h and then 100ng/ml colchicine (*Sigma*) was added to each culture. The cultures were incubated for a further 1.5h, and the cells pelleted by centrifugation at 200g for 10'. The medium was removed by aspiration and the cells resuspended in 10ml 0.075M hypotonic KCl pre-warmed to 37°C. The tubes were incubated at 37°C for 24' and then centrifuged at 200g for 10'. The pellet was resuspended by adding 10ml chilled 3:1 methanol: acetic acid fixative drop-wise. The cells were incubated at 4°C for 30' and pelleted at 200g for 10'. The fixative was replaced and the cells incubated again for 30' at 4°C. The centrifugation and incubation were repeated once more, and finally the cells were resuspended in 1.5ml fresh fixative.

The cells were dropped onto slides using a pasteur pipette from a height of ~20cm with three drops per slide. The slides were air-dried and stained with 5µg/ml propidium iodide (*Sigma*) for 5'. The slides were rinsed 3 x 20" in PBS. The mitotic index was

determined by counting 500 cells using an Olympus BH-2 phase microscope at 10x magnification.

2.1.1.5 Colchicine titration curve

Colchicine is a mitotic inhibitor that prevents microtubule formation (Lawce and Brown, 1991). To evaluate the effect of colchicine (*Sigma*) on chromosome contraction, 6 x 10ml cultures were either inoculated with 0.5ml whole blood or lymphocytes that were enriched using Ficoll-hypaque (section 2.1.1.1). The white blood cells were cultured in RPMI 1640 supplemented with 20% FCS, 1.8% PHA and 1% penicillin-streptomycin-fungizone solution. The cultures were incubated at 37°C/5% CO_2 for 68.5h and then colchicine was added so that the final concentration was 50ng/ml, 100ng/ml, 200ng/ml, 300ng/ml, 400ng/ml and 500ng/ml, respectively. The cultures were incubated at 37°C/5% CO_2 and then harvested as described in section 2.1.1.4.

RPMI 1640 supplemented with 15% FCS, 1% glutamine, 1.8% PHA and 1% penicillin-streptomycin-fungizone solution was also inoculated with either whole blood or enriched lymphocytes and then cultured at 37°C/5% CO_2 for 68h (P. Pearce, AgResearch Grasslands Research Centre, New Zealand, pers. comm.). Colchicine was either added 1.5h before the cells were harvested (as in Lin *et al.*, 1976) or it was added at a final concentration of 600ng/ml after 12' in 0.075M hypotonic KCl (following the protocol of Pearce, pers. comm.). Other time intervals were not attempted.

2.1.1.6 Synchronisation with thymidine and elongation with 5-BrdU

Cells can be synchronised using excess thymidine to inhibit DNA synthesis by negative feedback regulation. The cells accumulate at mid-S phase of the cell cycle until the thymidine is removed from the medium. Normal chromosome condensation during metaphase can be prevented by the addition of 5-bromodeoxyuridine (5-BrdU) to

the culture medium after the thymidine arrest. 5-BrdU binds preferentially to GC-rich DNA and causes differential elongation of the chromosomes (Lawce and Brown, 1991). The sites of 5-BrdU incorporation can be visualised as chromosome bands because they have decreased staining with Hoechst 33258, Giemsa and propidium iodide (Bickmore and Craig, 1997).

To produce R-bands and to determine whether synchronisation with thymidine and 5-BrdU increased the mitotic index or the quality of the metaphase spreads, 10ml cultures of RPMI 1640 supplemented with 15% FCS, 1% glutamine, 1.8% PHA and 1% penicillin-streptomycin-fungizone solution were inoculated with 0.5ml whole blood.

In general, the culture tubes were incubated with loose lids on their sides at 37°C/5% CO_2 for 70h and then arrested for 16h with 300μg/ml thymidine. The cells were pelleted by centrifugation at 400g for 10' and rinsed with new medium. After being pelleted again, the cells were resuspended in medium containing 20μg/ml 5-BrdU and cultured for 6 hours prior to harvesting (Schollmeyer *et al.*, 1981; Buckle and Craig, 1986).

To determine the efficacy of 5-BrdU alone for chromosome synchrony and elongation, the culture tubes were incubated with loose lids on their sides at 37°C/5% CO_2 for 62h before 200μg/ml 5-BrdU (*Sigma*) was added. The blood was cultured for a further six hours prior to harvesting (Dutrillaux and Viegas-Pequinot, 1981; Eichenbaum and Krumins, 1983).

The cells were pelleted by centrifugation at 400g for 10', resuspended in 8ml 0.075M hypotonic KCl, and incubated at 37°C for 24'. After 12', 600ng/ml colchicine was added to the tubes and mixed thoroughly. When the incubation was complete, 2ml 3:1 methanol:acetic acid fixative was added to the tubes and they were centrifuged at 400g for 10'. The cells were washed three more times in 5ml fixative by centrifugation at 400g for 5' before test slides were made from cells resuspended in 1ml fixative.

2.1.2 Fibroblast cell lines

Although the conditions for white blood lymphocyte culture were optimised to maximise the mitotic index (15-20%), only a small proportion of the metaphase spreads were satisfactory for gene mapping. Consequently, sheep, cattle and goat fibroblast cell lines were established to prepare metaphase chromosomes for comparative mapping studies by fluorescence *in situ* hybridisation. The major advantage of fibroblasts over lymphocytes is that they grow as a monolayer, and the metaphase cells can be preferentially recovered because they do not adhere to the growing surface as strongly as the interphase cells.

Sheep cell lines were initiated from ear punches (section 2.1.2.2) taken from three Rambouillet rams (#601, #606, #608) at Texas A&M University with the assistance of Dr. Dan Gallagher who developed the culture and harvesting methods described in sections 2.1.2.2-2.1.2.6. A fourth sheep cell line was initiated at the University of Adelaide from a Tukidale ram. Cattle cell lines were prepared from a Limousin steer (#771) and from a Simmental cow that has a 14:20 Robertsonian translocation. A fibroblast cell line was also established for an Angora goat.

2.1.2.1 Fibroblast growth medium

Fibroblast cells were cultured in Dulbecco's modified Eagle's medium (DMEM, *CSL*) with 10% v/v foetal calf serum (*CSL*) and 1% v/v penicillin-streptomycin-fungizone (*CSL*). Immediately before use, 1% v/v 200mM glutamine (*CSL*) was added.

2.1.2.2 Initiation of fibroblast cell lines

Before taking an ear punch, the area was swabbed with ethanol and shaved. The ear punch was transferred to a 15ml Falcon tube of growth medium and stored at 4°C. To prepare the ear punch for culture, it was sterilised by a brief wash in 100% ethanol in a petri dish. Any remaining hair was removed with sterile forceps before one epidermis

was cut away from the cartilage. The remaining epidermis and cartilage were stored in growth medium at 4°C in case the cells failed to grow or there was a massive infection of the culture.

The epidermis was rinsed successively in six drops of growth medium and in the final drop it was cut into tiny pieces (~1mm) with a sterile scalpel. The pieces were transferred with a pasteur pipette to two 25cm^2 culture flasks (*Falcon*) containing 1ml growth medium. To help keep each piece on the bottom, a small scratch was made with the pasteur pipette. The pieces were arranged in a grid-like pattern with ~20 pieces/flask. The medium was replaced and was sufficient to completely cover the bottom of the flask without causing the pieces to float.

The flasks were incubated with loose lids at 37°C/5% CO_2. The skin pieces were examined every day beneath an Olympus CK2 dissecting microscope and any contaminated pieces were removed from the flask. The medium was also replaced every day until fibroblasts were established (up to 2 weeks), and then the volume in the flask was increased to 5ml and the cells were grown to confluence.

2.1.2.3 Passaging Fibroblasts

The medium was removed and the cells were washed twice with 1-5ml HBSS (*Life Technologies*) to remove any trace serum that would inhibit trypsin. To recover cells from a 25cm^2 flask, 0.5ml Multicell trypsin/EDTA (*Surgical and Medical*) was used whilst 2.5ml was used for both 75cm^2 (*Falcon*) and 175cm^2 flasks (*Nunc*). The flasks were incubated at 37°C for 5' and then the cells were dislodged by vigorously tapping the flask against one hand. The cells were viewed under the Olympus CK2 dissecting microscope to confirm that they were in solution before 4.5ml growth medium was added to the flask to stop the trypsinisation. The solution was transferred to a 15ml Falcon tube and another 4.5ml growth medium was used to rinse the growing

surface of the flask to collect any remaining cells. The cells were pelleted by centrifugation at 250g for 5' in the swing bucket rotor of a Sorvall RT 6000D.

The supernatant was removed leaving ~200µl above the pelleted cells. The cells were resuspended by flick mixing and the volume was adjusted to 1-2ml with growth medium. The cells were then transferred to new flasks and incubated at 37°C/5% CO_2 until confluent. The cells were passaged several times into 5ml growth medium in 25cm^2 flasks to remove any contaminating epithelial cells. When the cell line was ready for a chromosome harvest (section 2.1.2.6) or to be frozen for long-term storage in liquid nitrogen (section 2.1.2.4), the cells from two 25cm^2 flasks were passaged into two 175cm^2 flasks containing 30ml growth medium and incubated at 37°C until they were >70% confluent.

2.1.2.4 Freezing fibroblasts for long-term storage in liquid nitrogen

The cells were trypsinised as described in section 2.1.2.3. After pelleting the cells by centrifugation at 250g for 5', the medium was removed and the cells were resuspended in 10ml growth medium containing 10% sterile dimethyl sulphoxide (DMSO, *Sigma*). The cells were mixed well to ensure a homogeneous solution and then split into 10 x 1ml aliquots in Nunc tubes. The cells were frozen slowly by wrapping them in insulating material and chilling them at 4°C, before freezing them at -80°C and then transferring them to liquid nitrogen. One tube was recovered from liquid nitrogen (section 2.1.2.5) within days of freezing to check the viability of the frozen cells.

2.1.2.5 Recovering cells from liquid nitrogen

The cells were removed from liquid nitrogen and thawed quickly in warm water (25-37°C). The vial was dipped in 70% ethanol and 500µl cells were added to two 25cm^2 flasks containing 4.5ml growth medium with fresh glutamine. The flasks were incubated with loose lids at 37°C/5% CO_2 overnight. The medium was then removed

and the flasks were rinsed with 2x5ml HBSS to remove any DMSO carried over from the freezing solution (section 2.1.2.4), which can inhibit cell growth. The medium was replaced and the flasks were incubated at 37°C/5% CO_2 until the cells were confluent, at which time they were passaged as in section 2.1.2.3.

2.1.2.6 Harvesting fibroblasts for G-banded metaphase chromosomes

The cells from 2 x 175cm^2 flasks were passaged (2.1.2.3) into 4 x 75cm^2 flasks containing 10ml growth medium with glutamine and 45µg/ml 5-BrdU which arrests the cells at mid-S phase of the cell cycle. The flasks were incubated with loose lids at 37°C/5% CO_2 for 24 hours. The medium was removed and the cells rinsed twice with 5ml HBSS. The medium and glutamine were replaced and 10^{-5}M thymidine (*Sigma*) added to release the cells from arrest at mid-S phase. The flasks were incubated at 37°C/5% CO_2 for 6 hours to enable the cells to progress to metaphase.

Colchicine was used to break the spindle fibres, but the amount of colchicine and length of incubation affected chromosome contraction. Stringy chromosomes were desired for FISH, and it was found that 0.5µg/ml colchicine and a 15' incubation at 37°C/5% CO_2 produced ideal metaphase spreads.

The medium was removed and 2ml 0.075M hypotonic KCl pre-heated to 37°C was added. As soon as the KCl was added to the flasks, timing of the next 25' period was commenced. Each flask was vortexed vigorously for 1' making small wavelets all over the growing surface of the flask to dislodge the metaphase cells. The cells from every tube were transferred to a single 15ml Falcon tube. Another 2ml 0.075M KCl was added to the flasks and the vortexing step was repeated. The cells were transferred to the same Falcon tube as before.

The cells were centrifuged at 250g in the swing bucket rotor for 5'. The medium was aspirated leaving ~200µl above the pellet. The cells were flick mixed, then 12ml 0.075M KCl was added and the tube was incubated in a 37°C waterbath for the

remainder of the 25'. Fresh fixative (3:1 methanol:acetic acid) was added to fill the tube (~3ml), mixed well and the cells pelleted again by centrifugation for 5' at 400g. The medium was removed again and the cells were flick mixed prior to adding 12ml fixative. After mixing well, the tube was centrifuged at 400g for 5'. This step was repeated twice before the cells were resuspended in an appropriate volume to make slides. Generally, the cells were resuspended in 0.5ml that was sufficient for 15-20 slides of well-spread metaphase chromosomes.

Glass microscope slides (*Marienfeld*) were cleaned by soaking them in 5% Decon90 for 2 hours followed by a rinse in running water for ~1h. The slides were rinsed briefly in deionised water followed by 3 washes in 100% ethanol and then air-dried.

The cells were dropped onto the slides using a pasteur pipette from a height of ~20cm with one drop at each end of the slide. The slides were air-dried and the best chromosome morphology was obtained when the drying time was 45-60 seconds. The slides were artificially aged at 37°C for 2 days and then stored with desiccant at -20°C.

2.2 Super-fast primed *in situ* targeting

Super-fast primed *in situ* targeting (SPRINT) is essentially a polymerase chain reaction (PCR) on a slide and was developed in an attempt to find a cheaper and more rapid way of physically localising single-copy genes to metaphase chromosomes than by existing techniques (eg. fluorescence *in situ* hybridisation). The SPRINT reaction conditions were optimised for human satellite III (section 2.2.2) by altering the primer and $MgCl_2$ concentrations, dNTP ratios, *Taq* DNA polymerase and annealing temperature (Garbett, 1994). Digestion of clones as a source of SPRINT primers was tested for sheep satellite I (section 2.2.3) and various single-copy genes.

2.2.1 Slide Preparation

Slides of G-banded human metaphase spreads (provided by Dr. G. Webb) and R-banded sheep and cattle chromosomes prepared from white blood lymphocytes (section 2.1.1.6) were denatured in 70% formamide/2xSSC for 2' at 70°C. The slides were plunged into a coplin jar of 70% ethanol at -10°C to restrict snap-back reannealing of the chromosomal DNA. They were then dehydrated through an ethanol series (70%, 95%, 100%) and air-dried. In addition to fresh preparations, old slides of several human males and females that had been sealed with a silica gel drying agent at -20°C and carefully stored for up to 7 years were also used.

2.2.2 Optimised SPRINT reaction for Human Satellite III

Each 25µl reaction mixture contained 67mM Tris-HCl (pH 8.8), 16.6mM $(NH4)_2SO_4$, 0.2mg/ml gelatine, 0.45% Triton X-100, 50µM dATP, 50µM dGTP, 50µM dCTP, 50µM dTTP, 5µM fluorescein-12-dUTP (*Boehringer Mannheim*), 4.5mM $MgCl_2$, 2.5U Taq DNA polymerase (*Bresatec Ltd.*), and 2.5µg $(AAGGT)_5$ (synthesised by *Bresatec Ltd.*).

The reaction mixture was sealed under a 20x25mm parafilm coverslip with rubber cement before the slide was placed on a satellite slide attachment to the Omnigene temperature cycling system (*Hybaid*). The slide was incubated at 55°C for 20' and then the SPRINT reaction was stopped by immersing the slide in 50mM NaCl, 50mM EDTA (pH 8.0) at 60°C for 2'.

2.2.3 Generation of sheep satellite I primers by restriction enzyme digestion

The 560bp insert of a pGEM-7Zf(+) clone of sheep satellite I (courtesy of Ms. Stacey Rizza) was amplified by PCR using 10ng/µl universal forward sequencing primer and 10ng/µl reverse sequencing primer in a 50µl reaction consisting of 50mM KCl, 100mM Tris-HCl (pH 8.3), 125µM dGTP, 125µM dATP, 125µM dTTP, 125µM

dCTP and 0.5U AmpliTaq DNA polymerase (*Perkin Elmer*). The plasmid insert was amplified over 32 cycles (94°C 1', 45°C 1', 72°C 1') using a Perkin Elmer thermal cycler.

The PCR product was purified directly through a Wizard™ PCR prep column (*Promega*) according to the manufacturer's instructions. A 10µg aliquot of the ~790bp amplified fragment was precipitated with 1/20 volume 4M sodium chloride and 0.6 volumes isopropanol. The DNA was digested with 10U *Sau3AI* (*Promega*) in a 10µl reaction (6mM Tris-HCl (pH 7.5), 6mM $MgCl_2$, 50mM NaCl, 1mM DTT) for 2h at 37°C to produce fragments from 50-150bp. This digestion mixture was used as the source of primers for SPRINT.

The SPRINT reaction was identical to that used to amplify human satellite III except that the slide was heated to 90°C for 5' to denature the Sau3AI fragments prior to the incubation at 55°C.

2.2.4 Detection and microscopy

The slides were washed 2 x 2' in 2xSSC and 2 x 2' in PBS before the chromosomes were counterstained by immersing the slides in 5µg/ml propidium iodide for 5'. The slides were rinsed 3 x 20" in PBS and then 100µl ρ-phenylenediamine dihydrochloride, pH 11 (PPD11) antifade mountant was added. The mountant was sealed under a 24 x 60mm glass coverslip with nail varnish. The slides were viewed using an Olympus Vanox microscope with blue epifluorescence used to detect the fluorescent SPRINT signal, combined with dull fluorescence of the G-banded or R-banded chromosomes. Photographs were taken on Fugichrome ASA 400 colour positive film that was programmed to ASA 2560 (1.6 x ASA 1600) for dark field.

2.3 BAC library construction

Bacterial artificial chromosomes (BAC) are a cloning system based on the *E. coli* F plasmid that can stably maintain inserts from 100 - 500kb (Shizuya *et al.*, 1992). BAC clones are easily mapped by fluorescence *in situ* hybridisation (FISH) and consequently, they are a useful tool for efficiently generating accurate, high-resolution physical maps (Cai *et al.*, 1995). A sheep BAC library was constructed from high molecular weight DNA encapsulated in agarose microbeads. To achieve at 95% probability of finding any unique sequence within the BAC library a total of 59,913 clones with an average insert size of 150kb are required (section 1.3.3.5).

2.3.1 Preparation of high molecular weight DNA

High molecular weight DNA was obtained for BAC library construction by encapsulating either white blood cells or sperm heads in agarose microbeads prior to enzymatic digestion. The microbeads protect the DNA against shearing so that megabase lengths can be recovered.

2.3.1.1 Isolation of white blood cells

A 500ml blood sample mixed with 25ml anti-coagulation solution was collected from a purebred Suffolk ram (#30, Texas A&M University) and 300ml was used immediately. The blood was transferred to 6 x 50ml Falcon tubes and centrifuged at 1500g for 15' in the swing bucket rotor of a Sorvall RT 6000D. The buffy coat was collected from each tube and mixed with 5ml Hank's balanced sodium salts (HBSS (pH 7.0), *Life Technologies*). The volume was adjusted to 30ml and then 10ml aliquots were carefully overlayed onto 5ml Histopaque (*Sigma*) in 15ml Falcon tubes. The tubes were centrifuged for 40' at 350g with no brake. The white blood cell layer was transferred to a new tube, washed with 10ml HBSS and centrifuged for 10' at 150g with brake = 1. The pellet was resuspended in 5ml HBSS and the white blood cell concentration was

determined using a haemocytometer. Cell quality was monitored by staining with trypan blue solution (ie. dead cells were stained blue).

2.3.1.2 Preparation of Microbeads

Before preparing the microbeads, 20ml mineral oil and 5ml 1% low melting point agarose (*FMC. Seaplaque*) in HBSS were pre-warmed to 45°C. A 400ml beaker containing 120ml HBSS, that was continuously stirred, was chilled by packing the beaker in ice. The cells in HBSS were also pre-warmed and then mixed with the agarose solution. Then the mineral oil was added quickly and mixed very vigorously by swirling for ~5 seconds. This mixture was poured into the cold HBSS and stirred for 5'.

The microbeads and oil were transferred to 50ml Falcon tubes and centrifuged at 500g for 10'. The supernatant was removed and the microbeads from each tube were resuspended in 10ml HBSS, then pooled in 2 x 50ml Falcon tubes and centrifuged again at 500g for 10'. The supernatant was removed and then SDE solution was added to a final volume of 50ml. The tubes were rocked occasionally for 10' before centrifugation at 500g for 10'. The microbeads were resuspended in ESP solution containing 1mg/ml proteinase K (*Boehringer Mannheim*) and rocked at 50°C overnight. The tubes were cooled to room temperature and then centrifuged at 500g for 10'. The incubation with ESP solution containing 1mg/ml proteinase K was repeated for a further ~10 hours at 50°C.

The microbeads were pelleted at 500g for 10' and resuspended in 25ml TE (pH 8.0) and split into two 12.5ml aliquots. One aliquot was pelleted and resuspended in 0.5M EDTA for long-term storage at 4°C. The other aliquot was treated with TE containing 1mM phenylmethyl sulfonyl fluoride (PMSF, *Sigma*) to destroy the proteinase K. The tube was left on ice for 1h and inverted infrequently before being centrifuged at 500g for 10' at 4°C. The microbeads were again resuspended in TE containing 1mM PMSF and left on ice for another hour. The tubes were centrifuged

again and the microbeads were resuspended in TE containing 0.1mM PMSF and incubated at 4°C overnight. After pelleting the microbeads at 500g, they were washed three times with 13ml TE for 1 hour on ice to remove the PMSF. The microbeads were pelleted at 500g for 10' at 4°C and all the TE was carefully removed.

2.3.1.3 Isolation of Sperm Heads

To evaluate the use of semen rather than white blood cells as the source of DNA for BAC library construction, 5 straws of semen were obtained from the purebred Angus bull, Y6, that was the animal used to construct the bovine BAC library (Cai *et al.*, 1995). The bovine library was being expanded while the ovine library was being constructed.

The ends of the straws were cut and the semen was collected by capillary action into eppendorf tubes. The semen was mixed, then split into 250µl aliquots and 1ml solution A was added. The semen was pelleted by centrifugation at 3000rpm for 5', then resuspended in solution A and centrifuged again. These washes were repeated 3 times before each pellet was resuspended in 1.5ml solution B and incubated at 50°C for 30'. The sperm heads were washed 3 times with HBSS (pH 7.0) by centrifugation at 5000rpm for 5' and then pooled into 5ml. The cell number was determined using a haemocytometer and then the sperm cells were encapsulated in agarose microbeads as described (section 2.3.1.2).

2.3.1.4 Confirmation of microbead quality

The quality of the high molecular weight DNA encapsulated in the microbeads was checked on a contour-clamped homogeneous electric field (CHEF) gel prior to partial digestion with HindIII (section 2.3.2). Conditions were selected that prevented high molecular weight DNA (>1 Mb) from progressing into the gel, so that the extent of DNA shearing could be determined. Enough microbeads were loaded with a cut tip

(~2mm diameter) to fill a lane on a 1% agarose gel in 0.5 x TBE. After the microbeads were loaded, the lane was capped with 1% low melting point agarose. The DNA was electrophoresed in 0.5 x TBE using a CHEF-Mapper (*Biorad*) at 6V/cm for 18h at 14°C with a fixed angle of 120°. Both the initial and final switch times were 40 seconds. The gel was stained with 0.5mg/ml ethidium bromide for 10' and destained for 30'.

2.3.2 Partial digest of genomic DNA

To determine the amount of *HindIII* required to obtain DNA fragments with an average size of 150kb, partial digests were prepared using 0U, 0.5U, 2.5U, 5U, 10U, 20U and 40U *HindIII* (*Promega*). Each 100µl reaction contained 50µl microbeads, 6mM Tris-HCl (pH 7.5), 6mM $MgCl_2$, 100mM NaCl, 1mM DTT, 4mM spermidine, 1mg/ml BSA and an appropriate amount of *HindIII*. The reactions were prepared in duplicate and incubated on ice for 15' and then at 37°C for 15'. The digests were loaded onto a 1% agarose CHEF gel in 0.5 x TBE. The total volume of both duplicates was loaded in a single lane and a tissue wick was used to remove excess fluid, leaving the microbeads in the well. A lambda concatomer (ProMega lambda ladder, *Promega*) was loaded either side of the digests and the lanes were capped with 1% low melting point agarose. The CHEF gel was run at 6V/cm for 18h at 14°C with a fixed angle of 120° in 0.5 x TBE. Both the initial and final switch times were 40 seconds. The partials were repeated using 8U, 4U, 2U, 1U, 0.75U, 0.5U, 0.25U, 0.125U and 0.0625U *HindIII*, since in the original series all digests using greater than 10U had been to completion.

2.3.3 Size selection

The amount of *HindIII* required for an average insert size of 150kb was determined from the partial digests (2.3.2). Two thirds of the microbeads were digested with 0.75U *HindIII* and the remainder were digested with 1U *HindIII*. The reactions were prepared as described (section 2.3.2) and 300µl microbeads were loaded in a

single lane of a 1% agarose CHEF gel in 1 x TAE with 50µl microbeads and a lambda concatomer loaded on both sides. The CHEF gel was run at 6V/cm for 20h at 14°C with a fixed angle of 120° in 1 x TAE. Both the initial and final switch times were 90 seconds.

The apparatus was dismantled and the lane containing DNA for size selection was cut away using a glass coverslip and stored in 1 x TAE at 4°C. The remainder of the gel was stained in 0.5µg/ml ethidium bromide for 10' and destained for 15'. The unstained piece and stained pieces were reassembled on a UV light box (300nm) and five cuts were made so that there were four gel pieces that spanned the region of digested DNA. The size-selected gel pieces were cut in half lengthwise and placed in 1.5ml eppendorfs. The surface of each piece was washed briefly three times with TE.

The gel pieces were weighed, melted by incubating in a 70°C waterbath for 5' and then allowed to equilibrate at 45°C for 5' before 1U GELase (*Epicentre Technologies*)/100mg gel slice was added to the top of the molten agarose. The GELase was mixed in by pipetting up and down once slowly with a wide-bore tip. The reaction was incubated at 45°C for 1h and then placed on ice to ensure that the agarose was completely digested.

A 2% agarose gel in 1 x TAE was prepared and 20µl size-selected DNA was loaded with 500ng, 250ng, 100ng, 80ng, 60ng, 40ng, 20ng, 10ng and 5ng of uncut lambda DNA. The DNA was electrophoresed at 100V for 0.5h and the gel was stained in 0.5µg/ml ethidium bromide. The concentration of the size-selected DNA was determined from the gel by comparison to the lambda standards. The three largest size-selections were used in the ligation (section 2.3.5).

2.3.4 BAC vector preparation

A stab of the vector, pBeloBAC11 in JM109 *E. coli* cells was obtained from Dr. R. Wing (Texas A&M University) and kindly donated by Dr. D.H. Shizuya and Dr. M.

Simon (California Institute of Technology). pBeloBAC11 is a single copy vector, so large amounts of cells are required to produce µg amounts of vector DNA. The vector was cultured from the stab by inoculating 5ml LB containing 30µg/ml chloramphenicol. This was grown for 6 hours at 37°C with shaking and then 500µl aliquots were used to inoculate 8 x 500ml of medium. These cultures were incubated at 37°C with shaking for 20 hours.

Plasmid DNA was prepared by the large-scale alkaline lysis method (Sambrook *et al.*, 1989). The cells were centrifuged for 30' at 5200g using 250ml tubes in the GSA rotor of a Sorvall RC50 centrifuge. The cells from each tube were resuspended in 6.25ml solution I with 2.5mg/ml lysozyme (ie., 25ml solution I per litre of cells). Next, 12.5ml Solution II was added to each tube (ie., 50ml/L cells), mixed well by inversion and placed on ice for 10' before 9.25ml solution III was added (ie. 37ml/L cells). The tubes were mixed gently by swirling and left on ice for 10' before they were centrifuged at 8000g for 30'. The supernatant was decanted from the tubes, filtered through Kimwipes and then incubated for 15' at room temperature with 0.1mg/ml RNase (*Boehringer Mannheim*). Approximately 200ml of supernatant was recovered prior to purification.

The plasmid DNA was purified using a Qiagen 500 kit following the manufacturer's instructions. Each 'tip 500' column was equilibrated with 10ml QBT buffer before 50ml supernatant was added and allowed to enter the resin by gravity flow. The columns were washed with 30ml QC buffer and then the DNA was eluted in 15ml QF buffer. The DNA from each column was precipitated in 0.7 volumes isopropanol and pelleted by centrifugation at 15000g for 30' at 4°C. The pellets were rinsed with ice cold 70% ethanol and airdried. Each pellet was resuspended in 2.5ml TE (pH 8) and pooled.

To remove any residual chromosomal DNA that may interfere with cloning, the pooled DNA was again purified down a Qiagen 500 column and precipitated with

isopropanol as before. The pellet was resuspended in 8ml TE and a cesium chloride gradient was prepared as the final purification step. The resuspended DNA was mixed with 8.81g CsCl to a final density of 1.6g/ml. The volume of each centrifuge tube was adjusted to 12ml with 1.5g/ml CsCl and 1.7g/ml CsCl before 200µl 10mg/ml ethidium bromide was added and mixed by inversion. The Beckman quick-seal tubes were capped and centrifuged in a Sorvall TV-865 ultra-centrifuge rotor at 27000g for 24 h at 20°C.

A 21 gauge needle (*Neolus*) was used to puncture an airhole in the top of the centrifuge tube. A needle attached to a 5ml syringe was then used to puncture the tube below the vector band visualised under UV (312 nm). The band was collected and the supernatant was split between 3 eppendorf tubes, with ~500µl/tube. An equal volume of CsCl saturated isoamyl alcohol was added to each tube and mixed by inversion. The pink layer containing the ethidium bromide was removed and the procedure was repeated until no pink was visible. The DNA was pooled into an Oakridge tube and precipitated with 2.5 volumes ethanol and 0.1 volumes 2M NaCl at 4°C overnight. The DNA was pelleted by centrifugation at 10000rpm for 20'. The pellet was rinsed with 70% ethanol, air-dried and then resuspended in 200µl ddH_20 overnight at 4°C.

2.3.4.1 Digestion and Dephosphorylation of pBeloBAC11

Prior to use in BAC library construction, a test digest of 5µl pBeloBAC11 (~750ng) with *HindIII*, followed by re-ligation and transformation by heat shock into DH10ß *E. coli* cells was performed to check the vector's quality.

To make the library, 20µg pBeloBAC11 was digested overnight at 37°C with 40U *HindIII* (*Promega*) in a 200µl reaction containing 6mM Tris-HCl (pH 7.5), 6mM $MgCl_2$, 100mM NaCl, 1mM DTT, 4mM spermidine and 1mg/ml BSA. After digestion, the volume was increased to 500µl and the DNA was extracted twice with 25:24:1

phenol:chloroform:isoamyl alcohol and once with chloroform before being ethanol precipitated and resuspended in 50µl ddH_20.

The digested vector was dephosphorylated at 37°C for 45' with 5U calf intestinal alkaline phosphatase (*Promega*) in a 100µl reaction containing 50µM Tris-HCl (pH 9.3), 1mM $MgCl_2$, 100µM $ZnCl_2$ and 1mM spermidine. The reaction was stopped with 0.5% sodium dodecyl sulphate, 5mM EDTA and 10µg proteinase K at 55°C for 30'. The DNA was extracted twice with 25:24:1 phenol:chloroform:isoamyl alcohol and once with chloroform before being ethanol precipitated and resuspended in 200µl ddH_20.

2.3.5 Ligation and transformation

To create a library of clones with an average insert size of 150kb, the size-selected DNA and pBeloBAC11 were ligated and then transformed into DH10β *E. coli*. Each 100µl reaction contained 40ng insert DNA, 20ng vector DNA, 30mM Tris-HCl (pH 7.8), 10mM $MgCl_2$, 10mM DTT, 1mM ATP and 4U T4 DNA ligase (*Promega*). The vector and insert DNA were incubated at 40°C for 5' without mixing. The buffer and enzyme were then added slowly and mixed once using a wide bore pipette. The ligations were incubated at 16°C overnight.

The ligated DNA was transformed into DH10β electromax competant *E. coli (Life Technologies)* by electroporation. On ice, 2µl ligation was mixed with 20µl DH10β cells and transferred to a microelectroporation chamber (*Life Technologies*). The chamber was placed into a cell-porator-safe (*Life Technologies*) that was filled with wet ice and the cells were electroporated using a cell porator with voltage booster (*Life Technologies*) set to 400V, 330µF, 4000 Ohms, low ohms impedance and fast charge rate. The cells were transferred to 250µl SOC medium pre-heated to 37°C and were incubated at 37°C for 1h with vigorous shaking. The entire 250µl was spread onto

14cm diameter LB plates containing 12.5µg/ml chloramphenicol, 62.5µg/ml Xgal (*Promega*) and 500µM IPTG (*Promega*). The plates were incubated at 37°C for 20h.

2.3.5.1 Determination of insert sizes

To determine which of the size selections produced clones with the largest inserts, ten colonies for each of the three selections were cultured overnight at 37°C in 5ml LB containing 12.5µg/ml chloramphenicol. The plasmid DNA was extracted by the alkaline lysis method (Sambrook *et al.*, 1989), precipitated with isopropanol and resuspended in 50µl TE at 65°C for 5'.

The inserts were excised from the vector by digestion with *NotI* and electrophoresed on a 1% agarose CHEF gel in 0.5 x TBE as described (section 2.4.1). From the CHEF gel, it was determined that the second largest size-selected gel slice produced clones with the largest inserts, so this DNA was used to construct the BAC library. Another ligation was prepared and the entire ligation was electroporated into DH10β electromax competent *E.coli* and plated onto LB agar containing 12.5µg/ml chloramphenicol, 62.5µg/ml Xgal and 500µM IPTG as described (section 2.3.5).

2.3.6 Culturing colonies for the BAC library

The BAC clones were cultured in 1200µl microtubes (*Evergreen Scientific*) arranged in boxes in a 12 x 8 format. Each tube was filled with 300µl LB containing 12.5µg/ml chloramphenicol. Each tube was inoculated with a different white colony picked manually from the plates with toothpicks. The toothpicks were left standing in the culture boxes until all the boxes to be cultured for the day were filled and then removed. The boxes were incubated at 37°C with shaking for 18-24h.

2.3.7 BAC DNA extraction and preparation of single-pools

Once the colonies had been cultured, the aliquots for DNA extraction were removed prior to the addition of glycerol storage solution to the tubes (section 2.3.8). Using a 12-tip multichannel pipette (*Lab Systems*), a 100µl aliquot was taken from each tube in a box, pooled in an autoclaved plastic reservoir (*Jomar Diagnostics*) and then transferred to a 50ml Falcon tube. The cells from the "single-pool" were pelleted by centrifugation at 1500g for 5' in the swing bucket rotor of a Sorvall RT 6000D. The supernatant was decanted and the BAC DNA was extracted from the cells by a standard alkaline lysis mini-preparation (Sambrook *et al.*, 1989). The cells were resuspended in 200µl solution I containing 5mg/ml lysozyme and transferred to a 1.5ml eppendorf. The tubes were incubated at room temperature for 5' and then on ice for 5' before 400µl solution II was added and mixed by inversion. After a 5' incubation on ice, 300µl solution III was added, mixed well, and the tubes incubated at -80°C for 10'. The tubes were thawed at room temperature and centrifuged at 14000g for 5' to pellet the cellular debris. The supernatant was decanted into a new 2ml eppendorf and 1µl 10mg/ml RNase (*Boehringer Mannheim*) was mixed in thoroughly by inversion before the tubes were incubated at 37°C for 30'. One volume of 25:24:1 phenol:chloroform:isoamyl alcohol was added to each tube, mixed by inversion and centrifuged for 5' at 14,000g. The upper aqueous phase was transferred to a new tube and extracted once with an equal volume of chloroform. After centrifugation at 14,000g for 5', the upper aqueous phase was again transferred to a new tube and 540µl isopropanol (0.6 volume) was added. The tubes were incubated at -80°C for a minimum of 15', thawed at room temperature and the DNA was pelleted by centrifugation at 14,000g for 15'. The supernatant was decanted and the pellet was rinsed with 500µl 70% ethanol and centrifuged at 14,000g for 5'. The pellet was air-dried and then resuspended in 200µl TE (pH 8.0) overnight at 4°C. After the DNA was eluted, it was split into 2x100µl

aliquots and one aliquot of each single-pool was retained at Texas A&M and the other was returned to the University of Adelaide and stored at -20°C.

2.3.7.1 Single-pool working stocks for PCR

A 1/10 dilution of every single-pool was made (10µl BAC DNA + 90µl ddH20) as a working stock for amplification by the polymerase chain reaction. To confirm that each working stock contained DNA, they were all amplified by PCR using two primers designed from pBeloBAC11. The reaction mixture consisted of 20mM Tris-HCl (pH 8.4), 50mM KCl, 125µM dGTP, 125µM dATP, 125µM dTTP, 125µM dCTP, 25pmol forward primer (5' GAA AAT ACC GCA TCA GGC 3'), 25pmol reverse primer (5' GGA TCC CCG GGT ACC GAG 3'), 1.5mM $MgCl_2$, 1µl single-pool DNA working stock and 0.5U Taq DNA polymerase (*Life Technologies*). The reactions were amplified using an Omnigene thermal cycler (*Hybaid*) and the following cycle conditions:

1 cycle: 94°C 4'
35 cycles: 94°C 1', 50°C 1', 72°C 1'
1 cycle: 72°C 10'

2.3.7.2 Preparation of superpools

"Super-pool" DNA was produced by combining 20µl from 8 single-pool DNA working stocks so that the entire library was represented in 78 tubes (Figure 2.1). This combination had previously been shown to produce robust PCR results (Cai *et al.*, 1995) and allows PCR to be performed in 96-well plate format. The superpools were stored in a microtitre plate with strip caps (*Integrated Sciences*) at -20°C. Although this meant the superpools were subjected to multiple freeze-thaws, it avoided the problem of evaporation of the BAC DNA that was encountered when the plate was stored at 4°C.

H	G	F	E	D	C	B	A	
1-8	9-16	17-24	25-32	33-40	41-48	49-56	57-64	1
65-72	73-80	81-88	89-96	97-104	105-112	113-120	121-128	2
129-136	137-144	145-152	153-160	161-168	169-176	177-184	185-192	3
193-200	201-208	209-216	217-224	225-232	233-240	241-248	249-256	4
257-264	265-272	273-280	281-288	289-296	297-304	305-312	313-320	5
321-328	329-336	337-344	345-352	353-360	361-368	369-376	377-384	6
385-392	393-400	401-408	409-416	417-424	425-432	433-440	441-448	7
449-456	457-464	465-472	473-480	481-488	489-496	497-504	505-512	8
513-520	521-528	529-536	537-544	545-552	553-560	561-568	569-576	9
577-584	585-592	593-600	601-608	609-616	617-624			10

Figure 2.1: Layout of superpool DNA for BAC library screening by the polymerase chain reaction. The 78 superpools each contain 8 plates of 96 BAC clones.

2.3.8 Storage of the BAC library

After the aliquots for DNA extraction had been removed, an equal volume of 2 x BAC storage solution was added to every tube. The cells and glycerol solution were mixed thoroughly by pipetting up and down, and then 100µl aliquots from every well were transferred to three gamma-sterilised microtitre plates (*Technoplas*). The plates were sealed with tape and stored at -80°C. The cells remaining in the culture tubes were stored at -20°C as a working copy of the library. One -80°C copy of the library was retained at Texas A&M University, one copy was stored at Utah State University and the third copy was returned to the University of Adelaide.

2.3.9 Preparation of row and column DNA

The plates stored at the University of Adelaide were used to generate three additional stocks of single-pool DNA as well as DNA from all the rows and columns. The microtitre plates were thawed at room temperature and the cells were mixed thoroughly by pipetting up and down. All the solution was dispensed back into the microtitre wells and what remained in the tip was sufficient to inoculate microtubes

containing 800µl LB with 12.5µg/ml chloramphenicol. The cultures were incubated at 37°C with shaking for 18-24h. For single-pool DNA, 300µl aliquots were taken from every tube in a box and pooled. A 100µl aliquot was then taken from each tube of a row or each tube of a column and pooled. The BAC DNA was extracted by a standard alkaline lysis mini-preparation (Sambrook *et al.*, 1989) as described (section 2.3.7). After the aliquots were taken for DNA extraction, the cells were stored (section 2.3.8).

One copy of the BAC library was stored at the University of Sydney and two copies of the library were stored in different -80°C freezers at the University of Adelaide. A working copy of the library was also stored at -20°C. One copy of the single-pool DNA and the row and column DNA was sent to Utah State University and one copy of the single-pool DNA was sent to AgResearch, New Zealand.

2.4 BAC library characterisation

To determine the amount of the genome covered by the BAC library, clones were sized (section 2.4.1) and genes and microsatellite markers were amplified by the polymerase chain reaction (PCR) (section 2.4.2).

2.4.1 Sizing BAC inserts

A total of 139 BAC clones were selected at random from the library and sized by CHEF gel electrophoresis. In addition to this, BAC clones that were isolated from the library by PCR-based screening (section 2.4.2) were also sized. BAC DNA for sizing was prepared by a standard alkaline lysis mini-preparation (section 2.3.7) from 5ml overnight cultures in LB with 12.5µg/ml chloramphenicol. The DNA was resuspended in 50µl TE overnight at 4°C.

The insert DNA was excised by digestion with *NotI* which has two recognition sequences in pBeloBAC11 that flank the *HindIII* cloning site. Each 25µl digest consisted of 3µl BAC DNA, 6mM Tris-HCl (pH 7.9), 6mM $MgCl_2$, 150mM NaCl,

1mM DTT, 4mM spermidine and 10U *NotI* (*Promega*). The reactions were incubated at 37°C for 5h and then the total digest was loaded onto a 1% agarose CHEF gel in 0.5 x TBE. A slice of a lambda concatomer (*Promega*) and 250ng *HindIII* lambda ladder (*Promega*) were loaded in a single lane.

To separate fragments from 100 - 500kb, the BAC DNA was electrophoresed using a CHEF DRII apparatus (*Biorad*) at 6V/cm for 18h at 14°C with a fixed angle of 120°. The initial switch time was 9.8s with a linear ramp to the final switch time of 35.3s. To separate fragments from 25 - 200kb, the gel was run for 16h with an initial switch time of 1.4s and a final switch time of 13.5s.

2.4.2 PCR-based screening strategy

The ovine BAC library was screened by PCR with a variety of primers for microsatellite markers and genes of interest to establish the amount of the genome that the library covered. Primers for PCR were kindly donated by Dr. Jill Maddox (University of Melbourne) and Dr. Noelle Cockett (Utah State University). Primers for genes of interest were made available by Prof. Cindy Bottema or were designed from published sequences using Oligo 4.0 software (*Molecular Biology Insights, Inc.*). Primers containing hairpins or which could form dimers at the 3' end were avoided.

Primers were selected so that at least two BAC clones for every sheep chromosome were isolated from the library. Where possible, a marker near the centromere and a second marker near the telomere were selected from the ovine genetic linkage map (Maddox *et al.*, 1996; de Gortari *et al.*, 1998). The BAC clones containing these markers would then anchor the physical and genetic linkage maps for each of the chromosomes.

To minimise the number of PCR reactions that needed to be performed in order to isolate an individual clone containing a sequence of interest, the BAC DNA was pooled. The first screen was of the 78 superpools which each contained 8 plates of 96 BAC

clones (section 2.3.9). In the second step, the 8 single-pool working stocks (section 2.3.8) that corresponded to a positive superpool were screened. Then finally the row and column DNA from the positive plate were screened. The intersection of the row and column identifies the location of the positive BAC clone (Figure 2.2).

2.4.2.1 Optimisation of PCR conditions

The PCR conditions for every set of primers were optimised so that a single product of the expected size was amplified. Reactions were performed in Omnigene 96 well microtitre plates (*Hybaid*). A standard 25µl reaction consisted of 20mM Tris-HCl (pH 8.4), 50mM KCl, 125µM dGTP, 125µM dATP, 125µM dTTP, 125µM dCTP, 25pmol forward primer, 25pmol reverse primer, 1.5mM $MgCl_2$, 100ng DNA template and 0.5U Taq DNA polymerase (*Life Technologies*). The reactions were overlayed with mineral oil and amplified using an Omnigene thermal cycler (*Hybaid*) with satellite attachments with the following conditions:

Simulated tube control

Calibration factor: 500

1 cycle: 94°C 4'

35 cycles: 94°C 1', T_A 1', 72°C 1'

1 cycle: 72°C 10'

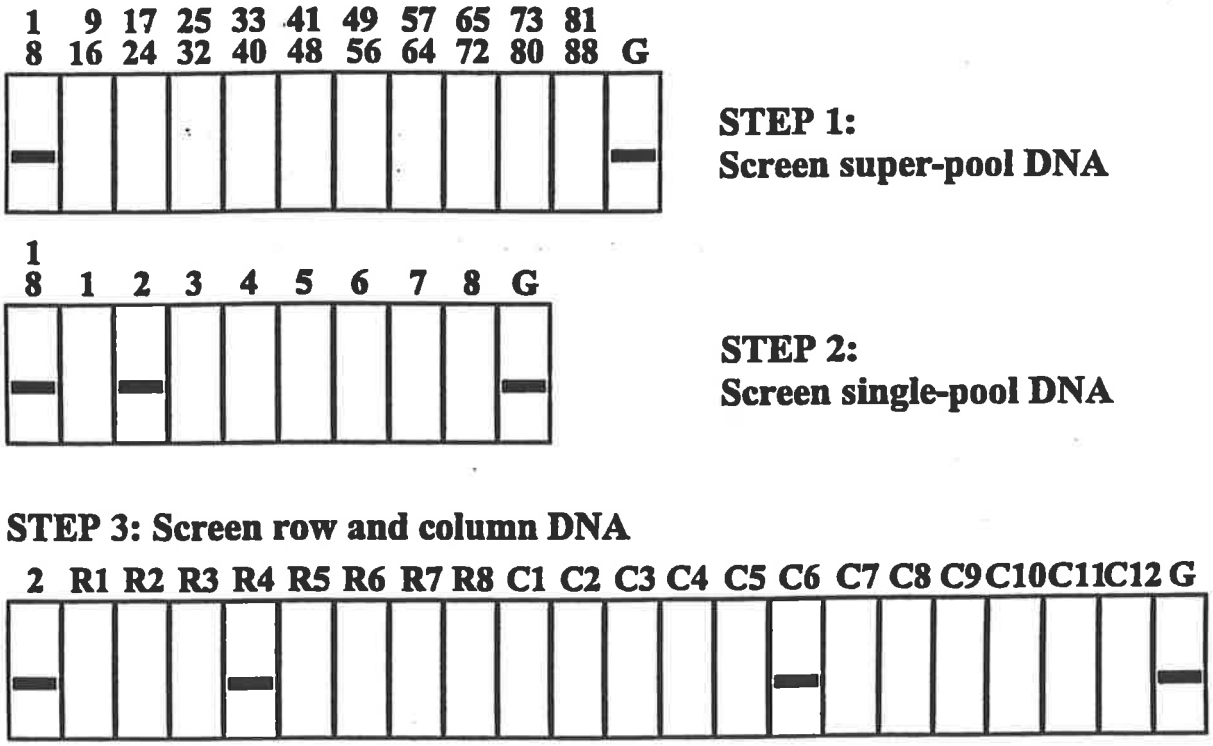


Figure 2.2: PCR-based screening strategy to identify a clone from the ovine BAC library that contains a sequence of interest. The 78 super-pools each consisting of 8 plates of 96 BAC clones are screened first. The second step is to screen the 8 single-pools (96 clones/pool) that correspond to a positive reaction. Finally, DNA from the rows and columns is screened and the intersection of the positive row and positive column is the address of the BAC of interest (eg. 2R4C6). (G: genomic control).

Initially, an annealing temperature (T_A) of 55°C was used. This temperature was altered if a single product could not be produced by adjusting the stringency of the PCR through either the magnesium or primer concentration. If no bands were produced under the standard conditions, then either the annealing temperature was decreased or a magnesium titration was performed (1.5mM, 2mM, 2.5mM, 3mM, 3.5mM, 4.5mM). If multiple bands were present, then either the temperature was increased and a magnesium titration or a primer titration (1x,1/2x, 1/5x, 1/10x, 1/20x, 1/50x) was performed. The optimum conditions for the primer pairs that were used to screen the BAC library are listed in Table 2.2.

2.4.2.2 Analysis of PCR products by agarose gel electrophoresis

PCR products were separated by electrophoresis on 2% agarose gels in 1 x TAE for 0.5 - 0.9h at 7.5 V/cm. A 'wide mini-sub cell' electrophoresis unit (*Biorad*) was used routinely with either 15, 20 or 30 lanes depending on the number of samples. The PCR products were sized against a pGEM ladder cut with *HinfI, RsaI* and *SinI* (*Promega*). Gels were stained in 0.5µg/ml ethidium bromide for 10' and photographed under UV illumination (312nm) using an integrating camera and Gel-Documentation 1000 System (*Biorad*). The images were captured using Molecular Analyst software and printed on photographic paper with a Hewlett Packard 890CXI Inkjet printer.

Table 2.2: Primers and reaction conditions used to screen the ovine BAC library.

Ch.	Locus	Primer sequence	pmol	$[Mg^{2+}]$	T_A	Citation[1]
15	ADCYC	5' AAAGTGACACAACAGCTTCTCCAG 3' 5' AACGAGTGTCCTAGTTTGGCTGTG 3'	20	1.5mM	58°C	X16451
6	BM415	5' GCTACAGCCCTTCTGGTTTG 3' 5' GAGCTAATCACCAACAGCAAG 3'	5	1.5mM	55°C	G18413
12	BM719	5' TTCTGCAAATGGGCTAGAGG 3' 5' CACACCCTAGTTTGTAAGCAGC 3'	12.5	1.5mM	55°C	G18427
3	BM827	5' GGG CTG GTC GTA TGC TGA G 3' 5' GTT GGA CTT GCT GAA GTG ACC 3'	12.5	1.5mM	55°C	U06763
19	BM1303	5' CTTGGGAAAAATTGCCAGC 3' 5' CTCTGCCGCTTGCCTCTC 3'	12.5	1.5mM	55°C	G18424
6	BM1329	5' TTGTTTAGGCAAGTCCAAAGTC 3' 5' AACACCGCAGCTTCATCC 3'	12.5	1.5mM	55°C	G18422
1	BM1824	5' GAGCAAGGTGTTTTTCCAATC 3' 5' CATTCTCCAACTGCTTCCTTG 3'	25	1.5mM	55°C	G18394
18	BM3413	5' TCCCTGGTAACCAATGAATTC 3' 5' CAATGGATTTGACCCTCCC 3'	25	1.5mM	55°C	G18492
16	BM4107	5' AGCCCCTGCTATTGTGTGAG 3' 5' ATAGGCTTTGCATTGTTCAGG 3'	12.5	1.5mM	55°C	G18519
8	BM4208	5' TCAGTACACTGGCCACCATG 3' 5' CACTGCATGCTTTTCCAAAC 3'	5	1.5mM	55°C	G18509
1	BM6438	5' TTGAGCACAGACACAGACTGG 3' 5' ACTGAATGCCTCCTTTGTGC 3'	1.25	1.5mM	55°C	G18435
26	BM6526	5' CATGCCAAACAATATCCAGC 3' 5' TGAAGGTAGAGAGCAAGCAGC 3'	25	1.5mM	55°C	G18454
15	CD3D	5' CTGCTTTGCTGGACATGAG 3' 5'CTGATAGAGCTGGTCATTCCC 3'	25	1.5mM	51°C	X52933
21	CD5	5' CAACTAGTCATTCTTGAGGGATTTC 3' 5' ATACGGAAAAAATACCAGATTAGACACT 3'	25	3mM	55°C	Matthew and Maddox, 1994a
5	CSF2		25	1.5mM	53°C	Maddox, unpublished
X	CSRD241	5' TAACACGGGCAAACAGCCTTTCA 3' 5' GTCTGTTTGGTAAACAGATAGCA 3'	20	1.5mM	50°C	Davies *et al.*, 1996
14	CSRD270	5' TTGCCTGAGCCAATTCTTTACAGT 3' 5' TGCTTGCAGCTGTGGTTGTATTAG 3'	25	1.5mM	55°C	Drinkwater *et al.*, 1997
10	CSRD287	5' ACAAAGGGCATGACTGAGCAACTA 3' 5' GGAGAGGATTCAAGTGCATCAGTA 3'	2.5	1.5mM	60°C	Drinkwater *et al.*, 1997
26	CSSM043	5' AAAACTCTGGGAACTTGAAAACTA 3' 5' GTTACAAATTTAAGAGACAGAGTT 3'	25	1.5mM	55°C	U03824
20	DQA1		12.5	1.5mM	61°C	Maddox, unpublished
20	DQA2		12.5	1.5mM	51°C	Maddox, unpublished
20	DQB		25	2.5mM	65°C	Maddox, unpublished
20	DRB1		25	1.5mM	55°C	Maddox, unpublished
20	DRB2		25	1.5mM	58°C	Maddox, unpublished
20	DYB		25	1.5mM	50°C	Maddox, unpublished

Table 2.2 continued.

Ch.	Locus	Primer sequence	pmol	$[Mg^{2+}]$	T_A	Citation[1]
2	FN1	5'TCCAACCCAGAGAAACAAGCGT 3' 5' GATCCCAATCCAAATCCAAAGT 3'	5	1.5mM	50°C	G42685
15	FSHB	5' TGGGATATAGACTTAGTGGC 3' 5' CAGTTTCTAAGGCTACATGGT 3'	20	1.5mM	50°C	S64745
11	GH2		2.5	2mM	60°C	Bottema, unpublished
12	HUJ614	5' CGCCAGGCATGGTGAAGTCG 3' 5' CCCAGCACAGTACAGGCTGC 3'	25	1.5mM	55°C	M93653
3	IFNG	5' TTGTGACTGTTAGCTAGATGTGTT 3' 5' ATACACATATTATGCCCATCTTTT 3'	25	3.5mM	52°C	Z54144
3	IGF1	5' GTGGTTGTTGTCAGAGATGAGT 3' 5' CTCCTTCTGTTCCCCTCCTGG 3'	1	1.5mM	45°C	U01338
3	IL1A	5' ACATTGTGAGGGATACAAAGATGT 3' 5' GAGATTCTTAGAGTCACAGGAAGC 3'	25	2mM	55°C	AF117652
13	IL2RA	5' AGCAGAGGTACAGGTGGTAAGCA 3' 5' GATAGCCTTGGAGAAGGTAGCGTAT 3'	20	1.5mM	50°C	Matthew and Maddox, 1994b
5	IL3	5' AGCCTCTCTATCTTGCATCTGCTC 3' 5' TCATCTGAGTTCAAGGAGCCCT 3'	25	1.5mM	55°C	Z18291
5	IL5	5' GACAGTTTCCTACGATGCTTATTGTC 3' 5' TCTCCTCCAGAAACTTTCACTACAG 3'	25	1.5mM	45°C	U17053
4	IL6	5' CTAAAAGTATGAGCGTTAGGAC 3' 5' GAGGTAAGCCTACACATTTCA 3'	25	1.5mM	50°C	X62501
7	ILSTS005	5' GGAAGCAATGAAATCTATAGCC 3' 5' TGTTCTGTGAGTTTGTAAGC 3'	25	1.5mM	50°C	L23481
9	ILSTS008	5' GAATCATGGATTTTCTGGGG 3' 5' TAGCAGTGAGTGAGGTTGGC 3'	12.5	1.5mM	55°C	L23483
9	ILSTS011	5' GCTTGCTACATGGAAAGTGC 3' 5' CTAAAATGCAGAGCCCTACC 3'	12.5	1.5mM	55°C	L23485
24	ILSTS043	5' TATTCAAAGTCACAGACGCC 3' 5' TAGGGAATGGCGAATTTGCC 3'	25	2mM	55°C	L37218
5	IRF1		25	1.5mM	53°C	Maddox, unpublished
11	KAP1.1	5' TATGACAAAGCAAGCCAGGG 3' 5' GAGTTGGGTTGAGAGGAGGG 3'	25	1.5mM	63°C	X01610
1	KAP6.1	5' CCAATGGCATGAAGGTGT 3' 5' AAAAAGGGAAGGGTTGGTG 3'	25	1.5mM	50°C	M95719
3	KRT2.13	5' GCCTGTAGGCGTGAGGGTTTT 3' 5' AAGGGCCAAGAGTCATTCACAT 3'	25	1.5mM	55°C	X72379
3p	LGB	5' AGCAACACACCCAGCACCAG 3' 5' CAGGAGGCACTTCATGGCTG 3'	2.5	1.5mM	59°C	X12817
3	MAF23	5' GTGGAGGAATCTTGACTTGTGATAG 3' 5' GGCTATAGTCCATGGAGTCGCAG 3'	1.25	2mM	58°C	M38719
23	MAF35	5' TCAAGAATTTTGGAGCACAATTCTGG 3' 5' AGTTACAAATGCAAGCATCATACCTG 3'	25	1.5mM	50°C	M80518
4	MAF50	5' GTAGACTACTCATGAAAATCAGGTCTTAGG 3' 5' GGGACATGCAGCTATACACTTGAG 3'	12.5	1.5mM	55°C	M77377
22	MAF92	5' TAGAATGTCATGTTCTCAGCATTCCC 3' 5' AACCCATGAATCATCTCTAACTACCTC 3'	25	2mM	60°C	M80527
17	MAF209	5' GATCACAAAAAGTTGGATACAACCGTGG 3' 5' TCATGCACTTAAGTATGTAGGATGCTG 3'	20	1.5mM	50°C	M80358
16	MAF214	5' GGGTGATCTTAGGGAGGTTTTGGAGG 3' 5' AATGCAGGAGATCTGAGGCAGGGACG 3'	0.5	1.5mM	55°C	M88160

Table 2.2 continued.

Ch.	Locus	Primer sequence	pmol	$[Mg^{2+}]$	T_A	Citation[1]
1p	McM58	5' CTGGGTCTGTATAAGCACGTCTCC 3' 5' CAGAACAATAAACGCTAAACCAGAGC 3'	5	1.5mM	55°C	L34283
19	McM111	5' CTGCCTGTCTGTAAAATGGAGCT 3' 5' ACTATCCCAGCCTCCTTGAGG 3'	25	1.5mM	55°C	L35310
23	McM136	5' GCA CAC ACA TAC ACA GAG ATG CG 3' 5' AAA GAG GAA AGG GTT ATG TCT GGA 3'	1.25	1.5mM	53°C	L38978
4	OarHH64	5' CGTTCCCTCACTATGGAAAGTTATATATGC 3' 5' CACTCTATTGTAAGAATTTGAATGAGAGC 3'	25	1.5mM	55°C	L12558
6	OarJMP8	5'CGGGATGATCTTCTGTCCAAATATGC 3' 5' CATTTGCTTTGGCTTCAGAACCAGAG 3'	25	1.5mM	55°C	U35059
26	OarJMP58	5' GAAGTCATTGAGGGGTCGCTAACC 3' 5' CTTCATGTTCACAGGACTTTCTCTG 3'	25	1.5mM	57°C	U35058
25	OarVH72	5' CTCTAGAGGATCTGGAATGCAAAGCTC 3' 5' GGCCTCTCAAGGGGCAAGAGCAGG 3'	25	1.5mM	55°C	L12548
21	OarVH110	5' CTCTAGAGGATCACAGAGAGTCGG 3' 5' GCAGAAACATTTTTTTCCTTCAATATAGTTTCCC 3'	25	1.5mM	55°C	L12550
17	OarVH116	5' AGTGTGACTAGAGAACTAAATTTTGAAGGTC 3' 5' TATTTTTCCATCAAAAAGAACTCTATAGGGC 3'	25	1.5mM	55°C	L12551
4	OBS	5' GTCTGGAGGCAAAGGGCAGAGT 3' 5' CCACCACCTCTGTGGAGTAG 3'	2.5	1.5mM	64°C	U50365
21	OPCML	5' CCTGACTATAATGTACAGATCCCTC 3' 5' GCAGAATGACTAGGAAGGATGGCA 3'	25	1.5mM	50°C	X12672
25	PRF	5' CCTCTGCACATTCTTGTGGA 3' 5' GATAGGCGTCTGTGGCTGTG 3'	25	1.5mM	50°C	Maddox *et al.*, 2000
19	PRL	5' GGAAAGTGAACATGA 3' 5' GCCCTCTCTTCTACAA 3'	20	4.5mM	50°C	Bottema, unpublished
9	RJH1	5' TGATTTAGATGCTTTGCTAATGCCA 3' 5' GGATTCTTTACCACTAGCCCCACCT 3'	1.25	1.5mM	55°C	U47049
7	RNASE6		25	2mM	50°C	Bottema, unpublished
11	SCYA	5' TTTCTTCCTCCTATTTCCCCTTCT 3' 5' ACCCACTTCTTCTTGGGGTC 3'	25	2mM	55°C	Maddox, unpublished
Y	SRY	5' TGCCAGGAGGTATTGAGGGG 3' 5' CAGAGGAGCAGTTATTTTGG 3'	25	2mM	63°C	Bottema, unpublished
18	TGLA122	5' AATCACATGGCAAATAAGTACATAC 3' 5' CCCTCCTCCAGGTAAATCAGC 3'	25	1.5mM	55°C	Crawford *et al.*, 1995
14	TGLA357	5' GCAGAGTCTGAGTTTAAACTTCTCTAACACC 3' 5' GAGGGCAAAAAGGTTTGGGGTGTATGG 3'	25	1.5mM	55°C	Crawford *et al.*, 1995
14	UWCA5	5' CTT CCC TGT GTT AGG CTA TGT 3' 5' CAA GAG AAG ACA AGA CTA CAG A 3'	12.5	1.5mM	55°C	L15338
X/Y	ZFX/ZFY		2.5	2mM	55°C	D84097

[1]Genbank Accession numbers when available

2.4.2.3 Screening superpool DNA

To minimise the number of freeze-thaws to which the superpool DNA was subjected, each week 2µl aliquots of every superpool were dried down in Omnigene mictrotitre plates (*Hybaid*) at 37°C and the plates stored at 4°C until used in PCR. A genomic DNA control (100ng) and a negative (no DNA) control were also included on the plate. A master-mix containing all the reagents for 85 PCR reactions was prepared, mixed well and then split so that each reaction was 25µl. The reactions were overlayed with mineral oil and amplified using the optimum conditions for the primer pair (Table 2.2). Positive superpools were identified by comparing the size of any PCR products generated by the superpool DNA to the genomic DNA control after agarose gel electrophoresis (section 2.4.2.2).

2.4.2.4 Screening single pools

The eight single pools that corresponded to a positive superpool were screened by PCR using 1µl of the single pool working stock as the template for PCR. Reactions containing a 1µl aliquot of the positive superpool, 100ng genomic DNA and a no DNA control were also prepared.

2.4.2.5 Screening row and column DNA

A 1µl aliquot of BAC DNA from the 12 columns and 8 rows that corresponded to a positive single-pool were amplified by PCR. Reactions for the positive superpool, single pool, genomic DNA control and no DNA control were also included in this screen. The individual BAC containing the sequence of interest was located at the intersection between the positive row and positive column.

2.4.2.6 Confirmation of individual BAC

Once the location of the individual BAC was determined, the box containing the working stock of cells was recovered from the -20°C freezer. A sterile 100ml bottle containing 50ml LB with 12.5µg/ml chloramphenicol was inoculated with the appropriate BAC clone and incubated with shaking at 37°C for 18-24h. The BAC DNA was extracted by a standard alkaline lysis miniprep (Sambrook *et al.*, 1989) as described (section 2.3.7) and resuspended in 50µl TE. The superpool, single-pool, row, column, individual BAC and genomic DNA control were all amplified by PCR to ensure that the clone that had been isolated generated a product of the expected size. The identity of the PCR product, and therefore the BAC, was confirmed by sequencing prior to physical mapping.

2.4.3 Sequencing

To avoid making incorrect physical mapping assignments, the identity of every BAC clone isolated from the library was confirmed by sequencing prior to fluorescence *in situ* hybridisation. PCR product was generated from the individual BAC, and this product was used as the template for cycle sequencing reactions with γ32P-dATP labelled primers.

2.4.3.1 Template for sequencing

PCR products were generated using 1µl of the individual BAC (section 2.4.2.5) as the template and the same PCR conditions that were used to screen the BAC library (Table 2.2). To generate sufficient product for sequencing, 6 x 25µl reactions were prepared and the success of the amplification was checked by agarose gel electrophoresis as usual. The remaining product was pooled and precipitated with 0.6 volumes isopropanol and 1/20 volume 4M NaCl. The tube was incubated at -80°C for 30' and then centrifuged at 14,000g for 10' to pellet the DNA. The pellet was rinsed

with 70% ethanol and centrifuged again for 5' at 14,000g. The DNA was air-dried and resuspended in 40µl TE.

2.4.3.2 End-labelling primers

Each template was sequenced with both the forward and reverse primers which were labelled with γ^{32}P-dATP (*Amersham*) using reagents supplied in a Sequitherm Excel II sequencing kit (*Epicentre Technologies*). Each 25µl end labelling reaction consisted of 25pmol primer, 33mM Tris-acetate (pH 7.8), 66mM potassium acetate, 10mM magnesium acetate, 500µM DTT, 3U T4 Polynucleotide Kinase and 1µl γ^{32}P-dATP. The reaction was incubated at 37°C for 30' and then stopped by heating to 70°C for 5'.

2.4.3.3 Cycle sequencing

For each primer, a 17µl pre-mix was prepared in a PCR tube and consisted of 1.5pmol end-labelled primer, 1 x Sequitherm EXCEL™ II sequencing buffer, 1µl purified PCR product (~75ng) and 5U Sequitherm EXCEL™ II DNA polymerase. The termination mixes were 15µM each of dATP, dCTP, dTTP, 7-deaza-dGTP and 0.0255mM ddGTP or 0.45mM ddATP or 0.525mM ddTTP or 0.3mM ddCTP, respectively. Four tubes (G, A, T, C) containing 2µl of the appropriate termination mix were prepared for each primer and combined with 4µl pre-mix. The sequencing reactions were overlayed with mineral oil and placed in a Perkin Elmer DNA thermal cycler. A two-step thermal cycling profile was used for all the sequencing reactions:

1 cycle: 95°C for 5'

30 cycles: 95°C 30", 72°C 1'.

When the cycles were complete, 3µl sequencing gel loading buffer (95% v/v formamide, 10mM EDTA, 0.1% xylene cyanol, 0.1% bromophenol blue, pH 11.0) was added to every tube.

2.4.3.4 Sequencing gel electrophoresis

A Life Technologies Model S2 sequencing gel apparatus was used. The small plate was coated with Rainkote (*Maxwell Chemicals*), both plates cleaned thoroughly with 5% SDS, and rinsed with ddH_20 and ethanol. The plates were assembled with 0.4 mm spacers and two 24 well sharks-tooth combs. A sequencing gel was prepared by mixing 100ml de-gassed 6% acrylamide with 560µl freshly prepared 10% ammonium persulphate (*Sigma*) and 22.5µl TEMED (*Biorad*) to catalyse the polymerisation. The gel was poured, allowed to set for 60', and then pre-run in 1 x TBE at 70W for 60' or until the gel temperature reached 50°C. The sequencing reactions were denatured at 75°C for 5' and then snap chilled on ice. Immediately before loading 4µl each sample on the gel, the wells were cleaned out with 1 x TBE running buffer. The gel was run for 75' at 70W or until the first dye front had just run off the bottom of the gel. This length run was sufficient to read 200-300 bases. For a longer read, the samples were re-loaded in another set of lanes after the first 75'. The apparatus was dismantled and the gel was transferred to a 13 x 16 inch sheet of Whatman paper, vacuum dried at 80°C for 60', and exposed to X-ray film (*Fuji*) for 1-3 days. The film was developed for 3' in a working stock of Phenisol (*Ilford*) and fixed for 3' in a working stock of Hypam X-ray rapid fixer that contained hardener (*Ilford*). The film was rinsed thoroughly in running water and air-dried. The sequence was read from the bottom to the top, input into the Oligo 4.0

programme and aligned with previously published sequence using Seqed version 1.0.3 or BLAST 2.0 (Altschul *et al.,* 1997).

2.5 Physical mapping

To estimate the number of chimaeric clones in the BAC library, all the BAC clones that were isolated by PCR-based screening were physically mapped by fluorescence *in situ* hybridisation.

2.5.1 Preparation of CotI DNA

CotI DNA is required to suppress repetitive DNA in probes that are used for FISH. In previous studies, it has been shown that whilst there is some cross-hybridisation of repetitive DNA between species, some repeats are unique to a species (Jobse, *et al.*, 1995; Modi *et al.*, 1996). Consequently, CotI DNA was prepared for sheep, cattle and goat. Whenever a young animal died, the thymus and liver were recovered as sources of DNA and the tissue was stored at -20°C. DNA was extracted from thymus (Ausubel *et al.*, 1992) or liver (Towner, 1991) to prepare CotI DNA.

2.5.1.1 DNA Extraction from thymus

A thymus was recovered from storage at -20°C and thawed at room temperature. The ~12g thymus was cut into 2g amounts, frozen in liquid nitrogen and ground into a fine powder with a mortar and pestle. Each 2g of powdered tissue sample was incubated with 24ml digestion buffer and 0.1mg/ml pronase (*Boehringer Mannheim*) in a 50ml Falcon tube at 50°C overnight with shaking. After the overnight incubation, 1 volume of 25:24:1 phenol:chloroform:isoamyl alcohol (PCI) was added and mixed thoroughly by inversion. The tubes were centrifuged for 10' at 1600g in the swing bucket rotor of a Sorvall RT 6000D centrifuge. If there was a definite interface following this centrifugation, then the upper aqueous phase was transferred to a new

tube. If the interface was not definite, then only the top 10ml of the upper aqueous phase was recovered, replaced with 10ml digestion buffer and the tube centrifuged again at 1600g for 10'. After this spin, the aqueous phase was transferred to new tubes so that each tube contained ~20ml solution. The extraction with PCI was repeated and followed by an extraction with an equal volume of chloroform. The DNA was precipitated by adding 1/20 volume 4M sodium chloride and 0.6 volume isopropanol and incubating the tubes at 4°C for 15'. The DNA was spooled around a 1ml pipette tip and the supernatant was poured off. The spooled DNA was then rinsed in 70% ethanol, air-dried and resuspended in 10ml TE overnight at room temperature with gentle rocking. After the DNA was thoroughly resuspended, it was treated with RNase as described (section 2.5.1.3).

2.5.1.2 DNA Extraction from liver

If a thymus was unavailable, then a liver was thawed at room temperature and cut into 2.5g amounts. The tissue was frozen in liquid nitrogen and ground to a fine powder with a mortar and pestle. Each 2.5g powdered tissue sample was transferred to 10ml CTAB (*Sigma*) buffer and incubated at 65°C for 15' with occasional shaking. After this incubation, 10ml chloroform was added to each tube and mixed by inversion before centrifugation at 3500g for 10'. The upper aqueous phase was transferred to a new tube and 0.2 vol 5% CTAB was added before the chloroform extraction was repeated. The upper aqueous phase was transferred to a new tube and an equal volume of CTAB precipitation buffer was added. The tubes were incubated at 4°C for 15' and then centrifuged at 3500g for 15' to pellet the nucleic acid. The supernatant was decanted and 2ml TE containing 1M NaCl was added to the pellet. The nucleic acid was redissolved at 65°C for 10'. The DNA was precipitated by adding 2 volumes 100% ethanol. After the tubes were mixed by inversion, they were immediately centrifuged at 3500g for 15' at 4°C. The supernatant was decanted and the pellets rinsed with 70%

ethanol. The tubes were centrifuged again at 3500g for 10', the supernatant decanted, and the pellets air-dried. The DNA was dissolved in 10ml TE overnight at room temperature with gentle rocking and then treated with RNase.

2.5.1.3 RNase Treatment

A 10µl aliquot of 10mg/ml DNase-free RNase was added to each tube and incubated at 37°C for 30'. An equal volume of 25:24:1 phenol: chloroform: isoamyl alcohol (PCI) was added to every tube, mixed thoroughly and centrifuged at 3500g for 10'. The upper aqueous phase was transferred to a new tube and the PCI extraction was repeated. Again, the upper aqueous phase was transferred to a new tube and an equal volume of chloroform was added. The tubes were mixed by inversion and then centrifuged at 3500g for 10'. The aqueous phase was recovered and the DNA was precipitated with 1/20 volume 4M sodium chloride and 0.6 volume isopropanol at 4°C for 15'. The DNA was either spooled around a 1ml pipette or pelleted by centrifugation at 3500g for 15'. After rinsing with 70% ethanol, the DNA was air-dried and then resuspended in a total of 20ml TE by rocking at room temperature overnight.

The DNA quality and quantity was checked by spectrophotometry at 260nm and 280nm. If the OD 260:280 ratio was <1.6, then the DNA was re-extracted with PCI and chloroform before continuing. If the ratio was >1.8, then the RNase treatment was repeated.

2.5.1.4 Shearing DNA by autoclaving

The DNA was pooled in a sterile autoclave-resistant tube and then autoclaved for 1' at 121°C, 103KPa. The sheared DNA was transferred to a 50ml Falcon tube and precipitated with 1/20 volume 4M NaCl and 0.6 volume isopropanol. The DNA was pelleted by centrifugation at 3500g for 15', the pellet rinsed with 70% ethanol and centrifuged for a further 5' at 3500g. The supernatant was decanted and the pellet was

air-dried before being resuspended in a total of 5ml TE by rocking overnight at room temperature.

The amount of shearing was determined by loading 1µl sheared DNA on a 1% agarose gel in 1xTAE and electrophoresing at 100V for 1h. The gel was stained with 0.5µg/ml ethidium bromide for 5'. The majority of fragments should range from 100bp to ~1kb.

2.5.1.5 S1 nuclease treatment of sheared DNA

The volume of the sheared DNA was adjusted to 5ml with TE and then 0.5ml 20xSSC was added. The DNA was denatured in a boiling water bath for 10' and then chilled on ice for 2'. The tubes were incubated at 68°C for 10' to allow the repeats to hybridise and then chilled on ice for 2'. Next, 550µl S1 nuclease buffer (Life Technologies), 83µl 3M NaCl (50mM final concentration) and 15,000U S1 nuclease (Life Technologies) was added to the tube and incubated at 37°C for 60'. The DNA was extracted once with an equal volume of PCI and centrifuged at 3500g for 10'. The upper aqueous phase was transferred to a new tube and an equal volume of chloroform was added. After mixing thoroughly, the tube was again centrifuged at 3500g for 10' and the aqueous phase was recovered. The DNA was precipitated as previously with 1/20 volume 4M NaCl and 0.6 volume isopropanol. After pelleting and rinsing the CotI, DNA it was resuspended in 1ml TE. The DNA was quantitated by spectrophotometry at 260nm and 280nm and 1µl was electrophoresed on a 1% agarose gel at 100V for 1h and stained with 0.5µg/ml for 10'.

2.5.2 Biotinylation of BAC DNA

Biotin-labelled dATP was incorporated into 1µg BAC DNA by nick translation using the reagents from a Bionick labelling system (*Life Technologies*). A trace amount of tritium-labelled dATP was also included in the reaction so that the tubes containing

the labelled DNA could be established by scintillation counting. In each labelling reaction, there was 1μCi ^{3}H-dATP (*Amersham*), 20μM dCTP, 20μM dGTP, 20μM dTTP, 10μM dATP, 10μM biotin-14-dATP, 50mM Tris-HCl (pH 7.8), 5mM $MgCl_2$, 10mM β-mercaptoethanol, 10μg/ml nuclease-free BSA, 1μg BAC DNA, 2.5U DNA polymerase I and 0.0375U DNase I. The reagents were mixed well and the reaction incubated at 16°C for 1h before 5μl stop buffer was added.

Whilst the reaction was proceeding, a 5cm Sephadex G50 column (*Pharmacia*) was prepared in a short pasteur pipette and the column was washed twice with TE. A rack of 15 eppendorf tubes was prepared and the first eppendorf tube was placed beneath the column. Once the column was dry, the 55μl nick reaction was layered on the top followed by 455μl TE. Once the total volume was collected, the second eppendorf was placed below the column and 100μl TE was layered on top. Fractions 2-10 were all 100μl and fractions 11-15 were 500μl to ensure that all the unincorporated nucleotides were recovered from the Sephadex.

The amount of tritium in each fraction is an indication of the amount of biotin in each fraction. Consequently, 2μl aliquots of each fraction were mixed thoroughly with 150μl ddH_20 and 1.5ml scintillation fluid in 3ml opaque scintillation tubes (*Beckman*). The amount of tritium in each tube was established using a Beckman LS3801 β-counter. The tubes in the initial 10 fractions that had the most counts contained the labelled BAC DNA and these fractions were pooled for use in FISH. The amount of incorporation was calculated based on the counts.

2.5.3 Fluorescence *in situ* hybridisation

In a total volume of 100μl, 100ng biotinylated BAC DNA, 8μg CotI DNA from the species being mapped and 50μg salmon sperm DNA were mixed and precipitated with 5μl 4M NaCl and 60μl isopropanol at -80°C for 20'. The DNA was pelleted by centrifugation at 18,000g for 10' and rinsed with 70% ethanol. The pellet was air-dried

and the DNA resuspended in 20µl hybridisation solution (10% dextran sulphate (*Pharmacia*) in 50% formamide, 2 x SSC) using a cut tip. The DNA was either resuspended at 4°C overnight or at 37°C for several hours and mixed thoroughly by vortexing at regular intervals.

The DNA was denatured at 72°C for 5' and then snap chilled on ice for 2' before incubating at 37°C for 60' to pre-anneal any repetitive DNA in the probe to the CotI DNA, thereby suppressing the repeats prior to hybridisation to the chromosomes.

Slides of metaphase chromosomes (section 2.1.2.6) were pre-heated to 72°C in an empty coplin jar and then denatured in 70% formamide/2xSSC at 72°C for 2'. The slides were then dehydrated through an ethanol series (ice-cold 70%, 70%, 95%, 100%) and air-dried.

The probe mix was placed on the slide with a cut tip beneath a 22x22mm coverslip and sealed with rubber cement. Two probes were used per slide since there were two drops of cells on each slide and generally 5 slides were processed at once. The slides were incubated in a humid chamber at 37°C overnight.

The coverslips were removed and the slides were washed 3x5' in 50% formamide/2xSSC at 40°C and 3x5' in 2xSSC at 40°C. The slides were then transferred to 4xSSC/0.05% Tween20 (*Sigma*) at room temperature for 3'. After this wash, the slides were removed one at a time, tapped dry and 100µl 4xSSC/1% BSA was placed on each slide beneath a 24x50mm coverslip and incubated at 37°C for 30' in a humid chamber. Next the slides were incubated at 37°C for 30' with 60µl of a 1/200 dilution of FITC-avidin (*Vector Laboratories*) in 4xSSC/1% BSA. After this incubation, the slides were washed 3x5' in 4xSSC/0.05% Tween20 at 40°C. A 1/100 dilution of biotinylated anti-avidin (*Vector Laboratories*) in 4xSSC/1% BSA was prepared and each slide was incubated with 60µl of this solution at 37°C for 30'. The slides were washed 3x5' in 4xSSC/0.05% Tween20 at 40°C. The incubation with FITC-avidin and the washes in 4xSSC/0.05% Tween20 were repeated before the slides were stained in

5µg/ml propidium iodide in PBS for 5'. The slides were washed 3x20" in PBS, tapped dry and 100µl PPD11 antifade mountant was placed on each slide and sealed with nail varnish beneath a 24 x 60mm coverslip.

2.5.4 Dual-FISH

To confirm that the BAC clones for microsatellite markers which were selected to be on the same chromosome from the sheep genetic linkage map were physically on the same chromosome, dual-FISH was performed. Dual-FISH was also used to confirm the location of previously unmapped genes by hybridisation with a BAC for the chromosome to which the gene had been assigned by karyotyping. The procedure was identical to a single localisation except that both probes were included in the hybridisation mix along with sufficient CotI to suppress each probe. For instance, if each probe required 8µg CotI for suppression in separate localisations, then in a dual localisation 16µg CotI was needed.

2.5.5 Fluorescence microscopy and photography

Slides were viewed with an Olympus BX70 microscope. Chromosomes stained with propidium iodide were visualised under green epifluorescence. Fluorescein was detected with blue epifluorescence combined with dull fluorescence of the chromosomes.

Colour photographs were taken on Fugichrome ASA 400 colour positive film which was programmed to ASA 1600 for dark field with an additional correction of 1.6 (ASA 2560).

Black and white photographs for karyotyping were taken on KODAK technical pan film which was programmed to ASA 800. For photographs taken with green epifluorescence, an additional correction of 2.0 (ASA 1600) was used. The negatives were developed in HC110 developer dilution B for 11' at 20°C and fixed for two

minutes with Hypam fixer containing hardener (*Ilford*). The negatives were printed on glossy multigrade IV paper (*Ilford*) at a contrast level of one. The prints were developed for 90" in Ilford paper developer (diluted 1:14) and fixed for 60" in Hypam fix without hardener (diluted 1:9).

2.6 Construction of a BAC restriction map for keratin associated protein genes

A BAC clone was isolated from the library that contained the gene for keratin associated protein 1.1 (KAP1.1, Powell *et al.*, 1983). A large-scale plasmid preparation was prepared by standard alkaline lysis method (Sambrook *et al.*, 1989) to generate sufficient DNA for plasmid end rescue (section 2.6.1) and restriction mapping (section 2.6.2). The BAC DNA was treated with 10µg/ml DNase free RNase, extracted once with 25:24:1 phenol:chloroform:isoamyl alcohol, once with chloroform, precipitated with 0.6 volumes isopropanol and resuspended in 1ml TE.

2.6.1 Plasmid end rescue

Each end of the BAC insert was isolated by plasmid end rescue (Figure 2.3) as described in Cai *et al.* (1995). Separate 250ng aliquots of the BAC clone were digested to completion with 10U *SacI, BamHI, SphI* and *EcoRI* in 25µl reactions at 37°C for 5h. The digested products were self-ligated under dilute conditions (Collins and Weissman, 1984) in a 40µl reaction with 10U T4 DNA Ligase (*Promega*) and 32ng digested DNA at 16°C overnight. The right-hand end of the BAC was recovered by transformation into electrocompetent *E. coli* (section 2.6.1.2) and the left-hand end was recovered by inverse PCR (section 2.6.1.3).

2.6.1.1 Preparation of electrocompetent *E. coli*

A -80°C glycerol stock of DH10β *E. coli* was used to inoculate 50ml SOB and the culture was incubated with shaking at 150rpm at 37°C overnight. A 1ml aliquot of SOB

was saved as a blank for spectrophotometry and the remaining 749ml was inoculated with 0.75ml from the overnight culture. The cells were incubated at 37°C until they reached mid-log phase which was equivalent to an OD_{550nm} of 0.5-0.6.

The cells were transferred to sterile Nalgene centrifuge tubes and pelleted at 3000g for 10' in the JA10 fixed angle rotor of a Beckman J2-HS centrifuge. The supernatant was decanted and the cells were resuspended in 250ml 10% glycerol by gently swirling the cells in a small volume of glycerol and then adding the remainder. The cells were pelleted again by centrifugation at 3000g and the wash in 10% glycerol repeated. After the cells were pelleted by centrifugation, the supernatant was decanted and the cells resuspended in the glycerol remaining in the bottom of the tubes. The cells were transferred to a single 50ml Falcon tube and pelleted by centrifugation at 3500g for 10' in the fixed angle rotor of a Sorvall RT6000D centrifuge. Finally, the cells were resuspended in 3ml 10% glycerol, split into 125µl aliquots and frozen at -80°C.

2.6.1.2 Transformation of plasmid end rescued products

A 2µl aliquot of each of the ligated products that were generated following digestion with *SacI, BamHI* and *SphI* was transformed into 40µl electrocompetent DH5-α using a 0.2cm cuvette in a *Biorad* gene pulser attached to a *Biorad* pulse controller set to 2.5KV, 25µFD and 200 Ohms. The cells were incubated in 1ml SOC medium at 37°C for 60' and then 100µl was plated on LB agar containing 12.5µg/ml chloramphenicol. The plates were incubated overnight at 37°C. White positive transformants were selected and cultured in 5ml LB containing 12.5µg/ml chloramphenicol overnight at 37°C. The BAC DNA was extracted by a standard alkaline lysis mini-preparation (section 2.3.7). The DNA was linearised by digestion with the same enzyme that was used for sub-cloning, and the smallest insert was selected for sequencing following electrophoresis on 1% agarose. The BAC DNA was

sequenced manually using the M13 reverse primer (5' CAC ACA GGA AAC AGC TAT GAC C 3') in a cycle sequencing reaction as described (section 2.4.3).

2.6.1.3 Inverse PCR

An aliquot of each of the ligated products that were generated following digestion with *SacI* and *EcoRI* was used as the template for inverse PCR. IPCR1 (5' GGA TCC CCG GGT ACC GAG 3') and IPCR2 (5' CTA GAG TCG ACC TGC AGG 3') were designed from pBeloBAC11 (Cai *et al.*, 1995) and were immediately adjacent to the *HindIII* cloning site used to generate the BAC library. Each 100µl reaction consisted of 20mM Tris-HCl (pH 8.4), 50mM KCl, 125µM dGTP, 125µM dATP, 125µM dTTP, 125µM dCTP (*Life Technologies*), 1.5mM $MgCl_2$, 25pmol IPCR1, 25pmol IPCR2, 2µl ligated product and 1U *Taq* DNA polymerase (*Life Technologies*). The products were amplified for 35 cycles (94°C 1', 56°C 1', 72°C 1') with a final extension at 72°C for 10'. Products were sequenced directly with IPCR1 and IPCR2 as described (section 2.4.3).

2.6.2 Large-scale Restriction mapping

The restriction enzymes *NotI, XhoI, ClaI, SfiI* and *NruI* were used individually and in pairs to digest 20µl aliquots of the large-scale preparation for the BAC clone containing the gene for KAP1.1. Each 30µl reaction consisting of 25mM Tris-Acetate (pH 7.8), 100mM potassium acetate, 10mM magnesium acetate, 4mM spermidine, 10U each enzyme and the BAC DNA was incubated at 37°C for 5h. The products were electrophoresed on a 1% agarose CHEF gel using the conditions to separate fragments from 25-100kb (section 2.4.1).

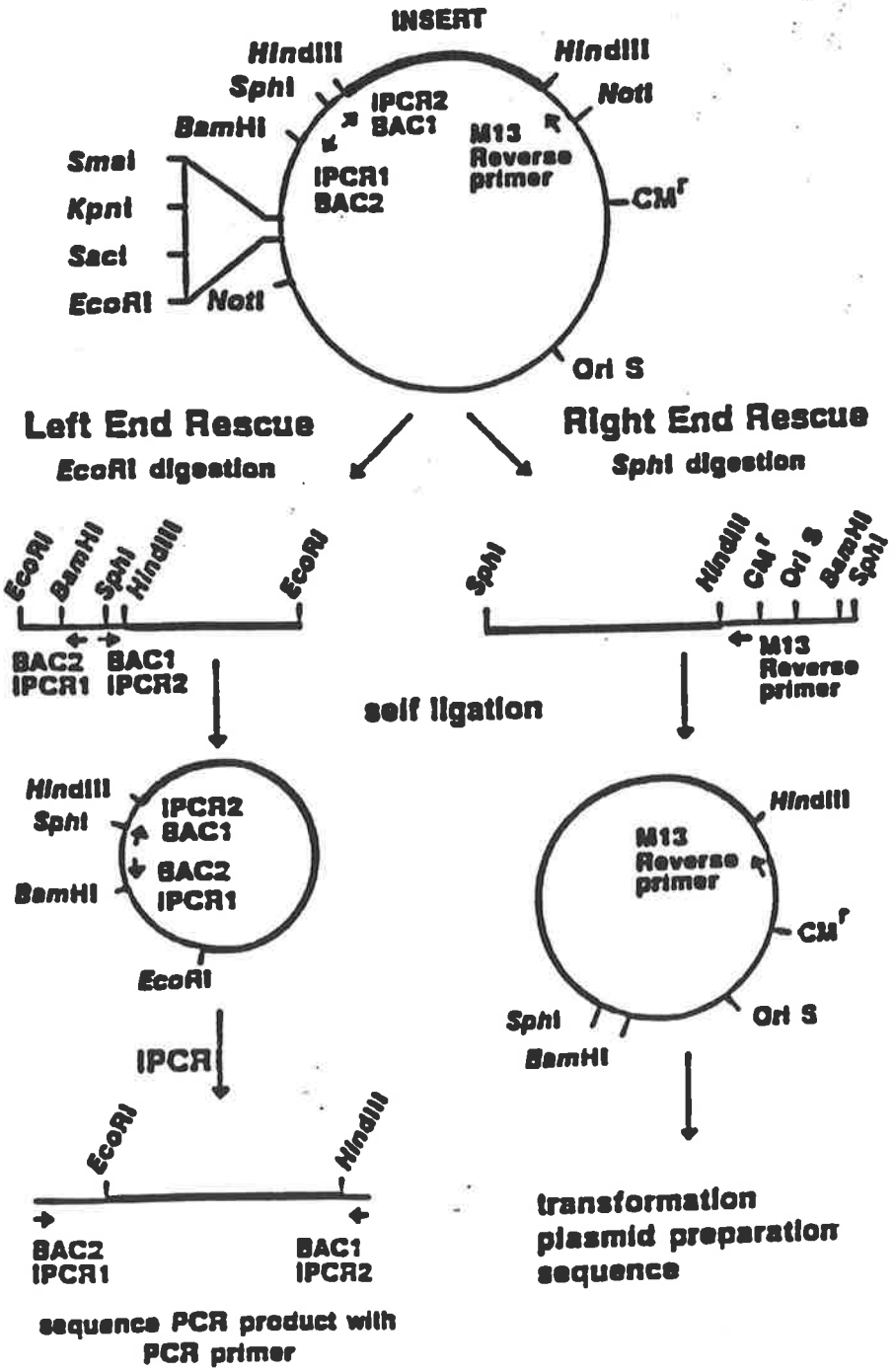


Figure 2.3: A strategy for BAC end rescue. BAC DNA is digested and self-ligated to create subclones containing some vector plus one end of the insert. The left end of a BAC clone is isolated by inverse PCR and then the product is sequenced with internal primers. The right end is recovered by transformation and this subclone is sequenced with M13R. (From Cai *et al.*, 1995).

The gel was rocked in denaturing solution for 15' and then in neutralising solution for 15'. The DNA was transferred to a Hybond-N nylon membrane (*Amersham*) in 20xSSC (Southern, 1975). Capillary transfer of the DNA onto the membrane was allowed to proceed for 24h. The membrane was rinsed briefly in 2xSSC, blotted dry and then the DNA was fixed by exposure to UV light (312nm) for 1'.

2.6.2.1 Radiolabelling probe DNA

The Southern blot of the BAC for KAP1.1 was probed with various other KAP genes and repetitive elements to establish how many other genes and markers were present in the BAC clone. A $(GT)_{30}$ oligonucleotide was end-labelled with γ^{32}P-dCTP as for sequencing (section 2.4.3.2). Plasmid DNA probes for keratin associated protein genes (donated by Dr. Barry Powell) and for the BovA Alu-like element (courtesy of Mrs. Jan Cook) were labelled with α^{32}P-dATP using a nick translation kit (*Promega*). Each 50µl nick consisted of 10µM dCTP, 10µM dGTP, 10µM dTTP, 50mM Tris-HCl (pH 7.2), 10mM MgSO4, 100µM DTT, 1µg plasmid DNA, 1.85MBq α32P-dATP, 5U DNA polymerase I and 20pg DNase I. The nicks were incubated at 15°C for 60' and then the reaction was stopped with 25mM EDTA. The DNA was denatured by incubating with 10µl 500mM NaOH for 5' and then neutralised with 10µl 500mM HCl immediately before the probe was added to the hybridisation solution (section 2.7.3).

2.6.2.2 Probe hybridisation

The membrane was pre-hybridised in 20ml hybridisation solution at 50°C for 1h prior to the addition of the radiolabelled probe which was incubated with the membrane at 50°C overnight. The membrane was then washed 4 x 15' with 2xSSC/0.1% SDS and 1 x 20' with 0.5% SSC/0.1% SDS at 65°C. The membrane was wrapped in plastic and exposed to X-ray film for 1-3 days. The same membrane was probed multiple times by

removing the old probe in 0.1% SSC/0.1% SDS at 90°C for 15' prior to pre-hybridisation at 50°C.

2.6.2.3 Shot-gun sequencing

To further characterise the BAC clone containing KAP1.1, shot-gun sequencing was performed. Libraries of BAC subclones were prepared by digesting 1µg aliquots of BAC DNA with 40U *HindIII*, *Sau3AI*, *Tsp509I,* or *PstI* at 37°C for 2h. The digested DNA was ligated to *HindIII*, *BamHI, EcoRI*, or *PstI* cut pBluescript, respectively. Each 25µl ligation mixture was incubated at room temperature for 2h and then 1µl aliquots were transformed into 20µl DH10β *E. coli* by electroporation. A 1:500 dilution of the tranformation was plated onto LB agar containing 50µg/ml ampicillin, 50µg/ml IPTG and 40µg/ml X-gal. After incubation at 37°C overnight, white colonies were picked into 96-well plates, with 100µl LB containing 50µg/ml ampicillin in each well.

For the four-base cutters (*Sau3AI* and *Tsp509I*), the inserts were amplified with M13F and M13R in 50µl PCR reactions with an annealing temperature of 50°C. Each reaction was inoculated directly with 1µl cells diluted 1:100 in water. All products >500bp were purified for sequencing using a QIAGEN PCR purification kit according to the manufacturer's instructions.

For the six-base cutters (*HindIII* and *PstI*), DNA from each sub-clone was prepared using an Autogen robot. The DNA was digested with *HindIII* or *PstI* to check for the presence of empty clones. Every clone containing an insert was sequenced.

The ABI™ BigDye™ terminator cycle sequencing ready reaction kit with AmpliTaq DNA polymerase, FS (*Perkin Elmer*) was used to sequence 5.5µl PCR product or 2.5µl Autogen DNA with M13F or M13R primers in 10µl reactions with 4µl BigDye™ mix. The sequencing reactions were performed on a GeneAmp PCR System 9700 (*Perkin Elmer*) over 25 cycles (96°C 10", 50°C 5", 60°C 4'). Sequences were purified on Centri-sep™ sephadex columns (*Princeton Separations*), vacuum dried and

then resuspended in 1.5µl gel loading buffer (5:1 deionized formamide: 25mM EDTA with 50mg/ml blue dextran). Samples were denatured at 96°C for 2' and then loaded on a 0.2mm 4.25% Seaquate^{R} polyacrylamide gel (*Sooner Scientific*) with 48 lane sharks-tooth comb in an ABI377 automated sequencer (*Perkin Elmer*). Data were collected for 7h using run module "Seq Run 36E 2400". The gels were tracked and then the chromatograms were analysed and assembled into contigs in Sequencher 3.1.1 (*Gene Codes Corp.*).

CHAPTER 3

SUPER-FAST PRIMED *IN SITU* TARGETING

SUPER-FAST PRIMED *IN SITU* TARGETING

3.1 Introduction

The ovine genetic linkage map now consists of 519 markers with an average spacing of ~6.4cM between adjacent linked markers (de Gortari *et al.*, 1998), but the physical map is less well developed. Only 91 loci have been physically assigned to specific chromosomal regions by *in situ* hybridisation in sheep and just 25 of these have been mapped by genetic linkage analysis as well (ArkDB, AgResearch 6/8/99). These 25 loci act as anchors for 17 chromosomes meaning that there are still 11 chromosomes that have not been anchored directly, although the genetic linkage maps for every sheep chromosome have been oriented and anchored indirectly through the homologous bovine chromosomes (Kappes *et al.*, 1997).

In order to be able to use comparative mapping data from the extensively developed human and mouse genomic maps to select candidate genes to positionally clone economically important genes in livestock species, the ovine physical map needs to be developed further. The United States Department of Agriculture distributes PCR primers for genetic linkage analysis in sheep. It would be advantageous to develop a technique that would allow these PCR primers, or the product generated by these primers, to be used for physical mapping to rapidly increase the number of linked markers on the ovine physical map. The use of a non-radioisotopic hybridisation method (oligonucleotide primed *in situ* DNA synthesis or PRINS) was investigated because the technique potentially yields experimental results more rapidly than either traditional radioactive methods or fluorescence *in situ* hybridisation (FISH).

FISH is the hybridisation of biotin-labelled (Brigati *et al.*, 1983) or digoxygenin-labelled (de Frotos *et al.*, 1989) probes to metaphase chromosomes whereas PRINS is a rapid detection method based on the sequence specific annealing of unlabelled oligonucleotides to chromosomes *in situ*. These oligonucleotides are primers for chain

elongation catalysed by thermostable polymerases (eg. *Taq*) using the target chromosomal DNA as the template (Koch *et al.*, 1989). The new chain is labelled by the incorporation of either a biotin-labelled or digoxygenin-labelled nucleotide as a substrate for chain elongation. Since the priming oligonucleotide is not labelled, high concentrations of the primer can be used to allow rapid hybridisation without the production of extensive background.

Physical mapping by FISH generally requires long probes (>3kb) and plasmid-derived sequences are often insufficient because the target is too small and the intensity of the fluorescent signal is dependent on probe length. Differences in signal strength after PRINS though are only due to differences in the number and organisation of target sequences because short oligonucleotide primers give rise to as much chain elongation as longer primers (Hindkjær *et al.*, 1994).

In this chapter, a modification of the conventional PRINS reaction is presented. The super-fast primed *in situ* targeting (SPRINT) reaction uses a fluorescein-labelled nucleotide to label the newly synthesised strand so that it can be visualised directly (Koch *et al.*, 1992). The microsatellite core sequence, $(AAGGT)_5$, of human satellite III, and restriction endonuclease digested fragments of sheep satellite I were used as primers for SPRINT to demonstrate the efficacy of this physical mapping technique for repetitive sequences. The limitations of using this technique for single-copy sequences are discussed.

3.2 Results

3.2.1 Optimisation of white blood lymphocyte culture for metaphase spreads

Prior to the physical mapping experiments it was necessary to optimise the production of good quality metaphase chromosome spreads. A metaphase spread deemed suitable for physical mapping by *in situ* hybridisation was one with elongated

("stringy") chromosomes that did not overlap and produced G-bands or R-bands when treated appropriately.

The size of the pellets recovered during harvesting was an indication of how well each of the cultures grew. Cultures inoculated with Ficoll-hypaque enriched lymphocytes (section 2.1.1.1) grew better than those cultures inoculated with lymphocytes enriched by ammonium chloride treatment (section 2.1.1.2) or cultures inoculated with white blood cells from the buffy coat at the plasma-blood interface (section 2.1.1.3). Interestingly, cultures inoculated with whole blood grew as well as the Ficoll-hypaque enriched lymphocytes.

The best growth medium was RPMI supplemented with 15% FCS, 1% glutamine, 1.8% PHA and 1% penicillin-streptomycin-fungizone solution. In this medium, 8% of non-synchronised cells were in metaphase. Arrest with BrdU alone also resulted in 8% metaphase cells, but a mitotic index of 15-20% was routinely achieved by arresting the cells with thymidine and releasing them with BrdU. However, in all cases (ie. non-synchronised and synchronised), ~95% of the metaphase cells were not suitable for physical mapping because they were too contracted, and consequently, the chromosomes were poorly banded. This is likely to be due to the length (1.5h) of the colchicine treatment because decreasing the concentration of colchicine did not affect the severity of chromosome contraction. When 200-600ng/ml colchicine was added for only 12', 30-45% of the metaphase cells contained well-spread, elongated chromosomes suitable for physical mapping. When <200ng/ml colchicine was used, the chromosomes in some cells were still attached to the spindle fibres.

Elongated, R-banded chromosomes for SPRINT were produced by inoculating supplemented RPMI with whole blood, arresting the cells at mid-S phase with thymidine, releasing them with BrdU, and then inhibiting microtubule formation with 600ng/ml colchicine for 12'. Only 5-7% of these cells were suitable for physical mapping.

3.2.2 Optimisation of SPRINT signal

A human satellite III probe has been shown to hybridise strongly to the q12 bands of chromosomes 1, 9, 16 and Y, moderately to the p13 bands of chromosomes 13, 14, 15, 21 and 22 and weakly to the centromeres of 5, 7, 10, 17, 20 and Y by traditional *in situ* hybridisation techniques (Gosden *et al.*, 1975; Grady *et al.*, 1992). Satellite III was chosen to optimise the SPRINT signal because of the three levels of signal intensity that have previously been observed. Under optimum conditions SPRINT signal should be observed on all target chromosomes.

Three *Taq* DNA polymerases (*Perkin Elmer, Bresatec Ltd.* and *Boehringer Mannheim*) were tested and it was found that all three enzymes were able to produce strong SPRINT signal for human satellite III provided that at least 2.5 units of enzyme were used. The sensitivity of the SPRINT reaction for each of the enzymes was dependent on the Mg^{2+} concentration. The optimum Mg^{2+} concentration was determined empirically by amplification of satellite III in a standard PCR. It was found that the different Taq DNA polymerases required different salt concentrations to produce the same results. The surface area to volume ratio of a tube differs vastly from that created by a coverslip on a slide. Consequently, the conditions had to be checked on the slides and were found to match the results of PCR. *Taq* polymerase from *Bresatec Ltd.* produced the strongest SPRINT signal for human satellite III at 4.5mM Mg^{2+} (Table 3.1) and both *Amplitaq* from *Perkin Elmer* and *Taq* from *Boehinger Mannheim* gave a strong signal at 1.5mM Mg^{2+}.

For the *Bresatec Ltd.* Taq polymerase, maximum sensitivity and specificity was achieved when the reaction was performed at 55°C. When higher annealing temperatures were used (60°C, 65°C), the intensity of the SPRINT signal decreased and some centromeres, presumably the shorter regions of satellite III, no longer produced any signal (Table 3.1). The age of the slides had no significant effect on the success of

the SPRINT reaction, with old slides producing as much signal as freshly prepared slides.

The minimum amount of each unlabelled dNTP required for SPRINT was found to be 50μM. Several ratios of dTTP to fluorescein-12-dUTP were tested (1:1, 2:1, 5:1 and 10:1). All of these ratios produced specific SPRINT signal, but since fluorescein-12-dUTP was the most expensive component of the reaction, the ratio that required the least amount of fluorescein (10:1) was selected (Table 3.1). It was also possible to generate specific SPRINT signal with just 5μM fluorescein-12-dUTP and no competitor dTTP.

It was found that 2.5μg of the probe $(AAGGT)_5$ was the minimum amount of primer needed to generate SPRINT signal on all target chromosomes (Table 3.1). However, it is predicted that this amount would need to be optimised for each new oligonucleotide used.

3.2.3 Localisation of human satellite III on G-banded chromosomes

The addition of 5-bromodeoxyuridine to the cultured lymphocytes enabled the metaphase chromosomes to be banded using the alkaline PPD11 antifade mountant. The G-bands or R-bands and fluorescent SPRINT signal were visualised simultaneously under blue epifluorescence (Figure 3.1). Human satellite III was localised strongly to the variable bands 1q12, 9q12, 16q12 and Yq12, to a lesser extent to bands p13 of chromosomes 13, 14, 15, 21 and 22 and weakly to the centromeres of chromosomes 5, 7, 10, 17, 20 and Y. These results are consistent with prior reports using traditional *in situ* methods (Gosden *et al.*, 1975; Grady *et al.*, 1992).

Table 3.1: Optimisation of SPRINT conditions. SPRINT conditions were optimised as described in section 2.2 using Taq polymerase from *Bresatec Ltd.* Mg^{2+} concentration was optimised first, followed by temperature, dNTP ratio, dTTP:dUTP and finally primer concentration.

Variable	Conditions			
Mg^{2+} (mM)	1.5	2.5	4.5	7.5
	1, 9[a]	1, 9, 16, Y[b]	All target[c]	1, 9, 16, Y
Temp (°C)	50	55	60	65
	All target	All target	1, 9, 16, Y	1, 9
dNTPs (μM)	20	50	100	200
	No signal	All target	All target	All target
dTTP:dUTP	1:1	2:1	5:1	10:1
	All target	All target	All target	All target
Primer (ng)	500	1000	2500	5000
	No signal	1, 9	All target	All target

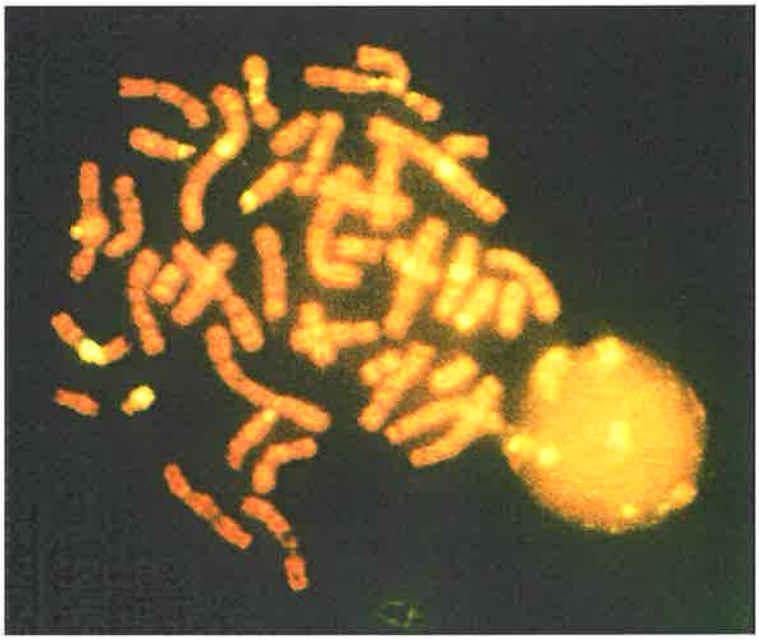
[a]1, 9 = SPRINT signal on the q12 bands of 1 and 9.

[b]1, 9, 16, Y = SPRINT signal on the q12 bands of chromosomes 1, 9, 16 and Y.

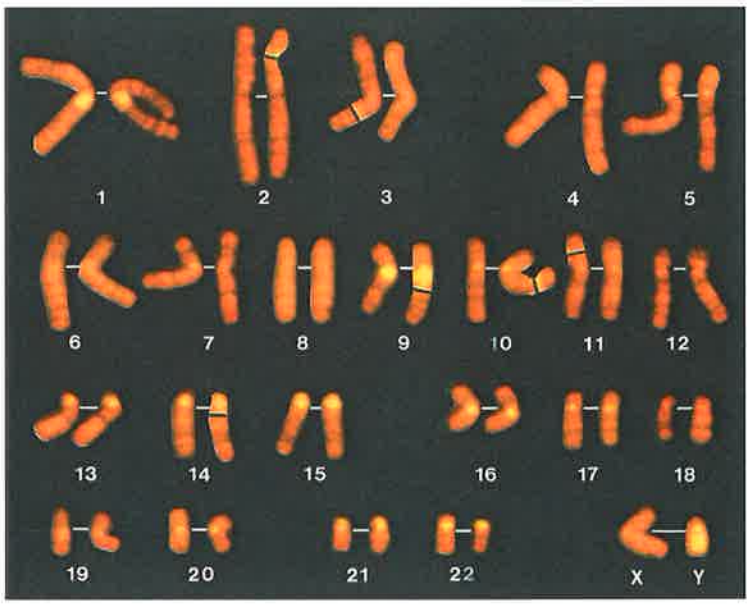
[c]All target = Strong SPRINT signal on the q12 bands of chromosomes 1,9, 16 and Y, moderate signal on the p13 bands of chromosomes 13, 14, 15, 21 and 22, and weak signal at the centromeres of 5, 7, 10, 17, 20 and Y.

Figure 3.1: Localisation of human satellite III by SPRINT on metaphase chromosomes. (A) Satellite III located on R-banded metaphase chromosomes of a human male. (B) Karyogram of a PPD11 G-banded human late prophase viewed under blue epifluorescence. G-positive chromosomal arms are stained bright orange with propidium iodide. Satellite sites are labelled by SPRINT with fluorescein and show yellow. The position of the centromeres is marked by a white line between the chromosome pairs. The positions of overlaps between the chromosome arms are marked with green lines. The metaphase spread for the karyogram was prepared in 1988 and the slides were stored under dessicating conditions at -20°C. SPRINT conditions were as described in section 2.2.

A



B



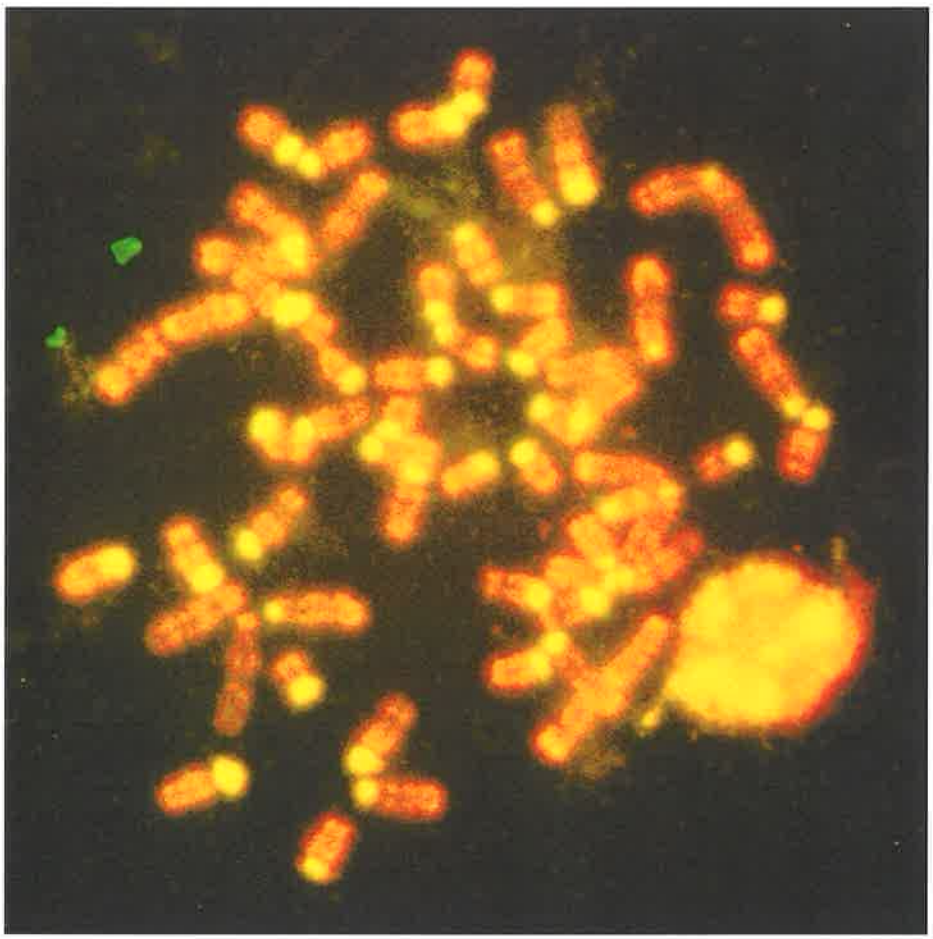


Figure 3.2: Localisation of ovine satellite I by SPRINT on metaphase chromosomes. G-positive chromosomal arms are stained bright orange with propidium iodide. Satellite sites are labelled by SPRINT with fluorescein and show yellow. SPRINT conditions were as described in section 2.2.

3.2.4 Localisation of sheep satellite I by SPRINT

When 2.5 - 10μg of the undigested ovine satellite I PCR product was used as a primer for SPRINT, no signal could be detected. The PCR product was digested with *Sau3AI* to increase the effective primer concentration and specific SPRINT signal was found in regions near the centromeres of all autosomes (Figure 3.2). Chromosome 1, which is the largest of the naturally-fused metacentric chromosomes, was only lightly labelled and the sex chromosomes were unlabelled. These results are consistent with previous reports using *in situ* hybridisation methods (Novak, 1984; Burkin *et al.*, 1996).

3.2.5 Attempted localisation of single copy sequences by SPRINT

Various primers for microsatellites and genes were tested at increasing concentrations and at different annealing temperatures in an attempt to detect single copy sequences by SPRINT (Table 3.2). No consistent signal was detected on metaphase chromosomes and was rarely visible in interphase cells (<5%) by direct incorporation of fluorescein. Incorporation of biotin by PRINS was also tried with both a single and double amplification using FITC-avidin and biotinylated goat anti-avidin. The fluorescent background increased with successive rounds of amplification but no signal was discernable. PCR products for each of the microsatellites and genes were cloned by blunt end ligation into *SmaI* cut pBluescript so that large amounts of the product (1-50μg) could be used as a SPRINT primer. The whole clones, excised inserts and restriction digested fragments of the clones were tested as SPRINT primers, but no signal was detected above the high generalised background that was associated with using the cloned products.

Table 3.2: Single copy sequences used in SPRINT reactions. Oligonucleotide primers, cloned PCR product generated by those primers and restriction digested fragments of the cloned product were all tested as sources of primers for SPRINT. Various concentrations of each source of primer were tried at each annealing temperature (TA).

Locus	SPRINT Primer Type	[Primer]	T_A
ILSTS005	Oligonucleotide primers Cloned PCR product	1µg 5µg 10µg 50µg	50 55 60 70
MAF35	Oligonucleotide primers Cloned PCR product	1µg 5µg 10µg 50µg	50 55 60 70
BM1824	Oligonucleotide primers Cloned PCR product *DdeI* digested fragments	1µg 5µg 10µg 50µg	50 55 60 70
FSHB	Oligonucleotide primers Cloned PCR product *DdeI* digested fragments	1µg 5µg 10µg 50µg	50 55 60 70
ADCYC	Oligonucleotide primers Cloned PCR product	1µg 5µg 10µg 50µg	50 55 60 70
OPCML (OCAM)	Oligonucleotide Cloned PCR product	1µg 5µg 10µg 50µg	50 55 60 70
KAP1.1	*AluI*, *Sau3AI*, *Tru91* digested genomic clone	1µg 5µg 10µg	50 55 60 70

3.3 Discussion

Probes which target repetitive sequences, like human satellite III and ovine satellite I, have been localised to metaphase chromosomes by FISH (Grady *et al.*, 1992) and PRINS using biotin-labelled or digoxygenin-labelled probes or nucleotides (Gosden and Lawson, 1994). The hybridised FISH probe or the newly synthesised DNA strand generated by PRINS is detected using fluorescein-conjugated antibodies. For a large-scale physical mapping project though, the long hybridisation time of several hours required for FISH limits the number of markers that can be processed. Moreover, the immunochemistry involved in the detection of biotin-labelled or digoxygenin-labelled probes is laborious and expensive.

The extent of human satellite III and ovine satellite I detected on chromosomes by SPRINT using a fluorescein-labelled nucleotide was comparable to that produced by PRINS and FISH but with the background fluorescence much diminished. The components of the reaction mixture have been optimised so that signal can be detected on all target chromosomes without the need for further signal amplification using antibodies or by using cycles of denaturation, annealing and extension (cycling PRINS; Gosden and Hanratty, 1993). The strong, specific labelling obtained in 20 minutes illustrates that the SPRINT technique is both fast and sensitive, and is ideal for rapid physical mapping of repetitive sequences. The main disadvantage of SPRINT is that the thermal kinetics of the reaction are not easily predicted, so that the procedure needs to be optimised for each new probe sequence used. The requirement for very high primer concentrations and high *Taq* concentrations suggests that the organisation or spatial arrangement of the chromatin must somehow limit the SPRINT reaction.

All attempts to physically map single copy sequences by SPRINT were unsuccessful. Numerous sets of primers for microsatellite sequences and genes were used in SPRINT reactions. Various primer concentrations (1-50μg), annealing temperatures (50-70°C) and cycling profiles were tested, but no fluorescent signal could

be visualised by standard fluorescence microscopy on metaphase chromosomes or in interphase nuclei. Biotin-labelled nucleotides were also used in conventional PRINS reactions, but the background was too high to discern any specific signal. On several occasions, upon completion of the SPRINT reaction, the reagent mixture was pipetted onto agarose gels to assess whether diffusion of the product away from the chromosomes was responsible for the lack of signal. Only unincorporated nucleotides were detected on these gels (data not shown). To evaluate whether PCR amplicons are being generated in the SPRINT reaction it should be possible to denature the labelled products away from the chromosomes and detect them by gel electrophoresis. For improved sensitivity, radioactive nucleotides could be used, but such experiments were not attempted herein.

Since the PRINS technique was developed (Koch *et al.*, 1989), there has only been one report describing the localisation of a single copy gene by PRINS (Cinti *et al.*, 1993), and the signal was detected using a Confocal Laser Scanning Microscope (CLSM). The human factor IX gene was localised to the X chromosome by CLSM, but was barely detectable by standard fluorescence microscopy. CLSM allows long exposures, the signal-to-noise ratio can be manipulated and the signal can be enhanced. Since the factor IX gene had previously been physically mapped to Xq27 (Nguyen *et al.*, 1987), it may have been less difficult to detect the signal than it would have been for a previously unmapped gene. There does not appear to be anything unusual about either the organisation of the X chromosome or the factor IX gene that would enable it to be more readily detected by PRINS than other single copy genes.

Porcine α interferon has been localised to chromosome 1 by DISC-PCR (direct *in situ* single copy PCR) which is essentially cycling PRINS except that a colorimetric reaction with horseradish peroxidase was used instead of detection with fluorescently labelled nucleotides (Troyer *et al.*, 1994). Another innovation that was used was a thermocoupler attached to a glass slide so that the temperature on the reaction surface

was monitored directly. This method of controlling the thermal cycles is probably more precise than the "simulated" cycle control used for the Hybaid Omnigene slide block.

As for the human factor IX gene, porcine α interferon had previously been physically mapped to 1q2.5 (Yerle and Gellin, 1989; Sarmiento *et al.,* 1993) which is an easily distinguished chromosome. Out of 290 spreads that were analysed a total of 59 signals were detected. Only 21 were localized to a position on chromosome 1 and just 10% of these (ie. 2 signals) were on both chromatids. Given that only 1% of all the spreads had signal on both chromatids, it would be very difficult to assign a previously unmapped gene to a particular chromosome by this method.

It seems likely that single-copy sequences cannot readily be detected by SPRINT because the newly synthesised strand incorporating fluorescently labelled nucleotides is too short (only a few hundred bases) to be visualised on metaphase chromosomes by standard fluorescent microscopy. Since the signal was not visualised in interphase nuclei, where the chromatin is much less contracted than at metaphase, alternative strategies such as using extended chromatin fibres (Heiskanen *et al.*, 1994) to improve the resolution were not attempted.

Whilst it is unlikely that SPRINT could ever be used for a large-scale physical mapping project for single-copy sequences, it is an ideal method for rapidly localising repetitive DNA sequences to metaphase or interphase cells. SPRINT could be used to investigate centromere organisation and the evolution of repeats in closely related species. In human genetics, PRINS with chromosome-specific repeats is being used to identify supernumerary chromosomes (Koch *et al.*, 1993), to detect aneuploidy (Adinolfi et al., 1993; Pellestor *et al.*, 1995) and to sex foetuses (Adinolfi *et al.*, 1993). Enlarged euchromatic regions, indicating trisomy for a region of the genome containing coding information and which are, therefore, associated with a high risk of phenotypic abnormality, have been detected using the Alu repeat as a primer for PRINS (Callen *et al.*, 1997). In livestock, SPRINT with a panel of chromosome specific repeats could be

used to test for aneuploidy, which is a major cause of fertilisation failure (Pellestor *et al.*, 1995). Another cause of reduced fertilisation in livestock is Robertsonian translocations (eg. cattle 1:29 and cattle 14:20 translocations). It may be possible to detect such translocations in sperm cells by SPRINT using chromosome specific repeats and two different fluorophors (Volpi and Baldini, 1993). If there were a translocation, then the two signals would always appear together. Sexing pre-implantation embryos (Thomsen *et al.*, 1992) by SPRINT with X or Y specific repeats would be a valuable tool for the dairy and meat industries.

Since it was not possible to use SPRINT to add loci to the sheep physical map, alternative strategies to improve the map were investigated. A library of large-insert bacterial artificial chromosome clones was developed as a source of probes for physical mapping by fluorescence *in situ* hybridisation (FISH).

To accurately localise single-copy probe sequences physically on chromosomes, preparations of elongated metaphase spreads are required. Even after optimising the culture conditions for ovine lymphocytes, only 5-7% of all the cells were suitable for making assignments to particular bands. When this study began, fibroblast cell lines for sheep were not available in the laboratory. Fibroblasts grow as a monolayer and can be passaged several times without any change in their karyotype. In the metaphase stage, fibroblasts adhere less tightly to the growth support than during other stages of the cell cycle (Lawce and Brown, 1991). As a consequence of this, they can be easily enriched to obtain a very high mitotic index (>80%). Given the poor yield of chromosomes suitable for mapping from lymphocyte cultures, it was advantageous to develop ovine fibroblast cell lines as a source of metaphase spreads for mapping by FISH.

CHAPTER 4

CONSTRUCTION AND CHARACTERISATION OF AN OVINE BACTERIAL ARTIFICIAL CHROMOSOME LIBRARY

CONSTRUCTION AND CHARACTERISATION OF AN OVINE BACTERIAL ARTIFICIAL CHROMOSOME LIBRARY

4.1 Introduction

In order to identify regions of the genome that contain genes for economically important traits in livestock species and to isolate those genes so that they can be utilised in breeding programmes, high density genome maps are required. Large-insert clones are used to efficiently obtain accurate, high resolution physical maps of eukaryotic genomes (Cai *et al.*, 1995). One aim of this project was to integrate the ovine genetic and physical maps by localising type II microsatellite markers to metaphase chromosomes by fluorescence *in situ* hybridisation (FISH). Consequently, an ovine bacterial artificial chromosome (BAC) library was constructed as a source of large-insert probes for FISH.

A DNA library is a collection of clones that has a high probability of containing at least one copy of every DNA sequence in the genome. A library of clones with large DNA inserts (100kb - 1Mb) is a valuable tool for any genome mapping project. Recently, yeast artificial chromosomes (YAC), P1-derived artificial chromosomes (PAC), and bacterial artificial chromosomes (BAC) were developed (Burke *et al.*, 1987; Ioannou *et al.*, 1994; Shizuya *et al.*, 1992). These vectors can carry ~500kb, 75-100kb and 100-300kb of genomic DNA, respectively. YAC libraries have been constructed for humans (Albertson *et al.*, 1990), mice (Larin *et al.*, 1993; Kusumi *et al.*, 1993) rats (Cai *et al.*, 1997), cattle (Libert *et al.*, 1993; Smith *et al.*, 1996), sheep (Broom and Hill, 1994) and pigs (Rogel-Gaillard *et al.*, 1997). The primary advantage of YAC technology is that the very large clones enable long segments of mammalian chromosomes to be investigated by chromosome walking (Silverman *et al.*, 1989). However, there are some difficulties associated with YAC technology. First, cloning efficiencies are low which means that it is a very laborious and expensive process to

construct a library which completely covers the genome (Smith *et al.*, 1990). Second, YAC clones are often chimaeric, consisting of DNA from different parts of the genome that have accidentally been combined into a single clone (Green and Olson, 1990). For instance, it has been estimated that at least one third of the clones in the bovine YAC library are chimaeric (Libert *et al.*, 1993). This is a serious hindrance to physical mapping and chromosome walking because the true location of the gene of interest cannot be readily determined.

The PAC cloning system is based on the P1 bacteriophage with the large-insert DNA packaged into phage particles that infect *Escherichia coli* with high efficiency. Human PAC libraries have been constructed (Ioannou *et al.*, 1994; Gingrich *et al.*, 1996) but until recently no PAC libraries were available for livestock species. A porcine PAC library is now available (Al-Bayati *et al.*, 1999). Although PAC clones are stable, the *in vitro* packaging system limits the use of this cloning system.

BAC libraries have been constructed for a number of plant species including sorghum (Woo *et al.*, 1994), rice (Wang *et al.*, 1995), soybean (Marek and Shoemaker, 1997) and lettuce (Frijters *et al.*, 1997) as well as for humans (Kim *et al.*, 1996), mice (Research Genetics inc.), cattle (Cai *et al.*, 1995), goats (Schibler *et al.*, 1998a) and horses (Godard *et al.*, 1998) using cloning systems based on the *E. coli* F factor. BAC clones are stable, easy to manipulate and are not chimaeric (Shizuya *et al.*, 1992, Ioannou *et al.*, 1994, Woo *et al.*, 1994, Cai *et al.*, 1995). The only disadvantage of BAC technology for chromosome walking is that the DNA inserts are somewhat smaller than those maintained by YAC clones. Despite this, BAC clones are becoming the vector of choice for physically mapping genes to specific chromosomal locations and for isolating the genes by positional cloning. Additionally, BAC clones can be used in functional studies to investigate gene structure, expression and regulation. Although the BAC vector does not contain a selection system or reporter genes suitable for expression in eukaryotic cell lines (Kim *et al.*, 1998), methods for modifying BAC clones so that the

genes and regulatory elements can be studied in cell lines have recently been described. BAC clones can be retrofitted with expression cassettes by restriction digestion (Mejía and Monaco, 1997), homologous recombination with a shuttle vector (Yang *et al.*, 1997) or Cre-mediated recombination (Chatterjee and Coren, 1997; Kim *et al.*, 1998). BAC DNA can be modified by nested-deletion (Chatterjee and Coren, 1997) or site-specific mutagenesis (Boren *et al.*, 1996), and then expressed in eukaryotic cells (Baker and Cotten, 1997; Chen *et al.*, 1997). BAC clones have recently been used in transgenic animal studies to determine gene function (Antoch *et al.*, 1997) and to demonstrate the importance of long-range regulatory elements in gene expression and the developmental or stage-specific activation of genes (Porcu *et al.*, 1997; Nielsen *et al.*, 1997).

The construction and characterisation of an ovine BAC library are described in this chapter. The quality of the library was determined by PCR-based screening and fluorescence *in situ* hybridisation (FISH). A BAC library covering the ovine genome will be a key resource for comparative gene mapping studies, for identifying quantitative trait loci (QTL) by positional cloning and for functional studies to understand gene expression and regulation.

4.2 Results

4.2.1 Quality of high molecular weight DNA

The quality of DNA prepared by encapsulating white blood cells in agarose microbeads was confirmed by CHEF gel electrophoresis prior to library construction. The majority of the DNA was >1Mb and was retained in the well or at the compression band (Figure 4.1).

4.2.2 Size-selection for BAC library construction

The ovine BAC library was constructed by ligating *HindIII* digested sheep DNA fragments into the *HindIII* cloning site of the pBeloBAC11 vector. This restriction

enzyme was selected to avoid the bulk of the centromeric DNA that is resistant to digestion with *HindIII* and remained in the compression band when the fragments were separated by CHEF gel electrophoresis (Figure 4.1).

The partial digests with 0.75U and 1U *HindIII* gave a majority of fragments in the size range from 50 - 250kb (Figure 4.1). The DNA fragments in this size range were excised from low melting point agarose so that the average insert size of the library would be 150kb. The CHEF gel for size selection was electrophoresed in 1 x TAE rather than 0.5 x TBE because borate ions inhibit ligation. The switch time was also altered to 90" because a 40" switch time previously had been found to adversely affect the ligation efficiency of the fragments, presumably because the 5' overhang or phosphate group were lost. The lambda ladder, however, was not resolved when a 90" switch time was used and the spread of fragments that were excised from the gel was based on their mobility in the previous CHEF gel that used a 40" switch time. This problem is likely to account for the smaller than expected insert sizes that were recovered from the library (section 4.2.3).

The region of the gel corresponding to fragments in the desired size range was cut into four pieces. DNA from the three gel pieces containing the largest fragments was ligated to pBeloBAC11 and transformed by electroporation into DH10β *E. coli.* Colonies produced from each ligation were sized and it was determined that the DNA in the second largest gel piece would be used for library construction. Many of the inserts obtained from the largest gel piece were actually smaller than those from the second largest piece and may indicate that the majority of fragments in that gel slice exceeded the cloning capability of the BAC vector. The transformation efficiency for the BAC clones was 1.5 x 10^6 cfu/μg or ~300 transformants from 1μl of ligation product (ie. ~600 colonies/plate). To achieve a 95% probability of finding any unique sequence within the library, a total of 59,913 colonies with an average insert size of 150kb are required (section 1.3.3.5). Two 100μl ligations were required to produce sufficient

colonies for the BAC library and a total of 59,904 white colonies (624 x 96-well plates) were picked.

Although a second round of size selection has been shown to give a more uniform insert size distribution across the library it is accompanied by a 10 fold reduction in transformation efficiency (Cai *et al.*, 1995). Consequently, the ovine BAC library was generated from DNA that was only subjected to a single size selection to maximise the efficiency of library construction.

4.2.3 Determination of average insert size

The average insert size of the BAC clones was determined by CHEF gel electrophoresis (Figure 4.2). In initial sizing experiments, double bands were observed that migrated through the gel about 7kb apart and hindered accurate sizing of the BAC clones (Figure 4.2a). Faintly staining high molecular weight bands were also visible. The doublet migrated through the gel about a well-width apart. Initially, the samples were loaded whilst the gel was dry and then the wells were capped with low-melting point agarose which was allowed to set before the gel was placed in the electrophoresis buffer. It was thought that perhaps some of the DNA was being trapped at the back of the well so that it was migrating through the extra agarose used to cap the well. Alternatively, the molten agarose may have been denaturing some of the DNA with the different conformations migrating through the gel matrix at different rates. To eliminate these possibilities as the source of the second band, the samples were mixed with gel loading buffer and the gel was loaded in the electrophoresis buffer. However, the additional band and faint high molecular weight bands persisted. It was determined that the doublet was due to incomplete digestion of a *NotI* site on the vector since pBeloBAC11 is 7.4kb. The faintly staining high molecular weight bands were due to supercoiled and open-circular BAC DNA (Wang and Lai, 1995). The problem was overcome by reducing the amount of DNA used in the *NotI* digest to ~100ng (Figure

4.2b). Linearised BAC DNA prepared by standard alkaline lysis is not stable and some degradation was visualised as a smear on the CHEF gels. A total of 139 clones were selected at random and the average insert size was determined to be 103kb with a range from 25kb to 410kb (Figure 4.3). This corresponds to 2 genome equivalents (59,904 x 103kb = ~6 x 10^9) assuming that the ovine genome is 3 x 10^9bp. Therefore, there is an 88% chance (since $P = 1 - (1-I/GS)^N$, section 1.3.3.5) of finding any given sequence of interest in the library.

Only one of the BAC clones that were selected for sizing failed to have an insert suggesting that less than 1% of the clones in the library are empty clones. Many of the clones that were sized produced a single insert band. However, 32 inserts had at least one internal *NotI* site. *NotI* sites are indicative of CpG islands which are associated with all house-keeping genes and some tissue specific genes in mammalian genomes (Gardiner-Garden and Frommer, 1987; Aïssiani and Bernardi, 1991).

4.2.4 PCR-based screening

The BAC library was screened by PCR with primers for 80 microsatellite markers and genes (Table 4.1). Primers were selected so that at least two BAC clones were isolated for every sheep chromosome. At least one positive superpool was found for 69 loci. IL5 and IRF1 were both found within a single BAC clone. DQB, DQA1 and DQA2 were not found in the library. It is expected that they were clustered within 100kb of each other, and would, therefore represent a single missing BAC clone (J. Maddox, pers. comm.). If these three genes are treated as a single locus, then it can be estimated that there is an 88.5% chance of finding any sequence of interest in the ovine BAC library. This agrees with the estimation based on the average insert size (section 4.2.3).

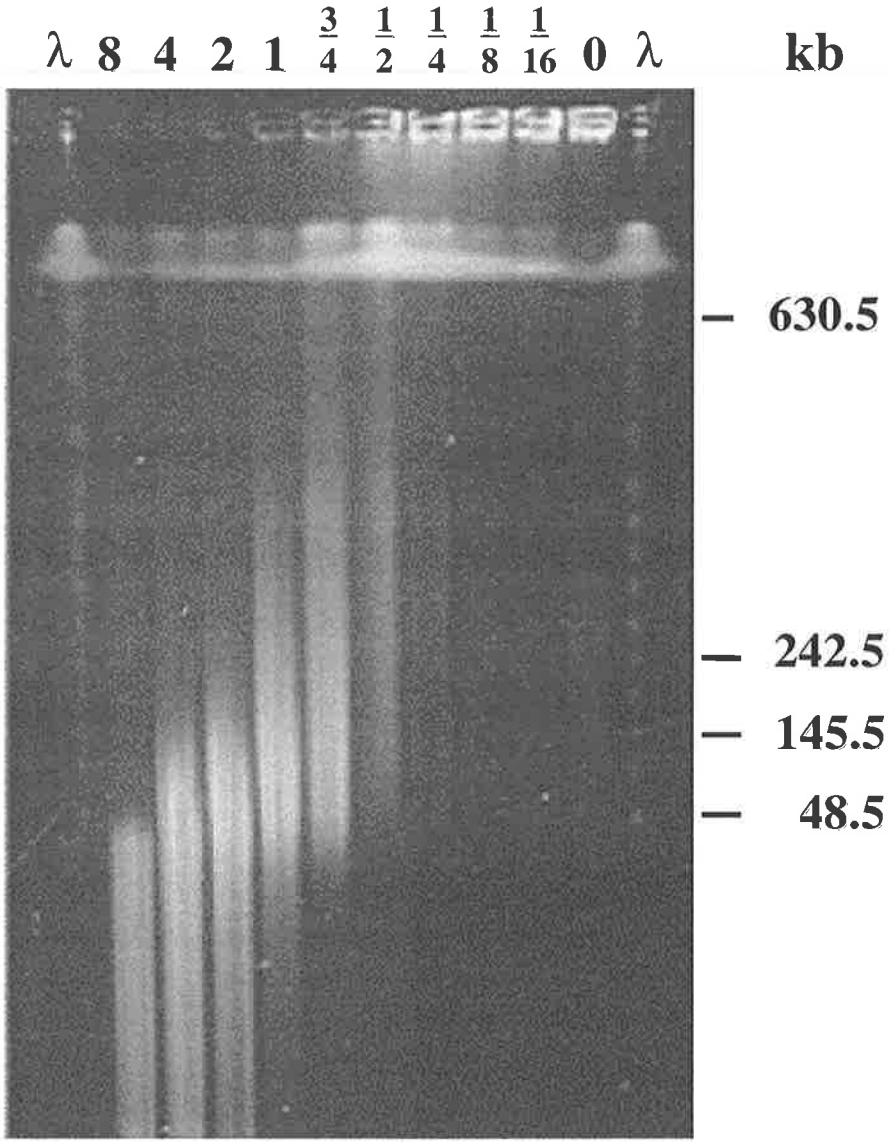


Figure 4.1: Partial digestion series of ovine DNA with *HindIII* to generate fragments with an average size of 150kb. The amount of enzyme (0-8U) used in each digest is indicated. The fragments were sized against lambda concatomers. The digests were loaded onto a 1% agarose CHEF gel in 0.5xTBE and the gel was run at 6V/cm for 18h at 14°C.

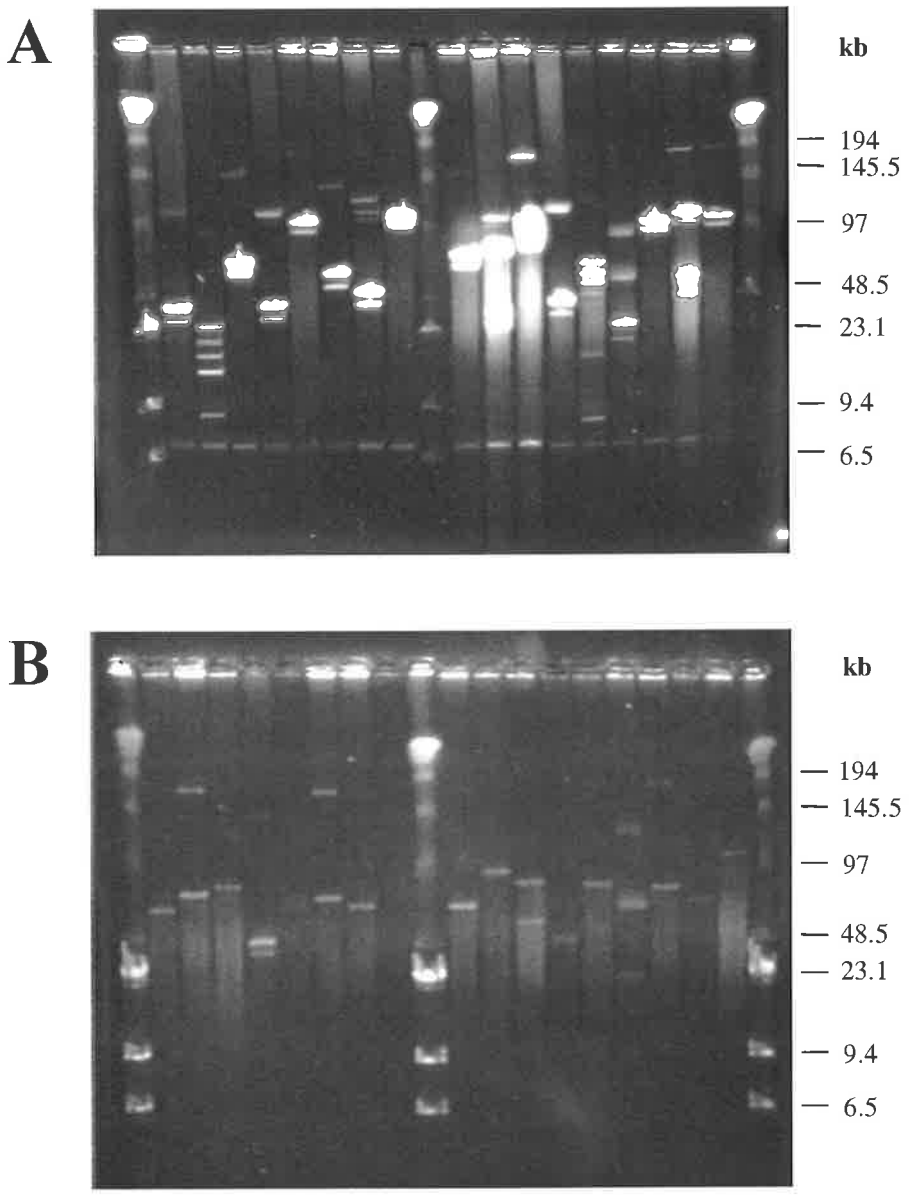


Figure 4.2: Pulse-field gel electrophoretic analysis of insert size in the ovine BAC library. Inserts were removed from the vector by digestion with *NotI*. The DNA was separated on the basis of size by electrophoresis using a CHEF-DRII apparatus (*Biorad*). Samples were loaded on a 1% agarose gel in 0.5 x TBE and run at 14¡C for 16h at 6V/cm with an initial pulse time of 1.4s and a linear ramp to a final pulse time of 13.5s. The size standard is composed of lambda concatamers (*Promega*) and a *HindIII* lambda ladder (*Promega*). A) Example of the doublets and faint high-molecular weight bands caused by incomplete digestion of BAC clones with *NotI*. B) Complete digestion of BAC clones when 100ng DNA was used in the digest enables accurate sizing. Degraded DNA is visible as a smear.

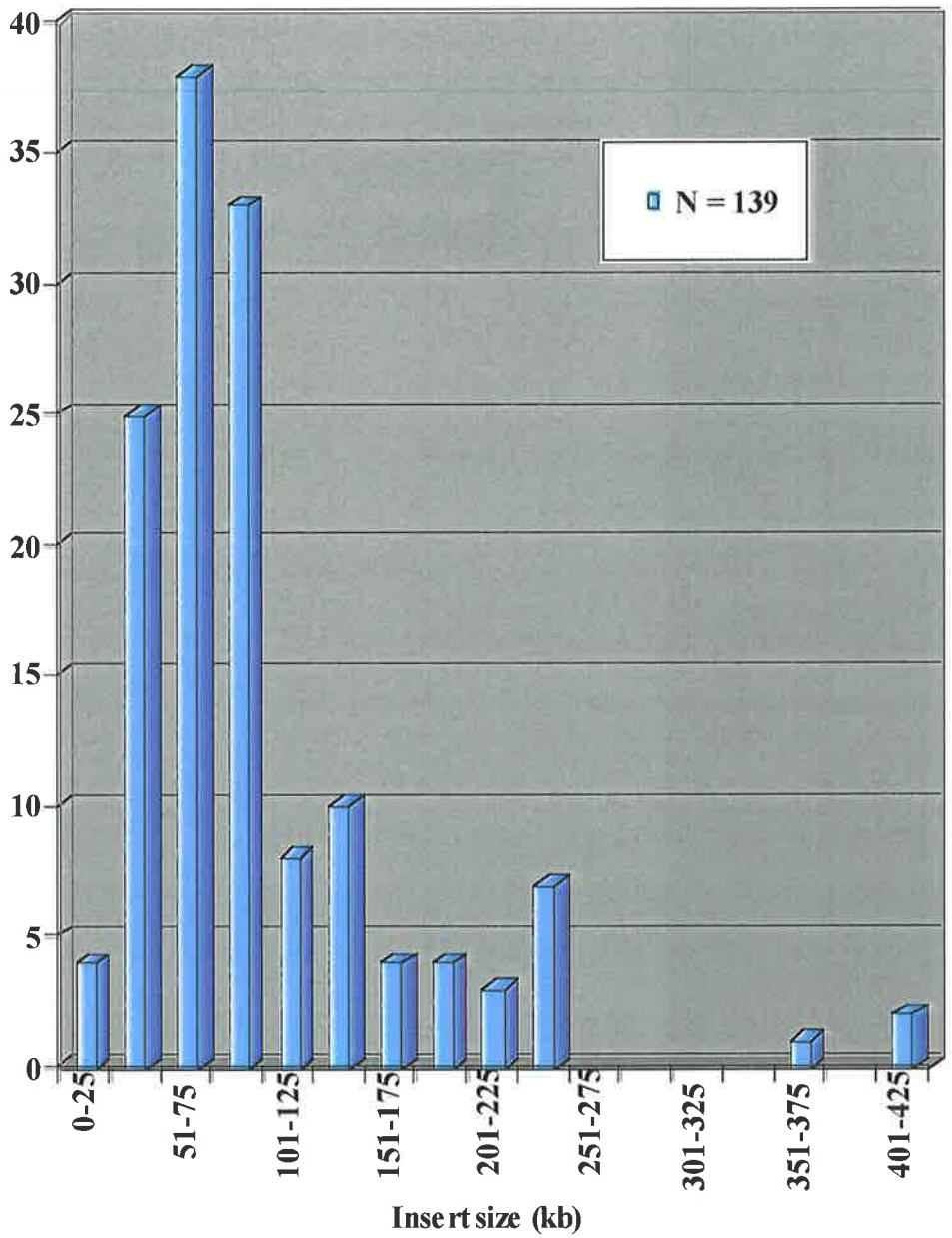


Figure 4.3: Distribution of insert sizes in the ovine BAC library

An individual BAC clone was isolated for each locus that had at least one positive superpool. PCR products generated from the individual clones were sequenced manually prior to physical mapping and the sequences were aligned with published data.

4.2.5 Sequencing

The identity of every BAC clone that was isolated by PCR-based screening was confirmed by sequencing prior to physical mapping. PCR products generated from each clone were used as the template for sequencing since the BAC vector is a single-copy vector and it is not possible to recover sufficient plasmid for sequencing from a standard 50ml preparation. The PCR products generated from the BAC clones were sequenced manually in both directions and then aligned with published sheep or cattle sequences (Table 4.1). Many of the microsatellites used to construct the sheep genetic linkage map were originally isolated from cattle and for 33 loci that were sequenced only cattle data were available in the GENBANK database. Every clone that was isolated shared at least 84% sequence identity with the published sequences over a minimum of 64bp and this was considered sufficient to confirm that the correct clones had been isolated. Any differences in the alignments which were due to microsatellite polymorphisms (ie. a different repeat length to the published allele) were not included in the calculation. As expected the best alignments (KAP1.1, MAF214, IL2RA, IL6, KRT2.13) were achieved when the sequence was compared to sheep data.

4.2.6 Estimation of chimaerism

The proportion of chimaeric clones in the BAC library was evaluated by physically mapping 45 of the 69 clones identified by PCR-based screening (section 4.2.4). The BAC clones that were physically mapped by FISH to ovine metaphase

chromosomes (Figure 4.5) ranged in size from 25kb to 250kb and no chimaeric clones were detected.

4.3 Discussion

There are a number of biases in the ovine BAC library because it was constructed using *HindIII* size-selected DNA fragments extracted from a Suffolk ram. For the Australian wool industry, it would have been desirable to use a fine-wool Merino as the DNA source, but no suitable animals were available for sampling in North America. This means that if a gene for a particular trait is isolated from the library, then the exact genotype that causes characteristics peculiar to the fine-wool Merino cannot be established using the library.

Since the BAC library was made from male DNA, both the X and Y chromosomes are represented but each sex chromosome is under-represented since there is only one copy of each in the genome and two copies of every autosome. Apart from the pseudo-autosomal region, there are very few genes on the Y chromosome. However, there are many important genes on the X chromosome that control X-linked traits. To increase the representation of the X chromosome, additional clones prepared from female DNA could be added to the library.

The restriction enzyme *HindIII* was selected for library construction to avoid the bulk of the repetitive centromeric DNA sequences, which may mean that regions immediately adjacent to the centromere are poorly represented in the library. However, a BAC clone containing the microsatellite BM6438 has been isolated from the library. BM6438 is linked to the horn-polled gene in cattle and has been mapped by genetic linkage analysis to the centromere of both cattle and sheep chromosome 1. The proportion of clones in the library containing centromeric sequences could be investigated using satellite DNA probes.

Table 4.1: Summary of results from screening the ovine BAC library by PCR with primers for microsatellites and genes. The number of positive superpools for each locus (NF: not found) and the size of the individual BAC clone (ND: not done) that was isolated for a locus are listed. PCR products generated from the BAC clones were sequenced prior to physical mapping and were compared to published sheep (*Ovis*) or cattle (*Bos*) sequences.

Locus	Sources	No. of Positive Superpools	Insert Size (kb)	GENBANK Sequence Accession #	Origin of Pub. Sequence	% Sequence Identity
ADCYC	N. Cockett, Utah State Uni., UT, USA	6	100	X16451	*Bos*	92% over 68bp
BM415	N. Cockett, Utah State Uni., UT, USA	NF	-	G18413	*Bos*	-
BM719	N. Cockett, Utah State Uni., UT, USA	2	100	G18427	*Bos*	95.7% over 70bp
BM827	N. Cockett, Utah State Uni., UT, USA	3	ND	U06763	*Bos*	90% over 111bp
BM1303	N. Cockett, Utah State Uni., UT, USA	3	125	G18424	*Bos*	91.7% over 60bp
BM1329	N. Cockett, Utah State Uni., UT, USA	2	235	G18422	*Bos*	87.3% over 95bp
BM1824	N. Cockett, Utah State Uni., UT, USA	3	250	G18394	*Bos*	88% over 137bp
BM3413	N. Cockett, Utah State Uni., UT, USA	1	405	G18492	*Bos*	91.7% over 118bp
BM4107	N. Cockett, Utah State Uni., UT, USA	2	155	G18519	*Bos*	91.2% over 91bp
BM4208	N. Cockett, Utah State Uni., UT, USA	4	30	G18509	*Bos*	96% over 108bp
BM6438	N. Cockett, Utah State Uni., UT, USA	1	70	G18435	*Bos*	90% over 196bp
BM6526	N. Cockett, Utah State Uni., UT, USA	5	80	G18454	*Bos*	90.5% over 95bp
CAPN2[3]	C. Bottema, Uni. of Adelaide, SA, Aust.	NF	-	J05065	*Bos*	-
CD3D	J. Maddox, Uni. of Melbourne, VIC, Aust.	2	240	AF117651	*Ovis*	94% over 105bp
CD5	J. Maddox, Uni. of Melbourne, VIC, Aust.	NF	-	M97195	*Ovis*	-
CSF2	J. Maddox, Uni. of Melbourne, VIC, Aust.	1	ND	[1]Unpublished	*Ovis*	91% over 136bp
CSRD241	N. Cockett, Utah State Uni., UT, USA	4	30	[2]Not in Genbank	*Bos*	$(GT)_{17}$

Table 1 continued.

Locus	Sources	No. of Positive Superpools	Size (kb)	GENBANK Sequence Accession #	Origin of Pub. Sequence	% Sequence Identity
CSRD270	N. Cockett, Utah State Uni., UT, USA	1	85	[2]Not in Genbank	*Bos*	$(GT)_{30}$
CSRD287	N. Cockett, Utah State Uni., UT, USA	1	230	[2]Not in Genbank	*Bos*	$(GT)_{26}$
CSSM043	N. Cockett, Utah State Uni., UT, USA	1	40	U03824	*Bos*	92.3% over 130bp
DQA1	J. Maddox, Uni. of Melbourne, VIC, Aust.	NF	-	Z28418	*Ovis*	-
DQA2	J. Maddox, Uni. of Melbourne, VIC, Aust.	NF	-	Z28421	*Ovis*	-
DQB	J. Maddox, Uni. of Melbourne, VIC, Aust.	NF	-	Z28425	*Ovis*	-
DRB1	J. Maddox, Uni. of Melbourne, VIC, Aust.	1	55	M73984	*Ovis*	93.6% over 140bp
DRB2	J. Maddox, Uni. of Melbourne, VIC, Aust.	2	50	S83920	*Ovis*	93% over 183bp
DYB	J. Maddox, Uni. of Melbourne, VIC, Aust.	2	55	Z27401	*Ovis*	93% over 63bp
FN1	Kornblihtt *et al.*, 1983	2	80	-	*Bos*	97% over 105bp
FSHB	N. Cockett, Utah State Uni., UT, USA	1	215	S64745	*Ovis*	95% over 80bp
GAPR[3]	C. Bottema, Uni. of Adelaide, SA, Aust.	4	147	X12602	*Bos*	98% over 161bp
GHRH[3]	C. Bottema, Uni. of Adelaide, SA, Aust.	3	188	U29611	*Bos*	93% over 152bp
GH2	C. Bottema, Uni. of Adelaide, SA, Aust.	1	ND	M37310	*Ovis*	98% over 210bp
HUJ614	N. Cockett, Utah State Uni., UT, USA	1	ND	M93653	*Bos*	93% over 101bp
IFNG	J. Maddox, Uni. of Melbourne, VIC, Aust.	2	75	Z73273	*Ovis*	95% over 140bp
IGF1	C. Bottema, Uni. of Adelaide, SA, Aust.	1	ND	X17229	*Ovis*	98.4% over 66bp
IL1A	J. Maddox, Uni. of Melbourne, VIC, Aust.	NF	-	AF117652	*Ovis*	-
IL2RA	N. Cockett, Utah State Uni., UT, USA	2	75	[1]Unpublished	*Ovis*	100% over 115bp
IL3	J. Maddox, Uni. of Melbourne, VIC, Aust.	4	60	Z18897	*Ovis*	95% over 197bp
IL5	J. Maddox, Uni. of Melbourne, VIC, Aust.	3	95	U17052	*Ovis*	96% over 127bp
IL6	J. Maddox, Uni. of Melbourne, VIC, Aust.	1	195	X62501	*Ovis*	100% over 75bp
ILSTS005	N. Cockett, Utah State Uni., UT, USA	3	ND	L23481	*Bos*	90% over 102bp

Table 1 continued.

Locus	Sources	No. of Positive Superpools	Size (kb)	GENBANK Sequence Accession #	Origin of Pub. Sequence	% Sequence Identity
ILSTS008	N. Cockett, Utah State Uni., UT, USA	3	45	L23483	*Bos*	88.8% over 125bp
ILSTS011	N. Cockett, Utah State Uni., UT, USA	2	75	L23485	*Bos*	92.3% over 78bp
ILSTS043	N. Cockett, Utah State Uni., UT, USA	1	130	L37218	*Bos*	91.5% over 71bp
IRF1	J. Maddox, Uni. of Melbourne, VIC, Aust.	1	95	[1]Unpublished	*Ovis*	88% over 166bp
KAP1.1	Powell *et al.*, 1983	2	105	-	*Ovis*	100% over 266bp
KAP6.1	Y. Parsons, Uni. of Sydney, NSW, Aust.	2	ND	M95719	*Ovis*	97.5% over 123bp
KRT2.13	J. Maddox, Uni. of Melbourne, VIC, Aust.	3	30	X72379	*Ovis*	100% over 96bp
LDLR[3]	C. Bottema, Uni. of Adelaide, SA. Aust.	3	147	M11341	*Bos*	87% over 86bp
LGB	C. Bottema, Uni. of Adelaide, SA, Aust.	1	95	OALGB	*Ovis*	91% over 259bp
LPL[3]	C. Bottema, Uni. of Adelaide, SA, Aust.	1	200	M16966	*Bos*	92% over 133bp
MAF23	N. Cockett, Utah State Uni., UT, USA	1	ND	[2]Not in Genbank	-	$(GT)_{31}$
MAF35	N. Cockett, Utah State Uni., UT, USA	3	150	M80518	*Ovis*	91.8% over 73bp
MAF50	N. Cockett, Utah State Uni., UT, USA	NF	-	M77377	*Ovis*	-
MAF92	N. Cockett, Utah State Uni., UT, USA	3	90	M80527	*Ovis*	88.6% over 88bp
MAF214	N. Cockett, Utah State Uni., UT, USA	1	25	M88160	*Ovis*	100% over 131bp
McM58	N. Cockett, Utah State Uni., UT, USA	2	ND	[2]Not in Genbank	-	$(GT)_{16}$
McM111	N. Cockett, Utah State Uni., UT, USA	2	ND	L35310	*Ovis*	85% over 64bp
McM136	N. Cockett, Utah State Uni., UT, USA	2	ND	L38978	*Ovis*	95% over 70bp
MTNR1A	Messer *et al.*, 1997	NF	-	-	-	-
OarEL01	N. Cockett, Utah State Uni., UT, USA	1	85	G07424	*Ovis*	92% over 76bp
OarFCB48	N. Cockett, Utah State Uni., UT, USA	2	ND	M82875	*Ovis*	89.8% over 98bp
OarHH56	N. Cockett, Utah State Uni., UT, USA	2	ND	L13871	*Ovis*	98.6% over 125bp
OarHH64	N. Cockett, Utah State Uni., UT, USA	1	ND	L12558	*Ovis*	93% over 129bp

Table 1 continued.

Locus	Sources	No. of Positive Superpools	Size (kb)	GENBANK Sequence Accession #	Origin of Pub. Sequence	% Sequence Identity
OarJMP8	N. Cockett, Utah State Uni., UT, USA	2	135	OAU35059	*Ovis*	95.2% over 125bp
OarJMP58	N. Cockett, Utah State Uni., UT, USA	4	55	OAU35058	*Ovis*	96.3% over 136bp
OarVH72	N. Cockett, Utah State Uni., UT, USA	4	75	L12548	*Ovis*	90.6% over 64bp
OarVH110	N. Cockett, Utah State Uni., UT, USA	2	ND	L12550	*Ovis*	88.2% over 93bp
OarVH116	N. Cockett, Utah State Uni., UT, USA	1	250	L12551	*Ovis*	97.6% over 83bp
OB	C. Bottema, Uni. of Adelaide, SA, Aust.	2	ND	Y11369	*Bos*	96% over 96bp
OCAM	N. Cockett, Utah State Uni., UT, USA	2	125	BTOBCAM	*Bos*	90% over 130bp
PRF	J. Maddox, Uni. of Melbourne, VIC, Aust.	2	ND	AJ010711	*Ovis*	94% over 105bp
PRL	Wolf *et al.*, 1990	1	55	-	*Bos*	84% over 108bp
RJH1	N. Cockett, Utah State Uni., UT, USA	NF	-	U47049	*Ovis*	-
SCYA@[4]	J. Maddox, Uni. of Melbourne, VIC, Aust.	2	170	[1]Unpublished	*Ovis*	95% over 291bp
SRN[3]	C. Bottema, Uni. of Adelaide, SA, Aust.	1	176	AJ000518	*Bos*	100% over 211bp
SRY	Payen and Cotinot, 1994	4	90	OADSRYC	*Ovis*	97% over 67bp
TGLA122	N. Cockett, Utah State Uni., UT, USA	4	ND	[2]Not in Genbank	-	$(GT)_{15}$
TGLA357	N. Cockett, Utah State Uni., UT, USA	NF	-	[2]Not in Genbank	-	-
UWCA5	N. Cockett, Utah State Uni., UT, USA	2	ND	L15334	*Bos*	97% over 91bp
ZFX/ZFY	C. Bottema, Uni. of Adelaide, SA, Aust.	8	ND	D84097	*Bos*	91.7%

[1]Sequence for comparisons provided by Dr. J. Maddox.
[2]Sequence unavailable for comparison. PCR products contain the expected microsatellite repeats.
[3]BAC clones isolated and sequenced by L. Allen (1997).
[4]BAC was isolated with primers for SCYA, but was subsequently shown to be another member of this gene family.

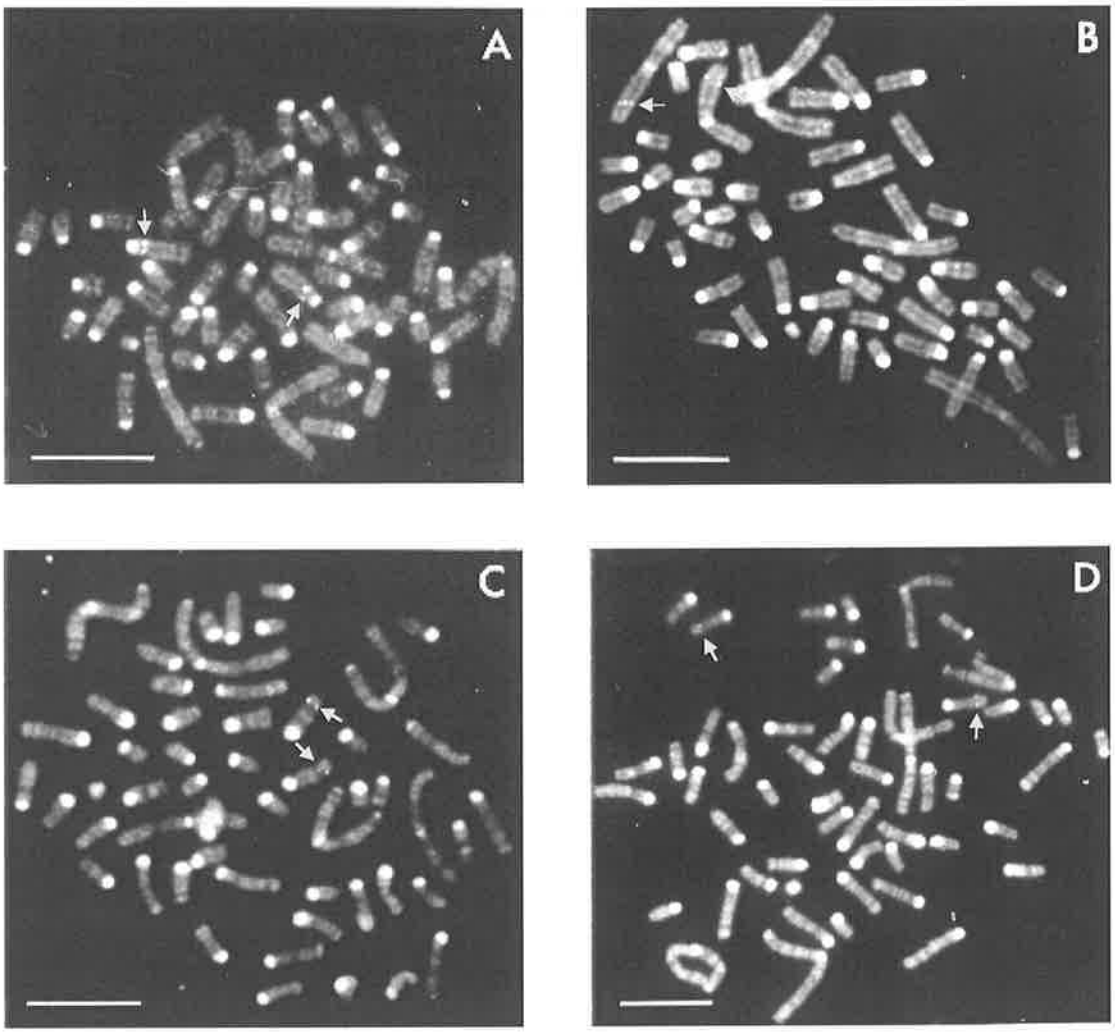


Figure 4.4: Physical localisation of ovine BAC clones to sheep metaphase chromosomes by fluorescence *in situ* hybridisation. A) BM1329 (235kb clone) localised to 6q1.6-q1.8. B) IFNG (75kb clone) localised to 3q1.4. C) BM4208 (30kb clone) localised to 8q2.3-q2.5. D) FSHB (215kb clone) localised to 15q3.3-q3.4. Arrows indicate fluorescent signal. Bars represent 10μm.

Similarly, a probe for the telomeric repeat, $(TTAGGG)_n$, could be used to isolate telomeric clones. To overcome some of the biases in the library due to the use of a single restriction enzyme, additional clones could be prepared using another enzyme (eg. *BamHI*).

The coverage of the ovine BAC library was evaluated by sizing inserts and by PCR-based screening. The presence of additional bands due to incomplete digestion of the BAC clones could lead to an over-estimation of insert size, and therefore, genome coverage, if this was the only method used to characterise the BAC library. Indeed, sizing was the only form of characterisation that was presented for the recently constructed chicken BAC library (Zimmer and Verrinder Gibbins, 1997). Although the size standard is not resolved in any of the figures that are presented by the authors, they claim that the library has an average insert size of 390kb ranging from 25 to 725kb. It remains to be seen whether the expected coverage based on size is supported by screening. Extensive PCR-based screening of the sheep BAC library for 80 markers and genes verified that the estimation of genome coverage made on the basis of size was accurate.

The difficulties in sizing the BAC clones by pulsed-field gel electrophoresis could be overcome by using flow cytometry (Huang *et al.*, 1996). In this method, large DNA fragments are labelled with a bis-intercalating dye and passed through a laser illuminated detection volume to produce photon bursts. The burst size is proportional to the number of dye molecules bound to the DNA, and therefore, is proportional to fragment length. Whilst sizing by pulsed-field gel electrophoresis requires ng amounts of linearised DNA and the gel needs to be run for hours to separate the fragments, the advantages of flow cytometry are that the size of undigested, supercoiled BAC DNA can be established in minutes using pg amounts of DNA.

The insert sizes in at least two of the BAC libraries that have been constructed previously were smaller than expected (Cai *et al.*, 1995; Frijters *et al.*, 1997) with the

average insert size of the clones below the size of the DNA that was selected to make the BAC libraries. The ovine BAC clones were also smaller than expected, but within the selected size range, with an average insert size of 100kb instead of the intended 150kb. This may be due to preferential cloning of smaller DNA fragments which co-migrate with the larger fragments as previously described (Cai *et al.*, 1995; Frijters *et al.*, 1997), or it may to be due to inaccurate sizing of the fragments that were excised to make the library, since the molecular weight ladder was not resolved under the conditions that were necessary to guarantee a good cloning efficiency.

The smaller than expected insert sizes means that at present there is an 88% chance of identifying an ovine BAC clone containing a sequence of interest (section 1.3.3.5). This obviously means that about 1 in 10 genes of interest are missing from the library and it should be expanded to ensure that there is a 99% of isolating any sequence. Prior to completion of the library, the CHEF gel conditions should be investigated to determine whether a different set of pulse times would enable the molecular weight marker to be resolved and yet still maintain good ligation and transformation efficiencies.

No chimaeric BAC clones were detected by FISH but, due to limits in sensitivity, this does not preclude the possibility of there being small regions of non-contiguous DNA within the clones. It has been demonstrated using Southern hybridisation to DNA from somatic cell hybrids that BAC clones, which appeared non-chimaeric, by FISH did indeed map to multiple chromosomes (T. Broad, AgResearch, pers. comm.). Furthermore, two microsatellites isolated from a single BAC clone have mapped to two different linkage groups (T. Broad, pers. comm.). It will remain to be seen whether chimeras that were not detected by FISH are discovered by these alternative methods in the sheep BAC library described herein.

The present coverage of the library is quite adequate for physical mapping studies because the exact sequence is not required to obtain hybridisation. Consequently, if a

locus is not present in the ovine library then either the cattle (Cai *et al.*, 1995) or goat (Schibler *et al.*, 1998a) BAC libraries could be screened and those clones used for comparative mapping. The expansion of the BAC library to 99% coverage is necessary though so that ovine BAC clones for any sequence can be isolated for functional studies where the exact sheep sequence is required to determine the physiological basis for phenotypic variation due to DNA sequence variation at the locus. Also, the BAC library would need to be expanded if the clones in the library were to be ordered so that contigs could be developed for every chromosome.

Recently, another ovine BAC library was constructed (Vaiman *et al.,* 1999b). The impetus was to generate a source of clones that could be used to develop an animal model for prion diseases by transgenesis in mice. A cell line derived from a Romanov ram with increased susceptibility to scrapie due to the genotype of the PRNP gene (VRQ/VRQ) was the source of DNA for library construction. This *HindIII* library has 90,000 clones with an average insert size of 123kb and represents three genome equivalents. Unfortunately, the same biases exist in this library as the one constructed in the present study, since a male from one breed was used as the source of DNA and the same enzyme (*HindIII*) was used for the entire library.

The two libraries combined, however, represent five genome equivalents, meaning there is theoretically now a 99% probability of finding any given sheep sequence in one of the libraries (section 1.3.3.5). Large BAC libraries (10x genome coverage) constructed with a single enzyme in mice have a higher than expected number of missing loci (N. Jenkins, NCI-Frederick Cancer Research and Development Center, pers. comm.). This is likely to be due to variability in the abundance of enzyme sites throughout the genome. It is expected that *HindIII*-rich and *HindIII*-poor regions of the genome would be more likely to be absent from the ovine BAC libraries because of the size selections made for library construction. It would be worthwhile screening the Romanov BAC library for the 11 markers and genes that were not found in the

Suffolk BAC library herein. If these loci were also absent from the Romanov library it would suggest that rather than by chance, some property of the sequence in these regions is making each locus recalcitrant to BAC cloning with *HindIII*.

Prior to physical mapping studies, PCR products generated from the BAC clones were sequenced to verify the identity of the clones. All the products shared >84% sequence identity with previously published sequences for the genes and markers screened. Sequencing alone, however, does not necessarily verify the identity of BAC clones isolated for genes from large families (Cai *et al.*, 1995; Schibler *et al.*, 1998a). Members of large gene families that arose by duplication events are often highly conserved sharing >90% sequence identity. The only confirmation of clone identity in such cases is coincident mapping data. A BAC clone for a SCYA chemokine was isolated from the ovine library and shared 95% sequence identity with the sequence from which the primers were derived. The BAC clone was expected to map to OAR11, but a microsatellite isolated from this clone was subsequently mapped by genetic linkage analysis to OAR13 (J. Maddox, pers. comm.). It is presumed that the BAC clone contains another member of the SCYA gene family. This result demonstrates the care that must be taken to avoid making incorrect map assignments when members of large gene families are involved.

The ovine BAC library, described herein, will be a powerful tool for comparative mapping and the positional cloning of genes for economically important traits. BAC clones containing genes of interest could be modified and expressed in mammalian cells or used in transgenesis programmes. The ovine BAC library will also enable genome organisation, gene regulation and chromosome evolution in closely related species, such as deer, to be investigated.

CHAPTER 5

PHYSICAL MAPPING IN SHEEP USING OVINE BAC CLONES

PHYSICAL MAPPING IN SHEEP USING OVINE BAC CLONES

Physical mapping is the localisation of genes or markers to specific chromosomal regions. A physical map is a useful tool in the search for genes for economically important traits because it can define the actual region that contains the candidate genes or markers that are closely linked to the trait of interest. The physical maps for sheep and cattle (eg. ArkDB at Roslin: http://www.ri.bbsrc.ac.uk:8080/cgi-bin/arkdb) are in their infancy compared to the dense physical maps that are available for human (eg. Whitehead/MIT physical map at http://carbon.wi.mit.edu:8000/cbi-bin/contig/phys_map) and mouse (eg. MGI-Whitehead/MIT physical map at http://www.informatics.jar.org/mitmaps.html). However, human and mouse physical mapping data can be utilised to improve the density of genes on the livestock maps. Physical localisation of large-insert clones containing human and mouse gene homologues (comparative mapping loci) helps to define critical regions containing economically important traits. Candidate genes that localise to the same region as the comparative mapping loci can then be extrapolated from the gene-rich human and mouse maps.

In 1998, the ovine physical map consisted of 178 genes and 54 markers (ArkDB, 11/9/98). There are currently 222 genes and 207 markers on the physical map (ArkDB, 6/8/99) and several physical mapping methods have been utilised to localise these loci to specific chromosomes. The majority of the physical assignments in sheep (338/429) have been made using somatic cell hybrids (Saidi-Mehtar *et al.*, 1981; Burkin *et al.*, 1993; Burkin *et al.*, 1998). The recently completed, cytogenetically characterised sheep x hamster somatic cell hybrid panel was used to verify the chromosomal assignments of 110 microsatellite markers (Burkin *et al.*, 1998), leading to a rapid increase in the number of markers on the physical map. Whilst this method localises loci to particular chromosomes, it does not reveal the precise band location. This limits the amount of comparative mapping information that can be taken from the human and mouse physical maps because gene order is not conserved at the whole chromosome level.

In situ hybridisation (ISH) of radioactive probes or fluorescence *in situ* hybridisation (FISH) of fluorescent probes to target sequences on the chromosomes is a more precise localisation method and enables loci to be mapped to discrete chromosomal bands. Only 10 of the 197 physical assignments that were made in 1999 were FISH localisations. Indeed, only 91 loci have been mapped by ISH or FISH in sheep and just 25 of these have been mapped by genetic linkage analysis as well (ArkDB, 6/8/99). Three of the new assignments (NRAMP1, BM1237 and IGF2) directly anchored the physical and genetic linkage maps for chromosomes 2, 7 and 21, respectively. There are still 11 chromosomes (3, 8, 11, 12, 16, 20, 22, 23, 25, 26 and X) for which the physical maps and genetic linkage maps have not been directly anchored by FISH.

One of the primary objectives of this project was to place more loci on the ovine physical map. Specifically, the aim was to use FISH to map ovine BAC clones containing loci that have been mapped by genetic linkage analysis (chapter 4) to anchor and orient all the linkage maps to their respective chromosomes. The other objective was to place more comparative mapping loci on the physical map in co-localisation experiments with the anchor BAC clones to avoid the problems associated with chromosome identification in Bovidae (Popescu *et al.*, 1996).

The production of a physical map of a species requires the unambiguous identification of all the chromosomes (Ansari *et al.*, 1993). The identification of the chromosomes in Bovidae (eg. sheep, cattle and goat) is quite difficult because the chromosomes are similar in shape and decrease gradually in size. The cattle (BTA) and goat (CHI) karyotypes consist of 29 acrocentric chromosomes and the sheep (OAR) karyotype is 3 metacentric chromosomes and 23 acrocentric chromosomes.

At the 1976 Reading conference (Reading, 1980) and the second international conference on standardization of domestic animal karyotypes in 1989 (ISCNDA, 1990), standard nomenclature for the cattle, sheep and goat chromosomes was developed based

on G-banding and R-banding patterns. There appears to be extensive conservation between the sheep, cattle and goat karyotypes on the basis of the chromosome banding patterns (Evans *et al.*, 1973; Hayes *et al.*, 1991; Hediger *et al.*, 1991a; Kaftanovskaya and Serov, 1994; Iannuzzi and Di Meo, 1995). This is largely supported by genetic linkage data (Kemp *et al.*, 1993; Vaiman *et al.*, 1996; de Gortari *et al.*, 1997; de Gortari *et al.*, 1998) and comparative mapping assignments (Hayes and Petit, 1993; Echard *et al.*, 1994; Vaiman *et al.*, 1996). Every goat chromosome and every arm of the sheep karyotype has a bovine homologue with some minor rearrangements (Table 5.1). BTA1 and BTA3 correspond to the metacentric OAR1, BTA2 and BTA8 correspond to OAR2, and BTA5 and BTA11 correspond to OAR3. The nomenclature for cattle and goat are identical (Popescu *et al.,* 1996), but the numbering system is different for sheep (Table 5.1).

There were several ambiguities in the Reading (1980) and ISCNDA (1990) karyotypes, and these have only recently been resolved (Ansari *et al.*, 1994; Ansari *et al.,* 1996; Popescu *et al.*, 1996; Ansari *et al.,* 1999). The ovine chromosome identifications that were contentious were 4, 6, 8, 9, 19, 20, 21, 23, 24, 25 and 26. Chromosome-specific reference markers have been assigned to every sheep and cattle chromosome to assist in chromosome identification (Ansari *et al.*, 1996; Popescu *et al.*, 1996; Ansari *et al.,* 1999)(Table 5.1). For instance, chromosome 4 and 6 have similar banding patterns and were interchanged in the ISCNDA (1990) karyotype. This led to the concurrent assignment of β-casein (CSN2) to both sheep chromosome 6 (Ansari *et al.*, 1992) by the Reading (1980) G-banded karyotype and sheep chromosome 4 (Hayes *et al.*, 1993) by the ISCNDA (1990) R-banded karyotype. The Reading (1980) convention is now used and β-casein is the reference marker for sheep, cattle and goat chromosome 6 (Table 5.1).

A similar problem was identified for BTA25 and BTA29. Insulin-like growth factor 2 (IGF2) and lactate dehydrogenase-A (LDHA) were mapped by *in situ*

hybridisation to BTA29 in an animal carrying the traditional 1:29 Robertsonian translocation (Schmutz *et al.*, 1996). LDHA had also been assigned to syntenic group U7 (Womack and Moll, 1986) and this linkage group was localised to BTA25 by FISH using two cosmids, BMC3224 and BMC8012 (Bishop *et al.*, 1994). However, when these cosmids were mapped again by FISH in cells carrying the 1:29 translocation they were localised to the translocated chromosome, BTA29 (Schmutz *et al.,* 1996). Consequently, the physical assignment and linkage group nomenclature for BTA25 and BTA29 have been interchanged to eliminate this inconsistency (Popescu *et al.*, 1996). The homologous chromosomes in sheep are OAR24 and OAR21, respectively.

The G-banding patterns for sheep and goat are identical, but there have been some minor rearrangements compared to the cattle karyotype. Goat and sheep chromosomes 8 (according to the Reading (1980) convention, but referred to as chromosome 9 by Kaftanovskaya and Serov (1994) and Iannuzzi and Di Meo (1995)) are shorter than the cattle homologue 9. Conversely, CHI14 and OAR9 (Reading (1980)) are longer than the cattle homologue BTA14 (Kaftanovskaya and Serov, 1994). Genetic linkage analysis demonstrated that there has been a translocation of the centromere of BTA9 to BTA14 to generate OAR9 (Crawford *et al.,* 1995; de Gortari *et al.*, 1998). The remainder of BTA9 is equivalent to OAR8 (Figure 5.1a).

There has also been a rearrangement in the band order and shape of the sheep and goat X chromosome compared with the cattle X chromosome (Evans *et al.*, 1973; Hayes *et al.*, 1991; Hediger *et al.*, 1991a; Ponce de León *et al.*, 1996; Hassanane *et al.*, 1998). Several models for the evolution of the sheep, cattle and goat X chromosomes from an ancestral buffalo- or goat-type chromosome have been proposed (Kaftanovskaya and Serov, 1994; Iannuzzi and Di Meo, 1995; Ponce de León *et al.*, 1996; Hassanane *et al.*, 1998). Chromosome painting probes showed that the bovine Xp arm is homologous to Xq3.2-Xq3.6 of sheep and goat chromosomes (following the band numbering of Ansari *et al.*, 1996 but referred to as Xq4 by Ponce de León *et al.*, 1996). In addition, bovine

Xq4.1-Xq4.3 (the terminal three bands) is homologous to Xp1.2-q1.2 in sheep and goat (Hassanane *et al.,* 1998), and it has been demonstrated by genetic linkage analysis and FISH that these bands correspond to the pseudo-autosomal region (Ponce de León *et al.*, 1996; Galloway *et al.,* 1996) (Figure 5.1b).

Some caprine microsatellite markers (eg. TGLA245, HEL13, OarVH98, OarCP9, OarHH22, BM4305) and the β-globin gene cluster (HBB) map to non-homologous chromosomes in cattle and sheep (Simi *et al.*, 1989; Vaiman *et al.*, 1996). Similarly, ovine markers BMS66, BMS719, BM3627, BM4439, BMS2840 and HEL6 map to linkage groups other than the expected cattle homologues. These differences may be due to as yet undefined translocations or more likely they are due to amplification of closely related sequences flanking a second locus (de Gortari *et al.*, 1998). Improving the density of genes and markers on the ovine physical map by FISH with BAC clones will help to define the boundaries of chromosomal rearrangements which will aid the search for positional candidate genes from the human and mouse maps.

5.1 Fluorescence *in situ* hybridisation

Standard sheep (Ansari *et al.*, 1996) and cattle ideograms (ISCNDA, 1990) were used in this study (Figure 5.2). Goat nomenclature followed cattle as recommended by Popescu *et al.* (1996), except for X, which followed sheep. A G-banded ovine karyotype was constructed using the sheep ideogram as a reference (Figure 5.3).

5.1.1 Biotinylation of BAC clones

Prior to physically mapping ovine BAC clones to metaphase sheep chromosomes by FISH, the identity of the BAC was confirmed by sequencing (section 4.5) to avoid making incorrect physical assignments. Once confirmed, the BAC clones were labelled with biotinylated-dATP by nick translation in the presence of trace tritiated-dATP. The % incorporation of tritium into the BAC DNA was measured by scintillation

counting as an estimate of the % of biotin incorporated. Routinely, 15-30% incorporation was achieved. However, if the DNA was poorly labelled (<10%), nick translation was repeated and if it failed for a second time, the DNA was re-isolated.

5.1.2 CotI DNA Preparation

CotI DNA was prepared from both sheep liver and thymus. A CTAB extraction method (Towner, 1991) was used to isolate the DNA from liver to avoid precipitating glycogen and other monosaccharides associated with the tissue matrix. A standard pronase digestion was used to extract DNA from thymus (Ausubel *et al.*, 1992). The recovery from thymus was 5mg DNA/g tissue (65mg total), whilst the yield from liver was only 0.2mg DNA/g tissue (3mg total). The yield from liver was inadequate for CotI DNA preparation since at least 20mg DNA was required as starting material. Spleen does not have the extensive tissue matrix associated with liver, and it was evaluated as an alternative source of DNA. A yield similar to thymus was recovered by pronase digestion (4.2mg DNA/g tissue).

The DNA was sheared by autoclaving to produce fragments with an average length of 600bp. The original protocol (A. Nesci, personal communication) recommended a 15' treatment at 121°C, 103Kpa. However, after 15' all of the DNA was <100bp. Various times from 1-10' were tried, and it was found that a 1' treatment produced fragments in the appropriate size range. After shearing, the DNA was treated with S1 nuclease and the final yield of CotI DNA was 13.35µg/µl (26.7mg total).

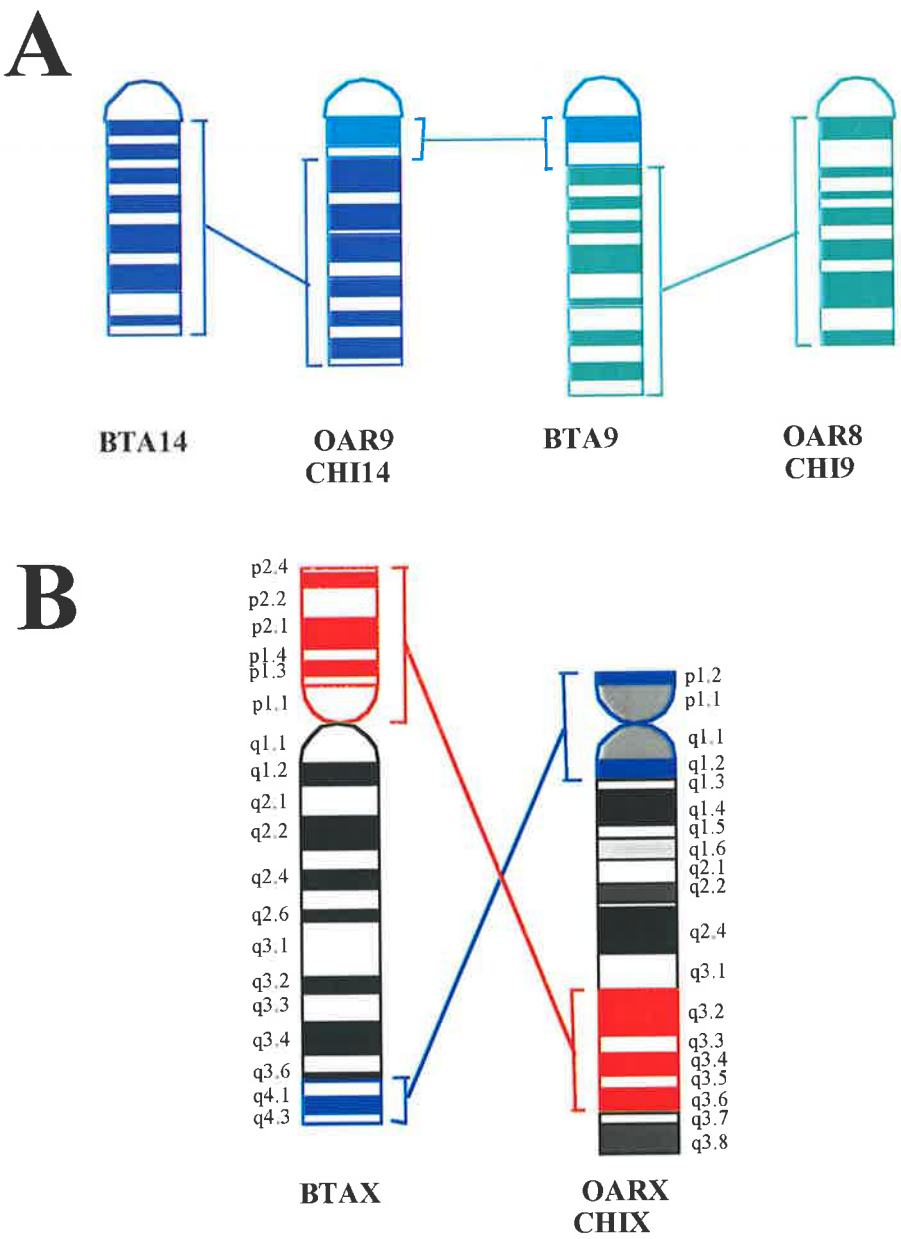


Figure 5.1: Rearrangements that have been described for cattle (BTA), sheep (OAR) and goat (CHI) chromosomes. A) Translocation of the centromeric region of BTA9 (light blue) to BTA14 (dark blue) to generate OAR8 and CHI14. The remainder of BTA9 corresponds to OAR9 and CHI9 (green). This translocation has been defined by genetic linkage analysis but the physical boundaries are yet to be defined by FISH. The bands included in the translocation in this schematic are as described by Kaftanovskaya and Serov (1994). B) Rearrangements that have occurred on the X chromosome. Microdissected chromosome paints were used to define these rearrangements (blue: Ponce de Le¢n *et al.*, 1996; red: Hassanane *et al.*, 1998). The blue region also corresponds to the pseudoautosomal region. The bands for BTAX are numbered according to ISCNDA (1990) whilst OARX follows the numbering system of Ansari *et al.*, 1996.

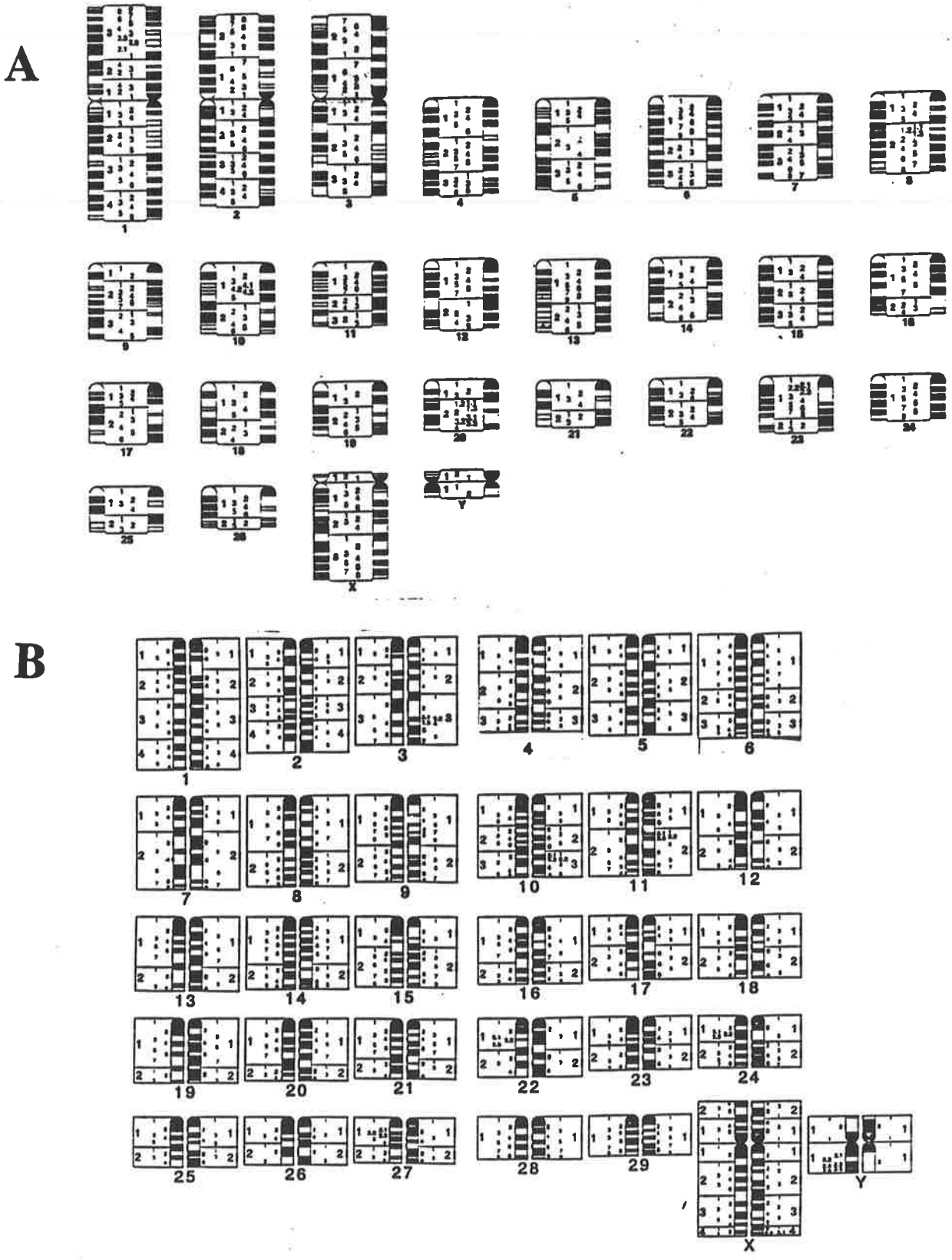


Figure 5.2: Ideograms used for chromosome identification. A) Ovine G-banded (left) and R-banded (right) karyotype (from Ansari *et al.*, 1996). B) Bovine G-banded (left) and R-banded karotype (from ISCNDA (1990) except that BTA4 and BTA6 have been interchanged to follow the Reading (1980) convention).

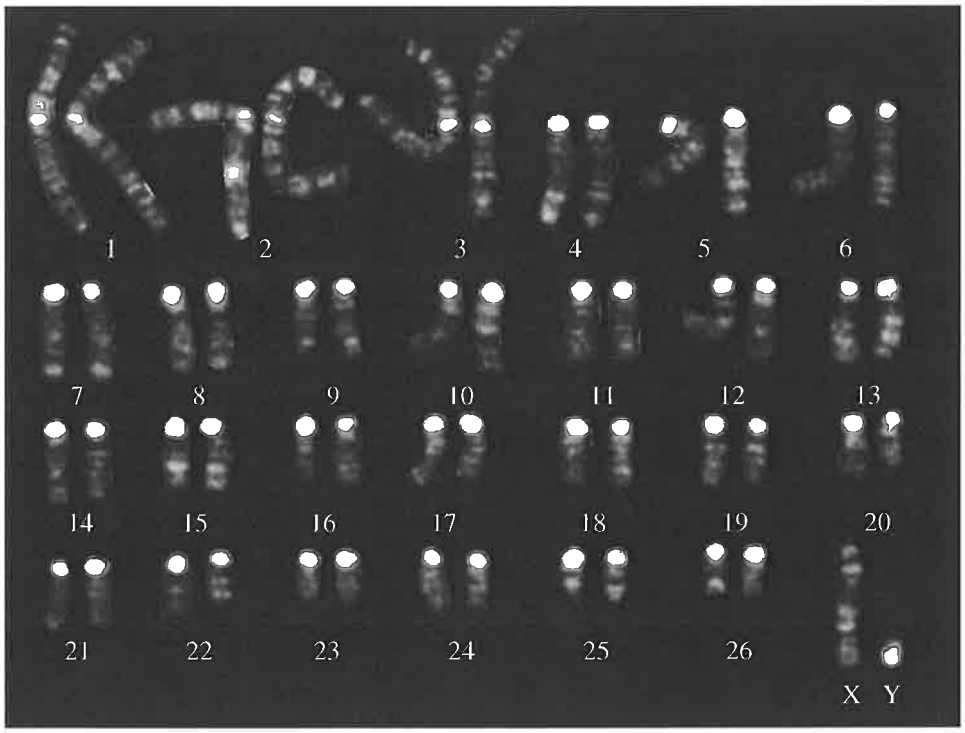


Figure 5.3: G-banded karyotype from a single metaphase spread of a normal ram. Bands were generated as described in section 2.1.2.6.

Table 5.1: Equivalent sheep, cattle and goat chromosomes according to the Texas nomenclature (Popescu *et al.*, 1996). The chromosome-specific reference markers recommended by Ansari *et al.* (1994) and Popescu *et al.* (1996) are shown.

Cattle	Sheep	Goat	Reference Markers
1	1q	1	SOD1, TF
2	2q	2	VIL1, FN1
3	1p	3	HSD3B, NGFB
4	4	4	INHBA
5	3q	5	IFNG, LALBA
6	6	6	CSNα, CSNβ
7	5	7	RASA, PDEA
8	2p	8	IFNA, INFW
9[a]	8	9	IGF2R, ESR
10	7	10	CYP19, HEXA
11	3p	11	LGB
12	10	12	RB1
13	13	13	IL2RA
14[a]	9	14	TG, CRH
15	15	15	FSHB
16	12	16	PIGR, LAMC1
17	17	17	FGG
18	14	18	GPI
19	11	19	GH1
20	16	20	MAP1B
21	18	21	IGHγ, GMBT16
22	19	22	LTG, RHO
23	20	23	(B)OLA
24	23	24	DSCI, MBP
25	24	25	ELN, PRKCB1
26	22	26	APT1, CYP17
27	26	27	DEFB
28	25	28	CGN1, TAC2R
29	21	29	LDHA, IGF2
X[b]	X	X	PGK1
Y	Y	Y	ZFY

[a]15cM of BTA9 has been translocated to BTA14 to form OAR9 (Figure 5.1a).
[b]Rearrangements in X chromosome structure have been defined (Figure 5.1b).

5.1.3 Suppression of repeats with CotI DNA

In the absence of CotI DNA, the repetitive DNA in the BAC clones hybridises uniformly to the chromosomes (ie. painting the chromosome arms) and masks the true location of the unique DNA in the clone. Consequently, the labelled BAC DNA was pre-annealed to ovine CotI DNA to suppress the repetitive DNA within the BAC clone and prevent it from hybridising to numerous sites in the genome. A CotI DNA concentration titration (2μg, 4μg, 6μg, 8μg, 10μg, 12μg) was performed. It was found that, in general, 8μg CotI DNA was sufficient to suppress the repeats within the BAC clone, whilst still producing strong hybridisation signal at a unique location on the chromosomes. Although 8μg was used initially, some BAC clones required less suppression (5μg) to reveal a hybridisation signal, whilst others required much more (25μg) to completely suppress the repeats (Table 5.2). Even with 25μg CotI, BAC 347R4C6 for OarVH116 was still under-suppressed and hybridised to numerous sites along the arms of several chromosomes, similar to a SINE painting pattern. It is possible that this clone contains a low-copy repetitive element that is not well represented in CotI DNA. Conversely, the BAC clones for BM3413 (168R8C5) and OBS (608R2C8) were over-suppressed with 5μg CotI DNA and no fluorescent signal was detected. It may be that these clones largely consist of repetitive DNA elements and contain very little single-copy DNA. BAC 210R5C6 for BM4107 was under-suppressed with 8μg CotI DNA and this FISH localisation needs to be repeated with more CotI DNA.

A dot blot of some of the BAC clones was probed with the BovA repetitive element (a SINE) and the microsatellite $(GT)_{30}$ to establish whether there was any relationship between the presence or absence of these repeats and the amount of CotI DNA required for adequate suppression (Figure 5.4). Given that the amount of DNA in each dot was equal (100ng), any difference in the intensity of the dots may indicate a difference in the length of the repeats (ie. a longer repeat may hybridise better than a

shorter repeat) or the number of repeats (ie. two separate repeats may hybridise more strongly than a single repeat). There was, however, no relationship between the amount of CotI DNA needed and the intensity of the dots probed with either the microsatellite ($r^2 = 0.0147$) or the SINE element ($r^2 = 0.0122$). Similarly, the intensity of the dots showed that there was no relationship between the presence of a microsatellite and a SINE in the same clone ($r^2 = 0.2369$). Only two BAC clones (ILSTS008 and BRN) failed to hybridise with either the microsatellite or the SINE. However, sequencing showed that the PCR product amplified from the clone ILSTS008 contained a $(CA)_7$ repeat. Similarly, the BAC for IGF1 contained a $(CA)_6AATA(CA)_8$ interrupted repeat, OarEL01 contained a $(CA)_7$ repeat and PRL contained a $(AC)_3TTT(CA)_3$ interrupted repeat but these also failed to hybridise to the $(GT)_{30}$ probe. It may be that the repeats are too short or that some secondary structure prevents hybridisation. The data from the dot blots, therefore, were not able to assist in predicting the amount of CotI DNA required for suppression.

5.1.4 Chromosome identification

Ovine fibroblast cell lines were the source of metaphase chromosomes for FISH. Since fibroblasts at metaphase do not adhere to the growing surface as strongly as interphase cells, the chromosome preparations were enriched for metaphases. In general, more than 90% of all cells on the slide were prophase or metaphase cells and 70% of these were suitable for physical mapping, being elongated and well-spread with few overlapping chromosomes.

For each BAC clone at least 20 metaphase cells were scanned under the microscope to establish the location of the fluorescent signal. Under optimum CotI DNA conditions, signal was observed in 90% of cells and on both chromatids of both chromosomes 70% of the time. A minimum of 3 well-banded metaphase spreads was scored from high-contrast black and white film to assign each BAC clone to a specific

chromosomal band (Table 5.2). Three BAC clones (158R2C6 for CD3D, 140R8C10 for DRB2 and 102R1C11 for PRL) could not be assigned because the metaphase spreads were poorly banded or not banded at all and these FISH localisations need to be repeated.

Propidium iodide (Lemieux *et al.*, 1992) stains the centromeres and telomeres brightly when the chromosomes are G-banded. This aided chromosome orientation which can be difficult using other stains (eg. Hoescht). However, it hindered band identification in some cases, particularly when the fluorescent signal was adjacent to the centromere (eg. BM6438) or the telomere (eg. OarJMP58). Indeed, it was impossible to visualise the signal on black and white film if the BAC was located at the centromere as for BM6438 and colour photographs were taken instead (Figure 5.5).

Occasionally, the chromosomes were simultaneously G-banded and R-banded. This was particularly noticable for the smallest chromosomes making identification problematic. It was, however, difficult to decide whether the small chromosomes were truly R-banded or whether they took on that effect because of the brightly staining centromeres and telomeres. In humans, background R-banding has been observed with the propidium iodide staining method even when the cells have not been arrested with BrdU and it may be an effect of the high pH of the PPD11 antifade mountant (G. Webb, pers. comm). To refine the localisation of several markers that mapped to bands adjacent to the centromere (eg. BM827 on OAR3, IL2RA on OAR13 and CSRD241 on OARX) or telomere (eg. MAF92 on OAR22),a different stain (eg. Hoescht) or banding method could be used.

In cases where chromosome identification was ambiguous or where there was no previous mapping information in sheep, cattle or goat, a co-localisation experiment was performed with two BAC clones that were expected to map to the same chromosome on the basis of linkage (section 5.3).

5.2 Integration of the ovine genetic linkage map and physical map

A total of 69 clones were isolated from the ovine BAC library with at least two BACs for every chromosome, except OAR8, 10, 11, 18 and 22. To date, 45 BACs (for 46 genes and markers) have been physically localised to discrete chromosomal bands by FISH (Figure 5.6). Each of these BACs mapped to the expected chromosome based on previous mapping information (Table 5.2). At least one BAC was mapped to each sheep chromosome, except OAR18. BAC clones 601R8C12 for TGLA122 and 168R8C5 for BM3413 have been isolated for OAR18 but these still need to be localised physically. McM58 (BAC 21R4C10) has been isolated for OAR1p, the only other cattle chromosome equivalent for which a BAC clone is yet to be localised by FISH. In addition, FISH localisations for a second BAC per chromosome still need to be performed for 13 chromosomes (OAR4, 8, 10, 11, 12, 14, 16, 17, 18, 19, 21, 22 and X).

Of the 69 clones that have been isolated, 50 clones are anchors for the genetic map and 19 are comparative mapping loci (ie. BAC clones isolated for gene sequences rather than microsatellites), with 9 of these comparative mapping loci acting as anchor loci as well. To date, 29 anchor and 17 comparative mapping loci (including 6 that also act as anchors) have been localised by FISH (Figure 5.6). The physically localised BAC clones directly anchor and orient the linkage maps with respect to the physical maps for every sheep chromosome, except OAR18 and OAR2 since neither FN1 nor LPL have been mapped by genetic linkage analysis. The genetic maps for OAR2 and OAR18 are already anchored by NRAMP1 and GMBT16, respectively (Georges *et al.,* 1991; Pitel *et al.,* 1995).

The location of GAPR, which was previously unmapped, was confirmed by dual localisation (section 5.3). Perforin (PRF) mapped to 25q1.4-2.1 which has conserved synteny with a segment of human chromosome 10. ILSTS011 mapped to 9q2.1 and this defines the boundary of the translocation between cattle 14 and 9.

5.3 Dual localisation of BAC clones

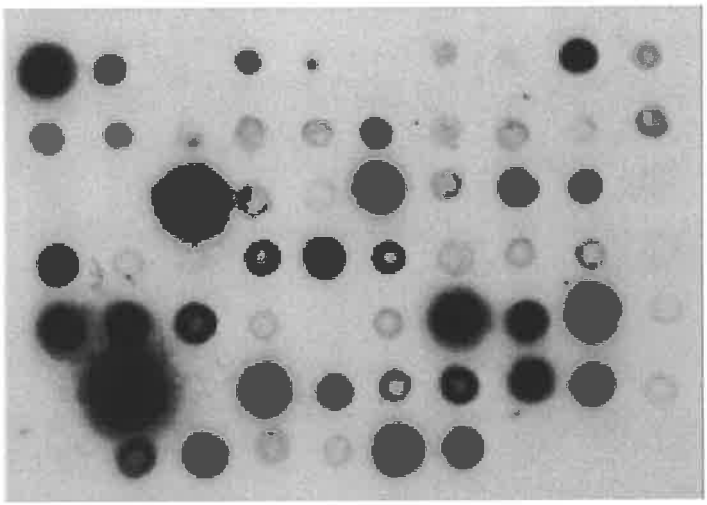
To avoid making incorrect mapping assignments, dual localisation experiments were performed if the chromosome identification by G-banding was ambiguous or if there was no previous mapping information in sheep, cattle or goat. For instance, it was confirmed that GAPR mapped to 5q3.5 by simultaneously localising LDLR to 5q1.3 (Figure 5.7). Dual localisations were also performed for LPL and FN1 on OAR2, RNASE6 (formerly SRN) and ILSTS005 on OAR7, and IL2RA and GHRH on OAR13. These co-localisation experiments verified that each of these pairs of BAC clones did map to the same chromosomes. Adequate suppression was only achieved if the amount of CotI DNA required to suppress the repeats in each clone was included in the mixture. For example, 8μg CotI DNA suppressed the repetitive DNA in the BAC clones for GAPR and LDLR when they were hybridised separately. Consequently, 16μg CotI DNA was required for adequate suppression in the dual FISH experiment.

Figure 5.4: Dot blots of BAC clones probed with repetitive DNA elements. Each dot contained 100ng DNA. A) The loading scheme indicating the loci known to be contained within each of the BAC clones. B) Hybridisation with $(GT)_{30}$. C) Hybridisation with the BovA SINE element. The probes were labelled with ^{32}P-dATP and the autoradiographs were exposed overnight.

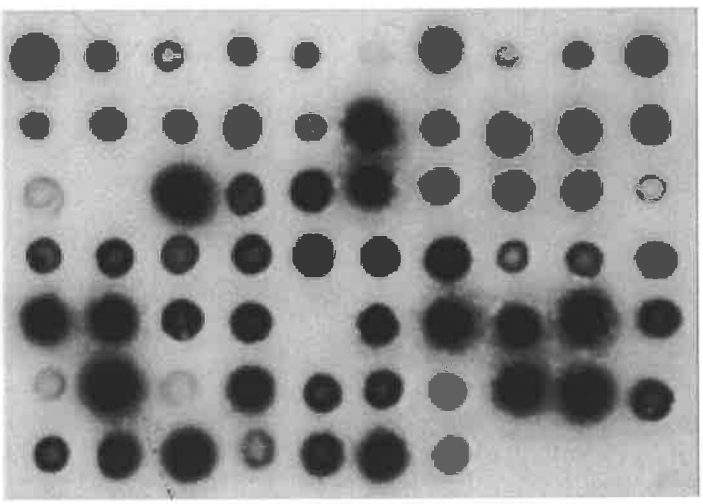
A

BM6438	DYB	LDLR	BM6526	IGF1	ILST008	IL3	BM1329	Perforin	IFNG
BM719	DRB1	564r6c6	IL2RA	MAF92	JMP8	ILSTS011	154r4c6	BM1824	FSHB
OarEL01		CSRD287	252r8c10	SCYA	361r6c9	IL5/IRF1	BM1227	BM4107	PRL
JMP58	MAF35	169c11r5	KRT2.13	ILSTS05	LPL	CSRD24	OarVH72	SRN	CSF2
CSRD270	KAP1.1	DRB2	547r4c7	BRN	LGB	CD3D	VH116	KAP1.1	BRN
578r2c8	OarHH64	KRTAP6	OB	574r7c5	CSSM43	OCAM	ADCYC	FN1	GAPR
IL6	FCB48	BM1303	ILSTS43	JMP8	BM3413	+ve	$(GT)_{30}$	Vector	-ve

B



C



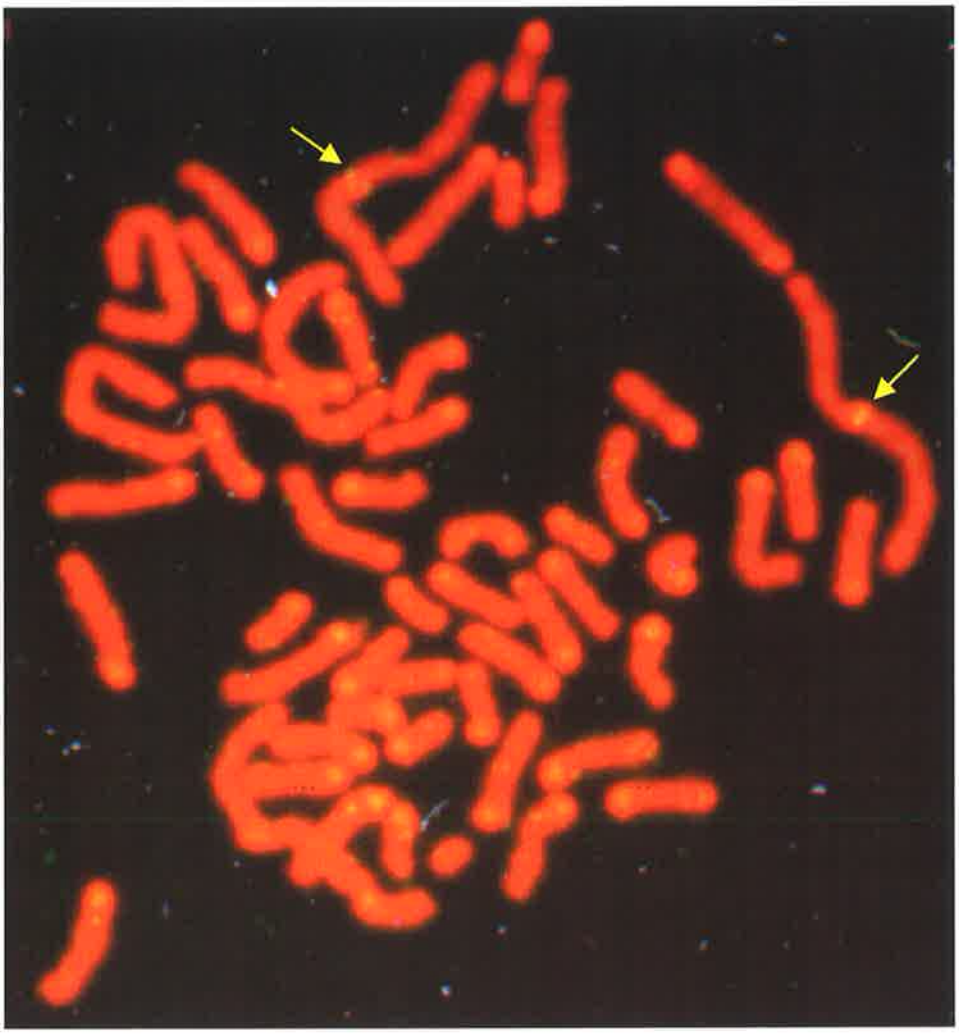


Figure 5.5: Localisation of a BAC clone for BM6438 to 1q1.1-1.3 by FISH to G-banded sheep metaphase chromosomes. The fluorescent signal produced by BM6438 was photographed under blue epifluorescence using Fugichrome colour positive film set at ASA 1600 with a further correction of 1.6.

Table 5.2: Physical map locations of ovine BAC clones localised to G-banded metaphase chromosomes by fluorescence *in situ* hybridisation. Published mapping information for each locus and the method that was used to make the chromosomal assignment are listed (L: linkage, SCH: somatic cell hybrids, FISH: fluorescence *in situ* hybridisation). The amount of CotI DNA required to suppress repeats within the BAC clones is indicated (OS: over-suppressed, US: under-suppressed). The number of chromosomes that were scored to assign the bands is in parentheses (NA: not assigned, ND: not done).

Locus	[1]Published Map Location	Mapping Method	BAC	CotI	Physical Location of BAC clone	
ADCYC	15	L	443R7C6	8µg	15q1.4-2.2	(4)
BM719	12	L	296R6C9	25µg	12q2.2-2.3	(7)
BM827	3	L	153R6C4	8µg	3p1.2-1.5	(4)
BM1303	19	L, SCH	139R7C8	25µg	19q1.2-1.3	(5)
BM1329	6	L, SCH	5R3C4	8µg	6q1.6-1.8	(8)
BM1824	1q	L, SCH	284R5C4	25µg	1q4.2	(8)
BM3413	18	L	168R8C5	5µg	OS	
BM4107	16	L	210R5C6	8µg	US	
BM4208	8	L, SCH	408R8C6	8µg	8q2.3-2.5	(8)
BM6438	1	L, SCH	321R4C8	25µg	1q1.1-1.3	(6)
BM6526	26	L, SCH	304R4C7	5µg	26q1.2-1.3	(6)
CD3D	[2]15	[2]L	158R2C6	8µg	NA	
CSF2	5q1.3-1.5	FISH	295R7C7	8µg	5q1.3-1.5	(5)
CSRD241	X	L	22R8C7	12µg	Xq1.1-1.4	(4)
CSRD270	14	L	306R8C4	8µg	14q1.3-1.4	(7)
CSRD287	10	L	100R4C3	25µg	10q1.3	(6)
CSSM043	26	L, SCH	475R8C2	8µg	ND	
DRB1	[2]20	[2]L	569R6C11	8µg	20q2.2	(10)
DRB2	[2]20	[2]L	140R8C10	8µg	NA	

Table 5.2 continued

Locus	[1]Published Map Location	Mapping Method	BAC	CotI	Physical Location of BAC clone	
DYB	[2]20	[5]L	78R3C6	8µg	20q3.2-3.3	(4)
FN1	2q4.1-4.4	FISH, SCH	597R6C12	8µg	2q4.1	(10)
FSHB	15q2.4-qter	L, ISH	94R8C7	12µg	15q3.2	(5)
[3,4]GAPR	UNMAPPED	-	560R5C11	8µg	5q3.4	(10)
GH2	11	HSA17 BY FISH	382R1C2		ND	
[4]GHRH	13	BTA13 BY L	327R8C12	8µg	13q2.4	(10)
HUJ614	12	L, SCH	364R5C10		ND	
IFNG	3q2.3	FISH	237R5C6	25µg	3q2.3	(8)
IGF1	3	SCH	402R7C8		ND	
IL2RA	13q1.2-1.5	L, FISH	43R5C3	25µg	13q1.3	(6)
IL3	5q1.3-1.5	FISH	319R7C9	8µg	5q1.5	(4)
IL5/IRF1	5q1.3-1.5	FISH	233R6C6	8µg	5q1.5	(7)
IL6	[2]4	[2]L	290R2C10	25µg	4q1.4-1.5	(4)
ILSTS005	7	L, SCH	220R8C7	8µg	7q3.6-qter	(6)
ILSTS008	9	L, SCH	337R8C6	8µg	9q2.6	(6)
ILSTS011	9	L, SCH	35R2C11	25µg	9q2.1	(12)
ILSTS043	24	L	292R3C5	8µg	24q1.2-1.3	(4)
KAP1.1	11	L	164R5C4	25µg	11q3.1	(5)
KAP6.1	1	L	516R8C6		ND	
KRT2.13	3	L	374R2C7	25µg	3q2.1	(6)
[4]LDLR	5	BTA7 BY L	192R4C7	8µg	5q1.3	(8)
LGB	3p2.7	FISH	103R8C4		ND	
[4]LPL	2	SCH	323R6C9	8µg	2p2.4	(10)
MAF23	3	L, SCH	68R6C1		ND	
MAF35	23	L, SCH	249R5C9	8µg	23q1.7-2.1	(5)
MAF92	22	L, SCH	87R3C2	8µg	22q2.3-qter	(10)
MAF214	16	L	68R6C1	8µg	16q1.7	(4)

Table 5.2 continued

Locus	[1]Published Map Location	Mapping Method	BAC	CotI	Physical Location of BAC clone	
McM58	1	L	21R4C10		ND	
McM111	19	L	140R4C7		ND	
McM136	23	[2]L	39R7C7	8µg	23q1.7-2.1	(4)
OarEL01	24	L, SCH	141R6C6	25µg	24q1.8	(6)
OarFCB48	17	L, SCH	220R5C3	8µg	17q2.1-2.3	(4)
OarHH56	20	L, SCH	605R5C8		ND	
OarHH64	4	L	537R7C4		ND	
OarJMP8	6q3.4-3.5	L, SCH, FISH	547R7C11	8µg	6q3.4-3.5	(6)
OarJMP58	26	L, SCH	535R4C6	8µg	26q2.1-qter	(6)
OarVH72	25	L, SCH	396R6C11	8µg	25q1.2-1.3	(5)
OarVH110	21	L, SCH	298R1C1		ND	
OarVH116	17	L	347R4C6	25µg	US	
OBS	4	SCH	608R2C8	5µg	OS	
OCAM	21	L, SCH	343R8C1	8µg	21q2.1-2.2	(4)
PRF	25	HSA10q22 BY ISH	187R7C3	25µg	25q1.4-2.1	(9)
PRL	19	BTA23 BY L	102R1C11	8µg	NA	
[4]RNASE6	7	SCH	79R1C6	8µg	7q1.5	(6)
SCY@	[2]11	[2]L	131R7C4		ND	
SRY	Y	SCH	574R7C5	5µg	Yp1.1-1.2	(6)
TGLA122	18	L	601R8C12		ND	(5)
UWCA5	14	L, SCH	59R6C5		ND	
ZFX/ZFY	Xp1.3/Yp1.2-1.3	SCH, FISH	602R4C4		ND	

[1]As listed in the ArkDB sheep and cattle genome databases (www.ri.bbsrc.ac.uk)
[2]Maddox *et al.* (1996)
[3]Unmapped in sheep, cattle, goat, human and mouse
[4]FISH performed by L. Allen (1997)
[5]Maddox *et al.* (1998)

5.4 Estimation of genome size

It was possible to estimate genome size by measuring the physical distance between anchor BAC clones and by comparing it to the genetic distance (Ferretti *et al.*, 1997). The genetic linkage maps (Maddox *et al.*, 1996; de Gortari *et al.*, 1998) for 10 chromosomes were anchored with 2 BAC clones (Figure 5.6). There are now 17 chromosomes that have 2 anchors if previously mapped anchors are included. The physical distances between anchors were evaluated (as in Ferretti *et al.*, 1997) by measuring the length spanning hybridisation peaks, defined by band assignments (Lichter *et al.*, 1990). When anchors were assigned to more than one band the central position was taken. The length of the interval was estimated as a percentage of the total karyotype.

To estimate genome size, only 9 pairs of anchors that each represented more than 1% of the genome were considered (Table 5.3). Consequently, RP0011-CSRD287 on OAR10 (0.62%), FSHB-ADCYC on OAR15 (0.88%), BM4107-MAF214 on OAR16 (0.62%), FGF2-OARFCB48 on OAR17 (0.53%), DYB-DRB1 on OAR20 (0.53%), IGF2-OCAM on OAR23 (0.18%) and MAF35-McM136 on OAR23 (0.35%) were not included in the calculation. The two anchors for OAR14 (GP1 and CSRD270) are not integrated into the same map, so a genetic distance between them could not be calculated. GP1 is an anchor for the second-generation ovine linkage map (de Gortari *et al.*, 1998) and CSRD270 was mapped by Maddox *et al.* (1996). All the genetic distances used to estimate genome size are from de Gortari *et al.* (1998), except for the anchors for OAR4 (TCRB-IL6) which are from Maddox *et al.* (1996). Both maps for OAR4 are 130cM long, so the relative distances between markers should be similar and should not influence the estimation of genome size greatly.

Figure 5.6: Physical assignments of 46 BAC clones on sheep chromosomes. Green indicates existing anchors; comparative mapping loci are in blue and the set of anchor BAC clones mapped herein are in red. Green underlining indicates the anchor already existed and comparative mapping loci underlined with red are also anchors.

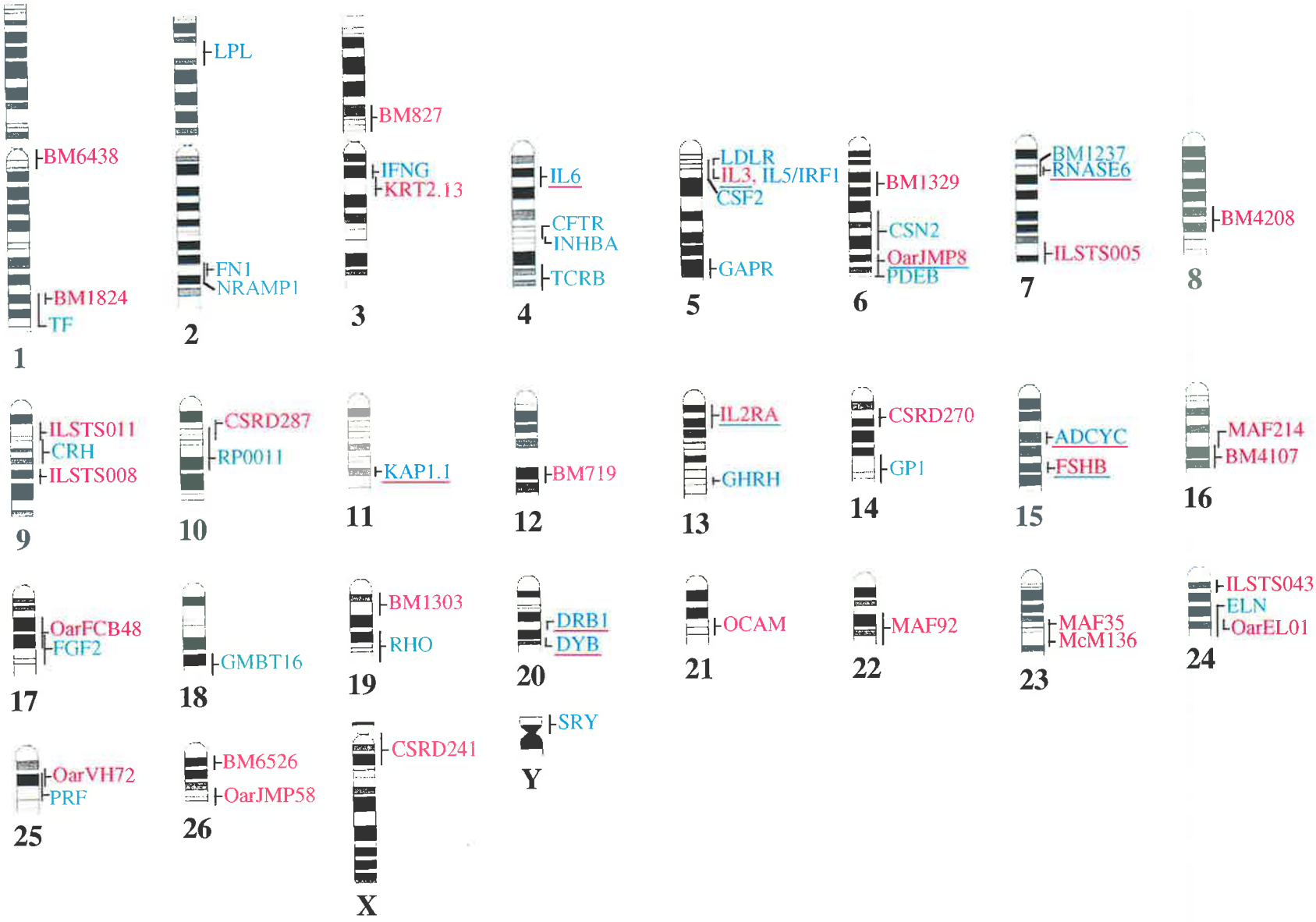

BM6438
BM1824
TF
1
LPL
FN1
NRAMP1
2
BM827
IFNG
KRT2.13
3
IL6
CFTR
INHBA
TCRB
4
LDLR
IL3, IL5/IRF1
CSF2
GAPR
5
BM1329
CSN2
OarJMP8
PDEB
6
BM1237
RNASE6
ILSTS005
7
BM4208
8
ILSTS011
CRH
ILSTS008
9
CSRD287
RP0011
10
KAP1.1
11
BM719
12
IL2RA
GHRH
13
CSRD270
GP1
14
ADCYC
FSHB
15
MAF214
BM4107
16
OarFCB48
FGF2
17
GMBT16
18
BM1303
RHO
19
DRB1
DYB
20
OCAM
21
MAF92
22
MAF35
McM136
23
ILSTS043
ELN
OarEL01
24
OarVH72
PRF
25
BM6526
OarJMP58
26
CSRD241
X
SRY
Y

Table 5.3: Comparison of the ovine physical and genetic maps to estimate genome size

OAR	Markers	Length (% genome)	[1]Genetic distance sex ave. (cM)	Inferred genome size sex ave. (cM)
1	BM1824-BM6438	4.26	153.2	3592
3	KRT2.13-BM827	2.09	42.1	2017
4	TCRB-IL6	3.09	[2]81.5	2642
6	OARJMP8-BM1329	2.90	82.5	2841
7	RNASE6-ILSTS005	2.54	88	3463
9	ILSTS008-ILSTS011	1.45	27.9	1922
19	RHO-BM1303	1.27	53.7	4227
24	OAREL01-ILSTS043	1.09	31.4	2884
26	OARJMP58-BM6526	1.09	34.5	3168
	TOTAL	19.78	594.8	3007

[1]de Gortari *et al.* (1998)
[2]Maddox *et al.* (1996)

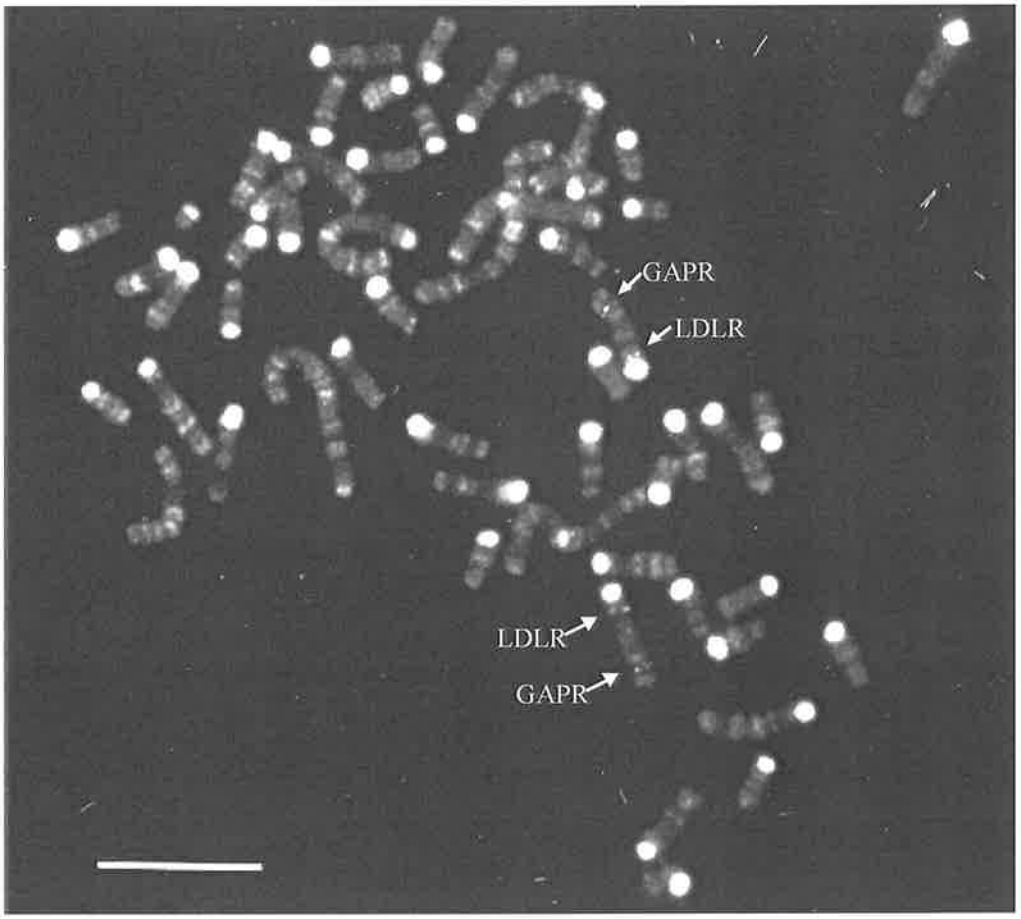


Figure 5.7: Dual localisation of ovine BAC clones for LDLR and GAPR to 5q1.3 and 5q3.5, respectively. Bar represents 10□M.

The intervals for the 9 chromosomes that could be considered cover 19.8% of the whole genome. The total genetic distance covered by the intervals between markers is 595cM, giving an estimate for the whole genome of 3006cM. Estimates of genome size obtained by considering individual intervals vary considerably (Table 5.3), but the mean of 2985cM (± 266 SEM) is similar to the minimum genome size of 2770cM predicted by chiasmata counts (Chapman and Bruere, 1977).

5.5 Discussion

The assignment of 45 BAC clones for 46 genes and markers to G-banded ovine metaphase chromosomes by FISH increases the number of physical assignments in sheep to 475 with 137 loci now mapped by ISH methods. This represents a 52% increase in the number of markers mapped to discrete chromosomal bands. The assignments of IL3, OarJMP8, IL2RA and FSHB to 5q1.5, 6q3.4-q3.5, 13q1.3 and 15q3.2, respectively, validate previous localisations (Hediger *et al.,* 1991b; Pearce *et al.,* 1994; Ansari *et al.,* 1995; Hawken *et al.,* 1996). The physical positions of IL3, IL2RA and FSHB have been refined and they are each now assigned to a single band.

An additional 41 new assignments by FISH were made but several BAC clones were not assigned because they were either over- or under-suppressed. In such cases, overlapping BAC clones could be isolated by plasmid end-rescue to build a contig of the immediate region by chromosome walking (Cai *et al.,* 1995). These newly identified clones may have sufficient unique DNA sequence to enable FISH hybridisation to a discrete location, thereby localising the marker or gene of interest.

For those clones that were localised to G-banded chromosomes, there was no obvious association between whether a BAC clone was isolated for a gene or marker and whether it mapped to a light or dark G-band. Several of the assignments were clustered in the same band or adjacent bands. INFG and KRT2.13 mapped to 3q1.4 and 3q2.1, respectively. LDLR mapped to 5q1.3, and the BAC clones for IL3, IL5 and IRF1

mapped to 5q1.5. CSF2 was localised to 5q1.3-q1.5. The order of these clustered markers could be ascertained by multiple colour FISH (Schläpfer *et al.,* 1998; Gallagher *et al.,* 1998; Gallagher *et al.,* 1999) or fibre-FISH on extended chromatin fibres (Brunner *et al.,* 1998; Liu *et al.,* 1998; Horelli-Kuitunen *et al.,* 1999).

A number of markers were localised to more than one band. Most of these broad localisations were due to the problems associated with band identification near the brightly staining centromeres and telomeres. In most cases these assignments could be refined by using an alternative staining strategy. Some localisations, however, are unlikely to be refined, because genes that are part of large families may be clustered at those locations. Genes that arise by duplication events share significant sequence identity, and therefore, may all hybridise to the fluorescent probe, causing a broader localisation than expected for a single gene.

The physical locations of 36 of the BAC clones anchor and orient all the ovine genetic linkage maps, except OAR2 and OAR18 (Maddox *et al.,* 1996; de Gortari *et al.,* 1998). OAR18 will be anchored once the BAC clones for TGLA122 and BM3413 have been localised by FISH because these markers are already on the genetic linkage map. LPL and FN1 for OAR2 have not been mapped by genetic linkage analysis. The BAC clones for these loci both hybridised strongly with the $(GT)_{30}$ probe indicating that in each of these clones there is at least one microsatellite (Figure 5.4). Similarly, BAC clones for the other comparative mapping loci, except GHRH and SRY which were not assessed, all hybridised to the $(GT)_{30}$ probe. By isolating polymorphic microsatellites from these clones (eg. by shot-gun subcloning (Cai *et al.,* 1995; Schibler *et al.,* 1998a)) and then scoring them across the IMF animals, the comparative mapping loci could be integrated into the genetic map.

The set of anchor BAC clones that were isolated in this study can be used for unambiguous identification of ovine chromosomes. The availability of this set of clones means that researchers with minimal cytogenetic experience of the sheep karyotype can

readily make new, accurate physical assignments in sheep. It is important that prior to the release of this set of clones as a mapping tool, the current assignments are verified in some way. So far, four pairs of BAC clones have been co-localised by dual FISH confirming that both BAC clones map to the same chromosome. The remaining pairs should be mapped by dual FISH to guarentee that these BAC clones map to the same chromosomes and there has been no mis-identification based on G-banding patterns. Alternatively, microsatellites isolated from these BAC clones could be mapped by genetic linkage analysis. For example, three microsatellites have been isolated from the BAC containing perforin (PRF) and all three mapped to OAR25, thereby validating the FISH assignment of the clone to 25q1.4-q2.1 (J. Maddox, pers. comm.).

When this study began, the most comprehensive genetic linkage map that was available for sheep was developed by Maddox *et al.* (1996). Whenever possible, markers were selected from this map to isolate BAC clones to develop the set for unambiguous chromosome identification. The physical mapping data revealed that the genetic linkage maps for 11 chromosomes (OAR5, 7, 8, 9, 10, 15, 17, 18, 20, 23 and 26) were inverted with respect to the centromere. In the second generation ovine genetic linkage map (de Gortari *et al.*, 1998), all 26 autosomes were anchored by direct physical assignment in sheep or indirectly using comparative mapping data from cattle. The new physical assignments of ovine BAC clones containing anchor loci verify directly that these ovine genetic maps are oriented correctly with respect to the centromere.

The addition of 36 anchor BAC clones to the ovine physical map allowed a preliminary evaluation of the alignment between the physical and genetic maps. Markers on 16 chromosomes defined intervals representing 24.4% of the total karyotype. Only 9 intervals that each represented >1% of the genome were considered in the estimation of genome size to minimise the bias due to differences in the rate of recombination along the length of the chromosomes. The rate of genetic recombination is lower at the centromere than at the telomere, so the relationship between the physical

and genetic maps is not uniform along the length of a chromosome (Bray-Ward *et al.*, 1996). Near the centromere, short genetic distances correspond to large physical distances, whereas near the telomere, large genetic distances cover short physical distances.

The Kosambi map distance covered by the 9 intervals is 595cM, giving an estimate for the whole genome of 3006cM (section 5.4). The interval on OAR1 covers 80% of the q-arm and the estimate of the entire genome is similar (3592cM). The smallest estimates were for OAR3 (2017cM) and OAR9 (1921cM). The intervals on these chromosomes do not include any telomeric markers and the region on OAR3 spans the centromere. The mean estimate obtained by considering individual intervals was 2976cM (± 246 SEM) and is similar to the minimum genome size of 2770cM predicted by chiasmata counts (Chapman and Bruere, 1977). The current length of the sheep linkage map is 3190cM and measures have been taken to minimise inflation of genetic distances due to typing errors (de Gortari *et al.*, 1998).

Although only a fraction of the total genome was considered in the present calculation, there is good correspondance with previous estimates of sheep genome size (26 autosomes, X, Y). Similar estimates have also been made for other species. The bovine genome was estimated to be 2800cM by chiasmata counts (Logue and Harvey, 1978) and the two bovine genetic maps are 2900cM and 3540cM, respectively (Kappes *et al.*, 1997; Barendse *et al.*, 1997). Ferretti *et al.* (1997) estimated that the bovine genome (29 autosomes, X, Y) was 4176cM (±1360 SEM) using intervals that covered more than 40% of the genome. The large standard errors may reflect imprecise FISH assignments and this estimate overlaps previous estimates at the extremes of the range. Pig (18 autosomes, X, Y), mouse (19 autosomes, X, Y), human (23 autosomes, X, Y) and goat (29 autosomes, X, Y) genome sizes have been estimated as 2000cM, 1600cM, 3800-4000cM and 3250cM respectively (Davisson and Roderick, 1989; Ellegren *et al.*, 1994; Ferretti *et al.*, 1997). As proposed by Ferretti *et al.* (1997), these estimates

suggest that there is a tendency for genome size to increase with the number of chromosomes, rather than the size in base pairs. A better alignment of the ovine physical and genetic maps will be possible once more ovine BAC clones are isolated and mapped by FISH. An alternative strategy would be to physically localise characterised bovine and caprine BAC clones (Cai *et al.*, 1995; Schibler *et al.,* 1998a) on ovine chromosomes. These additional assignments will enable a more robust estimate of genome size, particularly if each of the terminal genetic markers are physically mapped.

In closely related species, such as sheep, cattle and goats, it has been demonstrated that about 50% of microsatellites from one species are polymorphic in the others (Crawford *et al.,* 1995; Vaiman *et al.,* 1996; de Gortari *et al.,* 1998). The second generation ovine linkage map was developed using predominantly anonymous, bovine microsatellite markers (402/504) to enhance the comparative maps between these livestock species. This will assist in the map-based search for the genes controlling economically important traits by enabling direct comparisons of regions with conserved marker order between livestock species. One problem with this approach is that very few microsatellites associated with genes that are anchors to the human or mouse maps are integrated into the ovine genetic map. Consequently, the current strategy for genetic map development in sheep does not take advantage of the vast amounts of comparative mapping information available from the gene-rich human and mouse maps.

Zoo-FISH chromosome painting experiments have identified which chromosomal segments in cattle and sheep are homologous to individual human chromosomes (Hayes *et al.*, 1995; Solinas-Toldo *et al.,* 1995; Chowdhary *et al.,* 1996; Iannuzzi *et al.,* 1999). However, since so few type I anchor loci are mapped in sheep, it is not yet possible to determine whether gene order within these syntenic blocks is conserved or whether numerous minor rearrangements have occurred. Recently, bovine BAC clones have been used to build framework comparative maps for a number of chromosomes (Schläpfer *et*

al., 1998; Gallagher *et al.,* 1998; Sonstegard *et al.,* 1998; Larsen *et al.,* 1999; Gallagher *et al.,* 1999). Similarly, a comprehensive goat cytogenetic map was developed using 160 caprine BAC clones for genes (Schibler *et al.,* 1998b). These FISH experiments revealed numerous chromosomal rearrangements within small genomic regions, demonstrating the need for dense comparative maps to efficiently utilise human and mouse mapping data in the search for economically important genes.

The location of breakpoints in conserved synteny between the ovine and human comparative maps need to be defined to permit accurate selection of positional candidate genes (Sonstegard *et al.,* 1998). This study has demonstrated the ease with which ovine BAC clones for microsatellites and genes can be isolated and physically localised to discrete chromosomal bands by FISH. A concerted effort should now be made to isolate and physically localise more BAC clones containing sequences orthologous to human or mouse genes to help define chromosomal breakpoints, critical regions, and ultimately, identify positional candidate genes for traits of interest. In addition, these BAC clones should be simultaneously mapped in cattle and goats to develop fully integrated comparative maps for these species.

CHAPTER 6

COMPARATIVE MAPPING USING OVINE BAC CLONES

COMPARATIVE MAPPING USING OVINE BAC CLONES

Comparative mapping is the localisation of orthologous genetic loci in multiple species (Sun *et al.*, 1997). Comparative mapping is possible because conserved synteny of homologous genes has been observed in species from diverse mammalian orders (Lyons *et al.*, 1997). The principle objective of comparative mapping studies in livestock species such as sheep, cattle, and goats is to elucidate regions containing economically important genes, using information from other livestock maps and from the gene-rich human and mouse maps. Localising markers associated with quantitative trait loci to specific chromosomal regions allows candidate genes to be identified from the other maps based on location and function.

Sheep, cattle and goat chromosomes are highly conserved as demonstrated by very similar banding patterns (Iannuzzi and Di Meo, 1995; Hayes, 1995). There is, however, little similarity between the banding patterns of these chromosomes and those of human and mouse (Hayes, 1995). Lack of conservation of bands suggests that there have been many intrachromosomal and interchromosomal rearrangements since the divergence of these species from a common ancestor. By mapping many homologous genes across species, these rearrangements will be better defined, allowing the evolutionary history of chromosomes to be traced back to an ancestral type (Womack and Kata, 1995).

Livestock genetic linkage maps have largely been constructed using polymorphic, type II, microsatellite markers. In general, there is insufficient homology in the sequence flanking microsatellites for these regions to be amplified across different mammalian orders, and consequently, they are not useful as comparative mapping markers (O'Brien *et al.*, 1993). It is type I loci (genes) that are most useful for comparative mapping because these coding regions are conserved in mammals, so orthologous chromosomal regions can be defined across species. Unfortunately, few type I loci are included in the current livestock

linkage maps, largely due to the lack of polymorphisms within coding sequences. Alternative strategies have been employed to identify homologous regions between species and to integrate genes into livestock genomic maps for comparative mapping purposes.

Early comparative mapping efforts involving livestock species relied on synteny mapping using somatic cell hybrids (Womack and Moll, 1986; Womack and Kata, 1995). Although synteny mapping can reveal which genes are grouped together on a chromosome, it does not elucidate gene order. Comparison of human, mouse and cattle syntenic maps reveals segments of conserved synteny between these species. Interchromosomal rearrangements that occurred since the divergence of humans and cattle from a common ancestor can be inferred from the syntenic maps (Womack and Kata, 1995).

A more recent strategy to investigate interchromosomal rearrangements is ZOO-FISH with heterologous probes (Scherthan *et al.,* 1994). Human chromosome-specific painting probes have been hybridised to the chromosomes of numerous mammalian species to delineate regions of conserved synteny (reviewed by Wienberg and Stanyon, 1997). Human paints have been hybridised to cattle (Chowdhary *et al.,* 1996; Hayes, 1995; Solinas-Toldo *et al.,* 1995), pig (Fronicke *et al.*, 1996; Goureau *et al.,* 1996; Rettenberger *et al.,* 1995) and sheep chromosomes (Iannuzzi *et al.*, 1999) to reveal 46-50, 47, and 48 regions of conserved synteny, respectively. Burkin *et al.* (1997b) used sheep chromosome paints on Indian muntjac deer chromosomes and compared the hybridisation patterns to previous studies using human chromosome paints (e.g. Yang *et al.,* 1997) to predict conserved chromosomal segments between sheep and human genomes. Each of these studies demonstrated that several human chromosomes paint whole sheep, cattle and pig chromosomes, whilst others paint two or three different chromosomes. In several cases, complex rearrangements must have occurred to generate the hybridisation pattern that is visualised. For instance, OAR7 and BTA10 are painted by seven alternating segments of HSA5, 14 and 15, respectively

(Hayes, 1995; Iannuzzi *et al.,* 1999). Neither the centromeres nor the Y chromosome hybridise to the human paints indicating the presence of divergent repetitive elements. Human X, 13 and 17 are conserved intact and paint a single chromosome in every mammalian species that has been tested (Womack, pers. comm.). It has been hypothesised that the evolutionary arrangement of the X chromosome is restricted due to inactivation mechanisms (Ohno, 1973). There is, however, no obvious explanation for the conservation of HSA13 and 17. It is noteworthy that both of these chromosomes carry clusters of tumour suppressor genes which regulate expression of other adjacent genes (Papp *et al.*, 1996; Phelan *et al.*, 1995; Gardiner *et al.*, 1997; Panayiotidis *et al.,* 1997).

The hybridisation pattern of human paints on sheep chromosomes is very similar to that visualised for cattle (Iannuzzi *et al.,* 1999), but the comparison is more complex for distantly related artiodactyls. For example, pig chromosome specific paints were recently hybridised to bovine chromosomes revealing 44 segments of conserved synteny between these two species (Schmitz *et al.,* 1998).

Although chromosome painting reveals regions of conserved synteny and the interchromosomal rearrangements that have occurred, it cannot detect intrachromosomal rearrangements or reveal gene order within the conserved regions. Mapping comparative anchor tagged sequences (CATS) across numerous species has been proposed as one way to resolve this issue (O'Brien *et al.,* 1993; Lyons *et al.,* 1997; Chen *et al.,* 1999). A set of 410 evolutionarily conserved primer pairs, evenly distributed throughout the human and mouse genomes, were designed in exons with the amplification product spanning an intron. Polymorphisms are more likely in intronic sequence enabling CATS to be incorporated in linkage maps. CATS primers were tested on 20 mammals from 11 orders and successful amplification was achieved with 32-75% of the primers depending on the species. Although there has been limited success with these primers, for species with very few

mapped genes, any additional data is welcome. At present, the only comparative maps that have been developed using CATS are synteny maps (Lyons *et al.*, 1997). Again, these maps do not address the issue of whether gene order is conserved within regions of conserved synteny. As mapping tools that do not require polymorphisms (eg. radiation hybrids, FISH mapping with large-insert clones) become more widely available, CATS will probably be superseded.

There is already only limited reliance on CATS for comparative mapping. Instead, primer pairs to amplify conserved sequences (ie. exons) in other type I loci are being selected from human gene mapping databases (e.g. traced orthologous amplified sequence tags (TOASTs), Jiang *et al.*, 1998). Preliminary linkage or physical (FISH or RH) maps that incorporate genes for human homologues have been developed for several pig (Sun *et al.*, 1999; Van Poucke *et al.*, 1999) and cattle chromosomes (Sun *et al.*, 1997; Schläpfer *et al.*, 1998; Sonstegard *et al.*, 1998; Larsen *et al.*, 1999). These maps demonstrate that although synteny is conserved, there have been numerous intrachromosomal rearrangements altering gene order.

A recent, comprehensive study in goats (Schibler *et al.*, 1998b) mapped 202 type I loci by FISH (160 genes, 23 microsatellites near coding sequence, 4 comparative anchor tagged sequences (CATS), and 15 expressed sequence tags (ESTs)) to locations consistent with previous ZOO-FISH results in cattle. Goat BAC clones were isolated for the 160 genes using 62 primers derived from human sequence, 63 from cattle, 17 from sheep, 11 from goat and 7 from other species. In total, 255 genes were mapped in goats and human map locations are available for 249 of these. From these data, twice as many rearrangements between humans and goats were identified compared with results from ZOO-FISH studies. In fact, it seems that intrachromosomal rearrangements are as frequent as interchromosomal rearrangements; they are just not as easily detected. These data

highlight the need for extensive comparative maps because complex chromosomal rearrangements can occur. If there are insufficient markers to define the extent and nature of these rearrangements, candidate gene identification will be hindered.

BAC libraries are now available for sheep, cattle and goat, and several hundred clones for genes and markers have already been isolated from these libraries. By utilising these clones for FISH in other closely related species, framework comparative maps can be rapidly developed. In this study, the localisation of ovine BAC clones onto cattle, goat and deer chromosomes by FISH is demonstrated.

6.1 Suppression with sheep, cattle and goat CotI DNA

Suppression of repetitive DNA in the BAC clones was required to detect discrete FISH signals on Bovidae chromosomes. Without suppression, the arms of most chromosomes were painted.

For cattle, the same amount of ovine CotI DNA was used as for FISH on sheep chromosomes (Table 5.2). However, in all cases, suppression was inadequate with numerous, large grains along the arms of the chromosomes. Even pre-annealing the BAC DNA for up to 4h with increased CotI (up to 25μg) failed to improve suppression.

To test whether the repeats in the species being mapped might also affect suppression, bovine and caprine CotI DNA was prepared from thymus tissue. Similar yields of CotI DNA were recovered as for sheep (~12μg/μl). The average length of cattle fragments was 800bp, but the average fragment length for goat was less than 200bp. When bovine CotI DNA was pre-annealed to the ovine BAC clones and hybridised to bovine chromosomes, adequate suppression was achieved with a similar amount of CotI DNA as that used in sheep with ovine CotI DNA (Table 6.1). However, when goat CotI DNA was used for mapping on goat chromosomes, the repeats were poorly suppressed. Better

suppression was achieved when sheep CotI DNA was used instead, but in many cases the repeats were still somewhat undersuppressed, with small grains lightly painting the chromosomes. Ovine BAC clones were also mapped to red deer and fallow deer chromosomes and suppression was better with deer CotI DNA than sheep CotI DNA (Kasprzak, 1998).

6.2 Physical assignment of ovine BAC clones to bovine chromosomes

Ovine BAC clones for 12 microsatellite markers and 15 genes were mapped by FISH to homologous locations on bovine chromosomes (Figure 6.1, Table 6.1). Of these loci, 4 were not previously mapped in cattle (CSRD270, CSRD287, GAPR and OarJMP58). IL2RA was the only marker previously assigned to a discrete band by FISH (Yoo *et al.*, 1995) and the new BAC assignment verifies the localisation to 13q1.4. However, this marker has not been integrated into a genetic linkage map of cattle and the BAC clone will allow the isolation of polymorphic markers for such analysis. Likewise, CD3D which maps to 15q2.1 has not been mapped by genetic linkage analysis. The FISH assignments for the other 23 ovine BAC clones anchor the genetic maps for 18 bovine chromosomes. The BAC clones for CD3D, IGF1 and PRL were assigned in cattle, but have not yet been assigned in sheep. BAC clones to anchor BTA3, 9, 17, 19, 20, 21, 22, 24, 25, 28, 29 and X still need to be assigned because suppression with bovine CotI DNA needs to be optimised for these BAC clones.

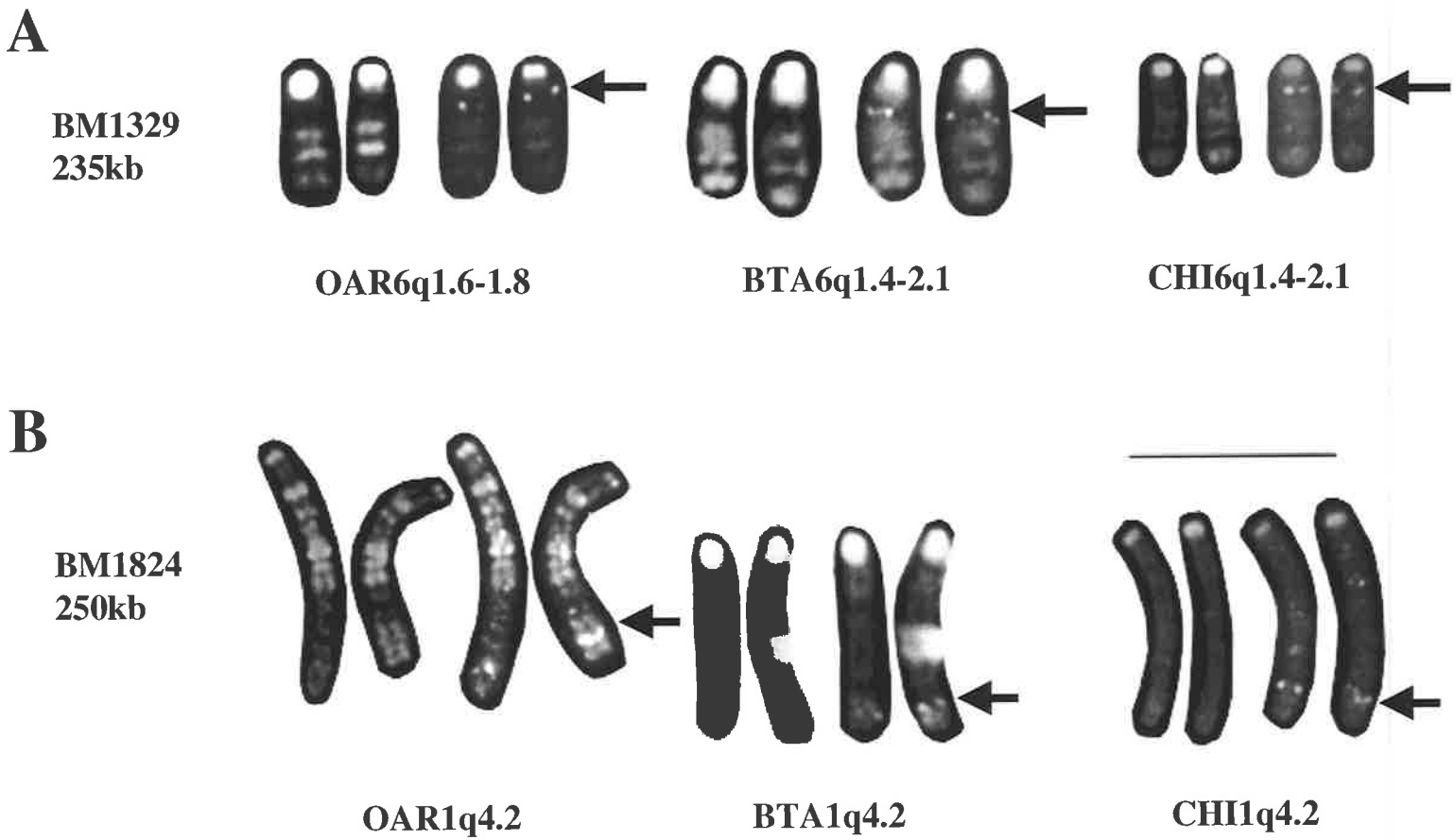


Figure 6.1: Comparative physical mapping by FISH in sheep (OAR), cattle (BTA), and goat (CHI) using ovine BAC clones. A) 5R3C4 for BM1329. B)284R5C4 for BM1924. Bands are shown on the left-hand pair of chromosomes and the arrow indicates signal on the right-hand pair. Bar represents 10μm.

Table 6.1: Physical map locations of ovine BAC clones localised to G-banded bovine metaphase chromosomes by fluorescence *in situ* hybridisation. Published mapping information in cattle for each locus and the method that was used to make the chromosomal assignment are listed (L: linkage, SCH: somatic cell hybrids, FISH: fluorescence *in situ* hybridisation). The amount of bovine CotI DNA required to suppress repeats within the BAC clones is indicated. The number of chromosomes that were scored to assign the bands is in parentheses.

Locus	[1]Published Map Location	Mapping Method	BAC	CotI	Physical Location in Cattle	
ADCYC	15	OAR15 BY L	443R7C6	8µg	15q1.4	(3)
BM719	16	L	296R6C9	25µg	16q2.3	(4)
BM827	11	L	153R6C4	8µg	11q1.2-1.3	(4)
BM1329	6	L	5R3C4	8µg	6q1.4-2.1	(6)
BM1824	1	L	284R5C4	25µg	1q4.4	(4)
BM6438	1	L	321R4C8	25µg	1q1.2-1.3	(3)
CD3D	15	SCH	158R2C6	12µg	15q2.1	(3)
CSF2	7	SCH	295R7C7	8µg	7q1.3-1.4	(5)
CSRD270	18	OAR14 BY L	306R8C4	8µg	18q1.3-1.4	(3)
CSRD287	12	OAR10 BY L	100R4C3	8µg	12q1.2	(4)
DRB1	23	L	569R6C11	8µg	23q2.2	(3)
DYB	23	L	78R3C6	8µg	23q2.4-qter	(4)
FN1	2	SCH, L	597R6C12	8µg	2q3.4-3.6	(3)
[3]GAPR	7	[2]OAR5q3.5 BY FISH	560R5C11	8µg	7q2.6-2.7	(6)
[3]GHRH	13	L, SCH	327R8C12	8µg	13q2.1	(5)
IGF1	5	SCH, L	402R7C8	8µg	5q3.2-3.4	(5)
IL2RA	13q14	FISH	43R5C3	25µg	13q1.4	(3)
IL6	4	L	290R2C10	25µg	4q1.4-1.6	(4)
ILSTS005	10	L	220R8C7	8µg	10q3.4-qter	(3)

Table 6.1 continued

Locus	[1]Published Map Location	Mapping Method	BAC	CotI	Physical Location in Cattle	
ILSTS008	14	L	337R8C6	8µg	14q1.6-1.8	(4)
ILSTS011	14	L	35R2C11	8µg	14q1.3-1.5	(4)
3LDLR	7	L	192R4C7	8µG	7q1.5	(6)
[3]LPL	8	L	323R6C9	8µg	8q2.3	(6)
MAF92	26	L	87R3C2	8µg	26q2.3	(3)
OarJMP58	27	OAR26 BY L	535R4C6	8µg	27q2.1-qter	(4)
PRL	23	L	102R1C11	8µg	23q2.1	(4)
RNASE6	10	L	79R1C6	8µg	79R1C6	(6)

[1]As listed in ArkDB sheep and cattle genome databases (www.ri.bbsrc.ac.uk)
[2]This study
[3]FISH performed by L. Allen (1997)

Table 6.2: Physical map locations of ovine BAC clones localised to G-banded caprine metaphase chromosomes by fluorescence *in situ* hybridisation. Published mapping information in goat for each locus and the method that was used to make the chromosomal assignment are listed (L: linkage, SCH: somatic cell hybrids, FISH: fluorescence *in situ* hybridisation). The amount of ovine CotI DNA required to suppress repeats within the BAC clones is indicated. The number of chromosomes that were scored to assign the bands is in parentheses.

Locus	[1]Published Map Location	Mapping Method	BAC	CotI	Physical Location in Goat	
BM719	16	L	296R6C9	25µg	16q2.2-2.4	(3)
BM1329	6q1.5	FISH	5R3C4	8µg	6q1.4-2.1	(6)
BM1824	1	BTA1 BY L	284R5C4	25µg	1q4.4	(4)
BM4208	9q2.6	FISH	408R8C6	8µg	9q2.5-2.6	(3)
BM6526	27q12.2	FISH	304R4C7	5µg	27q1.2.2-2.1	(3)
CD3D	15	BTA15 BY SCH	158R2C6	12µg	15q2.1	(3)

[1]As listed in GoatMap (http://locus.jouy.inra.fr/cgi-bin/lgbc/mapping/common/intro2.pl?BASE=goat)

Table 6.3: Ovine BAC clones that have been isolated for comparative mapping loci. Physical map locations were determined by FISH in sheep, cattle and goat. Mapped loci that were not assigned in this study are in parentheses. Comparative mapping information for these loci in humans, mice, rats and pigs is listed.

	Locus	Human	Mouse	Rat	Pig	Sheep	Cattle	Goat[1]
ADCYC	Adenylate cyclase 2	5p15.3-p15.2				15q1.4-2.2	15q1.4	
CD3D	Antigen CD3E, delta polypeptide	11q23	9				15q2.1	15q2.1
CSF2	Colony stimulating factor 2	5p23-31	11			5q1.3-1.5	7q1.3-1.4	
DRB1	MHC class II, DRB1	6			7	20q2.2	23q2.2	
DYB	MHC class II, DYB					20q3.2-3.3	23q2.4-qter	
FN1	Fibronectin 1	2q34	1C1-1C5		15q2.5	2q4.1	2q3.4-3.6	
GAPR	GTPase activating protein receptor					5q3.4	7q2.6-2.7	
GHRH	Growth hormone releasing hormone	20q11.2	2			13q2.4	13q2.1	
IFNG	Interferon gamma	12q24.1			5p1.1-q1.1	3q2.3		(5q2.2-2.3)
IGF1	Insulin-like growth factor 1	12q22-q23	10	7	5q2.5		5q3.2-3.4	(5q3.1)
IL2RA	Interleukin 2 receptor	10p15-14	2A2-A3			13q1.3	13q1.4	(13q1.5)
IL5	Interleukin 5	5p31.1	11A5-B1	10		5q1.5		
IL6	Interleukin 6	7p21-15	5	4	9p1.4-1.5	4q1.4-1.5	4q1.4-1.6	
KAP1.1	Keratin associated protein 1.1					11q3.1		
LDLR	Low density lipoprotein receptor	19p13.2	9			5q1.3	7q1.5	(7q1.5)
LPL	Lipoprotein lipase	8p22	8	16	14q1.2-1.4	2p2.4	8q2.3	
PRL	Prolactin	6p22.2	13	17	7p1.1-p1.2		23q2.1	
RNASE6	Seminal ribonuclease	14				7q1.5	10q1.5	
SRY	Sex determining region Y	Yp11.3			Yp1.2-1.3	Yp1.1-1.2		

[1]Physical assignments from GoatMap (http://locus.jouy.inra.fr/cgi-bin/lgbc/mapping/common/main.pl?BASE=cattle)

6.3 Physical assignment of ovine BAC clones to caprine and cervine chromosomes

To demonstrate that ovine BAC clones could also be mapped in goats, six ovine BAC clones for 5 microsatellites and 1 gene were assigned to homologous locations on caprine chromosomes by FISH (Figure 6.1, Table 6.2). Similarly, ovine BAC clones for 6 microsatellites and 6 genes have been hybridised to red and fallow deer (Kasprzak, 1998, data not shown).

6.4 Attempted localisation of ovine BAC clones to human and mouse chromosomes

Ovine BAC clones were hybridised to human and mouse chromosomes (provided by G. Webb) under various conditions. Experiments were performed 1) with and without suppression, 2) at various annealing temperatures (room temperature to 37°C), 3) with various amounts of probe, and 4) with and without amplification of the fluorescent signal using avidin and biotinylated anti-avidin. No signal was detected with any of the conditions tested.

6.5 Discussion

Suppression of repetitive elements within the ovine BAC DNA probe is critical for physical mapping to discrete chromosomal regions. Consequently, the BAC DNA was pre-annealed to sheep CotI DNA prior to hybridisation on metaphase chromosomes. Surprisingly, it was found that the chromosomes from closely related species affected the results. Better suppression was achieved when equal amounts (μg) of bovine CotI DNA was used on bovine chromosomes and cervine CotI DNA was used on deer chromosomes than when ovine CotI DNA was used. These results, however, cannot be taken as conclusivebecause the reciprocal experiment (bovine or cervine CotI DNA on ovine chromosomes) was not tested. The observations may also be confounded by the difference

in CotI DNA fragment length. Goat CotI DNA did not suppress well, but the fragment length was short (<200bp). The bovine and cervine CotI DNA fragments were longer (~800bp) than the ovine fragments (~600bp) and suppressed better. To determine whether fragment length is affecting the efficacy of suppression, it would be interesting to test whether bovine or cervine CotI DNA better suppress FISH experiments on goat chromosomes than ovine CotI DNA.

An alternative explanation for the difference in effectiveness of CotI DNA from different species, is that excess CotI DNA present in the hybridisation mixture after pre-annealing, suppresses repetitive regions on the chromosomes. This would, thereby, block non-specific hybridisation of the probe DNA to those chromosomal repeats sharing partial sequence identity with repetitive elements in the probe. In at least two ZOO-FISH painting experiments, a combination of CotI DNA from the species that the probe was derived from and the species that was being painted was used to suppress repeats (Goureau *et al.,* 1996; Schmitz *et al.,* 1998). This combinatorial might also improve suppression in single gene FISH.

Even without suppression, no hybridisation was detected when ovine BAC clones were mapped to mouse or human chromosomes, despite trying various conditions. One parameter that was not varied was hybridisation length. For ZOO-FISH with human chromosome-specific painting probes, the most important factor affecting signal intensity is the duration of hybridisation (Pinkel *et al.,* 1988; Scherthan *et al.*, 1994). Hybridisation of human paints on bovine chromosomes was allowed to proceed for 3 days at 37°C (Hayes, 1995). It will be interesting to determine whether a longer hybridisation period enables discrete fluorescent signals to be detected on human and mouse chromosomes when BAC clones are used as probes instead of whole chromosome paints. Conversely, if human or mouse BAC clones could be localised to ovine chromosomes there would be a huge pool of

comparative markers available for mapping since these libraries are extensively characterised. It does not seem likely, however, that physical mapping between such distantly related species would ever be routine.

In this study, 25 FISH assignments were made in cattle and 6 in goats using ovine BAC clones as probes, demonstrating the relative ease with which comparative mapping by FISH in closely related species can be undertaken using large insert clones. Both genes and microsatellites markers were mapped. Inclusion of some microsatellites on comparative maps is important to anchor the maps to existing linkage maps. However, to make full use of the gene-rich human and mouse maps, it is important that more type I loci are mapped in livestock. To date, 19 comparative mapping loci have been isolated from the ovine BAC library. Human map locations for 16 of these loci are available. Likewise, 11 have been mapped in mice, 5 in rats and 8 in pigs (Table 6.3). In this study, only CD3D was mapped in goats, however, 4 of this set of comparative mapping loci (IFNG, IGF1, IL2RA, LDLR) have previously been mapped in goats (Table 6.3).

Hundreds of BAC clones for genes and microsatellites have already been isolated from the cattle, goat and sheep BAC libraries (Schibler *et al.*, 1998b; S.K. Davis, pers. comm.). These clones represent a largely untapped resource for comparative mapping. As demonstrated in this study, it is possible to map BAC clones from one species in other closely related species. Consequently, by mapping all these clones in sheep, cattle and goats, framework comparative maps could be rapidly developed.

Extensive comparative maps will be essential in the search for major genes and quantitative trait loci that encode economically important traits. First, the boundaries of candidate regions need to be well defined so that positional candidates can be selected from the human and mouse maps. Then, fine-mapping of the critical region must be undertaken to identify the genotype causing the phenotype of interest. Stable, large-insert clones are

proving invaluable for fine-mapping by enabling deep contigs of the critical region to be developed for large-scale sequencing. The positional candidate cloning approach and the use of large-insert clones for contig development was successfully applied to identify the major genes causing spider lamb syndrome in sheep (Cockett., unpublished) and double muscling in cattle (Grobet *et al.*, 1997; Sonstegard *et al.*, 1998; Pirottin *et al.*, 1999). Contigs of large-insert clones are also being employed in the search for *POLL* in cattle (Stillwell *et al.*, 1999; Davis *et al.*, 1999), *PIS* in goats (Schibler *et al.*, 1998a; Vaiman *et al.*, 1999a) and *FecB* in sheep (Lumsden *et al.*, 1999). The availability of ovine BAC clones for fine-scale, comparative mapping in sheep will aid the search for genes that are responsible for economically important traits in this valuable livestock species.

CHAPTER 7

INVESTIGATION OF GENOME ORGANISATION USING BAC CLONES

INVESTIGATION OF GENOME ORGANISATION USING BAC CLONES

Bacterial artificial chromosomes are an ideal tool for the study of genome organisation. The clones are stable, non-chimaeric and of sufficient length (>100kb) to enable chromosome walking and long-range restriction mapping to be undertaken with relative ease. The keratin associated protein (KAP) genes were selected for this investigation because before this study there was a paucity of mapping information available for these genes in sheep. Keratin intermediate filament (KIF) genes and KAP genes have been associated with variation in wool fibre diameter (Parsons *et al.,* 1994) and staple strength (Rogers, 1994), and would be valuable genes to incorporate in advanced breeding programmes using marker assisted selection. Elucidating the organisation of these genes in the sheep genome would assist in understanding their regulation. Many of the type I and type II KIF genes are tightly clustered in two separate domains in humans (Romano *et al.,* 1988; Cerrato *et al.,* 1997) and mice (Nadeau *et al.,* 1989; Compton *et al.,* 1991), but the organisation of the KAP genes has not been described. There is some evidence of clustering of the KAP genes in sheep (Powell *et al.*, 1983). It was of interest to know whether members of different KAP families were grouped together as this may have regulatory implications.

7.1 Organisation of the Wool Follicle

Hair is composed of about 50-100 different keratin proteins derived from several multigene families that are expressed in tissue-specific and differentiation-specific patterns (Powell *et al.*, 1991). The mature hair follicle is a derivative of the epidermis (Figure 7.1). All the cell types of the hair shaft (ie. inner root sheath, cortex, cuticle and medulla) except for the outer root sheath (ORS) arise from germinative cells in the follicle bulb as distinctive differentiation programmes are activated. The ORS is continuous with the epidermis, and epidermal rather than hair-type keratins are expressed. In the inner root sheath (IRS),

trichohyalin granules are expressed first, in association with 8-10nm filaments. About 10-20% of cells from the follicle bulb progress to the cortex. Keratins are synthesised as the cells move upwards with expression maximal about a third of the way up the length of the follicle. During keratinocyte differentiation in the wool follicle, the acidic type I (KRT1.n) and basic type II (KRT2.n) low-sulphur KIF proteins are assembled into filaments which become embedded in a matrix of high-sulphur KAPs and high-glycine/tyrosine KAPs. The KIFs are the first detectable hair-specific keratins and the KAP families are activated at later stages of differentiation (Powell *et al.,* 1991).

The cortex contains two cells types (orthocortex and paracortex), with differing organisation, which are filled with KIFs and separated by intercellular material that appears in the terminal stages of differentiation. There are at least two KAP families that are expressed separately in the cortex (KAP4.n) and the cuticle (KAP5.n) (MacKinnon *et al.,* 1990; Powell *et al.,* 1991). Expression of the KAP4 family in the paracortex is induced by infusion of cysteine (Fratini *et al.,* 1994).

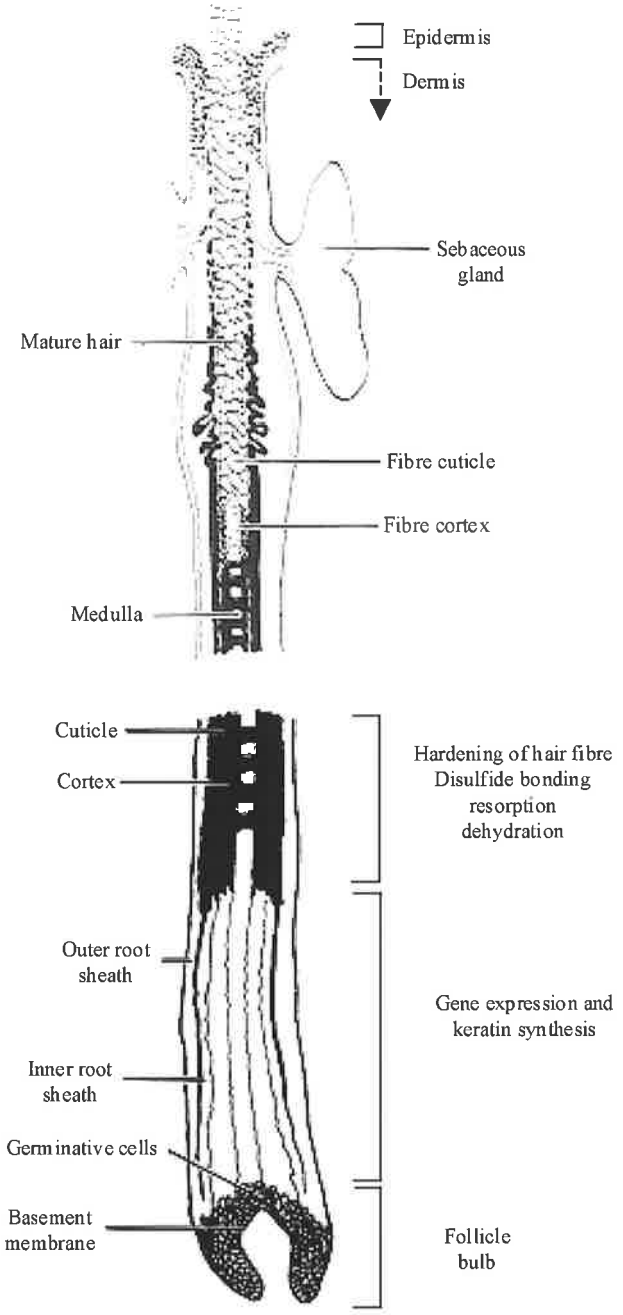


Figure 7.1: Organisation of the hair follicle. The regions where the main events of cell proliferation in the follicle bulb take place and the subsequent formation of the keratinized hair shaft are depicted. (From Powell and Rogers, 1994).

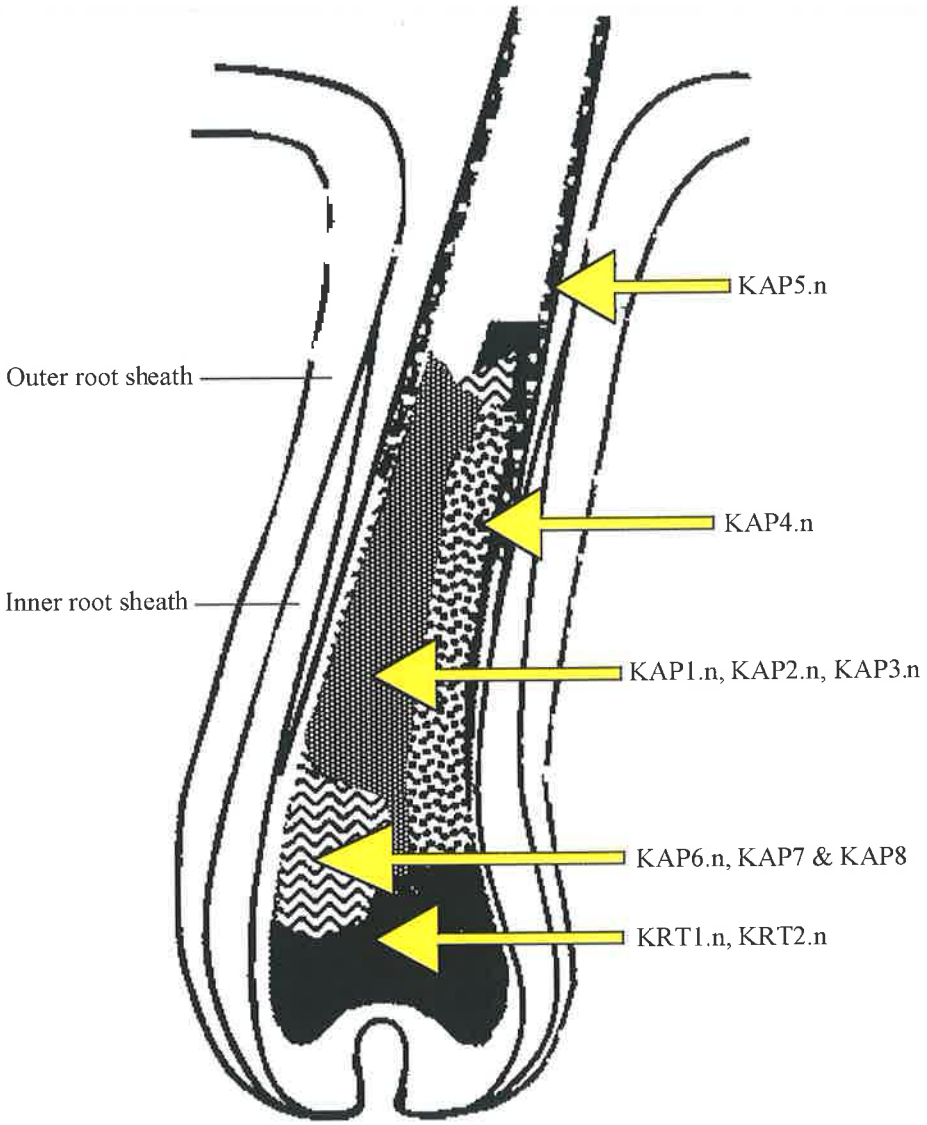


Figure 7.2: Expression of keratin (KRT) and keratin associated protein (KAP) genes in wool follicle differentiation. Schematic of overlapping gene expression was assembled by comparing RNA *in situ* hybridisation analyses of many follicle sections. (From Powell and Rogers, 1994).

7.2 Keratin associated protein genes

The KAP genes encode a heterogenous group of proteins that form the matrix in which the microfibrils of the wool fibre are embedded. The KAP gene families (Table 7.1) have been classified on the basis of their protein sequences as high glycine-tyrosine KAP genes (KAP 6.n, KAP7, KAP8), high sulphur KAP genes (KAP1.n, KAP2.n, KAP3.n) and ultra-high sulphur KAP genes (KAP4.n, KAP5.n) (Powell *et al.*, 1994). Recently, a number of KAP genes have been mapped by genetic linkage analysis to several different ovine chromosomes. KAP6.1, KAP7, KAP8 and trichohyalin were mapped to ovine chromosome 1 and are closely linked to one another (Parsons and Cooper, 1995; McLaren *et al.,* 1997). KAP1.1 and KAP1.4 are separated by 1.9kb (Powell *et al.*, 1983) and have been mapped to chromosome 11 with KAP1.3, KAP3.2 and KRT1.2 (McLaren *et al.,* 1997). All five of these genes are closely linked to one another on OAR11. KAP5.1 was mapped to ovine chromosome 21 (McLaren *et al.,* 1997). Genetic linkage analysis demonstrated that KAP protein genes from related families (eg. high sulphur KAP genes) are clustered. The high glycine-tyrosine KAP genes are the first KAPs to be expressed in the wool follicle, followed by the high sulphur KAP genes, and then the ultra-high sulphur KAPs (Figure 7.2). Consequently, it is possible that there is a relationship between the clustering of KAP genes and their expression in the wool follicle. Plasmid-based probes for each of the keratin associate protein gene families are available (Fratini *et al.*, 1994) and were used in this study to establish whether multiple different KAPs were present within a single BAC clone known to contain KAP1.1.

Table 7.1: Keratin associated protein gene families

Gene or Gene Family	Alternative name
KAP1.n	High-sulphur keratin B2 family
KAP2.n	High-sulphur keratin BIIIA family
KAP3.n	High-sulphur keratin BIIIB family
KAP4.n	Ultra high-sulphur keratin cortex family
KAP5.n	Ultra high-sulphur keratin cuticle family
KAP6.n	Type II high glycine/tyrosine keratin family
KAP7	High glycine/tyrosine type I component C2
KAP8	High glycine/tyrosine type I component F
KAP9.n	Ultra high-sulphur keratin cortex family
KAP10.n	
KAP11.1	Hacl-1 (mouse)
KAP12.n	

7.3 Results

7.3.1 Isolation and Physical Mapping of a BAC Clone containing KAP1.1

Primers specific for KAP1.1 (Table 2.2) that amplified a 329bp product were designed from the sequence for a genomic clone (Powell *et al.*, 1983; GENBANK #X01610) and were used to screen the ovine BAC library. A single BAC clone (164R5C4) was identified by PCR-based screening. The PCR product from the BAC clone was sequenced and shared 100% sequence identity with the published genomic sequence for KAP1.1 (Powell *et al.*, 1983). The BAC clone was mapped by FISH to ovine chromosome 11q3.1.

7.3.2 BAC End Rescue

To isolate overlapping clones, sequence from each end of the KAP-containing BAC clone was recovered by BAC end rescue (Figure 2.3). Inverse-PCR products containing the left-end of the BAC clone were obtained from *EcoRI* and *BamHI* self-ligations. A ~300bp *EcoRI* product was sequenced and a pair of primers were designed to re-screen the BAC library:

SWKL2F: 5' GAT GCA ACA CAG TTC 3'
SWKLR: 5' GTG ATA CTG GTG CAT GCG G 3'

Optimum conditions for amplification over 35 cycles were: 2.5pmol each primer, 2.5mM Mg^{2+} at T_A=50°C for 1'. A single ~50kb BAC clone (578R2C8) overlapping the left-end was recovered from the library. This BAC clone did not contain KAP1.1 (ie. it failed to amplify by PCR with KAP1.1 primers). *HindIII* digestion of the two BAC clones revealed three coincident bands (~8kb, 5kb and 3kb) indicating the maximum overlap was ~16kb.

An ~20kb *SacI* sub-clone was the smallest transformant recovered containing the right-end of the BAC clone. No sequence of sufficient quality for primer design was recovered from this sub-clone by manual sequencing. Ultimately, automated sequence for the right-end of the BAC clone was obtained directly from the BAC itself using 1µg template, 16µl BigDye mix and 50 thermal cycles (as recommended for BAC end sequencing by PE Biosystems). Primers from this sequence can now be designed to isolate a clone overlapping the right-end of the BAC clone.

7.3.3 Long-range restriction mapping of a KAP-containing BAC clone

To investigate whether additional KAP genes were clustered within the same BAC clone, long-range restriction mapping was performed using endonucleases that recognise rare restriction sites (Ceratto *et al.,* 1997). BAC DNA was digested in single and double restriction reactions with *NotI*, *XhoI*, *ClaI*, *SfiI* and *NruI* and the resulting fragments were separated by CHEF gel electrophoresis (Figure 7.3a). *NotI* and *NruI* cleave the vector twice, *Sfi* and *XhoI* cut once, and there are no *ClaI* recognition sites in pBeloBAC11 (Appendix 2). Only two *NotI* fragments were generated. Thus, there were no internal *NotI* sites in the BAC insert and it was estimated to be ~94kb from the CHEF gel. Digestion with *SfiI* yielded a single band, so there were no *SfiI* sites in the insert.

The BAC DNA was transferred to a nitrocellulose membrane and hybridised sequentially with the BAC clone itself, pBeloBAC11, KAP-gene family probes, BovA and $(GT)_{30}$ (Figure 7.3b-h). Hybridisation with the vector was useful for orientation and positioning of some of the restriction sites. *ClaI* digestion generated two fragments (~55kb and ~48kb) and the larger fragment was cleaved into two equal pieces by *SfiI* (visualised as a single band), one of which hybridised to the vector. This enabled the two *ClaI* sites to be placed unequivocally on the restriction map (Figure 7.4).

Digestion with *NruI* yielded 3 fragments (~89kb, 8.5kb and 5.4kb). The 5.4kb fragment hybridised strongly to pBeloBAC11, as expected. The 8.5kb fragment also hybridised, so this fragment must contain the larger of the two remaining portions of vector (ie. 1249bp) plus ~6.5kb of insert, thereby enabling the sole *NruI* site in the insert to be positioned.

Positioning the *XhoI* fragments on the restriction map was problematic. The fragment sizes for double digests with *XhoI/SfiI* (~26-23-19-7kb) and *XhoI/NruI* (~30-27-20-10-8-5.5kb) could not be explained. When these digests were repeated, different length fragments

were generated (~45-26-23-7kb for *XhoI/SfiI* and ~41-23-19-7-8.5-5kb for *XhoI/NruI*), indicating that at least one of the enzymes in the original digests was incorrect (data not shown). The *XhoI* sites either side of the vector were positioned using the *NotI/XhoI* and *XhoI/SfiI* digests. The order of the remaining two *XhoI* fragments, and therefore, the position of the last *XhoI* site could only be determined by examining the hybridisation patterns for the KAP family probes. The most consistent hybridisation pattern for KAP1.1 was obtained when the 6kb *XhoI* fragment was adjacent to a 27.5kb *XhoI* fragment that included ~2kb of vector. The location of a *XhoI* and *ClaI* site within ~1kb of each other agreed with previous sequence data for KAP1.1 (GENBANK #X01610).

Distinct hybridisation patterns were obtained for the KAP1.n, KAP2.n and KAP3.n probes. Several KAP probes hybridised weakly to fragments that were inconsistent with the most likely location of the particular gene, suggesting there was some cross-hybridisation between the gene families. KAP2.n was localised to a 6.5kb *NruI* fragment adjacent to the vector (Figure 7.4). Both the KAP1.n and KAP3.n probes hybridised to broad regions (>20kb), separated by ~25kb, indicating there may be several members of these gene families located together. A BovA repeat was localised to the same fragment as KAP2.n and a $(GT)_n$ repeat was localised to the *XhoI/NruI* fragment that was also positive for KAP1.n.

As expected, the KAP5.n probe did not hybridise to the ovine BAC clone. However, KAP4.n, KAP10.n and KAP12.n genomic probes all hybridised to the same fragment as KAP2.n and BovA. Given the size of these genes, it was not possible for them all to be co-located in a single 6.5kb fragment. To determine whether the probes were cross-hybridising to KAP2.n or BovA, the BAC DNA was digested with *Sau3AI*, *HaeIII* and *AluI* and then hybridised with each of the probes to generate fingerprints. The hybridisation patterns for KAP2.n and BovA were distinct. The hybridisation patterns for KAP4.n and KAP12.n

overlapped BovA. KAP10.n was not hybridised, but the forward sequence from this genomic clone formed a contig with KAP12.n, so it is likely that this clone also contains BovA (see below). There were additional positive bands for BovA, indicating there may be other short stretches of BovA within the BAC clone that did not previously hybridise. Indeed, 10 separate sequences containing BovA subclasses, that did not form contigs together were obtained by shot-gun sequencing (section 7.3.4).

7.3.4 Shot-gun Sequencing

To further characterise the KAP-containing BAC clone, shot-gun sequencing was performed. Four libraries of sub-clones for shot-gun sequencing were generated from *Sau3AI* (^GATC), *Tsp509I* (^AATT), *HindIII* (A^AGCTT) and *PstI* (CTGCA^G) digested fragments of BAC DNA cloned into complementary restriction sites in pBluescript (ie. *BamHI, EcoRI, HindIII* and *PstI*, respectively). A fifth library of *MspI* (C^CGG) fragments was cloned into *AccI* digested pMCS. A combination of four-base and six-base cutters was chosen to aid contig development with the larger *HindIII* and *PstI* fragments forming a framework upon which smaller fragments could be assembled. Different recognition sequences (GC-rich vs AT-rich) were chosen to minimise cloning biases due to base composition. The *Sau3AI* library was the most efficient (~1.7 x 10^{10} transformants/μg), followed by the *Tsp509I* and *MspI* libraries (5.71 x 10^{9} transformants/μg), and the *HindIII* and *PstI* libraries (~4.8 x 10^{9} transformants/μg).

Figure 7.3: Southern analysis of a KAP-containing ovine BAC clone. A) CHEF gel electrophoretogram of BAC DNA digested with combinations of *NotI* (N), *XhoI* (X), *ClaI* (C), *SfiI* (S), *NruI* (U). Digested DNA was separated on a 1% agarose gel in 0.5xTBE at 14°C for 16h at 6V/cm with an initial switch time of 1.4s and a final switch time of 13.5s. The DNA was sized against a lambda concatomer (*Biorad*) and a HindIII lambda ladder. B-H) Schematic representations of autoradiographs following hybridisation with (B) OBAC 164R5C4, (C) pBeloBAC11, (D) a KAP1.1 genomic clone, (E) a KAP2.12 genomic clone, (F) a KAP 3.3 genomic clone (G) BovA genomic clone, and (H) a $(GT)_{30}$ probe.

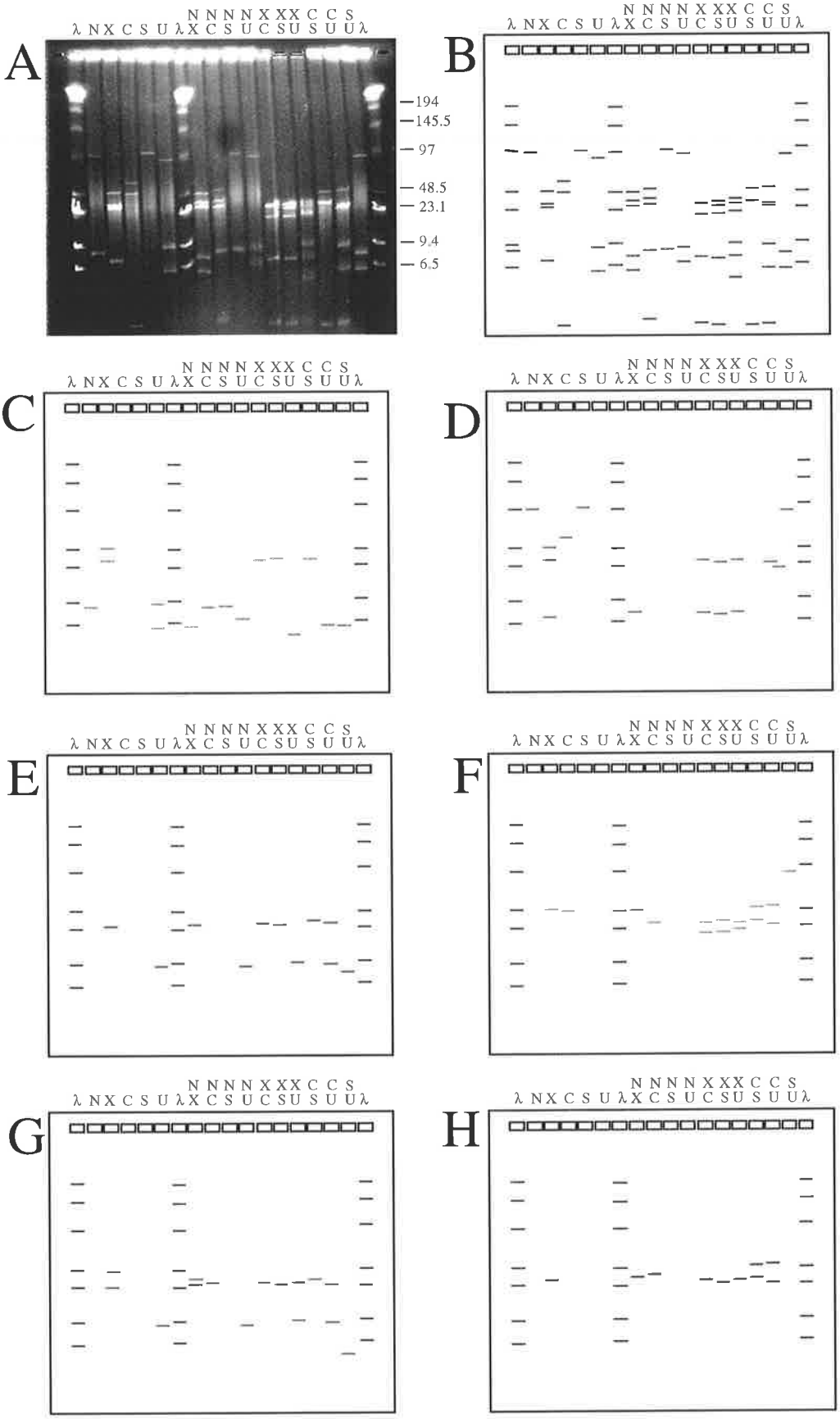
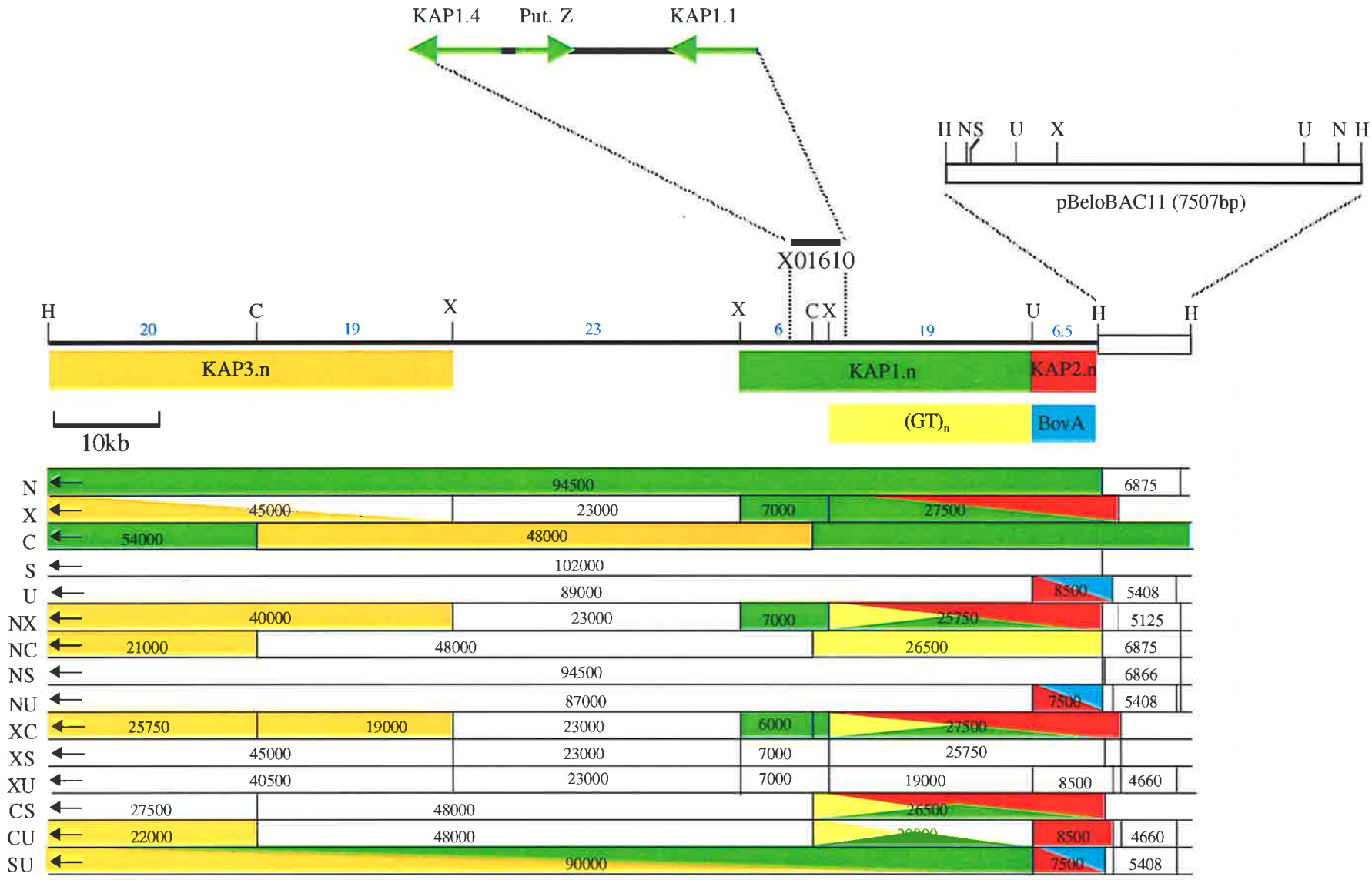

A
B
C
D
E
F
G
H
N N N N X X X C C S
λ N X C S U λ X C S U C S U S U U λ
194
145.5
97
48.5
23.1
9.4
6.5

Figure 7.4: Long-range restriction map of an ovine BAC clone containing KAP1.1. The *HindIII* (H) BAC clone was digested with *NotI* (I), *XhoI* (X), *ClaI* (C), *SfiI* (S) and *NruI* (U) and hybridised to KAP gene family probes. Regions containing at least one KAP3.n gene (orange), KAP1.n gene (green) and KAP2.n gene (red) were identified. The BAC also hybridised to BovA (blue) and $(GT)_n$ (yellow). Restriction fragment sizes are shown and hybridisation patterns are colour coded. Hatching indicates that only weak hybridisation was detected. KAP1.1, a putative Z gene and KAP1.4 were localised precisely and were oriented (arrows) by published sequence data (GENBANK #X01610).


KAP1.4
Put. Z
KAP1.1
H N S U X
U N H
pBeloBAC11 (7507bp)
X01610
H 20 C 19 X 23 X 6 C X 19 U 6.5 H H
KAP3.n
KAP1.n
KAP2.n
(GT)n
BovA
10kb
N 94500 6875
X 45000 23000 7000 27500
C 54000 48000
S 102000
U 89000 8500 5408
NX 40000 23000 7000 25750 5125
NC 21000 48000 26500 6875
NS 94500 6866
NU 87000 7500 5408
XC 25750 19000 23000 6000 27500
XS 45000 23000 7000 25750
XU 40500 23000 7000 19000 8500 4660
CS 27500 48000 26500
CU 22000 48000 8500 4660
SU 90000 7500 5408

It was estimated that 400 clones with an average insert size of 750bp were required for 95% sequence coverage of the BAC clone (N = ln (1-0.95)/ln (1-750/100,000) = 397, section 1.3.3.5). To obtain the 400 templates for sequencing, first the *Sau3AI*, *Tsp509I* and *MspI* libraries were amplified with universal primers (M13F and M13R) and products >500bp were purified (96 products for *Sau3AI*, 48 for *Tsp509I*, 48 for *MspI*). Secondly, the *HindIII* and *PstI* clones were sized (500bp – 20kb) and 96 clones from each library were sequenced directly. In addition, each of the KAP gene probes used for Southern analysis was sequenced.

The templates were sequenced with M13F and then non-redundant *HindIII* and *PstI* clones were sequenced with M13R. Overall 80% of the sequences were successful, but only 6/48 of the *MspI* fragments produced sequence. About 10% of all the sequences were either *E.coli* or the BAC vector. A total of 82,314 insert bases of good quality were obtained from 244 sequences, with an average of 337 insert bases/sequence. There were 52 orphan sequences and 46 contigs representing 39,025 non-redundant bases. A search for internal restriction sites revealed three subclones that were chimaeric or the result of partial digestion. Two of these subclones were from the *Tsp509I* library and the other was a *Sau3AI* subclone. These sequences were split at the internal recognition sites, but all remained orphans. The rest of the subclones were surprisingly rich in *TspI* sites. A total of 282 internal *Tsp509I* sites were identified, yielding an average of one *TspI* site every 138bp of sequence.

The forward sequences from the KAP4 and KAP5 gene family probes used for the Southern analysis formed a contig with a minimum of 95% homology. Similarly, the KAP10 and KAP12 probes formed a contig, but none of these four KAP probe sequences formed contigs with any sub-clone sequences.

7.3.5 Fine-scale comparative mapping

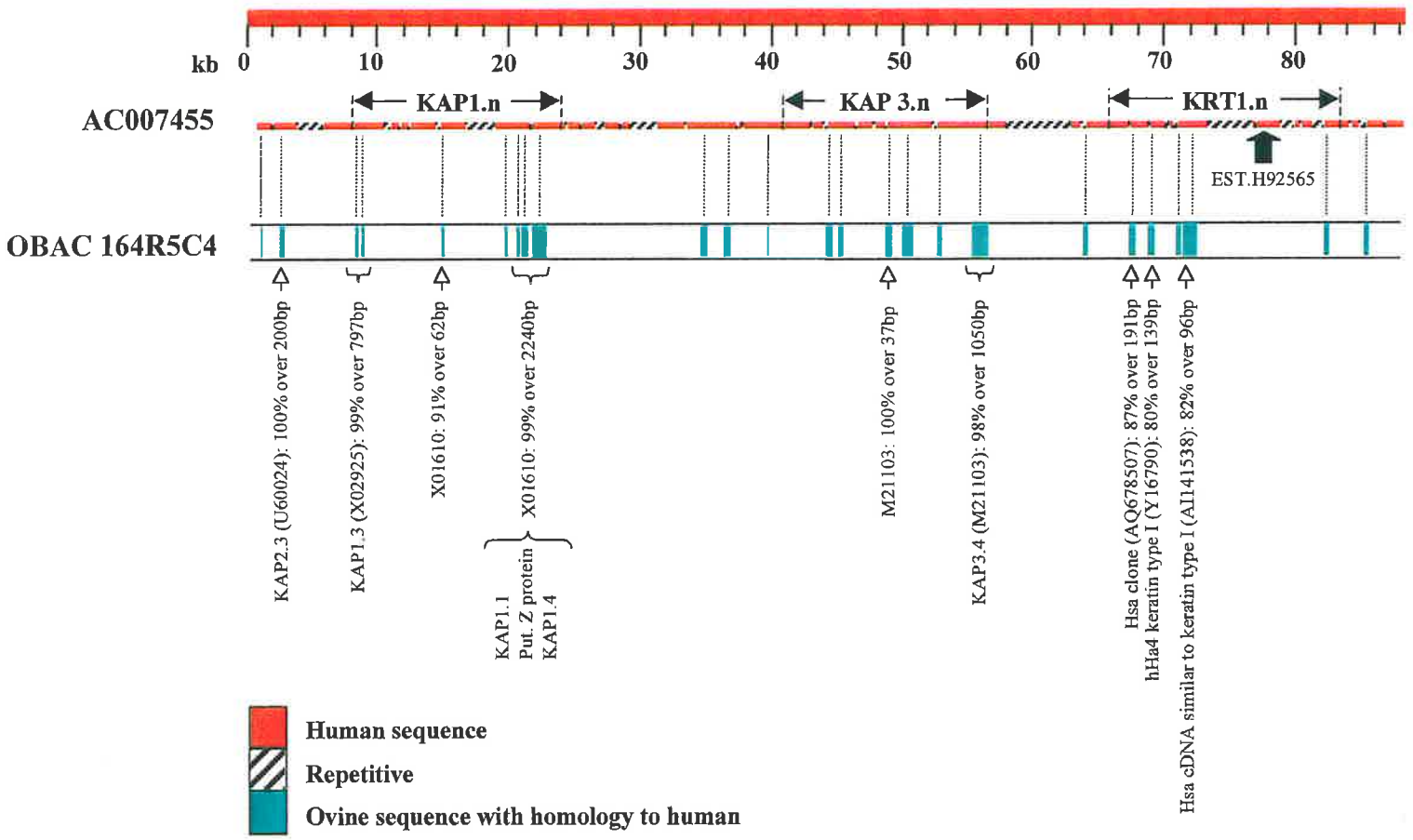
Every orphan sequence and consensus sequences for the contigs were imported into Repeat Masker (http://ftp.genome.washington.edu/cgi-bin/RepeatMasker). A total of 17,607 repetitive bases (45% of the sequence) were classified into 20 repeat sub-classes (Table 7.2). Long interspersed nuclear elements (LINE) represented 37.4% of the repeats, short interspersed nuclear elements (SINE) accounted for 6.7% of repeats and the remaining 1.2% was due to long terminal repeats (LTR). The most common repetitive element in the KAP BAC clone sequence was BovB (22%). In addition, a $(TTA)_6$ microsatellite was identified, but no $(GT)_n$ repeats were sequenced.

After masking the repeats, the resulting single-copy sequence was aligned to non-redundant (nr), high-throughput genome scan (htgs), general survey sequence (gss) and expressed sequence tag (dbest) databases of GENBANK by BLASTN. Some of the sequences aligned to different ovine KAP genes. The GENBANK sequences for these genes were imported into the data set to assist contig assembly. Contigs containing sequences similar to KAP1.1 and KAP1.4 (GENBANK #X01610), KAP1.3 (#X02925), KAP2.3 (#U60024), and KAP3.4 (#M21103) were recovered.

Table 7.2: Types of repetitive DNA elements in 39,025bp of sequence from a KAP containing BAC clone. Repeat sub-classes were identified using Repeat masker (http://ftp.genome.washington.edu/cgi-bin/RepeatMasker) and are ordered according to their contribution to the total sequence.

Repeat	Class	Sub-class	bp	# of seq.	% of seq.
LINE	BovB	BovB	8705	25	22.31
	L1	L1_BT	1815	4	4.65
		L1	796	3	2.04
		L1M4	767	3	1.97
		L1_Art	556	4	1.42
		L1Mec	404	1	1.04
		L1MC2	384	1	0.98
		L1M3	287	1	0.74
		L1M4	228	2	0.58
		L1MD2	224	1	0.57
		L1MB4	79	1	0.2
	L2	L2	348	3	0.89
LTR	Retroviral	MLT2D LTR	265	1	0.68
		MLT2CB	204	1	0.52
SINE	Artiodactyl	D repeat	147	1	0.38
	BovA	Bov-tA2	699	4	1.79
		Bov-tA3	491	1	1.26
		Bov-tA1	378	2	0.97
		BovA2	366	3	0.94
	MIR	MIR	543	4	1.39

Figure 7.5: Comparative sequence analysis of a BAC clone known to contain KAP1.1 reveals other genes within the ~100kb insert. Repetitive elements (hatched) in human sequence AC007455 were masked using Repeat Masker and then the remaining unique sequence (red) was compared to other sequences in the Genbank database by BLASTN. Three regions with homology to keratin associated protein (KAP) or keratin (KRT) genes are shown. Sequences from shot-gun subclones of OBAC 164R5C4 were aligned to AC007455 by BLAST2 (Green). Genbank accession numbers are given for alignments with significant homology to KAP or KRT genes.


kb
0
10
20
30
40
50
60
70
80
AC007455
KAP1.n
KAP 3.n
KRT1.n
EST.H92565
OBAC 164R5C4
KAP2.3 (U60024): 100% over 200bp
KAP1.3 (X02925): 99% over 797bp
X01610: 91% over 62bp
KAP1.1
Put. Z protein
KAP1.4
X01610: 99% over 2240bp
M21103: 100% over 37bp
KAP3.4 (M21103): 98% over 1050bp
Hsa clone (AQ678507): 87% over 191bp
hHa4 keratin type I (Y16790): 80% over 139bp
Hsa cDNA similar to keratin type I (AI141538): 82% over 96bp
Human sequence
Repetitive
Ovine sequence with homology to human

There were 25 cases (7 orphans and 18 contigs) covering 9965 bases where the sequence aligned (E value <0.001) to a human chromosome 17 BAC clone (GENBANK # AC007455). The human clone contains an EST (#H92565), and is mapped on HSA17 at $2040cR_{10000}$ and in the reference interval from D17S933-D17S800 at 58.3-62.9cM. After masking the repetitive DNA in the human sequence, the unique sequence was aligned by BLASTN to the nr database (Appendix III). Three regions which shared sequence identity with KAP1.n (~10-24kb), KAP3.n (~41-64kb) and KRT1.n (~66-85kb) genes were recovered (Figure 7.5). Within each of these regions, there were several alignments with similar likelihood, suggesting that there may be multiple genes from these families clustered together. BLAST2 with the GENBANK entry for KAP1.1 (X01610) revealed regions with 74%, 70%, 75%, and 70% sequence identity from 8757-9310bp, 15294-16017bp, 20154-21066bp and 21851-24234bp of AC007455, respectively. KAP3.4 (M21103) aligned to AC007455 from 41031-41396 at 73%, 45468-45706 at 70%, 50150-50716 at 74%, and 55809-56859bp at 77%. KRT1.2 (M23912) aligned to AC007455 from 69044-69419 at 65%, 87633-88029 at 65%, and 71162-71400 at 79%.

A human BAC clone (GENBANK #AC004231) that overlapped the right-hand end of AC007455 was also identified. By BLASTN, it was shown to contain sequence similar to keratin 13 (25-40kb away from AC007455), cytokeratin 20 (70-85kb from AC007455) and keratin 12 (90-100kb from AC007455).

By anchoring sequences from the sub-clones to the human BAC sequence, the most likely location of KAP genes within the ovine BAC could be determined (Figure 7.5), validating the arrangement of the KAP genes in the restriction map. There was strong support for the localisation of KAP1.3, KAP1.1, KAP1.4, and KAP3.4 within the single ovine BAC clone (>500bp sequence at >98% sequence identity). One contig had 200bp of

sequence which shared 100% sequence identity with KAP2.3. There was a short alignment at ~15kb with high homology to KAP1.1 (91% over 62bp), but this sequence did not form a contig to any other subclone sequences. Likewise, a very short segment of KAP3.4 (100% over 34bp) aligned at ~49kb, but did not form a contig with any other subclone sequences. These alignments may be at conserved motifs and suggest that related KAP1.n and KAP3.n genes are at these locations.

Two contigs, anchored to the human sequence at 69kb and 72kb, had homology to human type I keratins by BLASTN, but no similarity to ovine keratins could be demonstrated by this search. By BLAST2, however, the sequence anchored at 69kb and another sequence at 71kb shared 78% (over 135bp) and 82% (over 203bp) sequence identity with ovine KRT1.2 (GENBANK #M23912).

7.4 Discussion

Development of a restriction map for the KAP-containing ovine BAC clone was hampered by several factors. First, hybridisation with the BAC clone itself revealed that the DNA was not transferred uniformly to the nitrocellulose membrane, with the middle lanes being most affected and giving the weakest hybridisation signals. In subsequent Southern experiments with BAC clones, the DNA was allowed to transfer for 3 days, rather than overnight, and resulted in more intense and uniform hybridisation (data not shown).

Second, banding patterns for two of the double digests could not be explained. The optimum temperature for digestion with *SfiI* is 50°C, but all digests were performed at 37°C. It was initially thought that perhaps under these sub-optimal conditions *SfiI* had been able to cleave the relatively simple DNA of the vector, but had not been able to cleave a recognition site within the intact BAC insert. Then, after digestion with *XhoI*, *SfiI* had been able cut at a recognition site in one these smaller fragments because of the more relaxed

arrangement of the DNA. However, it was subsequently realised that the banding patterns for neither the *XhoI/SfiI* digestion nor the *XhoI/NruI* digestion could be explained, so at least one of the enzymes in these mixtures must have been incorrect.

Third, several of the enzyme combinations generated multiple fragments of similar size so it was difficult to localise the KAP genes to particular fragments. In addition, neither *SfiI* nor *NotI* cleaved the BAC insert, so these enzymes added no information to the restriction map of the insert. In future studies, test digests with a panel of rare-cutting restriction enzymes should be performed and only those enzymes that can be shown to cut the insert should be used in map development.

The restriction map of BAC clone 164R5C4 revealed there was at least one KAP2.n gene in the *NruI* fragment adjacent to the left end of the BAC clone and at least one KAP3.n gene within the *XhoI* fragment adjacent to the right end of the BAC clone. The KAP1.n genes spanned a *XhoI* fragment and a *XhoI/NruI* fragment adjacent to the KAP2.n fragment. Without sequence data, it was impossible to speculate as to how many KAP genes were present within these fragments or indicate their precise location. Fortuitously, KAP1.1, a putative Z protein and KAP1.4 could be positioned and oriented exactly, because the GENBANK entry for these genes (X01610) contained both a *XhoI* and *ClaI* site.

Developing a single contig of sequence for a 100kb BAC is a laborious and expensive process, but would be necessary to fully characterise the gene content and arrangement of regulatory elements within the clone unless comparative sequence data were available. A preliminary round of shot-gun sequencing was performed with a view to using sequence data from the human genome project to develop a comparative alignment of this region. The complete sequence of a human chromosome 17 BAC clone (AC007455) covering most of the ovine KAP BAC clone was released in May 1999. No genes were annotated in the

sequence, but there was an EST (H92565) at ~78kb. The EST was surrounded by repetitive DNA and no corresponding sheep sequence for this region was obtained from the BAC clone. This was one of the few stretches of unique DNA within the human BAC clone for which at least one shot-gun sequence from sheep was not obtained.

The organisation of the human BAC clone was investigated using Repeat Masker to categorise the repetitive DNA, and BLASTN to identify regions with similarity to known genes. This search revealed three regions with similarity to KAP1.n, KAP3.n and KRT1.n genes. Anchoring shot-gun sequences from the ovine BAC clone to the human sequence validated the location of KAP and KIF genes within the ovine clone predicted by BLASTN and by restriction mapping. KAP1.1 and KAP1.4 have previously been mapped to OAR11 (McLaren et al., 1997). This study demonstrated that genes highly similar to KAP1.3, KAP2.3 and KAP3.4 also map within this ~100kb ovine BAC clone. In addition, there is at least one as yet unidentified type 1 KIF gene within the clone. Conclusive identification of the KAP and KIF genes is difficult because of the high degree of similarity between family members. For example, there is 96% similarity between the GENBANK entries for KAP1.3 and KAP1.1 and KAP1.4 over 378bp. Also, the complex and inconsistent nomenclature that has been used to describe the KAP and KIF genes made GENBANK searches difficult. Indeed, it became apparent that some of the ovine KAP and KIF genes that have been described in the literature have not been submitted to GENBANK. It would be advantageous, now that a standard nomenclature for the KAP and KIF genes has been adopted (Powell and Rogers, 1994), to update the GENBANK entries to include the new names.

Fingerprinting the BAC clone demonstrated that the probes for KAP4.n and KAP12.n cross-reacted with BovA. Consequently, the hybridisation of these KAP genes to the BAC clone was a false positive result. No members of the KAP4.n, KAP10.n or

KAP12.n families have yet been mapped in sheep. Primers for these gene families could be designed from the genomic clones (KAP4.3, KAP10.1, KAP12.1) to isolate BAC clones for mapping studies.

By sequencing libraries of shot-gun subclones from the ovine BAC clone, 39,025 non-redundant bases were generated. There was 2.1x redundancy in the sequence and only ~41% of the BAC clone has been sequenced to date. After trimming the sequences, there was an average of 337 good bases. Consequently, 890 sub-clones would be needed for 95% sequence coverage of the BAC clone. This estimate presumes that it is possible to clone and sequence every fragment with equal likelihood. These data clearly demonstrate that this is not necessarily the case. Only 6/48 sub-clones from the *MspI* library could be sequenced. Presumably, the *MspI* (C^CGG) sub-clones which failed to sequence are GC-rich. To obtain sequence from GC-rich templates using the BigDye kit, the reaction conditions should be optimised by reducing the extension temperature and increasing the magnesium concentration (*Perkin Elmer Applied Biosystems*, BigDye kit instructions).

Only clones >500bp were sequenced, but it was demonstrated that *Tsp509I* recognition sites occurred on average once every 138bp of sequence. It would be worthwhile returning to this library and picking additional smaller clones for sequencing, since these are currently not represented in the contigs. For the other existing libraries though, it would not be worthwhile to pick additional clones because it is more likely that a previously sequenced sub-clone would be picked again than a novel one. Instead, to generate additional sequences from the BAC clone, it would be better to create sub-clone libraries with different enzymes.

Even after sequencing the additional sub-clones, it is likely that there will still be gaps in the sequence. Since comparative human sequence is available for the entire region, the sheep contigs could be crudely ordered, so predictions could be made about which contigs

should overlap. Consequently, gaps in the sequence could be filled by sequencing from large (eg. *HindIII* or *PstI)* sub-clones in the putatively overlapping contigs with primers designed from the end of existing unique sequence. The only remaining regions that would be difficult to assemble are those contigs that end with large blocks (>2kb) of repetitive DNA. Mutations in the repeat unit need to be identified within these blocks so the unique mutated bases can be incorporated at the 3' end of a sequencing primer (S.K. Davis, pers. comm.).

In humans, it has been estimated that whilst Alu (a SINE) and L1 (a LINE) account for >90% of all interspersed repeats and may constitute 13-18% of the total DNA in a chromosomal band, neither of these repetitive elements are randomly distributed (Chen and Manuelidis, 1989; Korenburg and Rykowski, 1988). Alu, which is GC-rich (56%), predominates in early replicating, GC-rich, reverse (G-light) bands. Conversely, L1, which is AT-rich (58%), is concentrated in late replicating, AT-rich, G positive (G-dark) bands. It has been proposed that the distribution of Alu and L1 (ie. GC vs AT content of a chromosomal band) may be a major cause of chromosome banding patterns produced by fluorescent dyes (Korenburg and Rykowski, 1988). The distribution of SINE and LINE elements in other mammalian genomes has not been extensively investigated. In mice, as in humans, there is a predominance of SINEs in R-bands and LINEs in G-bands (Boyle *et al.,* 1990). In pigs, this differential distribution also exists, but is much less pronounced (Thomsen and Miller, 1996).

In Bovidae, the main classes of dispersed repeats are BovA SINEs (Bov-A2 and Bov-tA) and BovB LINEs (Lenstra *et al.,* 1993). Indeed, these repeats have been shown to be specific for Ruminantia (Jobse *et al.,* 1995). The BovB repeat was originally described as a SINE, but has since been demonstrated to belong to a bovine dimer-driven (BDDF) LINE

family (Szemraj *et al.*, 1995). Like the human L1 repeat, BovB is commonly truncated at various 5' positions (Korenbrug and Rykowski, 1988; Szemraj *et al.*, 1995). Additionally, the distribution of the 5' ends of BovB in the genome is not random (Szemraj *et al.*, 1995), but the localisation of BovA or BovB to particular chromosomal bands has not been described. Given that BovB is AT-rich (57.2% estimated from GENBANK #M94327, bases 17286-19062; Szemraj *et al.*, 1995), it seems likely that it would follow the pattern of L1 distribution in humans, and localise to G positive (dark) bands. Similarly, BovA, like Alu, is GC-rich (57% estimated from GENBANK #X64124 (Bov-tA) and #X64126 (Bov-A2)), and therefore, would be expected to be concentrated in G-light bands.

The ovine and human sequences from the BAC clones containing the tissue-specific, late-replicating, high-sulphur KAP genes are AT-rich (57.6% and 60.6%, respectively) and LINEs are the predominant repeat class associated with these genes (22% BovB and 17% L1, respectively). Sheep and human BAC clones containing these genes map to corresponding G-dark bands (11q3.1 and 17q12 (Ceratto *et al.*, 1997), respectively). These sequence and physical mapping data provide some evidence that L1 and BovB are distributed similarly within the human and sheep genomes.

Given that it is estimated that interspersed repeats account for ~20% of DNA in a chromosomal band (Chen and Manuelidis, 1989; Korenburg and Rykowski, 1988), a striking feature of both the ovine and human sequences for the KAP-containing BAC clones was the localised, high concentration of repetitive DNA (45% and 41%, respectively) within this ~100kb region. Repeat-rich intergenic regions between KIF genes have been previously described (Romano *et al.*, 1988), but the nature of KAP intergenic regions has not been reported previously. It is interesting to note that whilst KIF genes generally have six introns, KAP genes are intronless (Powell *et al.*, 1983; MacKinnon *et al.*, 1990; Rogers

et al., 1998). It is speculated, herein, that since KAP genes lack introns, they require the accumulated repetitive DNA in the intergenic regions for proper regulation or function.

The type I and type II KIF genes in humans are clustered, separated by 5-10kb intergenic regions (Romano *et al.*, 1988; Rosenberg *et al.,* 1988; Cerratto *et al.* 1997; Rogers *et al.*, 1998). Acidic (type I) cytokeratins and hair keratins have been mapped to human chromosome 17q12-q21, while the basic (type II) keratins are clustered on chromosome human 12q11-q13 (Milisavljevic *et al.*, 1996). Homologous type I and type II keratins in mice are located on chromosomes 11 and 15, respectively (Nadeau *et al.*, 1989; Compton *et al.,* 1991). This study demonstrated that several high sulphur KAP1.n, KAP2.n and KAP3.n genes are also clustered on human chromosome 17q12-21 and map to a homologous region in sheep (11q3.1).

The roles of gene order and clustering, repetitive DNA and other elements, in regulation of the high sulphur KAP genes, could be investigated by introducing systematically modified BAC clones by trangenesis into an animal model. It has widely been suggested that hierarchical, temporal and spatial expression of KAP and KIF genes is due to the organisation and regulation of these genes in clustered domains (Compton *et al.*, 1991; McClaren *et al.*, 1997; Bawden *et al.*, 1998). It has recently been demonstrated using BAC clones for transgenesis that regulatory elements controlling several genes with related functions within a multigene locus can be located more than 50kb away from that locus (Nielsen et al., 1997; Nielsen et al., 1999). Transgenic studies with KAP- or KIF-containing BAC clones would help unravel the complex control of keratin gene expression in the wool follicle.

CHAPTER 8

GENERAL DISCUSSION

GENERAL DISCUSSION

The ultimate goal of livestock gene mapping projects is to isolate the genes responsible for traits of economic importance. In order to do this, dense genome maps are required. In this study, two different strategies for adding markers and genes to the ovine physical map were investigated. Initially, oligonucleotide primers and PCR products were tested as a source of probes for super-fast primed *in situ* targeting (SPRINT). The modified primed *in situ* hybridisation (PRINS) procedure relied on the visualisation of fluorescence emitted by excited molecules of fluorescein attached to nucleotides directly incorporated by *Taq* DNA polymerase into the newly synthesised strand and localised on metaphase chromosomes. While hybridisation signals generated by primers targeting highly repetitive regions, such as satellite DNA, were clearly visible, single copy sequences could not be detected by SPRINT. Consequently, this approach was abandoned in favour of the use of large-insert BAC clones as a source of probes for FISH. The idea of using PCR products directly for *in situ* hybridisation has recently been revisited using tyramide signal amplification for detection (Schriml *et al.*, 1999). Signal strength is magnified ~1000 fold compared with standard FISH, thereby enabling short PCR products (350-1000bp) for single-copy sequences to be detected. Tyramide has yet to be used to detect PCR products containing microsatellite sequences on metaphase chromosomes and such products would have to be carefully designed to ensure there was sufficient unique flanking DNA for specific hybridisation. It would be worthwhile further investigating tyramide technology for cytogenetic mapping in livestock.

In the absence of an effective method for physically localising short, single-copy sequences on metaphase chromosomes in sheep, an ovine BAC library was constructed as a source of stable, large-insert probes for FISH. At least two BAC clones were isolated for

every chromosome and to date, an additional 45 genes and markers have been added to the ovine physical map.

8.1 Use of BAC clones

These physically mapped clones can be used for unambiguous chromosome identification, and in addition to dual localisation experiments to confirm new physical assigments, they are presently being used to identify the chromosomal origin of extended chromatin fibres in fibre-FISH (C.D.K Bottema, pers. comm.).

The BAC clones can also be used for chromosome identification in other species and to examine regions of conserved synteny between species. The set of ovine BAC clones described herein, plus additional clones to represent all 34 deer genetic linkage groups (http://www.ri.bbsrc.ac.uk/cgi-bin/arkdb/browsers/browser.sh?species=deer) are currently being physically mapped in red deer to assign the genetic linkage maps to particular chromosomes (C.D.K Bottema, pers. comm.).

8.2 BAC library construction

In comparison to the recently developed dog (8x), pig (10x) and cattle (12x) BAC libraries, and human (25x) and rat (10x) PAC libraries (Woon *et al.*, 1998; Li *et al.*, 1999; Osoegawa *et al.*, 1998; bacpac.med.buffalo.edu), genome coverage for the sheep BAC library described herein is poor (2x) and the library has several intrinsic biases (discussed in chapter 4). Nevertheless, PCR based screening supports the estimate based on insert size that there is an 88.5% chance of finding any unique sequence within the library. In combination with the other 3x sheep BAC library (Vaiman *et al.*, 1999b), there is theoretically a 99% probability of finding any given sequence in one of the two libraries which is quite adequate for physical mapping studies.

Deep BAC libraries (>10x genome coverage) are being produced to construct ordered contigs and minimum tiling paths for large-scale sequencing of whole genomes (Mahairas *et al.*, 1999; Siegel *et al.*, 1999). To minimise the number of clones within each contig, clones with very large, yet stable, inserts are desirable. It is well documented that the average insert size of BAC clones in many of the libraries that have been produced are smaller than expected, ranging from 100-120kb (Cai *et al.*, 1995; Frijters *et al.*, 1997; Osoegawa *et al.*, 1998). Recently, several modifications to the BAC procedure have resulted in consistently larger clones with average insert sizes from 160-235kb (Osoegawa *et al.*, 1998; Li *et al.*, 1999). Rather than microbeads, agarose plugs are now being used because centrifugation to separate the microbeads from excess mineral oil is avoided. Additionally, it is presumed that DNA encapsulated in a plug migrates through the agarose CHEF gel more uniformly than DNA in microbeads since the potential for the DNA to exit and then re-enter the agarose matrix multiple times (ie. between adjacent microbeads) is greatly reduced (S.K. Davis, pers. comm.).

The principle improvement to the procedure that has resulted in increased insert size is the use of altered CHEF gel conditions for size selection. Prior to digestion, the plugs are electrophoresed under conditions that remove inhibitors (eg. residual salts from DNA preparation) and degraded DNA from the plugs. Intact DNA molecules remain in the plugs and are subsequently digested to give fragments ranging from 150-250kb as usual. The subsequent CHEF gel electrophoresis is critical because small fragments (<120kb), that are usually preferentially cloned, literally run backwards off the gel leaving only the fragments in the desired size range. These fragments are then compressed into a narrow band and extracted from the gel. Electroelution of the size selected DNA from the gel results in better integrity than digestion with an agarase (Strong *et al.*, 1997; Osoegawa *et al.*, 1998),

but the DNA is significantly diluted and must be concentrated by dialysis against TE containing polyethylene glycol (PEG8000) prior to ligation (Osoegawa *et al.,* 1998).

Although this new sizing method produces larger BAC clones, there is a concomitant loss in transformation efficiency, probably due to additional handling of the linearised, high molecular weight DNA. Under optimum conditions with the new method, the transformation efficiency is just $2.5x10^5$cfu/μg (Osoegawa *et al.,* 1998), compared with $1.5x10^6$cfu/μg that was achieved for the ovine BAC library herein (chapter 4). This 6-fold reduction in efficiency greatly increases the cost of library construction since the commercial electrocompetant cells are one of the most expensive components of the library. Precipitation of BAC DNA with tRNA to change the topological form (ie. induce supercoiling) has been reported to increase transformation efficiency 70-fold. However, the results presented were equivalent to those achieved for the sheep library, without precipitation (Zhu and Dean, 1999).

The ovine library herein was transformed by electroporation at 400V, equivalent to a field strength of 16kV/cm. It has previously been shown that the optimum field strength for transformation of large DNA fragments is 13-14kV/cm (Sheng *et al.,* 1995; Frijters *et al.,* 1997; Zimmer *et al.,* 1997) with smaller fragments preferentially transformed at higher field strengths. Consequently, it is apparent that the average insert size of the ovine clones could be improved without any loss in transformation efficiency simply by altering the field strength for electroporation. As proposed by Sheng *et al.* (1995), pulse regimens for electroporation should be investigated because open circular DNA molecules of different sizes require different amounts of time to pass through pores in the cells. Although a single-pulse is currently used, it is envisaged that an initial high voltage pulse to open the pore followed by a period of lower voltage to maintain the open pore would be even more effective for transformation of large BAC molecules.

The best strategy to adopt for BAC library construction (ie. large BAC inserts vs high efficiency) will ultimately depend on the objective behind library development. To simply add markers and genes to the physical and genetic maps, BAC clones in the range from 100-120kb are sufficient, and so, the most efficient and cost effective method of producing clones should be employed. For contig development, the cost of arraying the BAC clones will ultimately dwarf the cost of library construction, so modified procedures should be used to generate clones with the largest possible inserts to minimise the number of clones within each tiling path.

8.3 Ordering BAC clones

Ordered arrays of BAC clones are being developed by a number of methods, including BAC end sequencing (Mahairas *et al.,* 1999), radioactive fingerprinting by restriction landmark genome scanning (RLGS; Okazaki and Hayashizaki, 1997), and fluorescent fingerprinting (Gregory *et al.,* 1997; Marra *et al.,* 1997; Ding *et al.,* 1999; Cao *et al.,* 1999). As part of the human genome project, 450,000 BAC clones are being arrayed by sequencing each end of the clones and by fingerprinting each clone with a single enzyme (*HindIII*) to generate 'sequence-tagged connectors' (STC) every 3.3kb across the genome (Mahairas *et al.,* 1999). These STC are seeds for sequence extension from minimally overlapping BAC clones. Contig assembly using this strategy relies primarily on sequence data with fingerprints of each clone being used for confirmation to avoid generating chimeric contigs across repetitive regions.

Without sequence data, fingerprints generated by a single restriction enzyme are not sufficient for contig assembly because of the inability to distinguish small overlaps or assembly errors (Taylor *et al.,* 1996; Ding *et al.*, 1999). Both RLGS and fluorescent fingerprinting employ two or more restriction enzymes (Okazaki and Hayashizaki, 1997;

Gregory *et al.*, 1997; Ding *et al.*, 1999). Typically, BAC DNA is cleaved with a 6bp-cutter (eg. *HindIII*) to generate fragments that are subsequently labelled (radioactively or with different coloured fluorophors). The labelled fragments are then digested with one or more common endonucleases (eg. *Sau3A*, *EcoRV*). Only the subset of fragments that remain labelled after the second cleavage are visualised (~30 bands). RLGS has not been used as extensively as fluorescent fingerprinting because it depends on two-dimensional electrophoresis to separate the labelled fragments (Okazaki and Hayashizaki, 1997). Fluorescent fingerprints can be run on an automated sequencer and several BAC clones labelled with different coloured fluorophors can be multiplexed in a single lane (Gregory *et al.*, 1997). Precise sizing of fragments is achieved by including a size standard in every lane. The data are analysed with Genescan software (*Perkin Elmer*) and can be loaded directly into FPC software (ftp.sanger.ac.uk) for contig assembly (Soderlund *et al.*, 1997). FPC uses an algorithm to cluster the clones into contigs based on the probability that they must overlap due to the similarity of their banding patterns.

8.4 Use of BAC contigs

Apart from developing minimum tiling paths for sequencing, ordered arrays of BAC clones will be useful for positionally cloning genes of economic importance. If two flanking markers that have tight linkage to the gene of interest have been identified, then the array of BAC clones between those markers could be examined. By isolating novel markers from within the arrayed clones, the critical region containing the gene of interest could be reduced to a managable size for sequencing by shot-gun subcloning.

Sequencing the ends of arrayed BAC clones with known positions in the genome will be valuable in the production of the next generation of genetic linkage maps using single nucleotide polymorphisms (SNPs). In humans, SNPs have been shown to occur about once

every 1000bp, although expressed sequences were less polymorphic than random genomic sequences (Wang et al., 1999). It is expected that a similar abundance of SNPs will be found in livestock species. One strategy to identify livestock SNPs would be to amplify BAC end sequences in a breed panel and look for base variations between individuals. It will ultimately be important to identify SNPs that occur both between and within breeds. Since a SNP is only a two allele system, it is not yet known how many SNPs will be required at any one locus (ie. haplotype analysis) to provide as much information as the multiple allele microsatellites. The huge advantage of SNPs over microsatellites is that the oligonucleotides used for SNP detection can be arrayed on a chip so that a whole genome scan can be conducted for thousands of markers simultaneously (Sapolsky *et al.,* 1999; Wang *et al.,* 1999).

Currently, the ovine BAC library contains an insufficient number of clones for genome-wide contig development. Generating additional BAC clones for sheep with larger inserts will facilitate development of ordered arrays of clones for marker and gene identification in the future.

8.5 BAC clones and functional studies

Large, stable BAC clones are also desirable for functional studies because they are more likely to contain all the long-range regulatory elements required for position-independent expression than short plasmid inserts (Peterson *et al.*, 1995; Lien *et al.*, 1997; Nielsen *et al.*, 1999; Stinnakre *et al.*, 1999). One hindrance to the use of BAC clones for functional studies in mammalian systems is that the original BAC vector, pBeloBAC11, did not contain a selection system or reporter genes suitable for expression in mammalian cells (Kim *et al.,* 1998). Complex systems were used to modify the clones for functional studies in mammalian cells (Mejía and Monaco, 1997; Yang *et al.,* 1997; Chatterjee and Coren,

1997; Kim *et al.,* 1998). In addition to modifying the vector, systems were developed to make targeted modifications to the BAC insert, but these methods either require shuttle vectors or are not suitable for use in the host DH10ß *E. coli* cells (Boren *et al.,* 1996; Yang *et al.,* 1997; Chatterjee and Coren, 1997; Messerle *et al.,* 1997; Jessen *et al.,* 1998). A new BAC vector, pEBAC, based on the backbone of pBeloBAC11 also includes hygromycin and thymidine kinase genes for selection in eukaryotic cells as well as oriP and EBNA-1 genes to facilitate episomal maintenance (Narayanan *et al.,* 1999).

Targeted modification can be made to BAC clones by homologous recombination in DH10β cells using pEBAC in combination with a second plasmid, pGETrec (Narayanan *et al.,* 1999). Unlike previous methods which require a new shuttle vector to be designed for every modification, the pGETrec vector does not carry the modified sequences. Instead, the L-arabinose inducible expression plasmid carries the bacteriophage λ *gam* gene that inhibits recBCD nuclease which is active in DH10β cells. In the absence of recBCD nuclease activity, linear double-stranded DNA is not degraded. Consequently, PCR products containing the desired modifications (eg. a reporter gene, insertion, deletion or point mutation) can be introduced into the BAC clones by electroporation followed by homologous recombination (Narayanan *et al.,* 1999). By using an inducible system where the ability to introduce modifications is only transient, unwanted rearrangements due to recombination between repetitive elements is minimised. The modified BAC can be purified away from pGETrec by a standard plasmid preparation and subsequent electroporation into DH10β cells. The availability of this simple system for modifying BAC clones will facilitate functional studies to identify gene function and the regulatory elements that control their function, and will ultimately lead to the use of modified BAC clones in gene therapy or transgenesis programmes.

The utility of BAC clones for investigating genome organisation was exemplified in this study by considering the keratin associated protein (KAP) gene family. It was demonstrated that several KAP and keratin genes are clustered within a single 100kb ovine BAC clone. NotI digestion resected the insert intact (chapter 7), so it could easily be ligated to the new pEBAC vector for expression in mammalian cells. Future experiments will need to demonstrate that the cloned sequence is sufficient for position-independent, tissue-specific expression within the wool follicle. Once it is established that long-range regulatory elements are contained within the BAC, subsequent modifications can be performed to investigate gene function and regulation. For instance, by generating a series of deletions by homologous recombination, it would be possible to determine whether the unusually high percentage of repetitive DNA, or other motifs within the sequence, were important for gene function and would enable any regulatory elements to be identified. It would also be interesting to investigate whether gene order affects the expression cascade within the follicle by changing the position of genes within the BAC.

Even with sophisticated mapping tools such as large-insert clones, FISH, radiation hybrids, sequencing and fingerprinting, problems that will continue to plague genome mapping are the correct identification of members of large gene families and distinguishing genes from pseudogenes. As highlighted in this study for the SCYA and KAP genes, sequence data alone is sometimes insufficient to confirm the identity of genes that have arisen by duplication events because sequence conservation is very high. Avoiding errors in gene identification is vital for comparative mapping where inferences are made about the location and identity of neighbouring genes based on previous mapping information in other species. Only coincident mapping data is sufficient to confirm the same gene has been isolated. Simultaneous analysis of sequence data, physical and genetic linkage results

and fluorescent fingerprints will help avoid making incorrect gene designations and map assignments.

8.6 Summary

An ovine BAC library consisting of 59,904 clones with an average insert size of 103kb, which corresponds to about two genome equivalents was constructed as a resource for genome mapping in sheep. The rapid identification of 69 BAC clones by PCR-based screening, and the addition of 45 of these clones to the physical map by FISH, clearly demonstrates that the ovine BAC library is an efficient tool for map development. These new physical assignments represent a 52% increase in the number of markers mapped to discrete bands in sheep. Furthermore, these additional localisations enabled a preliminary evaluation of the alignment between the ovine physical and genetic maps. Examination of 9 intervals gave an estimate of the whole sheep genome of 2976 ± 246SEM which is similar to the previous estimate based on chiasmata counts (Chapman and Bruere, 1977).

Although PCR products generated from each BAC clone were sequenced prior to FISH and in all cases aligned with published sequence >85%, sequence data alone was not always sufficient to confirm the identity of members of large gene families (eg. SCYA). Coincident mapping information is the only way to guarantee that the correct gene has been isolated. Verification of all the physical assignments by dual FISH and genetic linkage analysis has been commenced.

The ovine BAC library will continue to be useful for comparative mapping and positional cloning of economically important traits. BAC clones containing 12 microsatellites and 15 genes were mapped by FISH to homologous locations on bovine chromosomes, and 5 microsatellites and 1 gene were physically mapped in goat.

Comparative analysis of human sequence for the keratin associated proteins on HSA17 was critical to the construction of a long-range restriction map of a 100kb BAC clone containing KAP1.1. At least four other KAP genes and a keratin gene were also located within the BAC clone which mapped to OAR11q3.1. To construct the fine map of the KAP region, more than 40% of the BAC was sequenced by shot-gun subcloning. An unusual feature of the sequence was that it was 45% repetitive. KAP genes are intronless and it is proposed that these genes require accumulated repetitive DNA in intergenic regions for proper regulation and function. Further investigation of the organisation and function of these clustered genes and transgenesis studies with modified BAC clones would help unravel the complex control of keratin gene expression in the follicle.

The development and characterisation of the ovine BAC library and the mapping data presented herein represent a significant contribution to the sheep mapping community. The ovine BAC library is a powerful new tool for sheep genome mapping and is available for collaborative research.

REFERENCES

REFERENCES

Abbott, C. and Povey, S. (1995). Somatic Cell Hybrids: The Basics. IRL Press, Oxford.

Adinolfi, M., Davies, A., Sharif, S., Soothill, P. and Rodeck, C. (1993). Detecton of trisomy 18 and Y-derived sequences in fetal nucleated cells obtained by transcervical flushing. Lancet 342: 403-404.

Aïssani, B. and Bernardi, G. (1991). CpG islands: features and distribution in the genome of vertebrates. Gene 106: 173-183.

Al-Bayati, H.K., Duscher, S., Kollers, S., Rettenberger, G., Fries, R. and Brenig, B. (1999). Construction and characterization of a porcine P1-derived artificial chromosome (PAC) library covering 3.2 genome equivalents and cytogenetical assignment of six type I and type II loci. Mamm. Genome 10 (6): 569-572.

Albertsen, H.M., Abderrahim, H., Cann, H.M., Dausset, J., Le Paslier, D. and Cohen, D. (1990). Construction and characterization of a yeast artificial chromosome library containing seven haploid genome equivalents. Proc. Natl. Acad. Sci. USA 87: 4256-4260.

Allen, L. (1997). Comparative mapping in sheep and cattle. Honours Thesis. University of Adelaide. South Australia.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25: 3389-3402.

Ansari, H.A., Pearce, P.D., Maher, D.W., Malcolm, A.A., Wood, N.J., Phua, S.H. and Broad, T.E. (1992). Resolving ambiguities in the sheep karyotype: a molecular cytogenetic approach. Proceedings of the 10th European colloquium on cytogenetics of domestic animals. 10: 21-25 (1992).

Ansari, H.A., Pearce, P.D., Maher, D.W., Malcolm, A.A. and Broad, T.E. (1993). Resolving ambiguities in the karyotype of domestic sheep (*Ovis aries*). Chromosoma 102: 340-347.

Ansari, H.A., Bosma, A.A., Bunch, T.D., Long, S.E. and Popescu, C.P. (1994). Clarification of chromosome nomenclature in the sheep (*Ovis aries*). Report of the committee for the standardization of the sheep karyotype. Cytogenet. Cell Genet. 67: 114-115.

Ansari, H.A., Pearce, P.D., Maher, D.W. and Broad, T.E. (1995). Human chromosome 10 loci map to three different sheep chromosomes. Mamm. Genome 6: 46-48.

Ansari, H.A., Maher, D.W., Pearce, P.D. and Broad, T.E. (1996). Resolving the ambiguities in the karyotype of domestic sheep (*Ovis aries*). Chromosoma 105: 62-67.

Ansari, H.A., Bosma, A.A., Broad, T.E., Bunch, T.D., Long, S.E., Maher, D.W., Pearce, P.D. and Popescu, C.P. (1999). Standard G-, Q-, and R-banded ideograms of the domestic sheep (*Ovis aries*): homology with cattle (*Bos taurus*). Report of the committee for the standardization of the sheep karyotype. Cytogenet. Cell Genet. 87: 134-142.

Antoch, M.P., Song, E.-J., Chang, A.-M., Vitaterna, M.H., Zhao, Y., Wilsbacher, L.D., Sangoram, A.M., King, D.P., Pinto, L.H. and Takahashi, J.S. (1997). Functional identification of the mouse circadian *Clock* gene by transgenic BAC rescue. Cell 89: 655-667.

Ashworth, L.K., Alegria-Hartman, M., Burgin, M., Devlin, L., Carrano, A.V. and Batzer, M.A. (1995). Assembly of high-resolution bacterial artificial chromosome, P1-derived artificial chromosome, and cosmid contigs. Anal. Biochem. 224 (2): 564-571.

Athwal., R.S., Smarsh, M., Searle, B.M. and Deo, S.S. (1985). Integration of a dominant selectable marker into human chromosomes and transfer of marked chromosomes to mouse cells by microcell fusion. Somatic Cell Mol. Genet. 11: 177-187.

Ausubel, F.M., Brent R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. and Struhl, K. (eds). (1992). Preparation of genomic DNA from mammalian tissue. Ch 2.2. In: Current protocols in molecular biology. Vol. 1. Wiley-Interscience, New York. pp. 2.2.1-2.2.3.

Baker, A. and Cotten, M. (1997). Delivery of bacterial artificial chromosomes into mammalian cells with psoralen-inactivated adenovirus carrier. Nucleic Acids Res. 5 (10): 1950-1956.

Barendse, W., Vaiman, D., Kemp, S.J., Sugimoto, Y., Armitage, S.M., Williams, J.L., Sun, H.S., Eggen, A., Agaba, M., Aleyasin, S.A., Band, M., Bishop, M.D., Buitkamp, J., Byrne, K., Collins, F., Cooper, L., Coppettiers, W., Denys, B., Drinkwater, R.D., Easterday, K.,

Elduque, C., Ennis, S., Erhardt, G., Ferretti, L., Flavin, N., Gao, Q., Georges, M., Gurung, R., Harlizius, B., Hawkins, G., Hetzel, J., Hirano, T., Lenstra, J.A., Leveziel, H., Lewin, H.A., Leyhe, B., Lil, L., Martin Burriel, I., McGraw, R.A., Miller, J.R., Moody, D.E., Moore, S.S., Nakane, S., Nijman, I.J., Olsaker, I., Pomp, D., Rando, A., Ron, M., Shalom, A., Teale, A.J., Thieven, U., Urquhart, B.G.D., Vage, D.-I., Van de Weghe, A., Varvio, S., Velmala, R., Vilkki, J., Weikard, R., Woodside, C., Womack, J.E., Zanotti, M. and Zaragoza, P. (1997). A medium-density genetic linkage map of the bovine genome. Mamm. Genome 8: 21-28.

Bauman, J.G.J., Wigant, J., Borst, P. and van Duijn, P. (1980). A new method for fluorescence microscopical localization of specific DNA sequences by *in situ* hybridisation of fluorochrome-labelled RNA. Exp. Cell Res. 128: 485-490.

Bawden, C.S., Powell, B.C., Walkder, S.K. and Rogers, G.E. (1998). Expression of a wool intermediate filament keratin transgene in sheep fibre alters structure. Transgenic Res. 7 (4): 273-287.

Beckman, J.S. and Soller, M. (1983). Restriction fragment length polymorphisms in genetic improvement: methodologies, mapping and costs. Theoretical and Applied Genetics 67: 35-43.

Bickmore, W. and Craig, J. (1997). Chromosome bands: patterns in the genome. Springer-Verlag, Heidelberg, Germany.

Bishop, M.D., Kappes, S.M., Keele, J.W., Stone, R.T., Sunden, S.L.F., Hawkins, G.A., Solinas Toldo, S., Fries, R., Grosz, M.D., Yoo, J. and Beattie, C.W. (1994). A genetic linkage map of cattle. Genetics 136: 619-639.

Boren, J., Lee, I., Callow, M.J., Rubin, E.M. and Innerarity, T.L. (1996). A simple and efficient method for making site-directed mutants, deletions, and fusions of large DNA such as P1 and BAC clones. Genome Res. 6: 1123-1130.

Boyle, A.L., Ballard, S.G. and Ward, D.C. (1990). Differential distribution of long and short interspersed element sequences in the mouse genome: chromosome karyotyping by fluorescence *in situ* hybridization. Proc. Natl. Acad. Sci. USA 87 (19): 7757-7761.

Bray-Ward, P., Bowlus, C., Choi, J., Paslier, D.L., Weissenbach, J. and Gruen, J.R. (1996). FISH-Mapped CEPH YACs spanning 0 to 46 cM on human chromosome 6. Genomics 36 (1): 104-111.

Brigati, D.J., Myerson, D., Leary, J.J., Spalholz, B., Travis, S.Z., Fong, C.K.Y., Hsuing, G.D. and Ward, D.C. (1983). Detection of viral genomes in cultures cells and paraffin-embedded tissue sections using biotin-labelled hybridization probes. Virology 126: 32-50.

Broad T.E. and Hill D.F. (1994). Mapping the sheep genome: practice, progress and promise. Br. Vet J. 150 (3): 237-52.

Broad, T.E., Lewis, P.E., Burkin, D.J., Gleeson, A.J., Carpenter, M.A., Jones, C., Pearce, P.D., Maher, D.W. and Ansari, H.A. (1995). Thirteen loci physically assigned to sheep

Chromosome 2 by cell hybrid analysis and *in situ* hybridization. Mamm. Genome 6: 862-866.

Broad, T.E., Lambeth, M., Burkin, D.J., Jones, C., Pearce, P.D., Maher, D.W. and Ansari, H.A. (1996). Physical mapping confirms that sheep chromosome 10 has extensive conserved synteny with cattle chromosome 12 and human chromosome 13. Anim. Genet. 27: 249-253.

Broad, T.E., Hayes, H. and Long, S.E. (1997). Cytogenetics: physical chromosome maps. In: Piper, L. and Ruvinsky, A. (eds) "The genetics of sheep." CAB International, pp. 241-296.

Broad, T.E., Hill, D.F., Maddox, J.F., Montgomery, G.W. and Nicholas, F.W. (1998). The sheep gene map. ILAR 39: 160-170.

Broom, M.F. and Hill, D.F. (1994). Construction of a large-insert yeast-artificial chromosome library from sheep DNA. Mamm. Genome 5: 817-819.

Brunner, R.M., Goldammer, T., Furbass, R., Vanselow, J. and Schwerin, M. (1998). Genomic organization of the bovine aromatase encoding gene and a homologous pseudogene as revealed by DNA fiber FISH. Cytogenet. Cell Genet. 82: 37-40.

Buchanan, F.C., Littlejohn, R.P., Galloway, S.M. and Crawford, A.M. (1993). Microsatellites and associated repetitive elements in the sheep genome. Mamm. Genome 4: 258-264.

Buckle, V.J. and Craig, I.W. (1986). *In situ* hybridisation. Ch. 6. In: Davies, K.E. (ed). Human genetic diseases: a practical approach. IRL Press, Oxford.

Buitkamp, J., Ewald, D., Schalkwyk, L., Weiher, M., Masaanda, J., Sazanov, A., Lehrach, H. and Fries, R. (1998). Construction and characterisation of a gridded chicken cosmid library with four-fold genomic coverage. Anim. Genet. 29 (4): 295-301.

Burke, D.T., Carle, G.F. and Olson, M.Y. (1987). Cloning of large sements of exogenous DNA into yeast by means of artificial chromosome vectors. Science 238: 806-812.

Burkholder, G.D. (1993). The basis of chromosome banding. Applied Cytogenetics 19 (6): 181-186.

Burkin, D.J., Morse, M.G., Broad, T.E., Pearce, P.D., Ansari, H.A., Lewis, P.E. and Jones, C. (1993). Mapping the sheep genome: production of characterized sheep x hamster cell hybrids. Genomics 16: 466-472.

Burkin, D.J., Broad, T.E. and Jones, C. (1996). The chromosomal distribution and organization of sheep satellite I and II centromeric DNA using characterized sheep-hamster somatic cell hybrids. Chromosome Res. 4: 49-55.

Burkin, D.J., O'Brien, P.C.M., Broad, T.E., Hill, D.F., Jones, C.A., Wienburg, J. and Ferguson-Smith, M.A. (1997a). Isolation of chromosome-specific paints from high-resolution flow-sorted karyotypes of the sheep (*Ovis aries*). Chromosome Res. 5: 102-108.

Burkin, D.J., Yang, F., Broad, T.E., Wienberg, J., Hill, D.F. and Ferguson-Smith, M.A. (1997b). Use of the Indian muntjac ideogram to align conserved chromosomal segments in sheep and human genomes by chromosome painting. Genomics 46: 143-147.

Burkin, D.J., Broad, T.E., Lambeth, M.R., Burkin, H.R. and Jones, C. (1998). New gene assignments using a complete characterized sheep-hamster somatic cell hybrid panel. Animal Genetics 29: 48-54.

Cai, L., Taylor, J.F., Wing, R.A., Gallagher, D.S., Woo, S.S. and Davis, S.K. (1995). Construction and characterization of a bovine bacterial artificial chromosome library. Genomics 29: 413-425.

Cai, L., Schalkwyk, L.C., Schoeberlein-Stehli, A., Zee, R.Y.L., Smith, A., Haaf, T., Georges, M., Lehrach, H. and Lindpaintner, K. (1997). Construction and characterization of a 10-genome equivalent yeast artificial chromosome library for the laboratory rat *Rattas norvegicus*. Genomics 39: 385-392.

Cai, W.W., Reneker, J., Chow, C.W., Vaishnav, M. and Bradley, A. (1998). An anchored framework BAC map of mouse chromosome 11 assembled using multiplex oligonucleotide hybridization. Genomics 54 (3): 387-97.

Callen, D.F., Yip, M.Y. and Eyre, H.J. (1997). Rapid detection of euchromatin by Alu-PRINS: use in clinical cytogenetics. Chromosome Res. 5 (2): 81-85.

Cao, Y., Kang, H.L., Xu, X., Wang, M., Dho, S.H., Huh, J.R., Lee, B.J., Kalush, F., Bocskai, D., Ding, Y., Tesmer, J.G., Lee, J., Moon, E., Jurecic, V., Baldini, A., Weier, H.U., Doggett, N.A., Simon, M.I., Adams, M.D. and Kim, U.J. (1999). A 12-Mb complete coverage BAC contig map in human chromosome 16p13. 1-p11.2. Genome Res. 9 (8): 763-774.

Carey, A. (1997). Use of commercially available radiation hybrid panels. Ch.3. In: Dracopoli, N.C., Haines, J.L., Korf, B.R., Moir, D.T., Morton, C.C., Seidman, C.E., Seidman, J.G. and Smith, D.R. (eds). Current Protocols in Human Genetics. John Wiley & Sons, Inc., New York, USA. 3.5.1-3.5.10.

Casperson, T., Farber, S., Foley, G.F., Kudynowski, J., Modest, E.J., Simonsson, E., Wagh, U. and Zech, L. (1968). Chemical differentiation along metaphase chromosomes. Exp. Cell Res. 49: 219-222.

Ceratto, N., Dobkin, C., Carter, M., Jenkins, E., Yao, X.-L., Cassiman, J.-J., Aly, M.S., Bosco, P., Leube, R., Langbein, L., Feo, S. and Romano, V. (1997). Human type I cytokeratin genes are a compact cluster. Cytogenet. Cell Genet. 77: 169-174.

Chapman, H.M. and Bruere, A. (1977). Chromosome mophology during meiosis of normal and Robertsonian translocation-carrying rams (*Ovis aries*). Can. J. Genet. Cytol. 19: 93-102.

Chatterjee, P.K. and Coren, J.S. (1997). Isolating large nested deletions in bacterial and P1 artificial chromosomes by *in vivo* P1 packaging of products of Cre-catalysed recombination

between the endogenous and a transposed *loxP* site. Nucleic Acids Res. 25 (11): 2205-2212.

Chen, M., Compton, S.T., Coviello, V.F., Green, E.D. and Ashlock, M.A. (1997). Transient gene expression from yeast artificial chromosome DNA in mammalian cells is enhanced by adenovirus. Nucleic Acids Res. 25 (21): 4416-4418.

Chen, S., Lao, W., Ni, Z. and Shi, L. (1982). Constructions of gene libraries of *Rana nigromaculata* and *Mus musculus* with cosmid pHC79. Sci Sin [B]. 25 (3): 261-267.

Chen, T.L. and Manuelidis, L. (1989). SINEs and LINEs cluster in distinct DNA fragments of Giemsa band size. Chromosoma 98 (5): 309-316.

Chen, Z.-Q., Lautenberger, J.A., Lyons, L.A., McKenzie, L. and O'Brien, S.J. (1999). A human genome map of comparative anchor tagged sequences. J. Hered. 90: 477-484.

Chowdhary, B.P., Fronicke, L., Gustavsson, I. and Scherthan, H. (1996). Comparative analysis of the cattle and human genomes: detection of ZOO-FISH and gene mapping-based chromosomal homologies. Mamm. Genome 7: 297-302.

Cianfriglia, M. (1979). Cytogenetic characterization of Chinese hamster-sheep somatic cell hybrids. Bolletino della Societa Italiano di Biologia Sperimentale (Napoli). 55: 2201-2207.

Cinti, C., Santi, S. and Maraldi, N.M. (1993). Localization of single copy gene by PRINS technique. Nucleic Acids Res. 21 (24): 5799-5800.

Clarke, L. and Carbon, J. (1976). A colony bank containing synthetic Col E1 hybrid plasmids representative of the entire *E. coli* genome. Cell 9: 91-99.

Claverie, J.M. (1994). A streamlined random sequencing strategy for finding coding exons. Genomics 23 (3): 575-581.

Collins, F.S. and Weissman, S.M. (1984). Directional cloning of DNA fragments at a large distance from an initial probe: A circularization method. Proc. Natl. Acad. Sci. USA 81: 6812-6816.

Comincini, S., Drisaldi, B. and Ferretti, L. (1997). Isolation of coding sequences from bovine cosmids by means of exon trapping. Mamm. Genome 8 (7):486-490.

Compton, J.G., Ferrara, D.M., Yu, D.W., Recca, V., Freedberg, I.M. and Bertolino, A.P. (1991). Chromosomal localization of mouse hair keratin genes. Ann. N.Y. Acad. Sci. 642: 32-43.

Cox, D.R., Burmeister, M., Price, E.R., Kim, S. and Myers, R.M. (1990). Radiation hybrid mapping: A somatic cell genetic method for constructing high resolution maps of mammalian chromosomes. Science 250: 245-250.

Cox, G.S., Gutkin, D.W., Haas, M.J. and Cosgrove, D.E. (1998). Isolation of an Alu repetitive DNA binding protein and effect of CpG methylation on binding to its recognition sequence. Biochim. Biophys. Acta 1296: 67-87.

Crawford, A.M., Dodds, K.G., Ede, A.J., Pierson, C.A., Montgomery, G.W., Garmonsway, H.G., Beattie, A.E., Davies, K., Maddox, J.F., Kappes, S.W., Stone, R.T., Nguyen, T.C., Penty, J.M., Matthews, M.E., Hulme, D.J., Beh, K.J., McGraw, R.A. and Beattie, C.W. (1995). An autosomal genetic linkage map of the sheep genome. Genetics 140: 703-724.

Davies, K.P., Maddox, J.F., Harrison, B. and Drinkwater, R.D. (1996). Ovine dinucleotide repeat polymorphism at eight anonymous loci. Anim. Genet. 27: 381-382.

Davis, S.L.F., Taylor, J.F., Stillwell, D.R., Gill, C.A. and Davis, S.K. (1999). Comparative sequence analysis of the down syndrome critical region in humans and cows identifies genes encoding rare transcripts. Proceedings of the Genome Sequencing and Biology meeting, Cold Spring Harbor, USA.

Davisson, M.T. and Roderick, T.H. (1989). Linkage map. In: Lyon, M.F. and Searles, A.G. (eds.). Genetic variants and strains of the laboratory mouse. Gustay Fischer Verlag, Stuttgart. pp. 416-427.

de Frotos, R., Kimura, K. and Peterson, K.R. (1989). *In situ* hybridization of Drosophila polytene chromosomes with digoxygenin-dUTP labeled probes. Trends Genet. 5: 366.

de Gortari, M.J., Freking, B.A., Kappes, S.M., Leymaster, K.A., Crawford, A.M., Stone, R.T. and Beattie, C.W. (1997). Extensive genomic conservation of cattle microsatellite heterozygosity in sheep. Anim. Genet. 28: 274-290.

de Gortari, M.J., Freking, B.A., Cuthbertson, R.P., Kappes, S.., Keele, J.W., Stone, R.T., Leymaster, K.A., Dodds, K.G., Crawford, A.M. and Beattie, C.W. (1998). A second-generation linkage map of the sheep genome. Mamm. Genome 9: 204-209.

Deloukas, P., Schuler, G.D., Gyapay, G., Beasley, E.M., Soderlund, C., Rodriguez-Tome, R., Hui, L., Matise, T.C., McKusick, K.B., Beckmann, J.S., Bentolila, S., Bihoreau, M., Birren, B.B., Browne, J., Butler, A., Castle, A.B., Chiannilkulchai, N., Clee, C., Day, D.J., Dehejia, A., Dibling, T., Drouot, N., Duprat, S., Fizames, C., Fox, S., Gelling, S., Green, L., Harrison, P., Hocking, R., Holloway, E., Hunt, S., Keil, S., Lijnzaad, P., Louis-Dit-Sully, C., Ma, J., Mendis, A., Miller , J., Morissette, J., Muselet, D., Nusbaum, H.C., Peck, A., Rozen, S., Simon, D., Slonim, D.K., Staples, R., Stein, L.D., Stewart, E.A., Suchard, M.A., Thangarajah, T., Vega-Czarny, N., Webber, C., Wu, X., Hudson, J., Auffray, C., Nomura, N., Sikela, J.M., Polymeropoulos, M.H., James, M.R., Lander, E.S., Hudson, T.J., Myers, R.M., Cox, D.R., Weissenbach, J., Boguski, M.S. and Bentley, D.R. (1998). A physical map of 30,000 human genes. Science 282 (5389): 744-746.

Dib, C., Faure, S., Fizames, C., Samson, D., Drouot, N., Vignal, A., Millasseau, P., Marc, S., Hazan, J., Seboun, E., Lathrop, M., Gyapay, G., Morissette, J. and Weissenbach, J. (1996). A comprehensive genetic map of the human genome based on 5,264 microsatellites. Nature 380: 152-154.

Dietrich, W., Katz, H., Lincoln, S.E., Shi, H.-E., Friedman, J., Dracopoli, N. and Lander, E. (1992). A genetic map of the mouse suitable for typing intraspecific crosses. Genetics 131: 423-447.

Dietrich, W.F., Miller, J., Steen, R., Merchant, M.A., Damron-Boles, D., Husain, Z., Dredge, R., Daly, M.J., Ingalls, K.A., O'Connor, T.J., Evans, C.A., DeAngelis, M.M., Levinson, D.M., Kruglyak, L., Goodman, N., Copeland, N.G., Jenkin, N.A., Hawkins, T.L., Stein, L., Page, D.C. and Lander, E.S. (1996). A comprehensive genetic map of the mouse genome. Nature 380: 149-152.

Ding, Y., Johnson, M.D., Colayco, R., Chen, Y.J., Melnyk, J., Schmitt, H. and Shizuya, H. (1999). Contig assembly of bacterial artificial chromosome clones through multiplexed fluroescence-labeled fingerprinting. Genomics 56: 237-246.

Drinkwater, R.D., Harrison, B., Davies, K.P. and Maddox, J.F. (1997). Ovine anonymous dinucleotide repeat polymorphism at the CSRD264, CSRD269, CSRD270, CSRD287, CSRD2108, CSRD2138 and CSRD2164 loci. Anim. Genet. 28: 70-71.

Dutrillaux, B. and Viegas-Pequinot, E. (1981). High-resolution R- and G- banding in the same preparation. Hum. Genet. 57: 93-95.

Echard, G., Broad, T.E., Hill, D. and Pearce, P. (1994). Present status of the ovine gene map (*Ovis aries*); comparison with the bovine map (*Bos taurus*). Mamm. Genome 5 (6): 324-332.

Eichenbaum, S. and Krumins, E. (1983). A simple and reliable method of chromosome banding for prenatal cytogenetics using a bromodeoxyuridine pulse. Prenat. Diagn. 3: 291-296.

Ellegren, H. (1993). Abundant $(A)_n \bullet (T)_n$ mononucleotide repeats in the pig genome: linkage mapping of the porcine APOB, FSA, ALOX12, PEPN and RLN loci. Anim. Genet. 24: 367-372.

Ellegren, H., Chowdhary, B., Johansson, M. and Andersson, L. (1994). Integrating the porcine physical and linkage map using cosmid-derived markers. Anim. Genet. 25: 155-164.

Englander, E.W., Wolffe, A.P. and Howard, J. (1995). Nucleosome positioning by human Alu elements in chromatin. J. Biol. Chem. 270 (17): 10091-10096.

Evans, H.J., Buckland, R.A. and Sumner, A.G. (1973). Chromosome homology and heterochromatin in goat, sheep and ox studied by banding techniques. Chromosoma 42: 383-402.

Ferretti, L., Urquhart, B.G., Eggen, A., Olsaker, I., Harlizius, B., Castiglioni, B., Mezzelani, A., Solinas Toldo, S., Thieven, U., Zhang, Y., Morgan, A.L., Teres, V.M., Schwerin, M., Martin-Burriel, I., Chowdhary, B.P., Erhardt, G., Nijman, I.J., Cribiu, E.P., Barendse, W., Leveziel, H., Fries, R. and Williams J.L. (1997). Cosmid-derived markers anchoring the bovine genetic map to the physical map. Mamm. Genome 8 (1): 29-36.

Fratini, A., Powell, B.C., Hynd, P.I., Keough, R.A. and Rogers, G.E. (1994). Dietary cysteine regulates the levels of mRNAs encoding a family of cysteine-rich proteins of wool. J. Invest. Dermatol. 102: 178-185.

Frijters, A.C.J., Zhang, Z., van Damme, M., Wang, G.-L., Ronald, P.C. and Michelmore, R.W. (1997). Construction of a bacterial artificial chromosome library containing large *EcoRI* and *HindIII* genomic fragments of lettuce. Theor. Appl. Genet. 94: 390-399.

Fronicke, L., Chowdhary, B.P., Scherthan, H. and Gustavsson, I. (1996). A comparative map of the porcine and human genomes demonstrates ZOO-FISH and gene mapping-based chromosomal homologies. Mamm. Genome 7: 285-290.

Gallagher, D.S., Yang, Y.P., Burzlaff, J.D., Womack, J.E., Stelly, D.M., Davis, S.K. and Taylor, J.F. (1998). Physical assignment of six type I anchor loci to bovine chromosome 19 by fluorescence *in situ* hybridization. Anim. Genet. 29 (2): 130-134.

Gallagher, D.S., Schlapfer, J. Burzlaff, J.D., Womack, J.E., Stelly, D.M., Davis, S.K. and Taylor, J.F. (1999). Cytogenetic alignment of the bovine chromosome 13 genome map by fluorescence *in-situ* hybridization of human chromosome 10 and 20 comparative markers. Chromosome Res. 7 (2): 115-119.

Galloway, S.M., Hanrahan, V., Dodds, K.G., Potts, M.D., Crawford, A.M. and Hill, D.F. (1996). A linkage map of the ovine X chromosome. Genome Res. 6: 667-677.

Garbett, C.A. (1994). Isolation of microsatellites from the bovine genome. Honours thesis. Flinders University of South Australia.

Gardiner, A.C., Corcoran, M.M. and Oscier, D.G. (1997). Cytogenetic, fluorescence *in situ* hybridisation, and clinical evaluation of translocations with concomitant deletion at 13q14 in chronic lymphocytic leukaemia. Genes Chromosomes Cancer 20 (1): 73-81.

Gardiner-Garden, M. and Frommer, M. (1987). CpG islands in the vertebrate genome. Mol. Biol. 196: 261-282.

Georges, M., Gunawardana, D.A., Threadgill, D.W., Lathrop, M., Olsaker, I., Mishra, A., Sargeant, L.L., Schoeberlein, A., Steele, M.R., Terry, C., Threadgill, D.S., Zhao, X., Holm, T., Fries, R. and Womack, J.E. (1991). Characterization of a set of variable number tandem repeat markers conserved in Bovidae. Genomics 11 (1): 24-32.

Gingrich, J.C., Boehrer, D.M., Garnes, J.A., Johnson, W., Wong, B.S., Bergmann, A., Eveleth, G.G., Lanlois, R.G. and Carrano, A.V. (1996). Construction and characterization of human chromosome 2-specific cosmid, fosmid, and PAC clone libraries. Genomics 32: 65-74.

Godard, S., Schibler, L., Oustry, A., Cribiu, E.P. and Guérin, G. (1998). Construction of a horse BAC library and cytogenetical assignment of 20 type I and type II markers. Mamm. Genome 9: 633-637.

Goodfellow, P.N. (1993) Microsatellites and the new genetic maps. Curr. Biol. 3 (3): 149-151.

Gosden, G.R., Mitchell, A.R., Buckland, R.A., Clayton, R.P. and Evans, H.J. (1975). The location of four human satellite DNAs on human chromosomes. Exp. Cell. Res. 92: 148-158.

Gosden, J. and Hanratty, D. (1993). PCR *in situ*: a rapid alternative to *in situ* hybridization for mapping short, low copy number sequences without isotopes. BioTechniques 15 (1): 78-80.

Gosden, J.R. and Lawson, D. (1994). Oligonucleotide primed *in situ* DNA synthesis (PRINS): an alternative to *in situ* hybridization for gene mapping and the investigation of genome organization. In J.R. Gosden (ed). Methods in molecular biology. Vol. 29: Chromosome analysis protocols. Humana Press, New Jersey.

Goss, S.J. and Harris, H. (1975). New method for mapping genes in human chromosomes. Nature 255 (5511): 680-684.

Goureau, A., Yerle, M., Schmitz, A., Riquet, J., Milan, D., Pinton, P., Frelat, G. and Gellin, J. (1996). Human and porcine correspondance of chromosome segments using bidirectional chromosome painting. Genomics 36: 252-262.

Grady, D.L., Ratliff, R.L., Robinson, D.L., McCanlies, E.C., Meyne, J. and Moyzis, R.K. (1992). Highly conserved repetitive DNA sequences are present at human centromeres. Proc. Natl. Acad. Sci. USA 89: 1695-1699.

Green, E.D. and Olson, M.V. (1990). Systematic screening of yeast artificial chromosome libraries by use of the polymerase chain reaction. Proc. Natl. Acad. Sci. USA 87: 1213-1217.

Green, E.D., Riethman, H.C., Dutchik, J.E. and Olson, M.V. (1991). Detection and characterization of chimeric yeast artificial-chromosome clones. Genomics 11 (3): 658-669.

Gregory, S.G., Howell, G.R. and Bentley, D.R. (1997). Genome mapping by fluorescent fingerprinting. Genome Res. 7 (12): 1162-1168.

Grobet, L., Martin, L.J., Poncelet, D., Pirottin, D., Brouwers, B., Riquet, J., Schoeberlein, A., Dunner, S., Menissier, F., Massabanda, J., Fries, R., Hanset, R. and Georges, M. (1997). A deletion in the bovine myostatin gene causes the double-muscled phenotype in cattle. Nat. Genet. 17 (1): 71-74.

Haldi, M., Perrot, V., Saumier, M., Desai, T., Cohen, D., Cherif, D., Ward, D. and Lander, E.S. (1994). Large human YACs constructed in a rad52 strain show a reduced rate of chimerism. Genomics 24 (3): 478-484.

Haley, J., Crawford, R., Hudson, P., Scanlon, D., Tregear, G., Shine, J. and Niall, H. (1987). Porcine relaxin. Gene structure and expression. J. Biol. Chem. 262 (25): 11940-11946.

Hanahan, D., Jessee, J. and Bloom, F.R. (1991). Plasmid transformation of *Escherichia coli* and other bacteria. Methods Enzymol. 204: 63-113.

Hanke, J.H., Hambor, J.E. and Kavathas, P. (1995). Repetitive Alu elements form a cruciform structure that regulates the function of the human CD8α T cell-specific enhancer. J. Mol. Biol. 246: 63-73.

Harris, H. and Watkins, J.F. (1965). Hybrid cells derived from mouse and man; artificial heterokaryons of mammalian cells from different species. Nature 205: 640-646.

Hassanane, M.S., Chaudhary, R. and Chowdhary, B.P. (1998). Microdissected bovine X chromosome segment delineates homoeologous chromosomal regions in sheep, goat and buffalo. Chromosome Res. 6 (3): 213-217

Hassett, A.L., Locker, J., Rupp, G., Kunz, H.W. and Gill, T.J. 3d (1989). Molecular analysis of the rat MHC. II. Isolation of genes that map to the RT1.E-grc region. J Immunol. 142 (6): 2089-96

Hawken, R.J., Broom, M.F., van Stijn, T, C., Lumsden, J.M., Broad, T.E. and Maddox, J.F. (1996). Mapping the ovine genes encoding IL3, IL4, IL5, and CSF2 to sheep chromosome 5q13-q15 by FISH. Mamm. Genome 7: 858-859.

Hawken, R., Alexander, L., Murtaugh, J., Flickinger, G., Beattie, C., Yerle, M., Pinton, P., Robic, A., Delcros, C., Milan, D., Gellin, J. and Schook, L. (1999). Construction and characterization of a porcine radiation hybrid panel. Plant and Animal Genome VII Conference Proceedings. San Diego, CA, USA.

Hayes, H., Petit, E. and Dutrillaux, B. (1991). Comparison of RBG-banded karyotypes of cattle, sheep, and goat. Cytogenet. Cell Genet. 57: 51-55.

Hayes, H. and Petit, E.J. (1993). Mapping of the β-lactoglobulin gene and of an immunoglobulin M heavy chain-like sequence to homoeologous cattle, sheep, and goat chromosomes. Mamm. Genome 4: 207-210.

Hayes, H., Petit, E., Bouniol, C. and Popescu, P. (1993). Localisation of the alpha S20 casein gene (CASAS2) to the homologous cattle, sheep and goat chromosomes 4 by *in situ* hybridization. Cytogenet. Cell Genet. 64: 281-285.

Hayes, H. (1995). Chromosome painting with human chromosome-specific DNA libraries reveals the extent and distribution of conserved segments in bovine chromosomes. Cytogenet. Cell Genet. 71: 168-174.

Heaton, M.P., Loez-Corrales, N.L., Smith, T.P.L., Kappes, S.M. and Beattie, C.W. (1997). Directed cosmid isolation of bovine markers for physical assignment by FISH. Animal Biotechnology 8 (2): 167-177.

Hediger, R., Ansari, H.A. and Stranzinger, G.F. (1991a). Chromosome banding and gene localizations support extensive conservation of chromosome structure between cattle and sheep. Cytogenet. Cell Genet. 57: 127-134.

Hediger, R., Johnson, S.E. and Hetzel, D.J.S. (1991b). Localisation of the beta-subunit of follicle stimulating hormone in cattle and sheep by *in situ* hybridization. Anim. Genet. 22: 237-244.

Heiskanen, M., Karhu, R., Hellsten, E., Peltonen, L., Kallioniemi, O.P. and Palotie, A. (1994). High resolution mapping using fluorescence *in situ* hybridization to extended DNA fibres from agarose-embedded cells. BioTechniques 17 (5): 928-933.

Hetzel, D.J.S. (1991). The use of reference families for genome mapping in domestic livestock. In: Schook, L.B., Lewin, H.A. and McLaren, D.G. (eds). Gene-mapping techniques and applications. Marcel Dekker, Inc., New York. pp. 51-64.

Hindkjær, J., Koch, J., Terkelsen, C., Brandt, C.A., Kϕlvraa, S. and Bolund, L. (1994). Fast, sensitive multicolor detection of nucleic acids *in situ* by PRimed *In situ* labeling (PRINS). Cytogenet. Cell Genet. 66: 152-154.

Horelli-Kuitunen, N., Aaltonen, J., Yaspo, M.L., Eeva, M., Wessman, M., Peltonen, L. and Palotie, A. (1999). Mapping ESTs by fiber-FISH. Genome. Res. 9: 62-71.

Huang, Z., Petty, J.T., O'Quinn, B., Longmire, J.L., Brown, N.C., Jett, J.H. and Keller, R.A. (1996). Large DNA fragment sizing by flow cytometry: application to the characterization of P1 artificial chromosome (PAC) clones. Nucleic Acids Res. 24 (21): 4202-4209.

Hubert, R.S., Mitchell, S., Chen, X.N., Ekmekji, K., Gadomski, C., Sun, Z., Noya, D., Kim, U.J., Chen, C., Shizuya, H., Simon, M., de Jong, P.J.and Korenberg, J.R. (1997). BAC and

PAC contigs covering 3.5 Mb of the Down syndrome congenital heart disease region between D21S55 and MX1 on chromosome 21. Genomics 41 (2): 218-226.

Iannuzzi, L. and Di Meo, G.P. (1995). Chromosomal evolution in bovids: a comparison of cattle, sheep and goat G- and R-banded chromosomes and cytogenetic divergences among cattle, goat and rive buffalo sex chromosomes. Chromosome Res. 3: 291-299.

Iannuzzi, L., Di Meo, G.P., Perucatti, A. and Incarnato, D. (1999). Comparison of the human with the sheep genomes by use of human chromosome-specific painting probes. Mamm. Genome 10: 719-723.

Ioannou, P.A., Amemiya, C.T., Garnes, J., Kroisel, P.M., Shizuya, H., Chen, C., Batzer, M.A. and De Jong, P.J. (1994). A new bacteriophage P1-derived vector for the propagation of large human DNA fragments. Nat. Genet. 6: 84-90.

ISCNDA (1990). The second international conference on standardization of domestic animal karyotypes. Di Berardino, D., Hayes, H., Fries, R., Long, S. (eds). Cytogenet. Cell Genet. 53: 65-78.

Ishiura, M., Ohashi, H., Hazumi, N., Uchida, T. and Okada, Y. (1989). Simplified cosmid vectors for gene transfer to cultured mammalian cells: isolation of the gene for elongation factor 2 from the mouse. Gene 85 (2): 427-433.

Jelinek, W.R. and Schmid, C.W. (1982). Repetitive sequences in eukaryotic DNA and their expression. Ann. Rev. Biochem. 51: 813-844.

Jessen, J.R., Meng, A., McFarlane, R.J., Paw, B.H., Zon, L.I., Smith, G.R. and Lin, S. (1998). Modification of bacterial artificial chromosomes through chi-stimulated homologous recombination and its application in zebrafish transgenesis. Proc. Natl. Acad. Sci. USA 95 (9): 5121-5126.

Jiang, Z., Priat, C. and Galibert, F. (1998). Traced orthologous amplified sequence tags (TOASTs) and mammalian comparative maps. Mamm. Genome 9: 577-587.

Jobse, C., Buntjer, J.B., Haggsma, N., Breukelman, H.J., Beintema, J.J. and Lenstra, J.A. (1995). Evolution and recombination of Bovine DNA repeats. J. Mol. Evol. 41: 277-283.

Kaftanovskaya, E.M. and Serov, O.L. (1994). High-resolution GTG-banded chromosomes of cattle, sheep and goat: a comparative study. J. Hered. 85: 395-400.

Kappes, S.M., Keele, J.W., Stone, R.T., McGraw, R.A., Sonstegard, T.S., Smith, T.P.L., Lopez-Corrales, N.L. and Beattie, C.W. (1997). A second-generation linkage map of the bovine genome. Genome Res. 7: 235-249.

Kasprzak, L. (1998). Comparative mapping of deer genomes. Honours Thesis. University of Adelaide. South Australia.

Kemp, S. J., Brezinsky, L. and Teale, A.J. (1993). A panel of bovine, ovine and caprine polymorphic microsatellites. Anim. Genet. 24: 363-365.

Kim, S.Y., Horrigan, S.K., Altenhofen, J.L., Arbieva, Z.H., Hoffman, R. and Westbrook, C.A. (1998). Modification of bacterial artificial chromosome clones using Cre recombinase: introduction of selectable markers for expression in eukaryotic cells. Genome Res. 8: 404-412.

Kim, U.J., Birren, B.W., Slepak, T., Mancino, V., Boysen, C., Kang, H.L., Simon, M.I. and Shizuya, H. (1996). Construction and characterization of a human bacterial artificial chromosome library. Genomics 34: 213-218.

Kioussis, D., Wilson, F., Daniels, C., Leveton, C., Taverne, J. and Playfair, J.H.L. (1987). Expression and rescuing of a cloned human tumour necrosis factor gene using an EBV-based shuttle cosmid vector. EMBO J. 6: 355-361.

Koch, J.E., Kϕlvraa, S., Peterson, K.B., Gregersen, N. and Bolund, L. (1989). Oligonucleotide-priming methods for the chromosome-specific labelling of alpha satellite DNA *in situ*. Chromosoma 98: 259-265.

Koch, J., Mogensen, J., Pedersen, S., Fischer, H., Hindkjær, J., Kϕlvraa, S. and Bolund, L. (1992). Fast one-step procedure for the detection of nucleic acids *in situ* by primer-induced sequence-specific labeling with fluorescein-12-dUTP. Cytogenet. Cell Genet. 60: 1-3.

Koch, J., Fischer, H., Askholm, H., Kindkjær, J., Pedersen, S., Kϕlvraa, S. and Bolund, L. (1993). Identification of a supernumerary der(18) chromosome by a rational strategy for the cytogenetic typing of small marker chromosomes with chromosome-specific DNA probes. Clin. Genet. 43: 200-203.

Kojima, M., Ohata, K. and Yasue, H. (1997). Structural organization and chromosomal assignment of the swine endothelin-1 gene. J Vet Med Sci. 59(6): 431-5.

Korenburg, J.R. and Rykowski, M.C. (1988). Human genome organization: Alu, Lines, and the molecular structure of metaphase chromosome bands. Cell 53: 391-400.

Kornblihtt, A.R., Vibe-Pedesen, K. and Baralle, F.E. (1983). Isolation and characterization of cDNA clones for human and bovine fibronectins. Proc. Natl. Acad. Sci. USA 80: 3218-3222.

Kotani, H., Sato, S., Fukami, M., Hosouchi, T., Bakazaki, N., Okumura, S., Wada, T., Liu, Y.G., Shibata, D. and Tabata, S. (1997). A fine physical map of *Arabidopsis thaliana* chromosome 5: construction of a sequence-ready contig map. DNA Res. 4 (6): 371-378.

Kurdi-Haidar, B., Levine, F., Roemer, K., LaPorte, P. and Friedmann, T. (1993). Provirus-anchored long range mapping of mammalian genomes. Genomics 15: 305-310.

Kusumi, K., Smith, J.S., Segre, J.A., Koos, D.S. and Lander, E.S. (1993). Construction of a large-insert yeast artificial chromosome library of the mouse genome. Mamm. Genome 4 (7): 391-392.

Larin, Z., Monaco, A.P., Meier-Ewert, S. and Lehrach, H. (1993). Construction and characterization of yeast artifical chromosome libraries from the mouse genome. Methods Enzymol. 225: 623-633.

Larionov, V., Kouprina, N., Nikolaishvili, N. and Resnick, M.A. (1994). Recombination during transformation as a source of chimeric mammalian artificial chromosomes in yeast (YACs). Nucleic Acids Res. 22 (20): 4154-61.

Larsen, N.J., Hayes, H., Bishop, M., Davis, S.K., Taylor, J.F. and Kirkpatrick, B.W. (1999). A comparative linkage and physical map of bovine chromosome 24 with human chromosome 18. Mamm. Genome 10 (5): 482-487.

Latt, S.A., Willard, H.F. and Gerald, P.S. (1976). Brdu-33258 Hoeschst analsis of DNA replication in human lymphocytes with supernumerary or structurally abnormal X chromosomes. Chromosoma 57: 135-153.

Lawce, H.J. and Brown, M.G. (1991). Harvesting, slide-making, and chromosome elongation techniques. Ch. 3. In: Barch, M.J. (ed). The ACT Cytogenetics Laboratory manual. Second edition. Raven Press, New York. pp. 31-105.

Lehrman, M.A., Goldstein, J., Russell, D.W. and Brown, M.S. (1987). Duplication of seven exons in LDL receptor gene caused by Alu-Alu recombination in a subject with familial hypercholesterolemia. Cell 48: 827-835.

Lemieux, N., Dutrillaux, B. and Viegas-Péquignot, E. (1992). A simple method for simultaneous R- or G-banding and fluorescence *in situ* hybridization of small single-copy genes. Cytogenet. Cell Genet. 59: 311-312.

Lenstra, J.A., van Boxtel, J.A.F., Zwaagstra, K.A. and Schwerin, M. (1993). Short interspersed nuclear element (SINE) sequences of the *Bovidae*. Anim. Genet. 24:33-39.

Levinson, G. and Gutman, C.A. (1987). Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Mol. Biol. Evol. 4 (3): 203-221.

Li, R., Mignot, E., Faraco, J., Kadotani, H., Cantanese, J., Zhao, B., Lin, X., Hinton, L., Ostrander, E.A., Patterson, D.F. and de Jong, P.J. (1999). Construction and characterization of an eightfold redundant dog genomic bacterial artificial chromosome library. Genomics 58 (1): 9-17.

Libert, F., Lefort, A., Okimoto, R., Womack, J. and Georges, M. (1993). Construction of a bovine genomic library of large yeast artificial chromosome clones. Genomics 18: 270-276.

Lichter, P., Tang, C.J., Call, K., Hermanson, G., Evans, G.A., Housman, D. and Ward, D.C. (1990). High-resolution mapping of human chromosome 11 by *in situ* hybridization using cosmid clones. Science 247 (4938). 64-69.

Lien, L.L., Lee, Y. and Orkin, S.H. (1997). Regulation of the myeloid-cell-expressed human gp91-phox gene as studied by transfer of yeast artificial chromosome clones into embryonic stem cells: suppression of a variegated cellular pattern of expression requires a full complement of distant *cis* elements. Mol. Cell Biol. 17 (4): 2279-2290.

Lin, C.C., Newton, D.R., Smink, W.K. and Church, R.B. (1976). A rapid and simple method for the isolation and culture of leukocytes for chromosome analysis in domestic animals. Can. J. Anim. Sci. 56: 27-31.

Lin, C.C., Newton, D.R. and Church, R.B. (1977). Identification and nomenclature for G-banded bovine chromosomes. Can. J. Genet. Cytol. 19: 271-282.

Ling, L.L., Ma, N.S., Smith, D.R., Miller, D.D. and Moir, D.T. (1993). Reduced occurrence of chimeric YACs in recombination-deficient hosts. Nucleic Acids Res. 21 (25): 6045-6046.

Linton, M.F., Farese, R.V., Chiesa, G., Grass, D.S., Chin, P., Hammer, R.E., Hobbs, H.H. and Young, S.G. (1993). Transgenic mice expressing high plasma concentrations of human apolipoprotein B100 and lipoprotein(a). J. Clin. Invest. 92: 3029-3037.

Littlefield, J.W. (1964). Selection of hybrids from matings of fibroblasts in bitro and their presumed recombinants. Science 145: 709-710.

Liu, W.S., Harbitz, I., Gustavsson, I. and Chowdhary, B.P. (1998). Mapping of the porcine erythropoietin gene to chromsome 3p16-p15 and ordering of four related subclones by fiber-FISH and DNA-combing. Hereditas 128: 77-81.

Logue, D.N. and Harvey, M.J.A. (1978). Meiosis and spermatogenesis in bulls heterozygous for a presumptive 1/29 Robertsonian translocation. J. Reprod. Fertil. 54: 159-165.

Lyons, L.A., Laughlin, T.F., Copeland, N.G., Jenkins, N.A., Womack, J.E. and O'Brien, S.J. (1997). Comparative anchor tagged sequences (CATS) for integrative mapping of mammalian genomes. Nat. Genet. 15: 47-56.

MacKinnon, P.J., Powell, B.C. and Rogers, G.E. (1990). Structure and expression of genes for a class of cysteine-rich proteins of the cuticle layers of differentiating wool and hair follicles. J. Cell. Biol. 111: 2587-2600.

Maddox, J.F. and Matthew, P. (1994a). A polymorphic microsatellite detected in the ovine CD5 gene. Anim. Genet. 25: 371.

Maddox, J.F. and Matthew, P. (1994b). A polymorphic dinucleotide repeat microsatellite detected in the ovine interleukin 2 receptor alpha. Anim. Genet. 25: 200.

Maddox, J.F., Davies, K.P., Cuthbertson, R., Hulme, D.J., Henry, H., Hawken, R.J., Drinkwater, R., Dodds, K.G. and Crawford, A.M. (1996). Updating the sheep linkage map. Proceedings of the XXVth International Society of Animal Genetics conference, Tours, France.

Mahairas, G.G., Wallace, J.C., Smith, K., Swatzell, S., Holzman, T., Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and Hood, L. (1999). Sequence-tagged connectors: a sequence approach to mapping and scanning the human genome. Proc. Natl. Acad. Sci. USA. 96 (17): 9739-9744.

Mandel, J. (1993). Questions of expansion. Nat. Genet. 4: 8-9.

Marra, M.A., Kucaba, T.A., Dietrich, N.L., Green, E.D., Brownstein, B., Wilson, R.K., McDonald, K.M., Hillier, L.W., McPherson, J.D. and Waterson, R.H. (1997). High throughput fingerprint analysis of large-insert clones. Genome Res. 7: 1072-1084.

Marek, L.F. and Shoemaker, R.C. (1997). BAC contig development by fingerprint analysis in soybean. Genome 40: 420-427.

Markie, D., Ragoussis, J., Senger, G., Rowan, A., Sansom, D., Trowsdale, J., Sheer, D. and Bodmer, W.F. (1993). New vector for transfer of yeast artificial chromosomes to mammalian cells. Somat. Cell Mol. Genet. 19 (2): 161-169.

Matsumoto, N., Soeda, E., Ohashi, H., Fujimoto, M., Kato, R., Tsujita, T., Tomita, H., Kondo, S., Fukushima, Y. and Niikawa, N. (1997). A 1.2-megabase BAC/PAC contig spanning the 14q13 breakpoint of t(2; 14) in a mirror-image polydactyly patient. Genomics 45 (1): 11-6.

McCarthy, L.C., Terrett, J., Davis, M.E., Knights, C.J., Smith, A.L., Critcher, R., Schmitt, K., Hudson, J., Spurr, N.K. and Goodfellow, P.N. (1997). A first-generation whole genome-radiation hybrid map spanning the mouse genome. Genome. Res. 7 (12): 1153-1161.

McCormick, S.P., Ng, J.K., Taylor, S., Flynn, L.M., Hammer, R.E. and Young, S.G. (1995). Mutagenesis of the human apolipoprotein B gene in a yeast artificial chromosome

reveals the site of attachment for apolipoprotein(a). Proc. Natl. Acad. Sci. USA 92 (22): 10147-51.

McLaren, R.J., Rogers, G.R., Davies, K.P., Maddox, J.F. and Montgomery, G.W. (1997). Linkage mapping of wool keratin and kerain-associated protein genes in sheep. Mamm. Genome 8: 938-940.

Mejía, J.E. and Monaco, A.P. (1997). Retrofitting vectors for *Escherichia coli*-based artificial chromosomes (PACs and BACs) with markers for transfection studies. Genome Res. 7: 179-186.

Messer, L.A., Wang, L., Tuggle, C.K., Yerle, M., Chardon, P., Pomp, D., Womack, J.E., Barendse, W., Crawford, A.M., Notter, D.R. and Rothschild, M.F. (1997). Mapping of the melatonin receptor 1a (MTNR1A) gene in pigs, sheep and cattle. Mamm. Genome 8: 368-370.

Messerle, M., Crnkovic, I., Hammerschmidt, W., Ziegler, H. and Koszinowski, U.H. (1997). Cloning and mutagenesis of a herpesvirus genome as an infectious bacterial artificial chromosome. Proc. Natl. Acad .Sci. USA 94 (26): 14759-14763.

Mezzalani, A., Zhang, Y., Redaelli, L., Castiglioni, B., Leone, P., Williams, J.L., Solinas Toldo, S., Wigger, G., Fries, R. and Ferretti, L. (1995). Chromosomal localization and molecular characterization of 53 cosmid-derived bovine microsatellites. Mamm. Genome 6: 629-635.

Modi, W.S., Gallagher, D.S. and Womack, J.E. (1996). Evolutionary histories of highly repeated DNA families among the artiodactyla (mammalia). J. Mol. Evol. 42 (3): 337-349.

Moffat, A.S. (1998). Improving gene transfer into livestock. Science 282 (5394): 1619-1620.

Moir, D.T. and Smith, D.R. (1998). Strategies for large-insert cloning and analysis. Ch. 5 In: Dracopoli, N.C., Haines, J.L., Korf, B.R., Moir, D.T., Morton, C.C., Seidman, C.E., Seidman, J.G. and Smith, D.R. (eds). Current Protocols in Human Genetics. 5.0.1-5.0.13. John Wiley & Sons, Inc. New York, USA.

Montgomery, G.W. and Crawford, A.M. The sheep linkage map. In: Piper, L. and Ruvinsky, A. (eds) "The genetics of sheep". CAB International, pp. 297-352.

Muller, M. and Brem, G. (1998). Transgenic approaches to the increase of disease resistance in farm animals. Rev. Sci. Tech. 17 (1): 365-378.

Murphy, G.L., Connell, T.D., Barritt, D.S., Koomey, M. and Cannon, J.G. (1989). Phase variation of gonococcal protein II: regulation of gene expression by slipped-strand mispairing of a repetitive DNA sequence. Cell 56: 539-547.

Nadeau, J.H., Berger, F.G., Cox, D.R., Crosby, J.L., Davisson, M.T., Ferrara, D., Fuchs, E., Hart, C., Hunihan, L., Lalley, P.A., Langley, S.H., Martin, G.R., Bichols, L., Phillips, S.J., Roderick, T.H., Roop, D.R., Ruddle, F.H., Skow, L.C. and Compton, J.G. (1989). A family

of type I keratin genes and the homeobox-2 gene complex are closely linked to the Rex locus on mouse chromosome 11. Genomics 5: 454-462.

Narayanan, K., Williamson, R., Zhang, Y., Stewart, A.F. and Ioannou P.A. (1999). Efficient and precise engineering of a 200 kb beta-globin human/bacterial artificial chromosome in E. coli DH10B using an inducible homologous recombination system. Gene Ther. 6 (3): 442-447.

Nguyen, C., Pontarotti, P., Birnbaum, D., Chimini, G., Rey, J.A., Mattei, J.F. and Jordan, B.R. (1987). Large scale physical mapping in the q27 region of the human X chromosome: the coagulation factor IX gene and the mcf.2 transforming sequence are separated by at most 270 kilobase pairs and are surround by several 'HTF islands'. EMBO 6 (11): 3285-9.

Nielsen, L.B., McCormick, S.P.A., Pierotti, v., Tam, C., Gunn, M.D., Shizuya, H. and Young, S.G. (1997). Human apoliprotein B transgenic mice generated with 207- and 145-kilobase bacterial artificial chromosomes. J. Biol. Chem. 272 (47): 29752-29758.

Nielsen, L.B., McCormick, S.P.A. and Young, S.G. (1999). A new approach for studying gene regulation by distant DNA elements in transgenic mice. Scand. J. Clin. Lab. Invest. 59, Suppl 229: 33-40.

Novak, U. (1984). Structure and properties of a highly repetitive DNA sequence in sheep. Nucleic Acids Res. 12 (5):2343-2350.

O'Brien, S.J., Seuánez, H.N. and Womack, J.E. (1988). Mammalain genome organization: an evolutionary view. Annu. Rev. Genet. 22: 323-351.

O'Brien, S.J., Womack, J.E., Lyons, L.A., Moore, K.J., Jenkins, N.A. and Copeland, N.G. (1993). Anchored reference loci for comparative genome mapping in mammals. Nat. Genet. 3: 103-112.

Ohno, S. (1973). Ancient linkage groups and frozen accidents. Nature 244: 259-262.

Okazaki, Y. and Hayashizaki, Y. (1997). High-speed positional cloning based on restriction landmark genome scanning. Methods 13 (4): 359-377.

Oostra, B.A. and Verkerk,.A.J.M.H. (1992). The fragile X syndrome: isolation of the FMR-1 gene and characterization of the fragile X mutation. Chromosoma 101: 381-387.

Orti, R., Megarbane, A., Maunoury, C., Van Broeckhoven, C., Sinet, P.M. and Delabar J.M. (1997). High-resolution physical mapping of a 6.7-Mb YAC contig spanning a region critical for the monosomy 21 phenotype in 21q21.3-q22.1. Genomics 43 (1): 25-33.

Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tateno, M., Catanese, J.J. and de Jong, P.J. (1998). An improved approach for construction of bacterial artificial chromosome libraries. Genomics 52 (1):1-8.

Panayiotidis, P., Ganeshaguru, K., Rowntree, C., Jabbar, S.A., Hoffbrand, V.A. and Foroni, L. (1997). Lack of clonal BCRA2 gene deletion on chromosome 13 in chronic lymphocytic leukaemia. Br. J. Haematol. 97 (4): 844-847.

Papp, J., Csokay, B., Bosze, P., Zalay, Z., Toth, J., Ponder, B. and Olah, E. (1996). Allele loss from large regions of chromosome 17 is common only in certain histological subtypes of ovarian carcinomas. Br. J. Cancer 74 (10): 1592-1597.

Pardue, M.L. and Gall, J.G. (1969). Molecular hybridization of radioactive DNA to the DNA of cytological preparations. Proc. Natl. Acad. Sci. USA 64 (2): 600-4.

Pardue, M.L. and Gall, J.G. (1970). Chromosomal localization of mouse satellite DNA. Science 168: 1356-1358.

Parsons, Y.M., Cooper, D.W. and Piper, L.R. (1994). Evidence of linage between high-glycine-tyrosine keratin gene loci and wool fibre diameter in a Merino half-sib family. Anim. Genet. 25: 105-108.

Parsons, Y.M. and Cooper, D.W. (1995). Linkage of wool proteins in the sheep genome. Proc. Aust. Assoc. Anim. Breed. Genet. 11: 469-472.

Payen, E.J. and Cotinot, C.Y. (1994). Sequence evolution of SRY gene within *Bovidae* family. Mamm. Genome 5: 723-725.

Pearce, M., Blake, D.J., Tinsley, J.M., Byth, B.C., Campbell, L., Monaco, A.P. and Davies, K.E. (1993). The utrophin and dystrophin genes share similarities in genomic structure. Hum. Mol. Genet. 2 (11): 1765-72.

Pearce, P.D., Ansari, H.A., Maher, D.W., Broad, T.E., Cambridge, L.M., Lewis, P.E., Burkin, D.J. and Jones, C.A. (1994). The assignment of fourteen new loci to ovine chromosomes. Proceedings of the 5th Australasian Gene Mapping Workshop, Armidale, Australia.

Pellestor, F., Girardet, A., Lefort, G., Andréo, B. and Charlieu, J.P. (1995). PRINS as a method for rapid chromosomal labeling on human spermatozoa. Mol. Reprod. Dev. 40: 333-337.

Peterson, K.R., Clegg, C.H., Huxley, C., Josephson, B.M., Haugen, H.S., Furukawa, T. and Stamatoyannopoulos, G. (1993). Transgenic mice containing a 248kb yeast artificial chromosome carrying the human β-globin locus display proper developmental control of human globin genes. Proc. Natl. Acad. Sci. USA 90: 7593-7597.

Peterson, K.R., Li, Q.-L., Clegg, C.H., Furukawa, T., Navas, P.A., Norton, E.J., Kimbrough, T.G. and Stamatoyannopoulos, G. (1995). Use of yeast artificial chromosomes (YACs) in studies of mammalian development: production of β-globin locus YAC mice carrying human globin developmental mutants. Proc. Natl. Acad. Sci. USA 92 (12): 5655-5659.

Phelan, C.M., Liu, L., Ruttledge, M.H., Muntzing, K., Ridderheim, P.A. and Collins, V.P. (1995). Chromosome 17 abnormalities and lack of TP53 mutations in paediatric central nervous system tumours. Hum. Genet. 96 (6): 684-690.

Pinkel, D., Landegent, J., Collins, C., Fuscoe, J., Segraves, R., Lucas, J. and Gray J (1988). Fluorescence *in situ* hybridization with human chromosome-specific libraries: detection of trisomy 21 and translocations of chromosome 4. Proc. Natl. Acad. Sci. USA 85 (23): 9138-9142.

Pirottin, D., Poncelet, D., Grobet, L., Royo, L.J., Brouwers, B., Masabanda, J., Takeda, H., Fries, R., Sugimoto, Y., Womack, J.E., Dunner, S. and Georges, M. (1999). High-resolution, human-bovine comparative mapping based on a closed YAC contig spanning the bovine *mh* locus. Mamm. Genome 10: 289-293.

Pitel, F., Cribiu, E.P., Yerle, M., Lahbib-Mansais, Y., Lanneluc, I., Lantier, F. and Gellin, J. (1995). Regional localization of the NRAMP gene to chromosome 2q41-42 by *in situ* hybridization. Cytogenet. Cell Genet. 70: 116-118.

Ponce de León, F.A., Ambady, S., Hawkins, G.A., Kappes, S.M., Bishop, M.D., Robl, J.M. and Beattie, C.W. (1996). Development of a bovine X chromosome linkage group and painting probes to assess cattle, sheep, and goat X chromosome segment homologies. Proc. Natl. Acad. Sci. USA 93 (8): 3450-3454.

Pontecorvo, G. (1975). Production of mammalian somatic cell hybrids by means of polyethylene glycol treatment. Somatic Cell Genetics 1: 397-400.

Popescu, C.P., Long, S., Riggs, P., Womack, J., Schmutz, S., Fries, R. and Gallagher, D.S. (1996). Standardization of cattle karyotype nomenclature: report of the committee for the standardization of cattle karyotype. Cytogenet. Cell Genet. 74: 259-261.

Porcu, S., Kitamura, M., Witkowska, E., Zhang, Z., Mutero, A., Lin, C., Chang, J. and Gaensler, K.M.L. (1997). The human β globin locus introduced by YAC transfer exhibits a specific and reproducible pattern of developmental regulation in transgenic mice. Blood 90 (11): 4602-4609.

Poustka, A., Rackwitz, H.-R., Frischauf, A.-M., Hohn, B. and Lehrach, H. (1984). Selective isolation of cosmid clones by homologous recombination in *Escherichia coli*. Proc. Natl. Acad. Sci. USA 81: 4129-4133.

Powell, B.C., Sleigh, M.J., Ward, K.A. and Rogers, G.E. (1983). Mammalian keratin gene families: organization of genes coding for the B2 high-sulphur proteins of sheep wool. Nucleic Acids Res. 11: 5327-5346.

Powell, B., Nesci, A. and Rogers, G. (1991). Regulation of keratin gene expression in the hair follicle. Ann. N.Y. Acad. Sci. 642: 1-20.

Powell, B.C. and Rogers, G.E. (1994). Differentiation in hard kertain tissues: hair and related structures. In: Leigh, F., Watt, R.F. and Lane, B. (eds). *Keratinocyte Handbook*. Cambridge University Press, Cambridge. pp.401-436.

Powell, B.C., Walker, S.K., Bawden, C.S., Sivaprasad, A.V. and Rogers, G.E. (1994). Transgenic sheep and wool growth: possibilities and current status. Reprod. Fertil. Dev. 6: 615-623.

Purandare, S.M. and Patel, P.I. (1997). Recombination hot spots and human disease. Genome Res. 7: 773-786.

Ramsay, G. (1998). DNA chips: state-of-the art. Nat. Biotechnol. 16 (1): 40-44.

Reading conference (1980). Proceedings of the first international conference for the standardisation of banded karyotypes of domestic animals. Hereditas 92: 145-162.

Rettenberger, G., Klett, C., Zechner, U., Kunz, J., Vogel, W. and Hameister, H. (1995). Visualization of the conservation of synteny between humans and pigs by heterologous chromosomal painting. Genomics 26: 372-378.

Richards, R.I. and Sutherland, G.R. (1992). Dynamic mutations: a new class of mutations causing human disease. Cell 70: 709-712.

Richards, R.I. and Sutherland, G.R. (1994). Simple repeat DNA is not replicated simply. Nat. Genet. 6: 114-116.

Rijnkels, M., Kooiman, P.M., Platenburg, G.J., van Dixhoorn, M., Nuijens, J.H., de Boer, H.A. and Pieper, F.R. (1998). High-level expression of bovine alpha s1-casein in milk of transgenic mice. Transgenic Res. 7(1): 5-14.

Riley, J.H., Morten, J.E. and Anand, R. (1992). Targeted integration of neomycin into yeast artificial chromosomes (YACs) for transfection into mammalian cells. Nucleic Acids Res. 20 (12): 2971-6.

Roach, J.C. (1995). Random subcloning. Genome Res. 5 (5): 464-473.

Robic, A., Milan, D., Woloszyn, N., Riquet, J., Yerle, M., Nagel, M., Bonnet, M., Pinton, P., Dalens, M. and Gellin, J. (1997). Contribution to the physically anchored linkage map of the pig. Anim. Genet. 28 (2): 94-102.

Rogel-Gaillard, C., Bourgeaux, N., Save, J.-C., Renard, C., Coullin, P., Pinton, P., Yerle, M., Vaiman, M. and Chardon, P. (1997). Construction of a swine YAC library allowing an efficient recovery of unique and centromeric repeated sequences. Mamm. Genome 8: 186-192.

Rogers, G.R. (1994). Identification of a quantitative trait locus for wool strength in romney sheep (Lincoln: Lincoln University).

Rogers, M.A., Winter, H., Wolf, C., Heck, M. and Schweizer, J. (1998). Characterization of a 190-kilobase pair domain of human type I hair keratin gene. J. Biol. Chem. 273 (41): 26683-26691.

Romano, V., Bosco, P., Rocchi, M., Costa, G., Leube, R.E., Franke, W.W. and Romeo, G. (1988). Chromosomal assignments of human type I and type II cytokeratin genes to different chromosomes. Cytogenet. Cell Genet. 48: 148-151.

Rosen, J.M., Li, S., Raught, B. and Hadsell, D. (1996). The mammary gland as a bioreactor: factors regulating the efficient expression of milk pretein-based transgenes. Am. J. Clin. Nutr. 63 (4): 627S-632S.

Rosenberg, M., Chaudhury, A., Shows, T.B., Le Beau, M.M. and Fuchs, E. (1988). A group of type I keratin genes on human chromosome 17: characterization and expression. Mol. Cell Biol. 8 (2): 722-736.

Royal, A., Garapin, A., Cami, B., Perrin, F., Mandel, J.L., LeMeur, M., Brégégègre, F., Gannon, F., LePenec, J.P., Chambon, P. and Kourilsky, P. (1979). The ovalbumin gene region: Common features in the organisation of three genes expressed in chicken oviduct under hormonal control. Nature 279: 125-132.

Ryan, A.M., Gallagher, D.S. and Womack, J.E. (1993). Somatic cell mapping of omega and trophoblast interferon genes to bovine syntenic group U18 and *in situ* localization to chromosome 8. Cytogenet. Cell Genet. 63: 6-10.

Saidi-Mehtar, N., Hors-Cayla, M.-C. and Van Cong, N. (1981). Sheep gene mapping by somatic cell hybridization: four syntenic groups: ENO1-PGD, ME1-PGM3, LDHB-PEPB-TPI and G6PD-PGK-GALA. Cytogenet. Cell Genet. 30: 193-204.

Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, B. and Erlich, H.A. (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239: 487-491.

Sakamoto, K., Fordis, C.M., Corsico, C.D., Howard, T.H. and Howard, B.H. (1991). Modulation of HeLa cell growth by transfected 7SL RNA and Alu gene sequences. J. Biol. Chem. 266: 3031-3038.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989). Molecular cloning: a laboratory manual. 2nd edition. Cold Spring Habor Laboratory Press.

Sapolsky, R.J., Hsie, L., Berno, A., Ghandour, G., Mittmann, M. and Fan, J.B. (1999). High-throughput polymorphism screening and genotyping with high-density oligonucleotide arrays. Genet. Anal. 14: 187-192.

Sarmiento, U.M., Sarmiento, J.I., Lunney, J.K. and Rishi, S.S. (1993). Mapping of the porcine alpha interferon (IFNA) gene to chromosome 1 by fluorescence *in situ* hybridization. Mamm. Genome 4: 62-63.

Scherthan, H., Cremer, T., Arnason, U., Weier, H.-U., Lima-de-Faria, A. and Frönicke, L. (1994). Comparative chromosome painting discloses homologous segments in distantly related mammals. Nature Genet. 6: 342-347.

Schibler, L., Vaiman, D., Oustry, A., Guinec, N., Dangy-Caye, A.-L., Billault, A. and Cribiu. E.P. (1998a). Construction and extensive characterization of a goat bacterial

artificial chromosome library with threefold genome coverage. Mamm. Genome 9: 119-124.

Schibler, L., Vaiman, D., Oustry, A., Giraud-Delville, C. and Cribiu, E.P. (1998b). Comparative gene mapping: a fine-scale survey of chromosome rearrangements between ruminants and humans. Genome Res. 8: 901-915.

Schläpfer, J., Yang, Y., Rexroad, C. and Womack, J.E. (1997). A radiation hybrid framework map of bovine chromosome 13. Chromosome Res. 5 (8): 511-519.

Schläpfer, J., Gallagher, D.S., Burzlaff, J.D., Womack, J.E., Stelly, D.M., Taylor, J.F. and Davis, S.K. (1998). Comparative mapping of bvoine chromsomes 13 by fluorescence *in situ* hybridization. Anim. Genet. 29 (4): 265-272.

Schmitz, A., Oustry, A., Vaiman, D., Chaput, B., Frelat, G. and Cribiu, E.P. (1998). Comparative karyotype of pig and cattle using whole chromosome painting probes. Hereditas 128: 257-263.

Schmutz, S.M., Moker, J.S., Gallagher, D.S., Kappes, S.M. and Womack, J.E. (1996). *In situ* hybridization mapping of LDHA and IGF2 to cattle chromosome 29. Mamm. Genome 7: 473.

Schollmeyer, E., Schafer, D., Frisch, B. and Schlerermacher, E. (1981). High resolution analysis and differential condensation in RBA-banded human chromosomes. Hum. Genet. 59: 187-193.

Schriml, L.M., Padilla-Nash, H.M., Coleman, A., Moen, P., Nash, W.G., Menninger, J., Jones, G., Ried, T. and Dean, M. (1999). Tyramide signal amplification (TSA)-FISH applied to mapping PCR-labeled probes less than 1 kb in size. Biotechniques 27 (3): 608-613.

Schuler, G.D., Boguski, M.S., Stewart, E.A., Stein, L.D., Gyapay, G., Rice, K., White, R.E., Rodriguez-Tome, P., Aggarwal, A., Bajorek, E., Bentolial, S., Birren, B.B., Butler, A., Castle, A.B., Chiannilkulchai, B., Chu, A., Clee, C., Cowles, S., Day, P. J., Dibling, T., Drouot, N., Dunham, I., Duprat, S., East, C., Edwards, C., Fan, J.-B., Fang, N., Fizames, C., Garrett, C., Green, L., Hadley, D., Harris, M., Harrison, P., Brady, S., Hicks, A., Holloway, E., Hui, L., Hussain, S., Louis-Dit-Sully, C., Ma, J., MacGilvery, A., Mader, C., Maratukulam, A., Matise, T.C., McKusick, K.B., Morissette, J., Mungall, A., Muselet, D., Nusbaum, H.C., Page, D.C., Peck, A., Perkins, S., Piercy, M., Qin, F., Quackenbush, J., Ranby, S., Reif, T., Rozen, S., Sanders, C., She, X., Silva, J., Slonim, D.K., Soderlund, C., Sun, W.-L., Tabar, P., Thangarajah, T., Vega-Czarny, N., Vollrath, D., Voyticky, S., Wilmer, T., Wu, X., Adams, M.D., Auffray, C., Walter, n.A.R., Brandon, R., Dehejia, A., Nagase, T., Ishikawa, K., Nomura, N., Phillips, C., Polymeropoulos, M.H., Sandusky, M., Schmitt, K., Berry, R., Swanson, K., Torres, R., Venter, J.C., Sikela, J.M., Beckmann, J.S., Weissenbach, J., Myers, R.M., Cox, D.R., James, M.R., Bentley, D., Deloukas, P., Lander, E.S. and Hudson, T.J. (1996). A gene map of the human genome. Science 274: 540-546.

Sehested, J. (1974). A simple method of R-banding of human chromosomes, showing a pH dependent connection between R- and G-bands. Human-genetik 21: 55-58.

Serikawa, T., Kuramoto, T., Hilbert, P., Mori, M., Yamada, J., Dubay, C.J., Lindpainter, K., Ganten, D., Guénet, J.-L., Lathrop, G.M. and Beckman, J.S. (1992). Rat gene mapping using PCR-analysed microsatellites. Genetics 131: 701-721.

Sheng, Y., Mancino, V. and Birren, B. (1995). Transformation of *Escherichia coli* with large DNA molecules by electroporation. Nucleic Acids Res. 23 (11): 1990-1996.

Shizuya, H., Birren, B., Kim, U.-J., Mancino, V., Slepak, T., Tachiiri, T. and Simon, M. (1992). Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. Proc. Natl. Acad. Sci. USA 89: 8794-8797.

Shuxian, L., Harris, C.P. and Leong, S.A. (1993). Comparison of fluorescence *in situ* hybridization and primed *in situ* labeling methods for detection of single-copy genes in the fungus *Ustilago maydis*. Experimental Mycology 17: 301-308.

Siegel, A.F., Trask, B., Roach, J.C., Mahairas, G.G., Hood, L. and van den Engh, G. (1999). Analysis of sequence-tagged-connector strategies for DNA sequencing. Genome. Res. 9 (3): 297-307.

Silverman, G.A., Ye, R.D., Pollock, K.M., Sadler, J.E. and Korsmeyer, S.J. (1989). Use of yeast artificial chromosome clones for mapping and walking within human chromosome segment 18q21.3. Proc. Natl. Acad. Sci. USA 86: 7485-7489.

Simi, L.B., Sasi, R., Lingrel, J.B. and Lin, C.C. (1989). Mapping of the goat beta-globin gene cluster to a region of chromosome 7 by *in situ* hybridization. J. Hered. 80 (3): 246-249.

Sinnett, D., Richer, C. and Baccichet, A. (1998). Isolation of stable bacterial artificial chromosome DNA using a modified alkaline lysis method. Biotechniques 24 (5): 752-754.

Smith, D., Smyth, A.P. and Moir, D.T. (1990). Amplification of large artificial chromosomes. Proc. Natl. Acad. Sci. USA 87: 8242-8246.

Smith, T.P.L., Alexander, L.J., Sonstegard, T.S., Yoo, J., Beattie, C.W. and Broom, M.F. (1996). Construction and characterization of a large insert bovine YAC library with five-fold genomic coverage. Mamm. Genome 7: 155-156.

Soderlund, C., Longden, I. and Mott, R. (1997). FPC: a system for building contigs from restriction fingerprinted clones. Comput. Appl. Biosci. 13 (5): 523-535.

Solinas-Toldo, S., Lengauer, C. and Fries, R. (1995). Comparative genome map of human and cattle. Genomics 27: 489-496.

Sonstegard, T.S., Lopez-Corrales, N.L., Kappes, S.M., Beattie, C.W. and Smith, T.P. (1997). Comparative mapping of human chromosome 2 identifies segments of conserved synteny near the bovine mh locus. Mamm. Genome 8 (10): 751-755.

Sonstegard, T.S., Kappes, S.M., Keele, J.W. and Smith T.P. (1998). Refinement of bovine chromosome 2 linkage map near the mh locus reveals rearrangements between the bovine and human genomes. Anim. Genet. 29(5):341-7.

Southern, E.M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98: 503-517.

Stillwell, D.R., Davis, S.K., Taylor, J.F. and Davis, S.L.F. (1999). Genetic and physical mapping of the POLL locus to bovine chromosome 1 in a *Bos indicus* x *Bos taurus* cross. Proceedings of the Plant and Animal Genome VII conference, San Diego, USA.

Stinnakre, M.-G., Soulier, S., Schibler, L., Lepourry, L., Mercier, J.-C. and Vilotte, J.-L. (1999). Position-independent and copy-number-related expression of a goat bacterial artificial chromosome α-lactalbumin gene in transgenic mice. Biochem. J. 339: 33-36.

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

Strausberg, R.L., Feingold, E.A., Klausner, R.D. and Collins, F.S. (1999). The mammalian gene collection. Science 386: 455-457.

Strong, S.J., Ohta, Y., Litman, G.W. and Amemiya, C.T. (1997). Marked improvement of PAC and BAC cloning is achieved using electroelution of pulsed-field gel-separated partial digests of genomic DNA. Nucleic Acids Res. 25 (19): 3959-3961.

Sun, H.S., Cai, L., Davis, S.K., Taylor, J.F., Doud, L.K., Bishop, M.D., Hayes, H., Barendse, W., Vaiman, D., McGraw, R.A., Hirano, T., Sugimoto, Y. and Kirkpatrick, B.W. (1997). Comparative linkage mapping of human chromosome 13 and bovine chromosome 12. Genomics 39: 47-54.

Sun, H.-F.S., Ernst, C.W., Yerle, M., Pinton, P., Rothschild, M.F., Chardon, P., Rogel-Gaillard, C. and Tuggle, C.K. (1999). Human chromosome 3 and pig chromosome 13 show complete synteny conservation but extensive gene-order differences. Cytogenet. Cell Genet. 85: 273-278.

Szemraj, J., Plucienniczak, G., Jaworski, J. and Plucienniczak, A. (1995). Bovine *Alu*-like sequences mediate transposition of a new site-specific retroelement. Gene 152: 261-264.

Szybalski, W., Szybalski, E.H., and Ragni, G. (1962). Genetic studies with human cell lines. Nat. Cancer Institute Monograph 7: 75-88.

Tautz, D., Trick, M. and Dover, G.A. (1986). Cryptic simplicity in DNA is a major source of genetic variation. Nature 322: 652-656.

Tautz, D. (1989). Hypervariability of simple sequences as a general source for polymorphic DNA markers. Nucleic Acids Res. 17 (16): 6463-6471.

Tautz, D. (1992). Slippage synthesis of simple sequence DNA. Nucleic Acids Res. 20 (2): 211-215.

Taylor, K., Hornigold, N., Conway, D., Williams, D., Ulinowski, Z., Agochiya, M., Fattorini, P., de Jone, P., Little, P.F.R. and Wolfe, J. (1996). Mapping of the human Y chromosome by fingerprinting cosmid clones. Genome Res. 6: 235-248.

Taylor, J.F., Coutinho, L.L., Herring, K.L., Gallagher, D.S. Jr., Brenneman, R.A., Burney, N., Sanders, J.O., Turner, J.W., Smith, S.B., Miller, R.K., Savell, J.W. and Davis, S.K. (1998). Candidate gene analysis of GH1 for effects on growth and carcass composition of cattle. Anim. Genet. 29 (3): 194-201.

Terkelson, C., Koch, J., Kϕlvraa, S., Hindkjær, J., Pederson, S. and Bolund, L. (1993). Repeated primed *in situ* labeling: formation and labeling of specific DNA sequences in chromosomes and nuclei. Cytogenet. Cell Genet. 63: 235-237.

The Huntington's collaborative research group. (1993). A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. Cell 72: 971-983.

Thomsen, P.D., Hindkjær, J. and Christensen, K. (1992). Assignment of a porcine male-specific DNA repeat to Y-chromosomal heterochromatin. Cytogenet. Cell Genet. 61: 152-154.

Thomsen, P.D. and Miller, J.R. (1996). Pig genome analysis: differential distribution of SINE and LINE sequences is less pronounced than in the human and mouse genomes. Mamm. Genome 7: 42-46.

Toldo, S.S., Fries, R., Steffen, P., Beibergs, H.L., Barendse, W., Womack, J.E., Hetzel, D.J.S. and Stranzinger, G. (1993). Physically mapped, cosmid-derived microsatellite markers as anchor loci on bovine chromosomes. Mamm. Genome 4: 720-727.

Towner, P. (1991). Purification of DNA. Ch. 3. In: Brown, T.A. (ed). Essential molecular biology: a practical approach. Vol 1. Oxford university press, New York. pp. 47-68.

Troyer, D.L., Xie, H., Goad, D.W. and Skinner, D.Z. (1994). Use of a new technique to map the porcine α interferon gene to Chromosome 1. Mamm. Genome 5: 112-114.

Tucker, E.M., Dain, A.R., Wright, L.J. and Clarke, S.W. (1981). Culture of sheep x mouse hybridoma cells *in vitro*. Hybridoma 1: 77-86.

Tucker R.M. and Burke D.T. (1997). Two vectors for the insertion of mammalian selectable genes into yeast artificial chromosome cloned DNA. Gene 199 (1-2): 25-30.

Vaiman, D., Bahri-Darwich, I., Mercier, D., Yerle, M., Eggen, A., Leveziel, H., Guérin, G., Gellin, J. and Cribiu, E.P. (1993). Mapping of new bovine microsatellites on cattle chromosome 15 with somatic cell hybrids, linkage analysis, and fluoresence *in situ* hybridization. Mamm. Genome 4: 676-679.

Vaiman D., Mercier, D., Moazami-Goudarzi, K., Eggen, A., Ciampolini, R., Lépingle, A., Velmala, R., Kaukinen, J., Varvio, S.L., Matin, P., Levéziel, H. and Guérin, G. (1994). A

set of 99 cattle microsatellites: characterization, synteny mapping, and polymorphism. Mamm. Genome 5: 288-297.

Vaiman, D., Schibler, L., Bourgeois, f., Oustry, A., Amigues, Y. and Cribiu, E.P. (1996). A genetic linkage map of the male goat genome. Genetics 144: 279-305.

Vaiman, D., Schibler, L., Oustry-Vaiman, A., Pailhoux, E., Goldammer, T., Stevanovic, M., Furet, J.P., Schwerin, M., Cotinot, C., Fellous, M. and Cribiu, E.P. (1999a). High-resolution human/goat comparative map of the goat polled/intersex syndrome (PIS): the human homologue is contained in a human YAC from HSA3q23. Genomics 56 (1): 31-39.

Vaiman, D., Billault, A., Tabet-Aoul, K., Schibler, L., Vilette, D., Oustry-Vaiman, A., Soravito, C. and Cribiu, E.P. (1999b). Construction and characterization of a sheep BAC library of three genome equivalents. Mamm. Genome 10: 585-587.

Valdes, A.M., Slatkin, M. and Freimer, N.B. (1993). Allele frequencies at microsatellite loci: The stepwise mutation model revisited. Genetics 133: 737-749.

Van Dilla, M.A. and Deaven, L.L. (1990). Construction of gene libraries for each human chromosome. Cytometry 11 (1): 208-18.

Van Poucke, M., Törnsten, A., Mattheeuws, M., Van Zeveren, A., Peelman L.J. and Chowdhary, B.P. (1999). Comparative mapping between human chromosome 3 and porcine chromosome 13. Cytogenet. Cell Genet. 85: 279-284.

Vogt, F. (1990). Potential genetic functions of tandem repeated DNA sequence blocks in the human genome are based on a highly conserved "chromatin folding code". Hum. Genet. 84: 301-336.

Volpi, E.V. and Baldini, A. (1993). MULTIPRINS: a method for multicolor primed *in situ* labelling. Chromosome Res. 1:257-260.

Wall, R.J., Kerr, D.E. and Bondioli, K.R. (1997). Transgenic dairy cattle: genetic engineering on a large scale. J. Dairy Sci. 80 (9): 2213-2224.

Walter, M.A., Spillet, D.J., Thomas, P., Weissenbach, J. and Goodfellow, P.N. (1994). A method for constructing radiation hybrid maps of whole genomes. Nature Genet. 7: 22-28.

Wang, D.G., Fan, J.-B., Siao, C.-J., Berno, A., Young, P., Sapolsky, R., Ghandour, G., Perkins, N., Winchester, E., Spencer, J., Topaloglou, T., Hubbell, E., Robinson, E., Mittmann, M., Morris, M.S., Shen, N., Kilburn, D., Rioux, J., Nusbaum, C., Rozen, S., Hudson, T.J., Lipshutz, R., Chee, M. and Lander, E.S. (1999). Large-scale identification, mapping and genotyping of single-nucleotide polymorphisms in the human genome. Science 280: 1077-1082.

Wang, G.L., Holsten, T.E., Song, W.Y., Wang, H.P. and Ronald, P.C. (1995). Construction of a rice bacterial artificial chromosome library and identification of clones linked to the Xa-21 disease resistance locus. Plant J. 7: 525-533.

Wang, H.C. and Federoff, S. (1972). Banding juman chromosomes treated with trypsin. Nature New Biol. 235: 52-53.

Wang, K., Boysen, C., Shizuya, H., Simon, M.I. and Hood, L. (1997). Complete nucleotide sequence of two generations of a bacterial artificial chromosome cloning vector. BioTechniques 23: 992-994.

Wang, M., Chen, X.N., Shouse, S., Manson, J., Wu, Q., Li, R., Wrestler, J., Boya, D., Sun, Z.G., Korenberg, J. and Lai, E. (1994). Construction and characterization of a human chromosome 2-specific BAC library. Genomics 24: 527-534.

Wang, M. and Lai, E. (1995). Pulsed field separation of large supercoiled and open-circular DNAs and its application to bacterial artificial chromosome cloning. Electrophoresis 16: 1-7.

Wang, Z., Weber, J.L., Zhong, G. and Tanksley, S.D. (1994). Survey of plant short tandem DNA repeats. Theor. Appl. Genet. 88: 1-6.

Warburton, D., Gersen, S., Yu, M.-T., Jackson, C., Handelin, B. and Housman, D. (1990). Monochromosomal rodent-human hybrids from microcell fusion of human lymphoblastoid cells containing an inserted dominant selectable marker. Genomics 6: 358-366.

Warford, A. and Lauder, I. (1991). *In situ* hybridisation in perspective. J. Clin. Pathol. 44: 177-181.

Weissenbach, J., Gyapay, G., Dip, C., Vignal, A., Morissette, J., Millasseau, P., Baysseix, G. and Lathrop, M. (1992). A second-generation linkage map of the human genome. Nature 359: 794-801.

Weissenbach, J. (1998). The human genome project: from mapping to sequencing. Clin. Chem. Lab. Med. 36 (8): 511-514.

Wienburg, J. and Stanyon, R. (1997). Comparative painting of mammalian chromosomes. Curr. Opin. Genet. Dev. 7: 784-791.

Wolf, J.B., David, V.A. and Deutch, A.H. (1990). Identification of a distal regulatory element in the 5' flanking region of the bovine prolactin gene. Nucleic Acids Res. 18 (16): 4905-4912.

Womack, J.E. and Moll, Y.D. (1986). Gene map of the cow: conservation of linkage with mouse and man. J. Hered. 77: 2-7.

Womack, J.E. (1991). Strategies and technologies for comparative mapping. In: Schook, L.B., Lewin, H.A. and McLaren, D.G. (eds). Gene-mapping techniques and applications. Marcel Dekker, Inc., New York. pp. 3-20.

Womack, J.E. and Kata, S.R. (1995). Bovine genome mapping: evolutionary inference and the power of comparative genomics. Curr. Opin. Genet. Dev. 5: 725-733.

Womack, J.E., Johnson, J.S., Owens, E.K., Rexroad III, C.E., Schläpfer, J. and Yang, Y. (1997). A whole genome radiation hybrid panel for bovine gene mapping. Mamm Genome 8: 854-856.

Wong, G.K., Yu, J., Thayer, E.C. and Olson, M.V. (1997). Multiple-complete-digest restriction fragment mapping: generating sequence-ready maps for large-scale DNA sequencing. Proc. Natl. Acad. Sci. USA 94 (10): 5225-30.

Woo, S.-S., Jiang, J., Gill, B.S., Paterson, A.H. and Wing, R.A. (1994). Construction and characterization of a bacterial artificial chromosome library of Sorghum bicolor. Nucleic Acids Res. 22: 4922-4931.

Woon, P.-Y., Osoegawa, K., Kaisaki, P.J., Zhao, B., Catanese, J.J., Gauguier, D., Cox, R., Levy, E.R., Lathrop, G.M., Monaco, A.P. and de Jong, P.J. (1998). Construction and characterization of a 10-fold genome equivalent rat P1-derived artificial chromosomal library. Genomics 50: 306-316.

Wyman, A.R. and Wertman, K.F. (1987). Host strains that alleviate underrepresentation of specific sequences: Overview. Methods Enzymol. 152: 173-180.

Yang, F., Muller, S., Just, R., Ferguson-Smith, M.A. and Wienberg, J. (1997). Comparative chromosome painting in mammals: human and the Indian muntjac (*Muntiacus muntjak vaginalis*). Genomics 39 (3): 396-401.

Yang, X.W., Model, P. and Heintz, N. (1997). Homologous recombination based modification in *Escherichia coli* and germline transmission in transgenic mice of a bacterial artificial chromosome. Nature Biotechnology 15: 859-865.

Yang, Y.P. and Womack, J.E. (1998). Parallel radiation hybrid mapping: a powerful tool for high-resolution genomic comparison. Genome Res. 8 (7): 731-6.

Yerle, M. and Gellin, J. (1989). Localization of leucocyte interferon gene in the q2.5 region of pig chromosome 1 by *in situ* hybridization. Genetics, Selection, Evolution 21: 249-52.

Yerle, M., Pinton, P., Robic, A., Alfonso, A., Palvadeau, Y., Delcros, C., Hawken, R., Alexander, L., Beattie, C., Schook, L., Milan, D. and Gellin, J. (1998). Construction of a whole-genome radiation hybrid panel for high-resolution gene mapping in pigs. Cytogenet. Cell Genet. 82: 182-188.

Yokobata, K., Trenchak, B. and de Jong, P.J. (1991). Rescue of unstable cosmids by *in vitro* packaging. Nucleic Acids Res. 19 (2): 403-434.

Yoo, J., Ponce de Leon, F.A., Stone, R.T. and Beattie, C.W. (1995). Cloning and chromosomal assignment of the bovine interleukin-2 receptor alpha (IL-2Rα) gene. Mamm. Genome 6 (10): 751-753.

Zhu, H. and Dean, R.A. (1999). A novel method for increasing the transformation efficiency of *Escherichia coli*-application for bacterial artificial chromosome library construction. Nucleic Acids Res. 27 (3): 910-911.

Zimmer, R. and Verrinder Gibbins, A.M. (1997) Construction and characterization of a large-fragment chicken bacterial artificial chromosome library. Genomics 42 (2): 217-226.

APPENDICES

APPENDIX I: SOLUTIONS

Anti-Coagulation Solution (pH 6.1)

	500ml	[Final]
Sodium citrate	125g	25% w/v
Glucose	150g	30% w/v

Adjust pH to 6.1 and store at 4°C.

SDE Solution (pH 8.0)

	1L	[Final]
Sodium dodecyl sulphate	10g	1% w/v
EDTA-di sodium salt	9.3g	25mM

Adjust pH to 8.0.

ESP Solution (pH 8.0)

	1L	[Final]
Sarcosyl	10g	1% w/v
EDTA-di sodium salt	9.3g	25mM

Adjust pH to 8.0.

Immediately before use in microbead preparation, mix 47.5ml ESP and 2.5ml 20mg/ml proteinase K (1mg/ml final concentration).

LB Medium

	1L	[Final]
Bactotryptone	10g	1%
Bactoyeast extract	5g	0.5%
NaCl	10g	1%
Agar (for plates)	15g	1.5%

Adjust pH to 7.5 and autoclave.

Chloramphenicol (CM)

Stock solution: 50mg/ml in 100% ethanol. Store at -20°C

Use 30µg/ml for pBeloBAC11 plasmid preparation and 12.5µg/ml for BAC transformations and plasmid minipreps.

QBT Buffer

	1L	[Final]
NaCl	43.83g	750mM
MOPS (free acid)	10.46g	50mM
Isopropanol	150ml	15% v/v
Triton-X 100	15ml	0.15% v/v

Adjust pH to 7.0.

QC Buffer

	1L	[Final]
NaCl	58.44g	1.0M
MOPS (free acid)	10.46g	50mM
Isopropanol	150ml	15% v/v

Adjust pH to 7.0.

QF Buffer

	1L	[Final]
NaCl	73.05g	1.25M
Tris base	6.06g	50mM
Isopropanol	150ml	15% v/v

Adjust pH to 8.5.

0.5M EDTA (pH 8.0)

	1L	[Final]
EDTA-di sodium salt	186.1g	0.5M
NaOH	20g	

Adjust to pH 8.0 with NaOH and autoclave.

1M Tris-HCl (pH 8.0)

	1L	[Final]
Tris base	121.1g	1M
HCl	42ml	

Adust to pH 8.0 with HCl and autoclave.

$T_{10}E_1$ (pH 8.0)

	1L	[Final]
Tris-HCl	10ml of 1M stock	10mM
EDTA	2ml of 0.5M stock	1.0mM

Autoclave.

4M NaCl

	1L	[Final]
NaCl	233.76g	4M

Autoclave.

Solution A for sperm cell isolation

	1L	[Final]
Tris	10ml of 1M stock	10mM
EDTA	20ml of 0.5M stock	10mM
NaCl	25ml of 4M stock	100mM

Adjust pH to 7.0 and autoclave.

Solution B for sperm cell isolation

	1L	[Final]
Tris	10ml of 1M stock	10mM
NaCl	25ml of 4M stock	100mM
EDTA	20ml of 0.5M stock	10mM
Sodium dodecyl sulphate	5g	0.5%w/v
ß-mercaptoethanol	2ml	0.2%v/v

Adust pH to 8.0.

Phenylmethyl sulfonyl fluoride (PMSF)

	Amount	[Final]
PMSF	250mg vial	100mM
Isopropanol	14.35ml	

Prepare in fume-hood and store at 4°C.

10 x TBE Electrophoresis buffer

	1L	**[Final]**
Tris base	108g	0.45M
Boric acid	55g	0.45M
EDTA	40ml of 0.5M stock	10mM

2 x BAC Storage solution

	1L	**[Final]**
Glycerin (>99% pure)	650ml	65%
1M Tris-HCl (pH 8)	25ml	0.025M
$MgSO_4$ heptahydrate	24.6g	0.1M

Dissolve $MgSO_4$ separately in 50ml ddH_20 before adding to glycerin mix.

Autoclave and store at 4°C.

10N NaOH

	100ml
NaOH	40g

Solution I

	1L	**[Final]**
Glucose	9g	50mM
0.5M EDTA (pH 8.0)	20ml	10mM
1M Tris-HCl (pH 8.0)	2.5ml	25mM

Autoclave.

Immediately before use add 5mg/ml lysozyme.

Solution II

	1L	**[Final]**
10N NaOH	20ml	0.2N
SDS	10g	1%w/v

Solution III

	1L	**[Final]**
Potassium acetate	294.42g	3M
Glacial acetic acid	285ml	

25:24:1 Phenol:Chloroform:Isoamyl alcohol

	500ml
TE Buffered phenol	250ml
Chloroform	240ml
Isoamyl alcohol	10ml

Cover with fresh TE and store at 4°C.

Digestion buffer

	1L	**[Final]**
4M NaCl	25ml	100mM
1M Tris-HCl (pH 8.0)	10ml	10mM
0.5M EDTA (pH 8.0)	50ml	25mM
SDS	5g	0.5% w/v

Adjust volume to 1L with ddH_20 and autoclave.

CTAB Buffer

	1L	**[Final]**
CTAB	20g	2% w/v
1M Tris-HCl (pH 8.0)	100ml	0.1M
0.5M EDTA (pH 8.0)	40ml	20mM
4M NaCl	350ml	1.4M
PVP (40kd)	10g	1%w/v

5% CTAB

	100ml
CTAB	5g

CTAB precipitation buffer

	1L	**[Final]**
CTAB	10g	1%w/v
1M Tris-HCl (pH 8.0)	50ml	50mM
0.5M EDTA (pH 8.0)	20ml	10mM

10mg/ml DNase-free RNase

	25ml	**[Final]**
RNase A	250mg	10mg/ml
Tris base	30mg	10mM
NaCl	20mg	15mM

Heat to 100°C for 15ml and allow to cool slowly to room temperature.

Split into 1ml aliquots and store at -20°C.

10mg/ml ethidium bromide stock

	100ml
Ethidium bromide	1g

Store in dark bottle.

50xTAE

	1L	**[Final]**
Tris base	242g	2M
EDTA	100ml of 0.5M stock	50mM
Glacial acetic acid	57.1ml	

20xSSC

	1L	**[Final]**
NaCl	175.3g	3M
Sodium citrate	88.2g	0.3M

Adjust pH to 7.0 and autoclave.

4xSSC/0.05% Tween20

	1L	**[Final]**
20xSSC	200ml	4x
Tween20	0.5ml	0.05%v/v

Adjust volume to 1L with ddH_20.

4xSSC/1% BSA

	100ml	**[Final]**
20xSSC	20ml	4x
BSA	1g	1%

Filter sterilise and split into 1.5ml aliquots. Store at -20°C.

PBS

	1L	**[Final]**
NaCl	8g	137mM
KCl	0.2g	2.7mM
Na_2HPO_4	1.44g	10mM
KH_2PO_4	0.24g	1.8mM

Adjust pH to 7.4 and autoclave.

Propidium iodide

	10ml	**[Final]**
Propidium iodide	50mg	5mg/ml

Dissolve in PBS.

To prepare chromosome stain, dilute 50µl propidium iodide stock solution in 50ml PBS (5µg/ml final).

PPD11 Antifade Mountant

	100ml	**[Final]**
p-phenylenediamine free base	100mg	1mg/ml
Glycerol	90ml	9 parts
PBS	10ml	1 part

Adjust pH to 11 and store at -20°C.

10 x TBE

	1L	**[Final]**
Tris	108g	0.9M
Boric acid	55g	0.45M
0.5M EDTA (pH 8.0)	40ml	20mM

Filter through Whatman 541 paper.

40% acrylamide solution

	500ml	**[Final]**
Acrylamide	190g	38% w/v
Bis-acrylamide	10g	2% w/v

Heat to 37°C to dissolve.

Filter through Whatman 541 paper and store in a dark bottle at 4°C.

6% acrylamide sequencing gel solution

	500ml	**[Final]**
Urea	210g	7M
40% acrylamide solution	75ml	6% v/v
10 x TBE	50ml	1xTBE

Filter through Whatman 541 paper and store in a dark bottle at 4°C for up to 14 days.

2M $MgSO_4$/$MgCl_2$

	50ml	**[Final]**
$MgCl_2.6H_20$	10.17g	1M
$MgSO_4.7H_20$	12.32g	1M

Filter sterilise.

2M Glucose

	50ml
D-Glucose	18g

Filter sterilise.

SOC Medium

	490ml	**[Final]**
Bactotryptone	10g	2% w/v
Bactoyeast extract	2.5g	0.5% w/v
NaCl	0.29g	0.058% w/v
KCl	0.1g	0.02% w/v

Adjust pH to 7.0.

Aliquot 49ml into bottles and autoclave.

Immediately before use add 0.5ml 2M $MgSO_4/MgCl_2$ and 0.5ml 2M Glucose.

SOB Medium (without Mg^{2+})

	800ml	**[Final]**
Tryptone	16g	2% w/v
Yeast extract	4g	0.5% w/v
NaCl	0.47g	0.058% w/v
KCl	0.15g	0.02% w/v

Adjust pH to 7.0.

Autoclave in two flasks - 750ml in a 2L flask and 50ml in a 250ml flask.

Denaturing Solution

	1L	**[Final]**
NaCl	87.66g	1.5M
NaOH	20g	0.5M

Neutralising Solution

	1L	**[Final]**
NaCl	175.3g	3M
Tris-HCl	60.5g	0.5M

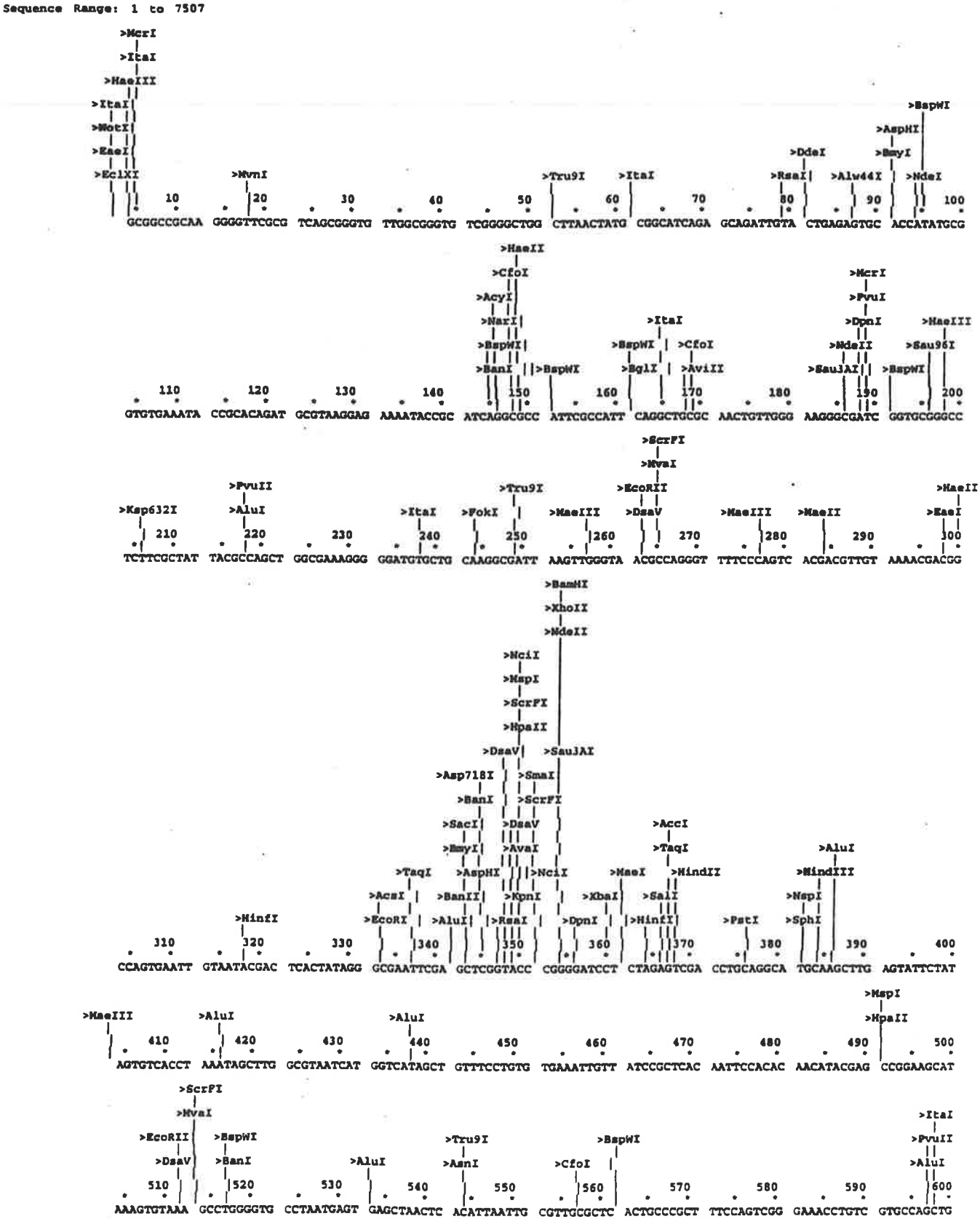
Adjust pH to 7.0.

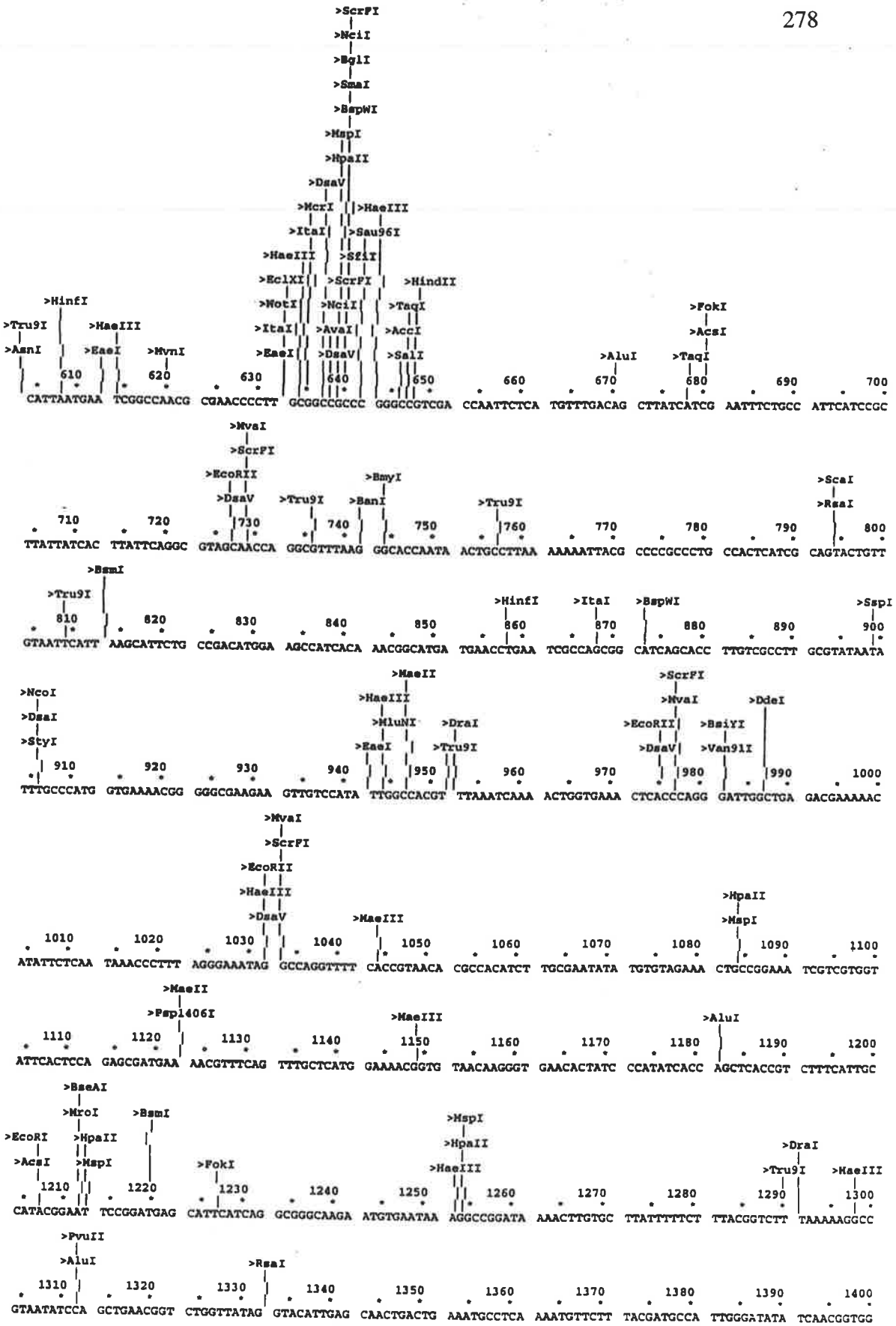
Probe Hybridisation Solution

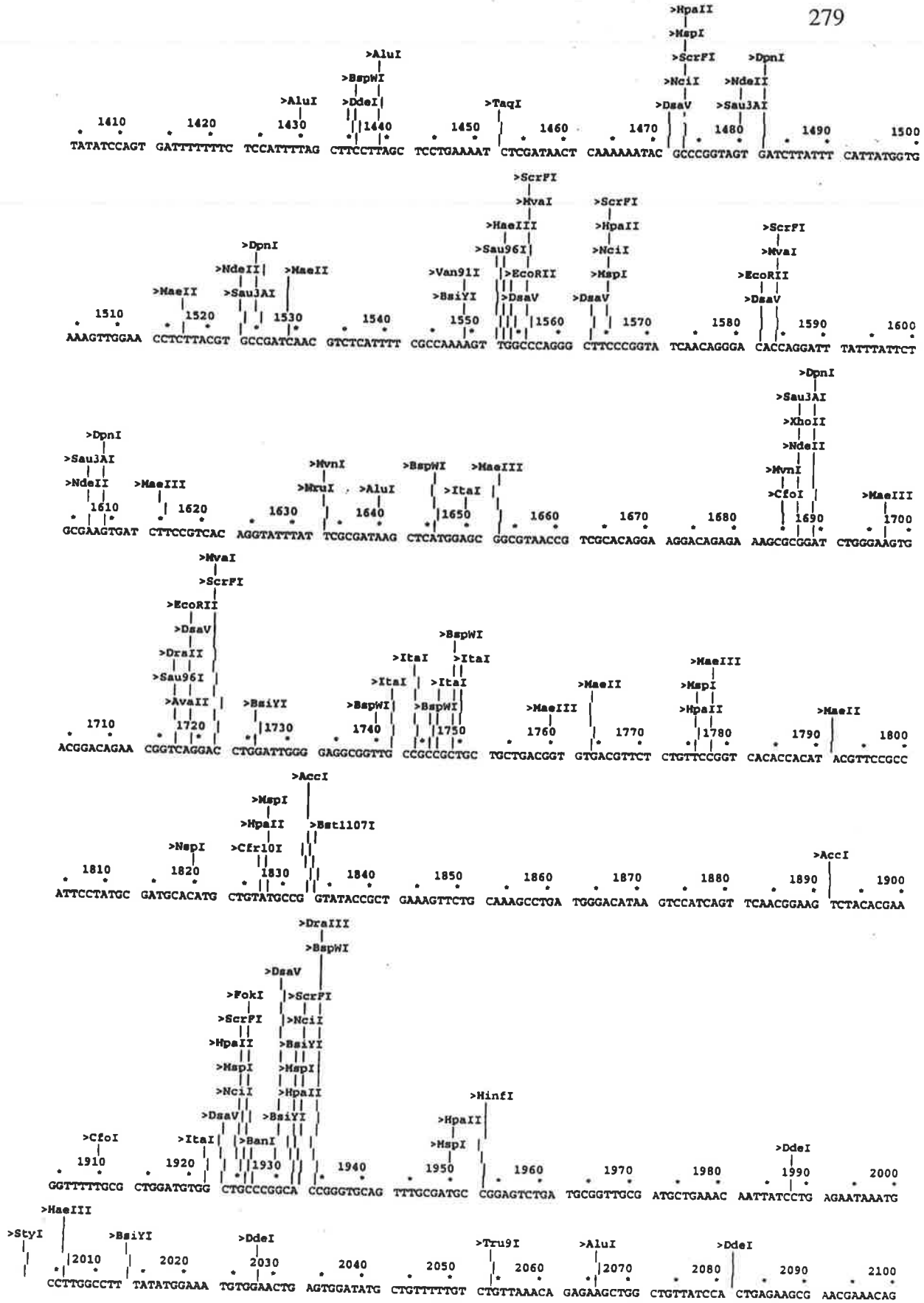
	100ml	**[Final]**
N-laurylsarcosine-Na salt	0.1g	0.1% w/v
SDS	0.02g	0.02% w/v
Blocking reagent (*Boehringer*)	1g	1% w/v
20xSSC	25ml	5xSSC

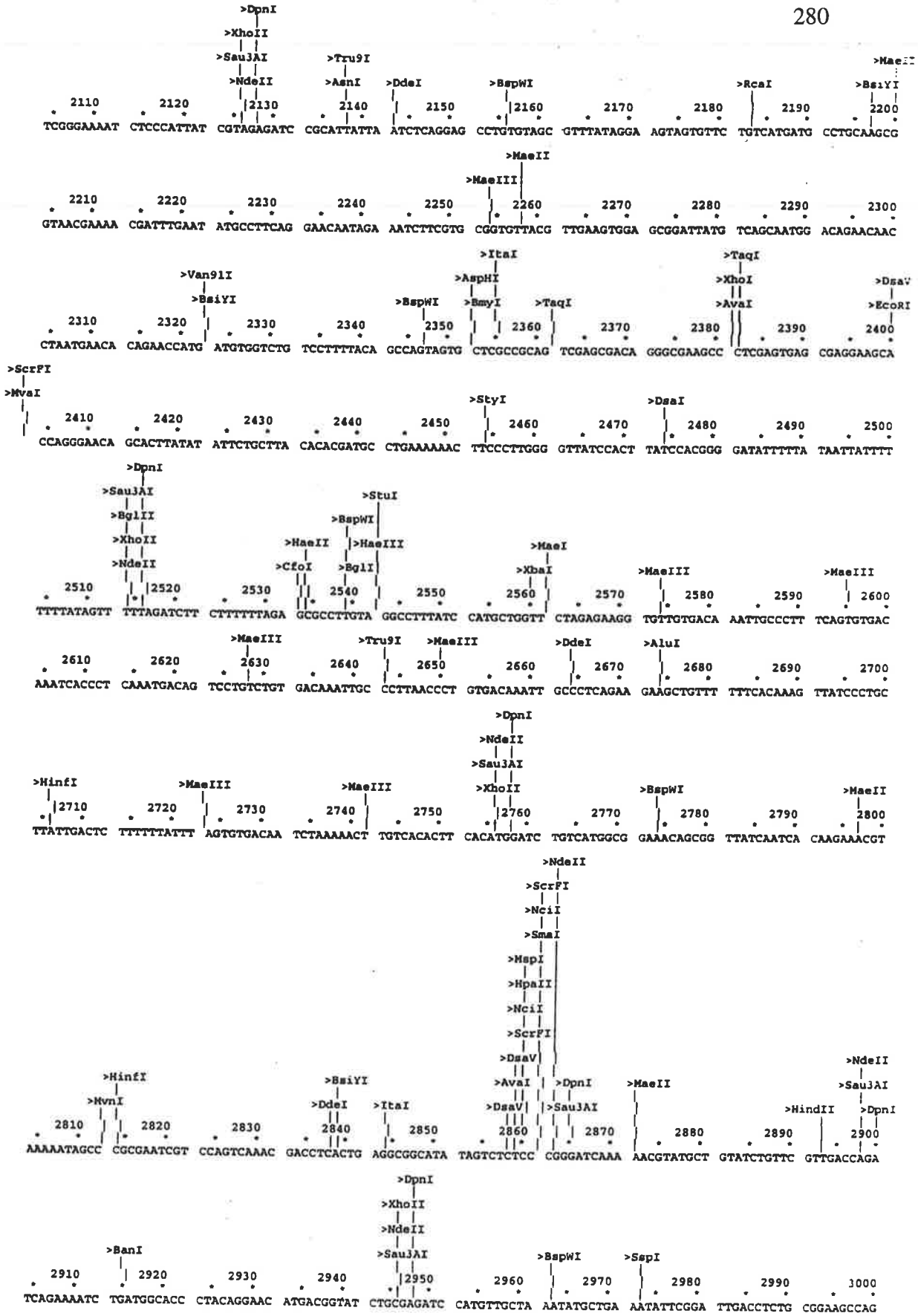
Heat the solution to 37°C and stir vigorously until the Blocking reagent has dissolved.

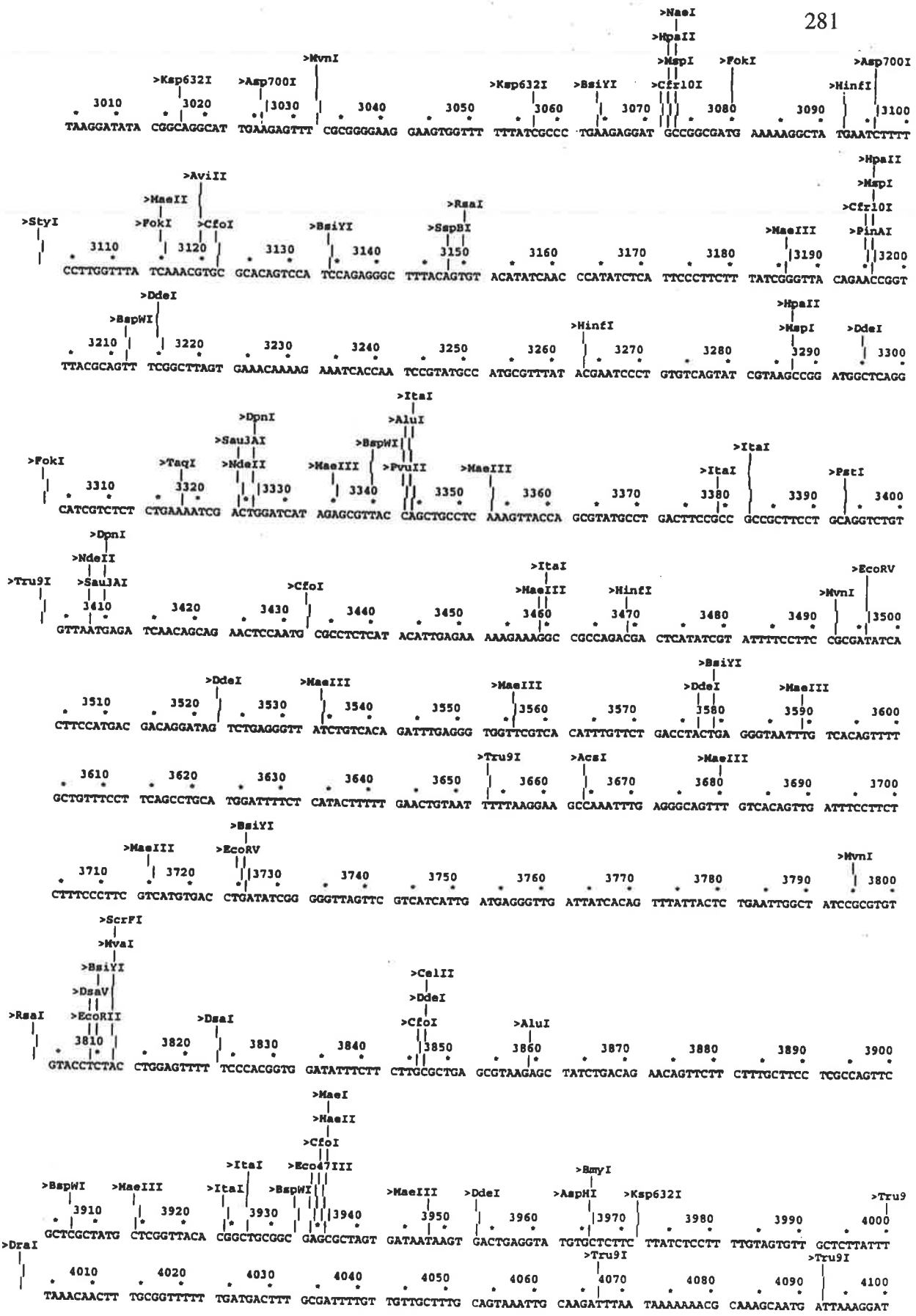
APPENDIX II: pBeloBACII

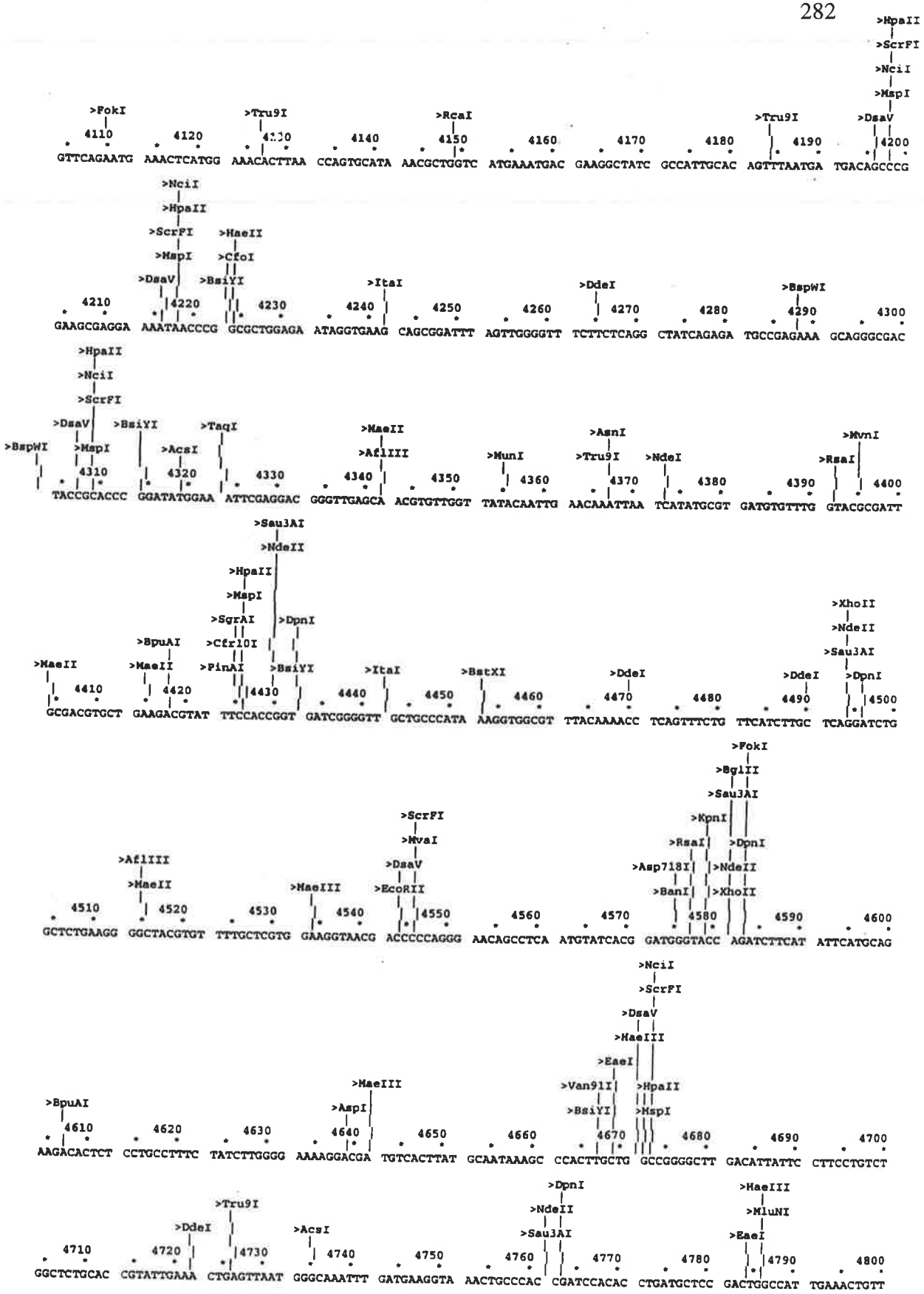


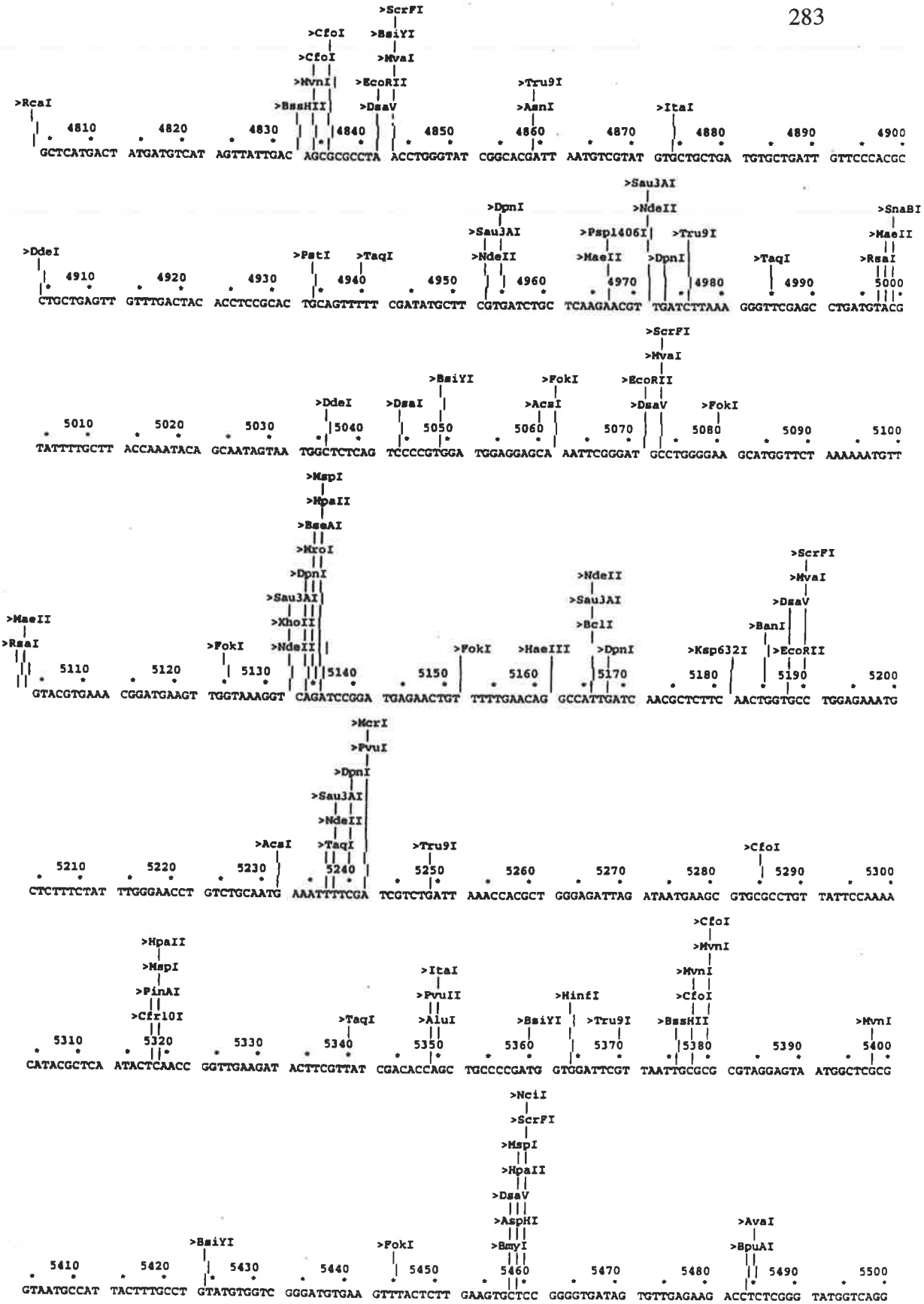

>ScrFI
>NciI
>BglI
>SmaI
>BspWI
>MspI
>HpaII
>DsaV
>McrI
>HaeIII
>ItaI
>Sau96I
>HaeIII
>SfiI
>EclXI
>ScrFI
>HindII
>HinfI
>NotI
>NciI
>TaqI
>FokI
>Tru9I
>HaeIII
>ItaI
>AvaI
>AccI
>AcsI
>AsnI
>EaeI
>EaeI
>DsaV
>SalI
>AluI
>TaqI
610 620 630 640 650 660 670 680 690 700
CATTAATGAA TCGGCCAACG CGAACCCCTT GCGGCCGCCC GGGCCGTCGA CCAATTCTCA TGTTTGACAG CTTATCATCG AATTTCTGCC ATTCATCCGC
>MvaI
>ScrFI
>EcoRII
>BmyI
>ScaI
>DsaV
>Tru9I
>BanI
>Tru9I
>RsaI
710 720 730 740 750 760 770 780 790 800
TTATTATCAC TTATTCAGGC GTAGCAACCA GGCGTTTAAG GGCACCAATA ACTGCCTTAA AAAAATTACG CCCCGCCCTG CCACTCATCG CAGTACTGTT
>BsmI
>Tru9I
>HinfI
>ItaI
>BspWI
>SspI
810 820 830 840 850 860 870 880 890 900
GTAATTCATT AAGCATTCTG CCGACATGGA AGCCATCACA AACGGCATGA TGAACCTGAA TCGCCAGCGG CATCAGCACC TTGTCGCCTT GCGTATAATA
>HaeII
>NcoI
>HaeIII
>ScrFI
>DsaI
>MluNI
>DraI
>NvaI
>DdeI
>StyI
>EaeI
>Tru9I
>EcoRII
>BsiYI
>DsaV
>Van91I
910 920 930 940 950 960 970 980 990 1000
TTTGCCCATG GTGAAAACGG GGGCGAAGAA GTTGTCCATA TTGGCCACGT TTAAATCAAA ACTGGTGAAA CTCACCCAGG GATTGGCTGA GACGAAAAAC
>MvaI
>ScrFI
>EcoRII
>HaeIII
>HpaII
>DsaV
>HaeIII
>MspI
1010 1020 1030 1040 1050 1060 1070 1080 1090 1100
ATATTCTCAA TAAACCCTTT AGGGAAATAG GCCAGGTTTT CACCGTAACA CGCCACATCT TGCGAATATA TGTGTAGAAA CTGCCGGAAA TCGTCGTGGT
>HaeII
>Psp1406I
>HaeIII
>AluI
1110 1120 1130 1140 1150 1160 1170 1180 1190 1200
ATTCACTCCA GAGCGATGAA AACGTTTCAG TTTGCTCATG GAAAACGGTG TAACAAGGGT GAACACTATC CCATATCACC AGCTCACCGT CTTTCATTGC
>BseAI
>MroI
>BsmI
>HspI
>EcoRI
>HpaII
>HpaII
>DraI
>AcsI
>MspI
>FokI
>HaeIII
>Tru9I
>HaeIII
1210 1220 1230 1240 1250 1260 1270 1280 1290 1300
CATACGGAAT TCCGGATGAG CATTCATCAG GCGGGCAAGA ATGTGAATAA AGGCCGGATA AAACTTGTGC TTATTTTTCT TTACGGTCTT TAAAAAGGCC
>PvuII
>AluI
>RsaI
1310 1320 1330 1340 1350 1360 1370 1380 1390 1400
GTAATATCCA GCTGAACGGT CTGGTTATAG GTACATTGAG CAACTGACTG AAATGCCTCA AAATGTTCTT TACGATGCCA TTGGGATATA TCAACGGTGG

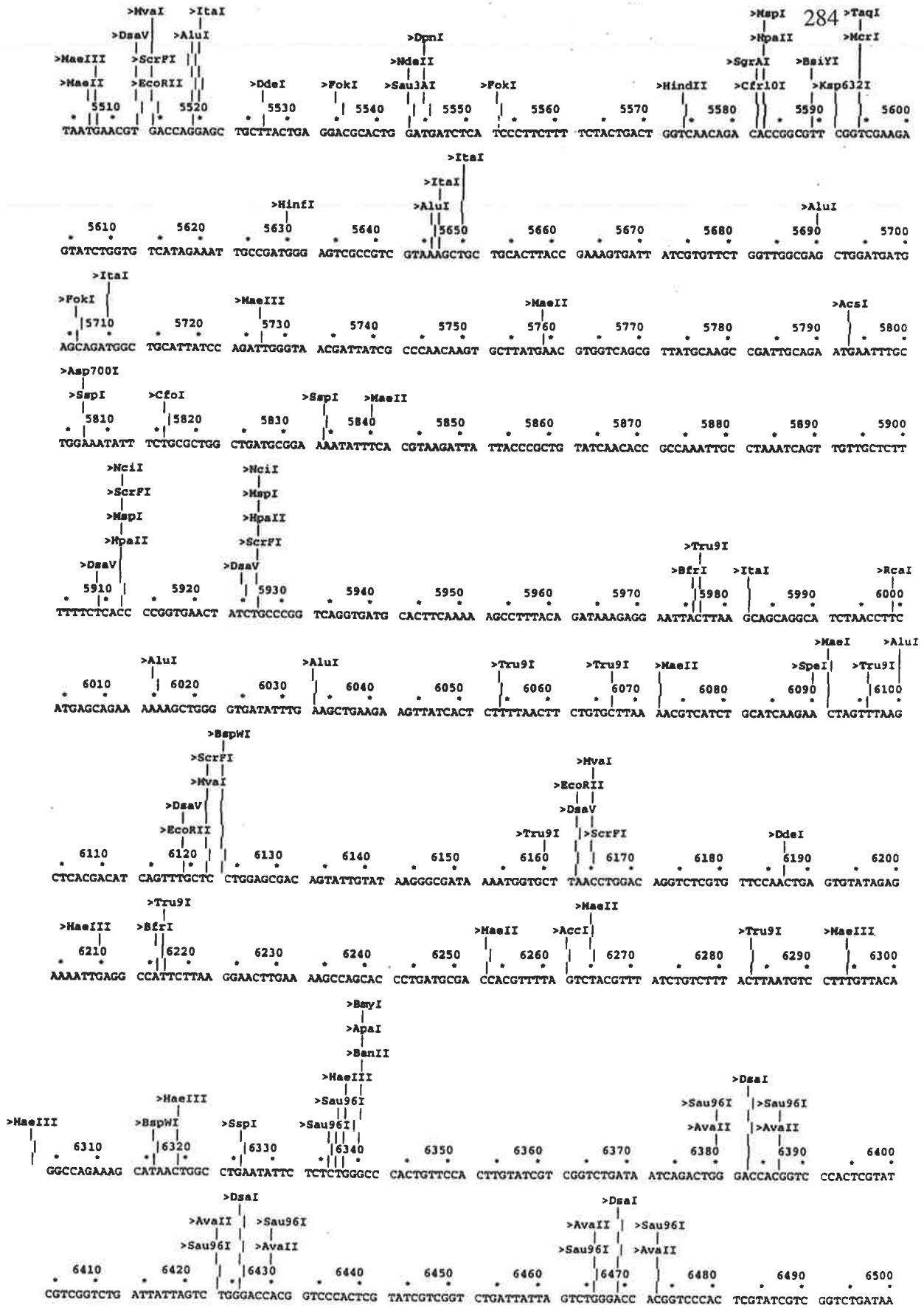

1410 1420 1430 1440 1450 1460 1470 1480 1490 1500
TATATCCAGT GATTTTTTTC TCCATTTTAG CTTCCTTAGC TCCTGAAAAT CTCGATAACT CAAAAAATAC GCCCGGTAGT GATCTTATTT CATTATGGTG
1510 1520 1530 1540 1550 1560 1570 1580 1590 1600
AAAGTTGGAA CCTCTTACGT GCCGATCAAC GTCTCATTTT CGCCAAAAGT TGGCCCAGGG CTTCCCGGTA TCAACAGGGA CACCAGGATT TATTTATTCT
1610 1620 1630 1640 1650 1660 1670 1680 1690 1700
GCGAAGTGAT CTTCCGTCAC AGGTATTTAT TCGCGATAAG CTCATGGAGC GGCGTAACCG TCGCACAGGA AGGACAGAGA AAGCGCGGAT CTGGGAAGTG
1710 1720 1730 1740 1750 1760 1770 1780 1790 1800
ACGGACAGAA CGGTCAGGAC CTGGATTGGG GAGGCGGTTG CCGCCGCTGC TGCTGACGGT GTGACGTTCT CTGTTCCGGT CACACCACAT ACGTTCCGCC
1810 1820 1830 1840 1850 1860 1870 1880 1890 1900
ATTCCTATGC GATGCACATG CTGTATGCCG GTATACCGCT GAAAGTTCTG CAAAGCCTGA TGGGACATAA GTCCATCAGT TCAACGGAAG TCTACACGAA
1910 1920 1930 1940 1950 1960 1970 1980 1990 2000
GGTTTTTGCG CTGGATGTGG CTGCCCGGCA CCGGGTGCAG TTTGCGATGC CGGAGTCTGA TGCGGTTGCG ATGCTGAAAC AATTATCCTG AGAATAAATG
2010 2020 2030 2040 2050 2060 2070 2080 2090 2100
CCTTGGCCTT TATATGGAAA TGTGGAACTG AGTGGATATG CTGTTTTTGT CTGTTAAACA GAGAAGCTGG CTGTTATCCA CTGAGAAGCG AACGAAACAG

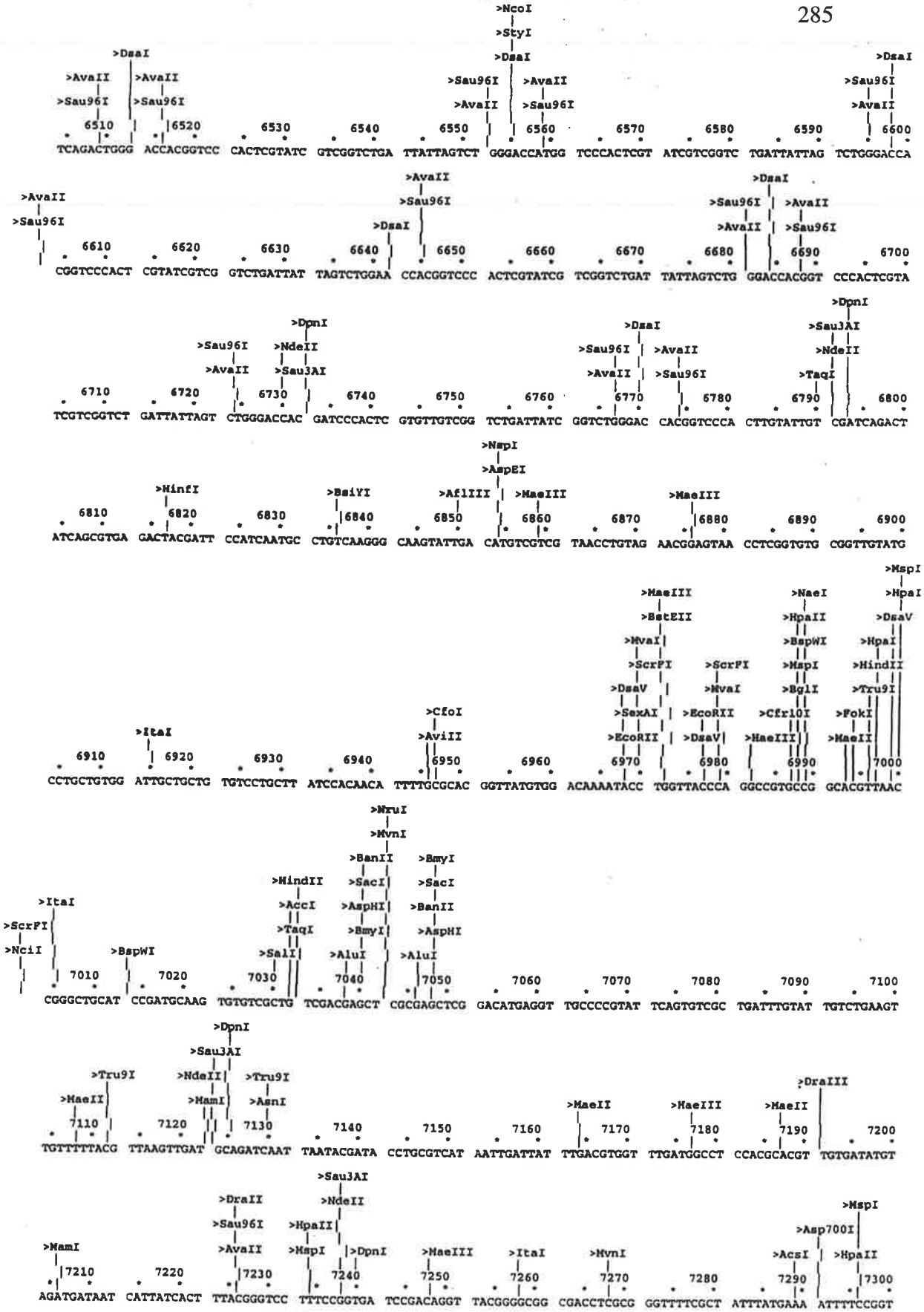

>DpnI
>XhoII
>Sau3AI
>NdeII
>Tru9I
>AsnI
>DdeI
>BspWI
>RcaI
>MaeII
>BsiYI
2110 2120 2130 2140 2150 2160 2170 2180 2190 2200
TCGGGAAAAT CTCCCATTAT CGTAGAGATC CGCATTATTA ATCTCAGGAG CCTGTGTAGC GTTTATAGGA AGTAGTGTTC TGTCATGATG CCTGCAAGCG
>MaeII
>MaeIII
2210 2220 2230 2240 2250 2260 2270 2280 2290 2300
GTAACGAAAA CGATTTGAAT ATGCCTTCAG GAACAATAGA AATCTTCGTG CGGTGTTACG TTGAAGTGGA GCGGATTATG TCAGCAATGG ACAGAACAAC
>ItaI
>TaqI
>Van91I
>AspHI
>XhoI
>DsaV
>BsiYI
>BspWI
>BmyI
>TaqI
>AvaI
>EcoRI
2310 2320 2330 2340 2350 2360 2370 2380 2390 2400
CTAATGAACA CAGAACCATG ATGTGGTCTG TCCTTTTACA GCCAGTAGTG CTCGCCGCAG TCGAGCGACA GGGCGAAGCC CTCGAGTGAG CGAGGAAGCA
>ScrFI
>MvaI
>StyI
>DsaI
2410 2420 2430 2440 2450 2460 2470 2480 2490 2500
CCAGGGAACA GCACTTATAT ATTCTGCTTA CACACGATGC CTGAAAAAAC TTCCCTTGGG GTTATCCACT TATCCACGGG GATATTTTTA TAATTATTTT
>DpnI
>Sau3AI
>BglII
>XhoII
>NdeII
>StuI
>BspWI
>HaeII
>HaeIII
>CfoI
>BglI
>MaeI
>XbaI
>MaeIII
>MaeIII
2510 2520 2530 2540 2550 2560 2570 2580 2590 2600
TTTTATAGTT TTTAGATCTT CTTTTTTAGA GCGCCTTGTA GGCCTTTATC CATGCTGGTT CTAGAGAAGG TGTTGTGACA AATTGCCCTT TCAGTGTGAC
>MaeIII
>Tru9I
>MaeIII
>DdeI
>AluI
2610 2620 2630 2640 2650 2660 2670 2680 2690 2700
AAATCACCCT CAAATGACAG TCCTGTCTGT GACAAATTGC CCTTAACCCT GTGACAAATT GCCCTCAGAA GAAGCTGTTT TTTCACAAAG TTATCCCTGC
>DpnI
>NdeII
>Sau3AI
>HinfI
>MaeIII
>MaeIII
>XhoII
>BspWI
>MaeII
2710 2720 2730 2740 2750 2760 2770 2780 2790 2800
TTATTGACTC TTTTTTATTT AGTGTGACAA TCTAAAAACT TGTCACACTT CACATGGATC TGTCATGGCG GAAACAGCGG TTATCAATCA CAAGAAACGT
>NdeII
>ScrFI
>NciI
>SmaI
>MspI
>HpaII
>NciI
>ScrFI
>DsaV
>HinfI
>BsiYI
>AvaI
>DpnI
>MaeII
>NdeII
>MvnI
>DdeI
>ItaI
>DsaV
>Sau3AI
>Sau3AI
>HindII
>DpnI
2810 2820 2830 2840 2850 2860 2870 2880 2890 2900
AAAAATAGCC CGCGAATCGT CCAGTCAAAC GACCTCACTG AGGCGGCATA TAGTCTCTCC CGGGATCAAA AACGTATGCT GTATCTGTTC GTTGACCAGA
>DpnI
>XhoII
>NdeII
>BanI
>Sau3AI
>BspWI
>SspI
2910 2920 2930 2940 2950 2960 2970 2980 2990 3000
TCAGAAAATC TGATGGCACC CTACAGGAAC ATGACGGTAT CTGCGAGATC CATGTTGCTA AATATGCTGA AATATTCGGA TTGACCTCTG CGGAAGCCAG

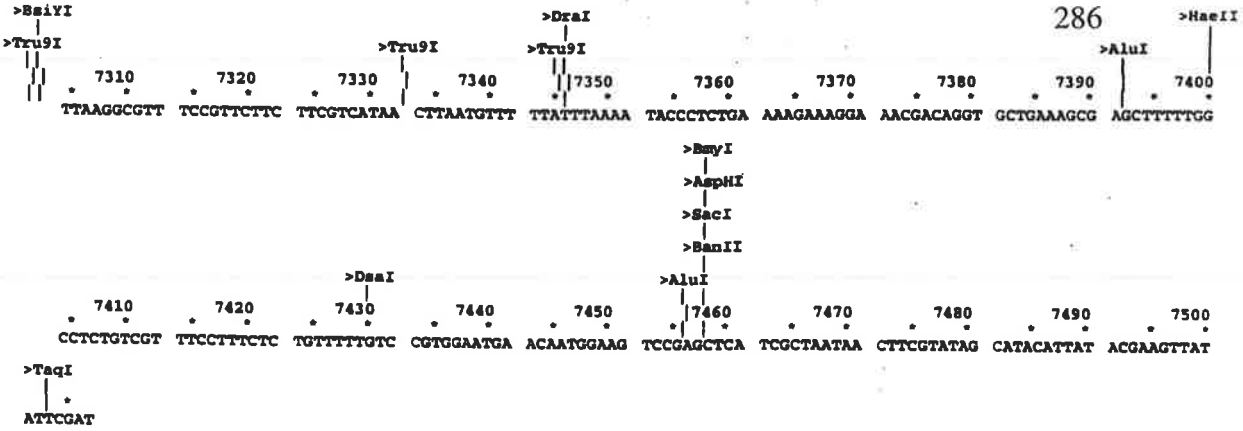

>NaeI
>HpaII
>MvnI
>MspI
>FokI
>Asp700I
>Ksp632I
>Asp700I
>Ksp632I
>BsiYI
>Cfr10I
>HinfI
3010 3020 3030 3040 3050 3060 3070 3080 3090 3100
TAAGGATATA CGGCAGGCAT TGAAGAGTTT CGCGGGGAAG GAAGTGGTTT TTTATCGCCC TGAAGAGGAT GCCGGCGATG AAAAAGGCTA TGAATCTTTT
>HpaII
>AviII
>MspI
>HaeII
>RsaI
>Cfr10I
>StyI
>FokI
>CfoI
>BsiYI
>SspBI
>HaeIII
>PinAI
3110 3120 3130 3140 3150 3160 3170 3180 3190 3200
CCTTGGTTTA TCAAACGTGC GCACAGTCCA TCCAGAGGGC TTTACAGTGT ACATATCAAC CCATATCTCA TTCCCTTCTT TATCGGGTTA CAGAACCGGT
>DdeI
>HpaII
>BspWI
>HinfI
>MspI
>DdeI
3210 3220 3230 3240 3250 3260 3270 3280 3290 3300
TTACGCAGTT TCGGCTTAGT GAAACAAAAG AAATCACCAA TCCGTATGCC ATGCGTTTAT ACGAATCCCT GTGTCAGTAT CGTAAGCCGG ATGGCTCAGG
>ItaI
>DpnI
>AluI
>Sau3AI
>BspWI
>ItaI
>FokI
>TaqI
>NdeII
>HaeIII
>PvuII
>HaeIII
>PstI
3310 3320 3330 3340 3350 3360 3370 3380 3390 3400
CATCGTCTCT CTGAAAATCG ACTGGATCAT AGAGCGTTAC CAGCTGCCTC AAAGTTACCA GCGTATGCCT GACTTCCGCC GCCGCTTCCT GCAGGTCTGT
>DpnI
>NdeII
>ItaI
>Tru9I
>Sau3AI
>EcoRV
>CfoI
>HaeIII
>HinfI
>MvnI
3410 3420 3430 3440 3450 3460 3470 3480 3490 3500
GTTAATGAGA TCAACAGCAG AACTCCAATG CGCCTCTCAT ACATTGAGAA AAAGAAAGGC CGCCAGACGA CTCATATCGT ATTTTCCTTC CGCGATATCA
>BsiYI
>DdeI
>MaeIII
>MaeIII
>DdeI
>MaeIII
3510 3520 3530 3540 3550 3560 3570 3580 3590 3600
CTTCCATGAC GACAGGATAG TCTGAGGGTT ATCTGTCACA GATTTGAGGG TGGTTCGTCA CATTTGTTCT GACCTACTGA GGGTAATTTG TCACAGTTTT
>Tru9I
>AcsI
>MaeIII
3610 3620 3630 3640 3650 3660 3670 3680 3690 3700
GCTGTTTCCT TCAGCCTGCA TGGATTTTCT CATACTTTTT GAACTGTAAT TTTTAAGGAA GCCAAATTTG AGGGCAGTTT GTCACAGTTG ATTTCCTTCT
>BsiYI
>MaeIII
>EcoRV
>MvnI
3710 3720 3730 3740 3750 3760 3770 3780 3790 3800
CTTTCCCTTC GTCATGTGAC CTGATATCGG GGGTTAGTTC GTCATCATTG ATGAGGGTTG ATTATCACAG TTTATTACTC TGAATTGGCT ATCCGCGTGT
>ScrFI
>MvaI
>BsiYI
>DsaV
>CelII
>DdeI
>RsaI
>EcoRII
>DsaI
>CfoI
>AluI
3810 3820 3830 3840 3850 3860 3870 3880 3890 3900
GTACCTCTAC CTGGAGTTTT TCCCACGGTG GATATTTCTT CTTGCGCTGA GCGTAAGAGC TATCTGACAG AACAGTTCTT CTTTGCTTCC TCGCCAGTTC
>MaeI
>HaeII
>CfoI
>ItaI
>Eco47III
>BmyI
>BspWI
>MaeIII
>ItaI
>BspWI
>MaeIII
>DdeI
>AspHI
>Ksp632I
>Tru9
3910 3920 3930 3940 3950 3960 3970 3980 3990 4000
GCTCGCTATG CTCGGTTACA CGGCTGCGGC GAGCGCTAGT GATAATAAGT GACTGAGGTA TGTGCTCTTC TTATCTCCTT TTGTAGTGTT GCTCTTATTT
>DraI
>Tru9I
>Tru9I
4010 4020 4030 4040 4050 4060 4070 4080 4090 4100
TAAACAACTT TGCGGTTTTT TGATGACTTT GCGATTTTGT TGTTGCTTTG CAGTAAATTG CAAGATTTAA TAAAAAAACG CAAAGCAATG ATTAAAGGAT


>HpaII
>ScrFI
>NciI
>MspI
>FokI
>Tru9I
>RcaI
>Tru9I
>DsaV
4110 4120 4130 4140 4150 4160 4170 4180 4190 4200
GTTCAGAATG AAACTCATGG AAACACTTAA CCAGTGCATA AACGCTGGTC ATGAAATGAC GAAGGCTATC GCCATTGCAC AGTTTAATGA TGACAGCCCG
>NciI
>HpaII
>ScrFI
>HaeII
>MspI
>CfoI
>DsaV
>BsiYI
>ItaI
>DdeI
>BspWI
4210 4220 4230 4240 4250 4260 4270 4280 4290 4300
GAAGCGAGGA AAATAACCCG GCGCTGGAGA ATAGGTGAAG CAGCGGATTT AGTTGGGGTT TCTTCTCAGG CTATCAGAGA TGCCGAGAAA GCAGGGCGAC
>HpaII
>NciI
>ScrFI
>DsaV
>BsiYI
>TaqI
>BspWI
>MspI
>AcsI
>MaeII
>AflIII
>MunI
>AsnI
>Tru9I
>NdeI
>MvnI
>RsaI
4310 4320 4330 4340 4350 4360 4370 4380 4390 4400
TACCGCACCC GGATATGGAA ATTCGAGGAC GGGTTGAGCA ACGTGTTGGT TATACAATTG AACAAATTAA TCATATGCGT GATGTGTTTG GTACGCGATT
>Sau3AI
>NdeII
>HpaII
>MspI
>SgrAI
>DpnI
>BpuAI
>Cfr10I
>XhoII
>NdeII
>Sau3AI
>MaeII
>MaeII
>PinAI
>BsiYI
>ItaI
>BstXI
>DdeI
>DdeI
>DpnI
4410 4420 4430 4440 4450 4460 4470 4480 4490 4500
GCGACGTGCT GAAGACGTAT TTCCACCGGT GATCGGGGTT GCTGCCCATA AAGGTGGCGT TTACAAAACC TCAGTTTCTG TTCATCTTGC TCAGGATCTG
>FokI
>BglII
>Sau3AI
>ScrFI
>KpnI
>MvaI
>RsaI
>DpnI
>AflIII
>DsaV
>Asp718I
>NdeII
>MaeII
>MaeIII
>EcoRII
>BanI
>XhoII
4510 4520 4530 4540 4550 4560 4570 4580 4590 4600
GCTCTGAAGG GGCTACGTGT TTTGCTCGTG GAAGGTAACG ACCCCCAGGG AACAGCCTCA ATGTATCACG GATGGGTACC AGATCTTCAT ATTCATGCAG
>NciI
>ScrFI
>DsaV
>HaeIII
>EaeI
>MaeIII
>BpuAI
>Van91I
>HpaII
>AspI
>BsiYI
>MspI
4610 4620 4630 4640 4650 4660 4670 4680 4690 4700
AAGACACTCT CCTGCCTTTC TATCTTGGGG AAAAGGACGA TGTCACTTAT GCAATAAAGC CCACTTGCTG GCCGGGGCTT GACATTATTC CTTCCTGTCT
>DpnI
>HaeIII
>Tru9I
>NdeII
>MluNI
>DdeI
>Sau3AI
>EaeI
>AcsI
4710 4720 4730 4740 4750 4760 4770 4780 4790 4800
GGCTCTGCAC CGTATTGAAA CTGAGTTAAT GGGCAAATTT GATGAAGGTA AACTGCCCAC CGATCCACAC CTGATGCTCC GACTGGCCAT TGAAACTGTT


>ScrFI
>CfoI
>BsiYI
>MvaI
>CfoI
>MvnI
>EcoRII
>Tru9I
>RcaI
>BssHII
>DsaV
>AsnI
>ItaI
4810 4820 4830 4840 4850 4860 4870 4880 4890 4900
GCTCATGACT ATGATGTCAT AGTTATTGAC AGCGCGCCTA ACCTGGGTAT CGGCACGATT AATGTCGTAT GTGCTGCTGA TGTGCTGATT GTTCCCACGC
>Sau3AI
>DpnI
>NdeII
>SnaBI
>Sau3AI
>Psp1406I
>Tru9I
>MaeII
>DdeI
>PstI
>TaqI
>NdeII
>MaeII
>DpnI
>TaqI
>RsaI
4910 4920 4930 4940 4950 4960 4970 4980 4990 5000
CTGCTGAGTT GTTTGACTAC ACCTCCGCAC TGCAGTTTTT CGATATGCTT CGTGATCTGC TCAAGAACGT TGATCTTAAA GGGTTCGAGC CTGATGTACG
>ScrFI
>MvaI
>BsiYI
>FokI
>EcoRII
>DdeI
>DsaI
>AcsI
>DsaV
>FokI
5010 5020 5030 5040 5050 5060 5070 5080 5090 5100
TATTTTGCTT ACCAAATACA GCAATAGTAA TGGCTCTCAG TCCCCGTGGA TGGAGGAGCA AATTCGGGAT GCCTGGGGAA GCATGGTTCT AAAAAATGTT
>MspI
>HpaII
>BseAI
>MroI
>ScrFI
>DpnI
>NdeII
>MvaI
>Sau3AI
>Sau3AI
>DsaV
>MaeII
>XhoII
>BclI
>BanI
>RsaI
>FokI
>NdeII
>FokI
>HaeIII
>DpnI
>Ksp632I
>EcoRII
5110 5120 5130 5140 5150 5160 5170 5180 5190 5200
GTACGTGAAA CGGATGAAGT TGGTAAAGGT CAGATCCGGA TGAGAACTGT TTTTGAACAG GCCATTGATC AACGCTCTTC AACTGGTGCC TGGAGAAATG
>McrI
>PvuI
>DpnI
>Sau3AI
>NdeII
>AcsI
>TaqI
>Tru9I
>CfoI
5210 5220 5230 5240 5250 5260 5270 5280 5290 5300
CTCTTTCTAT TTGGGAACCT GTCTGCAATG AAATTTTCGA TCGTCTGATT AAACCACGCT GGGAGATTAG ATAATGAAGC GTGCGCCTGT TATTCCAAAA
>CfoI
>HpaII
>MvnI
>MspI
>ItaI
>MvnI
>PinAI
>PvuII
>HinfI
>CfoI
>Cfr10I
>TaqI
>AluI
>BsiYI
>Tru9I
>BssHII
>MvnI
5310 5320 5330 5340 5350 5360 5370 5380 5390 5400
CATACGCTCA ATACTCAACC GGTTGAAGAT ACTTCGTTAT CGACACCAGC TGCCCCGATG GTGGATTCGT TAATTGCGCG CGTAGGAGTA ATGGCTCGCG
>NciI
>ScrFI
>MspI
>HpaII
>DsaV
>AspHI
>AvaI
>BsiYI
>FokI
>BmyI
>BpuAI
5410 5420 5430 5440 5450 5460 5470 5480 5490 5500
GTAATGCCAT TACTTTGCCT GTATGTGGTC GGGATGTGAA GTTTACTCTT GAAGTGCTCC GGGGTGATAG TGTTGAGAAG ACCTCTCGGG TATGGTCAGG

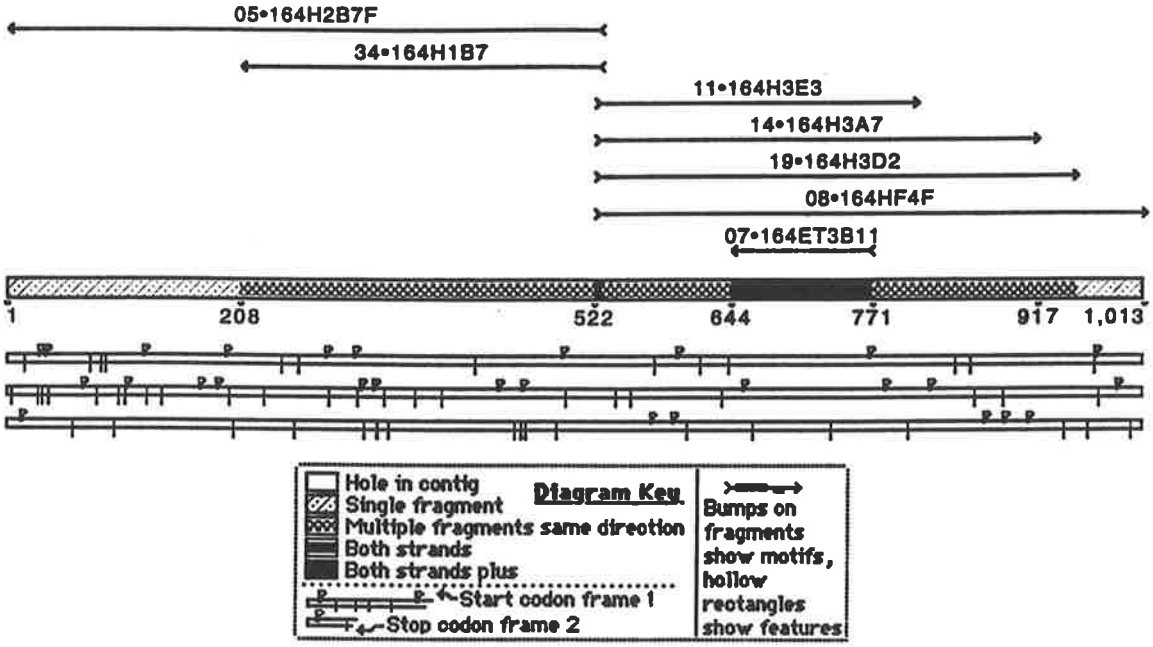

>MvaI >ItaI >DsaV >AluI >MaeIII >ScrFI >MaeII >EcoRII >DdeI >FokI >DpnI >NdeII >Sau3AI >FokI >MspI >HpaII >TaqI >McrI >SgrAI >BsiYI >HindII >CfrI0I >Ksp632I
5510 5520 5530 5540 5550 5560 5570 5580 5590 5600
TAATGAACGT GACCAGGAGC TGCTTACTGA GGACGCACTG GATGATCTCA TCCCTTCTTT TCTACTGACT GGTCAACAGA CACCGGCGTT CGGTCGAAGA
>ItaI >ItaI >HinfI >AluI >AluI
5610 5620 5630 5640 5650 5660 5670 5680 5690 5700
GTATCTGGTG TCATAGAAAT TGCCGATGGG AGTCGCCGTC GTAAAGCTGC TGCACTTACC GAAAGTGATT ATCGTGTTCT GGTTGGCGAG CTGGATGATG
>ItaI >FokI >MaeIII >MaeII >AcsI
5710 5720 5730 5740 5750 5760 5770 5780 5790 5800
AGCAGATGGC TGCATTATCC AGATTGGGTA ACGATTATCG CCCAACAAGT GCTTATGAAC GTGGTCAGCG TTATGCAAGC CGATTGCAGA ATGAATTTGC
>Asp700I >SspI >CfoI >SspI >MaeII
5810 5820 5830 5840 5850 5860 5870 5880 5890 5900
TGGAAATATT TCTGCGCTGG CTGATGCGGA AAATATTTCA CGTAAGATTA TTACCCGCTG TATCAACACC GCCAAATTGC CTAAATCAGT TGTTGCTCTT
>NciI >ScrFI >MspI >HpaII >DsaV >NciI >MspI >HpaII >ScrFI >DsaV >Tru9I >BfrI >ItaI >RcaI
5910 5920 5930 5940 5950 5960 5970 5980 5990 6000
TTTTCTCACC CCGGTGAACT ATCTGCCCGG TCAGGTGATG CACTTCAAAA AGCCTTTACA GATAAAGAGG AATTACTTAA GCAGCAGGCA TCTAACCTTC
>AluI >AluI >Tru9I >Tru9I >MaeII >MaeI >SpeI >AluI >Tru9I
6010 6020 6030 6040 6050 6060 6070 6080 6090 6100
ATGAGCAGAA AAAAGCTGGG GTGATATTTG AAGCTGAAGA AGTTATCACT CTTTTAACTT CTGTGCTTAA AACGTCATCT GCATCAAGAA CTAGTTTAAG
>BspWI >ScrFI >MvaI >DsaV >EcoRII >MvaI >EcoRII >DsaV >Tru9I >ScrFI >DdeI
6110 6120 6130 6140 6150 6160 6170 6180 6190 6200
CTCACGACAT CAGTTTGCTC CTGGAGCGAC AGTATTGTAT AAGGGCGATA AAATGGTGCT TAACCTGGAC AGGTCTCGTG TTCCAACTGA GTGTATAGAG
>Tru9I >HaeIII >BfrI >MaeII >AccI >MaeII >Tru9I >MaeIII
6210 6220 6230 6240 6250 6260 6270 6280 6290 6300
AAAATTGAGG CCATTCTTAA GGAACTTGAA AAGCCAGCAC CCTGATGCGA CCACGTTTTA GTCTACGTTT ATCTGTCTTT ACTTAATGTC CTTTGTTACA
>BmyI >ApaI >BanII >HaeIII >HaeIII >Sau96I >HaeIII >BspWI >SspI >Sau96I >DsaI >Sau96I >Sau96I >AvaII >AvaII
6310 6320 6330 6340 6350 6360 6370 6380 6390 6400
GGCCAGAAAG CATAACTGGC CTGAATATTC TCTCTGGGCC CACTGTTCCA CTTGTATCGT CGGTCTGATA ATCAGACTGG GACCACGGTC CCACTCGTAT
>DsaI >AvaII >Sau96I >Sau96I >AvaII >DsaI >AvaII >Sau96I >Sau96I >AvaII
6410 6420 6430 6440 6450 6460 6470 6480 6490 6500
CGTCGGTCTG ATTATTAGTC TGGGACCACG GTCCCACTCG TATCGTCGGT CTGATTATTA GTCTGGGACC ACGGTCCCAC TCGTATCGTC GGTCTGATAA


TCAGACTGGG ACCACGGTCC CACTCGTATC GTCGGTCTGA TTATTAGTCT GGGACCATGG TCCCACTCGT ATCGTCGGTC TGATTATTAG TCTGGGACCA
CGGTCCCACT CGTATCGTCG GTCTGATTAT TAGTCTGGAA CCACGGTCCC ACTCGTATCG TCGGTCTGAT TATTAGTCTG GGACCACGGT CCCACTCGTA
TCGTCGGTCT GATTATTAGT CTGGGACCAC GATCCCACTC GTGTTGTCGG TCTGATTATC GGTCTGGGAC CACGGTCCCA CTTGTATTGT CGATCAGACT
ATCAGCGTGA GACTACGATT CCATCAATGC CTGTCAAGGG CAAGTATTGA CATGTCGTCG TAACCTGTAG AACGGAGTAA CCTCGGTGTG CGGTTGTATG
CCTGCTGTGG ATTGCTGCTG TGTCCTGCTT ATCCACAACA TTTTGCGCAC GGTTATGTGG ACAAAATACC TGGTTACCCA GGCCGTGCCG GCACGTTAAC
CGGGCTGCAT CCGATGCAAG TGTGTCGCTG TCGACGAGCT CGCGAGCTCG GACATGAGGT TGCCCCGTAT TCAGTGTCGC TGATTTGTAT TGTCTGAAGT
TGTTTTTACG TTAAGTTGAT GCAGATCAAT TAATACGATA CCTGCGTCAT AATTGATTAT TTGACGTGGT TTGATGGCCT CCACGCACGT TGTGATATGT
AGATGATAAT CATTATCACT TTACGGGTCC TTTCCGGTGA TCCGACAGGT TACGGGGCGG CGACCTCGCG GGTTTTCGCT ATTTATGAAA ATTTTCCGGT

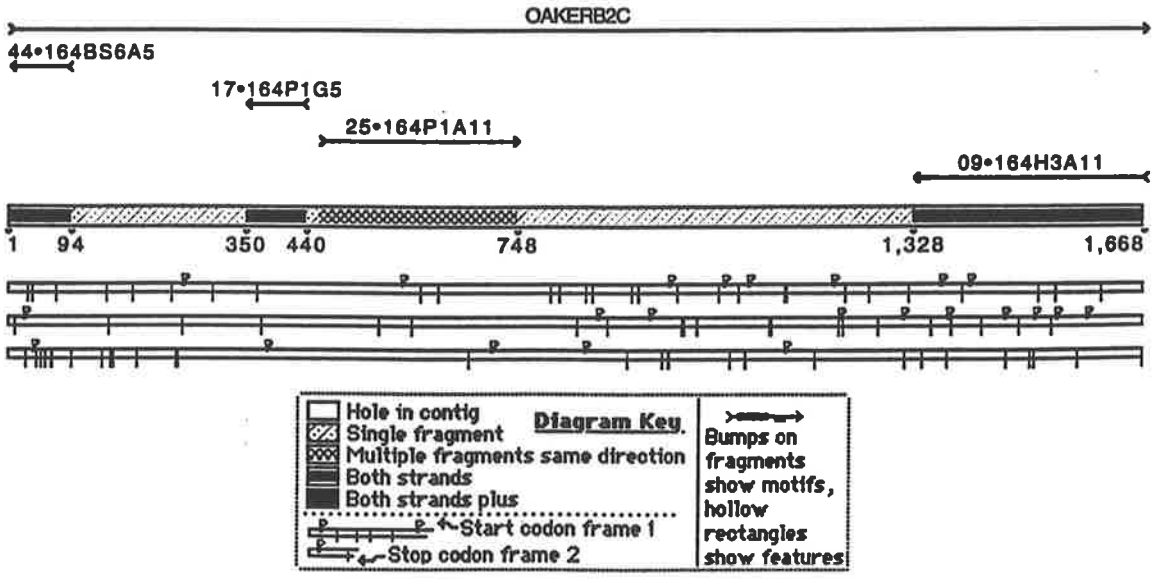

>BsiYI
>Tru9I
>Tru9I
>DraI
>Tru9I
>AluI
>HaeII
7310 7320 7330 7340 7350 7360 7370 7380 7390 7400
TTAAGGCGTT TCCGTTCTTC TTCGTCATAA CTTAATGTTT TTATTTAAAA TACCCTCTGA AAAGAAAGGA AACGACAGGT GCTGAAAGCG AGCTTTTTGG
>BmyI
>AspHI
>SacI
>BanII
>AluI
>DsaI
7410 7420 7430 7440 7450 7460 7470 7480 7490 7500
CCTCTGTCGT TTCCTTTCTC TGTTTTTGTC CGTGGAATGA ACAATGGAAG TCCGAGCTCA TCGCTAATAA CTTCGTATAG CATACATTAT ACGAAGTTAT
>TaqI
ATTCGAT

APPENDIX III: KAP CONTIGS AND SEQUENCE ALIGNMENTS

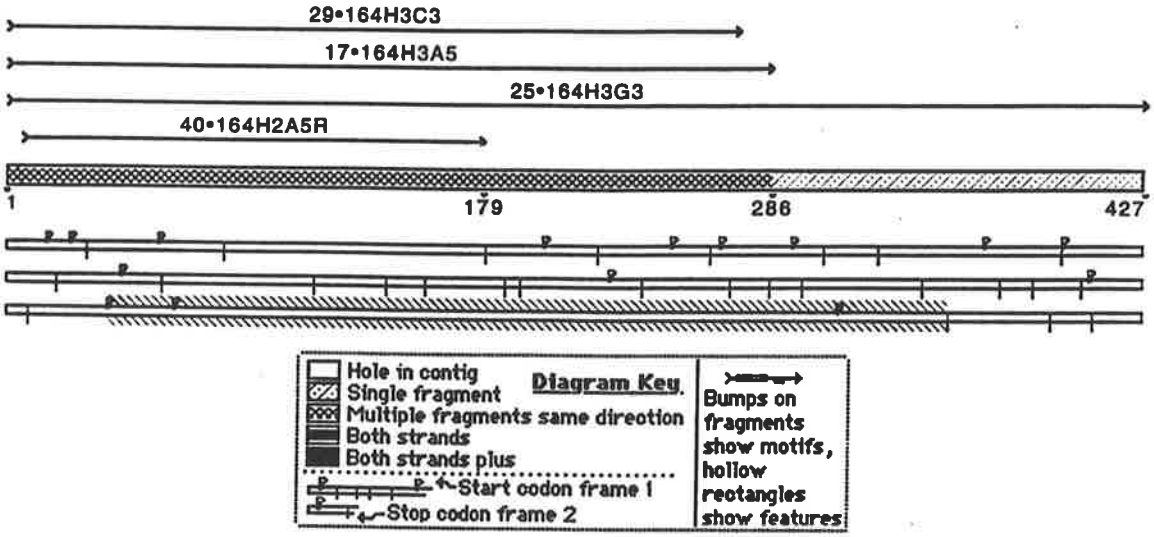
Contig[0167]
Sequencher™ "164 MAIN PROJECT"



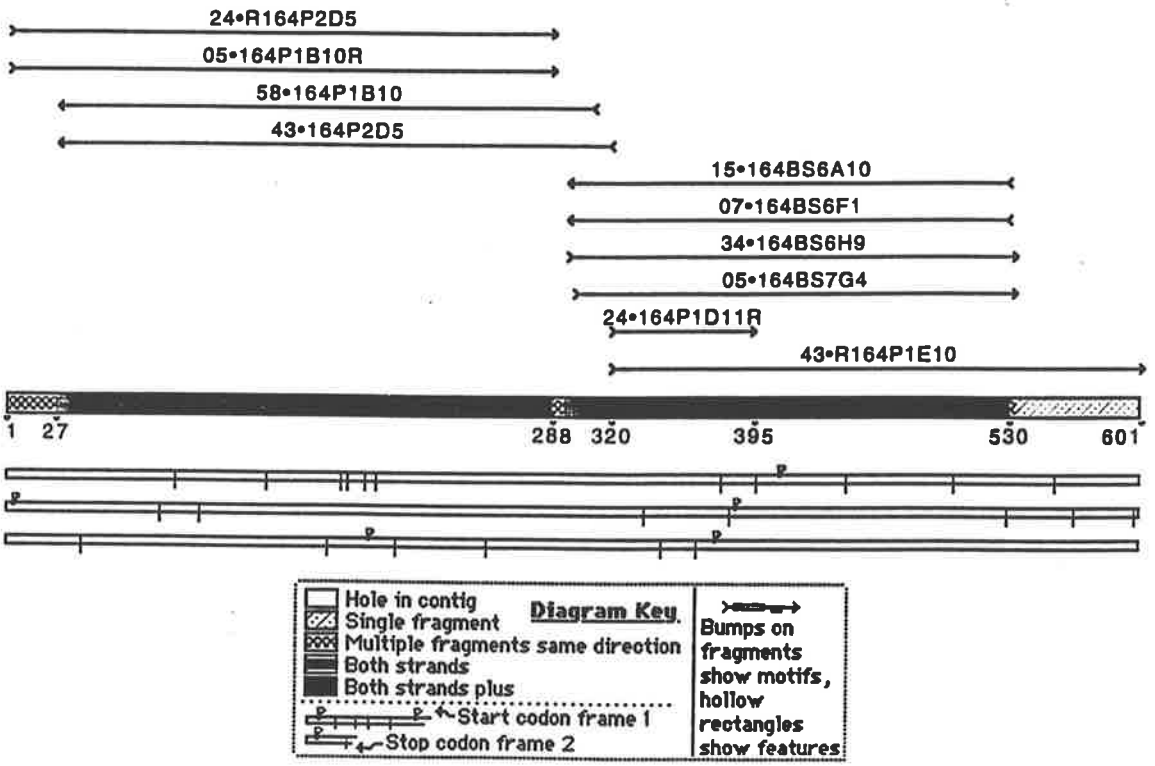
Contig[0169]
Sequencher™ "164 MAIN PROJECT"



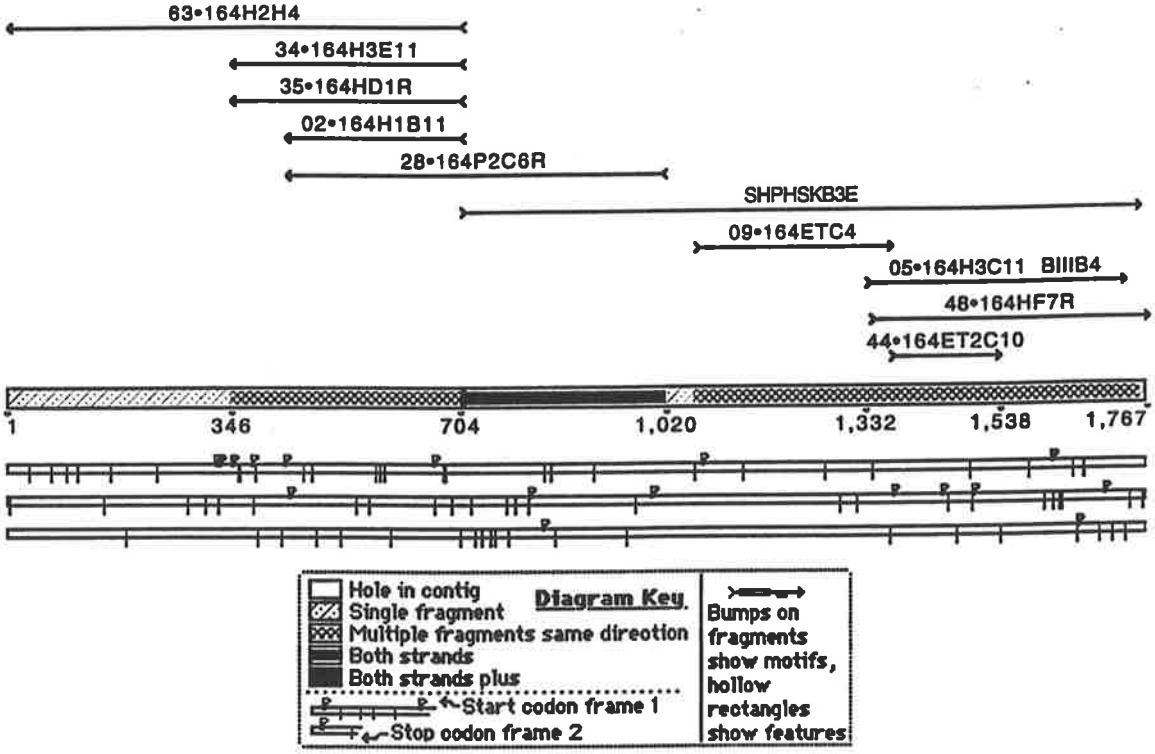
Contig[AA]
Sequencher™ "164 MAIN PROJECT"



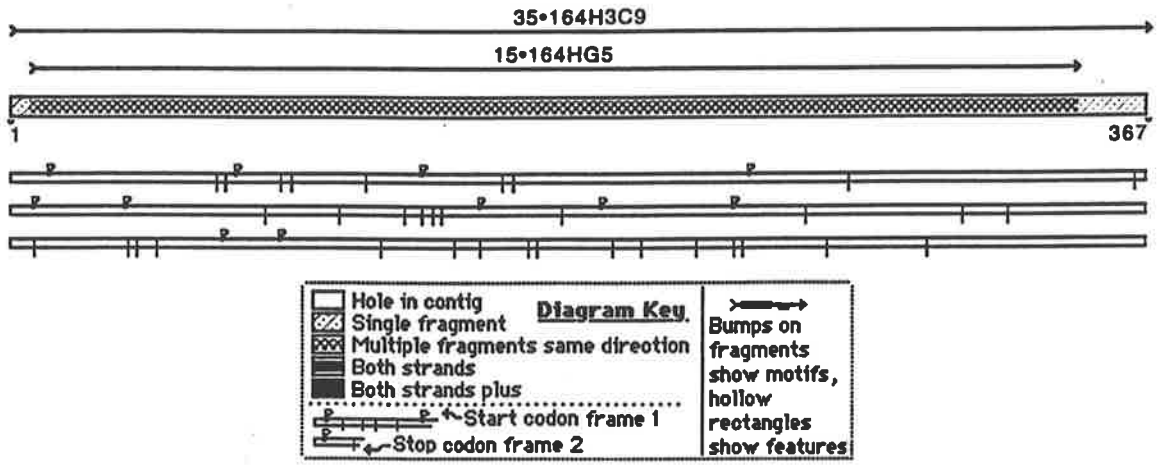
Contig[F]
Sequencher™ "164 MAIN PROJECT"



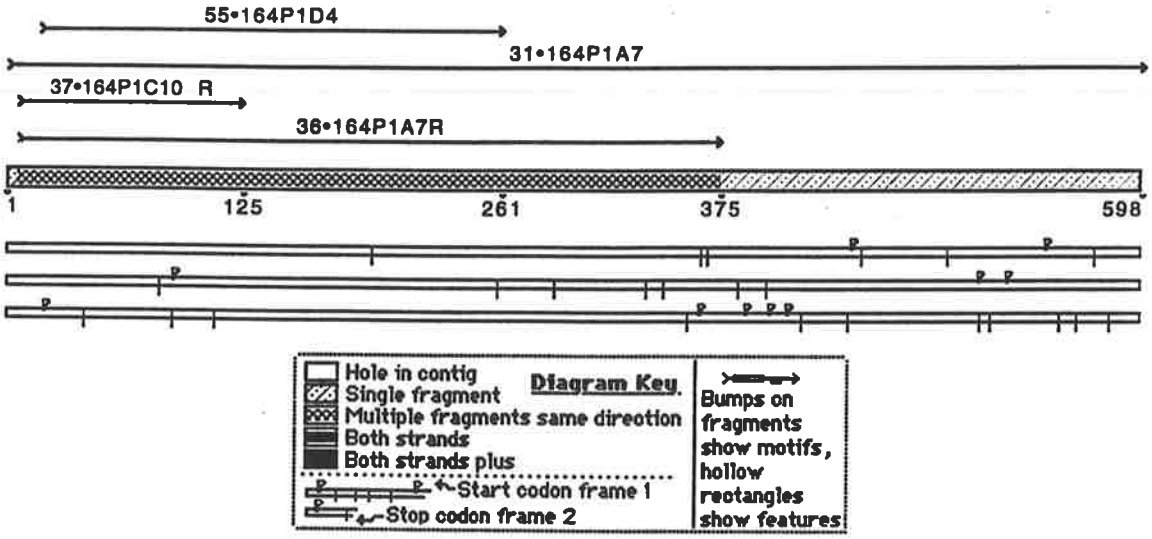
Contig[G]
Sequencher™ "164 MAIN PROJECT"



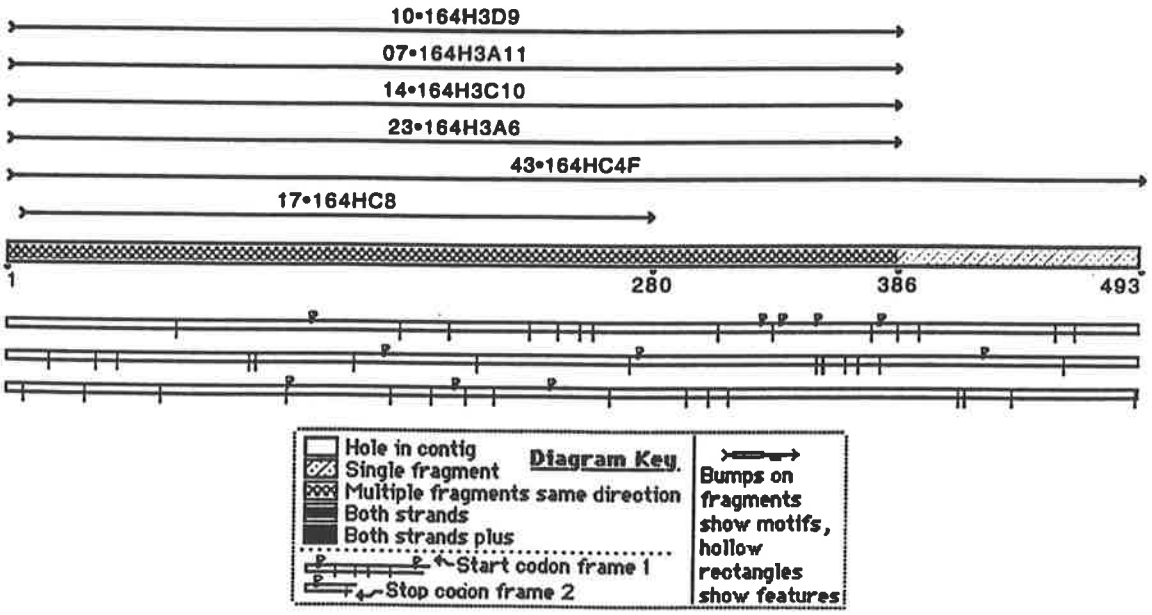
Contig[GG]
Sequencher™ "164 MAIN PROJECT"



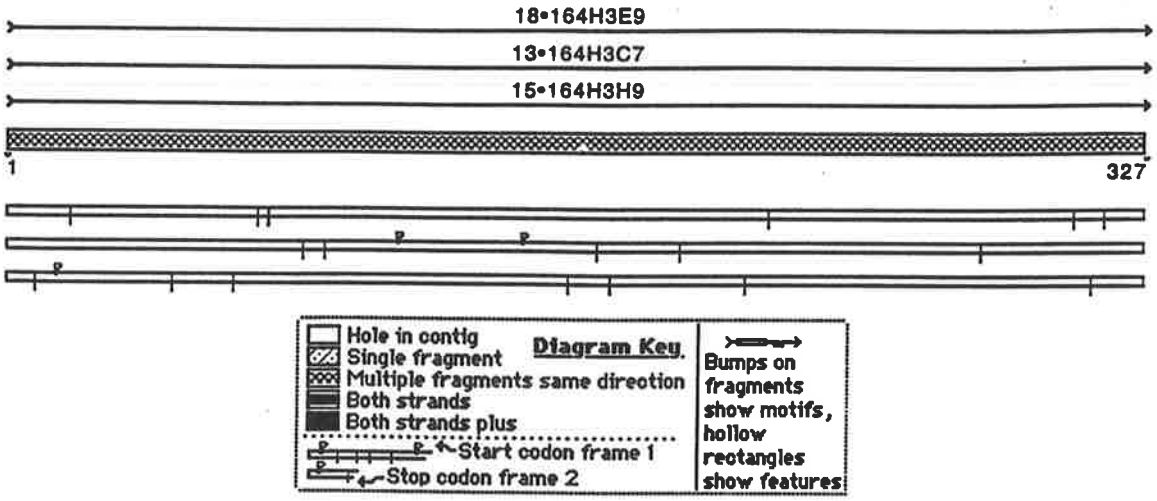
Contig[JJ]
Sequencher™ "164 MAIN PROJECT"



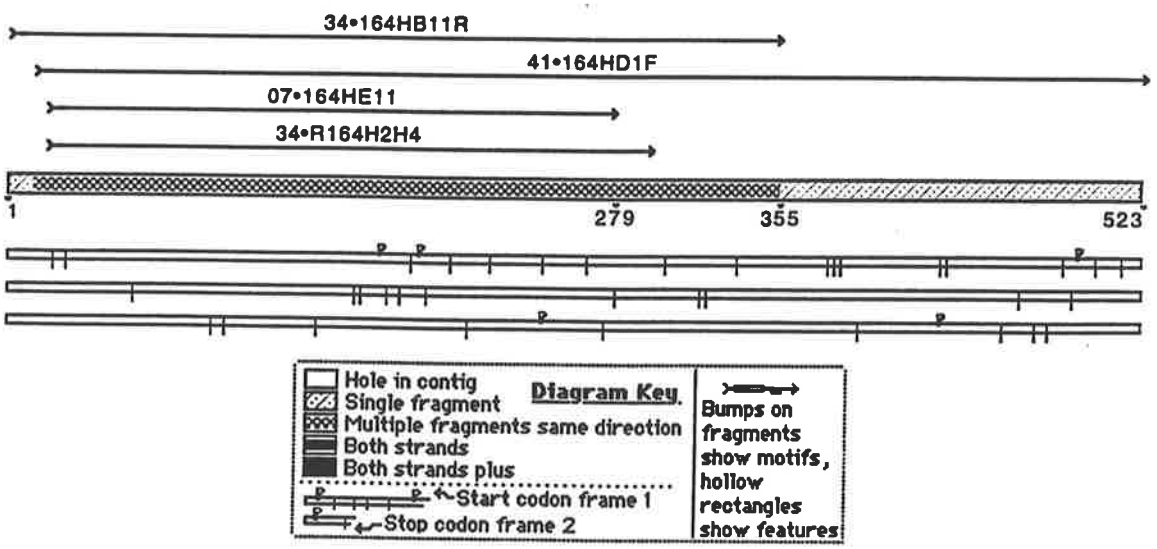
Contig[J]
Sequencher™ "164 MAIN PROJECT"



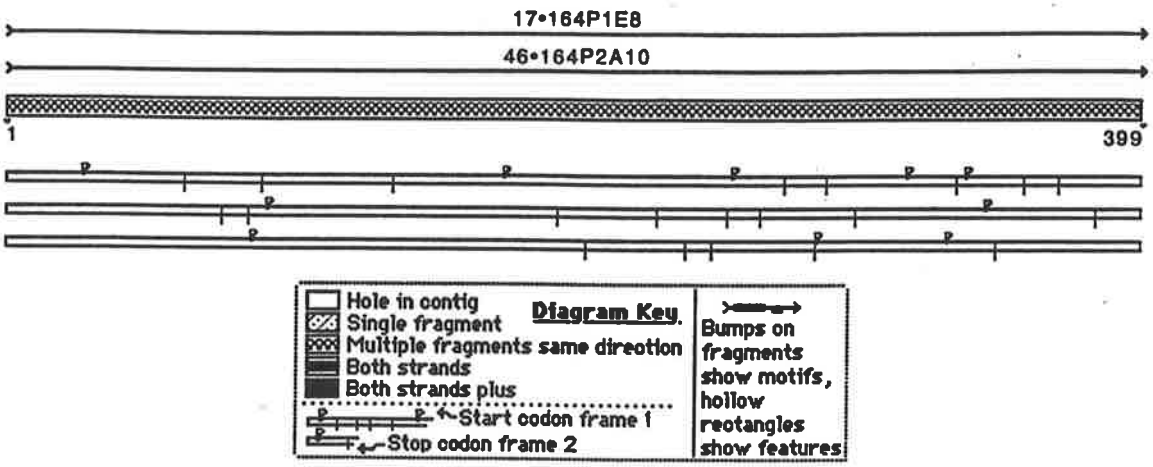
Contig[KK]
Sequencher™ "164 MAIN PROJECT"



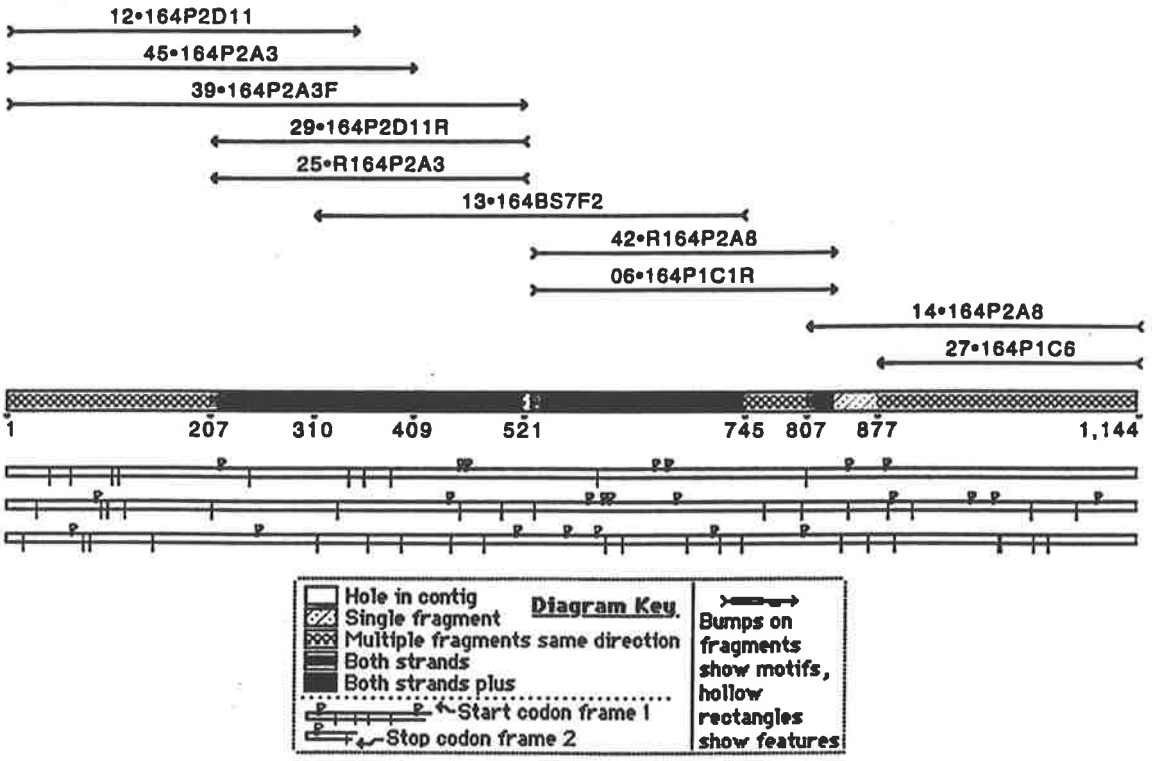
Contig[L]
Sequencher™ "164 MAIN PROJECT"



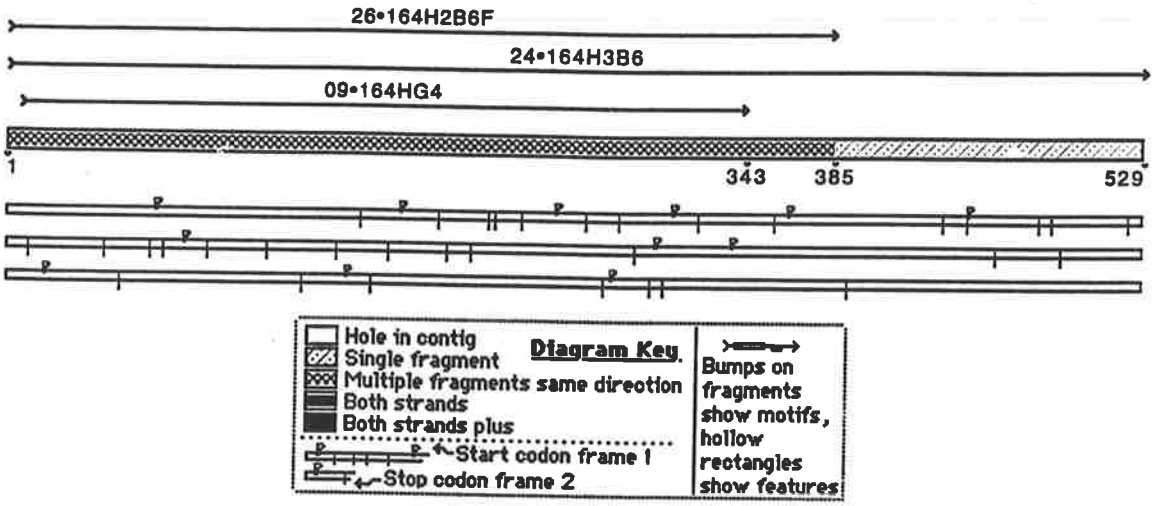
Contig[OO]
Sequencher™ "164 MAIN PROJECT"



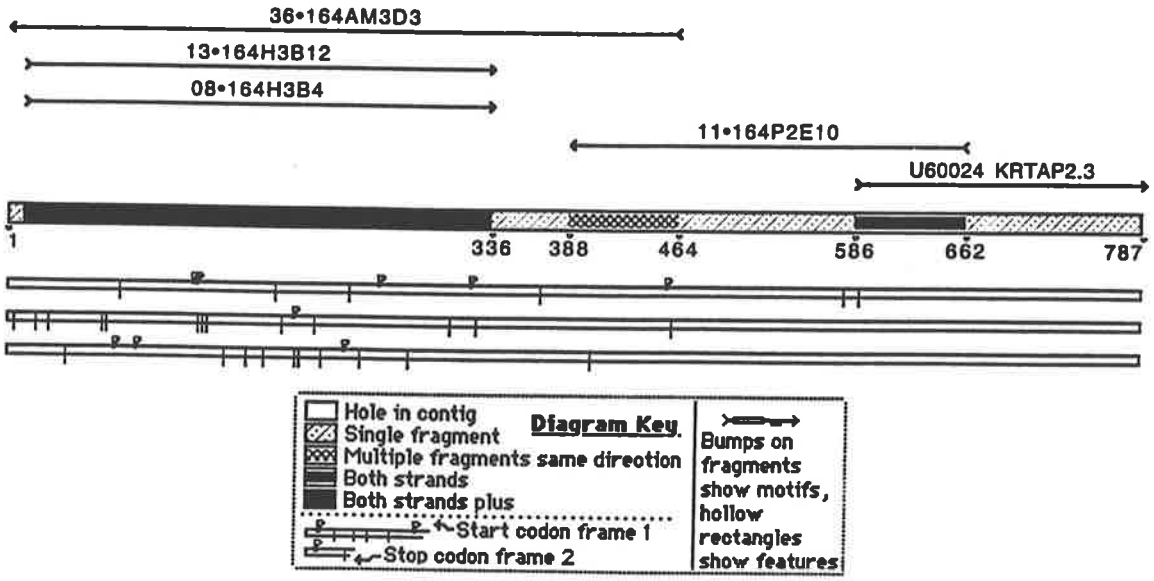
Contig[PP]
Sequencher™ "164 MAIN PROJECT"



Contig[P]
Sequencher™ "164 MAIN PROJECT"



Contig[R] KAP2.3
Sequencher™ "164 MAIN PROJECT"



Score = 46.0 bits (28), Expect = 0.002
Identities = 29/30 (96%), Positives = 29/30 (96%)

```
164ET2A7:  37    aatttcaaaactttaaaaggaatgtaaatt 66
                  |||||||||||||||| ||||||||||||
AC7455:      1491 aatttcaaaactttaagaggaatgtaaatt 1462
```

Score = 142 bits (89), Expect = 1e-32
Identities = 144/196 (73%), Positives = 144/196 (73%), Gaps = 1/196 (0%)

```
164HB1:  10    aacaaataactcagggacaggccaattaccattcagagacaagttgttttaaagcaaaca 69
               ||||||||| ||||||| |||||||||| |||||||||||| || |   | | |||| |
AC7455:  15190 aacaaataattcagggagaggccaattatcattcagagacaaattatccaaca-caaata 15248

164HB1:  70    aggaagagacattaagaattatgcaaatagcagccagtagaggcttataaaaggcccact 129
               || ||||||| |   |||||||| ||| || |||  |   ||||||||||||||||||| |
AC7455:  15249 agcaagagacttctggaattatgtaaacagtagctggcccaggcttataaaaggcccaat 15308

164HB1:  130   gcggaagtgtccatcaaagctcaaaaacttctcttaacaaccctcctctcaatctaaccc 189
               | || ||    || |||| |||| ||||| |||  | ||||||     |||  | | | |
AC7455:  15309 gtggcagccatcaccaaaactcagaaactcctccaagcaacccagacttcataccagctc 15368

164HB1:  190   ctgacactatggcctg 205
               |  |||| ||| ||||
AC7455:  15369 ccaacaccatgacctg 15384
```

Score = 120 bits (75), Expect = 8e-26
Identities = 144/198 (72%), Positives = 144/198 (72%), Gaps = 16/198 (8%)

```
164BS3F11:  91    taagaactcttagcata-gcgttgaaatataaggagaattactaaaatcaaaaccaactt 149
                  |||||||| || ||||| | |||||  | ||| |||||| |||||| ||||| ||| ||
AC7455:       64469 taagaacttttgcataagtgttgatgtgtaaagagaatcactaaattcaaacccagctc 64528

164BS3F11:  150   tgctcaatgaataataaattctaagagactattttaatcctctatcttaag-tctccagg 208
                  ||| ||  || |||||||| |||||||||||||||||  | ||  | || | ||| | ||
AC7455:       64529 tgcccagggagtaataaatcctaagagactattttaactcacttacatatgttcttctgg 64588

164BS3F11:  209   tttggttttggttggttggttgttttcccattatcagctatttaaggaagaagtactcaa 268
                  |||||||| | ||||         || | | |||||| ||  | |||||| || ||||||
AC7455:       64589 tttggtttggtttgg---------tttctactatcaggtacct-aggaagtagcactcaa 64638

164BS3F11:  269   ga----gttaaaaaataa 282
                  ||    |  |||||||||
AC7455:       64639 gattaggagaaaaaataa 64656
```

```
Score =  156 bits (98), Expect = 1e-36
Identities = 189/262 (72%), Positives = 189/262 (72%), Gaps = 36/262 (13%)

164BS3F12:  1     gatcaggctccttcagcccacctgctgtgatacctgccacccaccctgctgtgtgcctga 60
                   |||||| |||||||||||||||||||||| ||||| | ||            ||||||
AC7455:       45563 gatcagcctccttcagcccacctgctgtga-acctggcccc-----------tgcctg- 45608


164BS3F12:  61    ctcctgcgtgcccgactcctacgtgccaacctgttggctgctcaacagctgccaccctac 120
                      |||| |||| |||||||| |||||| ||||||| ||||||||||  ||||||||  |
AC7455:       45609 --gctgcatgcctgactcctatgtgccatcctgttgactgctcaacaaatgccacccagc 45666


164BS3F12:  121   tccaaacctgagcggaatctctgtcacaacctgcgtccagccctgtgagggtgaagcaaa 180
                   ||||| |||||||||  |||||||||| |||||| ||||       ||| ||||| ||
AC7455:       45667 tccaaccctgagcgggctctctgtcaccacctgcatcca-------gagtgtgaaccacc 45719


164BS3F12:  181   atcctactagccaaagacaatctcaatgagctttgcccgcgttgccctaagggactgcaa 240
                    | || |||||||||            |||| |||||| | || |||| |||  || ||
AC7455:       45720 ttgctgctagccaaa-----------gagc-ttgcccacattaccctgaggaccttcag 45766


164BS3F12:  241   ctgtcatcaagagctgctcagc 262
                     ||||| ||| ||||||||||
AC7455:       45767 tagtcattaagcgctgctcagc 45788

Score =  114 bits (71), Expect = 1e-23
Identities = 126/175 (72%), Positives = 126/175 (72%), Gaps = 18/175 (10%)

164BS3F12:  1     gatcaggctccttcagcccacctgctgtgatacctgccacccaccctgctgtgtgcctga 60
                  |||||| |||||||||||| ||||||||| ||||||| |||||||||
AC7455: 41240 gatcagcctccttcagcccatctgctgtgacacctgccccccaccctg---------- 41287


164BS3F12:  61    ctcctgcgtgcccgactcctacgtgccaacctgttggctgctcaacagctgccaccctac 120
                      ||||  ||| ||  |||| |||||||| || |||||||||||| ||| ||||| ||
AC7455:       41288 ---ctgcaagcctgatacctatgtgccaacttgctggctgctcaacaactgtcacccgac 41344


164BS3F12:  121   tccaaacctgagcggaatctctgtcacaacctgcgtccagcc---ctgtgagggt 172
                   |||    ||||| || |||    | || ||||  || |||||   ||||||| ||
AC7455:       41345 tcccggactgagtgggatcaacctgaccacctatgttcagcctggctgtgagagt 41399

Score =  110 bits (69), Expect = 1e-22
Identities = 138/183 (75%), Positives = 138/183 (75%), Gaps = 20/183 (10%)

164BS3F12:  282   ttttgaacaatgatttcaaat-agaagtgg---aaaaaatgctatcatgggccctcttaa 337
                   |||||||||||||||||||| |||| |||   |||||||||||| |||||  || ||||
AC7455:       45865 ttttgaacaatgatttcaaatgagaaatgggaaaaaaaatgctaccatggatcccttac 45924


164BS3F12:  338   tt------taaataaacaccactctttcccaaaccatggggctgactcaatgactaatgc 391
                   ||      ||||||||||||||  | |||||||||     ||| |||||    ||||||
AC7455:       45925 ttcaaaactaaataaacaccactactccccaaacca-catactgcctcaacatttaatgc 45983


164BS3F12:  392   cataataacc---cgactc-tgacacatgtatgcatcatggaacccataaatatgtccag 447
                    ||||| |||   |||||| ||||||    ||||||||| |||| | |||||  |||||
AC7455:       45984 tataattacccaacgactcttgacac----atgcatcat-gaacttaaaaatacatccag 46038
```

```
164BS3F12:  448   cag 450
                  |||
AC7455:      46039 cag 46041

Score = 72.9 bits (45), Expect = 3e-11
Identities = 104/160 (65%), Positives = 104/160 (65%), Gaps = 15/160 (9%)

164BS3F12:  6     ggctccttcagcccacctgctgtgatacctgccacccaccctgctgtgtgcctgactcct 65
                  || | ||  |||||||||||||||| | ||| | ||||||||||    | ||| |  |||
AC7455:      56222 ggttactggagcccacctgctgtgacaactgtcccccaccctgccacattcctcagccct 56281


164BS3F12:  66    gcgtgcccgactcctacgtgccaacctgttggctgctcaacagctgccaccctactccaa 125
                  ||||||||              ||||| |  |||||||||  |||||| || ||||||
AC7455:      56282 gcgtgccc---------------acctgcttcctgctcaactcctgccagccaactccag 56326


164BS3F12:  126   acctgagcggaatctctgtcacaacctgcgtccagccctg 165
                   ||||       ||    |||| |||| |   ||||||||
AC7455:      56327 gcctggagaccctcaacctcaccaccttcactcagccctg 56366

Score = 66.6 bits (41), Expect = 2e-09
Identities = 102/160 (63%), Positives = 102/160 (63%), Gaps = 15/160 (9%)

164BS3F12:  6     ggctccttcagcccacctgctgtgatacctgccacccaccctgctgtgtgcctgactcct 65
                  || | ||  |||||| |||||||||| | ||| | ||||||||||    | ||| |  |||
AC7455:      50466 ggttactggagcccatctgctgtgacaactgtcccccaccctgccacattcctcagccct 50525


164BS3F12:  66    gcgtgcccgactcctacgtgccaacctgttggctgctcaacagctgccaccctactccaa 125
                  ||||||||              |||||  |  |||||||||  |||||| || |||||
AC7455:      50526 gcgtgccc---------------acctgcttcctgctcaactcctgccagccaactccgg 50570


164BS3F12:  126   acctgagcggaatctctgtcacaacctgcgtccagccctg 165
                   ||||       ||    |||| |||| |   ||||||||
AC7455:      50571 gcctggagaccctcaacctcaccaccttcactcagccctg 50610


Score =  207 bits (130), Expect = 6e-52
Identities = 237/333 (71%), Positives = 237/333 (71%), Gaps = 18/333 (5%)

R164P2D8:  2     cagcaacntgcccaagagaaacacgaccacacacagctctctttatccaggagaaagtac 61
             ||||||  || | ||||||||   ||| | |    |||||||||| ||||||||||||
AC7455: 44534 cagcaatatgtcaaagagaaatgtgactataggtggctctctttattcaggagaaagtat 44593


R164P2D8:  62    atgctacttgacaatacatgctctacttgacaacacatgcttctagattgtctacttggt 121
                 |||  |||||||||| || |||||||         || | | |  ||||| |  || ||
AC7455:      44594 atgacacttgacaatatattctctact---------atccctgtcaattgtatcttttgt 44644


R164P2D8:  122   tgtatatggcaacagttttctctgtggttgttatagcttaagataaaatgagtatacaca 181
                 ||| |.       | ||||||||||||||  ||||||  ||| |||||||||| || |||
AC7455:      44645 tgtttttct----ctgttttctctgtggtgattatagtctaaaataaaatgagtgtataca 44700


R164P2D8:  182   atgagtctaggaaaaactgctcatgtttcaaaggtgaaaacaaatcatgtttctttaagt 241
                 |||||| |   ||||  |   | ||||||  ||| | | ||||||  | | |||    ||||
AC7455:      44701 atgagtattaaaaaaggtatccgtgtttctaagttaagaacaaa-ccttttttaaaaagt 44759
```

```
R164P2D8:  242   cagactttaaataccaatgagcaacatcagcacaggaaaaatacagactggttcattaac 301
                 |||||||    |||  || ||||   |||||| || | |||||  | |||  ||||||||
AC7455:      44760 cagactt----tactgattagcagtgtcagcagagaagaaatattggctgcatcattaac 44815

R164P2D8:  302   cttgcattaaaatgtagaatgtttggcaactca 334
                   ||| ||| ||||| ||||||||||||||||||
AC7455:      44816 tttgaattgaaatgcagaatgtttggcaactca 44848
```

Score = 188 bits (118), Expect = 2e-46
Identities = 165/209 (78%), Positives = 165/209 (78%), Gaps = 2/209 (0%)

```
164BSG9:  6     tgaatcctagacaggtaagatgatttaccttaagtagcctggtgaccaacttgttaagaa 65
                ||| || |||| |||| |||||||||| |||||  | |   ||   |||   |||||| ||
AC7455:   53203 tgattcatagagaggtgagatgatttgccttagatggttcgggagccagtgtgttaaaaa 53144

164BSG9:  66    ctaaatataggacctactgggactgagttacctttttatgtggtttaaaatgtaagtttag 125
                |||||||||||||||| ||||| |||| ||||||||||||||||  ||||| |||||||
AC7455:   53143 ctaaatataggacctattgggattgagctacctttttatgtggttagaaatggaagtttaa 53084

164BSG9:  126   agtatgtgttagagatattaaacatactttatttttagaaagaaaaggtacccttttcct 185
                | ||||||| ||||||| ||| |||  |||||||||| ||| ||   ||| | |||| ||
AC7455:   53083 actatgtgtcagagatactaagcat--tttatttttaaaaacaacgcgtatctttttttct 53026

164BSG9:  186   tcaagcaagtgcaggaaaaacatggtgat 214
                | |||  | || ||||||||||||||||
AC7455:   53025 ttaagtgaatgtaggaaaaacatggtgat 52997
```

```
Score =  131 bits (82), Expect = 2e-29
Identities = 147/194 (75%), Positives = 147/194 (75%), Gaps = 17/194 (8%)

164ET2G1:  4     tttttgctgatcttttaggaaaataacaaaaaactctgacttcccaaagttcccca-gca 62
                 |||||||||||||||||||||||||||||  ||| |||||||| |  |||||  ||| ||
AC7455:    85472 tttttgctgatcttttaggaaaataacaggaaa-tctgactttctgaagttttccaagct 85530

164ET2G1:  63    tacc-agcatagctatgaa----------aagcctaactagcctccccctgttaaccaa- 110
                 |    ||||||||||||||          ||| |||  |||||| |||||||||    |
AC7455:    85531 ttgtgagcatagctatgaagattaataaaaagtctacttagccttcccctgttacttgac 85590

164ET2G1:  111   acacacataagccccttggaggcttttaaaggaggatatcttggtgtttgtctcatttgt 170
                 ||||||| || ||||||||    | |||||| |||| ||||||||||||  ||||||||
AC7455:    85591 acacacacaaaccccttggt---tcttaaagaaggaaatcttggtgtttgcttcatttgt 85647

164ET2G1:  171   gtgtgcaacttctt 184
                 |||| |||  ||||
AC7455:    85648 gtgtacaatctctt 85661
```

```
Score = 1426 bits (899), Expect = 0.0
Identities = 1732/2453 (70%), Positives = 1732/2453 (70%), Gaps = 188/2453 (7%)

Contig3:  2147  aataaatgcacacacatctggaacaaaaaatacc---tcttaaaatttaatttttttcagt 2203
                || ||||  ||||| | || |||||||| | | |   | |||  ||||||||||  | |
AC7455:   21851 aaaaaattaacaca-acctagaacaaaacaaaacatattttaccatttaatttttaaatt 21909


Contig3:  2204  --tc-ttaagtatgtgagcacaacatacaaaacacaatagga-gaaaacgtagaacacat 2259
                  || ||| | ||  |||   || | |||||| ||||||||| | |||  |||||
AC7455:   21910 cttcattaggcattagagtgtaatacacaaaatacaataggatgcaaatatagaa----- 21964


Contig3:  2260  tgagaacaatatttttgagacaatt-acaagctcccggcactcacctctatcattcagct 2318
                  ||||  |  ||||  | |  ||| |||||||||| | ||| |||||||| ||| |  |
AC7455:   21965 --agaaacacgttttccacattatttacaagctcccagaacttacctctat-atttaatt 22021


Contig3:  2319  ggtggcagtgataggtcacacacaagattactta-tttcagagaagaataaagtagata- 2376
                ||| ||| |||| | | |||||||| ||||   | ||||  ||| ||||||||||||
AC7455:   22022 ggtagcaatgatggattacacacaaaattatgcaatttccaagaggaataaagtagatgg 22081


Contig3:  2377  cagtgcacccttcaaaatgtcagagcttatcgatgctttttcttcagacttagttcaact 2436
                |||| ||   |||||| ||||| | ||  | |||  ||||||   || |   |||| |||
AC7455:   22082 cagtacatttttcaaagtgtcataactcctagatatttttcccgagcccg-gttccact 22140


Contig3:  2437  attaaatacaaaaag-----------atgtcag------cttctcctgttccctttga- 2478
                | ||||| ||||||           |   ||       | | ||||  || ||||  |
AC7455:   22141 actaaatccaaaaaataaaataaaataaaccaaaaaacacatttcctagtctctttacag 22200


Contig3:  2479  caatcataaggactgcttcatgagaggcactacagcttaaaagttctatttaatcattta 2538
                |||  |  | ||||  ||||||||| | |||||||||| |||||||| ||||||||| |
AC7455:   22201 caaataacaagactctttcatgagaagtactacagcttgaaagttctctttaatcatgtt 22260


Contig3:  2539  taagtggtatagacactcatgaatcatgcctttttactccacccattgaa----------- 2587
                | |||||||| || |  ||||||||| |||||||||||||| ||| ||
AC7455:   22261 tgagtggtattgatatccatgaatcaagccttttactccacctattcaataaacatttat 22320


Contig3:  2588  --aatcagnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnngtggtg 2645
                  ||||||                                                |||
AC7455:   22321 tgaatcagattcagtaaacatttattaaggttctaccacaaactaaggactctgctggat 22380


Contig3:  2646  atctttctgtatactatatca-tttggttcacaaggaaatatttctggcctggaacacaa 2704
                | ||| ||  |||||||| ||| |||||||| |||| | | |||||||| |||||||||  |
AC7455:   22381 accttcctaaatactatgtcagtttggttcccaagaacagatttctggtctggaacagga 22440


Contig3:  2705  gtaatatttacacatttcaaattggttgaaatcaaaggaaaaatgat-ggtgatttattt 2763
                 ||||| | ||||||| |||||||||||||||||||||||||  || || ||| ||| | |
AC7455:   22441 ataatagtcacacattccaaattggttgaaatcaaaggaactataatcggtaattcaata 22500


Contig3:  2764  -gatgttggaataaatga-gttggaagcttttaattatgtgtgaccaaaggtac-ctttc 2820
                 ||  |||||| |||  | | || || ||   |||||||| ||||||||||||  | |||
AC7455:   22501 agacattggaacaaaaaaaggtgaaatctgcgaattatgtatgaccaaaggtattccttc 22560
```

```
Contig3:  2821  aattcagtggatgtcttgatgttctaaattcaagagaattcagtctatatattcaagagg 2880
AC7455:  22561  agttcagtagatatccagatatccaaaattcaataaaatttagtctatatatctaggagg 22620

Contig3:  2881  gggnnnnnnntagctaaa---aacagcacaatta---ctcttagccttagagacctggac 2934
AC7455:  22621  aa-aaaataatagttaaatataacagcacaattattactcttagccttagaggtatggat 22679

Contig3:  2935  agagaatcacatggaataactacctctacaaaatgagggtgttgacatggacaaaaatgg 2994
AC7455:  22680  agattgtcacgtggaataactat-tatgtaaaatgagagcattcatttggactaaaatgg 22738

Contig3:  2995  gagggctctttgaacatggaaaaccacaaacctattccaaagagcaaaggaccaatatga 3054
AC7455:  22739  aagtactctttgaacatgggaaaccacaaacatattccaaagagcaaaggactgatatga 22798

Contig3:  3055  ggagtccaaatgaaactataaaacgaagtaatttctaaactgctgctcataatgtatcat 3114
AC7455:  22799  ggagtccagaagaaactataaaac--agtaatttctaaactattgctcataatgtgtcac 22856

Contig3:  3115  aggtatttcaatgtaagccaatcagaggacccactgaaaaacaaatttatacaaaaacaa 3174
AC7455:  22857  aagtatttcaacataagccaatcagag-atccacggagtaacaaatttaaacaaaaacaa 22915

Contig3:  3175  ttaagatatgccaagaagtctataaatagcagctagcgcagatatataaaaggcctgaca 3234
AC7455:  22916  agaaaatatgttaacaattttgtaaacagcagctggttcagatatataaaaggaccaacg 22975

Contig3:  3235  cagaagtcaccacccaa--------aatcctccaagcattacaattctcagcccaactcc 3286
AC7455:  22976  tggaagtcaccaccaaacctcagaaactcctccaagcaacctaactcttaacccaacttc 23035

Contig3:  3287  tgacaccatggcctgttgctccaccagcttctgtggatttcccacttgctccactggtgg 3346
AC7455:  23036  tgacaccatgacctgctgccagaccagcttctgtggatatcccagcttctccatcagtgg 23095

Contig3:  3347  gacctgtggttccaacttttgccagccaacctgctgccagaccagctgctgccagccaac 3406
AC7455:  23096  gacctgtggctccagctgctgccagccaagctgctgtgagaccagctgctgccagccacg 23155

Contig3:  3407  ctccattcagaccagctgctgc--------cagc----cgacctcta-----tccaga-c 3448
AC7455:  23156  cagctgccagactagcttctgcggatttcccagcttctcaaccagtgggacctgcagctc 23215

Contig3:  3449  cagctgctgccagccaacttccatccaaaccagctgctgccaaccgatctccatccagac 3508
AC7455:  23216  cagttgctgccagccaagctgctgtgagaccagctgctgccagccaagctgctgtgagac 23275

Contig3:  3509  cagctgctgccagccaacctgcctccagaccagtggctgtgagacgggctgtggcattgg 3568
AC7455:  23276  cagctgctgccagccaagctgctgccagatcagctcctgcggaactggctgtggcattgg 23335
```

```
Contig3:  3569  tggcagcattggctatggccaagtgggtagcagcggagctgtgagcagccgcaccaagtg 3628
AC7455:  23336 tggtggcatcagctatggccaggagggcagcagtggagctgtgagcacccgtatcaggtg 23395

Contig3:  3629  gtgccgccctgactgccgcgtggagggcaccagcctgcctccctgctgtgtggtgagctg 3688
AC7455:  23396 gtgccgcccagacagtcgtgtggagggcacctacctacccccctgctgtgtggtg----- 23450

Contig3:  3689  cacatccccgtcctgctgccagctgtactatgcccaagcctcctgctgccgcc--catcc 3746
AC7455:  23451 --------------------agctgcac---gcccc--catcctgctgccaactgcacca 23485

Contig3:  3747  tactgtggacagtcctgctgccgcccagcctgctgctgccagcccacctgcattgagccc 3806
AC7455:  23486 tgcccaggcc--tcctgctgccgcccgtcctactgtggacagtcctgctgccgc---cca 23540

Contig3:  3807  gtctgtgagcccacctgctgtgagcccacctgctgaaagcaaggttgctcatttaaaatt 3866
AC7455:  23541 gtctg---------ctgctgtgagcccacttgctgaaagccagtttgcttattttcaatt 23591

Contig3:  3867  gcccaagacacagtatctctgaataatttatcgcctcaaccacccatggacagctaacaa 3926
AC7455:  23592 gcctaggtcacagtgtctctgaactgttcatc-ccttgaccacctctggaccactaacaa 23650

Contig3:  3927  gctcttagctcccatttgggttt-ttgttatgggcgctacagagtatatgag-------t 3978
AC7455:  23651 gttctcagac----tttgcattgcttgtgatggagactactaagtatatgagctcacaat 23706

Contig3:  3979  tctatctgatttcattctacaatgaatatctgtactttccacgg------caga--tgct 4030
AC7455:  23707 tctatctgattccattctacaatgaataccttgacccttcactggggacacagaaatgct 23766

Contig3:  4031  gcgta-ccacctg-tcatcatcaaattgctttggctatactatttctgattttgatgcaa 4088
AC7455:  23767 acaaagccacctgctgatcatcaatttgcttgggatatactatttctgatatttctgcag 23826

Contig3:  4089  ggttgaatattgctgacatattgtggaatttatccttttgaactgcgcacaagcctatt- 4147
AC7455:  23827 gattaaaaattactgacatgttgtggaatttatccatgagaactatccacaagtctaatg 23886

Contig3:  4148  tctcc-tgctttctgatctatttttagcttctgtttgtccccaaatttttacaacgtcaa 4206
AC7455:  23887 tttccatgctttataatctatttt-atctt--gtttacctaaaatttttttgcaacatcaa 23943

Contig3:  4207  aggcacctgagtatagatagtatatcttcag-atgtcaccaatgagaatagaagctcttc 4265
AC7455:  23944 agacaccaaattatagccaagtgacattcctcaagtcaccagagagaatggaagctcatc 24003

Contig3:  4266  acccaatgttcagcttctaagaaggagactagacttttccatatttcaacatctgattcc 4325
AC7455:  24004 acccaacattcagcttctaagaagtaggctggacttt-ccacattttaacatctgat-cc 24061
```

```
Contig3:  4326  atcc-------------tcctctt-----tgcctgttggctatttcaattatatctgtga 4367
                ||||             || |  |     |||||| ||| ||||||| ||||||||||||
AC7455:   24062 atcccttggtttttggatcataatgatcttgcctgctggatatttcagttatatctgtga 24121

Contig3:  4368  ttcagtgtcttctgtaatttttttaataaatgttacatatggggcaaagaaatcacagcct 4427
                | || ||||||||||| |||| ||||||||| ||| |||   |||||||||| ||  | ||
AC7455:   24122 tacaatgtcttctgtcatttcttaataaatattatatactaggcaaagaaaccattgtct 24181

Contig3:  4428  ttgtgtccacttacacacaaaatgggttaatttcttctgattaaaagatgcct 4480
                |||| | |||||  |||  |  ||| |||| | ||| | ||| ||  ||| ||
AC7455:   24182 ttgttttcacttgtacatcacctggtttaactgcttataattcaagaatgtct 24234


Score =  603 bits (380), Expect = e-170
Identities = 742/1058 (70%), Positives = 742/1058 (70%), Gaps = 85/1058 (8%)

Contig 0167:  22    aaaataatgatcatgaaacaatcttatatttttattaggaaggaaaatgcttagataaca 81
                    ||| | |||   |  | |||||||||| |||||||||| |   ||||||||||||| |||
AC7455:       72749 aaagtgatggctacaatacaatcttatgtttttattagaattaaaaatgcttagatgaca 72690

Contig 0167:  82    tagtgaaaaaataggttaaa--actaaatgcatacgaaactatgatta---cta------ 130
                    ||    ||| | || ||| |  | |   |||||| |  | ||||||||   ||
AC7455:       72689 taaatgaaagaaagtttacacaagtgtgtgcatatgttagtatgattatagctgtgtgga 72630

Contig 0167:  131   -----ccattataagaaagaaaactggaag----ttccccaaa-taccaacaatggatat 180
                         | ||||||||||||||||||||||    | ||||||| ||   || |||| | |
AC7455:       72629 aaaatctattataagaaagaaaactggaagaaagtaccccaaaatattgactatggctgt 72570

Contig 0167:  181   ttggggtatggaattatgagtga--tttttccctttcctgtgccaaattgnnnnnnnngg 238
                    ||||   ||| ||| || |||||  |||| || |||  || ||||||||         |
AC7455:       72569 ttggaagatgaaatcataagtgacctttttccctttaatgagccaaatt-taaaaaaaga 72511

Contig 0167:  239   gaatctagtacttttataattaa-----------tatcct---------caagtgtttct 278
                    |||||||||||||||||||| ||           |||| |         | | | ||| |
AC7455:       72510 gaatctagtacttttataatcaaaaaatacattttatcatattttattcctaatattttt 72451

Contig 0167:  279   tctcagaatgatatatacacacctatcttccatgatatatgacattttacatgaattagc 338
                     ||||||| |||| |  |||| |||| ||| | ||||||||||||||  ||||  ||
AC7455:       72450 cctcagaaagataca--cacaactattttctaagatatatgacatttctcatgggttcat 72393

Contig 0167:  339   taatctatttgtattcaaggagttaaaaaatattttccttctgggtttaggcaaaatcaa 398
                    |||| ||||||||||  ||||||||||||||||||||   |||||||| | | | | ||
AC7455:       72392 taatttatttgtatttgaggagttaaaaaatattttcagcctgggtttatgtacattgaa 72333

Contig 0167:  399   actactggcattgttttagacaattatcttggcatcaaatgtttattatattaacagtaa 458
                    ||||  |  || | ||||||  |||  |||| ||| |||||  |||| ||| ||||||||
AC7455:       72332 actatggttatgggtttagatgatttccttgccataaaatgaatattctatcaacagtaa 72273

Contig 0167:  459   tgagctacaagtctgtatttaccattctaaaggaatgaaa-acatttta-ttggagagca 516
                    | ||| | |   || |||||||||  |||||||||| ||| || ||||| |||||||  |
AC7455:       72272 tcagccatagagctatatttaccacgctaaaggaataaaacacgttttatttggagatga 72213
```

```
Contig 0167:  517   gcaagaaagctttacagactcctagaatatcacaaataagcaagcaagagaggaaatgac 576
                    ||||||| ||||||  | || | || ||| ||||||||||  | | ||||||||||
AC7455:       72212 gcaagaatgctttatgggcttccagtataccacaaataagccggaaggagaggaaatagt 72153

Contig 0167:  577   aaaacagagatccttcatgcatgcttagaggacaattttagaatacaaaagcgttcttaa 636
                    ||| |||| |  ||    | || | | |||| |||||  || |||  | ||| | ||  |
AC7455:       72152 aaagcagaaagact--gaggatacctggagggcaattccaggatatgagagcctcctgga 72095

Contig 0167:  637   cttgaatttcaagtttggatgtccctgatcgaagcaccgtctcctctcaaggattctgct 696
                     |||   |||  |||||||||||||||| | ||||| || |||||| || ||||| |||||
AC7455:       72094 tttgtgcttccggtttggatgtccctggttgaagccccatctccttttctaggatgctgct 72035

Contig 0167:  697   ggcttcccttcctccatctgctgtgttgggtgtttaaccttcagggacagctgttttcaa 756
                    |||||   |||||| | |||||       ||  ||||| ||||    |||| ||||||||
AC7455:       72034 ggctttgattcctctacctgct-------gtccttaacattcacaagcagcagttttcaa 71982

Contig 0167:  757   ctgtgcaaattatgtacgccgagcatggctcacaagtgttgtttgacgtgcttgttgcag 816
                    | ||||| || |  || || |||||||||||||||||||| | || |||| ||| |  |
AC7455:       71981 cagtgcagatgacataggctgagcatggctcacaagtgttgctcgaggtgcatgtggtcg 71922

Contig 0167:  817   aacatgggttacagggaagcct-gtgaaaaatccatttttaactaatctcaacatggaga 875
                    ||||||||| |||||| |||||| | |||||||| ||||| ||| ||  | |||| || ||
AC7455:       71862 aacatgggctacaggaaagcctggggaaaaatcgattttcaaccaagattaacaagg-ga 71863

Contig 0167:  876   agctggataaaatgtcaactcagnnnnnnntcatgccaaaatttgcccttcaaattatat 935
                    |  | ||||||||| ||||||||        | || |||||||||||| |||||||| |
AC7455:       71862 atgttgataaaatggcaactcagaaaaac--cgtgtcaaaatttgccctacaaattatgt 71805

Contig 0167:  936   tttgaagaactctattttcagaataacatttgaatgaa----------------------a 974
                    | |||| ||  |  | | || |||| || |||| ||||                      |
AC7455:       71804 tctgaaaaaaac--tctgcacaatagca-ttgactgaagcctgtttcttctctgcaccca 71748

Contig 0167:  975   gaattgccacttcatgttcactatagcagagacagttg 1012
                    |||||||||||||||||||||| |||||| | ||||||
AC7455:       71747 gaattgccacttcatgttcactgtagcagtggcagttg 71710


Score =  632 bits (397), Expect = e-179
Identities = 821/1203 (68%), Positives = 821/1203 (68%), Gaps = 110/1203 (9%)

Contig 0169:  337   cttctccaagcatcccagctctcagcctaacccctgacaccatggcctgctgttccacca 396
                    || || | |||  | | ||| | || | | |  ||| || |   || |||||   ||||
AC7455:       8899  ctcctgccagccacgctgctgtgagaccagctgctgccagccaagc-tgctgccagacca 8957

Contig 0169:  397   gcttctgtggatttcccatctgttccactgctgggacctgtggctccagctgctgccgat 456
                    |||||||||||||||| | ||  || || | |||||| |||| ||| ||||||||||
AC7455:       8958  gcttctgtggatttcctagcttctcaaccggtgggacttgtgactctagctgctgccagc 9017

Contig 0169:  457   caacctgcagtcagaccagctgctgccagccaacytccatccagaccagctgctgccagc 516
                    ||| |||| || | || ||||||||||||||||  | |  ||||||||||
AC7455:       9018  caagctgctgtgaaactagctgctgccagccaagctgctaccagaccagct--------- 9068
```

```
Contig 0169:  517   caacctgcctccagaccagtggctgtgagaccggctgtggcattggtggcagcaytggct 576
                          ||| | ||  | |||||||            ||||||||||  ||| |||||
AC7455:       9069  -------cctgcgga--actggctgt------------ggcattggtggtggcattggct 9107

Contig 0169:  577   atggccaggtgggtagcagcggagctgtgagcagccgcaccaggtggtgccg-cctgact 635
                    ||||||||| ||| ||||| |||||||||||||| ||| | |||||||||||| || ||||
AC7455:       9108  atggccaggagggcagcagtggagctgtgagcacccgtatcaggtggtgccgcccagact 9167

Contig 0169:  636   gccgcgtggagggcaccagcctgcctccctgctgygtggtgagctgcacatccccgtcct 695
                    |||| |||||||| ||| ||||||| |||||||| |||||||||||||| |||| ||||
AC7455:       9168  gccgtgtggagggtacctgcctgccccctgctgcgtggtgagctgcacacccccatcct 9227

Contig 0169:  696   gctgccagctgtactatgcccaggcctcctgctgccgcccatcctactgtggacagtcct 755
                    |||||||||||| || | ||| ||||||||||||||||||||||||||||||||||||||
AC7455:       9228  gctgccagctgcaccacgccgaggcctcctgctgccgcccatcctactgtggacagtcct 9287

Contig 0169:  756   gctgccgcccagcctgctgctgccagcccacctgcactgagcccgtctgtgagcccacct 815
                    | |||||||||| |||||||||| |      |||            || ||||||||| |
AC7455:       9288  gttgccgcccagtctgctgctgcta------ctg------------ctctgagcccactt 9329

Contig 0169:  816   gctcccaacccatctgttaaaaacctactgatggaaattctaagacaatggcacttcaaa 875
                    | |   ||||            ||| ||| |||  || || | |  ||||||| | |||
AC7455:       9330  gttgaaaacc------------tccttctgctggggatcctgataagatggcaccttaaa 9377

Contig 0169:  876   attagccagtgtacagtcccaacaaacttctaaggactagtacccataacccagcttacc 935
                    | ||||||   || | ||| ||||| ||||| |   | |||||| |||||   |||| |
AC7455:       9378  actagccaaattagaatcctaacaatcttctgaactccagtacctataactgggcttgca 9437

Contig 0169:  936   acctcatgtcagaaggctacataatttccctaacatggaaaattgatttataatnnnnnn 995
                    |||||   ||| |  || |||||| |||||||  |  | ||||| ||||| |||
AC7455:       9438  acctctcatcacacagccacataaattccctagga-agtaaattcatttacaatggaaga 9496

Contig 0169:  996   nnnnnnntatttcatagatctggatatcaacagaagtcttacagtgtgaaaagaaatgtg 1055
                          | |||| |||| |||| | ||| |  ||||| |||| ||||||||||  |
AC7455:       9497  ccaaaaatttttcctagacctggttgtcagccaaagtcctacaatgtgaaaagagttaga 9556

Contig 0169:  1056  taatactttactataaatatcttctgaaatgtttcaccaactcacttggcttttgggttt 1115
                    || || ||||||||||||||  ||||||| ||||| ||  |   ||||   ||   |||
AC7455:       9557  tactattttactataaatatcacctgaaatatttcaacagtt--attgggacttaaattt 9614

Contig 0169:  1116  aataaaatttttcat-tcttcagtgatgaaatcattg-ttggactttcttttctttggtg 1173
                    ||||||| ||||||| |||||| ||||||| |  ||| ||| ||||||||| ||||||
AC7455:       9615  aataaaagttttcatctcttcaatgatgaagttgttgtttgaactttctttgctttggaa 9674

Contig 0169:  1174  tccattattgatattggggtactgggatttctcaaaatgtcaataaatgtagtgtagcac 1233
                    |  ||||   |||||||||||||||||||| ||||||| |||  | |||   ||   |||
AC7455:       9675  tttattaacaatattggggtactgggatttttcaaaatatca--acatgacatg--acac 9730
```

```
Contig 0169:  1234  tgcggtttttctttattactgtgcttttagcagtggcctttaaaaggtggttttccttgt 1293
                    |  |||||||||| || |  ||  ||||||    ||   ||||| ||    ||| | ||||
AC7455:       9731  tatggtttttctgtacttttgctcttttaatgatgatatttaacagcaattttcctttgt 9790

Contig 0169:  1294  ct--------------------atctacagtcaaagtatatgattaaaagaatacgtcag 1333
                     |                    | | ||| |||   |||||||||||| |||   | ||
AC7455:       9791  tttaccattaaatggtccaatattttgcagccaagacatatgattaaaggaacttgacaa 9850

Contig 0169:  1334  tatttagacwctattccctgaagtcctcaagtatgcaaattaatgggttaggtgtcccag 1393
                     ||||||   || || ||| | ||||||   | |||||| |  |||| ||||| | ||
AC7455:       9851  catttaggaactgtttcctaaggtcctctcatgtgcaaagtggtgggctaggtattcc-- 9908

Contig 0169:  1394  aataaccaattattatgtcwtttggcttcttgaggtgtcaaccaaaaatattca---gtt 1450
                       |||  ||       |  ||  ||||||||||||  |   |||||  ||||||    |
AC7455:       9909  ---aacagat-------tgattcagcttcttgaggt-tttgccaaatctattcattcatc 9957

Contig 0169:  1451  cattcatatatgattctgtggacagttctttaaactatcgaggatataaaagtaatgcat 1510
                    ||||||||||||||   |||  ||| || |||      || | |||||| |   || ||||
AC7455:       9958  cattcatatatgaggttgtatacaattatttctgtcctccaagatataca---aaagcat 10014

Contig 0169:  1511  aat 1513
                    |||
AC7455:       10015 aat 10017
```

```
Score =  318 bits (200), Expect = 3e-85
Identities = 317/420 (75%), Positives = 317/420 (75%), Gaps = 6/420 (1%)

Contig AA:  1      aagcttgaatttttcatgattt-ggatgtagaagaaattatggcatgtgggttccttgat 59
                   |||||||||||||| |||| | || |||  ||  |||||||||| ||  ||| |||||  |
AC7455:     67592  aagcttgaatttatcattacttaggaattaagagaaattatg-cacatggcttcctgaag 67650

Contig AA:  60     gtcatgtgtccctcagctgtagggctcagcacttttctcagtgtccagcagagtaagcat 119
                     | ||||||  ||  |  |    | |  |||||| | ||| || ||||| ||||||| |
AC7455:     67651  tccgtgtgtcgatcctccatgtatccccacactttgcacagcgttcagcaaagtaagcgt 67710

Contig AA:  120    tggctgtatctcttttgctgtaaaaggtacgaatctgaactctctcttcgccagctggta 179
                   ||| | |  ||||||||||| ||||||||||   |||||| || |||||||||||| ||
AC7455:     67711  tggattt--ctcttttgctgtgaaaggtacgagagtgaactgtcccttcgccagctgtta 67768

Contig AA:  180    gagactgacattagtggcctgcatgggattctgggagagctgactgtatgcagatctgat 239
                   ||| |||||||  ||  |||||||||||| ||||  || |||||  | |||| |||||||
AC7455:     67769  gaggctgacatcagcagcctgcatgggatcctggaggaactgaccctgtgcaaatctgat 67828

Contig AA:  240    ctggaggcccatgtanagtctctgaaggatgatctcctttgccttaagaaaagccatgaa 299
                   |||||||||||||  |||||||||||| ||||||||||||||||||||| |||||||
AC7455:     67829  ctggaggcccatgtggagtctctgaaggaagatctcctttgccttaagaaaaaccatgaa 67888

Contig AA:  300    gaggtaaggaaaatggcaaaatcactgaaggaccctcaaagtgacaaaagttaagctggg 359
                   |||||||||||| ||||||   | ||||||||| |  | ||| ||   | |||   ||
AC7455:     67889  gaggtaaggaaagtggcaactccgctgaaggactccgagagtaac-gtacttacattgac 67947
```

```
Contig AA:  360    agagctcatgttga-gtatattcatagaacttgagtagatgttgactgatgtggaaaaaa 418
                   |||     |||||| || | |  || ||| ||| | ||||||||  | |  ||||| |||
AC7455:     67948 agatgatctgttgatgtttgtaaatcgaatttgggcagatgttgtttaagctggaataaa 68007


Score =  402 bits (253), Expect = e-110
Identities = 337/412 (81%), Positives = 337/412 (81%), Gaps = 3/412 (0%)

Contig F:  190    gcatgattcttcagtgacaaaacggagtg-gtcttgcccttacataccaggctctgctgt 248
                  || ||| |  |||||||||| ||||||    | |||  ||| ||  |||||||| ||||||
AC7455:    69460 gcctgacttatcagtgacacaacggacactgcctttgcctcacgcaccaggctttgctgt 69401


Contig F:  249    gcttgaagctcaatttccagagcgttggctgtgcgcttcagttccaagatctccgtctgg 308
                  ||||| |||||||||||||||||  |||||||||| |||||||||||||||||| |||||
AC7455:    69400 gcttggagctcaatttccagagcactggctgtgcgtttcagttccaagatctccatctgg 69341


Contig F:  309    cagccctgcagctgctctgcactggatagctgctgctgattcagttcctctgtctgaaat 368
                  |||||||||||||||| || ||||| || |||||||||||||| || ||||||||||
AC7455:    69340 cagccctgcagctgctccgcgctggacagttgctgctgattcagctcttctgtctgaaac 69281


Contig F:  369    acaggcatgattagaatatgcaaattaggcagaagtctggatgcctgtcattttattgct 428
                  |||| ||  |||||| || ||||  |||||||||||||| | ||||| |||||| |   |
AC7455:    69280 acagacaccattagagaattcaaaaaaggcagaagtctgtaagcctgacatttttcaat 69221


Contig F:  429    ccaagtgccaacctgaacagcgaaccattcttccacgtctctgcggttgttggcgagcac 488
                  |   || |  |||||||||| |||||||||| | |||||||| |||||||| |||||
AC7455:    69220 ctgggtactgacctgaacagccaaccattcttcagcttctctgcgattgttggcaagcac 69161


Contig F:  489    agtttcatactgacaacgcatctcatccaagactctgttaagatccgtggtgggggcaag 548
                   |||||| ||||||| |||||||||||| ||| ||||| || ||   ||||||||| ||
AC7455:    69160 cgtttcacactgacagcgcatctcatccaggaccctgttgaggtcaagggtgggggc-ag 69102


Contig F:  549    tgtttagct-cacactgagccggtcgccagcttgtccacgaagcacgttgac 599
                  |||  |||| ||||||||| |||||||||   ||| ||||||||| ||||||
AC7455:    69101 tgtccagctccacactgaggcggtcgccaagctgttcacgaagcaagttgac 69050


Score =  357 bits (224), Expect = 6e-97
Identities = 303/368 (82%), Positives = 303/368 (82%), Gaps = 14/368 (3%)

Contig GG:  1     aagcttaatgccatgggaagaggtcaacaaggacttaatgaacttaacatactcacactt 60
                  ||||| | ||||| | | |||||||||||||| |  |||||||||||  |||||||||| |
AC7455:    82917 aagctcagtgccaagtggagaggtcaacaagaatgtaatgaacttagtatactcacacct 82976


Contig GG:  61    gcttaataatggatgctgttagaataatgacgacagatccttttgaagacataaggtgaa 120
                  ||||||||||||||||||||||         ||||| ||||||||||||||||  ||||
AC7455:    82977 gcttaataatggatgctgttag---------gacagttccttttgaagacataatttgaa 83027


Contig GG:  121   gctctgaatctaat-gataagtagtasttgtgatgggtaataacttagtaacagcctaaa 179
                  ||||||  || ||| || |||||||| || |||||||||||||||| ||||||||||||
AC7455:    83028 gctctgggtccaatagacaagtagtagttatgatgggtaataacttactaacagcctaaa 83087
```

```
Contig GG:  180    cacatatttttatgagacaagtagctcttctcaggcaagtaaaaagcagactaatgagat 239
                     || || ||||||||||||| || ||||||| | ||||||||||||||||||||| ||||
AC7455:     83088 tgcacatgtttatgagacaaataactcttcttaagcaagtaaaaagcagactaataagat 83147


Contig GG:  240    gttcctgtgggatttgatatataaacattagaaaggaaaacttcctgcctactgtgacac 299
                   ||||||   |||||||  |||||||| | |||||||||||| |||  |||  ||||||||
AC7455:     83148 gttcctacaggatttg--atataaacctgagaaaggaaaacatcccacct-gtgtgacac 83204


Contig GG:  300    catcttgagcaaaggctatataagtacttcaaagaggaaaaagccaagaggtttgggaat 359
                   ||   | ||   | ||||||||||||  ||||||||||||||||  |||||||  || |||||
AC7455:     83205 cacgtagaagagaggctatataagcacttcaaagaggaaaacaccaagagccctt-ggaat 83263


Contig GG:  360    ttagactc 367
                   | ||||||
AC7455: 83264 tcagactc 83271


Score = 1072 bits (674), Expect = 0.0
Identities = 1022/1327 (77%), Positives = 1022/1327 (77%), Gaps = 40/1327 (3%)

Contig G:  448    tcaattaaaagtcagttaatttattaaagtccccaaagctaaaagagagttactaaaaat 507
                  |||  | || || | ||   |  | |||  || ||||||||||||| |||    |||| |
AC7455:    55571 tcatgttaatgtaattttgatcttcaaaagcctcaaagctaaaagaaagt----aaaatt 55626


Contig G:  508    aagtta---tctttaaaagaaaaatagttcatcaaaccaattgataacagataaaggtaa 564
                  |||||    ||||||||| || |||| || |||| || ||||| ||  |||||||  |
AC7455:    55627 aagttcaactctttaaaacaagaatactttatcagacaaattggtgtat-agataaagtgac 55685


Contig G:  565    agaaaataaaaactgaccagtacgagggaaaatacctacaccctacagctgtactgtcac 624
                  |||   ||||||   | ||| ||  | ||  |  || |||   |||||    |||||||
AC7455:    55686 agat--taaaaataaagcagcactggagagcaggcccacatattacagactcactgtcat 55743


Contig G:  625    gca-tattccccatcagggctgatgctggaattataatagcagaagccactgacgatacc 683
                  ||| ||||||||| |||||||| || || || ||||||| |||||| |  ||  ||||||
AC7455:    55744 gcaatattccccagcagggctggtgttgaaaatataataacagaaggctttggtgatacc 55803


Contig G:  684    tcataaagcttgtgttccagtgacaaatagcaagagtaatttgttcagatttagcagtga 743
                  |||||||||||||||||||||||||||||||||| |||||  |||||||||||||  ||
AC7455:    55804 tcataaagcttgtgttccagtgacaaatagcaagggtaatcagttcagatttagcaacga 55863


Contig G:  744    actcaaacaaaggttaattaggaaaataaagactctttacaagcatctaaatgtcaaacc 803
                  ||||||| ||| |||||||||||||||   |||||||||| |  |||||||||||||
AC7455:    55864 actcaaataaaagttaattaggaaaatggggactctttaca-gagtctaaatgtcaaaca 55922


Contig G:  804    agtcaaaggataatgggtacactaacataaacaacaac-aggaaagaagccttccggttg 862
                  | | |||| |||||||||||||||||||||||| ||| ||||||||||||   || ||
AC7455:    55923 aattaaagaataatgggtacactaacataaacaagaaccaggaaagaagcctcttggctg 55982


Contig G:  863    gaccaacacccactggagagggtatataagagcctcagagcagaaagagagasattctcacg 922
                  ||||||| |||| |||||||||| ||||||||| ||||| |   ||||| |||| |||
AC7455:    55983 gaccaactcccagtggagagggtacataagagccccagagaaaggagagatattcacacc 56042
```

```
Contig G:  923   ccaggagactcgtc-ttctttaaaaccaaatcaataaaaaacacagcttcccaacaccat 981
AC7455:    56043 tcagaagccttatccttcttttcaatcaaacccataaataccacagactctaatagccat 56102

Contig G:  982   ggcttgctgtgcccgcctctgctgcagcgtccccaccacccccgccaccaccatctgctc 1041
AC7455:    56103 ggattgctgtgcctctcgaggctgcagtgtccccaccgggcctgccaccaccatctgctc 56162

Contig G:  1042  ctctgacaaattctgcagatgtggagtctgyctgcccagcacctgsccacacacagtctg 1101
AC7455:    56163 ctctgacaaatcctgccgctgtggagtctgcctgcccagcacctgcccacacacagtttg 56222

Contig G:  1102  gttcctggagccaacctgctgtgacaaccgccccccaccttgccacattcctcagccctc 1161
AC7455:    56223 gttactggagcccacctgctgtgacaactgtcccccaccctgccacattcctcagccctg 56282

Contig G:  1162  tgtgcccacctgcttcctgctcaactcttcccagcccaccccaggcctggaaagcatcaa 1221
AC7455:    56283 cgtgcccacctgcttcctgctcaactcctgccagccaactccaggcctggagaccctcaa 56342

Contig G:  1222  cctcacaacctacactcagcccagctgtgagccctgcatcccaagctgctgctgaccgac 1281
AC7455:    56343 cctcaccaccttcactcagccctgctgtgagccctgcctcccaagaggctgctaatggat 56402

Contig G:  1282  ggctgcctcacccastgcc------tgacagagtcaacccagaagctttagtgctcacct 1335
AC7455:    56403 ggctactttgctcagtgcctgagattgaaaaagtcaacatagaagctttagcattcacct 56462

Contig G:  1336  gtctcagtacctgcaactaattatgtctccgctttcaaagttggaacaa----ggcatta 1391
AC7455:    56463 atctcagtacctacaactaat--gtactctgctttagaaattggaacaaggatggtacta 56520

Contig G:  1392  tcacagacaaccctcacaaaaaacaa---accaagagactttcgatggccgtgtagtgga 1448
AC7455:    56521 ccacaatcacccctgcaaaaaaaaagagaccaagaaactttcaatgaccat-------t 56573

Contig G:  1449  catcagtgaacaaggacagctggagtaggtagatgcctacaggtttcccagcgttgttca 1508
AC7455:    56574 cagctataaccaactgcagtttgaatcagtggatgcctata-gcttcctgaagctgttcg 56632

Contig G:  1509  gttccttcgtgttaaattgtatctttct-ttggtgctttgggaattctgtttccagtctt 1567
AC7455:    56633 attccttcatattaaagtgtctctttctgtgggtggtttgggaattctgttttcagtctt 56692

Contig G:  1568  gaatcgtatctttctggaaattgaggagcttcttcatgattattctaataaagtttacat 1627
AC7455:    56693 gggtggtatctttctgaaaattaaggaagttcttcatgattatcctaataaattttacat 56752

Contig G:  1628  ctctggcataacataaatgtctataggtatttccatttatttttgttaacacatcaaatg 1687
AC7455:    56753 ctctggcatagcacaagtgtctacaattacttctgtttatttctgttcattcattaaatg 56812
```

```
Contig G:  1688  tatttcttaactccaaaggtcascattttaas--aaactgagaacatctgcagaactgaa 1745
                 ||||||||||||  ||  |   ||||| ||   ||||||  |||||  |   |||||||
AC7455:    56813 cgtttcttaactctgaacattggcatttaaagaaaaactggaaacattagtgaaactgaa 56872


Contig G:  1746  aagagtg 1752
                 ||  |||
AC7455:    56873 aatggtg 56879


Score =  412 bits (259), Expect = e-113
Identities = 376/472 (79%), Positives = 376/472 (79%), Gaps = 15/472 (3%)

Contig JJ:  3     gcagacagaatctgtggaatgcaccgtggcggagactgaggcccagtacagctcggagct 62
                  |||||||||||||| |||||||||||||||| || || ||||||||||||||||  ||||
AC7455:     71170 gcagacagaatctctggaatgcaccgtggcagaaaccgaggcccagtacagctcccagct 71229


Contig JJ:  63    ggcccagatacagtgtctgatcgataatgtggagaaccagctggctgagatccgctgcga 122
                  |||||| || |||||||||||||||||  |||||||||||||||| ||||||||||||||
AC7455:     71230 ggcccaaattcagtgtctgatcgataacctggagaaccagctggccgagatccgctgcga 71289


Contig JJ:  123   cctggagcggcagaaccaggagtatcaggtgctgctggacaccaaggcccggctggagtg 182
                  ||||||||| |||||||||||||| |||||||| ||||||   |||||||||||||| |
AC7455:     71290 cctggagcgacagaaccaggagtaccaggtgctcctggacgtgaaggcccggctggaggg 71349


Contig JJ:  183   tgagatcaacacgtaccgggggctgctggacagcgaggacagcaggcaagtccacacaat 242
                  |||||||||||||||| |||| ||||||||||||||||||||||||||||| | |||| |
AC7455:     71350 tgagatcaacacgtactggggcctgctggacagcgaggacagcaggcaagttc-cacagt 71408


Contig JJ:  243   ccctcttaaaatccccccagagtcacacagaggctcttgaaggaccatcctgkttcccaaa 302
                    ||| | |||||||||||||||||| |||||||  || |  |  |||||| ||||| ||
AC7455:     71409 tactcatgaaatccccccagagtcacaaagaggctgatgtaaaagaatcctgtttccccaa 71468


Contig JJ:  303   gatctggccatattcaaacctccttaaaggagktaaaggatattctt-aaagagtgatga 361
                    ||||| ||  |||| ||||||   |  | | | | |||| ||||| ||| |||| |||
AC7455:     71469 agtctggacaagttcagacctcc-ccagtgtgtttagggatcttcttaaaaaagtgctga 71527


Contig JJ:  362   aaagcagc-aaggctagtcacatgtca--attgaaatggtcaaaatggtctaaattttcc 418
                  ||||  ||  ||| ||   ||||| ||  |  || ||  | |||||  |||||
AC7455:     71528 aaagtggctgaggttaagaacatgacactagagatatttttaaaatattctaa------- 71580


Contig JJ:  419   aggcggaaaaaatataatatgtaacatttnaagctgg-gggttccttttgttc 469
                  |||| ||  || ||| |  |||||||||| | |||||    | |||||||||
AC7455:     71581 aggcagattaattatta-gtgtaacatttcatgctggcatctccctttgttc 71631
```

```
Score =  225 bits (141), Expect = 5e-57
Identities = 280/502 (55%), Positives = 280/502 (55%), Gaps = 26/502 (5%)

Contig J:  8     agcaggagataagtgtacactcaggtaatttagattatatagggg-ttcattggacacata 67
               ||||||| ||  |||||| || |  ||  | ||| | ||  ||| |||   ||  ||| |
AC7455:    35305 agcaggacatg-gtgtactcttaaatatctgagaatttactgggattcgcagggtacaca 35247


Contig J:  68    ggacttgaacacaaggcagtgctgtt-------attccatccgtagtgaacagcagcagt 120
               || ||||||  |||||||| |  ||        |||||| |  ||| ||||||||||||
AC7455:    35246 gggcttgaatgcaaggcagaggagtaggaaggcattccagctatagggaacagcagcagt 35187


Contig J:  121   aaatgcacagaaatggaagagaaaaaactgaaaacaagaagcaatgagtagtgcagtgct 180
               ||||||| |||| ||||| |||| |||  |     | ||||| |||| ||||||  ||
AC7455:    35186 aaatgcatagaagtggaaaagaacaaatggttgttaggaagccatgaatagtgcgctgtg 35127


Contig J:  181   tctaggaaataattatgtgagataaaagctaaacacagagagtcttgactcctgggatga 240
               |||||  || |||| ||||||||||||||  || |   |||||||||| ||  ||||||
AC7455:    35126 tctagagaacaattgtgtgagataaaagc--aagatgtagagtcttgac-cccaggatga 35070


Contig J:  241   ggagtttgaactt-aatcaggtaga-ttattgaaagatgctggtatttgtttttctaaag 298
               || |||||||||| ||| | |||||   || || |    ||||||| | ||||||||| |
AC7455:    35069 ggtgtttgaacttgaattaagtagacccatggagaaccactggtatatatttttctaagg 35010


Contig J:  299   caggtaagta---atttaac----------nnnnnnnnnnnnnnnnnnnnnnnnnnnnn 345
               |||| |  ||    ||||||
AC7455:    35009 caggaagttacctctttaactccgtcatttgtaaatatggctaatgttactatcacattg 34950


Contig J:  346   nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn 405

AC7455:    34949 ataggatgtttaaagaataaattaaagaaagcattaaaatgtgcaatgcagtgcctggaa 34890


Contig J:  406   nnnnnnnnnnnnnnnncaatggttcttataggtgtggtacttattattaagtaagctga 465
                           |||| |||| ||||||||| |  || ||| ||| ||||| ||
AC7455:    34889 catagtaagtcctcaataaatgattctaataggtgtgattgttgttactaaataagcaga 34830


Contig J:  466   gggaagctggagctcttttctt 487
               |||| ||||||||| |||||||
AC7455:    34829 gggaggctggagctttttttctt 34808
```

```
Score =  139 bits (87), Expect = 2e-31
Identities = 124/155 (80%), Positives = 124/155 (80%), Gaps = 8/155 (5%)

Contig KK:  2      agcttgatcaggaatgagtccatccttcttttccttttcttctctaaaaaggggaatcta 61
                   ||||| | ||||||||||| ||||||||||||||||||||||||| |||  ||||| |||
AC7455:     82921 agcttaaccaggaatgagttcatccttcttttccttttcttctctagaaa-agggaaacta 82863


Contig KK:  62     ttgatatctaataattattggtaaacttagaaatcag-------tattgtattttggtat 114
                   || || | ||||||||||||||||||||||||||||       |||| || |||| | |
AC7455:     82862 ttaatgtttaataattattggtaaacttagaaatcagaagtcaatattttagtttgttct 82803


Contig KK:  115    gtccaacagaatttcccatttccattgttcggaca 149
                   | ||||  |  || ||||||| || |||  |||||
AC7455:     82802 gcccaaatggtttccccattttcagtgtcaggaca 82768


Score =  274 bits (172), Expect = 9e-72
Identities = 352/532 (66%), Positives = 352/532 (66%), Gaps = 39/532 (7%)

Contig L:  10     atcaagctttaagagtgattcaacattctcaaagtcacaaagtta-----ttaaaaaatt 64
                  || ||| |||||| || |||||||||||| ||||||| |||||||     ||| |||||
AC7455:    51299 atgaagttttaagcgttattcaacattctgaaagtcataaagttaattacttagaaaat- 51357


Contig L:  65     ccaggattcagctccttgtatatctgtttgactttaaagcccaattctttctatcacatc 124
                   ||| |||||||  |||||||| ||||    ||| |||||||||||| || ||| ||||||
AC7455:    51358 -cagaattcagcctcttgtatgtctg----actctaaagcccaatttttcctaccacatc 51412


Contig L:  125    acagaacccccagtgctaacagagccagaaatttaataaacacaggtatgacattaactn 184
                  | ||||||  || | ||||   | ||||||  ||||| | ||||| ||||| || |||
AC7455:    51413 atagaaccttcaatactaatgaaaccagaagcttaatgagcacagttatgaaataaaccc 51472


Contig L:  185    nnnnnnnnnnnnnnnnnnnnnnnnnnnnnt-ctttaacaaaatcatccaatccagctg 243
                                                | ||||||||||||  |  |||    |
AC7455:    51473 aaaatatgcaagagaaataaaacatgtaagattctttaacaaaatccac--atcggatta 51530


Contig L:  244    taatggacaaggatatatatataggcttatagattaaaagtggcagatattcagttctaa 303
                  |||| || || ||    || || || | |||||||||| |||||  |||||| ||||||
AC7455:    51531 taatagaaaaagag---tacattggttcatagattaaaattggcaagtattcaattctaa 51587


Contig L:  304    gaaccgggtcatatagtgagcctacc-attttgacttcatttttccttggagaaaactta 362
                   || | || ||| | |   |  |||| |||| ||||||||||  ||| |||| |||| |
AC7455:    51588 -aatctgggcattttgaaggtgtacctatttcaacttcattttctcttagagagaactca 51646


Contig L:  363    cagtggtccattctagtagtagggcaataggaatatccttctccctgccacaaca----- 417
                   ||||||| | ||||||| | |||| ||| | || | | |   ||||||| | |
AC7455:    51647 aagtggtcaactctagtaacaaggcaccagggacatacctttatttgccaccatactcag 51706


Contig L:  418    ----------ctaccccagtaatgaagagaacagcagtattttccaggtgattgtctgag 467
                            ||||| ||||||| || ||||   || || ||| ||||    ||||| ||
AC7455:    51707 cacagatgttctaccacagtaatcaatagaatgccaatacttttcagg----tgtctcag 51762
```

```
Contig L:  468   ggctagagct-aaagtctgattaaacatggnttaatctactacttaatttta 518
                 |   |||||| ||| ||||||||   |||  |||||    ||||||||||||
AC7455:    51763 gttgagagctaaaactctgattacgtatgctttaattcgttacttaatttta 51814


Score =  257 bits (161), Expect = 9e-67
Identities = 305/417 (73%), Positives = 305/417 (73%), Gaps = 30/417 (7%)

ContigOO:  1     ctgcagcagaggcgggcacagcaagccatggtgttggaaatctggtttgctttcggtttc 60
                 ||||||| | |   |||||||||| ||||||   |||   ||||||| |||||| |||||
AC7455:    50373 ctgcagctgcgagaggcacagcaatccatggcaatgggcgtctggttagctttcagtttc 50314


ContigOO:  61    taggaaagaaaggtgaggtttct-aggatgaaatgtctcttctcactttgggactcttat 119
                 |  ||        ||||| |||| |||   ||||||||| |||    || || |||||||
AC7455:    50313 ttaga--------tgaggattctgaggtacaaatgtctcctcttttgttagggctcttat 50262


ContigOO:  120   atactcctctcgttgagtgtcggtccaatcagaagacatcctcttatttgcatttatgcc 179
                 |||| |  |    || |  | |||||||   |||| | |||| ||  |||  |||||| |
AC7455:    50261 ataccctcccaagtgggcattggtccaactggaagccttcctgtt--ttg--tttatggc 50206


ContigOO:  180   gcc-ttttccactaagattctctaatcagkttggtatttacttg-ccattaagagttcct 237
                  || |||| |||| |||||||||||||| |||  |||||| || |||||||||||||
AC7455:    50205 acccttttcactgagattctctaatcagtttgtcatttacatgtccattaagagttccc 50146


ContigOO:  238   gagcttattagggtagtcatgtttttgactcattgacttgtcatt---------agatga 288
                 |  ||||||| | || |||| |||||||||||||||||||||||         |||| |
AC7455:    50145 gctcttattaagttaatcatatttttgactcattgacttgtcatttttgtcacaagatca 50086


ContigOO:  289   tca-tttttaactttacagggttctggaatgccaaatcttgcatga--attatggagaat 345
                 |||  ||||  ||| ||||| |||||| |||||||| |||  ||||   ||||||  |||
AC7455:    50085 tcacattttgtcttcacaggtttctggcatgccaaaccttatatgacttttatgggtaat 50026


ContigOO:  346   gaatccaagtgacaattcagc--tagtttcagaggctaaa-tattctcttcattccc 399
                  ||||||||||||||||||||  || |||  ||||| ||| | |||| | |||||
AC7455:    50025 caatccaagtgacaattcagctgtaattttggaggccaaactcttcttgtttttccc 49969


Score =  306 bits (192), Expect = 5e-81
Identities = 355/492 (72%), Positives = 355/492 (72%), Gaps = 22/492 (4%)

ContigPP:  661   aaaagatggacatcaatggcagtcctgataccaatcattctttcctccaacc-atgtcat 719
                 || | || ||  |||||||||| | ||||||||  ||||||||| | ||| | ||   ||
AC7455:    36579 aataaattgatgtcaatggcagccttgataccagccattctttctttcaatctatactat 36638


ContigPP:  720   aaaggagtttccagaaatttt-aaagatcttgcaattcaggaagtaaagaccaagtgttt 778
                 |||||| |   | |  | ||| ||||||| |||||||||||| | |||||||| |||||
AC7455:    36639 aaaggaatgcactggcagtttcaaagatcatgcaattcaggaaatgaagaccaaatgttt 36698


ContigPP:  779   ttacttgcaag-catttttattaggatgacaaagtggaaagatatctgtatcctgctttt 837
                 ||| || | || |||||||||||||||||||| |  |||| | ||| ||||||||||||
AC7455:    36699 ttatttactagacatttttattaggatgacaaactttaaaggtttctatatcctgctttt 36758
```

```
ContigPP:  838   gtgaagggatgaagcacttcattgggcttaaagaaaatcgggttgccatgaagtatgaac 897
AC7455:    36759 gtgaagggatgaagaattccaacaggctttaggaaaattgagtggccctaaggtgtgaag 36818

ContigPP:  898   cctcccaccacttgacatccctgccta--cagraggtactgtcttgggaggggtttcccg 955
AC7455:    36819 cttcc-------taacatccctgctgactcaggaggtgacactgtggaaggtgtttccca 36871

ContigPP:  956   acagcaacccagaaatgcaccgagttgttgggaggtggatgctagtaaggttcccacaga 1015
AC7455:    36872 acagccacccacaaattcactgagttgttgggagatggatattcgtaaggcacaaagtgg 36931

ContigPP:  1016  ctcacamcaagcagtgactaacacacatactactgacctttt--atcacagtccagcaag 1073
AC7455:    36932 ctcacagcaaaacatgactaacgcacataaaagtgactatctcgttctcagtccaggcac 36991

ContigPP:  1074  ttggtagcaagcacgtatttcatctatgtattggag-cttccaaatc------ttcagcc 1126
AC7455:    36992 atcattgtaagca-aaatttttcctgtatgttggagcctcccaaatctgcacgtccagct 37050

ContigPP:  1127  ctgctccctctc 1138
AC7455:    37051 ctgctccctctc 37062


Score =  126 bits (79), Expect = 2e-27
Identities = 164/261 (62%), Positives = 164/261 (62%), Gaps = 12/261 (4%)

Contig P:  3     gcttctgactcctctatgttgtacaattttctacttcattttagaggctgagtcattttg 62
AC7455:    39553 gcttctgattcctctgcattgtgcaattttctac-----tttagaacttgagccatctta 39607

Contig P:  63    tttg--aggatgagtcaaaagaatgt----atttttaaaaagatttcaaaggaaaatatt 116
AC7455:    39608 tctgaaaaaatgggtcaagaaaatatgtacatgtttttaagcctttctaagtaagccatt 39667

Contig P:  117   tctaaa-gatttaccaccttaatatttttcctcctttaaaagaatgctaaaatagcattc 175
AC7455:    39668 tctaaaggttttaccagtttagtatttttcatattctaaagtaattctaaatcatcatac 39727

Contig P:  176   taagtttcatgcctttacnnnnnnnnnnnnnnnnnnnnnnngattctagtgatttaagt 235
AC7455:    39728 taaattttctacctttatctacttttaaaactttatttttagagttttagtggtttagat 39787

Contig P:  236   gtaactgggggatctccctat 256
AC7455:    39788 gaaactcggaaatgtctctat 39808
```

```
Score =  529 bits (332), Expect = e-148
Identities = 392/448 (87%), Positives = 392/448 (87%), Gaps = 2/448 (0%)

Contig R:  339  gggccatcagctcctggccagggtatataaagggcccagagcaggaggaagacactcaca 398
               ||| || | || || ||||||||||||||||||||||||  |||||| |||| |||||
AC7455:    2814 gggtcacctgcacccggccagggtatataaagggcccagagggggaggaggacattcaca 2873

Contig R:  399  cctgagaacactcaactcctctcaccrcctcarccccactcggccccamacaccaccatg 458
               ||||||||||  || ||||||||  |  | || |||    | ||| || ||||||||||
AC7455:    2874 cctgagaacatccagctcctctccacagcccaacccacaccagcctcagacaccaccatg 2933

Contig R:  459  accgggctcctgttgcgggccccaccttctcctccctcagctgtggcggaggctgcctcc 518
               ||||| |||||| ||||| | ||||||||||||||| ||||  || |||||||||  ||
AC7455:    2934 accgg-ctcctgctgcgg-ctccaccttctcctccctgagctacgggggaggctgctgcc 2991

Contig R:  519  agccctgctgctaccgcgacccctgctgctgccgcccagtgtcctgccagaccaccgtga 578
               |||||||||||| |||||||||||||||||||||| ||| ||||||||||||||||
AC7455:    2992 agccctgctgctgccgcgacccctgctgctgccgccccgtgacctgccagaccaccgtgt 3051

Contig R:  579  gccgccccgtgaccttcgtgccccgctgcacgcgccccatctgcgagccctgccgccgcc 638
               ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
AC7455:    3052 gccgccccgtgacctgcgtgccccgctgcacgcgccccatctgcgagccctgccgccgcc 3111

Contig R:  699  ccacatcctgccaggccgtggtgtgccgcccctgctgctgggccaccacgtgttgccagc 758
               || | || |||  ||| ||||||||| | |||||||||||||||||| || |||||||
AC7455:    3172 cctcgtcgtgcacggctgtggtgtgcaggccctgctgctgggccaccacctgctgccagc 3231

Contig R:  759  ctgtctctgtgcagtgcccctgctgccg 786
               |||| |||||||||| ||||||||||||
AC7455:    3232 ctgtgtctgtgcagtccccctgctgccg 3259
```

SUBMISSION OF RESULTS

Gill, C. A., Davis, S. K., Taylor, J. F., Cockett, N. E. & Bottema, C. D.K. (1999). Construction and characterization of an ovine bacterial artificial chromosome library. *Mammalian Genome*, *10*(11), 1108-1111.