School of Biomedical Sciences

Characterisation of the Central Region of the Sheep Major Histocompatibility Complex

Jinyi Qin

This thesis is presented for the Degree of
Doctor of Philosophy
of
Curtin University of Technology

November 2008

Declaration of Authenticity

To the best of my knowledge and belief this thesis contains no material previously published by any other person except where due acknowledgment has been made.

This thesis contains no material which has been accepted for the award of any other degree or diploma in any university.

Signature: …….

Date: ……..

ACKNOWLEDGEMENTS

First of all, I would like to thank my supervisors Associate Professor David Groth and Professor John Wetherall from Curtin University of Technology, for their kindness and full support in my study. Thank you David for working hard to generate enough research funds throughout these years, your patience and talent in molecular genetics. Thank you John for your kindness to letting me having this chance to be involved in molecular genetic research and for your great help in editing and finalizing my thesis.

Thank you Adrian Paxman for your wise advice, support and assistance to my work throughout these studies.

Thank you to all of the people at the School of Biomedical Science and at the Technology Park laboratories for their friendship and technical support.

Lastly, but most importantly to Deby Rori, my wife, I dedicate this thesis to you. Thank you so much for your love. I may not have been able to complete this project without your support and encouragement.

Abstract

The major histocompatibility complex (MHC) is a chromosomal region encoding molecules controlling adaptive immune response in vertebrates. In farm animals, many associations between MHC loci and productivity traits including disease susceptibility have been described. However, current knowledge about the structure and function of the MHC in domestic animals, especially sheep, is very limited. Characterization of the sheep MHC may potentially facilitate breeding for enhanced disease-resistant animals through use of marker assisted selection. The main aim of this project has been to provide insights into the organization of the genomic content of the central region of the sheep MHC.

The work described herein has utilized subcloning of a sheep BAC genomic library in conjunction with DNA sequencing to generate a map of the central region of the sheep MHC covering ≈700 kbp. Within this map the relative order and identity of twenty five recognized loci were established. For some loci the intergenic distances were also determined. The final map is the most accurate map of this region reported to date and shows a high degree of similarity to the analogous region of the human MHC. This work has been published and a copy of the paper is included in Appendix 1. During the course of this work detailed genomic sequences were obtained for several sheep central region loci. Complete nucleotide sequences were generated for the complement factor B locus (CFB) and the TNFα locus and a comparative analysis of these sequences confirmed their homology with other vertebrate orthologues. Extensive partial sequences for complement components C2 and C4 were also obtained and reported to GenBank. In addition, a previously identified short tandem repeat locus designated BfMs believed to be in the CFB locus was mapped to an intron within the adjacent SKI2VL locus.

Single nucleotide polymorphisms (SNPs) were identified by analysing homologous sequences from a minimum of five individual sheep. In total 33 SNPs were discovered distributed over eleven distinct loci. Allele frequencies for SNPs from ten of these loci were determined and reported for a panel of 71 sheep comprising 58 unrelated sheep from the Rylington Merino flock plus a further 13 unrelated parental animals from a three generation half sibling sheep pedigree.

The availability of an independently confirmed pedigree constructed from a three generation half sibling sheep family permitted the identification by deduction of central region MHC haplotypes based on a panel of SNPs derived from 10 loci. This is the first reporting of haplotypes covering this region of the sheep MHC. Analysis of SNP panel genotypes in the cohort of 71 unrelated sheep using the expectation maximization algorithm permitted the prediction of a group of approximately 20 haplotypes, which accounted for more than 90% of the expected haplotype distribution. Four of these predicted haplotypes were also present in the known haplotype cohort deduced from the sheep pedigree. Analysis of pairwise linkage disequilibrium between SNP loci in the cohort of 71 unrelated sheep showed a centre-most region displaying relatively high levels of linkage disequilibrium which was bounded by two regions displaying more variable linkage disequilibrium. It is hypothesised that this mid region of the central region of the sheep MHC may be a block like structure characterized by low recombination similar to those that have been widely described in the human and mouse genomes.

The discoveries reported in this thesis provide a more accurate and detailed description of the central region of the sheep MHC together with a panel of SNPs, which reflect the diversity of this important genomic region which is known to be associated with immune responsiveness. The description, for the first time, of central region haplotypes provides a practical means of seeking candidate loci associated with disease resistance and productivity traits. The application of molecular techniques will enhance the rate at which the genomic composition of this region is elucidated and the work described in this thesis will contribute to final characterization of this important complex in health and disease.

Table Contents

		Page
Declaration of Authenticity		ii
Acknowledgements		iii
Abstract		iv
Contents		vi
List of Figures		X
List of Tables		Xi
List of Abbreviations		Xiii
Chapter 1	**The Major Histocompatibility Complex in Mammals Emphasising Sheep and Cattle**	
1.1	Introduction	1
1.2	Genomic Composition of the Human MHC	3
1.2.1	Linkage disequilibrium and block like structures	5
1.2.2	Conserved Human MHC Haplotypes	6
1.2.3	Central region of the human MHC	7
1.2.4	Recombinogenic loci	7
1.3	Comparative Aspects of the MHC in sheep and cattle	8
1.3.1	Cattle MHC	10
1.3.2	Sheep MHC	11
1.4	Disease Associations with MHC Loci	12
1.4.1	Human MHC	12
1.4.2	Disease and productivity associations with sheep and cattle MHC loci	12
1.5	Selection on sheep MHC loci	16
1.6	Relevance of this study	17
Chapter 2	**General Materials & Resources**	
2.1	Sources of sheep DNA	19
2.2	DNA purification	20
2.2.1	Sheep DNA	20
2.2.2	Plasmid and cosmid DNA	20

2.2.3 Bacterial artificial chromosome (BAC) DNA 21
2.2.4 DNA amplicons from PCR 21
2.2.5 Final purification of DNA products 21
2.3 Subcloning of Cosmid and BAC DNA into Plasmid Vector pGEM-4Z 21
2.3.1 The pGEM®-4Z Vector 21
2.3.2 Restriction enzyme digestion of DNA isolates 22
2.3.3 DNA ligation 23
2.3.4 Transformation 23
2.3.5 Recombinant Selection 23
2.4 Generation of DNA probes for identification of locus specific clones and subclones 24
2.5 DNA probe labelling and hybridisation 27
2.5.1 Prime-a-Gene® Labelling Probe 27
2.5.2 Labelling overlapping oligonucleotide (Overgo) 27
2.5.3 Hybridisation 28
2.5.3.1 Hybridisation of the sheep cosmid library 28
2.5.3.2 Hybridisation of the sheep BAC library 28
2.5.4 Post hybridisation washing and film development 29
2.6 DNA sequencing 29
2.6.1 LI-COR DNA 4000 sequencing 29
2.6.2 ABI 3100 automated sequencing 30
2.6.3 DNA Sequences Analysis 30
2.7 Detection of SNP and microsatellite loci 31
2.8 Sources of materials and composition of reagents 31

Chapter 3 A Map of the Central Region of the Sheep Major Histocompatibilty Complex (MHC)

3.1 Introduction 34
3.2 Materials and methods 35
3.2.1 Strategy for identification of sheep MHC central region orthologues of human and mouse genes 35
3.2.2 Labelling of DNA and oligonucleotide probes 39
3.2.2.1 Labelling of DNA 39
3.2.2.2 Labelling overgo overlapping oligonucleotide probes 39
3.2.3 BAC and cosmid library screening 39

3.2.4 Subcloning of the BAC clones 40
3.2.5 Restriction analysis of cosmid clones containing complement and CYP21 genes 40
3.2.6 DNA Sequencing 40
3.2.7 SNP discovery and genotyping 41
3.3 Results 42
3.3.1 Identification of BAC clones containing sheep central region MHC loci 42
3.3.2 Cosmid mapping of the MHC associated complement genes in sheep 46
3.3.3 Detection of SNPs 47
3.4 Discussion 49

Chapter 4 Sequence Analysis of Sheep and Cattle TNF, CFB, C2, C4 Genes

4.1 Introduction 52
4.2 Materials and methods 54
4.2.1 Cosmid and BAC clones 54
4.2.2 Sequencing and sequence analysis 54
4.2.3 SNP discovery and genotyping 54
4.3 Results 55
4.3.1 Sequence of sheep and cattle CFB genes 55
4.3.2 Polymorphism of sheep CFB 56
4.3.3 Sequence analysis of sheep and cattle C2 genes 69
4.3.4 Sequence analysis of sheep and cattle TNF 71
4.4 Discussion 75

Chapter 5 Genotyping of Polymorphic Loci within the Central Region of the Sheep MHC

5.1 Introduction 79
5.2 Materials and methods 80
5.2.1 Animals genotyped 80
5.2.2 SNP genotyping 81
5.2.2.1 Genotyping using the pyrosequencing method 81
5.2.2.2 RFLP Analysis (PCR followed by RFLP) 83
5.2.2.3 Allele specific PCR (ARMS: amplification refractory mutational sequence) 83

5.2.2.4 Microsatellite assay 84
5.2.3 Analysis of SNP and microsatellite genotypic data 84
5.2 Results 85
5.3.1 Genotyping of individual loci and allele frequencies 85
5.3.2 Genetic characterization of sheep populations 91
5.3.3 Identification of haplotypes from the sheep MHC central region 95
5.3.4 Intragenic haplotypes 95
5.4 Discussion 96

Chapter 6 Concluding Comments and Discussion 100

Bibliography 106

Appendix 1 Publication 118

FIGURES

		Page
1.1	Extended Map of the Human MHC (7.6Mb)	4
1.2	Genomic Structures of Cattle and Sheep MHCs Compared with the Core Human MHC (≈4Mb)	9
2.1	pGEM®-4Z Vector circle map and multiple cloning region	22
3.1	Map of the central region of the Sheep MHC	44
3.2	Identification of a further nine loci within the central region of the sheep MHC	45
3.3	Identification of SNPs using DNA sequencing method	47
4.1	Alignment of Genomic and cDNA for human sheep and cattle Complement factor B gene	57
4.2	Alignment of CFB proteins from sheep and cattle	68
4.3	Alignment of C2 proteins from sheep, cattle and human	69
4.4	Alignment of genomic and CDS sequences of sheep TNF with their human and cattle orthologues	71
4.5	Alignment of peptide sequences of sheep TNF with their human, cattle, rat and mouse orthologues	75
5.1	Pedigree chart for three generation half sibling sheep family	81
5.2	Three genotypes for the TNFaSNPS1 locus detected by the pyrosequencing method	90
5.3	Allele frequencies for the BfMs (SKIV2L) microsatellite locus in the central region of the sheep MHC for 71 unrelated animals	90
5.4	Deduced haplotypes for the 25 animals from a three generation half-sibling sheep family	95

TABLES

2.1	Standardized PCR Reaction mixes	25
2.2	Standardized PCR conditions for amplification of sheep genomic DNA, plasmid DNA, cosmid DNA and BAC DNA	25
2.3	Consensus primers for exonic sequences used to generate PCR amplicons for use as probes to screen Cosmid clones	26
2.4	The sources of important materials used in the study	32
2.5	The composition of important reagents used in experimental work	33
3.1	List of consensus primers from human and mouse exonic sequences used to generate sheep amplicons for use as probes to screen BAC clones	36
3.2	Sheep DNA sequences of amplicons generated using the primers listed in Table 3.1.	37
3.3	Locus specific overgo primers used to identify sheep genes in BAC clones	38
3.4	GenBank accession numbers for the ten BAC clones used in Figure 3.1	38
3.5	Primers for sequence walking of plasmid subclones derived from individual BAC clones	41
3.6	PCR primers for identifying locus specific SNPs	42
3.7	Identified sequences of genes within the central region	43
3.8	SNPs identified in the sheep MHC central region	48
3.9	SNPs identified in the sheep CFB and TNFa loci	48
4.1	Primers for generating and sequencing overlapping contigs of CFB and TNFa	56
5.1	Primers for SNP typing using the pyrosequencing method	82
5.2	Allele specific oligonucleotide primers for TNFaS1 genotyping using the modified ARMS assay	84
5.3	Genotypes for the 25 animals from a three generation half-sibling sheep family	86
5.4	Genotypes for the panel of 58 unrelated sheep together with the 13 parental animals from the half sibling pedigree	87
5.5	Allele frequencies for 12 SNP loci in the central region of the sheep MHC for 71 unrelated animals	89
5.6A	Allele frequencies for the BfMs (SKIV2L) microsatellite locus in the central region of the sheep MHC for 71 unrelated animals	89

5.6B	Observed and expected heterozygosities for cohorts of sheep	91
5.7	Inferred distribution of haplotypes present in the composite group of 71 unrelated sheep based on the SNPstats implementation of the EM algorithm	93
5.8	Estimation of pairwise linkage disequilibrium between 10 distinct SNP loci in a cohort of 71 unrelated sheep using SNPstats software	94

ABBREVIATIONS

aa	amino acid
≈	approximately
bp	base pairs (nucleotides)
BSA	bovine serum albumin
C2	complement component C2 (gene in italics)
C4	complement component C4
CFB	complement component factor B
CYP21	steroid hydroxylase gene(s). In humans the CYP21A locus is a pseudogene and its paralogue is designated CYP21B.
dNTP	deoxy nucleotide tri-phosphate
dATP	deoxyadenosine triphosphate
dCTP	deoxycytidine triphosphate
dGTP	deoxyguanosine triphosphate
dTTP	thymidine triphosphate
ddNTP	dideoxynucleotide triphosphate
DNA	deoxyribonucleic acid
dsDNA	double stranded DNA
EtBr	ethidium bromide
E.coli	*Escherichia coli*
EDTA	ethylenediamine tetraacetic acid (disodium salt)
H2	murine major histocompatibility complex
HLA	human lymphocyte antigens
hpH_2O	highly purified water
hrs	hours
IPTG	isopropyl-β-D-thiogalactopyranoside
kbp	kilo base pairs (nucleotides)
kDa	kiloDalton
LB	Luria-Bertolli broth
LBA	Luria-Bertolli agar
MAS	marker assisted selection
Mb	mega base pairs (nucleotides)
g, mg, μg, ng	gram, milligram, microgram, nanogram
M, mM, μM, pM	molar, millimolar, micromolar, picomolar
min	minute
ml, μl	millilitre, microlitre
mm, μm, nm	millimetre, micrometre, nanometre

MCS	multiple cloning site
MHC	major histocompatibility complex
NA	not available
OHCC1	MHC class I associated microsatellite locus
PCR	polymerase chain reaction
PCNM	positively charged nylon transfer membrane
RDBP	reverse direction binding protein gene
RE	restriction enzyme
RELP	restriction fragment length polymorphism
rpm	revolutions per minute
RT	room temperature
SDS	sodium dodecyl sulphate
SNPs	single nucleotide polymorphisms
SSC	sodium chloride / sodium citrate buffer
SSCP	single stranded conformational polymorphism
ssDNA	single stranded DNA
TAE	tris / acetate buffer
TBE	tris borate buffer
TC	*T. colubriformis*
TE	tris EDTA buffer
TEMED	N, N, N', N'-tetramethylethylenediamine
TNF	tumour necrosis factor
UV	ultraviolet
V,kV,mV	volts, kilovolts, millivolts
x-gal	5-bromo-4-chloro-3-idolyl-β-D-galactoside

Chapter 1

The Major Histocompatibility Complex in Mammals Emphasising Sheep and Cattle

The major histocompatibility complex (MHC) is a chromosomal region originally identified by the presence of loci determining the outcome of allograft transplantation. It is now understood that MHC encoded molecules expressed on lymphocytes and macrophages play an essential role in controlling adaptive immune responses in vertebrate species. MHCs have been identified in many species of jawed vertebrates representing the major orders of the Phylum Vertebrata. The well defined MHCs of the jawed vertebrates have not been observed in jawless vertebrates or invertebrates. The human and mouse MHCs have been the most extensively characterized and mapped (Kulski et al., 2002), although a plethora of current genome projects is rapidly contributing to the characterisation of MHCs in other species. The human MHC has become the reference standard for comparative studies of the MHC in other mammals and comprises three well defined contiguous subregions covering ≈4Mb. The subregions of mammalian MHCs are referred to as the class I, class II and central regions, although it is known that considerable structural diversity exists between species. This chapter will compare the composition and organization of the human MHC with that of two ungulates of economic importance namely sheep and cattle and discuss the identification and use of polymorphic markers from this region for association studies with genes influencing productivity traits in these domesticated species.

1.1 Introduction

The classically defined major histocompatibility complex (MHC) comprises a cluster of highly polymorphic genes encoding molecules responsible for controlling the adaptive immune response in vertebrates (Beck and Trowsdale, 2000; Kulski *et al*., 2002). In general terms, MHCs in vertebrates are gene dense regions containing diverse loci some of which are involved in effector mechanisms associated with immune responsiveness (Kumanovics *et al*., 2003; consortium, 1999). It is not surprising therefore that many associations between disease states and loci within MHCs have been reported, especially for humans (Milner and Campbell, 2001; Shiina *et al*., 1999). Classical MHCs, identified by the presence of histocompatibility loci, have been found in all vertebrates species studied

except for the jawless fishes (Agnathans). An MHC is regarded as being a *sine qua non* for the generation of adaptive immune responses in turn characterized by the synthesis of immunoglobulins (antibodies) and activated lymphocytes (Klein, 2004).

The term MHC derives from the historical discovery, initially in mice and subsequently humans, that the acute rejection of primary allografts and the accelerated rejection of subsequent allografts (ie. tissue = histo compatibility) was determined by a single locus of major effect. The early definition of the MHC was only possible due to the pioneering efforts of George Snell and his colleagues in breeding, and making available to others, several lines of highly inbred mice within which skin transplants were not rejected. It subsequently became apparent that this major histocompatibility locus was not a single gene but rather a complex of linked genes each expressing a high degree of polymorphism. Thus the acronym MHC was formed. The discovery of the MHC in inbred mice and humans and the delineation of its biological functions represents one of the major advances in the development of biomedical science and has led directly to the awarding of two Nobel Prizes. The first of these was awarded to Baruj Benacerraf, Jean Dausset and George Snell in 1980 for their combined work in identifying and characterising the MHC, while the second was awarded to Peter Doherty and Rolf Zinkernagel in 1996 for their seminal discovery of the role played by the polymorphic histocompatibility genes within the MHC in determining the specificity and collaborative nature of interactions between lymphocytes and antigen presenting cells (http://nobelprize.org/nobel_prizes/medicine/laureates/index.html).

Detailed characterization of the MHCs in humans and several experimental animal species has shown that two main classes of histocompatibility gene exist, each of which contains polymorphic loci, and which encode glycoproteins expressed on lymphoid and myeloid cell membranes. These are now referred to as class I and class II loci and the complexes as classical MHCs.

Class I loci encode a single glycoprotein of approximately 44kd which is associated with a $\beta 2$ microglobulin polypeptide ($\approx$11.8kd) encoded by a non MHC gene. Class I molecules are present on T lymphocytes and macrophages and are integral in the generation of immune responses against infectious agents, especially viruses. In humans these are referred to as the HLA-A, B, C genes. The HLA designation stands for human lymphocyte antigen since the molecules were first identified as membrane associated proteins present on lymphocytes using

serological methods (based on alloantibodies). In contrast, class II loci encode heterodimeric membrane associated molecules resulting from the expression of two very closely linked MHC genes (a and b). These molecules are preferentially expressed on antigen presenting cells such as B lymphocytes and macrophages. In the human MHC these loci are called HLA DQ, DR, DP etc. Despite their different peptide compositions, class I and class II molecules show similarities in their mode of function and three-dimensional structures and it is clear they share a phylogenetic ancestry based on multiple gene duplication events followed by nucleotide divergence (Beck and Trowsdale, 2000; Kumanovics *et al*., 2003). Class I and II pseudogenes are also common (Horton *et al*., 2004).

In all classical MHCs present in vertebrate species many non histocompatibility loci are also present. Many of these are associated with effector immunological functions but others are not related to the immune system. In humans the class I and class II regions are separated by a gene dense region which has been referred to as the class III region or more commonly as the central region. The central region of the human MHC contains loci which contribute to both the innate and adaptive immune responses (Schwaiger *et al*., 1996). This region is highly conserved in several mammalian species with respect to structure, homology and gene mapping and it has been postulated therefore that it has a distinct and older evolutionary history to that of the class I or II regions (Beck and Trowsdale, 2000; Kumanovics *et al*., 2003).

1.2 Genomic Composition of the Human MHC

In this review the human MHC will be used as an exemplar for comparison with the less well characterized MHCs of sheep and cattle, both ungulate mammalian species of considerable commercial importance. The human MHC occurs on the short arm of chromosome 6 (6p21.31) spanning approximately ≈7.6Mb of gene dense DNA of which the central 4Mb is regarded as the core or classical MHC (Horton *et al*., 2004). Following Beck and Trowsdale (2000), the extended MHC may be conveniently subdivided into five regions designated extended class I, class I, central region, class II and extended class II. Each region contains numerous loci and the main features of this complex are summarized in Figure 1.1 and discussed further below.

1.2.1 Linkage disequilibrium and block like structures

One of the earliest defining characteristics of the HLA genes present in the human MHC was the presence of extensive linkage disequilibrium (more correctly gametic phase disequilibrium) exhibited by specific alleles encoded by these polymorphic loci. For example the class I HLA-A1 allele and the HLA-B8 allele, both serologically identified and now typed using DNA methods, manifest strong linkage disequilibrium. Later, this pair of alleles was shown to also manifest strong linkage disequilibrium with the physically more distant HLA-DR3 class II allele. Such linkage disequilibrium was the first clue that the human MHC contained conserved 'blocks' of loci and that combinations of these blocks reflected the existence of conserved haplotypes within human populations. It is important to note that although the mouse MHC was more fully characterized for many years than that of the human, studies of the MHC in this species were usually based on mice belonging to highly inbred strains. As a consequence the existence of a block like genomic structure and conserved extended haplotypes was not initially detected and appreciated in this species. This major advance in understanding the evolutionary history of the MHC was discovered only from genomic characterisation of the MHC in the outbred human species. The first clues and general statement of the block like composition of the human MHC were described by Dawkins and colleagues in Perth who referred to them as "supratypes" (see review by Dawkins *et al.* (1983)).

It is now more generally appreciated that the class I, class II and central regions of the human MHC consist of four blocks of loci within which recombination is less frequent or rare and between which recombination is much more frequent (Stenzel *et al*., 2004; Daly *et al*., 2001). These blocks are illustrated in Figure 1.1 together with the main defining loci for each of the block like genomic regions. It is likely that the five subregions comprising the extended human MHC also share this block like structure (Yunis *et al*., 2003). The concept of a more general block like structure for the human genome has been developed by several groups. For example Gabriel and colleagues (Gabriel *et al*., 2002) showed that the human genome can be parsed into haplotypic blocks based on extensive SNP typing. The haplotypic blocks observed correlated across populations and were characterized by sizeable regions exhibiting little intrablock recombination. Each block comprised relatively few common haplotypes. The existence of these blocks has important implications for mapping of disease and other loci within the human genome. If the block like composition of human DNA occurs in other species, as

seems likely, then identification and mapping of disease and productivity related loci in the genomes of farm animals, and especially their MHCs will require less polymorphic markers and hence become a more tractable task. Some of the work reported in this thesis will contribute to this hypothesis. Although the arguments described above are supported by considerable confirmed research, a note of caution is also warranted. There is also evidence for gene conversion occurring within the human and mouse MHCs and this mechanism (entailing DNA repair mechanisms) can mimic linkage disequilibrium between and within loci, especially where repetitive sequences occur. The presence of gene conversion within the MHC may add an additional level of complexity to our understanding of the presence of conserved regions within the MHCs of mammalian species. This topic has been reviewed by Martinsohn and colleagues (1999).

1.2.2 Conserved Human MHC Haplotypes

A vast amount of MHC typing in human populations has occurred over the past 50 years or so and the outcomes of these efforts have been published in the proceedings of a series of International Histocompatibility Workshops. These workshops reflect an extraordinary degree of international and ethnic collaboration. Collectively they have permitted the identification of a collection of individuals carrying combinations of haplotypic blocks which define what are believed to be ancestral MHC haplotypes. In some instances the individuals identified are homozygous for these haplotypes and Epstein Barr virus transformed cell lines have been prepared from lymphocytes obtained from these persons. These valuable cell lines are now available for more intensive characterisation. The Sanger Institute UK is currently undertaking an MHC haplotype project with the objective of providing complete nucleotide sequences of the haplotypes present in many of these cell lines (http://www.sanger.ac.uk/HGP/Chr6/MHC/). Very recently, comprehensively annotated reference sequences for eight human chromosomes representative of the European population were published which identified in excess of 44000 polymorphisms including indels (Horton *et al*., 2008).

It is generally hypothesised and accepted, but not proven, that variation within the MHC, both polymorphic and epistatic, is a selective response to pathogens (Beck and Trowsdale, 2000). These authors have argued that maintenance of MHC diversity is due to heterozygote advantage or frequency dependent selection for the rarer alleles or perhaps both mechanisms acting concurrently. Either

model could explain the preservation of MHC haplotypes with combinations of alleles that generate synergy in adaptive immunity or its effector mechanisms. Other possibilities are that rare recombination within haplotypic blocks has helped promote linkage disequilibrium within blocks and that the relatively short evolutionary history of humans from ancestral families has been insufficient to allow for linkage equilibrium to be achieved. These last two hypotheses may be complementary.

1.2.3 Central region of the human MHC

Relative to other regions of the human MHC, the class III or central region is the most gene dense and has the least number of pseudogenes. The variability of the genes located in this region is very diverse and, although many loci are associated directly or indirectly with immunity, there are numerous other loci for which there is no obvious association with immunity. The central region of the human MHC is generally considered to span ≈700kbp between the NOTCH4 and BAT1 loci (Xie *et al*., 2003). One region of special interest within the class III region is a duplicated region containing paralogous complement component C4 loci which Yu and colleagues have termed the RCCX module (Yu, 1998; Yu *et al.,* 2002).

1.2.4 Recombinogenic loci

The block like structure of the human MHC described above implies that there are regions of sequence between blocks in which recombination occurs more frequently. Consequently a number of researchers have sought to locate and identify the main characteristics of these regions. A seminal report from Jeffreys group (Jeffreys *et al*., 2001) utilised genotyping of sperm to identify punctate meiotic recombination in the human MHC. A 216kbp region in the class II region of the human MHC, previously characterized for familial crossovers, was analysed. This analysis showed extended regions of strong linkage disequilibrium interrupted by relatively short regions of linkage disequilibrium breakdown. These latter regions corresponded precisely with meiotic crossover hot spots. This work was extended in an elegant study by McVean and colleagues (McVean *et al*., 2004) who showed that there was a perfect correspondence between the recombination hot spots identified by Jeffreys' group and recombination sites based on estimates of linkage disequilibrium derived from population genetic data. In a commentary on Jeffreys' paper Goldstein (Goldstein, 2001) referred to these block like regions manifesting a lack of recombination as "islands of linkage disequilibrium".

Although the discussion above relates to the MHC, it is clear from McVean's paper (McVean *et al*., 2004) and also that of Gabriel and colleagues (Gabriel *et al*., 2002) that the "islands of linkage disequilibrium" concept is applicable to the broader genome and is a general phenomenon.

Since hotspots for meiotic recombination act much like major intersections in the genome, there has been some speculation about the nature of these regions. Several studies have found that patterns of recombination are poorly conserved between humans and chimpanzees despite the similarity of the two genomes. For example, Coop and Myers (2007) have argued that recombination hotspots are in fact transient features of genomes which may be disrupted by mechanisms entailing biased gene conversion. Recombination hotspots within the sheep MHC have not been characterized to date with the exception of an apparent hotspot in the class I region reported by Bot *et al*. (2004). If, as seems likely, intra-MHC recombinogenic hotspots are not conserved between species, it will be necessary to treat cautiously analogies between human and sheep blocks for the purpose of disease and productivity association studies.

1.3 Comparative Aspects of the MHC in sheep and cattle

The structures of mammalian MHCs are characterized by differences based on a common structure. Thus all mammalian MHCs may be sub-divided into broad regions on a single chromosome and manifest a general genomic organization as shown in Figure 1.2. However, these regions do show evidence of rearrangements, duplications, inversions and translocations which collectively reflect the active evolutionary history of this genomic region in several different species (reviewed by Liu *et al*., 2006). Furthermore as summarized by Kumanovics and colleagues (Kumanovics *et al*., 2003), sequence analysis has shown that loci within the mammalian MHCs display both high and low levels of conservation. The least conserved loci are those comprising the species specific expansions of class I, class II and perhaps complement C4 genes whereas many, if not most, of the non-histocompatibility loci manifest greater conservation.

The main focus of this review is the comparative structure of the MHCs of the ungulate species cattle and sheep. At the present time the major regions comprising these MHCs have been identified, although they remain relatively poorly characterized relative to the MHCs of human and mouse (Dukkipati *et al.*, 2006). The broad structure of the MHCs of these two species relative to that of the core 3.5Mb region of the human MHC is shown in Figure 1.2. Very recently a

collaborative study by Wu and colleagues (Wu *et al*., 2008) published a high resolution radiation hybrid map of a sheep chromosomal region homologous to human chromosome 6, including the MHC, and compared these to corresponding chromosomal regions in cattle, horse and dog. This study confirmed the broad structure of the sheep MHC relative to the other mammals included but showed discrepancies with the higher resolution mapping of the central region described later in this thesis. This report is referred to again in Chapter 3.

1.3.1 Cattle MHC

The cattle MHC was originally identified in an analogous manner to the human MHC using serological identification of lymphocyte antigens. These were called bovine lymphocyte antigens (BoLA) and for this reason the cattle MHC has been historically referred to as the BoLA complex. The cattle MHC is now more usually referred to as *Bota* (*Bos taurus*) for conformity of nomenclature; it is located on chromosome 23 (Gelhaus *et al*., 2006). The class II region is the most centromeric and is subdivided into class IIa and class IIb sub-regions separated by approximately 15Mb. In the class IIb region the order of the constituent loci is reversed relative to that of the class IIa region while an insertion of approximately 1.7Mb is present in the telomeric end of the class IIa region (Womack *et al.*, 1997).

The cattle MHC differs from its human counterpart in several respects including variation in the number of expressed class I loci, and number of DQ class II genes. At least four classical class I loci are found in cattle, yet none of them is expressed constantly in all animals or all breeds (Kelley *et al*., 2005; Holmes *et al*., 2003). The class II region contains single copies of DOβ, DRα and DZα genes, and multiple copies of DRβ, DQα and DQβ loci (Andersson 1988; Andersson and Rask, 1988; Andersson *et al*., 1988). The DQ locus is the most polymorphic in cattle and is usually duplicated in approximately half of the common haplotypes in this species (Glass *et al*., 2000). The DP subregion is absent in cattle but two other class II loci, DYA/DYB, are present and expressed (Kelley *et al*., 2005).

The class III region has been identified and is known to be similar in length to that of humans but with the order of the majority of loci inverted relative to human. According to Liu and colleagues (Liu *et al*., 2006) the SKIV2L locus (superkiller viralicidic activity 2 previously referred to as SKI2W by Dangel *et al*., 1995) is missing from its syntenic location and appears to have been translocated to this

duplicated region. The human SKIV2L locus encodes a helicase enzyme which is related to the yeast SKI2 gene and was previously referred to as the SKI2W locus in the literature (http://www.genenames.org). Further Liu and colleagues (2006) have proposed that the central region of the cattle MHC has a short inverted duplicated region at its telomeric end. This duplication is not present in the human MHC. This view of the cattle central region is very likely incorrect: interrogation of version 4 of the cattle genome assembly (UCSC Genome Bioinformatics - http://genome.ucsc.edu/) does not show the reported rearrangement, and, as will be described in this thesis, it is not observed in the corresponding map for the central region of the sheep MHC.

1.3.2 Sheep MHC

The sheep MHC was originally identified in an analogous manner to the human MHC using serological identification of ovine lymphocyte antigens (Millot, 1978) and for this reason has been historically referred to as the OLA complex. As with cattle, the sheep MHC is now usually referred to as Ovar-Mhc (from *Ovis aries*) following the nomenclature system for MHCs of vertebrates proposed by Klein and others (1990); it has been mapped to chromosome 20q15-q23 (Jugo and Vicario, 2001). The sheep MHC is less well characterized than its bovine counterpart. Once again, distinct well separated class IIa and IIb regions are present together with a telomeric class I region (Figure 1.2). A detailed review of the composition of the sheep MHC has been published by Dukkipati and colleagues (Dukkipati *et al*., 2006) in which the nature and polymorphisms of the histocompatibility loci are described. These authors also describe the class III or central region and describe the presence in this region of loci homologous with those in the human MHC central region, although their genomic mapping is presently unknown.

It seems clear that the MHCs of various mammals show much basic structural similarity on which is superimposed a number of species specific variations such as those described in cattle above. Although the sheep MHC seems remarkably similar to that of humans, some structural variations are likely to be discovered as more detailed maps are generated. For example evidence will be described in this thesis showing a local inversion between the complement component C4 genes and the closely linked cytoplasmic enzyme 21-hydroxylase locus, both of which are duplicated in sheep and humans. In humans these genes occur in the order C4-21OH-C4-21OH, whereas in sheep the cosmid mapping data show the order of these loci to be C4-21OH-21OH-C4, implying a local inversion relative to that of

the humans. A similar inversion seems present in version 4 of the cattle genome assembly (UCSC Genome Bioinformatics - http://genome.ucsc.edu/).

The work to be described in this project will show that the sheep central region is similar in length to that of the corresponding human region (≈700kbp) long and that the order of loci present in sheep is similar to that in its human counterpart. More detailed mapping of this region of the sheep MHC is described later in this thesis.

1.4 Disease Associations with MHC Loci

1.4.1 Human MHC

Many disease predisposing loci have been mapped to the human MHC (eg Lie and Thorsby, 2005). This realization reflects a long history of disease association studies commencing with the observation of a very high relative risk for the HLA-B27 allele with ankylosing spondylitis, a debilitating rheumatic disease more prevalent in males (Bowness, 2002). It is of interest that despite the strength of this association and its long history a broadly accepted explanation for this observation has not been forthcoming and no specific disease locus has been identified. This observation illustrates very well the difficulty which may arise in deriving a genetic explanation for even a strong disease association observed in an outbred population and is a cogent lesson for those seeking disease predisposing loci in domesticated animals such as sheep and cattle. The variety and importance of many of these associations explains and justifies the plethora of research in this field. It is very likely that there will be important consequences arising from the transfer of this research to other species including cattle and sheep – the main focus of this project.

1.4.2 Disease and productivity associations with sheep and cattle MHC loci

The high degree of polymorphism of many MHC loci and the discovery of disease associations in humans prompted many researchers to seek to similar associations in a variety of mammalian species and in particular farm animals. This research was encouraged by a concurrent stream of research delineating the role of major histocompatibility loci in regulating adaptive immune responses in mammals and other MHC loci participating in a variety of immunological effector mechanisms. One of the earliest associations to be discovered in farm animals was the strong

association of Marek's disease with the chicken MHC (Briles *et al*., 1977). Marek's disease is a common disease in chickens caused by a virus of the herpes genus. It affects both commercial and backyard poultry and may result in death or severe production loss. Thus the observation that susceptibility to this viral infection was associated with certain heterozygous genotypes within the chicken MHC was of value to the poultry industry but also led the way to seeking similar applications in other species of economic importance. However, the extensive investment in time and money needed to generate alloantisera for serological identification and typing of polymorphic MHC loci in these species proved to be a major barrier to progress. More recently of course, the use of DNA technology has circumvented some, but not all, of these barriers.

There is a vast literature implicating MHC loci as determinants of resistance or susceptibility to a wide variety of pathogens in various vertebrate species and this literature is not reviewed comprehensively in this chapter. Rather, a selective review of the role of MHC loci in determining predisposition to nematode infestation in sheep and to a lesser extent in cattle will be undertaken since this illustrates as well as any other example the rationale for the characterisation of the central region of the sheep MHC as a prelude to using SNPs to identify haplotypes characteristic of this region and eventually of course to more extended haplotypes. As discussed above, the strong linkage disequilibrium over extended regions of the human MHC reflects a block like structure within which recombination is rare and this structure probably is present in other mammalian MHCs. Hence when associations are sought between MHC loci and disease states, haplotypical associations are usually more meaningful than an allelic association.

In Australia, the CSIRO was among the first to seek genetic markers in the sheep MHC associated with productivity traits such as resistance to parasites and bacterial pathogens. Gastrointestinal nematodes comprise the most important parasites for sheep and of these the species of the order Strongylida are most prevalent in Australia and New Zealand (Dukkipati *et al.,* 2006). Antihelminthic chemotherapy has been the main stay for controlling such infections but is becoming increasingly expensive and has led to the development of drug resistance (Roos, 1997; Hoekstra *et al*., 1997). Gastrointestinal parasites of the *Trichostrongylus* species result in significant economic loss in Australian sheep, both in summer and winter rainfall areas. The earliest work seeking associations between OLA antigens and resistance to gastrointestinal nematodes in sheep was undertaken at the Commonwealth Scientific and Industrial Research Organization

(CSIRO) McMaster laboratory at the University of Sydney and by Stear and colleagues at the John Curtin School of Medical Research, Australia National University (Stear *et al*., 1984; Schwaiger *et al*., 1995). These studies have been reviewed by Hohenhaus and Outteridge (1995).

Dineen and colleagues (1965) and later Windon and colleagues (1984) attempted to induce immunity to this parasite using a radiation attenuated larval vaccine followed by challenge with live worms to study the immune response of lambs to *T. colubriformis.* Genetic variation in the magnitude of the immune response was observed and selection and assortative matings based on faecal worm egg counts resulted in lines of sheep with varying resistance to this parasite. This work was then extended by Outteridge and colleagues (1988) who undertook the daunting task of preparing and using alloantisera for typing class I loci in sheep. They identified a class I lymphocyte antigen designated SY1 which was associated with resistance to *T. colubriformis*. Later, using DNA probes for human DQA, DQB and DRB cDNA sequences, which cross reacted with sheep orthologues, this group was able to extend this work to show statistically significant associations between class II alleles and faecal worm egg count in both *T. colubriformis* susceptible and resistant flocks of sheep. These studies marked a watershed in this research and are important since they were the first to show an MHC association with parasite immunity in a ruminant species.

Following this promising start, MHC disease associations in sheep then fell into a quiescent period and failure to reproduce some published research, together with the cost of this type of research, resulted in a questioning of the utility of the approach by many research groups. However some research did continue and in 1994 Karlsson and his colleagues established and reported on a major initiative in Western Australia called the Rylington Merino Project (Karlsson, 1994). Stear and colleagues (1995) extended the importance of the MHC in parasite immunity by reporting an association between a class II locus and parasite immunity which was later confirmed in 2005 (Stear *et al*., 2005). The major objective of this project was to select for enhanced resistance to worm infestation in sheep while improving, or at least maintaining, other traits of economic importance such as wool quality and growth rate. The Curtin University group was able to contribute to this project and reported a statistically significant association between worm susceptibility and a 19kbp fragment of sheep DNA which hybridised with both complement C4 and DRB sequences (Wetherall and Groth, 1992). This study was

the first to implicate the presumptive class III or central region loci of the sheep MHC in parasite immunity.

A more recent study from the Curtin University group (Bot *et al*., 2004) reported on associations of five MHC markers with production traits such as bodyweight, fleece weight, faecal scouring, fibre strength and fibre diameter. Two alleles at microsatellite loci close to the complement factor B locus (central region) and the Class II DRB1 locus showed significant associations with increased fleece weight than the most frequent alleles at each of these loci. This study has confirmed the importance of the sheep MHC, including the central region, as a source of loci for marker assisted selection in this species. Additional associations between the sheep MHC and other productivity traits (marbling of muscle and birth weight) have also been described but remain to be confirmed (reviewed by Dukkipati *et al*. (2006)). McRae and colleagues (2002) undertook to measure the extent of linkage disequilibrium in sheep rather than assume it existed to a similar extent as that of humans and mice. They used short tandem repeat genotypes (microsatellite genotypes) to measure the extent of linkage disequilibrium in two populations of domestic sheep. High levels of linkage disequilibrium were observed over distant loci which declined as a function of marker distance. Thus there seems to be evidence showing that linkage disequilibrium mapping will be as useful in sheep as it has been in other species such as humans and mice.

Associations between productivity traits and MHC loci have also been sought and found in cattle, the most studied species to date. For example, Schook and Lamont (1996) reported an association between MHC markers and growth rate, milk yield, milk fat, milk protein, and fertility in cattle (Schook and Lamont, 1996) and Kaminski *et al*. (2004) reported associations between SNPs of several candidate genes with milk protein biosynthesis also in cattle. Thus it seems likely that more extensive studies permitting the identification of haplotypic regions of the sheep MHC associated with nematode infestation and other productivity traits will increase the utility of marker assisted selection as a means of enhancing productivity in this economically important species of farm animal. The characterisation of the central region of the sheep MHC, which forms a major part of the project described in this thesis, is one significant step in achieving this goal.

The evidence that resistance to gastrointestinal nematodes is controlled by genetic elements is well established in the more easily studied mouse inbred lines. Thus Behnke and colleagues (2003) studied the responses of mice to the

gastrointestinal nematode *Heligmosomoides polygyrus* and concluded that multiple loci regulate the responses including both MHC associated and non MHC loci. In another study from the same laboratory, which reflects more closely the natural exposure of grazing animals to these nematodes, genetic variation in resistance to repeated infections with the parasite was shown to occur in inbred mouse strains selected for the mouse genome project (Behnke *et al*., 2006). These findings support the thrust of experimental work seeking similar results in sheep and cattle.

1.5 Selection on sheep MHC loci

The critical role of MHC class I and class II loci in mediating adaptive immune responses, together with their extreme polymorphism, has been explained by the mechanism of balancing selection driven in large part by responses to pathogens including parasites (Hedrick and Thompson, 1983). This topic has been reviewed more recently by Meyer and Thompson (2001) and by Bernatchez and Landry (2003).

Several recent studies have shown that there is good evidence that balancing selection also applies to the sheep MHC and that disease association studies in this species are therefore supported by evolutionary theory. For example, Gutierrez-Espeleta and colleagues (Gutierrez-Espeleta *et al*., 2001) investigated the resistance of big horn sheep to infectious disease and genetic variation at MHC class II loci. They found extensive genetic variation at the sheep MHC-DRB locus thereby showing that low MHC variation does not appear to be the basis of the high infectious disease susceptibility and subsequent decline in numbers of desert bighorn sheep and supporting the importance of balancing selection for maintaining variation within the MHC.

Charbonnel and Pemberton (2005) undertook a long term genetic survey of MHC and neutral loci of the free living Soay sheep population on the island of Hirta (Scotland) to demonstrate the presence of balancing selection and reasons for its existence. These sheep are subject to population size variation resulting from winter malnutrition exacerbated by infection with the nematode *Teladorsagia circumcincta*. For the latter 5 years of the 13 year study they observed lower levels of temporal genetic differentiation at MHC loci relative to neutral loci. They interpreted this as strong evidence for balancing selection at the MHC loci acting through spatial and temporal heterogeneity in selection pressure.

Santucci and colleagues (Santucci *et al*., 2007) investigated polymorphism at three short tandem repeat (STR) loci (OLADRB (U00204), OLADRBps (AH003856), OMHC1 (U21953)) within the MHC in several breeds of domestic sheep and one of their wild relatives, the Mouflon (*Ovaris orientalis musimon*) and compared these with genetic diversity at non-MHC STR loci across the sheep genome. They observed higher levels of polymorphism at MHC linked STR loci relative to the non-MHC linked loci. For most breeds studied, the distribution of alleles at MHC linked STR loci were more even than would be expected for strictly neutral loci and sampled from populations manifesting an equilibrium between genetic drift and mutation. They interpreted their results as evidence for balancing selection at these loci resulting from "hitchhiking" due to close linkage with alleles at both class I and class II MHC loci.

Finally, Schwensow and colleagues (Schwensow *et al*., 2007) compared neutral versus adaptive genetic variation and resistance to a nematode parasite in a free ranging population of fat tailed dwarf lemurs (a primate) under natural selection conditions. In this extensive report the authors characterized their animals by the presence of "supratypes" defined by the presence of similar antigen binding motifs from the MHC-DRB exon 2 locus. They observed that one supratype was associated with reduced parasite burden and concluded that their results supported the hypothesis that significant associations exist between specific MHC types and parasite infestation in naturally occurring populations (Schwaiger, *et al*., 1995).

1.6 Relevance of this study

As discussed above, it seems likely that associations between MHC loci and productivity traits in sheep, including immunity to gastrointestinal parasites, exist and may provide a basis for marker assisted selection facilitating breeding programs for superior stock. However the complexity of the MHC and the extensive polymorphism of many loci within its confines make identifying candidate markers a daunting task. Since these general goals were first enunciated several decades ago, the development and typing of other genetic markers such as polymorphic short tandem repeats (STRs) and single nucleotide polymorphisms (SNPs) has provided impetus for renewed efforts in this field. Of even more importance, the discovery of genomic blocks characterized by restricted recombination, means that the search for candidate markers can

become more efficient. It may become possible to identify a genomic block covering many kilobases with a relatively small representative panel of markers. By combining such blocks, an extensive chromosomal region could in theory be screened for useful associations more quickly and with less cost. The concurrent development of software capable of inferring haplotypic combinations of these loci from genotypic data further enhances this approach.

The main thrust of this study has been to identify and map the central region of the sheep MHC and to then identify polymorphic short tandem repeat loci and single nucleotide polymorphisms as a first step to identification of candidate loci associated with productivity traits. The potential applications of the project are predicated therefore on two hypotheses. Firstly, that the sheep and human MHC central regions will share conserved characteristic loci. Secondly, that a group of adjacent loci within the central region of the sheep MHC will define a genomic region with restricted recombination which will permit the identification of a relatively small number of common haplotypes within a population.

The main objectives of the project were as follows.

a applying a new strategy for generating DNA probes to identify genes from clones of sheep cosmid and BAC libraries containing central region loci of the sheep MHC.

b generate a physical map of the central region containing identified loci,

c identify SNPs and/or polymorphic STR loci by inspection of amplified sequences from a panel of individual sheep, and

d to use the panel of polymorphic markers so obtained to identify central region haplotypes in the sheep MHC.

Chapter 2

General Methods and Resources

This chapter describes the general experimental resources and methods used throughout this project. Laboratory methods restricted to the work described in experimental chapters are included in those chapters. Critical laboratory reagents and the materials required for their implementation are summarized in three tables at the end of this chapter.

2.1 Sources of sheep DNA

Sheep from two panels of animals maintained as part of the Rylinton Merino Project (Western Australia) were used as donors of blood for extraction of individual DNA samples. The gastrointestinal parasite resistance status of these animals, as measured by worm egg counts (WEC), was known. The two panels comprised:

(i) Twenty-five sheep from an independently confirmed pedigree. The pedigree consisted of a half sibling family covering three generations. Sheep blood was freshly collected into 5 ml EDTA tubes, mixed and kept at -20°C until the DNA was extracted. Semen straws from sires were obtained from the Department of Agriculture and Food of Western Australia (DAFWA), and had been stored in liquid nitrogen prior to DNA extraction. The family relationships of these animals were confirmed by testing with panel of tetranucleotide microsatellite markers representing distinct chromosomes (Dr David Groth – personal communication).

ii A panel of fifty eight (58) unrelated merino sheep provided by the Centre for High Throughput Agricultural Genetic Analysis (CHAGA) at Murdoch University, WA.

2.2 DNA purification

DNA was extracted from sheep blood and semen from the various sheep resources. The quality and quantity of the extracted DNA was determined by agarose gel electrophoresis (0.7%). All DNA was archived and stored at –20°C freezer for future use.

2.2.1 Sheep DNA

A Qiagen Tissue kit was used to extract genomic DNA from whole sheep blood. The standard DNA extraction protocol was applied with slight modification. Whole blood (500μl) was mixed with 1ml highly purified water and centrifuged at 13000rpm for 5min. The supernatant was removed and the procedure repeated. A pellet, composed mainly of white blood cell nuclei was then used for extraction of genomic DNA as per manufacturer's instructions.

A manual proteinase K/SDS based method was used to extract genomic DNA from sheep semen. Sheep semen (200μl) was mixed with 1ml of 1×TE buffer and centrifuged at 13000 for 5 mins. The pellet was resuspended in lysis buffer (800μl of WBC lysis buffer, 16 μl of 100% 2-Mercapto-ethanol, 8μl of 20mg/ml proteinase K) and incubated at 50°C until dissolved. The sample was extracted in five consecutive steps using one volume of phenol, phenol/chloroform (twice) and chloroform (twice) respectively. Sheep genomic DNA was then precipitated after the addition of 1 volume of isopropanol and incubation at RT for 2 hours. DNA fibre spooled out using a 10μl tip and redissolved in 100μl of 1×TE buffer.

2.2.2 Plasmid and cosmid DNA

QIAprep® Spin Miniprep Kit was used to extract plasmid and cosmid DNA from bacteria.

A single plasmid or cosmid colony was inoculated into 5ml of LB broth culture media containing ampicillin (100μg/ml for plasmid and 50μg/ml for cosmid). The culture was incubated at 37°C for 16 hours with constant agitation and then bacteria harvested for DNA extraction. The manufacturer's standard protocol was employed except that cosmid DNA was eluted with 50μl of Tris-HCl (pH8.0) at 60°C.

2.2.3 Bacterial artificial chromosome (BAC) DNA

QIAGEN Large-construct Kit was used to extract BAC DNA from bacteria.

A single BAC colony was inoculated in 3ml of LB broth culture media containing 12.5μg/ml chloramphenicol. The culture was incubated at 37°C for 8 hours with constant agitation. A portion of the culture (600μl) was transferred to an identical pre-warmed media (500ml) and incubated at 37°C for another 16 hours with constant agitation. The culture was then harvested and BAC DNA extracted according to the manufacturer's standard protocol.

2.2.4 DNA amplicons from PCR

PCR products were purified using the QIAGEN PCR clean up kit. The manufacturer's standard protocol was followed with the following modifications.

A pool of three 20μl PCR products was adjusted to a pH of 6.5 prior to loading onto the PCR cleanup column. Elution buffer (20μl) was used to elute amplified DNA from the columns.

2.2.5 Final purification of DNA products

Either ethanol or isopropanol was used to precipitate DNA following laboratory modifications such as RE digestion or ligation. Alcohols (2.5 volumes of 100% ethanol or 1 volume of 100% isopropanol) were added to the DNA solution and incubated at -20°C overnight. A DNA pellet was obtained following centrifugation at 13000rpm for 30 minutes. The pellet was then washed with 70% ethanol, air-dried for 10-15 minutes, and resuspended in either sterile hpH_2O or 1× TE buffer.

2.3 Subcloning of Cosmid and BAC DNA into Plasmid Vector pGEM-4Z

2.3.1 The pGEM®-4Z Vector

The subcloning of cosmid and BAC DNA was performed using the pGEM®-4Z vector supplied by Promega Life Science. This vector was chosen for this work because of its high copy number, presence of two sequencing primers binding site (M13 Forward and M13 Reverse) which allow bi-directional sequencing of insert DNA, colour screening (white/blue) for recombinants, and extensive enzyme repertoire in the multiple cloning region.

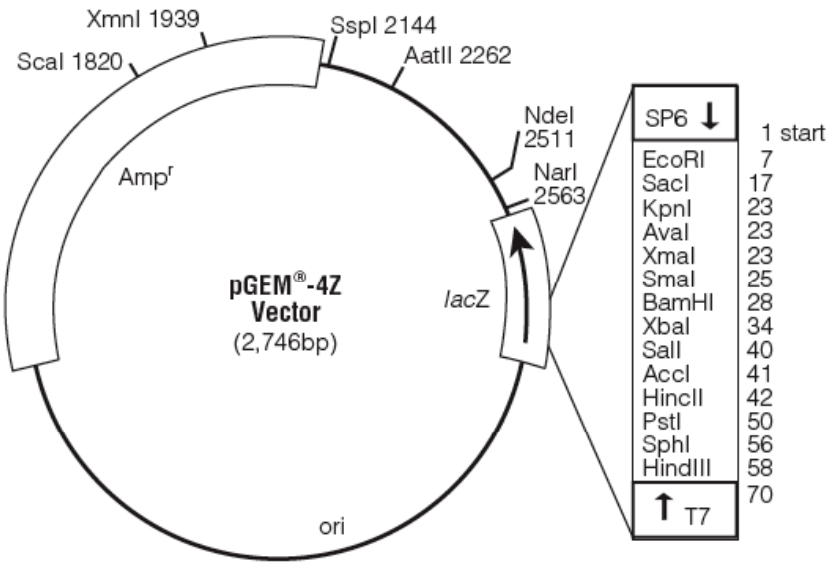


Figure 2.1: pGEM®-4Z Vector circle map and multiple cloning region. SP6 and T7 RNA polymerase promoters allows production of functional β–galactosidase that facilitates recombinants selection by colour screening. (Promega)

Target DNA and vector DNA was linearised through RE digestion, followed by purification. Vector DNA was treated with shrimp alkaline phosphatase (see section 2.3.2) to reduce self-ligation. Target DNA and vector DNA was then ligated at a 3:1 molar ratio of the target DNA to vector DNA. The recombinant plasmid DNA was transformed into ELECTROMAX™ DH5α-E™ *E. coli* Cells (Invitrogen) and plated onto agar plates containing 100 ug/ml Ampicillin. Finally, positive recombinants were identified by two selection methods: gel electrophoresis and PCR.

2.3.2 Restriction enzyme digestion of DNA isolates

A standard reaction including 17μl of DNA, 2μl of 10 × restriction buffer and 1μl of restriction enzyme was incubated at 37°C for 1-3 hours. Acetylated BSA (0.01 volume of 10mg/ml) was added when performing an overnight digestion. All restriction enzymes and their buffers were purchased from Promega Life Science.

Cosmid DNA was digested with Not1 and separated on 0.5% agarose gel. The large fragment that contains insert DNA was excised and purified using QIAquick Gel Extraction Kit according to manufacturers' standard protocol. Purified insert DNA, plasmid vector pGEM-4Z, and BAC DNA (section 2.2.3) were digested with BamHI and EcoRI respectively.

Vector PGEM®-4Z DNA was digested in a final volume of 50μl containing 5μg of vector PGEM®-4Z, 20 units of restriction enzyme, 0.5μl of BSA, 5μl of 10×buffer and the remainder hpH_2O. Digestion was followed by purification (section 2.2.5) and de-phosphorylation with shrimp alkaline phosphatase in a 50 μl reaction containing 40μl of DNA digest, 5μl (=5 unit) of shrimp alkaline phosphatase and 5μl of buffer. The reaction was incubated at 37°C for 15mins and terminated by heating up to 65°C for 15 minutes. Treated DNA was kept at -20°C for further cloning reaction.

2.3.3 DNA ligation

Ligation was performed at 15°C overnight in a 10μl reaction comprising 150ng of pGEM®-4Z vector, 50ng of insert DNA, 1μl 10 × ligation buffer and 1μl T4 DNA ligase (Promega). The reaction was purified by DNA precipitation (section 2.2.5) and the DNA pellet was resuspended in 5μl of hpH_2O. A portion of this ligation mix (1-2μl) was used for following transformation.

2.3.4 Transformation

Bacteria were transformed using a BioRad GenePulse II electroporation system. The ligation mix (1-2μl) was mixed with 15 μl of ELECTROMAX™ DH5α-E™ *E. coli* cells, briefly incubated on ice and transferred to an ice-cold electroporation cuvette. Electroporation was carried out using 1.8 kilovolts, 25μF capacitance, and 200 to 500 range. Successful transformation was indicated by time constants between 4.2 to 4.6 milliseconds. The transformation mix was then immediately mixed with 1 ml of SOC media and incubated with gentle agitation at 37°C for 1hour. A portion of above culture (50-100μl) was plated onto IPTG, X-Gal, ampicillin (100μg/ml ampicillin) LBA plates and incubated at 37°C for 16 – 18 hours.

2.3.5 Recombinant Selection

Following growth of transformation culture on IPTG, X-Gal, ampicillin (100μg/ml ampicillin) LBA plates, single white colonies were selected. Positive colonies containing correct insert DNA were selected using either of the these two methods:

The first method was using plasmid DNA extraction, followed by gel electrophoresis. Single white colonies were regrown in 5.5 ml of LB broth culture media (containing 100μg/ml ampicillin) for plasmid DNA extraction (section 2.2.2). Positive plasmid DNA was identified by their size comparison to a standard DNA marker together with unmodified pGEM®-4Z vector DNA on 0.7% agarose gel.

The second method was by PCR of single white colonies using M13 forward and M13 reverse primers, followed by gel electrophoresis. PCR was carried out for every single white colony in a 10μl reaction volume including 1μl of bacteria culture (i.e. a single white colony dissolved in 10μl sterile hpH_2O). Standard PCR reaction mix (Table 2.1) and cycling condition (Table 2.2) were applied. Positive amplicons were identified on 1.5% agarose gel by their size comparison to a standard DNA marker as well as with amplicons amplified from unmodified pGEM®-4Z vector DNA.

Residual bacteria from positive colonies on the original IPTG, X-Gal, ampicillin (100μg/ml) LBA plates were then picked off and cultured in LB media for further manipulation. A portion of the cultures was used to make glycerol stocks and these were stored at–70°C.

2.4 Generation of DNA probes for identification of locus specific clones and subclones

One of the main objectives for this study was to identify cosmid clones and BAC clones containing the central region of sheep MHC. However, very few sheep DNA specific probes for this particular region are available. Furthermore, there is very limited sequence data of sheep MHC in the NCBI database. Therefore, a new strategy was developed for generating DNA probes to identify specific gene loci within the sheep MHC region.

Sheep genomic DNA was amplified using consensus oligonucleotide primers identified from sequence alignment between mouse mRNA and human genomic DNA sequence of gene loci within the MHC region. Primers used in this study are summarized in Table 2.3. Standard PCR reaction (Table 2.1) and PCR cycling conditions (Table 2.2) were applied with minor modifications with respect to $MgCl_2$ concentrations and primer annealing temperatures.

Table 2.1: Standardized PCR Reaction mixes. A 10μl reaction contains:

Reaction component	Concentration
dNTPs	0.2mM each dNTP
$MgCl_2$	1.5mM
10 × buffer	1× buffer
BSA	0.2mg/ml
Primers	10pmol/μl each primer
Taq DNA polymerase	0.1Unit/μl
DNA template	50 ng

Table 2.2: Standardized PCR conditions for amplification of sheep genomic DNA, plasmid DNA, cosmid DNA and BAC DNA.

Step	Cycle Number	Temperature (°C)	Time (sec)
Denaturation	1×	95	6 minutes
Denaturation	1×	95	1 minutes
Annealing	1×	See Table 2.3	30 sec
Extention	1×	72	1 minutes
Denaturation	30×	95	20 sec
Annealing	30×	See Table 2.3	30 sec
Extention	30×	72	1 minutes
Denaturation	1×	95	20 sec
Annealing	1×	See Table 2.3	30 sec
Extention	1×	72	5 minutes
Hold		12	Infinitely

Table 2.3: Consensus primers for exonic sequences used to generate PCR amplicons for use as probes to screen Cosmid clones.

Gene	GenBank Acc. No.	Primer (5′to3′)	AT (°C)
BAT2	4337095(h) 9910175(m)	F: GATCARCTTCTCTTTCAAAG R: AGAGGCGAGTATTTTGCCAG	50
BAT3	4337095(h) 33147081(m)	F: GTTATYCACCTGGTGGAACG R: GGAAGATTGAAGGTTCCAAC	55
BAT4	4337095(h) 34784641(m)	F: GCTCACCTGCTGTCACTGTC R: AGTCAGAACTCGAGGTTCATGT	55
BAT5	4337095(h) 31341550(m)	F: ACGTGGGCAGCCATGTCCTA R: CAGCTGTCYGGCATGACCTT	52
BING4	2648017(h) 31982691(m)	F: GTGTCTTTATGGAGTCCAGC R: GTGCCTGTAGAATCTACTGC	52
CAT56	5926700(h) 28193528 (m)	F: GGTATTCTTGAYAAGAGACCAGC R: CAGRGGTCCTCCAGGATCAA	55
CDSN*	26787969(h) 28511280(m)	F: CAGCTCCAACCWGCGKCC R: GAGAAGTAYTTGCCCTCAGAGA	55
Col11a2*	3646023(h) 6753481(m)	F: TGGGYGTGGTCCAGCTCAC R: CTGRCAGCCATCTCTGAA	52
DAXX*	2961406(h) 6681134(m)	F: GAGTTCTGCAACATCCTCTC R: CTSAGAGGCAGTGTTTTCAG	52
G7c	4529886(h) 7381104(m)	F: CTCTGCGTTTTGARCCATA R: CCATGCTCTCCCCAACAAT	55
G6D	4337095(h) 15638949(m)	F: ACAGTCCTGGCAAGAGMCAG R: CCCATCATTGCAATCAAGTGG	55
Lmp2*	34634(h) 309424(m)	F: GGAGTCAGRACYCCAGRGCTT R: MGTGTTYGACAAGCTSTCCCC	52
LTA	4337095(h) 7106342(m)	F: GAGGARAAGAGCTGGACCTC R: CAGAACTCACTGCTCTGGAGA	52
MSH5*	4337095(h) 7305280(m)	F: CTCTCTGGAGAAGCTTKAGGC R: GATCCATCTGTGNGTGCTGTG	55
BAT8	4529886(h) 3986763(m)	F: TTCATAGCTCTTTGGGGGACA R: CCATCTCCCTCAAGRYTCTC	55
NOTCH4	29804415(h) 6754873(m)	F: TATGAGGGACAGAACTGCTCA R: CTGTGTGYCCAGGCAGACACT	55
RING1*	3646023(h) 31982045(m)	F: GCGATGTAGATGGTGTACTGCT R: AGTACYTGGCCCTGCGCATT	55
RING3	20068658(h) 47059182(m)	F: CCACAGGCTGCCGGAATGGCA R: AATGCAGGGTTGCTGGGGCT	59
TAP2*	1054740(h) 1399928(m)	F: GAGCTGAACTCRCGGCTGAG R: GCGGGKGTTGTACACCTTCT	55
TAPBP	18375876(h) 15929261(m)	F: CTGYCTTGYRTCCCACTTCT R: CCAGGGTGACCTCAGCRCTG	55
TNFα	4337095(h) 7305584(m) 512016(ovine)	F: ACTTTATTTCTCGCCACTG R: YGTGAAAACGGAGCTGAAC	55

Note: All primers were ordered from GeneWorks Pty Ltd. Mixed base codes: R(AG) Y(CT) M(AC). AT: Annealing temperature. The primers for loci identified with an asterisk were not locus specific and were not further used.

PCR amplicons were cloned into pGEM®-T easy vector using pGEM®-T easy vector systems (Promega). A 10μl reaction containing 150ng PCR product (either purified or not purified) and 50ng PGEM-T easy vector was carried out according

to the manufacturer's standard protocol with minor modification. The reaction was incubated at RT for 1 hour and left at 4°C for 1.5 hours, followed by purification (section 2.2.5). The DNA pellet was resuspended in 5μl hpH_2O. A portion of the DNA solution (1μl) was used for transformation (section 2.3.4).

Following recombinant selection (section 2.3.5), the positive plasmid recombinant DNA was purified from bacterial culture and sequenced using LI-COR DNA 4000 sequencer (section 2.6.1). The identity of sequences was confirmed by searching the NCBI database for orthologous gene loci. Recombinant DNA with desired gene identity was used for probing sheep cosmid and BAC libraries.

Two distinct probes were used in this study: probes generated from cloning of PCR amplicons (see Table 3.2 in chapter 3) and those from overlapping oligonucleotides. The following two labelling systems were employed respectively for these two different DNA probe systems.

2.5 DNA probe labelling and hybridisation

DNA probes were radioactively labelled, and denatured prior to being added into the hybridisation buffer for overnight hybridisation to the sheep cosmid and BAC libraries.

2.5.1 Prime-a-Gene® Labelling Probe

Prime-a-Gene® Labelling system (Promega) was used to label probes generated from cloning of PCR amplicons. A standard labelling reaction containing 25ng of DNA probe was prepared according to the manufacturers' protocol. The reaction was incubated at RT for 1hour.

2.5.2 Labelling overlapping oligonucleotide (Overgo)

Overlapping oligonucleotide primers were generated from two different resources: DNA probe sequences obtained from cloning of PCR amplicons and also those derived from conserved coding sequences obtained by alignment of mouse mRNA sequences and human genomic DNA sequences of gene loci within the sheep MHC region. Overgo primers were designed using the Overgo Maker Program (http://www.genome.wustl.edu/tools/?overgo.html). These sequences and primers are summarised in Table 3.2 and Table 3.3 (Chapter 3) respectively, and therefore are not repeated here.

Overgo primers were labelled using the method of Gustafson *et al*. (2003) with minor modification. Briefly, a 10μl reaction was set up for each probe comprising 1μl each of 10μM forward and reverse primer, 1μl each of 250μM dTTP and dGTP, 1μl of 3000Ci/mmol each of ^{32}P dATP and ^{32}P dCTP, 1μl (2U) Klenow fragment DNA polymerase, 1μl 10×buffer and 2μl hpH_2O. After incubation at 37°C for 30 minutes, 1μl each of cold 250μM dATP and dCTP was added and incubated for another 15 minutes. Following incubation, reactions for all probes were pooled and purified to eliminate unincorporated nucleotides through a Sephadex G-15 gravity column using PBS as running buffer.

2.5.3 Hybridisation

Radioactively labelled DNA probes were denatured at 96°C for 5 minutes before immediately added into hybridisation buffer for overnight hybridisation. Two distinct hybridisation methods were used in this study.

2.5.3.1 Hybridisation of the sheep cosmid library

Positively charged nylon transfer membranes (PCNMs) were overlaid on agar containing sheep cosmid library for 5 mins and removed on a sheet of filter paper with colony side facing up. After air drying for 15 mins, PCNMs were placed in a 60°C oven until thoroughly dry. The PCNMs were then treated with 10% SDS for 2 mins, followed by transfer buffer for 10 mins and finally with 1 × TAE buffer for 1 mins. After drying completely at 60°C, the PCNMs were sealed in plastic together with 20ml hybridisation buffer and prehybridised at 60°C for 30 mins. Following prehybridisation, approximately 10ml hybridisation buffer was removed and replaced with 5ml of fresh hybridisation buffer (60°C) containing the denatured DNA probes. The plastic bag was resealed, and incubated at 60°C overnight with gentle agitation.

2.5.3.2 Hybridisation of the sheep BAC library

The dry ovine CHORI-243 BAC library filters (PACBAC Resources) were wetted with hpH_2O, rolled in a sheet of nylon muslin and placed in a hybridisation bottle. Following the addition of 20ml hybridisation buffer, the bottle was incubated in a hybridisation oven (Hybaid) at 53°C for 40 mins. After prehybridisation, the buffer was removed and replaced with 10ml of fresh hybridisation buffer containing the denatured DNA probes. The bottle was then incubated at 53°C overnight with constant rotation.

2.5.4 Post hybridisation washing and film development

After overnight hybridisation, filters were washed twice with washing buffer (1× SSC, 0.1% SDS) at 60°C for cosmid PCNM or 53°C for BAC filters for 30 mins, followed by a brief rinse in 1× SSC buffer. Wet PCNMs or BAC library filters were sealed in plastic and exposed to Kodak X-ray film using an intensifying screen at -80°C for 2 days. The film was developed in Kodak developer for 2 mins, briefly rinsed in tap water for a few seconds and then fixed in Kodak fixer with gentle agitation until the background turned clear. The film was then cleansed in tap water and dried at 37°C. After autoradiography, PCNMs and filters were treated with transfer buffer for 10 mins, followed by neutralisation buffer (1× TAE) twice. Moist PCNMs and filters were sealed in plastic and stored at -20°C for future hybridisation.

A pool of positive clones was identified after the primary hybridisation. A secondary screening was performed by PCR amplification of positive clones from the pool using primers listed in Table 2.3. Positive cosmid and BAC clones were then subcloned into pGEM®-4Z vector (section 2.3).

2.6 DNA sequencing

The main thrust of this project entailed sequencing of cloned DNA fragments generated by PCR amplification and recombinant plasmid DNA fragments generated by subcloning of sheep cosmid and BAC DNA. This was performed in two ways. Slab gel-based DNA sequencing was performed on a LI-COR DNA 4000 automated sequencer using an infra-red dye labelled primer. Capillary-based DNA sequencing was performed on an ABI 3100 automated sequencer using an unlabelled primer and four individual dye labelled dNTPs. Due to the high cost of infra-red dye labelled primers, the LI-COR DNA 4000 automated sequencer was only used for fragment analysis (i.e. microsatellite assay in this study) and sequence analysis of plasmid DNA and cosmid DNA from both ends. Sequencing of large DNA fragment using internal primers was all performed on the ABI 3100 automated sequencer.

2.6.1 LI-COR DNA 4000 sequencing

Sequencing reaction were performed using a SequiTherm EXCEL™ II DNA Sequencing Kit (Epicentre) according to the manufacturers' protocol. Infra-red

dye labelled M13 forward/reverse primers and T3/T7 primers were used for sequencing plasmid DNA and cosmid DNA respectively. The PCR reaction was initiated at 95 °C for 5 minutes, followed by 30 cycles of 95 °C for 30 seconds, 51°C for 15 seconds and 70°C for 1 minute. Following denaturation, a portion of product (0.7-1.0μl) was resolved on a 4% TBE denaturing 66 cm polyacrylamide gel running at 45°C and 2000 Volts overnight. Sequencing data were collected in real time by computer.

IRD-labeled M13 Forward primer: CACGACGTTGTAAAACGAC

IRD-labeled M13 Reverse primer: GGATAACAATTTCACACAGG

IRD-labeled T3 primer: CGCAATTAACCCTCACTAAAG

IRD-labeled T7 primer: GCATAATACGACTCACTATAG

2.6.2 ABI 3100 automated sequencing

Plasmid DNA (section 2.2.2) was further cleaned by ethanol precipitation (section 2.2.5) and resuspended in 20μl sterile hpH_2O. A 10 μl sequencing reaction including 2μl Dye terminator mix, 1μl (3.2 pmoles) primer, 1μl 5× sequencing buffer, 3 μl DNA template (350ng) and 3μl H_2O was carried out for 25 cycles as follows: 96 °C for 10 seconds, 50-60°C for 15 seconds and 60°C for 4 minutes. After amplification, the PCR product was precipitated with 25μl of 100% ethanol after the addition of 1μl of 3M sodium acetate pH5.2 and 1μl of 125mM EDTA (disodium salt). The DNA pellet was air-dried and submitted to CHAGA ABI sequencing centre.

For sequencing directly from cosmid and purified PCR product, 800ng and 25ng DNA template was used respectively. Annealing temperature was adjusted according to sequencing primers' TM value.

2.6.3 DNA Sequence Analysis

For the LI-COR sequencing method, sequences were checked and scanned manually on the computer. Any ambiguous sequence was resequenced or sequenced using both forward and reverse primers if required for accuracy. Sequence contigs were determined using the Vector NTI program (Invitrogen). Sequence identity was achieved by searching the GenBank database. Sequence translation and restriction digestion maps were obtained using DNA Strider™ 1.3 software (Marck, 1988).

2.7 Detection of SNP and microsatellite loci

Single nucleotide polymorphisms (SNPs) were identified by comparing locus specific sequences from several unrelated sheep. In brief, oligonucleotide primers spanning the gene of interest (within the central region of sheep MHC) were used to generate genomic PCR products which were then sequenced directly. Sequenced PCR products were generated from at least five randomly selected unrelated Australian merino sheep. SNP's were confirmed when at least two heterozygotes and one of each homozygote were observed. Microsatellite loci were sought by interrogating sequence files with RepeatMasker software (Smit *et al.,* 1996).

Details of the methods used to generate SNP and microsatellite genotypes for the animals comprising the two panels available for this study are described in more detail in Chapter 5.

2.8 Sources of materials and composition of reagents

The sources of significant materials used in this project are summarized in Table 2.4 and the compositions of significant reagents are summarized in Table 2.5.

Table 2.4: The sources of important materials used in the study

Agarose Amersham Pharmacia Biotech	**Boric Acid** Solon Ind. Pkwy.
BSA Roche Diagnostics Corporation	**[α-^{32}P]dATP & [α-^{32}P]dCTP** PerkinElmer Life Sciences
DNA marker (1Kb) Gibco Life Technologies	**dNTPs** Invitrogen Life Technologies
EDTA MERCK Pty. Limited	**ElectroMAX DH5α-E cells** Invitrogen Life Technologies
Electroporation Cuvettes (0.1 cm) Invitrogen Life Technologies	**100% Ethanol** MERCK Pty. Limited
GIBCO ultra PURE Water Invitrogen Life Technologies	**Hybond-N+ Membrane** Amersham Biosciences
Instant Black&White Film FP-3000B Fujifilm	**IPTG** PROGEN Industries Limited
100% Isopropanol MERCK Pty. Limited	**Kodak Scientific Imaging Film** Kodak
LongRanger Gel Solution BioWhittaker Molecular Applications	**Mastercycler gradient** Eppendorf
Mastercycler ep gradient S Eppendorf	**Ovine CHORI-243 BAC library** PACBAC Resources (USA)
PCR clean up kit Promega cat no. A9281	**PCR buffer & $MgCl_2$** Invitrogen Life Technologies
pGEM-4Z Vector Promega Life Science	**pGEM-5Zf(+) Vector** Promega Life Science
pGEM-T Easy Vector System I Promega Life Science	**Prime-a-Gene Labeling System** Promega Life Science
Platinum Taq DNA Polymerase Invitrogen Life Technologies	**PSQ HS96 pyrosequencing kit** Biotag, Uppsala, Sweden
PTC-100TM Programmable Thermal Controller MJ Research, Inc	**Pulse controller Plus & Gene PulseR II** Bio-RAD
QIAEX II Gel Extraction Kit (150) QIAGEN cat no. 20021	**QIAprep Spin Miniprep kit (50)** QIAGEN cat no. 27104
QIAquick Gel Extraction Kit (50) QIAGEN cat no. 28704	**Restriction Enzyme** Promega
SDS Solon Ind. Pkwy.	**SequiTherm EXCEL II sequencing kit** Epicentre Technologies
Shrimp Alkaline Phosphatase Promega Life Science	**Sodium Chloride** Solon Ind. Pkwy.
Sodium Hydroxide CHEM-SUPPLY PTY LTD	**Streptavidin Sepharose Beads** Amersham, Biosciences, Uppsala, Sweden
TEMED ICN Biomedicals Inc.	**T4 DNA ligase** Promega Life Science
Tri-sodium Citrate MERCK Pty. Limited	**Tris** Solon Ind. Pkwy.
Urea Spectrum Chemical Mfg. Corp.	**X-Gal** PROGEN Industries Limited

Table 2.5: The composition of important reagents used in experimental work.

Agaros gel loading buffer Bromophenol blue (0.25%) 12.5mg Xylene cyanol (0.25%) 12.5mg Glycerol 1.5ml Dissolved in hpH_2O, and adjusted to a final volume of 5.0ml.	**Ampicillin/IPTG/X-Gal LBA Plate** 100mM IPTG 100µl 50mg/ml X-Gal 40µl Ampicillin 10mg Dissolved in 100ml 55°C LBA and pour onto 6 petri dish.
1× Annealing Buffer Tris-Acetate 20mmol/L $MgAc_2$ 2mmol/L	**10% APS** APS 0.5g Dissolved in 5ml of hpH_2O and stored at 4°C.
2× Binding Buffer (pH7.6) Tris-HCl 10mmol/L NaCl 2M EDTA 1mmol/L Tween20 0.1%	**10% Ethidium Bromide (w/v)** EtBr 0.1g hpH_2O 1 ml Dissolved in hpH_2O, and wrapped with light proof tape to avoid sunlight.
0.25 M HCl solution 32%HCl 56 ml Made up to a final volume of 2 liter with dpH_2O, stored at RT.	**Hybridisation Buffer** 20 × SSPE 250 ml BSA 10 g SDS 70 g 0.5M EDTA (pH8.0) 5 ml Made up to a final volume of 1 Liter.
0.1M IPTG IPTG 1.2g Dissoved in sterile hpH_2O, and djusted to a final volume of 50 ml stored at -20°C.	**1 N NaOH** NaOH 40g Dissolved in hpH_2O, and made up to 1 Litre in hpH_2O
PBS Buffer NaCl 8g Kcl 0.2g Na2HPO4 1.44g KH2PO4 0.24g Dissolved in hpH_2O, and pH was adjusted to 7.4. Made up to a final volume of 1 Liter. Sterilized by autoclaving.	**4% Polyacrylamide Gel** Urea 25.2g Long Ranger 4.8ml 10 × TBE 7.2ml Dissolved in hpH_2O and adjusted to a final volume of 60ml. 400µl of 10%APS and 40µl of TEMP were added in and mixed gently without introducing air bubbles.
6 × Loading buffer (for agarose gel) Bromophenol blue 0.25% Xylene cyanol FF 0.25% Ficoll (type 400) 15% SDS 0.1%	**20 × SSPE Buffer** NaCl 175.3g $NaH_2PO_4 \bullet H_2O$ 27.6g EDTA 7.4g Dissolved in hpH_2O, and pH was adjusted to7.4.Made up to a final volume of 1 litre.
10% SDS solution SDS 100g Dissolved in dpH_2O and adjusted to 1 liter.	**20 × SSC Buffer** NaCl 175.3g Sodium citrate 88.2g Dissolved in hpH_2O, and pH was adjusted to7.0 with a final volume of 1 litre.
0.5× SSC, 0.1% SDS 20× SSC 50 ml 10%SDS 20 ml Dissolved in dpH_2O and adjusted to a final volume of 2 liter.	**50× TAE Buffer** Tries 242g $Na_2EDTA{\cdot}2H_2O$ 37.2g Glacial acetic acid 57.1ml Dissolved in hpH_2O, adjust to1 litre.
10 × TBE Buffer Tris 162g Boric Acid 27.5g $EDTA{\cdot}2H_2O$ 9.3g Dissolved in hpH_2O, adjust to1 litre.	**1 × TE Buffer (pH 8.0)** Tris-HCl 1.21g EDTA 0.372g Dissolved in hpH_2O, and pH was adjusted to 8.0. Made up to final volume of 1 Liter.
Transfer Buffer NaOH 0.4 M NaCl 0.6 M Dissolved in dpH_2O and adjusted to 1 liter.	**X-Gal (50mg/ml)** X-gal 100mg Dissolved in N,N'-dimethylformanide, and adjusted to a final volume of 2 ml in a light proof tube and stored at -20°C.

Chapter 3

A Map of the Central Region of the Sheep Major Histocompatibilty Complex (MHC)

This chapter describes the generation of a physical map spanning approximately 700 kbp of the central region of the sheep MHC. The map was obtained by ordering overlapping clones from a sheep bacterial artificial chromosome (BAC) library and cosmid libraries of genomic DNA. Subsequently, specific loci within these clones were identified using subcloning and sequencing. The work described herein follows and extends an earlier related study undertaken by supervisors Drs Groth and Wetherall using cosmid clones to map complement C4 and CYP21 genes in sheep which at the time this project was undertaken was unpublished. Since the cosmid study enhances the BAC map obtained it is included in this chapter and is identified by text in italic type. A manuscript containing the work presented in this chapter has been accepted for publication in the journal BMC Genomics.

3.1 Introduction

As was reviewed in Chapter 1, the extreme polymorphism and extensive associations with diverse diseases have made the MHCs of mammals a region of significant interest for agricultural research, especially in species of economic importance such as ungulates. In comparison with other mammals, the sheep and cattle MHCs are poorly characterized. A study by Paterson and colleagues has shown that the broad structure of the sheep MHC region appears to be similar to its human counterpart (Paterson *et al*., 1998), with a central region separating class I and II regions. More recently Liu and colleagues (Liu *et al*., 2006) also used BAC clones to provide a broad description of the sheep MHC and confirmed that a central region flanked by class I and class II regions existed which contained loci orthologous with those present in the human MHC central region. This work provides a provisional framework for the more detailed mapping of this region.

Although serological identification of sheep class I proteins was undertaken many years ago (Outteridge *et al*., 1988), characterization of polymorphic loci in the sheep MHC has been relatively slow. Several microsatellite loci have also been mapped to the sheep MHC; microsatellites occur within the OLADRB and OLADRBps loci in the class II region (Schwaiger *et al*., 1996; Blattman and Beh,

1992) while OMHC1 (Groth and Wetherall, 1994) is a dinucleotide microsatellite within the class I region and the BfMs microsatellite locus (Groth and Wetherall, 1995) maps close to the CFB locus. All of these markers have been associated with resistance in sheep to intestinal parasites (Paterson *et al*., 1998; Bot *et al*., 2004) but in general terms, these studies have lacked sufficient power for identification of potential disease predisposing candidate genes.

In order to gain a better knowledge of the central region of the sheep MHC and its contributions to health and disease we sought to construct an ordered BAC contig map of this region based on characterization of clones from a freely available sheep BAC library (CHORI 243) using comparative analysis of loci typically present in the central regions of the human and mouse MHCs. This study has resulted in the first low-resolution physical map covering ≈700kbp of the sheep central region together with a panel of SNPs. Both should facilitate identifying partial MHC haplotypes in sheep for future disease association studies and provide insights to the presence of subregions within the sheep MHC.

3.2 Materials and Methods

3.2.1 Strategy for identification of sheep MHC central region orthologues of human and mouse genes

Primers amplifying exonic regions of sheep orthologues of human and mouse MHC central region genes were identified from an alignment of genomic human sequence with its orthologous mouse cDNA sequence. The loci chosen and the primers identified are summarised in Table 3.1. These primers were used to amplify sheep DNA and the amplicons were cloned into the pGEM-T easy vector and sequenced. The resulting sheep DNA sequences, together with the pairs of overgo primers derived from these sequences, which were used to screen the sheep CHORI-243 BAC library, are summarized in Table 3.2. Overgo primers were designed using the Overgo Maker Program and summarized in Table 3.3 (http://www. genome.wustl.edu/tools/?overgo.html).

During the course of this study end sequences of many of the BAC clones of the CHORI-243 library were published in GenBank. Five of these sequences permitted identification of additional genes within the sheep MHC central region while the other five BACs were subcloned and re-screened for specific loci. Finally a contig map was generated from analysis of the 10 BAC clones which spanned ≈700kbp of the putative central region of the sheep MHC. This map was the result of

identifying genes in sequences derived from 5 subcloned BACs and identification of genes in the other 5 BACs with published end sequences. The accession numbers of the 10 BAC clones used in this study are summarized in Table 3.4.

In a related earlier study, undertaken by Drs Groth and Wetherall also at Curtin University, a sheep (male) genomic cosmid library (Clontech) was screened with a variety of probes to identify cosmids containing the complement gene region of the sheep MHC. Direct sequencing of cosmid BfC4.2.20 resulted in sequence identification of the C2/CFB subregion.

SNP discovery was performed by designing primers to exonic regions of the sequenced loci. These primer sets were designed to amplify a ≈500bp region. SNPs were then identified from alignments of these sequences from at least five unrelated animals or occasionally by alignments with corresponding sequences from GenBank. SNPs were confirmed when at least two heterozygotes and one of each homozygote for each locus were observed.

Table 3.1: List of consensus primers from human and mouse exonic sequences used to generate sheep amplicons for use as probes to screen BAC clones.

Locus	MHC region	GenBank Acc No	Primer (5′to3′)	Product Size (bp)
Bat3	central	23093110(h) 33147081(m)	F: GTTATYCACCTGGTGGAACG R: GGAAGATTGAAGGTTCCAAC	181
BAT4	central	4337095(h) 34784641(m)	F: GCTCACCTGCTGTCACTGTC R: AGTCAGAACTCGAGGTTCATGT	309
CAT56	class I	18181758(h) 28193528 (m)	F: GGTATTCTTGAYAAGAGACCAGC R: CAGRGGTCCTCCAGGATCAA	251
G7c	central	18673924 (h) 7381104(m)	F: CTCTGCGTTTTGARCCATA R: CCATGCTCTCCCCAACAAT	112
G6D	central	4337095(h) 51243039(m)	F: ACAGTCCTGGCAAGAGMCAG R: CCCATCATTGCAATCAAGTGG	169
BAT8	central	4529886(h) 3986763(m)	F: TTCATAGCTCTTTGGGGGACA R: CCATCTCCCTCAAGRYTCTC	190
NOTCH4	central	20152668(h) 6754873(m)	F: TATGAGGGACAGAACTGCTCA R: CTGTGTGYCCAGGCAGACACT	241
RING3	class II	20068658(h) 47059182(m)	F: CCACAGGCTGCCGGAATGGCA R: AATGCAGGGTTGCTGGGGCT	265
TAPBP	class II	18375876(h) 15929261(m)	F: CTGYCTTGYRTCCCACTTCT R: CCAGGGTGACCTCAGCRCTG	259
TNFα	central	4337095(h) 7305584(m)	F: ACTTTATTTCTCGCCACTG R: YGTGAAAACGGAGCTGAAC	116

All primers were ordered from GeneWorks Pty Ltd. Mixed base codes: R(AG) Y(CT) M(AC).

Table 3.2: Sheep DNA sequences of amplicons generated using the above primers. Sequences in red colour were used for generating overgo primers for screen sheep BAC library.

Locus	Sheep DNA Sequence
BAT3	GGAAGATTGAAGGTTCCAACCACGACATAGCTGTTGGCATTCCGGTCAT GAACAGAGGCCCCAGGCCCCGAGTACCAGGCGTGGGTCCCCCACCATG GGTGCTGAGGTGGACCCTGTTCCAGAAGATGCTCCGGAAGGGAGCTGA GTCTGAGGAGGAGCCCGTTCCACCAGGTGAATAACA
BAT4	GGAATTCACTAGTGATTAGTCAGAACTCGAGGTTCATGTATGTCCGCAG ATCCCGTCCCAGGCCTTGTCCTTCTCTTCCTGCCTTTTCTCCTCCCTGCGG TTCAGTGTGGTCACCCTAGGGGCTCTCTCCCTCCCAGCCACTGCCCGTAT ATCCCGAGCTGGGAAACGTGTAACTCGGGGCTGAGGTGCTGGTCTGTAG CCTAACCCCTCCTGGTCCCTCTTGAGGACAGTGGGGATAGGGTTGGCAC GGCCCTCACCCCGGGGTCCTAGCCCATTCCAGGCTCCCAGCCCCCCTCAG CAGCAGCTTGA
CAT56*	CAGAGGTCCTCCAGGATCAAGGGGACCAATGATCCCACCACTGCTGAGT CTCCCACCTCCTCCCCGGGGCAGAGGCCCAATTCGGGGAGGCCTAGGCC CCAGGTCTGGCCCATATGGTCGTGGTTGGTGGGGGGTCAATGCCGAGCC TCCTTTCCCTGGACCAGGCCACGGGGGTCCTTCCAGGGGAGGCTTTCAC AAAGAACAGAGAAATCCTCGAAGGCTCAAAAGCTGGTCTCTTGTCAAG AATACCA
C2	CCTGGAAACAAAGTAGCAGATTAGGCTGGACCAGGGCTCCTGAAGGGG CCAGAGGCTGGGAACAAGGGGTGGGAGCTCTGCCCAGGGGATCGGGAA TCTGGTGCGGGCCCTTCCTTACCCACACTGACCCTCCACAGTGTTCTGTC CTCGGCATTGCTTAAGCAGTGAGCGGCTGTCAGGACCCACTGGTCGGA GATGAGGGCCCCCCGGCAGGTCTCCTGGCTCTTGGGCTGCAGGGGGAC AGGGAATCTCGGGGACTGCTATGTCTCTGGCTTGTGGCCACACCTCTG
G6d	CATCATTGCAATCAAGTGGAGACAGAAGTGGTGGGAGACGTGACTTATA CGACCCACAGGGACTGCTGCGTTGGTGACCTGTGCAACGGTGCTGTGGC GAGCACTGCAGCCCCCATGAGCATCGTGGCTGCAGCAGTCACCACGCTG GCCTGTCTCTTGCCAGGACTGT
G7c	CTCTGCGTTTTGAGCCATACGAGGCAGTGGCTCTGGCCTCAGGAGGAGAA GTGATTTTCACCGAAGACCAGTATATTCAGGATGTGGCGGCCATTGTTGG GGAGAGCATGGA
BAT8.1	CATAGCTCTTTGGGGGACACCCTCGTAGTGAGGAGACCCTGCCCAAGGC CAACCCTGACTCCCTGGAGACTGCTGGCCCCTCATCCCCAGCCTCTGTCA CGGTCACTGTCGGCGATGAGGGGGCTGACACCCCTGTAGGGCCACACCA CTCATTGGGGATGAACCTGAGAACCTTGAGGGAGATGG
BAT8.2	CCATCTCCCTCAAGGTTCTCAGGTTCATCCCCAATGAGTGGTGTGGCCCC TACAGGGGTGTCAGCCCCCTCATCGCCGACAGTGACCGTGACAGAGGCT GGGGATGAGGGGCCAGCAGTCTCCAGGGAGTCAGGGTTGGCCTTGGGCA GGGTCTCCTCACTACGAGGGGTGTCCCCAAAGAGCTATGAAA
NOTCH4	TATGAGGGACAGAACTGCTCAAAGGAACCAGACGCGTGTCAATCCCAGC CCTGTCACAACCAGGGGACCTGCACCTCCAAACCCGGAGGCTTCCACTGT GCCTGCCCGCCGGCTTTTGTGGGGCTGCGCTGTGAGGGGGACGTGGATGA GTGTCTGGACCGGCCCTGTCACCCCACAGGCACTGCAGCCTGCCATTCTC TGGCCAACGCCTTCTACTGCCAGTGTCTGCCTGGACACACAG
RING3*	CCACAGGCTGCCGGAATGGCCATGCAAACTGATGTTTCCATAGAGCCTTC ATCACCACCTTGTGCAGATATTGCAACTGGTTGGTAACCCGTCCTGGCTTT TTGGGATTGGACACCTCAGGAGGTGGTGGGTTGGCAGGGGTTAGTTGCAA AGCAGGCACTGAAGCCATTGTGGGGCTCTCGAATCCCTCATACAATAGAG AGGGCTTTCGAATCCTTTTCCCAGGCGCTGCTGCTTCTGGGCCCAGCCCCA GCAACCCTGCATT
TAPBP*	CCAGGGTGACCTCAGCGCTGCGCCCCAGGGCGGGCAGGCTGGGGTGGTGG ACCCGGCAGGCATAGCGGGCCCATGCTGCTCACTGGTGACCGGGGGCGGC TGCAGGTGTGCAGAGAGGCTTACAGAGCCATCTGAGTGATGTCTCAGGGC CGAGAGCCACCTCTGGCCCTCAGCCTTCTGAAAGCGACCCTCTGGGCCACC CCGAAGCTCCCACTCCACCTCCAGGCCCTGGGAGGGGTAGAAGTGGGACA CAAGACAG
TNFa5	ACTTTATTTCTCGCCACTGACCAGTAGGCGGTTACAGGCATGACTCCCCTG GGGAGCGGAGGTTCAGTGATGTAGCGACAAATCAGTCACCAAATCAGCAT CATTTAGACAACTTG

Note: Loci identified with '' occur within the class I and II regions of the MHC. They were detected in this study but not pursued further.*

Table 3.3: Locus specific overgo primers used to identify sheep genes in BAC clones

Gene	Forward Primer (5'-3')	Reverse Primers (5'-3')
BAT1[R]	TGGCAGAGAACGATGTGGACAATG	TCATAGTCCAAGAGCTCATTGTCC
BAT1[R]	GAAGCAGGTCATGATGTTCAGTGC	CTCTTTGCTCAAGGTAGCACTGAA
BAT3[R]	CTGTTCATGACCGGAATGCCAACA	CCAACCATGACATAGCTGTTGGCA
BAT3	AAGGTTCCAACCACGACATAGCTG	ATGACCGGAATGCCAACAGCTATG
BAT4	CTTCTCTTCCTGCCTTTTCTCCTC	CACACTGAACCGCAGGGAGGAGAA
CAT56	GGGAGGCTTTCACAAAGAACAGAG	GAGCCTTCGAGGATTTCTCTGTTC
C2	AAGCAGTGAGCGGCTGTCAGGACC	TCATCTCCGACCAGTGGGTCCTGA
C2[R]	AATCCATGACTCCTGCATGGCATG	CACCCCAGATTGTATGCATGCCAT
C4	AGATGGTTCCTATGGGGCTTGGTT	GCTACTATCCCGATGTAACCAAGC
G6D	GCAATCAAGTGGAGACAGAAGTGG	TAAGTCACGTCTCCCACCACTTCT
G7C	TCACCGAAGACCAGTATATTCAGG	ACAATGGCCGCCACATCCTGAATA
HSPA1B[R]	TGTCCATCCTGACGATCGACGACG	TTCACCTCGAAGATGCCGTCGTCG
MHC414	AGCAGGAGGGGCAGGAGTATTGGG	AATCGCGTCTCCTGATCCCAATAC
MHC517	GACGAGGTTTCATTTTCAGTTTGG	CACGCGGTGATTAAACCCAAACTG
MSH5[R]	ATTCATGGTTCTGGCCCCACCTCT	TCTCAAGCTTCTCCAGAGAGGTGG
BAT8	ATCTCCCTCAAGGTTCTCAGGTTC	ACCACTCATTGGGGATGAACCTGA
NOTCH4	CAGAACTGCTCAAAGGAACCAGAC	GCTGGGATTGACACGCGTCTGGTT
RING3	TAGAGCCTTCATCACCACCTTGTG	CCAGTTGCAATATCTGCACAAGGT
TAPBP	TGCAGAGAGGCTTACAGAGCCATC	CCTGAGACATCACTCAGATGGCTC
TNF	ATTTCTCGCCACTGACCAGTAGGC	AGTCATGCCTGTAACCGCCTACTG
TNF[R]	TACCTCATCTACTCCCAGGTCCTC	AGCCTTGGCCTTTGAAGAGGACCT
TNXB	CACCTACTTTCCAGGTATTCCAGC	TTTCCATCCATGCGGCGCTGGAAT

Superscript R: Obtained from Gustafson et al., 2003.

Table 3.4: GenBank accession numbers for the ten BAC clones used in Figure 3.1. Clones shown in normal font were subjected to additional direct sequencing for confirmation and identification of specific loci. Clones shown in bold font are those for which the published end sequences in GenBank contained orthologues of genes from the human MHC central region.

Clone ID	GenBank Acc No (gi)
CH243-257B22	**77151234/77143173**
CH243-265D24	77125512/77125520
CH243-295I15	**77241001/77237208**
CH243-302L6	**77296914/77298047**
CH243-340A7	77275575/77263741
CH243-348O13	**76832624/76828474**
CH243-349H16	76828020/76826237
CH243-443I20	76938294/76938211
CH243-487C7	na
CH243-521I18	**77339223/77338595**

3.2.2 Labelling of DNA and oligonucleotide probes

3.2.2.1 Labelling of DNA

The Prime-a-Gene® Labelling system (Promega) was used to label probes generated from cloning of PCR amplicons. A standard labelling reaction containing 25ng DNA probe was prepared according to the manufacturers' protocol. The reaction was incubated at RT for 1 hour.

3.2.2.2 Labelling overgo overlapping oligonucleotide probes

Overgo primers were labelled using the method of Gustafson *et al*. (2003) with minor modifications. Briefly, a 10μl reaction was set up for each probe comprising 1μl each of 10μM forward and reverse primer, 1μl each of 250μM dTTP and dGTP, 1μl of 3000Ci/mmol each of ^{32}P dATP and ^{32}P dCTP, 1μl (2U) Klenow fragment DNA polymerase, 1μl 10×buffer and 2μl hpH_2O. After incubation at 37°C for 30 minutes, 1μl each of 250μM dATP and dCTP was added and incubated for another 15 minutes. Following incubation, reactions for all probes were pooled and purified through a Sephadex G-15 gravity column to eliminate unincorporated nucleotides.

3.2.3 BAC and cosmid library screening

Six high-density replica filters each containing 18,000 distinct BAC clones from the ovine CHORI-243 BAC library (PACBAC Resources) were probed with ^{32}P labelled Overgo primers for the presence of loci known to be within the central region of the human MHC. The Overgo primers were radioactively labelled using overgo technology (Gustafson *et al*., 2003) and hybridised in a buffered solution containing 20 × SSPE, 1%BSA, 7%SDS and 0.5M EDTA, and incubated at 53°C overnight, and washed in 1xSSC buffer containing 0.1% SDS. The hybridized filters were sealed in plastic and exposed to Kodak x-ray film over intensifying screens at -80°C for 2 days prior to development of the film.

Overlapping cosmid clones in the pWE16 vector containing sequences from the complement C4 gene(s) were selected from a sheep genomic cosmid library prepared from the liver of a single male (library constructed by Clontech, USA). These were identified by cross hybridisation with a 2.1 kbp bovine C4 cDNA probe designated pCUT78 (GenBank: U16749). This probe includes the a/g encoding region of a bovine C4 gene and cross-hybridises at high stringency with sheep C4 genes. DNA isolated from eleven cosmid clones containing sheep C4 DNA sequences extending over about 50kbp were digested with EcoRI, electrophoresed in 0.7% agarose gel and blotted to a Hybond N+ membrane

under alkaline conditions. These membranes were hybridised sequentially (5xSSPE, 7% SDS, 1% BSA, 5mM EDTA and washed at elevated stringency (0.1xSSC, 0.1% SDS, 65^{O}C)) with a family of probes specific for the loci of interest. DNA probes used in this study comprised pCUT78 probe (GenBank: U16749), a C4 alpha chain specific probe generated by PCR of sheep DNA described by Ren et al. (1993), human probes for CFB and C2 (gift of Dr Mike Carroll) and a CYP21A cDNA probe of human origin (ATCC: 57420).

3.2.4 Subcloning of the BAC clones

BAC DNA was extracted from colonies (QIAGEN - Large-construct Kit), digested with EcoR1 and BamH1 respectively and subcloned into the pGEM-4Z vector (Promega Life Science) following the manufacturer's instructions (see chapter 2.2). Positive plasmid colonies were identified after being PCR amplified using M13 primers and size compared on agarose gel.

3.2.5 Restriction analysis of cosmid clones containing complement and CYP21 genes

Cosmid clones (≈40kbp insert) gave many fragments when digested with EcoR1. In order to determine which of these fragments was on one end of the insert DNA fragment a special property of the vector was ultilised. The vector (pWE16) has both T3 and T7 RNA polymerase promoter sequences flanking the polylinker region into which the sheep DNA was ligated. DNA from the cosmid clones was digested with EcoRI and at 10min intervals a 10ml aliquot of the digest was removed and further digestion terminated by the addition of 1ml of 0.5M EDTA. Three aliquots (at 10min, 20min and 30min respectively) together with completely digested DNA were electrophoresed in agarose gel (0.7%), Southern blotted onto Hybond N+ and hybridised with ^{32}P kinase labelled T3 RNA promoter oligonucleotide. Using this strategy, one end of all inserts was identified thereby enabling the relative positions of all fragments to be deduced.

3.2.6 DNA Sequencing

Sequences were generated from plasmid clones in both directions using ABI Big Dye chemistry, a 3730 DNA Analyser and M13 primers (M13F: CCCAGTCACGACGTTGTAAA and M13R: AGCGGATAACAATTTCACAC). Similarly, pWE16 cosmid clones were sequenced using T3 and T7 primers as described in Chapter 2. The full length of the gene was then sequenced sequentially in both directions by walking. The primers used for sequence walking of plasmid subclones derived from individual BAC clones are summarized in Table 3.5. Sequence contigs were generated using Vector NTI contig express (Invitrogen). The identification of genomic DNA sequences was performed using BLAST software and the

GenBank databases, while the intron/exon organization of genes was derived using servers for Twinscan (http://mblab.wustl.edu/query.html) and GAP (http://genome.cs.mtu.edu/aat.html) and also DNA Strider software (Marck, 1988).

Table 3.5: Primers for sequence walking of plasmid subclones derived from individual BAC clones

NAME	Primer sequence (5'-3')
265D24BAMH1.1F1	TGGGAAACCTGTCTTTTACA
265D24ECOR1.7F1	AGAGGGTCATGGTTTGTTC
265D24ECOR1.19F1	AAAAGAGCTCTGCCCACT
340A7BAMHI1.15F1	AGATCTGGGGAAAAAGAGAG
340A7BAMHI1.6F1	CAGCAAAGAGACAGACAAGA
340A7ECOR1.3RF1	GACTCTGGCACCTTTATCTG
340A7ECOR1.9F1	ATGCCTAGTACCATCACCAC
340A7ECOR1.15F1	CAGCGTGGATTTCTTCAG
340A7.6F1	TGGTTTCTTTCGGAGAGTAG
342G24.8F1	GAGATTCAAACAGCATAACG
342G24.13F1	ATGGATCACTCCGCCTAT
487C7BAMH1.2F1	CATGTTGTTTGTGTGGTTGT
487C7BAMH1.3F1	TTATCTGGGTCCTGTGACAT
JQBAT5W1	TAGTACCATCACCACTGTGC

3.2.7 SNP discovery and genotyping

Oligonucleotide primers (Table 3.6) were designed from the DNA sequences of sheep MHC central region so that approximately 500bp fragments were amplified. PCR products generated from at least five randomly selected unrelated Australian Merino sheep were sequenced. SNPs were confirmed when at least two heterozygotes and one of each homozygote were observed. SNPs were genotyped by PCR and pyrosequencing using standard protocols with Pyro Gold reagents and the PSQ^{TM}96MA System (Biotage). In a few instances SNP genotypes were confirmed by restriction digestion of PCR amplified DNA fragments followed by 2% agarose gel electrophoresis.

Table 3.6: PCR primers for identifying locus specific SNPs. List of locus specific primers selected from sheep sequences used in SNP discovery. Each primer pair amplified ≈500bp fragments which were then sequenced.

Locus	Forward Primer (5'-3')	Reverse Primers (5'-3')
APOM	GGTGGGTAGATTAGGGAGTC	CACACTGGCTATGTTGACAC
BAT2	AGCGAGGAATACTTCTTTCC	GCTGAGCAAGCTCATAAAAT
BAT5	GTTTGAGCACCTTCTCTCTG	ATGGTGAGGGTTAAATTGG
C2	GACCCAGAAAGTGAAGATGT	ACAGAAGGATGGAGTGAGTG
C2	GTTCCCTACCCCACAGAT	CTGGGCCATTGTAGTTTTAC
G6B	CCTAGAAGAGGGAGAAAAGC	GGATCCTTGAGTCTAGTGGA
LTA	TTTTCCAAACCAAGACACTT	TAGAAGATGCTGCTGTTTCA
MSH5	GGAGCCTCCTAAGATTTGTC	AACACAATCACAAGACAGCA
MSH5	GTCCTGAGCAAATCAGTCAT	ACCTCTGATGATGAAACTGG
MSH5	TTGCAAAGAGTTCACACAAC	CAGGTGACAGCTCTAAGCA
NG22	TTGTGAGGGTGGTGATAAAT	GATGCTCTGGGAGATAGATG
TNXB	ACCATCACCACAACAAAGAT	GGACCTTGAAGGAGTCAAAT
TNF	GAGCGGAGGTTCAGTGAT	GGGTTCTTACCGGAATACTT

3.3 Results

3.3.1 Identification of BAC clones containing sheep central region MHC loci

The sheep CHORI 243 BAC library contains 18,000 × 6 distinct sheep BAC clones. Average insert size of CHORI-243 clone is 182Kbp. The initial screening of the sheep BAC library with radiolabelled, locus specific overgo probes for the 10 putative orthologous loci chosen to cover the entire length of the human MHC central region identified 89 positive colonies. Following a second screening by PCR using PCR primers of sheep origin specific for these 10 loci, 34 of the 89 BAC clones generated amplicons of the predicted sizes. Ten of these BAC clones were selected for further characterisation based upon both the number of loci present per clone and the availability in GenBank of end sequences, which anchored at least one end of the clone within a known gene. Subcloning and partial sequencing of 5 BAC clones, plus the information obtained from the end sequences of the remaining five clones, permitted the assembly of the contig map shown in Figure 3.1. This map covers ≈700kbp anchored by a clone (487C7) manifesting a 5' end sequence homologous with a human class I sequence and

several classical central region loci thereby identifying one end of the sheep central region. Another clone (257B22) contained a 3' sequence homologous with the Notch4 gene, which is close to the start of the class II region in the human MHC.

All locus specific sheep sequences obtained in this study were submitted to GenBank and a summary of the accession numbers allocated to these sequences is shown in Table 3.7.

Table 3.7: Identified sequences of genes within the central region

Locus	GenBank Accession Number
21OH	EF382834
Apom	EF139192
ATP6V1G2	EF139193
BAT2	EF197828
BAT3	EF197829
BAT5	EF197830
C2	EF446374
C4	EF197831
CFB	EF446375
G4	EF197832
G6B	EF197833
G6d	EF197834
G6f	EF197835
HSPA1A	EF197836
G7A	EF197837
LTA	EF446376
LTB	EF197838
MSH5	EF197839
NG22	EF197840
NG35	EF197841
NOTCH4	EF197842
SKI2W	EF197843
snRNP	EF197844
TNF-alpha	EF446377
TNXB	EF197845

3.3.2 *Cosmid mapping of the MHC associated complement genes in sheep*

Initial screening of approximately 100,000 colonies of the cosmid library with the complement C4 pCUT78 probe (specific for part of the a chain and all the γ chain regions of cattle C4 cDNA) at high stringency resulted in the identification of three positive clones (3.3.2.1, 4.2.1 and 10.3.2.4) each of which manifested distinct fragment patterns following digestion with EcoRI. A terminal 2.3 kbp EcoR1 fragment from cosmid 10.3.2.4 was chosen as a probe to rescreen the library for overlapping clones, which permitted extension of the chromosomal region covered and included three additional cosmids known to contain the sheep CFB and C2 loci. The contig map derived from these patterns is shown in Figure 2 included in the published paper (Appendix 1) describing the work reported in this Chapter.

A Southern blot of independent cosmid clones containing complement C4 sequence was hybridised with several C4 specific probes to ascertain the length of genomic DNA which reacted with these probes. As expected, the pCUT78 probe hybridised to all cosmid clones and identified in the majority of clones a 10.1kbp DNA fragment. However, clone C2.9.2.31 contained a 5.0kbp C4 positive fragment but lacked the C4 positive 10.1kbp fragment thereby demonstrating the presence of at least two distinct C4 loci. Clones 4.2.1 and D5.4.11 contained two distinct pCUT78 reactive fragments indicating that they contained both C4 genes. A probe generated by PCR of sheep DNA and specific for the C4 a chain (Ren et al., 1993) showed that the pCUT78 positive 10.1kbp fragment reacted with the C4 a chain probe. However clones 4.2.1 and D5.4.11 contained an additional 2.4kbp C4 a chain positive fragment and clone C2.9.2.31 contained two positive C4 a chain fragments (2.4kbp and 1.5kbp). These data show that the transcriptional orientation of the C4 gene shown on the right side of Figure 2 (included in the paper comprising Appendix 1) is left to right.

Interrogation of the Southern blot with the human CYP21A probe revealed the presence of two cross reacting sheep loci. One of these was present on a 3.0kbp fragment at the end of cosmid 21.1.2.2 and therefore follows the C4 gene with left to right transcription. The other CYP21 gene is present in an 8.4kbp fragment in cosmids C2.3.4.57 and 3.3.2.1 which do not contain the 3.0kbp fragment identifying its paralogue. These results show that in sheep both the C4 and CYP21 loci are duplicated and that their relative order is ..CYP21..C4..C4..CYP21.. Hybridisation of cosmid Bf4.2.20 with C2 and CFB specific human probes revealed the presence of both loci in this cosmid and hybridisation with a radiolabelled probe specific for the previously identified microsatellite locus BfMs showed that it was located to the right of the CFB gene in this cosmid.

Subsequent direct sequencing of this cosmid revealed a partial sequence of sheep C2 (≈5000bp) followed by a complete sequence of CFB (≈5000bp) and a partial sequence of the RDBP gene (≈376bp). These loci are very close to each other with approximate intra-locus distances of 600bp and 4bp respectively.

Thus the map for this subregion is ..C2.(600bp).CFB.(4bp).RDBP. BAC clone (521I18) was found to contain the C2, CFB, SKIV2L, C4 and CYP21 loci. Direct sequencing showed that the BfMs microsatellite locus was located within an intron in the SKIV2L gene which was adjacent to a CYP21 gene. This information has been summarised in Figure 3.1.

3.3.3 Detection of SNPs

Ten of the loci identified in Figure 3.2 were chosen so that they covered the presumed span of the sheep central region. These loci were amplified and the amplicons sequenced from genomic DNA from at least five individual sheep using primers shown in Table 3.6. As an example, a typical sequence electrophoretogram showing the presence of a heterozygous SNP within the BAT2 locus is shown in Figure 3.3. In total a panel of 33 SNPs was generated from 11 loci spanning the entire central region. Complete sequencing of CFB and TNFα genes has identified nine and five intragenic SNPs respectively. These are summarized in Table 3.8 and 3.9. Genotyping of these SNPs in a sheep family and identification of central region MHC sheep haplotypes is described in Chapter 5.

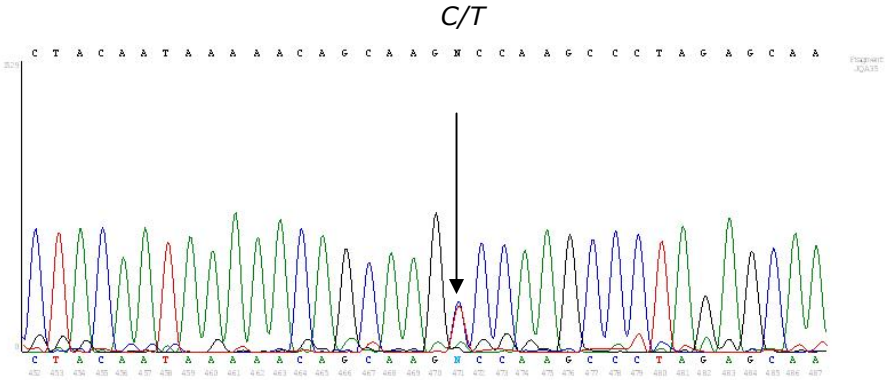


Figure 3.3: Identification of a SNP in the BAT2 gene after sequencing. A double peak, designated as 'N' with half signal strength indicates a heterozygote C/T SNP site.

Table 3.8: SNPs identified in the sheep MHC central region. Location and allelic variation of SNPs present in loci from the sheep MHC central region (excluding CFB and TNFα).

Locus Name	GenBank ID	SNP	Position	Sequence
APOMS1	EF139192	G/A	1423	TTGGCAGGAC(G/A)GCCAGCTCAT
BAT2S1	EF197828	C/T	1441	AAACAGCAAG(C/T)CCAAGCCCTA
BAT2S2	EF197828	A/G	1101	TCCCTGCTCC(A/G)CCAAGGGCCT
BAT5S1	EF197830	A/G	225	GCTTGGGGAC(A/G)GGGGTACCCT
C2S1	EF446374	A/G	1468	CACGGTGAGC(A/G)CTGGACTCAG
C2S2	EF446374	G/C	1809	CTAGCAAGCT(G/C)TGCCCTGGGC
C2S3	EF446374	C/T	1394	GGCCTATCTG(C/T)CTCCCCTGCA
C2S4	EF446374	C/T	320	GAGCGGAGGG(C/T)ATCCTGCTCA
G6bS1	EF197833	C/T	1396	AGGAGTTGAG(C/T)GACCTACATT
LTAS1	DU475543.1	A/G	259	CAGGTGGGAG(A/G)GTATATACTG
LTAS2	DU475543.1	C/T	486	CCTTCTCCTC(C/T)GAATTTACTG
LTAS3	DU475543.1	A/G	254	TGTCCCAGGT(A/G)GGAGGGTATA
JQMSH5S1	EF197839	A/T	3781	TATTTTAGAA(A/T)GAAGAATAGT
JQMSH5S2	EF197839	A/G	3844	ATGCAATGCT(A/G)CAGGTTAAAG
JQMSH5S3	EF197839	C/T	2169	GACAAAATTA(C/T)GGAATGTTCC
JQMSH5S4	EF197839	A/G	2832	TTCTGCTCCA(A/G)TTGGTCCATG
JQNG22S1	EF197840	A/G	926	GTACGGTGCC(A/G)GAGCTCCCAG
TNXBS1	EF197845	C/T	1293	CTCTACCTGT(C/T)TCTCAGAACC
TNXBS2	EF197845	A/G	1324	GAAGAAATCG(A/G)AGCAGAGCCT

Table 3.9: SNPs identified in the sheep CFB and TNFα loci from the sheep MHC central region.

SNP	Position*	Allele1 (freq)	Allele2 (freq)	Type*
ShCFB1	757	A	G	Intronic
ShCFB2	2245	C	T	Syn (S)
ShCFB3	2423	A	G	Intronic
ShCFB4	2922	C (0.08)	T (0.92)	Intronic
ShCFB5	2985	C	T	Intronic
ShCFB6	3149	C	G	NS (F->L)
ShCFB7	4222	C (0.17)	T (0.83)	Intronic
ShCFB8	4413	A	C	NS (K->T)
ShCFB9	5515	C	G	Intronic
TNFαS1	2408	A (0.38)	G (0.62)	3'UTR
TNFαS2	1590 (exon 4)	A	G	Syn (K)
TNFαS3	2149	A (0.22)	G (0.78)	3'UTR
TNFαS4	1951	C	T	3'UTR
TNFαS5	1389 (exon 4)	A (0.14)	G (0.86)	Syn (S)

Note: 'Position' Refers to base number in the genomic sequence. Allele frequencies in panel of 58 unrelated sheep shown in parentheses.*
Syn = synonymous substitution. NS = non synonymous substitution.

3.4 Discussion

Two main approaches were used in this work, which has permitted the construction of a physical map of the central region of the sheep MHC. A broad contig map based on analysis of BAC clones was initially generated. Subsequently characterisation of cosmid clones (identified in an earlier unpublished study) from a sheep genomic library permitted identification and more detailed mapping of the region containing the complement genes (C2, CFB and C4) and sex steroid hydroxylase (CYP21) loci. Finally a panel of SNPs was generated representing 10 loci spanning the mapped central region by direct sequencing of DNA from individual sheep.

The 10 BACs used in this study span ≈700kbp, which may represent most of the extent of the central region although the boundary with the class II region remains to be confirmed by the characterisation of a BAC clone containing both class II region genes and central region loci such as Notch4. The latter is known to be at the class II end of the human central region and this work therefore extends the sheep central region beyond the 600kbp reported by Liu and colleagues (2006). The final map produced contains 17 loci for which the relative positions are known plus 2 groups of loci for which the relative position is known but for which the intragroup order of loci is not known. It is hypothesised that their order will be similar to the corresponding syntenic group in the human MHC central region. The data show that the central regions of the sheep and human MHCs are very similar with respect to length, location and orientation between the class I and class II regions, and their gene composition. This result differs from the map reported by Liu and colleagues (2006 online supplementary data) for the cattle MHC central region in which the class I end of the central region contains a duplicated and inverted subregion of unknown length which places an additional C4 gene adjacent to the class I region. Inspection of the latest version (version 4) of the cattle genome assembly (UCSC Genome Bioinformatics - http://genome.ucsc.edu/) does not support the rearrangement proposed by Liu and colleagues (2006) but does support the broad map reported in this study.

Detailed analysis of twelve cosmid clones containing the complement C4, C2, and CFB loci plus the CYP21 loci revealed a localised inversion of this sub-region such that the order of these genes in sheep is C2…CFB…RDBP…SKIV2L…CYP21...C4 ..C4…CYP21…class II while in the well characterized human central region the order is C2…CFB...RDBP...SKIV2L...C4A...CYP21A...C4B...CYP21B(pseudo)...class

II. The mapping data do not show which pair of C4/CYP21 loci has inverted relative to the human order, although the left to right transcription of the furthermost C4 locus suggests that it is the CYP21 and C4 loci closest to CFB that have transposed. Further sequences will be required to confirm this prediction. Although it is clear that the sheep has two CYP21 loci, it is not known whether either one is a pseudogene. The cosmids described cover approximately 115kbp of sheep DNA and provide a detailed map of this region. Interrogation of version 4 of the cattle genome assembly (UCSC Genome Bioinformatics - http://genome.ucsc.edu/) shows that a similar local rearrangement of C4 and CYP21 loci exists in cattle, although the assembly lists three C4 loci whereas only two such loci were detected in sheep. This could reflect a species difference, or even an individual animal variation, since copy number variation of C4 loci has been observed in several mammalian species. The cosmid mapping showed the presence of the previously described microsatellite locus designated BfMs close to the CFB locus. Sequencing of a DNA fragment subcloned from a BAC clone (443I20) revealed that this polymorphic locus is actually located within the 5' end of the SKIV2L gene which is itself ≈729bp downstream from the 3' end of the CFB gene.

Dalrymple and colleagues (2007) have constructed a virtual map of the sheep genome based largely on end sequencing of the CHORI-243 BAC library used in this study, together with arguments based on structural comparisons and synteny with the better characterized genomes of other mammals. The virtual map published by this group for the MHC region is consistent with the independent mapping data provided in this study. Another extensive study published recently by Wu and colleagues (2008) utilized radiation hybrid mapping to provide a map of sheep chromosome 20, which includes the MHC. The partial map of the central region reported by these workers shows discrepancies with both this study and the comparative information in version 4 of the cattle genome assembly.

Although this study shows structural similarity between the human and sheep MHCs, it is known that the MHCs of mammals exhibit considerable structural diversity and it is likely that further studies will show species-specific differences between the central regions of the sheep MHC and those of other species including the human MHC. The localized inversion of the C4 and CYP21 loci in sheep described in this report is one example of this type of diversity. The map described in this work is the result of analysis of DNA from two male sheep (i.e. originators of the BAC and cosmid libraries). It is known that in humans there is

significant haplotype diversity, especially in the complement containing subregion of the central region, which is believed to reflect the ancestral haplotypes and block like structure of this genomic region (Yu *et al*., 2000). Hence it is likely that individual sheep will manifest a degree of structural diversity in this region, however this will only become apparent when chromosomes defined by panels of polymorphic markers such as the SNPs reported in this work are used to assess haplotype diversity.

The SNPs reported in Tables 3.8 and 3.9 should cover most of the length of the sheep central region and provide a panel of genetic markers for identifying central region haplotypes. As will be described further in Chapter 5, the SNPs identified above permitted the recognition of haplotypic combinations which displayed Mendelian inheritance in a three generation sheep half sib family.

In summary, this report provides a physical map covering ≈700kbp of the central region of the sheep MHC together with a SNP panel which will facilitate disease and productivity association studies and permit studies of block like subregions within this region. This map has been obtained by utilizing clones from the freely available sheep CHORI-243 BAC library and from a cosmid genomic library from Clontech. The sheep central region is remarkably similar to the human central region, although a local inversion of the duplicated C4 and CYP21A/B loci relative to humans is a significant point of difference. However, a more extensive inverted duplication reported by Liu and colleagues (2006) for the cattle MHC central region was not observed in sheep. In addition, a previously described dinucleotide tandem repeat locus (BfMs) has been located within an intron in the SKIV2L gene.

Chapter 4

Sequence Analysis of Sheep and Cattle TNF, CFB, C2, C4 Genes

This chapter describes the partial sequencing of cosmid and BAC subclones to characterize map complement genes and the TNF gene within a subregion of the central region of the sheep MHC. Detailed sequence data for the CFB and TNF genes are presented together with partial sequences for complement C2 and C4 genes. The CFB gene is present in a single cosmid clone and is located approximately 600bp upstream of the complement C2 gene. Sheep CFB contains 18 exons and encodes domains characteristic of CFB protein; multiple SNPs were identified. Comparative analysis of the sheep CFB gene facilitated determining the nucleotide sequence of the cattle CFB orthologue together with its predicted amino acid sequence. Both sheep and cattle CFB genes contain a 9 nucleotide deletion in exon 18 which results in the CFB proteins of these species containing 761 amino acids relative to 764 amino acids in the human orthologue. A comparative analysis of sheep and cattle C2 genes was also undertaken. Both C2 genes share two non-contiguous deletions of 6 and 3 bp in exon 18 relative to human C2. Unlike humans, cattle C2 does not contain a long retroviral sequence in the third intron. The tumour necrosis factor alpha gene (TNFα) was identified in a cosmid clone from a male sheep genomic library and sequenced completely. The gene is similar to that of cattle TNF spanning 1770bp with four exons. Both sheep and cattle TNF encode predicted proteins with 234 amino acids, which include an extra glutamine at the beginning of exon 2 relative to their human orthologue. Five SNPs were identified within a 5210bp genomic sequence encompassing sheep TNFα; two of these were synonymous and located in exon 4 while the remaining three were within the 3' untranslated region.

4.1 Introduction

The complement C2 and CFB genes encode paralogous proteins that are essential contributors to the classical and alternative pathways of complement activation respectively (Porter, 1985). Both proteins are single polypeptides exhibiting serine proteinase activity and are synthesised in many different organs and cell types. An important characteristic of these proteins is the presence in the N-terminal region of multiple so called 'Sushi' elements or short consensus repeats (SCRs) which are evolutionarily conserved and believed to confer important biological functions to

these proteins (McClure *et al.*, 2004). Indeed, Sushi domains have been discovered in other regulatory proteins of complement activation and it has been proposed that their biological functions require multiple copies of the elements (McClure *et al.*, 2004). Orthologues of the CFB gene have been described in several protochordate species which confirms their long evolutionary history and supports claims for their contribution to innate mechanisms of immunity (Whitehead and Sackstein, 1985) which may also include immunity to parasites. Contrary to earlier speculation CFB deficient mice are viable and reproduce and develop normally (Matsumoto *et al.*, 1997; Pekna *et al*., 1998). More recently, CFB deficiency has been associated with inflammation during allergen challenge in mice (Taube *et al.*, 2006) and variation in CFB and C2 genes has been shown to be associated with age related macular degeneration in humans (Gold *et al.*, 2006).

Tumour necrosis factor (TNF) and related proteins are cytokines with potent immunoregulatory activities and their dysfunction has been implicated in a variety of infectious and inflammatory diseases including septic shock (Beutler and Cerami, 1989; Ruuls and Sedgwick, 1999). Two genes of the TNF family (TNFα and TNFβ) have been mapped to the central region of the human major histocompatibility complex (MHC) (Ziegler *et al*., 1991) and orthologues of both have been identified in the central region of the sheep MHC (see Chapter 3).

Sheep TNFα was originally characterized from a cDNA library constructed from lipopolysaccharide stimulated alveolar macrophages and shown to comprise a single polypeptide containing 234 amino acids (Nash *et al.*, 1991). More recently, Alvarez-Busto and colleagues (2004) identified three sequence variants within a 273bp fragment encompassing part of exon 4 and the 3' untranslated region. The nucleotide sequences defining these alleles were compared with their corresponding cattle and goat orthologues.

In this Chapter, genomic sequences of sheep CFB, C2, C4 and TNF genes are reported and compared with their cattle and human orthologues. A panel of single nucleotide polymorphisms (SNPs) in sheep CFB and TNFα have been identified and presented in Table 3.9 (see Chapter 3).

4.2 Materials and Methods

4.2.1 Cosmid and BAC clones

A total of 35,000 colonies from a male sheep cosmid library (Clontec) were plated out and transferred onto nylon membranes (Hybond N+, Amersham). A cosmid clone containing sheep TNF sequence (TNF1.29) was identified using a ^{32}P labelled PCR product amplified from sheep genomic DNA using primers described in Table 3.1 (see chapter 3). A cosmid clone (BfC4.2.20) containing a complete sheep CFB sequence (and also a part of a complement C4 sequence) was previously identified by Dr Groth using a ^{32}P labelled human CFB cDNA probe provided by Dr M. Carroll (Oxford) as described in Chapter 3. The insert from both of these clones was digested and subcloned into pGEM-4Z vector (Promega Life Science).

Six filters each containing 18,000 distinct BAC clones from the sheep CHORI 243 BAC library (Children's Hospital Oakland Research Institute) were screened with a ^{32}P labeled probe for the sheep C2 gene generated using 'overgo' primers (Gustafson *et al.,* 2003) derived from a partial sequence of the C2 gene present in the BfC4.2.20 clone. A positive BAC clone (CHORI 243-443I20) was digested and subcloned into the pGEM-4Z vector.

4.2.2 Sequencing and sequence analysis

Sequences were generated using ABI Big Dye chemistry and a 3730 DNA Analyser. Primers were designed from first pass sequences and the remainder of the CFB gene was sequenced progressively in both directions. Final sequence contigs were generated using Vector NTI contig express (Invitrogen). The identification of putative genomic DNA was performed using BLAST software, while the intron/exon organization of genes was derived using servers for Twinscan (http://mblab.wustl.edu/query.html) and GAP (http://genome.cs.mtu.edu/aat.html) and also DNA Strider software (Marck, 1988). Multiple sequence alignments were generated using ClustalX 1.83 software (Thompson *et al.,* 1997).

4.2.3 SNP discovery and genotyping

Oligonucleotide primers (Table 4.1) were identified from the sheep CFB and TNFα sequence which, amplified approximately 500bp fragments covering the entire genomic sequence. PCR products generated from at least five randomly selected Australian Merino sheep were sequenced. SNP's were confirmed when at least two

heterozygotes and one of each homozygote were observed. The methods used for genotyping SNPs are described more fully in Chapter 5, where possible SNPs were genotyped by pyrosequencing of PCR amplicons using standard protocols with Pyro Gold reagents and the PSQ^{TM}96MA System (Biotage). In a few instances SNPs could not be routinely typed using pyrosequencing (e.g. TNFaS2 and S4, see Table 3.9 in Chapter 3) and these polymorphisms were not pursued further.

4.3 Results

4.3.1 Sequence of sheep and cattle CFB genes

The sheep CFB gene is 5857bp long and is similar to its human orthologue with 18 exons separating 17 introns. The deduced coding sequence was 2283bp encoding a peptide of 761 amino acids. At the time this work was completed, an annotated genomic and cDNA sequence of the cattle CFB locus was available from GenBank (NW 930219 and XM 591873) together with a predicted CFB peptide of 741 amino acids (XP 591873.2). Comparison of these sequences with human (NP_001701.2) and sheep CFB showed the presence of an additional cytosine base at residue position 790 in the 2284bp putative CFB cattle sequence. Re-sequencing of a 224bp PCR amplified segment of the cattle CFB gene spanning this region identified an error in the reported sequence. When the corrected cattle sequence was aligned to the sheep and human CFB sequences the similarity of the cattle CFB gene (and its deduced CDS) became apparent.

A multiple alignment of the genomic and CDS sequences for the human, sheep and cattle CFB genes was generated using ClustalX software (see Figure 4.1). The revised cattle and sheep CFB genes share a nine nucleotide deletion in exon 18 which results in a predicted protein for both species of 761 amino acids relative to the 764 amino acids present in human CFB. An alignment of the predicted CFB peptides from sheep and cattle together with the known amino acid sequences of mouse (NP032224.1), chimpanzee (NP001009169.1) and human CFB is shown in Figure 4.2. Both human and chimp CFB proteins have 764 amino acids whereas the other three species have 761 amino acids each. Mouse CFB has a three amino acid deletion near the amino terminal end of the peptide whereas both sheep and cattle genes share the nine residue deletion within exon 18 corresponding to a three amino acid deletion in the CFB proteins of these two species.

4.3.2 Polymorphism of sheep CFB

Comparisons of the CFB genomic sequence between individual sheep revealed nine SNP's, three of which occurred in exons (summerized in Table 3.9, chapter 3). Only SNPs which displayed a minimum allele frequency of ≥5% are reported. Two of these SNPs (ShCFB4/7) were genotyped in 58 sheep by RFLP and pyrosequencing.

Table 4.1: Primers for generating and sequencing overlapping contigs of CFB and TNFα

Locus	**Forward (5'-3')**	**Reverse (5'-3')**
BF1	CTTTAGGTGTGGGCATGA	GGACCCTGTCACTGAGAAAT
BF2	GAAGTTATATTTCTCAGTGACAGG	TCTGACCTGAGGGTACAGAT
BF3	CTTATCTGTACCCTCAGGTCA	CCACCTTCTCAATGAAATCT
BF4	TCAGAGATTTCATTGAGAAGG	CCTCTCAGTTCCAAAGGATT
BF5	GGAACTGAGAGGGGTACTTT	CAAACCTGCAGGAGAGAC
BF6	CGTCTCTCCTGCAGGTTT	GCCCCAGCTTCCATATCT
BF7	CCTTTTACAGATATGGAAGCTG	ATCCTTACCCACGCTGAC
BF8	GTCAGCGTGGGTAAGGAT	ACCCTTACAAGCCGCTCT
BF9	GCTTGTAAGGGTGAAGTTCT	GAACCTGGGGGTGACTAC
BF10	GTAGTCACCCCCAGGTTC	TGTCGCAGATGTTTAATTTG
BF11	CAAATTAAACATCTGCGACA	AAAGGGCTATGGAGATTTTC
TNFα1	AGAGTTGGAAATTCCCTTGT	GCTGATTTGGTGACTGATTT
TNFα2	GAGCGGAGGTTCAGTGAT	GGGTTCTTACCGGAATACTT
TNFα3	CAGGAAGGGAGAGTTGAAG	CAGTCTCCTACCAGACCAAG
TNFα4	TCTAGGGTCTCCCTGTGG	GTAGCCCACGTTGTAGGTAA
TNFα5	TGAATCCTCACACTGTCTCA	AAGATGAAGCTAAGCGTGTC
TNFα6	CTTCTCTCTCCATTCATCCA	CTTCTCCTTCCTCCTGGTT
TNFα7	GAGACTCACCTCTTCCCTCT	TCAGGGTCCCATATTTACAG
TNFα8	GGCATCAAGGATACTTCCTA	GATACATGCCCACACACAG
TNFα9	CTGGTTTCAGTCTTGGTTTC	TTCTGGAAGGAAGTGAAAAG

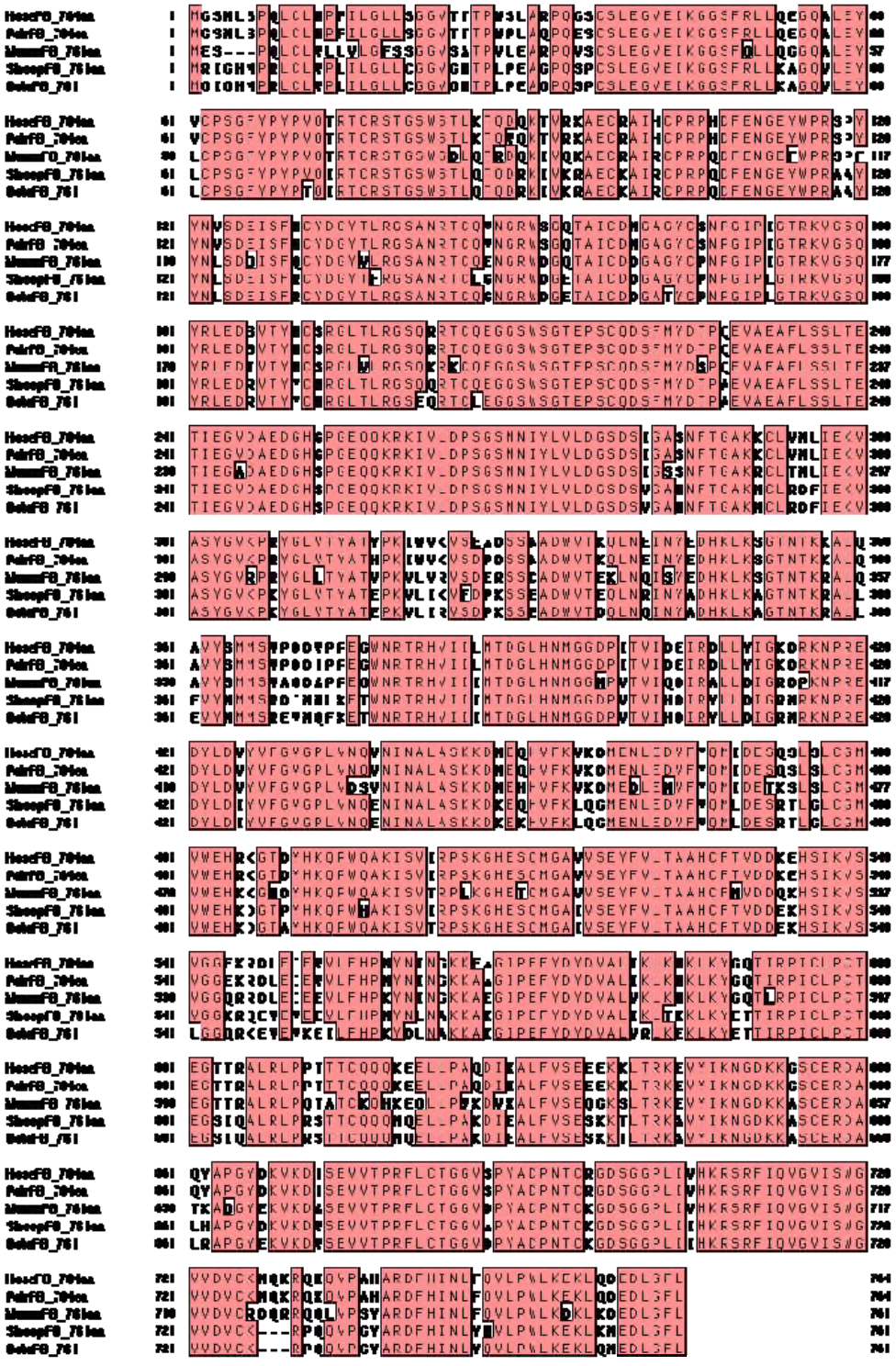


Figure 4.2: Alignment of CFB proteins from sheep and cattle

4.3.3 Sequence analysis of sheep and cattle C2 genes

A partial genomic sequence (2504bp) from the 3′ end of the sheep C2 gene was obtained and aligned with cattle and human orthologues. This sequence contained exons 10 to 18 of the sheep C2 gene and generated an open reading frame encoding the carboxy terminal 342 amino acids of the sheep C2 protein. A sequence containing the complete genomic cattle C2 was obtained from GenBank and its CDS extracted. A multiple sequence alignment of the genomic and CDS sequence of cattle C2 together with the partial sequence of sheep C2 was then obtained. Excellent conservation of exonic/intronic boundaries was observed and both sheep and cattle C2 showed non-contiguous deletions of 6 and 3 bps in exon 18 relative to that of human C2. Together with three other indels in the 5′ end of the cattle C2 gene, the predicted C2 protein of cattle is two amino acids shorter than its human orthologue (750aa versus 752aa).

An alignment of the predicted C2 peptides from sheep and cattle, together with the known amino acid sequence of human C2 is shown in Figure 4.3.

```
HumC2_ENSP0000036466 MGPLMVLFCLLFLYPGLADSAPSCPQNVNISGGTFTLSHGWAPGSLLTYS   50
BotaC2_750aa_AAI0335 MDPLMAVLCLLPLYPGLATAALSCPKNVNISGGSFTLSNGWNPGSILTYS   50
BotaC2_BC103357.1_75 MDPLMAVLCLLPLYPGLATAALSCPKNVNISGGSFTLSNGWNPGSILTYS   50
BotaC2_BTO26538.1_75 MDPLMAVLCLLPLYPGLATAALSCPKNVNISGGSFTLSNGWNPGSILTYS   50
SheepC2_342aa_jinyi  --------------------------------------------------    0

HumC2_ENSP0000036466 CPQGLYPSP-ASRLCKSSGQWQTPGATRSLSKAVCKPVRCPAPVSFENGI   99
BotaC2_750aa_AAI0335 CPLGHYPYPVVTRLCKSNGQWQIPRSTRST-KAICKPVRCPAPVSFENGV   99
BotaC2_BC103357.1_75 CPLGHYPYPVVTRLCKSNGQWQIPRSTRST-KAICKPVRCPAPVSFENGV   99
BotaC2_BTO26538.1_75 CPLGHYPYPVVTRLCKSNGQWQIPRSTRST-KAICKPVRCPAPVSFENGV   99
SheepC2_342aa_jinyi  --------------------------------------------------    0

HumC2_ENSP0000036466 YTPRLGSYPVGGNVSFECEDGFILRGSPVRQCRPNGMWDGETAVCDNGAG  149
BotaC2_750aa_AAI0335 YIPRLGSHPVGGNLSFECEDGFTLRGSAVRQCRPNGMWDGETAVCDNGAS  149
BotaC2_BC103357.1_75 YIPRLGSHPVGGNLSFECEDGFTLRGSAVRQCRPNGMWDGETAVCDNGAS  149
BotaC2_BTO26538.1_75 YIPRLGSHPVGGNLSFECEDGFTLRGSAVRQCRPNGMWDGETAVCDNGAS  149
SheepC2_342aa_jinyi  --------------------------------------------------    0

HumC2_ENSP0000036466 HCPNPGISLGAVRTGFRFGHGDKVRYRCSSNLVLTGSSERECQGNGVWSG  199
BotaC2_750aa_AAI0335 HCPNPGISVGAVRTGSRFGLGDKVRYRCSSNLVLTGSAERECQDDGVWSG  199
BotaC2_BC103357.1_75 HCPNPGISVGAVRTGSRFGLGDKVRYRCSSNLVLTGSAERECQDDGVWSG  199
BotaC2_BTO26538.1_75 HCPNPGISVGAVRTGSRFGLGDKVRYRCSSNLVLTGSAERECQDDGVWSG  199
SheepC2_342aa_jinyi  --------------------------------------------------    0

HumC2_ENSP0000036466 TEPICRQPYSYDFPEDVAPALGTSFSHMLGATNPT-QKTKESLGRKIQIQ  248
BotaC2_750aa_AAI0335 TEAICRQPYSYDFPEDVAPALGTSFSHLLATTNPIQQKKKQNLGRKIQIQ  249
BotaC2_BC103357.1_75 TEAICRQPYSYDFPEDVAPALGTSFSHLLATTNPIQQKKKQNLGRKIQIQ  249
BotaC2_BTO26538.1_75 TEAICRQPYSYDFPEDVAPALGTSFSHLLATTNPIQQKKKQNLGRKIQIQ  249
SheepC2_342aa_jinyi  --------------------------------------------------    0

HumC2_ENSP0000036466 RSGHLNLYLLLDCSQSVSENDFLIFKESASLMVDRIFSFEINVSVAIITF  298
BotaC2_750aa_AAI0335 RSGHLNLYLLLDASQSVSKDDFEIFKDSASRMVDRIFSFEIKVSVAIITF  299
BotaC2_BC103357.1_75 RSGHLNLYLLLDASQSVSKDDFEIFKDSASRMVDRIFSFEIKVSVAIITF  299
BotaC2_BTO26538.1_75 RSGHLNLYLLLDASQSVSKDDFEIFKDSASRMVDRIFSFEIKVSVAIITF  299
SheepC2_342aa_jinyi  --------------------------------------------------    0

HumC2_ENSP0000036466 ASEPKVLMSVLNDNSRDMTEVISSLENANYKDHENGTGTNTYAALNSVYL  348
BotaC2_750aa_AAI0335 ASKPKIIMSVLEDRSRDVTEVENSLRNINYKDHENGTGTNIYEALHAVYI  349
BotaC2_BC103357.1_75 ASKPKIIMSVLEDRSRDVTEVENSLRNINYKDHENGTGTNIYEALHAVYI  349
BotaC2_BTO26538.1_75 ASKPKIIMSVLEDRSRDVTEVENSLRNINYKDHENGTGTNIYEALHAVYI  349
```

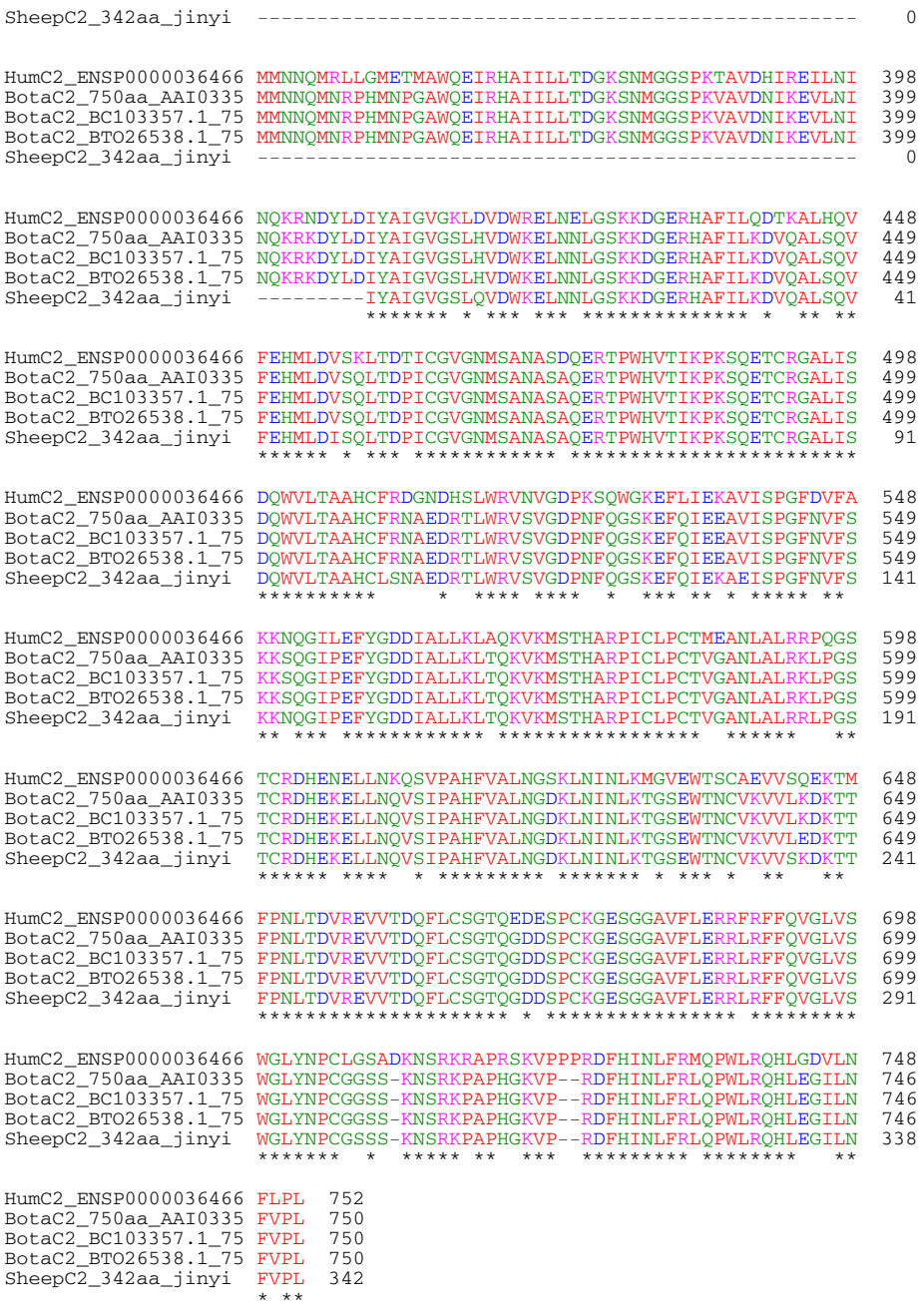


Figure 4.3: Alignment of C2 proteins from sheep, cattle and human

4.3.4 Sequence analysis of sheep and cattle TNF

The complete 1770bp sequence of the sheep TNF is described within a 5210bp genomic sequence which also contains 5′ and 3′ untranslated regions together with part of the related LTA gene (GenBank EF446376, EF446377). As was described in Chapter 3, sheep TNF is located within the central region of the MHC.

A multiple sequence alignment of genomic TNF and its coding region (CDS) together with the previously described corresponding sequences for the human and cattle orthologues was generated and is shown in Figure 4.4. The coding region of sheep TNF is 702 nucleotides long and predicts a peptide of 234 amino acids as shown in Figure 4.5. This peptide has the same length as that reported for the cattle TNF (NM_173966) peptide; both peptides have an additional glutamine at position 63 relative to that of human TNF protein.

```
HUMTNF_GEN  ATGAGCACTGAAAGCATGATCCGGGACGTGGAGCTGGCCGAGGAGGCGCTCCCAAGAAG   60
HUMTNF_CDS  ATGAGCACTGAAAGCATGATCCGGGACGTGGAGCTGGCCGAGGAGGCGCTCCCAAGAAG   60
SHTNF_GEN   ATGAGCACCAAAAGCATGATCCGGGATGTGGAGCTGGCGGAGGAGGTGCTCTCCAACAAA   60
SHTNF_CDS   ATGAGCACCAAAAGCATGATCCGGGATGTGGAGCTGGCGGAGGAGGTGCTCTCCAACAAA   60
BOVTNF_GEN  ATGAGCACCAAAAGCATGATCCGGGATGTGGAGCTGGCGGAGGAGGTGCTCTCCGAGAAA   60
BOVTNF_CDS  ATGAGCACCAAAAGCATGATCCGGGATGTGGAGCTGGCGGAGGAGGTGCTCTCCGAGAAA   60
            ********  ********************* *********** ******* **** ** * **

HUMTNF_GEN  ACAGGGGGGCCCCAGGGCTCCAGGCGGTGCTTGTTCCTCAGCCTCTTCTCCTTCCTGATC  120
HUMTNF_CDS  ACAGGGGGGCCCCAGGGCTCCAGGCGGTGCTTGTTCCTCAGCCTCTTCTCCTTCCTGATC  120
SHTNF_GEN   GCAGGGGGCCCCCAGGGCTCCAGAAGTTGCTGGTGCCTCAGCCTCTTCTCCTTCCTCCTG  120
SHTNF_CDS   GCAGGGGGCCCCCAGGGCTCCAGAAGTTGCTGGTGCCTCAGCCTCTTCTCCTTCCTCCTG  120
BOVTNF_GEN  GCAGGGGGCCCCCAGGGCTCCAGAAGTTGCTTGTGCCTCAGCCTCTTCTCCTTCCTCCTG  120
BOVTNF_CDS  GCAGGGGGCCCCCAGGGCTCCAGAAGTTGCTTGTGCCTCAGCCTCTTCTCCTTCCTCCTG  120
            ******* **************  * **** ** **********************  *

HUMTNF_GEN  GTGGCAGGCGCCACCACGCTCTTCTGCCTGCTGCACTTTGGAGTGATCGGCCCCCAGAGG  180
HUMTNF_CDS  GTGGCAGGCGCCACCACGCTCTTCTGCCTGCTGCACTTTGGAGTGATCGGCCCCCAGAGG  180
SHTNF_GEN   GTTGCAGGAGCCACCACGCTCTTCTGCCTGCTGCACTTCGGGGTAATCGGCCCCCAGAGG  180
SHTNF_CDS   GTTGCAGGAGCCACCACGCTCTTCTGCCTGCTGCACTTCGGGGTAATCGGCCCCCAGAGG  180
BOVTNF_GEN  GTTGCAGGAGCCACCACGCTCTGCTGCCTGCTGCACTTCGGGGTAATCGGCCCCCAGAGG  180
BOVTNF_CDS  GTTGCAGGAGCCACCACGCTCTTCTGCCTGCTGCACTTCGGGGTAATCGGCCCCCAGAGG  180
            ** ***** ************* *************** ** ** ***************

HUMTNF_GEN  GAAGAGGTGAGTGCCTGGCCAGCCTTCATCCACTCTCCCACCCAAGGGGAAATGAGAGAC  240
HUMTNF_CDS  GAAGAG------------------------------------------------------  186
SHTNF_GEN   GAAGAGGTGAGTCTCTGGCCAGCCTTGGCTCATTCTCCCACCCAGAGAGAAATGAGAGA-  239
SHTNF_CDS   GAAGAG------------------------------------------------------  186
BOVTNF_GEN  GAAGAGGTGAGTTTCTGGCCGGCCTTGGCTCATTCTCCCACCCAGAGAGAAATGAGAGA-  239
BOVTNF_CDS  GAAGAG------------------------------------------------------  186
            ******

HUMTNF_GEN  GCAAGAGAGGGAGAGAGATGGGATGGGTGAAAGATGTGCGCTGATAGGGAGGGATGAGAG  300
HUMTNF_CDS  ------------------------------------------------------------  186
SHTNF_GEN   --GAAAGAGGGTGAGAGATGGGGTAGG-GAAAGAGGTGTGCTGATGGGGAAGAATGGGGA  296
SHTNF_CDS   ------------------------------------------------------------  186
BOVTNF_GEN  --GAAAGAGGGTGAGAGATGGGGTGGA-GAAAGAGGTGTGTTGATGGGGAAAGATGGGGA  296
BOVTNF_CDS  ------------------------------------------------------------  186

HUMTNF_GEN  AGAAAAAAACATGGAGAAAGACGGGGATGCAGAAAGAGATGTGGCAAGAGATGGGGAAGA  360
HUMTNF_CDS  ------------------------------------------------------------  186
SHTNF_GEN   GGAAAGTAG--TGGAGAAAGATGGGGAGGCAGGAGGAGACATGG--AGAGAGTGGAGGGA  352
SHTNF_CDS   ------------------------------------------------------------  186
BOVTNF_GEN  GGAAAATAG--TGGAGAAAGATGGGGAGGCAGGAGGAGACATGG----AGAGAGATGGGG  350
BOVTNF_CDS  ------------------------------------------------------------  186

HUMTNF_GEN  GAGAGAGAGAAAGATGGAGAGACAGGATGTCTGGCACATGGAAGGTGCTCACTAAGTGTG  420
HUMTNF_CDS  ------------------------------------------------------------  186
SHTNF_GEN   GAAAGGGAGAAGGACAGAGAAATAAGACCCTTGGCACGTGGGGGATGCCCTCTAAATATT  412
SHTNF_CDS   ------------------------------------------------------------  186
BOVTNF_GEN  GGAAGAGGGAGAAGGGTGGAGAAATAACACTTGGCACATGGGGGATGCCCTCTAAATATT  410
BOVTNF_CDS  ------------------------------------------------------------  186
```

```
HUMTNF_GEN  TATGGAGTGAATGAATGAATGAATGAATGAACAAGCAGATATATAAATAAGATATGGAGA  480
HUMTNF_CDS  ------------------------------------------------------------  186
SHTNF_GEN   TGTTGGACGAATGAGTAAGCAGACGTACAGACATAAAGA---------GAGATGTGTAGA  463
SHTNF_CDS   ------------------------------------------------------------  186
BOVTNF_GEN  TGTTGGACGAATGAGTAAGCAGAGATACAGACATAAAGA---------GAGATGTGTAGA  461
BOVTNF_CDS  ------------------------------------------------------------  186

HUMTNF_GEN  CAGATGTGGGGTGTGAGAAGAGAGATGGGGGAA-GAAACAAGTGATATGAATAAAGATGG  539
HUMTNF_CDS  ------------------------------------------------------------  186
SHTNF_GEN   CAGATAAGGGGTGTGAGTGGAAAGATGGAGATTCAAAACAAGTGATCTGGATGAAGATGA  523
SHTNF_CDS   ------------------------------------------------------------  186
BOVTNF_GEN  CAGATAAGGGGTGTGAGTAGAAAGATGGAGACTCAAAACAAGTGATCTGGATGAAGATGA  521
BOVTNF_CDS  ------------------------------------------------------------  186

HUMTNF_GEN  TGAGACAGAAAGAGCGGGAAATATGACAGCTAAGGAGAGAGATGGGGGAGATAAGGAGAG  599
HUMTNF_CDS  ------------------------------------------------------------  186
SHTNF_GEN   TGAGACAGAAAGAGATACAGACAATATGAAGAGAGATAAGGAAAGATGAAGCTA------  577
SHTNF_CDS   ------------------------------------------------------------  186
BOVTNF_GEN  TGAGACAGAAAGAGATACAGACGATAGGAAAGAGAGATAAGGAAAGATGAAGCTA-----  576
BOVTNF_CDS  ------------------------------------------------------------  186

HUMTNF_GEN  AAGAAGATAGGGTGTCTGGCACACAGAAGACACTCAGGGAAAGAGCTGTTGAATGCTGGA  659
HUMTNF_CDS  ------------------------------------------------------------  186
SHTNF_GEN   --------AGCGTGTCTGGCACGTGGAAGGAACTCAATGAAAGAGCTGTTAAATGGATGA  629
SHTNF_CDS   ------------------------------------------------------------  186
BOVTNF_GEN  --------AGTGTGTCTGGCATGTGGAAGGAACTCAATGAAAGAGCTGTTAAATGGATGA  628
BOVTNF_CDS  ------------------------------------------------------------  186

HUMTNF_GEN  AGGTG---AATACACAGATGAATGGAGAGAGAAAACCAGACACCTCAGGGCTAAGAGCGC  716
HUMTNF_CDS  ------------------------------------------------------------  186
SHTNF_GEN   ATGAATGAAAAAAAATGGATGAATGGAGAGAGAAGACAAGACCCATCTGGGCAGAGAGTAC  689
SHTNF_CDS   ------------------------------------------------------------  186
BOVTNF_GEN  ATGAACAAAAAAAAATGGATGAATGGAGAGAGAAGACAAGACCCATCAGGGCAGAGAGTAC  688
BOVTNF_CDS  ------------------------------------------------------------  186

HUMTNF_GEN  AGGCCAGACAGGCAGCCAGCTGTTCCTCCTTTAAGGGTGACTC-CCTCGATGTTAACCAT  775
HUMTNF_CDS  ------------------------------------------------------------  186
SHTNF_GEN   AGGCTAGAGAAGCAGCCAGCAGTTTCTCGTTCAGGGGTGACTTGCTTTAACACTCATCCT  749
SHTNF_CDS   ------------------------------------------------------------  186
BOVTNF_GEN  AAGCTAGAGAAGCAGCCAGGAGTTTCTCCTTCAGGGGTGACTTGCTCTAACACTCATCCT  748
BOVTNF_CDS  ------------------------------------------------------------  186

HUMTNF_GEN  TCTCCTTCTCCCCAACAGTTCCCCAGGGACCTCTCTCTAATCAGCCCTCTGGCCCAGGCA  835
HUMTNF_CDS  ----------------TTCCCCAGGGACCTCTCTCTAATCAGCCCTCTGGCCCAGGCA  228
SHTNF_GEN   CCTTTTTCTCCCCAGCAGTCCCCAGCTGGCCCCTCCTTCAACAGGCCTCTGGTTCAGACA  809
SHTNF_CDS   ----------CAGTCCCCAGCTGGCCCCTCCTTCAACAGGCCTCTGGTTCAGACA  231
BOVTNF_GEN  CCTTTTTCTCCCCAGCAGTCCCCAGGTGGCCCCTCCATCAACAGCCCTCTGGTTCAGACA  808
BOVTNF_CDS  ----------CAGTCCCCAGGTGGCCCCTCCATCAACAGCCCTCTGGTTCAAACA  231
                      * ***    * ** ***  * * *** *******  **  **

HUMTNF_GEN  GTCAGTAAGTGTCTCCAAACCTCTTTCCTAATTCTGGGTTTGGGTTTGGGGGTAGGGTTA  895
HUMTNF_CDS  GTCA----------------------------------------------------  232
SHTNF_GEN   CTCAGTAAGTGTCTCCACAACCTCTTTTCTAATTCTGGACTTGGAGGGTGGGTGTGGGTG  869
SHTNF_CDS   CTCA----------------------------------------------------  235
BOVTNF_GEN  CTCAGTAAGTGTCTCCACAACCTCTCTTCTAATTCTGGACTTGGGGGGTGGGTGTGGGTG  868
BOVTNF_CDS  CTCA----------------------------------------------------  235
             ***

HUMTNF_GEN  GTACCGGTATGGAAGCAGTGGGGGAAATTTAAAGTTTTGGTCTTGGGGGAGGATGGATGG  955
HUMTNF_CDS  ------------------------------------------------------------  232
SHTNF_GEN   GGGTTCAGTCCAGGGATGGAAGCAGTAGGGGAAATTAAGGGTTTGGGATCGGAGGGGAGA  929
SHTNF_CDS   ------------------------------------------------------------  235
BOVTNF_GEN  GGGTTTAGTCAAGGGATGGAAGCAGTAGGGGAAATTTAGGGTTTGGGATCGGGGGGAGAA  928
BOVTNF_CDS  ------------------------------------------------------------  235

HUMTNF_GEN  AGGTGAAAGTAGGGGGGTATTTTCTAGGAAGTTTAAGGGTCTCAGCTTTTTCTTTTCTCT 1015
HUMTNF_CDS  ------------------------------------------------------------  232
SHTNF_GEN   AG--GAATGGAGGTTAAAGTATGGGATATTCTCTGGGAAGCTTACCTTTTCCTTTCCTCT  987
SHTNF_CDS   ------------------------------------------------------------  235
BOVTNF_GEN  GG---AATGGAGGTTAAAGTATGGGATATTTTCTGGGAAGCTTACCTTTTCCTTTCCTCT  985
BOVTNF_CDS  ------------------------------------------------------------  235

HUMTNF_GEN  CTCCTCTTCAGGATCATCTTCTCGAACCCCGAGTGACAAGCCTGTAGCCCATGTTGTAGG 1075
HUMTNF_CDS  ----------GATCATCTTCTCGAACCCCGAGTGACAAGCCTGTAGCCCATGTTGTAG-  280
SHTNF_GEN   TTCCTCCTCAGGGTCATCTTCTCAAGCCTCAAATAACAAGCCGGTAGCCCACGTTGTAGG 1047
```

```
SHTNF_CDS    ----------GGTCATCTTCTCAAGCCTCAAATAACAAGCCGGTAGCCCACGTTGTAG-  283
BOVTNF_GEN   TTTCTCCTCAGGGTCCTCTTCTCAAGCCTCAAGTAACAAGCCGGTAGCCCACGTTGTAGG 1045
BOVTNF_CDS   ----------GGTCCTCTTCTCAAGCCTCAAGTAACAAGCCGGTAGCCCACGTTGTAG-  283
                      * ** ******* * ** * * * ******* ******** *******

HUMTNF_GEN   TAAGAGCTCTGAGGATGTGTCTTGGAACTTGGAGGGCTAGGATTTGGGGATTGAAGCCCG 1135
HUMTNF_CDS   ------------------------------------------------------------  280
SHTNF_GEN    TAAGAGTTCTGAGGATGTATCTGGGGCATGAAGAAGCAGAGAGACAGGGCGGGGTA---- 1103
SHTNF_CDS    ------------------------------------------------------------  283
BOVTNF_GEN   TAAGAGTTCTGAGGATGTATCTGGGTCATGAAGAAGCAGAGAGACAGTGCAGGATAGGAT 1105
BOVTNF_CDS   ------------------------------------------------------------  283

HUMTNF_GEN   GCTGATGGTAGGCAGAACTTGGAGACAATGTGAGA---------AGGACTCGCTGAGCTC 1186
HUMTNF_CDS   ------------------------------------------------------------  280
SHTNF_GEN    -GGATTGGGGGGCTGAATCCAGACTGAGACAGTGT---------GAGGATTCACAAGCTC 1153
SHTNF_CDS    ------------------------------------------------------------  283
BOVTNF_GEN   TAGGGGGCTGAATCCAGACTGAGGGCAGACAGAGACAGTGTGAGAAGAATTCATGAGCTC 1165
BOVTNF_CDS   ------------------------------------------------------------  283

HUMTNF_GEN   AAGGGAAGGGTGGAGGAACAGCACAGGCCTTAGTGGGATACTCAGAACGTCATGGCCAGG 1246
HUMTNF_CDS   ------------------------------------------------------------  280
SHTNF_GEN    AAGGGGAGGGTGGAGGA-TACTTAGAACTTGAGGGGGCTACTCAGAACCTCATGGCCAGA 1212
SHTNF_CDS    ------------------------------------------------------------  283
BOVTNF_GEN   AAGGGGAGGGTGGTGGA-TACTTAGAACTTGAGGGGGATACTCAGAACCTCATGGCCAGA 1224
BOVTNF_CDS   ------------------------------------------------------------  283

HUMTNF_GEN   TGGGATGTGGGATGACAGACAGAGAGGACAGGAACCGGATGTGGGG--TGGGCAGAGCTC 1304
HUMTNF_CDS   ------------------------------------------------------------  280
SHTNF_GEN    TGGGATGTGGGAACACAGACAGAGGGGACAGGAACCGGATGTGGGGGGTGGGCAGAACTC 1272
SHTNF_CDS    ------------------------------------------------------------  283
BOVTNF_GEN   TGGGATGTGGGAACACAGACAGAGGGGACAGGAACCGGATGTGGGGGGCGGGCAGAACGC 1284
BOVTNF_CDS   ------------------------------------------------------------  283

HUMTNF_GEN   GAGGGCCAGGATGTGGAGAGT-GAACCGACATGGCCACACTGACTCTCCTCTCCCT---- 1359
HUMTNF_CDS   ------------------------------------------------------------  280
SHTNF_GEN    GAGGGCCAGGATGTGGAGAGTAGAACTGACAGGGTCGCACTGACTCACCCCTCCCTGTTC 1332
SHTNF_CDS    ------------------------------------------------------------  283
BOVTNF_GEN   GAGGGCCAGGATGTGGAGAGTAGAACTGACAGGGTCGCACTGA-TCACCCCTCCCTGTTC 1343
BOVTNF_CDS   ------------------------------------------------------------  283

HUMTNF_GEN   CTCTCCCTCCCTCCAGCAAACCCTCAAGCTGAGGGGCAGCTCCAGTGGCTGAACCGCCGG 1419
HUMTNF_CDS   ----------------CAAACCCTCAAGCTGAGGGGCAGCTCCAGTGGCTGAACCGCCGG  324
SHTNF_GEN    TTCCCCCTCCCTCCAGCCAACATCAGCGCTCCGGGGCAGCTCCGATGGGGGGACTCATAT 1392
SHTNF_CDS    ----------------CCAACATCAGCGCTCCGGGGCAGCTCCGATGGGGGGACTCATAT  327
BOVTNF_GEN   TTCCCCCTCCCTCCAGCCGACATCAACTCTCCGGGGCAGCTCCGGTGGTGGGACTCGTAT 1403
BOVTNF_CDS   ----------------CCGACATCAACTCTCCGGGGCAGCTCCGGTGGTGGGACTCGTAT  327
                           *  **       **  ***********  ***  * **

HUMTNF_GEN   GCCAATGCCCTCCTGGCCAATGGCGTGGAGCTGAGAGATAACCAGCTGGTGGTGCCATCA 1479
HUMTNF_CDS   GCCAATGCCCTCCTGGCCAATGGCGTGGAGCTGAGAGATAACCAGCTGGTGGTGCCATCA  384
SHTNF_GEN    GCCAATGCCCTCATGGCCAACGGCGTGGAGCTGAAAGACAACCAGCTGGTGGTGCCCACT 1452
SHTNF_CDS    GCCAATGCCCTCATGGCCAACGGCGTGGAGCTGAAAGACAACCAGCTGGTGGTGCCCACT  387
BOVTNF_GEN   GCCAATGCCCTCATGGCCAACGGTGTGAAGCTGGAAGACAACCAGCTGGTGGTGCCTGCT 1463
BOVTNF_CDS   GCCAATGCCCTCGTGGCCAACGGTGTGAAGCTGGAAGACAACCAGCTGGTGGTGCCTGCT  387
             ************ ******* ** *** *****  *** *****************  *

HUMTNF_GEN   GAGGGCCTGTACCTCATCTACTCCCAGGTCCTCTTCAAGGGCCAAGGCTGCCCCTCCACC 1539
HUMTNF_CDS   GAGGGCCTGTACCTCATCTACTCCCAGGTCCTCTTCAAGGGCCAAGGCTGCCCCTCCACC  444
SHTNF_GEN    GACGGGCTTTACCTCATCTACTCGCAGGTCCTCTTCAGGGGCCACGGCTGCCCTTCCACC 1512
SHTNF_CDS    GACGGGCTTTACCTCATCTACTCGCAGGTCCTCTTCAGGGGCCACGGCTGCCCTTCCACC  447
BOVTNF_GEN   GACGGGCTTTACCTCATCTACTCACAGGTCCTCTTCAGGGGCCAAGGCTGCCCTTCCACC 1523
BOVTNF_CDS   GACGGGCTTTACCTCATCTACTCACAGGTCCTCTTCAGGGGCCAAGGCTGCCCTTCCACC  447
             ** ** ** ************** ************* ****** ******** ******

HUMTNF_GEN   CATGTGCTCCTCACCCACACCATCAGCCGCATCGCCGTCTCCTACCAGACCAAGGTCAAC 1599
HUMTNF_CDS   CATGTGCTCCTCACCCACACCATCAGCCGCATCGCCGTCTCCTACCAGACCAAGGTCAAC  504
SHTNF_GEN    CCCTTGTTCCTCACCCACACCATCAGCCGCATTGCAGTCTCCTACCAGACCAAGGTCAAC 1572
SHTNF_CDS    CCCTTGTTCCTCACCCACACCATCAGCCGCATTGCAGTCTCCTACCAGACCAAGGTCAAC  507
BOVTNF_GEN   CCCTTGTTCCTCACCCACACCATCAGCCGCATTGCAGTCTCCTACCAGACCAAGGTCAAC 1583
BOVTNF_CDS   CCCTTGTTCCTCACCCACACCATCAGCCGCATTGCAGTCTCCTACCAGACCAGGGTCAAC  507
             *   ** ************************* ** ***************** *******

HUMTNF_GEN   CTCCTCTCTGCCATCAAGAGCCCCTGCCAGAGGGAGACCCCAGAGGGGGCTGAGGCCAAG 1659
HUMTNF_CDS   CTCCTCTCTGCCATCAAGAGCCCCTGCCAGAGGGAGACCCCAGAGGGGGCTGAGGCCAAG  564
SHTNF_GEN    ATCCTCTCTGCCATCAAAAGCCCTTGCCACAGGGAGACCCTAGAGGGGGCTGAGGCCAAG 1632
SHTNF_CDS    ATCCTCTCTGCCATCAAAAGCCCTTGCCACAGGGAGACCCTAGAGGGGGCTGAGGCCAAG  567
```

```
BOVTNF_GEN  ATCCTGTCTGCCATCAAGAGCCCTTGCCACAGGGAGACCCCAGAGTGGGCTGAGGCCAAG 1643
BOVTNF_CDS  ATCCTGTCTGCCATCAAGAGCCCTTGCCACAGGGAGACCCCAGAGTGGGCTGAGGCCAAG  567
             **** *********** ***** ***** ********** **** **************

HUMTNF_GEN  CCCTGGTATGAGCCCATCTATCTGGGAGGGGTCTTCCAGCTGGAGAAGGGTGACCGACTC 1719
HUMTNF_CDS  CCCTGGTATGAGCCCATCTATCTGGGAGGGGTCTTCCAGCTGGAGAAGGGTGACCGACTC  624
SHTNF_GEN   CCCTGGTACGAACCCATCTACCAGGGAGGGGTCTTCCAGCTGGAGAAGGGAGATCGCCTC 1692
SHTNF_CDS   CCCTGGTACGAACCCATCTACCAGGGAGGGGTCTTCCAGCTGGAGAAGGGAGATCGCCTC  627
BOVTNF_GEN  CCCTGGTACGAACCCATCTACCAGGGAGGAGTCTTCCAGCTGGAGAAGGGAGATCGCCTC 1703
BOVTNF_CDS  CCCTGGTACGAACCCATCTACCAGGGAGGAGTCTTCCAGCTGGAGAAGGGAGATCGCCTC  627
            ******** ** ******** * ****** ******************** ** ** ***

HUMTNF_GEN  AGCGCTGAGATCAATCGGCCCGACTATCTCGACTTGCCGAGTCTGGGCAGGTCTACTTT 1779
HUMTNF_CDS  AGCGCTGAGATCAATCGGCCCGACTATCTCGACTTGCCGAGTCTGGGCAGGTCTACTTT  684
SHTNF_GEN   AGTGCTGAGATCAACCTGCCGGAATACCTGGACTATGCCGAGTCTGGGCAGGTCTACTTT 1752
SHTNF_CDS   AGTGCTGAGATCAACCTGCCGGAATACCTGGACTATGCCGAGTCTGGGCAGGTCTACTTT  687
BOVTNF_GEN  AGTGCTGAGATCAACCTGCCGGACTACCTGGACTATGCCGAGTCTGGGCAGGTCTACTTT 1763
BOVTNF_CDS  AGTGCTGAGATCAACCTGCCGGACTACCTGGACTATGCCGAGTCTGGGCAGGTCTACTTT  687
            ** *********** * *** ** ** ** **** ************************

HUMTNF_GEN  GGGATCATTGCCCTGTGA 1797
HUMTNF_CDS  GGGATCATTGCCCTGTGA  702
SHTNF_GEN   GGGATCATCGCCCTGTGA 1770
SHTNF_CDS   GGGATCATCGCCCTGTGA  705
BOVTNF_GEN  GGGATCATCGCCCTGTGA 1781
BOVTNF_CDS  GGGATCATCGCCCTGTGA  705
            ******** *********
```

Figure 4.4: Multiple sequence alignment of genomic and CDS sequences of sheep TNF with their human and cattle orthologues.

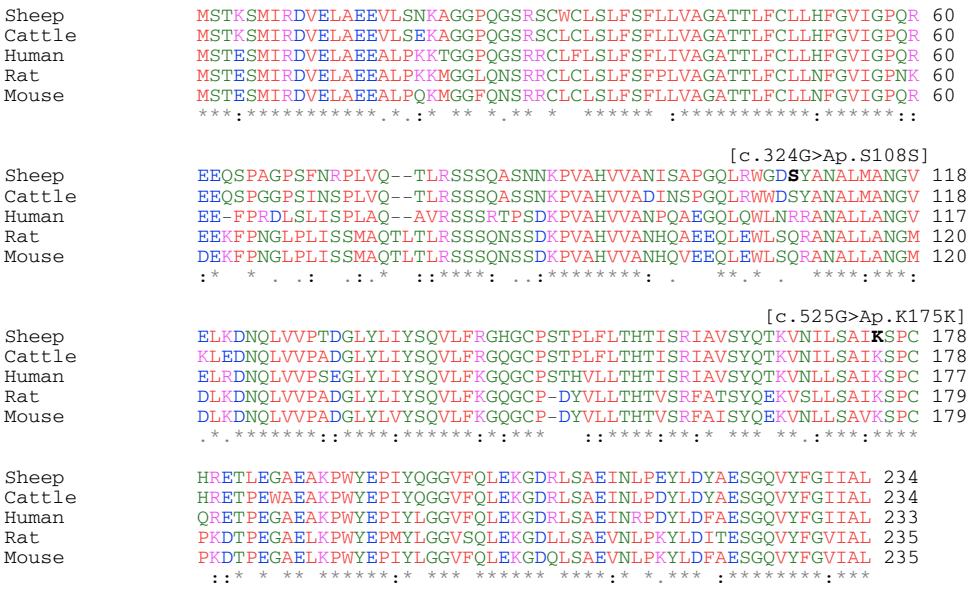


Figure 4.5: Multiple sequence alignment of peptide sequences of sheep TNF (JQ 2007) with their human (NP_000585.2), cattle (NP_776391.2), rat (CAA47146) and mouse (NP_038721.1) orthologues. *Translation of sheep TNF cds sequence into peptide showing 234 predicted amino acids and two synonymous SNPs (highlighted as bold black and annotated).*

Five SNPs were identified in a 3310bp region encompassing the sheep TNF locus including both 5′ and 3′ untranslated regions. In addition to the two SNPs designated S2 and S5 within exon 4 of the TNF, a further three SNPs were present within the 3′ untranslated region of this genomic sequence. The SNP panel is summarised in Table 3.9 (see chapter 3).

4.4 Discussion

The results described in this paper support the physical map of the central region of the sheep MHC reported in Chapter 3. In addition to the three complement loci, two other loci, SKIV2L and RDBP, were also identified in this region and the previously reported CA short tandem repeat locus designated BfMs (Groth and Wetherall, 1995) was shown to occur within the SKIV2L locus. A complete genomic sequence of sheep CFB is reported which encodes a protein of 761aa and manifests nine SNPs of which two are nonsynonymous. It was also observed that for two of the sheep CFB SNPs (BF5 and BF8) all four possible haplotypes were present in the sheep population tested. This is a surprising observation and implies gene conversion within the CFB locus. It has been observed that gene conversion events are

relatively prevalent within MHCs (Moore *et al.,* 1992). In addition, a corrected genomic sequence of cattle CFB is reported together with its predicted CDS and protein. Both sheep and cattle CFB proteins are three amino acids shorter than human CFB protein and their genes manifest a common 9bp deletion in exon 18. This deletion may be characteristic of ungulates. The panel of sheep CFB SNPs will be of value in determining MHC haplotypes within this genomic region.

A partial genomic sequence encoding exons 10-18 of a sheep C2 gene was obtained. Comparative analysis of this sequence with both cattle and human C2 genomic and CDS sequences confirmed the identity of the sheep C2. It was observed that the cattle C2 gene had non-contiguous deletions of 6 and 3 bps in exon 18 and that the predicted corresponding protein was two amino acids shorter than human C2 protein. The sheep C2 sequence also shared the same deletions in exon 18 as the cattle C2 gene suggesting once again that this may be a characteristic of ungulates.

Complete sequences for the RDBP and SKIV2L loci were not obtained although sufficient sequence of SKIV2L (approximately 3000bp) was obtained to confirm the presence of the polymorphic CA repeat locus originally described by Groth and Wetherall (1995). Similarly, an incomplete sequence of sheep C4 (8454bp) was obtained which was of the C4B isotype. Collectively, these data provide more detailed information concerning the structure and genetic variation of this region of the sheep MHC and to a lesser extent that of cattle. All sequences were submitted to GenBank.

Some earlier studies reported a predicted length of 233 amino acids for cattle and sheep TNF peptides (Gupta *et al.,* 2004). The multiple sequence alignment for TNF genomic DNA from human sheep and cattle, permitted the identification of an A->G mutation four nucleotides from the end of intron 1 which creates another AG splice site and results in three additional bases (CAG /Q) in exon 2. Exon 2 in sheep and cattle therefore encodes an additional glutamine resulting in a peptide one amino acid longer than human TNF protein.

This conclusion is supported by the study of Nash *et al.* (1991) who also obtained a 234 amino acid peptide from a cDNA construct for sheep TNF. Gupta and colleagues (2004) also reported the presence of an additional codon at the same location in cDNA sequence of TNF from Indian cattle (*Bos indicus*) but not from Indian water buffalo. These corresponded with the peptide length deduced from the

cDNA constructs thereby confirming the conclusion above. Odbileg at al (2005) reported peptide lengths of 233 amino acids based on cDNA sequences of 699bp for these species. We would predict therefore that the H->G mutation is not present in at least some members of these two species but genomic sequences will be required to confirm this prediction. The variable presence of the mutation creating the alternative splice site for intron 1 may well lead to an incorrect predicted TNF peptide unless attention is paid to the presence or absence of the H->G mutation in the fourth last position of intron 1. For example, a dog TNF sequence in the Ensembl database (ENSCAFG00000000517) is reported as 234aa rather than as 233aa. Interestingly, inspection of genomic sequences for rat and mouse TNF show a mutation in the third last nucleotide of intron 2 (C->A) together with the mutation in the fourth last position relative to human TNF which results in an additional lysine rather than a glutamic acid in exon 2 thereby creating a nonsynonymous SNP for these two species. However in rat and mouse TNF there is a double insertion (T, L) at positions 76 and 77 and a deletion (S) at position 150 which, results in an overall peptide length of 235 amino acids as previously reported (Accession Nos. NP_038721.1 and CAA47146).

It is of interest that the three SNPs reported by Alvarez-Busto and colleagues (2004) in sheep TNF were all present in the 3' untranslated region. A comprehensive search for SNPs in human *TNF* revealed SNPs within introns and the 3' untranslated region, but not within exons (http://pga.gs.washington.edu/data/tnf/). One of the SNPs discovered in the 3' untranslated region (called TNFαS1 in Table 3.9, in chapter 3) was the same as that reported by Alvarez-Busto and colleagues (2004). The three base deletion (ACA) in the 3' UTR reported by Nash *et al*. (1991), and confirmed by Alvarez-Busto *et al.* (2004), in sheep TNF was not present in the sequence reported in this study.

In summary, the work included in this chapter describes for the first time, the complete genomic sequences of sheep TNF and CFB genes together with substantial partial sequences for sheep C2 and C4 genes. Multiple sequence alignments have confirmed the homology of these loci with cattle and human orthologues and an explanation for the varying lengths of the TNF protein in several species is provided. The direct sequencing also permitted identification of small panels of SNPs for the CFB and TNFα loci, including two new synonymous SNPs in exon 4 of TNFα.

Since TNF is an important cytokine, which is likely to be involved in a variety of immune responses to infection, the SNPs reported in this chapter may be particularly relevant in disease association studies.

Chapter 5

Genotyping of Polymorphic Loci within the Central Region of the Sheep MHC

This chapter describes the genotyping of two sheep panels, namely a three generation half sib family comprising 25 sheep and a panel of fifty eight unrelated sheep from the Rylington Merino Project. Allele frequencies for the unrelated animals in these two panels are reported together with the unequivocal identification of haplotypes in thirteen parental animals from the three generation half sibship pedigree. Twenty two distinct haplotypes were identified out of a possible 26 haplotypes. Pairwise linkage disequilibrium between the SNPs specific for 10 distinct genes was estimated. A subregion within the sheep MHC central region manifesting high linkage disequilibrium was identified which is indicative of a putative conserved block like structure similar to analogous structures observed in the human MHC. A high level of intragenic haplotypic variation was observed within the TNFα locus.

5.1 Introduction

Microsatellite DNA markers have been extensively used in population and disease association studies in many species. Several polymorphic microsatellite loci have been identified in the sheep MHC, including OLADRB and OLADRBps within the class II region (Schwaiger *et al*., 1996; Blattman and Beh, 1992), OMHC1 (Groth and Wetherall, 1994) within the class I region, and BfMs in the class III region (Groth and Wetherall, 1995). Although microsatellite markers are very informative, they manifest high mutation rates and display both convergent and divergent evolutionary patterns. Furthermore, by their nature microsatellites are not easily adapted to a high throughput genotyping environment (Kwok and Gu, 1999).

Single nucleotide polymorphisms (SNPs) are ubiquitous and comprise much of the sequence variation observed in the human genome with an estimated density of 1 per kb in the human genome (Kwok and Gu, 1999; Geraghty, 2002). SNPs have lower mutation rates per generation than microsatellites, are stable and are more easily adapted to high throughput methods of analysis (Lazarus *et al*., 2002). SNPs may be synonymous (sSNP) when their presence in a coding region does not result in an amino acid substitution or non-synonymous (nsSNP) when

their presence in a coding region results in an amino acid substitution. Non synonymous SNPs occur less frequently than synonymous SNPs.

With the advent of automated typing technologies SNPs are relatively easy to detect and suitable for high throughput analysis at relatively low unit cost (Johnson and Todd, 2000). Patterns of linkage disequilibrium and haplotypic variation based on high-density panels of SNPs may help to identify recombination 'hot-spots' and hence assist in genome mapping (Lazarus *et al*., 2002). SNPs have been widely used for identifying genes associated with multifactorial disorders and quantitative trait loci (Kaminski *et al*., 2006) and a great deal of effort has been expended on their discovery and use in disease association studies.

Over recent years many methods have been developed for SNP genotyping. These include RFLP of PCR amplicons, single strand conformational polymorphisms, allele specific PCR (also referred to as ARMS - amplification refractory mutational sequence), DGGE, oligonucleotide microarray (e.g. APEX-Arrayed primer extension) as well as direct sequencing (Kaminski *et al*., 2006). In this study all SNP genotypes (based on the SNP panel described in Chapter 3) in the animal panels were obtained using the pyrosequencing method described below. However, as noted in the text, in some instances animals were also typed using the RFLP and allele specific PCR methods as adjunct methods providing checks for the accuracy of the pyrosequencing method. An analysis of MHC central region haplotypes based on the SNP panel was then undertaken using 25 animals from a three generation half sib sheep family together with a second panel of 58 unrelated sheep from the Rylington Merino Project. The SNP profiles for each sheep were supplemented with genotypes of the microsatellite locus BfMs which is located within the SKIV2L gene.

5.2 Materials and methods

5.2.1 Animals genotyped

DNA samples used in this study were isolated from 25 animals from the three generation half sibling family for which blood samples were available together with the panel of 58 unrelated animals; both groups of animals are described in Chapter 2. The twenty five members of the half sibling family were subdivided into two groups comprising the 13 parental animals and their 12 offspring and

are shown in Figure 5.1. The 13 parental sheep were not closely related and in this regard are similar to the panel of 58 unrelated sheep.

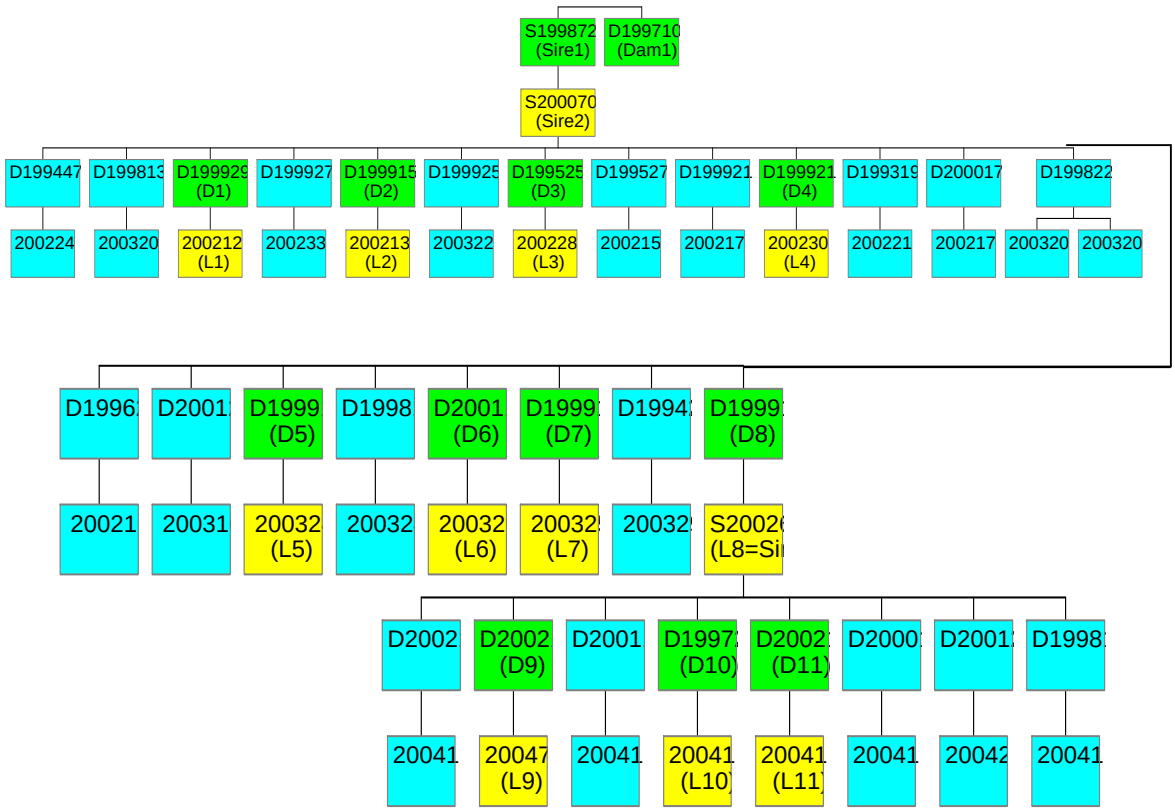


Figure 5.1: Pedigree chart for three generation half sibling sheep family. Animals shown in green and yellow colours had blood samples available for genotypic analysis from which haplotypes could be deduced. Animals shown in green comprise the 13 parental animals which were included as part of the cohort of 71 unrelated animals described in the text.

5.2.2 SNP genotyping

All SNPs were genotyped using the pyrosequencing method described below. However, initially some animals were also genotyped using restriction fragment length polymorphism analysis (RFLP) of PCR generated amplicons and allele specific amplification methods. These latter two methods are also described briefly below

5.2.2.1 Genotyping using the pyrosequencing method

The PSQ 96MA pyrosequencing protocol (Dr. Gershon's Lab, Dept of Psychiatry, University of Chicago) was followed. Primer3 (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi)

and Netprimer software (http://www.premierbiosoft.com/netprimer) were used to design primers to amplify an amplicon of less than 200bp containing a SNP site. Primers for pyrosequencing were designed using software provided by the company that developed the procedure (http://primerdesign.pyrosequencing.com/jsp/TemplateInput.jsp). A universal biotin-labeled primer, designated Univ2_B (manufactured by Sigma-Proligo) was used for each SNP locus tested as suggested by the pyrosequencing protocol.

Table 5.1 Primers for SNP typing using the pyrosequencing method

GENE	PRIMERS 5'-3'
Apom	JQAPOMS1F_2nd: GGGACACCGCTGATCGTTTAAAGGGAGGACCTGAGGATTC JQAPOMS1R_2nd: TTGAGTGAGCAAACCCAACC JQAPOMS1.7R: GCCCCTATGAGCTG
Bat2	JQBAT2S1F: GGGACACCGCTGATCGTTTACTCCCCTTTGTGCCAGTCTA JQBAT2S1R: CGGAGCTGTATGCCTCTCTC JQBAT2S1.1R: TGCTCTAGGGCTTGG
Bat5	JQBAT5S1F: CATGAGTGCAGACGGAAGAC JQBAT5S1R: GGGACACCGCTGATCGTTTAACAGGTGGGGTACAGAGCAC JQBAT5S1.1F: TGGGGCTTGGGGA
BF	BFS2F: TCTGTGTAGGCAGCCTCCTT BFS2R: GGGACACCGCTGATCGTTTAATCAAGGTCAGCGTGGGTAA BFS2.E1F GAGGGCACTGCAGAA
C2	JQC2S4F: GGGACACCGCTGATCGTTTACCGATGGCATAGATGTCTGA JQC2S4R: AAGGGGAGGAGTGGTAGGAA JQC2S4.1R: GGAGGGGAGCGGA
G6b	JQG6Bs1F_2nd: GGGTGTGAGAGACAGGGGTA JQG6bS1R: GGGACACCGCTGATCGTTTAGAAGGCACCATGGGATAGAA JQG6Bs1F.12F: GAGATAGCCTGGAATGTA
LTA	JQLTAS2F_2nd: GGGACACCGCTGATCGTTTAGTCAAGGGTGCAGAGGTGTT JQLTAS2R_2nd: CGACTGAGTTCTGGGAAAGG JQLTAS2.2R: TCTAGAGAAGGGACAGTAAA
MSH5	JQMSHS4F_2nd: TGTTCGTTGAGGGGAATCTC JQMSH5S4R: GGGACACCGCTGATCGTTTACAGTTTGCCTGAGGCTTTTC JQMSH5S.1F: GTCCTTTTCTGCTCCA
TNXB	JQTNXBS1S2F: GGGACACCGCTGATCGTTTAAGGAGAGATGCAGGCTGTGT JQTNXBS1R: CTTTCCTCCTTCCCTCCATC JQTNXBS1.1R: CACCCCTCTCTACCTGT JQTNXBS2.1R: CCAGGGAAGAAATCG
TNFα	TNFaS1F: GGGACACCGCTGATCGTTTAGAGCGGAGGTTCAGTGATGT TNFaS1R: GCCTTGGCTCAGATGTGTTT TNFS1.E1R CCTTTGTGCCTCCTT JQTNFaS2S3F: GGGACACCGCTGATCGTTTAGAGCGGAGGTTCAGTGAT JQTNFaS2S3R: GGGTTCTTACCGGAATACTT JQTNFaS2.1R: CCTTTGTGCCTCCTT JQTNFaS3.1R: CTCCCTTCCTGCCA JQTNFaS5F: GGGACACCGCTGATCGTTTATGATGGCAGAGAGGATGTTG JQTNFaS5R: ACTCACCCCTCCCTGTTCTT JQTNFaS5.1R: CGATGGGGGGACTC
Univ2_B	5'- Biotin - GGGACACCGCTGATCGTTTA -3'

A 25μl PCR was performed containing 100ng sheep genomic DNA, 5pmol untagged primer, 4.5pmol Univ2_B primer, 0.5pmol tagged primer, and 2.5 unit of Taq polymerase. All PCRs were performed at 60°C for annealing temperature, 20 sec for extension and 50 cycles. 10μl of PCR product was used for pyrosequencing.

The pyrosequencing procedure was performed at Royal Perth Hospital, WA, using a pyrosequencing primer diluted in 1 x Annealing Buffer, 2 x Binding Buffer, Streptavidin Sepharose Beads (Amersham Biosciences, Uppsala, Sweden), 0.2M NaOH, 10mM Tris PH7.6, 70% ethanol and a PSQ HS96 pyrosequencing reagent kit. Samples were analyzed and result collected on a PSQ HS96A instrument with proprietary pyrosequencing software (Biotag, Uppsala, Sweden).

5.2.2.2 RFLP Analysis (PCR followed by RFLP)

A PCR amplicon containing a SNP site was amplified and digested with 2 units of an appropriate restriction endonuclease enzyme that recognized and digested at the SNP site. The resulting DNA fragments were separated by electrophoresis on a 3% agarose gel. Analysis of the fragments observed permitted identification of the SNP allele(s) present and hence identification of the possible two homozygous and one heterozygous genotypes generated from the DNA sample analyzed.

5.2.2.3 Allele specific PCR (ARMS: amplification refractory mutational sequence)

A general approach for typing SNPs is called allele specific PCR in which the terminal 3' nucleotide is complementary to one of the two alleles characteristic of the SNP. ARMS is one variation of this general method and was used in this study. PCR primers were designed using a method of Drenkard *et al*. (2000) with minor modification in which three sets of primers had been designed to carry out nested PCRs. In addition, these oligonucleotides contained an additional mismatch at the penultimate nucleotide adjacent to the 3' end of the reverse primer. Standard PCR reaction mixes (see Table 2.1, Chapter 2) were used for all amplifications. In brief, a primary PCR was performed at an annealing temperature of 55°C for 25 cycles. Two secondary PCRs were then performed using a dilution (1 in 10000) of the primary PCR products. The annealing temperature and the number of PCR cycles vary for every SNP allele. In the case of TNFαS1 locus, detection of the G allele required 57°C and 25 cycles whereas

the A allele required 52°C and 30 cycles for assays based on the oligonucleotide primers shown in Table 5.2.

Table 5.2: Allele specific oligonucleotide primers for TNFαS1 genotyping using the modified ARMS assay

	Forward Primers (5'-3')	Reverse Primers (5'-3')
Primary PCR	AGCTTGGTGCAGGCTACTGT	TTCAGGAGGTCAAGGTGTCC
Allele-G PCR	AGCTTGGTGCAGGCTACTGT	GGCCTTTGTGCCTCCTTTAG
Allele-A PCR	AGCTTGGTGCAGGCTACTGT	GGCCTTTGTGCCTCCTTTAA

Nucleotides in red colour are modified mismatches relative to the DNA template. Allele-specific nucleotides are highlighted in blue colours.

5.2.2.4 Microsatellite assay

The BfMs dinucleotide microsatellite locus previously identified by Groth and Wetherall (1995) is located within an intron of the SKIV2L locus. Primers to amplify the CA repeats region were designed using the Primer 3 software package (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi). The forward primer (SKIV2LMF*:5' GCCTGATGATTAGTGGAAAA 3') was labeled with an infrared dye (purchased from LI COR - see section 2.6); the reverse primer was unlabelled (SKIV2LMR:5' AACCGAATTACCTAGGGTTC 3'). PCR was performed using standardized reaction mixes with an annealing temperature of 50°C for 30 cycles as described in Table 2.1. Amplicons of approximately 200bp present in an aliquot of the product mix (0.7-1.0μl) were separated on a 4% polyacrylamide mini-gel using the LI-COR 4000 sequencer (see Chapter 2.6.1) which permitted identification of the phenotype from which genotypes could then be deduced. Alleles were named based upon the number of repeating elements e.g. allele 18 was (CA)18.

5.2.3 Analysis of SNP and microsatellite genotypic data

The panel of 12 SNP markers representing 10 loci within the central region of the sheep MHC (see Chapter 3) was typed for both sheep panels. The complex genotype results obtained were analyzed using the web based interactive SNPstats software (http://bioinfo.iconcologia.net/index.php?module=Snpstats). Allele frequencies, haplotype frequencies, linkage disequilibrium and statistical significance were generated using this software. SNPstats estimates haplotype frequencies using an implementation of the expectation maximization (EM) algorithm previously written for the *haplo.stats* application (Sinnwell and Schaid,

2005). The population genetical analysis of genotypes for the two sheep populations was performed using the internet version of the GenePop application (Raymond and Rousset, 1995) developed and maintained by Curtin University (http://genepop.curtin.edu.au/).

An important convention was used when analyzing SNP data. Since the direction of transcription of the individual loci studied was usually unknown, or that the SNP was present in a non coding region, it is not known on which DNA strand the SNP allele occurs. As all mutations typed as SNPs were transitions, they have been identified therefore as either a T (for T or A) and as a C (for C or G). Use of this convention simplifies the analysis of haplotypic patterns and importantly does not result in any loss of information.

5.3 Results

5.3.1 Genotyping of individual loci and allele frequencies

The pyrosequencing method was robust and generated SNP genotypes for all loci and animals tested. A number of the genotypes obtained by pyrosequencing were confirmed using the alternative RFLP and allele specific PCR methods with no discrepancies being observed. A typical result of a pyrosequencing analysis for the TNFα SNPS1 locus is shown in Figure 5.2.

Genotypes for the panel of twelve SNPs together with the BfMs microsatellite locus were typed on 25 members of the three generation half sib sheep family for which blood samples were available for testing. These animals and their family relationships are identified by the green and yellow squares shown in Figure 5.1 and comprise a part of the larger pedigree included in this Figure. The genotypes for these 25 animals are shown in Table 5.3. Genotyping of the panel of 58 unrelated sheep was also performed and the genotypes for this group are shown in Table 5.4 which for convenience also includes the 13 unrelated parental animals from the sheep pedigree making a total of 71 sheep typed.

Allele frequencies for all SNPs plus the microsatellite locus were estimated from the genotypes for this panel of 71 sheep. These are summarized in Table 5.5 and Table 5.6A.

Table 5.5: Allele frequencies for 12 SNP loci in the central region of the sheep MHC for 71 unrelated animals.

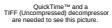
Locus	Allele 1 Frequency	Allele 2 Frequency
LTAS2	T (0.94)	C (0.06)
TNFaS1	C (0.62)	T (0.38)
TNFaS3	C (0.78)	T (0.22)
TNFaS5	C (0.86)	T (0.14)
BAT2S1	T (0.73)	C (0.27)
APOMS1	C (0.83)	T (0.17)
BAT5S1	C (0.70)	T (0.30)
G6bS1	C (0.75)	T (0.25)
MSH5S4	C (0.79)	T (0.21)
C2S4	C (0.96)	T (0.04)
CFBS2	T (0.83)	C (0.17)
TNXBS2	C (0.59)	T (0.41)

Table 5.6A: Allele frequencies for the BfMs (SKIV2L) microsatellite locus in the central region of the sheep MHC for 71 unrelated animals comprising 58 animals and 13 unrelated parentals from the half sibling pedigree.

Sheep_71											
Allele No.	15	16	17	18	19	20	21	22	23	24	Totals
Frequency	0.07	0.021	0.092	0.676	0.092		0.007	0.014	0.014	0.014	1
Frequency%	7	2.1	9.2	67.6	9.2	0	0.7	1.4	1.4	1.4	100

TNFα SNPS1-G/A (C/T)

TNFα SNPS1-A/A



TNFα SNPS1-G/G

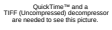


Figure 5.2 *Three genotypes for the TNFαSNPS1 locus detected by the pyrosequencing method. Top panel: Two alleles present (G/A heterozygote). Middle panel: Homozygous A/A genotype. Bottom panel: Homozygous G/G genotype.*

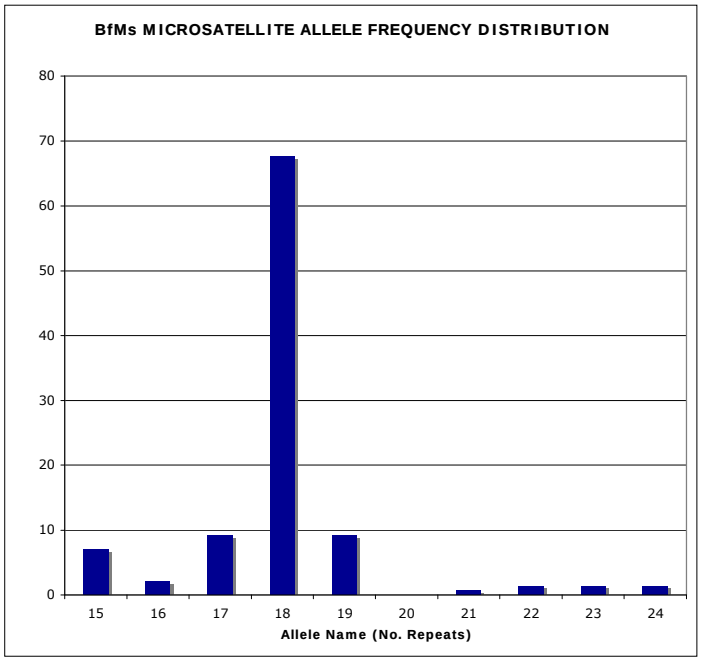


Figure 5.3: *Allele frequencies for the BfMs (SKIV2L) microsatellite locus in the central region of the sheep MHC for 71 unrelated sheep comprising 58 animals and 13 unrelated parental animals from the half sibling pedigree.*

5.3.2 Genetic characterization of sheep populations

Each of the two populations of sheep included in this analysis (ie panel of 58 unrelated and 13 parental sheep from the pedigree) was tested for Hardy Weinberg proportions and genetic differentiation using the Genepop population genetics software application (http://genepop.curtin.edu.au/). This analysis showed that based on genotypes from all 10 distinct genes plus the BfMs microsatellite locus, both populations of sheep manifested Hardy Weinberg proportions. The observed and expected heterozygosities for the two sheep cohorts are shown in Table 5.6B. On the basis of this analysis it was concluded that the combined population of 71 sheep could be considered as a population of unrelated animals.

Table 5.6B: Observed and expected heterozygosities for cohorts of sheep

LOCUS	Observed Heterozygosity	Expected heterozygosity	No. Chromosomes
Combined cohort (n = 71)			
LTAS2	6	7.6	142
TNFaS1	33	34.4	142
BAT2S1	23	28.5	142
APOMS1	20	20.1	142
BAT5S1	31	31.7	142
G6bS1	25	25.5	142
MSH5S4	22	21.4	142
C2S4	14	12.7	142
CFBS2	25	25.5	142
TNXBS2	32	35.2	142
BfMs	39	37.2	142
Cohort 1 (n = 58)			
LTAS2	5	6.6	116
TNFaS1	24	27.5	116
BAT2S1	17	22.9	116
APOMS1	16	16.7	116
BAT5S1	21	24.7	116
G6bS1	21	21.9	116
MSH5S4	20	19.2	116
C2S4	5	4.8	116
CFBS2	16	16.7	116
TNXBS2	26	28.4	116
BfMs	28	28.1	116
Cohort 2 (n = 13)			
LTAS2	1	1.0	26
TNFaS1	9	6.8	26
BAT2S1	6	5.8	26
APOMS1	4	3.5	26
BAT5S1	10	6.7	26
G6bS1	4	3.5	26
MSH5S4	2	1.9	26
C2S4	9	6.1	26
CFBS2	9	6.8	26
TNXBS2	6	6.7	26
BfMs	11	8.6	26

Table 5.8: Estimation of pairwise linkage disequilibrium between 10 distinct SNP loci in a cohort of 71 unrelated sheep using SNPstats software. The three usual measures of linkage disequilibrium are reported, namely the linkage disequilibrium parameter D, D' which reports the magnitude of D as a percentage of its maximum values for the two loci in question, the r statistic which corresponds to the correlation coefficient between alleles at each of the two loci, and a probability (p) value supporting the hypothesis that the r statistic observed is different from zero. There is a putative region of significant linkage disequilibrium between loci BAT2S1 and MSH5S4 identified by the red blocks.

D statistic

	LTAS2	TNFaS1	BAT2S1	APOMS1	BAT5S1	G6bS1	MSH5S4	C2S4	CFBS2	TNXBS2
LTAS2	.	-0.023	0.004	0.0013	-0.019	-0.003	-0.003	0.0036	-0.013	0.0142
TNFaS1	.	.	-0.100	-0.068	0.1981	-0.045	-0.074	0.0074	0.0235	0.056
BAT2S1	.	.	.	-0.046	-0.091	0.0821	-0.028	0.012	-0.009	0.0055
APOMS1	.	.	.	.	-0.056	-0.027	0.1086	0.0028	0.0279	-0.064
BAT5S1	.	.	.	.	.	-0.077	-0.061	0.02	0.0242	0.0439
G6bS1	.	.	.	.	.	.	-0.043	-0.021	-0.054	-0.006
MSH5S4	.	.	.	.	.	.	.	-0.018	0.0212	-0.059
C2S4	.	.	.	.	.	.	.	.	0.0756	-0.017
CFBS2	.	.	.	.	.	.	.	.	.	-0.033
TNXBS2	.	.	.	.	.	.	.	.	.	.

D' statistic

	LTAS2	TNFaS1	BAT2S1	APOMS1	BAT5S1	G6bS1	MSH5S4	C2S4	CFBS2	TNXBS2
LTAS2	.	0.997	0.0972	0.0285	0.9964	0.2422	0.3155	0.0705	0.9949	0.4464
TNFaS1	.	.	0.9106	0.9989	0.9997	0.4795	0.9994	0.1262	0.1686	0.2475
BAT2S1	.	.	.	0.9984	0.9992	0.4869	0.5524	0.168	0.1413	0.0354
APOMS1	.	.	.	.	0.9989	0.6963	0.7866	0.0344	0.2152	0.8604
BAT5S1	.	.	.	.	.	0.9992	0.9993	0.3034	0.1557	0.2354
G6bS1	.	.	.	.	.	.	0.999	0.9092	0.9991	0.0548
MSH5S4	.	.	.	.	.	.	.	0.996	0.1508	0.7348
C2S4	.	.	.	.	.	.	.	.	0.999	0.3919
CFBS2	.	.	.	.	.	.	.	.	.	0.3226
TNXBS2	.	.	.	.	.	.	.	.	.	.

r statistic

	LTAS2	TNFaS1	BAT2S1	APOMS1	BAT5S1	G6bS1	MSH5S4	C2S4	CFBS2	TNXBS2
LTAS2	.	-0.120	0.0386	0.0154	-0.171	-0.033	-0.037	0.0521	-0.134	0.1239
TNFaS1	.	.	-0.459	-0.369	0.8587	-0.216	-0.388	0.051	0.1133	0.2302
BAT2S1	.	.	.	-0.277	-0.432	0.4354	-0.161	0.0903	-0.048	0.0248
APOMS1	.	.	.	.	-0.317	-0.173	0.7493	0.0252	0.1764	-0.342
BAT5S1	.	.	.	.	.	-0.387	-0.333	0.1427	0.1218	0.1881
G6bS1	.	.	.	.	.	.	-0.260	-0.166	-0.303	-0.027
MSH5S4	.	.	.	.	.	.	.	-0.156	0.1297	-0.306
C2S4	.	.	.	.	.	.	.	.	0.601	-0.114
CFBS2	.	.	.	.	.	.	.	.	.	-0.156
TNXBS2	.	.	.	.	.	.	.	.	.	.

P-values

	LTAS2	TNFaS1	BAT2S1	APOMS1	BAT5S1	G6bS1	MSH5S4	C2S4	CFBS2	TNXBS2
LTAS2	.	0.0174	0.6455	0.8541	0.0413	0.698	0.6636	0.5348	0.111	0.1398
TNFaS1	.	.	0	0	0	0.01	0	0.5435	0.177	0.0061
BAT2S1	.	.	.	0.001	0	0	0.0551	0.282	0.5686	0.7679
APOMS1	.	.	.	.	2e-04	0.0395	0	0.7637	0.0355	0
BAT5S1	.	.	.	.	.	0	1e-04	0.0891	0.1468	0.025
G6bS1	.	.	.	.	.	.	0.0019	0.0487	3e-04	0.7517
MSH5S4	.	.	.	.	.	.	.	0.0631	0.1221	3e-04
C2S4	.	.	.	.	.	.	.	.	0	0.1739
CFBS2	.	.	.	.	.	.	.	.	.	0.0626
TNXBS2	.	.	.	.	.	.	.	.	.	.

5.3.3 Identification of haplotypes from the sheep MHC central region

Deduced haplotypes based on 10 distinct gene specific SNPs, plus the BfMs microsatellite locus, for the 13 unrelated members of the sheep pedigree cohort were analyzed. These are shown in Figure 5.4 where they have been colour coded and sorted for similarity. Four pairs of haplotypes were identical resulting in 22 distinct haplotypes from the 26 haplotypes present. The loci comprising these haplotypes spanned a major part of the presumed central region of the sheep MHC. In the corresponding region of the human MHC these loci span approximately 550kbp. From Figure 5.4 it seems likely that the distribution of SNP alleles over the length of the region covered is not random and that some chromosomes share conserved regions particularly at the LTAS2 end of the region.

Haplotype frequencies were also estimated in the cohort of 71 unrelated animals for which SNP genotypes were obtained using an implementation of the expectation maximization EM algorithm in the SNPstats software package. The BfMs locus was excluded from this analysis since SNPstats only allows for analysis of single nucleotide polymorphism data.

The resulting inferred haplotypes, shown in decreasing order of frequency, are listed in Table 5.7. The first six of these predicted haplotypes account for just over 50% of the total predicted haplotypes in this cohort and five of these six were also present in the known haplotypes from the 13 parental animals in the sheep half sibling pedigree. The SNPstats software also permitted estimates of pairwise linkage disequilibrium between the 10 SNP loci for this cohort. The results of this analysis are shown in Table 5.8 and demonstrate a chromosomal region bounded by the BAT2S1 and MSH5S4 loci which manifests relatively high linkage disequilibrium and may be indicative of a conserved region resulting from reduced recombination. The increasing Type I errors arising from successive pairwise linkage disequilibrium estimations will require further data to confirm this putative observation.

5.3.4 Intragenic haplotypes

Three of the SNPs typed in this study were located within the TNF□ gene and are designated TNFS1, S3 and S5. Alleles for these SNPs for the 25 sheep included in the pedigree are shown in Table 5.3. Inspection of Table 5.3 shows that six of the eight possible intragenic haplotypes were present in this sample of sheep.

5.4 Discussion

This chapter describes the genotyping of individual sheep using the panel of SNP markers described in Chapter 3. Genotyping was performed using three methods, namely allele specific PCR, RFLP/PCR and pyrosequencing. These methods were all useful however their efficiency, accuracy and cost varied considerably. Allele specific PCR is relatively inexpensive and suitable for high throughput analysis, but is difficult to optimize because in some cases the modified 3' primer end mutation does not provide enough discrimination power to distinguish between the two alleles (Drenkard *et al*., 2000). Even the introduction of a second mismatched nucleotide immediately adjacent to the 3' primer end did not always overcome this problem. Hence, although this method was used initially in this study for the TNFa SNPs it was not pursued further. In contrast, the RFLP/PCR method is more reliable but is more expensive and requires the presence of a restriction enzyme cleavage site that includes the SNP. This method was used primarily for the CFB SNPs but also was not pursued. In the final analysis both of these methods were only used to confirm genotypes obtained using the pyrosequencing method.

The disadvantages of the two SNP genotyping methods described above led to utilization of the pyrosequencing method for all of the genotyping reported in this study. Subject to the availability of equipment, pyrosequencing is an inexpensive robust and accurate genotyping technology suitable for high throughput. The assay is very reliable so long as the PCR step is optimized to give a single product and the pyrosequencer is adjusted for the appropriate stringency. Moreover, one PCR product can sometimes be used to detect two or more SNP loci. As indicated above, all of the genotypes reported in this study were typed using pyrosequencing, but some were independently confirmed using at least one of the other two methods.

In contrast to humans, microsatellite loci were sparse within the sheep regions sequenced during this study. It has been reported that the average density of microsatellite loci within the human genome is one in 2-3kbp and 297 microsatellite have been identified in the central region of the human MHC (Matsuzaka *et al*., 2001). Only 5 short tandem repeat loci were observed in the ≈69 kbp sequenced in this study and all of these were monomorphic in the panel of sheep tested. Therefore the BfMs microsatellite within the SKIV2L gene reported by Groth and Wetherall (1995) represents a relatively infrequent

polymorphic marker within this genomic region. It is of interest that another project from the Curtin University laboratory has successfully identified a panel of polymorphic tetranucleotide microsatellite loci in sheep, which permits individual identification with high discrimination. Thus perhaps the central region of the sheep MHC is unusual in this regard. In contrast, abundant SNPs in the central region of the sheep MHC have been observed with an estimated frequency of ≈1 per kilobase pair. Both synonymous and non-synonymous SNPs were observed.

Analysis of the genotypes obtained for 25 members of a three generation half sibling pedigree has permitted the identification of 22 of a possible 26 central region haplotypes in the thirteen parental sheep included in this group. Four pairs of haplotypes were identical. The distribution of genotypes at the BfMs locus for these thirteen animals manifested Hardy Weinberg proportions confirming other anecdotal information that these thirteen parental animals are unrelated. Furthermore, when these 26 haplotypes were sorted across loci based on similarity, the presence of relatively conserved genomic subregions was indicated. These results, albeit based on a small cohort of animals, suggests that the sheep central region may exhibit an organized structure similar to the block like structure reported for the analogous region of the human MHC (Stenzel *et al*., 2004).

The observations reported in this Chapter are consistent with two previous studies of haplotypic diversity in the central region of the human MHC. Stenzel *et al.* (2004) reported patterns of linkage disequilibrium based on 968 SNPs across ≈10Mb of chromosome 6p21 in five distinct populations. Well defined block like structures were observed together with extensive regions lacking any clear structure. Closer examination of their data suggest that the central region of the human MHC contains several block like sub regions characterized by lower recombination including one block at the TNF end of the region which may be similar to the putative block reported in this Chapter. Although most of the observations contributing to the definition of block like regions within the human MHC have been based on haplotypic typing in family groups, Jeffreys and colleagues have utilized human sperm typing to provide independent evidence for this phenomenon (Jeffreys *et al*., 2001; Jefferys and Neumann, 2002). Thus individual sperm typing demonstrated six meiotic recombination "hot spots" in a 216kbp region the class II region of the human MHC (Jeffreys *et al*., 2001). Further, Webb *et al.* (2008) working in Jeffreys' laboratory has shown that regions of the human genome manifesting high levels of recombination are associated with extreme breakdown of marker association.

There is strong evidence that recombination hot spots defining relatively conserved subregions of the genome are characteristic of mammalian genomes. Recombination hot spot loci are not conserved between humans and chimpanzees and indeed exhibit polymorphism within human males as shown by sperm typing. Although the mechanisms determining their formation and structure are still largely unknown, Coop and Myers (2007) have proposed that a phenomenon termed 'biased gene conversion' could explain their appearance and disruption. Clearly, this is a field in which much further research is required to elucidate the nature and formation of these sequences, which facilitate enhanced meiotic recombination. Apart from their theoretical importance, they are of practical value in reducing the number of genotypes required to undertake large studies of marker associated searches for disease genes and productivity traits.

The discovery of three SNPs within the 5.5kbp sequence encompassing the TNFα gene permitted the identification of six of a possible eight intragenic haplotypes at this locus. This result implies a high level of intragenic recombination over the ≈5.5kbp extent of this sequence. The explanation for this observation is not clear. It may be that this locus is included within a hot spot for recombination, or some other mechanism such as recurrent mutation or gene conversion is resulting in homogenization of this haplotype. A similar observation has been reported by Ackerman and colleagues (Ackerman *et al*., 2003), who analyzed haplotypic diversity of the TNF locus in two human African populations (Gambian and Malawian). They found extensive haplotypic diversity in both populations with poor correlation between SNPs and attributed this to a 'rich history of intragenic recombination'. There were also significant differences in the haplotype distributions of both populations with implications for disease gene mapping. Brown and colleagues (Brown *et al*., 2006) studied haplotypes comprising 5 SNPs in the promoter region of the TNFα locus in several equidae species and observed nine haplotypes.

These latter two studies are consistent with the results reported above for the sheep TNFα locus. Further studies will be required to identify the mechanisms responsible for the high degree of intragenic heterogeneity at this locus. In the sheep study we were looking for evidence of haplotypic blocks across several loci and for this reason the main part of this study was undertaken using only one of the three SNPs within the TNFα locus typed in the sheep panel.

In summary, genetic diversity within the central region of the sheep MHC has been investigated. Haplotypes have been identified from pedigrees and evidence obtained for conserved subregions characterized by elevated levels of linkage disequilibrium similar to those reported for the analogous region of the human MHC.

CHAPTER 6

CONCLUDING COMMENTS AND DISCUSSION

The main thrust of this study has been to map the central region of the sheep MHC and to then discover single nucleotide polymorphisms as a first step to identification of haplotypes containing candidate loci associated with productivity traits. The potential applications of the project are predicated therefore on two hypotheses. Firstly, that the sheep and human MHC central regions will share conserved characteristic loci. Secondly, that a group of adjacent loci within the central region of the sheep MHC will define a genomic region characterized by restricted recombination which will permit the identification of a small number of relatively common haplotypes within a sheep population.

The objectives of the project, as stated in Chapter 1, were as follows.

a applying a new strategy for generating DNA probes to identify genes from clones of sheep cosmid and BAC libraries containing central region loci of the sheep MHC.

b generate a physical map of the central region containing identified loci,

c identify SNPs and/or polymorphic STR loci by inspection of amplified sequences from a panel of individual sheep, and

d to use the panel of polymorphic markers so obtained to identify central region haplotypes in the sheep MHC.

In general terms, all these objectives have been met. A physical map identifying 25 loci within the central region of the sheep MHC was obtained. This work entailed sequencing and orientating cosmid and BAC clones containing a library of sheep genomic DNA. This was followed by complete sequencing of sheep TNFα and complement factor B (CFB) genes which were published in GenBank. Comparative analyses of these sequences were undertaken and their homology

with other mammalian species established. Finally, a panel of polymorphic markers was obtained which spanned much of the central region of the sheep MHC. Genotyping of a panel of 12 single nucleotide polymorphisms from 10 defined loci in a confirmed half sibling pedigree comprising 25 sheep permitted the deduction of individual haplotypes spanning some 600 kbp of the central region of the sheep MHC. In addition, partial genomic sequences were obtained for the complement C2 and for complement C4 genes and the homology of these sequences with their human counterparts was established. It is of interest that the polymorphic short tandem repeat locus, referred to as BfMs, previously discovered in the Curtin laboratory and believed to be within the complement factor B gene was mapped to an intron at the 5' end of the SKIV2L gene. All sequences were published in GenBank.

Prior to this study very little detailed sequence of the sheep MHC central region was known and during this project approximately 70 kbp of high quality sequence was generated and published in GenBank. The final physical map of the central region of the sheep MHC covered approximately 700 kbp and displayed remarkable synteny with its human counterpart. BAC clones containing MHC class I and class II loci were also identified in this work and these defined the physical limits of the central region. The physical map of the sheep MHC central region reported in this thesis corrects errors reported by Liu and colleagues (Liu *et al.*, 2006) and Wu and colleagues (Wu *et al.*, 2008) and is to the best of my knowledge currently the most accurate map for this region of the sheep genome.

Subcloning of the BAC clones permitted the identification of a CYP21 locus and a complement C4 locus between the SKIV2L gene and the TNXB gene however characterization of the intervening region was not achieved. A distinct cognate study from the Curtin laboratory (not undertaken by me) did however permit more accurate mapping of this sub-region and, while confirming the map obtained from the BAC clones, also identified an inversion of the CYP21 and C4 loci relative to the corresponding region in the human MHC. The final map of the sheep MHC central region has been published and a copy of this paper is included as Appendix 1.

The approach taken in this study to identify loci in the central region of the sheep MHC is based on assuming synteny with the central region of the human MHC. While other studies of the corresponding regions of other mammalian species

made this a reasonable assumption, it was always possible that the sheep central region could differ appreciably from its human counterpart perhaps as a result of chromosomal rearrangements which are known to have occurred during vertebrate evolution. However this difficulty did not arise and the synteny between sheep and human central regions ensured the experimental approach used was a very efficient way of generating a map of this previously poorly characterised region of the sheep chromosome. Further studies based on more intensive sequence analysis may of course reveal species specific, or even breed specific, differences with the corresponding region in other mammals. The localized inversion of the CYP21 and complement C4 loci is an example of such a difference.

The genomic sequence of the sheep CFB gene was approximately 5500 bp long and was very close (≈600bp) to its paralogous complement C2 gene. The exon/intron structure of sheep CFB was similar to human CFB and contained three short consensus repeat domains, characteristic of a number of complement loci and in particular CFB. A von Willbrand domain and a serine protease domain were also present. These observations provide conclusive evidence for its homology to human CFB. Comparison of the sheep CFB sequence with the human CFB sequences did not show any mutations which would preclude expression of the sheep gene. The sheep CFB sequence permitted a homologous cattle CFB genomic sequence in GenBank to be identified and analysed. A sequence error was present in this GenBank file which when corrected resulted in an excellent alignment with sheep CFB. Both sheep and cattle CFB sequences contained identical indels relative to the human CFB sequence and these may be a sequence signature for this class of mammal.

A complete genomic sequence for sheep TNFα was obtained and an extensive comparative analysis of this gene undertaken. As expected, sheep TNFα was similar in structure to its human orthologue. However both sheep and cattle TNFα proteins, predicted from the genomic sequences, have an additional glutamine at position 63 relative to human TNFα (234 amino acids v 233 amino acids). Analysis of the multiple sequence alignment for this locus revealed this variation was due to a mutation in the 3' splice site of the first intron. With the advantage of hindsight, a previously reported error for cattle TNFα in GenBank can now be understood. Interestingly, the TNFα loci of mouse and rat have an additional indel resulting in predicted and actual proteins with 235 amino acids.

It is unfortunate that only partial sequences comprising approximately 50% of the sheep complement C2 and C4 loci were obtained due to lack of time and resources to complete additional sequencing. However, the partial sequences obtained showed that they shared similar genomic structures and predicted amino acid sequences to their human orthologues. The predicted amino acid sequence of the sheep C4 locus showed it was of the C4B isotype.

Examination of a cattle complement C2 genomic sequence present in GenBank did not contain the long retroviral insert present in the third intron of human C2 which results in an unusually long genomic sequence for this gene. It is not known whether an analogous retroviral insert is present in sheep C2 but its absence in the cattle C2 gene suggests this is unlikely. The multiple sequence alignment of these C2 genes and the putative translated amino acid sequences of their C2 proteins predicts that the C2 proteins of sheep and cattle will be two amino acids shorter than human C2. This is due to the presence in both species of two non contiguous deletions in (putative) exon 18.

The panel of SNP markers were genotyped using pyrosequencing technology. This was applied to two populations of unrelated sheep – a cohort comprising 58 merino sheep from the Rylington Merino Project and a separate large multisire sheep pedigree comprising 25 full and half sib sheep of which 13 parental animals were not closely related. Both populations are managed by the WA Department of Agriculture. No mutations or recombinations were observed in this family group. Haplotypes and their frequencies were estimated in the two test flocks using the SNPstats software which is based on a maximum likelihood method. One of the main reasons this study was undertaken was to demonstrate the feasibility of this approach to the identification of haplotypes containing putative loci associated with disease and/or productivity traits. It was observed that there was significant linkage disequilibrium between adjacent SNPs in the combined group of 71 unrelated sheep. However the linkage disequilibrium appeared clustered into a main central block with subregions of low LD close to each end of the central region. This result provides preliminary evidence for a block like structure of loci in the sheep MHC central region as has been reported in several previous studies and in particular in the human MHC.

It would seem reasonable to hypothesize therefore that the entire sheep MHC is comprised of a series of adjacent blocks each characterised by a low rate of meiotic recombination and which may be identified by high levels of linkage

disequilibrium. From the analyses performed in this project it is clear that estimating linkage disequilibrium from SNPs is very sensitive to the quality and especially the quantity of the data available. To obtain reliable estimates of the linkage disequilibrium parameters large numbers of animals need to be included in the genotyping panels. Furthermore, predicting haplotypes by analysis of unrelated animal genotypes using statistical methods such as the expectation maximization algorithm seems to generate large numbers of haplotypes only some of which are actually observed in known haplotypes panels deduced from pedigree analysis. This may of course also be very dependent on the sample size analysed.

Characterisation of the sheep MHC is an important prelude to further understanding of the adaptive immune response in this economically important species, together with the mapping of candidate genes associated with parasite immunity and productivity traits. The identification of sheep MHC haplotypes will provide an important tool for further studies focussed on applications of marker assisted selection for disease resistance and productivity. Such studies in sheep will be relatively more difficult than equivalent studies in humans due to the peculiarities in the way sheep are bred. This is essentially a form of line breeding which entails non random mating. Consequently, while not inbred, sheep are more likely to be related than for example humans. One way of addressing this problem would be to compare different breeds of sheep, or flocks from geographically different locations which could be expected to also differ genetically.

This project has served to highlight the amount of research that remains to be completed before the sheep MHC is fully characterized. Only 25 loci were identified in the central region of the sheep MHC although it is clear that many more exist in the ≈700 kbp bounded by the markers characteristic of the class I and class II regions. The small number of SNPs discovered and animals genotyped have limited the identification of block like regions. Further definition of these regions will require much more dense panels of SNPs and is one of the more important targets for further research since their discovery will facilitate more efficient identification of haplotypic blocks with their potential candidate genes associated with important productivity traits. There is no reason to believe that the sheep MHC will not become as amenable to haplotypic block analysis as is currently achieved for the human MHC and its many disease associations. Future work should also include close examination of MHC associated breed

variation especially in those instances where one breed manifests a particularly desirable (or undesirable) trait such as enhanced resistance to parasites. Comparative studies of the MHC in many species, including sheep, will permit a clearer understanding of the function of a number of genes and the selective forces to which they are subjected. In the longer term, the evolutionary history of this fascinating chromosomal region will be elucidated.

In summary, this study has resulted in an improved map of the central region of the sheep MHC, the detailed genomic characterisation of two significant genes and a panel of polymorphic markers which permitted MHC haplotypes to be identified. These results will contribute to further genetic studies in sheep aimed at improving the productivity and management of this economically important species.

BIBLIOGRAPHY

Ackerman, H., Usen, S., Mott, R., Richardson, A., Sisay-Joof, F., Katundu, P., Taylor, T., Ward, R., Molyneux, M., Pinder, M. & Kwiatkowski, D. P. (2003). Haplotypic analysis of the TNF locus by association efficiency and entropy. *Genome Biology*, **4**:R24.

Alvarez-Busto, J., Ruiz-Nunez, A., Mazon, L. I. & Jugo, B. M. (2004). Detection of polymorphisms in the tumour necrosis factor alpha candidate gene in sheep. *European Journal of Immunogenetics*, **31**, 155-158.

Andersson, L. (1988). Organisation of the bovine MHC class II region as revealed by genomic hybridizations. *Animal Genetics*, **19**, 32-34.

Andersson, L. & Rask, L. (1988). Characterization of the MHC class II region in cattle: The number of DQ genes varies between haplotypes. *Immunogenetics*, **27**, 110-120.

Andersson, L., Lunden, A., Sigurdardottir, S., Davies, C. J. & Rask, L. (1988). Linkage relationships in the bovine MHC region. High recombination frequency between class II subregions. *Immunogenetics***, 27**, 273-280.

Beck, S. & Trowsdale, J. (2000). The human major histocompatibility complex: lessons from the DNA sequence. *Annual Reviews,* **01**, 117-137.

Behnke, J. M., Iraqi, F., Menge, D., Baker, R. L., Gibson, J. & Wakelin, D. (2003). Chasing the genes that control resistance to gastrointestinal nematodes. *Journal of Helminthology*, **77**, 99-110.

Behnke, J. M., Mugambi, J. M., Clifford, S., Iraqi, F. A., Baker, R. L., Gibson, J. P & Wakelin, D. (2006). Genetic variation in resistance to repeated infections with *Heligmosomoides polygyrus bakeri*, in inbred mouse strains selected for the mouse genome project. *Parasite Immunology*, **28**, 85-94.

Besier, R. B. (2004). New approaches to sheep parasite control - the potential for individual sheep management. *Animal Production in Australia,* **1**, 13-16.

Bernatchez, L. & Landry, C. (2003). MHC studies in nonmodel vertebrates: what have we learned about natural selection in 15 years? *Journal of Evolutionary Biology*, **16**, 363-377.

Beutler, B. & Cerami, A. (1989). The biology of cachectin/TNF--a primary mediator of the host response. *Annual Review of Immunology,* **7**, 625-655.

Blattman, A. N. & Beh, K. J. (1992). Dinucleotide repeat polymorphism within the ovine major histocompatibility complex. *Animal Genetics*, **23**, 392-392.

Bot, J., Karlsson, L. J. E., Greef, J. & Witt, C. (2004). Association of the MHC with production traits in Merino ewes. *Livestock Production Science,* **86**, 85-91.

Bowness, P. (2002). HLA B27 in health and disease: a double-edged sword? *Rheumatology*, **41**, 857 - 868.

Briles, W. E., Stone, H. A. & Cole, R. K. (1977). Marek's disease: effects of B histocompatibility alloalleles in resistant and susceptible chicken lines. *Science*, **195**, 193-195.

Brown, J. J., Ollier, W. E. R., Thomson, W., Matthews, J. B., Carter, S. D., Binns, M., Pinchbech G. & Clegg, P. D. (2006). TNF-a SNP haplotype frequencies in equidae. *Tissue Antigens*, **67**, 377-382.

Charbonnel, N. & Pemberton, J. (2005). A long-term genetic survey of an ungulate population reveals balancing selection acting on MHC through spatial and temporal fluctuations in selection. *Heredity*, **95**, 377-388.

Coop, G. & Myers, S. R. (2007). Live hot, die young: Transmission distortion in recombination hotspots. *PLoS Genetics,* **3**, e35.

Daly, M. J., Rioux, J. D., Schaffner, S. F., Hudson, T. J. & Lander, E. S. (2001). High-resolution haplotype structure in the human genome. *Nature Genetics*, **29**, 229-232.

Dalrymple, B. P., Kirkness, E. F., Nefedov, M., McWilliam, S., Ratnakumar, A., Barris, W., Zhao, S., Shetty, J., Maddox, J. F., O'Grady, M., Nicholas, F., Crawford, A., Smith, T., de Jong, P. J., McEwan, J., Hutton, O. V., Cockett, N. E. & International Sheep Genomics Consortium. (2007). Using comparative genomics to reorder the human genome sequence into a virtual sheep genome. *Genome Biology*, **8**, R152.

Dangel, A. W., Shen, L., Mendoza, A. R., Wu, L. & Yu, C. Y. (1995). Human helicase gene SKI2W in the HLA class III region exhibits striking structural similarities to the yeast antiviral gene SKI2 and to the human gene KIAA0052: emergence of a new gene family. *Nucleic Acids Research*, **23**, 2120-2126.

Dawkins, R. L., Christiansen, F. T., Kay, P. H., Garlepp, M., McCluskey, J., Hollingsworth, P. N. & Zilko, P. J. (1983). Disease associations with complotypes, supratypes and haplotypes. *Immunogical Reviews*, **70**, 1-22.

Dawkins, R., Leelayuwat, C., Gaudieri, S., Tay, G., Hui, J., Cattley, S., Martinez, P. & Kulski, J. (1999). Genomics of the major histocompatibility complex: haplotypes, duplication, retroviruses and disease. *Immunological Review*, **167**, 275-304.

Dineen, J. K., Donald, A. D., Wagland, B. M. & Turner, J. H. (1965). The dynamics of the host-parasite relationship. II. The response of sheep to primary and secondary infection with *Nematodirus spathiger*. *Parasitology*, **55**, 163-171.

Drenkard, E., Richter, B. G., Rozen, S., Stutius, L. M., Angell, N. A., Mindrinos, M., Cho, R. J., Oefner, P. J., Davis, R. W. & Ausubel, F. M. (2000). A simple procedure for the analysis of single nucleotide polymorphisms facilitates map-based cloning in Arabidopsis. *Plant Physiology,* **124**, 1483-1492.

Dressel, R., Walter, L. & Gunther, E. (2001). Genomic and functional aspects of the rat MHC, the RT1 complex. *Immunological Reviews,* **184**, 82-95.

Dukkipati, V. S., Blair, H. T., Garrick, D. J. & Murray, A. (2006). Ovar-Mhc-ovine major histocompatibility complex: role in genetic resistance to dieseases. *New Zealand Veterinary Journal,* **54**, 153-160.

Estler, H. C., Grewe, M., Gaussling, R., Pavlovic, M. & Decher, K. (1992). Rat tumor necrosis factor-alpha. Transcription in rat Kupffer cells and in vitro posttranslational processing based on a PCR-derived cDNA. *Biological chemistry Hoppe-Seyler,* **373**, 271-281.

Gabriel, S. B., Schaffner, S. F., Nguyen, H., Moore, J. M., Roy, J., Blumenstiel, B., Higgins, J., Defelice, M., Lochner, A., Faggart, M., Liu-Cordero, S. N., Rotimi, C., Adeyemo, A., Cooper, R., Ward, R., Lander, E. S., Daly, M. J. & Altshuler, D. (2002). The structure of haplotype blocks in the human genome. *Science,* **296**, 2225-2229.

Gelhaus, A., Hess, M., Forster, B., Goldammer, T., Schwerin, M. & Horstmann, R. D. (2006). YAC/BAC contig spanning the MHC class III region of cattle. *Cytogenetic and Genome Research,* **115**, 45-50.

Glass, E. J., Oliver, R. A. & Russell, G. C. (2000). Duplicated DQ haplotypes increase the complexity of restriction element usage in cattle. *The Journal of Immunology,* **165**, 134-138.

Geraghty, D. E. (2002). Genetic diversity and genomics of the immune response. *Immunological Review*, **190**, 5-8.

Gold, B., Merriam, J. E., Zernant, J., Hancox, L. S., Taiber, A. J., Gehrs, K., Cramer, K., Neel, J., Bergeron, J., Barile, G. R., Smith, R.T., AMD Genetics Clinical Study Group, Hageman, G. S., Dean, M. & Allikmets, R. (2006). Variation in factor B (BF) and complement component 2 (C2) genes is associated with age-related macular degeneration. *Nature Genetics,* **38**, 458-462.

Goldstein, D. B. (2001). Islands of linkage disequilibrium. *Nature Genetics*, **29**, 109-111.

Groth, D. & Wetherall, J. (1994). Dinucleotide repeat polymorphism within the ovine major histocompatibility complex class-I region. *Animal Genetics,* **25**, 61.

Groth, D. M. & Wetherall, J. D. (1995). Dinucleotide repeat polymorphism associated with the sheep Factor B complement gene. *Animal Genetics*, **26**, 282-283.

Gustafson, A. L., Tallmadge, R. L., Ramlachan, N., Miller, D., Bird, H., Antczak, D. F., Raudsepp, T., Chowdhary, B. P. & Skow, L. C. (2003). An ordered BAC contig map of the equine major histocompatibility complex. *Cytogenetic and Genome Research*, **102**, 189-195.

Gupta, P. K., Bind, R. B., Walunj, S. S. & Saini. M. (2004). High nucleotide and amino acid sequence similarities in tumor necrosis factor-a amongst Indian buffalo (*Bubalus Bubalis*), Indian cattle (*Bos Indicus*) and other ruminants. *European Journal of Immunogenetics*, **31**, 189-193.

Gutierrez-Espeleta, G. A., Hedrick, P. W., Kalinowski, S. T., Garrigan, D. & Boyce, W. M. (2001). Is the decline of desert bighorn sheep from infectious disease the result of low MHC variation? *Heredity*, **86**, 439-450.

Hanchard, N., Rockett, K., Udalova, I., Wilson, J., Keating, B., Koch, O., Nijnik, A., Diakite, M., Herbert, M. & Kwiatkowski, D. (2006). An investigation of transmission ratio distortion in the central region of the human MHC. *Genes and Immunity,* **7**, 51-58.

Hedrick, P. W. & Thomson, G. (1983). Evidence for balancing selection at HLA. *Genetics,* **104**, 449-456.

Hedrick, P. W., Gutierrez-espeleta, G. A. & Lee, R. N. (2001). Founder effect in an island population of bighorn sheep. *Molecular Ecology*, **10**, 851–857.

Hoekstra, R., Borgsteede, F. H., Boersema, J. H. & Roos, M. H. (1997). Selection for high levamisole resistance in *Haemonchus contortus* monitored with an egg-hatch assay. *International Journal for Parasitology*, **27**, 1395-1400.

Hohenhaus, M. A. & Outteridge P. M. (1995). The immunogenetics of resistance to *Trichostrongylus colubriformis* and *Haemonchus contortus* parasites in sheep. *British Veterinary Journal,* **151**, 119-140.

Holmes, E. C., Roberts, A. F., Staines, K. A. & Ellis, S. A. (2003). Evolution of major histocompatibility complex class I genes in cetartiodactyls. *Immunogenetics,* **55**, 193-202.

Horton, R., Wilming, L., Rand, V., Overing, R. C., Bruford, E. A., Khodiyar, V. K., Lush. M. J., Povey, S., Jr, C. Conover T., Wright, M. W., Wain, H. M., Trowsdale, J., Ziegler, A. & Beck, S. (2004). Gene map of the extended human MHC. *Nature Reviews,* **5**, 889-899.

Horton R., Gibson R., Coggill P., Miretti M., Allcock R. J., Almeida J., Forbes S., Gilbert J. G., Halls K., Harrow J. L., Hart E., Howe K., Jackson D. K., Palmer S., Roberts A. N., Sims S., Stewart C. A., Traherne J. A., Trevanion S., Wilming L., Rogers J., de Jong P. J., Elliott J. F., Sawcer S., Todd J. A., Trowsdale J. & Beck S. (2008). Variation analysis and gene annotation of eight MHC haplotypes: The MHC Haplotype Project. *Immunogenetics*, **60**, 1-18.

Hughes, A. L. (2002). Natural selection and the diversification of vertebrate immune effectors. *Immunological Review*, **190**, 161-168.

Jeffreys, A. J., Kauppi, L. & Neumann, R. (2001). Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex. *Nature Genetics,* **29**, 217-222.

Jeffreys, A. J. & Neumann, R. (2002). Reciprocal crossover asymmetry and meiotic drive in a human recombination hot spot. *Nature genetics*, **31**, 267-271.

Johnson, G. C. & Todd J. A. (2000). Strategies in complex disease mapping. *Current Opinion in Genetics & Development*, **10**, 330-334.

Jugo, B. M. & Vicario A. (2001). Lymphocyte antigens in sheep: linkage to the MHC class II DRB1 gene. *European Journal of Immunogenetics,* **28**, 451-458.

Kaminski, S., Rusc, A. & Malewski, T. (2004). Single nucleotide polymorphism database of candidate genes associated with cow milk protein biosynthesis. *Journal of Animal and Feed Science*, **13**, 53-66.

Kaminski, S., Brym, P., Rusc, A., Wojcik, E., Ahman, A. & Magi, R. (2006). Associations between milk performance traits in Holstein cows and 16 candidate SNPs identified by arrayed primer extension (Apex) microarray. *Animal Biotechnology,* **17**, 1-11.

Karlsson, L. J. E. (1994). Selecting sheep for enhanced resistance to gastro-intestinal parasites (Final No. Project DAW 100). Department of Agriculture.

Kelley, J., Walter, L. & Trowsdale, J. (2005). Comparative genomics of major histocompatibility complexes. *Immunogenetics,* **56**, 683-695.

Klein, J., Benoist, C., David, C. S., Demant, P., Lindahl, K. F., Flaherty, L., Flavell, R. A., Hämmerling, U., Hood, L. E., Hunt, S. W., Jones, P. P., Kourilsky, P., McDevitt, H. O., Meruelo, D., Murphy, D. B., Nathenson, S. G., Sachs, D. H., Steinmetz, M., Tonegawa, S., Wakeland, E. K. & Weiss, E. H. (1990). Revised nomenclature of mouse H-2 genes. *Immunogenetics*, **32**, 147-149.

Klein, J. (2004). Did viruses play a part in the origin of the adaptive immune system? *Folia Biologica (Praha)*, **50**, 87-92.

Kulski, J. K., Shiina, T., Anzai, T., Kohara, S. & Inoko, H. (2002). Comparative genomic analysis of the MHC: the evolution of class I duplication blocks, diversity and complexity from shark to man. *Immunological Review*, **190**, 95-122.

Kumanovics, A., Takada, T. & Lindahl, K. F. (2003). Genomic organization of the mammalian MHC. *Annual Reviews Immunologicals*, **21**, 629-657.

Kwok, P. Y. & Gu, Z. (1999). Single nucleotide polymorphism libraries: why and how are we building them?. *Molecular Medicine Today,* **5**, 538-543.

Lazarus, R., Vercelli, D., Palmer, L. J., Klimecki, W. J., Silverman, E. K., Richter, B., Riva, A., Ramoni, M., Martinez, F. D., Weiss, S. T. & Kwiatkowski, D. J. (2002). Single nucleotide polymorphisms in innate immunity genes: abundant variation and potential role in complex human disease. *Immunological Review,* **190**, 9-25.

Lie, B. A. & Thorsby E. (2005). Several genes in the extended human MHC contribute to predisposition to autoimmune diseases. *Current Opinion in Immunology,* **17**, 526-531.

Liu H., Liu K., Wang J. & Ma R. Z. (2006). A BAC clone-based physical map of ovine major histocompatibility complex. *Genomics*, **88**, 88-95.

Marck, C. (1988). 'DNA Strider': A 'C' program for the fast analysis of DNA and protein sequences on the Apple Macintosh family of computers. *Nucleic Acids Research*, **16**, 1829-1836.

Martinsohn J. T., Sousa A. B., Guethlein L. A. & Howard J. C. (1999). The gene conversion hypothesis of MHC evolution: a review. *Immunogenetics*, **50**, 168-200.

Matsuzaka, Y., Makino, S., Nakajima, K., Tomizawa, M., Oka, A., Bahram, S., Kulski, J. K., Tamiya, G. & Inoko, H. (2001). New polymorphic microsatellite markers in the human MHC Class III Region. *Tissue Antigens*, **57**, 397-404

Matsumoto, M., Fukuda, W., Circolo, A., Goellner, J., Strauss-Schoenberger, J., Wang, X., Fujita, S., Hidvegi, T., Chaplin, D. D. & Colten, H. R. (1997). Abrogation of the alternative complement pathway by targeted deletion of murine factor B. *Proceedings of the National Academy of Sciences of the United States of America,* **94**, 8720-8725.

McLure, C. A., Williamson, J. F., Stewart, B. J., Keating, P. J. & Dawkins, R. L. (2004). Genomic analysis reveals a duplication of eight rather than seven short consensus repeats in primate CR1 and CR1L: evidence for an additional set shared between CR1 and CR2. *Immunogenetics*, **56**, 631-638.

McLure, C. A., Dawkins, R. L., Williamson, J. F., Davies, R. A., Berry, J., Natalie, L. J., Laird, R. & Gaudieri, S. (2004). Amino acid patterns within short consensus repeats define conserved duplicons shared by genes of the RCA complex. *Journal of Molecular Evolution*, **59**, 143-157.

McRae, A. F., McEwan, J. C., Dodds, K. G., Wilson, T., Crawford, A. M. & Slate, J. (2002). Linkage disequilibrium in domestic sheep. *Genetics*, **160**, 1113-1122.

McVean, G. A. T., Myers, S. R., Hunt, S., Deloukas, P., Bentley, D. R. & Donnelly, P. (2004). The fine-scale structure of recombination rate variation in the human genome. *Science*, **304**, 581–584.

MHC Sequencing Consortium (1999). Complete sequence and gene map of a human major histocompatibility complex. *Nature,* **401**, 921-923.

Millot, P. (1978). The major histocompatibility complex of sheep (OLA) and two minor loci. *Animal Blood Groups and Biochemical Genetics,* **9**, 115-121.

Milner, C. M. & Campbell R. D. (2001). Genetic organization of the human MHC class III region. *Frontiers in bioscience : a journal and virtual library,* **6**, D914-926.

Meyer, D. & Thomson, G. (2001). How selection shapes variation of the human major histocompatibility complex: a review. *Annals of Human Genetics*, **65**, 1-26.

Moore, L. A., Tidyman, W. E., Arrizubieta, M. J. & Bandman, E. (1992). Gene conversions within the skeletal myosin multigene family. *Journal of Molecular Biology,* **223**, 383-387.

Nash, A. D., Barcham, G. J., Brandon, M. R. & Andrews, A. E. (1991) Molecular cloning, expression and characterization of ovine TNF alpha. *Immunology and Cell Biology,* **69**, 273-283.

Odbileg, R., Konnai, S., Ohashi, K. & Onuma, M. (2005). Molecular cloning and phylogenetic analysis of inflammatory cytokines of Camelidae (Llama and Camel). *Journal of Veterinary Medical Science,* **67**, 921-925.

Outteridge, P. M., Windon, R. G. & Dineen, J. K. (1988). An ovine lymphocyte antigen marker for acquired resistance to *Trichostrongylus colubriformis*. *International Journal for Parasitology*, **18**, 853-858.

Paterson, S., Wilson, K. & Pemberton, J. M. (1998). Major histocompatibility complex variation associated with juvenile survival and parasite resistance in a large unmanaged ungulate population (*Ovis aries* L.). *The National Academy of Sciences*, **95**, 3714-3719.

Pekna, M., Hietala, M. A., Landin, A., Nilsson, A. K., Lagerberg, C., Betsholtz, C. & Pekny, M. (1998). Mice deficient for the complement factor B develop and reproduce normally. *Scandinavian Journal of Immunology,* **47**, 375-380.

Porter, R. R. (1985). The complement components coded in the major histocompatibility complexes and their biological activities. *Immunological Reviews*, **87**, 7-17.

Raymond, M. & Rousset, F. (1995). GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *The Journal of Heredity*, **86**, 248-249.

Ren, X. D., Dodds, A. W. & Law, S. K. (1993). The thioester and isotypic sites of complement component C4 in sheep and cattle. *Immunogenetics*, **37**, 120-128.

Roos, M. H. (1997). The role of drugs in the control of parasitic nematode infections: must we do without? *Parasitology*, **114 Suppl**, S137-144.

Ruuls, S. R. & Sedgwick, J. D. (1999). Unlinking tumor necrosis factor biology from the major histocompatibility complex: lessons from human genetics and animal models. *American Journal of Human Genetics,* **65**, 294-301.

Santucci, F., Ibrahim, K. M., Bruzzone, A. & Hewit, G. M. (2007). Selection on MHC-linked microsatellite loci in sheep populations. *Heredity*, **99**, 340-348.

Schwaiger, F. W., Gostomski, D., Stear, M. J., Duncan, J. L., McKellar, Q. A., Epplen, J. T. & Buitkamp, J. (1995). An ovine major histocompatibility complex DRB1 allele is associated with low faecal egg counts following natural, predominantly Ostertagia circumcincta infection. *International Journal for Parasitology,* **25**, 815-822.

Schwaiger, F. W., Maddox, J., Ballingall, K., Buitkamp, J., Crawford, A. M., Dutia, B. M., Epplen, J. T., Ferguson, E. D., Groth, D., Hopkins, J., Rhind, S. M., Sargan, D., Wetherall, J. & Wright, H. (1996). The ovine major histocompatibility complex. In L. B. Schook & S. J. Lamont (Eds.), *The major histocompatibility complex region of domestic animal species* (pp.121-176). Boca Raton: CRC press.

Schwensow, N., Fietz, J., Dausmann, K. H. & Sommer, S. (2007). Neutral versus adaptive genetic variation in parasite resistance: importance of major histocompatibility complex supertypes in a free-ranging primate. *Heredity*, **99**, 265-277.

Schook, L. B. & Lamont, S. J. (1996). *The major histocompatibility complex region of domestic animal species*. Boca Raton: CRC press.

Semon, D., Kawashima, E., Jongeneel, C. V., Shakhov, A. N. & Nedospasov, S. A. (1987). Nucleotide sequence of the murine TNF locus, including the TNF-a (tumor necrosis factor) and TNF-b (lymphotoxin) genes. *Nucleic Acids Research*, **15**, 21.

Shiina, T., Tamiya, G., Oka, A., Takishima, N. & Inoko, H. (1999). Genome sequencing analysis of the 1.8Mb entire human MHC class I region. *Immunological Reviews,* **167**, 193-199.

Sinnwell, J. P. & Schaid, D. J. (2005). haplo.stats: Statistical Analysis of Haplotypes with Traits and Covariates when Linkage Phase is Ambiguous. R package version 1.2.2 http://mayoresearch.mayo.edu/mayo/research/biostat/schaid.cfm

Smit, A. F. A., Hubley, R. & Green, P. (1996-2004). RepeatMasker Open-3.0. at http://repeatmasker.org.

Stear, M. J., Innocent, G. T. & Buitkamp, J. (2005). The evolution and maintenance of polymorphism in the major histocompatibility complex. *Veterinary Immunology and Immunopathology,* **108**, 53-57.

Stear, M. J., Newman, M. J., Nicholas, F. W., Brown, S. C. & Holroyd, R. G. (1984). Tick resistance and the major histocompatibility system. *The Australian Journal of Experimental Biology and Medical Science,* **62,** 47-52.

Stenzel, A., Lu, T., Koch, W. A., Hampe, J., Guenther, S. M., De La Vega, F. M., Krawczak, M. & Schreiber, S. (2004). Patterns of linkage disequilibrium in the MHC region on human chromosome 6p. *Human Genetics*, **114**, 377-385.

Su, X. Z., Morris, D. D. & McGraw, R. A. (1991). Cloning and characterization of gene TNF alpha encoding equine tumor necrosis factor alpha. *Gene,* **107**, 319-321.

Taube, C., Thurman, J. M., Takeda, K., Joetham, A., Miyahara, N., Carroll, M. C., Dakhama, A., Giclas, P. C., Holers, V. M. & Gelfand, E. W. (2006). Factor B of the alternative complement pathway regulates development of airway hyperresponsiveness and inflammation. *Proceedings of the National Academy of Sciences of the United States of America,* **103**, 8084-8089.

The Sanger Institute (2006). The MHC haplotype project. http://www.sanger.ac.uk/HGP/Chr6/MHC/. Accessed November 1, 2006.

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. (1997). The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research,* **24**, 4876-4882.

Webb, A. J., Berg, I. L. & Jeffreys, A. (2008). Sperm cross-over activity in regions of the human genome showing extreme breakdown of marker association. *Proceedings of the National Academy of Sciences of the United States of America,* **105**, 10471-10476.

Wetherall, J. D. & Groth, D. M. (1992). The major histocompatibility complex and parasite immunity. In W. K. Yong (Eds.), *Animal Parasite Control Utilizing Biotechnology* (pp. 353-386). Boca Raton: CRC Press.

White, A. M., Yoshimura, T., Smith, A. W., Westwick, J. & Watson. M. L. (1997). Airway inflammation induced by recombinant guinea pig tumor necrosis factor-alpha. *The American Journal of Physiology*, **273**, L524-530.

Whitehead, A. S. & Sackstein, R. (1985). Molecular biology of the human and mouse MHC class III genes: phylogenetic conservation, genetics and regulation of expression. *Immunological Reviews,* **87**, 185-208.

Windon, R. G., Dineen, J. K., Gregg, P., Griffiths, D. A., & Donald, A.D. (1984). The role of thresholds in the response of lambs to vaccination with irradiated *Trichostrongylus colubriformis* larvae. *International Journal for Parasitology*, **14**, 423-428.

Womack, J. E., Johnson, J., Owens, E. K., Rexroad, C. E., III, Schlapfer, J. & Yang, Y. P. (1997). A whole-genome radiation hybrid panel for bovine gene mapping. *Mammalian Genome*, **8**, 854–856.

Wu, C. H., Nomura, K., Goldammer, T., Hadfield, T., Dalrymple, B. P., McWilliam, S., Maddox, J. F., Womack, J. F. & Cockett, N. E. (2008). A high-resolution comparative radiation hybrid map of ovine chromosomal regions that are homologous to human chromosome 6 (HSA6). *Animal Genetics*, **39**, 459-467

Xie, T., Rowen, L., Aguado, B., Ahearn, M. E., Madan, A., Qin, S., Campbell, R. D. & Hood, L. (2003). Analysis of the gene-dense major histocompatibility complex class III region and its comparison to mouse. *Genome Research*, **13**, 2621-2636.

Young, A. J., Hay, J. B. & Chan, J. Y. C. (1990). Primary structure of ovine tumor necrosis factor alpha cDNA. *Nucleic Acids Research,* **18**, 6723.

Yu, C. Y. (1998). Molecular Genetics of the Human MHC Complement Gene Cluster. *Experimental and Clinical Immunogenetics*. **15**, 213-230

Yu, C. Y., Blanchong, C. A., Chung, E. K., Rupert, K. L., Yang, Y., Yang, Z., Zhou, B. & Moulds, J. M. (2002). Molecular genetic analyses of human complement components C4A and C4B. In: Rose NR, Hamilton RG, Detrick B (Eds.), *Manuals of clinical laboratory immunology* (pp 117–131). 6th ed. Washington, DC.: ASM Press.

Yunis, E. J., Larsen, C. E., Fernandez-Vina, M., Awdeh, Z. L., Romero, T., Hansen, J. A. & Alper, C. A. (2003). Inheritable variable sizes of DNA stretches in the human MHC: conserved extended haplotypes and their fragments or blocks. *Tissue Antigens,* **62**, 1-20.

Ziegler, A., Field, L. L. & Sakaguchi, A. Y. (1991). Report of the committee on the genetic constitution of chromosome 6. *Cytogenetics and Cell Genetics,* **58**, 295-336.

List of websites used in this study:

ClustalW2: http://www.ebi.ac.uk/Tools/clustalw2/index.html

GAP: http://genome.cs.mtu.edu/aat.html

GENSCAN: http://genes.mit.edu/GENSCAN.html

NCBI: http://www.ncbi.nlm.nih.gov/

Netprimer: http://www.premierbiosoft.com/netprimer

Nobel Prize in Medicine:
http://nobelprize.org/nobel_prizes/medicine/laureates/index.html

Overgo Maker Program:
http://www.genome.wustl.edu/tools/?overgo.html

Primer 3: http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi

Primers Design for Pyrosequencing:
http://primerdesign.pyrosequencing.com/jsp/TemplateInput.jsp

RepeatMasker: http://www.repeatmasker.org/

The MHC Haplotype Project: http://www.sanger.ac.uk/HGP/Chr6/MHC/

SNPsats: http://bioinfo.iconcologia.net/index.php?module=Snpstats

Twinscan: http://mblab.wustl.edu/query.html

UCSC Genome Bioinformatics: http://genome.ucsc.edu/

APPENDIX 1

PUBLISHED PAPER

A map of the class III region of the sheep major histocompatibility complex

J Qin, C Mamotte, NE Cockett, JD Wetherall and DM Groth

***BMC Genomics* 2008, 9:409**

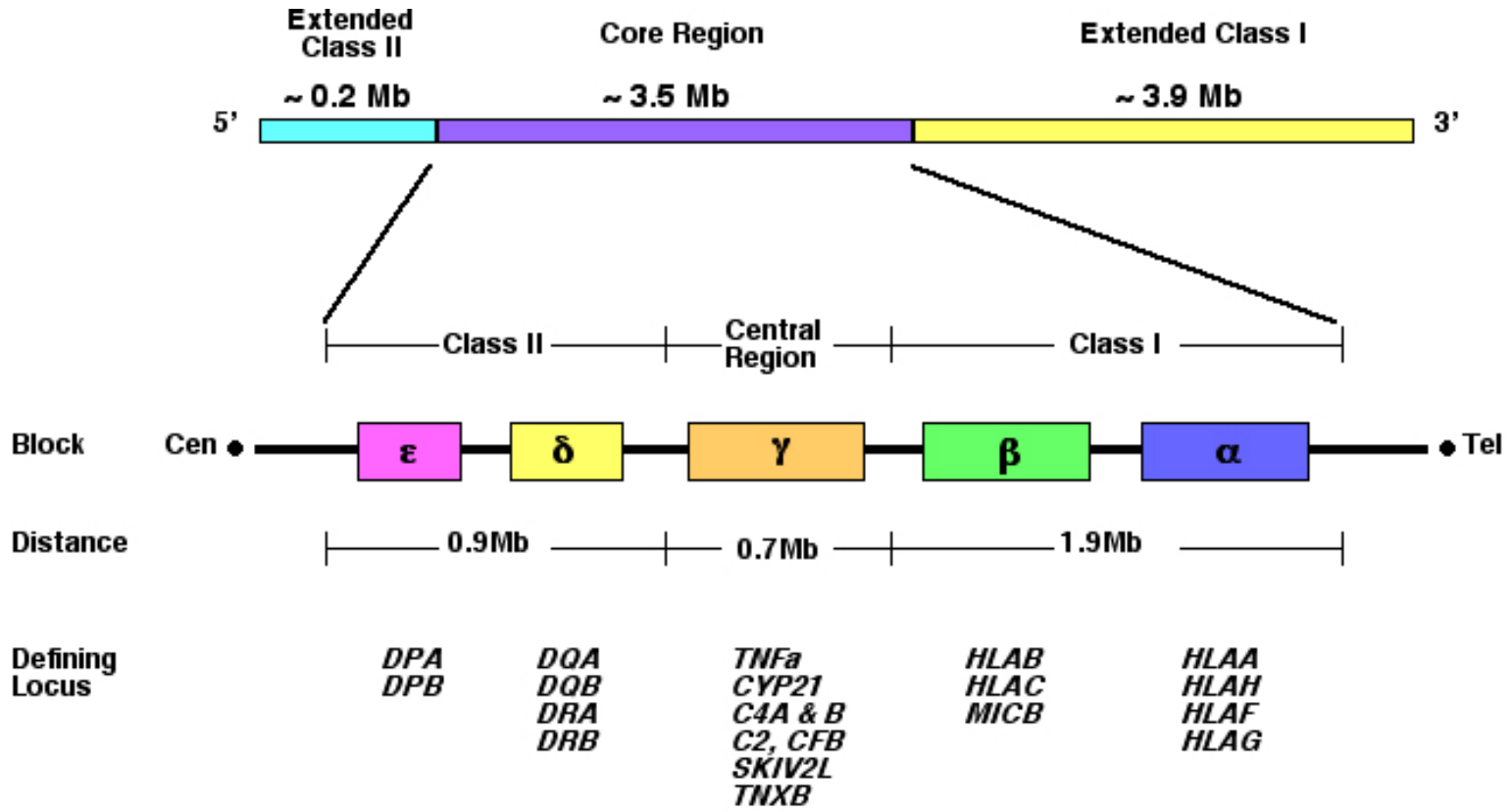


Figure 1.1 *Map of the extended human MHC. This figure shows the main subregions of the MHC together with blocks characterized by reduced recombination and approximate sizes of each block/subregion. The loci identified are those that have been used to characterize each block and subregion.*

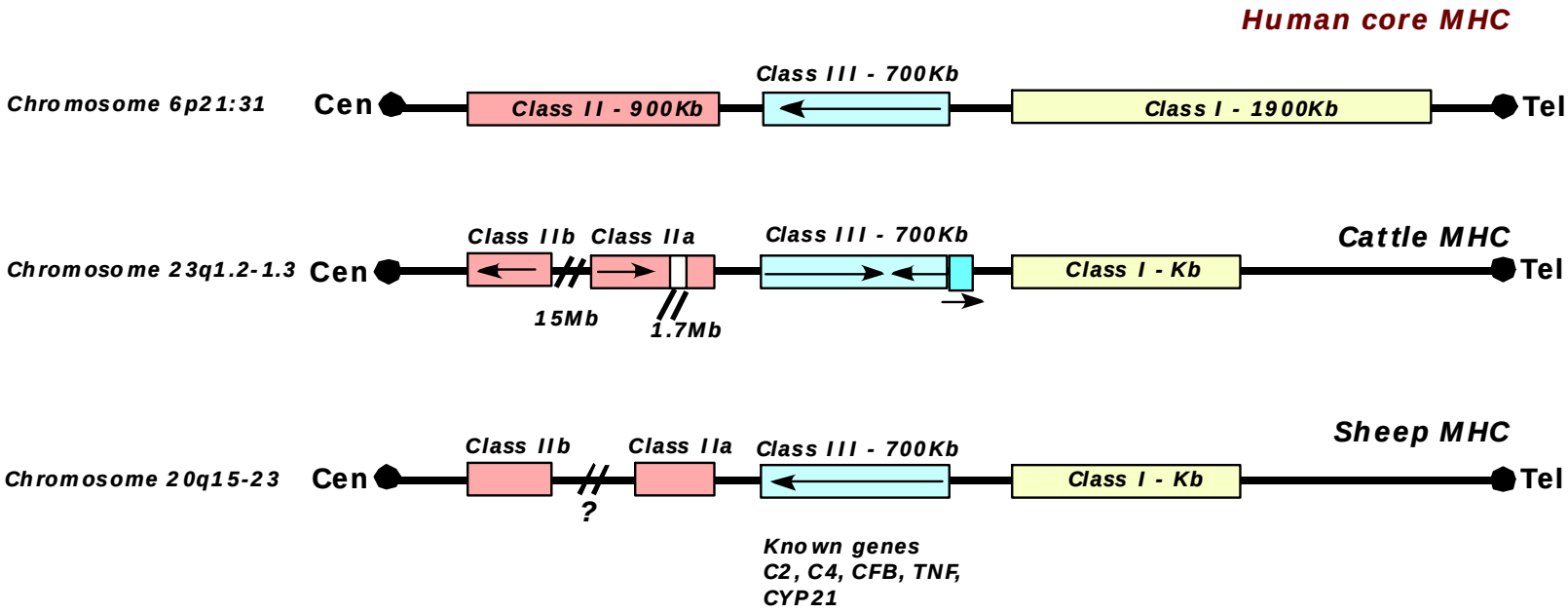


Figure 1.2: GENOMIC STRUCTURES OF CATTLE AND SHEEP MHCs COMPARED WITH THE CORE HUMAN MHC (- 4Mb)

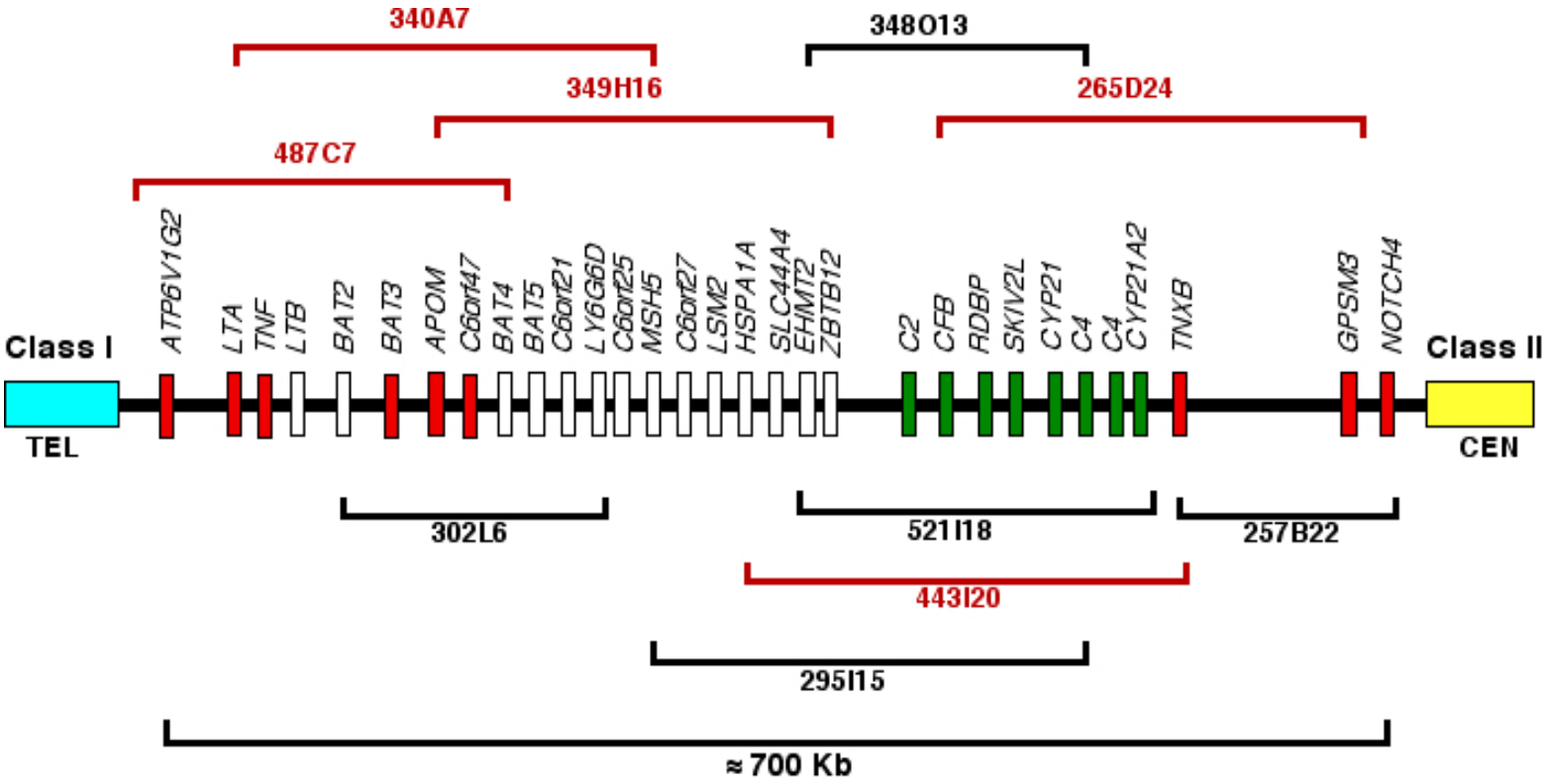


Figure 3.1: *The relative order of 10 BAC clones spanning ≈700Kb containing sheep orthologues of loci mapped to the central region of the human MHC. The presence of the class I and II regions has been assumed from the locus content of the terminal BAC clones. Individual loci identified within these BAC clones are shown as vertical bars. The distance between bars is not to scale. Bars shaded red and green are loci for which their relative positions are known from the sequencing data, whereas the locations of the empty bars have been assigned assuming a similar map to that of the human MHC.*

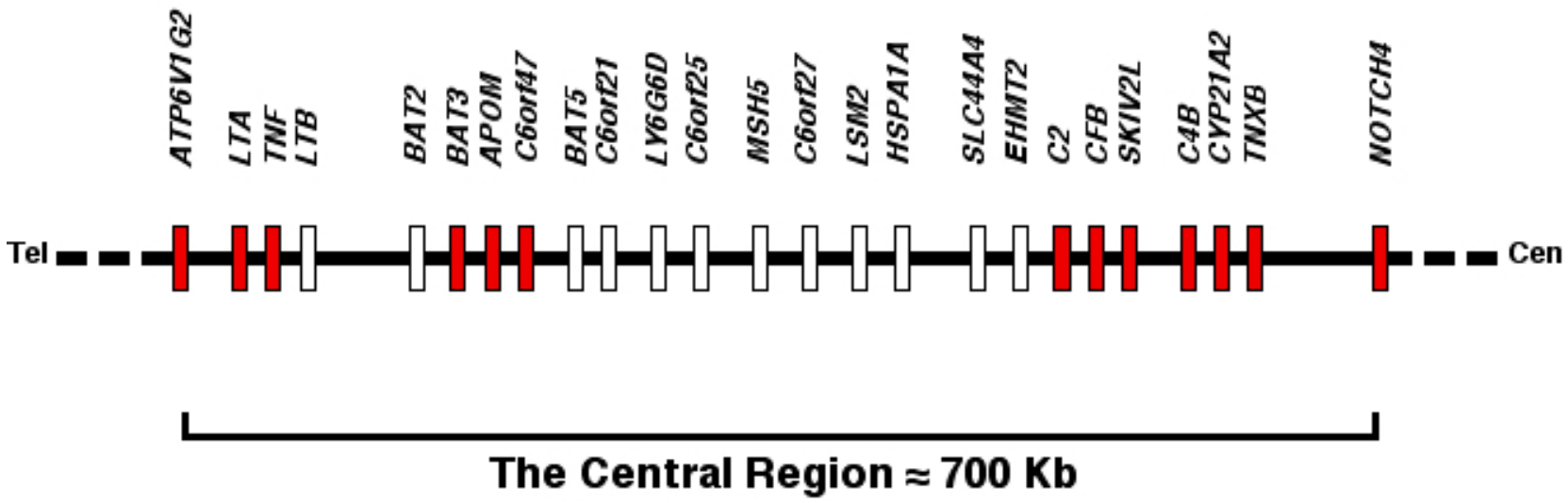


Figure 3.2: *Identification of a further nine loci within the central region of the sheep MHC. The identity of these loci was confirmed by sequence homology with human orthologues. As for Figure 3.1, the red bars are loci whose relative positions are known whereas the empty bars are loci whose position is inferred assuming a similar map to that of the human MHC.*

ALIGNMENT OF GENOMIC AND cDNA FOR HUMAN SHEEP AND CATTLE COMPLEMENT FACTOR B GENE

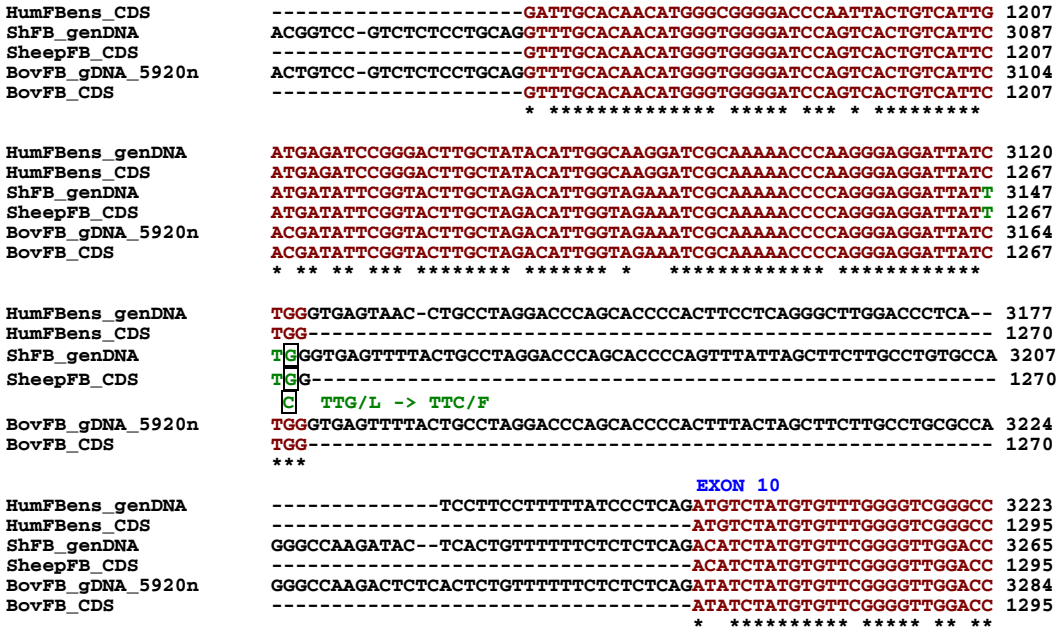
```
                        EXON 1
HumFBens_genDNA         ATGGGGAGCAATCTCAGCCCCCAACTCTGCCTGATGCCCTTTATCTTGGGCCTCTTGTCT 60
HumFBens_CDS            ATGGGGAGCAATCTCAGCCCCCAACTCTGCCTGATGCCCTTTATCTTGGGCCTCTTGTCT 60
ShFB_genDNA             ATGAGGATTGGTCACAACCCCCGACTCTGCCTGGTACCCTTGATCCTGGGCCTCTTGTGT 60
SheepFB_CDS             ATGAGGATTGGTCACAACCCCCGACTCTGCCTGGTACCCTTGATCCTGGGCCTCTTGTGT 60
BovFB_gDNA_5920n        ATGGGGATTGGTCACAACCCCCGACTCTGCCTGGTACCCTTGATCCTGGGCCTCTTGTGT 60
BovFB_CDS               ATGGGGATTGGTCACAACCCCCGACTCTGCCTGGTACCCTTGATCCTGGGCCTCTTGTGT 60
                        *** ***    ** ** ***** ********** * ***** *** ************ *

HumFBens_genDNA         GGAGGTAAGCGAGGGTAA---------CCTTCCCTTCCTGCTGTCTCCAGCATCCCTCCT 111
HumFBens_CDS            GGAG---------------------------------------------------------- 64
ShFB_genDNA             GGAGGTAAGTGAGATCACAGCCCTC-TCCCTCCACTCCTGCTTCCCCCACAATCTGTCCT 119
SheepFB_CDS             GGAG---------------------------------------------------------- 64
BovFB_gDNA_5920n        GGAGGTAAGTGAGATCACAGCCCCCCTCCCTCCACTCCTGCTTCCCCCACGATCCGTCCT 120
BovFB_CDS               GGAG---------------------------------------------------------- 64
                        ****
                                                        EXON 2
HumFBens_genDNA         TGGCCTTTTGGGGCCAGGCTTCATCAGCCTTTCTCTTCAGGTGTGACCACCACTCCATGG 171
HumFBens_CDS            ------------------------------------GTGTGACCACCACTCCATGG  84
ShFB_genDNA             CGGCCTTGTGAGGCTGGGCCTCACCAACTTTTCTCTTTAGGTGTGGGCATGACACCACTG 179
SheepFB_CDS             ------------------------------------GTGTGGGCATGACACCACTG  84
BovFB_gDNA_5920n        CGGCCTTGTGAGGCTGGGCCTCACCAACTTTTCTCTTTAGGTGTGGGCATGACGCCATTG 180
BovFB_CDS               ------------------------------------GTGTGGGCATGACGCCATTG  84
                                                            *****  **  ** ***  *

HumFBens_genDNA         TCTTTGGCCCGGCCCCAGGGATCCTGCTCTCTGGAGGGGGTAGAGATCAAAGGCGGCTCC 231
HumFBens_CDS            TCTTTGGCCCGGCCCCAGGGATCCTGCTCTCTGGAGGGGGTAGAGATCAAAGGCGGCTCC 144
ShFB_genDNA             CCTGAGGCTGGGCCTCAAAGCCCCTGCTCTCTGGAGGGAGTAGAGATCAAAGGTGGCTCC 239
SheepFB_CDS             CCTGAGGCTGGGCCTCAAAGCCCCTGCTCTCTGGAGGGAGTAGAGATCAAAGGTGGCTCC 144
BovFB_gDNA_5920n        CCTGAGGCTGGGCCCCAAAGCCCCTGCTCTCTGGAGGGAGTAGAGATCAAAGGTGGCTCC 240
BovFB_CDS               CCTGAGGCTGGGCCCCAAAGCCCCTGCTCTCTGGAGGGAGTAGAGATCAAAGGTGGCTCC 144
                         **  ***  **** **  *  ********************** ********** ******

HumFBens_genDNA         TTCCGACTTCTCCAAGAGGGCCAGGCACTGGAGTACGTGTGTCCTTCTGGCTTCTACCCG 291
HumFBens_CDS            TTCCGACTTCTCCAAGAGGGCCAGGCACTGGAGTACGTGTGTCCTTCTGGCTTCTACCCG 204
ShFB_genDNA             TTCCGGCTTCTCAAGGCGGGCCAGGTGCTGGAGTACTTGTGTCCTTCTGGCTTCTACCCA 299
SheepFB_CDS             TTCCGGCTTCTCAAGGCGGGCCAGGTGCTGGAGTACTTGTGTCCTTCTGGCTTCTACCCA 204
BovFB_gDNA_5920n        TTCCGGCTTCTCAAGGCGGGCCAGGTGCTGGAGTACTTGTGTCCTTCTGGCTTCTACCCA 300
BovFB_CDS               TTCCGGCTTCTCAAGGCGGGCCAGGTGCTGGAGTACTTGTGTCCTTCTGGCTTCTACCCA 204
                        ***** ****** * * ********  ********* ********************** 

HumFBens_genDNA         TACCCTGTGCAGACACGTACCTGCAGATCTACGGGGTCCTGGAGCACCCTGAAGACTCAA 351
HumFBens_CDS            TACCCTGTGCAGACACGTACCTGCAGATCTACGGGGTCCTGGAGCACCCTGAAGACTCAA 264
ShFB_genDNA             TATCCTGTGCAGATTCGCACCTGCAGATCCACAGGGTCCTGGAGCACCCTGCAGACTCAA 359
SheepFB_CDS             TATCCTGTGCAGATTCGCACCTGCAGATCCACAGGGTCCTGGAGCACCCTGCAGACTCAA 264
BovFB_gDNA_5920n        TACCCCACGCAGATTCGCACCTGCAGATCCACAGGGTCCTGGAGCACCCTGCAGACTCAA 360
BovFB_CDS               TACCCCACGCAGATTCGCACCTGCAGATCCACAGGGTCCTGGAGCACCCTGCAGACTCAA 264
                        ** **   *****  ** ************ ** ****************** ********

HumFBens_genDNA         GACCAAAAGACTGTCAGGAAGGCAGAGTGCAGAGGTTTGAGGGCAATGAGTGTGGGCAGT 411
HumFBens_CDS            GACCAAAAGACTGTCAGGAAGGCAGAGTGCAGAG-------------------------- 298
ShFB_genDNA             GACAGAAAGATTGTCAAGAGGGCTGAATGCAAAGGTTGGAGGGCAGTGAGGATGGGCATG 419
SheepFB_CDS             GACAGAAAGATTGTCAAGAGGGCTGAATGCAAAG------------------------- 298
BovFB_gDNA_5920n        GACAGAAAGATTGTCAAGAGGGCTGAATGCAAAGGTTGGAGGGCAGTGAGGATGGGCATG 420
BovFB_CDS               GACAGAAAGATTGTCAAGAGGGCTGAATGCAAAG------------------------- 298
                        ***  ***** ***** ** *** ** **** **

HumFBens_genDNA         GGCCTAAGGCAGAAACAGGGCAGGCGGCAGCAAGGTCAGGACTAGGAT------------ 459
HumFBens_CDS            ------------------------------------------------------------
ShFB_genDNA             GGCCGGCTGTGGGACGGAGCAGGTGGCAGCCAAGGTCCAGACTTGAATAAGTCTAGGCTT 479
SheepFB_CDS             ------------------------------------------------------------
BovFB_gDNA_5920n        GGCTGGCTGTGGGACGGAGCAGGTGGCAGCCAAGGTCCAGACTTGGATAAGTCTAGGCTT 480
BovFB_CDS               ------------------------------------------------------------

HumFBens_genDNA         -----GAGACTAGGCAGGGTGACAAGGTGGGCTGACCGGGAGTAGGAGCAGTTTTAGGGT 514
HumFBens_CDS            ------------------------------------------------------------
ShFB_genDNA             GAGTGGAGACCAGGCAGGCTGAAAAGTTGGGCTGAGCACGGGTCAGAGCTGAGCGAGGGT 539
SheepFB_CDS             ------------------------------------------------------------
BovFB_gDNA_5920n        GAGTGGAGACCAGGCAGGCTGACAAGTTGGGCTGAGCACGGGTCAGAGCTGAGCGAGGGT 540
BovFB_CDS               ------------------------------------------------------------

HumFBens_genDNA         GGCAGGCGGAAAGGGGGCAAGAAAAAGCGGAGTTAACCCTTACTAAGCATTTACCCTGGG 574
HumFBens_CDS            ------------------------------------------------------------
ShFB_genDNA             TGGAGGTGGGATGGAGAGAGAAGGAAGGGGAGCTAGACTTTACTAAGCATTTACCCTGTG 599
SheepFB_CDS             ------------------------------------------------------------
BovFB_gDNA_5920n        TGGAGGTGGGATGGAGAGAGAAGGAAGGGGAGCTAGACTTTACTAAGCATTTACCCTGTG 600
BovFB_CDS               ------------------------------------------------------------
```

```
HumFBens_genDNA       CTTCCAGGCAGCCCTGGAAGTCAAGAGAACACTCAGAAATGGGGAGGGAGAAGCAGTGGA 634
HumFBens_CDS          ------------------------------------------------------------
ShFB_genDNA           ---TCAGGCAGCCCTGGCAGTTCAAAGAACACTTAGAAATGGGGAGGAGGGAGCAGCAGT 656
SheepFB_CDS           ------------------------------------------------------------
BovFB_gDNA_5920n      ---TCAAGCAGTCCCGGTAGTTCAAAGAACACTTAGAAATGGGGAGCGGGGAGCAGCAGT 657
BovFB_CDS             ------------------------------------------------------------

HumFBens_genDNA       AATCCATATGGGTTGAGG----------AGTAGGTAAGATGCTGCTTCTGCGGGACTGGG 684
HumFBens_CDS          ------------------------------------------------------------
ShFB_genDNA           ACTCCTTATGGGTGGAGGCTCCTCCAGGGGTCGAGGGTAGTAGCTATCCTGTGGGGCCCA 716
SheepFB_CDS           ------------------------------------------------------------
BovFB_gDNA_5920n      ACTCCTTATGGGTGGAGGTTCCTCCAGGGGTTGAAGGTAGTAGCTATTCTGTGAGGCCCA 717
BovFB_CDS             ------------------------------------------------------------

HumFBens_genDNA       AATGCGCTGTTTCTCAGTGACATGGTCTCCGAGACCAGGAGGGATACACCTAAGG--CAG 742
HumFBens_CDS          ------------------------------------------------------------
ShFB_genDNA           GAAGTTATATTTCTCAGTGACAGGGTCCCTGAGACCAGAAAGGTTATCCTAAAGGGGCAG 776
                                                        (G)
SheepFB_CDS           ------------------------------------------------------------
BovFB_gDNA_5920n      GAAGTGGTATTTCTCAGTGACAGGGTCCCTGAGACCAGAAGGGTTATCCTAAAGGG-CAG 776
BovFB_CDS             ------------------------------------------------------------
                                                                    EXON 3
HumFBens_genDNA       CCTTTCCCTCTTGATGACTTCTACTTG--------TCCCCCCTTCTCAAAGCAATCCACT 794
HumFBens_CDS          ------------------------------------------------------CAATCCACT 307
ShFB_genDNA           CCCTTCCCTTCCAGTAACTTCTACTTGGTTTCCACCTCCTTGCCCCCAAAGCAATTCGCT 836
SheepFB_CDS           ------------------------------------------------------CAATTCGCT 307
BovFB_gDNA_5920n      CCCTTCCCTTCCAGTAACTTCTACTTGGTCTCCACCCCCTTGCCCCCAAAGCAATTCGCT 836
BovFB_CDS             ------------------------------------------------------CAATTCGCT 307
                                                                               **** * **
HumFBens_genDNA       GTCCAAGACCACACGACTTCGAGAACGGGGAATACTGGCCCCGGTCTCCCTACTACAATG 854
HumFBens_CDS          GTCCAAGACCACACGACTTCGAGAACGGGGAATACTGGCCCCGGTCTCCCTACTACAATG 367
ShFB_genDNA           GTCCCAGACCACAGGATTTTGAGAATGGGGAGTACTGGCCCCGCGCTGCCTACTACAATT 896
SheepFB_CDS           GTCCCAGACCACAGGATTTTGAGAATGGGGAGTACTGGCCCCGCGCTGCCTACTACAATT 367
BovFB_gDNA_5920n      GTCCCAGACCACAGGATTTTGAGAATGGGGAGTACTGGCCCCGGGCTGCCTACTACAATT 896
BovFB_CDS             GTCCCAGACCACAGGATTTTGAGAATGGGGAGTACTGGCCCCGGGCTGCCTACTACAATT 367
                      **** ******** ** ** ***** ***** ***********  ** ***********

HumFBens_genDNA       TGAGTGATGAGATCTCTTTCCACTGCTATGACGGTTACACTCTCCGGGGCTCTGCCAATC 914
HumFBens_CDS          TGAGTGATGAGATCTCTTTCCACTGCTATGACGGTTACACTCTCCGGGGCTCTGCCAATC 427
ShFB_genDNA           TGAGCGATGAGATCTCTTTCCGCTGCTATGATGGCTACACTTTTCGGGGCTCTGCCAATC 956
SheepFB_CDS           TGAGCGATGAGATCTCTTTCCGCTGCTATGATGGCTACACTTTTCGGGGCTCTGCCAATC 427
BovFB_gDNA_5920n      TGAGCGATGAGATCTCTTTCCGCTGCTATGATGGCTACACTCTTCGGGGCTCTGCCAATC 956
BovFB_CDS             TGAGCGATGAGATCTCTTTCCGCTGCTATGATGGCTACACTCTTCGGGGCTCTGCCAATC 427
                      **** ********** ********* ** ****** * **********

HumFBens_genDNA       GCACCTGCCAAGTGAATGGCCGATGGAGTGGGCAGACAGCGATCTGTGACAACGGAGGTG 974
HumFBens_CDS          GCACCTGCCAAGTGAATGGCCGATGGAGTGGGCAGACAGCGATCTGTGACAACGGAG--- 484
ShFB_genDNA           GCACCTGCCTAGGGAATGGTCGGTGGGATGGGGAAACGGCCATCTGTGATGATGGAGGTG 1016
SheepFB_CDS           GCACCTGCCTAGGGAATGGTCGGTGGGATGGGGAAACGGCCATCTGTGATGATGGAG--- 484
BovFB_gDNA_5920n      GCACCTGCCAAGGGAATGGTCGGTGGGATGGGGAAACGGCCATCTGTGATGATGGAGGTG 1016
BovFB_CDS             GCACCTGCCAAGGGAATGGTCGGTGGGATGGGGAAACGGCCATCTGTGATGATGGAG--- 484
                      ********* ** ****** ** ***  **** * ** ** ********  * ****

HumFBens_genDNA       AGAAGCATCCCCTC---CCCCTACATTGCTGTCTCCCTGACGGCGCCCAGCCCGAGGAGT 1031
HumFBens_CDS          ------------------------------------------------------------
ShFB_genDNA           AGAACCGTCCCCCCG-TCCACAACAGCTCCTTCTCCCTGACCACCTGCAGCCTGAGGAA- 1074
SheepFB_CDS           ------------------------------------------------------------
BovFB_gDNA_5920n      AGAACCGTCCCCCCTGTCCACAACAGCTCCTTCTCCCTGACCACCCGCAGCCTGAGGAA- 1075
BovFB_CDS             ------------------------------------------------------------

HumFBens_genDNA       GGGCACTCGGCTCCGGACACTGTAACTCTTGCTCTCTACCTTGCTCACGGGGCCTCAGGC 1091
HumFBens_CDS          ------------------------------------------------------------
ShFB_genDNA           ------------CCAGGGCCACAGCTCCTGGTGCTCTGCCTTGCCTTTTGCGCCCCAGGC 1122
SheepFB_CDS           ------------------------------------------------------------
BovFB_gDNA_5920n      ------------CCAGTGCCACAGCTCTCGGTTCTCTGCCTTGCCCTTTGGGCCCCAGGC 1123
BovFB_CDS             ------------------------------------------------------------


                                                    EXON 4
HumFBens_genDNA       TTCAGTGCTTACCTCGATGTCTCATACCTCTGCAGCGGGGTACTGCTCCAACCCGGGCAT 1151
HumFBens_CDS          ----------------------------------CGGGGTACTGCTCCAACCCGGGCAT 509
ShFB_genDNA           TTTGGCCCTCACCAGTGTGTCTTACACTTCTGCAGCGGGGTACTGCCCCAACCCGGGCAT 1182
SheepFB_CDS           ----------------------------------CGGGGTACTGCCCCAACCCGGGCAT 509
BovFB_gDNA_5920n      TTTGGCCCTCACCAGTGTGTCTTACACTTCTGCAGCGACGTACTGCCCCAACCCGGGCAT 1183
BovFB_CDS             ----------------------------------CGACGTACTGCCCCAACCCGGGCAT 509
                                                       **  ******* *************

HumFBens_genDNA       CCCCATTGGCACAAGGAAGGTGGGCAGCCAGTACCGCCTTGAAGACAGCGTCACCTACCA 1211
HumFBens_CDS          CCCCATTGGCACAAGGAAGGTGGGCAGCCAGTACCGCCTTGAAGACAGCGTCACCTACCA 569
ShFB_genDNA           CCCCCTTGGCACGAGGAAGGTGGGCAGCCAGTACCGCCTTGAAGATCGTGTCACCTACTA 1242
SheepFB_CDS           CCCCCTTGGCACGAGGAAGGTGGGCAGCCAGTACCGCCTTGAAGATCGTGTCACCTACTA 569
BovFB_gDNA_5920n      CCCCCTTGGCACGAGGAAGGTGGGCAGCCAGTACCGCCTTGAAGACCGTGTCACCTACTA 1243
BovFB_CDS             CCCCCTTGGCACGAGGAAGGTGGGCAGCCAGTACCGCCTTGAAGACCGTGTCACCTACTA 569
                      **** ******* ********************************  * ********* *
```

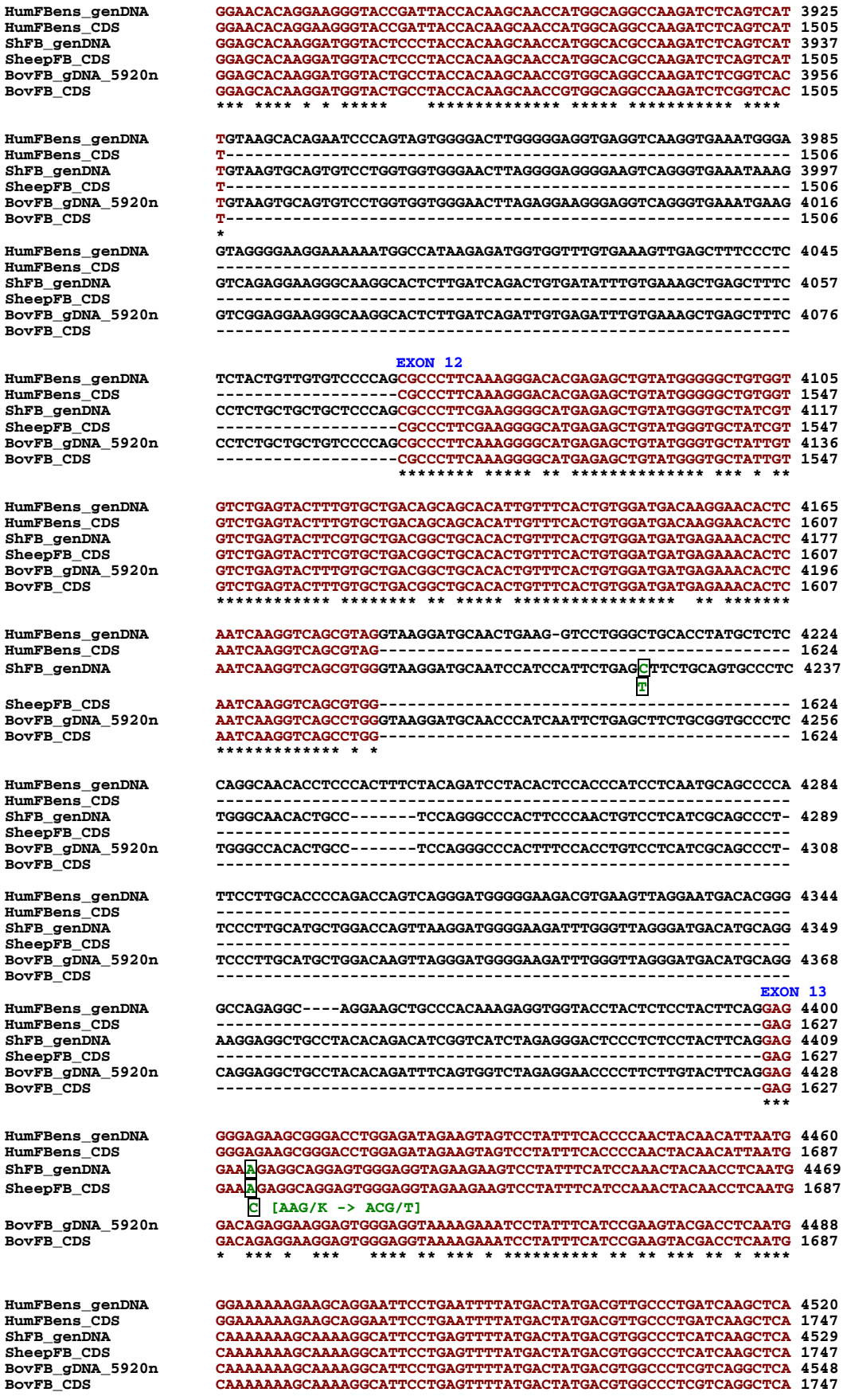
```
HumFBens_genDNA      CTGCAGCCGGGGGCTTACCCTGCGTGGCTCCCAGCGGCGAACGTGTCAGGAAGGTGGCTC 1271
HumFBens_CDS         CTGCAGCCGGGGGCTTACCCTGCGTGGCTCCCAGCGGCGAACGTGTCAGGAAGGTGGCTC 629
ShFB_genDNA          CTGCAACCGGGGGCTCACTCTACGTGGTTCCCAGCAGCGAACATGCCAGGAAGGTGGCTC 1302
SheepFB_CDS          CTGCAACCGGGGGCTCACTCTACGTGGTTCCCAGCAGCGAACATGCCAGGAAGGTGGCTC 629
BovFB_gDNA_5920n     CTGCAACCGGGGGCTCACTCTACGTGGTTCTGAGCAGCGAACATGCCTGGAAGGTGGCTC 1303
BovFB_CDS            CTGCAACCGGGGGCTCACTCTACGTGGTTCTGAGCAGCGAACATGCCTGGAAGGTGGCTC 629
                     ***** ********* ** ** ***** **  *** ****** ** * ************

HumFBens_genDNA      TTGGAGCGGGACGGAGCCTTCCTGCCAAGGTGACCTTTG---ACCTGTACCCCCAGGTCA 1328
HumFBens_CDS         TTGGAGCGGGACGGAGCCTTCCTGCCAAG------------------------------ 658
ShFB_genDNA          TTGGAGTGGAACAGAGCCTTCCTGCCAAGGTGACCCTTACTTATCTGTACCCTCAGGTCA 1362
SheepFB_CDS          TTGGAGTGGAACAGAGCCTTCCTGCCAAG------------------------------ 658
BovFB_gDNA_5920n     TTGGAGTGGAACGGAGCCTTCCTGCCAAGGTGACCCTTA----TCTGTACCCTCAAGTCA 1359
BovFB_CDS            TTGGAGTGGAACGGAGCCTTCCTGCCAAG------------------------------ 658
                     ****** ** ** ****************

HumFBens_genDNA      GATCCTGGTCTTCCATCCTACTGTCTTCTCTCCCCACCTCAACCCTGCTCTTTCCTCACT 1388
HumFBens_CDS         ------------------------------------------------------------
ShFB_genDNA          GACCCTGCTCTCCCATCCGCACATCCCCCCAGGAGCATACTGCCATCTCCCGCTGCCACC 1422
SheepFB_CDS          ------------------------------------------------------------
BovFB_gDNA_5920n     GACCCTGCTCTCCCATCTGCACATCCCCTCAGGAGCGCACTGCCTTCTCTCACTGCCACC 1419
BovFB_CDS            ------------------------------------------------------------

HumFBens_genDNA      TTGTTTAAACCTCCCTGTACAACTATCTCACTTCTGAGCCTTTTATACCCTGGAAACCCA 1448
HumFBens_CDS         ------------------------------------------------------------
ShFB_genDNA          ATCCAGCTCCTTCCTCCCTTCTTAATCTCCCCGAAGAGCTTCTGAGCCCTTCCCGCCCCA 1482
SheepFB_CDS          ------------------------------------------------------------
BovFB_gDNA_5920n     ATCCAGCTCCTTCCTCCCTTCTCTAAATCTCCCTGAAGAACTTCTAAGCCCTTCCCGCCC 1479
BovFB_CDS            ------------------------------------------------------------

HumFBens_genDNA      TGATCCCCCGTCTCTTTGGTCACTGTATCCCTGACACTCCCAGACATTTGACCTCATTTC 1508
HumFBens_CDS         ------------------------------------------------------------
ShFB_genDNA          GAAGCCCACCACCTCCTGT---CAGTGTCCCTGACCCTCCCAGACATTTGACCTGCTCTC 1539
SheepFB_CDS          ------------------------------------------------------------
BovFB_gDNA_5920n     CTGAAGCCCACCATCTCGTGT-CAGTGTCCCTGACCCTCCCAGACATTTGACCTGCTCTC 1538
BovFB_CDS            ------------------------------------------------------------

                           EXON 5
HumFBens_genDNA      TGAC-TCTCCCAG-ACTCCTTCATGTACGACACCCCTCAAGAGGTGGCCGAAGCTTTCCT 1566
HumFBens_CDS         -------------ACTCCTTCATGTACGACACCCCTCAAGAGGTGGCCGAAGCTTTCCT 704
ShFB_genDNA          TGACCTCTCCCAG-ACTCCTTTATGTACGACACTCCTGCAGAGGTGGCCGAAGCCTTCCT 1598
SheepFB_CDS          -------------ACTCCTTTATGTACGACACTCCTGCAGAGGTGGCCGAAGCCTTCCT 704
BovFB_gDNA_5920n     TGATCTCTCCCAGAACTCCTTTATGTACGACACTCCTGCAGAGGTGGCCGAAGCTTTCCT 1598
BovFB_CDS            -------------ACTCCTTTATGTACGACACTCCTGCAGAGGTGGCCGAAGCTTTCCT 704
                              ******* *********** ***  *************** *****

s
HumFBens_genDNA      GTCTTCCCTGACAGAGACCATAGAAGGAGTCGATGCTGAGGATGGGCACGGCCCAGGTTT 1626
HumFBens_CDS         GTCTTCCCTGACAGAGACCATAGAAGGAGTCGATGCTGAGGATGGGCACGGCCCAG---- 760
ShFB_genDNA          GTCTTCCCTGACAGAGACCATAGAAGGAGTTGATGCCGAGGATGGACACAGCCCAGGTTT 1658
SheepFB_CDS          GTCTTCCCTGACAGAGACCATAGAAGGAGTTGATGCCGAGGATGGACACAGCCCAG---- 760
BovFB_gDNA_5920n     GTCTTCCCTGACAGAGACCATAGAAGGAGTTGATGCCGAGGATGGACATAGCCCAGGTTT 1658
BovFB_CDS            GTCTTCCCTGACAGAGACCATAGAAGGAGTTGATGCCGAGGATGGACATAGCCCAG---- 760
                     ****************************** ***** ******** **  ******

HumFBens_genDNA      GAAGACAGAGA----AGGGAGGCAGGGCAGGGAACTGGGGGAAAATGGAGAAGGGACAGA 1682
HumFBens_CDS         ------------------------------------------------------------
ShFB_genDNA          GAAGGCAGAGGGG--AGGCGGGGCAGGGAGTTGGGCGTGGGGTGGAGAAGGGGCAGGAGA 1716
SheepFB_CDS          ------------------------------------------------------------
BovFB_gDNA_5920n     GAAGGCAGAGAGGGGAGGCGGGGCAGGGAGTTGGGCATGGGGTGGAGAAGGGGCAGGAGA 1718
BovFB_CDS            ------------------------------------------------------------

                                                    EXON 6
HumFBens_genDNA      ACTGTTAATGCTGGAGCCTGAGCCACTCTCCTGGCACCCAGGGGAACAACAGAAGCGGAA 1742
HumFBens_CDS         -------------------------------------------GGGAACAACAGAAGCGGAA 779
ShFB_genDNA          CCTATTTGTTCTGGAGCCTGAGCCTCTCTGATGGCATCCAGGGGAACAACAGAAGAGAAA 1776
SheepFB_CDS          -------------------------------------------GGGAACAACAGAAGAGAAA 779
BovFB_gDNA_5920n     CCTCTTTGTTCTGGAGCCTGAGCCTCTCTGGTGACATCCAGGGGAACAACAGAAGAGAAA 1778
BovFB_CDS            -------------------------------------------GGGAACAACAGAAGAGAAA 779
                                                    ************** * **

HumFBens_genDNA      GATCGTCCTGGACCCTTCAGGCTCCATGAACATCTACCTGGTGCTAGATGGATCAGACAG 1802
HumFBens_CDS         GATCGTCCTGGACCCTTCAGGCTCCATGAACATCTACCTGGTGCTAGATGGATCAGACAG 839
ShFB_genDNA          GATTGTCCTGGACCCCTCAGGCTCCATGAACATCTACCTGGTATTGGATGGATCAGACAG 1836
SheepFB_CDS          GATTGTCCTGGACCCCTCAGGCTCCATGAACATCTACCTGGTATTGGATGGATCAGACAG 839
BovFB_gDNA_5920n     GATTGTCCTGGACCCCTCAGGCTCCATGAACATCTACCTGGTATTGGATGGATCAGACAG 1838
BovFB_CDS            GATTGTCCTGGACCCCTCAGGCTCCATGAACATCTACCTGGTATTGGATGGATCAGACAG 839
                     *** *********** *************************  * **************

HumFBens_genDNA      CATTGGGGCCAGCAACTTCACAGGAGCCAAAAAGTGTCTAGTCAACTTAATTGAGAAGGT 1862
HumFBens_CDS         CATTGGGGCCAGCAACTTCACAGGAGCCAAAAAGTGTCTAGTCAACTTAATTGAGAAG-- 897
ShFB_genDNA          CGTGGGGGCCCACAACTTCACAGGGGCCAAGAATTGTCTCAGAGATTTCATTGAGAAGGT 1896
SheepFB_CDS          CGTGGGGGCCCACAACTTCACAGGGGCCAAGAATTGTCTCAGAGATTTCATTGAGAAG-- 897
BovFB_gDNA_5920n     CGTGGGGGCCCACAACTTCACAGGGGCCAAGAATTGTCTCAGAGATTTCATCGAGAAGGT 1898
BovFB_CDS            CGTGGGGGCCCACAACTTCACAGGGGCCAAGAATTGTCTCAGAGATTTCATCGAGAAG-- 897
```

```
                        * * ******  ************ ***** ** *****     * ** ** ******

HumFBens_genDNA         GGAATCCTCCTATCCCTGAACTCGGGGGAATGGAA---TCTCGCTGATCTTCCAGGACTA 1919
HumFBens_CDS            ------------------------------------------------------------
ShFB_genDNA             GGGGCCCCCTCCCCTTGAAGGTGGGAAC------------------------------- 1924
SheepFB_CDS             ------------------------------------------------------------
BovFB_gDNA_5920n        GGGGCCCCCTCCCCTTGAAGGTGGGAACCTTGAAAGGACCTCACTGCCCTGCCAGGACTA 1958
BovFB_CDS               ------------------------------------------------------------


HumFBens_genDNA         GCTCCCTGATCATTCCAGCCCCTCTGAACAACAGGGCCCCAGGAAAATCTCCAGGTCCTA 1979
HumFBens_CDS            ------------------------------------------------------------
ShFB_genDNA             ------------TCCCAGCCCCTCTGAGCCACAGGGTCCCAG-AAAATCTCCAGGTCCCA 1971
SheepFB_CDS             ------------------------------------------------------------
BovFB_gDNA_5920n        GTTCCCAAATACTCCCAGCCCCTTGGAGCCACAGGGTCCCAG-AAAATCTCCAGGTCCTA 2017
BovFB_CDS               ------------------------------------------------------------


HumFBens_genDNA         TTCTGTCCTCCTTCCCTTTTACTTGAAGCAGTTTCTTGACTGGTAATTCCTCCATGAACC 2039
HumFBens_CDS            ------------------------------------------------------------
ShFB_genDNA             TTGTGTCCTCCTTTCCTTTTACTTGACCTTTAACTTGGACTCCTCCATGACCCCAGGGTG 2031
SheepFB_CDS             ------------------------------------------------------------
BovFB_gDNA_5920n        TTGTGTCCTTCTTTCCTTTTACTTGACCTTTAACCTGGACTCCTCCATGACCTCAGGGTG 2077
BovFB_CDS               ------------------------------------------------------------


HumFBens_genDNA         -----TCAGCCCTTGAGCCTCTTACTGAGAGCCTCCCT----GTCCCAGCAAAGTCGCTG 2090
HumFBens_CDS            ------------------------------------------------------------
ShFB_genDNA             ATGAGTCAGTCCTTGAGTCTCTTACTAAGGACTTTCCTATCTGTCTCAGCCAAGCTGCTG 2091
SheepFB_CDS             ------------------------------------------------------------
BovFB_gDNA_5920n        ATGAGTCAGTTCTTGAGTCTCTTACTAAGGACTTTCCTATCTGTCTCAGCCCAGCTGCTG 2137
BovFB_CDS               ------------------------------------------------------------


HumFBens_genDNA         AAATCTCCCAATCACAGTATTCTATTTTCAATGCCATGGCGCCTTGTTCTCCTCACCCAC 2150
HumFBens_CDS            ------------------------------------------------------------
ShFB_genDNA             AATTCTCCCAATCGTA-TATTCTACTTTTGATGCTGTGACCTATTGCTTCCCTTGCCCAC 2150
SheepFB_CDS             ------------------------------------------------------------
BovFB_gDNA_5920n        AATTCTCCCAACCGTA-TATTCCTCTTTTGATGCTGTGACCTATTGCTCCCCTCGCCCAC 2196
BovFB_CDS               ------------------------------------------------------------

                          EXON 7
HumFBens_genDNA         AGGTGGCAAGTTATGGTGTGAAGCCAAGATATGGTCTAGTGACATATGCCACATACCCCA 2210
HumFBens_CDS            --GTGGCAAGTTATGGTGTGAAGCCAAGATATGGTCTAGTGACATATGCCACATACCCCA 955
ShFB_genDNA             AGGTGGCAAGCTACGGGGGTGAAGCCAAAATATGGGCTAGTGACATATGCCACAGAACCCA 2210
SheepFB_CDS             --GTGGCAAGCTACGGGGGTGAAGCCAAAATATGGGCTAGTGACATATGCCACAGAACCCA 955
BovFB_gDNA_5920n        AGGTGGCAAGCTATGGGGGTGAAGCCAAAATATGGTCTAGTGACATATGCCACAGAACCCA 2256
BovFB_CDS               --GTGGCAAGCTATGGGGGTGAAGCCAAAATATGGTCTAGTGACATATGCCACAGAACCCA 955
                          ******** ** ** ********** ****** ***************** * ****




HumFBens_genDNA         AAATTTGGGTCAAAGTGTCTGAAGCAGACAGCAGTAATGCAGACTGGGTCACGAAGCAGC 2270
HumFBens_CDS            AAATTTGGGTCAAAGTGTCTGAAGCAGACAGCAGTAATGCAGACTGGGTCACGAAGCAGC 1015
ShFB_genDNA             AAGTTTTGATCAAAGTGTTCGATCCAAAGAGCAGTGAAGCAGACTGGGTCACAGAGCAGC 2270
                                                    C [AGT/S -> AGC/S]
SheepFB_CDS             AAGTTTTGATCAAAGTGTTCGATCCAAAGAGCAGTGAAGCAGACTGGGTCACAGAGCAGC 1015
BovFB_gDNA_5920n        AAGTTTTGATCAGAGTGTCCGACCCAAAGAGCAGCGAAGCAGACTGGGTCACAGACCAGC 2316
BovFB_CDS               AAGTTTTGATCAGAGTGTCCGACCCAAAGAGCAGCGAAGCAGACTGGGTCACAGACCAGC 1015
                        ** *** * *** *****  **  ** * *****  * **************  * ****

HumFBens_genDNA         TCAATGAAATCAATTATGAAGGTCAGAGGTTAGGGAATGGTGGGAGGTTCACTTTGGGGT 2330
HumFBens_CDS            TCAATGAAATCAATTATGAAG---------------------------------------- 1036
ShFB_genDNA             TCAACCGAATCAACTACGCAGGTCAGAAGTCAGGGAAGGGTGGGAGGTTCACTTTAGGGT 2330
SheepFB_CDS             TCAACCGAATCAACTACGCAG---------------------------------------- 1036
BovFB_gDNA_5920n        TCAACCAAATCAACTACGCAGGTCAGAAGTCAGGGAAGGGTGCAAGGTTCACTTTGGGGT 2376
BovFB_CDS               TCAACCAAATCAACTACGCAG---------------------------------------- 1036
                        ****   ****** ** * **

HumFBens_genDNA         CAGGAGGTTCAGG----------------------------------------------- 2343
HumFBens_CDS            ------------------------------------------------------------
ShFB_genDNA             CAGAGAGGTCAGAAGTCAGGGAAGGGTGGGAGGTTCACTTTGGTCTGAAGTTTCAGAAGT 2390
SheepFB_CDS             ------------------------------------------------------------
BovFB_gDNA_5920n        CAGAGAGGTCTGAAGT---------------------------------TTCAGGAGT 2401
BovFB_CDS               ------------------------------------------------------------


HumFBens_genDNA         ---GTGGAGGGGGTCATGAGACTACCTTGAGGGCGACAGGGAGGACCACTTTGTAGTCAA 2400
HumFBens_CDS            ------------------------------------------------------------
ShFB_genDNA             GTTGTGGGGAGGGTGTAGTGAGACCATGTGGGGAGGCAGGGAGGACTACGTTTTAATCCA 2450
                                                      A
SheepFB_CDS             ------------------------------------------------------------
BovFB_gDNA_5920n        GTCGTGGGGAGGGTGTAGTGAGACCATGTGGTGGGGCAGGGAGGACTACTTTTTAATCCA 2461
```

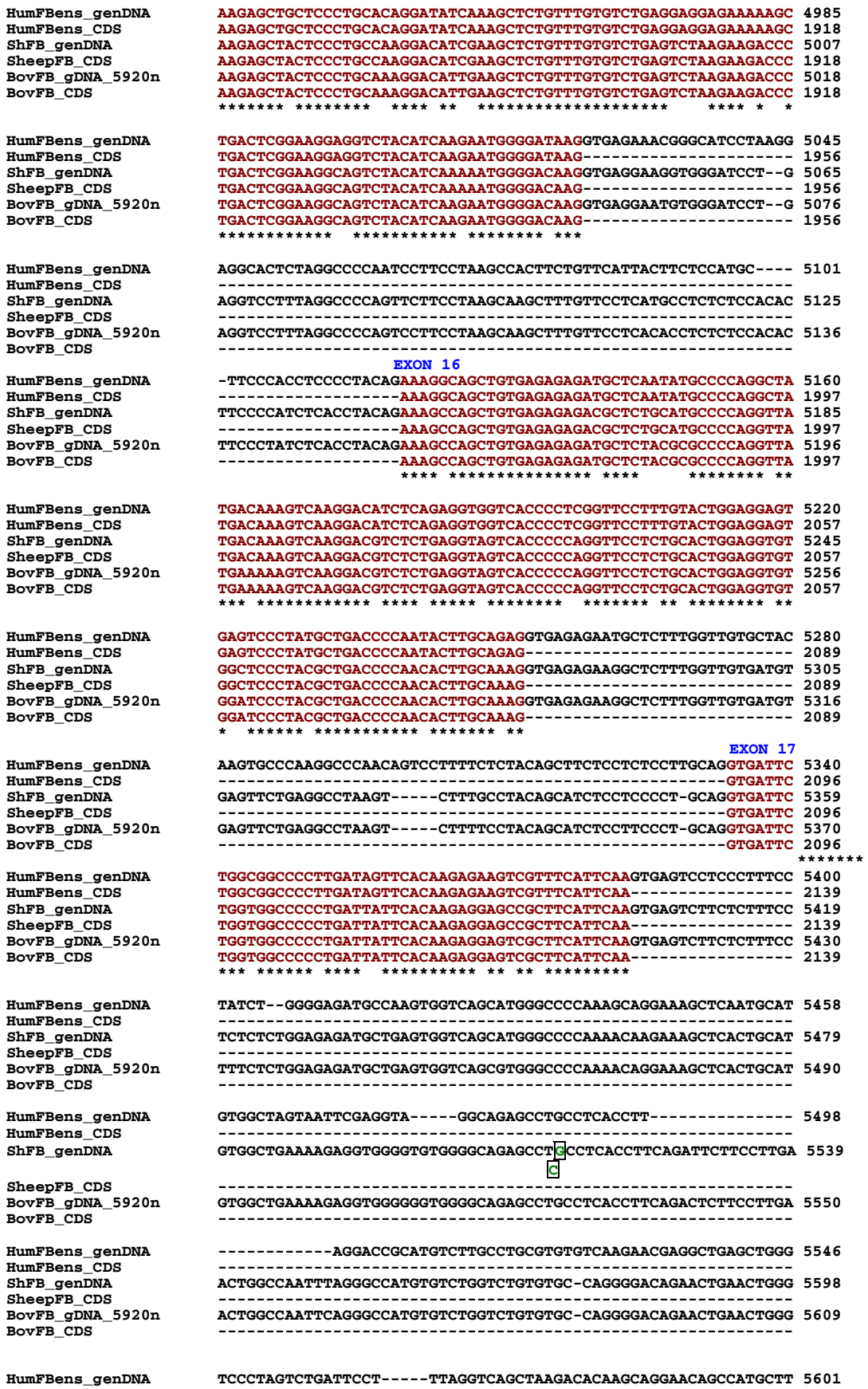
```
BovFB_CDS          ------------------------------------------------------------

HumFBens_genDNA    AAGTTGAACAGCAGGATCGTTGGGCAATGGAGGTTAGTGGGAACCTGTTGGGGGCTGGAA 2460
HumFBens_CDS       ------------------------------------------------------------
ShFB_genDNA        TGGTTGGACAGCAGGGTCATTGGAAAATGTGGTTTGGTGGGAATCCTTTGGAA-CTGAGA 2509
SheepFB_CDS        ------------------------------------------------------------
BovFB_gDNA_5920n   CGGTTGGACAGCAGGGTCATTGGAAAATGGGGTTTGGTGGGAACCCTTTGGAA-CTGAGA 2520
BovFB_CDS          ------------------------------------------------------------

HumFBens_genDNA    GGGCCACTTTGTGGTCAAAGGGAAGTCCGTGTAATGATGATTAACTTAAAAAGTTGAAAG 2520
HumFBens_CDS       ------------------------------------------------------------
ShFB_genDNA        GGGGTACTTTGTTGTCAAAGGGAAGACAGTGGGGTAACAACTCA-------AGTTGACAG 2562
SheepFB_CDS        ------------------------------------------------------------
BovFB_gDNA_5920n   GGGGCACTTTATTGTCAAAGAAAAGACAATGGGGTGACAACTCA-------AGTTGACAG 2573
BovFB_CDS          ------------------------------------------------------------

HumFBens_genDNA    ATGTGGGATTTCAGTTGCAGATTGGTCTCTGGGGTTAAAAGATGGCTTGGAAGACCAGGT 2580
HumFBens_CDS       ------------------------------------------------------------
ShFB_genDNA        ATGTGGGATGCCAGTTGCTGCCAGACCTCTGGGGTCAAG-GATGGCTTGGGAGACCAGG- 2620
SheepFB_CDS        ------------------------------------------------------------
BovFB_gDNA_5920n   ATGTGGGAACCCAGTTGCCGCTGGACCCCTGGGGTCAAG-GACGGGTTGGGAGACCAGGC 2632
BovFB_CDS          ------------------------------------------------------------

                                          EXON 8
HumFBens_genDNA    GAGGTGATGGTCTCTTCCCTCTCCACAGACCACAAGTTGAAGTCAGGGACTAACACCAAG 2640
HumFBens_CDS       ----------------------------ACCACAAGTTGAAGTCAGGGACTAACACCAAG 1068
ShFB_genDNA        ----TGATGACATCTTCCCTGTTCACAGATCACAAGCTGAAAGCAGGGACCAACACCAAG 2676
SheepFB_CDS        ----------------------------ATCACAAGCTGAAAGCAGGGACCAACACCAAG 1068
BovFB_gDNA_5920n   GAGGTGACGACACCTTCCCTGTTTCCAGATCACAAGTTGAAGGCAGGGACCAATACCAAG 2692
BovFB_CDS          ----------------------------ATCACAAGTTGAAGGCAGGGACCAATACCAAG 1068
                                                * ****** ****  ******* ** ******

HumFBens_genDNA    AAGGCCCTCCAGGCAGTGTACAGCATGATGAGCTGGCCAGATGACGTCCCTCCTGAAGGC 2700
HumFBens_CDS       AAGGCCCTCCAGGCAGTGTACAGCATGATGAGCTGGCCAGATGACGTCCCTCCTGAAGGC 1128
ShFB_genDNA        AGGGCTCTCCTGGAAGTGTACAACATGATGAGCAGAGATATAAACAACCTCAAAGAGACC 2736
SheepFB_CDS        AGGGCTCTCCTGGAAGTGTACAACATGATGAGCAGAGATATAAACAACCTCAAAGAGACC 1128
BovFB_gDNA_5920n   AGGGCTCTCCTGGAAGTGTACAACATGATGAGCAGGGAAGTAAACCAGTTCAAAGAGACC 2752
BovFB_CDS          AGGGCTCTCCTGGAAGTGTACAACATGATGAGCAGGGAAGTAAACCAGTTCAAAGAGACC 1128
                   * *** **** ** ******** ********** *        **         **   *

HumFBens_genDNA    TGGAACCGCACCCGCCATGTCATCATCCTCATGACTGATGGTCAGAAGGGACCTCTCTCC 2760
HumFBens_CDS       TGGAACCGCACCCGCCATGTCATCATCCTCATGACTGATG-------------------- 1168
ShFB_genDNA        TGGAACCGCACGCGCCACGTCATTATCATCATGACTGATGGTAAGAGGGG---CTTTTCT 2793
SheepFB_CDS        TGGAACCGCACGCGCCACGTCATTATCATCATGACTGATG-------------------- 1168
BovFB_gDNA_5920n   TGGAACCGCACGCGCCACGTCATCATCATCATGACTGACGGTAAGAGGGG---CTTTTCT 2809
BovFB_CDS          TGGAACCGCACGCGCCACGTCATCATCATCATGACTGACG-------------------- 1168
                   ********** ***** ***** *** ********** *

HumFBens_genDNA    TGTCCCAGCCTCCCCACCTTCTCAGACCAGCATGTGGCCCTTAAGTCCACTTGTAACACT 2820
HumFBens_CDS       ------------------------------------------------------------
ShFB_genDNA        CCTGCCTGACTCCCCACCCTCTCAAATCAGCTCGTGGCCCC-GAGCCCACACATAGCAC- 2851
SheepFB_CDS        ------------------------------------------------------------
BovFB_gDNA_5920n   CCTGCCTGACTCCCCACCCTCTCAAACCAGCCCGTGGCCCC-AAGTCCACATGTAACACT 2868
BovFB_CDS          ------------------------------------------------------------

HumFBens_genDNA    ATACCCATGGTTGGGGCCCTGAATGTGACTCATAGCTGGCTGTTCATCTCTCCTGTGACC 2880
HumFBens_CDS       ------------------------------------------------------------
ShFB_genDNA        AGACTTGTGGCTGGGGGATGGCGTGCTGCTCTTCCAGTGCCTTTACAAGGAGGGCCCCTC 2911
SheepFB_CDS        ------------------------------------------------------------
BovFB_gDNA_5920n   GGACTTGTGGTGGGGGATGGCGTGCTGCTCTTCCTTGTGCCTTTACAAGGAGGGACCCTC 2928
BovFB_CDS          ------------------------------------------------------------

HumFBens_genDNA    CTTCATAAGGAATTCTTCCTAAGCCCTGTGATCAACTATCTCTAACCCTTCCTCAACTTG 2940
HumFBens_CDS       ------------------------------------------------------------
ShFB_genDNA        CACTCCATGACCTCTCACTCATGGTCTCCTCCCCAACCTGCTCACTCAGCCCTGTAC--- 2968
                             T
SheepFB_CDS        ------------------------------------------------------------
BovFB_gDNA_5920n   CACTCCGTGATCTCTCACTCATGATCTCCTCTCCAACCTGCTCACTCAGCCCTGTAT--- 2985
BovFB_CDS          ------------------------------------------------------------

HumFBens_genDNA    CTCACCCTGCCATGTGTATCCCTGCCTTTAGCCAGTTTATCTTCCTTATCTCCTACCCTC 3000
HumFBens_CDS       ------------------------------------------------------------
ShFB_genDNA        GTCATTCCTTCCAGCCTCTTTGTCTTCCTGGACCCTCTCCCTTCACGTCTTCCCACCCTC 3028
                                  C
SheepFB_CDS        ------------------------------------------------------------
BovFB_gDNA_5920n   GTCATTCCTTCCAGCCTCTTTGTCTTCCTGGACCCTCTTCCTTCATGTCTCCCCACCCTC 3045
BovFB_CDS          ------------------------------------------------------------

                                          EXON 9
HumFBens_genDNA    ATGGTCCTGTCTCTTCTGCAGGATTGCACAACATGGGCGGGGACCCAATTACTGTCATTG 3060
```


HumFBens_CDS ----------GATTGCACAACATGGGCGGGGACCCAATTACTGTCATTG 1207
ShFB_genDNA ACGGTCC-GTCTCTCCTGCAGGTTTGCACAACATGGGTGGGGATCCAGTCACTGTCATTC 3087
SheepFB_CDS ----------GTTTGCACAACATGGGTGGGGATCCAGTCACTGTCATTC 1207
BovFB_gDNA_5920n ACTGTCC-GTCTCTCCTGCAGGTTTGCACAACATGGGTGGGGATCCAGTCACTGTCATTC 3104
BovFB_CDS ----------GTTTGCACAACATGGGTGGGGATCCAGTCACTGTCATTC 1207
HumFBens_genDNA ATGAGATCCGGGACTTGCTATACATTGGCAAGGATCGCAAAAACCCAAGGGAGGATTATC 3120
HumFBens_CDS ATGAGATCCGGGACTTGCTATACATTGGCAAGGATCGCAAAAACCCAAGGGAGGATTATC 1267
ShFB_genDNA ATGATATTCGGTACTTGCTAGACATTGGTAGAAATCGCAAAAACCCCAGGGAGGATTATT 3147
SheepFB_CDS ATGATATTCGGTACTTGCTAGACATTGGTAGAAATCGCAAAAACCCCAGGGAGGATTATT 1267
BovFB_gDNA_5920n ACGATATTCGGTACTTGCTAGACATTGGTAGAAATCGCAAAAACCCCAGGGAGGATTATC 3164
BovFB_CDS ACGATATTCGGTACTTGCTAGACATTGGTAGAAATCGCAAAAACCCCAGGGAGGATTATC 1267
HumFBens_genDNA TGGGTGAGTAAC-CTGCCTAGGACCCAGCACCCCACTTCCTCAGGGCTTGGACCCTCA-- 3177
HumFBens_CDS TGG---------- 1270
ShFB_genDNA TGGGTGAGTTTTACTGCCTAGGACCCAGCACCCCAGTTTATTAGCTTCTTGCCTGTGCCA 3207
SheepFB_CDS TGG---------- 1270
C TTG/L -> TTC/F
BovFB_gDNA_5920n TGGGTGAGTTTTACTGCCTAGGACCCAGCACCCCACTTTACTAGCTTCTTGCCTGCGCCA 3224
BovFB_CDS TGG---------- 1270
EXON 10
HumFBens_genDNA ----------TCCTTCCTTTTTATCCCTCAGATGTCTATGTGTTTGGGGTCGGGCC 3223
HumFBens_CDS ----------ATGTCTATGTGTTTGGGGTCGGGCC 1295
ShFB_genDNA GGGCCAAGATAC--TCACTGTTTTTTCTCTCTCAGACATCTATGTGTTCGGGGTTGGACC 3265
SheepFB_CDS ----------ACATCTATGTGTTCGGGGTTGGACC 1295
BovFB_gDNA_5920n GGGCCAAGACTCTCACTCTGTTTTTTCTCTCTCAGATATCTATGTGTTCGGGGTTGGACC 3284
BovFB_CDS ----------ATATCTATGTGTTCGGGGTTGGACC 1295

```
HumFBens_genDNA      TTTGGTGAACCAAGTGAACATCAATGCTTTGGCTTCCAAGAAAGACAATGAGCAACATGT 3283
HumFBens_CDS         TTTGGTGAACCAAGTGAACATCAATGCTTTGGCTTCCAAGAAAGACAATGAGCAACATGT 1355
ShFB_genDNA          TCTAGTGAACCAAGAGAACATCAATGCTTTGGCCTCCAAGAAGGATAAAGAGCAACACGT 3325
SheepFB_CDS          TCTAGTGAACCAAGAGAACATCAATGCTTTGGCCTCCAAGAAGGATAAAGAGCAACACGT 1355
BovFB_gDNA_5920n     TCTAGTGAACCAAGAGAACATCAATGCTTTGGCTTCCAAGAAGGATAAAGAGAAACACGT 3344
BovFB_CDS            TCTAGTGAACCAAGAGAACATCAATGCTTTGGCTTCCAAGAAGGATAAAGAGAAACACGT 1355
                     * * ********** ****************** ******** ** ** *** **** **

HumFBens_genDNA      GTTCAAAGTCAAGGATATGGAAAACCTGGAAGATGTTTTCTACCAAATGATCGGTAGGGA 3343
HumFBens_CDS         GTTCAAAGTCAAGGATATGGAAAACCTGGAAGATGTTTTCTACCAAATGATCG------- 1408
ShFB_genDNA          GTTCAAACTCCAGGGCATGGAAAACCTGGAGGATGTTTTCGTCCAAATGCTTGGTAAGAA 3385
SheepFB_CDS          GTTCAAACTCCAGGGCATGGAAAACCTGGAGGATGTTTTCGTCCAAATGCTTG------- 1408
BovFB_gDNA_5920n     GTTCAAACTCCAAGGCATGGAAAACCTGGAGGATGTTTTCGTCCAAATGCTTGGTAGGAA 3404
BovFB_CDS            GTTCAAACTCCAAGGCATGGAAAACCTGGAGGATGTTTTCGTCCAAATGCTTG------- 1408
                     ******* ** * *  ************** *********  ******* * *

HumFBens_genDNA      GATACA-------------AGGGAATAAAGAACACAACTCTCCTCAGGTTCCCCTGAAGT 3390
HumFBens_CDS         ------------------------------------------------------------
ShFB_genDNA          GCTATGGAAAGCTATAAGGAGATGGTCAGGAAACTCAGCTCTCCCCAGACCCCTCAAGGT 3445
SheepFB_CDS          ------------------------------------------------------------
BovFB_gDNA_5920n     GCTATGGAAGGCTATAAGGAGATGGTCAGGAAACTCAGCTCTCCCCAGACCCCTCAAGGT 3464
BovFB_CDS            ------------------------------------------------------------

HumFBens_genDNA      AATTCATTCTTCCTCTACACCTGAAGCTCTAGTTGCCTGGAAAGCCTTCTTCATTCCTCC 3450
HumFBens_CDS         ------------------------------------------------------------
ShFB_genDNA          CACTCATTCTTCCACTCCTTCTAAAGCCCTGGAAAGTCTGGAAGCCCTTTC----TCTCC 3501
SheepFB_CDS          ------------------------------------------------------------
BovFB_gDNA_5920n     CACTCATTCTTCCACTCCTCCTAAAGCCCTGGAAAGTCTGGAAGCCCTTTC----TTTCC 3520
BovFB_CDS            ------------------------------------------------------------

HumFBens_genDNA      TTCTCTACCTCAGTGTCACTATTCTTGTTTCCTGGCACTGTTCACT---TAACCTTAGAA 3507
HumFBens_CDS         ------------------------------------------------------------
ShFB_genDNA          TCCTCTACCTCCAGGTCACTATTCTTGCTTCTTGGTACTTTCCACTCCCTGATCTCAGAA 3561
SheepFB_CDS          ------------------------------------------------------------
BovFB_gDNA_5920n     TCCTCTACCTCCAGGTCACTATTCTTGCTTCCTGGTACTTTCCACTCCCTGATCTCAGAA 3580
BovFB_CDS            ------------------------------------------------------------

HumFBens_genDNA      TCACAGAGCTCTGAGCACTTCAGAGATCTTTCTATAGTCCTACATTTGACA--CGTGGAA 3565
HumFBens_CDS         ------------------------------------------------------------
ShFB_genDNA          TCACAAGTTCTGAAATGATGCAGACACCTTTCTCAAGTCCCTCCTTTTACAGATATGGAA 3621
SheepFB_CDS          ------------------------------------------------------------
BovFB_gDNA_5920n     TCATGAGTTCTGAAATGATGCCGAGACCTTTCTCAAGTCCCTCCTTTTACAGATATGGAA 3640
BovFB_CDS            ------------------------------------------------------------

HumFBens_genDNA      ACAGAAGCCAAAGGAGGTCAAGGGACAGCAAGTTAGCAACAAGGGTGGGCTTGAAAACAG 3625
HumFBens_CDS         ------------------------------------------------------------
ShFB_genDNA          GCTGGGGCCAACAGTGGTCAAGAAATGGCAAGTTAGCAGGA--GTTGGGCCTGAAAATGC 3679
SheepFB_CDS          ------------------------------------------------------------
BovFB_gDNA_5920n     GCTGGGGCCAAAAGAGGTCAAGAAATGGCAAGTTAGCAGGA--GTTGGGCCTGAAAATGC 3698
BovFB_CDS            ------------------------------------------------------------

HumFBens_genDNA      CCAGGCCTCTGACAGCTTGATCCCAAGTTCTTTCCCTTTTCAGTCCACCATAGCAGTTTT 3685
HumFBens_CDS         ------------------------------------------------------------
ShFB_genDNA          CCAGGTCTCTGAGTT-----------------------------------------TAGT 3698
SheepFB_CDS          ------------------------------------------------------------
BovFB_gDNA_5920n     CCAGGTCTCTGAGCT-----------------------------------------TAGT 3717
BovFB_CDS            ------------------------------------------------------------

HumFBens_genDNA      CTCCTAACACGAGGAAACAAATACCCGTGGTCTTTCCCTTTCTCCTTTTGGGCCTTTGCT 3745
HumFBens_CDS         ------------------------------------------------------------
ShFB_genDNA          CTCATGATTCCAGTGGTTCCTTGCCTTCCCTTCCCTCTGTTGGGCATCTGCTCCCCACAG 3758
SheepFB_CDS          ------------------------------------------------------------
BovFB_gDNA_5920n     CCCATGATTCCAGTGGTTCCTTGCCTTCCCTTCCCTCTTTTGGGCATCTGCTCCCCACAG 3777
BovFB_CDS            ------------------------------------------------------------

HumFBens_genDNA      CCCCATAGACTCCTACCCAAAAGGCTGCTGCCATTTGGGAATGAAGTGTTCCGAGTTTTC 3805
HumFBens_CDS         ------------------------------------------------------------
ShFB_genDNA          ACACACTGGCGTCCTTCCCCAAGACAGCTGCCATTTGGGAATGAACTGCCCC-AGCTTTT 3817
SheepFB_CDS          ------------------------------------------------------------
BovFB_gDNA_5920n     ACACACTGGCCTCCCTCCGCAAGACAGCTGCCATTTGGAAATGAACTGTCCC-AGCTTTT 3836
BovFB_CDS            ------------------------------------------------------------

                                      EXON 11
HumFBens_genDNA      AGCACATTCTCCTTCTCTGCCAGATGAAAGCCAGTCTCTGAGTCTCTGTGGCATGGTTTG 3865
HumFBens_CDS         ----------------------ATGAAAGCCAGTCTCTGAGTCTCTGTGGCATGGTTTG 1445
ShFB_genDNA          AGCACAATCTCCTTTCTTGCCAGATGAAAGCCGGACACTGGGTCTCTGTGGCATGGTTTG 3877
SheepFB_CDS          ----------------------ATGAAAGCCGGACACTGGGTCTCTGTGGCATGGTTTG 1445
BovFB_gDNA_5920n     AGCACAATCTCCTTTCTTGTCAGATGAAAGCCGGACACTGGGTCTCTGTGGCATGGTTTG 3896
BovFB_CDS            ----------------------ATGAAAGCCGGACACTGGGTCTCTGTGGCATGGTTTG 1445
                                           ********* * * *** ******************
```


HumFBens_genDNA GGAACACAGGAAGGGTACCGATTACCACAAGCAACCATGGCAGGCCAAGATCTCAGTCAT 3925
HumFBens_CDS GGAACACAGGAAGGGTACCGATTACCACAAGCAACCATGGCAGGCCAAGATCTCAGTCAT 1505
ShFB_genDNA GGAGCACAAGGATGGTACTCCCTACCACAAGCAACCATGGCACGCCAAGATCTCAGTCAT 3937
SheepFB_CDS GGAGCACAAGGATGGTACTCCCTACCACAAGCAACCATGGCACGCCAAGATCTCAGTCAT 1505
BovFB_gDNA_5920n GGAGCACAAGGATGGTACTGCCTACCACAAGCAACCGTGGCAGGCCAAGATCTCGGTCAC 3956
BovFB_CDS GGAGCACAAGGATGGTACTGCCTACCACAAGCAACCGTGGCAGGCCAAGATCTCGGTCAC 1505
HumFBens_genDNA TGTAAGCACAGAATCCCAGTAGTGGGGACTTGGGGGAGGTGAGGTCAAGGTGAAATGGGA 3985
HumFBens_CDS T 1506
ShFB_genDNA TGTAAGTGCAGTGTCCTGGTGGTGGGAACTTAGGGGAGGGGAAGTCAGGGTGAAATAAAG 3997
SheepFB_CDS T 1506
BovFB_gDNA_5920n TGTAAGTGCAGTGTCCTGGTGGTGGGAACTTAGAGGAAGGGAGGTCAGGGTGAAATGAAG 4016
BovFB_CDS T 1506
HumFBens_genDNA GTAGGGGAAGGAAAAAATGGCCATAAGAGATGGTGGTTTGTGAAAGTTGAGCTTTCCCTC 4045
HumFBens_CDS
ShFB_genDNA GTCAGAGGAAGGGCAAGGCACTCTTGATCAGACTGTGATATTTGTGAAAGCTGAGCTTTC 4057
SheepFB_CDS
BovFB_gDNA_5920n GTCGGAGGAAGGGCAAGGCACTCTTGATCAGATTGTGAGATTTGTGAAAGCTGAGCTTTC 4076
BovFB_CDS
EXON 12
HumFBens_genDNA TCTACTGTTGTGTCCCCAGCGCCCTTCAAAGGGACACGAGAGCTGTATGGGGGCTGTGGT 4105
HumFBens_CDS CGCCCTTCAAAGGGACACGAGAGCTGTATGGGGGCTGTGGT 1547
ShFB_genDNA CCTCTGCTGCTGCTCCCAGCGCCCTTCGAAGGGGCATGAGAGCTGTATGGGTGCTATCGT 4117
SheepFB_CDS CGCCCTTCGAAGGGGCATGAGAGCTGTATGGGTGCTATCGT 1547
BovFB_gDNA_5920n CCTCTGCTGCTGTCCCCAGCGCCCTTCAAAGGGGCATGAGAGCTGTATGGGTGCTATTGT 4136
BovFB_CDS CGCCCTTCAAAGGGGCATGAGAGCTGTATGGGTGCTATTGT 1547
HumFBens_genDNA GTCTGAGTACTTTGTGCTGACAGCAGCACATTGTTTCACTGTGGATGACAAGGAACACTC 4165
HumFBens_CDS GTCTGAGTACTTTGTGCTGACAGCAGCACATTGTTTCACTGTGGATGACAAGGAACACTC 1607
ShFB_genDNA GTCTGAGTACTTCGTGCTGACGGCTGCACACTGTTTCACTGTGGATGATGAGAAACACTC 4177
SheepFB_CDS GTCTGAGTACTTCGTGCTGACGGCTGCACACTGTTTCACTGTGGATGATGAGAAACACTC 1607
BovFB_gDNA_5920n GTCTGAGTACTTTGTGCTGACGGCTGCACACTGTTTCACTGTGGATGATGAGAAACACTC 4196
BovFB_CDS GTCTGAGTACTTTGTGCTGACGGCTGCACACTGTTTCACTGTGGATGATGAGAAACACTC 1607
HumFBens_genDNA AATCAAGGTCAGCGTAGGTAAGGATGCAACTGAAG-GTCCTGGGCTGCACCTATGCTCTC 4224
HumFBens_CDS AATCAAGGTCAGCGTAG 1624
ShFB_genDNA AATCAAGGTCAGCGTGGGTAAGGATGCAATCCATCCATTCTGAGCTTCTGCAGTGCCCTC 4237
T
SheepFB_CDS AATCAAGGTCAGCGTGG 1624
BovFB_gDNA_5920n AATCAAGGTCAGCCTGGGTAAGGATGCAACCCATCAATTCTGAGCTTCTGCGGTGCCCTC 4256
BovFB_CDS AATCAAGGTCAGCCTGG 1624
HumFBens_genDNA CAGGCAACACCTCCCACTTTCTACAGATCCTACACTCCACCCATCCTCAATGCAGCCCCA 4284
HumFBens_CDS
ShFB_genDNA TGGGCAACACTGCC-------TCCAGGGCCCACTTCCCAACTGTCCTCATCGCAGCCCT- 4289
SheepFB_CDS
BovFB_gDNA_5920n TGGGCCACACTGCC-------TCCAGGGCCCACTTTCCACCTGTCCTCATCGCAGCCCT- 4308
BovFB_CDS
HumFBens_genDNA TTCCTTGCACCCCAGACCAGTCAGGGATGGGGGAAGACGTGAAGTTAGGAATGACACGGG 4344
HumFBens_CDS
ShFB_genDNA TCCCTTGCATGCTGGACCAGTTAAGGATGGGGAAGATTTGGGTTAGGGATGACATGCAGG 4349
SheepFB_CDS
BovFB_gDNA_5920n TCCCTTGCATGCTGGACAAGTTAGGGATGGGGAAGATTTGGGTTAGGGATGACATGCAGG 4368
BovFB_CDS
EXON 13
HumFBens_genDNA GCCAGAGGC----AGGAAGCTGCCCACAAAGAGGTGGTACCTACTCTCCTACTTCAGGAG 4400
HumFBens_CDS GAG 1627
ShFB_genDNA AAGGAGGCTGCCTACACAGACATCGGTCATCTAGAGGGACTCCCTCTCCTACTTCAGGAG 4409
SheepFB_CDS GAG 1627
BovFB_gDNA_5920n CAGGAGGCTGCCTACACAGATTTCAGTGGTCTAGAGGAACCCCTTCTTGTACTTCAGGAG 4428
BovFB_CDS GAG 1627
HumFBens_genDNA GGGAGAAGCGGGACCTGGAGATAGAAGTAGTCCTATTTCACCCCAACTACAACATTAATG 4460
HumFBens_CDS GGGAGAAGCGGGACCTGGAGATAGAAGTAGTCCTATTTCACCCCAACTACAACATTAATG 1687
ShFB_genDNA GAAAGAGGCAGGAGTGGGAGGTAGAAGAAGTCCTATTTCATCCAAACTACAACCTCAATG 4469
SheepFB_CDS GAAAGAGGCAGGAGTGGGAGGTAGAAGAAGTCCTATTTCATCCAAACTACAACCTCAATG 1687
C [AAG/K -> ACG/T]
BovFB_gDNA_5920n GACAGAGGAAGGAGTGGGAGGTAAAAGAAATCCTATTTCATCCGAAGTACGACCTCAATG 4488
BovFB_CDS GACAGAGGAAGGAGTGGGAGGTAAAAGAAATCCTATTTCATCCGAAGTACGACCTCAATG 1687
HumFBens_genDNA GGAAAAAAGAAGCAGGAATTCCTGAATTTTATGACTATGACGTTGCCCTGATCAAGCTCA 4520
HumFBens_CDS GGAAAAAAGAAGCAGGAATTCCTGAATTTTATGACTATGACGTTGCCCTGATCAAGCTCA 1747
ShFB_genDNA CAAAAAAAGCAAAAGGCATTCCTGAGTTTTATGACTATGACGTGGCCCTCATCAAGCTCA 4529
SheepFB_CDS CAAAAAAAGCAAAAGGCATTCCTGAGTTTTATGACTATGACGTGGCCCTCATCAAGCTCA 1747
BovFB_gDNA_5920n CAAAAAAAGCAAAAGGCATTCCTGAGTTTTATGACTATGACGTGGCCCTCGTCAGGCTCA 4548
BovFB_CDS CAAAAAAAGCAAAAGGCATTCCTGAGTTTTATGACTATGACGTGGCCCTCGTCAGGCTCA 1747

```
                          ******* *  *** ******** ***************** *****  *** *****

HumFBens_genDNA       AGAATAAGCTGAAATATGGCCAGACTATCAGGTGAGAGCGTCCAGATCCCTGAGGAAAGG 4580
HumFBens_CDS          AGAATAAGCTGAAATATGGCCAGACTATCAG---------------------------- 1778
ShFB_genDNA           CAAAGAAGCTCAAGTATGAGACCACCATCAGGTGAGCCATCTGGATTCCGAGAGAAAAGG 4589
SheepFB_CDS           CAAAGAAGCTCAAGTATGAGACCACCATCAG---------------------------- 1778
BovFB_gDNA_5920n      AGGAAAAGCTCAAGTATGAGACCACCATCAGGTGAGCCATCTGGATTCTGAGAGAAAAGG 4608
BovFB_CDS             AGGAAAAGCTCAAGTATGAGACCACCATCAG---------------------------- 1778
                         * ***** ** ****     ** *****

HumFBens_genDNA       CTGGGAAAGGCTGGA---------GGACTGGGGTGAGGAGCAGGCCTGGTTTGCTGTTCT 4631
HumFBens_CDS          ------------------------------------------------------------
ShFB_genDNA           CTGGGAGAGGCAGAACTGAGACGGGAGCAGGCCCAGGGTTCACGATCCTTGAATTTCCCC 4649
SheepFB_CDS           ------------------------------------------------------------
BovFB_gDNA_5920n      CTGGGAGAGGCAGAACTGAGACAGTAGCAGGCCCAGGGTTCACGATCCTTGAATTTCCCC 4668
BovFB_CDS             ------------------------------------------------------------

                                    EXON 14
HumFBens_genDNA       CCTTGTCCTTTATAGGCCCATTTGTCTCCCCTGCACCGAGGGAACAACTCGAGCTTTGAG 4691
HumFBens_CDS          ---------------GCCCATTTGTCTCCCCTGCACCGAGGGAACAACTCGAGCTTTGAG 1823
ShFB_genDNA           ATCTCTCCTCAACAGGCCCATTTGTCTCCCCTGCACTGAGGGATCGATTCAAGCCTTGAG 4709
SheepFB_CDS           ---------------GCCCATTTGTCTCCCCTGCACTGAGGGATCGATTCAAGCCTTGAG 1823
BovFB_gDNA_5920n      ATCTCTCCTCAACAGGCCCATTTGTCTCCCCTGCACTGAGGGATCGATTCAAGCCTTGAG 4728
BovFB_CDS             ---------------GCCCATTTGTCTCCCCTGCACTGAGGGATCGATTCAAGCCTTGAG 1823
                                     ******************** ****** * * ** *** *****

HumFBens_genDNA       GCTTCCTCCAACTACCACTTGCCAGCAACAAAGTAAGACATACTTGGCAAGAGGA----T 4747
HumFBens_CDS          GCTTCCTCCAACTACCACTTGCCAGCAACAAA---------------------------- 1855
ShFB_genDNA           GCTTCCAAGATCAACTACGTGCCAGCAACAGAGTAAGACACGTTCAGGGAGAGCGGCT-T 4768
SheepFB_CDS           GCTTCCAAGATCAACTACGTGCCAGCAACAGA---------------------------- 1855
BovFB_gDNA_5920n      GCTTCCAAGGTCAACCACATGCCAGCAACAGAGTAAGAGACATTCAGGGAGAGCAGCTTG 4788
BovFB_CDS             GCTTCCAAGGTCAACCACATGCCAGCAACAGA---------------------------- 1855
                      ******     * ** ** ********** *

HumFBens_genDNA       AAGGATGAGATCCCAAGAGACAAGTGGGGCATGAGAGGGAGGTGCAATAGGAAGAGATGA 4807
HumFBens_CDS          ------------------------------------------------------------
ShFB_genDNA           GTAAGGGTGAAGTTCTGAGACAAGGAAGGGACGAGGGGGCG-TAGAGGGAGAACAGGTGA 4827
SheepFB_CDS           ------------------------------------------------------------
BovFB_gDNA_5920n      TTAAGGGTGAAGTTTTGAGACAAGGAAGGGATGAGGGGGAG------GGAGAGCAGGTCA 4842
BovFB_CDS             ------------------------------------------------------------

HumFBens_genDNA       TGCCTGGCCCAGAACCTAGCTCTAGAAGGGCTTAGGGGACATCTACTGAGTGACAAAGGC 4867
HumFBens_CDS          ------------------------------------------------------------
ShFB_genDNA           TGCCCTGCCCAGCCGCCAGCTCACAGGAAGGAGTGGAGGGCCACTTGATGACAAGCATGA 4887
SheepFB_CDS           ------------------------------------------------------------
BovFB_gDNA_5920n      TGCCCTGCCCAGCCGCCAGCTCACGGGAAGGAGTGGAGGGCCACTCGATGACAAGCACGA 4902
BovFB_CDS             ------------------------------------------------------------

                                                                          EXON 15
HumFBens_genDNA       AATGGGGAGATGACAGTGGTGGGAGCAGCTGAAGTGAC--GCAGTCTATTCGTCCAGAGG 4925
HumFBens_CDS          ------------------------------------------------------------AGG 1858
ShFB_genDNA           GGCGGGGAGGGGAGGGGAGGGGAGAGTGCTGCAATGACCGGCTATCCATCTGTAAAGTGC 4947
SheepFB_CDS           ------------------------------------------------------------TGC 1858
BovFB_gDNA_5920n      GGCGGGGAGGGGAGGGAGAGTGCTGCAATGACCCAGTC----TATCCATCTGTAAAGTGC 4958
BovFB_CDS             ------------------------------------------------------------TGC 1858
                                                                                    *
```


HumFBens_genDNA AAGAGCTGCTCCCTGCACAGGATATCAAAGCTCTGTTTGTGTCTGAGGAGGAGAAAAAGC 4985
HumFBens_CDS AAGAGCTGCTCCCTGCACAGGATATCAAAGCTCTGTTTGTGTCTGAGGAGGAGAAAAAGC 1918
ShFB_genDNA AAGAGCTACTCCCTGCCAAGGACATCGAAGCTCTGTTTGTGTCTGAGTCTAAGAAGACCC 5007
SheepFB_CDS AAGAGCTACTCCCTGCCAAGGACATCGAAGCTCTGTTTGTGTCTGAGTCTAAGAAGACCC 1918
BovFB_gDNA_5920n AAGAGCTACTCCCTGCAAAGGACATTGAAGCTCTGTTTGTGTCTGAGTCTAAGAAGACCC 5018
BovFB_CDS AAGAGCTACTCCCTGCAAAGGACATTGAAGCTCTGTTTGTGTCTGAGTCTAAGAAGACCC 1918
HumFBens_genDNA TGACTCGGAAGGAGGTCTACATCAAGAATGGGGATAAGGTGAGAAACGGGCATCCTAAGG 5045
HumFBens_CDS TGACTCGGAAGGAGGTCTACATCAAGAATGGGGATAAG---------- 1956
ShFB_genDNA TGACTCGGAAGGCAGTCTACATCAAAAATGGGGACAAGGTGAGGAAGGTGGGATCCT--G 5065
SheepFB_CDS TGACTCGGAAGGCAGTCTACATCAAAAATGGGGACAAG---------- 1956
BovFB_gDNA_5920n TGACTCGGAAGGCAGTCTACATCAAGAATGGGGACAAGGTGAGGAATGTGGGATCCT--G 5076
BovFB_CDS TGACTCGGAAGGCAGTCTACATCAAGAATGGGGACAAG---------- 1956
HumFBens_genDNA AGGCACTCTAGGCCCCAATCCTTCCTAAGCCACTTCTGTTCATTACTTCTCCATGC---- 5101
HumFBens_CDS ---------- 5101
ShFB_genDNA AGGTCCTTTAGGCCCCAGTTCTTCCTAAGCAAGCTTTGTTCCTCATGCCTCTCTCCACAC 5125
SheepFB_CDS ---------- 5125
BovFB_gDNA_5920n AGGTCCTTTAGGCCCCAGTCCTTCCTAAGCAAGCTTTGTTCCTCACACCTCTCTCCACAC 5136
BovFB_CDS ---------- 5136
EXON 16
HumFBens_genDNA -TTCCCACCTCCCCTACAGAAAGGCAGCTGTGAGAGAGATGCTCAATATGCCCCAGGCTA 5160
HumFBens_CDS ----------AAAGGCAGCTGTGAGAGAGATGCTCAATATGCCCCAGGCTA 1997
ShFB_genDNA TTCCCCATCTCACCTACAGAAAGCCAGCTGTGAGAGAGACGCTCTGCATGCCCCAGGTTA 5185
SheepFB_CDS ----------AAAGCCAGCTGTGAGAGAGACGCTCTGCATGCCCCAGGTTA 1997
BovFB_gDNA_5920n TTCCCTATCTCACCTACAGAAAGCCAGCTGTGAGAGAGATGCTCTACGCGCCCCAGGTTA 5196
BovFB_CDS ----------AAAGCCAGCTGTGAGAGAGATGCTCTACGCGCCCCAGGTTA 1997
HumFBens_genDNA TGACAAAGTCAAGGACATCTCAGAGGTGGTCACCCCTCGGTTCCTTTGTACTGGAGGAGT 5220
HumFBens_CDS TGACAAAGTCAAGGACATCTCAGAGGTGGTCACCCCTCGGTTCCTTTGTACTGGAGGAGT 2057
ShFB_genDNA TGACAAAGTCAAGGACGTCTCTGAGGTAGTCACCCCCAGGTTCCTCTGCACTGGAGGTGT 5245
SheepFB_CDS TGACAAAGTCAAGGACGTCTCTGAGGTAGTCACCCCCAGGTTCCTCTGCACTGGAGGTGT 2057
BovFB_gDNA_5920n TGAAAAAGTCAAGGACGTCTCTGAGGTAGTCACCCCCAGGTTCCTCTGCACTGGAGGTGT 5256
BovFB_CDS TGAAAAAGTCAAGGACGTCTCTGAGGTAGTCACCCCCAGGTTCCTCTGCACTGGAGGTGT 2057
HumFBens_genDNA GAGTCCCTATGCTGACCCCAATACTTGCAGAGGTGAGAGAATGCTCTTTGGTTGTGCTAC 5280
HumFBens_CDS GAGTCCCTATGCTGACCCCAATACTTGCAGAG---------- 2089
ShFB_genDNA GGCTCCCTACGCTGACCCCAACACTTGCAAAGGTGAGAGAAGGCTCTTTGGTTGTGATGT 5305
SheepFB_CDS GGCTCCCTACGCTGACCCCAACACTTGCAAAG---------- 2089
BovFB_gDNA_5920n GGATCCCTACGCTGACCCCAACACTTGCAAAGGTGAGAGAAGGCTCTTTGGTTGTGATGT 5316
BovFB_CDS GGATCCCTACGCTGACCCCAACACTTGCAAAG---------- 2089
EXON 17
HumFBens_genDNA AAGTGCCCAAGGCCCAACAGTCCTTTTCTCTACAGCTTCTCCTCTCCTTGCAGGTGATTC 5340
HumFBens_CDS ----------GTGATTC 2096
ShFB_genDNA GAGTTCTGAGGCCTAAGT-----CTTTGCCTACAGCATCTCCTCCCCT-GCAGGTGATTC 5359
SheepFB_CDS ----------GTGATTC 2096
BovFB_gDNA_5920n GAGTTCTGAGGCCTAAGT-----CTTTTCCTACAGCATCTCCTTCCCT-GCAGGTGATTC 5370
BovFB_CDS ----------GTGATTC 2096
HumFBens_genDNA TGGCGGCCCCTTGATAGTTCACAAGAGAAGTCGTTTCATTCAAGTGAGTCCTCCCTTTCC 5400
HumFBens_CDS TGGCGGCCCCTTGATAGTTCACAAGAGAAGTCGTTTCATTCAA---------- 2139
ShFB_genDNA TGGTGGCCCCCTGATTATTCACAAGAGGAGCCGCTTCATTCAAGTGAGTCTTCTCTTTCC 5419
SheepFB_CDS TGGTGGCCCCCTGATTATTCACAAGAGGAGCCGCTTCATTCAA---------- 2139
BovFB_gDNA_5920n TGGTGGCCCCCTGATTATTCACAAGAGGAGTCGCTTCATTCAAGTGAGTCTTCTCTTTCC 5430
BovFB_CDS TGGTGGCCCCCTGATTATTCACAAGAGGAGTCGCTTCATTCAA---------- 2139
HumFBens_genDNA TATCT--GGGGAGATGCCAAGTGGTCAGCATGGGCCCCAAAGCAGGAAAGCTCAATGCAT 5458
HumFBens_CDS ---------- 5458
ShFB_genDNA TCTCTCTGGAGAGATGCTGAGTGGTCAGCATGGGCCCCAAAACAAGAAAGCTCACTGCAT 5479
SheepFB_CDS ---------- 5479
BovFB_gDNA_5920n TTTCTCTGGAGAGATGCTGAGTGGTCAGCGTGGGCCCCAAAACAGGAAAGCTCACTGCAT 5490
BovFB_CDS ---------- 5490
HumFBens_genDNA GTGGCTAGTAATTCGAGGTA-----GGCAGAGCCTGCCTCACCTT---------- 5498
HumFBens_CDS ---------- 5498
ShFB_genDNA GTGGCTGAAAAGAGGTGGGGTGTGGGGCAGAGCCTGCCTCACCTTCAGATTCTTCCTTGA 5539
C
SheepFB_CDS ---------- 5539
BovFB_gDNA_5920n GTGGCTGAAAAGAGGTGGGGGGTGGGGCAGAGCCTGCCTCACCTTCAGACTCTTCCTTGA 5550
BovFB_CDS ---------- 5550
HumFBens_genDNA ------------AGGACCGCATGTCTTGCCTGCGTGTGTCAAGAACGAGGCTGAGCTGGG 5546
HumFBens_CDS ---------- 5546
ShFB_genDNA ACTGGCCAATTTAGGGCCATGTGTCTGGTCTGTGTGC-CAGGGGACAGAACTGAACTGGG 5598
SheepFB_CDS ---------- 5598
BovFB_gDNA_5920n ACTGGCCAATTCAGGGCCATGTGTCTGGTCTGTGTGC-CAGGGGACAGAACTGAACTGGG 5609
BovFB_CDS ---------- 5609
HumFBens_genDNA TCCCTAGTCTGATTCCT-----TTAGGTCAGCTAAGACACAAGCAGGAACAGCCATGCTT 5601

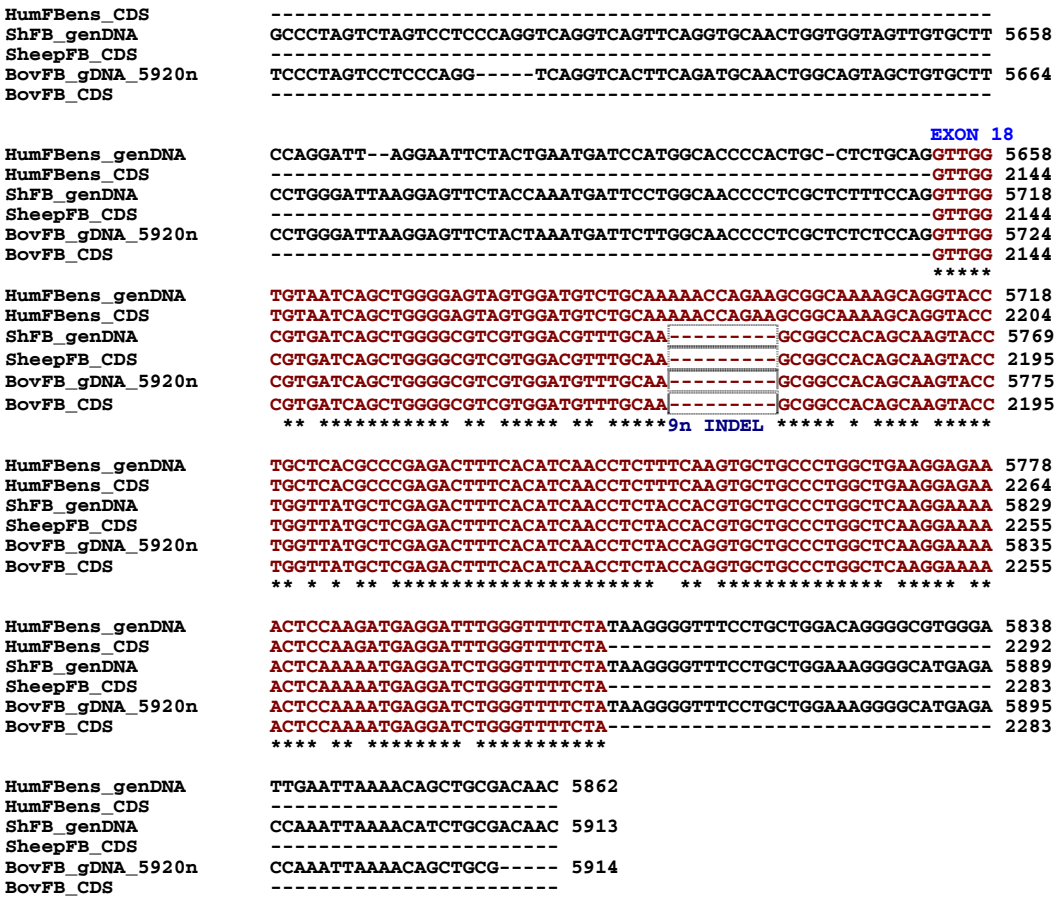
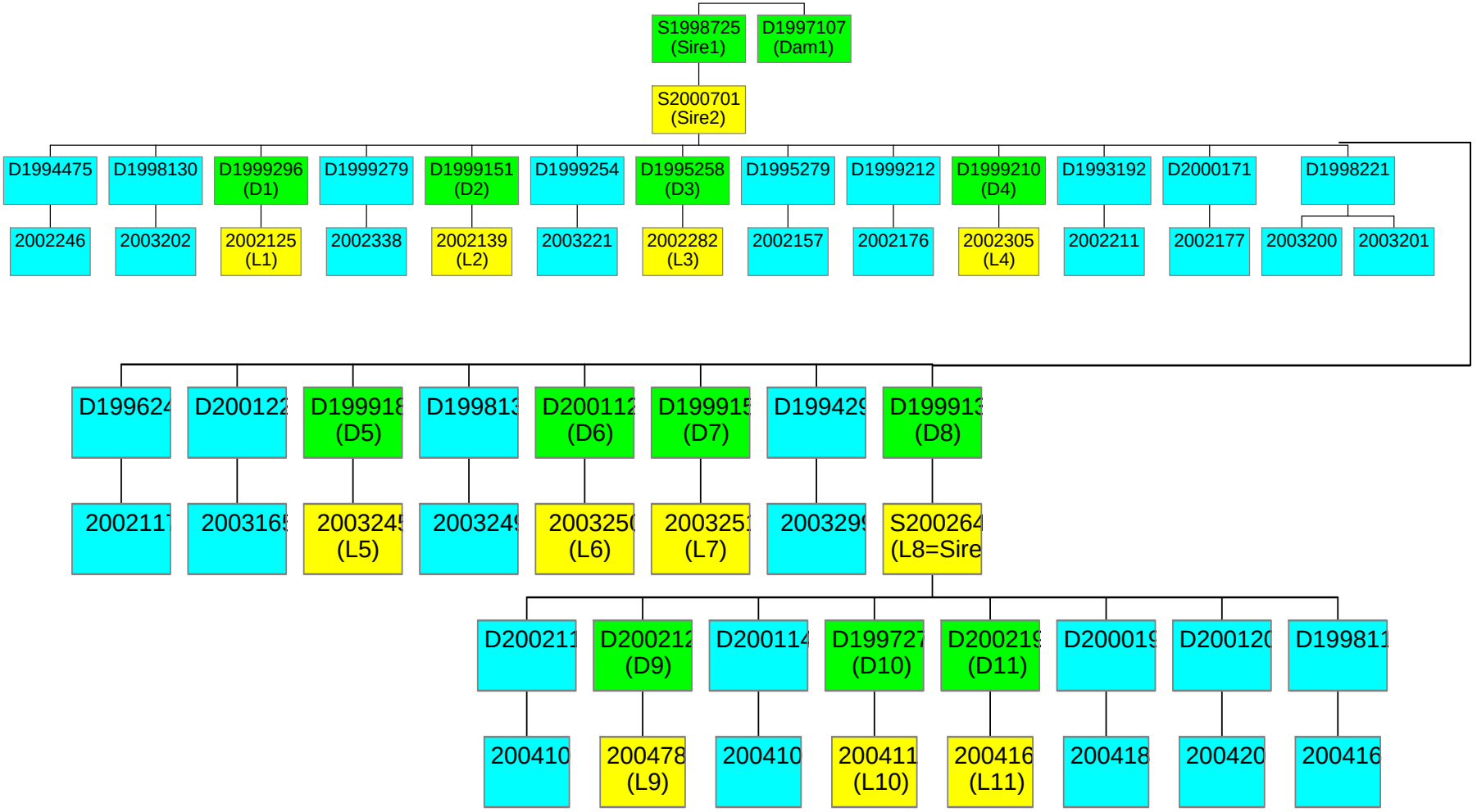


Figure 4.1: Complete genomic and CDS sequences of sheep and cattle CFB aligned to human CFB. SNPs are identified in green font and exonic sequences in red font. The 9bp deletion characteristic of sheep and cattle CFB is present in exon 18.

Sheep Family Tree



S: Sire
D:Dam

Animal ID	LTAS2	TNFa1	BAT2S1	APOMS1	BAT5S1	G6bS1	MSH5S4	C2S4	CFBS2	TNXBS2	BfMs (SKIV2L)
1999296	C	C	T	C	C	C	C	C	T	T	18
1998725	T	C	C	C	C	C	C	C	C	C	18
1999210	T	C	C	C	C	C	C	C	T	C	18
2001123	T	C	C	C	C	C	C	C	T	T	18
1999296	T	C	C	C	C	C	C	T	C	T	18
2002123	T	C	C	C	C	T	C	C	T	C	18
1999151	T	C	C	C	C	T	C	C	T	T	18
1999136	T	C	C	C	C	T	C	C	T	T	18
1999210	T	C	C	C	C	T	C	T	C	C	18
1995258	T	C	T	C	C	C	C	C	T	T	18
1999187	T	C	T	C	C	C	C	T	C	C	18
2002191	T	C	T	T	C	C	T	C	T	C	17
1999155	T	C	T	T	C	C	T	C	T	T	18
1997272	T	T	T	C	C	C	C	C	C	C	21
1999136	T	T	T	C	T	C	C	C	C	C	17
1999187	T	T	T	C	T	C	C	C	C	T	23
1999155	T	T	T	C	T	C	C	C	C	T	23
1997107	T	T	T	C	T	C	C	C	T	C	18
1999151	T	T	T	C	T	C	C	C	T	T	19
2001123	T	T	T	C	T	C	C	C	T	T	19
2002191	T	T	T	C	T	C	C	T	C	C	18
1997101	T	T	T	C	T	C	C	T	C	T	17
1997272	T	T	T	C	T	C	C	T	C	T	17
1998725	T	T	T	C	T	C	C	T	T	T	19
1995258	T	T	T	T	T	C	C	T	C	C	17
2002123	T	T	T	T	T	C	C	T	C	C	17

Figure 5.4: Deduced haplotypes for the 25 animals from a three generation half-sibling sheep family. Loci are ordered according to the physical map described in Chapter 3 except that the BfMs microsatellite locus (within the SKIV2L gene) is between the BFS2 and TNXBS2 loci. D for Dam, L for Lamb. Haplotypes have been sorted in order of similarity. The four pairs of identical haplotypes are shown in the red font.

Family	Locus	LTAS2		TNFa1		TNFa3		TNFa5		BAT2S1		APOMS1		BAT5S1		G6bS1		MSH5S4		C2S4		BFS2		TNXBS2		Micro
SIRE 1	1998725	T	T	C	T	C	C	C	C	C	T	C	C	C	T	C	C	C	C	C	T	C	T	C	T	18,19
Dam 1	1997107	T	T	T	T	C	C	C	T	T	T	C	C	T	T	C	C	C	C	T	C	C	T	T	C	17,18
SIRE 2	2000701	T	T	T	T	C	C	C	C	T	T	C	C	T	T	C	C	C	C	T	C	C	T	T	T	17,19
D1	1999296	C	T	C	C	C	T	C	C	T	C	C	C	C	C	C	C	C	C	C	T	T	C	T	T	18,18
L1	2002125	C	T	C	T	C	T	C	C	T	T	C	C	C	T	C	C	C	C	C	C	T	T	T	T	18,19
D2	1999151	T	T	T	C	C	C	C	C	T	C	C	C	T	C	C	T	C	C	C	C	T	T	T	T	18,19
L2	2002139	T	T	T	C	C	C	C	C	T	C	C	C	T	C	C	T	C	C	T	C	C	T	T	T	17,18
D3	1995258	T	T	C	T	C	C	C	T	T	T	C	T	C	T	C	C	C	C	C	T	T	C	T	C	17,18
L3	2002282	T	T	C	T	C	C	C	C	T	T	C	C	C	T	C	C	C	C	C	C	T	T	T	T	18,19
D4	1999210	T	T	C	C	C	C	C	C	C	C	C	C	C	C	C	T	C	C	T	C	C	T	C	C	18,18
L4	2002305	T	T	C	T	C	C	C	C	C	T	C	C	C	T	C	C	C	C	T	C	C	T	C	T	18,19
D5	1999187	T	T	T	C	C	C	T	C	T	T	C	C	T	C	C	C	C	C	C	T	C	C	T	C	18,23
L5	2003245	T	T	T	C	C	C	C	C	T	T	C	C	T	C	C	C	C	C	T	T	C	C	T	C	17,18
D6	2001123	T	T	C	T	C	C	C	C	C	T	C	C	C	T	C	C	C	C	C	C	T	T	T	T	18,19
L6	2003250	T	T	C	T	C	C	C	C	C	T	C	C	C	T	C	C	C	C	C	C	T	T	T	T	18,19
D7	1999155	T	T	C	T	C	T	C	T	T	T	T	C	C	T	C	C	T	C	C	C	T	C	T	T	18,23
L7	2003251	T	T	C	T	C	T	C	C	T	T	T	C	C	T	C	C	T	C	C	C	T	T	T	T	18,19
D8	1999136	T	T	T	C	C	C	T	C	T	C	C	C	T	C	C	T	C	C	C	C	C	T	C	T	17,18
SIRE(L8)	2002645	T	T	T	C	C	C	C	C	T	C	C	C	T	C	C	T	C	C	T	C	C	T	T	T	17,18
D9	2002123	T	T	C	T	C	C	C	T	C	T	C	T	C	T	T	C	C	C	C	T	T	C	C	C	17,18
L9	2004782	T	T	T	T	C	C	C	T	T	T	C	C	T	T	C	C	C	C	T	T	C	C	T	C	17,17
D10	1997272	T	T	T	T	C	C	C	C	T	T	C	C	T	C	C	C	C	C	T	C	C	C	T	C	17,21
L10	2004110	T	T	T	T	C	C	C	C	T	T	C	C	T	C	C	C	C	C	T	C	C	C	T	C	17,21
D11	2002191	T	T	T	C	C	T	C	T	T	T	C	T	T	C	C	C	C	T	T	C	C	T	C	C	17,18
L11	2004169	T	T	T	C	C	T	C	C	T	T	C	T	T	C	C	C	C	T	T	C	C	T	T	C	17,18

Table 5.3 Genotypes for the 25 animals from a three generation half-sibling sheep family. Loci are ordered according to the physical map described in Chapter 3 except that the BfMs microsatellite locus (within the SKIV2L gene) is between the BFS2 and TNXBS2 loci. D for Dam, L for Lamb.

Sheep ID	LTAS2	TNFaS1	BAT2S1	APOMS1	BAT5S1	G6bS1	MSH5S4	C2S4	CFBS2	TNXBS2	BfMs STR
Panel of 58 unrelated sheep											
3c	T/T	C/C	T/T	C/C	C/C	C/C	C/C	C/C	T/T	C/T	18/18
7c	T/T	C/C	C/C	C/C	C/C	T/T	C/C	C/C	T/T	C/C	18/18
8c	T/T	T/T	T/T	C/C	C/T	C/T	C/C	C/C	T/T	C/T	18/18
13c	C/T	C/C	C/T	C/C	C/C	C/T	C/C	C/C	T/T	C/T	18/18
14c	T/T	C/C	C/C	C/C	C/C	C/C	C/T	C/C	C/T	C/T	15/18
16c	T/T	C/C	C/C	C/C	C/C	C/T	C/C	C/C	T/T	C/T	18/18
21c	T/T	C/T	T/T	C/C	C/C	C/T	C/C	C/C	T/T	T/T	18/18
22c	T/T	C/T	C/T	C/C	C/T	C/C	C/T	C/C	C/T	C/T	15/18
30c	T/T	C/C	T/T	C/C	C/C	C/C	C/C	C/C	T/T	C/T	18/18
56c	T/T	C/T	C/T	C/T	C/C	T/T	C/C	C/C	T/T	C/T	18/18
502c	T/T	C/C	C/C	C/C	C/C	C/T	C/T	C/C	T/T	C/T	18/18
602c	T/T	T/T	C/T	C/C	C/T	C/C	C/C	C/C	C/T	C/T	18/22
618c	T/T	C/T	T/T	C/T	C/T	C/C	C/T	C/C	C/T	C/T	18/24
619c	T/T	C/C	T/T	T/T	C/C	C/C	T/T	C/C	T/T	C/C	18/18
621c	T/T	C/T	T/T	C/T	C/T	C/C	C/T	C/C	T/T	C/C	18/18
625c	T/T	T/T	T/T	C/C	T/T	C/C	C/C	C/C	C/T	T/T	19/24
632c	T/T	C/C	C/T	C/C	C/C	C/T	C/C	C/T	C/T	C/C	18/18
646c	T/T	C/T	T/T	C/T	C/T	C/C	C/T	C/C	C/T	C/C	15/16
649c	T/T	C/T	T/T	C/T	C/T	C/C	C/T	C/C	T/T	C/C	18/18
652c	T/T	T/T	T/T	C/C	T/T	C/C	C/C	C/C	T/T	T/T	18/19
659c	T/T	C/T	T/T	C/T	C/T	C/C	C/T	C/C	T/T	C/C	18/18
663c	T/T	T/T	T/T	C/C	C/T	C/T	C/C	C/C	T/T	C/T	18/19
672c	T/T	C/T	C/T	C/C	C/T	C/C	C/T	C/C	T/T	C/T	18/18
680c	T/T	T/T	T/T	C/C	T/T	C/C	C/C	C/C	T/T	C/C	16/18
684c	C/T	C/C	C/T	C/T	C/C	C/C	C/C	C/T	C/T	C/T	18/18
687c	T/T	C/C	T/T	C/T	C/C	C/C	C/T	C/C	C/T	C/C	15/18
688c	T/T	T/T	T/T	C/C	T/T	C/C	C/C	C/C	T/T	C/T	18/19
120s	T/T	C/T	T/T	C/C	C/T	C/C	C/C	C/C	T/T	T/T	18/19
125s	C/T	C/T	T/T	C/C	C/T	C/C	C/C	C/C	T/T	T/T	18/19
127s	T/T	C/T	C/T	C/C	C/T	C/T	C/C	C/T	C/T	C/T	17/18
145s	T/T	C/T	T/T	C/T	C/T	C/C	C/T	C/C	T/T	C/T	18/18
156s	T/T	C/C	T/T	C/T	C/C	C/T	C/T	C/C	T/T	C/T	17/18
161s	T/T	T/T	T/T	C/C	T/T	C/C	C/C	C/C	T/T	C/T	18/18
182s	T/T	T/T	T/T	C/C	T/T	C/C	C/C	C/C	C/T	C/C	15/18
184s	T/T	C/T	C/T	C/C	C/T	C/T	C/C	C/T	C/T	C/T	17/18
194s	T/T	C/T	T/T	C/C	C/C	T/T	C/C	C/C	T/T	T/T	17/19
197s	C/C	C/C	C/T	C/T	C/C	C/T	C/T	C/C	T/T	C/C	18/18

207s	T/T	C/T	C/T	C/C	C/C	T/T	C/C	C/C	T/T	C/T	18/19
213s	T/T	C/T	T/T	C/T	C/C	C/C	C/T	C/C	C/C	C/C	15/22
215s	T/T	T/T	T/T	C/C	T/T	C/C	C/C	C/C	C/T	C/C	15/18
218s	T/T	C/T	T/T	C/T	C/T	C/C	C/T	C/C	T/T	C/T	18/18
236s	T/T	C/T	T/T	C/C	C/T	C/T	C/C	C/C	T/T	C/T	17/18
243s	T/T	C/T	T/T	C/T	C/T	C/C	C/T	C/T	C/T	C/T	17/18
250s	T/T	C/T	C/T	C/C	C/T	C/C	C/C	C/C	T/T	T/T	18/19
271s	T/T	C/C	T/T	T/T	C/C	C/C	T/T	C/C	C/C	C/C	15/15
284s	C/T	C/T	T/T	C/C	C/T	C/C	C/C	C/C	T/T	T/T	16/18
291s	T/T	C/C	T/T	C/T	C/C	C/T	C/T	C/C	T/T	C/C	17/18
304s	T/T	C/C	C/C	C/C	C/C	C/T	C/C	C/C	C/T	T/T	18/18
308s	T/T	C/C	C/C	C/C	C/C	C/T	C/C	C/C	T/T	C/T	18/18
316s	C/T	C/T	T/T	C/C	C/T	C/C	C/C	C/C	T/T	T/T	18/19
502s	T/T	C/T	C/T	C/C	C/C	C/T	C/T	C/C	T/T	T/T	18/18
506s	T/T	C/C	T/T	C/T	C/C	C/C	C/T	C/C	T/T	C/C	18/18
511s	T/T	C/C	C/T	C/C	C/C	C/T	C/C	C/C	T/T	C/C	18/18
512s	T/T	C/C	C/C	C/C	C/C	C/T	C/C	C/C	T/T	C/T	18/18
516s	T/T	C/C	C/T	C/C	C/C	C/T	C/C	C/C	T/T	C/C	18/18
519s	T/T	C/C	C/T	C/C	C/C	C/T	C/C	C/C	T/T	C/C	18/18
521s	T/T	C/C	C/T	C/C	C/C	C/C	C/T	C/C	C/T	C/C	15/18
523s	T/T	C/C	T/T	C/C	C/C	C/T	C/C	C/C	T/T	C/C	18/18
13 parental members of half sibling family											
1998725	T/T	C/T	C/T	C/C	C/T	C/C	C/C	C/T	C/T	C/T	18/19
1997107	T/T	T/T	T/T	C/C	T/T	C/C	C/C	C/T	C/T	C/T	17/18
1999296	C/T	C/C	C/T	C/C	C/C	C/C	C/C	C/T	C/T	T/T	18/18
1999151	T/T	C/T	C/T	C/C	C/T	C/T	C/C	C/C	T/T	T/T	18/19
1995258	T/T	C/T	T/T	C/T	C/T	C/C	C/C	C/T	C/T	C/T	17/18
1999210	T/T	C/C	C/C	C/C	C/C	C/T	C/C	C/T	C/T	C/C	18/18
1999187	T/T	C/T	T/T	C/C	C/T	C/C	C/C	C/T	C/C	C/T	18/23
2001123	T/T	C/T	C/T	C/C	C/T	C/C	C/C	C/C	T/T	T/T	18/19
1999155	T/T	C/T	T/T	C/T	C/T	C/C	C/T	C/C	C/T	T/T	18/23
1999136	T/T	C/T	C/T	C/C	C/T	C/T	C/C	C/C	C/T	C/T	17/18
2002123	T/T	C/T	C/T	C/T	C/T	C/T	C/C	C/T	C/T	C/C	17/18
1997272	T/T	T/T	T/T	C/C	C/T	C/C	C/C	C/T	C/C	C/T	17/21
2002191	T/T	C/T	T/T	C/T	C/T	C/C	C/T	C/T	C/T	C/C	17/18

Table 5.4: Genotypes for the panel of 58 unrelated sheep together with the 13 parental animals from the half sibling pedigree. Loci are ordered as for Table 5.3 but only one TNF SNP is included

Haplotype	LTAS2	TNFaS1	BAT2S1	APOMS1	BAT5S1	G6bS1	MSH5S4	C2S4	CFBS2	TNXBS2	Frequency	Cum Freq
1	T	A	T	G	A	C	G	C	T	A	0.1445	0.1445
2	T	A	T	G	A	C	G	C	T	G	0.1175	0.262
3	T	G	C	G	G	T	G	C	T	G	0.1146	0.3766
4	T	G	T	A	G	C	A	C	T	G	0.079	0.4556
5	T	G	T	G	G	C	G	C	T	G	0.0694	0.525
6	T	G	C	G	G	C	G	C	T	A	0.0563	0.5814
7	T	A	T	G	G	T	G	C	T	A	0.0423	0.6236
8	T	A	T	G	A	C	G	T	C	A	0.0378	0.6614
9	T	G	T	A	G	C	A	C	C	G	0.0376	0.699
10	T	G	T	G	G	C	G	C	T	A	0.0282	0.7272
11	C	G	T	G	G	C	G	C	T	A	0.0281	0.7553
12	T	G	C	G	G	C	A	C	T	A	0.0219	0.7772
13	T	G	C	G	G	C	A	C	C	G	0.0203	0.7975
14	T	G	T	G	G	T	G	C	T	G	0.0202	0.8177
15	T	G	T	G	G	T	G	C	T	A	0.0169	0.8347
16	T	A	T	G	A	C	G	C	C	G	0.0168	0.8515
17	T	G	C	G	G	C	G	T	C	G	0.016	0.8675
18	T	A	T	G	A	C	G	C	C	A	0.0144	0.8819
19	T	A	T	G	G	C	G	C	C	G	0.0141	0.896
20	T	G	C	G	G	T	G	C	C	A	0.0098	0.9058
21	T	G	T	A	G	C	A	T	C	G	0.0092	0.915
22	T	G	T	A	G	C	A	C	C	A	0.0079	0.9229
23	T	G	C	G	G	T	G	C	T	A	0.0075	0.9303
24	T	G	T	G	G	C	G	T	C	G	0.0074	0.9377
25	C	G	T	G	G	C	G	C	T	G	0.0072	0.9449
26	C	G	C	A	G	C	G	T	C	A	0.007	0.9519
27	C	G	C	A	G	T	A	C	T	G	0.007	0.959
28	C	G	T	G	G	C	G	T	C	A	0.007	0.966
29	T	A	C	G	G	C	G	C	C	G	0.007	0.9731
30	T	A	T	A	G	T	G	C	T	A	0.007	0.9801
31	T	G	C	A	G	T	G	T	C	G	0.007	0.9872
32	T	G	T	A	G	C	G	T	C	G	0.007	0.9942
33	T	G	T	G	G	C	G	C	C	G	0.0058	1
34	T	A	T	G	A	C	G	T	C	G	0	1
35	T	G	C	G	G	T	G	T	C	G	0	1

Table 5.7: Inferred distribution of haplotypes present in the composite group of 71 unrelated sheep based on the SNPstats implementation of the EM algorithm