THE DQA_2 LOCUS OF THE OVINE MAJOR HISTOCOMPATIBILITY COMPLEX: CHARACTERISATION, POLYMORPHISM, AND ASSOCIATIONS WITH FOOTROT.

Sandra-Marie Slow

A thesis submitted
in fulfilment of the requirements
for the Degree of Doctor of Philosophy
at Lincoln University
1999.

Abstract of a thesis submitted in fulfilment of the
requirements for the Degree of Doctor of Philosophy

The DQA_2 locus of the Ovine Major Histocompatibility Complex: Characterisation, polymorphism, and associations with footrot.

by Sandra-Marie Slow

Footrot is one of the most economically important diseases affecting the sheep industry in New Zealand. Losses in production from the disease and expenditure on treatment have been estimated to cost the New Zealand agricultural industry up to 100 million dollars per annum. None of the treatment options currently available are completely effective in reducing the economic impact of the disease, and are undesirable because of the possible impact on consumer health and the environment. One option that has not been fully investigated is the possibility of controlling footrot by exploiting natural variation in resistance. Variation in natural resistance to footrot is genetically derived. A number of heritable factors may be involved, and many of these interact with the environment to affect disease status. This implies that genetic markers for footrot resistance may exist, allowing the selection of superior animals.

The major histocompatibility complex (MHC) is central to the immune response. Research has shown that genetic variation within this region is associated with natural resistance to footrot. Genes within the MHC may be able to act as a genetic marker to enable the selection of animals that are naturally more resistant to footrot. In this study association between variation within the ovine class II DQA_2 gene and resistance to footrot was investigated in four different flocks from three different sheep breeds; Awassi, Corriedale and Merino. Half-sib progeny were typed at the DQA_2 locus using *Taq*I RFLP and Southern hybridisation typing. The animals were subjected to a footrot challenge, and their condition subsequently recorded to investigate associations between the inherited DQA_2 haplotype and footrot status.

Within the Awassi flock, the D and C alleles associated with resistance ($P \leq 0.05$), while the L allele associated with susceptibility ($P \leq 0.05$). No significant associations were

observed within the Corriedale flock, a result attributed to the challenge, where drought conditions led to poor disease transmission, and unreliable disease classification. Two new banding patterns were observed in the Merinos, and were tentatively defined as alleles Q and R. In the two Merino flocks, the G and C alleles were observed to significantly associate with resistance ($P \leq 0.05$), while the E allele associated with susceptibility ($P \leq 0.05$).

The data from the four flocks were pooled, which enabled twelve DQA_2 alleles (B, C, D, E, F, G, H, I, J, K, L, and N), to be ranked in terms of their relative footrot susceptibility or resistance. The G allele appeared to be most strongly associated with resistance ($P \leq 0.01$), whilst the I allele appeared to be most strongly associated with susceptibility ($P \leq 0.01$).

Additional variation was detected at the ovine DQA_2 locus by cloning and sequencing a number of DQA_2 alleles, with sub-division of the G, C and F alleles into G1, G2, C1, C2, C3, F1 and F2. A total of fourteen different DQA_2 sequences were obtained. Alignment of all the sequences revealed that the E, G2 and H allelic sequences were conserved across different breeds. Sequences obtained from the F allele did however, show breed-specific differences, with two F allele sequences being identified. F1 came from a mixed breed, which was a Coopworth x Perendale cross, while F2 came from a Romney. In addition, the sequence data provided evidence of a DQA_2 duplication in some sheep, with the G1 allele sharing closer homology to a putative bovine DQA_3 gene, than with other ovine DQA_2 sequences. The putative DQA_3 gene appeared to associate with a DQA_1-null haplotype. Sequence analysis also suggested that DQA_2 alleles E and C2 may not be expressed because of the lack of a splice site at the end of exon 2. From the sequence information obtained the structure of the ovine DQA_2 antigen-binding groove could be predicted. This enabled the antigen-peptide binding ability of different alleles to be compared.

Keywords: major histocompatibility complex (MHC), genetic marker, footrot, DQA_2, disease resistance, RFLP and Southern hybridisation.

ACKNOWLEDGMENTS

Upon reflection I have realised that undertaking and completing a PhD is mostly an exercise in patience and perseverance, regardless of the subject material. With both of these attributes having deserted me at times during this project, it was up to the people around me to keep me on track and to prevent complete insanity taking hold. Without the unwavering support, encouragement, and technical expertise of many people this project would never have been completed. I would like to take this opportunity to express my sincere gratitude to everyone who has been a part of this project.

First and foremost a very big thank you to my supervisors Dr Jon Hickford, and Dr Rob M^{c}Farlane. Jon without your commitment, foresight, patience, and continual assistance this project would never have begun. Thank you for allowing me to ask stupid questions, and giving so freely of your time to drive long distances for sample collection, and your continual input during the preparation of this thesis. Rob, although we hardly ever discussed the project, it was nice to know that if difficulties arose I could talk to you, thank you for your support.

To Chris Frampton, thank you for all your work and help with the statistical analysis, and being patient when I didn't understand what it all meant.

I would also like to sincerely thank the New Zealand Vice Chancellors Committee William Georgetti Scholarship, and Lincoln University for generously funding me during this work.

Thank you to all the farmers that were involved, without your commitment, and vision this project would not have been possible. A special thank you to Edward and Margaret Orr, Rodney Patterson, Angela and Jamie Molloy, and Rolland and Jean Mulvey. Thank

you all for the access to your animals, your enthusiasm for the project, and your wonderful hospitality during our visits.

To the blood extraction team, Andrea, Noelle, Grant, Rachael, Huitong, and Tulay, were we getting good after a couple of hundred samples? Thanks to you all for your help time, and making a mundane job entertaining, if not a little bit insane some days. Also thank you all for, the fun lab meetings at the pub, putting up with my loud music and bad moods. Most of all thank you for your friendship.

To Martin Keeley the expert bleeder, what can I say, if I had to bleed all the animals myself I would still be gathering the samples, thanks for your support.

A special thank you to Jane and Steve Brennan, your support and friendship has been very special, thank you for all your kindness.

Chris thank you for the good times, all your encouragement and friendship.

To Leigh, Hayley and Toni, well guys what can I say, we all started university together, and without your friendship, chat sessions and encouragement it would have been extremely difficult to get this far-thanks.

Lastly a very special thank you and all my love to my family. Mum and Dad, thank you for always being there for me, giving me confidence to achieve my goals, and your unconditional love and friendship. To Tracey, Patrick, Chris, Niki, wee Rhyn, Grandma, Grandad, Pat, Murray and Baz, thank you for your love, support, and encouragement. Doing this would have been a lot tougher if not impossible without you all.

TABLE OF CONTENTS

TITLE PAGE	I
ABSTRACT	II
ACKNOWLEDGEMENTS	IV
TABLE OF CONTENTS	VI
LIST OF FIGURES	XI
LIST OF TABLES	XIII
ABBREVIATIONS	XIV
AMINO ACID ABBREVIATIONS	XVII
CHAPTER ONE: LITERATURE REVIEW	**1**
Introduction	**1**
1.1 The Immune System	**2**
The Specific Immune Response: Lymphocytes and the Major Histocompatibility Complex	3
The Functional Significance of the MHC	5
Gene Organisation of the MHC	7
The Structure of MHC Proteins	9
The Antigen-peptide Binding Region of Class II MHC Molecules	10
1) Hydrogen Bonding to the Peptide Backbone.	12
2) Bonds Which Determine Antigen-peptide Specificity	13
Evolution of the MHC: Part One-Origin	15
Evolution of the MHC: Part Two-What Drives Polymorphism?	17
1) The Disease Theory	17
2) The Behaviour Theory	19
3) Disease Theory vs Behaviour Theory	20

Generation of New Variants: Gene Duplication, Gene Conversion, and Point Mutations 23

MHC and Disease 26

The Ovine Major Histocompatibility Complex 28

The Ovine DQ Sub-region 30

1.2 Footrot **31**

Pathology of *Dichelobacter nodosus* 32

Genetic Markers for Footrot Resistance 35

The Ovine MHC and Footrot 37

1.3 Investigating Class II OLA Genes and Associations with Natural Footrot Resistance: Research at Lincoln University **38**

Aims of this Thesis 39

CHAPTER TWO: RFLP AND SOUTHERN HYBRIDISATION TYPING OF THE OVINE-MHC DQA_2 GENE AND ASSOCIATIONS WITH FOOTROT **40**

Introduction **40**

2.1 Flocks Investigated, Criteria, and Challenge Protocol: An Overview **40**

Criteria for the Selection of Animals 40

Flocks Analysed 41

1) Awassi NZ Ltd 42

Introduction 42

The Challenge Procedure 43

The Footrot Scoring System 45

2.2 Methods **49**

Sample Collection 49

DNA Extraction 49

DNA Digestion and Electrophoresis 49

Southern Transfer 50

Description of the Ovine DQA_2 Probe 50

Probe Labelling 50
Membrane Pre-hybridisation 51
Membrane Hybridisation 52
Washing the Membranes 52
Autoradiography 52
Stripping the Membranes 53
Allele Identification 53
Statistical Analysis 53
2.3 Results **55**
Results for the Analysis of the Individual Sire Lines within Each Flock 56
1) Awassi NZ Ltd 60
2) Corriedale Flock 3 (BCBG) 60
3) Merino Flock 1 61
-1997 Trial 62
-1998-99 Trial 62
4) Merino Flock 2 63
Results for the Pooled Data Analysis 63
2.4 Discussion **65**
Is Normal Field Exposure or a Natural Challenge Suitable to Differentiate between Resistant and Susceptible Animals? 65
Analysis of the Individual Sire Lines within Each Flock 68
Pooled Data Analysis 70
2.5 Conclusion **73**
CHAPTER THREE: SEQUENCE ANALYSIS OF THE OVINE-MHC DQA_2 GENE **74**
Introduction **74**
3.1 Methods **76**
Sheep 76
Genomic DNA 77
PCR Primers 77

PCR Amplification 77
PCR Amplification Using Elongase™ 78
Agarose Gel Electrophoresis 81
Cloning PCR Products 81
DNA Concentration Measurement and Sequencing 82
3.2 Results **83**
G Allele Sub-division 86
C Allele Sub-division 86
Exon 2 Sequence Alignment 86
3.3 Discussion **93**
Are the Sequences from the Ovine DQA_2 Gene? 93
Breed-specific Alleles 95
C Allele Sub-division 95
Evidence of an Ovine DQA_2 Gene Duplication 97
The DQA_1-null Haplotype 99
E and C2 Alleles 102
A Possible Theory for the Evolution of the Ovine DQA Genes 103
Functional Relatedness of the Sequences within Exon 2 105
3.4 Conclusion **108**

CHAPTER FOUR: ANTIGEN-BINDING GROOVE POLYMORPHISM AND ASSOCIATIONS WITH FOOTROT SUSCEPTIBILITY **109**
Introduction **109**
4.1 Methods **111**
4.2 Results **112**
Hydrogen Bonding to the Antigenic-peptide Backbone 117
Salt Bridge Formation 117
Binding the N-terminus of the Antigenic-peptide 117
Side-chain Pockets and Peptide Specificity 118
The First Pocket 118
The Fifth Pocket 118

4.3 Discussion **119**

Is it Valid to Model the Structure of OLA-DQ from HLA-DR1? 119

Why Study Specific Peptide Motifs in the Antigen-binding Groove of Class II Molecules? 120

The α-chain of the Antigen-binding Groove for Ovine DQA_2 Alleles B, D, G2, H, K, and Q 120

The α-chain of the Antigen-binding Groove for Ovine DQA_2 Pseudogene Alleles C2 and E 123

4.4 Conclusion **126**

CHAPTER FIVE: GENERAL SUMMARY AND FUTURE DIRECTIONS **127**

REFERENCES **130**

APPENDICES **142**

A: Major Suppliers of Reagents, Enzymes and Other Consumables 142

B: Commonly Prepared Solutions 143

C: DNA Extraction from Sheep Blood 145

D: Southern Transfer of DNA to Nylon Membranes 146

E: Plasmid Map for the Ovine DQA_2 Probe 148

F: RFLP and Southern Hybridisation DQA_2 Typing Results for all of the Animals from Four New Zealand Farms 149

G: The Relative Risk Calculation 160

H: GenBank Accession Numbers for Sequences Used in Chapter Three 161

LIST OF FIGURES

Fig. 1.1:	MHC restricted antigen presentation to T-cells.	5
Fig. 1.2:	Genomic map of the human MHC.	8
Fig. 1.3	Domain organisation of MHC class I and II proteins.	9
Fig. 1.4:	Schematic representation of the basic architecture of the antigen-binding groove.	13
Fig. 1.5:	An electron density map attained by X-ray crystallography of the human MHC class II antigen HLA-DR1.	14
Fig. 1.6:	Five proposed mechanisms to explain how MHC genes influence individual odour.	21
Fig. 1.7:	Primary selective forces proposed by Potts and Wakeland, (1993) to maintain MHC genetic diversity.	23
Fig. 1.8:	Simplified genomic map of the ovine Major Histocompatibility Complex.	29
Fig. 1.9:	Factors contributing to the incidence and severity of footrot in sheep in New Zealand.	32
Fig. 1.10:	Factors affecting susceptibility to footrot.	34
Fig. 2.1:	Geographical locations of the farms investigated.	42
Fig. 2.2:	Diagrammatic representation of the banding patterns of the 15 DQA_2 *Taq* I alleles defined by Escayg, (1995).	54
Fig. 2.3:	Results from a Southern blot	55
Fig. 2.4:	The banding pattern observed for two new DQA_2 alleles, Q and R.	61
Fig. 2.5:	The ranking of the ovine DQA_2 alleles based on the relative risk values obtained for each allele.	71

Fig. 3.1:	Sequence alignment of the fourteen alleles identified.	87
Fig. 3.2:	Homology tree constructed from exon 2 of the fourteen DQA_2 alleles.	92
Fig. 3.3:	Homology tree constructed from ovine DRA1, DQA_1, and DQA_2 sequences.	94
Fig. 3.4:	Alignment of C allele sequences.	96
Fig. 3.5:	Phylogenic tree constructed from ovine DQA_2, bovine DQA_2 and DQA_3 sequences.	101
Fig. 3.6:	Schematic diagram representing the possible evolution of ovine DQA genes.	104
Fig. 4.1:	Comparison of the amino acid sequences of HLA-DR1 (M60333), HLA-DQA_1 (L34089), OLA-DRA (Z11600), OLA-DQA_1 (M33304), and the OLA-DQA_2 and DQA_2 pseudogene alleles.	113
Fig. 4.2a:	Predicted structure of the ovine-DQ alpha chain (residues 24-79).	114
Fig. 4.2b:	Residues 24-79 of the alpha chain of ovine DQA_2.	115
Fig. 4.2c:	Position of the first and fifth pockets in ovine DQA_2.	116
Fig. A1:	Capillary DNA blot (Southern blot) apparatus.	147
Fig. A2:	pC17-2PH plasmid map.	148

LIST OF TABLES

Table 2.1:	Summary of flock data and the type of challenge implemented for each flock used in this study.	47
Table 2.2:	Summary of results for the analysis of individual sire lines.	57
Table 2.3:	Results of the pooled data analysis.	64
Table 3.1:	DQA_2 Southern typing results, and flock information for animals used in this study.	79
Table 3.2:	Tabulated sequencing results.	84
Table 3.3:	RFLP and Southern hybridisation DQA_1 typing results for animals within groups 1 and 2.	100

ABBREVIATIONS

α	Alpha
A	Adenine
APCs	Antigen Presenting Cells
β	Beta
β_2m	Beta-2-microglobulin
bp	Base pairs
C	Cytosine
cDNA	Complementary Deoxyribonucleic Acid
CLIP	Class II associated invariant chain peptide
cM	Centi-Morgans
cm	centimetres
cpm	Counts per minute
D. nodosus	*Dichelobacter nodosus*
DNA	Deoxyribonucleic Acid
dATP	Deoxy-Adenosine Triphosphate
dCTP	Deoxy-Cytidine Triphosphate
dGTP	Deoxy-Guanosine Triphosphate
dNTP's	Deoxy-ribonucleotide Triphosphate
dTTP	Deoxy-Thymidine Triphosphate
EDTA	Ethylenediaminetetraacetic Acid
ER	Endoplasmic Reticulum
G	Guanine
HLA	Human Leucocyte Antigens
hsp70	Heat Shock Protein 70
IDDM	Insulin-dependent Diabetes Mellitus
Ig	Immunoglobulin
kb	kilobases
kD	kilodaltons

λ	Lambda
LMP	Low Molecular Mass Proteins of the Proteosome
M	Molar
Mb	Megabase
mCi	millicurie
mg	milligram
MHC	Major Histocompatibility Complex
min	minutes
mL	millilitres
mM	millimolar
mm	millimetre
MW	Molecular Weight
ng	nanogram
nm	nanomole
OLA	Ovine Leucocyte Antigens
PBD	Peptide-binding Domain
PCR	Polymerase Chain Reaction
RFLP	Restriction Fragment Length Polymorphism
s	seconds
SDS	Sodium Dodecyl Sulphate
SSC	Standard Sodium Citrate
SSCP	Single Stranded Conformational Polymorphism
T	Thymine
TAP	Transporter Associated Protein
TBS	Tris Buffered Saline
T_c	Cytotoxic T-cell
TCR	T-Cell Receptor
T_h	Helper T-cell
U	Units
UV	Ultraviolet

μg	microgram
μL	microlitre
μM	micromolar
WBC	White Blood Cell

AMINO ACID RESIDUE ABBREVIATIONS

A	**Ala**	Alanine
C	**Cys**	Cysteine
D	**Asp**	Aspartic Acid
E	**Glu**	Glutamic acid
F	**Phe**	Phenylalanine
G	**Gly**	Glycine
H	**His**	Histidine
I	**Ile**	Isoleucine
K	**Lys**	Lysine
L	**Leu**	Leucine
M	**Met**	Methionine
N	**Asn**	Asparagine
P	**Pro**	Proline
Q	**Gln**	Glutamine
R	**Arg**	Arginine
S	**Ser**	Serine
T	**Thr**	Threonine
V	**Val**	Valine
W	**Trp**	Tryptophan
Y	**Tyr**	Tyrosine

CHAPTER ONE

LITERATURE REVIEW.

Introduction

Traditional breeding of domesticated animals for food production has primarily focused on production traits related to growth, reproduction, and carcass characteristics. Less emphasis has been placed on breeding for animal health. Intensive livestock production systems have frequently relied upon extensive and expensive antibiotic, vaccine and chemoprophylactic methods to maintain health (Hawken *et al.*, 1998), but these options are becoming less desirable. Emerging new diseases, antibiotic resistant strains, issues relating to biosecurity, food safety, and the increased consumer pressure for antibiotic-free, and "naturally" produced food means that the sole reliance upon pharmaceutical intervention for the maintenance of animal health cannot continue.

One option available is the possibility of controlling the severity and incidence of ill health in farm animals by exploiting natural genetic variation in disease resistance. There have been many documented instances of breed and individual differences in genetic disease resistance among farm animals (Soller & Andersson, 1998). Although good animal health is a prerequisite for optimal animal production, with losses as a result of disease reaching 10-20% of total production costs (Horin, 1998), integrating health-related traits into traditional breeding programs has proven difficult, with most early observations of natural disease resistance largely being ignored. There were several reasons for this. Firstly, antibiotics were discovered in the late 1920's, and their usefulness in combating disease had been clearly proven during World War II. Secondly, a number of animal vaccines had been developed, which were effective in controlling disease. Finally, the genetics of natural disease resistance seemed extremely complicated, and there was concern that the planned breeding programs to increase natural resistance would be too slow compromising productivity (Adams & Templeton, 1998).

Advances in technology which allow the identification and characterisation of genes involved in disease resistance, combined with changes in attitude towards the excessive use of pharmaceuticals, and the failure of some vaccines to control certain diseases has meant that genetic approaches to disease control have been reconsidered. Over recent years attention has focused on identifying genes that are either directly involved or are closely linked to genes, which provide natural disease resistance. Once identified the genes can act as "genetic markers", whereby the presence or absence can indicate potential disease susceptibility. Such information can then be integrated into breeding strategies. When trying to identify and characterise genes that may be involved in natural disease resistance, a suitable starting point is to investigate genes associated with the immune system.

1.1 The Immune System

The immune system has evolved to meet the need for an internal monitoring system, where its primary function is to protect against pathogenic microorganisms. It can be likened to a mobile nervous system, which has evolved to detect and eliminate foreign material within a host (Germain, 1994). The vast network of molecules and cells which constitute the immune system have a single goal: to distinguish self from nonself, destroying the latter. Its hallmarks are specificity, adaptation and memory. The various cellular components of the immune system patrol the body for information, interpret this input within the highly ordered structure of secondary lymphoid tissue, and provide effector function in the form of activated cells or secreted products (Germain, 1994).

The immune system's response to invasion by pathogenic organisms can be divided into two categories, the first is innate immunity, and the second is specific immunity. Innate immunity, or inborn immunity includes the non-specific engulfment and destruction of microorganisms by phagocytes, and the release of generalised antimicrobial molecules, such as lysozymes and defensins. It is the first-line of defence upon pathogen invasion. Pathogenic organisms have a high potential to adapt and evolve, such that innate immunity is not always sufficient to provide protection from invasion. The immune

system has evolved a second-line of defence that is far more specific. As its name implies, the specific immune response involves the specific recognition of foreign antigens. Specific immunity is a learnt process, is adaptive, and provides immunological memory.

The Specific Immune Response: Lymphocytes and the Major Histocompatibility Complex

As a result of the high evolutionary potential of pathogens, birds and mammals have evolved specialised cells called lymphocytes, which have a comparable potential for reproduction and genetic variability to that of pathogens. Lymphocytes are the basis of the specific immune response, and they transcribe the highest number of genes among all cells within an organism (Horin, 1998).

Lymphocytes patrol the body as quiescent cells until they are activated by antigen, they then divide and become effector cells. Activated lymphocytes engage in different effector functions: B-cells differentiate into plasma cells and secrete immunoglobulins; cytotoxic T cells (T_C) destroy virus infected cells; and helper T cells (T_H) secrete cytokines which can enhance antibody production by B cells, promote the expansion of T_C cells, and provide chemotactic signals which enlist other immune system cells, such as macrophages, and natural killer cells to the site of infection.

Lymphocytes bind antigen through cell surface receptors. B-cells use immunoglobulins as both receptor, and effector molecules. Immunoglobulins can interact with a broad range of chemical structures, but their ligands must generally be on the outside of the invader, secreted by the organism, or expressed intact on the surface of infected host cells (Germain, 1994). If only this detection system existed, pathogens could easily evade the host immune system by "hiding" within cells of the host. The immune system has therefore, evolved a way of providing an extracellular indication of cellular invasion. The system that has evolved is based on one simple fact: all pathogens differ from their hosts by at least one protein that can be uniquely represented as a short peptide sequence

distinct from all host sequences of similar length. As long as these peptides can be produced, captured and stably expressed at the cell surface, membrane-bound immune receptors can identify them, providing a general system for the detection of intracellular infection (Germain, 1994).

Pathogenic peptides are bound by Major Histocompatibility Complex (MHC) glycoproteins within an infected cell. MHC proteins with bound antigenic peptide are transported to the cells outer membrane, where the MHC-antigen peptide complex is expressed on the surface of the cell, acting as a "flag" indicating that the cell is infected. T cells, through their T cell receptor (TCR), bind the MHC-antigen peptide complex at the cell surface, initiating an immune response.

T_C-cells bind intracellularly derived peptides, such as virus proteins, which are transported and presented on the cell surface by peptide binding class I MHC molecules. MHC class I molecules are expressed by virtually all nucleated cells. The initial processing of antigenic proteins occurs within the cytoplasm, and peptide fragments are transported into the endoplasmic reticulum (ER). Peptides then form stable trimolecular complexes with heterodimers of MHC class I heavy chain and β_2-microglobulin, prior to expression on the cell surface (Chicz & Urban, 1994).

T_H-cell receptors bind antigenic peptides derived from internalised proteins as well as other proteins entering the endocytic route, and are presented by MHC class II molecules. Degraded peptide fragments in endocytic or lysosomal compartments bind to class II molecules that intersect these vesicles *en route* to the cell surface (Chicz & Urban, 1994). Expression of MHC class II molecules is restricted to specialised antigen presenting cells (APCs), including B-cells, macrophages, dendritic cells, the thymic epithelium, and activated T-cells (Weenink & Gautam, 1997). Co-receptors including CD4 and CD8 on the surface of T_H- and T_C-cells respectively, assist in the binding of the TCR and MHC molecule (Figure 1.1).

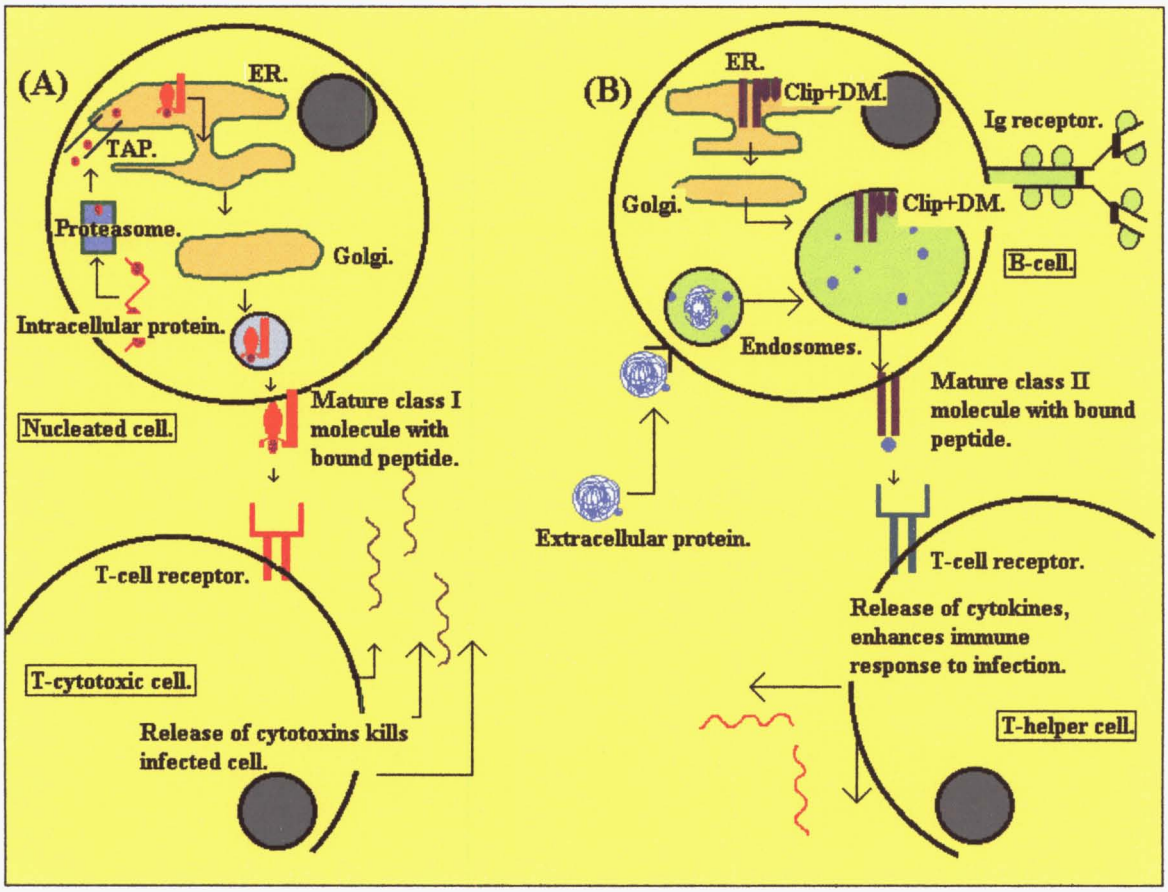


Figure 1.1: MHC restricted antigen presentation to T-cells. (A) MHC class I presentation to T-cytotoxic cells. Intracellular proteins are processed by the proteasome and transported to the ER by TAP, where peptide is loaded onto class I molecules. (B) MHC class II presentation to T-helper cells. Extracellular protein is engulfed by the antigen-presenting cell and is processed through the endosomal pathway. Naive class II molecules enter the ER and Golgi *en route* to the endosomal pathway. The invarient chain (CLIP) binds to MHC class II molecules blocking peptide binding in the ER and Golgi. Once in the endosome CLIP is removed and peptide is loaded onto the class II molecules, facilitated by the MHC-encoded molecule DM (Adapted from Weenink & Gautam, (1997)).

The Functional Significance of the MHC

The MHC is considered to be the centre of the "immune universe" (Trowsdale, 1995; Edwards & Hendrick, 1998). Class I and II MHC-encoded molecules serve as peptide-binding transport and display proteins, evoking effector responses upon recognition by

the MHC-antigen specific T-lymphocytes. Pathogens possess non-host-like proteins that can be uniquely presented as a short (8-15 amino acids) peptide sequence distinct from host sequences. Complete surveillance of the cellular environment is therefore, mediated by two distinct populations of MHC molecules, class I and class II. Together, the two pathways enable the immune system to monitor both the extracellular and intracellular environment, presenting fragments of many proteins to the circulating T-cell population. This phenomenon is known as MHC restriction (Weenink & Gautam, 1997). Other gene products, some also encoded within the MHC, participate in either the formation (LMP-low molecular mass proteins of the proteosome), or translocation of peptides (TAP-transporter associated with antigen processing), or the trafficking of MHC molecules (Germain, 1994). In addition, complement genes are encoded within the MHC, which encode proteins that provide many effector functions critical to antibody-mediated immunity and inflammation (Edwards & Hendrick, 1998).

Activated T-cells constitute a large part of the specific immune response, either killing infected cells or enhancing and co-ordinating the immune response towards infected cells. T-cells cannot recognise antigenic proteins unless they are part of an MHC-antigen complex. Inability to produce MHC molecules effectively prevents any response from the circulating T-cells towards invading pathogens, having extreme deleterious effects on the organism's survival. Thus, the MHC is central to the fundamental immunological processes of antigen processing and presentation.

MHC molecules are highly polymorphic membrane glycoproteins. Individuals simultaneously express several polymorphic forms from a large pool of alleles (more than 50 at some loci) within a population. No two individuals, with the exception of identical twins have the same MHC genotype. Different alleles selectively bind different sets of peptides, and hence MHC polymorphism appears to be responsible for the variation seen between individuals in terms of susceptibility to different diseases (Weenink & Gautam, 1997). The MHC also contributes the major genetic component of important autoimmune diseases. For example, insulin dependent diabetes mellitus (IDDM)(Horn *et al.*, 1988; Singer *et al.*, 1998; Durinovicbello, 1998) and rheumatoid

arthritis (Seidl *et al.*, 1997; Toussirot *et al.*, 1999). In addition, there have also been numerous publications with correlations between MHC haplotype and resistance/susceptibility to autoimmune disease (Powis & Trowsdale, 1991).

MHC genes therefore, make ideal candidates when trying to identify potential genetic markers for natural disease resistance in farm animals. When searching for genetic markers any candidate gene(s) however, need to be extensively characterised at the DNA level, with a comprehensive understanding of the genetic organisation, function, and allelic polymorphism.

Gene Organisation of the MHC

The MHC has been most extensively characterised in man and mouse. In humans it is referred to as the HLA (Human Leucocyte Antigens), and is located on the short arm of chromosome 6, occupying approximately 4 million base pairs (Milner & Campbell, 1992). The murine MHC, referred to as H-2 is located on chromosome 17. All vertebrates examined to date appear to possess a MHC with similar genetic organisation and complexity. In addition, characteristics of MHC-determined functions are found in invertebrates such as sponges and colonial tunicates (Kaufman *et al.*, 1984). Figure 1.2 shows a schematic representation of the gene organisation of the HLA, emphasising the polygenic nature of the class I- and II- loci.

Generally the MHC consists of three major families of molecules, designated class I, II, and III. Class I and II MHC genes encode the glycoproteins that are involved in MHC restriction (Milner & Campbell, 1992). They are often referred to as "classical transplantation antigens", as they were initially discovered when the ability to transplant both neoplastic and normal tissue in mice was found to depend on the sharing of similar MHC haplotypes (Kaufman *et al.*, 1984; Zinkernagel & Doherty, 1997). Class III MHC genes encode a raft of different genes, some of which are involved in the immune response (complement components C2, and C4, tumour necrosis factor α and β), but they do not encode the "classical" MHC glycoproteins involved in antigen presentation.

The class I region encodes HLA-A, B, and C loci, and spans approximately 1.8 Mb (Milner & Campbell, 1992). The class II or HLA-D region contains sub-regions DR, DP, and DQ, and spans approximately 0.75 Mb (Milner & Campbell, 1992). Each D sub-region contains α and β genes whose products form the class II antigens (Powis & Trowsdale, 1991). Other class I (HLA-E, -J, -H, -G, and -F), and II (HLA-DN, -DO) genes have been identified but their role in antigen presentation remains to be defined, and they are generally referred to as non-classical genes. Some of these genes may in fact be pseudogenes, and do not give rise to functional proteins (Lawrance & Quaranta, 1991). The overall size of the human MHC is approximately 3 Mb (Trowsdale, 1988).

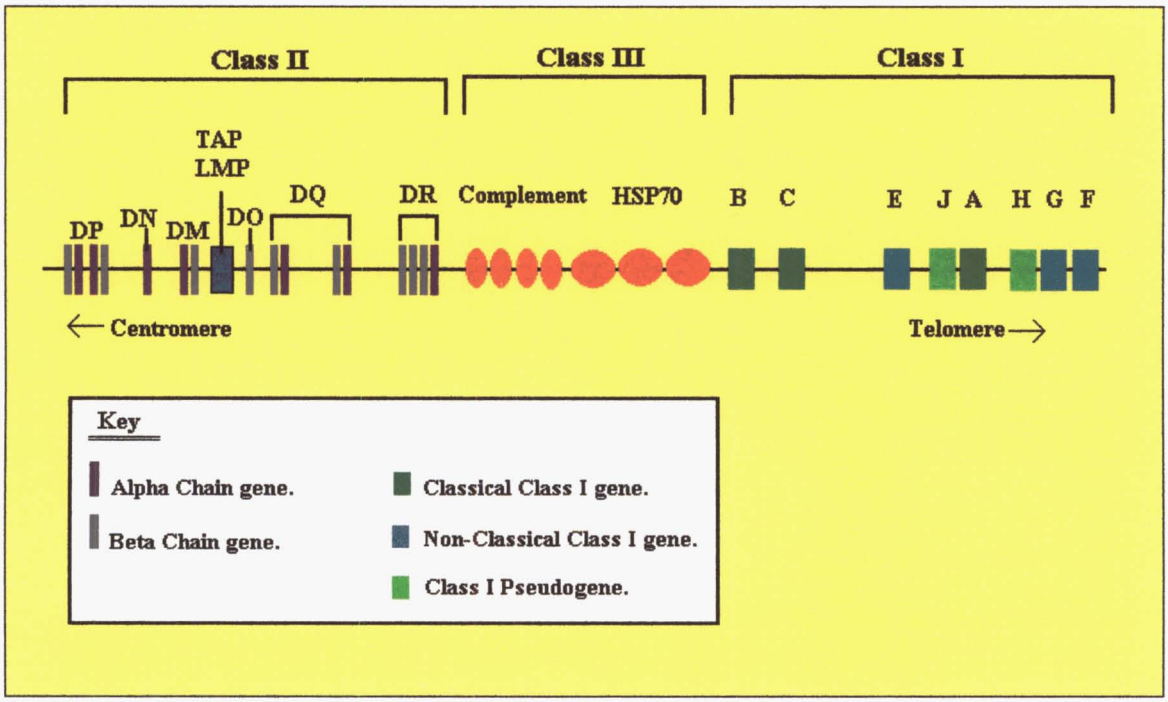


Figure 1.2: Genomic map of the human MHC (adapted from Trowsdale, 1995; Ober *et al.*, 1998).

Recombination tends to take place in certain locations in the human MHC. Non-random recombination positions (hotspots) have been observed in class I and III regions (Trowsdale, 1995). In the class II region however, recombination has never been reported between DRA and DQA genes, suggesting that this region is inherited as an intact block (Trowsdale, 1995).

Scattered throughout class I and II MHC genetic regions are pseudogenes and gene fragments. Pseudogenes, are thought to have once been functional genes, but have degenerated over time such that they no longer produce functional proteins, a result of aberrant splice sites, mutations or premature stop codons (Trowsdale, 1988). As a gene family, it is remarkable that the MHC contains a large number of highly related sequences on one chromosome, but not on others. One reason for having all of these sequences, including pseudogenes, together may be to facilitate gene conversion for exchange of sequence information, enabling the maintenance of the extreme polymorphism that has been observed (Trowsdale, 1988).

The Structure of MHC Proteins

The MHC class I and II antigens are expressed as heterodimeric molecules, showing structural similarities to the immunoglobulin C regions. Unlike the immunoglobulin genes, MHC genes do not somatically recombine to provide structural diversity, instead their genes are highly polymorphic (Srivastava *et al.*, 1991).

A major breakthrough in the understanding of MHC restriction and antigen presentation came with the solution of the X-ray crystallographic structure of firstly class I MHC molecules (HLA-A2; Bjorkman *et al.*, 1987), and later class II MHC molecules (HLA-DR1; Brown *et al.*, 1993). Such analysis revealed that MHC molecules were composed of two distinct functional regions, a polymorphic extracellular portion consisting of two antigen-binding domains, and a membrane proximal portion consisting of two conserved domains (Figure 1.3).

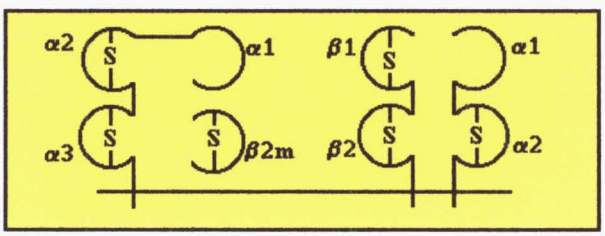


Figure 1.3: Domain organisation of MHC class I and II proteins
(Germain, 1994).

In class I MHC, the 45 kD heavy chain (class I α) contains both the polymorphic peptide binding domains (α1 and α2), and one conserved domain (α3). The 12 kD light chain consisting of the non-MHC encoded molecule β2-microglobulin (β2m) makes up the other conserved domain (Germain, 1994; Hughes & Yeager, 1997).

In class II MHC, each of the two chains, α (33 kD), and β (29 kD), includes one polymorphic peptide binding domain (α1, and β1), and one conserved domain (α2, and β2) (Germain, 1994).

The Antigen-peptide Binding Region of Class II MHC Molecules

The most studied domain of MHC molecules is that which constitutes the antigen-peptide binding region, or the antigen-binding groove. The binding groove defines the MHC molecule's function, by determining the nature of the antigenic peptide that is to be bound and presented to the circulating T-cells.

For effective peptide capture and presentation it is possible to define certain parameters within which the antigen-binding groove must operate. The maximum T-cell stimulus needs to be obtained from the smallest amount of available antigen, when all antigenic peptides containing the same linear sequence are displayed in the same conformation. This would ensure that all MHC molecules containing a particular peptide are suitable for interaction with the set of identical receptors on individual T-cells. Achieving this type of structural homology suggests that the peptide needs to be tethered within the binding groove at several places along its length (Germain, 1994). In addition, by anchoring the peptide in several places, prolonged interaction between the groove and the peptide could be maintained, such that a stable MHC-peptide complex can be transported to the cell surface, and be effectively displayed. If the MHC-peptide complex dissociated too quickly, or the peptide was only loosely bound within the groove the peptide could be lost during intracellular transit, or upon arrival at the cell surface (Germain, 1994).

The immune system must also be able to sample a broad range of peptides to ensure that at least one unique sequence is presented for each pathogen. Each individual expresses only a small number of different histocompatibility proteins. Thus, each histocompatibility protein must be able to bind a large number of different peptides to ensure an immune response against many possible pathogens (Stern *et al.*, 1994). This condition could be met by having a very degenerate binding site, focused only on conserved features common to all peptides. However, when the requirements for fixed peptide orientation and stable binding are considered, a highly degenerate binding site seems unlikely (Germain, 1994). As a compromise, MHC molecules seemed to have sacrificed the broadest possible binding capacity to maximise effective antigen display. The limitations on the breadth of peptide capture are largely overcome by the expression of multiple loci, and of different alleles at a given locus in the individual and the species (Germain, 1994). In addition, it has been suggested that the extensive polymorphism of histocompatibility genes, particularly within exon 2, which encodes the antigen binding groove, may be the result of the selection of alleles that can present peptides from particular pathogens (Stern *et al.*, 1994). This theory illustrates the possible action of various selective forces that could be operating on MHC genes to increase genetic diversity, and is considered to be part of the "co-evolutionary arms race", between pathogen and host. The selective forces that drive MHC diversity has been the topic of much speculation and debate, and will be discussed in further detail later.

A further complication of MHC-peptide binding is that if the stable capture of peptide within the groove depended upon multiple conserved bonds to main-chain atoms, restricted peptide specificity would result, owing to steric interference of particular side chains of some peptides. Rapidly evolving micro-organisms could take advantage of this type of limitation by fixing mutations that incorporated inhibitory side chains at suitable places in their proteins, thus preventing peptide binding (Germain, 1994).

So how does the binding groove of MHC molecules maintain a reasonably degenerate binding site in order to accommodate possible pathogen evolution, and yet maintain a degree of peptide specificity, which would be required for effective capture and display?

Crystallographic studies of human class I (Bjorkman *et al.*, 1987; Maden *et al.*, 1991) and II (Brown *et al.*, 1993; Stern *et al.*, 1994) molecules have helped answer how the binding groove of MHC molecules meets these seemingly contradictory requirements.

The binding system that has evolved consists of two parts: 1) Hydrogen bonding to the antigenic-peptide backbone; 2) Bonds that determine antigenic-peptide specificity, whereby specific residues within the groove interact with antigenic-peptide side chains.

1) Hydrogen Bonding to the Peptide Backbone

The architecture of the groove of both classes is composed of eight strands of antiparallel β sheet as a floor, and two antiparallel helical regions at the sides (Figure 1.4). As both class I and II molecules have a similar role in antigen presentation, it was assumed at first, that the antigen-binding groove of both classes was the same. While this remains fundamentally correct, there are some important differences, which determine the types of antigenic peptide each class presents (Brown *et al.*, 1993).

For both class I and class II MHC molecules, a broad peptide binding capacity arises from hydrogen bonds between conserved residues of the MHC molecule, and backbone or terminal residues of the associated peptide (Brown *et al.*, 1993; Stern *et al.*, 1994). Unlike class I, class II molecules form conserved bonds with main-chain atoms of the peptide, not the terminal groups. The end of the groove is open, such that the peptide can extend out (Brown *et al.*, 1993). Consequently, class II MHC molecules bind peptides of a greater range of lengths than class I, typically 12-24 residues, although whole proteins with appropriately exposed regions could be bound (Germain, 1994) (Figure 1.5).

Hydrogen binding of the antigen-peptide backbone by conserved residues within the MHC molecule is independent of antigenic-peptide sequence. MHC residues involved in hydrogen bond formation have been shown to be conserved in most human and mouse class II molecules investigated, suggesting a universal method for peptide binding by class II molecules (Brown *et al.*, 1993; Stern *et al.*, 1994).

2) Bonds Which Determine Antigen-peptide Specificity

The second part of the binding of antigenic-peptides, determines the specificity of the antigen-peptide to be bound. Polymorphic residues in both class I and II molecules are clustered within the peptide-binding region (Stern *et al.*, 1994). The polymorphism observed in different MHC molecules within the binding groove can subtly alter the structure of the groove such that the peptide specificity is changed. As a consequence, different MHC molecules are able to bind and present different antigenic-peptides to the immune system, with some molecules being able to present antigenic peptides from a particular pathogen more effectively than others (Stern *et al.*, 1994).

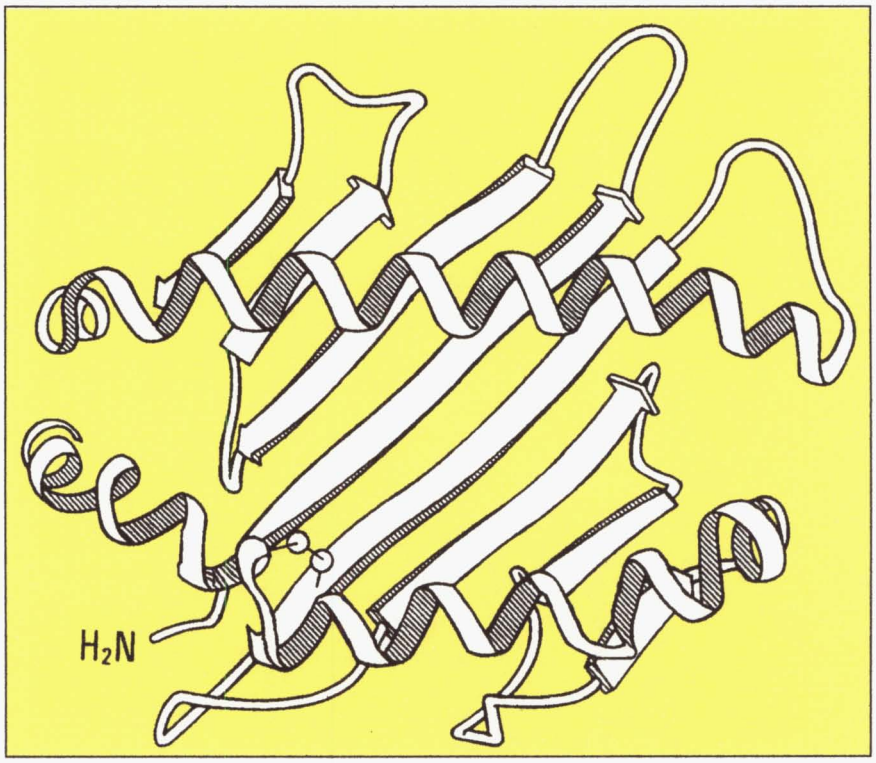


Figure 1.4: Schematic representation of the basic architecture of the antigen-binding groove (Adapted from Lawrance & Quaranta, 1991).

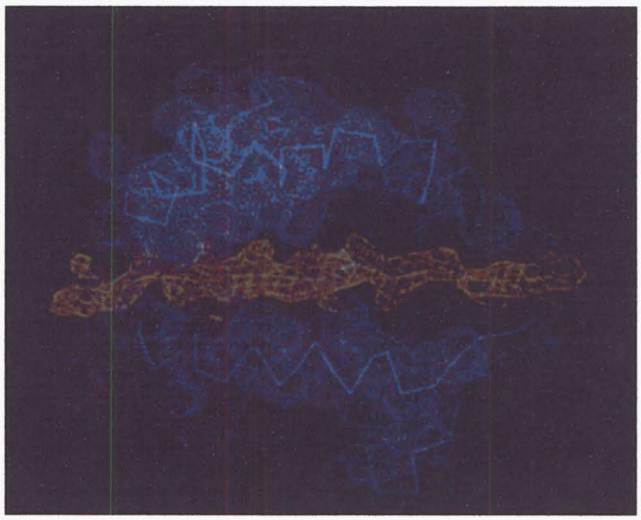


Figure 1.5: An electron density map attained by X-ray crystallography of the human MHC class II antigen HLA-DR1. HLA-DR1 is complexed with a collection of endogenous bound peptides. Electron density of the peptides is shown in orange, and the van der Waals surface of DR1 in blue. This is "top view", presumably as seen by T-cells. The α1 domain helix is on top, β1 domain helix on the bottom, and the peptide N-terminal is to the left (Brown *et al.*, 1993).

Specificity in antigen-peptide binding involves polymorphic residues in the MHC molecule and specific side chains of the peptide. Within the antigen-binding groove the side chains of the polymorphic residues form depressions, or pockets (Brown *et al.*, 1993; Stern *et al.*, 1994). The distinct chemical and size characteristics of these pockets in different MHC molecules results in strong preferences for interacting with certain amino acid side chains. The "anchor" residues that fit optimally into these pockets occur with high frequency in specific positions in peptides associating tightly with particular MHC molecules (Stern *et al.*, 1994). Two to three correct anchors are required for optimal binding, and for any given MHC allele, anchor positions are at fixed distances from one another, and involve only a few specific amino acids (Stern *et al.*, 1994).

Five pockets have been identified for the human class II molecules, HLA-DR1, and HLA-DQ (Brown *et al.*, 1993; Stern *et al.*, 1994; Paliakasis *et al.*, 1996). As long as at least two anchor residues are correctly bound within a pocket, the class II groove can accommodate non-optimal binding of other side chain residues within the other pockets. In this way there is still some flexibility in the range of peptides that can be bound by

any given MHC molecule, and is particularly true of class II molecules as they bind much longer peptides (Stern *et al.*, 1994; Paliakasis *et al.*, 1996).

Evolution of the MHC: Part One-Origin

The MHC molecules are members of the immunoglobin super family, and represent the most diversified germ-line genes known. The members of this family have domains belonging to three basic types called C1, C2 and V. In the case of class I, the α3 domain and the single domain of β2-microglobulin (β2m) are C1 domains, while α2 and β2 domains of class II are the C1 domains. The conserved (C) domains of TCR and Ig are also C1 domains, suggesting that all of these molecules share a common origin. However, although TCR and Ig possess V (Variable) domains that show high homology, the MHC molecules lack V domains. Rather, the peptide binding domains of MHC molecules are a type unique to the MHC. The origin of these domains has been an area of intense speculation because they are both unique to the MHC and critical to its function (Hughes & Yeager, 1997).

One hypothesis suggested that MHC peptide-binding domains (PBD), evolved from immunoglobulin superfamily V domains (Hashimoto *et al.*, 1990), but was generally unaccepted because sequence similarity is low between these domain types. Another hypothesis (Flajnik *et al.*, 1991) suggested that the MHC PBD might have evolved from the PBD of the heat shock protein 70 (hsp70), which has some sequence similarity to the class I α1 and 2 domains. It was thought that a recombination event between an immunoglobulin superfamily member and hsp70 gave rise to the class I molecule, and that class I evolved before class II. However, with the description of the structure of a bacterial hsp70 PBD, little similarity was observed with class I molecules, suggesting that MHC molecules did not evolve in this way (Hughes & Yeager, 1997).

Other investigators (Hood *et al.*, 1985) noticed that β2m and the class II α chain were similar to each other such that they may have shared a common ancestor, and that the class I α chain and the class II β chain also shared a common ancestor. A phylogenic

analysis of the C1 type domains from representatives of these groups of molecules provide support for this theory, although not strongly, but better than the other proposed models (Hughes & Nei, 1993). This analysis gave rise to yet another hypothesis, that the first MHC molecules had a class II-like structure, and that class I evolved later as a result of an exon shuffling event (Kaufman, 1988). Providing further support for this hypothesis was the observation that class II α1 and class II β1 are much more similar in sequence to each other than would be expected than if they were derived from different portions of an ancestral hsp70 PBD (Hughes & Nei, 1993).

Klein and Sato also supported the theory that class I MHC molecules originated from class II. They proposed that the ancestral MHC gene encoded a single PBD and a single membrane anchored C1 domain, and that it initially coded for proteins comprised of homodimers, forming a complete peptide-binding groove (Klein & Sato, 1998). Duplication of the ancestral gene and the subsequent divergence of the copies produced the class II A and B genes encoding the α- and β-chains of the class II heterodimer. An ancestral class I gene may have then been generated by exon shuffling, for example, by the joining of the PBD-encoding exon of one gene to a complete class II gene. The newly generated class I α-protein came to be stabilised by its association with β2m, which may have been available because it carried out, at that time a different function (Klein & Sato, 1998).

As yet however, MHC protein sequences provide no clear indication as to which of the two classes may have emerged first, an observation consistent with the notion that they may have arisen in a short order. The matter of the "first born" is still open to speculation, and both options, class I derived from class II, or vice versa, have their proponents (Klein & Sato, 1998).

Evolution of the MHC: Part Two-What Drives Polymorphism?

Without some form of selective pressure to maintain polymorphism, many HLA alleles would have been lost by random genetic drift. The high frequency of apparently new class I and II alleles in certain populations suggests that very powerful selective forces operate on the MHC (Riley & Olerup, 1992). The nature of these selective forces, and their relative influence has been the subject of debate for at least the last thirty years.

Where genealogies of MHC allele sequences have been constructed, from a number of species, it has been observed that the divergence of allelic MHC lineages predates the speciation event, giving rise to separate taxa. It is clear that the maintenance of polymorphism is not simply an equilibrium between the selection against harmful alleles and their introduction by mutation. Neither is it an equilibrium between the introduction of neutral alleles by mutation and their loss by drift (Wills, 1991). This suggests the action of some form of balancing selection over long periods of evolutionary time. In addition, comparisons of allelic sequences present within mice and human populations indicate that the rate of nonsynonymous substitution exceeds the rate of synonymous substitution at the antigen presenting site, thus favouring new MHC variants and increasing diversity (Hughes & Nei, 1991). Furthermore, in human populations the large numbers of alleles present at MHC loci show a relatively even distribution, leading to higher levels of heterozygosity than may be explained under neutral theory, again suggesting some form of balancing selection (Paterson, 1998).

1) The Disease Theory

From an evolutionary point of view, the most striking characteristic of the MHC is the high polymorphism of some loci. This polymorphism was discovered long before the molecule's peptide-presenting function was known, and for a long time it was very difficult for biologists to come up with a satisfactory explanation for it. Soon after the phenomenon of MHC restriction was documented, Doherty and Zinkernagel (1975) proposed a theory (the "disease theory") to explain the maintenance of MHC polymorphism, which was related to the recently discovered antigen-presenting function

of these molecules. They proposed that the genetic diversity of the MHC genes results from pathogen driven selection that favours either MHC heterozygotes (heterozygote advantage, or overdominant selection), or relatively rare MHC genotypes (negative frequency-dependent selection), or both.

Negative frequency-dependant selection predicts that pathogens evolve to minimise immune recognition by the most prevalent MHC alleles in a population. New MHC alleles have a selective advantage because pathogens have not had time to adapt to cells carrying the new MHC alleles and, as a result, individuals with these alleles have increased resistance to pathogens (Bodmer, 1972).

The overdominant selection (heterozygous advantage) model is based on the observation that within some populations the number of MHC heterozygotes is higher than expected. Heterozygotes have a selective advantage by being able to bind and present more antigens. These individuals therefore, have a superior ability to resist pathogens and consequently polymorphism would be maintained in the population (Hughes & Nei, 1989).

Support for the disease theory was provided when it was observed that there were reduced frequencies of the class I HLA-Bw53 allele and class II DRB1*501 allele in West African children suffering from severe malaria (Hill *et al.*, 1991). These alleles provide less individual, but greater overall protection than the sickle-cell haemoglobin allele. The protective MHC alleles occurred in low frequency among severe cases of malaria compared to mild cases. Furthermore, the class I allele is very rare in non-malarial regions, while reaching a frequency of 25% in malarial West Africa. It therefore, appears that certain MHC alleles confer resistance to malaria and are increasing in the population for that reason. The phenomenon has also been observed in other species. In chickens with a particular MHC haplotype, 95% survive infection by Marek's disease (a type of viral leukaemia), whereas individuals with other haplotypes have nearly 0% survival (Kaufman & Wallny, 1996).

2) The Behaviour Theory

One explanation for the evolution of sexual reproduction is that sex is an advantage in the coevolutionary arms race between pathogens and hosts. An alternative hypothesis for the maintenance of MHC diversity is the "behaviour theory", whereby mating preferences could be responsible for driving MHC polymorphisms. Mate preference (disassortative mating) and selective abortion of histocompatible foetuses have been proposed as the mechanisms by which the high heterozygosity is maintained at MHC loci (Potts *et al.*, 1991; Wedekind *et al.*, 1995; Ober, 1995).

Several studies in mice have revealed that females prefer to mate with MHC dissimilar males. When nine populations of half-wild mice carrying known histocompatibility types were established in enclosed but natural environments, 27 % fewer MHC-homozygotes than expected were observed (Potts *et al.*, 1991). Of these 6 % were due to selective fertilisation or abortion, and the remainder arose from mating choices made by females. Females initially tended to establish territorial relations with MHC-dissimilar males, but they also devoted half of their mating activity to contracting extraterritorial liaisons with further MHC-incompatible males.

Studies on human populations have been contradictory, although surveys still reveal a significant excess of couples with a history of spontaneous abortions sharing antigens for one or more HLA loci (Ober *et al.*, 1997).

There are two potential reasons why females prefer to mate with MHC-dissimilar males. MHC-dependent mating preferences may enhance immunological resistance of a female's progeny to pathogens and parasites (e.g. heterozygous advantage). MHC disassortative mating preferences may also reduce inbreeding, which would not only increase MHC heterozygosity, but overall genetic heterozygosity (Penn & Potts, 1998).

How do females recognise the MHC identity of potential mates? Evidence accumulated indicates that MHC genes influence individual odours, and MHC molecules have been found to occur in urine and sweat. It has been observed that mice can distinguish the

odours of mice that differ genetically only at classical MHC loci, rats can discriminate the odour of rats that differ genetically only in the MHC region, and women prefer the odour of men who have dissimilar genes (Penn & Potts, 1998). In mice and humans olfactory receptor-like genes have been located within the class I region of the MHC (Fan *et al.*, 1995). It has been proposed that these olfactory receptor-like genes may be involved in the detection of MHC-determined odour types. However, further investigation is required to determine whether these genes are truly olfactory receptors, whether they are polymorphic, what types of odour they detect, and whether this group of receptors play any specific role in MHC-linked odour determined mating behaviour (Fan *et al.*, 1995).

Several hypotheses (Figure 1.6), have been put forward to account for how MHC genes may influence odour and mating preferences (reviewed Penn & Potts, 1998).

3) Disease Theory vs Behaviour Theory

The idea that mating preference made a considerable contribution to MHC diversity initially received a lot of scepticism, as the connection with the acknowledged role of the MHC in immune recognition was not obvious. The questions that arose were, what advantage would there be in having an MHC-based mating preference, and how would this have evolved, such that it would fit with the role of the MHC in the immune system?

In 1993 Potts and Wakeland helped to reconcile the two theories, showing how both pathogen driven selection and the avoidance of inbreeding could enable a MHC-disassortative mating preference to evolve in vertebrates that were able to detect individual odours in an MHC-allele specific manner.

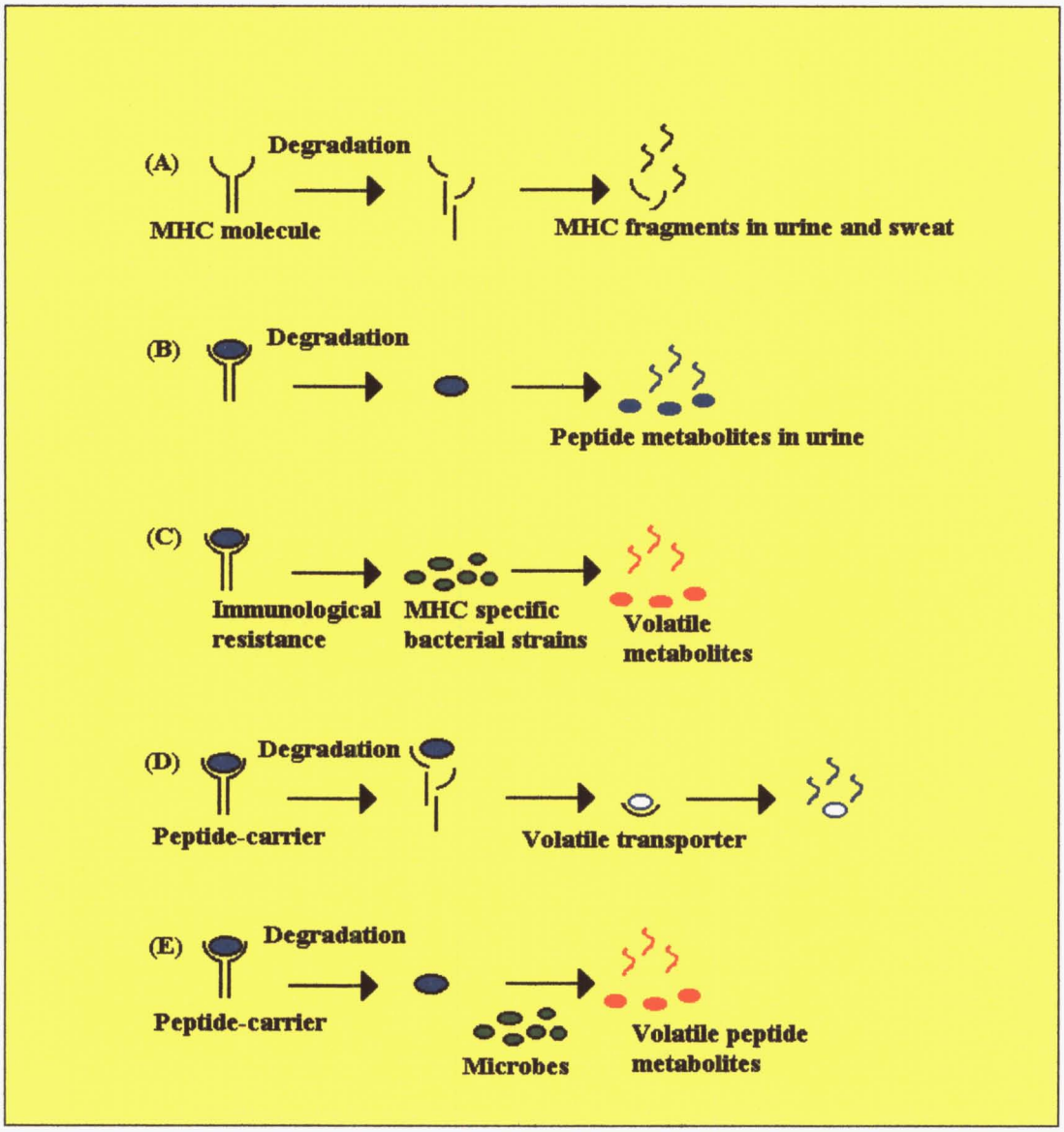


Figure1.6: Five proposed mechanisms to explain how MHC genes influence individual odour. (A) The MHC molecule hypothesis: Fragments of MHC molecules in urine and sweat provide the odourants. (B) The peptide hypothesis: MHC molecules may alter the pool of peptides in urine whose metabolites provide the odourants. (C) The microflora hypothesis: MHC genes may alter odour by shaping allele-specific populations of commensal microbes. (D) The carrier hypothesis: MHC molecules could be altered to carry volatile aromatics. (E) The peptide-microflora hypothesis: MHC molecules alter odour by changing the peptides that are available to commensal microbes. (Adapted from Penn & Potts, 1998).

Potts and Wakeland (1993) proposed that pathogen driven selection favours genetic diversity of the MHC through both heterozygous advantage (overdominance), and frequency dependent selection. This in turn favours the evolution of MHC disassortative mating preferences because, such matings would preferentially produce MHC-heterozygous progeny that have increased fitness, and would enjoy enhanced disease resistance. MHC-based urinary odours make MHC-directed mating preferences possible. These mating preferences professedly further extend MHC polymorphism, making these loci increasingly useful for kin recognition and therefore, avoidance of inbreeding becomes an additional selective force favouring MHC-based disassortative mating (Figure 1.7) (Potts & Wakeland, 1993).

Potts and colleagues (1994) went on to show that in house mice the MHC-based mating preferences were sufficient to account for the genetic diversity of their MHC genes. They concluded that inbreeding avoidance was the most important function of MHC-based mating preferences, and that although some form of pathogen-driven selection was responsible for generating genetic diversity at MHC loci this selective pressure was secondary.

Such theories are still controversial. It is clear, however, that both pathogen driven selection and reproductive selection through mating preference have played a part in generating the unprecedented diversity of MHC genes. The relative importance that each selective pressure has in maintaining MHC diversity would be variable depending on the species, strength of inbreeding, and the sensitivity to MHC based odours.

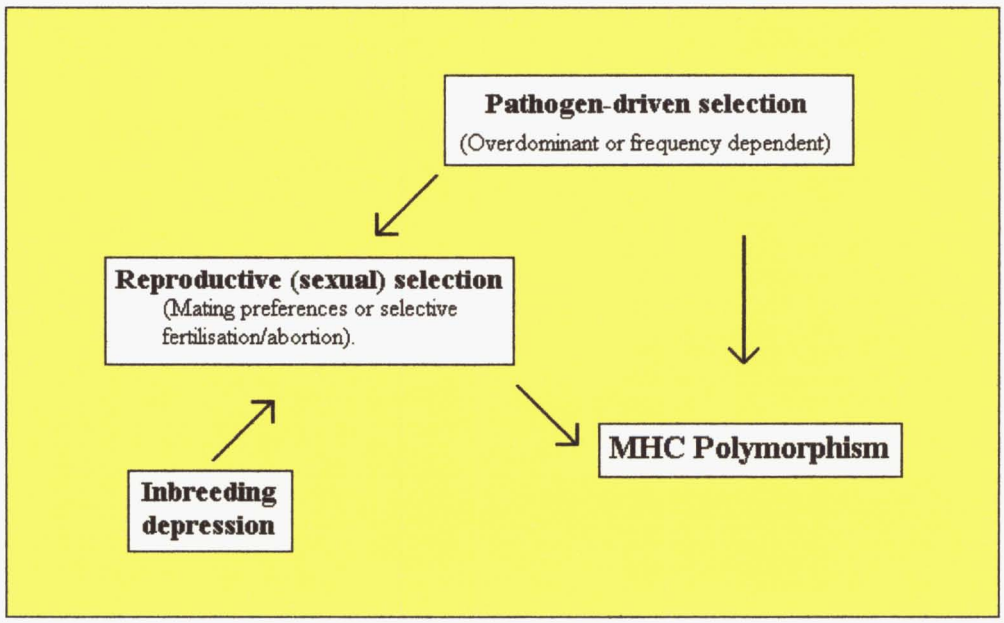


Figure 1.7: Primary selective forces proposed by Potts & Wakeland, (1993) to maintain MHC genetic diversity. Arrows indicate the direction of selective forces that favour MHC polymorphisms, either directly, or indirectly because they favour MHC-based disassortative mating preferences. Reproductive mechanisms such as selective fertilisation or selective abortion are viable alternatives to mating preferences, but the evidence for these post-mating mechanisms is weak. Furthermore, pre-mating mechanisms will generally be more efficient at preferentially producing progeny with favoured genotypes, because they avoid the costs associated with mating and aborted zygotes. The relative importance of the three selective forces in a particular species depends on variables such as strength of inbreeding-and pathogen load, likelihood of inbreeding, and sensitivity to MHC-based odours (Potts & Wakeland, 1993).

Generation of New Variants: Gene Duplication, Gene Conversion, and Point Mutations

The sequences of MHC genes that encode the peptide-binding region, have been the focus of much of the population genetic research on MHC diversity. Gene duplication is thought to have underlain the creation of class I and II ancestral genes, and has been a major force in MHC evolution ever since. No vertebrate has been found with only one MHC locus, the genes always come in groups, in which individual members are related to one another in their sequences. A quantitative evaluation of the relationship by dendrogram-construction algorithm reveals a clustering indicative of an origin by duplication from a single common ancestor (Klein *et al.*, 1998).

Current evidence favours a role for gene conversion, or microrecombination in the generation of polymorphism. Gene conversion is defined as the transfer of a gene segment from a donor gene to a homologous acceptor gene without the donor gene being changed in the process (Högstrand & Böhme, 1998). Gene conversion causes a phenomenon called "concerted evolution", meaning that duplicated genes do not evolve independently. Thus, duplicated gene copies may be more similar within species than between species even if the duplication occurred before speciation (Andersson *et al.*, 1991). Interlocus gene conversion increases genetic diversity for a locus or a set of loci by introducing new variants from different loci. Therefore, all loci are assumed to be polymorphic, and the polymorphism at different loci evolves in unison as a result of concerted evolution (Nei *et al.*, 1997).

Concerted evolution through gene conversion has been reported for human and mouse class I genes (Weiss *et al.*, 1983; Jaulin *et al.*, 1985; Wheeler *et al.*, 1990), and for the first domain exon of class II β loci (Andersson *et al.*, 1991). Indirect evidence for microrecombination comes from alleles that differ from one another only by small segments, and the demonstration that the majority of intron sequences on both sides of the peptide-binding region have similar evolutionary histories, a pattern that makes single-crossover recombination unlikely. The average size of the putative exchanges among human class I alleles is estimated to be between 15 and 20 nucleotides. An example of microrecombination with functional consequences is the HLA-B*5301 sequence, which is thought to confer resistance to malaria and is identical to the common, non-protective B*3501 allele, except at the exchanged codons 77 to 83 (Edwards & Hendrick, 1998).

Pseudogenes present within the MHC may play an important role in the creation of new genes through gene conversion. Some investigators believe that pseudogenes have been preserved in the MHC as they provide a source of genetic material for the exchange at functional loci, thus creating new genes from the ruins of old (Clayton & Gee, 1993). Others however, believe that pseudogenes are truly "dead" genes, and are in the process

of being completely lost from the genome. Creating new genes from pseudogenes has been likened to bringing a "corpse back to life" (Klein *et al.*, 1998).

Not all MHC loci appear to have evolved by concerted evolution through gene conversion. Concerted evolution has not been observed in cattle class I genes (Garber *et al.*, 1993), and the relative importance of point mutation versus conversion in generating diversity is still controversial. Many detractors insist that although gene conversion has played an important role in the generation of some MHC diversity, it occurs too infrequently to account for most MHC diversity, and the continual reintroduction of new alleles into the population.

Mutation is another mechanism that contributes to MHC variation. Some MHC loci have been conserved for many millions of years. Duplication of these genes, and subsequent mutations within the peptide-binding region that give an advantage under selective pressure has been maintained in a population, thus increasing MHC variability. One early proposal to explain MHC polymorphism was the hypothesis of a high mutation rate (Bailey & Kohn, 1965). A high mutation rate even in the absence of selection, may cause extensive polymorphism, and has been observed for microsatellites (Hughes & Hughes, 1995).

Some scientists believe that mutation is the most important mechanism by which genetic diversity of the MHC is maintained. However, this remains controversial because the mutation rate is not considered by some to be high enough in MHC genes to account for all of the diversity (Hughes & Hughes, 1995). In addition, a high mutation rate should affect all sites equally, and not be confined to the peptide-binding region codons, where most polymorphism is observed (Hughes & Hughes, 1995).

It is clear however, that whatever the mechanism that is most important in generating diversity, gene duplication, gene conversion, and point mutation, all play a part, and more than one mechanism many be operating. For instance, in mice class I genes, it has been reported that conversion-like events have taken place, and on one occasion was

accompanied by a point mutation (Wheeler *et al.*, 1990). The relative importance of each mechanism many therefore be dependent on the species, and MHC loci which are under investigation.

MHC and Disease

The MHC glycoproteins were originally described as the major obstacle to transplantation, although it seemed unlikely that their presence was solely to frustrate transplantation surgeons. Subsequent identification of their role in antigen processing and presentation, has highlighted their importance in infectious disease, autoimmunity and neoplastic disease, where loss of MHC expression may allow tumours to escape immune surveillance. It has been over twenty years since the first reports appeared linking genes in the HLA region with disease. The literature now contains many hundreds of examples confirming an increased frequency of different HLA antigens in a diverse spectrum of conditions, ranging from narcolepsy to diabetes (Powis & Trowsdale, 1991). Little progress however, has been made in elucidating the basis for such associations, and they may result from the HLA products themselves, their regulation, or from linked non-HLA genes (Trowsdale, 1987).

One possibility is that the selective nature of the peptide-binding groove of the MHC ensures that only certain alleles are capable of presenting pathogenic peptides. Evidence also suggests that some of the immune response phenomenon associated with MHC alleles could be a result of competition for peptide by MHC alleles (George *et al.*, 1995; Singh *et al.*, 1997).

Disease linkage may not necessarily be related to the peptide binding function of MHC molecules. Juvenile chronic arthritis shows some association with polymorphisms of the DQA promoter region (George *et al.*, 1995). Aberrant expression of genes within the HLA-DQA region could therefore be related to disease pathogenesis. In addition, the density of MHC molecules on the cell surface can also influence immune responsiveness. If not enough antigen is presented to stimulate T-cells, it can cause T-cell anergy, and loss of function. This may be a result of several factors, including a

lower ability by particular MHC molecules to stably bind and express antigenic peptides, problems in the transport of MHC-peptide complexes to the cell surface, or because of a short half-life of the MHC-peptide complex while at the cell surface (Germain, 1994; George *et al.*, 1995).

Studies associating HLA with disease should be interpreted with caution as a result of the phenomenon of "linkage disequilibrium". Linkage disequilibrium occurs when genetic loci are found together in a population at a higher frequency than would be predicted by their individual frequencies. One important consequence of this phenomenon is that observed disease associations with certain HLA alleles may in fact reflect a stronger association with another gene which is in linkage disequilibrium with the first (Powis & Trowsdale, 1991). For example, MHC-DR genes are in strong linkage disequilibrium with MHC-DQ genes. Any disease found closely associated with DR genes may in fact be more closely associated with DQ genes as a result of linkage disequilibrium, and vice versa.

The associations observed between the HLA and disease are never absolute. Even in the diseases, which show the strongest relative risk, individuals exist who, have the disease but do not possess the associated antigen. The converse is also true: the vast majority of individuals with a specific HLA antigen do not develop a disease association with that antigen. Furthermore, none of the diseases segregate in a classical Mendelian manner. This simply indicates that MHC-associated diseases develop as a result of complex genetic and environmental factors. There are most certainly other genes contributing to each disease, and it is therefore best to think of the MHC genes as disease susceptibility genes (Powis & Trowsdale, 1991).

The Ovine Major Histocompatibility Complex

The sheep equivalent of the HLA is the Ovine Leucocyte Antigen (OLA). Although, studies of the genetic organisation of the MHC region in species other than man and mouse are limited, experiments in sheep have detected the presence of several MHC genes.

Class I and II homologues have been identified and are polymorphic, suggesting that the OLA has a similar organisation, and is as complex as that seen in man. The genes encoding the OLA have been found on chromosome 20. Three class I polymorphic loci, designated OLA-A, -B, and –C have been described. The class II region has been examined using cDNA probes, with results indicating that there are at least four "classical" class II loci equivalent to human DQA, DQB, DRA, and DRB genes, and two "non-classical" loci homologous to HLA-, DNA, and DOB. In addition, two other class II loci have been found in ruminants DIB and DYB, whose function remains undetermined (Scott *et al.*, 1987; Amills *et al.*, 1998).

The OLA class II region has at least seven and ten different α-chain and β-chain genes, respectively, and 14 β-chain related sequences (Amills *et al.*, 1998). No HLA-DP equivalent sub-region has been detected in sheep, nor in any other ruminant species to date, and it is thought to have been completely lost (Scott *et al.*, 1987; Deverson *et al.*, 1991; Amills *et al.*, 1998).

One of the most notable differences in the genomic organisation of the MHCs of ruminants compared with humans and mice is the splitting of the class II region into two sub-regions which are separated by at least 15cM (centiMorgans). The class IIa sub-region comprises two clusters of genes, DR and DQ. No recombination has been reported between DR and DQ in sheep (Escayg, 1995). In sheep and cattle DR and DQ are in close proximity to each other and are in strong linkage phase disequilibrium. The class IIb region includes the DMA, DMB, LMP2, LMP7, and TAP genes, which are

involved in antigen processing and transport, and other class II-like genes (non-classical genes), such as DNA, DOB, DIB, and DYB (Figure 1.8) (Amills *et al.*, 1998).

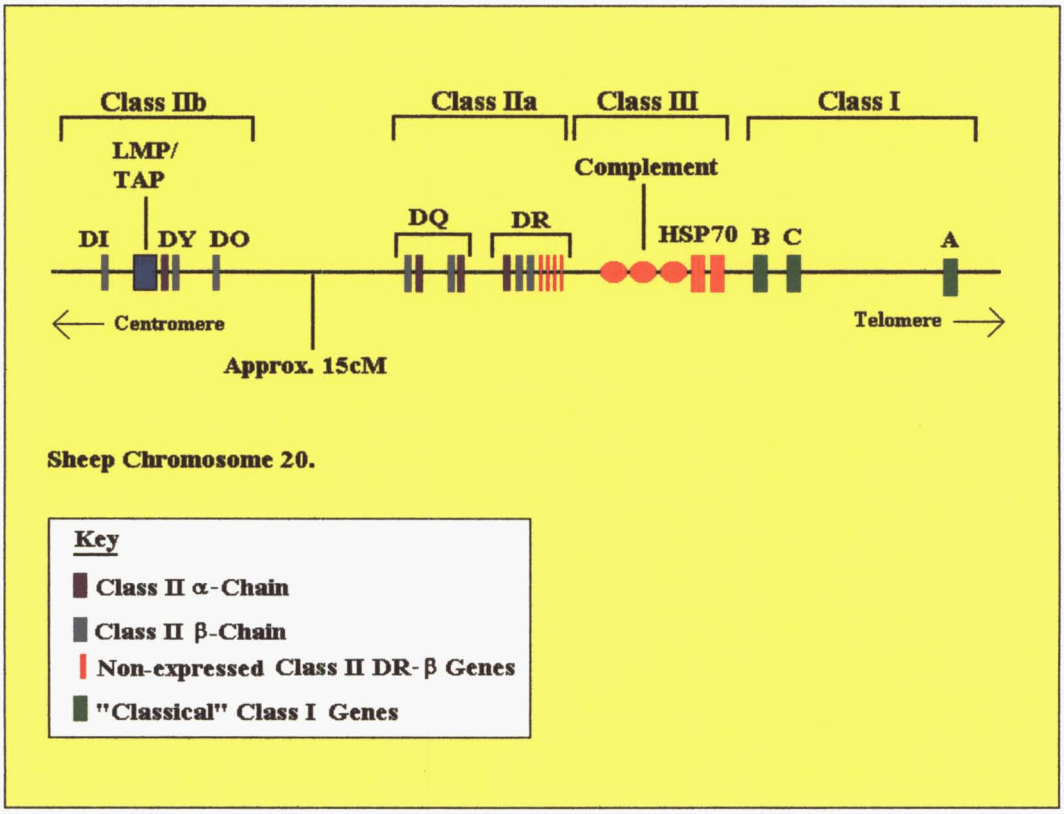


Figure 1.8: Simplified genomic map of the ovine Major Histocompatibility Complex. Three class I genes are transcribed in sheep, and are polymorphic with five distinct class I sequences identified to date. The class II region is divided into two, class IIa and IIb. Class IIa contains the "classical" class II genes DR and DQ, which are closely linked. In sheep one DRα gene has been identified, while the number of DRβ genes varies depending on the haplotype (as many as six genes have been observed), only two are thought to be transcribed (Amills *et al.*, 1998). Class IIb contains the genes involved in antigen processing and transport, LMP and TAP, and other non-classical class II genes, which have only been found in ruminants such as DYA, DYB and DIB. Largely these genes are monomorphic, and their function remains to be elucidated.

DY and DIB genes have only been found in ruminants and exhibit a low level of polymorphism. Interestingly, the distance between the class IIa and IIb sub-regions varies among individuals, and may be a result of the existence of a polymorphic recombinational hotspot (Amills *et al.*, 1998).

Analysis of DRB polymorphism has been useful for inferring the evolutionary history of the MHC in ruminant species, suggesting that selection, genetic drift, and population bottlenecks have played an important role in determining the repertoire of ruminant MHC class I and II alleles. In addition, evidence in ruminants suggests that interallelic exchange (microrecombination) of short sequence motifs has been of importance in the generation of allelic variability (Amills *et al.*, 1998).

The Ovine DQ Sub-region

The ovine DQ region encompasses 130kb, with the DQA_1 and DQA_2 sub-regions located 22kb apart. The number of DQA and DQB genes in sheep appears to vary depending on haplotype (Amills *et al.*, 1998). Using RFLP typing techniques it has been observed that up to 20% of animals analysed carry a DQA_1-null haplotype, and it is one of the most common haplotypes observed in the flocks studied to date (Fabb *et al.*, 1993; Escayg *et al.*, 1996; Snibson *et al.*, 1998). As a result of the common occurrence of the DQA_1-null haplotype the significance of this locus in sheep has been brought into question. Unlike humans both DQA genes are transcribed. There is evidence both in vitro (Wright & Ballingall, 1994) and *in vivo* (Scott *et al.*, 1991a) for DQA_1 and DQA_2 transcription. However, cell surface expression of DQ products has only been detected for the DQ1 locus. Inability to detect surface expression of DQ2 products may however, be a result of the lack of a suitable monoclonal antibody, and further investigation is required (Snibson *et al.*, 1998).

In humans the DR sub-region has been most characterised, and an extremely high level of polymorphism has been observed with DRB alleles. Most disease associations have been associated with the DR sub-region, and for this reason is considered to be the most important in humans. In contrast, in ruminants the DQ sub-region is thought by some investigators to be of primary importance for antigen presentation (Escayg *et al.*, 1996). The fact that a DP sub-region has not been observed in ruminants, combined with the high level of polymorphism in both DQA and DQB genes (van Oorschot *et al.*, 1994),

and the expression of both DQA_1 and DQA_2 loci suggests that the DQ sub-region is of greater importance (Escayg *et al.*, 1996; Ballingall *et al.*, 1997; Snibson *et al.*, 1998). In addition, whilst the DR sub-region in ruminants is polymorphic, the extreme level of polymorphism observed in humans has not been matched, with conflicting reports as to the level of polymorphism at the DR locus in sheep (Scott *et al.*, 1987; Scott *et al.*, 1991b; Grain *et al.*, 1993; Dutia *et al.*, 1994; Escayg *et al.*, 1996). This suggests that DR may not have such a significant role in antigen presentation in sheep (Escayg *et al.*, 1996).

Again, similarly to the HLA strong linkage disequilibrium is observed between OLA-DR and OLA-DQ genes, with no recombination reported between the two regions (Escayg *et al.*, 1996; Amills *et al.*, 1998). Any disease found closely associated with DR may in fact be more closely related to DQ, and vice versa. From the evidence obtained to date however, it appears that unlike humans where DR genes seem to perform the primary antigen-presenting function, both DR and DQ in ruminants may be equally important, and any disease associations between either sub-region may depend on the pathogen.

1.2 Footrot

Footrot is one of the most economically important diseases affecting the sheep industry in temperate climates that enjoy good rainfall. The disease, predominantly caused by the bacterium *Dichelobacter nodosus* (formerly *Bacteroides nodosus*), is highly contagious, and is most commonly found in sheep and goats (Skerman *et al.*, 1988; Egerton & Raadsma, 1989; Ghimire & Egerton, 1996). In New Zealand, footrot has considerable economic impact with an estimated cost to the agricultural industry of over 100 million dollars per annum. Economic impact is determined by the prevalence and length of severe infections, with on farm and environmental factors contributing to this (Figure 1.9).

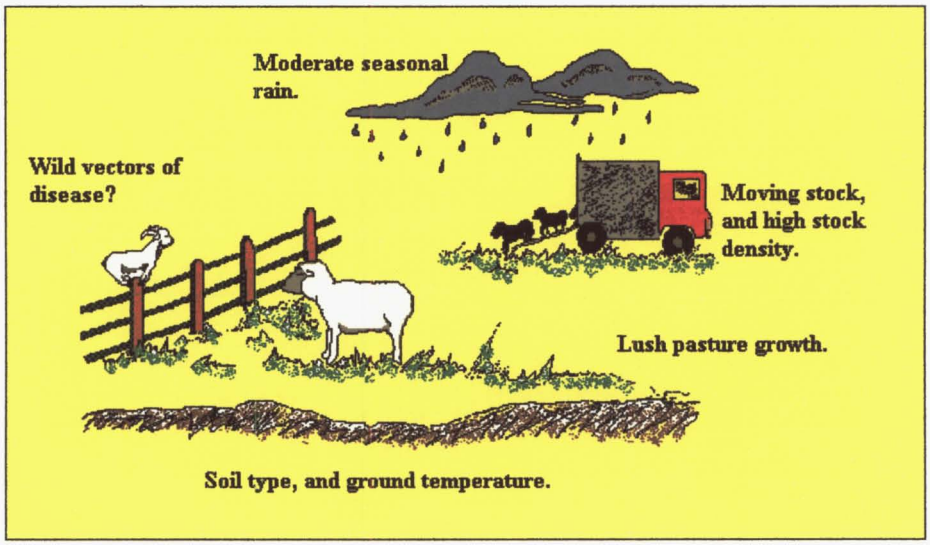


Figure 1.9: Factors contributing to the incidence and severity of footrot in sheep in New Zealand. Bacteria invade the hoof under appropriate environmental conditions-usually where there is moderate seasonal rain (usually above 500mm), and when lush pasture is produced during the more humid months of the year, and average ground temperature is above 10°C. The disease has a seasonal occurrence with outbreaks predominantly in the late autumn and spring. However, they can occur in summer on irrigated pastures, or during mild winters. High stocking densities contribute to the spread of the disease. In addition, improved pasture, and poorly drained soil with a high pH, increases the incidence and spread of the disease. Transmission does not occur easily after the pasture dries, and is substantially reduced during drought. Stock movement may also promote outbreaks, as animals may be carrying the bacterium, but show no disease symptoms. Overgrowth, or defects in the shape of the hoof can render animals more susceptible, and may act as a potential reservoir for the bacterium, providing a source of infection when environmental conditions allow transmission. Goats can act as disease vectors, but little is known about the role of other animals such as rabbits and opossums. Sheep breed also affects the severity of the disease, with coarser wool breeds being more resistant than finer wool breeds.

Pathology of *Dichelobacter nodosus*

Ovine footrot is a mixed bacterial dermatitis and is considered to result from the synergistic action of several bacterial species that are individually incapable of causing the disease (Stewart, 1989). *D. nodosus* is considered to be the specific causative pathogen since it is the only organism of the footrot flora capable of reproducing the disease when applied as a pure culture (Thomas, 1962). This anaerobic bacillus lives only in the diseased hoof, and survives for no longer than seven to fourteen days in faeces, soil or pasture (Stewart, 1989). *D. nodosus* invades the hoof/tissue junction, and

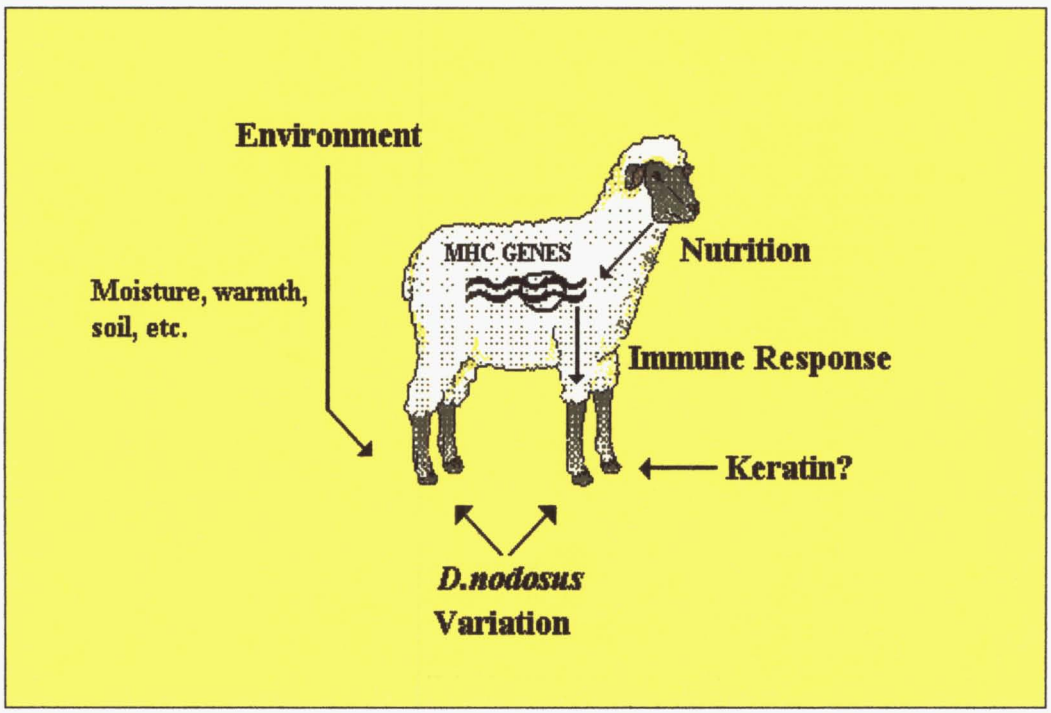
degrades proteins in this region causing lesions or "rot". Generally *D. nodosus* cannot invade the interdigital skin unless it has been damaged. Often however, the bacterium responsible for foot scald, *Fusobacterium necrophorum* colonises the moist epidermal surface at the skin-horn junction first, which enables subsequent entry of *D. nodosus* (Egerton *et al.*, 1969; Kimberling, 1988).

The type of footrot infection can be divided into two broad categories: virulent and benign, and is based upon an animals response to infection (Egerton & Raadsma, 1989). Virulent footrot is characterised by extensive separation of the hoof, which commences at the heel, and extends to the sole and toe, with the formation of a distinctive, foul-smelling, necrotic exudate (Stewart, 1989). Virulent infection causes severe lameness, with underrunning of the hoof, and severe inflammation. Usually both digits, and more than one foot are affected. If left untreated the animal may die, as it is unable to graze properly. In addition to causing severe lameness, infected animals also lose body weight, have decreased wool growth often accompanied by wool break, lambing percentages decrease, and overall stock value decreases (Stewart, 1989). Generally, virulent footrot is not as common in coarse wool breeds (e.g. Romneys) compared to fine wool breeds (e.g. Merinos), where the disease can cause severe hoof damage (Patterson & Patterson, 1991).

Benign footrot is a less persistent condition, in which a slight dermatitis is the only lesion observed. There is little or no underrunning or accumulation of necrotic exudate (Boundy, 1983). Lameness is usually only mild, not readily apparent, and rapidly disappears as the season dries off. The effect on production is minimal.

Between breed, as well as within breed differences in the susceptibility of sheep to footrot are well-recognised (Emery *et al.*, 1984). A number of heritable factors may be involved, and many of these interact with the environment to affect disease status (Figure 1.10). Susceptibility to footrot can be influenced by a number of factors. The ability to mount a successful immune response is important when infection occurs. The immune response can be affected by an individuals genetic make-up, part of which is

determined by an individuals complement of MHC genes. Additionally, nutrition and the environment can affect the immune response. In humans poor nutrition has been associated with poor immune responsiveness (reviewed Chandra, 1989). Also, if the individual is placed under stress by environmental conditions, immune response is compromised. Other factors that affect an animal's susceptibility to footrot include the strain of *D.nodosus*, with some strains being more virulent than others (Stewart, 1989). There may also be a link with variation in keratin genes, which are involved in wool growth; however, investigation is required to prove if there is any association. The expressed phenotype (ie. footrot susceptibility or resistance) is therefore, a complex interaction between genes (of the individual and pathogen), and the environment.



P = G x E

Phenotype = Genes x Environment.

Figure 1.10: Factors affecting susceptibility to footrot.

Genetic Markers for Footrot Resistance

Currently there are several treatment options for footrot including: Topical agents such as antibiotics; footbathing in chemical solutions of either formalin, zinc sulphate, or copper sulphate; and the use of vaccines to boost an individuals own immune response (Skerman *et al.*, 1988). A combination of two or more types of treatment may be necessary to get the disease to manageable levels. At present, footbathing combined with vaccination is the most widely used method of treating severe footrot in New Zealand. Regardless of approach, the treatment of footrot is time consuming and expensive, including expenditure on labour, equipment, vaccines, antibiotics and chemicals.

Most of the above treatment options have their limitations, and once footrot is established on farm, it is extremely difficult to get rid of it. The use of antibiotics and chemicals to treat the disease is becoming less accepted because of the possible impact on consumer health and the environment. Nevertheless it should be noted that parenteral treatment with antibiotics is a practice commonly followed by some producers, and recommended by some researchers (M^{c}Farlane.R, personal communication). In addition, the vaccines that are currently available are not as effective as originally hoped, providing very little protection in the finer wool breeds.

One option that has not been fully investigated is the possibility of controlling footrot by exploiting natural variation in resistance. Natural resistance to footrot has been a well-documented phenomenon (Egerton *et al.*, 1983; Emery *et al.*, 1984; Stewart *et al.*, 1985; Skerman, 1986; Outterridge *et al.*, 1989; Whittington & Nicolls, 1995). Selecting animals with a higher natural footrot resistance provides an alternative strategy for reducing the impact of the disease, and decreases the need for chemical intervention.

Nevertheless, a problem that arises is how do you select animals with a naturally higher disease resistance? Improving disease resistance in farm animals by traditional phenotype selection has been difficult. Phenotypic selection relies on good field measurement of disease status to identify animals that are naturally more resistant. If such assessment relies on natural exposure to disease, then the levels of exposure may be

greatly dissimilar among individuals, such that most of the phenotypic variation will be a result of the differences in the degree of challenge. If an artificial challenge is used to assess disease resistance the costs of setting up a suitable system and the direct costs of productivity loss as a result of disease development are high (Soller & Andersson, 1998).

This type of strategy also raises important ethical questions, because it involves exposing at least some members of the population to the pathogen responsible for the disease (Nicholas, 1991). Furthermore, in the framework of a commercial breeding program disease exposure and disease development could affect productivity such that selection for disease resistance occurs at the expense of selection for productivity (Soller & Andersson, 1998).

An alternative strategy to identify naturally disease resistant animals would be a genomic approach. In genomic approaches to the improvement of disease resistance, the criterion for selection is shifted from phenotypically expressed disease status to allele status at the DNA level (Soller & Andersson, 1998). Identifying at least part of the genetic variation in resistance, by the means of one or more genetic markers would provide a powerful supplement to the classical phenotypic breeding strategies. If such markers could be developed, it would be possible to readily identify animals that are naturally more resistant to the disease, at an early age, without the need for exposing animals to the disease (Rood & Yong, 1989; Drinkwater & Hetzel, 1990; Nicholas, 1991).

One way of identifying a genetic marker for a specific trait, such as disease resistance is to use the candidate gene approach. Candidate genes for a trait are genes that are directly involved in the physiology or development of the trait. Thus candidate genes are identified on the basis of their known function.

The broad spectrum of resistance to footrot that has been observed is likely to be a function of variability in some component of the immune response. Supporting this is

the observation that animals show variation in response to vaccination, implying that variation in immune response is important in determining an animal's disease susceptibility (Skerman *et al.*, 1982; Lewis *et al.*, 1989; Raadsma *et al.*, 1990). In addition, no undesirable genetic correlation has been detected between resistance to footrot and other major production objectives (Raadsma & Egerton, 1993).

As a result of its central role in antigen presentation, and highly polymorphic nature, the OLA presents a particularly good candidate for being the genetic region driving this variation in natural resistance. In addition, studies investigating parasite resistance in sheep has illustrated the potential ability of OLA genes to act as genetic markers enabling the selection of naturally resistant animals (Outteridge, 1991; Hulme *et al.*, 1993; Hohenhaus & Outteridge, 1995; Schwaiger *et al.*, 1995).

The Ovine MHC and Footrot

The first evidence of a positive correlation between resistance to ovine footrot and particular lymphocyte antigens was produced by Outteridge *et al.*, (1989), who demonstrated that class I molecules SY1b and SY6 may be important in some aspect of footrot resistance. It was shown that SY1b and SY6 were more commonly found in sheep without footrot. SY6 was found to be associated with a high sub-group specific response to vaccination with *D.nodosus* strain 198 (serotype A, subgroup 1). SY1b had been weakly associated with natural resistance to footrot.

The extracellular nature of *D.nodosus* immediately suggested that class II OLA would be responsible for the presentation of antigenic determinants of this pathogen. The availability of cDNA probes for class II genes from human and mice made it possible to screen the sheep genome for orthologous class II genes. Human probes corresponding to the ovine DQA and DRB regions, and Southern hybridisation, were employed by Litchfield *et al.*, (1993), to perform RFLP analysis of class II OLA molecules, and to determine if any associated with footrot susceptibility/resistance under artificial challenge with virulent footrot strains. Two potentially useful associations were found.

The first was the identification of a significant ($P<0.05$) association between a DRB/*Taq*I 8.3 kb fragment, and increased susceptibility to footrot. Secondly, a DRB/*Taq*I 2.65 kb fragment was reported as having an (unspecified) effect on antibody response after vaccination.

These associations held sufficient promise of eventual identification of OLA-derived genetic markers for footrot resistance, prompting further study.

1.3 Investigating Class II OLA Genes and Associations with Natural Footrot Resistance: Research at Lincoln University

To determine whether genetic polymorphism within OLA-class II loci was associated with footrot susceptibility/resistance, Andrew Escayg, using RFLP analysis and various OLA-class II probes investigated the allelic polymorphism within different flocks of sheep as part of his PhD thesis. The DQA, DQB, DRA and DRB loci of approximately 250 animals were investigated, resulting in the identification of $8DQA_1$, $15DQA_2$, 5DQB and 3DRA alleles (Escayg, 1995).

The high level of polymorphism observed at the DQA_2 locus favoured further investigation into the possible correlation of DQA_2 alleles with variation in susceptibility/resistance to footrot. A significant association ($P<0.05$) was observed within one flock of Corriedale sheep in North Canterbury, when the flock had been subjected to a standardised footrot field challenge (Escayg, 1995; Escayg *et al.*, 1997). The sire had been found to be heterozygous at the DQA_2 locus with alleles E and H (*Taq*I RFLP and Southern hybridisation) present. Segregation analysis of the sires' progeny revealed that the E allele had a frequency of 0.1 in resistant animals, and 0.7 in susceptible animals, while allele H had a frequency of 0.9 in resistant and 0.3 in susceptible animals (Escayg, 1995; Escayg *et al.*, 1997).

As the presence of these alleles correlated with either footrot susceptibility or resistance, they had the potential to act as genetic markers. This would allow an accurate prediction

of an individual animal's footrot susceptibility from its genotype. Selecting animals that carry the resistant allele provides a rational genetic basis on which to formulate breeding strategies, which would enable a rapid increase of the natural resistance to footrot in a flock (Rood & Yong, 1989).

Whilst this work provided some evidence that the DQA_2 locus could be involved in natural resistance and was a good candidate as a genetic marker for footrot susceptibility/resistance, further investigation was required to determine if similar associations could be found within other lines, flocks and breed of sheep.

The Aim of this Thesis was to:

1) Investigate the association of particular OLA-DQA_2 alleles and susceptibility/resistance to footrot in other lines, flocks and breed of sheep.

2) To further investigate OLA-DQA2 polymorphism by cloning and sequencing as many of the 15 DQA_2 alleles defined by Escayg (Escayg, 1995; Escayg *et al.*, 1996), enabling the identification of any new alleles, and to determine whether the same allele typed using RFLP and Southern Hybridisation has exactly the same sequence amongst different breeds.

3) To correlate ovine DQA_2 polymorphism with variation in the antigen-binding groove by modelling based on the known structure of human-DR molecules.

CHAPTER TWO

RFLP AND SOUTHERN HYBRIDISATION TYPING OF THE OVINE-MHC DQA_2 GENE AND ASSOCIATIONS WITH FOOTROT.

Introduction

Research has demonstrated an association between the DQA_2 E allele and increased susceptibility to footrot (Escayg, 1995; Escayg *et al.*, 1997). This was in a single sire-line of Corriedale sheep, and its applicability to other breeds and lines was unconfirmed. In this chapter RFLP and Southern hybridisation analysis was used to determine whether similar associations existed between DQA_2 alleles and footrot susceptibility or resistance within other lines, flocks, and breeds of sheep.

This chapter begins with an overview of the animals investigated, and includes the criteria used for selection, descriptions of the specific flocks analysed, the challenge protocol, and how the severity of infection was recorded.

2.1 Flocks Investigated, Criteria, and Challenge Protocol: An Overview

Criteria for the Selection of Animals:

Large paternal half-sib families with accurate and complete progeny records are required. The sire must also be heterozygous at the DQA_2 locus before a segregation analysis can be performed, and necessitates the need for the sire to be typed first. In addition, to obtain accurate pedigree information each sire line must be single-sire mated. Thus, single sire matings, and good paternity data were a prerequisite before half-sib families could be included in this study.

The distinction between disease resistance and susceptibility among individual animals can only be achieved under footrot challenge. Suitable flocks must be either undergoing a natural challenge, or be deliberately exposed to footrot.

The challenge, whether it is natural or artificial needs to be uniform to be able to reliably differentiate between susceptible and resistant individuals. The challenge conditions must be intensive, and maintained over a sufficient period of time (Escayg, 1995). However, economic losses incurred as a result of a footrot challenge precludes its tolerance in commercial flocks. In addition, the practices, which favour the spread and transmission of footrot, are contrary to good farm management practices. Consequently, studies are limited to properties with either a high prevalence of footrot, or to those in which a challenge can be deliberately imposed (Skerman *et al*, 1988).

As a result of the above criteria, many commercial flocks are unsuitable for such investigations. However, flocks which could be useful are those which are used to produce stud animals. Within stud flocks accurate pedigrees are usually kept, and all animals can be identified through ear tags. Farmers that were interested in improving resistance to footrot within their flocks, and who either had a recurring footrot problem, or were not adverse to challenging animals with the disease were contacted. Flocks that met the necessary criteria were then investigated further.

Flocks Analysed:

Three sheep breeds were analysed in detail, and these originated from six different geographical locations (Figure 2.1). The breeds analysed were Awassi, Corriedale and Merino. Initially samples were collected from the sires and were typed at the DQA_2 locus to identify heterozygous animals suitable for analysis.

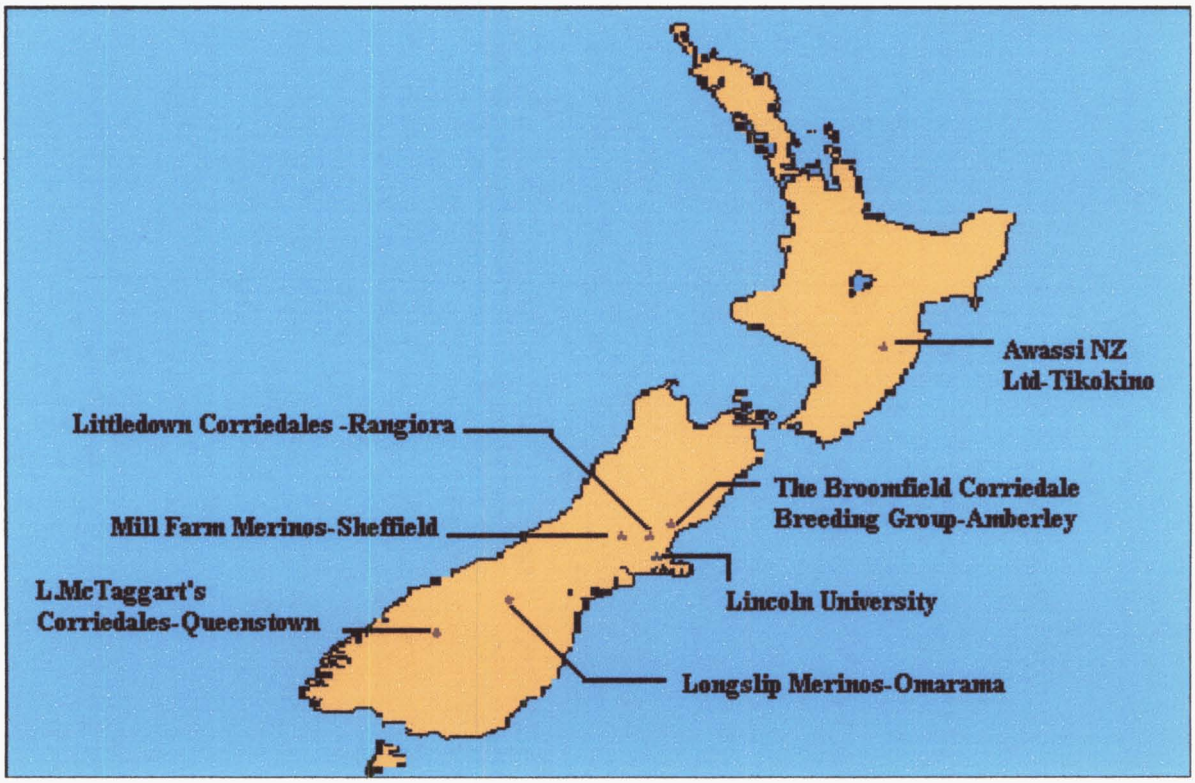


Figure 2.1: Geographical locations of the farms investigated.

1) Awassi NZ Ltd.

Introduction

Awassi sheep are large framed, fat tailed sheep, mostly bred for meat production in the Middle East countries. It is the dominant breed in Syria, Lebanon, Iraq, Jordan, Kuwait, and much of the Saudi Arabian peninsular, where it is highly valued for the production of meat, carpet wool, skins and dairy products (Lightfoot, 1988; Kingwell *et al.*, 1995).

The Awassi flock used in this study is situated at Tikokino in the North Island. The foundation flock was established in 1990/91 and initially consisted of fourteen purebred animals, seven of each sex. Footrot was a major limitation to production, a large proportion, (up to 60-70 %) at any one time was affected although some considerably worse than others.

Awassi originating in arid and semi-arid areas have probably not been extensively exposed throughout their evolution to footrot, nor been bred in an environment which is conducive to *D. nodosus* survival and transmission. However, in one case where a footrot outbreak has been recorded in Israel, Awassi sheep were more susceptible compared to Awassi-East Friesian crosses (Shimshony, 1988).

As a result of Awassi NZ Ltd being a commercial farm it was impossible to deliberately challenge animals with footrot. Nevertheless, the high natural occurrence of the disease within the flock was considered to be sufficient for this investigation.

Five other flocks were investigated, Corriedale flocks 1, 2 and 3, and Merino flocks 1 and 2, details of each flock are listed in Table 2.1.

The Challenge Procedure:

The process by which an individual's footrot susceptibility or resistance is determined is by deliberately exposing the animal to the disease and is referred to as a "challenge". The challenge can be accomplished by either artificial or natural procedures.

Artificial challenges are often performed by applying pure culture of one specific serogroup of *D. nodosus* to wet interdigital skin in one or more feet. The culture is maintained by bandaging the foot for up to five days before returning the animals to the pasture. The incidence of the infection is then monitored over an appropriate period (Escayg, 1995).

Alternatively, a natural challenge can be achieved by establishing conditions that are conducive to the spread and transmission of footrot. Typically, a number of severely infected animals are introduced to the flock, which are subsequently all grazed together. This allows infection to spread from infected to uninfected animals. This may be promoted by maintaining high stocking densities on well-irrigated pasture. In addition, unrelated "control" animals can be introduced to the flock during a footrot challenge

from farms that have no history of footrot. This provides an internal means of determining the severity of the challenge, while also providing an indication of the relative resistance of different lines.

As a result of the large numbers of animals, and the scattered distribution of farms involved in this study, artificial challenges were impractical not to mention expensive and technically demanding. There was also the risk that differences in disease susceptibility under experimental induction may not have reflected differences under natural conditions (Escayg, 1995). In addition, whilst other studies have used artificial challenges (Emery *et al.*, 1984), they have often challenged animals with only one strain or serogroup of *D. nodosus*. Such a challenge does not adequately mimic "on farm" conditions. Recent analysis has shown that up to five different serogroups of *D. nodosus*, and up to seven different strains from each serogroup have been identified on one hoof in some New Zealand flocks (Zhou & Hickford, 2000). Natural challenges were therefore used to assess animals' footrot resistance or susceptibility. The disadvantages of such an approach however, is that the effectiveness of the challenge is highly dependent on the ability to establish and maintain conditions that favour the spread of the bacterium, and that the initial exposure to *D. nodosus* cannot be standardised. Nevertheless, this can be balanced against the fact that a natural challenge is likely to reflect the responsiveness of animals under normal field conditions, and the initiation of such a challenge is less technically demanding and consequently more widely applicable.

Not all flocks could be formally challenged, instead the normal incidence of footrot on the farm was used to determine an animal's relative susceptibility or resistance. This has been termed "normal field exposure". The challenge procedure implemented for each flock is described in Table 2.1.

The Footrot Scoring System:

After a footrot challenge it was necessary to inspect each animal for signs of infection. Each farmer had their own system for scoring the severity of footrot infection, and scoring was usually carried out on farm and records were kept. The scoring systems employed by the farmers had a range to be able to delineate the severity of infection.

For the purposes of this study footrot status was recorded as a "yes" or a "no". An animal that was "yes" meant that infection had underrun the horn of the hoof, and had been persistent for more than two weeks in one or more feet. An animal that was "no" meant that either there was no sign of footrot infection, or infection was a mild scald that was not persistent, or infection had not underrun the horn of the hoof.

The rational behind the "yes" or "no" criteria was that for the immune system and hence MHC genes to be involved in the observed footrot resistance or susceptibility the infection had to have been underrun. It has been observed that the humoral immune response against *D. nodosus* has not been elevated until underrunning lesions of the hoof have developed (Emery *et al.*, 1984). In addition, it takes two weeks to get an immune response by the specific (MHC-based) immune system, therefore at least two weeks of persistent infection is required (Escayg *et al.*, 1997). After two weeks, if an appropriate MHC-restricted immune response against *D. nodosus* is mounted the severity of infection should diminish.

As each farmer had their own footrot scoring system, it was possible that there could be some observer bias when categorising animals into the "yes" and "no" groups. Every effort was made to limit the impact of any observer bias, and to standardise the scoring between farms. The scoring system that was employed by the farmer was discussed with them at length. Accurate notes had been kept as to how each scoring system operated, and the criteria for each category, such that animals that had a particular score could be placed in either our "yes" or "no" categories. This type of approach meant that statistical

analysis could be simplified, and farmers did not have to change to a new footrot scoring system.

In Merino flock 1, during the 1998-99 challenge, some animals exhibited strange hoof morphology, similar to having had a severe footrot infection, but no actual sign of footrot was found. This made defining them within the simple "yes" and "no" categories difficult, and they were given an intermediate score. In such cases where these animals occurred, statistical analysis was performed twice, one where these animals were grouped with the susceptible animals (the "yes" group), and one where they were grouped with the resistant animals (the "no" group).

Table 2.1: Summary of flock data and type of challenge implemented for each flock used in this study

Flock Name	Location	Number of Sires Investigated	Number of Half-Sibs Analysed	Number of Progeny per Line	Type of Challenge
Awassi NZ Ltd * (J and A Molloy)	Tikokino	10	4		
			-Sire 1046/91	58	Normal Field Exposure (Tikokino-1997)
			" 1036/91	35	
			" 1039/91	48	
			" 4071/94	20	
Corriedale Flock 1 † (McTaggart)	Queenstown	3	3		
			-Sire 216/92	15	Not Challenged
			" 6/93	2	
			" 7688	8	
Corriedale Flock 2 † (Dixon)	Rangiora	3	2		
			-Sire 281/93	64	Not Challenged
			" 89/93	14	
Corriedale Flock 3 ** (Orr-BCBG)	Amberly	15	8		
			-Sire 3547/92	29	Normal Field Exposure (Te Anau-1998)
			" 8661/95	31	
			" 7073/95	20	
			" 7472/94	30	
			" 7271/94	41	
			" 8215/95	22	
			" 5179/93	14	
			" 3433/ 92	20	

Merino Flock 1‡ (R Patterson)	Omarama	48 (1996) 23 (1998)	6 -Sire 10	49	Natural Challenge (Longslip-1997)
			" 9	34	
			" 380	33	
			-Sire 8	38	Natural Challenge (Lincoln University-1998-1999)
			" 101	64	
			" 4403	33	
Merino Flock 2 ‡‡ (R and J Mulvey)	Sheffield	0	1	27 Rams	Natural Challenge (Mill Farm-1998)
				28 Ewes	Normal Field Exposure (Mill Farm-1998)

* Both ewe and ram lambs were analysed from the Awassi NZ Ltd Flock. All animals were approximately 8-10 months old.

† Both Corriedale flocks 1 and 2 proved to be unsuitable for continued analysis after the initial screening of the sire samples. Very little genetic variation was observed at the DQA_2 locus, this combined with the small number of progeny in some of the sire lines would make the statistical analysis unreliable. As a result, these two flocks were not analysed further.

** Ram lambs of the Broomfield Corriedale Breeding Group (BCBG) have traditionally been challenged more extensively than ewe lambs, and for the purposes of this study, progeny samples were obtained from the ram lambs only. In Canterbury in 1997 and 1998, the region experienced severe drought. Environmental conditions did not allow the maintenance or transmission of *D. nodosus* within the flock. Instead the animals were moved to Te Anau for grazing, where environmental conditions allowed a footrot challenge. This challenge was not formally implemented, and was a reflection of the normal incidence of the disease on the farm. Thus, normal field exposure was used to differentiate between resistant and susceptible animals.

‡ Initially samples from 48 sires were collected and typed at the DQA_2 locus in 1996, and a further 23 sire samples were collected in 1998. Progeny (both ewes and rams) from different sire lines were challenged in two separate trials. One was performed in 1997 at Longslip, (animals approximately 8-10 months old) and the other was performed in 1998-99 at Lincoln University (animals approximately 6-8 months old). In both trials the progeny were subjected to an intense natural challenge. Animals with footrot were introduced into the flock and a high stocking density on well-irrigated pasture promoted transmission. Animals were challenged three to four weeks, and control animals were included in each trial.

‡‡ The sire had died several months before analysis of this flock, making it impossible to get a DNA sample. Initial screening of this line involved the collection of 55 progeny samples. It was hoped that by typing the progeny, the sire's haplotype could be determined. This flock was divided into two groups, one group consisting of the ram lambs, the other the ewes. The ram lambs were formally challenged using the natural challenge procedure, and control animals were included. The ewe lambs were not formally challenged, and the incidence of footrot was solely a result of the normal incidence on farm. As a result of the different challenges both groups were analysed separately

2.2 Methods

Sample Collection: Peripheral blood was collected by venopuncture of the jugular vein using 10 mL vaccutainer tubes (Becton Dickenson Vaccutainer Systems Europe, England), with lithium heparin as an anticoagulant. Tubes were inverted several times to ensure mixing and prevent the formation of blood clots. Samples were transported on ice and stored at 4°C for no more than 48 hours after collection.

DNA Extraction: A high salting out procedure was used for DNA extraction from nucleated cells, as described by Montgomery and Sise (1990) (see Appendix C).

DNA Digestion and Electrophoresis: DNA Samples (10 µg) were digested using *Taq*I restriction endonuclease enzyme (3-4 units/µg), in buffer supplied by the manufacturer (Boehringer Mannheim/Roche Diagnostics, Mannheim, Germany). Digestion was assisted by the addition of spermidine to a final concentration of 4 mM. Digests were incubated at the recommended temperature for 10 - 12 hours. Completeness of the DNA digestions was assessed on 1% agarose (Seakem LE agarose, FMC Bioproducts, Rockland Maine, U.S.A.) minigels containing 1X TBE gel buffer (0.089 M Tris, 0.089 M Boric Acid (BDH, Poole, Dorset, England), 0.002 M Na_2EDTA, 200 ng/mL eithidium bromide), in a mini submarine agarose gel unit (mini-sub cell GT, BioRad Laboratories, Hercules, California, U.S.A), and run at approximately 10 volts/ cm of gel/ hour, with ultraviolet illumination of the gel enabling verification of completed digests.

Digested samples were then run on larger preparative 0.8 % agarose gels, (Max Submarine Unit Model HE 99X, Hoefer Scientific Instruments, San Francisco, U.S.A., or Sub-cell GT, BioRad) using a 15 or 20 well comb, at approximately 1.5 volts/cm of gel for 18-22 hours. A molecular weight marker of Lambda phage DNA digested with *Hind* III (50 ng/µL) (Gibco BRL, Gaithersburg, MD, USA), was run adjacent to the electrophoresed DNA enabling size calibration of DNA fragments. Electrophoretic mobility was visualised by UV illumination of the gel and photographed with a

transparent ruler alongside the gel. This allowed the migration of any band of DNA to be measured by comparison with the λ *Hind* III molecular weight marker, and was kept as a permanent copy.

Southern Transfer: The digested DNA was transferred from the agarose gels to nylon membranes, either Hybond-N^{+TM} (Amersham International, Buckinghamshire England), or Zeta-Probe® (BioRad) by an alkaline capillary transfer method that was a modified version of the procedure originally proposed by Southern (Southern, 1975) (see Appendix D).

Description of the Ovine DQA_2 Probe: The ovine DQA_2 probe was obtained from a clone (pC17-2PH), which contained an ovine DQA_2 exon 2 gene insert. The insert could be excised from the clone by digestion with *Eco*R1/*Pst*1 restriction enzymes, yielding a 320 bp fragment. A miniprep of the clone was received from Mr Escayg prior to the beginning of this study (see Appendix E for plasmid map).

Probe Labelling: The probe was radiolabelled using the polymerase chain reaction (PCR). PCR primers were designed from the published sequence of Scott *et al*, (1991a) for the beginning and end of the 320 bp DQA_2 insert in the pC17-2PH clone. The primer sequences were, DQA_2 insert forward 5' AAT TCG AGC TCG GTA CCC CTG ACC 3', and DQA_2 insert reverse 5' GGT CGA CTC TAG AGG ATC CCC ATT 3' (Life Technologies, NY, USA).

Each 25 µL PCR reaction contained 2 µL of diluted (1:10) pC17-2PH DNA estimated to have a concentration between 50-100 ng, 0.35 µM of each primer, 100 nM of the dNTP's dATP, dTTP, dGTP and 25 nM dCTP (Gibco/BRL), 1 x PCR buffer containing 1.5 mM $MgCl_2$, 1.5 U Taq polymerase (Qiagen GmbH, Hilden, Germany), and 3 µL α-^{32}P-dCTP (6000 Ci/mmol) (New England Nuclear, Boston, MA, U.S.A.).

Reactions consisted of 30 cycles of 94°C denaturation for 1 min, 60°C annealing for 1 min, and 72°C extension for 2 min, followed by a final extension step of 72°C for 7 min. All reactions were carried out on a Perkin Elmer 2400 thermocycler (Perkin Elmer Corporation, Branchburg, NJ, U.S.A.), using thin walled 200 µL tubes. A negative control containing no DNA was included in every reaction.

Radiolabelled probe was separated from unincorporated α-^{32}P-dCTP by centrifugation through a small column of sephadex G-50 (Pharmacia, NJ, U.S.A., medium grade) prepared in a 1 mL plastic syringe. The activity of the labelled probe was determined by scintillation counting (specific activity needed to be greater than 4 x 10^7 c.p.m (Escayg A., personal communication)), and normally used the same day, but could be stored at -20°C until required (storage never exceeded 3 days).

Membrane Pre-hybridisation: Membranes were placed in thick walled flat polythene bags with 45-50 mL of pre-hybridisation solution that had been pre-heated to 65°C. All air bubbles were removed, the bag sealed and placed in a shaking water bath at 65°C for at least two hours. One membrane per bag was maintained. The composition of the pre-hybridisation solution differed depending on whether the samples had been transferred onto Hybond-N^{+}™ nylon membrane or Zeta-Probe® nylon membrane. Essentially Zeta-Probe® membrane required a higher concentration of SDS to prevent high background readings. The composition of each pre-hybridisation solution was as follows:

Hybond-N^{+}™ Membranes:	SSPE (x 20)	12.5 mL
	Denhardts reagent (x 50)	5.0 mL
	SDS (10%)	2.5 mL
	Sterile deionised water	30.0 mL
Zeta-Probe® Membranes:	SSPE (x 20)	12.5 mL
	Denhardts reagent (x 50)	5.0 mL
	SDS (10%)	25.0 mL
	Sterile deionised water	7.5 mL

Membrane Hybridisation: To each tube of radiolabelled probe 0.5 mL of fragmented herring sperm DNA (Boehringer Mannheim) (2 mg/mL stock solution) was added, and denatured in a boiling water bath for 10 minutes. The tube was then immediately placed on ice to prevent the DNA strands reannealing. One tube of denatured probe/ herring sperm mixture was added to a bag containing one pre-hybridised membrane. Again any air bubbles were removed, the bag resealed and returned to the shaking 65°C water bath for approximately 18-20 hours.

Washing the Membranes: Membranes were removed from the plastic bags, and placed into a plastic container of a suitable size. In general, up to four membranes were placed in each box, and the washing procedure was the same for both the Hybond-N^{+}™ and Zeta-Probe® membranes. Five hundred mL of each washing solution was sufficient to wash 4 membranes. All washing solutions were preheated to 65°C and the membranes washed with gentle agitation as follows:

1) One wash at 65°C in 2 x SSC, 0.1% SDS for 15 minutes.
2) One wash at 65°C in 2 x SSC, 0.1% SDS for 30 minutes.
3) One wash at 65°C in 0.1 x SSC, 0.5% SDS for 10 minutes.

After the last wash membranes were quickly rinsed in 500 mL of 0.1 x SSC, 0.5% SDS solution. Excess liquid was removed from the membranes by briefly blotting the membrane with clean Whatman 3 MM filter paper, care was taken to avoid completely drying the membrane, as this would have prevented subsequent stripping of the bound probe.

Autoradiography: Each membrane was individually sealed in a thin walled plastic bag, and placed in a film cassette with a BioMax™ MS intensifying screen (Kodak Rochester, New York, USA) and BioMax™ MS film (Kodak). The cassettes were stored at -80°C for 3-4 days, after which time the film was developed in an automated developer (100 Plus, All-Pro Imaging Corp, U.S.A.).

Stripping the Membranes: Membranes were placed in a plastic box of a suitable size. Typically, 4 membranes were simultaneously stripped in the same box. One litre of 0.5% SDS was heated to boiling and poured on the membranes. After gentle agitation for approximately 20 minutes the solution was discarded, and a further litre of 0.5% SDS heated to boiling was added, and the process repeated. Excess solution was removed by blotting the membranes between clean Whatman 3MM filter paper. The membranes were then individually sealed in thin walled plastic bags and stored at 4°C until required. The completeness of stripping was determined by placing a representative membrane in a film cassette with a BioMax™ MS intensifying screen and film. The cassette was stored at -80°C for 3-4 days, and the film subsequently developed. A lack of radioactivity was noted indicating the membrane had been stripped of probe.

Allele Identification: DQA_2 alleles from the above flocks were identified and named according to the definitions of Escayg (Escayg, 1995, Escayg *et al*, 1996). A diagrammatic representation of the banding patterns of the 15 DQA_2-*Taq*I alleles is shown in Figure 2.2. Banding patterns that did not fit those of the "known" DQA_2 alleles defined by Escayg, (1995) were tentatively described as a new allele, with subsequent retyping of the sample to confirm the new banding pattern. The DQA_2 typing results for all of the animals investigated are shown in Appendix F.

Statistical Analysis: To determine whether DQA_2 alleles associated with either footrot resistance or susceptibility the data collected from the four flocks underwent three separate analyses using Systat™ computer software. Initially, the individual sire lines within each flock were analysed. For each allele present within each line, a relative risk of footrot was calculated (see Appendix G for formula). The significance of these relative risks was determined using Pearson's Chi-square test. When allele frequencies were < 5.0 Fishers exact test was used.

In the second analysis data from all of the sire lines from the four flocks was pooled, with the exceptions of the animals from Merino flock 1 1998-99 challenge, and the ewes

from Merino flock 2. The Chi-squared analysis of the relative risk of each allele was then repeated for the pooled data.

Finally, multiple logistic regression analysis was performed on the pooled data from the four flocks. This was used to assess whether certain DQA_2 alleles or allele combinations could successfully predict the likelihood of footrot occurring or not.

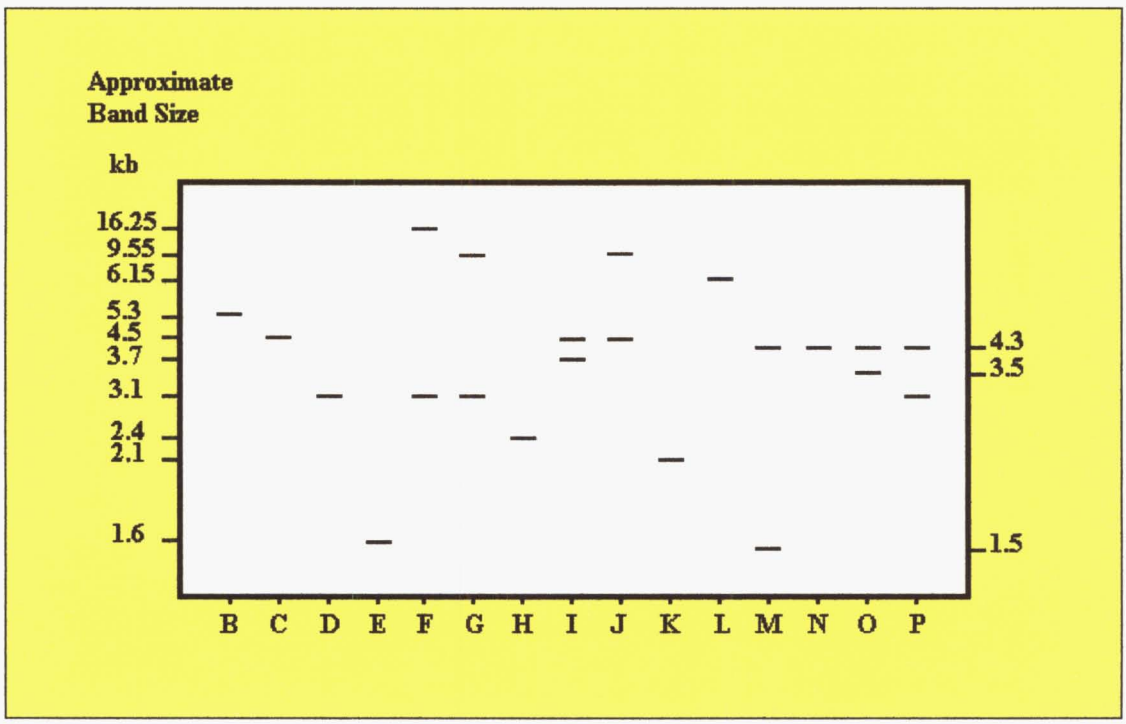


Figure 2.2: Diagrammatic representation of the banding patterns of the 15 DQA_2 *Taq*I alleles defined by Escayg (1995).

2.3 Results

Figure 2.3 depicts the results obtained for a typical Southern blot.

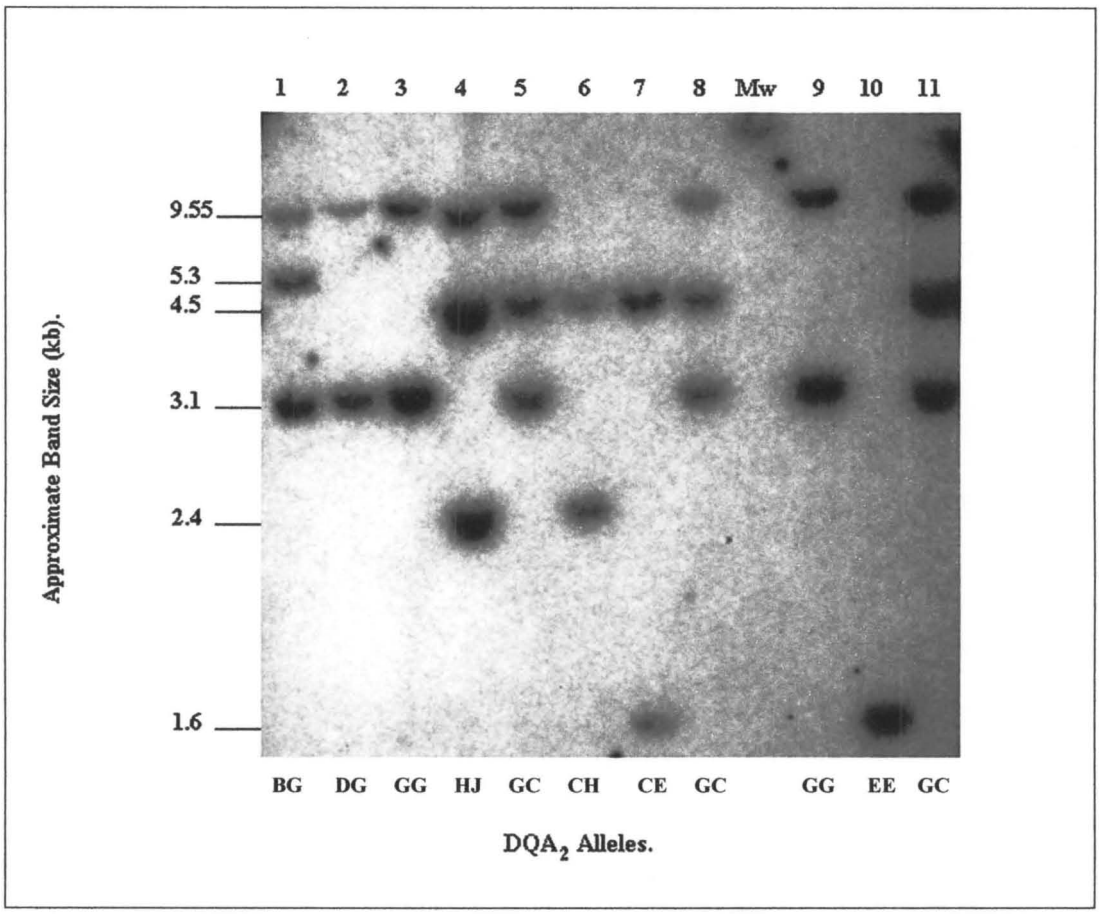


Figure 2.3: Results from a Southern blot. An ovine DQA_2 probe, which consisted of an exon 2 DQA_2 gene insert from clone pC17-2PH was labelled with α-^{32}P-dCTP and hybridised to genomic DNA samples that had been digested with the restriction enzyme *Taq*I, revealing the above banding patterns. A λ *Hind* III molecular weight marker that was electrophoresed with the genomic samples allowed the approximate size of each band to be determined. DQA_2 alleles were assigned according to the definition of Escayg, (1995). (The above samples come from Corriedale Flock 3) Lanes 2, 3 and 9 have the same banding pattern, which means that either G,G or D,G alleles were present. Whether an animal typed as a G,G homozygote or a D,G heterozygote was determined by comparing relative band intensity (see Appendix F for typing results).

Results for the Analysis of the Individual Sire Lines within Each Flock

In the initial analysis, each sire line within the four different flocks investigated was examined separately. The relative risk for each sire allele was calculated and Chi-squared analysis was performed to determine whether there was any association with footrot susceptibility or resistance (Table 2.2). A low relative risk value indicates low footrot susceptibility, while a high relative risk value indicates high footrot susceptibility. Progeny samples that had typed the same as the sire at the DQA_2 locus were not included in this analysis, as it could not be determined which allele had been inherited from the sire.

DQA_2 alleles inherited from the ewes that had a high frequency within a particular sire group were also examined to determine if there were any associations with footrot susceptibility or resistance. The prevalence of footrot (F) within each line was calculated. F is defined as the percentage of infected animals at the final inspection (Table 2.2).

Table 2.2: Summary of results for the analysis of individual sire lines

Flock	Sire Line	Sire Alleles	Number of Progeny	F (% Footrot)	Relative Risk[††]	p.value	N[‡ (Ψ)]
Awassi NZ Ltd.	1046/91	D,E	44	36 %	D = 0.56	0.15	20 (5)
					E = 1.84	0.15	24 (11)
	Ewe Alleles [†]	B			B = 0.89	1.00	9 (3)
		H			H = 0.66	0.37	15 (4)
	1039/91	H,K	40	48 %	H = 1.20	0.68	17 (7)
					K = 0.85	0.68	23 (8)
	Ewe Alleles [†]	D			D = 0	**0.033***	7 (0)
		L			L = 1.47	0.444	8 (4)
	1036/91	F,E	31	37.5 %	E = 0.81	0.605	18 (8)
					F = 1.20	0.605	13 (7)
	Ewe Alleles [†]	C			C = 0.40	**0.038***	12 (3)
		K			K = 1.86	0.106	12 (8)
	4071/94	L,C	16	37.5 %	C = 0.15	**0.035***	9 (1)
					L = 6.45	**0.035***	7 (5)
	Ewe Alleles [†]	K			K = 0	0.093	5 (0)
Corriedale Flock 3	3547/92	H,D	27	26 %			
	8661/95	H,D	21	24 %			
	7973/95	C,E	18	16.5 %			
	7472/94	G,C	16	0 %	N/A		
	8215/95	G,E	19	5 %	No significant associations were observed in any of the sire lines.		
	5179/93	J,E	10	10 %			
	3433/92	G,C	14	7 %			
	7271/94	G,D	21	19 %			

Merino Flock 1 (1997 Trial)	380	G,Q	29	10 %	G = 0.23	0.22	20 (1)
					Q = 4.4	0.22	9 (2)
	Ewe Alleles †	C			C = 0.5	1.00	14 (1)
	10	E,G	45	26.5 %	E = 3.36	**0.034***	17 (8)
					G = 0.30	**0.034***	28 (4)
	Ewe Alleles †	C			C = 1.95	0.187	19 (7)
		D			D = 0.42	0.419	8 (1)
	9	C,E	25	32 %	C = 0.47	0.389	14 (3)
					E = 2.14	0.389	11 (5)
	Ewe Alleles †	H			H = 1.72	0.570	4 (2)
		R			R = 0.76	1.00	4 (1)
Merino Flock 1 (1998-99 Trial)	8	C,G	24	33 %	C = 0.40	0.167	17 (4)
					G = 1.15	0.167	7 (4)
	Ewe Alleles †	D			D = 1	1.00	6 (2)
Group 1		L			L = 3	0.091	4 (3)
	101	C,Q	43	21 %	C = 5.17	0.065	26 (8)
					Q = 0.19	0.065	17 (1)
	Ewe Alleles †	H			H = 1.25	1.00	8 (2)
		L			L = 0.77	1.00	8 (1)
	4403	D,H	30	20 %	D = 1.11	1.00	14 (3)
					H = 0.90	1.00	16 (3)
	Ewe Alleles †	C			C = 1.79	0.657	16 (4)
		J			J = 7.14	**0.034***	2 (2)

Group 2	8	C,G	24	12.5 %	C = 0.21	0.194	17 (1)
					G = 4.8	0.194	7 (2)
	Ewe Alleles †	D			D = 0	0.546	6 (0)
		L			L = 10	0.061	4 (2)
	101	C,Q	43	16 %	C = ----	**0.031***	26 (7)
					Q = 0	**0.031***	17 (0)
	Ewe Alleles †	H			H = 1.78	0.579	8 (2)
		L			L = 0.7	1.00	8 (1)
	4403	D,H	30	10 %	D = 2.24	0.586	14 (2)
					H = 0.45	0.586	16 (1)
	Ewe Alleles †	C			C = 1.78	1.00	16 (2)
		J			J = 7.1	0.193	2 (1)
Merino Flock 2 Group 1-Ewes		I,L	27	22 %	I = 0.91	1.00	14 (5)
					L =1.09	1.00	13 (3)
Group 2-Rams		I,L	24	83 %	I = 1.08	1.00	14 (12)
					L = 0.93	1.00	10 (8)
	Ewe Alleles †	C			C = 0.42	**0.018***	5 (2)

† DQA_2 alleles inherited from the dams that occur frequently within the group.

‡ N = The number of animals within the sire line that had the allele.

(Ψ) = the number of animals that have the allele and footrot.

* Significant association, p. value ≤ 0.05 (Pearsons Chi-square test).

†† see Appendix G for relative risk calculation.

1) Awassi NZ Ltd

Progeny DNA samples from four sire lines were collected. All of the banding patterns observed in this flock corresponded to the alleles defined by Escayg, (1995), with no new alleles being identified. The DQA_2 alleles I, J, M, N, O and P were not observed in the Awassi flock, and the G allele was rare.

Four significant associations were observed in the Awassi lines (p value ≤ 0.05 Pearsons Chi-square test). In sire line 1039/91 the D allele significantly associated with low relative risk, hence low footrot susceptibility. Although, the D allele did not occur at a high frequency in this line, of the seven animals that inherited the D allele, none of the animals got footrot. The C allele in lines 1036/91 and 4071/94 associated significantly with low relative risk, while the L allele in line 4071/94 associated significantly with high relative risk, therefore high footrot susceptibility.

Overall from the relative risks obtained for each Awassi sire line, alleles D and C tend toward decreased footrot susceptibility, while alleles F, H and K tend toward low to intermediate footrot susceptibility, and the E and L alleles tend toward intermediate to high footrot susceptibility. The B allele also tends toward low footrot susceptibility, but the allele does not occur frequently within the lines analysed.

2) Corriedale Flock 3 (BCBG)

Progeny samples from eight sire lines were collected. No new alleles were identified, with all of the banding patterns observed corresponding to the alleles defined by Escayg, (1995). Alleles K, L, M, N, O and P were not observed in this flock, and the B allele was relatively rare.

No significant associations between DQA_2 alleles and footrot susceptibility or resistance were observed in any of the eight sire lines investigated. The level of footrot observed amongst the eight lines was extremely low with one line 7472/94 having no incidence of

footrot at all, suggesting the challenge was not sufficient or uniform amongst the flock to adequately delineate resistant and susceptible animals. In addition, once the progeny that had typed the same as the sire had been removed from the analysis, the progeny numbers within some of the lines were small.

3) Merino Flock 1

Progeny samples were collected from three sire lines that were challenged at Longslip in 1997, and an additional challenge was set up in 1998-99 at Lincoln University with progeny samples being collected from three further lines. The results of each trial were analysed separately. During the initial screening of the sires and subsequent progeny typing, two new banding patterns were observed within this flock. The banding patterns were tentatively defined as two new alleles Q and R (Figure 2.4).

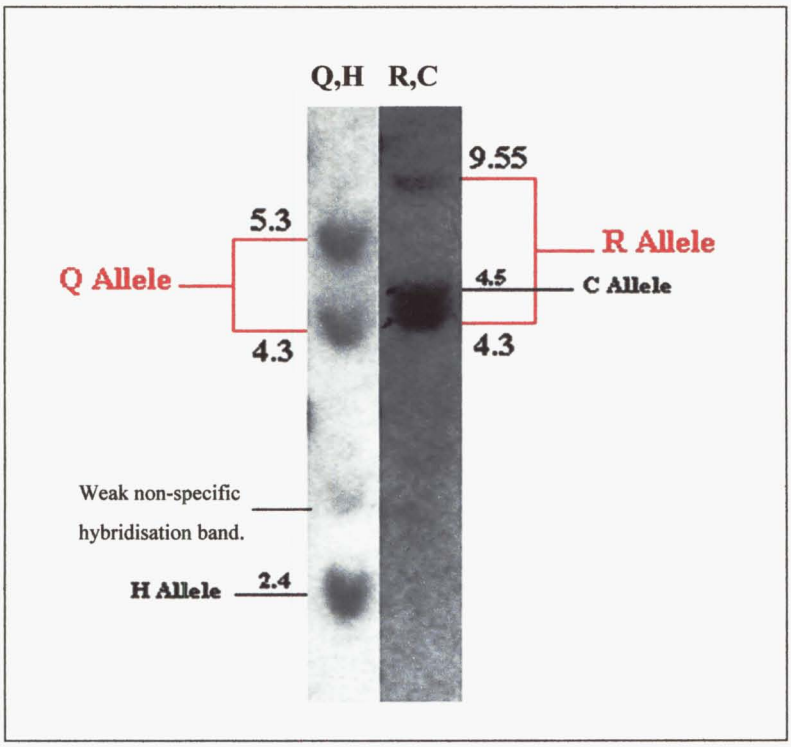


Figure 2.4: The banding pattern observed for two new DQA_2 alleles Q and R. The Q allele banding pattern consists of a 5.3 kb and a 4.3 kb fragment, while the R allele consists of a weak 9.55 kb fragment and a 4.3 kb fragment. The new Q and R alleles were only observed in Merinos.

1997 Trial

Two significant associations were observed in sire line 10. The G allele significantly associated with low relative risk, therefore low footrot susceptibility, while the E allele associated with high relative risk, therefore high footrot susceptibility. No significant associations were observed in sire lines 9 or 380.

Overall from the three Merino lines challenged in 1997, the G allele tends toward low footrot susceptibility, while the C allele tends toward low to intermediate footrot susceptibility, and the E allele tends toward intermediate to high footrot susceptibility. The D allele also tends toward low footrot susceptibility, but the allele does not occur frequently within the lines analysed.

1998-99 Trial

Alleles F, I, M, O and P were not observed in any of the animals analysed in the 1998-99 challenge. As stated earlier, during the 1998-99 challenge some animals within each line exhibited strange hoof morphology, similar to having had a severe footrot infection, but no active sign of footrot was found. In such cases where these animals occurred, statistical analysis was performed twice, one where these animals were grouped with the susceptible animals (Group 1), and one where they were grouped with the resistant animals (Group 2).

Three significant associations were observed within the three Merino lines analysed. In sire line 101 two significant associations were observed in Group 2, where the Q allele significantly associated with low relative risk. The relative risk value for the C allele could not be calculated. However, of the seven animals that had footrot in this line all inherited the C allele. The C allele therefore seemed to significantly associate with increased footrot susceptibility in this line. This is contrary to that observed in sire line 8, where the C allele has a low relative risk, although not to a significant level. The L allele also has low relative risk in line 101, but high risk in line 8.

In line 4403 the J allele significantly associates with high relative risk in Group 1, but this association becomes non-significant in Group 2. This result however, needs to be interpreted with caution as the J allele occurs at a very low frequency.

4) Merino Flock 2

Progeny were collected from one sire line, and divided into two groups, ewes (Group 1) and rams (Group 2), with 27 and 24 animals in each group respectively. As stated earlier, a sample could not be collected from the sire as it had died, and the DQA_2 haplotype was inferred from the progeny results, with the most likely being I, L.

The rams were formally challenged with footrot, while the delineation of susceptible and resistant animals in the ewes was achieved through normal field exposure. As a result of the splitting of this line into two groups, the ewes and rams were analysed separately. A new banding pattern was observed for two animals, which corresponded to the R allele that had been previously identified in the Longslip Merinos.

One significant association was observed in Group 2, where the C allele significantly associated with low relative risk. No significant associations were observed within Group 1, and none of the alleles inherited from the ewes occurred at a high enough frequency to determine any useful associations. Allele I had lower relative risk than the L allele within Group 1, while in Group 2 the L allele had lower relative risk than the I allele. These results were not significant.

Results for the Pooled Data Analysis

Data from Awassi NZ Ltd, Corriedale flock 3, the animals challenged in 1997 from Merino flock 1, and the rams from Merino flock 2, were pooled together to determine whether there was any association between DQA_2 alleles and footrot resistance or susceptibility. Progeny that had been removed from the analysis of the individual sire lines because they typed the same as the sire were included in this analysis, giving a total

of five hundred and forty animals. Table 2.3 shows the results obtained for each of the DQA_2 alleles.

Table 2.3: Results of the pooled data analysis

Allele	Relative Risk	p. value	Allele Frequency ‡‡ (N ‡)
B	1.68	**0.013** *	4.8 % (52)
C	0.64	**0.006** **	20 % (196)
D	0.74	0.097	13.7 % (140)
E	1.19	0.270	13.6 % (140)
F	1.31	0.242	3.0 % (31)
G	0.55	**0.001** **	16.8 % (177)
H	0.96	0.813	10.6 % (111)
I	2.96	**0.000** **	1.9 % (19)
J	0.57	0.131	3.4 % (37)
K	1.24	0.324	4.9 % (47)
L	1.5	**0.042** *	4.4 % (44)
N	1.9	0.555	0.3 % (3)

* Significant association, p. value ≤ 0.05 (Pearsons Chi-square test)

** Highly significant association, p. value ≤ 0.01 (Pearsons Chi-square test)

‡‡ The allele frequency does not add to 100 % because alleles Q and R were left out of the analysis. The combined allele frequency of the Q and R alleles is 2.6 %. Alleles Q and R were not included in the analysis because they had only been tentatively assigned as new alleles, and they were only observed within a small number of animals within the two Merino flocks.

‡ (N) = The number of animals that had the allele.

Alleles G and C associated with low relative risk, hence low footrot susceptibility to a highly significant level. Alleles I, B and L associated with high relative risk, hence high footrot susceptibility. The association observed between the I allele was highly significant, while the B and L alleles association was to a significant level. None of the other DQA_2 alleles were found to significantly associate with either a high or low relative risk.

Multiple logistic regression analysis was performed on the pooled data, to determine whether ovine DQA_2 alleles could predict the likelihood of footrot occurring or not (data not shown). This analysis failed to find DQA_2 alleles that could adequately predict the likelihood of footrot occurring.

2.4 Discussion

Investigation of the ovine MHC DQA_2 locus from Awassi, Corriedale and Merino breeds using an ovine derived probe covering the exon 2 region of the ovine DQA_2 gene detected similar banding patterns with the restriction enzyme *Taq*I, to the alleles defined by Escayg (1995). Of the 15 alleles that had been previously defined M, O and P were not observed in any of the flocks examined, but two new banding patterns were observed in the Merinos, which were tentatively defined as alleles Q and R.

Is Normal Field Exposure or a Natural Challenge Suitable to Differentiate between Resistant and Susceptible Animals?

While the mechanisms by which individuals are afforded greater protection from *D. nodosus* infection are yet to be defined, the ability to increase resistance through breeding implicates the involvement of underlying genetic components (Stewart *et al.*, 1986). Exposing individuals to the bacterium can delineate between resistant or susceptible animals, and allows the selection of animals that are more resistant for subsequent breeding. This has been the breeding strategy for both Corriedale flock 3 and Merino flock 1. Evidence of increased footrot resistance within Corriedale flock 3 has been reported by Skerman & Moorhouse (1987). However, the problem when relying on a natural challenge or normal field exposure to delineate between footrot resistance or susceptibility, is that it is difficult to estimate whether individuals are being exposed uniformly to the bacterium.

Within flocks a spectrum of resistance to footrot is often observed. Some animals are chronically infected throughout the challenge period; others have an intermediate resistance and vary in the time taken to become infected after exposure, whilst there are others that do not become infected at all (Egerton *et al.*, 1983). Variability within the skin integument may influence the degree to which tissue destruction by *D. nodosus*' proteolytic enzymes occurs, and therefore influences an individuals ability to resist infection (Emery *et al.*, 1984; Leslie, 1994). This innate form of resistance is the first

line of defence and can be undermined through scarification of the interdigital skin, which occurs during an artificial challenge (Emery *et al.*, 1984). Only once the disease becomes established does the strength and specificity of an individuals immune response influence the duration and magnitude of the infection (Leslie, 1994).

T-cell dependent antibody production involving ovine MHC class II mediated antigen recognition appears to be the main protection afforded by vaccination. Variability in K-agglutinin titre between individuals has been observed, suggesting variation in the ability of some individuals to recognise the antigenic determinants of specific pilus strains (Stewart, *et al.*, 1982; Egerton *et al.*, 1987). An inverse relationship between K-agglutinin titre and clinical severity of footrot, implicates the responsiveness to these pili epitopes as an important component in resisting *D. nodosus* infection (Stewart *et al.*, 1982; Egerton *et al.*, 1987; Leslie, 1994). It is important therefore when investigating associations between ovine MHC loci and footrot resistance that the challenge is uniform and severe, such that the difference between resistant and susceptible animals is not the result of differences in the interdigital skin barrier. In addition, the challenge period needs to be sufficiently long enough in duration (2-3 weeks) in order to promote a MHC-mediated immune response (Leslie, 1994; Escayg, 1995).

The prevalence of footrot within each of the individual sire lines from the four flocks investigated varied from 0 % to 83 %. The question that arises is were all the sire lines within the flocks uniformly challenged?

Corriedale flock 3 and Merino flock 1 had a low prevalence of footrot within each of the sire lines. Animals from Corriedale flock 3 had only encountered *D. nodosus* through normal field exposure, as environmental conditions did not allow a natural challenge to be performed. Although, this flock has shown to be more resistant to footrot (Skerman & Moorhouse, 1987), a higher prevalence of footrot within some of the sire lines was expected. This suggests that the challenge to which the animals were subject, was not uniform or severe enough to be able to adequately delineate resistant or susceptible animals. Although a few individuals did get footrot, it is possible that the difference

between those animals that got footrot and those animals that did not, was a result of other genetic factors (eg interdigital skin barrier), or differences in environmental conditions (eg nutrition), rather than variation at the DQA_2 locus. This was reflected in the results, as no associations between DQA_2 alleles and footrot resistance or susceptibility were observed. The low level of footrot in each line, and the small numbers within some of the lines made it difficult to determine if variation at the DQA_2 locus may have had any effect on an individuals footrot susceptibility or resistance. This flock therefore illustrates the requirement for a good challenge before association studies can be performed.

Merino flock 1 was naturally challenged during both the 1997 and 1998-99 trials. Control animals that had been included in both trials had severe footrot, suggesting that both the 1997 and 1998-99 challenge was suitable to be able to adequately delineate resistant and susceptible animals. The low prevalence of footrot within each line suggests that these animals are reasonably resistant to footrot and it is likely to be the outcome of selecting for increased resistance to footrot over a number of years. Some significant associations between DQA_2 alleles and footrot resistance and susceptibility were observed within Merino flock 1, suggesting that the DQA_2 locus may be involved in this increased resistance. Other genetic factors however, are likely to be involved.

The increased resistance of this flock becomes apparent when they are compared to Merino flock 2, which have not been selected for footrot resistance. In Merino flock 2 a natural challenge resulted in 83 % footrot being observed in the ram lambs. Whilst it is possible that Merino flock 2 could have encountered more virulent strains of the bacteria, accounting for the higher incidence of footrot amongst this flock, it seems unlikely to be able to account for such a large difference ($\approx$ 50 %). The resistance within Merino flock 1 therefore, is likely to be the outcome of selecting for increased footrot resistance over a number of years, and unlike Corriedale flock 3, the low prevalence of footrot within this flock is unlikely to be the result of a poor challenge.

Another aspect that needs to be taken into account is that a major variable when using natural challenges or normal field exposure to delineate between resistant or susceptible animals, is the different strains of *D. nodosus* that may be present on each farm. Other variables such as differences in environmental conditions and management practices within a farm or between farms need also to be considered.

These variables may be particularly important when pooling data from different lines and farms. It is difficult to determine whether each of the flocks investigated has been challenged with similar serogroups or strains of *D. nodosus*, particularly when some strains are thought to be more virulent than others (Egerton & Raadsma, 1989).

Such problems with the challenge procedure are difficult to overcome. Nevertheless recent typing of *D. nodosus* from the hooves of infected animals throughout New Zealand has shown that similar serogroups are present throughout the country, and more than one serogroup with multiple strains were detected on each farm analysed (Zhou & Hickford, 2000). It follows that whilst the animals were challenged in different locations, under different environmental conditions, they should have encountered similar serotypes of *D. nodosus*. It is important not to discount, however, the impact that different environmental conditions may have on disease status, or that some of the animals may have been exposed to more benign or virulent strains. This may help to explain why in some flocks the prevalence of footrot was very low (eg Corriedale flock 3), and the results, especially from the pooled data analysis need to be interpreted with some caution.

Analysis of the Individual Sire Lines within Each Flock

Within the individual sire lines of the flocks analysed, several DQA_2 alleles were found to significantly associate with footrot susceptibility or resistance. A relative risk was calculated, with a high relative risk (> 1), indicating higher footrot susceptibility, and a low relative risk (< 1), indicating lower footrot susceptibility. However, comparing across lines within the same flock, revealed in some instances that alleles in one line that

had a low relative risk had a high relative risk in another line. This was particularly true for the C and L alleles.

In Merino flock 1 the C allele had a low relative risk in sire lines 380, 9 and 8, but had high relative risks in sire lines 10, 101, and 4403. The L allele had a high relative risk in sire line 8 and a low relative risk in sire line 101. How can these seemingly contradictory results be explained?

The RFLP and Southern hybridisation typing of the DQA_2 gene, using the restriction enzyme *Taq*I may not necessarily detect all of the polymorphism at the DQA_2 locus. Previously, two different C alleles were detected at the DQA_2 locus when animals were typed using the restriction enzyme *Pst* I (Escayg, 1995). In addition, when animals that had been previously typed at the DQA_2 locus using *Taq*I and Southern hybridisation typing were retyped using PCR-SSCP, up to three variations of the C allele were observed (Ridgway H.J., personal communication).

It is possible that one of the C alleles may associate with footrot resistance whilst another may associate with footrot susceptibility. As a result of the inability to be able to detect variation within the C allele using *Taq*I and Southern hybridisation typing, it would be possible to get results similar to those observed. Whether a high or low relative risk was observed would depend on which variation of the C allele was most prevalent within a particular line, and would help to explain the seemingly contradictory results.

Further typing using PCR-SSCP of the DQA_2 locus has also revealed further variation with division of the F and J alleles into F1, F2, J1 and J2 (Ridgway H.J., personal communication).

Furthermore, other factors such as animal weight, strains of *D. nodosus*, and different environmental conditions during the challenge may also help to account for some of the contradictory results. The possible impact that other genes, different bacterial strains, and environmental conditions have on an individuals footrot resistance/susceptibility

could be determined by generating an experimental flock, and challenging with the same strain(s) of *D. nodosus*.

In addition, there may be more value in seeking associations between MHC haplotype rather than individual alleles (Escayg *et al.*, 1997). It may be that a linked gene that co-segregates with the DQA_2 locus is responsible for increased footrot resistance or susceptibility. A significant association ($P< 0.05$) between the DQA_1 C allele and footrot resistance has been observed (Leslie, 1994). By haplotyping animals at MHC loci (eg DRA, DQA_1, DQA_2, & DQB), a clearer indication of how MHC polymorphism impacts on an individual's immune response to footrot resistance or susceptibility could be gained.

Pooled Data Analysis

In the different flocks analysed, some DQA_2 alleles were more common than others. Data from Awassi NZ Ltd, Corriedale flock 3, the animals challenged in 1997 from Merino flock 1, and the rams from Merino flock 2 were pooled. This was done to provide enough data on each of the DQA_2 alleles that were defined by Escayg, (1995), in order to rank them in terms of their relative footrot resistance or susceptibility. Figure 2.5 shows the ranking, which has been based on the relative risk values obtained for each allele. Twelve of the 15 DQA_2 alleles that were defined by Escayg (Escayg, 1995; Escayg *et al.*, 1996), could be ranked.

Contrary to Escayg's results (Escayg, 1995; Escayg *et al.*, 1997), the E allele did not associate with the highest relative risk. The I allele associated with the highest footrot susceptibility, whilst the G allele associated with the lowest susceptibility. Similar to Escayg (Escayg, 1995; Escayg *et al.*, 1997), the H allele associated with a lower footrot susceptibility than the E allele, but it was not the most protective allele.

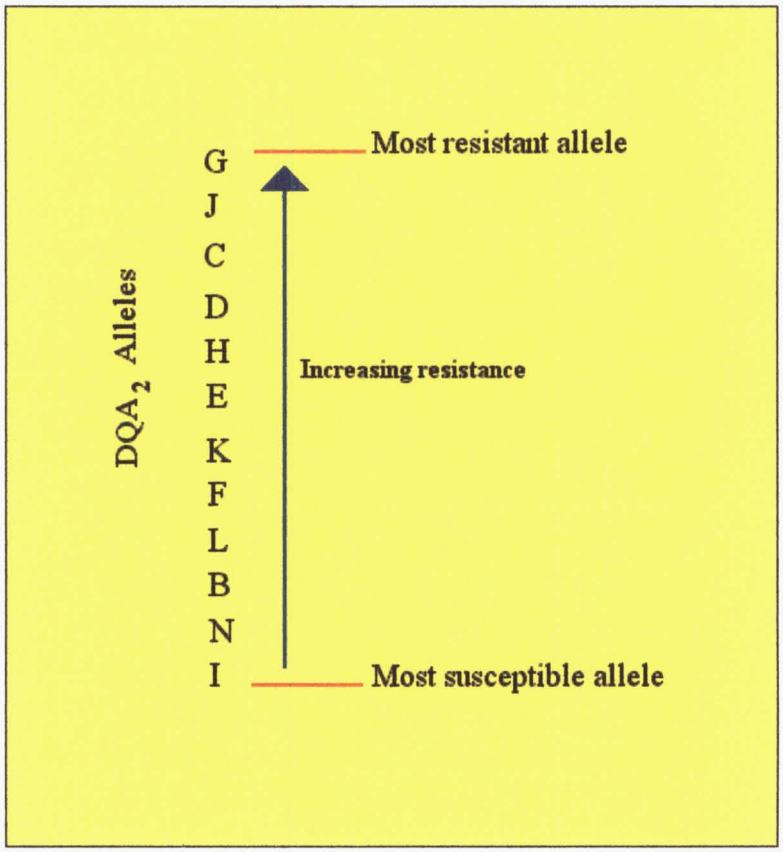


Figure 2.5: The ranking of the ovine DQA_2 alleles based on the relative risk values obtained for each allele (Table 2.3).

Pooling the data from the different flocks together in order to obtain this ranking meant that some assumptions were made. Firstly, it was assumed that the allele typed by RFLP and Southern hybridisation was the same across the different flocks and breeds analysed. As stated above this may not be correct as PCR-SSCP analysis has detected variation within the C, F and J alleles. Nevertheless, sequencing various DQA_2 alleles from different breeds has shown that the sequence of E, G and H alleles were conserved across various breeds (Chapter 3). In addition, although the data had been pooled, alleles B, H, I, J, K and N still had a low frequency. As a result, of the low frequency of these alleles, and the low incidence of footrot within some of the individual sire lines, the statistical analysis can be skewed. This may mean that some of the relative risk values may be either too high or too low, which may change the ranking of these alleles. The

only way to resolve this issue, would be to find other flocks in which these alleles occurred at a higher frequency, and challenge them with footrot.

It was assumed that all of the animals within the various lines and flocks were unrelated. Again this assumption may be incorrect. Inbreeding occurs frequently within flocks, especially within smaller flocks, where the ewes in particular, can be highly related to one another. This can cause a phenomenon termed the "founder effect".

Animals that are closely related to one another are genetically similar. As increased resistance or susceptibility to footrot is likely to be under multigenetic control, animals that are closely related may share several genes that contribute to increased footrot resistance or susceptibility. It is possible therefore, that some of the associations observed between the DQA_2 alleles and footrot resistance or susceptibility may not be a result of the alleles inherited at the DQA_2 locus. Instead it may be a result of other genes that are involved in footrot resistance or susceptibility that have been co-inherited by several progeny, because the ewes and sires have been closely related. The only way to resolve this issue, would be to generate an experimental flock, in which all of the animals are known to be unrelated. In this way the effect that DQA_2 alleles have on footrot susceptibility or resistance could be precisely determined.

Nevertheless, although increased footrot susceptibility or resistance may be afforded by other genes, the results of this study still suggests that the DQA_2 gene may either be directly involved, or that it may be closely linked to a gene that may provide greater resistance or susceptibility. Although the multiple logistic regression analysis failed to provide significant results, some of the DQA_2 alleles within the pooled data set were relatively rare. This combined with the low level of footrot within some flocks, means that there may be insufficient data for such an analysis to provide a useful result. The DQA_2 gene may have the ability to be a reasonable genetic marker to footrot resistance or susceptibility, but further investigation is required before its role as a useful genetic marker to footrot resistance or susceptibility can be fully assessed.

2.5 Conclusion

Several studies including the present one, have investigated the ovine MHC genetic region, in pursuit of a genetic marker associated with resistance or susceptibility to footrot (Outteridge *et al.*, 1989; Litchfield *et al.*, 1993; Escayg, 1995; Escayg *et al.*, 1997). All of these studies have found significant associations with resistance at different loci within the ovine MHC, implicating its involvement directly with class I and class II through epitope recognition and presentation, or indirectly through linkage to other genes which may be directly involved.

This current study revealed significant associations between ovine DQA_2 alleles and footrot susceptibility or resistance. Twelve of the 15 DQA_2 alleles defined by Escayg, (1995), could be ranked with the G allele affording the most protection, and the I allele affording the least protection.

As a result of possible inbreeding within the flocks analysed in this study, the associations between the DQA_2 locus and footrot resistance or susceptibility may not be solely a result of the DQA_2 alleles that have been inherited, but may be from other genetic components. Nevertheless, from the number of significant associations observed it seems that the DQA_2 gene may be a useful genetic marker to assess an animal's potential footrot resistance or susceptibility.

As a result of alleles B, I, J, K, L, and N being relatively rare within the flocks examined, and the low level of footrot within some of the flocks, the relative susceptibility or resistance of these alleles may be biased. Further disease association studies are required in other flocks that have a higher frequency of these alleles, before their true affect can be assessed. In addition, further analysis would reveal how robust the observed associations are in a wider genetic population.

CHAPTER THREE
SEQUENCE ANALYSIS OF THE OVINE-MHC DQA_2 GENE.

Introduction

Genetic selection for disease resistance provides an avenue for improving the health of farm animals. This may increase productivity and reduce the need for pharmaceutical intervention, thereby reducing costs, and possibly delaying the appearance of resistant pathogens.

The genes of the Major Histocompatibility Complex (MHC) may be markers of disease resistance. However, before MHC genes can become useful as markers of disease resistance they must be well characterised at the DNA level, with a comprehensive understanding of their genetic organisation, function and allelic polymorphism. In addition, an understanding of the genetic organisation of the MHC may lead to an improvement in typing methods such that a rapid, accurate and cost effective typing system can be developed.

The genetic organisation and variation within the MHC of humans and mice has been well characterised. Much less emphasis however, has been placed on the MHC of other species. It has been only since the mid-nineteen eighties that researches began to unravel the story behind the ovine-MHC class II region.

The class II ovine-MHC is situated on sheep chromosome 20, and shows extensive homology with corresponding loci in the HLA-D region of humans (chromosome 6). Southern blot hybridisation using human cDNA probes has revealed the presence of HLA-DQA, DQB, DRA, DRB, DNA, and DOB homologous loci (Scott *et al.*, 1987). The DQA locus has two DQA genes per haplotype, revealed from data derived from genomic clones (Scott *et al.*, 1991a), and cDNA clones (Fabb *et al.*, 1993). This has been further supported by a published genomic map of the ovine DQ sub-region, revealing two DQ loci, each consisting of a DQA and DQB gene arranged in a tail to tail

orientation (Wright & Ballingall, 1994). Unlike human HLA-DQ genes there is evidence that both ovine DQA genes (DQA_1, & DQA_2) are transcribed (Scott *et al.*, 1991a; Fabb *et al.*, 1993; Wright & Ballingall, 1994).

Consistent with other vertebrates, ovine-MHC DQ genes have been shown to be polymorphic, using RFLP and sequence analysis (Scott *et al.*, 1991a; Fabb *et al.*, 1993; Wright & Ballingall, 1994; Escayg *et al.*, 1996; Snibson *et al.*, 1998). Until 1996 no attempt was made to resolve the RFLP banding patterns observed for the DQ locus into alleles. Escayg *et al.*, (1996), defined 8 DQA_1 and 15 DQA_2 alleles from related and unrelated animals of three different breeds. Interestingly a DQA_1-null allele has also been observed (Scott *et al.*, 1991a; Fabb *et al.*, 1993; and Escayg *et al.*, 1996), raising the question as to the functional significance of the DQA_1 gene. Most recently Snibson *et al.*, (1998) have developed an ovine-DQA typing system using SSCP and sequence analysis, of exon 2, finding 10 DQA_2 and 6 DQA_1 alleles.

Until 1998 (Snibson *et al.*,), the lack of detailed sequence analysis of the defined DQA_2 alleles has meant that researchers investigating the role of the ovine-MHC in disease, have been limited to RFLP, and Southern hybridisation to type animals. This method is expensive, cumbersome and time consuming. With the availability of sequence information it is possible to design PCR primers, and develop less expensive, rapid and accurate typing techniques, such as PCR-RFLP, and PCR-SSCP.

It is clear however, that whilst Snibson *et al.*, (1998) have greatly increased the amount of sequence information available, in regards to the ovine DQA_2 gene, more effort is required to fully characterise the extent and nature of this polymorphism. They defined only ten alleles compared to fifteen that Escayg *et al.*, (1996), defined using RFLP and Southern hybridisation typing, giving rise to questions, such as, are these alleles the same, and how polymorphic is the DQA_2 locus? Both Escayg and Snibson have acknowledged that there may be more DQA_2 alleles, but because they occur at low frequency in the breeds and flocks investigated to date, they may not have been observed. It would therefore, be of benefit to sequence many DQA_2 alleles from

different flocks and breeds of sheep. This may allow a better determination of the extent of the polymorphism at the DQA_2 locus, the identification of any new alleles, and reveal whether the sequence of the same allele is conserved across different breeds.

Snibson *et al.* (1998) has sequence data pertaining to only exon 2 of the DQA_2 gene. Whilst, this is considered to be the most polymorphic exon, giving rise to most of the different alleles, it would be beneficial to assess other parts of the gene outside exon 2, perhaps identifying further new alleles. In addition, it would enable a more accurate prediction, of whether the alleles identified to date can produce a functional protein.

In the present study nucleotide sequence data is presented for part of exon 2, the intervening intron, and part of exon 3 of the DQA_2 gene. The animals sequenced were previously typed by RFLP and Southern hybridisation at the ovine-MHC DQA_2 locus, using the definitions of Escayg *et al.* (1996). Both DQA_2 homozygous and heterozygous animals were used, with samples obtained for various alleles from five different breeds.

3.1 Methods

Sheep: All animals utilised had previously been typed at the major histocompatibility DQA_2 locus, using Southern hybridisation and RFLP analysis, and could be divided into four groups (Table 3.1). Group 1, consisted of ten animals from four New Zealand breeds, (Corriedale, Romney, Texel and Awassi), and typed as homozygous at the DQA_2 locus. Sequence information from this group was obtained first (Coup, Ohlsson, & Slow, unpublished), acting as a mini database to which other sequences were compared. The three other groups consisted of half-sib families containing sire and progeny DNA from three different New Zealand breeds. All had been typed as heterozygous at the DQA_2 locus. Group 2, the Corriedale half-sib was from the Broomfield Corriedale Breeding Group (BCBG), Amberly, consisting of a sire and nine progeny. Group 3, a Merino half-sib was from Longslip, Omarama, and group 4, an Awassi half-sib was from Awassi NZ Ltd, Tikokino. Both the Awassi and Merino half-sibs consisted of a sire and four progeny samples. Samples were selected to cover as

many as possible of the 15 DQA_2 alleles that were defined by Escayg (Escayg, 1995; Escayg *et al.* (1996)), and to ascertain whether the allele defined, using the Southern hybridisation RFLP technique had identical DNA sequences in different breeds.

Genomic DNA: Genomic DNA was isolated using the high salt method described by Montgomery and Sise (1990). DNA was diluted for PCR amplification to a concentration of 50 ng/µL using sterile dH_2O.

PCR Primers: PCR primers were designed from a published DQA_2 sequence (Scott *et al.*, 1991a), to produce an 890bp amplimer, consisting of exon 2, exon 3, and the intervening intron of the DQA_2 gene. The primer sequences were, DQA_2 forward 5' CCA GTA CAC CCA GGA ATT TGA TGG 3', and DQA_2 reverse 5' CCA GTG CTC CAC TTT GCA GTC 3' (Life Technologies).

PCR Amplification: Each 25 µL reaction contained 100 ng ovine genomic DNA, 0.2 µM each primer, 100 nM dNTPs (Gibco/BRL) 1 x PCR buffer containing 1.5 mM $MgCl_2$, and 1.5 U *Taq* polymerase (Qiagen). Reactions consisted of 30 cycles of 94°C denaturation for 1 min, 60°C annealing for 1 min, and 72°C extension for 2 min, followed by a final extension step of 72°C for 7 min. All reactions were carried out on a Perkin Elmer 2400 thermocycler (Perkin Elmer Corporation), using thin walled 200 µL tubes. A negative control containing no DNA was included in every reaction.

After the first sequencing results were obtained, small base pair differences between clones of the same sequence were observed. These may have been a result of errors made by the enzymes during PCR amplification, or sequencing. To determine whether this was a result of PCR error, an enzyme mix consisting of *Taq/pyrococcus* species GB-D DNA mixture (Elongase™, Gibco BRL), was used. The Elongase™ enzyme mix has a "proof reading" function, which minimises errors during PCR amplification. However, some reactions using the Elongase™ enzyme mix proved difficult to optimise, and in this situation only *Taq* polymerase was used (Table 3.1).

PCR Amplification Using Elongase™: Each 50 µL reaction contained 100 ng of genomic DNA, 10 mM dNTP mix (Life Technologies), 0.2 µM each primer, 1 x Elongase™ PCR buffer containing 1.5 mM $MgCl_2$, and 2 µL Elongase™ enzyme mix. Reactions consisted of a pre-amplification denaturation step of 94°C for 30s, followed by 35 cycles of 94°C denaturation for 30s, 60°C annealing for 30s, and 68°C extension for 1min. All reactions were performed in a Perkin Elmer 2400 thermocycler using thin-walled 200 µL tubes.

Table 3.1: DQA_2 Southern typing results, and flock information for animals used in this study

Group	Tag Number	Breed	DQA_2 Alleles	PCR enzyme	Vector
1	90/257	Romney	F,F (Escayg)	Taq polymerase	pCR 2.1
	89/858	Corriedale	C,C (Escayg)	Taq polymerase	pCR 2.1
	86/878	Corriedale	G,G (Escayg)	Taq polymerase	pCR 2.1
	IMF001	Texel	D,D (Escayg)	Taq polymerase	pCR 2.1
	91/9344	Corriedale	C,C (Escayg)	Taq polymerase	pCR 2.1
	91/9551	Corriedale	E,E (Escayg)	Taq polymerase	pCR 2.1
	1657	Mixed	F,F (Escayg)	Taq polymerase	pCR 2.1
	883	Mixed	B,B (Escayg)	Taq polymerase	pCR 2.1
	88/771	Corriedale	G,G (Escayg)	Taq polymerase	pCR 2.1
	96/211	Awassi	K,K (Slow)	Taq polymerase	pCR 2.1
2	Sire 20/88	Corriedale	E,H (Escayg)	Taq polymerase	pCR 2.1
	91/9019	Corriedale	G,H (Escayg)	Taq polymerase	pCR 2.1
	91/9135	Corriedale	E,H (Escayg)	Taq polymerase	pCR 2.1

	91/9307	Corriedale	C,H (Escayg)	Taq polymerase	pCR 2.1
	91/9489	Corriedale	C,H (Escayg)	Taq polymerase and Elongase™	pCR 2.1
	91/9562	Corriedale	G,H (Escayg)	Taq polymerase	pCR 2.1
	91/9228	Corriedale	C,E (Escayg)	Taq polymerase	pCR 2.1
	91/9289	Corriedale	C,H (Escayg)	Taq polymerase	pCR 2.1
	91/9326	Corriedale	E,G (Escayg)	Taq polymerase	pCR 2.1
	91/9566	Corriedale	E,J (Escayg)	Taq polymerase	pCR 2.1
3	Sire 10	Merino	E,G (Slow)	Elongase™	pGem-T
	4572	Merino	E,Q (Slow)	Elongase™	pGem-T
	4618	Merino	E,L (Slow)	Taq polymerase	pGem-T
	4620	Merino	B,G (Slow)	Elongase™	pGem-T
	4621	Merino	F,G (Slow)	Elongase™	pGem-T
4	Sire 1039	Awassi	H,K (Slow)	Taq polymerase	pGem-T
	96/156	Awassi	B,K (Slow)	Taq polymerase	pGem-T
	96/100	Awassi	E,K (Slow)	Taq polymerase	pGem-T
	96/178	Awassi	L,H (Slow)	Taq polymerase	pGem-T
	96/168	Awassi	D,H (Slow)	Taq polymerase	pGem-T

Agarose Gel Electrophoresis: 1 µL of PCR product (10 µL with Elongase™ PCR), was run on a 1% agarose (Seakem LE agarose, FMC Bioproducts) minigels, containing 1 x TBE gel buffer (0.089 M Tris, 0.089 M Boric acid, 0.002 M Na_2EDTA, 200 ng/mL ethidium bromide), and run for approximately 1 hour at 10 volts/cm. 5 µl (50 ng/µL) of 100 bp marker (Gibco/ BRL) was run adjacent to the electrophoresed DNA enabling molecular weight determination, and an estimate of DNA concentration by comparing relative band intensities, when visualised by ultraviolet transillumination of the gel.

Cloning PCR Products: An Invitrogen TA Cloning™ kit (Invitrogen, San Diego, USA), or a pGEM®-T Easy Vector Systems kit (Promega Corporation, WI, U.S.A.) was used to clone the PCR products, according to the manufacturer's instructions. For both kits 1-3 µL of PCR product was ligated into the vector and transformed into INVαF' cells. 50-200 µL of transformed cells were spread on LB plates containing either 50 µg/mL of kanamycin (Invitrogen vector) (Sigma, St Louis, MO, USA) or 50 µg/mL ampicillin (Promega vector) (Sigma), and 40 µl of 40 mg/mL X-gal (Sigma). After overnight incubation at 37°C, 5-10 colonies for each sample were selected for plasmid isolation and restriction analysis. Colonies were grown overnight (225 rpm, 37°C), in 5 mL Terrific broth, with either kanamycin (50 µg/mL), or ampicillin (50 µg/mL), depending on the vector being used. Plasmid DNA was extracted using the Quantum Prep® Plasmid Miniprep kit (Bio Rad Laboratories), following the manufacturer's instructions. Plasmid DNA was eluted in 100 µL of sterile dH_2O, and was rotary evaporated to a final volume of approximately 50 µL. The Invitrogen and Promega vector's insertion sites were flanked on each side by an *Eco*RI recognition sequence, facilitating the excision of the insert. A 5 µL aliquot of plasmid DNA was digested with 3-4 units *Eco*RI (Boehringer Mannheim), and visualised on 1% agarose minigels, with a 100 bp molecular weight ladder (Gibco BRL), enabling the selection of colonies containing the correct sized inserts.

DNA Concentration Measurement and Sequencing: DNA concentration and purity was ascertained by the measurement of A_{260} and A_{280} in a spectrophotometer. Samples were prepared for sequencing by adjusting the concentration to 200 ng/μL using sterile dH_2O. Sequencing was contracted to the Auckland University School of Biological Sciences, and performed on ABI 377 sequencer using Sanger dideoxy chain-termination methodology. Analysis of sequence data was performed using DNAMAN© for Windows ™, version 4a (Lynnon BioSoft, Quebec, Canada).

3.2 Results

One hundred and sixty-three clones containing an 890 bp insert were selected and sequenced from the 30 animals that were investigated. Numerous clones from each sample were sequenced to minimise the likelihood of PCR and sequencing errors.

Analysis of the resulting sequences revealed that after approximately 600 base pairs, nucleotide resolution had deteriorated, making the sequences unreadable from this point onwards. To be able to obtain the full-length sequence of the 890 bp insert, all of the clones were sequenced in both the forward and reverse directions. One hundred and fifty-four forward sequences and 145 reverse sequences were obtained. These sequences were aligned giving 130 full-length sequences of the 890 bp insert.

Alignment of all 130 sequences (results not shown) enabled the identification of the clones that contained the same sequence, and because all of the animals had been previously typed using Southern hybridisation the allele could also be identified (Table 3.2). In addition, this alignment revealed that the sequence of alleles typed by RFLP and Southern hybridisation was conserved across the different breeds, with the exception of the F allele, where differences were observed between the mixed-breed F (Tag 1657), denoted as F1, and the Romney F (Tag 90/257), denoted as F2. However, when sequences that were considered to be from the same allele were aligned small nucleotide differences were sometimes observed at the same position amongst the sequences. In such cases, the most frequent nucleotide at these particular positions was considered to be correct.

DNA sequences for alleles B, C, D, E, F, G, H, J, K, and Q as defined by Escayg *et al.*, (1996), were obtained for part of exon 2, intron 2, and part of exon 3. Sequence analysis revealed that alleles C and G could be sub-divided, with the identification of three C sequences, denoted C1, C2 and C3, and two G sequences, denoted G1 and G2. This gave a total of fourteen sequences.

Table 3.2: Tabulated sequencing results

Tag Number	Breed	Southern Hybridisation Type	Number of Clones Sequenced	Number of Sequences (F+R)	Allele Defined
90/257	Romney	F,F	2	2 (F), 3 (R)	F2
89/858	Corriedale	C,C	2	3 (F), 5 (R)	C2
86/878	Corriedale	G,G	2	2 (F), 2 (R)	G2
IMF001	Texel	D,D	2	1 (F), 1 (R)	D
91/9344	Corriedale	C,C	1	2 (F), 2 (R)	C2
91/9551	Corriedale	E,E	1	2 (F), 1 (R)	E
1657	Mixed	F,F	1	2 (F), 2 (R)	F1
883	Mixed	B,B	2	4 (F), 3 (R)	B
88/771	Corriedale	G,G	1	1 (F), 1 (R)	G1
96/211	Awassi	K,K	2	2 (F), 1(R)	K
Sire 20/88	Corriedale	E,H	11	7 (F), 7 (R)	E, H
91/9019	Corriedale	G,H	8	7 (F), 7 (R)	H, G1, G2 *
91/9135	Corriedale	E,H	9	8 (F), 8 (R)	E, H
91/9307	Corriedale	C,H	8	7 (F), 7 (R)	H
91/9489	Corriedale	C,H	11	11 (F), 10 (R)	H
91/9562	Corriedale	G,H	7	7 (F), 7 (R)	G1, G2

91/9228	Corriedale	C,E	7	7 (F), 4 (R)	E, C1
91/9289	Corriedale	C,H	8	8 (F), 8 (R)	C2, C3, H *
91/9326	Corriedale	E,G	8	8 (F), 7 (R)	E, G2
91/9566	Corriedale	E,J	7	7 (F), 6 (R)	E, J
Sire 10	Merino	E,G	4	4 (F), 3 (R)	E, G2
4572	Merino	E,Q	7	5 (F), 5 (R)	E,Q
4618	Merino	E,L	7	7 (F), 7 (R)	E
4620	Merino	B,G	7	5 (F), 5 (R)	G2
4621	Merino	F,G	9	8 (F), 8 (R)	G2
Sire 1039	Awassi	H,K	8	7 (F), 8 (R)	H, K
96/156	Awassi	B,K	4	4 (F), 4 (R)	K
96/100	Awassi	E,K	7	7 (F), 6 (R)	E
96/178	Awassi	L,H	5	4 (F), 3 (R)	H
96/168	Awassi	D,H	5	5 (F), 4 (R)	H

* Animals that appear to have three DQA-like alleles.

• Note: The G and C allele variations; G1, C1, and C3 were only observed in Corriedale animals.

G Allele Sub-division

Two DNA sequences were observed for allele G. These were designated as G1 and G2. They were 91 % homologous to each other. Both G sequences were obtained from a number of different animals from the Corriedale breed. Closer inspection of the two G sequences revealed that the allele denoted as G1 appeared to lack a splice site at the end of exon 2 (Figure 3.1).

C Allele Sub-division

Three DNA sequences were obtained for animals that typed as C by Southern hybridisation. Alignment of the three C sequences revealed 98 % homology between them. However, the C1 and C2 alleles appear to lack a splice site at the end of exon 2 (Figure 3.1).

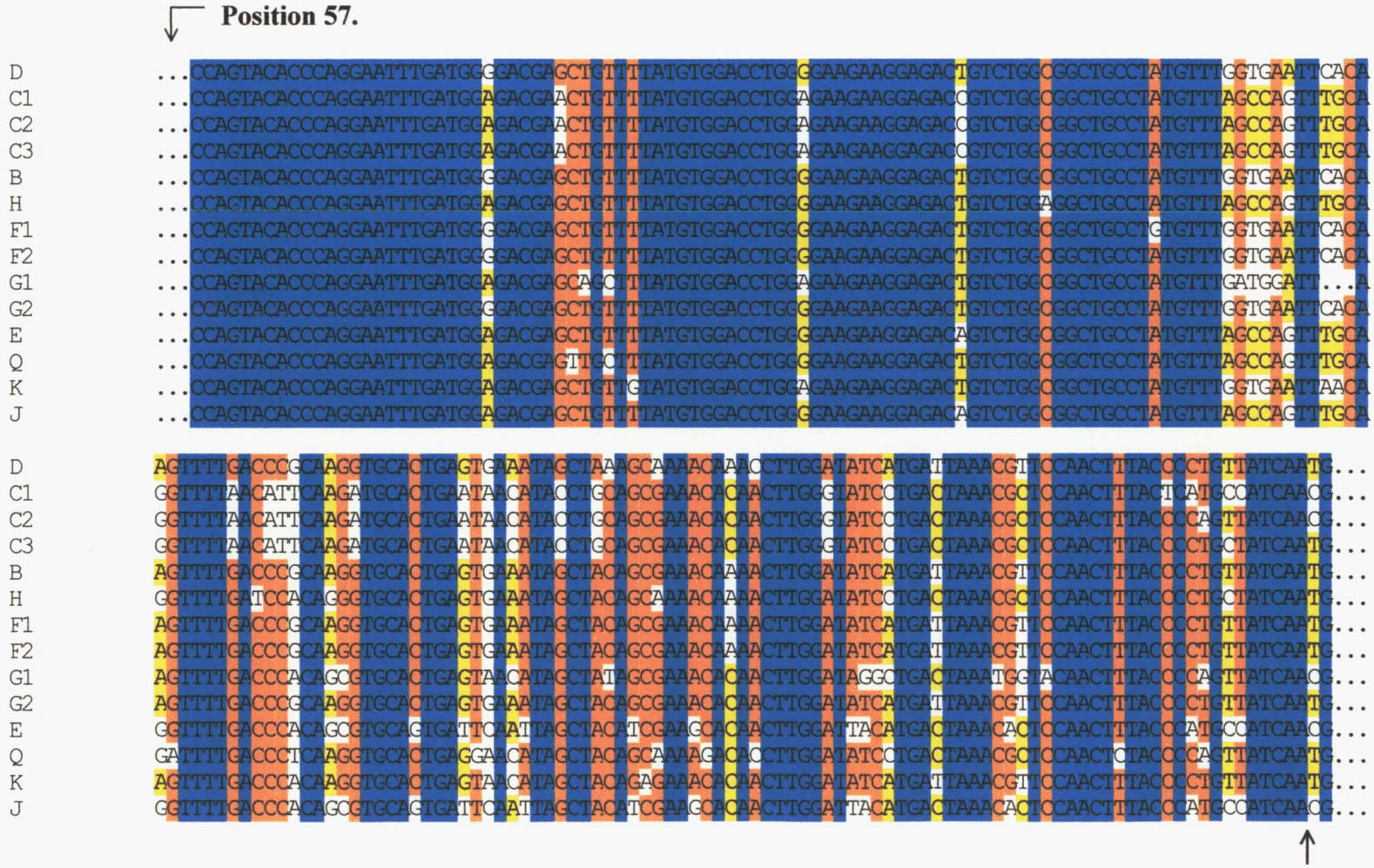
Other allelic sequences that also appear to lack a splice site at the end of exon 2, were the E and J alleles.

Interestingly, with two animals (9019/91-H, G1, G2 and 9289/91-C2, C3, H) three DQA-like sequences were identified. This suggested that these animals either had a duplicated DQA_2-like locus, or some PCR amplification error had occurred.

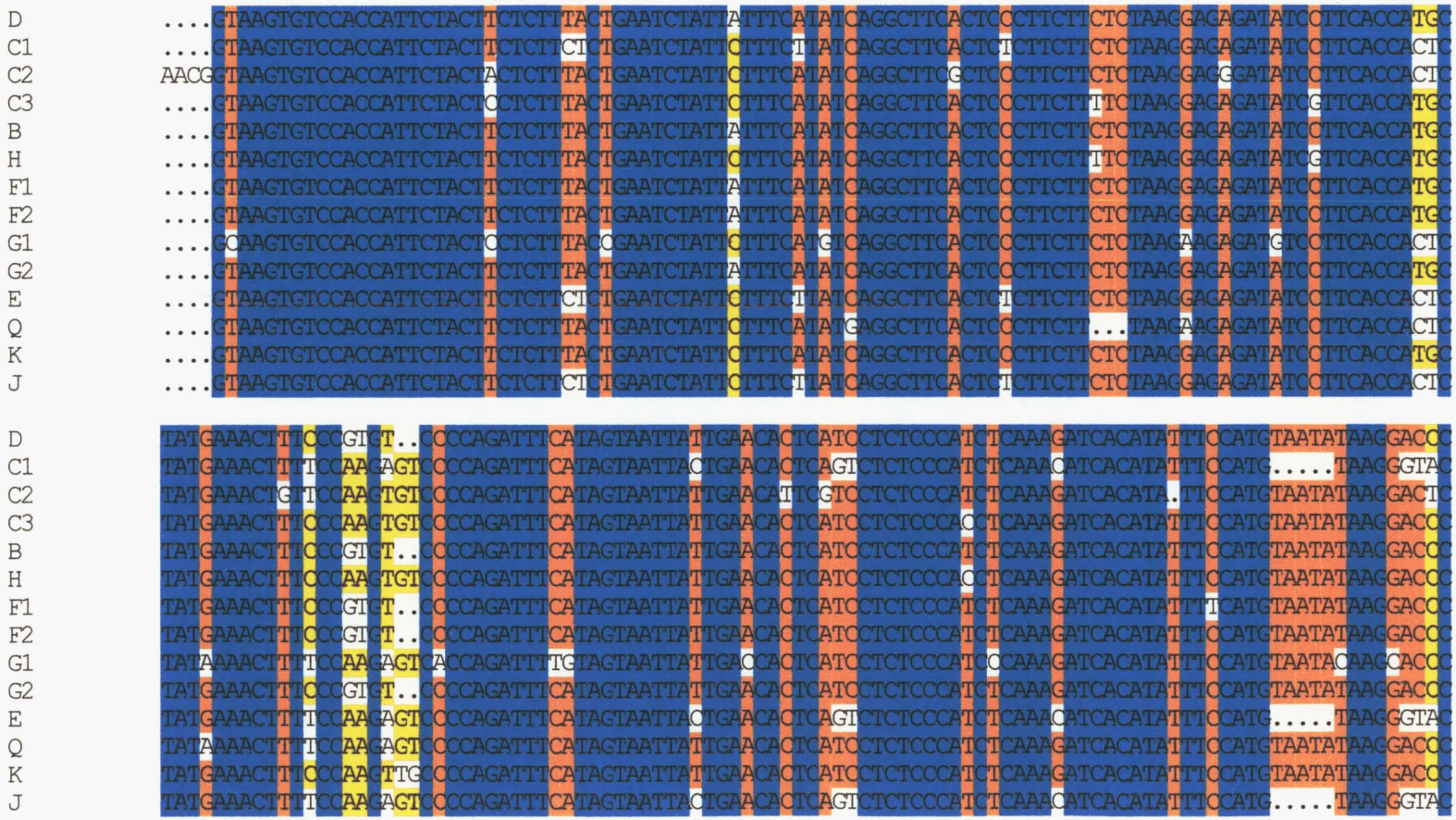
Exon 2 Sequence Alignment

Alignment of the fourteen allelic sequences obtained (Figure 3.1), reveals that like the class II MHC genes of other species, most polymorphism occurs within exon 2. The exon 2 alignment revealed two main clusters of alleles (Figure 3.2). The first cluster consists of alleles B, D, F1, F2 and G2, and shows high homology (99-100 %), while the second cluster consists of alleles E and J, which are 100 % homologous. These alleles are 85 % homologous to the B, D, F and G2 cluster. Alleles C, K, and Q show varying levels of homology to each other, and to the clusters.

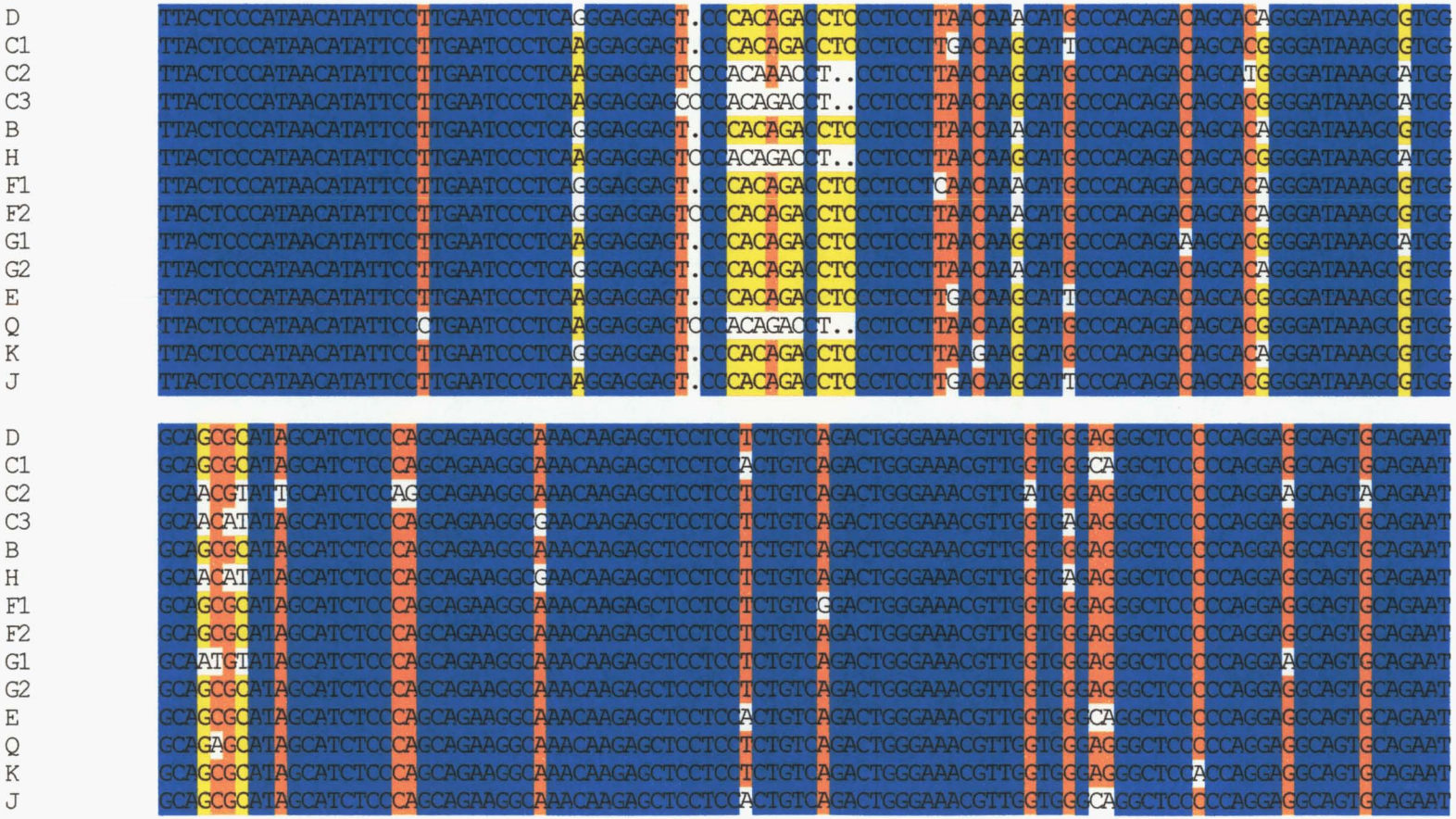
Exon 2

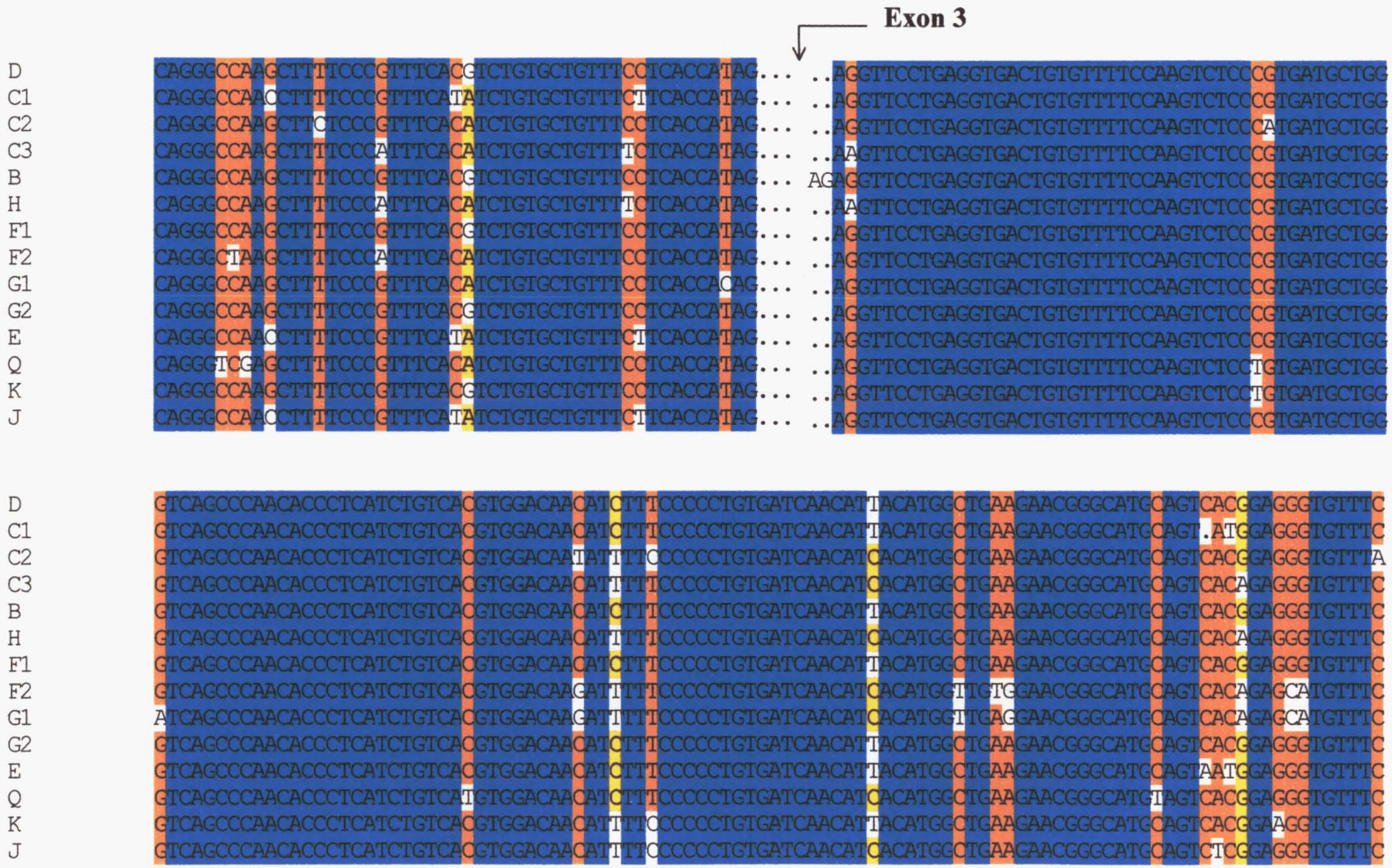

Position 57.
D
C1
C2
C3
B
H
F1
F2
G1
G2
E
Q
K
J
Alleles with CG sequence at the end of exon 2 appear to lack normally conserved splice site.

Intron



Intron contd.




Exon 3
D
C1
C2
C3
B
H
F1
F2
G1
G2
E
Q
K
J

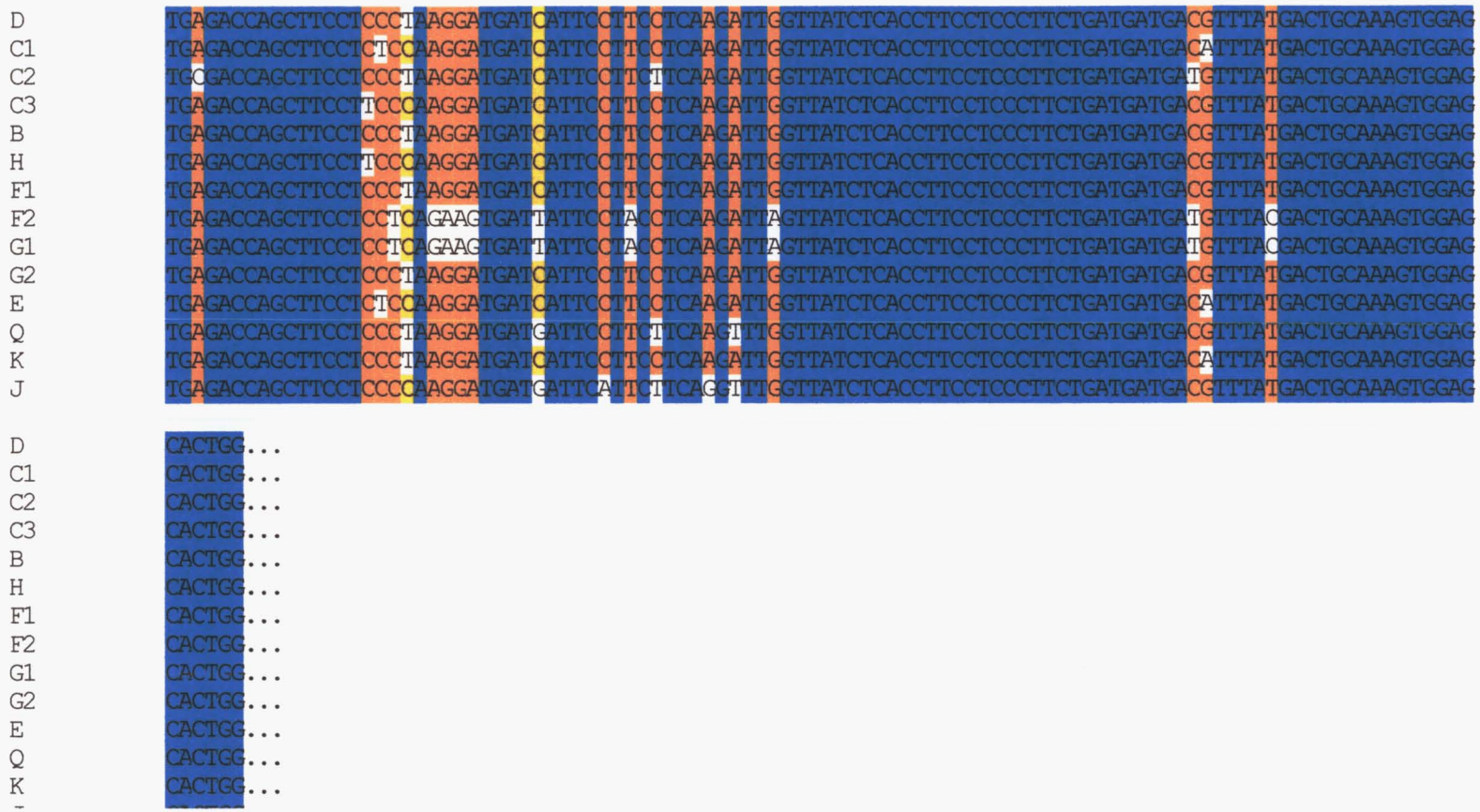


Figure 3.1: Sequence alignment of the fourteen alleles identified. Areas of 100% sequence homology are coloured blue, areas showing >70% sequence homology are coloured orange, and areas with >50% sequence homology are coloured yellow. Areas where <50% sequence homology occurs are uncoloured. Sequence G1 has a three base pair deletion in exon 2, while sequence B has a two base pair insertion in exon 3. The intron varies in length depending on the allele. Sequence data covers approximately 890 base pairs of exon 2, intron 2 and exon 3. Exon and intron boundaries were determined from the published sequence of the ovine DQA_2 clone C17-2 (Scott *et al*, 1991(a).

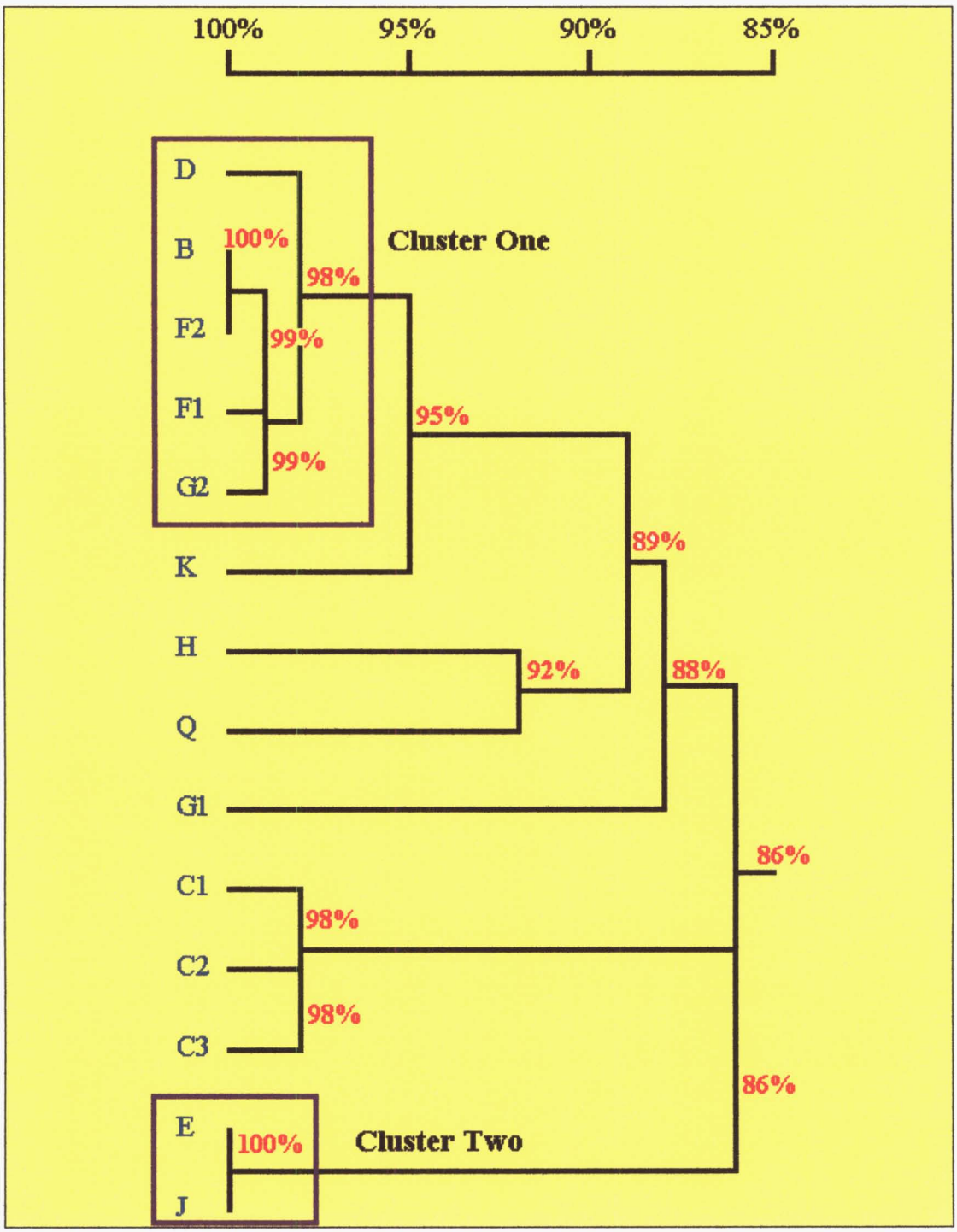


Figure 3.2: Homology tree constructed from exon 2 of the fourteen DQA_2 alleles (DNAMAN©). Two clusters of alleles have been identified, with alleles D, B, F2, F1, and G2 forming the first cluster, while alleles E and J form the second cluster. Other DQA_2 alleles have varying levels of homology to either cluster, but do not appear to belong to either.

3.3 Discussion

Are the Sequences from the Ovine DQA_2 Gene?

Ovine DQA and DRA genes share close homology at both the functional and nucleotide sequence level (Fabb *et al.*, 1993). As a result it was possible that the PCR reaction used in this work may have amplified DQA_1, DQA_2, DRA and potentially other A loci genes, giving rise to the question as to which gene did the sequences come from?

A prerequisite for the specific amplification of a particular gene using PCR is to design primers that will only amplify the gene of interest. The PCR primers employed in this study had been designed from a published DQA_2 sequence (Scott *et al.*, 1991a), and were expected to a produce a DQA_2 amplimer of approximately 890 bp. The primers were not thought to be able to amplify the ovine DRA gene, because they did not share enough sequence homology to enable the primers to anneal when a relatively high annealing temperature of 60°C was used during amplification. It was still possible however, that the DQA_1 gene could be amplified. Analysis of a published DQA_1 sequence (Scott *et al.*, 1991a) revealed that whilst close homology existed between the primers and the DQA_1 gene, such that primer binding and amplification could occur, the expected amplimer would be approximately 920 bp in length.

As a result of the difference in the expected amplimer size between the two ovine DQA genes it was possible to ascertain which gene was being amplified. Using agarose gel electrophoresis with a standard molecular weight marker, the size of the amplimer could be estimated prior to ligation. The visualisation of a PCR product that was slightly smaller than 900bp suggested that the most likely gene that was being amplified came from the DQA_2 locus.

In addition, further support that the sequences are DQA_2 encoded was provided when the sequences were aligned with other ovine DQA_2, DQA_1, and DRA sequences that were available from GenBank (version 102-106). Figure 3.3 shows that the sequences shared closest homology to other ovine DQA_2 sequences rather than DQA_1 or DRA sequences.

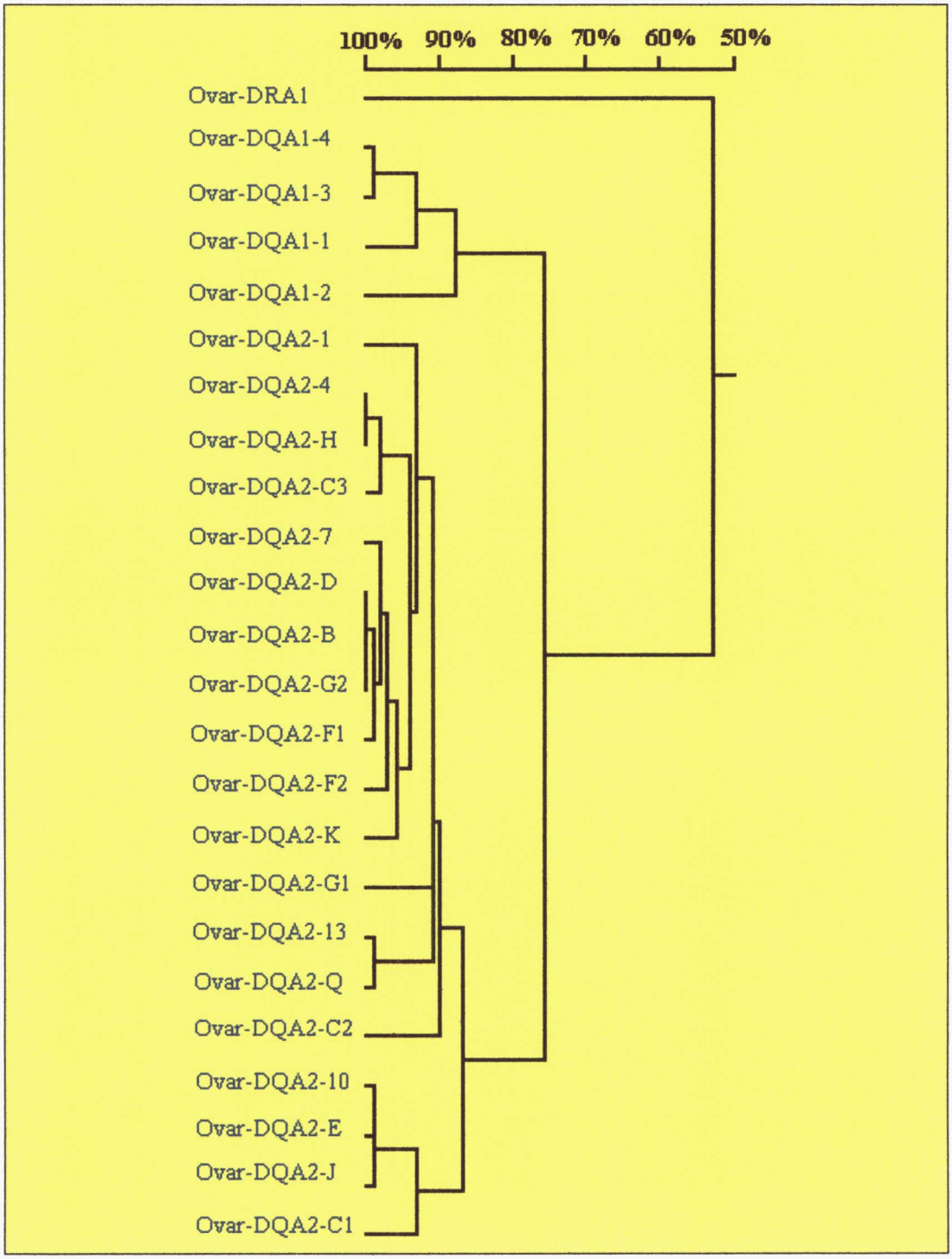


Figure 3.3: Homology tree constructed from ovine DRA1, DQA_1, and DQA_2 sequences. The sequences for the alleles identified share closest homology to other DQA_2 sequences retrieved from GenBank, indicating they are from the ovine DQA_2 gene rather than DQA_1 or DRA (see Appendix H for GenBank accession numbers).

One hundred and thirty full-length sequences were obtained from the clones, which came from various animals from a number of different breeds. Alignment of all the sequences obtained revealed fourteen different sequences that shared closest homology to other sequences from the ovine DQA_2 locus. As each animal had been previously typed at the DQA_2 locus using Southern hybridisation, it was possible to define which sequence corresponded to specific alleles. Sequences were obtained for alleles B, C, D, E, F, G, H, J, K, and Q, with subsequent sub-division of the C, F and G alleles into C1, C2, C3, F1, F2, G1, and G2.

Breed-specific Alleles

Alignment of all the sequences revealed that the E, G2 and H alleles typed by RFLP and Southern hybridisation was conserved across the different breeds. Sequences of the B, C, D, G1, J, K and Q alleles were only obtained from one breed, a result of some alleles being quite rare in particular flocks. It is impossible therefore to determine from the available data whether the sequences of these alleles are conserved across different breeds, and further sequencing is required to resolve this issue.

Sequences obtained from the F allele did however, show breed specific differences, with two F allele sequences being identified. F1 came from a mixed breed, which was a Coopworth x Perrindale cross, while F2 came from a Romney. The two F sequences shared closed homology (96 %). The small differences observed between the two F sequences may be either a result of PCR/sequencing errors, or a true reflection of the possible genetic differences between the two breeds, that RFLP and Southern hybridisation typing does not differentiate. The sub-division of the F allele however, is at best tentative, as very few F sequences were obtained. Further sequencing is required to be able to fully establish whether this is a breed specific difference or a result of PCR/sequencing errors.

C Allele Sub-division

Previously Escayg *et al.,* (1996), identified variations within the C allele using RFLP and Southern hybridisation typing. The restriction enzyme *Pst*1, identified at least

two C alleles, and it was therefore, not surprising that the sequence data obtained revealed three C allele sequences. Interestingly, all three C sequences were identical until the 179th base position in exon 2 (Figure 3.4). The C1 allele from this point then shares virtually 100% homology with the E sequence, while the C3 allele from this point shares virtually 100% homology with the H sequence. The C1 allele sequence was only observed once within animal 9228/91, which had been Southern typed as C, E, and similarly the C3 allele sequence was only observed once in animal 9289/91, which had typed as C, H. This suggests that the combined C1 and C3 allele sequences identified could be an artefact. One possibility is that it is a result of *in vitro* PCR recombination, and therefore they may not be true DQA_2 sequences.

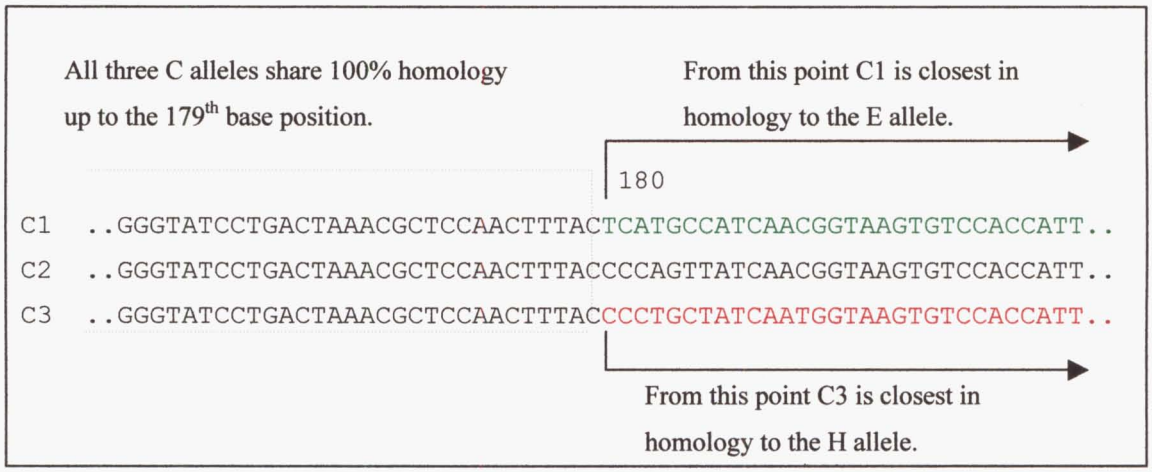


Figure 3.4: Alignment of C allele sequences. This illustrates the possibility that C1 and C3 allelic sequences are artifacts of *in vitro* PCR recombination.

In vitro PCR recombination occurs when DNA polymerases incompletely extend a segment of DNA, and the partially amplified sequence acts as a primer during subsequent amplification cycles (Bradley & Hillis, 1997). If the partially amplified segment hybridises to an alternative form of a template (i.e. when amplifying a heterozygous sample), then subsequent PCR products will be recombinants of the two original template sequences (Bradley & Hillis, 1997). The amplified product therefore, may not represent a sequence that actually exists in a single continuous stretch of DNA within the organism.

In addition, where high incidences of *in vitro* PCR recombination have been reported the polymerase enzyme that has been utilised has been *Taq* polymerase (from *Thermus aquaticus*), and has often been termed as "template switching" (Odelberg *et al.,* 1995; Bradley & Hillis, 1997; Zaphiropoulos, 1998). When alternative polymerases are employed, the incidence of recombination has been observed to markedly decrease (Bradley & Hillis, 1997).

Both the C1 and C3 sequences have shown a pattern that would be expected if *in vitro* PCR recombination had occurred. It is also interesting to note that in animals where the C1 and C3 sequences were observed (9228/91-C1, and 9289/91-C3) the only polymerase employed for amplification was *Taq* polymerase. The chances therefore, of *in vitro* PCR recombination occurring during amplification may be higher than if an alternative polymerase was utilised. It appears likely therefore, that the C1 and C3 alleles are artefacts of PCR amplification. If this is the case the only C allele sequence obtained is C2.

Although this seem contrary to Escaygs work where two different C alleles were identified, the number of C sequences obtained during this study were very small. Of the nineteen clones selected and sequenced from animals 91/9307 and 91/9489 which had been typed as C,H using Southern hybridisation, no C sequences were identified. This tends to suggest that the PCR may not amplify all of the DQA_2 alleles equally, and it is possible that not all variations of the C allele may have been amplified or sequenced. In addition, consistient with Escaygs findings recent work has indicated that there may be up to three different variations of the C allele, when animals that had been previously typed at the DQA_2 locus using Southern hybridisation, were typed using a new PCR-SSCP method (Single Stranded Conformational Polymorphism) (Ridgway H.J., personal communication). It is still probable therefore, that there is more than one varient of the C allele, and it is clear more sequencing is required to fully resolve this issue.

Evidence of an Ovine DQA_2 Gene Duplication

The apparent presence of three DQA_2 sequences for animals 91/9019 and 91/9289 suggests that more than one gene is being amplified. All of the sequences however,

appeared to share closest homology with other DQA_2 sequences rather than DQA_1 or DRA. This suggested the presence of haplotypes containing two DQA_2-like sequences in some animals, thus providing evidence for a possible DQA_2 gene duplication in some sheep.

Whilst the presence of three DQA_2 alleles in animal 91/9289 (C2, C3, H) can be explained as being an artefact of amplification, whereby the C3 allele is thought to be a result of *in vitro* PCR recombination, the presence of three alleles in animal 91/9019 (G1, G2, H) appears to be a true result. It was clear from the sequence analysis that there were two different G allele sequences (G1 and G2). The apparent absence of a splice site at the end of exon 2 in the G1 allele suggested that the G1 allele may not encode a functional protein, and may come from a DQA_2-like pseudogene. This suggests that if some animals have a duplicated DQA_2 haplotype, then the duplicated gene could be a DQA_2-like pseudogene.

Gene duplication is thought to have underlain the creation of the MHC. However, the number of MHC genes does not seem to be increasing steadily during evolution, suggesting that the expansions by gene duplication is followed by contraction, either by deletion or inactivation (Trowsdale, 1987; Hughes & Yeager, 1997; Klein & Sato, 1998). In the course of this some of the duplicated genes have lost their function and have become pseudogenes (Trowsdale, 1987). The presence of pseudogenes within the MHC has been a well-documented phenomenon, and has been observed in most species studied to date (Trowsdale, 1987; Milner and Campbell, 1992; Clayton & Gee, 1993). Pseudogenes therefore, are genes that have sequence defects that prohibit their expression. Degeneration of MHC genes over evolutionary time, such that they lose splice sites, or have premature stop codons gives rise to pseudogenes.

Closer inspection of the other allelic sequences revealed that in addition to the G1 allele, the C2, E and J alleles also appeared to lack a functional splice site at the end of exon 2, and may also be from the DQA_2-like pseudogene. This enabled the allelic sequences to be divided into two groups: DQA_2 alleles, and DQA_2-like pseudogene alleles. Alleles B, D, G2, H, K and Q have the expected splice site at the end of exon 2 and are therefore defined as DQA_2 alleles, whilst alleles G1, C2, E and J appear to

lack the expected splice site, and are therefore defined as DQA_2-like pseudogene alleles. Not enough sequence information was considered to be available to confidently assign the F alleles to either group, and because C1 and C3 alleles may be artefacts of PCR amplification they were also not assigned to either group. It is important to note however, that although the G1, C2, J and E alleles appear to lack the expected splice site at the end of exon 2, they may not necessarily be pseudogene sequences. An alternative splice site may be utilised enabling the production of a functional protein. To be able to establish whether these are pseudogene sequences requires expression studies to be conducted such that their functionality can be fully established.

Snibson *et al.*, (1998); and Hickford *et al.*, (unpublished), have also reported haplotypes that possibly contain a DQA_2 duplication in some sheep. This DQA_2 duplication is thought to be exclusively associated with a DQA_1-null allele, thus retaining the pattern of two DQA loci per haplotype.

The DQA_1-null Haplotype

In sheep several authors have documented the presence of a DQA_1-null allele, whereby up to 20 % of animals lack one or both DQA_1 genes (Scott *et al.*, 1991a; Fabb *et al.*, 1993; Escayg *et al.*, 1996). It has been shown that the DQA_1-null haplotype segregates with the DQA_2 alleles F, G, I, J, M, O and P (Hickford *et al.*, unpublished). DQA_1 haplotypes were available for animals within Groups 1 and 2 (Table 3.3) (Escayg, 1995, Escayg *et al.*, 1996). In all instances where the DQA_2-like pseudogene sequence G1 occurred the DQA_1-null haplotype was observed, thus supporting the above hypothesis of an ovine DQA_2 duplication.

Similar results have been observed in cattle, whereby animals that are homozygous null for DQA_1 have haplotypes with two diverse DQA_2-like sequences, one of which belongs to the DQA_2 locus, and the other to a putative third locus, designated BoLA-DQA_3 (Ballingall *et al.*, 1997). In addition, Snibson *et al.*, (1998) have shown that some ovine DQA_2 sequences share a closer similarity to cattle DQA_3 sequences than to other ovine DQA_2 sequences. This suggests that the duplicated ovine DQA_2 gene may be analogous to the cattle DQA_3 gene. Alignment of the ovine-DQA_2, and

DQA_2-like pseudogene sequences with BoLA DQA_3 sequences (Figure 3.5), revealed that the only sequence that clustered with the putative BoLA-DQA_3 gene was the G1 allele. This suggests that the G1 allele comes from a gene that is analogous to the putative DQA_3 gene found in cattle.

Table 3.3: RFLP and Southern hybridisation DQA_1 typing results for animals within groups 1 and 2

Group	Tag Number	DQA_1 Alleles
1	90/257	Null, Null
	89/858	A,C
	86/878	Null, Null
	Imf 001	E, E
	91/9344	A, A
	91/9551	C, C
	1657	Null, Null
	883	B, B
	88/771	Null, Null
	96/211	Not Available
2	Sire 20/88	C, E
	91/9019	Null, E
	91/9135	C, E
	91/9307	E, E
	91/9489	E, E
	91/9562	Null, E
	91/9228	A, C
	91/9289	A, E
	91/9326	Null, C
	91/9566	Null, C

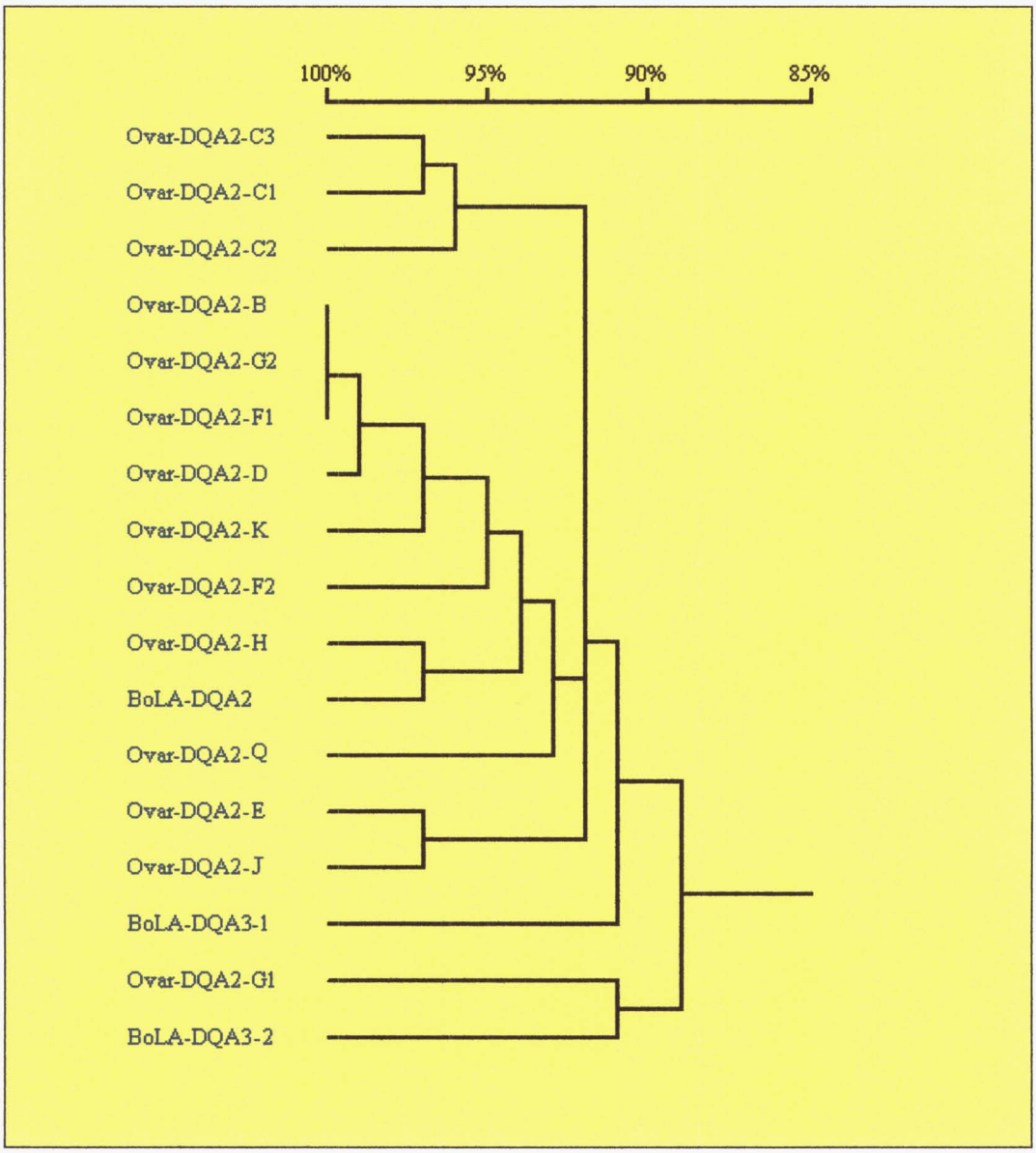


Figure 3.5: Phylogenic tree constructed from ovine DQA_2 , bovine DQA_2 and DQA_3 sequences. The G1 allele clusters with the BoLA-DQA_3 sequences, whilst all the other alleles cluster with DQA_2 (see Appendix H for GenBank accession numbers).

The G1 sequence however, appears to lack the expected splice at the end of exon 2, suggesting that the putative DQA_3 gene in sheep is a pseudogene. This is contrary to that observed in cattle, where the putative DQA_3 gene appears to be transcribed (Ballingall *et al.*, 1998). In addition, further analysis of cattle DQA_3 sequences has shown that the DQA_3 sequences are distinct from DQA_1 and DQA_2, and that they can be sub-divided into two distinct families. Sequence data has suggested that the cattle DQA_3 gene is either an extremely divergent family, or that a DQA_4 gene exists

(Ballingall *et al.*, 1998). Furthermore, Gelhaus *et al.*, (1999), suggests that there may also be a DQA_5 gene in cattle, which is confined to certain haplotypes. Further investigation is required in sheep to determine whether there is similar complexity to that observed in cattle DQA genes.

E and C2 Alleles

Whilst most evidence suggests that the DQA_2 duplication coincides with the DQA_1-null haplotype (Snibson *et al.*, 1998; Hickford *et al.*, unpublished). The DQA_1-null haplotype does not segregate with E or C alleles. In this study however, the sequence data suggests that the E and C2 alleles may be from a DQA_2-like pseudogene. This gives rise to two possible scenarios. Firstly, if the C and E alleles are from a DQA_2-like pseudogene, then no sequence has been obtained from the DQA_2 gene. This suggests that contrary to other studies, the DQA_2 duplication is not exclusively associated with the DQA_1-null allele, and that there may be a possible DQA_2-null haplotype. The second scenario is that the E and C2 alleles may be from the DQA_2 locus, but over time have accumulated an aberrant splice site such that they may no longer be able to produce a functional protein. This suggests that they have been incorrectly defined as coming from the DQA_2-like pseudogene, rather than the DQA_2 gene.

Of the two possible scenarios, the second seems the most likely. Unlike the G1 allele, the E and C2 alleles clustered with the other DQA_2 sequences, rather than the putative DQA_3 sequences. This suggests that it is more likely that the E and C alleles came from the DQA_2 locus. If this is the case then similar to the other studies the DQA_2 gene duplication seems to be exclusively associated with the DQA_1-null allele.

Whichever argument is correct, animals that have an E or C2 haplotype appear to only have a functional DQ product that is DQA_1 encoded. The existence of a DQA_1-null haplotype has often brought into question the functional significance of the DQA_1 gene. Of the 10-20 % of the animals that lack a DQA_1 gene, the genetic diversity of the DQA_2 gene is thought to provide the necessary diversity for appropriate MHC function (Scott *et al.*, 1991a; Fabb *et al.*, 1993; Escayg *et al.*, 1996;

Snibson *et al.*, 1998). The question that then arises is why have a DQA_1 gene at all? What is the advantage? If however, there are some animals that may not be able to produce a functional DQA_2 encoded product, such as animals that have a C or E DQA_2 haplotype, then the existence of a DQA_1 gene may be necessary to provide adequate MHC restriction. In such cases, if a DQA_1 gene did not exist there would be no functional DQ product. This may be a severe disadvantage, as the necessary MHC restriction for immune system initiation towards particular pathogens may not be provided by other class II loci. The existence of animals that may not be able to produce a functional DQA_2 product provides a possible explanation therefore for the requirement of a DQA_1 locus in sheep.

A Possible Theory for the Evolution of the Ovine DQA Genes

From the available sequence data from this study it seems that three DQA haplotypes may be possible: 1) DQA_1 / DQA_2; 2) DQA_1-null / DQA_2 / DQA_3 (pseudogene); 3) DQA_1 / DQA_2 (pseudogene). In each case two DQA loci per haplotype have been maintained, but in two out of the three haplotypes there may only be one gene that produces a functional protein.

Given that all three haplotypes have been observed in sheep the question arises as to which is the ancestral state. The close homology between the DQA_2 and the putative DQA_3 locus tends to suggest that they arose from a gene duplication event, from a DQA_2-like ancestral gene. In addition, the presence of a duplicated DQA_2-gene in both cattle and sheep suggests that the duplication event took place prior to the divergence of the separate species, approximately 15 million years ago (Klein & Figueroa, 1986). It is possible that following the gene duplication event, these genes have undergone independent evolution to produce the present DQA_1 and DQA_2 alleles. This would mean that the DQA_1 gene is relatively new, and that it arose from the DQA_3 gene (Figure 3.6). Supporting this theory for the evolution of DQA genes are the observations that the presence of all three DQA genes (ie. DQA_1, DQA_2, and DQA_3) have never been found in any sheep or cattle haplotypes (Ballingall *et al.*, 1997; Snibson *et al.*, 1998), and three DQA genes have never been mapped in sheep (Scott *et al.*, 1991a).

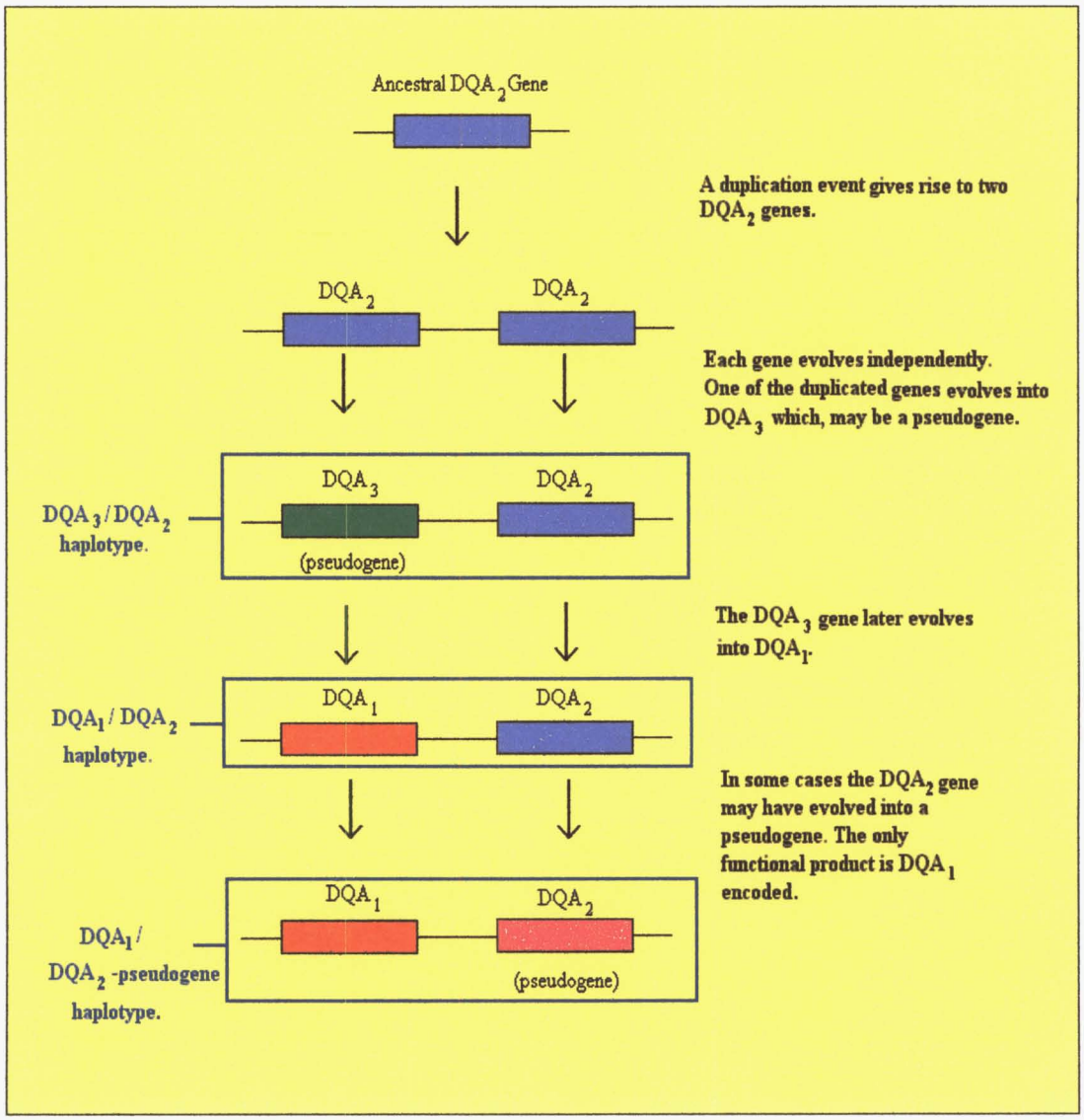


Figure 3.6: Schematic diagram representing the possible evolution of ovine DQA genes. A gene duplication event of an ancestral DQA_2 gene gives rise to two DQA_2-like genes. Both genes evolve through the actions of various selective pressures which, is independent from each other, giving rise to DQA_2 and DQA_3 genes. Some animals retain a DQA_2 / DQA_3 haplotype, while continued selective pressure in some populations causes the DQA_3 gene to evolve into the DQA_1 gene, providing the most common haplotype observed to date, the DQA_1/DQA_2 haplotype. A further possible step is that the DQA_2 gene in some animals has accumulated an aberrant splice site, such that it becomes a pseudogene, giving a third haplotype of DQA_1/ DQA_2 pseudogene. This suggests that the DQA_1 gene is evolutionally younger than DQA_2.

However, this theory for DQA gene evolution may be controversial because the results from this study suggest that the putative DQA_3 gene is likely to be a pseudogene. Currently, opinion is divided as to whether pseudogenes are sources of functional new genes, or simply redundant genes slowly degenerating into oblivion. If the latter is true then the DQA_1 gene could not have evolved from DQA_3 gene, as proposed here. However, it has been reported recently, that the ribonuclease gene cluster may have evolved from pseudogenes (Trabesinger-Reuf *et al.*, 1996). This supports the proposal that pseudogenes may not necessarily be redundant or "junk" DNA, and that gene conversion events may be able to repair pseudogenes to form fully functional genes. It is possible therefore that the DQA_3 gene may have been repaired, through a gene conversion event, giving rise to a functional DQA_1 gene.

It is important to note, however, that this is only one possible theory. The sequence data obtained in this study provides a "snap shot" of the DQA genes at this point in time, and the possible evolution of the DQA genes can only be inferred from the patterns observed. Subsequent studies may find a different pattern that may provide support for an alternative theory for the evolution of the ovine DQA genes.

Functional Relatedness of the Sequences within Exon 2

Consistent with other studies (Scott *et al.*, 1991a; Snibson *et al.*, 1998), most of the polymorphism observed between the different ovine DQA_2 alleles occurred within exon 2. This exon is considered the most important in terms of the function of the resulting histocompatibility protein, as it encodes the antigen-binding groove. Residues within the antigen-binding groove define the specificity of the bound pathogenic peptide (Stern *et al.*, 1994). Different alleles that show high homology within exon 2 are therefore expected to have a similar peptide specificity, whilst alleles that are less homologous are likely to have a different peptide specificity, and may not be able to effectively bind peptides from the same pathogen.

In humans many studies have found associations between human MHC loci and disease. Sometimes however, subsequent studies fail to find the same association (Powis and Trowsdale, 1991). Such inconsistencies are not surprising when the consequences of natural selection that favour heterozygosity in the peptide-binding

groove are considered. Past recombination events, and/or convergent evolution means that alleles that are not closely related in overall sequence, may come to resemble each other in the peptide-binding groove (Hughes *et al.*, 1996). As a result, an alternative strategy of searching for MHC-disease associations is to group alleles into functional categories based on shared antigen-binding groove sequence/motifs, rather than on the basis of shared primer sites in less functionally important regions (Hughes *et al.*, 1996).

The ovine DQA_2 alleles sequenced in this chapter were grouped into such "functional" categories by aligning exon 2 (Figure 3.2). This revealed two clusters of alleles that share close homology. The first cluster consists of alleles B, D, F1, F2 and G2, while the second cluster consists of alleles E and J.

Within each cluster, the resulting histocompatibility proteins would be expected to have similar antigen-peptide specificity, and therefore the same function in terms of their ability to present peptides from particular pathogens. This was observed in the results obtained from RFLP and Southern hybridisation typing of ovine DQA_2 alleles and associations with footrot susceptibility or resistance (Chapter 2). Alleles D and G tended to associate with footrot resistance, although not always to a significant level, whilst the E allele tended to associate with a higher footrot susceptibility.

In Chapter 2, the B and F alleles however, tended to associate with higher footrot susceptibility. Nevertheless, the frequency of these alleles within the flocks analysed, and the low prevalence of footrot may have artificially inflated the relative risk of each allele, and further analysis is required to be able to fully assess their effect.

As a result of the high homology between the J allele and E allele sequences it is expected they would have similar footrot susceptibility. However, in Chapter 2, the J allele tended towards greater footrot resistance. Nevertheless, as with the B and F alleles, the frequency of the J allele within the flocks analysed was low, which may cause the relative susceptibility of this allele to be biased. Furthermore, typing animals at the DQA_2 locus using PCR-SSCP, has shown that the J allele may be sub-divided into J1 and J2 (Ridgway H.J., personal communication). It is possible that

one J allele associates with lower footrot susceptibility compared to the other J allele. It is clear therefore, that more detailed sequence analysis, and further disease association studies are required before the effect that the J allele(s) has on footrot susceptibility or resistance can be fully assessed.

Alleles C, K and Q show varying levels of homology to each other and to either cluster, suggesting that their peptide specificity may also be different. The C and Q alleles have shown some associations with footrot resistance (Chapter 2), but there have been no significant associations with the K allele and footrot susceptibility or resistance to date.

Homology at the functional level has therefore been shown to be related to homology at the sequence level. Such a relationship supports the suggestion that there is not a single "protective" allele, and other "susceptible" alleles, but rather a continuum of protectiveness, that is correlated with a specific conformation of the antigen-binding groove, which is in turn determined by specific exon 2 sequences (Hughes *et al.*, 1996).

Whilst this study has provided useful sequence information for the ovine DQA_2 gene, several issues remain unresolved. It is clear that further sequence information is required for the F and C alleles to determine if the variation observed here exists, or is a result of PCR/sequencing errors. In addition, more sequencing may identify other alleles that may come from the putative DQA_3 gene. This would help to determine whether the DQA_3 exclusively associates with the DQA_1-null haplotype, and determine whether DQA_4 or DQA_5 genes exist in sheep. Expression studies are also required to assess which of the DQA genes are capable of producing a functional protein.

3.4 Conclusion

DNA sequencing provides an important tool for characterising unknown gene fragments. In this study, numerous clones containing an 890 bp DNA fragment, which were shown to be from the ovine DQA_2 gene were sequenced. Sequences corresponding to alleles B, C, D, E, F, G, H, J, K, and Q, were identified with subsequent sub-division of the C, F and G alleles into C1, C2, C3, F1, F2, G1, and G2, giving a total of 14 sequences.

The results provided some evidence of a possible DQA_2 gene duplication in some sheep, which has been described as a putative DQA_3 gene. The DQA_3 gene described here is thought to be a pseudogene because of the apparent lack of a splice site at the end of exon 2, and this gene is thought to be exclusively associated with a DQA_1-null haplotype. From the sequence information obtained, three DQA haplotypes in sheep seemed possible: DQA_1/DQA_2; $DQA_2/DQA_3/DQA_1$-null; DQA_1/DQA_2 (pseudogene). A possible theory for the evolution of ovine DQA genes has been proposed, whereby the DQA_1 gene may have evolved from the putative DQA_3 gene. This suggests that the DQA_1/DQA_2 haplotype is evolutionally younger than the DQA_2/DQA_3 haplotype.

Like class II sequences from other species most polymorphism within the ovine DQA_2 gene occurs within exon 2. The results also suggest that functional homology is related to sequence homology, and that categorising alleles into functional groups based on antigen-binding groove sequences may prove to be more successful when investigating MHC-disease associations.

Overall this study shows that the number of expressed DQA genes may vary depending on the haplotype, suggesting that the ovine DQ sub-region is more complex than that observed in man. However, more extensive sequence analysis and expression studies are required to be able to fully characterise the apparent complexity of the ovine DQ sub-region.

CHAPTER FOUR
ANTIGEN-BINDING GROOVE POLYMORPHISM AND ASSOCIATIONS WITH FOOTROT SUSCEPTIBILITY.

Introduction

The antigen-binding groove of histocompatibility proteins is responsible for binding antigenic peptides, and is considered to be one of the most important domains of histocompatibility proteins. Polymorphic residues cluster within the antigen-binding groove, a result of the polymorphism within exon 2 of the histocompatibility genes.

Residues within the antigen-binding groove must be able to bind the residues of the antigenic-peptide backbone and accommodate antigenic-peptide side chains, to enable adequate peptide capture for display to the immune system. It is therefore, the residues present within the antigen-binding groove that determine how effectively a particular antigenic-peptide will be bound (Brown *et al.*, 1993; Stern *et al.*, 1994). By studying the amino acid residues within the antigen-binding groove, it may be possible to determine the impact that polymorphism has on antigenic-peptide binding ability, and may provide a possible explanation as to why certain MHC alleles afford disease resistance, whilst others disease susceptibility.

Structural analysis of the human class II DR1 histocompatibility protein (HLA-DR1) (Brown *et al.*, 1993; Stern *et al.*, 1994), revealed the physical nature of the class II antigen-binding groove. Specific positions within the groove were identified that were important for either binding the antigenic-peptide backbone, or in forming pockets which accommodate antigenic-peptide side chains, providing peptide binding specificity. Amino acid residues involved in the binding of the peptide backbone were highly conserved in HLA-DR sequences, indicating the binding of similar antigenic-peptide fragments, whilst amino acid residues that formed the pockets were polymorphic, defining the specific antigenic-peptide to be bound.

It has not been possible to obtain structural information for histocompatibility proteins from all class II loci, as it has proven difficult to obtain proteins of adequate concentration and purity for crystallographic studies. However, Paliakasis *et al.* (1996) modelled the structure of human DQ histocompatibility proteins (HLA-DQ) from the published structure of DR, as they share close homology at both the amino acid level, and at the functional level. Whilst some novel structural features particular to DQ molecules were observed, the basic structure of the antigen-binding groove was conserved. Positions within the groove that had been identified as being important for either peptide backbone binding or the formation of polymorphic pockets in HLA-DR1 were also found to be important in HLA-DQ's binding groove. In some cases, specific amino acid residues at these positions had been conserved in both the HLA-DR and –DQ sequences analysed.

Ovine DR (OLA-DR) sequences also share close identity to HLA-DR (88 %) at the amino acid level, suggesting they have a similar structure. In addition, OLA-DR and –DQ sequences share close homology, and reasonable amino acid identity (at least 60%), suggesting that they may also have a similar structure. It follows, that the positions in the antigen-binding groove of HLA-DR molecules that are important for peptide backbone binding and pocket formation may also be similar in OLA-DR and –DQ molecules.

As a result of the close homology between human and ovine DR and DQ sequences, it is not unreasonable to model the structure of the OLA-DQA_2 binding groove from the structure of HLA-DR1. This may suggest how allelic polymorphism could alter the peptide binding capacity at the positions that have been identified as being important for peptide binding and specificity.

In this study, the DNA sequences obtained for DQA_2 alleles B, D, G2, H, K, and Q, and DQA_2 pseudogene alleles C2 and E (Chapter 3), were translated into amino acid sequence. The structure of the DQA_2 antigen-binding groove was modelled based on the published structure of HLA-DR1 (Brown *et al.*, 1993; Stern *et al.*, 1994). Although the

sequences for alleles C2 and E were thought to be DQA_2 pseudogene sequences because of the apparent lack of a splice site at the end of exon 2, (Chapter 3), it is possible that an alternative splice site may be used, enabling them to produce a functional protein. They were therefore included in this analysis to see if they could produce a histocompatibility protein with a similar groove structure, and peptide binding specificity of the other DQA_2 alleles.

The purpose of this modelling exercise was to determine whether polymorphism observed within the ovine DQA_2 chain affected the antigen-binding groove, such that, its peptide specificity and peptide-binding ability may be altered. This could provide a possible explanation as to why some alleles of the ovine DQA_2 locus have been observed to associate with footrot susceptibility, while others associate with resistance.

4.1 Methods

DNA sequences obtained for exon 2 of DQA_2 and DQA_2 pseudogene alleles B, C2, D, E, G2, H, K, and Q (Chapter 3), were translated into amino acid sequences, using DNAMAN© version 4a (Lynnon Biosoft), based on the sequence of Scott *et al.*, (1991a). Both the DNA and amino acid sequence for HLA-DR1, OLA-DR, and OLA-DQA_1 were obtained from GenBank, and the co-ordinates for the structure of human DR1 was obtained from the Swiss-Prot protein database. Swiss model Pdb viewer (Glaxo Wellcome, Geneva Switzerland) was used to manipulate the human DR1 molecule to obtain the putative 3D structure of the ovine DQA_2 molecule. The sequence of the G2 allele was used as the template.

4.2 Results

Figure 4.1 shows the alignment for amino acid residues 24-79 of the α-chain of the translated human DRA1 (M60333), human DQA_1 (L34089), ovine DRA (Z11600) and ovine DQA_2 and DQA_2 pseudogene sequences (B, C2, D, E, F2, G2, H, K, and Q). The position of amino acid residues in the binding groove identified by Stern *et al.* (1994) and Paliakasis *et al.* (1996), as being important for peptide-binding, and pocket formation in human DR and DQ molecules, have been highlighted.

There is 52-61% identity in amino acid residues in the α1 domain of ovine DQA_2 sequences compared to DR1. Within the ovine DQA_2 and DQA_2 pseudogene sequences there is 74-98% identity in the same domain, with the E and C2 alleles being the least conserved.

The structure of the α-chain of the antigen-binding groove, covering amino acid residues 24-79, was modelled for the ovine-DQA_2 and DQA_2 pseudogene sequences using the co-ordinates from HLA-DR1 (Stern *et al.*, 1994) and the G2 sequence as a template. HLA-DR1 has three amino acid residues less than HLA-DQ and ovine DQ sequences (Paliakasis *et al.*, 1996), and the amino acid residues have been numbered according to HLA-DQ.

Figure 4.2a shows the modelled structure obtained for ovine DQA_2. In accordance with HLA-DR1, the side of the modelled DQA_2 groove has an α-helical structure, whilst the floor of the groove has a β-sheet structure (Stern *et al.*, 1994). Figure 4.2b shows a simplified diagram of the DQA_2 groove, where amino acid residues have been numbered according to HLA-DQ (Paliakasis *et al.*, 1996). The hydrogen bonds that may be formed between groove amino acid residues, enabling the maintenance of the helical and strand structures were calculated by the Swiss model Pdb viewer program.

In HLA-DR1 and DQA_1 the amino acid residues which are thought to form the pockets that define peptide specificity have been identified (Stern *et al.*, 1994; Paliakasis *et al.*,

1996). As the HLA molecules share reasonable identity with OLA molecules at the amino acid level, it is thought that the positions of pocket residues are the same for ovine DQA_2. Figure 4c shows the positions of amino acid residues that may be involved in the formation of the first and fifth pockets in ovine DQA_2.

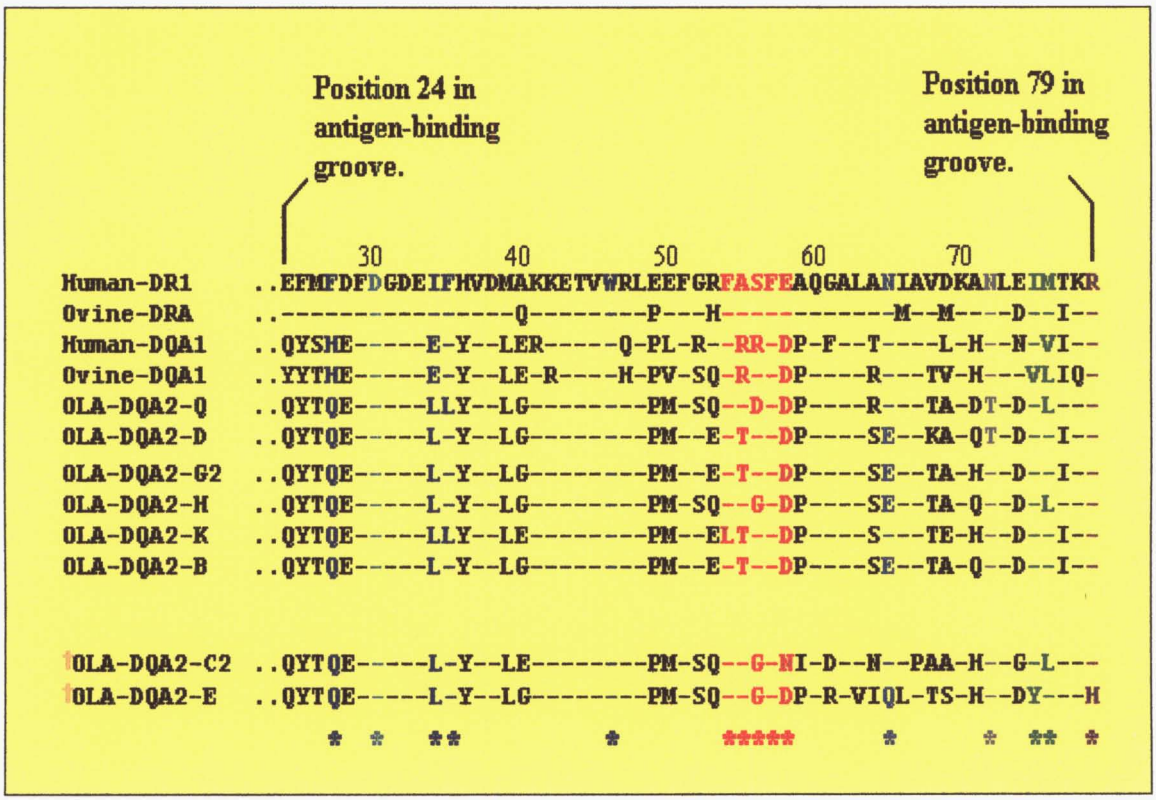


Figure 4.1:Comparison of the amino acid sequences of HLA-DR1 (M60333), HLA-DQA_1 (L34089), OLA-DRA (Z11600), OLA-DQA_1 (M33304), and the OLA-DQA_2 and DQA_2 pseudogene alleles. Amino acids are in the one letter code. DR1 has three amino acids less than DQ, and the numbering refers to DQ. Code: (†) DQA_2 pseudogene sequences; (-) identity; (*) Position of residues that form the first pocket; (*) Position of a residue involved in the formation of an intramolecular salt bridge; (*) Position of residues that are involved in binding the antigenic-peptide backbone, and allow the orientation of groove side chains, such that they do not sterically interfere with N-terminal extensions of the peptide; (*) Position of a residue involved in binding the antigenic-peptide backbone; (*) Position of a residue involved in antigenic-peptide backbone binding, and formation of the fifth pocket; (*) Position of residues that form the fifth pocket; (*) Position of a residue that is involved in antigenic-peptide backbone binding, formation of the fifth pocket, and forming a salt bridge with the β-chain of the histocompatibility heterodimer.

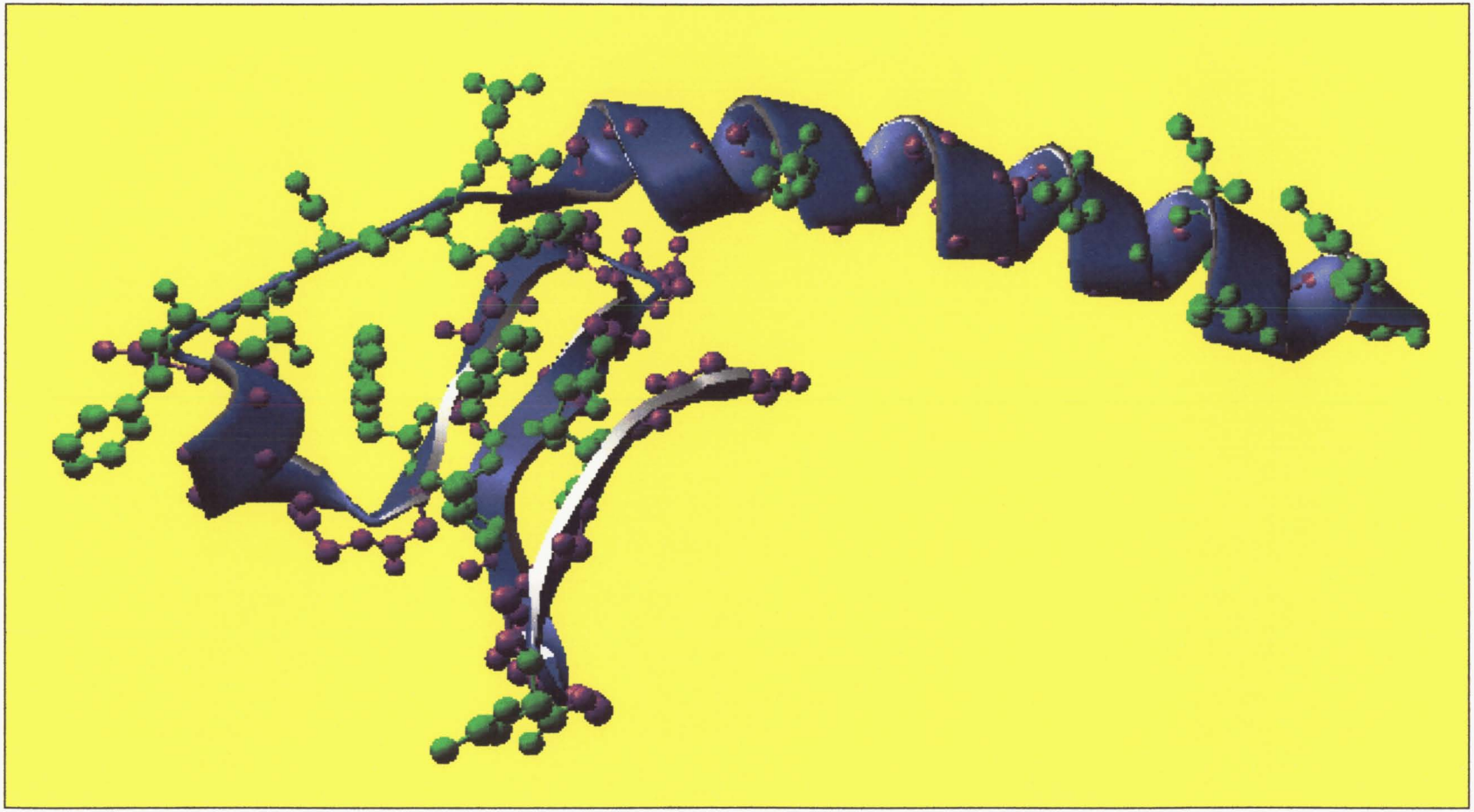


Figure 4.2a: Predicted structure of the ovine-DQ alpha chain (amino acid residues 24-79). The main chain is coloured purple, while the side chains of the amino acid residues which have been identified as being important for peptide binding and pocket formation in HLA-DR1 and HLA-DQA_1 (Stern *et al.*, 1994; Paliakasis *et al.*, 1996), are shown in green. The ribbon structure has been superimposed over the main chain, revealing the secondary helical and strand structures.

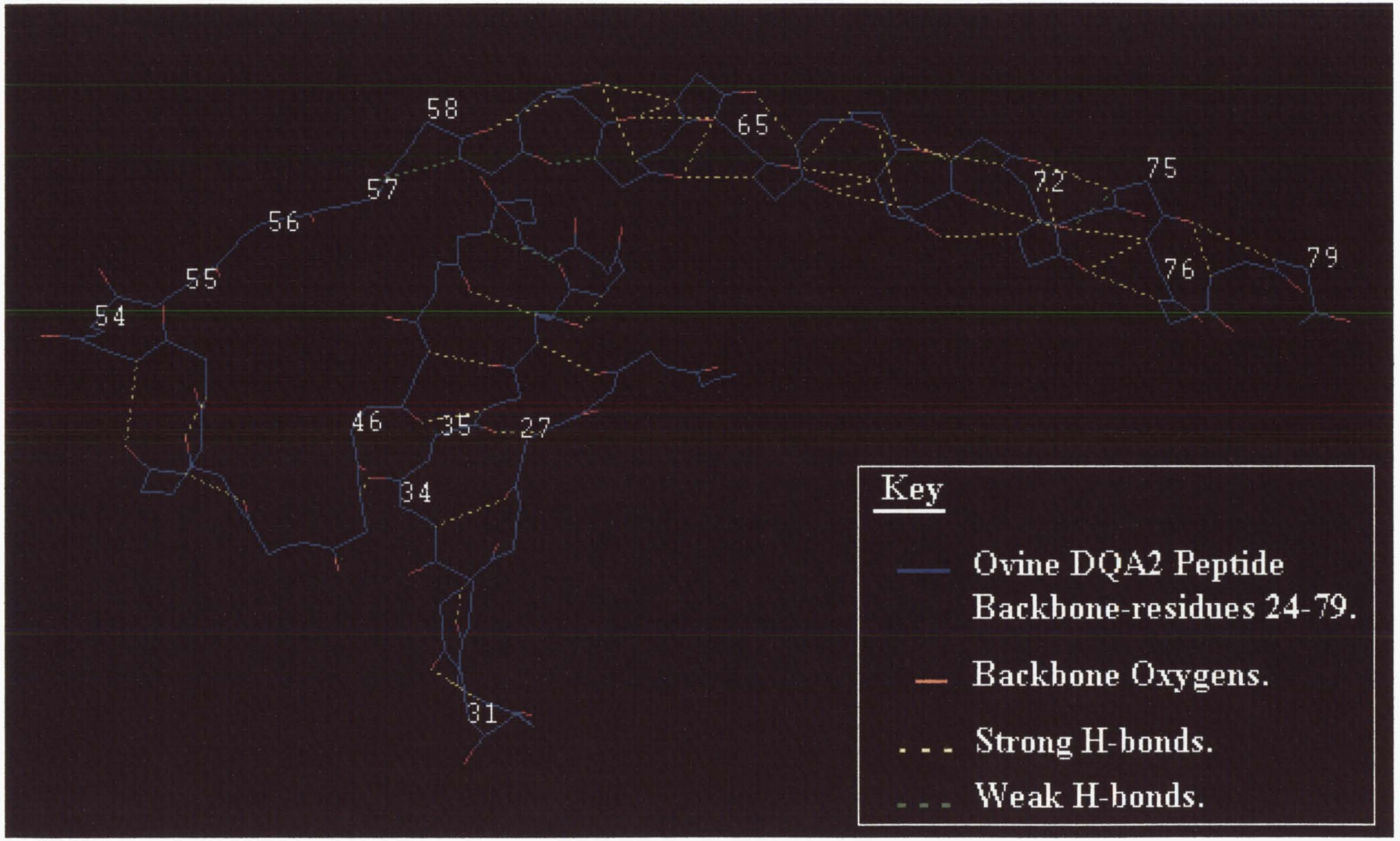


Figure 4.2b: Amino acid residues 24-79 of the alpha chain of ovine DQA_2. Carbon and nitrogen atoms are coloured blue, while oxygen atoms are coloured red. Hydrogen bonds formed between groove residues have been calculated, and the residues numbered according to DQ. Side chains have been omitted for simplicity.

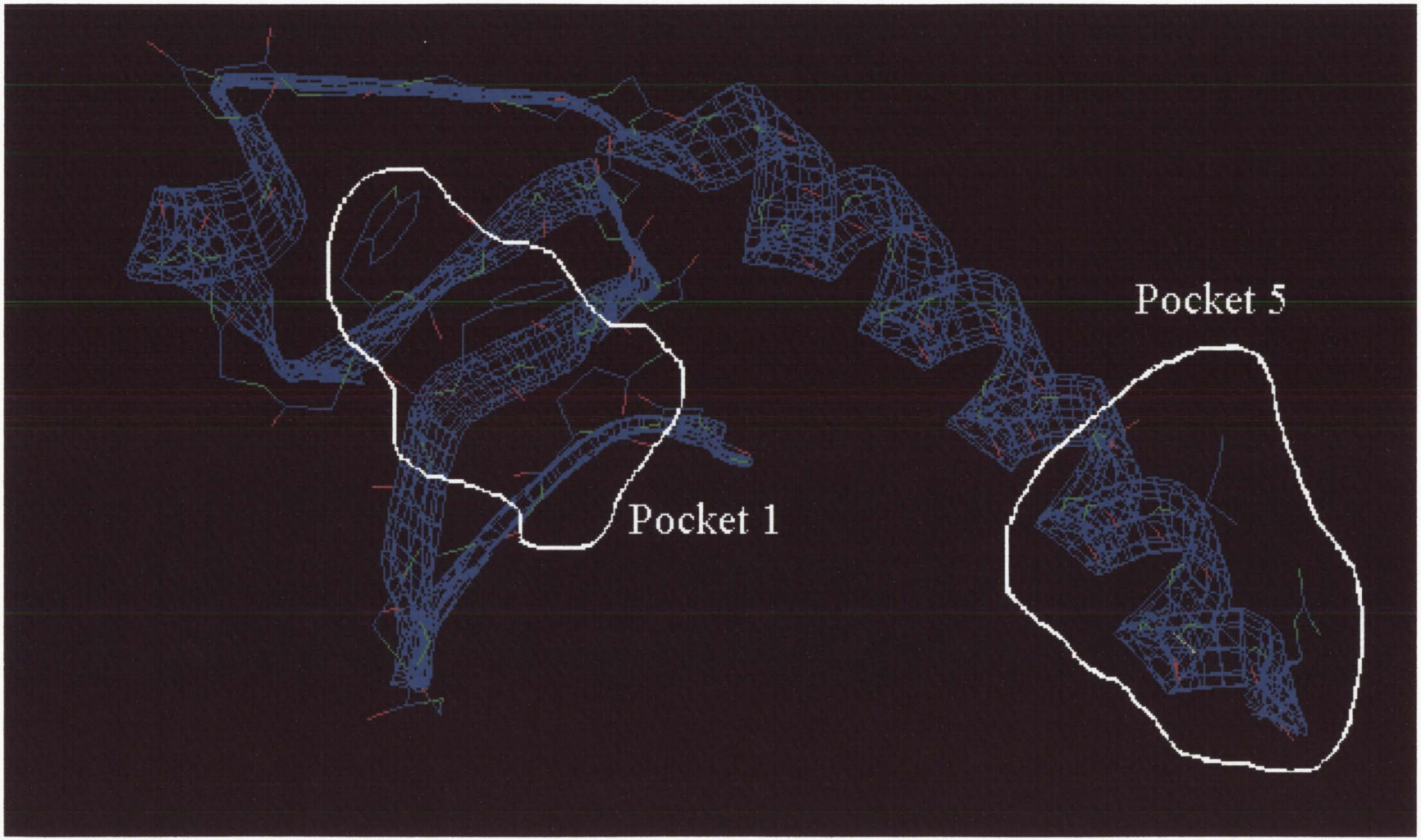


Figure 4.2c: Position of the first and fifth pockets in ovine DQA_2. The side chains of the amino acid residues which, contribute to each of the pockets have been included.

Hydrogen Bonding to the Antigenic-peptide Backbone

Amino acid residues that form hydrogen bonds to the foreign peptide's backbone in HLA-DR1, specifically α-65Asn, α-72Asn, and α-79Arg, were found to be conserved in the same positions in HLA-DQ, (Stern *et al.*, 1994; Paliakasis *et al.*, 1996). This was not observed for the ovine DQA_2 or DQA_2 pseudogene alleles. At position α-65, alleles D, G2, H, and B have glutamic acid, and the E allele has glutamine. At position 72, asparagine is conserved among most of the ovine DQA_2 alleles, the exceptions being the Q and D alleles, which have threonine at this position. At position 79 arginine has been conserved for all ovine alleles except E, which has histidine at this position.

Salt Bridge Formation

The formation of salt bridges between class II α1 and β1 domains contributes to the stability of the overall structure, and has been found to be invariant in human DR and DQ molecules (Brown *et al.*, 1993; Stern *et al.*, 1994; Paliakasis *et al.*, 1996). One of these salt bridges occurs between amino acid residues α-7 and α-30. Whilst data was not available for α-7 in the ovine sequences, aspartic acid was invariant at position α-30 for all human and ovine DR and DQ sequences analysed.

As well as forming a hydrogen bond Arg-79 is important for the formation of a salt bridge with β-57 in human molecules, stabilising the two chains that form the antigen-binding groove (Stern *et al.*, 1994). The only ovine allele in which Arg-79 was not conserved was the E allele sequence.

Binding the N-terminus of the Antigenic-peptide

Human DQ amino acid residues α54-58 (α51-55 in DR) have been identified as being important for the orientation of HLA side chains, such that they do not sterically interfere with the N-terminus of the bound peptide (Stern *et al.*, 1994; Paliakasis *et al.*, 1996). Additionally, α-54 and α-56 are involved in the binding of the antigenic-peptide backbone.

In the ovine DQ molecules this region appeared as one of five different combinations. The most common motif was Phe-Thr-Ser-Phe-Asp, and was observed for alleles D, G2, and B. At position α-54 phenylalanine was conserved in all alleles except K, which had leucine at this position, with the rest of the motif identical to the one above (Leu-Thr-Ser-Phe-Asp). Another common motif for this area in DQA_2 alleles C2, E, H, and Q is Phe-Ala-Gly-Phe-Asp, although alleles C2 and Q each had a slight variation of this motif. Alleles C2, E, and H, have alanine at position α-55 instead of threonine, and a glycine instead of serine at α-56, while the Q allele has alanine and aspartic acid.

Side-chain Pockets and Peptide Binding Specificity

Amino acid residues from the α-chain which have been found to contribute to the formation of the first and fifth pockets in human DR and DQ molecules (Stern *et al.*, 1994; Paliakasis *et al.*, 1996) were identified in the ovine DQA_2 and DQA_2 pseudogene sequences (Figure 4.2c).

The First Pocket

Residues from the α-chain, which are thought to contribute to the first pocket (α-10, α-27, α-34, α-35, and α-46), indicate that this pocket is hydrophobic in ovine DQA_2 and DQA_2 pseudogene molecules, and is similar to that observed for human DR and DQ (Stern *et al.*, 1994, Paliakasis *et al.*, 1996). Among the different ovine DQA_2 molecules, much of this pocket is conserved for all α-chain amino acid residues, the exception being the Q and K alleles where phenylalanine is replaced by leucine at position 35.

The Fifth Pocket

Amino acid residues from the α-chain, which are thought to contribute to the fifth pocket, (α-72, α-75, α-76, and α-79 (Stern *et al.*, 1994; Paliakasis *et al.*, 1996)), indicate that this pocket has a hydrophobic interior which is flanked by either charged or neutral residues for all ovine DQA_2 and DQA_2 pseudogene molecules. The exception is the E allele, where isoleucine at α-75 is substituted by tyrosine, and the normally conserved arginine at position α-79 is replaced by histidine.

4.3 Discussion

Is it Valid to Model the Structure of OLA- DQ from HLA-DR1?

In this study considerable amino acid identity between the human and ovine DR and DQ sequences was observed, suggesting that the structure of the antigen-binding groove is similar for ovine DQA_2 (Figure 4.2a).

Although the basic structure of the groove is likely to be conserved across different species, several assumptions have been made in this study. It has been assumed that areas identified in human DR and DQ molecules (Brown *et al.*, 1993; Stern *et al.*, 1994; Paliakasis *et al.*, 1996) as being important for peptide backbone binding and pocket formation are the same for ovine DQA_2 molecules. In human DR and DQ molecules, the pockets that provide antigen peptide specificity have been assigned on the basis of depressions in the binding groove (Paliakasis *et al.*, 1996), and it was assumed that similar depressions occur in the same place in ovine DQ (Figure 4.2c). However, all the depressions in the antigen-binding groove of human DQ molecules may not belong to pockets, and the number of pockets and the amino acid residues constituting each pocket may vary depending on the nature of the antigenic peptide bound (Stern *et al.*, 1994; Paliakasis *et al.*, 1996). It follows, that the number of pockets and the amino acid residues that constitute each pocket in ovine DQA_2 may also vary depending on the peptide that is bound. The pockets that have therefore been defined in this study may not occur in real ovine DQ molecules.

The only way to completely resolve these issues would be to obtain ovine DQ glycoproteins of appropriate concentration and purity, such that the actual structure and the interactions between amino acid residues in the antigen-binding groove and the antigenic peptide could be observed. However, whilst modelling the structure of ovine DQA_2 from human DR1 and DQ is not ideal because of the assumptions that are made, some interesting information can still be gleaned from such a study. In addition, the assumptions can be balanced against the fact that both the ovine and human DR and DQ

molecules share close homology at both the functional and amino acid level, and that it is likely that the structure of class II MHC molecules is conserved in different species.

Why Study Specific Peptide Motifs in the Antigen-binding Groove of Class II Molecules?

In humans, several studies have linked increased susceptibility to particular diseases to specific peptide motifs within histocompatibility proteins. One example of this is the non-conservation of aspartic acid at position 57 in the β-chain, which associates with increased susceptibility to insulin dependent diabetes mellitus (IDDM) (Todd *et al.*, 1988). Another example is the increased susceptibility to rheumatoid arthritis with the presence of particular HLA-DRB alleles (Seidl *et al.*, 1997; Toussirot, *et al.*, 1999). It has been shown that the presence of the amino acid motif (Gln) Arg/ Lys Arg Ala Ala at position 70-74 within the binding groove of DRB alleles, specifically the presence of lysine at position 71, is responsible for increased susceptibility to rheumatoid arthritis.

These studies in humans have illustrated the importance of specific amino acid residues within the antigen-binding groove and how antigenic-peptide binding ability can be altered with only one or two amino acid changes in some alleles. By analysing specific amino acid residues within the histocompatibility proteins of different alleles in sheep, it is hoped that a better understanding can be gained for why particular alleles associate with increased susceptibility to particular diseases.

The α-chain of the Antigen-binding Groove for Ovine DQA_2 Alleles B, D, G2, H, K and Q

Although close amino acid identity was observed between ovine DQA_2 and human DR1, and DQA_1, most of the amino acid variation in the B, D, G2, H, K and Q sequences occurred in areas that had been identified as being important for binding the antigen peptide backbone and pocket formation. This could suggest that the same amino acid residues identified in human molecules as being important for peptide binding and pocket formation are not the same in ovine molecules. Alternatively however, it may

suggest that ovine DQA_2 molecules are likely to have different peptide binding specificities to human class II molecules. This would be expected, as both humans and sheep would encounter different pathogens, and as a result, MHC molecules that have different peptide binding specificities would need to be produced in each species, thus perhaps accounting for the variation observed.

Amino acid residues found to be important for the formation of hydrogen bonds to the antigenic peptide backbone are invariant in human DR and DQ molecules, specifically α-65, α-72, and α-79 (Brown *et al.*, 1993; Stern *et al.*, 1994). This was not observed in the ovine DQA_2 alleles, with some variation occurring in these areas, although amino acid residues at these positions for the K allele were the same as those observed for human DR and DQ. Most of the amino acid variation in the other DQA_2 alleles seem to be neutral substitutions, whereby their ability to form a hydrogen bond with the peptide backbone was unlikely to be affected, although the strength may be altered.

It is possible that the ovine K allele may hydrogen bond with the antigenic backbone slightly differently from the B, G2, and H alleles, and in turn, the D and Q alleles may also have a slightly different hydrogen bonding potential. Whether this means that some alleles cannot successfully bind antigenic peptides, or that they might bind particular peptides better than others, is purely speculative and remains to be resolved.

As a result of the histocompatibility proteins being a heterodimeric molecule, salt bridges are required to stabilise the α- and β-chains. The formation of salt bridges in class II molecules is important for the stability of the groove, keeping the groove in the correct conformation to allow antigenic peptides to be bound (Stern *et al.*, 1994). Amino acid residues in the α-chain that have been found to contribute to salt bridge formation are α-7, α-30 and α-79. Whilst data was not available for α-7, in all of the ovine DQA_2 sequences the amino acid residues at positions α-30 and α-79 were invariant when compared to human DR1 and DQ. This suggests that amino acid residues at these positions in ovine DQ molecules are also important for the formation of salt bridges, and groove stability, and have therefore been conserved.

At the N-terminal side of the class II binding site in human DR and DQ molecules α-chain, amino acid residues α54-58 (α51-55 in DR), have been identified as being important for the orientation of HLA side chains such that they do not sterically interfere with the N-terminal extensions of the peptide (Stern *et al.*, 1994; Paliakasis *et al.*, 1996). Unlike class I molecules, class II are open at the ends and as a result, they can accommodate larger peptides. Most often, residues that form the N-terminal end of class II molecules are smaller residues than those observed for class I, which are commonly tyrosine and tryptophan.

Generally all of the ovine DQA_2 alleles have amino acid residues with similar characteristics to those found in human DR1 and DQ molecules. Phenylalanine at position α-57 has been conserved among all DQA_2 and DQA_2-like pseudogene alleles, and is also invariant in human DR and DQ (Paliakasis *et al.*, 1996). Although some variation in the amino acid sequence was observed between the different DQA_2 alleles, most of the substitutions are neutral, such that binding of the N-terminus of the antigenic peptide is likely to be unaffected. The K allele may however, be less restrictive on the size of N-terminus residues in the antigenic peptide because of the Phe→Leu substitution at position α-54, while the Q allele may prefer slightly smaller residues in this region because of an Asp→Glu substitution at position α-58.

Within the peptide-binding groove several smaller cavities, or "pockets", accommodate side chains of the bound peptide, five of which have been identified in human DR and DQ molecules (Stern *et al.*, 1994; Paliakasis *et al.*, 1996). Many of the amino acid residues that line these pockets are highly polymorphic, and it is this polymorphism that is thought to be responsible for the different peptide binding specificities of different class II proteins. Amino acid residues from the α-chain which have been identified in human DR and DQ molecules that contribute to the first pocket are α-10, α-27, α-34, α-35, and α-46, (Brown *et al.*, 1993; Stern *et al.*, 1994; Paliakasis *et al.*, 1996). It was assumed that these amino acid residues are also important in ovine DQA_2 molecules. The first pocket motif was highly conserved in all DQA_2 alleles, creating a largely hydrophobic pocket, that is capable of binding antigenic peptides that have large

hydrophobic amino acid residues, and has similar characteristics to that observed in human DR and DQ molecules (Stern *et al.*, 1994; Paliakasis *et al.*, 1996).

Amino acid residues from the α-chain which have been defined as contributing to the fifth pocket in human DR and DQ molecules are α-72, α-75, α-76 and α-79 (Brown *et al.*, 1993; Stern *et al.*, 1994; Paliakasis *et al.*, 1996). In sheep the fifth pocket of the DQA_2 alleles was largely found to have a hydrophobic interior, with flanking neutral or charged amino acid residues, and had similar characteristics to those observed in human DR and DQ molecules. As a result of the close similarities between the ovine DQA_2 alleles within these two pockets, all of the alleles would be expected to have similar peptide binding specificities.

The α-chain of the Antigen-binding Groove for Ovine DQA_2 Pseudogene Alleles C2 and E

The C2 and E alleles are not thought to be capable of producing a functional protein because they appear to lack a splice site at the end of exon 2 (Chapter 3). These sequences were included in this analysis to determine whether if an alternative splice site was used, could they produce a histocompatibility protein that had a similar structure and peptide specificity of the "true" DQA_2 alleles?

Outside the regions considered to be important for peptide backbone binding or pocket formation, the amino acid sequence for the C2 and E alleles was the least conserved compared to HLA-DR1 and DQ. This suggests that the overall structure of the groove could be different from the other alleles, and they may not be able to produce a protein that has the typical groove structure.

However, several amino acid residues in the regions that have been identified as being important for peptide backbone binding and pocket formation in HLA-DR and DQ have been conserved in the C2 allele. In addition, the C2 allele shares closest identity to the DQA_2 K allele within these regions. This suggests that if the C2 allele was expressed

and polymorphism outside the peptide binding and pocket areas produced a protein with the expected groove structure, then it would have similar peptide binding ability and specificity to that of the K allele.

The E allele, however, is quite different. The non-conservation of Arg-79 within the E allele may have a severe impact on the ability of the E allele to form a hydrogen bond with the peptide backbone, and also in forming a salt bridge with β-57.

Arg-79 is important for the formation of a salt bridge with β-57 underneath the bound peptide, linking the α1 and β1-chain helical regions and stabilising the two chains that form the groove (Stern *et al.*, 1994). The substitution of histidine at position 79 in the E allele may mean that the ability to form a salt bridge with the β-chain is altered. The side-chain of arginine is long, enabling close proximity to β-57, whilst histidine has a shorter side-chain, which could prevent it from being close enough to β-57 to form a salt bridge. In addition, histidine may not always be charged, depending on the local environmental conditions. Without charge it would not be able to form the salt bridge with β-57.

In studies of human DQ molecules loss of this salt bridge through polymorphism in the β-chain (residue 57), has correlated with increased susceptibility to IDDM (Todd *et al*, 1988; Horn *et al*, 1988), specifically non-conservation of aspartic acid at β-57. Similar studies in mice have found that the non-conservation of proline at position 56 and aspartic acid at position 57 in the DQβ chain also causes increased susceptibility to IDDM (Singer *et al.*, 1998), with histidine and serine being found at positions 56 and 57 in Non-obese diabetic (NOD) mice. It follows, that polymorphism in the α-chain, which could alter salt bridge formation, may have a similar affect on susceptibility to particular diseases. This has been observed in sheep whereby animals that have the E allele have an increased susceptibility to footrot, (Escayg *et al.*, 1997). If the E allele used an alternative splice site, such that it produced a protein, then the resulting protein may not be able to form such a stable heterodimer as the other alleles. This could possibly be a

result of the loss of the salt bridge between α-79 and β-57, causing instability in the groove, a lower peptide binding ability, and therefore increased disease susceptibility.

In addition, the fifth pocket in the E allele has slightly different characteristics, where isoleucine at α-75 is substituted by tyrosine, and the normally conserved arginine at position α-79 is replaced by histidine. This would make the fifth pocket of the E allele more reactive and hydrophillic than the other alleles. In addition pocket size may be smaller because of the ringed structure of tyrosine and histidine, preventing the binding of peptides with similarly large residues.

All of these observations imply that the E allele may not be able to produce a protein that has a typical groove structure, and may not have the same peptide binding specificity as other ovine DQA_2 alleles. This further supports the theory that the E allele is not able to produce a functional protein.

4.4 Conclusion

Outside the areas identified as being important for peptide binding and pocket formation human and ovine DR and DQ sequences are highly homologous, with large blocks of invariant amino acid residues, further supporting the notion that human and ovine class II molecules have a similar structure. However, polymorphic amino acid residues outside the modelled pockets and peptide binding residues were observed for all ovine DQA_2 alleles, particularly alleles C2 and E, which shared the least homology.

Whilst the C2 allele is likely to be from a DQA_2 pseudogene (Chapter 3), amino acid residues that are important for peptide binding and pocket formation are similar to B, D, F2, G2, H, K and Q alleles, indicating it could have a similar peptide binding specificity and peptide binding ability if it was expressed. Amino acid residues outside the pocket and peptide binding areas are however, quite different, and it is possible that this may impact on peptide binding specificity and on the structure of the groove itself.

The E allele is also different, and like C2 is thought to be a DQA_2 pseudogene sequence. If the E allele is expressed, the structure of the groove and peptide binding specificity may be altered as a result of amino acid variation outside the designated pockets and peptide binding areas. In addition, the ability to form a salt bridge with the β-chain may be altered, compromising groove stability, such that if a protein was produced by the E allele, peptide binding could be markedly altered. Pocket five also has different characteristics than the other ovine DQA_2 alleles suggesting that the E allele has different peptide binding specificity from the other alleles.

It is clear from the above study that interpreting potential differences in peptide binding and peptide binding specificity is difficult, and is somewhat arbitrary. While most of the results are purely speculative, some interesting information can still be gleaned from such analysis. As techniques improve however, it is hoped that it will be possible to obtain ovine class II glycoproteins of sufficient quantity and purity, such that the actual structure and the interactions between amino acid residues in the antigen binding groove and the antigenic peptide can be observed.

CHAPTER FIVE

GENERAL SUMMARY AND FUTURE DIRECTIONS.

This thesis was broadly divided into two parts. Firstly, whether associations existed between variation at the ovine MHC DQA_2 locus and susceptibility or resistance to footrot was investigated. This involved sheep from a number of different lines from four flocks, and three breeds. Secondly, the DQA_2 gene was further characterised by cloning and sequencing a number of alleles, which subsequently allowed the structure of the ovine DQA_2 antigen-binding groove to be predicted. This enabled the antigen-peptide binding ability of different alleles to be compared.

The results presented in this thesis showed that some significant associations existed between particular ovine MHC DQA_2 alleles and footrot resistance or susceptibility (Chapter 2). At the molecular level similarities and differences were observed between ovine MHC DQA_2 allelic sequences (Chapter 3). This inturn revealed that some DQA_2 products may be better at binding particular antigenic peptides than others (Chapter 4). These subsequent chapters provided a rational genetic basis for the observed associations between the ovine MHC DQA_2 alleles and footrot resistance or susceptibility, and suggested that this type of approach was a reasonable line of inquiry.

Twelve of the 15 previously defined DQA_2 alleles could be ranked in terms of their relationship to footrot susceptibility or resistance. Alleles M, O, and P did not occur in any of the flocks analysed, so they could not be ranked. Allele I appeared to be most strongly associated with susceptibility, while the G allele appeared to be the most strongly associated with resistance. In some of the lines analysed, particularly Corriedale flock 3, the prevalence of footrot was low, suggesting the challenge may have been inadequate to differentiate between susceptible and resistant animals. A low prevalence of footrot, combined with a low frequency of some of the DQA_2 alleles in various lines makes the statistical analysis unreliable. This may make the ranking of some of the alleles incorrect.

In future studies the footrot challenge could be improved, so that it is severe enough to adequately differentiate between footrot susceptibility and resistance. The major limitation to the effectiveness of the challenge, particularly within Corriedale flock 3, was insufficient rainfall. This problem could be addressed by conducting future challenges on paddocks, which allow artificial irrigation. In addition, increasing the duration of the trial may also counteract the level of variable exposure, making the challenge more uniform.

The ranking of the DQA_2 alleles needs to be confirmed, which could be achieved by performing further disease association studies in more flocks. Furthermore, if flocks could be identified that had a higher frequency of the rarer DQA_2 alleles a more accurate ranking could be achieved for these alleles. An ideal situation would be to have several flocks of different breeds exposed to the same challenge. In this way all of the animals would have the same level of exposure to *D. nodosus*, and the relative susceptibility of different breeds could be assessed.

Although RFLP and Southern hybridisation analysis has been shown to be a useful technique for the identification of alleles at the ovine DQA_2 locus, it is a time consuming, and expensive technique. Research within our group has developed a new typing system using PCR-SSCP. Not only is this technique less expensive and time consuming, preliminary analysis using this technique has identified additional polymorphism at the DQA_2 locus, with the subdivision of the C, F and J alleles. Further investigation is required to determine associations between these alleles and footrot susceptibility or resistance. The animals that were used in this investigation could be retyped using PCR-SSCP, which may allow the new alleles to be ranked. This ranking could also be confirmed by further disease association studies in other flocks and breeds.

Additional variation at the DQA_2 locus was detected by cloning and sequencing a number of DQA_2 alleles. However, further sequencing is required, so there is data available for all of the DQA_2 alleles, including the new alleles that have been identified

through PCR-SSCP. In addition, obtaining sequences from animals of different breeds would help to resolve whether there are any breed-specific DQA_2 alleles.

Future efforts should also be directed towards resolving the organisation of the ovine DQ subregion. Evidence suggests that a putative DQA_3 gene exists, and it shares close sequence homology to DQA_2. This gene appears to associate with a DQA_1-null haplotype, and preliminary data suggests that it may be a pseudogene because of the lack of a splice site at the end of exon 2. Further sequence analysis should readily determine whether the putative DQA_3 gene is a pseudogene, and is associated exclusively with the DQA_1-null haplotype. If sequences were found that were consistent with an expressed DQA_3 gene then Northern blot analysis, or a Ribonuclease protection assay (RPA), using DQA_3 gene sequences as a probe, would determine whether this locus is transcribed. The map of the ovine DQ subregion, constructed from overlapping cosmid clones, linking DQA_1, DQB_1, DQA_2, and DQB_2 (Wright and Ballingall, 1994) should be extended to allow the placement of the putative DQA_3 gene, if it is confirmed that this gene exists.

Sequence analysis also suggested that alleles E and C2 may not be expressed. If this is the case the only functional product would be DQA_1 encoded. Northern blot analysis, or RPA could confirm whether these DQA_2 alleles are transcribed. It would also be of benefit to type animals at the DQA_1 locus to determine if there is any association between the DQA_1 haplotype and footrot resistance or susceptibility, in those animals which are not thought to be able to produce a functional DQA_2 product.

The search for other markers of footrot resistance or susceptibility within the ovine MHC class II region should also be continued. It is possible other markers may be identified in different flocks or breeds that may prove to be more closely linked to footrot resistance than the DQA_2 locus.

REFERENCES.

Adams, L.G. & Templeton, J.W. (1998). Genetic resistance to bacterial diseases of animals. *World organisation for animal health scientific and technical review*, **17 (1),** 200-219.

Amills, M., Ramiya, V., Norimine, J. & Lewin, H.A. (1998). The major histocompatibility complex of ruminants. *World organisation for animal health scientific and technical review*, **17 (1),** 108-120.

Andersson, L., Gustafsson, K., Jonsson, A.K. & Rask, L. (1991). Concerted evolution in a segment of the first domain exon of polymorphic MHC class II β loci. *Immunogenetics*, **33,** 235-242.

Bailey, D.W. & Kohn, H.I. (1965). Inherited histocompatibility changes in progeny of irradiated and unirradiated inbred mice. *Genetic Research,* **6(3),** 330-340.

Ballingall, K.T., Luyai, A. & McKeever, D.J. (1997). Analysis of genetic diversity at the DQA loci in African cattle: evidence for a BoLA-DQA3 locus. *Immunogenetics,* **46,** 237-244.

Ballingall, K.T., Marasa, B.S., Luyai, A. & McKeever, D.J. (1998).Identification of diverse BoLA DQA3 genes consistent with non-allelic sequences. *Animal Genetics,* **29 (2),** 123-129.

Bjorkman, P.J., Saper, M.A., Samraoui, B., Bennett, W.S., Strominger, J.L. & Wiley, D.C. (1987). Structure of the human class I histocompatibility antigen, HLA-A2. *Nature,* **329,** 506-512.

Bodmer, W.F. (1972). Evolutionary significance of the HL-A system. *Nature*, **237,** 139-145.

Boundy, T. (1983). Footrot and foot conditions. In: *Diseases of sheep.* W.B. Martin (ed). Oxford: Blackwell Scientific Publications, 98-103.

Bradley, R.D. & Hillis, D.M. (1997). Recombinant DNA sequences generated by PCR amplification. *Molecular Biology Evolution,* **14 (5),** 592-593.

Brown, J.H., Jardetzky, T.S., Gorga, J.C., Stern, L.J., Urban, R.G., Strominger, J.L. & Wiley, D.C. (1993). Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. *Nature* **364,** 33-39.

Chandra, R.K. (1989). Nutritional regulation of immunity and risks of infection in old age. *Immunology,* **67,** 141-147.

Chicz, R.M. & Urban, R.G. (1994). Analysis of MHC-presented peptides: applications in autoimmunity and vaccine development. *Immunology Today,* **15 (4),** 155-159.

Clayton, J. & Gee, H. (1993). The evolutionary angle. *Nature,* **365,** 111-112.

Deverson, E.V., Wright, H. Watson, S., Ballingall, K., Huskisson, N., Diamond, A.G. & Howard, J.C. (1991). Class II major histocompatibility complex genes of the sheep. *Animal Genetics,* **22,** 211-225.

Doherty, P.C. & Zinkernagel, R.M. (1975). A biological role for the major histocompatibility antigens. *Lancet,* **1 (7922),** 1406-9.

Drinkwater, R.D. & Hetzel D.J.S. (1990). Developing DNA markers for use in animal breeding. *Proceedings of the Australian Society of Animal Breeding and Genetics,* 195-200.

Durinovicbello, I. (1998). Autoimmune diabetes-The role of T cells, MHC molecules and autoantigens (review). *Autoimmunity,* **27(3),** 159-177.

Dutia, B.M., McConnell, I., Ballingall, K.T., Keating, P. & Hopkins, J. (1994). Evidence for the expression of two distinct MHC class II DRβ like molecules in sheep. *Animal Genetics,* **25,** 235-241.

Edwards, S.V. & Hendrick, P.W. (1998). Evolution and ecology of MHC molecules: from genomics to sexual selection. [Review] [53 refs] *Tree,* **13 (8),** 305-311.

Egerton, J.R. & Raadsma, H.W. (1989). Breeding sheep resistance to footrot. *Footrot and foot abscess of ruminants.* J.R. Egerton, W.K. Young & G.G. Riffkin (eds). CRC Press Inc Florida.

Egerton, J.R., Cox, P.T., Anderson, B.J., Kristo, C., Norman, M. & Mattick, J.S. (1987). Protection of sheep against footrot with a recombinant DNA-based fimbrial vaccine. *Veterinary Microbiology,* **14,** 393-409.

Egerton, J.R., Ribeiro, L.A., Kieran, P.J. & Thorley, C.M. (1983). Onset and remission of ovine footrot. *Australian Veterinary Journal,* **60 (11),** 334-336.

Egerton, J.R., Roberts, D.S. & Parsonson, I.M. (1969). The aetiology and pathogenesis of ovine foot-rot. *Journal of Comparative Pathology,* **79,** 207-217.

Emery, D.L., Stewart, D.J. & Clarke, B.L. (1984). The comparative susceptibility of five breeds of sheep to footrot. *Australian Veterinary Journal,* **61 (3),** 85-88.

Escayg, A.P. (1995). The ovine MHC class II region: Polymorphism, organisation, and association with footrot. PhD thesis, Lincoln University, Canterbury, New Zealand.

Escayg, A.P., Hickford, J.G.H, Montgomery, G.W., Dodds, K.G. & Bullock, D.W. (1996). Polymorphism at the ovine major histocompatibility complex class II loci. *Animal Genetics,* **27,** 305-312.

Escayg, A.P., Hickford, J.G.H. & Bullock, D.W. (1997). Association between alleles of the ovine major histocompatibility complex and resistance to footrot. *Research in Veterinary Science,* **63,** 283-287.

Fabb, S.A., Maddox, J.F., Gogolin-Ewens, K.J., Baker, L., Wu, M.J. & Brandon, M.R. (1993). Isolation, characterization and evolution of ovine major histocompatibility complex class II DRA and DQA genes. *Animal Genetics,* **24,** 249-255.

Fan, W., Liu, Y.C., Parimoo, S. & Weissman, S.M. (1995). Olfactory receptor-like genes are located in the human major histocompatibility complex. *Genomics*, **27,** 119-123.

Flajnik, M.F., Canel, C., Kramer, J. & Kasahara, M. (1991). Which came first, MHC class I or class II. *Immunogenetics* **33,** 295-309.

Garber, T.L., Hughes, A.L., Letvin, N.L., Templeton, J.W. & Watkins, D.I. (1993). Sequence and evolution of cattle MHC class I cDNAs: concerted evolution has not taken place in cattle. *Immunogenetics,* **38,** 11-20.

Gelhaus, A., Forster, B., Wippern, C. & Horstmann, R.D. (1999). Evidence for an additional cattle DQA locus, BoLA-DQA5. *Immunogenetics,* **49 (4),** 321-327.

George, A.J.T., Ritter, M.A. & Lechler, R.I. (1995). Disease susceptibility, transplantation and the MHC. *Immunology Today,* **16 (5),** 209-211.

Germain, R.N. (1994). MHC-dependent antigen processing and peptide presentation: providing ligands for T lymphocyte activation. *Cell*, **76,** 287-298.

Ghimire, S.C. & Egerton, J.R. (1996). Transmission of footrot in migratory sheep and goats of Nepal. *Small Ruminant Research,* **22,** 231-240.

Grain, F., Nain, M.C., Labonne, M.P., Lantier, F., Lechopier, P., Gebuhrer, L., Asso, A., Maddox, J. & Betuel, H. (1993). Restriction fragment length polymorphism of DQB and DRB class II genes of the ovine major histocompatibility complex. *Animal Genetics,* **24,** 377-384.

Hashimoto, K., Nakanishi, T. & Kurosawa, Y. (1990). Isolation of carp genes encoding major histocompatibility complex antigens. *Proceedings of the National Academy of Sciences of the United States of America,* **87 (17),** 6863-6867.

Hawken, R.J., Beattie, C.W. & Schook, L.B. (1998). Resolving the genetics of resistance to infectious disease. *World organisation for animal health scientific and technical review,* **17 (1),** 17-25.

Hickford, J.G.H., Ridgway, H.J. & Escayg, A.P. (unpublished). Evolution of the ovine MHC DQA region.

Hill, A.V.S., Allsopp, C.E.M., Kwiatkowski, D., Anstey, N.M., Twumasi, P., Rowe, P.A., Bennett, S., Brewster, D., McMichael, A.J. & Greenwood, B.M. (1991). Common West African HLA antigens are associated with protection from severe malaria. *Nature,* **352,** 595-600.

Högstrand, K. & Böhme, J. (1998). Intrachromosomal gene conversion frequency in the H2 differs between haplotypes. *Immunogenetics,* **48,** 47-55.

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.

Hood, L., Kronenberg, M. & Hunkapiller, T. (1985). T cell antigen receptors and the immunoglobulin supergene family. *Cell,* **40(2),** 225-229.

Horin, P., (1998). Biological principles of heredity and resistance to disease. *World organisation for animal health scientific and technical review,* **17 (1),** 302-314.

Horn, G.T., Bugawan, T.L., Long, C.M. & Erlich, H.A. (1988). Allelic sequence variation of the HLA-DQ loci: Relationship to serology and to insulin-dependent diabetes susceptibility. *Proceedings of the National Academy of Science U.S.A.,* **85,** 6012-6016.

Hughes, A.L. & Hughes, M.K. (1995). Natural selection on the peptide-binding regions of major histocompatibility complex molecules. *Immunogenetics,* **42,** 233-243.

Hughes, A.L. & Nei, M. (1989). Nucleotide substitutions at major histocompatibility complex class II loci: evidence for overdominant selection. *Proceedings of the National Academy of Science U.S.A*, **86,** 958.

Hughes, A.L. & Nei, M. (1991). Polymorphism and the evolution of the major histocompatibility complex loci in mammals. In: *Evolution at the Molecular Level.* R.K. Selander, A.G. Clark & T.S. Whittan (eds). Sinauer Associates Inc Publishers.

Hughes, A.L. & Nei, M. (1993). Evolutionary relationships of the classes of major histocompatibility complex genes. *Immunogenetics,* **37 (5),** 337-346.

Hughes, A.L. & Yeager, M. (1997). Molecular evolution of the vertebrate immune system [Review] [71 refs] *BioEssays* **19 (9),** 777-786.

Hughes, A.L., Yeager, M. & Carrington, M. (1996). Peptide binding function and the paradox of HLA disease associations. *Immunology and Cell Biology,* **74 (5),** 444-448.

Hulme, D.J., Nicholas, F.W., Windon, R.G., Brown, S.C. & Beh, K.J. (1993). The MHC class II region and resistance to an intestinal parasite in sheep. *Journal of Animal Breeding and Genetics,* **110,** 459-472.

Jaulin, C., Perrin, A., Abastado, J.P., Dumas, B., Papamatheakis, J. & Kourilsky, P. (1985). Polymorphism in mouse and human class I H-2 and HLA genes is not the result of independent point mutations. *Immunogenetics,* **22,** 453-470.

Kaufman, J. & Wallny, H.J. (1996). Chicken MHC molecules, disease resistance and the evolutionary origin of birds. *Current Topics in Microbiological Immunology,* **212,** 129-141.

Kaufman, J. (1988). Vertebrates and the evolution of the major histocompatibility complex (MHC) class I and II molecules. *Verhandlung Deutschen Zoologischen Gesellschaft*, **81,** 131-144.

Kaufman, J.F., Auffray, C., Korman, A.J., Shackelford, D.A. & Strominger, J. (1984). The class II molecules of the human and murine major histocompatibility complex. *Cell*, **36,** 1-13.

Kimberling, C.V. (1988). In: *Jensens and Swift's diseases of Sheep. 3rd ed.* Philadelphia: Lea and Febiger, 317-321.

Kingwell, R.S., Abadi Ghadim, A.K. & Robinson, S.D. (1995). Introducing Awassi sheep to Australia: an application of farming system models. *Agricultural Systems,* **47,** 451-471.

Klein, J. & Figueroa, F. (1986). Evolution of the major histocompatibility complex genes. *Critical Reviews in Immunology,* **6,** 295-386.

Klein, J. & Sato, A. (1998). Birth of the major histocompatibility complex. *Scandinavian Journal of Immunology,* **47 (3),** 199-209.

Klein, J., Sato, A. & O'hUigin, C. (1998). Evolution by gene duplication in the major histocompatibility complex. *Cytogenetics and Cell Genetics,* **80,** 123-127.

Lawrance, S. & Quaranta, V. (1991). Molecular bases of MHC diversity. In: *Immunogenetics of the Major Histocompatibility Complex.* Srivastava, R., Ram, B.P., Tyle, P. (eds). VCH Publishers Inc, New York.

Leslie, S.E. (1994). Genetic analysis of the ovine major histocompatibility complex and association with footrot. Master of Applied Science Thesis, Lincoln University, Canterbury, New Zealand.

Lewis, R.D., Meyer, H.H., Gradin, J.L. & Smith, A.W. (1989). Effectiveness of vaccination in controlling ovine footrot. *Journal of Animal Science,* **67,** 1160-1166.

Lightfoot, J. (1988). Awassi fat tail sheep. *Journal of Agriculture-Western Australia,* **28 (4),** 107-113.

Litchfield, A.M., Raadsma, H.W., Hulme, D.J., Brown, S.C., Nicholas, F.W. & Egerton, J.R. (1993). Disease resistance in Merino sheep.II. RFLPs in class II MHC and their association with resistance to footrot. *Journal of Animal Breeding and Genetics,* **110,** 321-334.

Madden, D.R., Gorga, J.C., Strominger, J.L. & Wiley, D.C. (1991). The structure of HLA-B27 reveals nonamer self-peptides bound in an extended conformation. *Nature,* **353,** 321-325.

Milner, C.M. & Campbell, R.D. (1992). Genes, genes and more genes in the human major histocompatibility complex. *BioEssays,* **14 (8),** 565-570.

Montgomery, G.W. & Sise, S.A. (1990). Extraction of DNA from sheep white blood cells. *New Zealand Journal of Agricultural Research,* **33,** 437-441.

Nei, M., Gu, X. & Sitnikova, T. (1997). Evolution by the birth-and-death process in multigene families of the vertebrate immune system. *Proceedings of the National Academy of Science USA,* **94,** 7799-7806.

Nicholas, F.W. (1991). Genetic markers: potential as selection criteria. In: *Breeding for disease resistance in sheep*, G.D. Gray and R.R. Woolaston (eds). Australian Wool Coporation, Melbourne.

Ober, C. (1995). Current topic: HLA and reproduction: lessons from studies in the Hutterites. *Placenta,* **16,** 569-577.

Ober, C., Hyslop, T., Elias, S., Weitkamp, L.R. & Hauck, W.W. (1998). Human leukocyte antigen matching and fetal loss: results of a 10 year prospective study. *Human Reproduction,* **13 (1),** 33-38.

Ober, C., Weitkamp, L.R., Cox, N. Dytch, H., Kostyu, D. & Elias, S. (1997). HLA and mate choice in humans. *American Journal of Human Genetics*, **61,** 497-504.

Odelberg, .J., Weiss, R.B., Hata, A. & White, R. (1995). Template-switching during DNA synthesis by *Thermus aquaticus* DNA polymerase 1. *Nucleic Acids Research,* **23 (11),** 2049-2057.

Outteridge, P.M. (1991). Ovine lymphocyte antigens (OLA) and resistance to parasites. In: *Breeding for Disease Resistance in Sheep.* G.D Gray & R.R. Woolaston (eds). Australian Wool Corporation, Melbourne, 103-108.

Outteridge, P.M., Stewart, D.J., Skerman, T.M., Dufty, J.H., Egerton, J.R., Ferrier, G. & Marshall, D.J. (1989). A positive association between resistance to ovine footrot and particular lymphocyte antigen types. *Australian Veterinary Journal,* **66 (6),** 175-179.

Paliakasis, K., Routsias, J., Petratos, K., Ouzounis, C. Kokkinidis, M. & Papadopoulos, G.K. (1996). Novel structural features of the human histocompatibility molecules HLA-DQ as revealed by modelling based on the published structure of the related molecule HLA-DR[1]. *Journal of Structural Biology* **117,** 145-163.

Paterson, S. (1998). Evidence for balancing selection at the major histocompatibility complex in a free-living ruminant. *The Journal of Heredity*, **89 (4),** 289-294.

Patterson, R. & Patterson, H. (1991). The selection and breeding of Merino sheep for footrot resistance. *Proceedings of the New Zealand Society of Animal Production,* **51,** 283-286.

Penn, D. & Potts, W. (1998). How do major histocompatibility complex genes influence odour and mating preferences? [Review] [107 refs] *Advances in Immunology,* **69,** 411-436.

Potts, W.K. & Wakeland, E.K. (1993). Evolution of MHC genetic diversity: a tale of incest, pestilence and sexual preference. *Trends in Genetics,* **9,** 408-412.

Potts, W.K., Manning, C.J. & Wakeland, E.K. (1991). Mating patterns in seminatural populations of mice influenced by MHC genotype. *Nature,* **352,** 619-621.

Potts, W.K., Manning, C.J. & Wakeland, E.K. (1994). The role of infectious disease, inbreeding and mating preferences in maintaining MHC genetic diversity: an experimental test. *Philosophical Transactions of the Royal Society of London Series B,* **346,** 369-378.

Powis, S.H. & Trowsdale, J. (1991). HLA and disease. [Review] [30 refs] *British Journal of Clinical Practice,* **45 (2),** 116-120.

Raadsma, H.W. & Egerton, J.R. (1993). Current research on an alternative footrot control strategy. *Wool Technology and Sheep Breeding,* **41,** 21-28.

Raadsma, H.W., Egerton, J.R., Outteridge, P.M., Nicolas, F.W., Brown, S.C. & Litchfield, A.M. (1990). An investigation into genetic aspects of resistance to footrot in Merino sheep. *Wool Technology and Sheep Breeding,* **38,** 7-12.

Riley, E. & Olerup, O. (1992). HLA polymorphisms and evolution. *Immunology Today,* **13 (9),** 333-335.

Rood, J.I. & Yong, W.K. (1989). Application of biotechnology to the diagnosis of footrot. In: *Footrot and foot abscess of ruminants.* J.R. Egerton, W.K. Yong & G.G. Riffkin (eds). CRC Press Florida, 235-248.

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 (7),** 815-822.

Scott, P.C., Choi, C.L. & Brandon, M.R. (1987). Genetic organization of the ovine MHC class II region. *Immunogenetics,* **25,** 116-122.

Scott, P.C., Gogolin-Ewens, K.J, Adams, T.E. & Brandon, M.R. (1991a). Nucleotide sequence, polymorphism, and evolution of the ovine MHC class II DQA genes. *Immunogenetics,* **34,** 69-79.

Scott, P.C., Maddox, J.F., Gogolin-Ewens, K.J. & Brandon, M.R. (1991b). The nucleotide sequence and evolution of ovine MHC class II B genes: DQB and DRB. *Immunogenetics,* **34,** 80-87.

Seidl, C., Koch, U., Brunnler, G., Buhleier, T., Frank, R., Moller, B., Markert, E., Kollerwagner, G., Seifried, E. & Kaltwasser, J.P. (1997). HLA-DR/DQ/DP interactions in rheumatoid arthritis. *European Journal of Immunogenetics,* **24 (5),** 365-376.

Shimshony, A. (1988). Footrot in Awassi and their crosses with East Friesian sheep. *New Zealand Veterinary Journal,* **37,** 44.

Singer, S.M., Tisch, R., Yang, X.D., Sytwu, H.K., Liblau, R. & McDevitt, H.O. (1998). Prevention of diabetes in NOD mice by a mutated I-AB transgene. *Diabetes,* **47 (10),** 1570-1577.

Singh, N., Agrawal, S. & Rastogi, A.K. (1997). Infectious diseases and immunity: special reference to major histocompatibility complex. *Emerging Infectious Diseases,* **3 (1),** 41-49.

Skerman, T.M. & Moorhouse, S.R. (1987). Broomfield Corriedales: a strain of sheep selectively bred for resistance to footrot. *New Zealand Veterinary Journal,* **35,** 101-106.

Skerman, T.M. (1986). Genetic variation and inheritance of susceptibility to footrot in sheep. In: *Footrot in Ruminants-Proceedings of a Workshop Melbourne 1985.* D.J. Stewart, J.E. Peterson, N.M. McKern & D.L. Emery (eds). CSIRO Division of Animal Health, Australia.

Skerman, T.M., Erasmuson, S.K. & Morrison, L.M. (1982). Duration of resistance to experimental footrot infection in Romney and Merino sheep vaccinated with *Bacteroides nodosus* oil adjuvant vaccine. *New Zealand Veterinary Journal,* **30,** 27-31.

Skerman, T.M., Johnson, D.L., Kane, D.W. & Clarke, J.N. (1988). Clinical footscald and footrot in a New Zealand Romney flock: phenotypic and genetic parameters. *Australian Journal of Agricultural Research,* **39,** 907-16.

Snibson, K.J., Maddox, J.F., Fabb, S.A. & Brandon, M.R. (1998). Allelic variation of ovine MHC class II DQA1 and DQA2 genes. *Animal Genetics,* **29,** 356-362.

Soller, M. & Andersson, L. (1998). Genomic approaches to the improvement of disease resistance in farm animals. *World organisation for Animal Health Scientific and Technical Review*, **17 (1),** 329-345.

Southern, E.M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *Journal of Molecular Biology,* **98,** 503-517.

Srivastava, R., Ram B.P., & Tyle, P. (1991). Immunogenetics of the major histocompatibility complex: an introduction. In: *Immunogenetics of the Major Histocompatibility Complex.* Srivastava, R., Ram, B.P., Tyle, P. (eds). VCH Publishers Inc, New York.

Stern, L.J., Brown, J.H., Jardetzky, T.S., Gorga, J.C., Urban, R.G., Strominger, J.L. & Wiley, D.C. (1994). Crystal structure of the human class II MHC protein HLA-DR1 complexed with an influenza virus peptide. *Nature* **368,** 215-221.

Stewart, D.J. (1989). Footrot of sheep. In: *Footrot and foot abscess of ruminants.* J.R. Egerton, W.K. Yong & G.G. Riffkin (eds). CRC Press, Florida, 5-45.

Stewart, D.J., Clark, B.L., Peterson, J.E., Griffiths, D.A. & Smith, E.F. (1982). Importance of pilus-associated antigen in *Bacteroides nodosus* vaccines. *Research in Veterinary Science,* **32,** 140-147.

Stewart, D.J., Emery, D.L., Clark, B.L., Peterson, J.E., Iyer, H. & Jarrett, R.G. (1985). Differences between breeds of sheep in their responses to *Bacteroides nodosus* vaccines. *Australian Veterinary Journal,* **62 (4),** 116-120.

Stewart, D.J., Peterson, J.E., Vaughan, J.A., Clark, B.L., Emery, D.L., Caldwell, J.B. & Kortt A.A. (1986). The pathogenicity and cultural characteristics of virulent, intermediate and benign strains of *Bacteroides nodosus* causing ovine footrot. *Australia Veterinary Journal,* **63 (10),** 317-326.

Thomas, J.H. (1962). The bacteriology and histopathy of footrot in sheep. *Australian Journal of Agricultural Research,* **13,** 725-732.

Todd, J.A, Acha-Orbea, H., Bell, J.I., Chao, N., Fronek, Z., Jacob, C.O., McDermott, M., Sinha, A.A, Timmerman, L., Steinman, L & McDevitt, H.O. (1988). A molecular basis for MHC class II-associated autoimmunity. *Science,* **240 (4855),** 1003-1009.

Toussirot, E., Auge, B., Tiberghien, P., Chabod, J., Cedoz, J.P. & Wendling, D. (1999). HLA-DRB1 alleles and shared amino acid sequences in disease susceptibility and severity in patients from Eastern France with rheumatoid arthritis. *Journal of Rheumatology,* **26 (7),** 1446-1451.

Trabesinger-Reuf, N., Jermann, T., Zankel, T., Durrant, B., Frank, G. & Benner, S.A. (1996). Pseudogenes in ribonuclease evolution: a source of new biomacromolecular function? *Federation of European Biochemical Societies-Letters,* **382,** 319-322.

Trowsdale, J. (1987). Genetics and polymorphism: class II antigens. *British Medical Bulletin,* **43 (1),** 15-36.

Trowsdale, J. (1988). The human MHC: an overview. *Animal Genetics,* **19,** 5-9.

Trowsdale, J. (1995). "Both man & bird & beast": comparative organisation of MHC genes. [Review] [156 refs] *Immunogenetics,* **41,** 1-17.

van Oorschot, R.A.H., Maddox, J.F., Adams, L.J. & Fabb, S.A. (1994). Characterization and evolution of ovine MHC class II DQB sequence polymorphism. *Animal Genetics,* **25,** 417-424.

Wedekind, C., Seebeck, T., Bettens, F. & Paepke, A.J. (1995). MHC-dependent mate preferences in humans. *Proceedings of the Royal Society of London-Series B,* **260,** 245-249.

Weenink, S.M. & Gautam, A.M. (1997). Antigen presentation by MHC class II molecules. *Immunology and Cell Biology,* **75,** 69-81.

Weiss, E.H., Mellor, A., Golden, L., Fahrner, K., Simpson, E., Hurst, J. & Flavel, R.A. (1983). The structure of a mutant H-2 gene suggests that the generation of polymorphism in H-2 genes may occur by gene conversion-like events. *Nature,* **301,** 671-674.

Wheeler, C.J., Maloney, D., Fogel, S. & Goodenow, R.S. (1990). Microconversion between murine H-2 genes integrated into yeast. *Nature,* **347,** 192-194.

Whittington, R.J & Nicolls, P.J. (1995). Effects of the severity and duration of lesions on the primary and anamnestic humoral responses of sheep to *Dichelobacter nodosus* and observations of natural resistance to footrot. *Research in Veterinary Science,* **59,** 128-135.

Wills, C. (1991). Maintenance of multiallelic polymorphism at the MHC region. [Review] [101 refs] *Immunological Reviews*, **124,** 165-220.

Wright, H. & Ballingall, K.T. (1994). Mapping and characterization of the DQ subregion of the ovine MHC. *Animal Genetics,* **25,** 243-249.

Zaphiropoulos, P.G. (1998). Non-homologous recombination mediated by *Thermus aquaticus* DNA polymerase 1. Evidence supporting a copy choice mechanism. *Nucleic Acids Research,* **26 (12),** 2843-2848.

Zhou, H. & Hickford, J.G.H. (2000). Extensive diversity in New Zealand *Dichelobacter nodosus* strains from infected sheep and goats. *Veterinary Microbiology,* **71,** 113-123.

Zinkernagel, R.M. & Doherty, P.C. (1997). The discovery of MHC restriction. *Immunology Today,* **18(1),** 14-17.

APPENDICES.

Appendix A: Major Suppliers of Reagents, Enzymes, and other Consumables

Amersham International, Buckinghamshire, England.

Becton Dickinson Vaccutainer Systems Europe, England.

Biolab Scientific Ltd, Christchurch, New Zealand.

Bio-Rad Laboratories Pty Ltd., Richmond, California, U.S.A.

Boehringer Mannheim/Roche Diagnostics, Mannheim Germany.

FMC Bioproducts, Rockland, ME, U.S.A.

Gibco/BRL, Gaithersburg, MD, U.S.A.

Hoefer Scientific Instruments, San Francisco, CA, U.S.A.

Invitrogen, San Diego, U.S.A.

Kodak Company, Rochester, NY, U.S.A.

Life Technologies, NY U.S.A.

New England Nuclear, Dupont NEN, Boston, MA, U.S.A

Perkin Elmer Corporation, Branchburg, New Jersey, U.S.A.

Pharmacia Biotech, NJ, U.S.A

Promega Corporation, Madison, WI, U.S.A

Qiagen, GmbH, Hilden, Germany.

Sigma Chemical Co, St Louis, MO, U.S.A.

All other chemicals unless otherwise stated were of AnalaR® grade and obtained from **British Drug Houses Chemicals,** Poole, Dorset, England.

Appendix B: Commonly Prepared Solutions

20X SSC (for 1 litre)

NaCl	175.3 g
Sodium citrate	88.2 g

Adjust to pH 7.0. Sterilise by autoclaving.

20X SSPE (for 1 litre)

NaCl	175.3 g
$NaH_2PO_4.H_2O$	27.6 g
EDTA	7.4 g

Adjust to pH 7.4. Sterilise by autoclaving.

10% Sodium Dodecyl Sulphate (SDS) (for 1 litre)

SDS	100 g

Heat to 68°C to assist dissolution.

10X TBE (for 1 litre)

Tris base	108 g
Boric acid	55 g
$EDTANa_2$	9.3 g

Sterilise by autoclaving.

50X Denhardts Reagent (for 100 mL)

Ficoll	1 g
Polyvinylpyrrolidone	1 g
BSA	1 g

Tris-Buffered Saline (TBS) (for 1 litre)

NaCl	8.0 g
KCl	0.38 g
Tris base	3.0 g
Phenol red	0.015 g

Adjust pH to 7.4. Sterilise by autoclaving.

Red Blood Cell Lysing Solution (for 1 litre)

NH_4Cl	8.02 g
$KHCO_3$	1.0 g
0.5M EDTA	200 µL

Sterilise by autoclaving.

TE Buffer (for 1 litre)

1M Tris (pH 8.0)	10.0 mL
0.5M EDTA (pH 8.0)	200 µL

Sterilise by autoclaving.

Terrific Broth (for 1 litre)

To 900 mL of deionised water add:

bacto-tryptone	12 g
bacto-yeast extract	24 g
glycerol	4 mL

Sterilise by autoclaving then add 100 mL of a sterile solution of salts consisting of:

KH_2PO_4	2.31 g
K_2HPO_4	12.54 g

Make up to a volume of 100 mL using deionised water and sterilise by autoclaving.

Appendix C: DNA Extraction from Sheep Blood

White blood cells (WBC) were collected from whole blood, following the lysis of the red blood cells (150 mM NH_4Cl, 10 mM $KHCO_3$, 0.1 mM Na_2EDTA (BDH)), and centrifugation at 3000 rpm/10 minutes/4°C. WBC pellets were washed twice in Tris-buffered saline (140 mM NaCl, 0.5 mM KCl, 25 mM Tris), pH 7.0 (BDH), centrifuged (1000 rpm/10 minutes/4°C), then resuspended in TE, (10 mM Tris-HCl pH 8.0, 0.1 mM Na_2EDTA), with subsequent lysis and digestion with proteinase K (10 g/mL) (Boehringer Mannheim/Roche Diagnostics), 0.5% sodium dodecyl sulphate (SDS) (BDH), and 25 mM Na_2EDTA, at 50°C for 1½ hours.

Cellular debris was precipitated with the addition of a saturated salt solution (5 M NaCl (BDH)), centrifuged at 2500 rpm/10 minutes/4°C. Two volumes of absolute ethanol was added to the supernatant, the DNA spooled, washed in 70% ethanol at 4°C overnight. The DNA was air dried and resuspended in 250-500 µL sterile dH_2O.

DNA purity was ascertained by the measurement of A260/A280 ratios, (expected to be 1.5-1.8), with typical yields of 0.5 mg - 1.0 mg DNA.

Appendix D: Southern Transfer of DNA to Nylon Membranes

1) The DNA was depurinated by inverting the gel and immersing for 20 minutes in 250 mM HCl with gentle agitation.
2) The HCl was discarded and the gel immersed in 0.4 mM NaOH for at least 20 minutes.
3) While the gel was soaking, 22 sheets of Whatman 3 MM filter paper were cut to the exact dimensions of the gel.
4) Positively charged nylon membrane (Hybond-N^{+}™ (Amersham) or Zeta-Probe® (BioRad)) was cut to the exact dimensions of the gel. The top left-hand corner of the membrane was removed for orientation purposes.
5) The membrane was soaked in sterile water with gentle agitation for 1 minute.
6) A blotting stack was set up as shown in Figure 7.1.
7) A 21 cm x 20 cm sponge was placed in a deep tray, approximately 5-10 cm. The tray was filled saturating the sponge with 0.4 mM NaOH. The level of 0.4 mM NaOH did not immerse the sponge. Four 3 MM papers pre-soaked with 0.4 mM NaOH was laid on the sponge such that they were perfectly aligned, and all air bubbles were excluded.
8) The gel still inverted was laid on top of the paper, then the membrane was placed on top of the gel, ensuring that it was aligned with the rest of the stack. Care was taken to remove all the air bubbles.
9) Four more papers, pre-soaked with sterile water were placed on op of the membrane, followed by 12 pieces of dry paper, and a stack of paper towels. No weight was placed on the transfer stack. The blot was left for 18-24 hours to allow complete transfer of the DNA.
10) After transfer was complete, the blot was dismantled, and the nylon membrane placed between two sheets of 3 MM filter paper.
11) The sandwiched membrane was baked at 80°C for at least 1 hour, prior to neutralising the membrane in a solution of 0.5 M Tris (pH 7.0), and 1 M NaCl for 15 minutes. The membrane was left to air dry between two sheets of filter paper before storage at 4°C in plastic bags.

12) The remains of the gel were soaked in 1x TBE containing eithidium bromide for at least 30 minutes, then visualised under U.V light to ascertain that DNA transfer was complete.

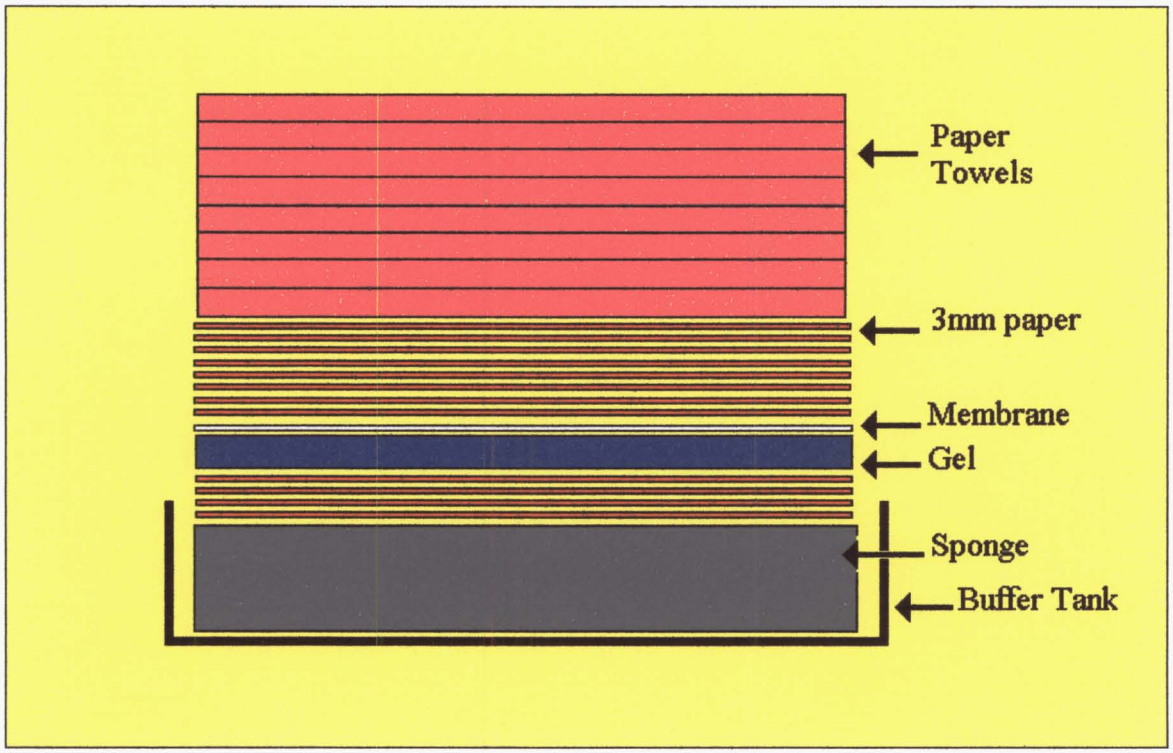


Figure A1: Capillary DNA blot (Southern blot) apparatus. DNA was transferred to nylon membranes by a modified version of the procedure originally proposed by Southern (Southern, 1975). A plastic tray between 5 and 10 cm deep had a high-density open-cell sponge placed in it. The tray was then filled with 0.4 mM NaOH, care being taken to completely saturate the sponge. Four sheets of Whatman 3mm paper pre-soaked in 0.4 mM NaOH were aligned on the sponge, and all air bubbles were removed. The gel was then aligned on paper such that the wells were face down. Pre-soaked membrane was next aligned on the gel such that no air or liquid was present at the interface. Four more sheets of paper pre-soaked in sterile water were applied to the stack, followed by 14 dry papers. Finally a stack of paper towels were added to the top of the stack. Transfers were generally left for 18-24 hours.

Appendix E: Plasmid Map for the Ovine DQA_2 Probe

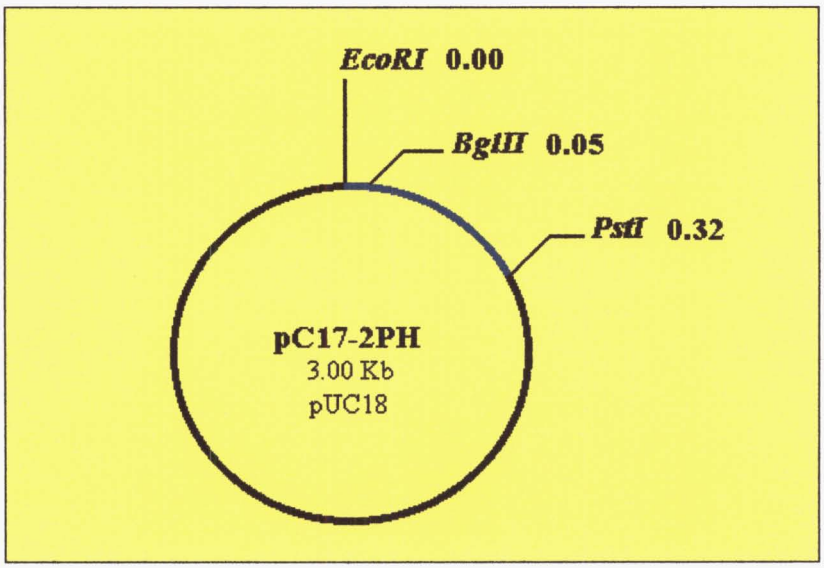


Figure A2: pC17-2PH plasmid map.

Plasmid name: pC17-2PH.

Plasmid size: 3.00 Kb.

Constructed By: Ming-Jie Wu.

Construction Date: 1991.

Comments/References: 280 bp *Pvu*II/*Hinf*I C17-2 fragment blunted prior to ligation into *Sma*I site (lost in cloning) of pUC18. GenBank accession # M33305 Scott *et al* (1991a) Immunogenetics 34:69-79.

(Map supplied by Dr J. Maddox, Centre for Animal Biotechnology, School of Veterinary Science, University of Melbourne, Parkville, Australia).

Appendix F: RFLP and Southern Hybridisation DQA_2 Typing Results for all of the Animals from Four New Zealand Farms

† **

1) Awassi NZ Ltd.											
Awassi Sires.		Progeny									
Tag Number	DQA2 Alleles		Tag Number	Sire	DQA2 Alleles	Tag Number	Sire	DQA2 Alleles	Tag number	Sire	DQA2 Alleles
ANZ 1032/91	K,K		76/96	1046	B,D	286/96	1046	??	89/96	1046	D,H
ANZ 1035/91	B,E		65/96	1046	B,E	13/96	1046	D,D	1/96'	1046	D,H
ANZ 1036/91	E,F		190/96	1046	B,D	87/96	1046	D,D	64/96	1046	B,D
ANZ 1039/91	H,K		72/96	1046	C,E	33/96	1046	K,E	69/96	1046	D,E
ANZ 1046/91	D,E		73/96	1046	D,E	14/96	1046	D,E	79/96	1046	E,H
ANZ 1048/91	E,K		256/96	1046	H,E	88/96	1046	E,E	5/96'	1046	D,H
ANZ 4071/94	L,C		91/96	1046	D,E	27/96	1046	D,E	177/96	1046	F,D (F,F)
ANZ 4079/94	??		24/96	1046	H,E	15/96	1046	B,E	287/96	1046	E,L
ANZ 1038/91	G,K		9/96'	1046	D,H	82/96	1046	C,D	26/96	1046	B,E
ANZ 4073/94	K,K		16/96	1046	D,D	241/96	1046	B,D	298/96	1046	??
			284/96	1046	??	63/96	1046	C,E	32/96	1046	D,?
			204/96	1046	E,E	6/96'	1046	E,?	28/96	1046	E,?
			78/96	1046	D,H	7/96'	1046	D,E	22/96	1046	E,E
			31/96	1046	D,F (F,F)	8/96'	1046	E,H	25/96	1046	E,F
			299/96	1046	E,G	86/96	1046	E,H	90/96	1046	E,H
			224/96	1046	C,D	313/96	1046	E,G	265/96	1046	E,?
			23/96	1046	C,E	199/96	1046	B,D	83/96	1046	E,H
			67/96	1046	??	200/96	1046	E,H	4/96'	1046	E,H
			10/96'	1046	D,H	291/96	1046	B,E	21/96	1046	D,F (F,F)
			176/96	1046	D,F (F,F)						
			169/96	1039	D,K	103/96	1039	H,L	84/96	1039	K,K
			251/96	1039	G,K	107/96	1039	K,K	98/96	1039	D,H

			92/96	1039	H,H	75/96	1039	B,K	97/96	1039	H,K
			47/96	1039	K,K	108/96	1039	H,H	156/96	1039	B,K
			70/96	1039	??	99/96	1039	C,K	215/96	1039	C,K
			112/96	1039	L,H	229/96	1039	G,H	289/96	1039	H,K
			80/96	1039	B,H	305/96	1039	D,H	193/96	1039	H,K
			102/96	1039	F,K	250/96	1039	K,K	101/96	1039	H,K
			44/96	1039	B,H	131/96	1039	F,K	168/96	1039	D,H
			217/96	1039	D,H	66/96	1039	F,E	296/96	1039	B,H
			283/96	1039	F,K	43/96	1039	H,L	178/96	1039	H,L
			187/96	1039	C,K	57/96	1039	H,L	216/96	1039	D,K
			210/96	1039	K,K	85/96	1039	L,K	230/96	1039	K,K
			223/96	1039	L,H	292/96	1039	C,K	106/96	1039	K,L
			48/96	1039	B,F	211/96	1039	K,K	100/96	1039	E,K
			81/96	1039	H,K	306/96	1039	D,K	300/96	1039	H,H
			61/96	1036	C,F	77/96	1036	K,E	40/96	1036	C,E
			116/96	1036	F,K	185/96	1036	C,E	125/96	1036	C,E
			35/96	1036	C,F	49/96	1036	B,F	123/96	1036	E,E
			117/96	1036	E,K	52/96	1036	K,E	128/96	1036	E,F
			290/96	1036	C,F	36/96	1036	F,K	51/96	1036	E,F
			54/96	1036	E,K	130/96	1036	C,E	275/96	1036	B,E
			45/96	1036	C,E	124/96	1036	K,E	50/96	1036	E,K
			118/96	1036	F,K	264/96	1036	F,C	56/96	1036	C,E
			38/96	1036	D,F (F,F)	62/96	1036	B,E	120/96	1036	F,K
			41/96	1036	E,K	37/96	1036	E,K	55/96	1036	F,D (F,F)
			281/96	1036	E,F	127/96	1036	C,F	129/96	1036	F,E
			46/96	1036	C,F	39/96	1036	B,E			

2) Corriedale flock 2-Orr											
Sires			Progeny								
Tag Number	DQA2 Alleles		Tag number	Sire	DQA2 Alleles	Tag number	Sire	DQA2 Alleles	Tag number	Sire	DQA2 Alleles
3547/92	H,D		4312	3433/92	D,G (G,G)	3436	8661/95	C,E	3871	7271/94	D,D
5128/93	H,G		4147	8215/95	J,E	3438	8661/95	H,G	3872	7271/94	D,D
5091/93	H,G		3546	7973/95	G,C (J,D G,J)	3441	8661/95	C,D	3879	7271/94	G,H
5179/93	J,E		3408	8661/95	J,D (G,C G,J)	3444	8661/95	J,C (J.J)	3884	7271/94	G,H
8009/95	H,G		4145	8215/95	G,C (J,D G,J)	3446	8661/95	C,C	3897	7271/94	B,D
8215/95	G,E		3757	8661/95	C,H	3448	8661/95	D,G (G,G)	3902	7271/94	G,D (G,G)
8661/95	H,D		3187	6510 + 6511	G,D (G,G)	3755	8661/95	D,?	3907	7271/94	D,E,H??
3433/92	GC(JD, GJ)		4219	5179/93	G,D (G,G)	3782	8661/95	H,H	3910	7271/94	G,H
6505/88	GC (JD, GJ)		4263	5179/93	J,D (G,C G,J)	3786	8661/95	G,D (G,G)	3920	7271/94	D,H
6510/94	GC, (JD, GJ)		4176	8215/95	D,E	3452	7973/95	D,E	3928	7271/94	G,D (G,G)
6511/91	H,G		3104	6510 + 6511	D,H	3457	7973/95	C,H	3933	7271/94	H,H
7973/95	C,E		3556	5128/93	H,G	3469	7973/95	C,J (J,J)	3938	7271/94	B,D
4R817	C,C		4063	R4187	C,C	3472	7973/95	G,C (D,J G,J)	3939	7271/94	D,D
7271/94	G,D (G,G)		4310	3433/92	C,D	3480	7973/95	H,E	3943	7271/94	B,G
7472/94	GC (JD, GJ)		3274	3547/92	H,J	3490	7973/95	J,E	3945	7271/94	D,G (G,G)
			3593	5128/93	H,E	3493	7973/95	C,E	3948	7271/94	B,D
			3176	6511 +6510	H,E	3495	7973/95	G,C (D,J G,J)	3950	7271/94	G,H
			3588	5128/93	H,?	3509	7973/95	C,D	3959	7271/94	G,D (G,G)
			3741	7472/94	C,E	3517	7973/95	C,C	3967	7271/94	GC (DJ, GJ)
			4194	5179/93	C,E	3528	7973/95	C,E	3969	7271/94	GC (DJ, GJ)
			3055	6510 + 6511	GC (JD, GJ)	3530	7973/95	C,H	3971	7271/94	G,D (G,G)
			3158	6510 + 6511	GC (JD, GJ)	3536	7973/95	J,E	3973	7271/94	D,H
			3593	5128/93	H,E	3550	7973/95	G,E	3974	7271/94	G,H
			4265	5179/93	E,G	3633	7472/94	GC (JD,GJ)	3977	7271/94	D,D

			3325	3547/92	H,G	3638	7472/94	GC (JD, GJ)	3292	3547/92	B,H
			4209	5179/93	C,E	3641	7472/94	GC (JD, GJ)	3293	3547/92	H,J
			3625	7472/94	J,E	3643	7472/94	C,C	3302	3547/92	H,G
			4339	3433/92	G,G (G,D)	3644	7472/94	G,C (JD,GJ)	3313	3547/92	G,D (G,G)
			3728	7472/94	C,C	3653	7472/94	G,C (JD,GJ)	3314	3547/92	G,D (G,G)
			3510	7973/95	J,E	3657	7472/94	C,H	3316	3547/92	J,D (G,C G,J)
			3028	6510 + 6511	C,E	3661	7472/94	G,E	3322	3547/92	H,D
			4151	8215/95	C,E	3671	7472/94	C,H	3336	3547/92	H,J
			4066	R4817/95	C,C	3674	7472/94	G,C (JD,GJ)	3340	8661/95	J,C (J,J)
			3620	7472/94	C,E	3681	7472/94	D,G (G,G)	3342	8661/95	C,C
			3154	6510 + 6511	GC (JD, GJ)	3683	7472/94	C,J (J,J)	3349	88661/95	D,G (G,G)
			3624	7472/94	C,J (J,J)	3685	7472/94	C,C	3352	8861/95	H,E
			3089	6510 + 6511	GC (JD, GJ)	3699	7472/94	G,H	3385	8661/95	C,E
			4104	R4817/95	E,E	3707	7472/94	H,E	3390	8661/95	D,C
			4118	8215/95	D,G (G,G)	3720	7472/94	C,E	3392	8661/95	H,E
			3227	3547/92	J,D (G,C G,J)	3725	7472/94	D,G (G,G)	3393	8661/95	J,C (J,J)
			4243	5179/93	C,E	3729	7472/94	??	3399	8661/95	D,D
			4291	3433/92	C,H	3736	7472/94	G,G (D,G)	3400	8661/95	C,C
			4006	R4817/95	GC (JD, GJ)	3738	7472/94	G,E	3401	8661/95	D,G (G,G)
			3387	8661/95	H,J	3743	7472/94	G,E	3411	8664/95	D,G (G,G)
			3299	3547/92	D,G (G,G)	3750	7472/94	C,C	3414	8661/95	D,G (G,G)
			4184	8215/95	B,G	4113	8215/95	C,E	3422	8661/95	H,G
			3484	7973/95	J,C	4135	8215/95	G,C (D,J G,J)	3424	8661/95	H,C
			3083	6510 + 6511	C,E	4141	8215/95	G,H	3426	8661/95	D,J (G,C G,J)
			3594	5128/93	D,H	4143	8215/95	G,H	3429	8661/95	D,J (G,C G,J)
			3860	7271/94	GC (JD, GJ)	4155	8215/95	G,C (D,J G,J)	3814	7271/94	GC (DJ, GJ)
			3488	7973/95	C,C	4156	8215/95	G,C (D,J G,J)	3832	7271/94	G,I
			3933	7271/94	H,H	4158	8215/95	??	3837	7271/94	GC (DJ, GJ)
			4302	3433/92	G,H	4159	8215/95	G,C (D,J G,J)	3841	7271/94	G,G (D,D)

			4074	R4817/95	GC (JD, GJ)	4160	8215/95	H,E	3842	7271/94	D,D
			3508	7973/95	C,J	4161	8215/95	E,G	3843	7271/94	H,?
			3460	7973/95	C,E	4163	8215/95	G,D (G,G)	3851	7271/94	G,D (G,G)
			4110	8215/95	J,E	4164	8215/95	G,G (D,G)	3856	7271/94	GC (DJ, GJ)
			3856	7271/94	GC (JD, GJ)	4175	8215/95	D,G (G,G)	3857	7271/94	C,C
			3702	7472/94	D,G (G,G)	4187	8215/95	D,E	3860	7271/94	GC (DJ, GJ)
			4428	3433/92	D,G (G,G)	4195	5179/93	C,E	3866	7271/94	J,H
			4345	3433/92	C,H	4231	5179/93	J,H	3222	3547/92	B,H
			3552	7973/95	C,J (J,J)	4235	5179/93	D,E	3232	3547/92	D,D
			3367	8661/95	J,D (G,C G,J)	4238	5179/93	C,H	3234	3547/92	D,G (G,G)
			4111	8215/95	C,E	4242	5179/93	G,H	3248	3547/92	B,H
			3969	7271/94	GC (JD, GJ)	4262	5179/93	E,E	3252	3547/92	B,D
			3264	3547/92	D,G (G,G)	4271	5179/93	E,G	3253	3547/92	G,H
			3246	3547/92	C,H	4285	3433/92	C,H	3266	3547/92	D,G (G,G)
			4064	R4817/95	C,H	4286	3433/92	GC (DJ, GJ)	3269	3547/92	B,D
			3082	6510 + 6511	J,H	4290	3433/92	G,D (G,G)	3283	3547/92	C,D
			3018	6510 + 6511	G,H	4306	3433/92	L,D	3284	3547/92	C,H
			3208	3547/92	H,C	4311	3433/92	G,D (G,G)	3286	3547/92	D,H
			3706	7472/94	G,G (D,G)	4316	3433/92	D,G (G,G)	3791	7271/94	GC (DJ, GJ)
			3745	7472/94	GC (JD, GJ)	4328	3433/92	H,H	3793	7271/94	D,G (G,G)
			3359	8661/95	D,G (G,G)	4331	3433/92	D,H	3794	7271/94	D,H
			3967	7271/94	GC (JD, GJ)	4411	3433/92	GC (DJ, GJ)	3802	7271/94	C,D
			4288	3433/92	G,D (G,G)	4422	3433/92	H,H	3808	7271/94	C,D
			4099	R4817/95	E,?	4425	3433/92	C,H	3810	7271/94	G,D (G,G)
			4136	8215/95	G,C (D,J G,J)	4472	3433/92	BC	3811	7271/94	D,D
			4051	R4817/95	C,C	4474	3433/92	G,E	3812	7271/94	C,D
			3310	3547/92	G,D (G,G)	3202	3547/92	B,H	3321	3547/92	J,D (G,C G,J)
			3275	3547/92	D,J (G,C G,J)	3212	3547/92	D,H	3394	8661/95	H?
			4256	5179/93	G, H	3215	3547/92	D,D			

3) Merino flock 1.											
Sires collected 1996.						Sires Collected 1998.					
Tag Number	DQA2 Alleles		Tag Number	DQA2 Alleles		Tag Number	DQA2 Alleles				
380	Q,G		20	D,E		Green 4735	G,C (J,D G,J)				
366	G,C (J,D G,J)		18	C,C		Orange 100	G,C (J,D G,J)				
343	H,J		17	G,C (J,D G,J)		Orange 101	C,Q				
28	D,G (G,G)		16	G,C (J,D G,J)		Orange 96	B,G				
27	G,G (G,D)		351/1529	G,C (J,D G,J)		Green 4400	B,R				
25	G,G (G,D)		1097/1155	C,J (J,J)		Orange 97	G,D (G,G)				
24	G,C (J,D G,J)		1009	G,C (J,D G,J)		Orange 98	G,C (J,D G,J)				
23	E,G		930/373	C,J (J,J)		Orange 89	G,C (J,D G,J)				
19	G,C (J,D G,J)		761/1023	G,M		Z-6746-3	C,D				
13	C,C		1174/587	D,L		Orange 103	C,C				
4	C,C		376	C,C		Orange 87	D,D				
3	G,G (G,D)		370	E,E		Yellow 4710	J,L				
15	E,G		369/2777	H,C		Z-5421-1	C,C				
14	D,G (G,G)		353	L,L		Orange 90	J,N				
12	C,C		334	C,E		309	G,C (J,D G,J)				
11	D,L		210	C,D		Orange 95	D,D				
10	E,G		32	G,G (G,D)		Orange 94	H,L				
9	C,E		31	B,B		Orange 99	G,G (G,D)				
8	C,D		30	C,C		Orange 93	G,C (J,D G,J)				
7	B,G		29	G,C (J,D G,J)		Orange 91	G,C (J,D G,J)				
6	C,D		26	B,E		Z-5331-1	B,C				
5	C,E		22	??		Orange 85	J,N				
2	D,G (G,G)		21	??		367	G,C (J,D G,J)				
1	G,G (G,D)		162	C,E							

Progeny Collected During 1997			Tag Number	Sire	DQA2 Alleles	Tag Number	Sire	DQA2 Alleles	Tag Number	Sire	DQA2 Alleles
			4262	380	E,G	4268	380	B,Q	4273	380	B,Q
			4263	380	D,G (G,G)	4269	380	C,G (D,J G,J)	4274	380	C,Q
			4264	380	C,G (D,J G,J)	4270	380	D,G (G,G)	4275	380	G,Q
			4265	380	J,Q	4271	380	C,G (D,J G,J)	4276	380	C,G (D,J G,J)
			4267	380	C,G (D,J G,J)	4272	380	J,Q	4277	380	G,N
			4278	380	C,G (D,J G,J)	4282	380	B,G	4287	380	C,G (D,J G,J)
			4279	380	C,Q	4284	380	C,G (D,J G,J)	4288	380	G,L
			4280	380	C,G (D,J G,J)	4285	380	C,G (D,J G,J)	4289	380	H,Q
			4281	380	B,Q	4286	380	C,Q	4290	380	C,G (D,J G,J)
			4292	380	G,Q	4297	380	G,H	4294	380	G,H
			4293	380	D,G (G,G)						
			4572	10	E,Q	4593	10	G,H	4583	10	B,G
			4573	10	C,E	4594	10	C,G (D,J G,J)	4584	10	E,D
			4574	10	D,E	4595	10	C,G (D,J G,J)	4585	10	D,G (G,G)
			4575	10	C,G (D,J G,J)	4596	10	D,G (G,G)	4586	10	D,G (G,G)
			4576	10	B,E	4597	10	G,D (G,G)	4588	10	E,K
			4577	10	H,G	4599	10	C,G (D,J G,J)	4589	10	E,I
			4578	10	E,F	4600	10	C,G (D,J G,J)	4590	10	C,G (D,J G,J)
			4579	10	C,G (D,J G,J)	4601	10	L,G	4591	10	C,E
			4580	10	C,E	4602	10	D,G (G,G)	4592	10	C,G (D,J G,J)
			4581	10	E,G	4603	10	C,E	4605	10	G,Q
			4582	10	D,G (G,G)	4604	10	G,L	4606	10	C,G (D,J G,J)
			4607	10	C,G (D,J G,J)	4611	10	L,E	4614	10	B,G
			4608	10	E,Q	4612	10	C,G (D,J G,J)	4615	10	D,E
			4610	10	E,E	4617	10	C,E	4619	10	C,G (D,J G,J)
			4621	10	F,G	4618	10	E,L	4620	10	B,G

			4624	10	C,G (D,J G,J)						
			5013	9	C,C	5027	9	E,G	5039	9	E,I
			5014	9	C,C	5028	9	C,D	5041	9	E,D
			5015	9	C,E	5029	9	E,J	5043	9	C,E
			5016	9	E,L	5030	9	C,E	5044	9	C,E
			5017	9	C,C	5031	9	B,E	5045	9	C,H
			5018	9	C,E	5032	9	C,C	5046	9	C,R
			5019	9	C,J	5033	9	C,H	5049	9	E,L
			5020	9	C,C	5034	9	E,R	5050	9	C,D
			5022	9	E,H	5035	9	C,H	5038	9	E,G
			5024	9	C,R	5036	9	E,R	5026	9	C,L
			5025	9	C,E	5037	9	C,E			
			4709	309	C,D	4722	309	C,G (D,J G,J)	4733	309	C,G (D,J G,J)
			4710	309	C,G (D,J G,J)	4723	309	C,G (D,J G,J)	4734	309	??
			4711	309	C,C	4724	309	C,G (D,J G,J)	4735	309	??
			4712	309	C,G (D,J G,J)	4727	309	C,G (D,J G,J)	4736	309	??
			4713	309	C,C	4729	309	??	4737	309	??
			4714	309	C,C	4730	309	??	4738	309	??
			4715	309	C,G (D,J G,J)	4731	309	C,C	4739	309	C,G (D,J G,J)
			4717	309	??	4732	309	G,L	4740	309	C,G (D,J G,J)
			4718	309	??	4720	309	??	4741	309	??
			4719	309	B,C	4721	309	??	4743	309	C,J (J,J)
			4744	309	C,H	4745	309	??			

* Not all the samples collected from sire 309 were typed as it became apparent that this sire was a C allele homozygote, as the sire had not been typed before the progeny were collected.

Progeny Collected During 98-99			Tag Number	Sire	DQA2 Alleles	Tag Number	Sire	DQA2 Alleles	Tag Number	Sire	DQA2 Alleles
			8276	8	G,C (D,J G,J)	8372	8	C,C	8373	8	G,C (D,J G,J)
			8303	8	C,C	8139	8	G,E	8395	8	C,C

			8361	8	G,C (D,J G,J)	8164	8	C,H	8469	8	G,C (D,J G,J)
			8672	8	B,C	8165	8	C,H	8345	8	G,B
			8279	8	G,C (D,J G,J)	8352	8	C,L	8351	8	G,L
			8317	8	G,D (G,G)	8075	8	G,C (D,J G,J)	8008	8	G,H
			8353	8	C,C	8157	8	C,D	8039	8	G,C (D,J G,J)
			8354	8	C,C	8320	8	G,C (D,J G,J)	8076	8	G,C (D,J G,J)
			8380	8	C,D	8470	8	G,C (D,J G,J)	8234	8	G,C (D,J G,J)
			8401	8	G,C (D,J G,J)	8104	8	G,C (D,J G,J)	8304	8	G,C (D,J G,J)
			8409	8	C,C	8387	8	L,C	8212	8	D,G (G,G)
			8516	8	J,N	8647	8	C,H	8319	8	L,C
			8306	8	C,C	8442	8	D,C			
			8346	101	C,D	8301	101	C,L	8178	101	Q,D
			8535	101	Q,H	8302	101	G,L	8231	101	I,K
			8659	101	Q,R	8323	101	G,C (D,J G,J)	8241	101	C,Q
			8676	101	L,J	8383	101	C,K	8259	101	J,N
			8044	101	C,C	8404	101	C,H	8145	101	L,C
			8088	101	C,H	8412	101	C,Q	8156	101	C,C
			8089	101	Q,H	8424	101	L,K	8169	101	G,K
			8093	101	Q,L	8436	101	H,K	8506	101	Q,E
			8114	101	Q,H	8443	101	J,K	8507	101	C,E
			8115	101	L,C	8499	101	C,Q	8515	101	E,K
			8478	101	Q,G	8542	101	Q,D	8571	101	H,K
			8521	101	K,K	8344	101	C,H	8531	101	Q,D
			8528	101	C,C	8092	101	C,C	8278	101	Q,E
			8599	101	C,L	8116	101	L,K	8347	101	C,C
			8607	101	C,K	8124	101	C,C	8518	101	C,D
			8615	101	C,L	8141	101	L,C	8543	101	C,D
			8619	101	C,C	8582	101	L,E	8189	101	C,H
			8635	101	C,D	8594	101	L,R	8500	101	C,Q

			8079	101	L,Q	8123	101	C,Q	8213	101	Q,Q
			8136	101	H,Q	8043	101	D,Q	8214	101	Q,Q
			8170	101	K,H	8188	101	C,C	8327	101	K,Q
			8437	101	H,K						
			8564	4403	C,D	8309	4403	C,H	8244	4403	D,J
			8620	4403	C,H	8148	4403	D,D	8110	4403	C,H
			8055	4403	G,H	8197	4403	C,D	8316	4403	D,H
			8158	4403	C,H	8205	4403	G,D (G,G)	8200	4403	D,J (G,C G,J)
			8233	4403	C,D	8273	4403	D,H	8208	4403	C,H
			8262	4403	C,H	8287	4403	G,H	8473	4403	C,H
			8315	4403	C,H	8363	4403	B,D	8603	4403	D,H
			8448	4403	C,H	8419	4403	G,H	8308	4403	C,D
			8471	4403	D,E	8559	4403	H,H	8016	4403	Q,H
			8558	4403	D,N	8614	4403	D,L	8117	4403	G,D (G,G)
			8566	4403	C,H	8656	4403	C,H	8618	4403	C,D

4) Merino flock 2 * From the progeny DQA2 typing results, the sire was thought to have the DQA2 alleles L, I.

			Tag Number	Sex	DQA2 Alleles	Tag Number	Sex	DQA2 Alleles	Tag Number	Sex	DQA2 Alleles
			E2-100	F	L,C	E2-107	F	C,I (I,I)	E2-117	F	L,N
			E2-101	F	G,L	E2-108	F	L,D	E2-118	F	L,H
			E2-103	F	C,I (I,I)	E2-111	F	C,I (I,I)	E2-119	F	C,I (I,I)
			E2-125	F	D,I	E2-142	F	C,I (I,I)	E2-66	F	L,D
			E2-130	F	L,E	E2-148	F	L,I	E2-68	F	I,E
			E2-135	F	L,E	E2-52	F	L,C	E2-70	F	C,I (I,I)
			E2-136	F	L,D	E2-53	F	C,I (I,I)	E2-89	F	I,I (C,I)
			E2-138	F	C,I (I,I)	E2-56	F	L,D	E2-93	F	I,N
			E2-139	F	J,L	E2-63	F	I,H	E2-95	F	I,I (C,I)
			E2-97	F	L,N						

			E2-102	M	D,I	E2-129	M	L,H	E2-59	M	L,I
			E2-110	M	I,I (C,I)	E2-132	M	I,N	E2-67	M	C,I (I,I)
			E2-121	M	L,H	E2-141	M	C,L	E2-69	M	L,D
			E2-122	M	J,I	E2-143	M	L,L	E2-71	M	I,R
			E2-123	M	C,I (I,I)	E2-144	M	L,D	E2-72	M	L,L
			E2-126	M	I,D	E2-145	M	D,I	E2-73	M	C,L
			E2-127	M	L,H	E2-150	M	I,D	E2-75	M	C,I (I,I)
			E2-80	M	I,N	E2-55	M	L,L	E2-77	M	I,E
			E2-86	M	I,R	E2-98	M	??	E2-94	M	L,I

† It is necessary to note that some animals have an ambiguous result, where more than one possible combination of alleles is possible. For example animals that type as G, C may in fact have alleles D, J or G, J as the banding pattern observed is the same for all of these allelic combinations. In such cases were this banding pattern was observed within sire samples the most likely allelic combination could be identified by typing progeny samples. In progeny samples where this banding pattern occurred, the allelic combination present was determined by looking at the alleles that could have been inherited from the sire. In cases where more than one allelic combination was possible the most likely combination is written first, while other possible combinations are written in brackets. Animals where the most likely alleles present could not be determined adequately were left out of the analysis.

** For alleles G, F, I and J it was difficult to determine whether these animals were homozygous or heterozygous at the DQA_2 locus. For example, an animal that typed as a G,G may in fact have alleles G,D as the observed banding pattern is the same, similarly F,F may be F,D, I,I may be I,C, and J,J may be J,C . Generally the most likely alleles were determined by assessing the relative band intensities. Where progeny samples had been collected from sires whose alleles had been determined in this manner it was possible to confirm the DQA_2 alleles present from the progeny results. Again, when more than one allelic combination is possible, the most likely have been written first, while other possible alleles have been written in brackets. In cases where the DQA_2 alleles are unknown, or where samples were not typed are indicated by ??.

Appendix G: The Relative Risk Calculation

$$\frac{\left[\frac{\text{Number of animals with the allele and footrot.}}{\text{Total number of animals with the allele.}}\right]}{\left[\frac{\text{Number of animals without the allele with footrot.}}{\text{Total number of animals without the allele.}}\right]} = \textbf{Relative Risk.}$$

Appendix H: GenBank Accession Numbers for Sequences Used in Chapter Three

Sequence Name	GenBank Accession	Author
Ovar-DRA1	Z11600	Ballingall K.T
Ovar-DQA1-1	M33304	Scott et al
Ovar- DQA1-2	L49463	Snibson K.J
Ovar-DQA1-3	L49464	Snibson K.J
Ovar-DQA1-4	Z28518	Ballingall K.T
Ovar-DQA2-1	M93433	Fabb et al
Ovar-DQA2-4	M33305	Scott et al
Ovar-DQA2-7	L49415	Snibson K.J
Ovar-DQA2-10	L49416	Snibson K.J
Ovar-DQA2-13	L49411	Snibson K.J
Ovar-DQA2-B	AF129117	Slow
Ovar-DQA2-C1	———	Slow
Ovar-DQA2-C2	AF129119	Slow
Ovar-DQA2-C3	———	Slow
Ovar-DQA2-D	AF129120	Slow
Ovar-DQA2-E	AF129122	Slow
Ovar-DQA2-F1	———	Slow
Ovar-DQA2-F2	———	Slow
Ovar-DQA2-G1	———	Slow
Ovar-DQA2-G2	———	Slow
Ovar-DQA2-H	AF129124	Slow
Ovar-DQA2-J	———	Slow
Ovar-DQA2-K	AF129116	Slow
Ovar-DQA2-Q	AF129118	Slow
BoLA-DQA2-2	D50059	Aida Y
BoLA-DQA3-1	Y14022	Ballingall et al
BoLA-DQA3-2	Y14021	Ballingall et al