Assessing MHC class I diversity in dairy cattle populations

By

Gemma Frances Codner

Submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

School of Veterinary Medicine

University of Glasgow

I would also like to acknowledge the Institute for Animal Health, the Biotechnology and Biological sciences research council and Merial for funding this project.

Abstract

The gene dense major histocompatibility complex (MHC) region, present in all jawed vertebrates, encodes molecules involved in self-non-self discrimination and the binding and presentation of antigenic peptides to T cells during the adaptive immune response. Variation in MHC genes is thought to be driven largely by pathogen-mediated selection, with diversity at MHC loci believed to benefit populations by allowing responses to rapidly evolving disease pathogens. However, in economically important dairy cattle, there are concerns that intensive selection for production and fitness traits may override natural selection. It had been hypothesised that these focussed dairy breeding practices may lead to a reduction in MHC diversity and leave cattle populations susceptible to new disease pathogens.

The purpose of this study was to estimate current levels of MHC class I diversity in the UK Holstein-Friesian dairy cattle population, primarily through the assessment of diversity in bull populations with genetic input into the UK herd. In a sample of Canadian Holstein bulls, levels of class I allelic diversity were low given the size of the population sampled, but no significant loss of diversity over a twenty year period of selection was detected.

Simulations of gene flow implicated trait selection as an influential force shaping diversity in the Canadian Holstein bull population. Haplotypes detected at high frequency were often negatively associated with selection traits indicating the action of heterozygote advantage. A SNP-based assay has been designed to facilitate rapid detection of common haplotypes and thus enable breeders to make more efficient selective breeding decisions whilst also maintaining MHC diversity in cattle populations.

Investigations of class I diversity were expanded to incorporate the British Friesian bull population which were shown to have a markedly different pattern of class I diversity to that observed in the Canadian Holstein sample. A number of novel allele sequences and haplotypes were detected in the British Friesian bulls, the characterisation of which has contributed to our knowledge of the mechanisms driving diversity in the cattle class I region.

MHC class I typing data from two bull populations and statistical analysis of trait associations with MHC haplotypes provides a comprehensive picture of MHC class I diversity in the wider UK herd and the selective forces integral to shaping diversity.

Contents

List of Tables

Chapter 1. Introduction

Chapter 2. Materials and Methods

Chapter 3. Analysis of MHC class I diversity in Canadian Holstein bulls

Chapter 4. Dissecting the impact of artificial selection upon MHC class I diversity

Chapter 5. Design and development of a SNP-based assay for the detection of common MHC class I haplotypes

Chapter 6. Investigating MHC class I diversity in the British Friesian population

Chapter 7. General discussion and further work

List of Figures

Chapter 1. Introduction

Chapter 3. Analysis of MHC class I diversity in Canadian Holstein bulls

Chapter 4. Dissecting the impact of artificial selection upon MHC class I diversity

Chapter 5. Design and development of a SNP-based assay for the detection of common MHC class I haplotypes

Chapter 6. Investigating MHC class I diversity in the British Friesian population

Chapter 7. General discussion and further work

Acknowledgement

I would like to thank both my IAH supervisor Prof. Shirley Ellis and Prof. Mike Stear for their valuable advice and guidance throughout this PhD project. Prof. Mike Stear, Dr Richard Reeve and Dr Louise Matthews have all provided invaluable help with the statistical analysis of data and mathematical modelling involved in this project for which I am very grateful. Thanks must also go to all the members of F1B past and present for their help and support during my time in the laboratory, especially Dr Efrain Guzman and Dr John Hammond.

I would also like to thank my family and all my friends, both at the Institute and outside of work. During the last four years you have all borne the brunt of my frustrations, provided many laughs and a welcome distraction, but above all, unfaltering encouragement.

This thesis is dedicated to my grandfather Richard Codner.

Author's declaration

I declare that, except where explicit reference is made to the contribution of others, that this dissertation is the result of my own work and has not been submitted for any other degree at the University of Glasgow or any other institution.

Signature _____

Printed name _____

List of abbreviations and definitions

aa	amino acid
Ag	antigen
APC	antigen presenting cell
$ß_2m$	$ß_2$ microglobulin
BLASTn	basic local alignment search tool
BoLA	bovine leucocyte antigen
bp	base pairs
cDNA	complementary DNA
cM	centimorgan
CTL	cytotoxic T lymphocyte
°C	degrees Celsius
dATP	deoxyadenosine triphosphate
dCTP	deoxycytidine triphosphate
dGTP	deoxyguanosine triphosphate
dTTP	deoxythymidine triphosphate
dNTP	deoxynucleotide
ddATP	dideoxyadenosine triphosphate
ddCTP	dideoxycytidine triphosphate
ddGTP	dideoxyguanosine triphosphate
ddTTP	dideoxythymidine triphosphate
ddNTP	dideoxynucleotide
DMSO	dimethylsulphoxide
DNA	deoxyribose nucleic acid
EDTA	ethylenediaminetetra-acetic acid
ER	endoplasmic reticulum
EtBr	ethidium bromide
FAM	fluorescein amidite
FLR	fluorescent-labelled reference strand
g	force of gravity
gDNA	genomic DNA
HA	heteroduplex analysis
H-2	murine histocompatibility complex
HLA	human leucocyte antigen

IPD	Immuno Polymorphism Database
kb	kilobase
LB	Luria-Bertani broth
L	litre
Mb	megabase
$MgCl_2$	magnesium chloride
MHC	major histocompatibility complex
µg	microgram
µl	microlitre
mg	milligram
ml	millilitre
mM	millimolar
M	molar
MW	molecular weight
NK	natural killer cell
nt	nucleotide
PBMC	peripheral blood mononuclear cells
PBC	peptide binding cleft
PBS	phosphate buffered saline
PCR	Polymerase chain reaction
PCR-SSP	PCR-sequence specific primers
RBC	red blood cell
RNA	ribose nucleic acid
ROX	6-carboxyl-X-rhodamine
RSCA	reference strand-mediated conformational analysis
SDS	sodium dodecyl sulphate
TAE	tris-acetate EDTA
TAP	transporter associated with antigen processing
TBE	tris-borate EDTA
TCR	T cell receptor
TE	tris-EDTA
Tris	2-amino-2-(hyromethyl)propane-1,3 diol
U	unit
UTR	untranslated region
UV	ultraviolet

V	voltage
X-gal	5-bromo-4-chloro-3-indolyl-ß-D-galactopyranoside

Bases

A	Adenine
C	Cytosine
G	Guanine
U	Uracil
T	Thymine

Amino Acids

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

Introduction

1.1. Adaptive immunity & MHC discovery

Early in the twentieth century, studies of tumour graft fate in mice (Little & Tyzzer, 1916) and skin transplants between identical twins (Bover, 1927), indicated that genetic factors were involved in allograft rejection. The genetic region, now known as the major histocompatibility complex (MHC), was first defined in mice by Gorer (1936) and later characterised by Snell (1958) as a set of polymorphic genes encoding molecules involved in graft rejection. The demonstration of leukocyte agglutination induced by incubation with serum of a multi-transfused individual led to the discovery of MHC in human (Dausett & Nenna, 1952). The importance of dissecting the host response to non-self and pathogenic infections led to an increasing number of studies investigating the adaptive immune response and its components in a range of species.

Antibodies, T cell receptors and the major histocompatibility complex (MHC) are defining components of the adaptive immune system (Flajnik, 2004) with the recognition and binding of signature pathogenic sequences by specific lymphocyte receptors integral to initiating an adaptive immune response (Pancer et al., 2004). The absence of these components in the jawless fish and invertebrates suggests that the adaptive immune system originated with the emergence of the jawed vertebrates, approximately 500 million years ago (Danchin et al., 2004).

1.2. MHC molecules and their function

It is estimated that almost half of the loci encoded on the gene dense MHC region are dedicated to immune functions (Kelley et al., 2005). Most notable are those genes encoding molecules integral to both self-non-self discrimination and the binding and presentation of antigenic peptides to T cells (Townsend & Bodmer, 1989a). The MHC region is subdivided into three classes based upon gene product function (Rhodes & Trowsdale, 1999). Genes in the class I region encode constitutively expressed transmembrane glycoproteins which are involved in presentation of endogenous antigens to CD8+ T cells. An additional function of MHC class I molecules is their interaction, either activating or inhibitory, with Natural Killer (NK) cell receptors. The MHC class II region also encodes glycoproteins, however these are restricted in their distribution to the cell membranes of macrophages, dendritic cells, B cells and thymic epithelium. MHC class II molecules present exogenously derived peptides to T cells expressing the $CD4^+$ co-receptor (T helper cells). In contrast to the class I and class II, the class III region encodes a number of secreted proteins of innate immune function and inflammation i.e. complement proteins and cytokines.

The ability of the immune system to distinguish self from non-self is reliant on a process named MHC restriction. This process was first demonstrated by Zinkernagel and Doherty (1974) who were able to show that murine cytotoxic T lymphocytes would only kill cells presenting viral antigens in the context of an MHC molecule of the same genotype. During development of an individual's immune system, immature T cells within the thymus are exposed to self peptide-MHC complexes in a process known as clonal selection, as shown in Figure 1.1. T cells deficient in T cell receptors (TCRs) or not exposed to self peptide-self MHC complexes undergo apoptosis. Additionally, negative selection eliminates 'forbidden clones' which are T cells bearing TCRs with too strong an affinity for self peptide-self MHC. Only 5 % of T cells survive selection and these will have a sufficient affinity for self peptide-self MHC complexes to drive positive selection. Autoimmune disease is based upon the failure of clonal selection and

the inability of the immune system to distinguish self from non-self, leading to the destruction of the body's own cells.

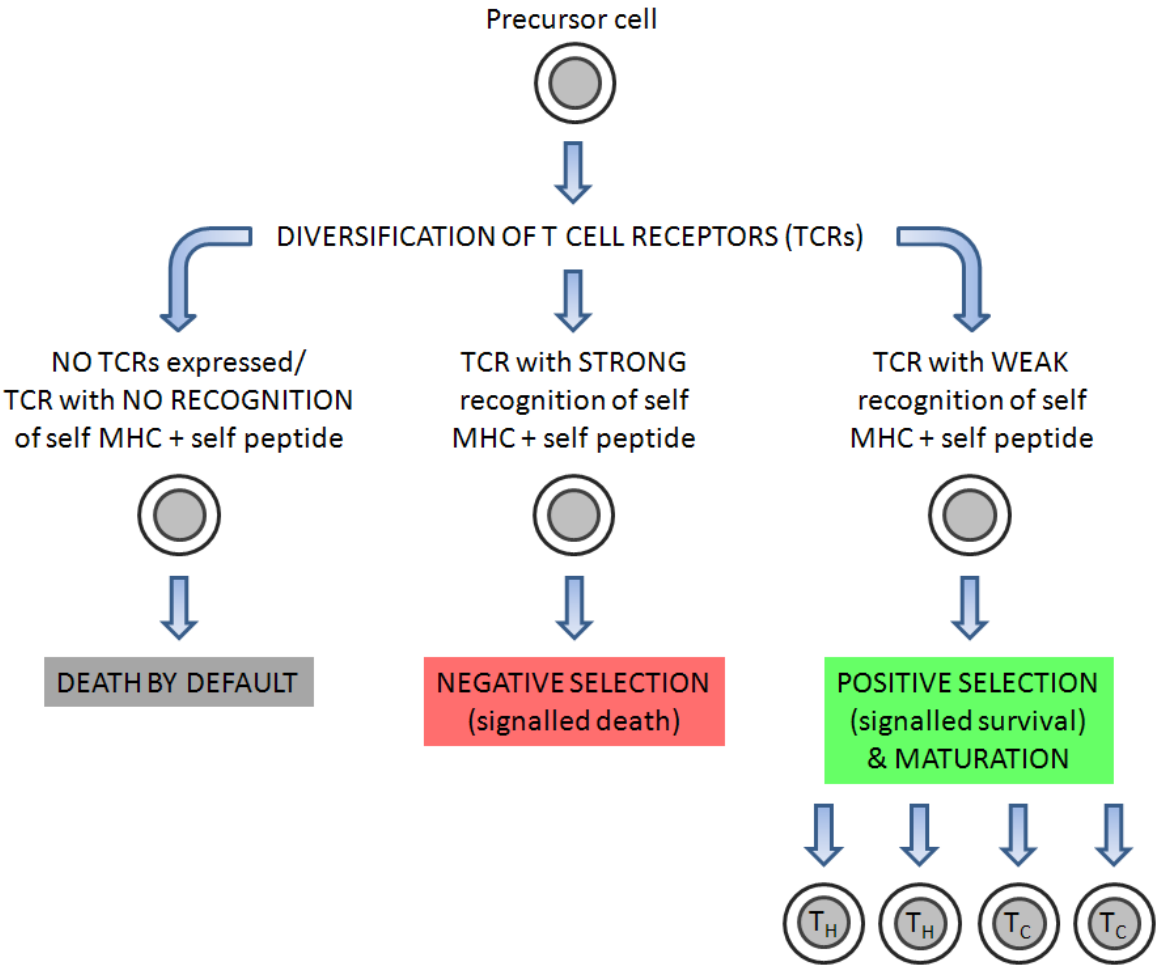


Figure 1.1. Positive and negative selection in the thymus (adapted from Alberts et al., 1994).

1.3. MHC molecules

Both MHC class I and class II molecules are vital in antigen presentation to T cells. However, these molecules differ in both structure and the type of T cells to which they present antigen.

1.3.1. MHC class I molecules

MHC class I molecules are heterodimers consisting of a heavy alpha-chain (45 kDa) and a light chain, β_2 microglobulin (12 kDa). The membrane-spanning heavy chain comprises a cytoplasmic domain, transmembrane domain and three extracellular domains, α_1, α_2, and α_3, with which the β_2 microglobulin chain is covalently associated (Flutter & Gao, 2004). The highly conserved α_3 and β_2 microglobulin domains consist of two anti-parallel β-sheets joined by a disulphide bond and are structurally homologous to the immunoglobulin constant domains (Becker & Reeke, 1985; Bjorkman et al., 1987; Cunningham et al., 1975). The α_3 and β_2 microglobulin domains are proximal to the cell membrane and support the hypervariable α_1 and α_2 domains which form the peptide binding cleft of class I molecules. The tertiary structures of the α1 and α2 domains are similar, each consisting of four anti-parallel β-strands with a long α-helical region at the C-terminus. The eight β-strands of the α_1 and α_2 domains form one large β-sheet, constituting the floor of the peptide binding cleft, and the two α-helices form the peripheral boundaries of the cleft in which peptides are bound. The binding of peptides stabilise MHC class I molecules and facilitate intracellular transport, although it has been demonstrated that at low temperatures empty MHC class I molecules are able to assemble and are expressed on the cell membrane, but these molecules are not stable at physiological temperatures. Figure 1.2A. presents the structure of the MHC class I molecules peptide binding cleft and a diagrammatic representation of the cell membrane bound MHC class I molecule with peptide.

Peptides bound by MHC class I molecules are typically eight to ten amino acids in length and are secured within the binding cleft by the interaction of invariant class I residues with the bound peptide at both the carboxy- and amino-termini (Zhang et al., 1998). Within the peptide binding cleft of the MHC class I molecule are a series of structural pockets, unique to each MHC allele as a result of polymorphism in the α_1 and α_2 domains, which accommodate peptide side chains, facilitate peptide binding and dictate antigen specificity (Matsumura et al., 1992). The structural pockets along the peptide binding cleft give rise to a specific peptide binding motif for each allele, defined by a series of anchor residues along the length of the peptide. Anchor residues binding a specific allele, of which there are usually two or three, are highly conserved with those positions in all peptide sequences binding to that allele being relatively invariant or having conserved properties. In humans, these anchor positions are usually located at the C terminus and position P2 of the peptide. For example, alleles of the *HLA-B27* subtype bind peptides with an arginine residue located at P2 and aliphatic or polar residues (lysine, isoleucine, leucine, arginine or valine) at the C terminus (Rammensee et al., 1995).

Peptides loaded into MHC class I molecules are derived from cytosolic self and foreign proteins which are cleaved by the proteosome into peptide chains approximately 8-10 amino acids in length. These are then transported into the endoplasmic reticulum (ER) lumen via two transporters associated with antigen processing (TAP) proteins. Once inside the ER, the processed peptides bind with, and stabilise the folding of MHC class I subunits with the help of chaperone proteins before transport to the cell membrane (Townsend et al., 1989b). On expression at the cell surface, CD8 T cell receptors (TCR) recognise cell membrane-bound MHC class I molecules that have an associated peptide. Once the CD8+ TCR has bound to the ligand, selective lysis of infected cells may be induced. This interaction also initiates proliferation and differentiation of cytotoxic T cells and the establishment of memory T helper cells (Sher & Ahmed, 1995). Figure 1.3A. shows a simplified diagrammatic representation the class I antigen presentation pathway.

1.3.2. MHC class II molecules

MHC class II molecules are composed of two non-covalent transmembrane glycoprotein chains, a 30-34 kDa alpha-chain and a 26-29 kDa beta-chain, which are joined by a disulphide bond. Each of these chains comprise a cytoplasmic domain, transmembrane domain and two extracellular domains; α_1 and α_2 on the alpha-chain, and β_1 and β_2 on the beta-chain. Similar to MHC class I molecules, the domains proximal to the cell surface, α_2 and β_2, have immunoglobulin-like structures, and the α_1 and β_1 domains are similar to the MHC class I α_1 and α_2 domains. The interaction of the α_1 and β_1 domains form the MHC class II binding cleft, comprising eight anti-parallel β-sheets bounded by an α-helix from each contributing domain, in a similar manner to class I molecules. The invariant residues that secure the carboxy- and amino-termini of the peptide in the peptide binding cleft of class I molecules are absent in class II molecules enabling the presentation of antigens at least 13 amino acids in length. Peptide binding within the class II cleft is facilitated by the insertion of antigenic side chains into structural pockets and hydrogen bonding along the length of the peptide (nine residues), with the additional residues at the carboxy- and amino-termini having relative conformational freedom. Polymorphic residues located within the structural pockets of the peptide binding cleft dictate peptide specificity (Jardetzky et al., 1996). Figure 1.2B. presents the structure of the MHC class II molecules peptide binding cleft and a diagrammatic representation of the cell membrane bound MHC class I molecule with peptide.

The pathway of peptide presentation for MHC class II molecules is somewhat different to that of class I. Following translocation into the ER lumen, MHC class II subunits are bound to an invariant chain (MHC II-Ii) molecule which prevents antigen binding by preventing folding or physically interfering. The invariant chain also directs the MHC class II molecule to low-pH endosomal compartments. Once fused with an acidic endosome the invariant chain molecule is degraded by proteases, leaving a short peptide called CLIP. In the presence of foreign peptides, CLIP is released and the antigen loaded into the peptide binding cleft

by chaperones. The MHC class II-antigen complex is displayed on the extracellular surface of the cell for detection by CD4+ T cell receptors. Figure 1.3B. shows a simplified diagrammatic representation the class II antigen presentation pathway.

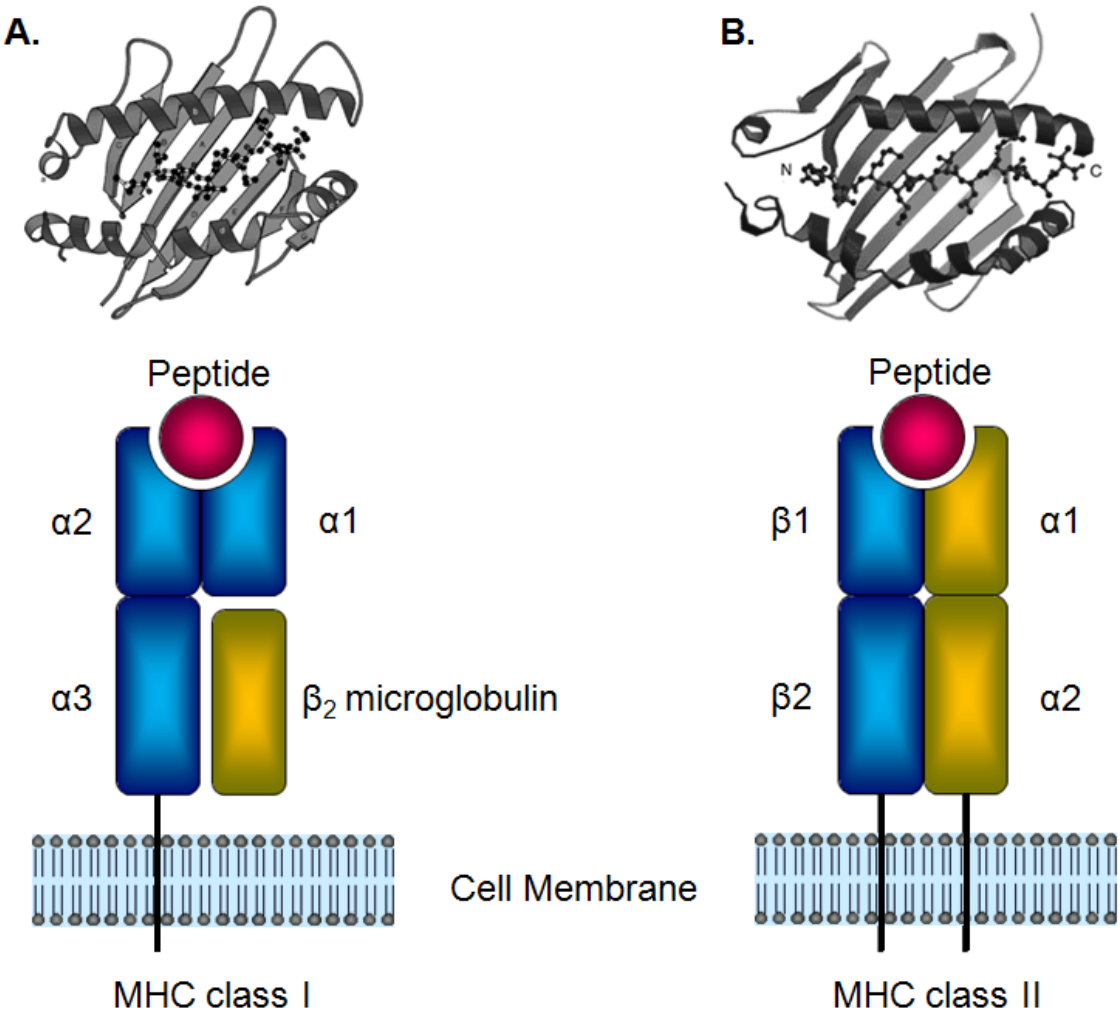


Figure 1.2. Schematic diagrams of the peptide binding cleft with peptide and overall structure of cell membrane bound **A.** MHC class I and **B.** MHC class II molecules.

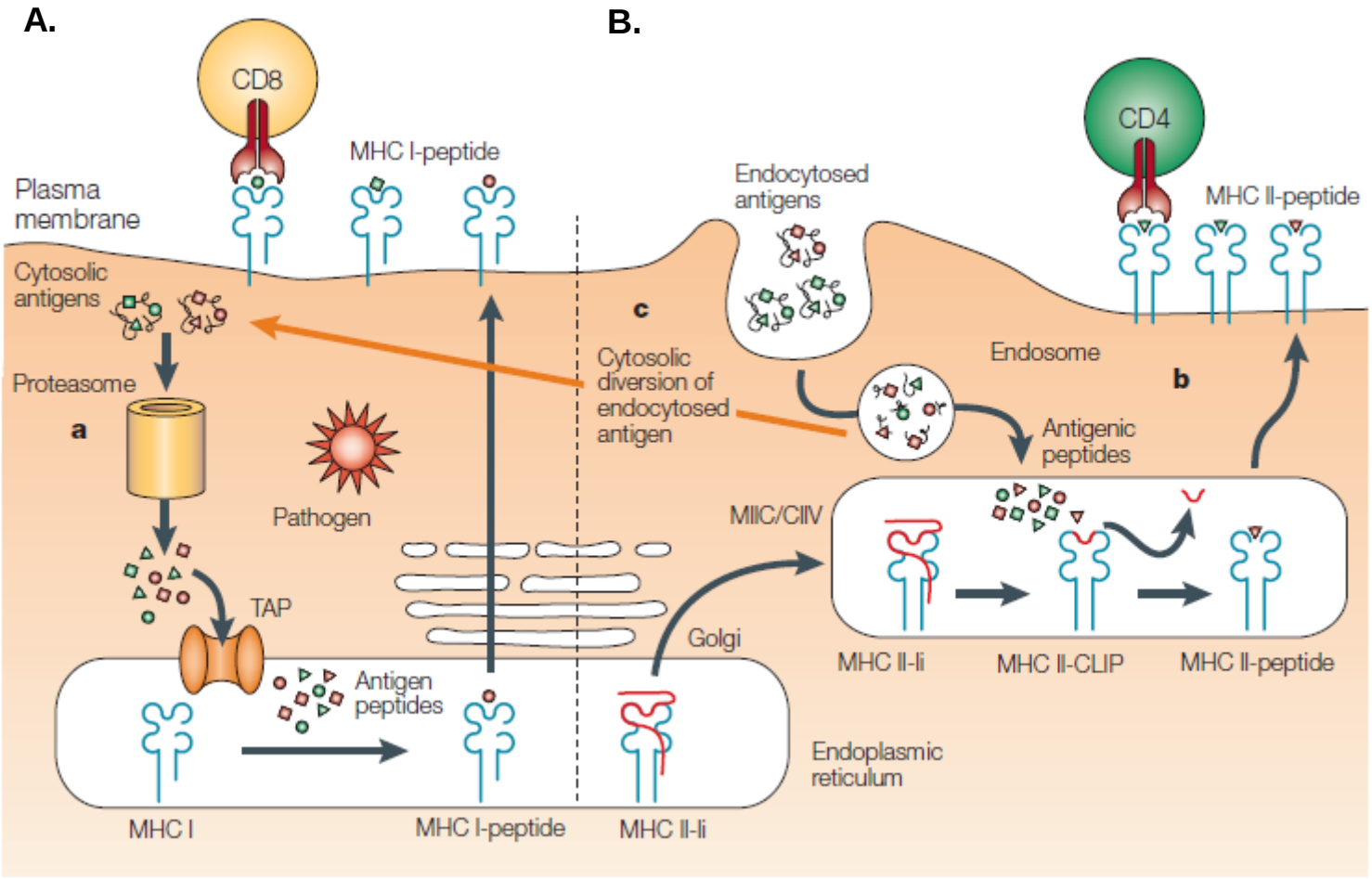


Figure 1.3. Schematic diagram of antigen presentation pathways by **A.** MHC class I and **B.** class II molecules (Heath & Carbone, 2001).

1.3.3. Non-classical MHC molecules

Both the class I and class II region of the MHC are known to encode non-classical MHC molecules. Although structurally similar, non-classical MHC molecules are distinguishable from their classical counterparts by limited polymorphism and cell surface expression. MHC class I non-classical genes in human include *HLA-E, -F*, and *-G* and in mice the non-classical genes were termed *Qa* and *TL*. HLA-E molecules in human, and Qa-1^b molecules in mice interact with the NK cell receptors NKG2A and NKG2C and modulate NK cell function (Braud et al., 1998; Vance et al., 1998). HLA-G is expressed at the maternal-fetal interface in humans and is believed to play a role in the immune tolerance of pregnancy (reviewed in Braud et al., 1999). The murine functional homologue of HLA-G, Qa-2, is encoded by the *Q7* and *Q9* genes and is expressed during pre-implantation development (Comiskey et al., 2003; Warner et al., 1987). Whilst *HLA-E* and *-G* have been extensively studied, the function of *HLA-F* is yet to be elucidated.

Also encoded within the class I region are the *MIC* (MHC class I chain related) loci and *HFE* gene. The expression of MIC proteins on the cell surface without binding to $ß_2$ microglobulin and independently of TAP indicates this protein is not involved in peptide binding (Groh et al., 1996). This protein is heat stress inducible, recognised by T cells which express gamma-delta T cell receptors and is believed to be integral in maintenance of healthy gut epithelium in human (Groh et al., 1998). MIC has not been found in mice, but two molecules encoded on the murine MHC T locus, T10 and T22, display similar characteristics to human MIC molecules (Crowley et al., 2000). Also, a family of MHC class I-like genes called *Mill* which are closely related to the *MIC* family has been located close to the mouse leukocyte receptor complex on chromosome 7 (Kasahara et al., 2002). The *HFE* locus encodes a protein which associates with $ß_2$ microglobulin, predominantly expressed in the small intestinal tract and is involved in iron metabolism but does not appear to have an immunological role (Feder et al., 1996; Lebrón et al., 1998; Waheed et al., 1999).

Genes encoding non-classical class II molecules have also been reported. In humans, the *HLA-DM* and *HLA-DO* genes encode proteins structurally similar to classical class II molecules, but these two molecules are restricted in their expression. HLA-DM is retained intracellularly within the MHC class II compartments (Sanderson et al., 1994). Here, HLA-DM catalyses the removal of CLIP and the subsequent loading of peptides to the class II molecule (Denzin & Cresswell., 1995). It has also been reported that HLA-DM stabilises the empty class II molecule prior to peptide loading and edits the peptide to ensure only long-lived peptides are presented at the cell surface (Denzin et al., 1996; Sloan et al., 1995). The murine equivalent of HLA-DM, known as H2-M, is also involved in the processing of antigens for $CD4^+$ T cells (reviewed in Alfonso & Karlsson, 2000). HLA-DO is expressed mainly in B cells and is able to form stable complexes with HLA-DM, thus modulating its function (Kropshofer et al., 1998). Examination the function of the HLA-DO homologue in mice, H2-O, shows that these non-classical molecules may promote the presentation of antigens internalised by the B-cell receptor on the surface of B cells (Alfonso et al., 2003a; Alfonso et al., 2003b).

1.4. Genomic organisation of the MHC

Current knowledge of MHC genetics is based primarily on studies of the human MHC or human leukocyte antigen (HLA) and the mouse MHC (H-2 genes). From the centromere to the telomere, the MHC genes are arranged in the order of class II, class III and class I. In human, the MHC is known to span 4 megabases on chromosome 6p and contains over 120 genes, 40 % of which are involved in the immune response (MHC Sequencing Consortium, 1999). Figure 1.4 shows the basic genomic organisation of the human MHC on chromosome 6.

The *HLA-DP*, *-DQ* and *-DR* loci are found as pairs within the class II region and encode α- and ß-chains of the heterodimeric molecule. Of these genes, *-DR* is the most polymorphic with over 900 alleles known. Also present within the class

II region are the non-classical *HLA-DM* and *-DO* loci which are involved in antigen recruitment by class II molecules, as well as the *LMP* and *TAP* loci. The *LMP* loci encode constituents of the proteosome complex, and the TAP proteins are involved in antigen peptide transport from the cytoplasm across the ER membrane, before being loaded onto class I molecules via the tapasin protein (Rhodes & Trowsdale, 1999).

The human class I region hosts three classical genes (*HLA-A, -B* and *-C*) encoding the heavy chain of antigen-presenting molecules, together with homologous non-classical genes or class Ib genes (*HLA-E, -F* and *-G*) and a number of pseudogenes. Physical mapping studies have identified and characterised 12 pseudogenes, four of which are full-length, with the remainder being abbreviated genes or single intron-exon fragments (Geraghty et al. 1992). Phylogenetic analyses suggest that the human class I pseudogenes originated from duplication of class Ia genes (Hughes, 1995).

Between these two regions are the class III region genes which encode molecules involved in innate immune system functions e.g. inflammation and regulation of immunity using cytokines. Components of the complement system, C2 and C4, are also encoded within the class III region. However, as with the class I and class II regions, many genes of the class III region have no clear link to either the adaptive or innate immune response (Klein & Sato, 1998).

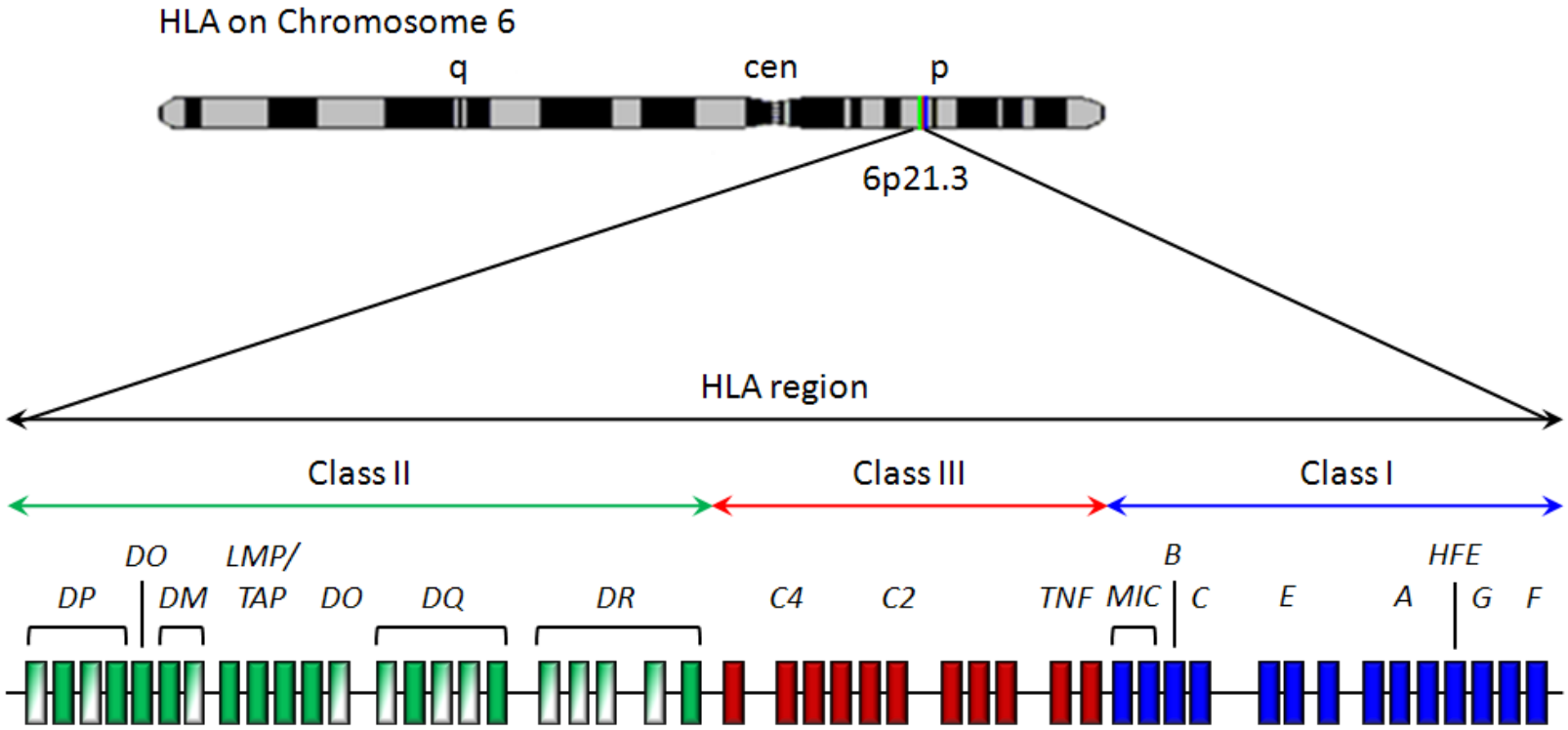


Figure 1.4. Basic genomic organisation of the human MHC on chromosome 6 (adapted from the MHC Sequencing Consortium, 1999). Solid fill blocks represent genes encoding the α-strand and gradient filled blocks represent genes encoding the ß-strand.

1.5. Evolution of the MHC

Genomic characterisation of MHC regions in a growing number of species of jawed vertebrates has contributed to our understanding of how such a complex and gene dense region, a major function of which is dedicated to immunity, has evolved. Analysis of the human genome sequence has shown that there are at least three regions paralogous to the MHC region (chromosome 6) i.e. regions that have arisen through gene duplication and translocation, on chromosomes 1, 9 and 19 (Kasahara et al., 1996; Kasahara, 1999; Katsanis et al., 1996). It has been shown that at least 80 gene families share copies on at least two of these four paralogous regions. Two hypotheses for the evolution of these paralogous regions exist. The first hypothesis proposed that these paralogous regions evolved from the large-scale block duplication from a common ancestor chromosomal segment, with two rounds of duplication required to give rise to four paralogous regions (Abi-Rached et al., 1999; Kasahara et al., 1996; Kasahara, 1999). The second hypothesis argues that selective pressure exerted on independently duplicated genes has resulted in the clustering of these genes into the functionally distinct MHC region (Hughes, 1998). The mechanisms required to facilitate sorting of such large amounts of genetic material into four paralogous regions appear inconceivable and so the hypothesis of en-bloc duplication is more widely accepted.

The emergence of the MHC region in the ancestral jawed vertebrates suggests that the duplications contributing to the evolution of MHC paralogous regions occurred prior to the divergence of this group approximately 500 million years ago. It has been proposed that two rounds of gene duplication were required to give rise to the paralagons seen in the jawed vertebrates, with one round of duplication prior to the emergence of vertebrates and the second pre-dating the divergence of the jawed vertebrates from jawless fishes (reviewed in Kasahara et al., 2004). The hypothesised timings of duplications contributing to the MHC paralagons appear to synchronise with the proposed whole genome duplication

events in early vertebrate evolution as proposed from karyotype and *HOX* gene analysis (as reviewed in Ohta, 1999).

Further information as to the process of MHC evolution was provided by the comparative analysis of the mouse and human genome sequences. This analysis showed that the class II and class III regions of the two species appeared orthologous, whilst the presence of non-MHC genes in the class I region, located in homologous regions across species, provided a framework within which major perturbations are permitted (Amadou, 1999). These non-MHC 'framework' or anchor genes demarcate evolutionary blocks which are well conserved in mammalian species (reviewed in Kulski et al., 2002).

1.6. MHC class I haplotypes in mammalian species

The collection of all MHC genes on a single chromosome is termed a haplotype, and the MHC genes from both chromosomes constitute the MHC genotype of the individual (Warner et al., 1988). Thus in humans each haplotype contains three classical MHC class I genes. The molecular characterisation of the MHC in an increasing number of species has shown that the consistent expression of three class I genes on a class I haplotype is relatively rare and that the expression of a variable number of class I loci is a more common class I haplotype configuration. Phylogenetic analysis, haplotype and transcript expression data in two heterozygous rams and their progeny, provided evidence that sheep have four classical class I genes, three of which are expressed in varying combinations on any class I haplotype (Miltiadou et al., 2005). Variable expression of up to three classical class I genes (*SLA-1*, *SLA-2* and *SLA-3*) dependent upon haplotype has been demonstrated in pig, with some haplotypes expressing duplicate *SLA-1* loci, or lacking expression of either *SLA-1* or *SLA-3* as a result of inter-locus recombination (Ho et al., 2009; Renard et al., 2001; Tanaka-Matsuda et al., 2009). Horse class I haplotypes have also been shown to express at least three of seven putative class I loci (Ellis et al., 1995; Tallmadge et al., 2010).

The most extreme example of class I haplotype variation is seen in the rhesus macaque. The macaque *Mamu-A* and *Mamu-B* loci are the orthologues of *HLA-A* and *HLA-B*, but an *HLA-C* orthologue is lacking in macaque species (Otting et al., 2005). Rhesus macaque class I haplotypes have been shown to consistently express more than one copy of a *Mamu-A* gene, with consistent high expression of the polymorphic *Mamu-A1* locus found in varying combinations with one or two genes designated *Mamu-A2, -A3, -A4, -A5, -A6* and *–A7* which exhibit more limited polymorphism and lower levels of expression (Otting et al., 2005; Otting et al., 2007). Defining the number of *Mamu-B* genes has proved more difficult, however it is estimated that there are approximately fourteen *Mamu-B-like* loci on the chromosome, of which up to three are transcribed (Otting et al., 2005; Otting et al., 2008). Variable gene content and gene expression contribute to the unparalleled class I haplotype diversity in rhesus macaques.

1.7. Polymorphism and generation of diversity within the MHC

Genes encoding MHC proteins are among the most polymorphic genes found in jawed vertebrates, with some loci recording heterozygosity values as high as 80% (Hughes & Yeager, 1998). There are currently 1001 *HLA-A*, 1605 *HLA-B* and 690 *HLA-C* alleles described (http://hla.alleles.org/nomenclature/stats.html) and the discovery of more alleles is likely to continue. It has been demonstrated that MHC polymorphism in humans is generated by point mutations and intra-locus recombination and this variation can be mapped to specific regions of the peptide binding cleft of MHC class I and class II molecules (Trowsdale and Parham 2004). Polygeny, polymorphism and heterozygosity at MHC loci give rise to huge variation, enabling the presentation of a wider range of antigenic peptides (Clarke & Kirby, 1966; Doherty & Zinkernagel, 1975).

The hypothesis that MHC polymorphism is maintained through some form of positive selection is supported by four independent lines of evidence: (a) allele frequencies do not fit the neutral expectation; (b) synonymous nucleotide

substitution is far exceeded by non-synonymous nucleotide substitution within the peptide binding cleft of MHC molecules; (c) polymorphisms have been maintained for long periods of time; and (d) homogenisation of intron sequences relative to exon sequences over evolutionary time suggests balancing selective forces are acting to maintain diversity (Hughes & Yeager, 1998). Two main types of mechanism are proposed to explain the high diversity at MHC loci: disease-based and reproductive mechanisms, although these two mechanisms are not mutually exclusive.

1.7.1. Pathogen-mediated selection mechanisms

The integral role of MHC molecules in antigen presentation during the immune response underpins the most widely presented theory that pathogen-mediated selection drives diversity at MHC loci. However, few convincing examples have been found to support pathogen-mediated selection at the MHC. In 1991, Hill et al. observed statistically high frequencies of the *HLA-B*5301* allele in West African populations and reported that expression of this allele conferred increased resistance against severe malaria. However, the actual frequencies, despite being statistically significant, were quite low and the observed association was not found in other African regions. The differential immunity to Marek's disease virus (MDV) in chickens is linked to specific MHC haplotypes with the B21 haplotype conferring resistance whilst the B19 haplotype is the most susceptible. However, the chicken MHC (also known as the 'minimal essential MHC') is considerably smaller and simpler than mammalian MHCs with highly conserved haplotypes (Kaufman et al., 1999; Kaufman, 2000), thus making mapping disease resistance to MHC loci an easier task. Linkage disequilibrium, the large number of genes encoded in the MHC region and interaction between MHC loci makes it difficult to assign disease susceptibility to an individual gene or locus (Goldsworthy et al., 2000). Diminished exposure to pathogenic selection compared to terrestrial mammals was hypothesised to cause the low levels of MHC diversity observed in the southern elephant seal and some whale species (Slade, 1992; Trowsdale et al., 1989). However, a number of studies have since

shown MHC loci in cetaceans have undergone positive selection to give rise to levels of MHC diversity comparable to that of terrestrial mammals and the marine environment does not provide diminished exposure to pathogens (Du et al., 2010; Purcell et al., 2008; Van Bressem et al., 1999).

1.7.1.1. Heterozygote advantage

Pathogen-mediated selection is hypothesised to operate through a balancing selection mechanism. The theory of heterozygote advantage, also known as the overdominance hypothesis, proposes that individuals heterozygous at MHC loci are able to present a wider range of peptides then either homozygote, thus conferring a fitness advantage (Doherty & Zinkernagel., 1975). Data supporting heterozygote advantage in the MHC are relatively scarce; examples include disease progression in HIV-infected patients (Carrington et al. 1999), mastitis susceptibility in Holstein cattle (Takeshima et al., 2008) and risk of parasitism in naturally-infected lambs (Stear et al., 2005). A study conducted by Potts and Wakeland (1993) indicated that a small heterozygote fitness advantage can have a huge impact upon polymorphism; in a population of 10^5 individuals with a mutation rate of 10^{-6} a selective advantage of heterozygotes over homozygotes of only 1 % is sufficient to maintain over 20 alleles, whereas under neutral conditions the same population would maintain fewer than 5 alleles. Whilst emphasising the proficiency of heterozygote advantage in maintaining polymorphism, this study also highlights that the selective forces involved may be so weak as to be undetectable in experimental contexts. In contrast, it has also been argued that heterozygote advantage alone may be insufficient to maintain MHC polymorphism and additional frequency-dependent selection mechanisms could contribute to the variability observed (Apanius et al., 1997; Borghans et al., 2004; De Boer et al., 2004).

1.7.1.2. Negative frequency-dependent selection

Under the negative frequency-dependent selection or rare allele advantage model of selection, the fitness values associated with a haplotype are not fixed, but fluctuate in proportion with haplotype frequency (Clarke & Bodmer., 1966). It is proposed that pathogen antigenicity is selected in order to evade immune detection by the most common MHC haplotypes, thus favouring rare haplotypes and decreasing the fitness of common haplotypes. The fitness advantage of the rare haplotypes results in their positive selection and drives an increase in frequency, whilst the lower fitness of the common haplotype will lead to a decrease in frequency. As the previously rare haplotype increases in frequency, pathogens become adapted to exploit the now common haplotype, resulting in a decline in its fitness. The antagonistic co-evolution of pathogen and host results in a cycling of fitness values of different genotypes in both populations, and the maintenance of high genetic diversity. As with heterozygote advantage, frequency-dependent selection is hard to detect. Examining the equations used to describe changes in allele frequency over time for both heterozygote advantage and frequency-dependent selection reveals strong similarity, making it difficult to identify the selection model responsible when alleles are distributed in equilibrium in a population (Denniston & Crow, 1990; Takahata & Nei, 1990). However, it has been highlighted that the dynamics of frequency-dependent selection differ from heterozygote advantage, and under certain conditions can be more effective at maintaining polymorphism (Hedrick, 1972; Hedrick & Kim, 2000).

1.7.1.3. Variable selection throughout time and space

The least examined mechanism of pathogen-mediated selection hypothesised to drive polymorphism at MHC loci is that of variable selection throughout time and space. The frequency of a given pathogen/pathogenic strain will alter over time and space, thus resulting in different alleles being selected at different times.

Unlike frequency-dependent selection, where the fitness values for a given allele fluctuate according to its frequency, the fitness values vary according to pathogen frequency i.e. as a pathogen increases in frequency, the selective advantage and fitness value for alleles conferring resistance will increase, whilst the fitness value of susceptible alleles will decrease (Hedrick et al., 1987). Whilst the resistant alleles have higher fitness values in this instance, heterozygous individuals in populations exposed to a variety of pathogens at one time may benefit from an overall selective advantage (Hedrick, 2002).

1.7.2. Reproductive mechanisms and selection

It has also been hypothesised that reproductive mechanisms impact upon MHC diversity, however the line between reproductive and pathogen-mediated selective mechanisms remains blurred. Reproductive mechanisms influencing MHC diversity occur at both the pre- and post-mating stages i.e. the effects of genotype on mate choice and whether the pregnancy will be successfully carried to term. It is believed that reproductive mechanisms serve to avoid inbreeding, increase diversity and so give rise to offspring able to combat infection from a wider range of pathogens.

1.7.2.1. Dissassortative mating

A number of studies have demonstrated preferential MHC-based dissassortative mating as a mechanism of maintaining diversity and avoiding inbreeding in a variety of taxa. Experiments using both wild and laboratory mouse populations demonstrated the preference for choosing a mate of a different MHC genotype (Egid & Brown, 1989; Potts et al., 1991; Yamazaki et al., 1976). Evidence of dissassortative mating was also detected in the Hutterites, a religiously isolated human population, in which couples shared fewer alleles across 5 MHC loci than would be expected under random mating (Génin et al., 2000; Ober et al., 1997).

A small number of studies have been able to link mate choice decisions with physical traits found to be correlated with MHC genotype. In ring-necked pheasants it has been shown that female mate preference is influenced by spur length, with which both viability and MHC genotype were correlated (Von Schantz et al., 1996 & 1997). Similarly, antler size in white-tailed deer was shown to provide an indirect signal to prospective mates as to fitness in terms of parasitic load and MHC genotype (Ditchkoff et al., 2001). The use of physical cues associated with MHC genotype suggests that MHC-based mate choice favours individuals with greater fitness as prospective parents, and also matings that produce heterozygous offspring.

1.7.2.2. Odour-based mate choice

The proximity of olfactory receptor genes to the MHC has instigated a number of studies investigating odour-based mate choice in response to MHC type in mice populations (Yamazaki et al., 1976; Yamazaki et al., 1979). The results of odour-based mate choice studies using human test subjects are highly controversial with limited reproducibility and as a result, compelling evidence in support of odour-based mate choice is lacking (Roberts et al., 2008; Thornhill et al., 2003; Wedekind et al., 1995; Wedekind & Furi, 1997). However, when examining olfactory cues in relation to MHC and kin recognition in fish, it has been found that Arctic char are able to discriminate kin from non-kin in fluvarium experiments, preferring water scented from siblings (Olsen et al., 1998). However, the ability to discriminate between kin and non-kin by scent is not apparent in fish that have been isolated from siblings since hatching (Olsen et al., 2002). This corroborates findings in mice, in which MHC-based mate preferences are believed to be part of a familial imprinting process (Beauchamp et al., 1988; Penn & Potts, 1998; Yamazaki et al., 1988). The ability to recognise kin in this manner enables the avoidance of an inbreeding depression.

1.7.2.3. Post-mating mechanisms

Evidence for post-mating mechanisms influencing MHC diversity has been found in human populations where it has been reported, somewhat controversially, that when parents are MHC-matched, there is a higher rate of spontaneous abortion in the early stages of pregnancy (Komlos et al., 1977; Schacter et al., 1979). Significantly higher rates of pregnancy wastage were also observed in pigtailed macaques when mating pairs shared MHC alleles (Knapp et al., 1996). Further evidence from the Hutterites showed that couples sharing MHC haplotypes had longer intervals between the births of their offspring implicating foetal wastage as a possible cause (Ober, 1995). However, it is important to note that maternal-foetal interactions cannot account for high MHC diversity in oviparous taxa (Hedrick, 1999; Piertney & Oliver, 2006).

1.8. MHC diversity within populations

Diversity within the class I and class II genes enables a wider range of peptides to be bound and recognised. This variation contributes to observed differences in resistance and susceptibility to infectious and parasitic disease (Carrington et al., 1999; Hendel et al., 1999; Stear et al., 2005; Takeshima et al., 2008). Examination of MHC diversity in wild animal populations is of interest in terms of population fitness in the face of immune challenges with populations that have undergone a demographic constriction or population bottleneck of particular interest in terms of diversity. The African cheetah notoriously experienced a population bottleneck approximately 10,000 years ago and has since been subject to high levels of inbreeding. Studies based on allozymes and two-dimensional gel electrophoresis highlighted the relative genetic uniformity of cheetah MHC class I (Yuhki & O'Brien 1990). O'Brien et al., (1985) had previously demonstrated the lack of diversity in cheetah MHC loci using reciprocal skin grafts between 12 unrelated and 2 sibling cheetahs, all of which failed to be

rejected by the recipient. This observed lack of diversity at the cheetah MHC is considered to be responsible for the extreme susceptibility of this species to viral infections (Yuhki & O'Brien, 1990). Similarly, Aldridge et al. (2006) reported a low degree of sequence variation in MHC class I gene transcripts in the endangered Hawaiian monk seal population. The application of denaturing gradient gel electrophoresis assays to assess MHC variability in highly polymorphic domains of the putative peptide binding cleft (PBC) also revealed a lack of inter-individual variability. In contrast, a recent study of the San Nicolas fox demonstrated that having undergone a bottleneck approximately 10-20 generations ago, reducing the population to <10 individuals, high levels of heterozygosity at MHC were maintained despite monomorphism at neutral hypervariable markers (Aguilar et al., 2004). This study serves to re-affirm the importance of MHC diversity as an important marker for assessing population fitness.

1.9. Domestic cattle

Cattle represent an economically important livestock species on a global scale. Domesticated approximately 10,000 years ago, cattle have accompanied migratory human populations and have provided a vital source of dietary protein, leather hides and draught power for both agricultural and transport purposes. In 2008, the global cattle population yielded 7 million tonnes of hide, over 60 million tonnes of meat and over 575 million tonnes of whole milk (FAOSTAT, 2010). The growing global population is causing an increased demand for foodstuffs providing the necessary dietary protein, exemplified by the importation of whole milk powder to China this year, which is already estimated to be approximately sixty percent higher than during the same period last year (FAO Global Food Outlook, 2010). With milk and beef products ranked in the top five of forty five agricultural important commodities in the developing world, cattle represent a major cash commodity.

1.10. Origin of modern day domestic cattle

The exact origin of modern day domesticated cattle is controversial (Bradley et al., 1996). Modern day cattle comprise two principal types, zebu (*Bos indicus*) and taurine (*Bos taurus*), both derived from the domestication of the wild aurochsen (*Bos primigenius*) during the Neolithic period (Beja-Pereira et al., 2006). It was originally hypothesised that both types of cattle were derived from a single domestication event approximately 10,000 years ago. However, mitochondrial DNA (mtDNA) analysis revealed a divergence between the zebu and taurine cattle breeds, suggesting two independent domestication events in the great Indus valley and Near East, from two genetically distinct populations of auroch (Loftus et al., 1994). Consequently it was hypothesised that European cattle, predominantly found to be of the T3 haplogroup, were derived from expansion of a small cattle population in the Near East approximately 10,000 years ago. Y-specific interspersed multilocus microsatellites, used as patrilinear markers, confirmed the existence of three main sire groups prior to domestication events, with taurine cattle belonging to haplogroups Y1 and Y2 whilst the Y3 haplogroup is exclusive to zebu cattle (Pérez-Pardal et al., 2010). Both Y-specific markers and mtDNA analysis reveal evidence of multiple domestication events, introgression from local auroch populations into domestic herds and introductions from northern Africa, in the domestication of modern day taurine cattle (Achilli et al., 2009; Beja-Pereira et al., 2006; Pérez-Pardal et al., 2010).

1.11. Bovine MHC (*BoLA*)

1.11.1. Overview of the genetic organisation of *BoLA*

The bovine MHC, also known as the bovine leucocyte antigen (*BoLA*), was located to bands q13-23 of cattle chromosome 23 (BTA23) by in situ hybridisation using cDNA from a pig MHC class I gene (Fries et al., 1986). As seen in humans and mice, *BoLA* consists of three gene classes, with class II located centromerically to class III and class I situated closest to the telomere (Sharif et al., 1998). However, unlike humans and mice, the bovine class II region is subdivided into class IIa and class IIb which are separated by a physical recombination distance of 17 centimorgans (cM; Andersson et al., 1988). Gene order comparisons between humans, mice and cattle indicate that the class II division is a result of single large chromosomal inversion (Band et al., 1998). The class IIa region is more extensively studied as it contains the main restriction element *DRB3* and exhibits a high degree of polymorphism compared with many vertebrate species (Takeshima et al., 2003). The class IIb region contains the *LMP2*, *LMP7* and *TAP2* genes together with the non-classicals *DMA, DMB, DOB, DOA, DIB* and *DYA*. Limited studies of the *BoLA* class III region have demonstrated the presence of complement proteins Bf and C4, heat shock protein HSP70, a 21 hydroxylase and tumour necrosis factor- alpha (TNF-α; Andersson et al., 1988; McShane et al., 2001; Skow et al., 1988; Teutsch et al., 1989). Figure 1.5 shows the genetic organisation of the bovine MHC in comparison to *HLA*.

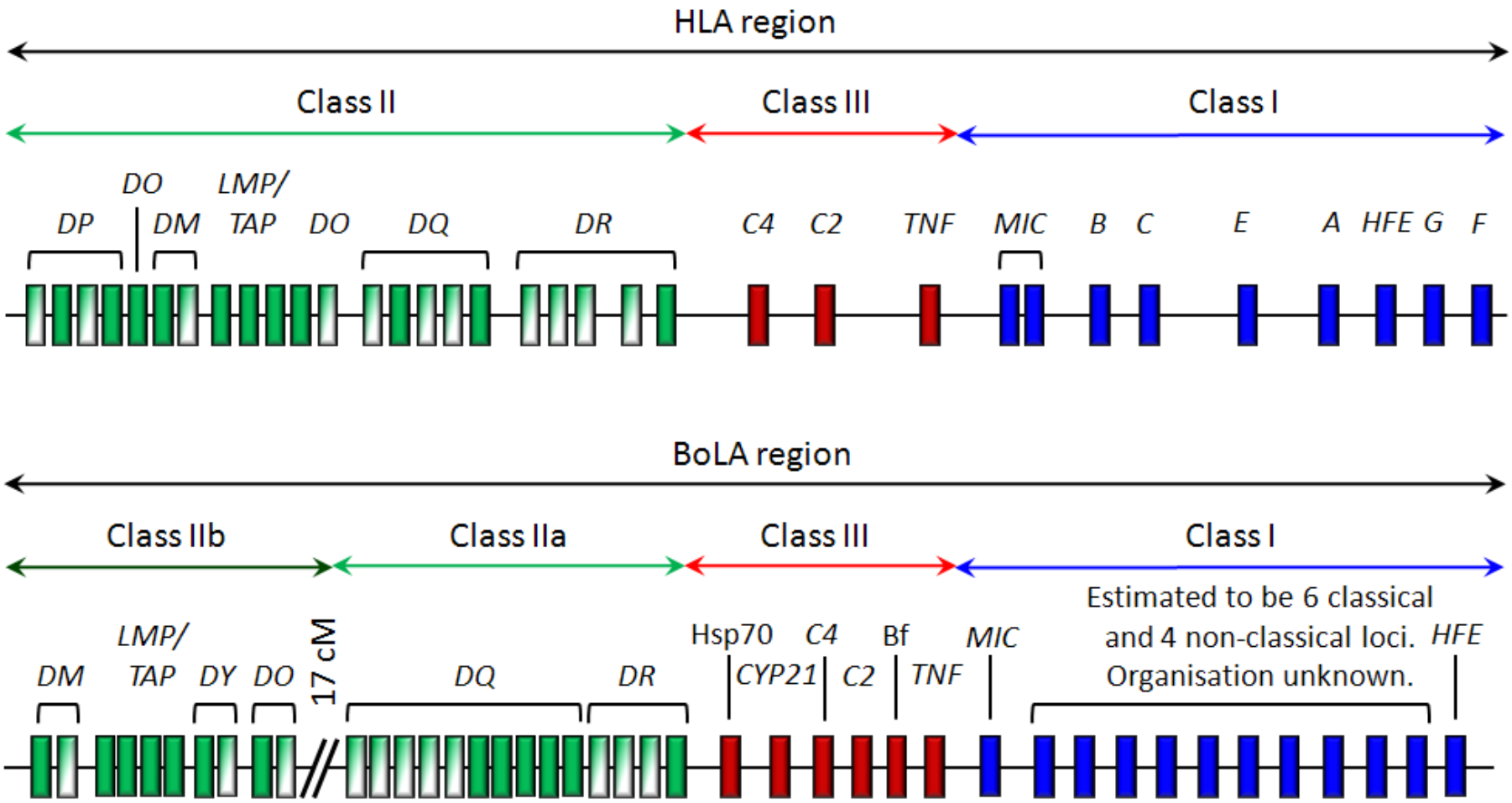


Figure 1.5. Diagrammatic representation comparing the genetic organisation of the bovine MHC (*BoLA)* on *Bos taurus* chromosome 23 with *HLA*. This is a simplistic interpretation of a selection of genes encoded on both *HLA* and *BoLA* and is not drawn to scale.

1.11.2. Characterising the *BoLA* class I region

1.11.2.1. Serological definitions

Characterisation of the class I region in cattle was initially carried out using alloantisera, produced by skin grafting or lymphocyte inoculations (Spooner et al., 1979). A series of International *BoLA* workshops resulted in the identification of 25 distinct specificities, with specificities segregating as alleles or pseudoalleles. The 25 specificities were initially thought to correspond to 25 alleles of a single locus, named *BoLA-A* (Spooner et al., 1979), although evidence for multiple loci encoding class I in cattle existed (Bensaid et al., 1988; Stear et al., 1982). By the fourth International *BoLA* workshop in 1990, the number of known specificities had increased to 50 (Bernoco et al., 1991). Further examination of serological methods showed that the sera were often only detecting a single gene on a haplotype (combination of alleles located on the same region of one chromosome) and that a number of serological specificities were cross reactive e.g. w6 sera is able to recognise three different class I haplotypes. Sera able to detect a number of different haplotypes became known as supertypes, which encompassed a number of subtypes or splits e.g. w6 and w8. It was later demonstrated that supertypic serological specificities indicate either the sharing of epitopes between alleles or the sharing of alleles at a less polymorphic locus (Davies et al., 1994).

1.11.2.2. Cloning cattle class I sequences

With the development of molecular methods, more detailed investigation of the *BoLA-A* locus became possible. Using a probe developed for the study of human MHC, screening of a cDNA library derived from a Hereford B-cell line identified two long clones, BL3-6 and BL3-7. The sequences of these clones appeared to

encode all the features of a molecule involved in antigen presentation (Ennis et al., 1988). Closer examination of the 3' untranslated region (3'UTR) sequences, and a pairwise comparison of the three extracellular heavy chain domains of BL3-6 and BL3-7 provided evidence that the clones were derived from two independent loci.

1.11.2.3. Further evidence for a second class I locus

Toye and colleagues (1990) provided additional evidence for the presence of at least two classical class I loci in cattle using serological methods. Genomic DNA isolated from an animal homozygous for a haplotype encoding both the w10 and KN104 specificities was transfected into mouse L cells, prior to screening with monoclonal antibodies (mAb) specific to either the w10 or KN104 allospecificities. Antibody screening showed two separate populations of cells, one expressing w10, and the other expressing KN104. Using immunoprecipitation to confirm the presence of full length bovine class I molecules, and cytotoxicity assays which showed differential killing according to allospecificity provided further proof that at least two independent class I loci exist in cattle.

It was later demonstrated that w10 and KN104 cDNA clones derived from an animal homozygous for the w10/KN104 specificities, when used as probes in Northern blots on mRNA derived from the above mouse L cell transfectants, were specific to a single population of transfectants (Bensaid et al., 1991). Using field inversion gel electrophoresis (FIGE) Bensaid and colleagues were able to map the w10 and KN104 genes to within 210 kb of one another. One dimensional isoelectric focusing (1D- IEF) of cell lysates derived from cattle typed serologically as w10 produced a banding pattern suggestive of at least two loci (Joosten et al., 1992).

1.11.2.4. Sequence and protein analysis provides evidence of a third BoLA class I locus

Cytoplasmic and transmembrane domain sequences of MHC class I molecules in other species have been shown to be indicative of locus (Gussow et al., 1987). Using these domains, Ennis et al. were able to hypothesise that clones BL3-6 and BL3-7 were encoded on independent loci. To better assess the number of alleles and loci expressed in cattle, Ellis et al. (1992) carried out analysis of the cytoplasmic and transmembrane domains of cattle class I molecules. It was shown that class I cDNA sequences derived from *Theileria annulata*-infected Holstein Friesian cell lines could be divided into two main groups according to the length of the expressed transmembrane domains i.e. either 35 or 37 amino acids in length, which had been previously reported. The published BL3-6 and w10 sequences (Ennis et al., 1988; Bensaid et al., 1991) each have 37 amino acid transmembrane domains and are assigned to the *'A'* locus. The BL3-7 sequence (Ennis et al., 1988), with a transmembrane domain of 35 amino acids in length, is assigned to the *'B'* locus. However, the 36 amino acid transmembrane domain of the KN104 sequence, with the 'missing' residue in a different position to those 'missing' in the putative *'B'* locus sequences, complicated the supposition of two independent loci. It was hypothesised that the KN104 sequence could represent a more divergent allele of the *'B'* locus, or a third *'C'* locus.

Further support of a third *'C'* locus was provided by the phylogenetic analysis of artiodactyl class I sequences (Garber et al., 1993). Once again, a subdivision into two groups was observed, but in this instance the KN104 sequence was shown to be intermediate of these two groups. With prior evidence that KN104 is encoded on a separate gene (Bensaid et al., 1991), the hypothesis that cattle have three class I loci was becoming more plausible. The application of reverse-transcriptase polymerase chain reaction (RT-PCR) and rapid amplification of cDNA ends (RACE) provided conclusive evidence of at least three transcribed class I loci in cattle (Garber et al.,1994). The expression of three loci was also confirmed at the protein level using 1D-IEF (Al-Murrani et al., 1994).

1.11.2.5. Haplotype analysis and gene expression analysis

A series of studies were conducted by Ellis et al., to further investigate the number of loci and alleles expressed on different class I haplotypes in cattle. A selective breeding strategy to produce homozygous class I cattle was initiated at the Institute of Animal Health, enabling molecular dissection of *BoLA* haplotype structures in *Bos taurus* cattle. A combination of cDNA cloning, sequence analysis, and expression studies were used to characterise the A10/KN104, A11, A14, A18 and A31 class I haplotypes over the course of these studies. The A10/KN104, A11 and A31 haplotypes were each shown to express two genes, whereas the A18 haplotype expressed a single gene. The characterisation of the A14 haplotype, expressing three genes, provides further conclusive evidence that cattle have at least three class I loci.

As a result of locus-specific characteristics, the class I coding sequences of humans and primates can be assigned relatively easily to a locus and phylogenetic analysis shows segregation into distinct groups corresponding to loci. These locus-specific characteristics are largely lacking in cattle, and phylogenetic analyses shown clustering of alleles into six weakly defined groups which are believed to correspond with six independent loci (Ellis et al., 1999; Ellis et al., 2005; Holmes et al. 2003). Comparing the phylogenetic analyses with fully characterised haplotypes shows that alleles expressed on a single haplotype fall into separate groups suggesting they are products of separate loci.

Although cattle are believed to have six loci, no single haplotype has been reported on which all six of these loci are expressed. Instead, cattle class I haplotypes have been shown to express between one and three genes in certain combinations e.g. genes 1 and 2 and genes 2 and 3 are often expressed together. Gene 1 is never expressed in combination with either gene 3 or 6, which may suggest that these genes are not distinct. However phylogenetic evidence shows no distinct relationship between these alleles at these putative loci (Holmes et

al., 2003) and the intron sequences are sufficiently divergent to facilitate gene-specific amplification (Birch et al., 2006).

1.11.2.6. Mapping the bovine MHC class I region

The culmination of data from mapping studies and the publication of the bovine genome sequence has indicated that the cattle class I region is larger in size than that of the pig, and has a greater number of genes (Ellis et al., 2005; http://www.hgsc.bcm.tmc.edu/projects/bovine/). Despite the publication of the Bovine genome sequence, a deficit of information regarding the organisation and exact number of class I loci still exists because the data used for the production of the sequence were derived from two related animals, and the sequence contains only three classical class I genes (The Bovine Genome Sequencing and Analysis Consortium, 2009). In addition, the two bovine genome assemblies (Btau_3.1 and 4.0) of the bovine MHC region differ substantially from one another, but also from other conventional MHC gene maps. Using a radiation hybrid panel, Brinkmeyer-Langford and colleagues (2009) were able to show that the Btau 4.0 genome build was in greater agreement with the radiation hybrid map than the Btau 3.1 genome build. The high resolution radiation hybrid map of the *BoLA* region enabled a comparison of marker order with other species, and when adjusted for the ruminant-specific large chromosomal inversion in the bovine class II region (see section 1.11.5.), revealed a high degree of conservation of both gene content and gene order between cattle and humans (Brinkmeyer-Langford et al., 2009).

Limited mapping studies have so far revealed that at least 9 genes are clustered within a region of 400 kb. A further 10 genes, which are likely to be pseudogenes or gene fragments, are distributed across the remainder of the class I region (Di Palma et al., 2002; http://www.hgsc.bcm.tmc.edu/projects/bovine/). Separate studies have shown that genes 1, 2 and 4 on the A14 haplotype (Di Palma et al., 2002) and genes 3 and 5 on the A10/KN104 haplotype (Bensaid et al., 1991) are

located within 210 kb of one another. However, the position of genes 1, 2 and 4 in relation to that of genes 3 and 5 is as yet unknown, as is the location of gene 6.

Mapping of the MHC class I genes in an A14 BAC library showed the classical class I genes were ordered gene 1, gene 4 and gene 2 oriented from the centromere to the telomere. An anchor gene, *ZNF173* and a gene called 'Z', now known as *N*04001,* were located at the telomeric end of this 400 kb cluster. A full length allele sequence of the *N*04001* allele has since been submitted to IPD and there are no obvious non-classical sequence signatures. However, this sequence cannot be detected in cDNA isolated from resting lymphocytes and when transfected into COS-7 cells is unable to present antigen from *Theileria parva*, *Theileria annulata* or bovine herpes virus-1 (Di Palma et al., 2001; Prof Ivan Morrison, personal communication). The BAC library mapping study also showed that the bovine non classical gene *NC1-N*00101*, formally called *gene X*, is located centromerically adjacent to gene 1 (Di Palma et al., 2002). The bovine genome sequence also shows that a *NC1* gene is situated close the classical class I genes (Birch et al., 2008a). However, there is a lack of overlap between the published bovine genome sequence and the BAC mapping study due to the DNA source from which the sequences were derived i.e. animals do not carry the same haplotype. Three additional non-classical class I genes, together with a possible four MIC genes have been located approximately 500 kb centromeric to the classical class I cluster in a similar configuration to that seen in the pig (Birch et al., 2008a; Birch et al., 2008b).

1.11.3. *BoLA* class I alleles

To date, approximately 80 *BoLA* class I sequences have been submitted to the IPD database (http://www.ebi.ac.uk/ipd/mhc/bola/index.html; Robinson et al., 2005). Over 60 % of the submitted sequences were derived from, or are known to be expressed in, Holstein-Friesian cattle, whilst knowledge of *BoLA* class I

sequence diversity in other breeds is scarce in comparison. *BoLA* class I nomenclature was agreed in 2004 by the *BoLA* nomenclature committee and is based upon the nomenclature used for *HLA* (Ellis et al., 2006). *HLA* allele names consist of 5-9 digits and are assigned according to the amino acid sequence, with the first 3 digits indicating allele 'group', the next two indicating coding change, the next two signifying non-coding changes and the last four digits, if required, indicate changes within the promoter or introns. As the assignment of alleles to loci still remains problematic in cattle (Ellis, 2004; Davies et al., 2004), the majority of alleles are prefixed *'N'* to indicate that they have yet to be assigned to a definitive locus, and they are numbered in a single series. Alleles differing by fewer than four amino acids in alpha 1 and 2 and no more than four across the remainder of the sequence are assigned to the same 'group'. Those alleles identified as non-classical alleles are prefixed by 'NC' and the non-classical locus number prior to the 'N'. The IPD database also serves to bridge the gap between the previously used serological specificities and allele names defined at the molecular level.

1.11.4. *BoLA* class I haplotypes

MHC class I typing was originally carried out using serological methods (Bernoco et al., 1991). The application of more powerful molecular techniques has shown that the serological specificities in the Holstein-Friesian breed define haplotypes with multiple class I loci. For ease of comparison, these haplotypes have been named after the serological specificities. Currently, sixteen *BoLA* class I haplotypes have been elucidated at the molecular level, including one haplotype from the *Bos indicus* lineage, A10/KN104 (Birch et al., 2006; Ellis et al., 2005; Stear et al., 1990; see Table 1.1). Detailed studies of cattle MHC haplotypes have shown that between one and three classical loci are expressed per haplotype (Ellis et al., 1999). Gene 2 has the highest number of alleles and is expressed on the majority of class I haplotypes. Gene deletion or gene 'silencing' are believed to be responsible for the variety in the number of genes transcribed per haplotype (Ellis et al., 2005). The expression of a variable

number of genes on cattle class I haplotypes mirrors the variability seen in other species e.g. primates, horses and sheep (see section 1.6), and does not reflect the consistent expression of three polymorphic loci as observed in human. The variability in allele combinations expressed on human class I haplotypes is lacking in cattle as cattle class I haplotypes appear to be highly conserved, with the inheritance of class I alleles in specific combinations. It is postulated that the close proximity of at least some of the cattle class I genes to one another is responsible for the preservation of class I haplotypes in this manner. However, the term 'conserved' can only be applied under the constraints of additional caveats i.e. haplotype definitions may only be conserved within breed groups as there is insufficient data to make a definitive statement as to haplotype conservation between breed groups.

Analysis of the non-classical genes on six molecularly defined class I haplotypes demonstrated that the most polymorphic non-classical locus NC1 is present on all haplotypes. Non-classical genes NC2 and NC4 were also detected on all haplotypes, with the exception of the A18 haplotype which lacked NC2 and the W12B haplotype on which the presence of NC4 was variable. NC3 was only detected on two of the six haplotypes analysed, but does not appear to correlate with classical class I gene configuration (Birch et al., 2008a).

HAPLOTYPE	GENE 1	GENE 2	GENE 3	GENE 4	GENE 5	GENE 6
A19		*N*01601*				*N*01401*
A17		*N*00602, N*00802*				*N*01502*
A11		*N*01801*	*N*01701*			
A20		*N*02601*	*N*02701*			
A33		*N*00501*	*N*00401*			
RSCA2[a]		*RSCA2.1*[b]	*N*03801*			
A10		*N*01201*	*N*00201*			
A14	*N*02301*	*N*02501*		*N*02401*		
A15	*N*00901*	*N*02501*		*N*02401*		
W12B	*N*01901*	*N*00801*				
A12 (A30)	*N*02001*	*N*00801*				
A31	*N*02101*	*N*02201*				
A10/KN1O4[c]			*N*00101*		*N*00301*	
A13	*N*03101*					
A18						*N*01301*
A18v						*N*01302*

Table 1.1. Transcribed classical MHC class I genes in different haplotypes. [a] A serological specificity has not been determined for this haplotype. [b] A formal allele name has not been assigned to this allele. [c] This haplotype was reported and analysed by Bensaid et al. (1991). It is possible additional genes are transcribed. Adapted from Birch et al., 2006.

1.11.5. *BoLA* class II region

The class I and class II regions are tightly linked within the bovine MHC. The high degree of homology of cattle class II with other mammalian class II regions has facilitated characterisation of this *BoLA* region. Southern blotting with human class II probes enabled the identification of the cattle *DO, DR* and *DQ* genes, but no evidence of a cattle *DP* homologue could be detected (Andersson & Rask., 1988). As sheep also lack a *DP* homologue, the absence of this gene could be characteristic of the ruminant class II region (Deverson et al., 1991). Using the identified *DO*, *DR* and *DQ* genes together with two additional loci, a family segregation analysis of linkage disequilibrium in the class II was carried out. This analysis indicated the presence of a large molecular distance separating the *DQ* and *DR* genes from the *DO* genes (Andersson et al., 1988). Further mapping enabled the assignment of class II loci to either the class IIa or class IIb regions, shown to be separated by a physical distance of 17 cM (Hess et al., 1999). Table 1.2 presents details of the genes encoded on the BoLA class II region.

The *DR* region, located in the class IIa region, contains the monomorphic *DRA* gene and at least three *DRB* genes. Of the *DRB* genes, only *DRB3* is actively transcribed, expressed at high levels and polymorphic. With over 120 alleles currently reported, the *DRB3* locus is the most polymorphic of the class II loci and as a result its associations with disease resistance and susceptibility have been extensively studied e.g. the *DRB3*1502* allele was significantly associated with lower somatic cell score, an indicator of mastitis infections, in Holstein cattle (Sharif et al., 1998). However, it is important to note that despite a number of papers linking certain MHC class II alleles to mastitis resistance/susceptibility there has been a lack of consistency in these studies (Dietz et al., 1997; Rupp et al., 2007). Each haplotype expresses one *DR* gene pair, *DRA* and *DRB3*.

Situated centromerically from the *DR* region in the class IIa region are the *DQ* genes. Five DQA genes of varying levels of polymorphism and five equally polymorphic DQB genes are known. The number and combination of *DQ* genes on a haplotype is highly variable. Three different haplotype configurations of the *DQA* genes have been reported: a single *DQA1* gene, *DQA1* with *DQA2*, and *DQA2* in combination with *DQA3* (Ballingall et al., 1997). Variation in the number of *DQB* genes expressed has also been reported (Marello et al., 1995). In addition, duplication of the DQ genes, with expression of both copies at the cell surface, has been reported in approximately half of the common class II haplotypes (Glass et al., 2000). Gene duplication coupled with marked variation of the *DQA* and *-B* genes on a given haplotype can give rise to a wide range of *DQ* products expressed at the cell surface. The variability of *DQ* may be a compensatory mechanism to account for the absence of *DP* genes in cattle.

The class IIb region encodes the *LMP* and *TAP* loci along with the non-classicals *DOB, DOA, DIB, DYA, DMA* and *DMB*. These genes have shown limited homology with human class II genes, and the *DYA* and *DIB* genes have no known homologues in human or mice (Stone & Muggli-Cockett., 1990; Vanderpoel et al., 1990). Both *DYA* and *DIB* are reported to be transcribed at low levels in bovine dendritic cells (DCs) and together these genes have been shown to transcribe class II molecules of unique function (Ballingall et al., 2001; Ballingall et al., 2004). This close association of *DIB* with *DYA* has prompted the suggestion that *DIB* be renamed *DYB*. The limited expression of this gene pair and unique polymorphism in the binding cleft of this transcribed molecule categorises these genes as non-classical.

BoLA REGION	GENE	MOLECULAR FEATURES
Class IIa	*BoLA-DRA*	DR-alpha chain. Monomorphic.
	BoLA-DRB1	Pseudogene with two alleles.
	BoLA-DRB2	DR-beta chain expressed at low levels. Monomorphic.
	BoLA-DRB3	DR-beta chain. Expressed at high levels. Highly polymorphic (120 alleles).
	*BoLA-DQA**	DQ-alpha chain. Highly polymorphic (39 alleles of 20 major types).
	*BoLA-DQB**	DQ-beta chain. Highly polymorphic (37 alleles of 20 major types).
Class IIb	*BoLA-DMA*	DM-alpha chain. Expressed. Monomorphic.
	BoLA-DMB	DM- beta chain. Expressed. Monomorphic.
	BoLA-DOA	DO-alpha chain. Monomorphic.
	BoLA-DOB	DO-beta chain. Monomorphic.
	BoLA-DYA	DY-alpha chain. Selectively expressed. Low polymorphism (3 alleles).
	BoLA-DYB	DY-beta chain. Selectively expressed. Monomorphic.

Table 1.2. Details of BoLA class II genes and molecules. * The *DQ* genes are duplicated on some class II haplotypes.

1.12. Breeds of domestic cattle

Originally defined as separate species, the two types of modern day domestic cattle are probably better defined as subspecies. These subspecies are distinguished phenotypically, with *Bos indicus* (zebu) cattle possessing a hump whilst *Bos taurus* (taurine) cattle are humpless, as shown in Figure 1.6. The complicated history of cattle domestication, coupled with the ability of zebu and taurine cattle to produce fertile hybrids, has given rise to genetic variability under the classification of a single species. Selection for performance and type has led to the diversification of cattle into separate breeds, with 482 cattle breeds currently recognised in Europe alone (Food and Agriculture Organization, 2000). Data collected in June 2008 showed that the majority of UK cattle (62.7 %) are dedicated to beef production, whilst a further third are dairy cattle. Closer examination of the UK dairy herd population structure reveals the dominance of Friesian, Holstein Friesian, British Friesian and Holstein breeds, combined under the classification of 'black and white' cattle, which constitute over 90 % of the UK dairy herd and 33 % of the total UK cattle population (DEFRA, 2008; Figure 1.7).

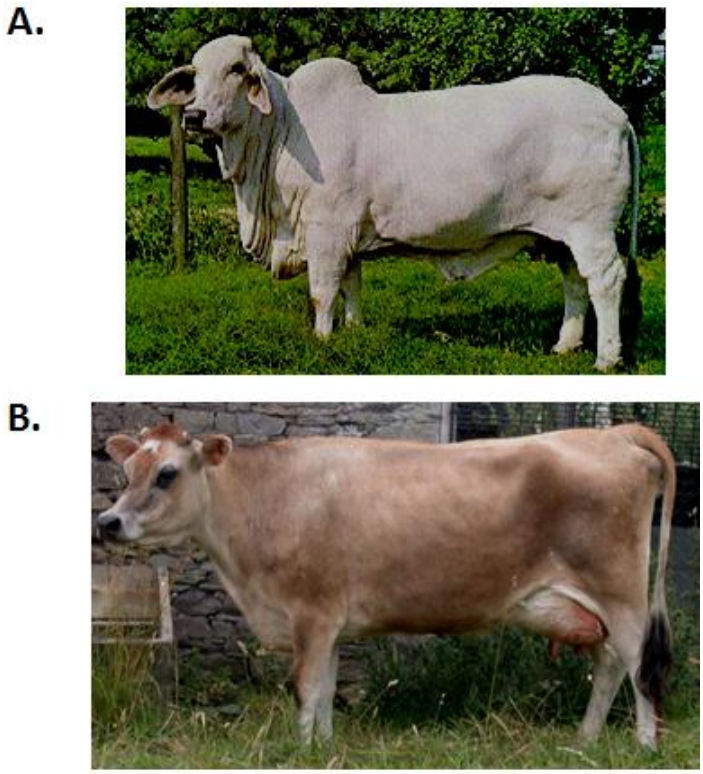


Figure 1.6. Examples of the two types of domestic cattle: **A.** humped zebu cattle (*Bos indicus*) and **B.** humpless taurine cattle (*Bos taurus*).

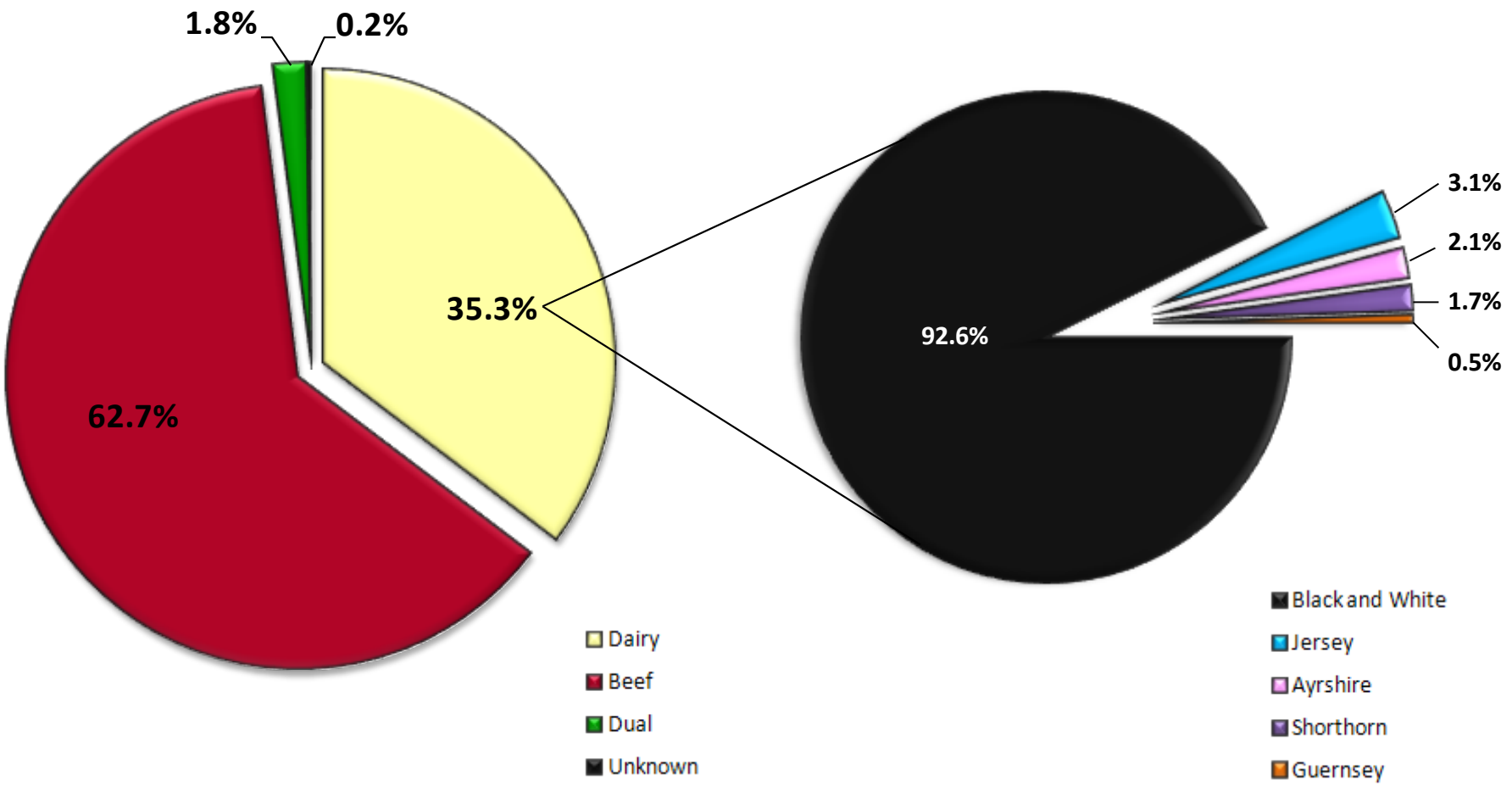


Figure 1.7. Breakdown of UK cattle population in terms of type and breed composition of UK dairy herd.

1.13. Holstein-Friesian dairy cattle history

The popular ‘black and white’ or Holstein-Friesian dairy cattle are believed to have originated in the Netherlands approximately 2000 years ago, having been selectively bred for their high milk yields and their efficient grass conversion. Holstein-Friesian cattle were first exported to North America, Canada and the UK in the 1800s until an endemic foot and mouth disease outbreak on the continent halted all exports in 1892. It was reported that 8,800 head of Holstein-Friesian cattle were exported to Canada and North America during this time. In contrast, exports to the UK were comparably few and insufficient in number to be included in the 1908 census. Since these initial imports, selective breeding programmes either side of the Atlantic have given rise to improved production but also segregation within the Holstein-Friesian breed. The terms Holstein and Friesian, although often used together as the all-encompassing name Holstein-Friesian, are more frequently used independently to designate breed ‘type’. The term Holstein is used to describe cattle of North American descent which are taller and more angular in stature, with higher production levels that require more intensive management systems. Friesian is used to denote animals of European ancestry, which are smaller, hardier cattle more suited to grassland management. However, it is important to note that whilst there is a recognised segregation between Holstein and Friesian, both are registered under a single herdbook, with the percentage lineage input denoted on the pedigree certificate and breed code of each animal.

1.14. Dairy cattle selection and breeding practices

Selective breeding in dairy cattle has focused first and foremost on production traits i.e. milk yield and milk components. Genetic improvement accounts for an estimated 55 % of the 3420 litres of milk, 129 kg of fat and 104 kg protein yield increases in the Holstein Friesian breed average over the past 20 years (Shook, 2006). Whilst the demand for high milk yields has remained stable, the emphasis

on different milk constituents fluctuates in breeding programmes to follow the demands of the economic market more closely e.g. increased emphasis on lowering milk fat content. More recently, selection has focused upon health and welfare in an effort to improve sustainability (Nielsen et al., 2006). Table 1.3 details range of selection traits used in dairy breeding programmes.

Dairy bull selection has been facilitated by extensive progeny-based genetic evaluations for over 60 traits. The data acquired is used to estimate the genetic merit of an animal in the form of estimated breeding values (EBVs) or predicted transmitting abilities (PTAs) published as readily accessible bull proofs. An EBV is an estimate of the total genetic merit of an animal in comparison to the population average, otherwise known as the genetic base. As an animal is only able to pass on half of its genes to its progeny, the PTA value of an animal is half that of its EBV. A diagrammatic representation of these selection indices in a breeding system is shown in Figure 1.8. Estimations of genetic merit are calculated using the best linear unbiased prediction method (BLUP) and mixed model equations (MME) devised by Henderson (1949), which enable the separation of genetic merit from environmental and management influence (Simm, 1998). The genetic base used for calculating EBVs is re-evaluated every five years to accommodate the effect of genetic gain by the population as a whole and improvements/changes in environmental conditions i.e. herd management systems e.g. modern day feeding regimes incorporate specialised concentrates which would not have been available in the 1950s when feeding regimes were largely forage-based. The EBVs of bulls from different proof releases are comparable, providing the values are adjusted to account for the differences in genetic base i.e. an EBV from a bull with a proof release in the 1980s can be compared with a bull with an EBV in the current proof run providing the genetic base used for comparison is the same for both bulls.

Holstein-Friesian breeding programmes worldwide rely on a small number of premium bulls, facilitated by the introduction of artificial insemination (AI) and cryopreservation technologies in the early 20th century (Foote, 1982). The estimated effective population size of Holstein-Friesian cattle is approximately

150 animals despite a global population of over 65 million head of cattle (Hayes et al., 2003). Large-scale breeding trials conducted in the 1970s by Stolzman and colleagues for the Food and Agriculture Organisation aimed to quantify the genetic merit of different ‘strains’ of Holstein Friesian cattle for specific selection traits. Ten countries (Canada, United States, Denmark, United Kingdom, Sweden, (West) Germany, the Netherlands, Poland, Israel, and New Zealand) submitted between 225 and 250 straws of semen from 40 unproven young sires to the trials. Evaluation of daughter performance in these trials indicated that sires from North America, Israel and New Zealand were superior, especially for yield traits, over their western European counterparts, instigating the propagation of North American genetics worldwide (Jasiorowski et al., 1983; Zarnecki et al., 1991; Zarnecki & Stolzman., 1986). Today, Canadian and North American bloodlines represent a dominant genetic input into UK breeding programmes e.g. in 2008 these bloodlines accounted for over 40 % of 250,000 pedigree dairy UK cattle registrations (Alison Maddrell, personal communication, Holstein UK). This partly represents a shift from the dual purpose Friesian to the predominantly milk producing Holstein.

TRAIT	DETAILS
Milk, Fat and Protein Yields	Expected yields of milk, fat and protein during a 305-day lactation in a herd of average management. Estimated breeding values (EBVs) for yield are measured in kilograms compared to the breed average/genetic base and are distributed normally. For milk yield, values range from -1500 to +2900 kg and the average EBV is +100 kg. Milk fat yield values range from -50 to +115 kg, with an average EBV of +10 kg. Milk protein yield EBV values range from -43 to +89 kg, with a breed average of +5 kg. Positive values are desirable.
Daughter Fertility	Measurement of the expected fertility of a bull's daughter across all lactations. Four measures of daughter fertility are used in the calculation: (1) the age at first insemination for virgin heifers, (2) the 56-day non-return rate for virgin heifers, (3) the interval between calving and first insemination for cows, and (4) the 56-day non-return rate for milking cows. The published Daughter Fertility proof based on a relative emphasis of 65 % to increase the cow non-return rate, 25 % to reduce the interval from calving to first service in cows and 10 % to reduce the age at first insemination for virgin heifers. The published proof is expressed as relative breeding value (RBV), with the breed average set to 100, and a range of five standard deviations i.e. the range in bull proof values will be from 115 to 85 with a normal distribution.
Herd Life	Expected number of lactations the daughter is expected to complete, compared to the average bull. This value is independent of production levels. Also expressed as an RBV, with values ranging from 85 (undesirable) to 115 (desirable) with a normal distribution around the breed average of 100.

Somatic Cell Score (SCS)	SCS is a measure of white blood cells in the milk, and is used as a marker of mastitis resistance. Animals susceptible to mastitis infections have a high SCS. The bull proof value estimates the SCS of daughters over the first three lactations, and is recorded on a logarithmic scale. The average bull is 3.00 but values range from 2.50 (desirable) to 3.50 (undesirable) with a normal distribution.
Conformation Traits	Expected relative superiority of first lactation daughters for each type trait. The final conformation score and the four major scorecard traits: legs and feet, mammary system, dairy strength and rump were used in the analysis. Genetic evaluations for conformation are standardized to a common scale for each trait such that the average bull proof is set to 0 and the standard deviation is set to 5. This gives a range of evaluations from –20 to +20 (i.e. four standard deviations away from the breed average) with a normal distribution.
Lifetime Profit Index (LPI)	Multi-component index used to estimate the relative lifetime profit of future daughters based upon their genetic potential for traits encompassed by three components, each with a different weighting; production traits (protein and fat yield; 54 %), health and fertility (SCS, udder depth, milking speed, daughter fertility and lactation persistency; 10 %), and durability (herd life, mammary system, feet and legs, dairy strength; 36 %). Detailed LPI formula for the Holstein breed is shown in appendix. LPI is recorded on a point scale relative to the breed average which is set to zero. It is estimated that for every +200 point difference in LPI above the breed average (0) translates to a profit of $50 more profit per daughter per year, which accumulates each year.

Table 1.3. Description of phenotypic traits and their measurements used for selection in dairy breeding schemes (adapted from Van Doormaal, 1997).

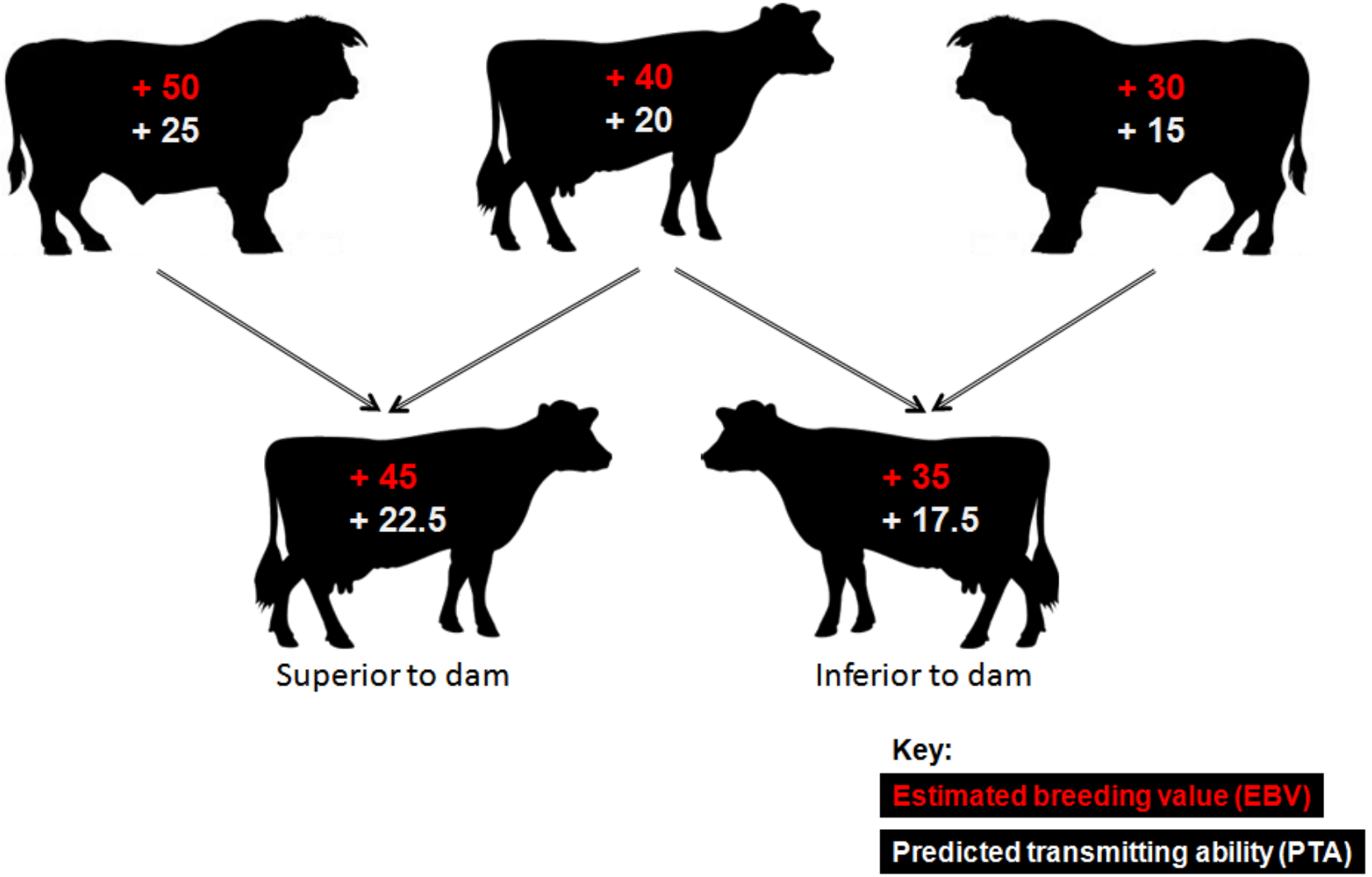


Figure 1.8. Diagrammatic representation of estimated breeding values and predicted transmitting abilities used in a breeding system.

1.14.1. Progeny testing and reproductive technologies

Traditionally, genetic evaluations encompass data from the bull in question, its siblings, relatives and its progeny. To obtain a progeny-tested bull proof, data must be gathered from a minimum number of sources i.e. a minimum number of offspring from a selection of different herd environments. The minimum criteria for an official bull proof are dependent upon the country in which the animal is proven. Table 1.4 presents the minimum criteria for a Canadian bull proof. The progeny testing process, as outlined in Figure 1.9, is both costly and time consuming. It is estimated that the cost of proving a young Holstein sire is approximately $25,000 to $35,000 (Funk, 2006) with a total time commitment of 64 months (Schaeffer, 2006). Approximately 12 % of all progeny tested bulls graduate to become a proven AI sire, resulting in an investment of $250,000 per successful candidate (Funk, 2006).

The rate of genetic improvement in cattle through the paternal line has been facilitated by the intense selection of bulls of high genetic merit by AI and progeny testing (Lohuis, 1995). Multiple ovulation transfer technology (MOET) is the maternal equivalent to AI and encompasses a series of reproductive techniques including the superovulation of a donor female, mating, embryo recovery and transfer of embryos into recipient females (Simm, 1998). MOET schemes enable genetically superior dams to have a greater number of offspring, the shortening of generation intervals and increased selection of the female line to obtain elite donors (Callesen et al., 1996). In practice, the rate of genetic gain in test MOET programmes has been lower than anticipated and the increase in inbreeding was underestimated (Nicholas, 1996). Although the majority of AI companies use embryo flushing from elite females, strict MOET schemes with closed nucleus herds have not been widely adopted (Funk, 2006).

TRAIT	DAUGHTERS	HERDS	RELIABILITY (%)
Production [1,2]	20	10	70
Somatic cell score	If production proof is official		
Type [2]	20	10	60
Herd life	If type proof is official		
Daughter fertility	20	10	45

Note 1: Production daughter minimums refer to daughters past 120 days in milk (DIM).

Note 2: Bull proofs for Production and Type are published only when both sets of criteria are met.

Table 1.4. Criteria for an official bull proof in December 2005. (Adapted from Canadian Dairy Network, 2005).

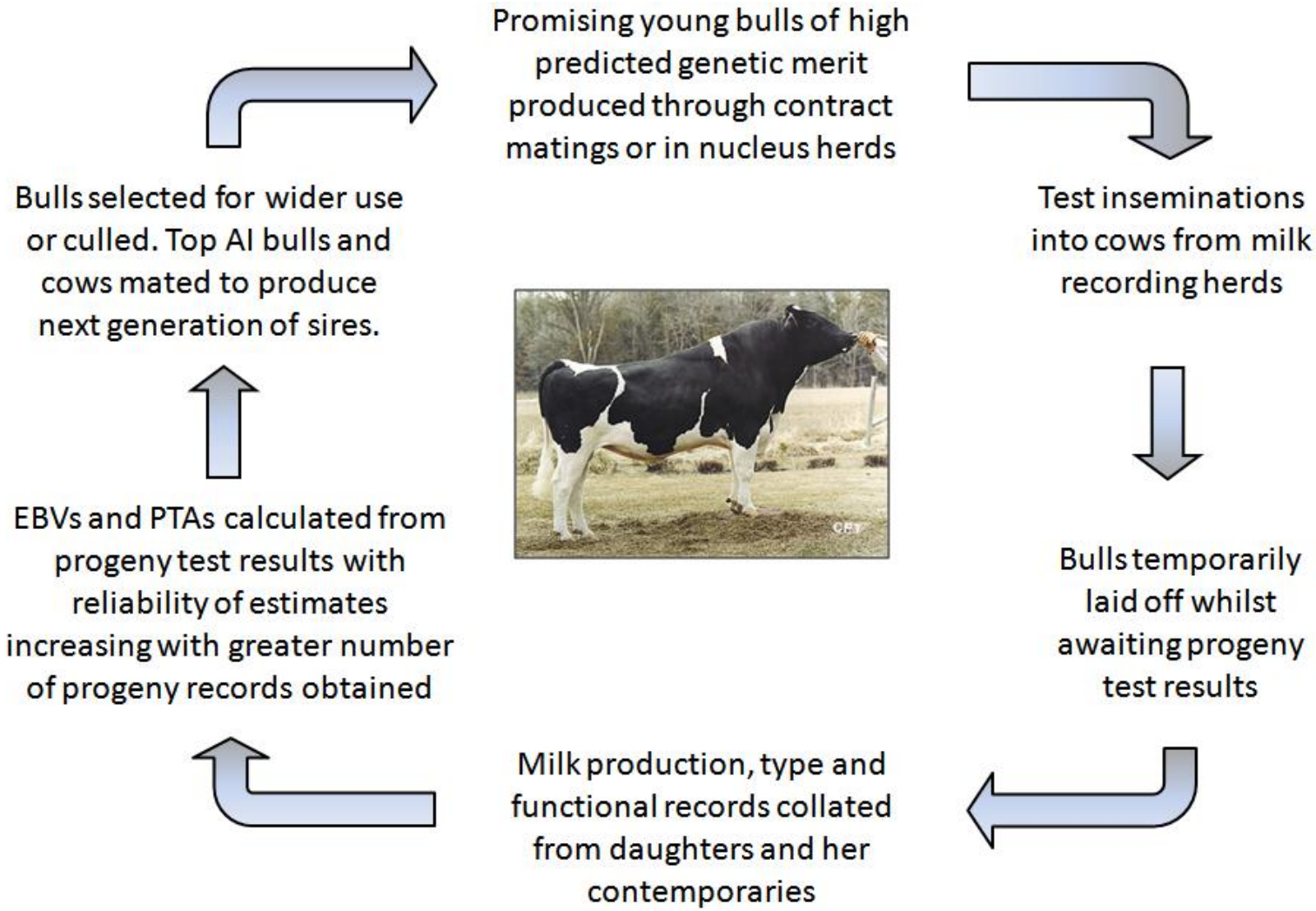


Figure 1.9. Outline of the progeny testing process (adapted from Simm, 1998).

1.14.2. Reliability estimates

Each PTA and EBV has an associated estimate of reliability or accuracy, with values ranging from 10 to 99 %. As an increasing amount of information is gathered from a wider range of sources, the associated reliability of the estimate increases i.e. the estimate of genetic merit becomes more accurate and is less likely to change. Table 1.5 presents the sources of data used to calculate genetic merit and the associated levels of reliability in bull proofs. It is not possible to predict the genetic merit of an animal with 100 % accuracy and therefore no proof is 100 % reliable.

1.14.3. Heritability & genetic correlations

In its simplest terms, the heritability of a trait is the proportion of superiority of the parents for a particular trait that is transmitted to the offspring. Expressed as proportions from 0 to 1 or as percentages, heritabilities vary according to trait and cattle breed. High heritabilities denote that a large proportion of genetic superiority of the parents will be transmitted to the progeny (Simm, 1998). Table 1.6 outlines the differences in heritabilities calculated for each trait for Canadian genetic evaluations.

Genetic and phenotypic correlations exist between selection traits in dairy cattle which, together with varying heritabilities, will influence the response to selective forces in a breeding programme. Correlations between traits can be either positive or negative. An undesirable positive correlation between increased milk yield and increased susceptibility to mastitis has been widely reported, with the average correlation estimated at approximately 0.30 (as reviewed by Heringstad et al., 2000). Using this estimated correlation, the impact of selecting for production alone was shown to increase the incidence of mastitis by 0.02 cases per animal (Strandberg & Shook, 1989). Increasing consideration of health and welfare in selection programmes has resulted in

greater awareness of correlations between production, conformation and functional traits such as longevity and fertility. For example a favourable correlation between body condition score and fertility has been reported with cows with good genetic merit for body condition score taking less time to reach first insemination after calving (Kadarmideen, 2004; Veerkamp et al., 2001; Wall et al., 2003).

% RELIABILITY	CATEGORY	INFORMATION SOURCE
10-29	Extremely low	Young animals with small number of pedigree indices only.
30-40	Very low	Most animals with pedigree indices, with reliable information from parents.
41-55	Low	A few animals with pedigree indices from very reliably-tested ancestors. Cows with one or two lactations.
56-65	Low to Moderate	Bulls with officially published PTAs (minimum 50%). Cows with 3 lactations.
66-75	Moderate	General maximum for most cows (3 lactations +). May be exceeded if a cow has many embryo transfer daughters in the UK. Also includes bulls proven abroad.
76-90	Moderate to High	Bulls with an initial progeny test through AI.
91-98	High	Proven AI bulls with a second crop of daughters in a wide cross section of herds.
99	Very High	Widely used proven AI bulls.

Table 1.5. Sources of information for calculation of genetic indices and approximate levels of reliability for genetic indices (adapted from Simm, 1998; Winters, 2009).

TRAIT	HERITABILITY
Milk Yield	0.41
Protein Yield	0.34
Fat Yield	0.37
Herd Life	0.10
Fertility	0.07
SCS	0.24
Conformation (Final Score)	0.26
Feet & Legs	0.15
Mammary system	0.25
Dairy Strength	0.36
Rump	0.23

Table 1.6. Heritabilities calculated for each trait in Canadian Holstein genetic evaluations (adapted from Canadian Dairy Network, 2010)

1.14.4. Genomic evaluations and advancements in genotyping technology

The release of the bovine genome (Bovine Genome Sequencing and Analysis Consortium, 2009) and the development of high-throughput genotyping technologies have made the acquisition of genotypic information for individual animals more readily accessible. In 2009 the United States, Canada and Ireland adopted the incorporation of genomic information into bull proofs in the form of genomic PTAs (gPTAs) or genomic EBVs (gEBVs). Combining genotypic information with phenotypic and pedigree data in genomic evaluations increases the reliability of genetic merit estimates but also enables the shortening of generation times (VanRaden et al., 2009). In the majority of cases, genotyping is performed using the Illumina BovineSNP50™ BeadChip, which has over 40,000 independent SNPs with a minor allele frequency greater than 2 % for genotyping Holstein cattle (Wiggans et al., 2009). There has been a rapid expansion of genotyping technologies available for cattle, as outlined in Table 1.7. As with traditional proofs, the extent and manner in which the new technology has been adopted is dependent upon both the breed and country in which the animal is proven.

The availability of progeny-derived data and comprehensive pedigree information together with the publication of the bovine genome sequence has also facilitated the genetic dissection of complex traits and fine mapping of quantitative trait loci (QTL) i.e. loci with a quantifiable effect upon a trait either directly, or as a result of linkage. Genome wide scan studies have utilised markers distributed across the genome coupled with statistical analysis of progeny-derived data to assess the effect of genetic markers upon traits of interest. For example, using 174 microsatellite markers distributed across the genome at approximately 20 cM intervals, Heyen and colleagues (1999) were able to identify marker effects upon selection traits on eleven chromosomes, with large marker effects for milk fat percentage found on chromosomes 3 and 14. The QTL for milk fat percentage was later refined by the identification of a mis-sense mutation (*K232A*) in the *DGAT1* gene on chromosome 14 which was demonstrated to have a significant effect upon

milk composition yields (Grisart et al., 2002). Genome wide scans have also been applied to assess diversity in cattle populations e.g. North American Holstein bull population (Vallejo et al., 2003), and identify disease related markers (Charlier et al., 2008). The identification of markers for genetic recessive defects has enabled the development of genotyping assays to detect disease carriers e.g. bovine leukocyte adhesion deficiency (BLAD) or complex vertebral malformation (CVM), and prevent the propagation of these disorders throughout the population.

COMPANY	PRODUCT(S)	TECHNOLOGY & SERVICES
Affymetrix	GeneChip Bovine Mapping 10k	10,000 SNPs from bovine genome sequencing project (92 %) and Australia's Commonwealth Scientific and Industrial Research Organisation (8 %)
Igenity	Igenity Dairy Profile	Measures genetic component of 15 desirable traits
Illumina	BovineSNP50 DNA Analysis BeadChip	54,001 highly informative SNPs uniformly distributed across the entire genomes of major cattle breeds
	High-density BeadChip	500,000 – 800,000 SNPs with information gathered from over 20 different breeds
Pfizer Animal Genetics	HD 50K for Angus, GeneStar series, SireTRACE, SureTRAK and genetic defect testing	50,000 SNPs for 14 traits
Metamorphix	Tru-marbling, Tru-tenderness, DNA certified beef programs, Horned polled diagnostics	Parentage verification and diagnostic testing
Genetic Visions	Genetic marker tests	Genetic tests for production traits, coat colour, animal health and viability
DNA Genotek	Performagene Livestock	Sample collection services

Table 1.7. Genotypic services currently available to cattle breeders (adapted from Strauss, 2010).

1.15. MHC diversity in cattle populations

Assessing the MHC haplotype in cattle was originally carried out serologically (Bernoco et al., 1991; Ellis et al., 1999). Following the Fifth International BoLA Workshop (BoLA5) in 1992, there were 50 accepted class I serologies (Davies et al., 1994). Sequence-based molecular analyses have since produced over 80 validated class I alleles which can be accessed on the IPD-MHC database (http://www.ebi.ac.uk/ipd/mhc/bola/) highlighting the high degree of polymorphism of the class I genes in cattle.

In addition to the complexity of the variable gene content and allelic diversity, inter-locus recombination and gene duplication has been shown to occur in the BoLA class I region. This 'unequal crossing-over' is likely to result in hybrid genes involving pseudogenes/gene fragments besides functional class I genes. Sequence comparison of genes 1 and 4 on the A14 haplotype showed the genes were distinct apart from homology throughout exon 3. The close proximity of these two genes to each other, and nine other class I genes, suggests recombination is responsible for the identified sequence homology (Birch et al., 2006; Di Palma et al., 2002). Phylogenetic analysis of the cytoplasmic and transmembrane domains of class I sequences show segregation of sequence into two subgroups, providing evidence of gene duplication (Holmes et al., 2003). In support of gene duplication in the cattle class I region is the expression of two gene 2 alleles on the A17 haplotype. This mix of strategies appear to be operating within the cattle class I region to maintain and generate diversity and so ensuring cattle populations are able to respond to a range of different pathogens.

1.16. Factors contributing to a hypothesised loss of MHC diversity in UK Holstein-Friesians

The widespread application of AI and reproductive technologies has had an enormous influence upon dairy cattle breeding, facilitating the widespread propagation of breeder-selected genotypes (Freeman & Lindberg, 1993). The combined use of MOET and AI has contributed to a small effective population size in the Holstein-Friesian breed i.e. a small number of premium sires and elite dams are relied upon for breeding. For example, Hannoverhill Starbuck (1979-1998) sired 200 000 calves in 45 countries, earning $25 million in semen sales (http://www.ciaq.com/ciaq/history/the-legend-of-starbuck/who-is-starbuck.html). In recent years, UK dairy breeding has relied heavily on Canadian and North American bloodlines, resulting in a reduced genetic pool from which MHC diversity is derived.

Closed breeding strategies and selective breeding for specific economically desirable traits has resulted in increased rates of inbreeding, ultimately leading to a decrease in genetic variance within populations (Kristensen & Sorensen, 2005). Selective breeding within the dairy industry has focused primarily on increased milk yields and disease resistance, primarily involving mastitis. A number of studies have shown a link between class II alleles and mastitis resistance (Mejdell et al., 1994; Oddgeirsson et al., 1988; Sharif et al. 1998). Thus the selection for disease resistance may have inadvertently selected for specific MHC alleles/haplotypes, again limiting MHC variation in that population.

The replacement of dead or injured animals of high genetic merit has been facilitated by the process of somatic nuclear transfer or cloning and so increasing the availability of semen or embryos from these animals (Galli et al., 2003). The prolific sire Hanoverhill Starbuck was ‘reincarnated’ two years after his death by the birth of his clone, Hanoverhill Starbuck II in 2000. Not restricted to the paternal line, cloning of cows has been more common

practice, accounting for over 70 % of the nuclear transfer clones registered in the United States over the last twenty years (Norman et al., 2004). The genetic superiority of cloned cows is reported to be less than one standard deviation above the breed average and suggests selection for cloning was based on non-yield traits i.e. showing quality conformation (Norman et al., 2004). Despite the availability of cloning technologies commercially in the dairy industry, uptake has not been widespread due to high costs and process inefficiency i.e. less than 1 % of cloned embryos generated are carried to term (Moore & Thatcher., 2006). It has been suggested that aberrant class I expression by trophoblast cells of somatic cell nuclear transfer (cloned) foetuses early in pregnancy induces immune-mediated abortion (Davies et al., 2004) and is in part responsible for the reported low success rates. Reincarnation and genetic retrieval will only serve to maintain a small gene pool size over an extended period of time and promote low levels of diversity.

The intense selection for production traits has impacted upon functional traits of dairy cattle as a result of reductions in genetic diversity and negative genetic correlations between traits e.g. selection for high production has been inadvertently coupled with an increased susceptibility to mastitis (Heringstad et al., 2000). A decline in fertility has also been observed in high-yielding dairy cattle which is speculated to be MHC-linked (Freeman & Lindberg, 1993). Aguilar and colleagues (1997) demonstrated that embryo transfer between MHC compatible donors and recipients resulted in higher rate of early fetal loses compared to MHC incompatible pairings.

In terms of pathogens and diseases, both bovine spongiform encephalopathy (BSE) and foot and mouth disease virus (FMDV) have had an impact upon genetic diversity of the UK cattle herds. BSE has been linked to specific genetic susceptibility i.e. polymorphisms in the prion gene (Juling et al., 2006), reducing the associated genotypic frequency and thus diversity of surviving/future cattle populations. In herds within which a case of BSE was confirmed, resulting in the cull of the animal which had tested positive to the

rapid BSE test together with her age-matched cohorts, a process known as ‘genetic recovery’ has often been employed to salvage the genetic variability of UK herds (Galli et al., 2003). Ovaries are recovered from donor cows designated for slaughter, for reasons such as disease, old age, or infertility, and the oocytes matured and later fertilised by a chosen sire. This technique may reduce the impact of BSE upon MHC diversity. The 2001 FMDV outbreak saw a cull of over 1 million cattle from the UK herd. Whilst the cull may have targeted some animals with a susceptible genotype, it also removed animals based simply on geographic location, subsequently exerting a more balanced selection pressure. Economic impacts of BSE, FMDV and supermarket chain influence may have contributed to an increase in diversity i.e. farmers selling up, giving rise to amalgamation of herds. Also, financial pressures may lead to the use of cheaper bulls, thus changing the basis of selection.

1.17. Project Aims

As a result of disease associations and reduced genetic complexity, the bovine class II region has been more extensively studied than the class I region. Although the bovine class I and class II regions are tightly linked, current levels of cattle MHC class I gene diversity have yet to be quantified. Previous studies have investigated class I diversity in dairy cattle using serological methods. The task of assessing diversity with molecular-based methods will be addressed by this project primarily in the Holstein-Friesian breed alone. The project aims to quantify diversity at the BoLA class I loci and assess whether the use of a very small number of highly selected bulls is leading to:

a) loss of MHC diversity and/or

b) some alleles/haplotypes becoming increasingly common (which would in turn lead to increased homozygosity) directly as a result of selection rather than inbreeding.

BoLA class I diversity in the UK Holstein-Friesian herd will be assessed from four perspectives:

1. Canadian Holstein AI bull populations will be investigated to ascertain current levels of diversity, and to identify common class I alleles/haplotypes in premium sires. Studies conducted in the 1980s provide a retrospective comparison and a means of assessing changes in the level of MHC diversity.

2. Statistical analyses will be carried out in order to establish whether the pattern of MHC diversity is a result of trait-based selective forces i.e. whether any class I haplotypes are associated with a particular production trait.

3. Common alleles/haplotypes will form the basis of a SNP assay to be used for the assessment of diversity in a representative sample from the UK herd.

4. Bull populations derived from different geographic locations/founder populations will be analysed in order to assess the impact of selection upon MHC diversity i.e. these bulls will be subject to the same selective pressures, but may harbour different patterns of MHC diversity as a result of the founder population gene pool. Cattle derived from these bulls will also be subject to analysis.

Materials and methods

2.1. Materials

2.1.1. Buffers and solutions

MilliQ ion-exchange purified water (Microbiological Services Department, IAH, Compton) was used for the preparation of all buffers and media. Chemicals were obtained from BDH Laboratory Supplies, Poole, England. All buffer reagents are listed in Table 2.1.

BUFFER	REAGENTS
Phosphate buffered saline (PBS)	0.15 M NaCl, 2.5 mM KCl, 10 mM Na_2HPO_4, 2 mM KH_2PO_4 (pH7.4)
Gel loading buffer (6x)	0.25% bromophenol blue, 0.25% xylene cyanol FF, 30% glycerol in water
Tris-acetate/EDTA (50xTAE)	2 M Tris base, 1 M glacial acetic acid, 0.05 M EDTA (pH8.0)
Tris-CL/EDTA (TE)	10 mM Tris-Cl, 1 mM EDTA (pH 8.0)
Tris-CL/EDTA $(TE)^{-1}$	10 mM Tris-Cl, 0.1 mM EDTA (pH 8.0)
Tris-borate/EDTA (10xTBE)	0.9 M Tris base, 0.9 M boric acid, 0.02 M EDTA (pH8.0)
Extraction buffer	10 mM Tris-HCl pH 8.0, 10 mM EDTA pH 8.0, 1% SDS, 100 mM NaCl

Table 2.1. Buffer solutions and the reagents used and their formulation.

2.1.2. Bacterial Growth Media

All bacterial growth media reagents used are presented in Table 2.2. Luria-Bertani (LB) agar plates were prepared by adding 15 g of Bacto-agar to 1 litre of LB medium. This was then supplemented with 80 µg/ml of X-gal (5-bromo-4-chloro-3-indolyl- beta-D-galactopyranoside) prior to plating.

GROWTH MEDIA	COMPOSITION
Luria-Bertani medium	1 % bactotryptone, 0.5 % bacto yeast extract, 0.5 % NaCl
SOC medium	2 % bactotryptone, 0.5 % bacto yeast extract, 10 mM NaCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 20 mM glucose

Table 2.2. Bacterial growth media reagents.

2.1.3. Commercial Kits

All commercial kits used are tabulated in Table 2.3. and each were used in accordance with manufacturer's protocols.

COMMERCIAL KIT	MANUFACTURER
QiaAmp DNA blood mini kit	Qiagen UK Ltd., Crawley, UK
Qiagen PCR purification kit	
Qiagen gel extraction kit	
Qiagen miniprep kit	
Dynabeads mRNA DIRECT kit	Invitrogen, Paisley, UK
SuperScript II Reverse Transcriptase	
pGEM®-T Easy Vector System I	Promega UK Ltd., Southampton, UK
Big Dye Terminator v3.1 Cycle Sequencing Kit	Applied Biosystems, Warrington, UK

Table 2.3. Commercial kits details

2.2. General methods

2.2.1. Blood sampling and extraction of peripheral blood mononuclear cells (PBMCs)

Venous blood was collected in syringes containing 1101 U of Heparin sodium (Leo Laboratories Ltd., Bucks) per 10 ml volume of blood collected. Glass pipettes were used to carefully layer the blood over histopaque 1083 (Sigma, St Louis, USA) at a ratio of 5:3 in a 50 ml Falcon tube. Samples were then spun in a refrigerated benchtop centrifuge at 2000 *g*, at 4ºC for 45 minutes (min) without the brake applied. Following centrifugation, the samples had separated into a layer of plasma and platelets uppermost, with a cloudy white interface of lymphocytes situated over the bottom layer containing histopaque and red blood cells (RBC). Ensuring no RBC were taken up, lymphocytes were carefully removed and transferred to a clean 25 ml Universal tube using a pipette. The PBMCs were then washed twice by

centrifugation with phosphate buffered saline (PBS, Media department, IAH, Compton) at 800 *g* for 3 min ensuring the pellet was thoroughly resuspended between washes. Finally the cells were pelleted (~5 x 10^6 cells) and the liquid removed before snap freezing on dry ice. Samples were then stored at -80ºC until required for genomic DNA extraction (gDNA).

2.2.2. Isolation of genomic DNA from PBMCs

The QIAmp DNA blood mini kit was used to isolate genomic (gDNA) from PBMCs. In brief, 5 x 10^6 cells (either frozen pellet or fresh) were lysed with 20 µl Qiagen proteinase K (20 mg/ml) followed by the addition of a high salt containing binding buffer. This solution is then passed through a QIAamp spin column containing a silica gel membrane on to which the DNA binds. Residual contaminants were removed by washing the spin columns twice before eluting the DNA from the column in 200 µl elution buffer (10mM TrisCl, 0.5 mM EDTA) and stored at 4ºC.

2.2.3. Extraction of gDNA from semen

One unit of extended semen (~5 x 10^7 sperm cells) was diluted with 1 ml of PBS in a 1.5 ml microcentrifuge tube. The solution was then centrifuged at 5000 *g* for 5 min to pellet sperm cells. The supernatant was aspirated and the pellet resuspended in 1 ml PBS before another centrifugation step of 5 min at 5000 *g*. The supernatant was aspirated and the sperm pellet lysed with 450µl of pre-warmed (60ºC) extraction buffer. To this solution, 50 µl of 0.5 M dithiothreitol and 15 µl of proteinase K (15 mg/ml) were added before incubation for 12 h at 60-65ºC. Following incubation, 160 µl of saturated NaCl (5M) solution was added and the mixture was shaken vigorously for 2 min. The solution was centrifuged for 10 min at 15 000 *g*. To precipitate the DNA 1 ml ethanol was added to the supernatant and mixed gently before extracting the purified DNA with glass pipette and resuspended in TE buffer.

2.2.4. Isolation of mRNA from PBMCs

The Dynabead mRNA DIRECT kit (Invitrogen) was used to isolate messenger RNA (mRNA) from pBMCs. The principle of this kit relies on the hybridisation of the poly A (adenine) tail, present at the 3' end of the majority of eukaryotic mRNAs, to short sequences of oligo dThymine (dThymine; dT) covalently bound to a solid matrix. In this kit the dTs are bound to magnetic beads. Approximately 5 x 10^6 cells were lysed in Dynal lysis binding buffer (100 mM Tris-HCl pH 7.5, 500 mM LiCl, 10 mM EDTA pH 8 1% LiDS, 5 mM dithiothreitol, DTT). In order to shear the DNA, the lysate was passed repeatedly through a 21-gauge needle using a 1 ml syringe. The lysate was then mixed with the oligo dT-coated beads to isolate the mRNA. The beads were then fixed by a magnetic field and washed with a series of buffers to remove non-poly A RNAs and other residual components. The mRNA was then eluted from the magnetic beads in 20 µl of 10 mM Tris-HCl. 5 x 10^6 cells yielded approximately 2 µg RNA for use in cDNA synthesis as quantified using the nanodrop. Sterile conditions were adhered to throughout this procedure to prevent contamination with RNases.

2.2.5. First strand cDNA synthesis

The SuperScript II Reverse Transcriptase kit (Invitrogen) was used to synthesise first strand cDNA immediately following mRNA isolation. An RNA-dependent DNA polymerase reverse transcriptase (Superscript™ II RNase H^-) and poly dT primer act on the mRNA to produce a hybrid mRNA-cDNA molecule. First strand cDNA was produced by adding approximately 2 µg mRNA to 5 x first strand buffer (100 mM Tris-HCl pH8.4, 250 mM KCl, 15 mM $MgCl_2$), 10 mM dNTP mix (10 mM each dATP, dCTP, dGTP, dTTP, Promega, Southampton, UK), 0.1 M DTT and 50 U reverse transcriptase to a final volume of 25 µl followed by incubation at 42°C for 50 min. Proteins are

removed and the DNA is purified using the Qiagen PCR purification kit, whilst the mRNA is removed later, during the initial stages of PCR.

2.2.6. Polymerase chain reaction (PCR)

The same standard reaction mixture was used for each type of PCR. Amplification from either gDNA and cDNA templates (1.0 µl each) was carried out in a final volume of 25 µl containing 10 pM of each primer, 0.2 mM dNTPs, 2.5 mM $MgCl_2$, 1.5 U GoTaq DNA Polymerase in 1X GoTaq Flexi Buffer (Promega, Madison, WI, USA). All thermal cycling programmes were performed on a PTC-200 thermal cycler (MJ Research, USA). Separation on 1% agarose/Tris-acetate-ethylendiaminetetraacetic (TAE) acid gels was used to ascertain the presence or absence of a product (see 2.2.7.). All primer sequences are detailed in the Appendix.

2.2.6.1. Gene-specific PCR

Gene-specific PCR was used to amplify MHC class I genes 1, 2, 3 and 6 from gDNA templates (Murphy, 2004; Birch et al., 2006). The thermal cycling profile used for gene 1 and 6-specific primers was: 95ºC for 2 min, 30 cycles of 95ºC for 30 secs, 65ºC for 1 min, 72ºC for 1 min and followed by 72ºC for 7 min. The thermal cycling profile for gene 2 and 3-specific primers was the same except the annealing temperature used was 60ºC and the number of cycles increased to 35. Gene-specific PCR products were analysed by RSCA analysis (see section 2.2.8) and/or sequencing (see section 2.2.13) and/or the SNP extension assay (see section 2.2.14.).

2.2.6.2. Sequence-specific PCR (SSPs)

Sequence-specific PCR was used to check the presence of alleles which could not be detected due to the limits of group-specific PCR and/or RSCA analysis. Thermal cycling profiles used were the same as those used for gene 1 and 6-specific PCR with the exception of primer pair 4221.1 for which the thermal cycling profile for gene 2 and 3 specific PCR was used.

2.2.6.3. Generic PCR

Extensively used generic primers Bov7 and Bov11 (Pichowski et al. 1996) were used to amplify the region from the start of exon 2 to the beginning of exon 4 from both gDNA and cDNA templates for sequencing. For amplification of full-length MHC class I from cDNA, a mixture of primers Bov 21a/g and Bov 21-BSF (forward) and a mixture of Bov 3 and Bov 3-BSF (reverse) were used. These mixed primers were used because there are known polymorphisms at the target sites. The thermal cycling profile used in both cases was: 95°C for 2 min, 35 cycles of 96ºC for 30 secs, 55ºC for 1 min, 72ºC for 1 min, followed by 72ºC for 7 min.

2.2.7. Electrophoresis of DNA fragments

PCR products were premixed with 6 µl gel loading buffer and loaded onto 1% agarose gels containing 0.5 µg/ml Ethidium bromide. Gels were run in 1 x TAE buffer at 5 V/cm with the running time determined by the agarose gel concentration and the predicted PCR product size (Sambrook et al., 1989). Samples were run alongside a 1 kb plus DNA ladder (Sigma) to enable the products to be sized. Gels were visualised under UV light using a UVP transilluminator and the images recorded on a UVP GDS5000.

2.2.8. Reference strand-mediated conformation analysis (hybridisation to FLRs)

Reference strand-mediated conformation analysis (RSCA) was used to determine the class I haplotype of Holstein cattle and was performed essentially as described in Birch et al. (2006). The technique utilises fluorescently labelled reference (FLR) strands derived from known allele sequences, which had been cloned prior PCR amplification with a fluorescein amidite (FAM)-labelled 5' oligonucleotide (following protocol described in section 2.2.6.1.). For this application, two FLRs were used per gene (see Table 2.4). Gene-specific PCR products (3 µl) were hybridised in separate reactions with gene-associated FLRs (1 µl) to produce a mixed solution of hetero- and homoduplexes under the following thermal conditions: 95ºC for 4min, 55ºC for 5 min and 15ºC for 3 min. Two microlitres of the gene-specific PCR product/FLR hybrid solution were then mixed with 2 µl of ficoll loading dye together with 0.5 µl of GeneScan-2500 6-carboxyl-X-rhodamine (ROX) dye labelled size standard (Applied Biosystems, Warrington, UK) and loaded onto a 6.5 % non-denaturing polyacrylamide gel. Electrophoresis was carried out in an ABI Prism 377 DNA sequencer (Perkin Elmer, Foster City, CA, USA) for 2 h at 35 mA and 51ºC (pre-run) followed by 10 h at 60 mA and 40ºC. Laser detection of FAM-labelled duplexes and an included size standard enable the Genescan 2.1 (ABI) analysis package to assign a mobility value which was compared against known allele mobility values and allele identification was assigned. In the cases of novel mobility values, further investigation was carried out using sequencing (see section 2.2.13).

GENE	FLR 1	FLR 2
1	*N*02101*	*N*02301*
2	*N*01801*	*N*02501*
3	*N*01701*	*N*00201*
6	*N*01301*	*N*01401*

Table 2.4. Alleles used for production of FLRs (Birch et al. 2006)

2.2.9. DNA extraction from agarose gels

PCR products of the correct size were excised from the agarose gel using a scalpel blade and the DNA was recovered using the Qiagen gel extraction kit. This kit is based on a similar principle to that of the Genomic DNA isolation kit previously described. The agarose gel is digested enzymatically prior to the binding of the DNA to the silica membrane of the spin column. The bound DNA is washed with buffers of varying pH and salt concentrations prior to elution in 30 µl elution buffer (10 mM Tris-HCl pH 8.5). Eluted samples were then stored at -20ºC.

2.2.10. Ligation of DNA into pGEM-T Easy vector

In order to produce individual single clones for sequencing, purified DNA of interest was ligated into the pGEM-T Easy vector system according to the supplied protocol including the recommended controls (Promega UK Ltd., Southampton, UK). This cloning vector has a multiple cloning site within the ß-galactosidase gene with ‘sticky’ ends or thymidine overhangs, enabling the insertion of PCR products by the Taq enzyme used. Successful insertion of a

PCR product of interest into the vector disrupts the coding sequence of ß-galactosidase and enables recombinant clones to be detected by blue/white selection in presence of X-gal substrate. The pGEM-Teasy cloning site is flanked by two bacteriophage promoters lying in opposite directions (Short et al., 1988). These promoter sequences can be used as primer sequences for the sequencing of the inserted product. For each ligation reaction 3 µl of DNA (approximately 20 ng) was added to 2x T4 DNA ligase buffer (60mM Tris-HCl pH 7.8, 20 mM MgCl2, 20 mM DTT, 2 mM ATP, 10% polyethylene glycol), 50 ng of pGEM-T Easy vector and 3 U of T4 DNA ligase, mixed by pipetting and incubated overnight at 4ºC.

2.2.11. Transformation of TOP10 cells using recombinant pGEM-Teasy vector

Two microlitres of each ligation reaction was added to 50 µl of TOP10 competent cells (made in house by technician) in 1.5 ml Eppendorf tubes, and gently mixed by flicking the tube. The cells were incubated on ice for 20 min, heat shocked at 42ºC for 50 secs and then placed directly back on ice for 2 min to allow the plasmid to enter the cells. 950 µl of SOC medium was added to each of the tubes and the cells were incubated at 37ºC with shaking (180 rpm) for 1.5 hrs. Samples were plated out in duplicate (100 µl) on LB plates containing ampicillin and X-gal, the plates inverted and incubated at 37ºC overnight. Colonies harbouring recombinant vector, indicated by the white colour, were picked using a sterile pipette and placed in a 20 ml universal tube containing 5 ml LB and 100 µg/ml ampicillin. Samples were then grown overnight at 37ºC with shaking (180 rpm).

2.2.12. Minipreparations of purified DNA from TOP10 transformed cell cultures

The QIAprep spin miniprep kit (Qiagen) was used to extract purified DNA from transformed bacterial cultures. The process uses alkaline lysis and SDS detergent (Birnboim & Doly, 1979) to lyse bacterial cells and denature chromosomal DNA and proteins, releasing plasmid DNA into the solution. The denatured material was then removed by centrifugation and the supernatant applied to a Qiagen spin column. The DNA binds the silica matrix of the column and is washed by a series of buffers to get rid of impurities. The plasmid DNA is then eluted in 10 mM Tris-Cl (pH 8.5). These samples were then stored at -20ºC until further use.

2.2.13. Sequencing

Sequencing was used to both confirm heterozygosity and identify novel alleles i.e. alleles with novel mobility values detected by RSCA analysis (see section 2.2.8.). Cloned products of both gene-specific and generic PCR were used as templates for sequencing reactions. Reactions were set up with the BigDye® Terminator v3.1 Cycle Sequencing kit (Applied Biosystems) as follows: 1 µl BigDye 3.1 ready reaction mix , 1.5 µl BigDye 3.1 sequencing x 5 buffer, 0.8 pmol of primer (either M13 Forward or SP6 reverse), ~150-300 ng of mini-prepped DNA template made up to a total volume of 10 µl. These reactions were then subject to the following thermal cycling conditions: ramp of 1ºC/s to 96ºC, 96ºC for 1 min, followed by 25 cycles of 1°/s to 96ºC, 96ºC for 10 s, 1°/s to 50ºC, 50ºC for 5 s, 1°/s to 60ºC, 60ºC for 4 min. DNA was precipitated by the addition of 1 µl of 3 M ammonium acetate (pH 4.6) and 25 µl of 95 % ethanol prior to centrifugation at 3000 *g* at 4ºC for 30 min. Supernatents were removed and pellets were washed with 70 % ethanol under centrifugation at 3000 *g* at 4ºC for 10 min. Reactions were left to dry in the dark at room temperature. A representative number of clones in each case were

sequenced using an Applied Biosystems 3730xl DNA Analyser. The sequences obtained were then aligned using Vector NTI Version 10 (Invitrogen), subjected to a NCBI BLASTn (http://www.ncbi.nlm.nih.gov/BLAST/) search, and a sequence identity assigned.

2.2.14. SNP assay methods

2.2.14.1. Preparation of template for SNP-extension reactions

Gene 1 and 2-specific PCRs were carried out independently, the products of which were purified enzymatically using Exo-SAPit (GE Healthcare, UK). Five microlitres of the PCR product were mixed with 2 µl of Exo-SAPit and incubated for 45 min at 37ºC, followed by 15 min at 80ºC.

2.2.14.2. SNP-extension reactions

During the course of developing the assay, different Beckman Coulter SNP-extension kits were used. The SNP-extension reaction mixtures and thermal cycling conditions used were in accordance with the SNP-kit protocols. The Beckman Coulter Methods development kit used the same reaction mixtures and thermal cycling conditions as the initial Beckman Coulter GenomeLab™ SNP-Primer Extension kit.

2.2.14.3. SAP purification of SNP-extension reaction products

The products of the SNP-extension product were purified enzymatically using shrimp alkaline phosphatase (SAP). Three microlitres of the SNP-extension product is mixed with 0.7 µl 10 x SAP buffer and 1 U SAP made up to a total volume of 5 µl with milli-Q water. The mixture was incubated for 45 min at 37^{o}C before inactivating the SAP enzyme by incubating at 65^{o}C for a further 15 min.

2.2.14.4. SNP assay genotyping

One microlitre of the purified SNP-extension product was mixed with 0.5 µl of CEQ size standard 80 and 39 µl of sample loading solution (SLS) prior to loading on the Beckman CEQ-8000 DNA sequencer. The Beckman default separation method SNP-1, modified to use an injection time of 50 s, was used for analysis of the SNP-extension products. SNP assay graphs were analysed by eye to confirm correct sizes and record trace dye colour.

Analysis of MHC class I diversity in Canadian Holstein AI bulls

3.1. Introduction

Diversity in MHC genes is believed to benefit populations by allowing responses to rapidly evolving disease pathogens (Doherty & Zinkernagel, 1975). Furthermore, heterozygosity is considered advantageous because individuals with two different alleles of each gene can potentially present a greater range of antigenic peptides than homozygotes, thus are potentially resistant to a greater variety of pathogens (Clarke & Kirby, 1966). Cattle herds with greater MHC diversity should therefore be better equipped to fight infection or respond to vaccination.

Selection for both production and disease-resistance traits has been an intensive and long-standing process in dairy cattle breeding. The introduction of artificial insemination (AI) and advancement of reproductive technologies such as multiple ovulation embryo transfer (MOET) has perpetuated a heavy reliance on a narrow spectrum of bloodlines. Dairy breeding programmes worldwide rely on a small number of premium bulls, resulting in an effective population size within the Holstein-Friesian breed of approximately 150 animals (Hayes et al., 2003). Ninety nine percent of pedigree Canadian Holstein heifers born since 2003 are descendants of the prolific American Holstein sire Round Oak Rag Apple Elevation, exemplifying the dominance of specific bloodlines in breeding programmes (Van Doormaal et al., 2005). Canadian and North American bloodlines represent a worldwide dominant force in dairy breeding and in the UK accounted for over 40 % of 250,000 Holstein-Friesian pedigree dairy cattle registrations in 2008 alone (Alison Maddrell, personal communication, Holstein UK).

Current animal husbandry practices coupled with high levels of inbreeding may have led to a reduction in MHC diversity in UK dairy cattle. In order to address this theory, samples from premium Canadian Holstein bulls were obtained in order to ascertain the current levels of MHC class I diversity in an active bull population. Assessing MHC diversity using the bulls as a reference population gives a good indication of diversity in the wider UK herd due to the reliance on relatively few bulls for breeding. The bull samples included in the 2006 Canadian Holstein bull population had sired approximately 80,000 daughters in the UK herd to date. This part of the project also includes a retrospective comparison of the 2006 Canadian Holstein bull population with data published in 1989 in which 271 Canadian Holstein bulls were MHC class I typed using serological methods (Batra et al., 1989).

3.2. General materials and methods

Blood samples from 74 premium Canadian AI bulls standing with Semex (The Semex Alliance, Guelph, Canada) were donated to our laboratory in 2006, thus these samples will be referred to as the 2006 sample set. These bulls were chosen by the donating company on the basis of their high performance and reliability. Genomic DNA (gDNA) was extracted from all but one of the samples using methods described in Chapter 2, sections 2.2.1. and 2.2.2. Prior to the start of this project, 25 of these samples were typed by J Birch, as described below and the typing results from these 25 animals contribute to the final 2006 sample data set. Of the 49 samples I typed, one was a repeat sample and so was removed from all frequency calculations. The final total number of samples included in the 2006 Canadian bull sample was 72, which gives a fair representation of the population given the estimated effective population size of the Canadian Holstein population.

For the MHC class I typing of the 2006 Canadian Holstein bull sample, Reference strand-mediated conformation analysis (RSCA), a PCR-based

genotyping technique was used. This technique has been applied to the MHC-typing of a number of species including humans (Arguello et al., 1998; Ramon et al., 1998), primate species (Smith et al., 2005; Tanaka-Takahashi, 2007; Blasky et al., 2008), dogs (Kennedy et al., 2005), cats (Kennedy et al., 2003), fish (Lenz et al., 2008) and also cattle (Birch et al., 2006). The technique assigns identity to allele sequences amplified from gDNA on the basis of conformation-dependent mobility in a polyacrylamide gel. Mobility values are calculated for each test sample compared to known reference strand allele sequences, of which there are two per gene analysed. Comparison with two reference strand sequences gives a multi-dimensional co-ordinate for each test allele sequence, facilitating differentiation of closely related variants, as is commonly found in MHC genes.

RSCA had been previously customised for use for cattle class I typing. Gene-specific primer pairs were designed to amplify the region containing exon 2, intron 2 and exon 3 from cattle MHC class I alleles encoded on genes 1, 2, 3 and 6 (Murphy, 2004). Genes 4 and 5 are not included in this MHC class I typing technique due to insufficient variability within the gene 4 intron sequences to enable gene-specific amplification, and gene 5 alleles have thus far only been detected in *Bos indicus* cattle. Two well-characterised alleles were chosen from each of the four genes to be used as fluorescently-labelled reference strands (FLRs; Chapter 2. Table 2.4.). The allele sequences to be used as FLRs are amplified from gDNA using the gene-specific primers, prior to cloning, and sequencing to confirm the identity. Once the identity has been confirmed by sequence analysis, the cloned products are subjected to further gene-specific amplification, this time using FAM-labelled versions of the forward primers that amplify and fluorescently label the products, resulting in the production of the FLR.

Test samples are subjected to gene-specific amplification, for which a product may or may not be present, due to the variability in the number of genes expressed on cattle class I haplotypes. Products of gene-specific PCR are hybridised independently with each of the FLRs for that gene by a process

of denaturation and re-annealing carried out in a thermal cycler. The amplified product becomes annealed to the FLR giving rise to heteroduplexes, with sequence differences resulting in a bulged conformation at the site of mismatch, which affects migration. The hybridisation mixtures are mixed with loading dye and a ROX-labelled size standard before being loaded onto a 6.5 % non-denaturing polyacrylamide gel and subject to electrophoresis on a sequencer. Using laser detection, all duplexes containing a FAM label i.e. FLR homoduplexes and test allele sequence-FLR heteroduplexes, can be detected and a mobility value compared to the FLR homoduplex calculated. If the test sample sequence is identical to the FLR, the mobility value of the test allele sequence-FLR heteroduplex will be identical to that of the FLR homoduplex. MHC class I genes 1, 2, 3 and 6 were amplified from the 2006 Canadian Holstein bull samples using gene-specific primers, the products of which were then analysed by RSCA performed as described in Birch et al. 2006 (see also Chapter 2, sections 2.2.6.1 and 2.2.8.). Each sample was assigned an MHC haplotype prediction based upon mobility values corresponding to previously sequenced alleles. Limitations of gene-specific PCR and RSCA e.g. gene-specific primer pairs are unable to amplify *N*01201* on the A10 haplotype due to divergent intron sequences and the mobility value of *N*03101* on the A13 haplotype is too great to be detected on the RSCA gel, necessitates the use of SSP-PCR and sequencing to complement missing haplotype data and/or confirm ambiguous typing results. Sequences were derived from products amplified with either generic or gene-specific primers. For full description of the methods see Chapter 2, sections 2.2.6. and 2.2.9 to 2.2.13. Figure 3.1 outlines the experimental plan.

MHC class I typing data from 271 Canadian Holstein bulls sampled in 1986 (Batra et al., 1989) from four Canadian AI studs were used in a retrospective comparison of MHC haplotype frequencies. This sample shall be referred to as the 1986 sample, in line with the naming of the later sample. The four Canadian AI studs (United Breeders Inc, Guelph; Western Ontario Breeders Inc, Woodstock; Eastern Breeders Inc, Kemptville and CIAQ, Quebec) have since amalgamated to become part of the Semex Alliance conglomerate, and as such the 1986 and 2006 sample sets are representative of the same

population. Typing of the 1986 sample was carried out serologically as the availability of reagents for molecular typing at this time was limited, and serology was the most robust method for cattle class I typing available at the time. Data generated by the BoLA workshops and independent studies facilitate the linking of current molecular definitions of haplotypes to the sera used for typing in the 1980s, thus enabling the comparison between these two sample sets (see Appendix). Mathematical modelling and statistical analysis was applied to determine whether there had been a loss of diversity between the two samplings. The level of heterozygosity in the 2006 Canadian Holstein bull sample was also calculated and compared against the expected level of heterozygosity as determined by Hardy-Weinberg proportions. This part of the study was carried out with the generous help of Prof. M. Stear, Dr R. Reeve and Dr L. Matthews from Glasgow University. For a detailed description of the mathematical methods used, please see the Appendix.

The impact of genetic drift and inbreeding were also modelled to see if these forces were responsible for changes in MHC haplotype frequencies over time. As the time between the two samples represents approximately four generations of cattle, the pedigrees of all bulls in the 2006 sample were traced back to the great-grandparents and used in the simulations. Haplotypes were randomly assigned to the great grandparents based on the frequencies observed in 1986. Haplotypes were randomly assigned to progeny by mimicking Mendelian inheritance with each offspring having a 50 % chance of inheriting either of the two haplotypes in each parent. An estimated prediction of gene frequency for each starting haplotype after four generations could then calculated. The algorithm used was written in SAS and 5000 replicates were run to ensure a high degree of reliability for each of the simulated haplotype frequency prediction estimates. The flow of genes was simulated within the constraints of the pedigree structure detailed for the 2006 Canadian Holstein bull sample, thus taking into account the impact of inbreeding upon the frequency of MHC class I haplotypes after 20 years of selective breeding.

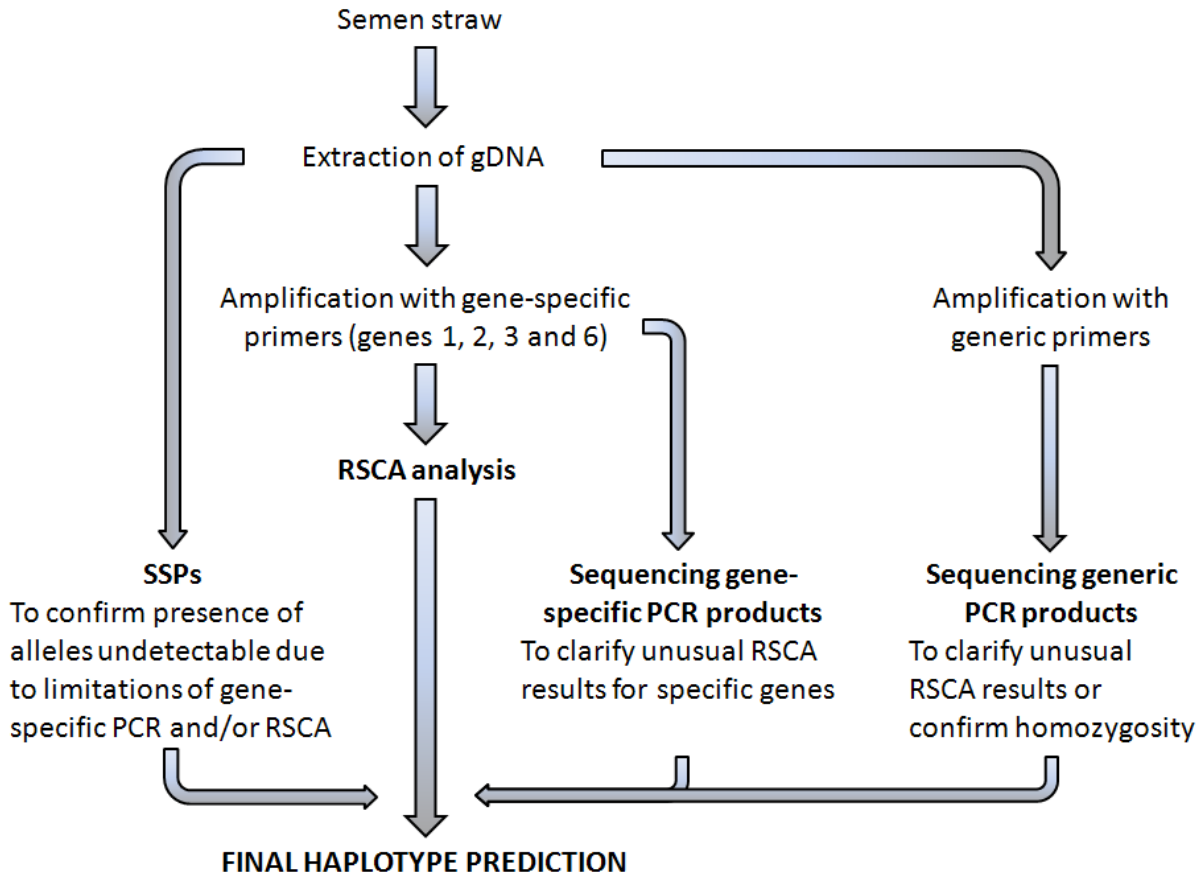


Figure 3.1. Diagrammatic representation of the experimental plan used to determine MHC class I haplotypes in the 2006 sample.

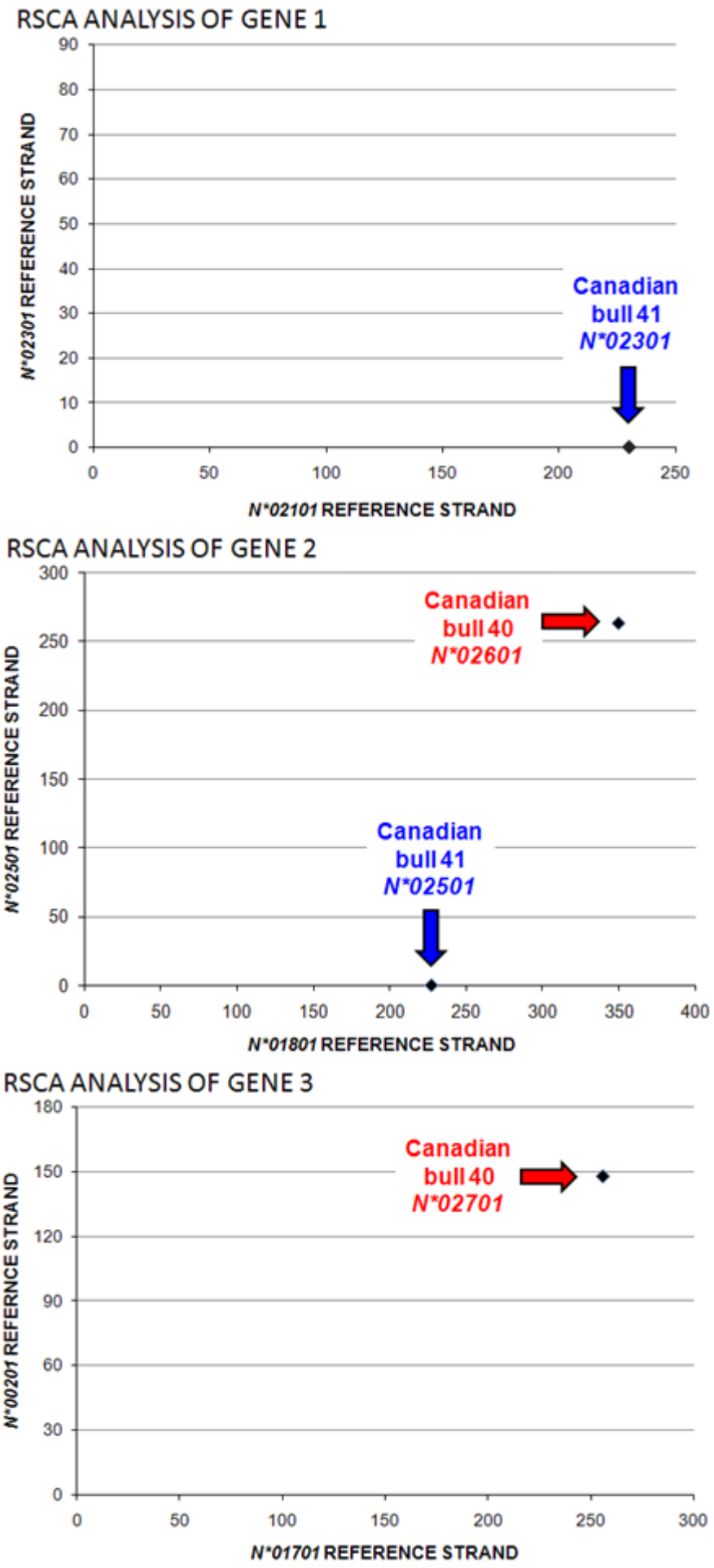
3.3. Results

The majority of samples were typed solely using RSCA analysis. Limitations of gene-specific PCR and/or RSCA, or ambiguous typing results warranted the application of additional methods as outlined below.

3.3.1. DNA sequencing of generic PCR products

DNA sequencing of generic PCR products was used for the clarification of unusual RSCA typing results or confirmation of homozygotes. Generic primers Bov7 and Bov11 (Chapter 2. section 2.2.6.3; Pichowski et al., 1996) were used to amplify the region from the start of exon 2 to the beginning of exon 4. PCR products were separated on 1 % agarose gels, excised and then purified using the Qiaquick gel extraction kit (Qiagen, Crawley) before ligation into pGEM-T Easy (Promega, Southampton). The recombinant vector was used for the transformation of competent *E. coli* TOP10 cells. A minimum of 8 clones in each case were sequenced using the Big Dye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) and an Applied Biosystems 3730xl DNA Analyser. The sequences obtained were then aligned using Vector NTI Version 11 (Invitrogen), subjected to a NCBI BLASTn (http://www.ncbi.nlm.nih.gov/BLAST/) search, and a sequence identity assigned. Figure 3.2 shows a worked example of two animals typed as homozygous by RSCA.

A.


RSCA ANALYSIS OF GENE 1
N*02301 REFERENCE STRAND
N*02101 REFERENCE STRAND
Canadian bull 41 N*02301
RSCA ANALYSIS OF GENE 2
N*02501 REFERENCE STRAND
N*01801 REFERENCE STRAND
Canadian bull 40 N*02601
Canadian bull 41 N*02501
RSCA ANALYSIS OF GENE 3
N*00201 REFERNCE STRAND
N*01701 REFERENCE STRAND
Canadian bull 40 N*02701

B.

CANADIAN BULL 41: SEQUENCING RESULTS SHOWS ANIMAL IS HETEROZYGOUS WITH HAPLOTYPE W12B/A14

```
          -20       *         0         *        20         *        40         *        60         *        80         *
N*02301 : MGPRTLFVLLLGALALTETRAGSHSLRYFYTAVSRPGLGEPRFISVGYVDDTQFVRFDSDAPNPREEPRAPWIEKEGPEYWDRETRISKENTLVYRESLNNLRGYYNQSEA
CB41    : ------------------------------------------------..............................................................

                100         *       120         *       140         *       160         *       180         *       200
N*02301 : GSHNIQAMYGCDVGSDGSFLRGYSQDAYDGRDYIALNEDLRSWTAADTAAQITKRKWEAEGYAESLRNYLEGRCVEWLRRYLENGKDALLRADPPMAHVTHHPSSEREVTL
CB41    : ..............................................................------------------------------------


          -20       *         0         *        20         *        40         *        60         *        80         *
N*02401 : MGPRTLLLLLSGVLVLTETRAGSHSLRYFSTAVSRPGLEEPRFIIVGYVDDTQFVRFDSDSPNPRAEPRAPWMEQEGPEYWDEQTRIVKDTAQTFRANLNTALGYYNQSEA
CB41    : ------------------------------------------------..............................................................

                100         *       120         *       140         *       160         *       180         *       200
N*02401 : GSHNIQAMYGCDVGSDGSFLRGYSQDAYDGRDYIALNEDLRSWTAADTAAQITKRKWEAEGYAESLRNYLEGTCVEWLRRYLENGKDTLLRADPPKAHVTHHSISGREVTL
CB41    : ..............................................................------------------------------------


             *         0         *        20         *        40         *        60         *        80         *
N*00801 : LLLLSGVLVLTETLAGSHSLRYFLTAVSRPGLGEPRFIIVGYVDDTQFVRFDSNTPNPRMEPRARWVEKEGPEYWDRETRNSKETAQTFRANLNTALGYYNQSEAGSHTVQ
CB41    : ---------------------------------------..............................................................

           100         *       120         *       140         *       160         *       180         *       200
N*00801 : EMYGCDVGPDGRLLRGFMQDAYDGRDYIALNEDLRSWTAADTAAQITKRKWEAAGDAETWRNYLEGRCVEWLRRYLENGKDALLRADPPKAHVTHHSISEREVTLRCWALG
CB41    : ..........................................................------------------------------------------------
```

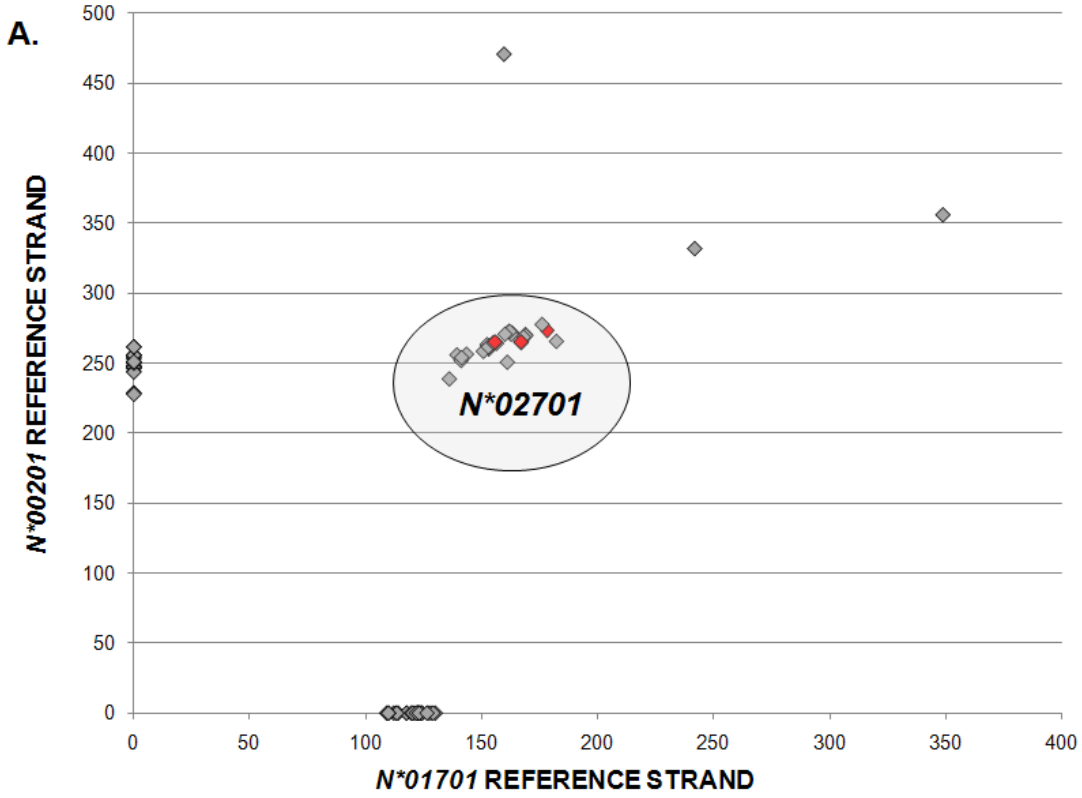
CANADIAN BULL 40: SEQUENCING RESULTS CONFIRM IT IS AN A20 HOMOZYGOTE

```
                *         0         *        20         *        40         *        60         *        80         *
N*02601 : LLLLPLGVLVLTENLAGSHSLRYLYTGVSRPGLGEPRFIAVGYVDDTQFTRFDSDAPNPREEPRVPWMEQEGPEYWDRETRISKETAQTFRVDLNTLRGYYNQSEAGSHTI
CB40    : ----------------------------------------.....................................................................

              100         *       120         *       140         *       160         *       180         *       200
N*02601 : QEMYGCDVGPDGRFLRGYEQYGYEGRDYIALNEDLRSWTAADTAAQITKRKWEAADYAESLRNYLEGRCVEGLRRYLENGKDTLLRADPPKAHVTRHPISEREVTLRCWAL
CB40    : .....................................................------------------------------------------

                *         0         *        20         *        40         *        60         *        80         *
N*02701 : LLLLPLGVLVLTETRAGSHSMRYFSTAVSRPGFGEPRYLEVGYVDDTQFVRFDSDARNPRMEPRTRWVKQEGPEYWDRNTRNAKGNAQSFRVGLNTLRGYYNQSEAGSHTL
CB40    : ----------------------------------------.....................................................................

              100         *       120         *       140         *       160         *       180         *       200
N*02701 : QWMSGCYVGPDGRLLRGFMQYGYDGRDYLALNEDLRSWTAADTEAQITKRKWEAEGYAEVQRNYLEGECVEWLRRHLENGKDTLLRADPPKAHVTHHPISDREVTLRCWAL
CB40    : .....................................................------------------------------------------------------
```

Figure 3.2. The use of DNA sequencing of generic PCR products for confirmation of RSCA analysis results. **A.** Mobility values detected during RSCA analysis of gene-specific PCR products from Canadian bulls 41 and 40. Alleles *N*02301* and *N*02501* were detected by RSCA analysis in Canadian bull 41, which was assumed to be an A14 homozygote. Alleles *N*02601* and *N*02701* were detected by RSCA analysis in Canadian bull 40, which was assumed to be an A20 homozygote. **B.** Generic PCR products from Canadian bull 41 and 40 were sequenced. Sequences identified as *N*02301*, *N*02401* and *N*00801* were detected in Canadian bull 41, indicating that it is an A14/W12B heterozygote (see section 3.3.3. for more information). Canadian bull 40 was confirmed as an A20 homozygote by sequencing with only alleles *N*02601* and *N*02701* detected. Dots indicate identity, dashes indicate gaps/lack of sequence data compared to reference.

3.3.2. DNA sequencing of gene-specific PCR products

A small number of samples yielded unusual results that warranted investigation by sequencing of individual genes. In such cases, amplification was carried out using gene-specific primers (Chapter 2. section 2.2.6.1) and the products processed as described in section 3.3.1. RSCA analysis of gene-specific PCR products, showed four animals appeared to have incomplete A20 haplotypes i.e. a gene 3 allele was detected with similar mobility values to that of *N*02701*, however the associated *N*02601* allele was not detected in these animals despite amplification with gene 2 specific primers. However, the gene 2 allele amplified in each case could belong to the second class I haplotype present in these heterozygous animals. Figure 3.3 shows the mobility values of the gene 3 allele detected with similar mobility values to that of *N*02701* and the amplification of both genes 2 and 3 from these animals.



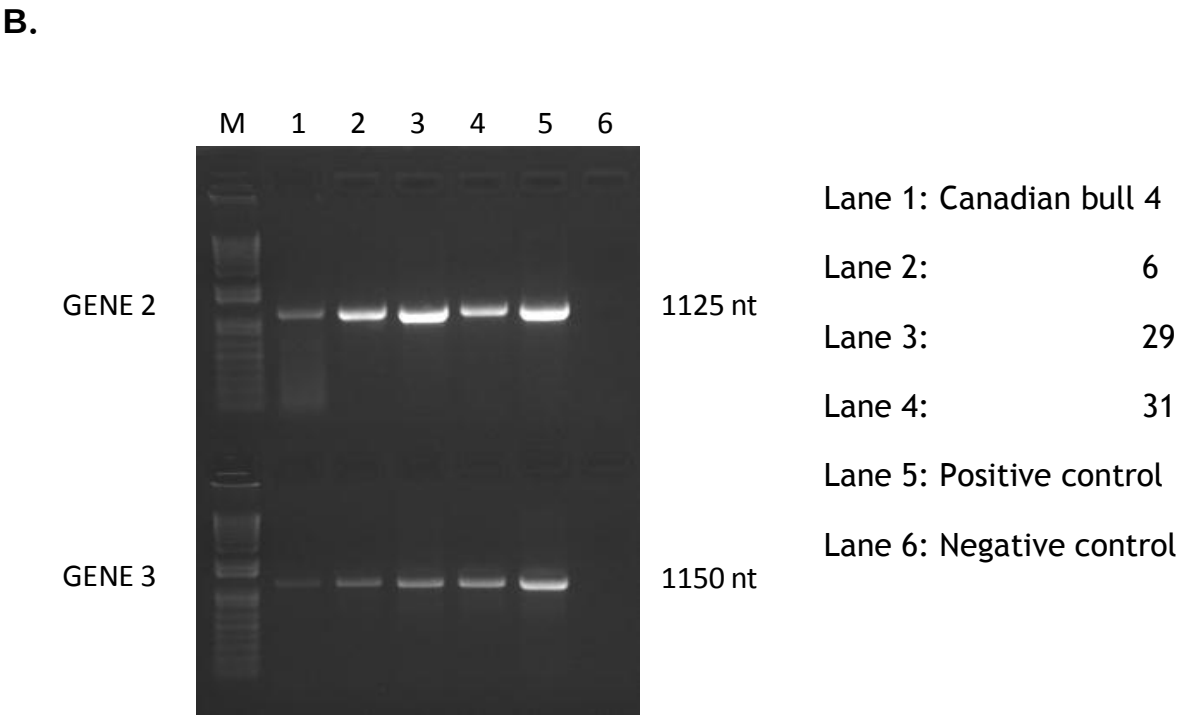


Figure 3.3. Unusual haplotypes identified in four of the Canadian bull samples. **A.** show the mobility values (marked in red) of the gene 3 allele detected in the Canadian bulls in relation to mobility values previously recorded for the *N*02701* allele, indicated by the grey shaded area. **B.** PCR amplification with gene 2 and 3-specific primers yield products for each of the four Canadian bulls.

Given the generally conserved nature of bovine MHC class I haplotypes, the absence of an associated allele is unusual and warranted further investigation. To confirm the identity of the gene 3 allele in this cohort of animals, gene 3-specific primers were used for amplification prior to sequencing. Following a BLASTn search for identification, individual clone sequences were shown to be identical across alpha 1 and alpha 2 with *N*03601* (see Figure 3.4). Although *N*03601* is only 91% identical at the nucleotide level to *N*02701* across alpha 1 and 2, it is the similarity in molecular weights that may account for the original miss-identification. RSCA is able to distinguish alleles differing by a single base, but the difference in molecular weight of these two alleles across exons 2 and 3 including introns sequences, is only 110 g which is less than the weight of the smallest nucleotide, cytosine.

```
                *         0         *        20         *        40         *        60         *        80         *
N*03601 : LLLLLSGVLVLTETRAGSHSLRYFYTGVSRPGLGEPRFIAVGYVDDTQFVRFDSDAPDPRMEPRARWVEQEGPEYWDRNTRNAKDAAQTFRVNLNTLRGYYNQSEAGSHTF
CB4_AF  : ----------------.....................................................................................
CB4_BF  : ----------------.....................................................................................
CB4_CF  : ----------------.....................................................................................
CB4_DF  : ----------------.....................................................................................
CB4_AR  : ----------------------------------------------------------------------------------------------------.....
CB4_BR  : ----------------------------------------------------------------------------------------------------.....
CB4_CR  : ----------------------------------------------------------------------------------------------------.....
CB4_DR  : ----------------------------------------------------------------------------------------------------.....

              100         *       120         *       140         *       160         *       180         *       200
N*03601 : QLMYGCDVGPDGRLLRGYEQYGYDGRDYIALNEDLRSWTAADTAAQITKRKVEAAGDAEGHRNYLEGRCVEWLRRYLENGKDTLLRADPPKAHVTHHPISGREVTLRCWAL
CB4_AF  : ..............................----------------------------------------------------------------------------
CB4_BF  : ..............................----------------------------------------------------------------------------
CB4_CF  : ..............................----------------------------------------------------------------------------
CB4_DF  : ...................................----------------------------------------------------------------------
CB4_AR  : ..........................................................................------------------------------
CB4_BR  : ..........................................................................------------------------------
CB4_CR  : ..........................................................................------------------------------
CB4_DR  : ..........................................................................------------------------------
```

Figure 3.4. Alignment of sequences derived from clones of gene 3-specific PCR products from Canadian bull 4 aligned with the *N*03601* reference sequence. Dots indicate identity, dashes indicate gaps/lack of sequence data compared to reference.

3.3.3 Sequence-specific primer pairs (SSP-PCR)

PCR amplification of *N*01901* on the W12B haplotype, *N*01201* on the A10 haplotype and *N*02401* on the A14 and A15 haplotypes is currently not possible due to the inability of the existing gene-specific primers to anneal. In addition, RSCA is unable to detect the *N*03101* allele on the A13 haplotype. This results in a deficit in haplotype information. Previous cattle MHC-typing has revealed that in all except chimeric animals, the MHC allele combinations of each haplotype remain conserved i.e. if the *N*00201* allele of the A10 haplotype is detected, the *N*01201* allele is also present. A panel of 15 sequence-specific primer pairs (SSPs) were previously designed for the identification of 15 different class I alleles from 10 serologically defined haplotypes (Ellis et al., 1998). With the availability of more sequence data, additional SSPs have also been developed in the laboratory (unpublished work by J Birch). From the Ellis et al. (1998) and laboratory SSP primer panels, SSPs for detection of the A13 haplotype and confirmation of the W12B and A10 haplotypes were applied to samples requiring confirmation of these haplotypes (see Table 3.1). The PCR mixes and thermal cycling conditions are outlined in Chapter 2 section 2.2.6. Sequencing of generic class I PCR products from the apparently A14/A14 homozygous Canadian bull 41 detected the *N*00801* allele of the W12B haplotype and amplification with primer pair 4221.1 confirmed the presence of the *N*01901* allele on the W12B haplotype in this animal (see Figure 3.5).

HAPLOTYPE	ALLELE	SSP NAME	CODING (5'-3')	NON-CODING (5'-3')
W12B	*N*01901*	4221.1	CCGGCCCGGCCTCGA	CACGTCGCAGCCGTACATG
A10	*N*01201*	A10.2	GGATCGGGAGACGCGAAACTT	CGTAGGCGTCCTGCCTGTAT
A14 & A15	*N*02401*	D18.1	CCCGCTTCATCACCGTT	ACTGTACCCGCGGAGGAAACT
A13	*N*03101*	A12.1	TCTGAAGTATTTCTACACCGCCGTA	CCTGGATATTGTGAGACCCTGCG

Table 3.1. Table of SSP pair sequences used to confirm presence of specific alleles.

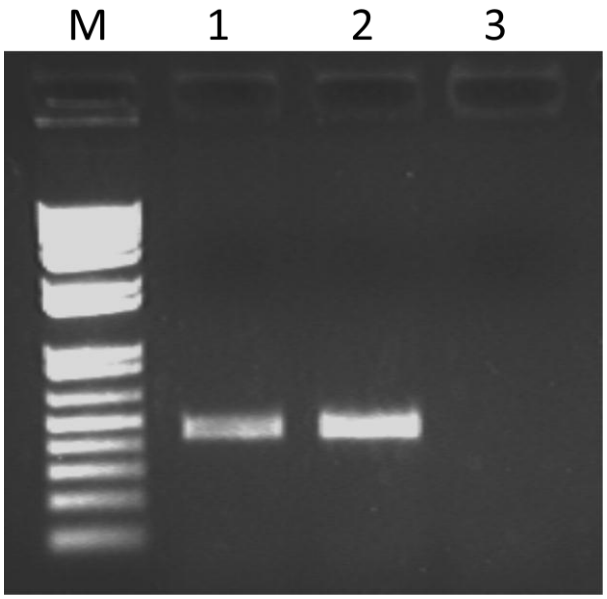


Figure 3.5. Amplification of the *N*01901* allele present on the W12B haplotype using 4221.1 SSPs in Canadian bull 41.

3.3.4. MHC class I allele frequencies in the 2006 sample

The haplotype data derived from 48 Canadian AI bull samples in this project were collated with data from 25 previously typed Canadian AI bull samples (raw data shown in the Appendix) and analysed to obtain frequencies of individual alleles within the sample. Table 3.2 and Figure 3.6 show the estimated allele frequencies for each gene for the sample of 72 Canadian Holstein bulls. Due to the variable nature of bovine haplotypes i.e. not all genes present, the frequency of having a null allele for each gene was also included. As such, the term null allele refers to the absence of an allele at that gene and not that the animal was not typed at that locus.

Gene 2 was shown to have the greatest number of alleles, with a total of eight detected in the 2006 sample set with frequencies ranging from 0.0069 to 0.3056. This is unsurprising given that a gene 2 allele is expressed on the majority of defined cattle class I haplotypes (Birch et al., 2006). Furthermore, the frequency of having a null allele was lowest for gene 2. Genes 1 and 3 were each shown to have a total of four class I alleles. The gene 3 allele frequencies were lower than those of the gene 1 alleles and as such the frequency of having a null allele was greater for gene 3 than gene 1. Genes 4 and 6 each had a single allele and the highest frequencies of a null allele at 0.6944 and 0.9306 respectively. Despite being the only gene 4 allele in the whole data set, *N*02401* is found at the highest frequency together with the gene 2 allele *N*02501*. The high frequency of these two alleles is a result of each being expressed on both the A14 and A15 haplotypes.

GENE	ALLELE	ASSOCIATED HAPLOTYPES	FREQUENCY
1	*N*00901*	A15	0.1806
	*N*01901*	W12B	0.1875
	*N*02301*	A14	0.1250
	*N*03101*	A13	0.0486
	Null		**0.4583**
2	*N*00801*	W12B	0.1875
	*N*01201*	A10	0.1181
	*N*01601*	A19	0.0694
	*N*01801*	A11	0.1181
	N*02501	**A14 & A15**	**0.3056**
	*N*02601*	A20	0.1111
	New22	New22	0.0069
	New24	New24	0.0069
	Null		0.0764
3	*N*00201*	A10	0.1181
	*N*01701*	A11	0.1181
	*N*02701*	A20	0.1111
	*N*03601*	New5	0.0278
	Null		**0.6250**
4	*N*02401*	A14 & A15	0.3056
	Null		**0.6944**
6	*N*01401*	A19	0.0694
	Null		**0.9306**

Table 3.2. Table of MHC class I allele frequencies found in 2006 sample analysed by RSCA, SSP amplification and sequencing. The most common allele for each gene is highlighted in bold.

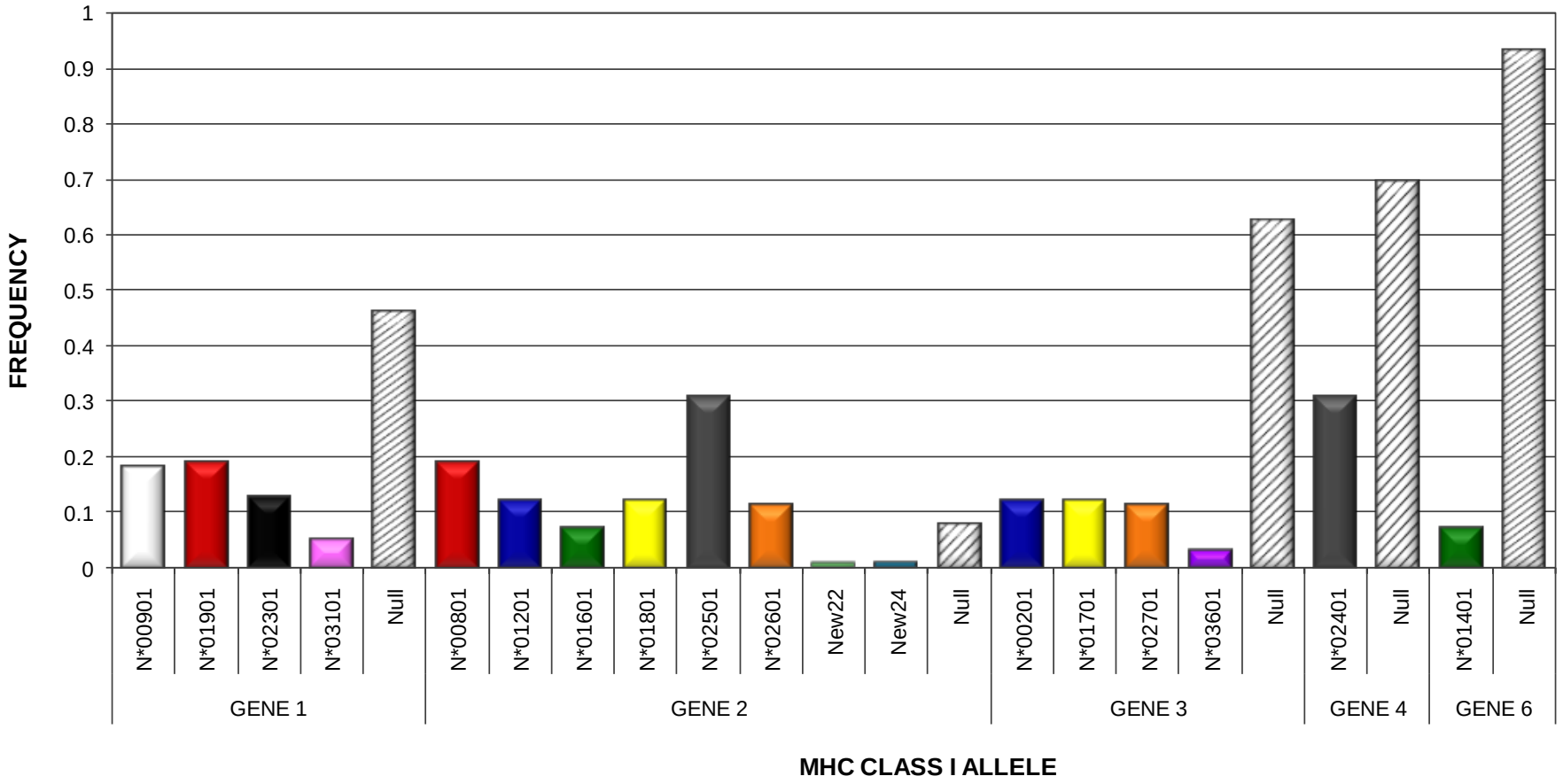


Figure 3.6. The allele frequencies observed for each of the MHC class I genes in the 2006 sample. Bar colour indicates the class I haplotypes on which the alleles are expressed e.g. *N*02401* and *N*02501* alleles shared by the A14 and A15 haplotypes are coloured in grey, whilst the differing gene 1 alleles of A14 and A15 are coloured in white and black respectively.

3.3.5. MHC class I haplotype frequencies

The 2006 sample set contained eleven MHC class I haplotypes with a range in frequency from 0.007 to 0.188 (see Table 3.3 and Figure 3.7). Of the eleven haplotypes, eight have been previously defined at the molecular level (Birch et al., 2006). The remaining three previously undefined haplotypes appeared to express only a single gene which is less common in bovine class I haplotypes as the majority of known haplotypes express at least two genes. The sample was dominated by two common haplotypes; A15 and W12B. The A14, A11, A10 and A20 haplotypes were all present at similar intermediate frequencies ranging from 0.111 to 0.125. The remaining 5 haplotypes were detected at relatively low frequencies, with New22 and the New24 haplotypes being detected at the lowest frequency of 0.07.

HAPLOTYPE	NO. OF ANIMALS	FREQUENCY
W12B	27	0.188
A15	26	0.181
A14	18	0.125
A10	17	0.118
A11	17	0.118
A20	16	0.111
A19	10	0.069
A13	7	0.049
New5	4	0.028
New22	1	0.007
New24	1	0.007

Table 3.3. Table of MHC class I haplotype frequencies found the 2006 sample set, including the number of animals in which each haplotype was detected, using RSCA, SSP amplification and sequence analysis.

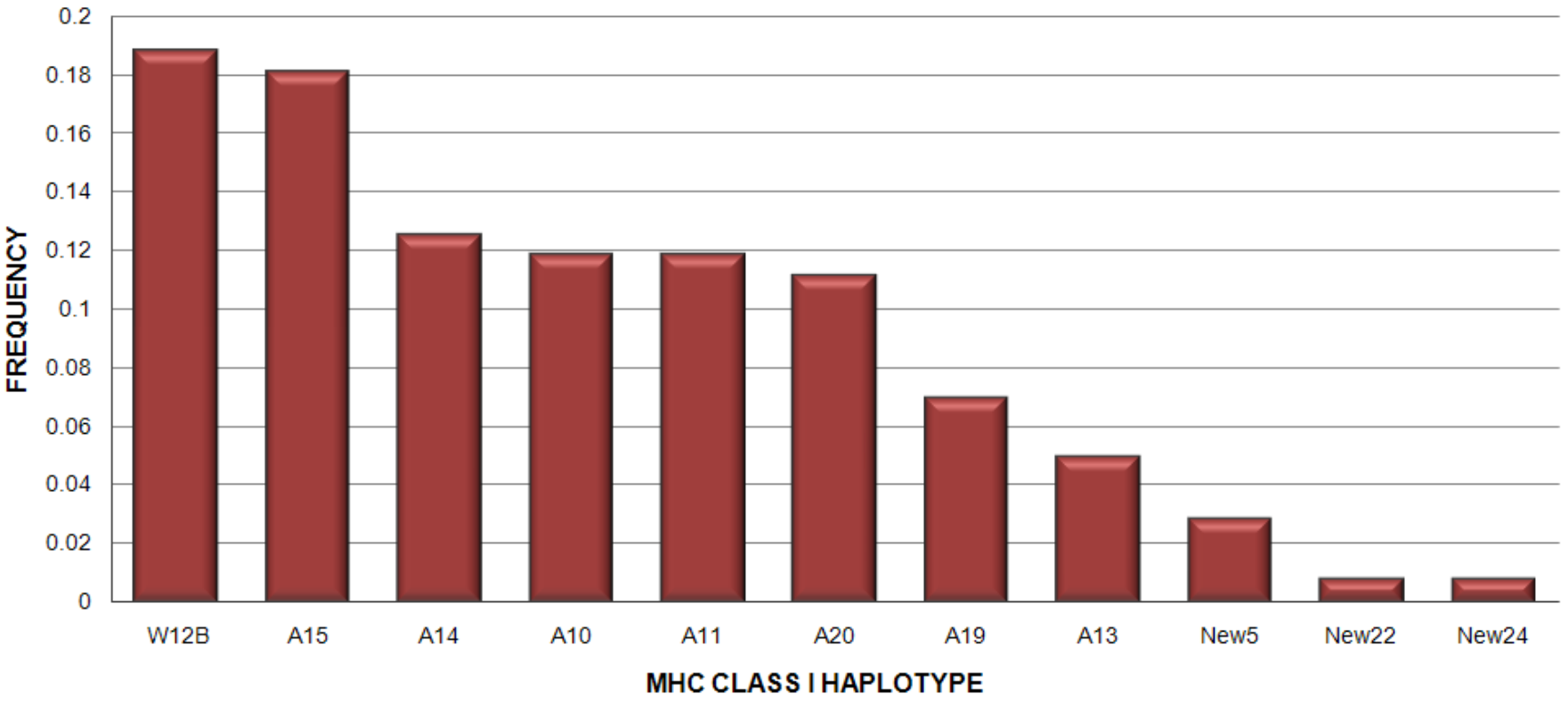


Figure 3.7. Distribution of MHC class I haplotypes in the 2006 Canadian Holstein bull sample set.

3.3.6. Haplotype frequencies over time

A previous study investigating *BoLA* class I diversity typed a sample of 271 Canadian Holstein bulls serologically (Batra et al., 1989; Figure 3.8). These data provide a suitable reference for comparing haplotype frequencies and patterns of class I diversity in the Canadian bull population over a 20 year period of selection. For reference, the 271 Canadian Holstein bulls typed by Batra et al. shall be referred to as the 1986 sample set. The most notable differences between the 1986 and 2006 sample sets are the frequencies of the A19 and W12B haplotypes (Figure 3.9). The 1986 sample was dominated by A19, present at a frequency of 0.26, but this has since reduced dramatically to 0.007. Conversely, W12B has undergone a nine-fold increase in frequency to become the most common haplotype in the 2006 sample at 0.188. The most common 2006 haplotype (W12B) has not yet reached the dominating frequency of the A19 haplotype in 1986 (0.26) and it is likely that the frequency of W12B may still be increasing over the next few years.

The A10, A11, A13, A14, A15 and A20 haplotypes have all maintained similar frequencies over time with only small fluctuations between the sample sets. Our data suggest that haplotypes present at low frequencies in 1986 (A17, CA24, CA42B, W16, W7 and W9.2) have been lost over time. Similarly, the 2006 sample has two haplotypes (New5 and New22) that were not present in the older sample which may reflect the introduction of new haplotypes into the bull population. However, as information linking serological to molecular definitions of bovine MHC haplotypes is lacking in some cases it is possible that CA24 found in the 1986 sample may in fact be the New5 or New22 haplotype detected in the 2006 sample.

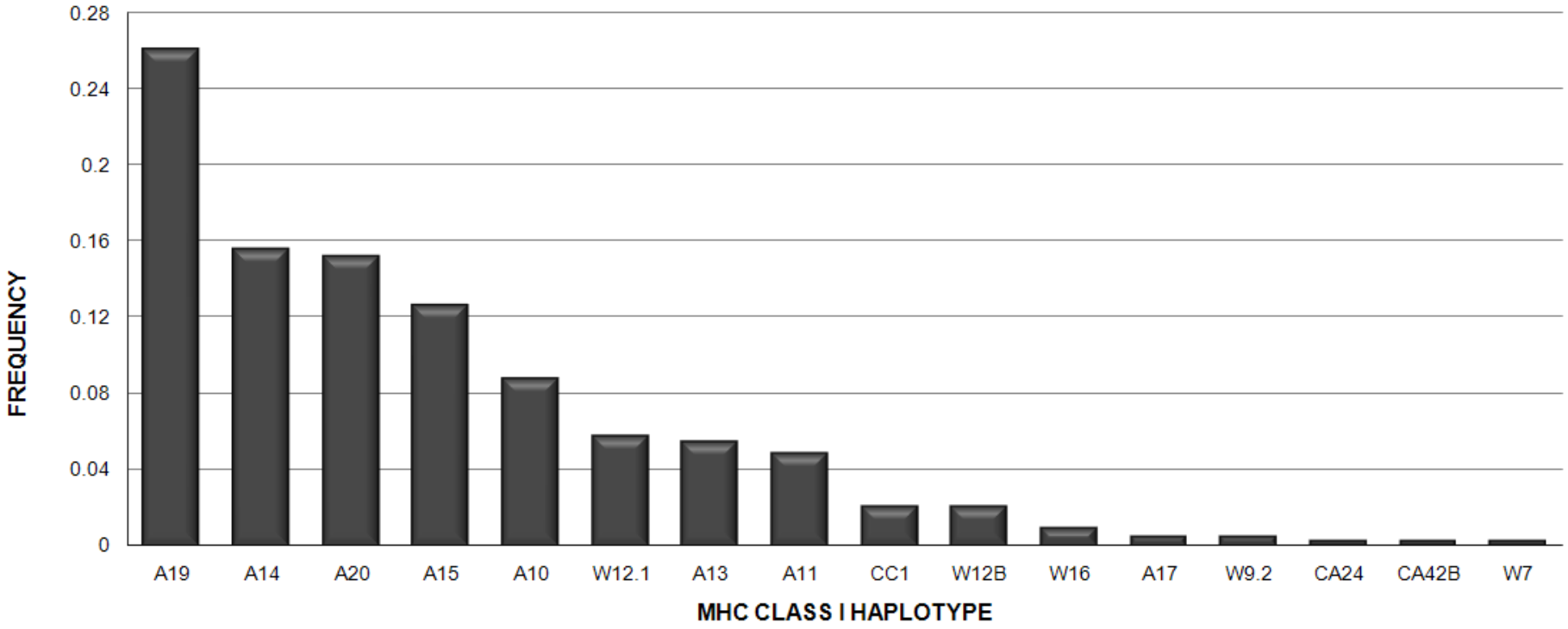


Figure 3.8. Distribution of MHC class I haplotypes in the 1986 Canadian Holstein bull sample set.

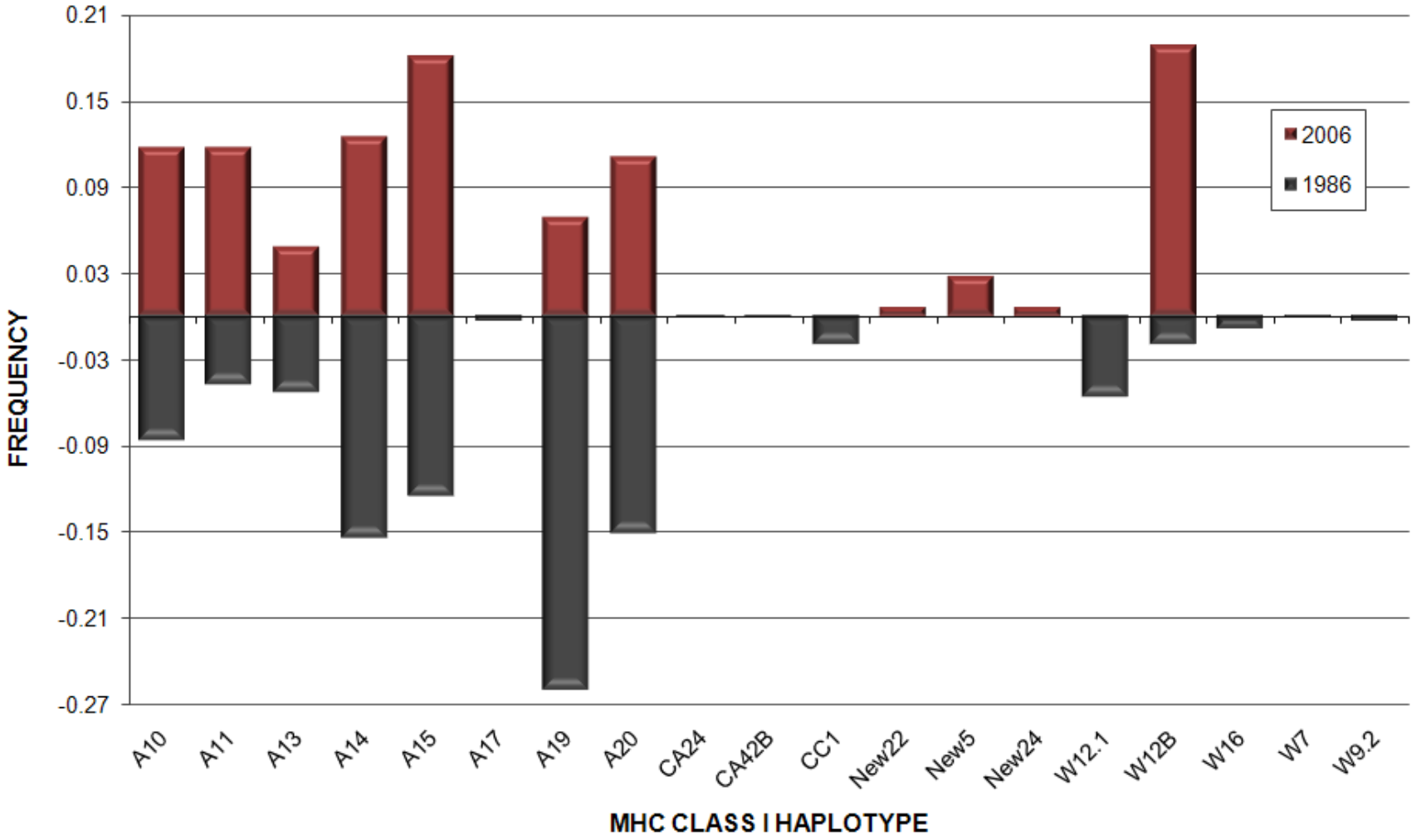
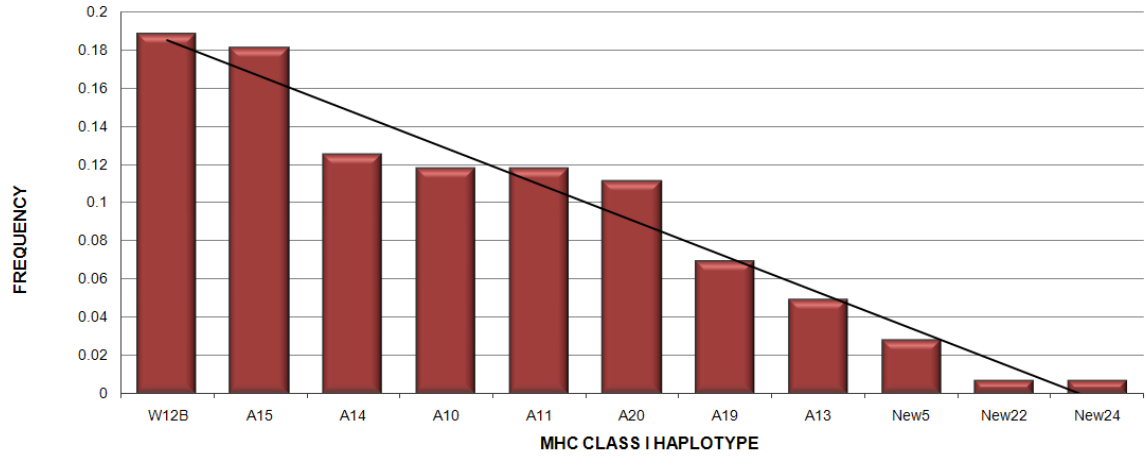


Figure 3.9. Comparison of the observed frequencies of MHC class I haplotypes in the 2006 and 1986 Canadian Holstein bull data sets.

3.3.7. Level of MHC class I diversity over time in Canadian Holstein bulls

The larger 1986 sample contained a total of 16 MHC haplotypes, compared to 11 haplotypes detected in the 2006 sample. Furthermore, the frequencies of the 11 haplotypes detected in the 2006 sample were distributed more evenly and gave a 'flatter' haplotype distribution (Figure 3.10A). A consequence of this 'flatter' distribution is that there are a higher number of heterozygotes (93 %) than expected (86 %) by Hardy-Weinberg calculations i.e. the 2006 sample had two common haplotypes, as opposed to being dominated by a single common haplotype as in 1986 resulting in greater heterozygosity in the 2006 sample than would be expected. The dominance of the A19 haplotype in the 1986 sample gave rise to a 'graduated' distribution (Figure 3.10B). Mathematical analysis (data not shown) was carried out to determine if there had been a loss in diversity between 1986 and 2006, or whether the observed differences were merely an artefact of sample size. This analysis revealed that the flatter distribution of haplotypes appears to maintain diversity despite the reduction in haplotype number.

A.



B.

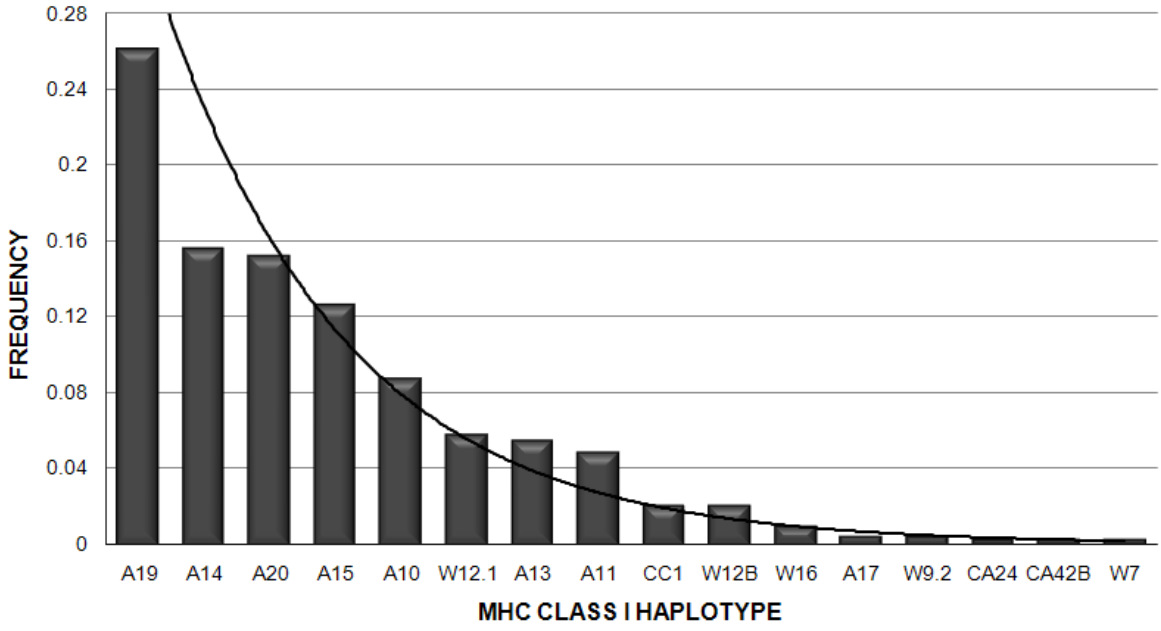


Figure 3.10. A comparison of the haplotype frequency distribution patterns in the Canadian Holstein population sampled in **A.** 2006 and **B.** 1986.

3.3.8. Mathematical modelling the effects of inbreeding and genetic drift in Canadian Holstein bulls

The effects of inbreeding and genetic drift were simulated in order to ascertain whether the observed changes in haplotype frequencies could be attributed to either of these processes. The process of genetic drift causes random fluctuations in gene frequency over time without regard to the phenotypic effect of the genotype in question and drives a population towards genetic uniformity over time, causing the loss of certain genotypes from a population, and in extreme cases leading to the fixation of certain genotypes i.e. a particular genotype is found at a frequency of 1.00 in the population. The effect of genetic drift is more marked in small populations, with the loss of genotypes occurring over a smaller number of generations e.g. it is estimated that it takes approximately six generations for an allele to be lost from a population when the effective population size is ten, compared to a timescale of twelve generations for an allele to be lost from a population with an effective population size of two hundred (Kimura & Ohta, 1969). Inbreeding in domestic animal populations have often been used to 'fix' certain desirable traits in a population, but increased relationships between individuals in a population causes the level of heterozygosity to decrease. It is estimated that the current level of inbreeding in the Holstein breed is approximately 12 % (Hansen, 1999). For the simulations of drift and inbreeding, the pedigree of the 2006 Canadian Holstein AI sample was traced back through four generations. This pedigree information was then used as a framework for the stochastic simulations of gene drift, using the haplotype frequency data from 1986 as the starting frequencies. The use of the pedigree information in these simulations dictates that any relationships between animals in the 2006 Canadian Holstein bulls i.e. shared ancestry, and thus the level of inbreeding in this population, is integrated to this model.

Table 3.4 presents the results of 5000 mathematical simulations of genetic drift and inbreeding in the Canadian Holstein population. An average of 10.5 haplotypes survived through four generations; a similar figure to the observed 10

haplotypes in 2006. The haplotype frequencies were estimated for the surviving haplotypes in each simulation and therefore overestimate the frequencies for those haplotypes that did not survive in all simulations. In the simulation, A19 had a mean frequency of 0.26 which was identical to the starting frequency in 1986 but considerably more than the observed frequency in 2006 of less than 0.07. The stochastic fluctuations caused by drift and inbreeding are as likely to increase as to decrease the haplotype frequency. Therefore it is not surprising that the simulated mean frequency was the same as the starting frequency. More relevant is the observation that the lowest simulated frequency in 5000 repeats was 0.11. This is considerably higher than the frequency observed in 2006. Consequently, the observed change in frequency of the A19 haplotype is unlikely to have been due to drift alone. Similarly, the observed frequency in 2006 of the W12B haplotype was 0.19. This is incompatible with the highest frequency observed in 5000 simulations, which was 0.12. These simulations demonstrate that the observed changes in the frequency of A19 and W12B were unlikely to be due to inbreeding or drift alone.

HAPLOTYPE	SIMULATIONS HAPLOTYPE SURVIVED (n)	SIMULATED HAPLOTYPE FREQUENCY			OBSERVED HAPLOTYPE FREQUENCY (2006)
		MINIMUM	MAXIMUM	AVERAGE	
A10	4999	0	0.250	0.087	0.118
A11	4964	0	0.236	0.048	0.118
A13	4989	0	0.194	0.054	0.049
A14	5000	0.034	0.340	0.156	0.125
A15	2000	0.021	0.284	0.125	0.181
A17	1656	0	0.063	0.012	0
A19	5000	0.111	0.444	0.260	0.069
A20	5000	0.028	0.354	0.151	0.111
CA24	886	0	0.076	0.011	0
CA42B	924	0	0.097	0.011	0
CC1	4273	0	0.118	0.022	0
W12.1	4988	0	0.188	0.057	0
W12B	4354	0	0.118	0.023	0.188
W16	3030	0	0.090	0.015	0
W7	888	0	0.090	0.011	0
W9.2	1601	0	0.069	0.012	0

Table 3.4. Results of 5000 simulations of genetic drift and inbreeding in the known pedigree to estimate MHC class I haplotype frequencies after four generations.

3.4. Discussion

In this study, the haplotypes of 72 Canadian AI bulls were established. The heavy reliance upon a small population of Canadian AI bulls for UK dairy breeding qualifies these bulls as a preliminary indicator of MHC diversity within the UK herd. To date 16 cattle MHC class I haplotypes have been defined at the molecular level, of which only eight were detected in the 2006 Canadian AI bull sample. A further three haplotypes, which have yet to be defined at the molecular level, were also detected.

Although data on *BoLA* class I allelic diversity in Holstein-Frieisian populations is relatively scarce, more data is available as to MHC class I haplotype diversity. A previous study using a sample of 271 animals from the Canadian Holstein bull population was conducted in 1986. This larger sample was shown to contain a greater number of class I haplotypes (16) than were observed in the 2006 Canadian Holstein bull sample set (11). Whilst the 1986 sample was dominated by a single haplotype at high frequency; A19 at 0.26, the 2006 sample contained two common haplotypes; W12B and A15 at frequencies of 0.188 and 0.181 respectively. The 2006 sample also showed a 'flatter' haplotype frequency distribution than the 1986 sample. Further statistical analysis showed that there was no apparent loss of diversity over the 20 year period of selection. The statistical analysis also demonstrated that the differences in haplotype frequencies are not simply an artefact of the differences in sample size i.e. 271 bulls in 1986 compared with 72 in 2006.

The study by Batra and colleagues also investigated class I diversity in a sample of Canadian Holstein cows as a comparison to the highly selected bull population. The Canadian Holstein cow sample of 179 animals was shown to contain only 10 class I haplotypes (Figure 3.11). The distribution of haplotype frequencies appeared intermediate between the 1986 and 2006 Canadian bull samples i.e. although dominated by a single haplotype (A10) the haplotype frequency distribution produced a 'flatter' curve than that of the 1986 Canadian

bull sample. The haplotypes found in the 1986 Canadian cow population reflect those seen in the 2006 Canadian bull sample i.e. seven of the haplotypes are found in both samples, with the remaining haplotypes in each case lacking molecular and/or serological definition and so it is possible that further overlap between the two samples exists. The similarity between the 1986 cow population and the 2006 bull population may reflect the fact that this Holstein cow population, part of the National Cooperative Dairy Cattle Breeding project, may represent the founding genetics of current bull populations or may reflect intensive selection for production traits alone in these two populations. However, the most important observation is that there has been no apparent loss in class I diversity over 20 years of selection.

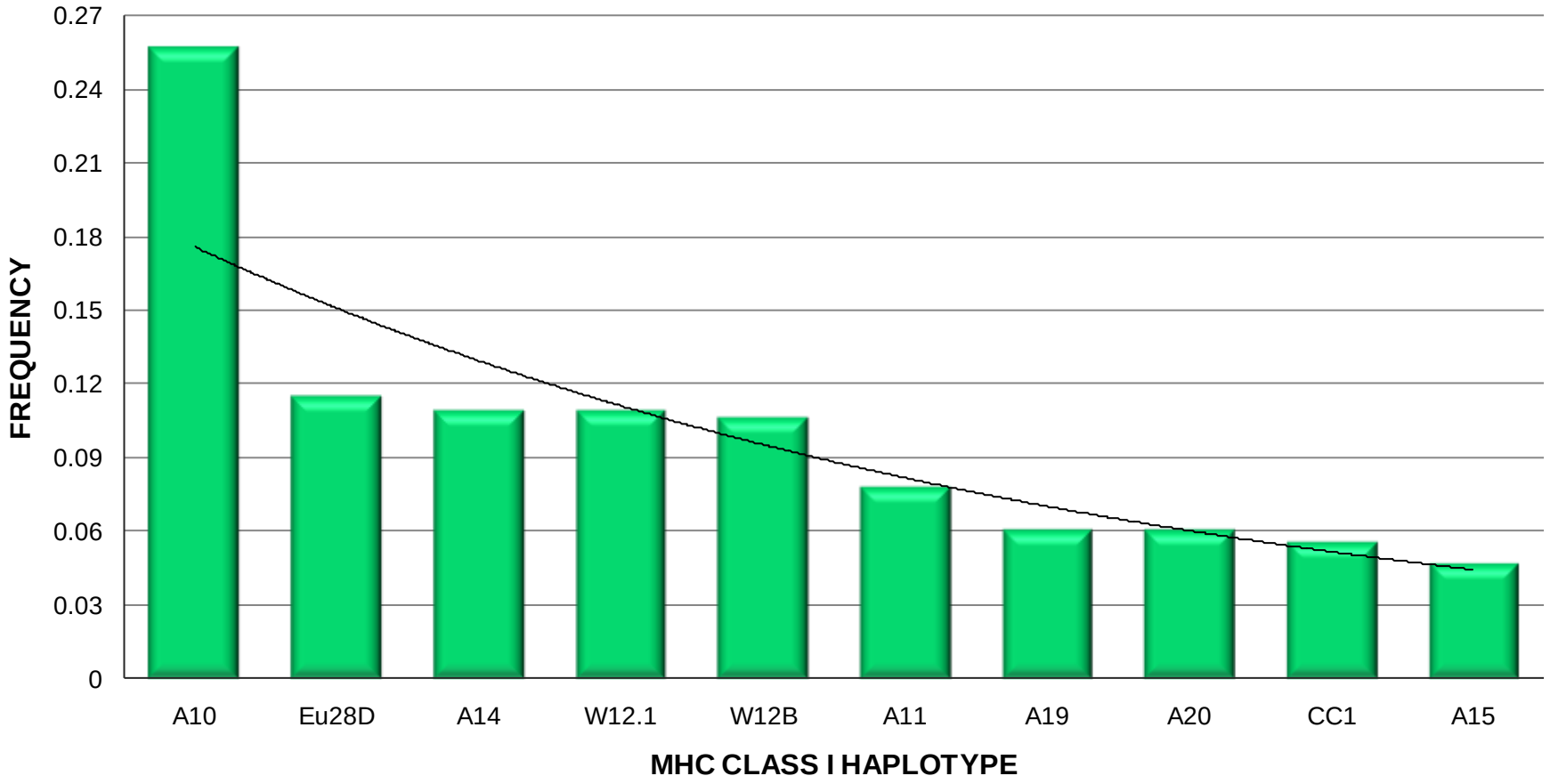


Figure 3.11. Observed frequencies of MHC class I haplotypes in 179 Canadian Holstein cows sampled in 1986 typed by serological methods (Batra et al., 1989)

The eleven haplotypes detected in the 2006 sample represent 18 alleles across 5 loci, which is a relatively small number given that over 80 *BoLA* class I alleles have been submitted to the IPD database. This may indicate a low level of diversity in this sample or in this breed. The *N*02401* and *N*02501* alleles associated with the A14 and A15 haplotypes are present in the 2006 Canadian AI bull population at a markedly higher frequency than any other allele (Table 3.2). As these two haplotypes share the *N*02401* and *N*02501* alleles, it is possible that selection is based directly upon one or both of these alleles or indirectly to a linked gene, and gives rise to the high frequency of these two haplotypes in the Canadian bull population. Currently, the role of the distinguishing gene 1 alleles in each of these haplotypes in terms of trait selection is unclear.

The distribution of MHC class I haplotypes may be a consequence of prolonged inbreeding i.e. the pattern of MHC diversity is not linked to trait selection markers. To investigate whether inbreeding and genetic drift were influencing MHC diversity in the 2006 sample, mathematical modelling was applied to simulate the effects of inbreeding and drift on the flow of MHC class I haplotypes observed in 1986 over four generations. The results from this model were incompatible with the observed gene frequencies indicating that these two processes alone were not responsible for the pattern of MHC diversity observed in the Canadian bull population in 2006. Trait selection is now implied as an instrumental force in shaping MHC diversity and will be dissected further in Chapter 4.

In the 2006 sample, three alleles that have yet to be assigned to a haplotype were detected. Of these three alleles, only the *N*03601* allele has been submitted to IPD. The alleles named New22 and New24 have not been submitted to IPD or Genbank as full length cDNA sequences could not be obtained i.e. only genomic DNA from these animals was available. The *N*03601* allele is relatively uncommon (data unpublished) but was found in four of the 72 animals in this sample. Further investigation into the pedigree of these animals revealed a common ancestor (Pawnee Farm Arlinda Chief) from whom this allele is likely to have been inherited.

In humans, there are marked differences in function, levels of expression and allelic diversity between the three classical class I loci. *HLA-A* and *HLA-B* are both reported to be involved in antigen presentation to CD8+ T cells (de Campos-Lima et al., 1997; McMichael, 1978; Solache et al., 1999), whereas *HLA-C* functions as the ligand for inhibitory killer-cell immunoglobulin-like receptors (KIR; Colonna et al., 1993). Expression of *HLA-C* on the cell surface is markedly lower than that of both *HLA-A* and *HLA-B* (Snary et al., 1977). The lower levels of cell surface expression of *HLA-C* has reduced the selection pressure and subsequently reduced allelic diversity at this locus i.e. there are currently 690 *HLA-C* alleles described in contrast to the 1001 and 1605 alleles reported for the *HLA-A* and *HLA-B* loci respectively (http://hla.alleles.org/nomenclature/stats.html). The high number of alleles reported and greater selection e.g. the selection pressure of HIV-1 on *HLA-B* is 4.4 fold greater than on *HLA-A* (Kiepiela et al., 2004), is leading to more rapid diversification of *HLA-B*. It is highly plausible that a similar situation exists in cattle i.e. one locus is under greater selection pressure, resulting in a higher number of alleles at that locus. The expression of a gene 2 allele on the majority of cattle haplotypes (Birch et al., 2006) and the high number of gene 2 alleles may indicate a higher rate of intra-locus recombination. The two 'new' (either rare or previously unseen) alleles, New22 and New24, which have not been submitted to IPD were both gene 2 alleles. Increased selection pressures and/or an increased mutation rate at gene 2 are likely to be responsible for the greater allelic diversity observed at this locus. Based on allelic diversity observed in this population genes 1 and 3 could be categorised as analogous to *HLA-A*, whereas genes 4 and 6 are more similar to *HLA-C*.

HLA studies can also be used as a means of comparing diversity at the population level. Whilst under the classification of the same species, cattle populations are separated into breeds on the basis of phenotype and geographic origin in a manner analogous to ethnicity in human populations. Assuming the large Holstein population is outbred, it would be expected that levels of diversity would be similar to that of a large ethnic groups in the human population e.g. American Caucasians. Examining *HLA* class I diversity in a sample of 265 American Caucasians revealed the presence of 97 alleles (28 *HLA-A*, 47 *HLA-B*,

and 22 *HLA-C* alleles) of which the most prevalent was detected at a frequency of 0.2717 (Cao et al., 2001). Similar levels of diversity were recorded in a sample of 564 African Americans, with a total of 145 alleles present at frequencies ranging from 0.00089 to 0.19592 (Tu et al., 2007). In contrast, the geographical isolation of the Chinese Drung ethnic minority, an aboriginal population of 5884 people, has given rise to a relatively limited MHC diversity. Across the three class I *HLA* loci, 32 alleles were detected in 86 individuals of the Drung population, with the most common allele present at a frequency of 0.770 (Chen et al., 2007). A study of the Mixe Amerindian natives of the south Mexican state of Oaxaca provide a less extreme example of limited class I diversity (Hollenbach et al., 2001). In this linguistically isolated indigenous population, currently numbering approximately 100, 000 persons, 30 alleles were detected across *HLA-A*, *HLA-B* and *HLA-C* in a sample of 52 individuals. In this population the *HLA-B* locus had the greatest range in frequencies with 13 alleles present at frequencies from 0.009 to 0.378.

Comparing *BoLA* diversity in the 2006 sample with *HLA* studies revealed a level of diversity more similar to that of the Chinese Drung and Mixe Amerindian populations rather than the large outbred American Caucasian and African American populations. Although the *N*02501* and *N*02401* alleles were both detected at the high frequency of 0.3056 (see Table 3.2.) in the Canadian AI bull sample, this is still not as high as the most prevalent alleles in either the Drung or Mixe Amerindian populations. In terms of the total number of alleles detected, the Canadian AI bull population is much more similar to the Drung and Mixe Amerindian studies as the total number of alleles detected in each of these studies was restricted to a small fraction of all known class I alleles, implying relatively low levels of diversity.

Dissecting the impact of artificial selection upon MHC class I diversity

4.1. Introduction

Dairy bulls are highly selected animals, based upon their ability to pass on desirable production, functional and conformation traits to their progeny; a number of these traits are not phenotypically expressed by the bull e.g. milk yield. Selection of premium breeding stock has been facilitated by extensive progeny-based genetic evaluations for over 60 traits, the data from which are used to produce publicly available bull proofs. Each bull must first undergo genetic evaluation or proving during which the performance records from at least 20 of his daughters distributed across a minimum of 10 herds are collated. In addition to the accumulation of data from progeny, data from siblings and parent pedigrees enable the estimation of genetic merit for a particular trait, with the reliability of these estimates growing with the number of progeny. The data acquired is used to estimate the genetic merit of an animal in the form of estimated breeding values (EBVs) which is an estimate of the total genetic merit of an animal in comparison to the population average, otherwise known as the genetic base.

The intensive selection of Holstein-Friesian bulls, resulting in a small effective population size (n = 150; Hayes et al., 2003) has prompted the hypothesis that dairy cattle may be subject to a reduction in MHC diversity. Detailed analysis of the MHC class I region in the Canadian Holstein bull population sampled in 2006 showed that there are 11 haplotypes in 72 highly selected dairy bulls with no evidence for a reduction in MHC diversity over the last 20 years (see Chapter 3). The relative frequency of some haplotypes has however changed significantly over this period; the most common allele in 1986 has decreased in frequency,

whilst a formerly rare allele has become the most common. In addition, there are fewer homozygotes among the selected bulls than expected. Mathematical simulations showed that the observed changes in MHC class I frequencies could not be attributed to genetic drift and inbreeding alone, implicating trait selection as an influential force acting upon MHC diversity.

Balancing selection, in one or more of its several forms, is widely proposed to be responsible for the generation and maintenance of diversity in MHC genes. This could encompass heterozygote advantage, frequency-dependent selection, or selection that varies in time and/or space (Apanius et al., 1997; Borghans et al., 2004; De Boer et al., 2004). Heterozygosity at MHC loci is considered advantageous because individuals with two different alleles can have the capacity to present a wider range of pathogen-derived peptides than homozygotes, thus may be resistant to a greater variety of pathogens (Clarke & Kirby, 1966; Doherty & Zinkernagel, 1975). It is believed that this diversity is driven and maintained by pathogen-mediated selection in natural populations (Parham et al., 1995). Data supporting heterozygote advantage in the MHC are relatively scarce; examples include disease progression in HIV-infected individuals (Carrington et al., 1999), mastitis susceptibility in Holstein cattle (Takeshima et al., 2008) and risk of parasitism in naturally-infected lambs (Stear et al., 2005). The availability of detailed pedigrees and progeny production records in dairy cattle provides an ideal system in which to dissect the impact of trait selection upon MHC diversity and its mechanism.

4.2. Acquisition of data

Blood samples from 72 Canadian Holstein bulls standing with Semex (The Semex Alliance, Guelph, Canada) were donated to our laboratory in 2006. Genomic DNA was extracted from each of these samples prior to amplification of MHC class I genes 1, 2, 3 and 6. Gene-specific PCR products were then analysed by RSCA performed as described in Birch et al. (2006) and each sample was assigned an

MHC haplotype prediction based upon mobility values corresponding to previously sequenced alleles. Limitations of gene-specific PCR and RSCA required the use of SSPs and sequencing of generic and/or gene-specific PCR products to complement missing haplotype data and/or confirm ambiguous typing results. Finalised typing results were then used to calculate the MHC class I haplotype frequencies for this sample, as shown in Table 4.1 (taken from Chapter 3).

HAPLOTYPE	FREQUENCY
W12B	0.188
A15	0.181
A14	0.125
A10	0.118
A11	0.118
A20	0.111
A19	0.069
A13	0.049
New5	0.028
New22	0.007
New24	0.007

Table 4.1. Table of MHC class I haplotype frequencies found the 2006 sample set analysed by RSCA, SSP amplification and sequencing.

Functional trait and type data for each of the MHC-typed Canadian AI bulls is freely available in the form of bull proofs. The data is based upon daughter performance and conformation and enables stockmen to identify suitable bulls for their own breeding schemes. Proofs for each of these bulls were released in January 2009 by the Canadian Dairy Network (CDN). As of April 2009, these bulls had a total of 222,178 daughters with a range of 53 to 33,350 daughters per bull. Estimated breeding values (EBVs) for production (milk yield (kg), milk protein (kg), milk fat (kg)) and functional traits (daughter fertility, somatic cell score, herd life, lifetime profit index) were recorded for these daughters in accordance with conventional practice (http://www.cdn.ca/query/individual.php; Table 4.2). An EBV represents the genetic potential of each bull relative to the genetic base, of which half is transmitted to his progeny i.e. if a bull has an EBV for milk yield of +2000 kg, then his daughters are expected to produce +1000 kg more milk than contemporaries sired by average bulls with a milk yield EBV of 0.

TRAIT	DETAILS
Milk, Fat and Protein Yields	Expected yields of milk, fat and protein during a 305-day lactation in a herd of average management. Estimated breeding values (EBVs) for yield are measured in kilograms compared to the breed average/genetic base. For milk yield, values range from -1500 to +2900 kg and the average EBV is +100 kg. Milk fat yield values range from -50 to +115 kg, with an average EBV of +10 kg. Milk protein yield EBV values range from -43 to +89 kg, with a breed average of + 5kg. Positive values are desirable.
Daughter Fertility	Measurement of the expected fertility of a bull's daughter across all lactations. Four measures of daughter fertility are used in the calculation: (1) the age at first insemination for virgin heifers, (2) the 56-day non-return rate for virgin heifers, (3) the interval between calving and first insemination for cows, and (4) the 56-day non-return rate for milking cows. The published Daughter Fertility proof based on a relative emphasis of 65 % to increase the cow non-return rate, 25 % to reduce the interval from calving to first service in cows and 10 % to reduce the age at first insemination for virgin heifers. The published proof is expressed as relative breeding value (RBV), with the breed average set to 100, and a range of five standard deviations i.e. the range in bull proof values will be from 115 to 85.
Herd Life	Expected number of lactations the daughter is expected to complete, compared to the average bull. This value is independent of production levels. Also expressed as an RBV, with values ranging from 85 (undesirable) to 115 (desirable) and a breed average of 100.

Somatic Cell Score (SCS)	SCS is a measure of white blood cells in the milk, and is used as a marker of mastitis resistance. Animals susceptible to mastitis infections have a high SCS. The bull proof value estimates the SCS of daughters over the first three lactations, and is recorded on a logarithmic scale. The average bull is 3.00 but values range from 2.50 (desirable) to 3.50 (undesirable).
Conformation Traits	Expected relative superiority of first lactation daughters for each type trait. The final conformation score and the four major scorecard traits: legs and feet, mammary system, dairy strength and rump were used in the analysis. Genetic evaluations for conformation are standardized to a common scale for each trait such that the average bull proof is set to 0 and the standard deviation is set to 5. This gives a range of evaluations from nearly –20 to +20 for bulls and cows (i.e. four standard deviations away from the breed average).
Lifetime Profit Index (LPI)	Multi-component index used to estimate the relative lifetime profit of future daughters based upon their genetic potential for traits encompassed by three components, each with a different weighting; production traits (protein and fat yield; 54 %), health and fertility (SCS, udder depth, milking speed, daughter fertility and lactation persistency; 10 %), and durability (herd life, mammary system, feet and legs, dairy strength; 36 %). Detailed LPI formula for the Holstein breed is shown in appendix. LPI is recorded on a point scale relative to the breed average which is set to zero. It is estimated that for every +200 point difference in LPI above the breed average (0) translates to a profit of $50 more profit per daughter per year, which accumulates year on year.

Table 4.2. Details of selection traits used in the gene substitution model.

4.3. Statistical analysis methodology used in analysis of MHC class I haplotype associations with selection traits

The statistical analysis of MHC class I haplotype associations with selection traits was carried out using the SAS computer package with the help of Prof. M. Stear. Means and standard deviations for the estimated breeding values (EBV) using these traits were calculated with the univariate procedure on SAS and are presented in Table 4.3. A gene substitution model was used to estimate the average effect of each haplotype, and a separate model was used for each trait EBV. The haplotypes were fitted as covariates. To reduce dependencies among the equations, the most common haplotype (W12B) was set to zero because the higher frequency means that its effects are estimated more precisely (see Table 4.1). In addition, the analyses were weighted in accordance with the reliability values associated for each animal for each selection trait analysed to take into account that the proof reliabilities are not uniform across the population i.e. greater emphasis was placed on those animals with more reliable proofs. As a comparison, the analyses were also run without weighting with reliability. The mixed procedure of SAS was used.

When a large number of comparisons are made, low probabilities may arise by chance alone. We carried out a meta-analysis of probabilities (Fisher, 1934) to determine whether the low probabilities were meaningful. The test statistic $-2 \Sigma \ln(p)$ is distributed as chi-square with twice as many degrees of freedom as probabilities tested.

TRAIT	MEAN ESTIMATE	STANDARD DEVIATION	STANDARD ERROR
Milk Yield (kg)	610.58	622.37	73.35
Milk Protein (kg)	21.46	16.33	1.92
Milk Fat (kg)	24.89	25.22	2.97
Daughter Fertility	98.76	4.88	0.57
Herd Life	101.24	5.54	0.65
Somatic Cell Score (SCS)	3.01	0.22	0.03
Lifetime Profit Index (LPI)	671.25	578.50	68.18

Table 4.3. Mean and standard error values for each trait EBV calculated using the univariate procedure on SAS.

4.4. Results of statistical analysis of trait associations with MHC class I haplotypes

All haplotypes were included in the analysis of class I haplotype associations with selection traits, with the common W12B haplotype set as the zero value. However, those haplotypes present in the Canadian bull population at low frequencies i.e. New5, New22 and New24, were repeatedly shown to have a marked affect upon selection traits but these estimates were not statistically significant. Therefore, the figures presented for these analyses omit these three haplotypes. It is also important to note that because these analyses use EBVs, the effect transmitted to the progeny will be half the value calculated because an EBV is twice the PTA (predicted transmitting ability) of an animal. Additionally the results presented in sections 4.4.1 to 4.4.4 were produced using the weighted gene substitution model according to proof reliability. The results presented in section 4.4.5 are those produced by the gene substitution model without weighting as a means of comparison.

The analysis calculated the means, standard deviations and standard errors for each of the traits analysed (see Table 4.3). The table presents large standard deviations for the traits e.g. for milk yield, the mean is 610.58 kg and the standard deviation is 622.37 kg, are reflective of the large range in EBVs recorded for these traits i.e. the range in EBVs for milk yield is 3733 kg. The data presented in Table 4.3 provides a representation of the EBV data used for the analysis of each trait i.e. the means for each of the traits and the spread of the data used, however the trait means are not used directly in the gene substitution analysis. When calculating the fixed effects under the gene substitution analysis i.e. calculation of the baseline provided by the common W12B haplotype, the intercept is calculated which provides a value for each trait for the W12B haplotype against which all other haplotypes are compared. The baseline is reset to zero for the common haplotype for every trait analysis for ease of comparison. Figure 4.1. shows the relationship between the overall population mean and the intercept calculated for the common W12B haplotype

which becomes the zero baseline, against which the trait data for all other haplotypes is compared.

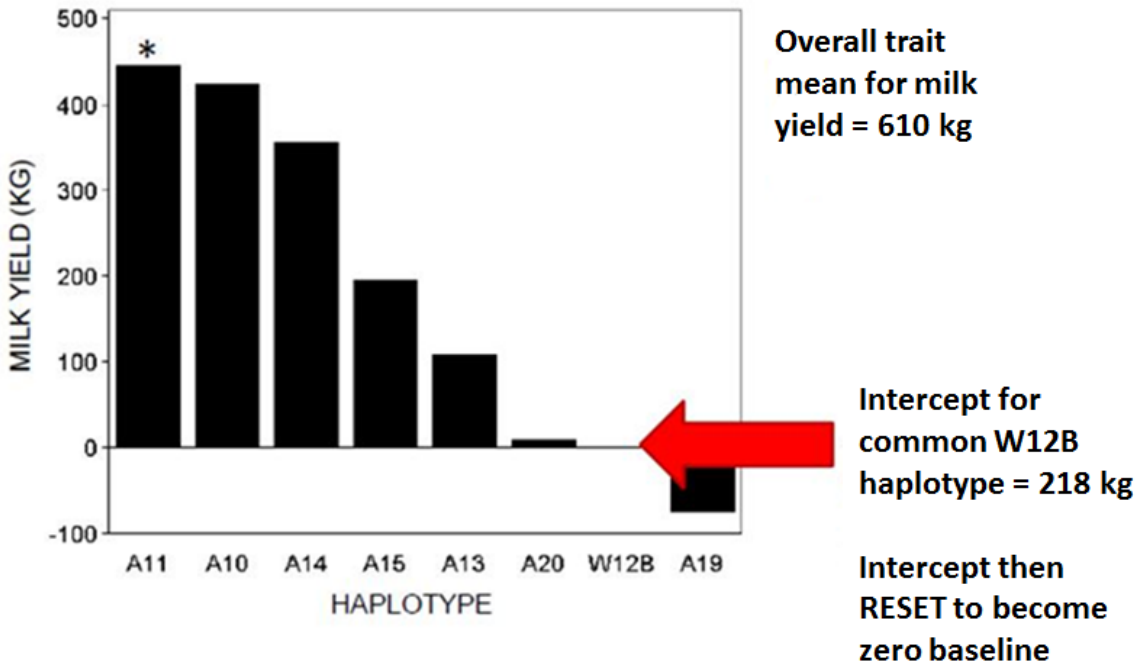


Figure 4.1. The relationship between the population mean and the intercept calculated for the common W12B haplotype, which is reset to become the zero baseline against which all other haplotype values are compared.

Seven traits were analysed, each involving the seven haplotypes present in five or more bulls to give a total of 49 analyses. One substitution (A19-W12B) was significant at the 1 % level and a further seven substitutions were significant at the 5 % level. Less than 3 probabilities below 5 % would be expected to occur by chance in 49 analyses. In addition, a meta-analysis was carried out on the 49 probabilities generated during the seven analyses, the sum of -2*ln(probability) was 146.01. In the absence of real associations, this sum is distributed as Chi-square with twice as many degrees of freedom as probability estimates (Fisher, 1970). For 98 degrees of freedom, this value is too high to be due to chance (p=0.001). This low probability indicates that the observed substitutions were not simply an artefact of multiple comparisons.

4.4.1. Milk Traits

The presence of the A11 haplotype was shown to be associated with improved milk yield in comparison with W12B, increasing daughter average milk yields by 445 kg per lactation (p = 0.0460). However, as the statistical analysis uses EBVs, this means that on average each daughter will produce an extra 220 kg of milk above the population average during each 305-day lactation. An increase in milk yield (approximately 210 kg) was also observed when replacing W12B with the A10 haplotype, however this effect was less statistically significant (p = 0.0528). Conversely, as the common W12B is set as the baseline, it means that the W12B is associated with poor yields in comparison to herd contemporaries with other MHC class I haplotypes, except for A19 (see Figure 4.2A).

A similar pattern of associations was observed between milk protein content (kg) EBV and MHC haplotype, as shown in Figure 4.2B. Presence of the A11 haplotype was significantly associated with an increase in milk protein content by 7 kg per lactation (p = 0.0166) compared to W12B, while the positive effect associated with A10 was less significant (p = 0.0718). Once again, the baseline provided by the W12B haplotype means that it is associated with lower milk protein yields.

Statistical analysis showed that each haplotype had an effect upon the milk fat yield in comparison to the W12B haplotype, as shown in Figure 4.2C. However, none of these associations were significant, which contradicts previous findings (Batra et al., 1989; Weigel et al., 1990).

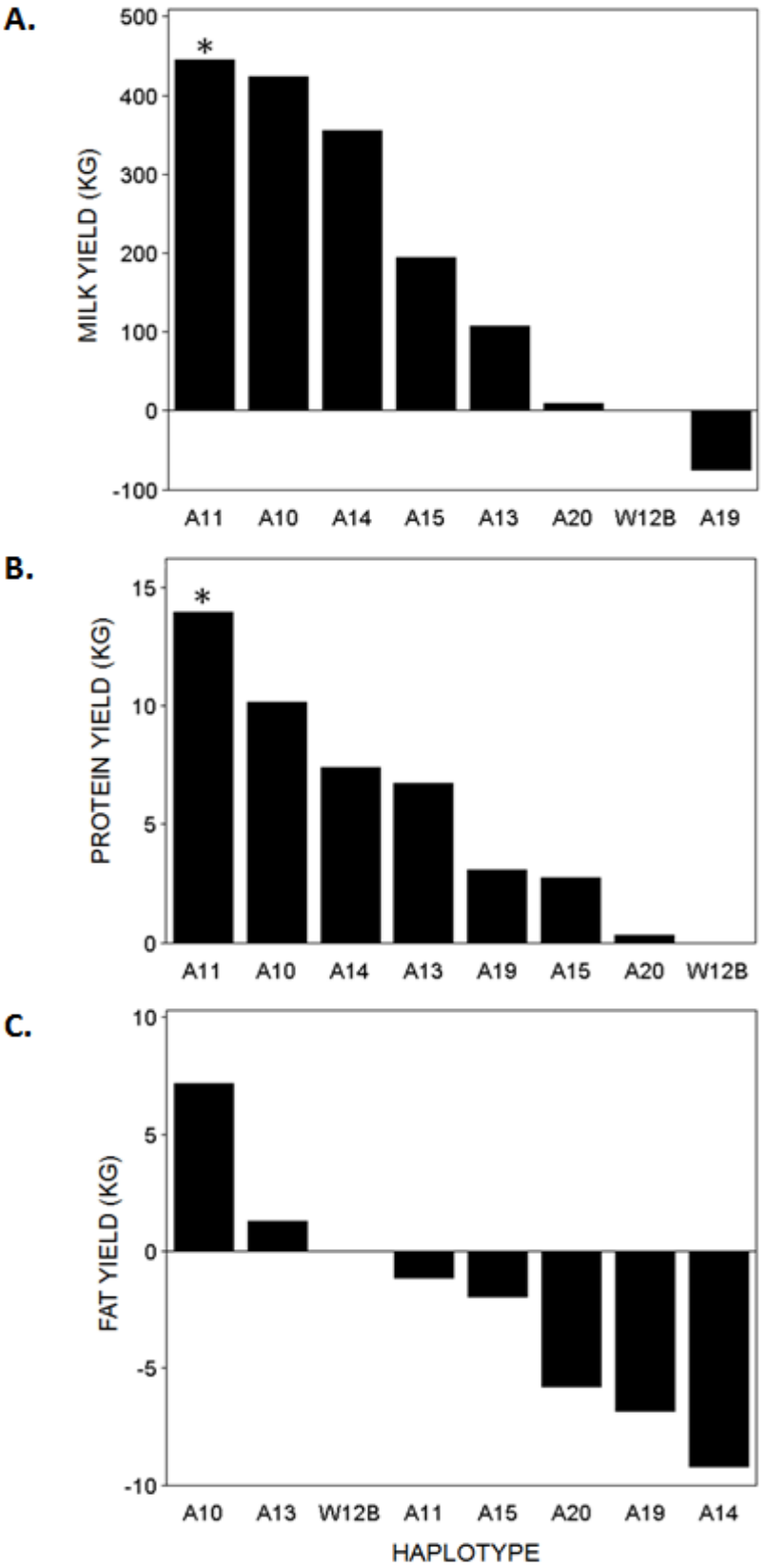


Figure 4.2. Effect of substituting one copy of W12B for another class I haplotype upon A. milk yield (kg per lactation), B. protein yield (kg per lactation), C. fat yield (kg per lactation). * Denotes significance at <0.05 level.

4.4.2. Functional & health traits

The gene substitution model showed that the replacement of W12B with one copy of A15 had a significantly detrimental effect on both herd life and fertility index with estimated effects of -3.88 and -3.17 ($p = <0.05$) respectively, as shown in Figure 4.3. The W12B haplotype was shown to be the most beneficial for fertility i.e. all other haplotypes were associated with lower estimates for fertility index.

Somatic cell scores (SCS) are used as an indicator of mastitis susceptibility or resistance, with resistant cows recording low SCS. SCS is the only trait examined for which a low number is advantageous. A19 was associated with a desirable decrease in SCS when substituted for W12B ($p<0.01$). The A10 and A14 haplotypes were also associated with a significant decrease in SCS ($p<0.05$). The common W12B haplotype was associated with the highest SCS and subsequently susceptibility to mastitis infection (see Figure 4.4.).

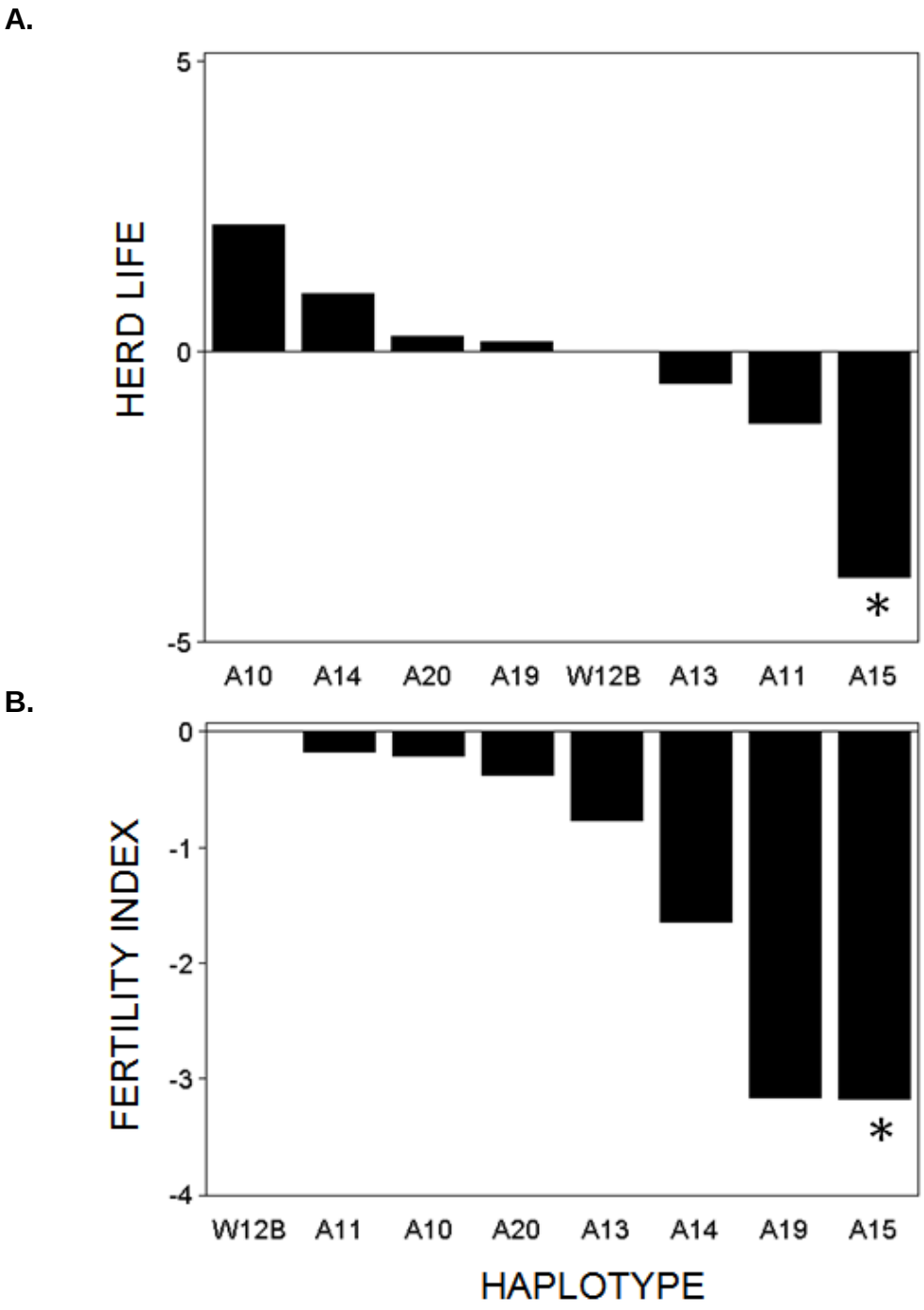


Figure 4.3. Effect of substituting one copy of W12B for another class I haplotype upon **A.** herd life index and **B.** fertility index. *Denotes significance at <0.05 level.

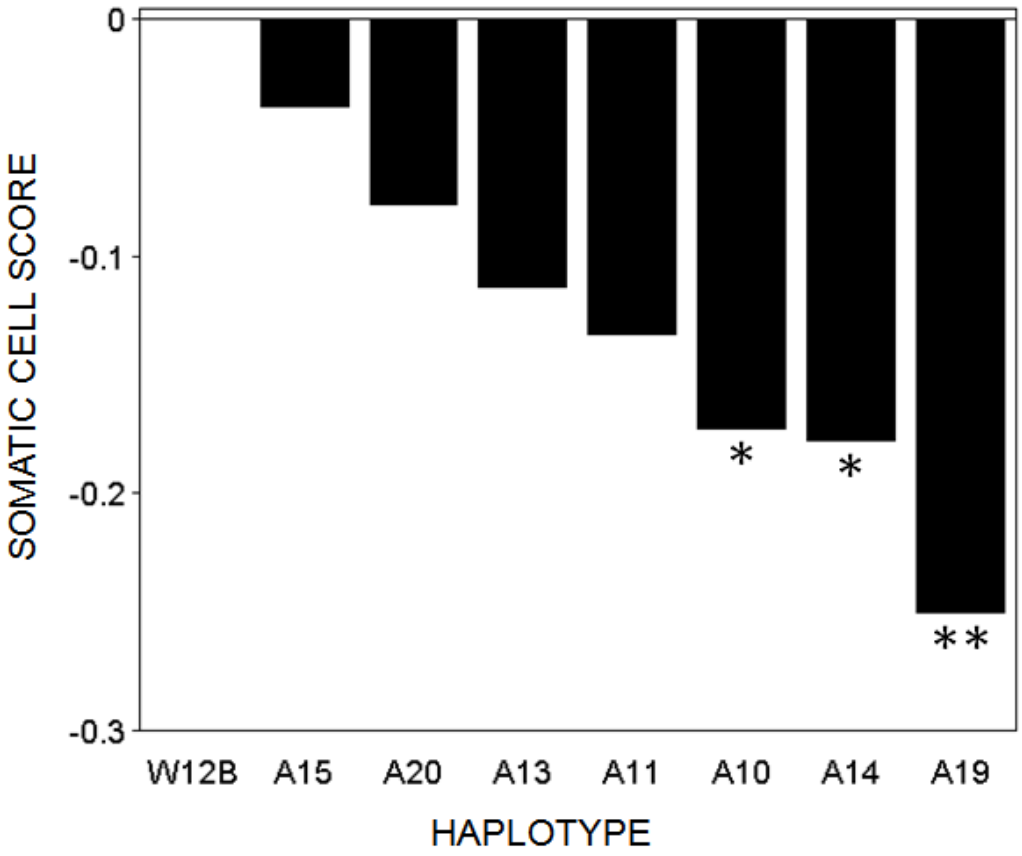


Figure 4.4. Effect of substituting one copy of W12B for another class I haplotype upon somatic cell score (SCS). *Denotes significance at <0.05 level. **Denotes significance at <0.01 level.

4.4.3. Conformation Traits

Of the five conformation traits examined with the gene substitution model, only rump angle was shown to have any statistically significant associations with MHC class I haplotype. Both the A11 and A15 haplotypes were shown to be associated with a decrease in rump score of -3.27 and -2.75 respectively, each with a p-value < 0.05 (see Figure 4.5.).

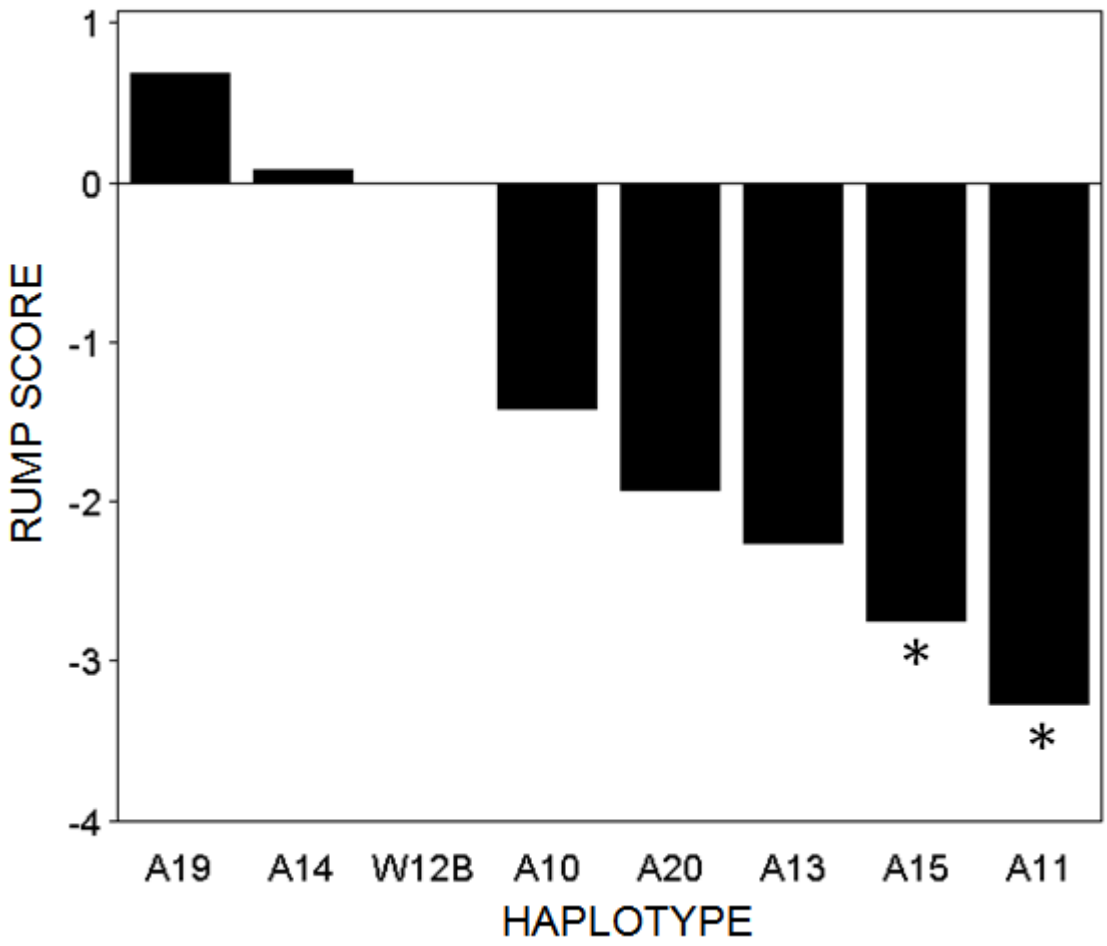


Figure 4.5. Effect of substituting one copy of W12B for another class I haplotype upon rump score. *Denotes significance at <0.05 level.

4.4.4. Economic Trait Selection: Lifetime profit index (LPI)

The gene substitution model data showed a significant association between the A10 haplotype and an increase in the Lifetime Profit Index (LPI; p= <0.05), see Figure 4.6. This reflects the cumulative associations of this haplotype with increased milk yield, milk protein yield and decreased SCS. This analysis also showed that the common W12B haplotype was not favourably associated with LPI.

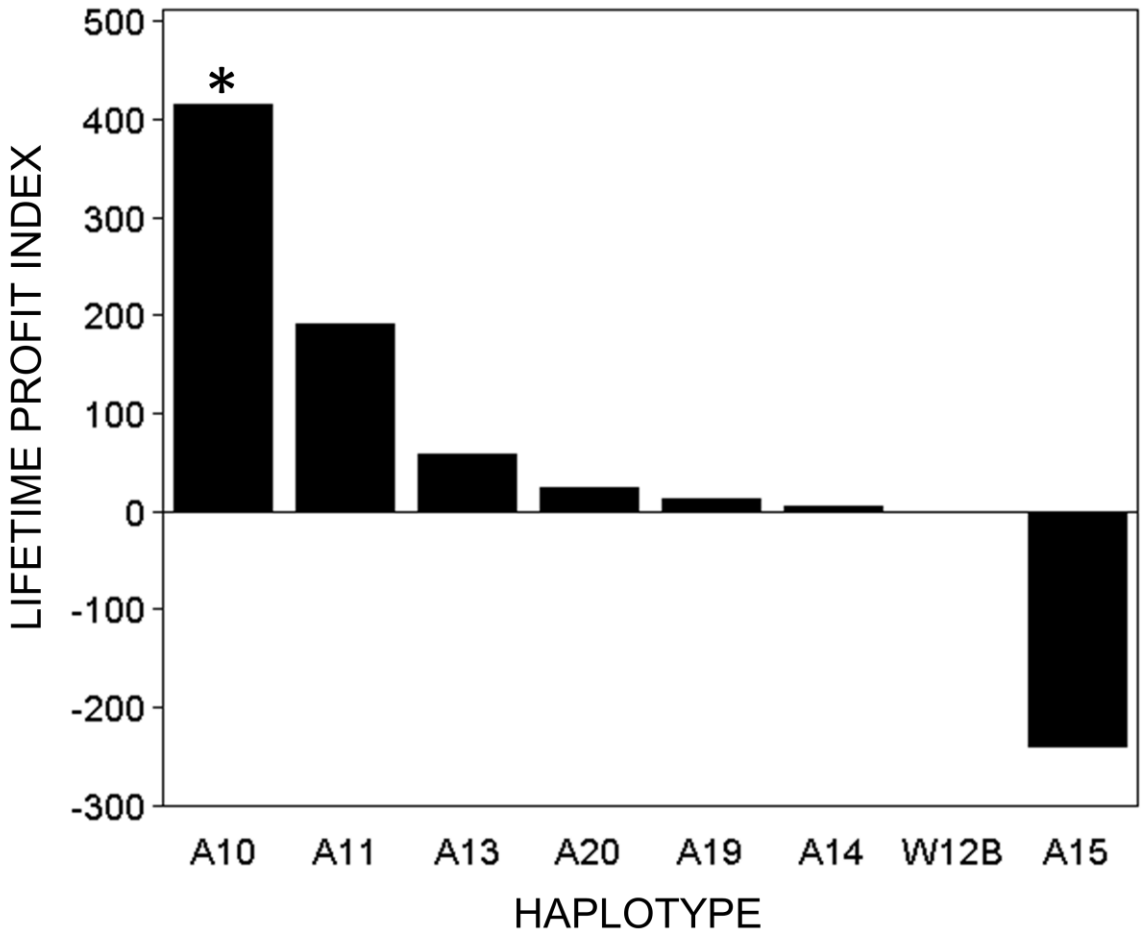


Figure 4.6. Effect of substituting one copy of W12B for another class I haplotype upon lifetime profit index (LPI). *Denotes significance at <0.05 level.

4.4.5. The effect of weighting upon trait associations

As outlined in the methods, the gene substitution model was run both with and without weighting according to the reliability of the bull proof estimates. Re-running the gene substitution analyses showed that removing the reliability weighting does have an impact upon the results; in general weighting increased the statistical significance of the associations. The analysis of conformation traits was excluded from the re-analysis.

Table 4.4 presents a comparison of the gene substitution results with and without the reliability weighting for production traits. The associations observed when the model was weighted are still evident without weighting i.e. A11 is still significantly associated with an increase in both milk yield and protein yield. In each case the estimated effect of substituting one copy of W12B with A11 is an increase of 445 kg but the p value is increased in the weighted model (p= 0.0460) compared to the unweighted model (p= 0.0447). In the weighted model the increase in milk yield associated with the A10 haplotype is just outside the limits of significance (p= 0.0528) but when the weighting is removed the p value increases to 0.0593 and the estimated effect is reduced. The A11 haplotype is also significantly associated with an increase in protein yield of approximately 14 kg in both the weighted and unweighted models, but the p value is increased with weighting i.e. 0.0166 compared to 0.0157. No significant associations between haplotype and milk fat yield were observed regardless of whether or not the analysis was weighted according to reliability.

The analysis of functional and health traits showed the same associations with and without weighting, see Table 4.5. The A15 haplotype was shown to be significantly associated with a decrease in both fertility index and herd life, however weighting the gene substitution model gives a greater negative effect upon both traits with a lower probability. The same three haplotypes, A10, A14 and A19, are shown to be significantly (p= <0.05) associated with a

desirable decrease in SCS both with and without weighting. Weighting the gene substitution model increases both the effect upon SCS and the level of significance. The most striking difference between the weighted and unweighted gene substitution models is observed when analysing the economic trait LPI, see table 4.6. With weighting, a significant association between the A10 haplotype and an increase in LPI of 416 points is observed ($p = 0.0386$). However, without weighting there is no significant association between LPI and class I haplotype.

		MILK YIELD				PROTEIN YIELD				FAT YIELD			
		WEIGHTED		UNWEIGHTED		WEIGHTED		UNWEIGHTED		WEIGHTED		UNWEIGHTED	
		EFFECT	P-VALUE	EFFECT	P-VALUE	EFFECT	P-VALUE	EFFECT	P-VALUE	EFFECT	P-VALUE	EFFECT	P-VALUE
HAPLOTYPE	A10	423.04	0.0528	411.53	0.0593	10.16	0.0718	9.72	0.0863	7.18	0.4326	6.18	0.4972
	A11	445.53	0.0460	445.51	0.0447	13.96	0.0166	14.07	0.0157	-1.15	0.9022	-1.60	0.8623
	A13	107.70	0.7017	99.13	0.7244	6.72	0.3577	6.60	0.3688	1.29	0.9137	0.63	0.9579
	A14	355.60	0.1030	369.59	0.0899	7.38	0.1896	7.55	0.1815	-9.23	0.3155	-9.28	0.3100
	A15	194.80	0.2709	200.63	0.2566	2.75	0.5477	2.94	0.5219	-1.96	0.7936	-1.99	0.7887
	A19	-75.42	0.7507	-51.13	0.8306	3.08	0.6168	3.70	0.5530	-6.83	0.4984	-6.91	0.4945
	A20	9.67	0.9616	10.36	0.9590	0.34	0.9482	0.31	0.9534	-5.77	0.4985	-5.99	0.4825

Table 4.4. Comparison of results generated using the gene substitution model with or without weighting according to proof reliability for production traits. Cells highlighted in pink denote associations significant at the <0.05 level.

		HERD LIFE				FERTILITY INDEX				SCS			
		WEIGHTED		UNWEIGHTED		WEIGHTED		UNWEIGHTED		WEIGHTED		UNWEIGHTED	
		EFFECT	P-VALUE	EFFECT	P-VALUE	EFFECT	P-VALUE	EFFECT	P-VALUE	EFFECT	P-VALUE	EFFECT	P-VALUE
HAPLOTYPE	A10	2.18	0.2320	1.71	0.3531	-0.21	0.9016	-0.39	0.8190	-0.17	0.0209	-0.17	0.0227
	A11	-1.24	0.5120	-1.25	0.5017	-0.17	0.9221	0.08	0.9648	-0.13	0.0802	-0.13	0.0734
	A13	-0.55	0.8177	-0.12	0.9610	-0.77	0.7297	-0.41	0.8533	-0.11	0.2397	-0.11	0.2568
	A14	1.01	0.5816	0.60	0.7453	-1.64	0.3390	-1.71	0.3210	-0.18	0.0178	-0.16	0.0309
	A15	-3.88	0.0102	-3.40	0.0264	-3.17	0.0257	-2.70	0.0576	-0.04	0.5394	-0.03	0.5655
	A19	0.17	0.9329	0.79	0.6980	-3.16	0.0877	-2.67	0.1638	-0.25	0.0027	-0.24	0.0035
	A20	0.26	0.8799	0.52	0.7610	-0.37	0.8138	-0.09	0.9559	-0.08	0.2534	-0.08	0.2635

Table 4.5. Comparison of results generated using the gene substitution model with or without weighting according to proof reliability for health and functional traits. Cells highlighted in pink denote associations significant at the <0.05 level.

		WEIGHTED		UNWEIGHTED	
		EFFECT	P-VALUE	EFFECT	P-VALUE
HAPLOTYPE	A10	416.06	0.0386	308.86	0.1282
	A11	191.53	0.3565	177.65	0.3857
	A13	58.60	0.8210	48.01	0.8551
	A14	4.23	0.9831	-55.52	0.7829
	A15	-239.10	0.1455	-219.84	0.1848
	A19	12.97	0.9516	35.81	0.8728
	A20	24.87	0.8927	20.53	0.9133

Table 4.6. Comparison of results generated using the gene substitution model with or without weighting according to proof reliability for LPI. Cells highlighted in pink denote associations significant at the <0.05 level.

4.5. Discussion

MHC class I typing results of this study revealed a total of 11 MHC class I haplotypes in a population of 72 Canadian Holstein AI bulls. Mathematical analysis of diversity showed that whilst there has been no significant loss of MHC class I diversity, there have been significant changes in class I gene frequencies over 20 years. Simulations of gene flow demonstrated that genetic drift and inbreeding alone were not responsible for the pattern of MHC diversity observed, implicating trait selection as an influential force (see Chapter 3; Codner et al., manuscript in preparation). Using the gene substitution model, we were able to dissect the influence of selection for specific agri-economic traits upon MHC class diversity in a widely used AI bull population and identify associations between selection traits and MHC haplotype.

Among production traits, both milk yield and milk protein yield were associated with the MHC class I haplotype A11 ($p = <0.05$). The A10 haplotype was also shown to be associated with increases in both these traits, but the p values of 0.0528 and 0.0718 respectively were just outside the limits of significance. Notably the A10 and A11 haplotypes have both increased in frequency since the 1988 sampling of Canadian Holstein bulls, suggesting selection for production traits has been influential in this change. Of note is that the common W12B haplotype is not favourably associated for either milk yield or milk protein content. No single haplotype was significantly associated with milk fat content, which apparently contradicts previous findings (Batra et al., 1989; Hines et al., 1986; Weigel et al., 1990). However, whilst Hines and colleagues found a significant association between the A11 haplotype and milk fat percentage, there was no such correlation with milk fat yield. It is possible that if our study had included a greater number of animals, the effects observed would have become significant.

The associations with milk production traits may reflect linkage disequilibrium between the MHC class I genes and other loci. For example, prolactin is a hormone that may influence milk related traits and the gene for prolactin is closely linked to the MHC on chromosome 23. Alternatively, MHC haplotypes could influence the incidence of subclinical disease which would affect production traits indirectly.

The A10, A14 and A19 haplotypes were associated with decreased somatic cell counts and consequently increased resistance to mastitis, which is a major disease of dairy cattle. This appears to contradict previous work which showed the A10 and A19 haplotypes were associated with increased somatic cell counts (Aarestrup et al., 1995). Associations between somatic cell score and MHC class I haplotype may be a direct effect of MHC class I genes, which could influence T cell-mediated or NK cell-mediated killing of pathogenic intracellular bacteria such as *Staphyloccus aureus* (Sordillo et al., 1997). Alternatively, the effect could be due to linked MHC class II genes; DRB3*1502 was significantly associated in an earlier study with lower SCS in Holsteins (Sharif et al., 1998). However, the effect of specific MHC class II alleles on mastitis appears to lack consistency (Dietz et al., 1997; Rupp et al., 2007).

The analysis of conformation traits was included because the physical make-up of the animal will have an effect across production, health and functional traits e.g. a lame cow has been reported to have a 360 kg decrease in milk yield over a 305-day lactation (Green et al., 2002), decreased fertility (Lee et al., 1989) and is likely to be removed from the herd earlier than her sound herd mates (Booth et al., 2004). The conformation proofs used in the gene substitution model were cumulative measures of type i.e. the score for rump incorporates type scores for rump angle, pin setting, pin height, pin width and loin strength. The use of the major scorecard conformation traits in the gene substitution model could mean that some associations with very specific aspects of conformation may be obscured by the other factors contributing to these composite scores. However, many conformation traits have low heritabilities in comparison to those recorded for production e.g. foot angle is estimated to have a heritability of 0.13,

whereas milk yield has a heritability of 0.35. The advantage of using composite conformation scores is that the predicted heritabilities range from 0.21 to 0.30 which are more reliable for the identification of any associations with MHC haplotype.

Of the five composite conformation traits analysed, a single significant association between MHC haplotype and rump score was observed. Both A11 and A15 were shown to be significantly associated with a decrease in rump score when substituted for a copy of W12B. Both of these haplotypes have increased in frequency over the last twenty years. The most interesting of these relationships is the decrease in rump score associated with the A15 haplotype. This haplotype has also been showed to be associated with undesirable decreases in both herd life and fertility. Those cattle with poor fertility will be removed from the herd earlier than fertile contemporaries. The independent relationships of fertility, herd life and rump score with A15 could be tightly linked as extremes in rump conformation traits i.e. extreme angularity or pin width settings, have been correlated with increased calving intervals (Wall et al., 2005) and reduced herd life (Berry et al., 2005; Sewalem et al., 2004). However, it is important to note that the A11 haplotype was also significantly associated with a decrease in rump score, but no association with fertility and/or herd life was observed. This suggests that the unfavourable decline in fertility and herd life associated with A15 may only be in part attributed to its association with a decrease in rump score. The significant associations of the A11 haplotype with both increased milk and protein yields may compensate for its negative associations with rump score and so maintaining these animals in the herd.

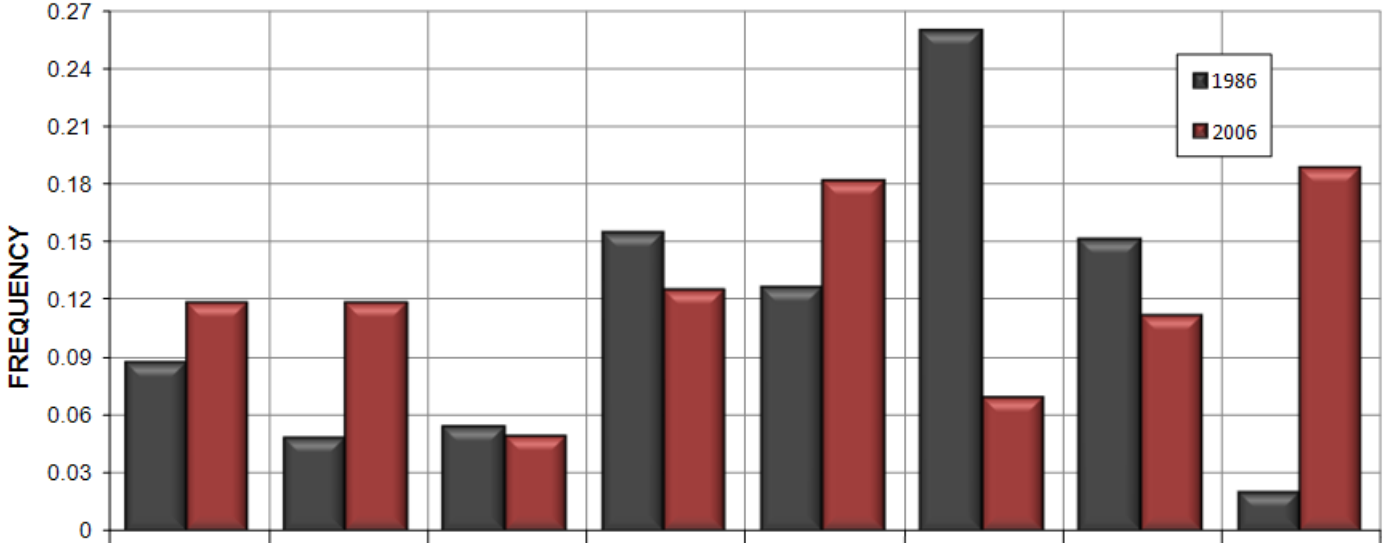
The A15 haplotype has increased in frequency over the last twenty years by a frequency of 0.055, despite the undesirable associations with rump angle, herd life and fertility. This could indicate that the unfavourable associations observed may have been a more recent development, becoming apparent as the haplotype increases in frequency. In human populations it has been reported that when parents are MHC-matched, there is a higher rate of spontaneous abortion in the early stages of pregnancy (Komlos et al., 1977; Schacter et al.,

1979). Further evidence in cattle has shown that embryo transfer between MHC compatible donors and recipients resulted in higher rate of early foetal loses compared to MHC incompatible pairings (Aguilar et al., 1997). As a haplotype increases in frequency, the likelihood of the dam being MHC-matched to the sire is increased and may contribute to the reduced fertility and herd life of this common haplotype, especially as it appears to have maintained a relatively high frequency over twenty years. However, the results of the gene substitution analysis showed that the common W12B haplotype was advantageous in terms of fertility index i.e. all other haplotypes were associated with lower fertility indexes. The rapid increase in frequency of this haplotype could be attributed to its favourable association with fertility. As with A15, it is plausible that the undesirable association with fertility is lagging behind the rapid increase in frequency of the W12B haplotype (0.168) and that this haplotype will become associated with a decline in fertility in the future.

The gene substitution models for production and functional traits were run both with and without weighting according to the bull proof reliability. With the removal of weighting, it is assumed that each bull proof is derived from the same number of progeny. Those bulls with a large number of progeny will have a more reliable prediction of genetic merit for each selection trait. In the majority of cases, the removal of weighting resulted in an decrease in the estimated effect and reduced significance e.g. the negative effect upon herd life associated with the A15 haplotype decreases from -3.88 ($p = 0.0102$) to -3.40 ($p = 0.0264$) upon removing the weighting. However, the association of the A11 haplotype with an increase in LPI is no longer significant when the weighting according to reliability is removed. A plausible explanation for the loss of this association with the removal of weighting is due to the fact that this is a cumulative index incorporating data from eleven selection traits across production, longevity and udder health. As observed with the majority of other traits, the removal of reliability weighting results in less significant associations and when applied across the spectrum of traits incorporated into the calculation of LPI then a loss of significance is to be expected.

The gene substitution analysis revealed that there were associations between selection traits and MHC haplotype, indicating that selective breeding is shaping MHC diversity in this cattle population. However, there were antagonistic effects among the traits. W12B, which has shown a large increase in frequency over the last 20 years to become the most common haplotype, is associated with relatively poor milk yield, milk protein, SCS and LPI but improved fertility. As with changes in haplotype frequency, the associations derived from the 2006 data differ from those observed in similar studies conducted in the 1980s, as summarised in Figure 4.7. Whilst a few associations seem to have been maintained over time e.g. the association of A14 with a favourable decrease in SCS, some associations appear to have been lost (A19 associated with increase in protein yield) or have developed more recently (association of A11 with increased milk yields) reflecting the fact that the associations observed are a static snap shot of a dynamic system.

Three hypotheses exist to explain the selective forces that determine MHC class I haplotype frequencies in dairy cattle: directional selection, inbreeding and drift or balancing selection. As selection was on-going in 1986, directional selection would increase the frequency of the most common haplotype and lead to a decline in diversity. In addition, the most common haplotype would be positively associated with production and fitness traits. The most common haplotype in 1986 is now rare, whilst the formerly rare W12B haplotype is now the most common. Furthermore, neither the A19 nor W12B haplotypes were associated with improved production or fitness i.e. the common W12B haplotype in 2006 was associated with decreased milk yield and decreased milk protein. These observations appear to rule out directional selection acting upon bovine MHC class I haplotypes. The effects of inbreeding and drift were simulated in a mathematical model and the results from this model were incompatible with the observed gene frequencies. These lines of evidence discount directional selection or inbreeding and drift as the responsible forces shaping cattle MHC class I haplotype diversity.



SELECTION TRAIT	A10	A11	A13	A14	A15	A19	A20	W12B
MILK YIELD		Increase *		Increase *				
PROTEIN YIELD		Increase *				Increase *		
		Increase **						
FAT YIELD		Decrease *						
		Increase **						
HERD LIFE					Decrease *			
FERTILITY INDEX					Decrease *			
SCS	Decrease *			Decrease *		Decrease **		
				Decrease **				
LPI	Increase *							

Figure 4.7. Summary of haplotype frequency changes between the two samples, together with associations found in this study and previous studies by Batra et al. (1989) highlighted in blue, and Weigel et al. (1990) highlighted in pink.

The observed changes in haplotype frequencies, the maintenance of diversity and the associations with the estimated breeding values were compatible with balancing selection. The deficiency of homozygous bulls (7 % as opposed to the 14 % expected in a population in Hardy-Weinberg equilibrium) also argues for heterozygote advantage (see Chapter 3). Both heterozygote advantage and frequency-dependent selection would increase the frequency of rare haplotypes and decrease the frequency of common haplotypes. Furthermore, heterozygote advantage means that the selective pressure will vary with the gene frequency. Rare haplotypes will be found predominantly in heterozygotes, but as these increases in frequency, they will occur in a greater number of homozygotes (Hardy, 1908). If homozygotes are less fit than heterozygotes then the average effect of a gene will decrease as the gene increases in frequency causing gene frequencies to change over time. Common alleles or haplotypes will become less common while rare haplotypes will increase in frequency, resulting in the long-term maintenance of diversity.

In terms of trait selection, the estimated breeding value is an estimate of the additive genetic effect, which is defined as the sum of the average effects of genes (Falconer & Mackay, 1996). Non-additive effects such as dominance and epistasis do not contribute to the additive genetic effect although they do contribute to the overall genotypic effect. The situation with heterozygote advantage is less clear. Conceptually heterozygote advantage is an interaction between different alleles at the same locus and is not part of the additive effect. In practice, a rare allele will be more likely to be found in heterozygous than homozygous individuals than a more common allele (Hardy, 1908), and if heterozygotes are fitter than homozygotes then the average effect of a rare allele will be increased by heterozygote advantage. Relatives are more likely to share the rare allele than unrelated members of the population and the breeding value will be boosted by heterozygote advantage.

The detailed pedigrees of dairy cattle and extensive progeny performance records provide a model system for identifying selective forces that act upon MHC diversity in an economically important livestock species. The results of

mathematical modelling and the gene substitution model provided evidence to implicate balancing selection, linked to production and functional traits, as an integral force in shaping MHC diversity in cattle. Balancing selection provides an explanation as to why those haplotypes present at high frequency, such as W12B, are not now particularly advantageous in terms of production traits. As the magnitude of effects varies with frequency, balancing selection accounts for the apparent temporal inconsistencies in the associations. Of all the balancing selection mechanisms known, heterozygote advantage appears to be the most likely form of selection acting on this population. Rare alleles, which are more likely to be found in heterozygous animals, appear to be favourable in terms of trait selection. Using this information, the process of selective breeding could be made more efficient by incorporating appropriate weighting of breeding values to favour rare alleles, which would both maintain MHC diversity in the population whilst mimicking the processes already on-going within the Holstein dairy bull population.

Design and development of a SNP-based assay for the detection of common MHC class I haplotypes

5.1. Introduction

Reference strand-mediated conformational analysis (RSCA) is currently the most robust technique available for MHC class I-typing cattle, using genomic DNA (gDNA) samples (Birch et al., 2006). However, with four independent gene-specific PCRs included, it is a time consuming technique further limited by the number of samples that can be run on each RSCA gel. The associated data analysis is also protracted and can be difficult to interpret due to shared mobility values between some alleles. Whilst studies investigating quantitative trait loci (QTLs), parentage and/or diversity in cattle populations have utilised a wide range of genetic markers including Short Tandem Repeats (STRs; van der Goor et al., 2009), microsatellites, and Restriction Fragment Length Polymorphisms (RFLPs; Rocha et al., 1992), the associated analysis methods are also hampered by being labour intensive and time consuming. Single Nucleotide Polymorphisms (SNPs), the simplest form of variation, are rapidly becoming the genetic markers of choice.

SNP-based techniques can be both easily automated and standardised across laboratories whilst having low genotyping error rates. The release of the bovine genome and the desire to identify quantitative trait loci (QTLs) for production traits has given rise to a rapid expansion in whole-genome genotyping (WGG) tools e.g. the BovineSNP Genotyping BeadChip (Matukumalli et al., 2009) together with assays that can be customised to specific markers e.g. the Illumina Goldengate assay (Fan et al., 2006). SNP markers are also being utilised in the development and commercialisation of genetic tests to enhance economic and production traits in cattle herds. Currently available genetic tests using SNP markers in the dairy industry focus on traits such as production, longevity,

fertility and recessive genetic diseases e.g. bovine leukocyte adhesion deficiency (BLAD).

Despite the availability of a wide range of commercial assays and chips, the 'en bloc' evolution of the MHC does not lend itself to the application of such methods. The generation of diversity by inter- and intra-locus recombination, gene duplication and deletion and a high number of pseudogenes in the bovine MHC results in high levels of conservation and sequence similarity. As a result, the number of unique SNPs is low and the design of SNP probes can be problematic. Any SNP markers used within the MHC would require a preliminary locus-specific step to focus on the area of interest. The identification of appropriate SNP markers in the cattle class I region to facilitate the detection of common MHC haplotypes will benefit MHC-typing work and form the basis of a diversity measure to be incorporated into a commercially available test for cattle breeders. The ability to rapidly type for the most common class I haplotypes would enable breeders to formulate breeding programmes that promote genetic diversity within their herds at MHC loci which it is predicted would benefit many aspects of herd health.

5.2. Feasibility and development criteria of a SNP-based assay

A commercial SNP test must fulfill the following criteria:

1. Identify the most frequently occurring MHC haplotypes in Holstein-Friesian cattle populations.
2. High-throughput
3. Cost- effective
4. Rapid
5. Suitable for inclusion within existing SNP test panels using the primer extension method e.g. the Igenity® dairy profile
6. Ease of sampling. Reduced need for specialist handling/refrigeration.

Considering all of these criteria, together with the time constraints of the project, it was decided that the SNP test should utilise existing technologies, apparatus and software at the Institute for Animal Health, Compton. The Institute supports two possible methods for SNP detection based upon the primer extension method. The Beckman Coulter GenomeLab™ SNP-Primer Extension kit is available for use with the Beckman Coulter CEQ-8000 enabling SNP typing with fluorescently labeled dideoxynucleotides. The matrix assisted laser desorption ionization- time of flight (MALDI-TOF) mass spectrometer allows the discrimination of bases corresponding to the SNP of interest on the basis of molecular weight. Another research group at the IAH have shown the successful application of the Beckman CEQ-8000 and associated kit for genome-wide typing of in-bred chicken lines. Availability of expertise and limited access to the mass spectrometer meant that customisation of the Beckman Coulter kit became the preferred method.

5.3. Customisation of the Beckman Coulter SNP extension kit for the detection of common MHC haplotypes

The Beckman Coulter SNP assay is based upon the single base extension method in which an interrogation primer, whose 3' end binds to the base preceding the SNP, is extended by the addition of a single fluorescently labelled dideoxynucleotide (ddNTP) complementary to the base at the SNP position. The templates for the SNP-extension reactions are purified PCR products amplified from the test gDNA sample. The SNP-extension reaction products are cleaned up enzymatically prior to analysis on the Beckman Coulter CEQ-8000 sequencer with an appropriate size standard marker. A trace is produced with the size of the product corresponding to that of the extended interrogation primer and the colour of the incorporated fluorescently labelled ddNTP enables allele identification.

The design and optimisation of SNP-specific interrogation primers is integral to the customisation of the Beckman Coulter assay. An important consideration before designing interrogation primers is the template produced by PCR amplification from the test gDNA sample. The template for the SNP-extension reaction may be amplified using either generic primers or gene-specific primers (Birch et al., 2006; Pichowski et al., 1996; Chapter 2 Materials and Methods). Generic primers have the advantage of amplifying all the class I genes in a single PCR, but equally these primers also amplify non-classical genes and potentially pseudogenes. Although using gene-specific primers requires a greater number of PCR reactions, amplification with gene 1 and 2-specific primers would be sufficient for the detection of one or more alleles found on the majority of class I haplotypes. Furthermore, gene-specific primers would not amplify non-classical genes (or hopefully pseudogenes). Therefore two separate gene-specific PCRs were carried out to generate the templates for the SNP-extension reactions. A diagrammatic overview of the assay is shown in Figure 5.1.

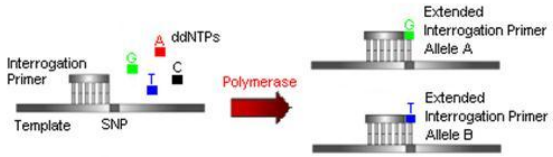
Bovine MHC class I gene-specific PCR

MHC class I gene-specific amplification from gDNA sample to focus on area of interest. PCR products span exons 2 and 3 in their entirety.

Clean-up: Purify PCR product template by incubating with ExoSAP-IT

SNP Extension reaction

Purified double-stranded PCR product added to thermal cycling reaction with SNP Interrogation primer, ddNTPs and Polymerase.



Clean-up : Purify SNP-extension reaction products with SAP

Production of traces

Purified SNP-extension product mixed with size standard and SLS and loaded onto Beckman CEQ8000 Sequencer. Traces produced are coloured according to the nucleotide located at the SNP position. The size standard is shown on the trace as red. All traces are analysed and haplotype identities assigned according to profile of SNPs in genes 1 and 2, i.e. common or uncommon.

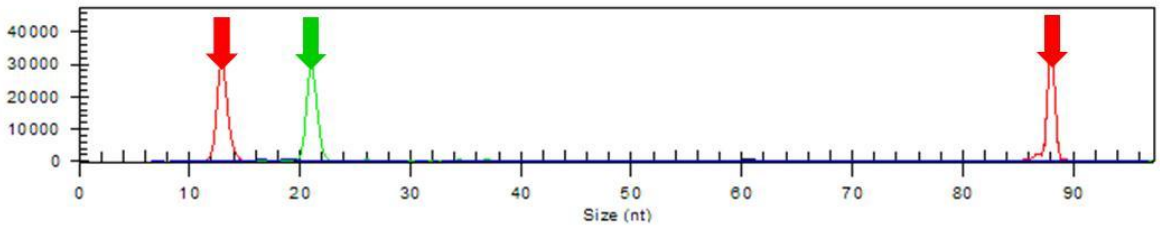


Figure 5.1. Outline of the Beckman Coulter SNP-extension assay.

5.4. Design and development of the SNP-based assay

5.4.1. Common MHC class I haplotypes

The introduction of artificial insemination (AI) techniques in dairy cattle herds has led to a heavy reliance upon Canadian and North American bulls in UK selective breeding programmes. Preliminary data generated from a cohort of 72 premium Canadian Holstein AI bulls (Chapter 3) highlighted the dominance of three common MHC class I haplotypes; A14, A15 and W12B. Due to a heavy reliance on Canadian genetics for UK dairy breeding programmes, it is believed that these three haplotypes are also prevalent in the wider UK herd. Although the W12B haplotype was the most common MHC haplotype in the Canadian Holstein bull population, the A14 and A15 haplotypes were the main area of focus for the development of the SNP test due to two unusual features of these haplotypes. Firstly, the A14 and A15 haplotypes share both a gene 2 (*N*02501*) and a gene 4 (*N*02401*) allele, and secondly these are the only two haplotypes defined in *Bos taurus* cattle that express a gene 4 allele. Mapping studies have shown the relative locations of the class I genes in an A14 BAC library, and as such the positions of the class I genes expressed on the A15 haplotype are inferred (Di Palma et al., 2001; see Figure 5.2).

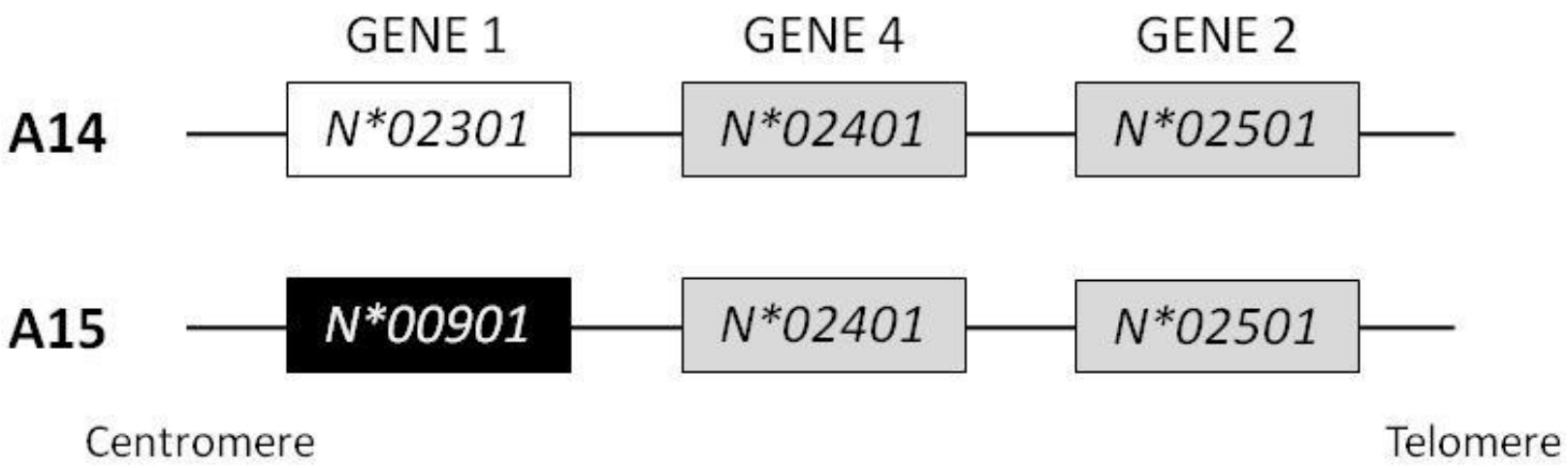


Figure 5.2. Diagrammatic representation of the class I genes expressed on the A14 and A15 haplotypes.

5.4.2. Choice of markers and designing appropriate interrogation primers

BoLA class I nucleotide alignments from the Immuno Polymorphism Database enable identification of SNPs unique to alleles of the A14 and A15 haplotypes: *N*02401* has seven unique SNPs, *N*02501* is unique at four sites and *N*00901* has a single SNP. Originally it was hoped that the A14 and A15 haplotypes could be identified using SNPs unique to each of the distinguishing gene I alleles and confirmed with a SNP unique to one of the shared alleles. The absence of a SNP unique to *N*02301* of the A14 haplotype meant this would not be possible. Furthermore, the lack of gene 4-specific primers meant that the seven SNPs present in the *N*02401* allele could not be used for identification. Instead, a SNP to distinguish *N*00901* and *N*02301* from all other class I alleles, used in conjunction with a SNP to differentiate between *N*02301* and *N*00901* was required. A SNP in the *N*02501* allele was then used for confirmation.

The process by which MHC genes evolve i.e. point mutation followed by recombination, has given rise to large amounts of allelic diversity and shared motifs. A lack of homology in the regions flanking the identified SNPs reduced the number that would be suitable for inclusion in the assay and hindered primer design. The presence of numerous pseudogenes and gene fragments within the bovine class I region, as in other species, and the lack of sequence data available for these pseudogenes may hinder this type of analysis, as the selected SNPs may not be unique as previously believed. However it is important to note that *BoLA* class I polymorphism is most extensively studied in Holstein-Friesian cattle; of the 80 class I alleles submitted to IPD over 62 % were derived from this breed and at least a further 6 % are known to be present in Holstein-Friesian cattle. As a result, identification of SNP markers in this breed is reliable as they are based upon the greatest amount of data available.

In order to facilitate multi-plexing in the future, i.e. analysing a number of different SNPs in a single reaction, it was decided that, where possible, primers

should differ in length by at least five nucleotides. If required, interrogation primers could be increased in length by the addition of a suitable homopolymeric 'tail' e.g. the addition of a series of adenine residues to the 5' end. Primers were named according to the codon in which the SNP was identified.

5.4.3. The A14 and A15 SNP assay outlined

The first assay used an initial gene 1-specific PCR step. An interrogation primer was designed to amplify a SNP at codon 55 which enabled the distinction of alleles *N*02301* and *N*00901* from all other known class I alleles. Multiplexing this assay with interrogation of a further SNP at codon 129 facilitated the distinction of *N*02301* and *N*00901*, thus enabling the identification of haplotypes A14 and A15 respectively. The two interrogation primers differed in length by 5 base pairs to allow product assignment.

The *N*02501* allele shared by the A14 and A15 haplotypes is identified by an interrogation primer for a SNP at codon 122 applied to the gene 2-specific PCR product in a second assay. This assay is used to both confirm the A14/A15 haplotypes identified in the first test, and to establish whether the sample animal is heterozygous or homozygous for the A14 and A15 haplotypes i.e. A14/A14, A15/A15, A14/A15, A14/other, A15/other and other/other. A summary of the customised A14/A15 bovine SNP test is shown in Figure 5.3.

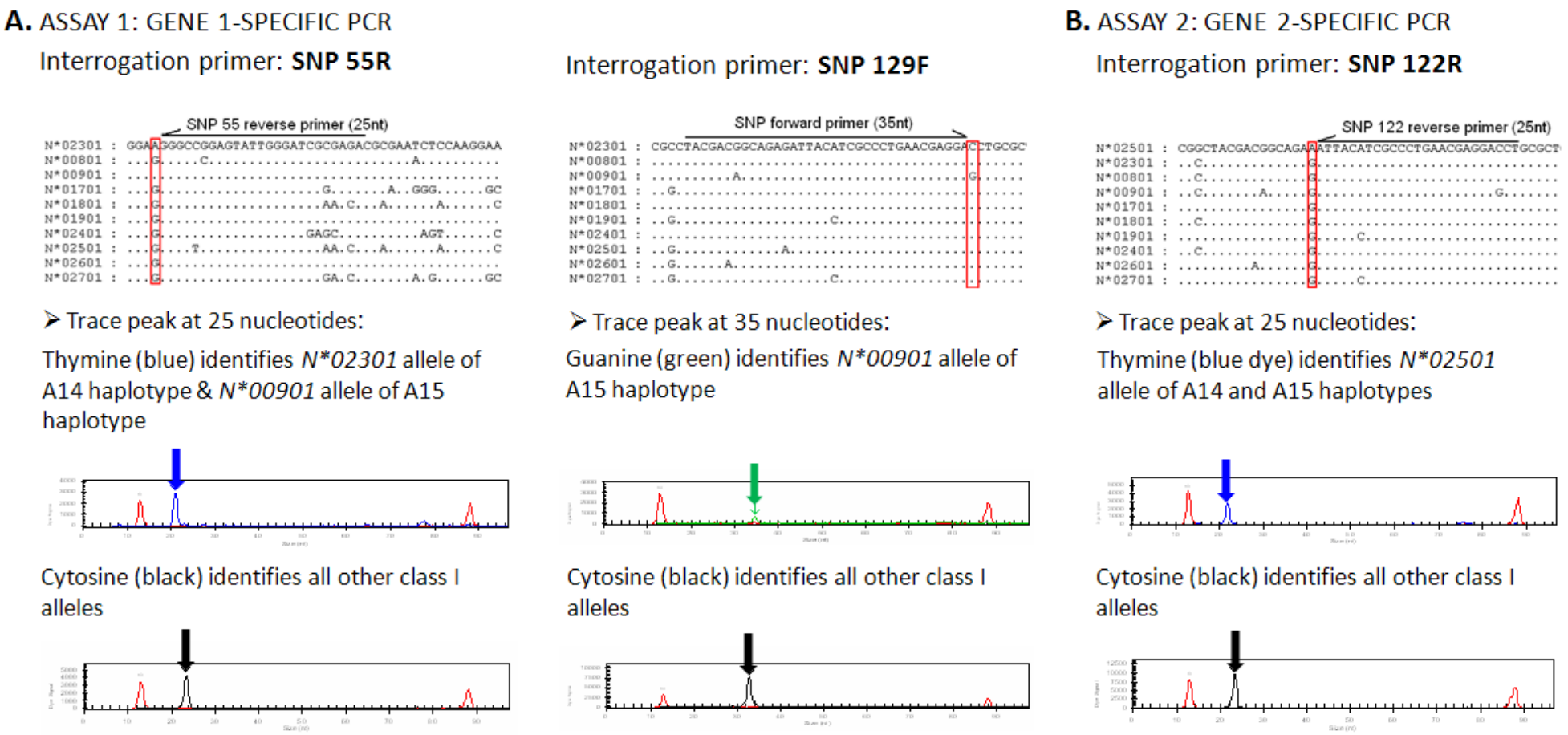


Figure 5.3. Outline of the A14/A15 SNP assay primers and expected results for both the **A.** Gene 1-specific assay and the **B.** Gene 2-specific assay. On sequence alignments, the dots indicate identity and dashes indicate gaps/lack of sequence data compared to the reference.

5.5. Optimising the Beckman Coulter SNP extension assay for detection of common MHC haplotypes A14 and A15

A panel of test animals of known haplotype was selected for testing the custom-designed interrogation primers (Table 5.1). Amplification of genes 1 and 2 from the test panel animal gDNA was performed and the products visualised on a 1 % agarose gel. A sample from which a gene 1 allele was absent/could not be amplified were included in the assays to confirm the effectiveness of the test i.e. control samples. Samples lacking a gene 2 allele are rare in the population and A10 is the only haplotype from which a gene 2 allele cannot be amplified using the gene 2-specific primers.

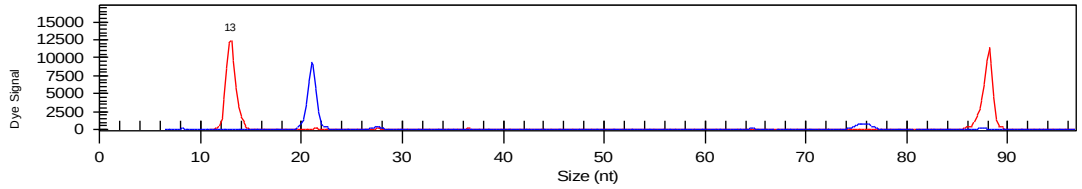
SAMPLE NO.	HAPLOTYPE	GENE 1	GENE 2
4903	A19/A19	×	✓
4929	A14/A19	✓	✓
4645	A15/A15	✓	✓
4954	A14/A15	✓	✓
4968	A13/A15	✓	✓

Table 5.1. Panel of samples from animals of known haplotype for preliminary optimisation and testing of the A14/A15 SNP assay.

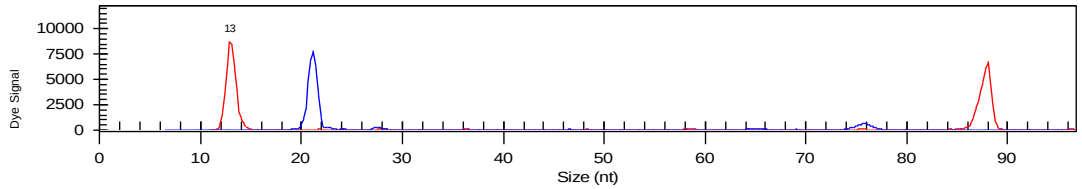
Initial optimisations were concerned with achieving 'readable' sample peaks. Preliminary runs following the protocol used by another Institute laboratory, in which 0.5 µl of sample was included, failed to yield sample peaks on the trace due to insufficient DNA. In order to establish an optimal sample volume for loading, a series of increasing amounts of sample DNA were included in the assay. Initial tests revealed a minimum volume of 2 µl was required to produce

readable peaks, and further testing identified an optimal sample volume of 3 µl. The effect of increasing sample volume upon signal strength is illustrated in Figure 5.4.

A.



B.



C.

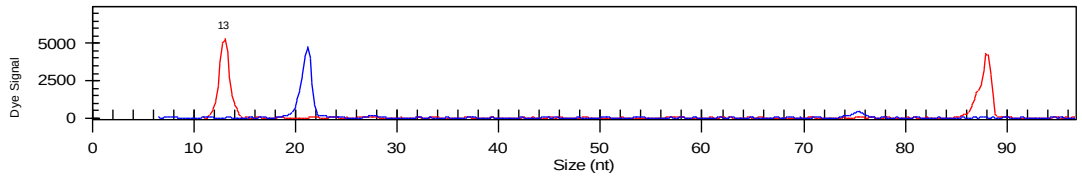
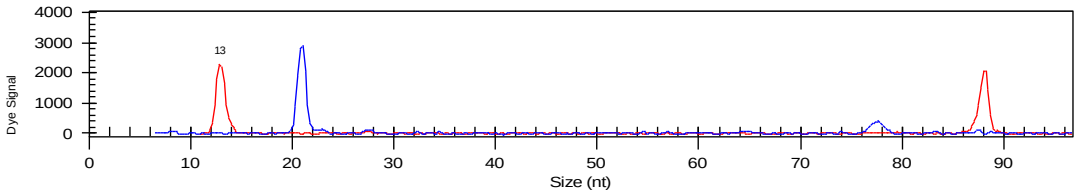


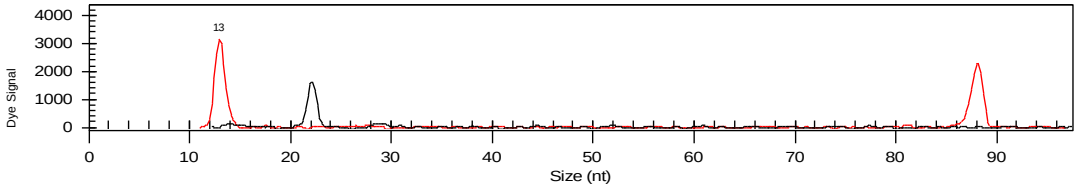
Figure 5.4. Traces from a series of SNP assays with interrogation primer SNP55R obtained using sample volumes of **A.** 3 µl, **B.** 4 µl, and **C.** 5 µl. As sample volume increases above the optimal volume of 3 µl, the signal strength of the sample peak decreases.

Identification of heterozygotes initially proved problematic during the optimisation process. In instances where the sample was known to have two alleles, the assay appeared unable to detect both alleles i.e. it was expected that the trace would have two peaks but consistently only one of the two expected peaks would be shown on the trace (Figure 5.5.). In the case of heterozygotes, the test repeatedly detected the same alleles preferentially. This could be a result of preferential amplification of specific alleles during initial gene-specific amplification, or preferential binding and amplification of specific sequences by the interrogation primer. Raw data traces for the heterozygote samples showed a weak signal which was insufficient for detection on the final output. To ensure that the test was able to detect the 'invisible' alleles a further control panel of typed animals was selected which had 'invisible' gene 1 and gene 2 alleles. A further three animals with previously detectable haplotypes were also included to verify the test was working. The results of this testing panel confirmed that the test was proficient in identifying the previously hidden alleles (Figure 5.5.). The SNP extension reaction was further optimised by altering the volume of DNA sample added to the reaction mixture, and experimenting with the thermal cycling conditions i.e. annealing temperature and number of cycles, to ensure the detection of both alleles in heterozygous animals in both the raw data and final traces.

A.



B.



C.

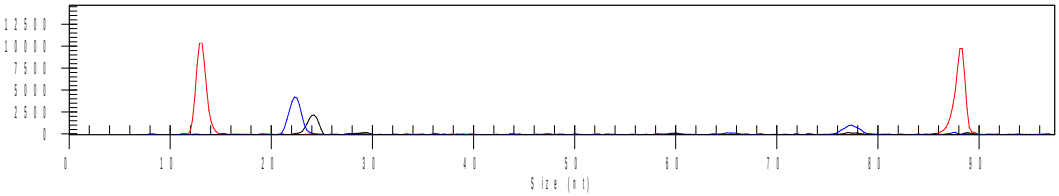


Figure 5.5. A comparison of SNP traces illustrating that in heterozygous animals with an A14 or A15 haplotype, the second haplotype is obscured and no peak is shown on the trace. Amplification of SNP 55 in an A15/A13 animal is shown in trace **A.** in which only a blue peak representing the thymine of the *N*02301* allele can be seen. The *N*03101* allele is obscured and no peak is produced. **B.** shows the amplification of the same SNP in an A13/W12B animal in which the cytosine of the *N*03101* allele is represented by the black peak. In this instance the peak can only be a result of the allele on the A13 haplotype because the gene 1 allele (*N*01901*) of the W12B haplotype does not amplify with gene 1-specific primers. **C.** shows the optimised assay detecting both alleles in an A15/A13 animal.

5.6. Blind trials of the optimised Beckman Coulter SNP extension assay for A14 and A15

Following optimisation of the A14/A15 assay, a blind trial of 40 samples was commenced. The first 20 samples of the blind trial were typed with an accuracy of 85 %, with the majority of incorrect calls being the result of undetectable haplotypes e.g. the A10 haplotype is undetectable using this assay because the gene 2 allele of the A10 haplotype (*N*01201*) does not amplify with the gene-specific primers (see Table 5.2). Having typed the first 20 samples of the blind trial, the kit reagents required re-ordering. However, by the time the second kit was ordered, the company had changed the kit components. Initial test runs failed to produce any peaks and it became clear that the SNP assay would have to be re-optimised in order to account for the new kit chemistry.

SAMPLE	GROUP 1			GROUP 2		HAPLOTYPE PREDICTION	ACTUAL HAPLOTYPE	RESULT
	PCR	SNP 55	SNP 129	PCR	SNP 122			
1	+	C	C	+	C	non A14/A15 with group1 & 2	A31/cA42c	✓ ✓
2	+	C	C	-	-	non A14/A15 with group 1	A13/New23	✓ ✓
3	+	T	C	+	T	A14	A14/New5	✓ ✗
4	+	T	C G	+	T	A14/A15 but A15 signal weak	A14/A15	✓ ✓
5	+	T	No peak	+	T C	A14 or A15 het	A14/A19	✗ ✗
6	+	T	C	+	T	A14	A10/A14	✗ ✓
7	+	T	C	+	T C	A14 het	A11/A14	✓ ✓
8	+	T	G C	+	T	A14/A15	A13/A15	✗ ✓
9	-	-	-	+	C	non A14/A15 with group 2 only	A11/W12B	✓ ✓
10	+	T	G	+	T	A15	A15/New5	✓ ✗
11	+	C	C	+	C	non A14/A15 with group 1 & 2	A31/W12B	✓ ✓
12	-	-	-	+	C	non A14/A15 with group 2 only	A19/A20	✓ ✓
13	+	T	C	+	T C	A14 het	A14/A20	✓ ✓
14	+	T C	C	+	T	A14 Het	A14/A31	✓ ✓
15	-	-	-	+	C	non A14/A15 with group 2 only	A17/A20	✓ ✓
16	+	T	G	+	T C	A15 het with group 2	A15/A17	✓ ✓
17	+	T	C	+	T	A14	A14/A14	✓ ✓
18	+	T	G	+	T	A15	A15/A15	✓ ✓
19	+	T	G	+	T C	A15 het with group 2	A15/A19	✓ ✓
20	+	T	G C	+	T	A14/A15	A14/A15	✓ ✓

Table 5.2. Results of first 20 samples in a blind trial of the A14/A15 SNP test.

5.7. Optimising the A14/A15 SNP assay using the Beckman Coulter SNPStart kit

Preliminary assays using the new SNPStart assay kit failed to produce readable sample peaks on the analysis traces. It was first assumed that an insufficient amount of SNP product had been added to the plate, and thus there was no sample peak. The next step was to run a series of sample volumes ranging from 0.5 µl to 6 µl on the CEQ-8000 to try and establish a minimum volume of SNP assay product required to produce a trace. Unfortunately, the trial of sample volumes did not produce sample peaks on the trace or shed light on the element that was causing these negative results.

After detailed conversations with a colleague who also uses the SNPStart kit, it was highlighted that the new kit had a different thermal cycling profile from the older kit. Using the old thermal cycling profile it was possible that the polymerase included in the kit mixture was being degraded/denatured by an initial denaturation step and so preventing the SNP-extension reaction. A panel of samples of known haplotype was amplified using the gene 2-specific primers followed by clean-up using ExoSAP-IT, in order to test the new kit with the new thermal cycling profile. The traces produced by SNP-extension reactions under the new thermal cycling conditions showed weak sample peaks. Having established that the SNP extension reaction was working sufficiently to produce weak sample peaks, there were a number of parameters that could be further optimised e.g. increase the primer concentration in the SNP-extension reaction or make alterations to the PCR clean-up process to ensure better quality of DNA.

An exhaustive list of parameters was tested in order to try and optimize the new SNPStart chemistry for use with the bovine MHC interrogation primers including:

- alterations to the SNP-extension reaction thermal cycling profiles
- increasing the interrogation primer concentrations and PCR template concentrations in the SNP-extension reaction

- increasing the inactivation time used for the ExoSAP-IT treatment of gene-specific PCR products
- replacement of all reagents throughout the process to ensure that none had degraded
- use of a different thermal cycler
- inclusion of DMSO in SNP-extension reaction
- use of commercially available sterile water
- use of non-autoclaved plates/tips/tubes
- use of different methods for the clean-up of SNP-extension reactions

5.8. The A14/A15 SNP assay using the Beckman Coulter GenomeLab Methods Development kit

Unfortunately, none of the alterations outlined proved successful for use with the MHC interrogation primers with bovine gDNA- derived PCR products. However, there was one option remaining: to 'resurrect' the SNP assay. The Beckman Coulter GenomeLab Methods Development kit, used for the sequencing of difficult templates, contains all the necessary components to concoct a master mix comparable to that found in the original GenomeLab™ SNP-Primer Extension kit. The Methods Development kit chemistry needed some optimisation to ensure readable peaks and, following advice from the Beckman technical advisor, the use of non-autoclaved plasticware and commercial sterile water was continued.

Having re-optimised the assay with the GenomeLab Methods kit chemistry, the second half of the 40 animal blind trial was re-convened. The results of the second half of the blind trial (see Table 5.3) again demonstrated that the selected markers could efficiently detect the common A14 and A15 haplotypes whilst also indicating whether the animal is hetero- or homozygous for these

haplotypes. As with the first 20 animals of the blind trial, incorrect calls were mainly a result of undetectable haplotypes e.g. the A18 haplotype which does not have a gene 1 or 2 allele. The only sample for which a conclusive result was not achieved was sample number 31 and this appeared to be due to repeated poor amplification of the sample gDNA with the gene 1-specific primers.

5.9. Designing interrogation primers for the identification of the W12B haplotype using the SNP assay

Having optimised the SNP assay for the detection of the A14 and A15 haplotypes, it was decided to use the assay to identify other common MHC class I haplotypes. Preliminary studies revealed that the W12B haplotype (Figure 5.6.) was present in the Canadian bull population at relatively high frequency i.e. the *N*01901* and *N*00801* alleles were present at a frequency of 0.1875 (see Chapter 3). However, due to the divergent intron sequence of the *N*01901* allele, the gene-specific primers were unable to amplify this allele. As a result the identification of this haplotype was based solely on SNPs within the *N*00801* allele. Using an online oligonucleotide design tool (http://www.basic.northwestern.edu/biotools/oligocalc.html) it was found that the majority of sequences flanking the identified SNPs were unsuitable for designing primers i.e. self-annealing or hairpin formation. This reduced the number of suitable SNPs for inclusion in the assay to three (see Figure 5.7.). Multi-plexing the assay with the existing gene 2 interrogation primer added further restriction in terms of primer length. The interrogation primers for detection of the *N*00801* allele must differ by 5 nucleotides in length from the 25 nucleotide interrogation primer used for the detection of the *N*02501* allele.

SAMPLE	GROUP 1			GROUP 2		HAPLOTYPE PREDICTION	ACTUAL HAPLOTYPE	RESULT
	PCR	SNP 55	SNP 129	PCR	SNP 122			
21	-	-	-	+	C	Non A14/A15, with gene 2 only	A11/New5	✓ ✓
22	-	-	-	+	C	Non A14/A15 with gene 2 only	A10/A11	✓ ✓
23	+	T	C	+	T	A14	A14/A18	✓ ✗
24	+	T	G	+	T	A15	A15	✓ ✓
25	+	T	C G	+	T	A14/A15	A14/A15	✓ ✓
26	+	C T	C	+	T	A14 het with gene 1	A13/A14	✓ ✓
27	+	T	C	+	T	A14	A14	✓ ✓
28	+	T	G	+	T	A15	A15	✓ ✓
29	+	T	C	+	T C	A14 het with gene 2	A14/W12B	✓ ✓
30	-	-	-	+	C	Non A14/A15, with gene 2 only	New3/6/7	✓ ✓
31	weak	T?	-	+	T C	A14 or A15 het	A15/A11	✗ ✗
32	+	T	G	+	T C	A15 het with gene 2	A11/A15	✓ ✓
33	+	C T	C G	+	T	A15 het with gene 1	A15/A31	✓ ✓
34	+	T	G	+	T C	A15 het with gene 2	A15/A20	✓ ✓
35	+	T	C	+	T	A14	A14	✓ ✓
36	+	C T	C G	+	T	A15 het with gene 1	A15/New23	✓ ✓
37	+	T	C	+	T	A14	A14	✓ ✓
38	+	T	G	+	T	A15	A15	✓ ✓
39	+	T	C G	+	T	A14/A15	A14/A15	✓ ✓
40	+	T	G	+	T	A15	A10/A15	✗ ✓

Table 5.3. Results of second 20 samples in a blind trial of the A14/A15 SNP test.

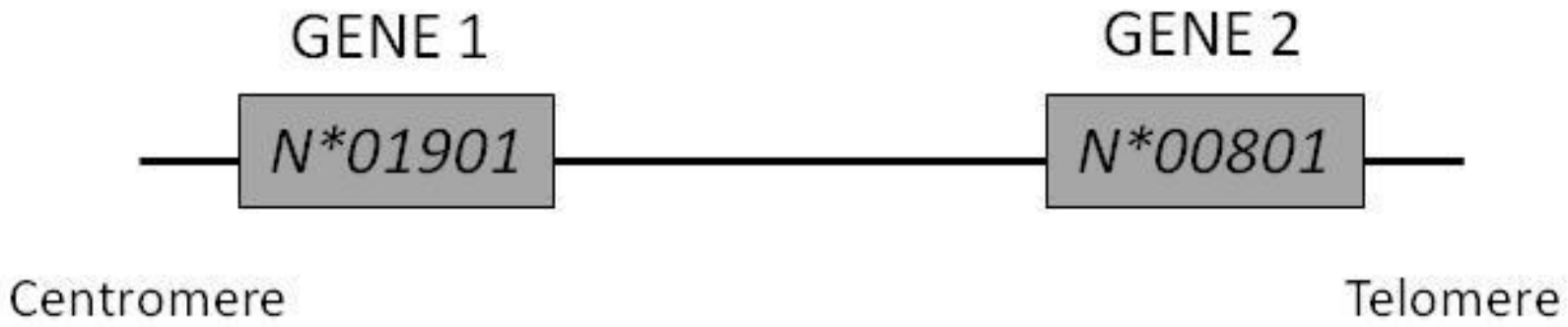


Figure 5.6. Diagrammatic representation of the W12B haplotype.

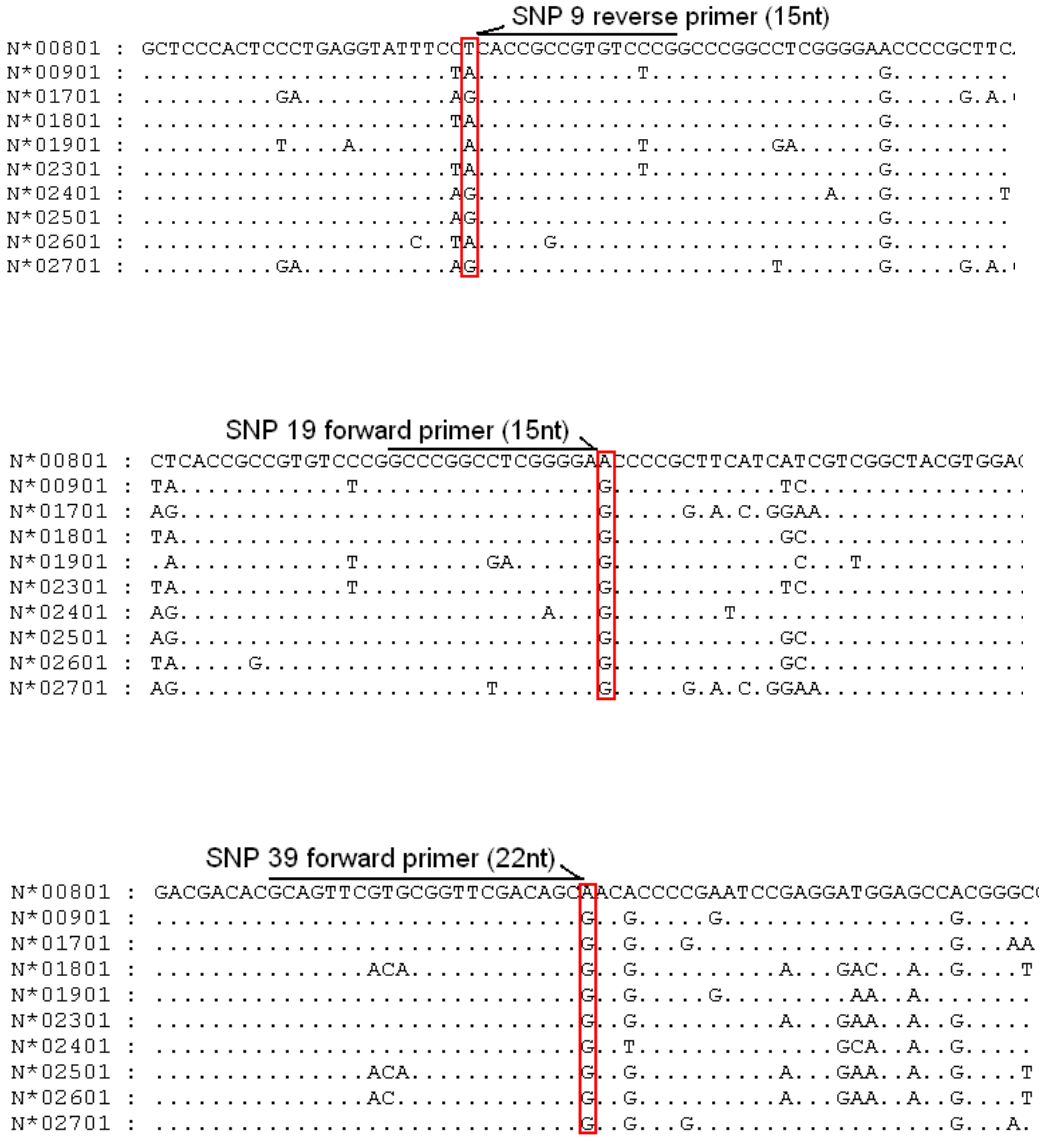


Figure 5.7. The SNPs identified in the *N*00801* allele of the W12B haplotype and their associated interrogation primers. Dots indicate identity, dashes indicate gaps/lack of sequence data compared to reference.

5.10. Optimising the SNP assay for the detection of the W12B haplotype

During initial tests the interrogation primer for SNP 39 was able to detect the *N*00801* allele with a strong signal, but it appeared unable to detect a second allele in heterozygous animals in the final trace. However, the raw data trace did show the presence of a second allele. This primer was subjected to further testing and optimization, and proved a valuable tool for detecting *N*00801*. The two short 15 nucleotide primers for SNPs 9 and 19 produced a signal that was obscured by the size standard peak at a size of 13 nucleotides. This problem was circumvented by the addition of 5 adenine residues on to the 5' end of these interrogation primers. By increasing the size of the SNP extended product, the sample peak was no longer obscured by the size standard peak. The testing and optimisation process for primers SNP9 and SNP19 showed that SNP9 was more accurate at detecting the *N*00801* allele, thus the optimisation of SNP19 primer was discontinued. Ultimately, following optimisation and testing, there are now two interrogation primers to detect the *N*00801* allele of the W12B haplotype to add to the existing A14/A15 assay, the details of which are summarized in Figure 5.8.

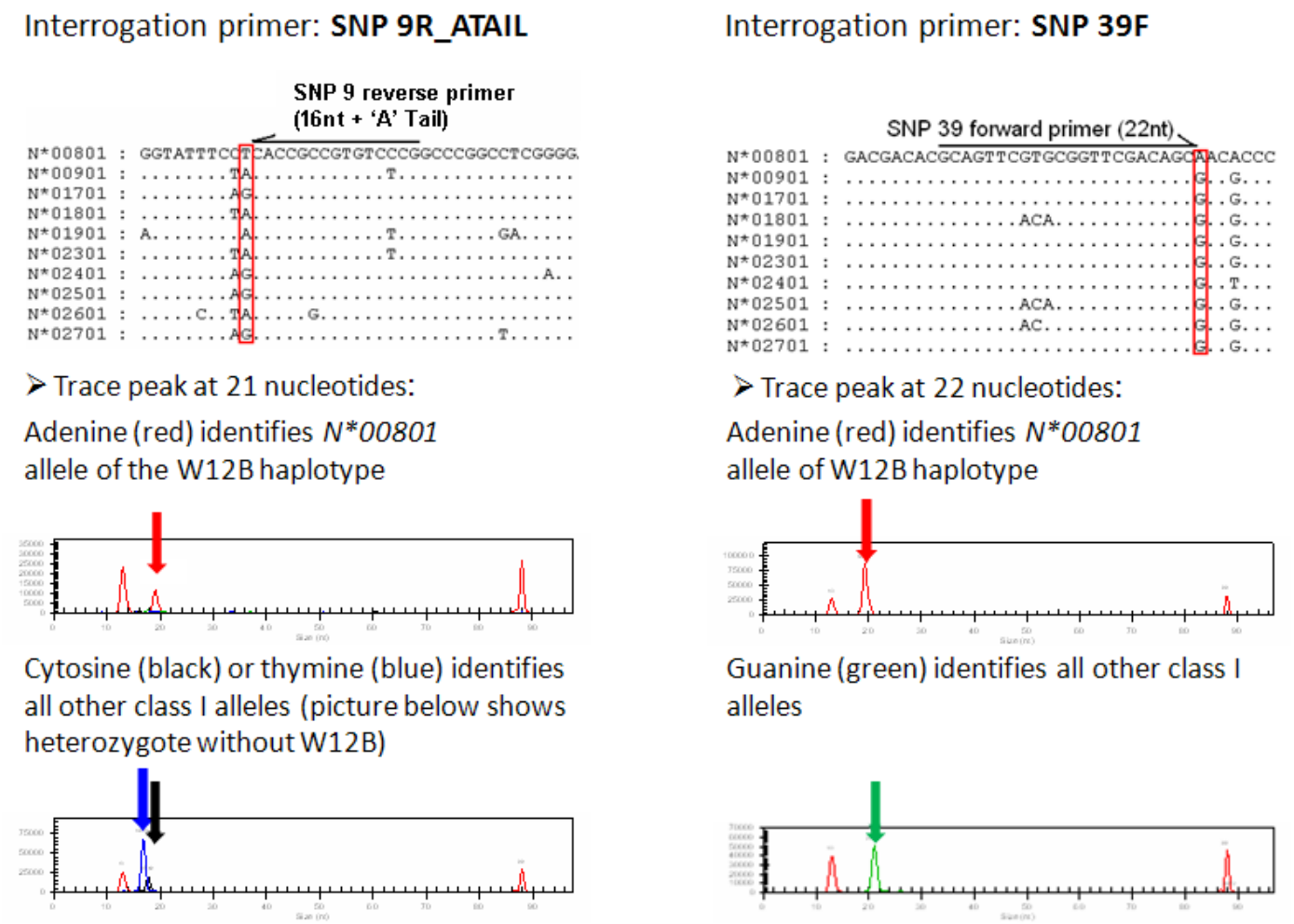


Figure 5.8. Summary of the W12B detection assay using two interrogation primers on a gene 2-specific PCR product template. On the sequence alignments, the dots indicate identity, dashes indicate gaps/lack of sequence data compared to reference.

5.11. Application of the SNP assay to the wider UK Holstein-Friesian population

5.11.1. Extracting gDNA from FTA card-stored blood samples

Collaboration with Dr K. Ballingall at the Moredun Research Institute gave access to a large number of samples from throughout the UK. These are dried blood samples stored on Whatman FTA cards from which 1.2mm discs were punched. The extraction of sufficient good quality DNA from these discs for analysis proved a challenge. This may be due to the initial preparation of the samples, the age/storage of the samples or the requirement for a large gene-specific fragment for PCR amplification prior to analysis.

Numerous attempts to extract good quality DNA following Whatman FTA protocols using Whatman reagents proved unsuccessful. A number of other techniques were tested and a procedure described by Zhou et al., (2006) proved the most efficient. The method involves the incubation of FTA card discs for 30 minutes in 20 mM sodium hydroxide with regular mixing, followed by two five minute washes using TE^{-1} buffer. The discs are then dried and used as the template substrate for PCR amplification.

During optimisation of the Zhou et al. DNA extraction process, a new enzymatic extraction kit from ZyGEM (ZyGEM Corporation Ltd, Hamilton, New Zealand) became available. This kit provided a high-throughput hands-off method for extracting DNA from FTA cards and/or whole blood samples by the lysis and removal of nucleosomes from cells. Trialing this method in the lab demonstrated it to be the most efficient method of extracting DNA from FTA cards, both in terms of the quality of the DNA extracted and the time taken.

Independent of the method of DNA extraction, the gene-specific PCR thermal cycling conditions also required optimisation to include an increased number of cycles. Once the DNA extraction and PCR conditions had been optimised using freshly blotted blood FTA cards, it was decided to check the optimised PCR on the older Moredun sample FTA discs. Amplification of MHC class I genes from the Moredun sample cards was achieved using the optimised DNA extraction method and PCR conditions. The PCR products amplified from FTA card discs were weaker than those obtained from gDNA samples and it was unclear whether there would be sufficient amounts of DNA for application of the SNP assay. The FTA card PCR products were then tested on the capillary SNP assay I had developed, demonstrating that analysis of these samples using this method was feasible.

5.11.2. Application of the SNP assay to assess diversity in the wider UK herd using DNA extracted from FTA-stored blood samples

A collection of samples from 95 Holstein-Friesian cows from 12 herds across the UK were collated. This sample set represented the first large-scale application of the SNP assay for class I molecular typing in cattle and intended to provide useful information as to the prevalence of 'common' MHC class I haplotypes in the wider UK Holstein-Friesian herd. Each sample was tested in duplicate to confirm the results achieved. Positive controls in the form of gDNA extracted from fresh blood and gDNA extracted from FTA blood cards from animals of known haplotype were also included in duplicate in each assay. Selected samples were also subjected to sequencing to prove that the assay was working efficiently and/or to clarify unusual results e.g. chimeric animals. The final SNP-typing results were compiled and the frequencies of A14, A15, W12B and less common haplotypes were calculated (Figure 5.9). In total, there were seven samples for which a conclusive result was not achieved or were apparently chimeric animals and so were omitted from the final frequency calculations; hence the final frequencies were calculated for 88 Holstein-Freisian cows in this study.

5.11.3. Comparing common haplotype frequencies in the Canadian bull population and the wider UK herd

Frequencies of the A14 and A15 haplotypes were comparable between the Canadian bull cohort and the wider UK Holstein-Friesian cow population, with the A15 haplotype showing the highest degree of similarity between the two populations (see Figure 5.9). The W12B haplotype, which was the most common haplotype in the Canadian bull samples, was not as prevalent in the wider UK Holstein-Friesian cow population. This lag in the frequency of common W12B haplotype may simply be a reflection of generation times in cattle breeding and is also testament to the rapidity in which the W12B haplotype has become prevalent i.e. increasing in frequency by 0.16 over 20 years. Previous studies conducted in the 1980s have shown that the A14 and A15 haplotypes were found at a much higher frequency in the Canadian bull population than W12B and this is reflected in its lower frequency in the UK herd sample taken in 2006. Most valuably, the Moredun sample panel provided the proof that the SNP assay is a valuable tool for the rapid detection of common MHC haplotypes in large-scale assessments of MHC class I diversity.

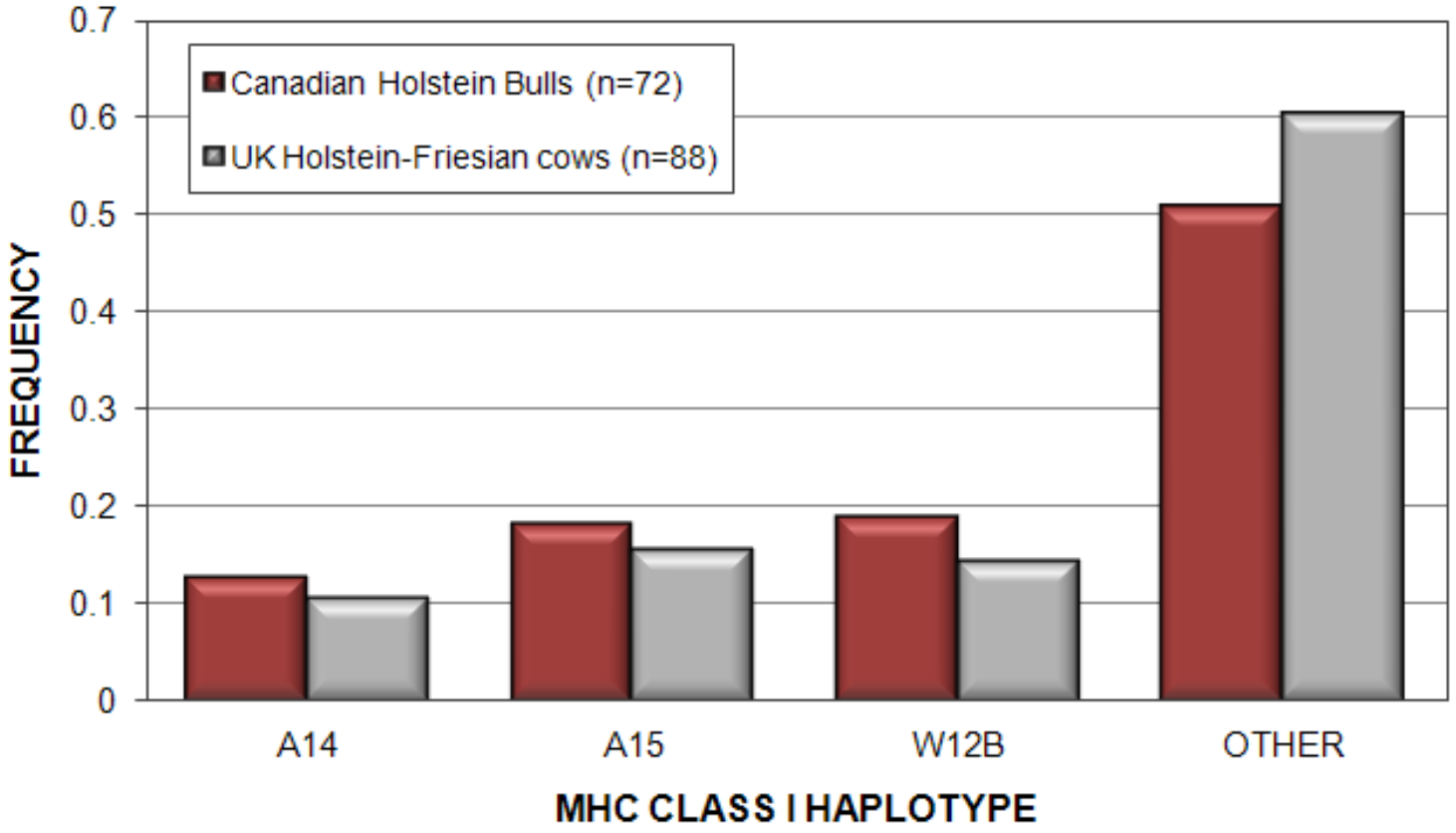


Figure 5.9. A graph to show the observed MHC class I haplotype frequencies in two Holstein-Friesian cattle populations.

5.12. The future of the SNP test

The assay has proved to be a reliable and rapid test for common class I haplotypes using samples from the wider UK herd. The problems encountered had little to do with the markers themselves, but related to the commercial kit chemistry and DNA extraction from old samples. However, in terms of commercialisation, the validated SNP markers would have to be adapted for use on a more commercial platform, for example the Igenity® SNP test uses the Sequenom iPLEX platform which also uses the primer extension method, but products are detected using MALDI-TOF mass spectrometry.

Future development of the test should focus on its expansion to incorporate SNP markers that enable the detection of other relatively common class I haplotypes in Holstein-Friesian cattle herds e.g. A11 and A20. Furthermore, if the number of initial MHC gene-specific PCRs were to be increased i.e. amplification using the gene 3 and gene 6 primers, then the assay could be extended to detect an even wider range of haplotypes e.g. A10 and A18. However, it is important to remember that in such a dynamic system, the most common haplotypes may only remain at such a high frequency for a relatively short period, thus small population MHC frequency studies would be required at regular intervals e.g. every 10 years. This test would work most efficiently in the commercial environment if it detected the most common haplotypes across a wide range of dairy breeds, and breeding decisions were based upon maintaining diversity i.e. putting a cow with a common haplotype to a bull with a less common haplotype. The accumulation of a wide range of samples, through associations with the different herd book societies/AI companies, would make it be possible to ‘group’ bulls according to MHC haplotype as a further addition to the standard progeny-derived trait proofs (Figure 5.10).

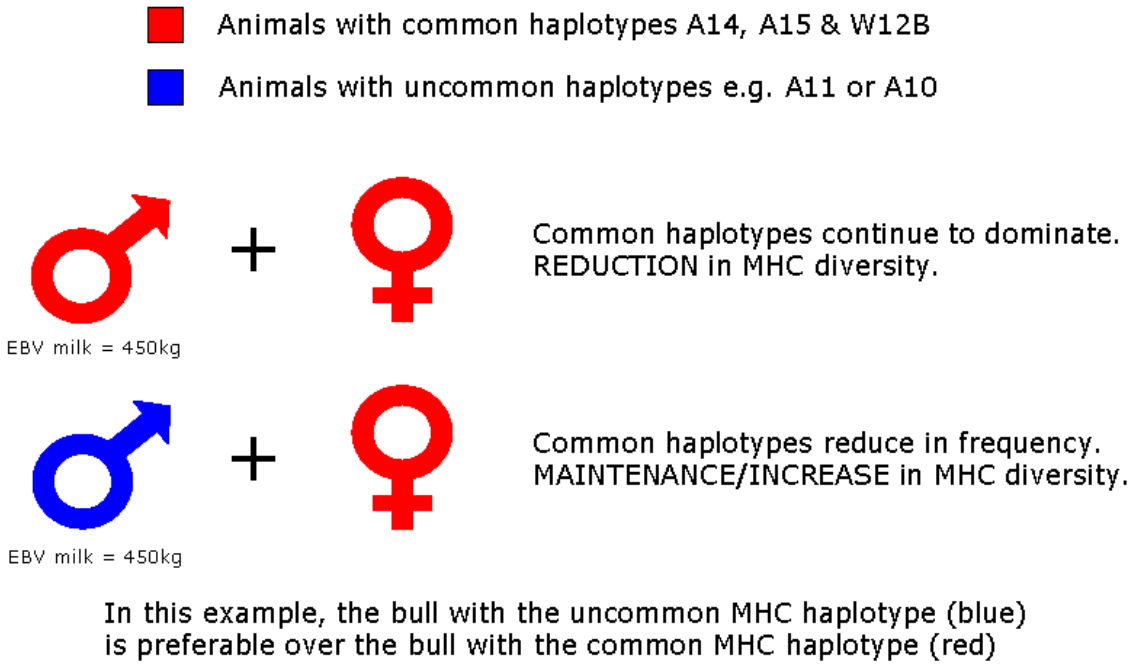


Figure 5.10. Incorporating MHC haplotypes data with standard bull proofs to help make informed breeding decisions which maintain, if not increase, MHC diversity in the herd.

Whilst promoting diversity at MHC loci, the ability to combine this MHC typing data with trait data may facilitate an increase genetic gain in terms of production and fitness traits in dairy cattle. The previous analysis of trait associations with MHC class I haplotypes in the Canadian Holstein AI bull population (see Chapter 4) showed that the common haplotypes were often unfavourably associated with a trait i.e. A15 was associated with decreased fertility, whilst the rarer haplotypes were often favourably associated with selection traits i.e. A11 was associated with an increase in milk yield. The apparent mode of balancing selection driving these associations was heterozygote advantage. Using the SNP assay to type animals for common haplotypes, should enable breeders to formulate breeding programmes that favour matings that give rise to heterozygous offspring i.e. ensuring cows with common haplotypes are mated to bulls of rare haplotypes. In this manner, it is possible make more efficient breeding decisions, the outcomes of which mimic

the selective forces already acting on AI bull populations, and thus increasing the health and productivity of the cattle population.

Another important consideration for development of the SNP test is the incorporation of SNP markers in MHC class II genes and/or other gene families that are involved in the immune response e.g. NK receptors. The MHC class II genes have been more widely studied than class I and although the literature is often contradictory, associations with mastitis resistance/susceptibility have been reported. However, unlike many class I alleles, the class II genes have fewer allele-specific SNP markers and so a more achievable aim would be the use of SNP markers to identify major types at a given class II locus. For example the *DRB3* locus has few allele-specific SNPs, but identification of major types and/or common alleles may be feasible using a series of SNPs, however the number of markers required may render the use of SNPs too expensive. In general, MHC class I and class II genes are tightly linked, thus breeding programmes promoting diversity at class I loci should also increase diversity at class II loci.

The NK cell receptors are encoded by a number of expanded gene families e.g. *KIR, NKG2* and *CD94,* all of which exhibit varying levels of polymorphism. In humans, most *KIR* have been shown to bind MHC class I molecules and there is clear evidence linking certain combinations of *KIR*/*MHC* with disease outcome, for example in hepatitis C and HIV infection. Expansion of the SNP test to encompass a wider range of immune gene families across the many facets of the immune response would be a valuable tool in ensuring that immune gene diversity is maintained despite the pressures of artificial selection for production traits.

Investigating MHC class I diversity in the British Friesian population

6.1. Introduction

The superiority of Holstein-Friesian cattle for milk production and their adaptability to numerous management systems has elevated them to become the dairy breed of choice, constituting over 90 % of the UK dairy herd today. Originating in the Rhine delta of the Netherlands approximately 2,000 years ago, Holstein Friesian cattle were exported during the nineteenth century to both the UK and North America. Initial exports became the founding populations for the development of the breed in each respective country. Prior to the 1980s, the primary focus of breeding programmes in North America and Canada was milk and fat yields with some consideration of conformation traits, whilst the emphasis of selection in the UK and Europe was for a dual-purpose animal i.e. cattle suitable for both dairy and beef production (Cunningham, 1983). During the 1980s, selection for beef traits was eliminated from the majority of European breeding programmes leading to an intensive dairy breeding strategy. Ancestral input coupled with generations of selective breeding has led to the development of a segregation of the Holstein and Friesian 'types' under the classification of a single breed. The name Holstein is now used to denote North American ancestry and the term Friesian indicates European bloodlines, with further segregation within that categorisation i.e. Dutch Friesian or British Friesian. Whilst all pedigree Holstein-Friesian cattle are registered on a single herdbook, ancestral input is indicated by the animal's breed code.

The UK herd of Holstein-Friesian cattle is currently estimated at 2.9 million head. However, the proportion of that total designated as British Friesian population is comparatively small, with a census estimate of 12,000 in 2008 (Wiener et al., 2004). Registration data also show the genetic input of British

Friesian bulls to the UK herd is considerably smaller than that of Holstein bulls e.g. in the two month period of July and August 2010 the top 20 Holstein bulls sired 6269 offspring, compared with only 135 offspring sired by the top 20 British Friesian bulls (Holstein UK: Animal data centre). However, the popularity of the Friesian type is reported to be increasing due to its suitability to a low-input management system with low associated costs, in comparison to its North American Holstein contemporaries (British Friesian Breeders Club website).

Using samples obtained from actively marketed British Friesian bulls and a British Friesian dairy cow herd ranked in the top 20 Friesian herds in the UK, the aim of this study was to compare the levels of MHC diversity in different Holstein-Friesian populations. As with the Canadian Holstein bulls sampled in 2006 (see Chapter 3) the British Friesian bull population sampled is contributing to the genetic make-up of the UK herd thus providing a more comprehensive picture of overall diversity levels in the UK herd. Although a 'type' segregation exists between the two bull populations they have each been exposed to similar selection pressures, and it is hypothesised that only small differences in terms of levels and patterns of MHC diversity exist between the two types.

6.2. Materials and methods

Semen straw samples from 46 British and Irish Friesian bulls were donated to our laboratory from four popular AI companies based in the UK (Genus ABS, Cogent Breeding Ltd, Avoncroft Genetics Ltd and UK Sire Services). Genomic DNA was extracted from each of the semen samples using the protocol outlined in section 2.2.3 of chapter 2. MHC class I genes 1, 2, 3 and 6 were amplified from these samples using gene-specific primers, the products of which were then initially analysed by RSCA performed as described in Birch et al. (2006). Each sample was assigned an MHC haplotype prediction based upon mobility values corresponding to previously sequenced alleles. Limitations of gene-specific PCR and RSCA require the use of SSPs and sequencing to complement missing haplotype data

and/or to confirm ambiguous typing results. Sequences were derived from products amplified with either generic or gene-specific primers. Difficulties in assigning MHC haplotypes to these samples using RSCA analysis meant that the majority were typed by sequencing gene-specific and/or generic PCR products. Final MHC class I typing results were used to calculate both the frequencies of haplotypes and alleles. Results from the 2006 Canadian Holstein bull sample (see Chapter 3) were used as a direct comparison of MHC diversity as the two populations have each undergone similar selection pressures and each have a genetic input into the UK herd. MHC class I diversity was compared using haplotype and allele frequencies together with statistical and phylogenetic analysis of all class I allele sequences detected in each population.

Blood samples were taken from 25 British Friesian cows from a farm in Gloucestershire (courtesy of B. Pullen, chairman of the British Friesian Breeders Club, UK). Peripheral blood mononuclear cells (PBMC) were isolated from venous blood using a standard Histopaque density gradient centrifugation method. The PBMC sample from each animal was divided, with half (approx 5 x 10^6 cells) used for gDNA extraction and the remainder snap frozen and stored at –80°C. MHC class I genes were amplified from these samples using gene-specific primers, the products of which were analysed by RSCA. Gene-specific or generic class I PCR products were sequenced whenever there was an ambiguous RSCA typing result or to confirm whether an animal was homozygous. In those samples that warranted further investigation into the MHC class I genes being transcribed, the snap frozen PBMCs were used for RNA extraction and cDNA production. For amplification of full-length MHC class I from cDNA, a mixture of primers Bov 21a/g and Bov 21-BSF (forward) and a mixture of Bov 3 and Bov 3-BSF (reverse) were used. Sequence data generated from cDNA were used to obtain new allele and haplotype data observed in both the British Friesian cows and bulls. Figure 6.1 shows a diagrammatic representation of the experimental processes involved in obtaining a class I haplotype from the British Friesian samples.

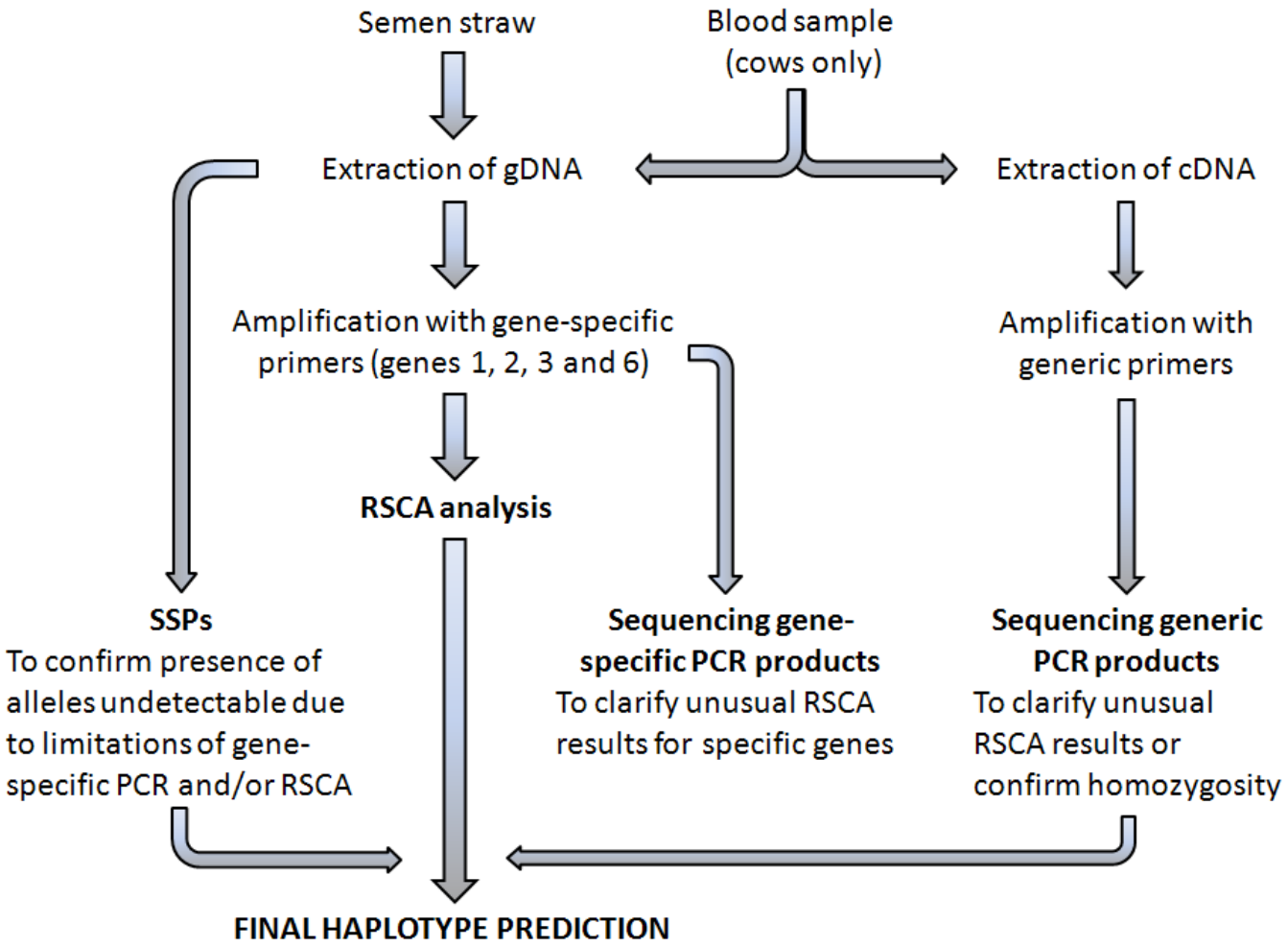


Figure 6.1. Diagrammatic representation of the experimental plan used to determine MHC class I haplotypes in the British Friesian samples.

6.3. Results

Assessing MHC class I diversity in the British Friesian samples using RSCA became problematic due to a large number of previously unseen class I alleles in this population. This encompasses both novel alleles, and alleles that have been previously reported in the literature but knowledge of associated RSCA mobility values are absent. These new alleles were expressed in combination to form previously uncharacterised haplotypes. The availability of cDNA from the British Friesian cow samples enabled the amplification of full length allele sequences for the majority of novel alleles detected and the characterisation of new haplotypes. All full length class I allele sequences detected in the British Friesian samples were submitted to IPD and assigned an official *BoLA* allele name.

6.3.1. Sequencing of generic class I PCR products from British Friesian cow cDNA samples

RSCA analysis indicated that the British Friesian samples contained a large number of novel alleles. The characterisation of full length allele sequences and novel class I haplotypes was achieved by sequencing generic PCR products from British Friesian cow cDNA samples each of which are described below.

6.3.1.1. BF_NEW1 haplotype (previously known as New24)

This haplotype was detected in the British Friesian cows, but not in the bull samples. This haplotype appears to be a single gene haplotype. Sequencing from cDNA to obtain full length sequences yielded a novel gene 2 allele sequence which was submitted to both Genbank (accession number GQ488022) and IPD (official allele name *N*05401*). Whilst comparing class I diversity between the British Friesian populations and the Canadian Holstein bull population it was

realised that the allele designated as New24 expressed on the New24 (Chapter 3) haplotype in the Canadian bulls is the *N*05401* allele found in the British Friesian cows. However, as full molecular characterisation of this haplotype was derived from British Friesian samples, it shall be known as BF_NEW1.

6.3.1.2. BF_NEW2 haplotype

The BF_NEW2 haplotype has only been found in the British Friesian samples and the availability of cDNA from British Friesian cows enabled the identification of the alleles expressed. RSCA analysis of gDNA samples gave unusual mobility values for both the gene 2 and gene 3-specific PCR products against their respective reference strands. British Friesian cow cDNA samples were used as sequencing templates to obtain full length allele sequences and enable identification of the alleles expressed. Both alleles expressed on this haplotype had been previously reported, although only a partial sequence was available on the Genbank database for the gene 2 allele. The gene 3 allele detected differed from *N*01701* of the A11 haplotype by 4 amino acids across alpha 1 and 2 and so is classed as a variant with the official IPD name *N*01702*. The gene 2 allele detected was 98 % similar at the nucleotide level to the *N*02601* allele of the A20 haplotype. The high degree of similarity of the gene 2 and gene 3 alleles of the BF_NEW2 haplotype to alleles of the A20 and A11 haplotypes respectively suggests this haplotype evolved through inter-locus recombination (see Figure 6.2).

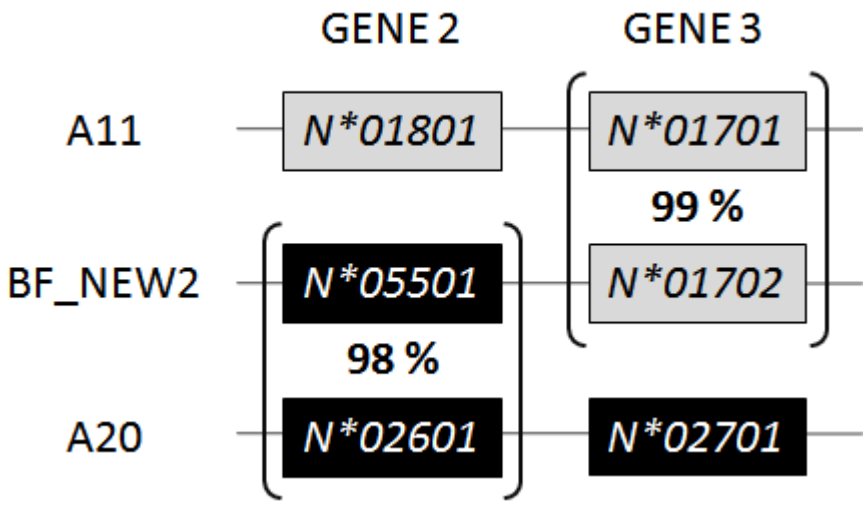


Figure 6.2. Diagrammatic representation of the alleles expressed on the BF_NEW2 haplotype. Similarity with alleles detected on the A11 and A20 haplotypes are denoted in the brackets.

6.3.1.3. BF_NEW3 haplotype

As with the BF_NEW2 haplotype, RSCA analysis of gene-specific PCR products from British Friesian gDNA gave unusual mobility values compared to the gene 2 and gene 3 reference strands. Initial sequencing reactions using gDNA templates and generic primers to amplify alpha 1 and 2 gave sequences which were difficult to assign an identity i.e. BLASTn searches gave more than one possibility as to the identity of each of the alleles on this haplotype. The inability to assign a definite identity to each of the alleles was a result of insufficient sequence available i.e. some class I sequences submitted to GenBank are identical across alpha 1 and 2 but differ across the remainder of the sequence. British Friesian cow cDNA samples were used as sequencing templates to obtain full length allele sequences and enable identification. The gene 3 allele sequence was identical to that of a previously submitted sequence (IPD allele name *N*05101*; Genbank accession number DQ121193; Babiuk et al., 2007). The gene 2 allele sequence

detected was 99 % similar to the *N*04401* sequence (Genbank accession number DQ121161) except in all of the animals examined the sequence appeared to be a variant consistently expressing two amino acid changes in alpha 2. The *N*04401* variant sequence from the British Friesian samples was submitted to Genbank, and assigned the official IPD name *N*04402*.

6.3.1.4. BF_NEW4 haplotype

Analysing the gene 6-specific PCR product from the British Friesian bull sample GEN1 by RSCA gave a mobility value similar to that of the *N*01502* allele found on the A17 haplotype. As a result, it was decided to sequence the gene 6-specific PCR product to confirm the predicted presence of the *N*01502* allele, and also sequence the gene 2-specific PCR product to confirm the presence of the *N*00602* and *N*00802* alleles, and hence the A17 haplotype in this animal. The nucleotide sequence of the gene 6-specific PCR product from GEN1 was in fact identified as *N*01501* and not *N*01502* as previously predicted by RSCA analysis. Although the *N*01501* and *N*01502* alleles differ by two amino acids across alpha 1 and 2, there is no difference in the molecular weight of the nucleotide sequences, thus accounting for the initial misidentification.

Both the *N*01501* and *N*01502* alleles have been previously reported in combination with two gene 2 alleles (*N*00602* and *N*00802*) on two closely related variants of the A17 haplotype. Sequencing of the gene 2-specific PCR products showed that the *N*00602* and *N*00802* alleles were absent, and only a single gene 2 allele sequence (Genbank Accession number DQ121177) was detected. This allele had been previously reported in a study of *BoLA* diversity in Charolais cross cattle (Babiuk et al., 2007). The availability of a full length sequence for this allele enabled its submission to IPD and assignation of an official allele name, *N*04501*. The presence of the single gene 2 allele *N*04501* expressed together with the *N*01501* allele on the BF_NEW4 haplotype was later

confirmed in a British Friesian cow cDNA sample. The configuration of alleles on the BF_NEW4 haplotype compared with that of A17 is shown in Figure 6.3.

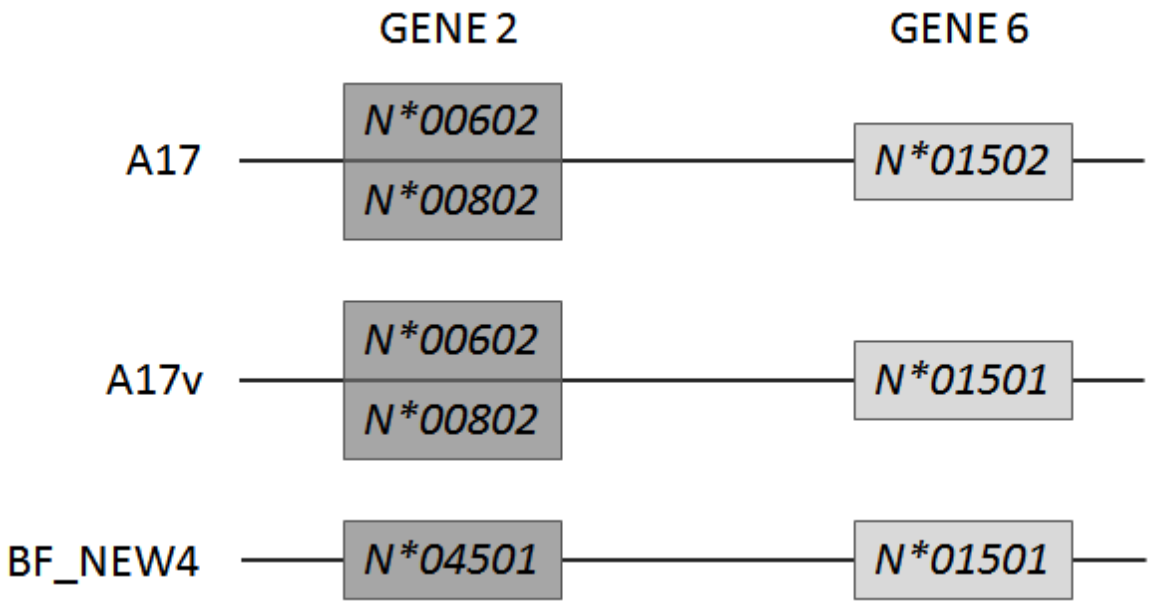


Figure 6.3. Diagrammatic representation of the alleles expressed on the BF_NEW4 haplotype compared with that of the conventional A17 haplotypes.

6.3.2. Sequencing of generic class I PCR products from British Friesian gDNA samples

RSCA analysis of British Friesian bull samples indicated the presence of new class I alleles and haplotypes. Sequencing generic PCR products from gDNA templates enabled identification of these previously reported sequences on these haplotypes. In some cases, alleles were found in varying combinations and the lack of cDNA from these animals meant that only partial characterisation of these haplotypes was possible (see sections 6.3.2.3 and 6.3.2.4).

6.3.2.1. A15v haplotype

RSCA analysis of British Friesian bull sample GEN7 showed a gene 2 allele assigned as *N*02501* based on its mobility value, but the mobility values of the gene 1 allele against the *N*02101* and *N*02301* reference strands were novel i.e. different to the mobility values of *N*00901* or *N*02301* (see Figure 6.4). SSPs were used to confirm the presence of the *N*02401* allele in this sample. Like the A14 and A15 haplotypes, this new haplotype has *N*02501* and *N*02401* but instead of having either *N*00901* or *N*02301*, it has a new gene 1 allele (Genbank Accession number DQ121149.1; Figure 6.5) which has been previously described (Babiuk et al., 2007). This allele has also been previously found in combination with *N*02401* and *N*02501* in a New Zealand Friesian animal in the laboratory sample repository. Closer examination of the DQ121149.1 sequence shows homology to *N*02301* in alpha 1 and homology to *N*00901* in alpha 2 (see Figure 6.6a). The leader, alpha 3, transmembrane domain and cytoplasmic domains sequences of each of gene 1 allele associated with *N*02501* and *N*02401* are identical. On the basis of amino acid differences, the DQ121149.1 is classed as a variant of *N*00901* and thus named *N*00902* (see Figure 6.6b).

The A15v haplotype is present at low frequency in the British Friesian bull sample, whilst the A15 haplotype remains absent from both the British Friesian samples. Again, this may reflect the divergence between the Holstein and Friesian ‘type’ with the separation of the A15 and A15v haplotypes occurring at the founder population stage.

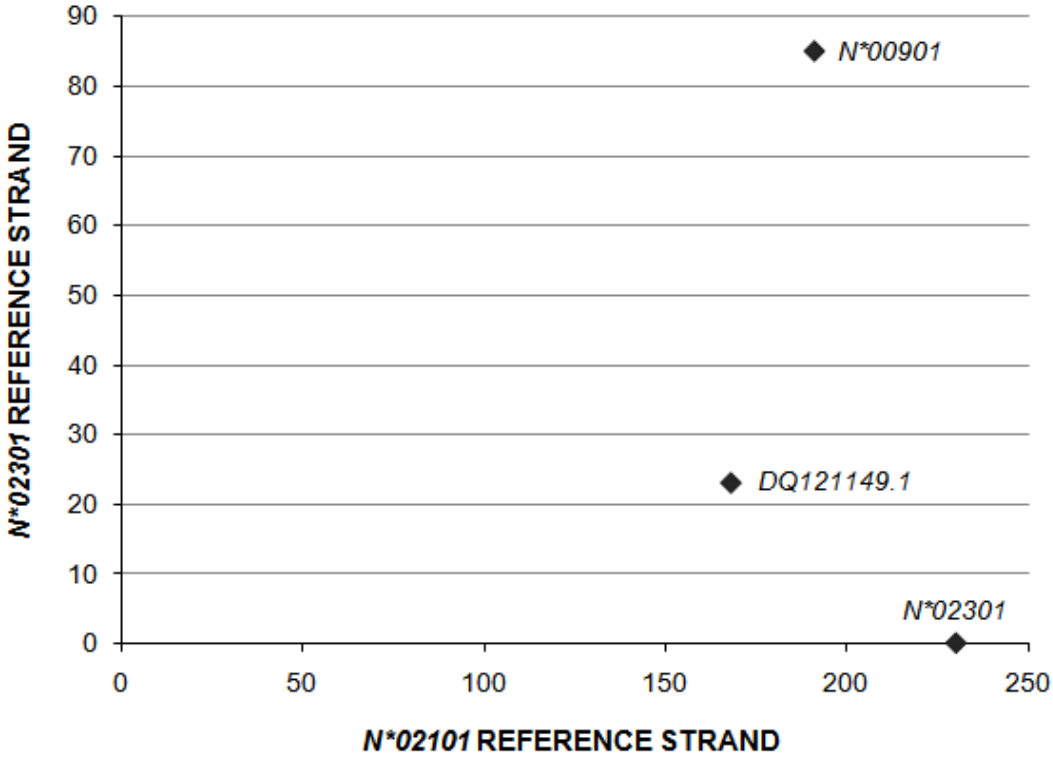


Figure 6.4. Plot of mobility values for gene 1 alleles N*00901, N*02301 and DQ121149.1 against the *N*02101* and *N*02301* allele reference strands.

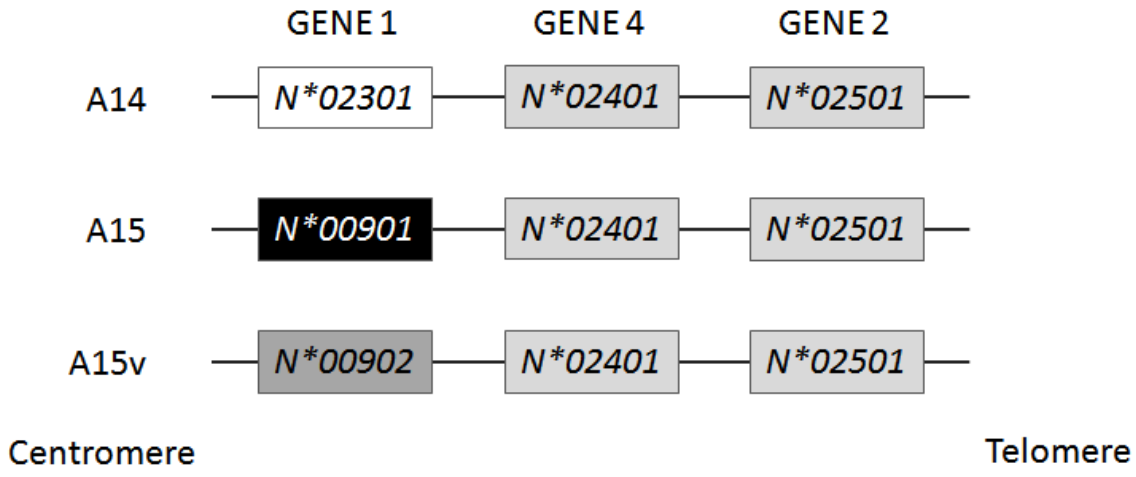
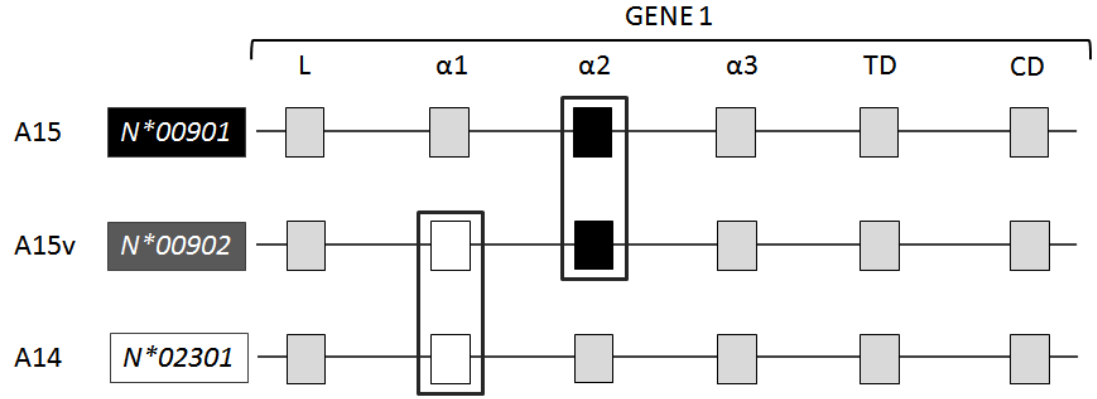


Figure 6.5. Comparison alleles expressed on the A14, A15 and A15v haplotypes. As the A14 haplotype has been previously mapped (Di Palma et al., 2002), the gene configuration of the A15 and A15v haplotypes is inferred due to the expression of the same genes on each of these haplotypes.

A.



B.

```
                                                                  Alpha 1
          -20       *         0         *        20        *        40        *        60        *        80        *
N*00901 : ----TLFVLLLGALALTETRAGSHSLRYFYTAVSRPGLGEPRFISVGYVDDTQFVRFDSDAPDPRMEPRAPWIEKEGPEYWDRETRISKENTLVYRGSLNNLRGYYNQSEA
N*00902 : MGPR.....................................................E.....E.....................
N*02301 : MGPR.....................................................E.....E.....................

          Alpha 2                                                                                                 Alpha 3
                100       *       120       *       140       *       160       *       180       *       200
N*00901 : GSHTLQLMYGCDVGPDGRLLRGYRQDAYDSRDYIALNEELRSWTAADTAAQITKRKWEAEGYAESLRNYLEGRCVEWLRRYLENGKDALLRADPPMAHVTHHPSSEREVTL
N*00902 : ..........................................................................................................
N*02301 : ...N..A.......S..SF....S.....G........D...................................................................

                                                                                  Transmembrane Domain
                *       220       *       240       *       260       *       280       *       300       *
N*00901 : RCWALGFYPKEISLTWQREGEDQTQDMELVETRPSGDGTFQKWAALVVPSGEEQRYTCHVQHEGLQEPLILRWEPPQTSFLIMGIIVGLVLLVVAVVAGAVIWRKKRSGEK
N*00902 : ..........................................................................................................
N*02301 : ..........................................................................................................

          Cytoplasmic Domain
               320       *
N*00901 : RQTYTQAASGDSDQGSDVSLTVPKV*
N*00902 : .........................
N*02301 : .........................
```

Figure 6.6. Comparison of the gene 1 alleles associated with A14, A15v and A15. **A.** Diagrammatic representation of exon homology between *N*00901*, *N*00902* and *N*02301*. **B.** Sequence alignments of *N*00901*, *N*00902* and *N*02301*. Dots indicate identity, dashes indicate gaps/lack of sequence data compared to reference and the asterisk represents the stop codon.

6.3.2.2. BF_NEW5 haplotype

This haplotype was found when using generic sequencing from gDNA to confirm homozygosity in a British bull sample believed to be an A14 homozygote. Generic sequencing confirmed the presence of the A14 haplotype alleles but also the presence of the *N*01601* allele. The gene 2 allele *N*01601* is usually found expressed on the A19 haplotype in combination with the gene 6 allele *N*01401*, but this allele could not be detected in this sample. This may simply be a result of the experimental process i.e. an insufficient number of clones were sequenced or a mutation within the generic primer binding site preventing the amplification and subsequent detection of this allele. However, an incomplete A19 haplotype in an animal of Friesian type has been previously reported, although in this instance it was the *N*01601* that was absent, whilst the *N*01401* allele was detectable (Ellis et al., 1998). Figure 6.7 shows a comparison of the gene combinations expressed on the A19, BF_NEW5 and the A19-like haplotype reported in Ellis et al., 1999. This indicates that the alleles of the A19 haplotype have previously segregated and that an incomplete A19 haplotype is plausible explanation for the haplotype configuration observed. Another possibility is that recombination between two single gene haplotype occurred to give rise to the A19 haplotype, which has now become fixed in the Holstein type.

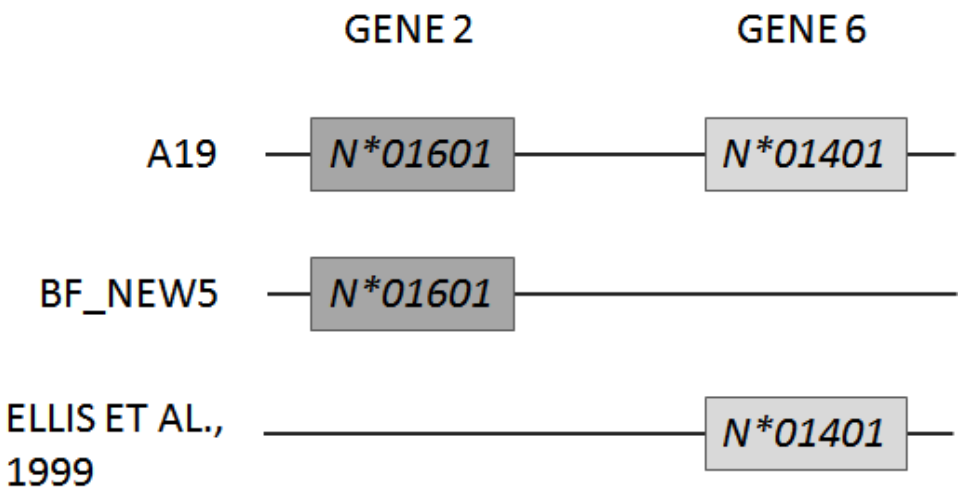


Figure 6.7. Comparison of gene combinations expressed on A19, BF_NEW5 and an A19-like haplotype reported in the literature.

6.3.2.3. BF_NEW6 haplotype

Bovine class I haplotypes have been shown to be highly conserved and the detection of a combination of alleles in more than one animal prompts the assumption that a new class I haplotype has been detected. However, due to the lack of cDNA samples for animals with this new haplotype, full characterisation has not yet been possible and all data presented for this haplotype was derived from gDNA analysis. Table 6.1 details the alleles detected in each of the six British bulls with the proposed BF_NEW6 haplotype. These sequences lack an official *BoLA* name because despite their submission to Genbank by another laboratory, these sequences have not been previously published in the literature and full length cDNA sequences could not be obtained during this investigation.

A gene 3 allele sequence was detected in six of the 46 British Friesian bulls, which when subjected to a BLASTn search, was 99 % similar to the Genbank accession sequence number DQ121187. However, in each of these six bulls the gene 3 allele sequence detected, independent of the primers used to amplify the template, had a single amino acid change from a lysine to an arginine residue at position 157 (see Figure 6.8). Alignment of the Genbank DQ121187 sequence with all bovine class I sequences available on IPD shows the presence of a lysine residue at position 157 to be highly unusual as the presence of an arginine residue at this position is conserved in all other class I sequences. The consistent detection of a an arginine residue at position 157 in the British Friesian bull gene 3 allele sequences is in keeping with the conservation of this residue in all other class I sequences. However, as a single nucleotide change is responsible for this coding change, the unusual lysine residue at position 157 in the submitted DQ121187 sequence could be a result of PCR error.

Sequencing generic PCR products (primers Bov7 and Bov11) in the six animals with the DQ121187-like gene 3 allele detected another allele sequence in apparent combination in five of the six animals. The lack of the second allele in

animal GEN26 could be a result of sequencing an insufficient number of clones, or that this animal carries an incomplete haplotype e.g. expressing only one allele when normally two alleles are expressed on this haplotype. Following a BLASTn search, three possible identities could be assigned to the second allele sequence; DQ121173, *N*05301* or *N*00402*. Difficulties in assigning a definitive identity to the second allele was a result of homology of the DQ121173, *N*05301* and *N*00402* sequences throughout the region amplified by Bov7 and Bov11 primers. Examination of the transmembrane and cytoplasmic domain of these three allele sequences enables assignment to a class I locus, with DQ121173 predicted to be encoded by gene 1, whilst *N*05301* and *N*00402* are both gene 3 alleles. In terms of haplotype structure, this presented two possibilities; a duplicated gene 3 with two alleles expressed or a single gene 3 expressed in combination with a gene 1 allele (see Figure 6.9). The expression of gene 1 and gene 3 together on the same haplotype represents a previously unseen class I haplotype gene combination, whereas gene duplication has been previously reported in bovine class I haplotypes e.g. A17. To solve the identity of the second allele and BF_NEW6 haplotype structure, further sequencing of generic PCR products amplified with primers Bov21 and Bov11 was carried out. Amplification with this primer pair generates a sequence including the leader sequence and the start of alpha 1, throughout which there are sufficient differences between DQ121173, *N*05301* and *N*00402* to enable an identification of the second allele, which was not possible with sequence amplified by primer pair Bov7 and Bov11. In the three animals from which Bov21 and Bov11 PCR products were sequenced, only two yielded sequence from the associated allele. In each case the associated allele sequence was hybridised with another allele sequence in the alpha 2 domain (see Figure 6.10). The inclusion of the start of alpha 1 and the leader sequence data, it can be assumed that the most likely identity of the associated allele is DQ121173, and that the BF_NEW6 haplotype consists of the novel gene 1 and 3 combination (see option 2 in Figure 6.9).

ANIMAL	HAPLOTYPE	ALLELE SEQUENCES DETECTED
GEN3	A18/**BF_NEW6**	*N*01301*, **DQ121173, DQ121187***
GEN20	A18v/**BF_NEW6**	*N*01302*, **DQ121173, DQ121187***
GEN21	A17/**BF_NEW6**	*N*00602, N*00802, N*01502,* **DQ121187*, DQ121173 hybrid**
GEN23	**BF_NEW6**/BF_NEW4	**DQ121173, DQ121187***, *N*04501, N*01501*
GEN26[$]	A10/**BF_NEW6**	*N*01201, N*00201,* **DQ121187***
GEN27	A31/**BF_NEW6**	*N*02101, N*02201,* **DQ121173 hybrid, DQ121187***

Table 6.1. Alleles detected in each of the British bulls with the proposed BF_NEW6 haplotype. * The DQ121187 sequence detected has a 1 amino acid change in alpha 2 compared to the sequence submitted to Genbank.[$] Possible incomplete haplotype detected in this animal.

```
          -20       *         0         *        20         *        40         *        60         *        80
DQ121187 : MGPRTLLLLLSGVLVLTETRAGSHSLRYFYTAVSRPGLGEPRYLEVGYVDDTQFVRFDSDAPNPRMEPRTRWVKQEGPEYWDQETRKAKDTAQTFRVDLNTLRGYYNQSE
GENE_3   : ----------------------.....................................................................................
BOV7&11  : ------------------------------------------.....................................................................
BOV21&11 : .....................................................................................................

          *       100         *       120         *       140         *       160         *       180         *      200
DQ121187 : AGSHTLQLMYGCYVGPDGRLLRGFTQYGYDGRDYLALNEDLRSWTAVETAAQISKRKMEAAGEEERFKNYLEGTCVEWLRRYLENGKDTLLRADPPKAHVTHHPISGHEV
GENE_3   : ..............................................................R...................----------------------
BOV7&11  : ..............................................................R....-------------------------------------
BOV21&11 : ..............................................................R....-------------------------------------
```

Figure 6.8. Alignment of DQ121187 with sequences derived from British Friesian bull samples using three different primer pairs; gene 3-specific, and two class I generic pairs (Bov7 and 11; Bov 21 and 11) consistently showing the lysine to arginine change at position 157. Dots indicate identity, dashes indicate gaps/lack of sequence data compared to reference.

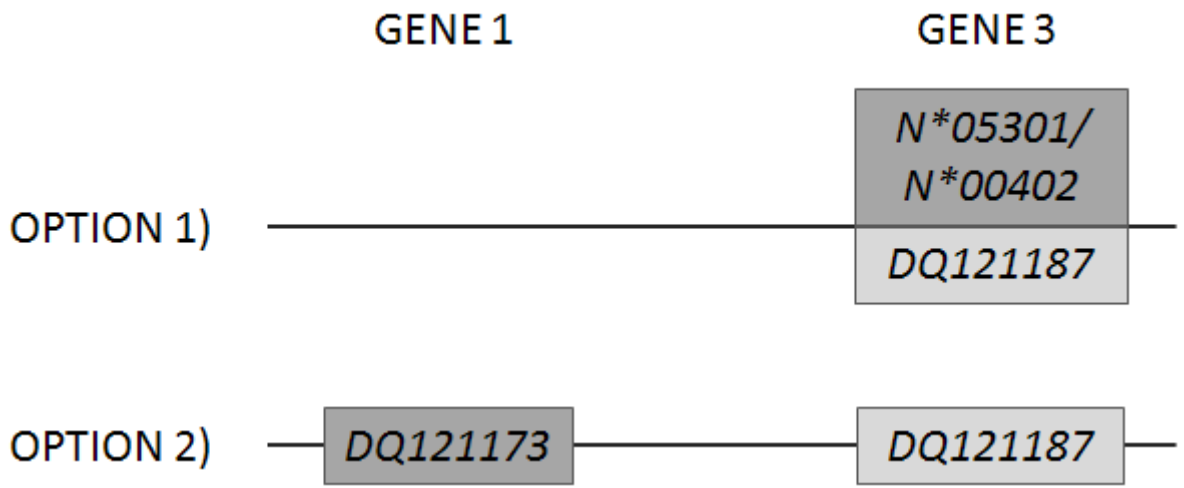


Figure 6.9. Possible gene combinations for haplotype BF_NEW6 depending upon allele identity.

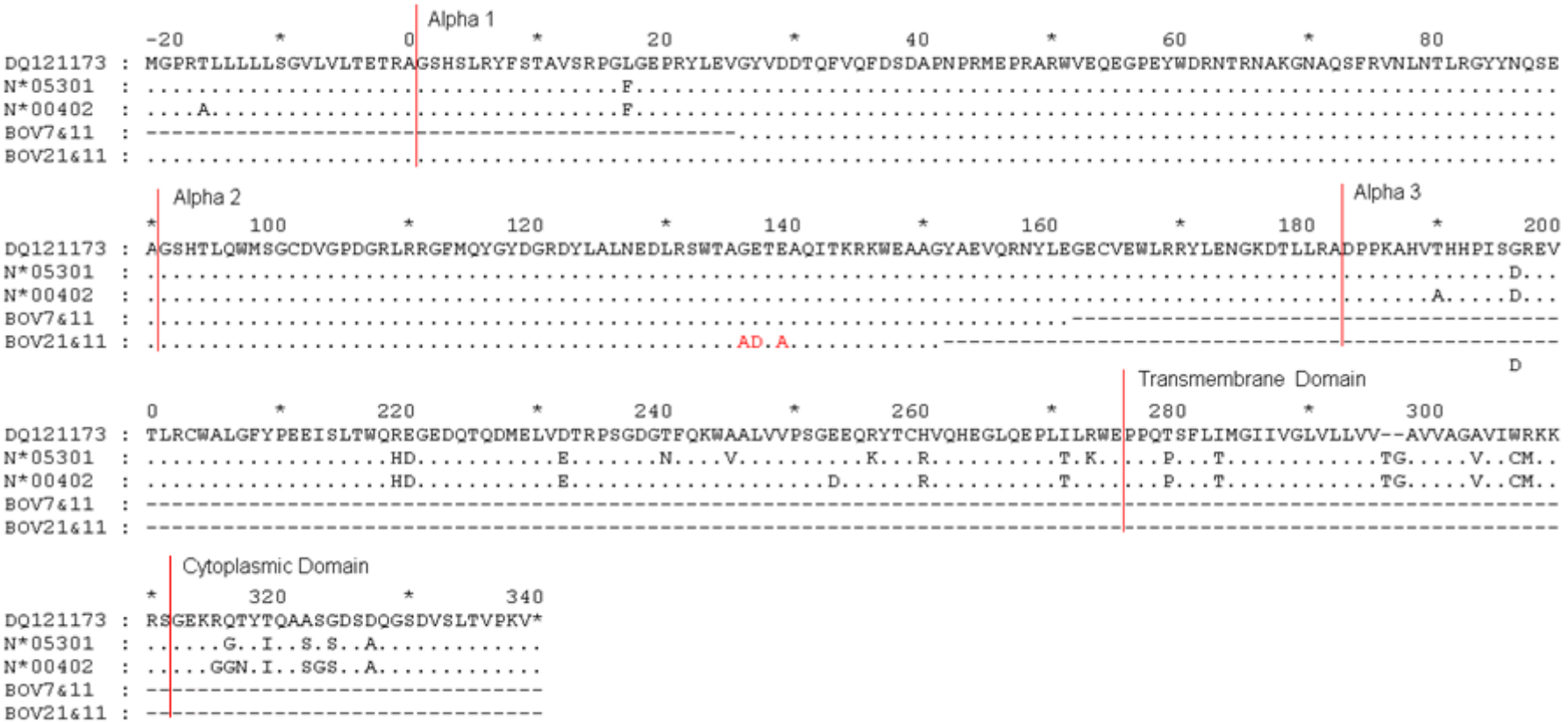


Figure 6.10. Alignment of sequences derived from British Friesian bulls with the BF_NEW6 haplotype using generic primer pairs Bov7 and Bov11, or Bov21 and Bov11. The letters in red indicate the hybrid sequence amplified using the generic Bov21 and 11 primer pair. Dots indicate identity, dashes indicate gaps/lack of sequence data compared to reference and asterisk represents the stop codon.

6.3.2.4. BF_NEW7 haplotype

As with the BF_NEW6 haplotype, this haplotype has only been characterised in gDNA samples. This haplotype was detected in eight of the 46 British Friesian bulls (see Table 6.2 for details of alleles detected in each of the eight animals with this haplotype). The *N*04801* alleles was detected in each of the eight animals with 100% similarity across alpha 1 and alpha 2. The detection of the *N*00402* and *N*01101* alleles thought to be associated with the *N*04801* allele was variable in these eight animals (see Figure 6.11). Full length cDNA sequences for the two variably expressed alleles, *N*00402* and *N*01101,* have been submitted to Genbank and the lack of non-classical characteristics e.g. a non-classical motif in the transmembrane domain or truncation/alternative splicing in the cytoplasmic domain, indicates that these two allele sequences are indeed classical. The variability of the *N*00402* and *N*01101* alleles on the BF_NEW7 haplotype could be explained by the fact that these are alleles are not expressed at high levels and are so are not always detected by sequencing a finite number of clones. In animal GEN22, which appeared to be a BF_NEW7 homozygote, only these three allele sequences were detected, providing evidence that this haplotype expresses three alleles.

ANIMAL	HAPLOTYPE	ALLELES DETECTED	SEQUENCING TEMPLATES
GEN8	A31/**BF_NEW7**	*N*02101, N*02201,* ***N*04801, N*00402***	Generic sequences from gDNA.
GEN9	A18/**BF_NEW7**	*N*01301,* ***N*04801***	Gene-specific and generic sequences from gDNA.
GEN10	A14/**BF_NEW7**	*N*02301, N*02501, N*02401,* ***N*04801***	Generic sequences from gDNA.
GEN11	A11/**BF_NEW7**	*N*01801, N*01701,* ***N*04801, N*00402***	Generic sequences from gDNA.
GEN22	**BF_NEW7**	***N*04801, N*00402, N*01101***	Generic sequences from gDNA.
GEN33	A18v/**BF_NEW7**	*N*01302,* ***N*04801, N*00402, N*01101***	Gene-specific and generic sequences from gDNA.
BLG4	BF_NEW4/ **BF_NEW7**	*N*04501, N*01501,* ***N*04801, N*01101***	Gene-specific and generic sequences from gDNA.
BLG5	A14/**BF_NEW7**	*N*02301, N*02501, N*02401,* ***N*04801, N*00402***	Generic sequences from gDNA.

Table 6.2. Alleles detected in each of the British bulls with the BF_NEW7 haplotype.

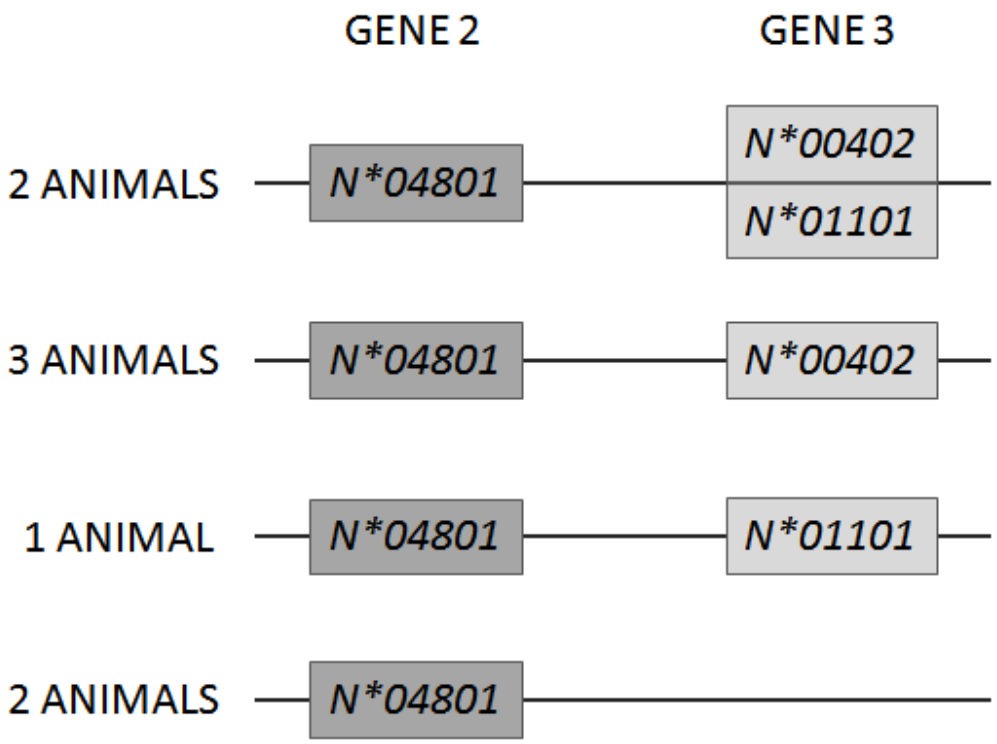


Figure 6.11. Diagrammatic representation of the BF_NEW7 haplotype with variable gene and allele expression from a total of eight British Friesian bulls.

6.3.3. Investigation of new haplotypes using SSP-PCR

With full length class I allele sequences available, it was possible to design sequence-specific primer pairs for PCR amplification of alleles of interest to confirm the presence of, or help characterise MHC class I haplotypes. The New5 haplotype, previously reported in the Canadian Holstein population (see Chapter 3), was originally believed to be a single gene haplotype expressing one gene 3 allele, *N*03601*. Sequencing cDNA from a British Friesian cow with the New5 haplotype revealed that this haplotype expressed an additional gene 3 allele (*N*03701*) and a gene 2 allele (*N*05601*). As the cow from which these additional allele sequences were detected also had another uncharacterised haplotype, investigation into the alleles expressed on the New5 haplotype was extended to incorporate more samples. Generic sequencing of a cDNA sample from a Canadian Holstein animal MHC-typed as New5/A15 detected both the *N*03601* and *N*03701* alleles, but *N*05601* was not present. This prompted speculation that the New5 haplotype has diverged into two lineages with the duplicated gene 3 alleles *N*03601* and *N*03701* expressed in both Canadian Holstein and British Friesian animals, whereas the gene 2 allele *N*05601* detected in British Friesian samples is absent in Canadian Holstein bulls (see Figure 6.12). Sequences derived from a crossed animal, i.e. 50 % from each lineage, did not include the *N*05601* allele, but this may reflect inheritance of this haplotype from the Canadian Holstein parent.

The inability to amplify the *N*03701* and *N*05601* alleles with gene-specific primers and the availability of full length cDNA sequences, enabled the design of sequence-specific primer pairs for each of these alleles (see Table 6.3). The *N*03701* and *N*05601* SSPs were applied to gDNA samples from both the Holstein and Friesian lineages to confirm whether the New5 haplotype had segregated with breed type. The results of SSP-PCR combined with sequencing data from both lineages indicates that the alleles expressed on the New5 haplotype are dependent upon breed type.

ALLELE	CODING (5'-3')	NON-CODING (5'-3')
N*03701	CGGGAGGCGGCAGGTGT	ACAAACTCCATGTCCTGCA
N*05601	GCTTCATCGCCGTCGGCTAT	TACTGCGTGAACCCGCT

Table 6.3. SSP sequences for detection of *N*03701* and *N*5601* from gDNA samples

LINEAGE	ANIMAL	SUBSTRATE	ALLELES DETECTED BY SEQUENCING	ALLELES DETECTED BY SSPs
FRIESIAN	BF100	cDNA	*N*03601, N*03701, N*05601*	*N*03701, N*05601*
		gDNA		*N*03701, N*05601*
	GEN29	gDNA	*N*03601, N*05601*	
HOLSTEIN	4936	cDNA	*N*03601, N*03701*	
		gDNA		*N*03701*
	CB4	gDNA	*N*03601*	*N*03701*
	CB6	gDNA	*N*03601*	*N*03701*
	CB29	gDNA	*N*03601*	*N*03701*
	CB31	gDNA	*N*03601*	*N*03701*
HOLSTEIN-FRIESIAN CROSS	GEN5	gDNA	*N*03601*	*N*03701*

Table 6.4. Alleles of the New5 haplotype detected by sequencing and SSP-PCR in samples from both the Holstein and Friesian type cattle.

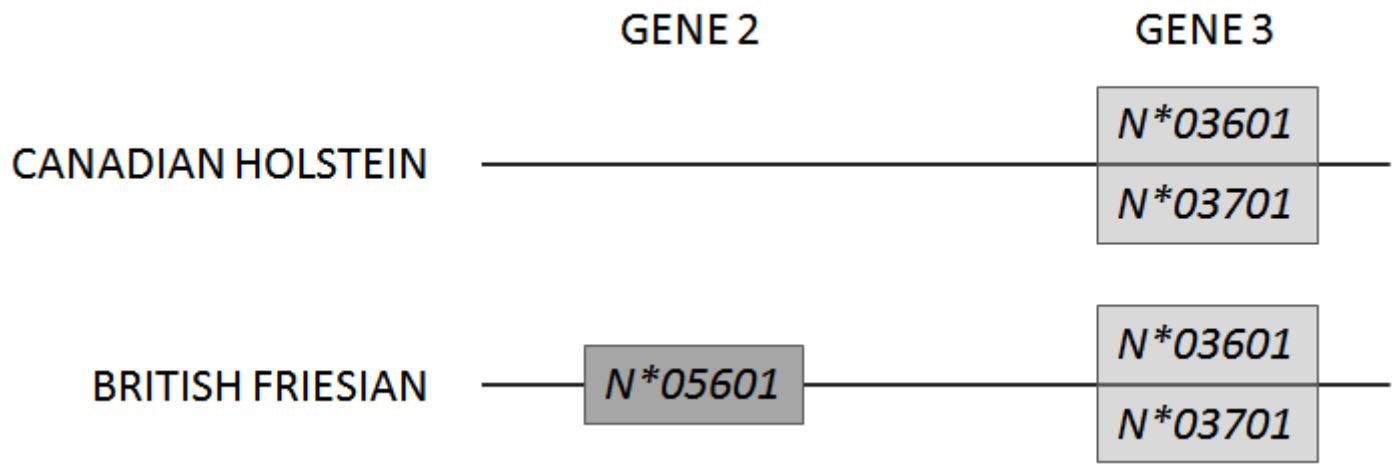


Figure 6.12. Diagrammatic representation of the genes expressed on the New5 haplotype in both lineages.

6.3.4. Analysis of MHC class I diversity in British Friesian bulls

6.3.4.1. MHC class I haplotype frequencies

The British Friesian bull sample (n= 46 animals) contained a total of sixteen MHC class I haplotypes. Table 6.5 presents the MHC class I haplotype frequencies detected in the British Friesian bulls. Haplotype frequencies ranged from 0.011 to 0.174, with A14 being the dominant haplotype in this sample. The A18v haplotype is detected at relatively high frequency in this sample (0.141). Haplotypes A31, BF_NEW7, A11, BF_NEW4, A18 and BF_NEW6 are detected at intermediate frequencies. The remaining haplotypes are distributed at frequencies from 0.043 to 0.011. The high number of haplotypes detected at low frequencies in this sample could be a reflection of the sample size being investigated. Of the sixteen haplotypes found in this sample, eight had not been previously defined at the molecular level. The availability of cDNA from British Friesian cow samples enabled the full molecular characterisation of six of these previously undefined haplotypes.

HAPLOTYPE	FREQUENCY
A14	0.174
A18v	0.141
A31	0.098
BF_NEW7	0.098
A11	0.087
BF_NEW4	0.076
A18	0.065
BF_NEW6	0.065
A10	0.043
A20	0.033
BF_NEW3	0.033
A15v	0.022
A17	0.022
New5	0.022
BF_NEW2	0.011
BF_NEW5	0.011

Table 6.5. MHC class I haplotype frequencies in the British Friesian bull population (n=46).

6.3.4.2. MHC class I allele frequencies in British Friesian bulls

Assuming that each animal has the capacity of expressing an allele from each parent for each gene, allele frequencies were calculated individually for each class I gene. However, in the British Friesian bull population, the presence of haplotypes with apparently duplicated genes i.e. A17 with two gene 2 alleles (*N*00602* and *N*00802*) and New5 with two gene 3 alleles (*N*03601* and *N*03701*) invalidates this assumption. As a duplicated gene represents another locus, there are two options for the calculation of allele frequencies. To include those animals with duplicated gene haplotypes, allele frequency calculations would have to include an additional gene to account for the possibility that every animal sampled could have gene duplication. However, as the A17 and New5 haplotypes have duplicated different genes, this would necessitate the inclusion of two additional genes, with each recording low allele frequencies and high null frequencies. Thus, the allele frequencies calculated may appear misleading and the number of class I genes greater than the six estimated by phylogenetic analysis. The remaining option is to remove those animals with duplicated gene haplotypes A17 and New5 from the final allele frequency calculations. This will give an underestimate of the frequency of the alleles on the other haplotype carried by each of these animals, but overall will produce a more reliable estimate of allele frequencies. The four animals carrying the A17 or New5 haplotypes were removed from the final calculations and allele frequencies in the British Friesian population were derived from 42 animals in total (see Table 6.6).

In addition to gene duplication, a haplotype with variable gene expression was also detected in the British Friesian bulls. The *N*04801* allele of the BF_NEW7 haplotype is consistently expressed but the variably expressed *N*01101* and *N*00402* alleles are assumed to be pseudogenes. As a result, Table 6.6 presents the frequency of *N*04801*, but omits an allele frequency for both *N*01101* and *N*00402*.

Gene 2, expressed on the majority of bovine class I haplotypes, was the most diverse locus expressing a total of ten alleles detected at frequencies ranging from 0.0119 to 0.2142. Six alleles were detected at gene 3, however the frequencies of each of these alleles were low i.e. ranged from 0.0119 to 0.0833. As a result of the low allele frequencies detected at gene 3, the frequency of having a null allele at this locus was greater than that recorded for both genes 1 and 6 despite expressing twice as many alleles as either of these loci. Gene 4 expressed a single allele, *N*02401*, which is found on both the A14 and A15v haplotypes. The *N*02401* and *N*02501* (also expressed on the A14 and A15v haplotypes) alleles were detected at the highest frequency (0.2142) in this sample. However, the frequency of having a null allele was the highest frequency recorded for each gene, ranging from 0.2857 at gene 2 to 0.7857 for gene 4.

GENE	ALLELE	ASSOCIATED HAPLOTYPE	FREQUENCY
1	*DQ121173*	BF_NEW6	0.0595
	*N*00902*	A15v	0.0238
	*N*02101*	A31	0.1071
	*N*02301*	A14	0.1905
	Null		**0.6191**
2	*N*01201*	A10	0.0357
	*N*01601*	BF_NEW5	0.0119
	*N*01801*	A11	0.0833
	*N*02201*	A31	0.1071
	*N*02501*	A14 & A15v	0.2142
	*N*02601*	A20	0.0357
	*N*04402*	BF_NEW3	0.0238
	*N*04501*	BF_NEW4	0.0833
	*N*04801*	BF_NEW7	0.1071
	*N*05501*	BF_NEW2	0.0119
	Null		**0.2857**
3	*DQ121187*	BF_NEW6	0.0595
	*N*00201*	A10	0.0357
	*N*01701*	A11	0.0833
	*N*01702*	BF_NEW2	0.0119
	*N*02701*	A20	0.0357
	*N*05101*	BF_NEW3	0.0238
	Null		**0.7501**
4	*N*02401*	A14 & A15v	0.2142
	Null		**0.7857**
6	*N*01301*	A18	0.0714
	*N*01302*	A18v	0.1548
	*N*01501*	BF_NEW4	0.0833
	Null		**0.6905**

Table 6.6. MHC class I allele frequencies recorded at each gene in the British Friesian bull population. The highest frequency detected at each gene is highlighted in bold.

6.3.4.3. Comparison of levels of MHC diversity in the British Friesian bulls with Canadian Holstein bulls

The 46 British Friesian bull samples contained 16 class I haplotypes in contrast to the 11 found in the 72 Canadian Holstein bulls sampled in 2006. Five class I haplotypes were found in both populations, but the remaining seventeen haplotypes were found exclusively in either the Canadian Holstein bulls (6) or the British Friesian bulls (11). Four of the five shared haplotypes (A10, A11, A14, and A20) were found at similar intermediate frequencies in the Canadian Holstein bulls, but in the British Friesian bulls were found distributed across the range of frequencies (see Figure 6.13). Despite the difference in sample size, five homozygote animals were detected in each of the samples.

Differences in the number of alleles expressed in each population were also evident. A minimum of 30 class I alleles, with a further two alleles with variable expression were detected in the British Friesian samples, in comparison to 19 class I alleles detected in the Canadian Holstein bulls. Using all classical full length allele sequences detected in each bull population, a comparison of diversity was carried out using the Tamura-Nei model (Tamura & Nei, 1993) and Tajima's neutrality index (Tajima, 1989) with the help of Dr J. Hammond at the Institute for Animal Health, Compton (see Table 6.7). The Tamura-Nei model examines the nucleotide distance (d) between different DNA sequences by calculating the number of transitional and transversional substitutions between each sequence, as well as the total number of sites at which nucleotide substitutions have occurred. This model showed that despite the difference in allele number there was no difference in the level of diversity between these two populations. Tajima's neutrality test, also known as Tajima's d, uses differences in sequence data to detect whether alleles in a population are under selection. Tajima's neutrality test provides a value for the difference between the average and total nucleotide diversity between allele sequences, and divides this estimate by its standard deviation. A negative Tajima's D value demonstrates that the total polymorphism is greater than the average

polymorphism estimate i.e. an excess of low frequency polymorphisms, which could be a result of purifying selection. A positive Tajima's D value indicates that the average polymorphism estimate is greater than the total polymorphism estimate, i.e. fewer than expected low frequency polymorphisms, which suggests the impact of balancing or over-dominant selection. For both the Canadian Holstein bull population and the British Frieisan bull population, a positive value was obtained for Tajima's D neutraility test implying the action of balancing or over-dominant selection upon these sequences. This corroborates the hypothesis that balancing selection was driving the pattern of MHC class I diversity in the Canadian Holstein bull population (see Chapter 4.). In terms of populations, there is no significant difference in the Tajima's d value obtained for each population and the lower value obtained for the Canadian Holstein bulls is likely to be a reflection of the smaller number of allele sequences present in this population.

POPULATION	TAMURA-NEI		TAJIMA'S NEUTRALITY TEST
	d	SE	D
CANADIAN HOLSTEIN	0.108	0.009	0.9412
BRITISH FRIESIAN	0.104	0.009	1.0932

Table 6.7. Levels of diversity in the Canadian Holstein and British Friesian bull populations assessed by the Tamura-Nei model and Tajima's neutrality test.

The full length allele sequences from each population were also subjected to phylogenetic analysis to determine whether a group of sequences were exclusive to either the Canadian Holstein or British Friesian bull populations (see Figure 6.13). A phylogenetic tree was constructed using the alpha 3, cytoplasmic and transmembrane domains for each sequence detected in the two bull populations

in line with previous work conducted by Holmes and colleagues (2003). The phylogenetic tree shows that each population contained alleles representative of five of the six putative loci i.e. alleles from all but gene 5 were detected in these populations. Furthermore, no single locus was exclusive to either population. However, the analysis did show that a tightly grouped cluster of gene 6 alleles comprising two alleles and their respective variants, indicated by the shaded box on Figure 6.14, were exclusive to the British Friesian bulls. Gene 4 was represented by a single allele, *N*02401*. Gene 3 and 4 alleles were detected in the Canadian bull samples, although none were exclusive to this population. In contrast, five gene 3 alleles were found exclusively in the British Friesian bull samples. The greatest number of alleles was detected at gene 2 and the allele sequences detected at this locus were the most divergent. Gene 1 had the highest degree of exclusivity, with only one of seven alleles detected in both bull populations.

A direct comparison of allele frequencies detected in each of the bull populations was also carried out. In order to calculate allele frequencies for each population, those animals with a duplicated gene haplotype were removed, leaving a total of 24 alleles in 42 British Friesian bulls and 17 alleles in 68 Canadian Holstein bulls. As it was previously assumed that the New5 haplotype was a single gene haplotype without duplication, the allele frequencies calculated for the Canadian Holstein bulls in this chapter differ from those presented in Chapter 3. Figure 6.15 shows the comparison of allele frequencies for each gene in the British Friesian and Canadian Holstein bull populations. The alleles found at the highest frequency in both the Canadian Holstein bulls and British Friesian bulls were the *N*02401* and *N*02501* alleles, expressed on the A14, A15 and A15v haplotypes. The high frequency of the *N*02401* and *N*02501* alleles in the British Friesian bulls is due to the high frequency of the A14 haplotype as these animals lack the A15 haplotype and the A15v haplotype is expressed at low frequency. The *N*02401* allele, present at high frequency in both populations, was the only gene 4 allele to be detected. The British Friesian bulls had a greater number of alleles at each locus than the Canadian Holstein bulls, apart from gene 1. Despite having a greater number of alleles at each locus, the British Friesian bulls had a higher frequency of having a null allele at

each locus except gene 6, compared to the Canadian Holstein bulls. This is in part due to the higher frequency of single gene haplotypes in the British Friesian bulls e.g. the A18 and A18v haplotypes have a combined frequency of over 0.2 in the British Friesian bull population. Analysis of heterozygosity in the British Friesian bull sample showed that despite a higher number of haplotypes in fewer animals than that of the Canadian Holstein sample, observed heterozygosity was estimated at 89 % whilst the expected heterozygosity calculated by Hardy-Weinberg proportions was 90 %. However, this difference in heterozygosity was not statistically significant.

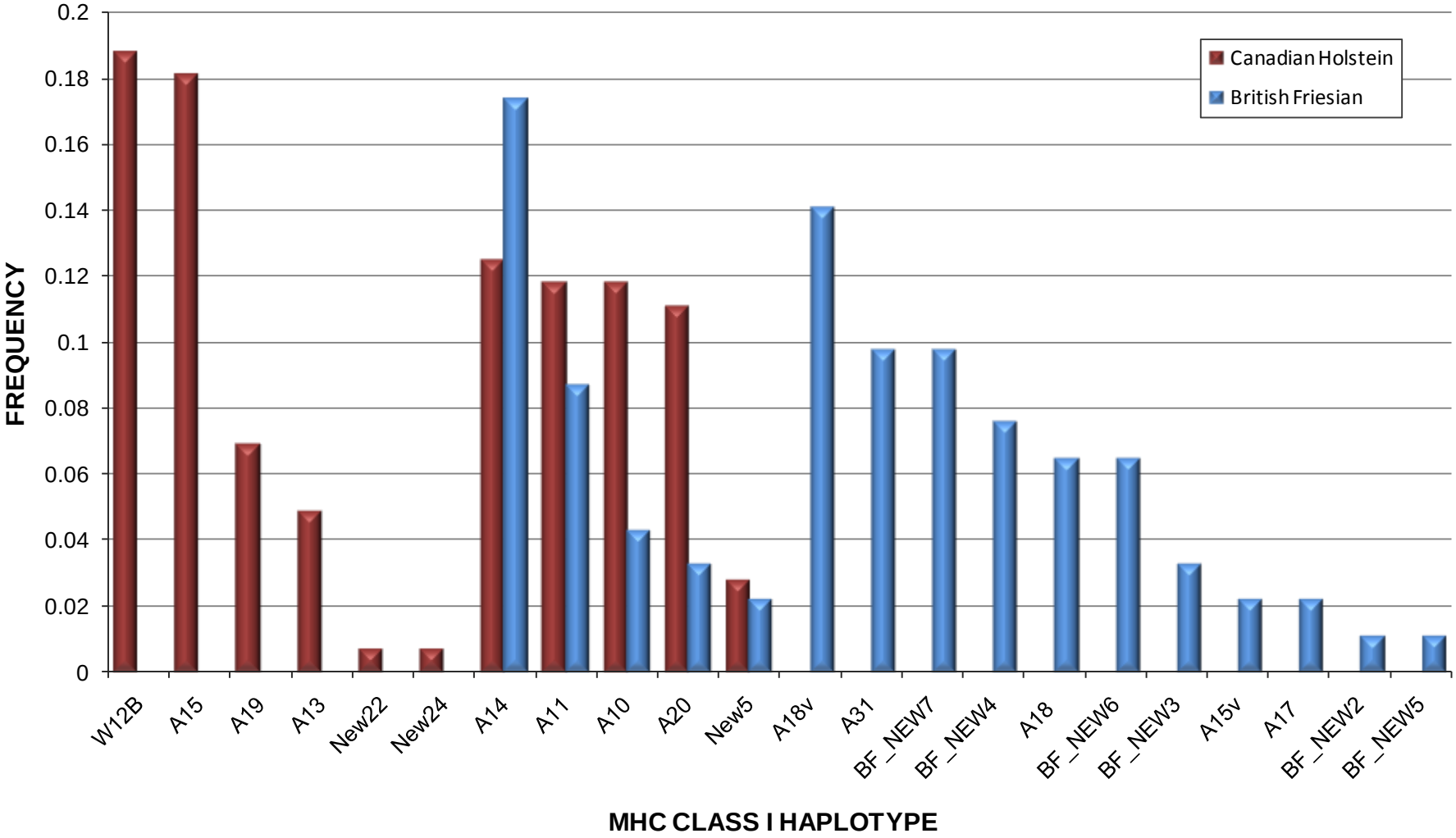


Figure 6.13. Comparison of MHC class I haplotype frequencies in the Canadian Holstein and British Friesian bulls.

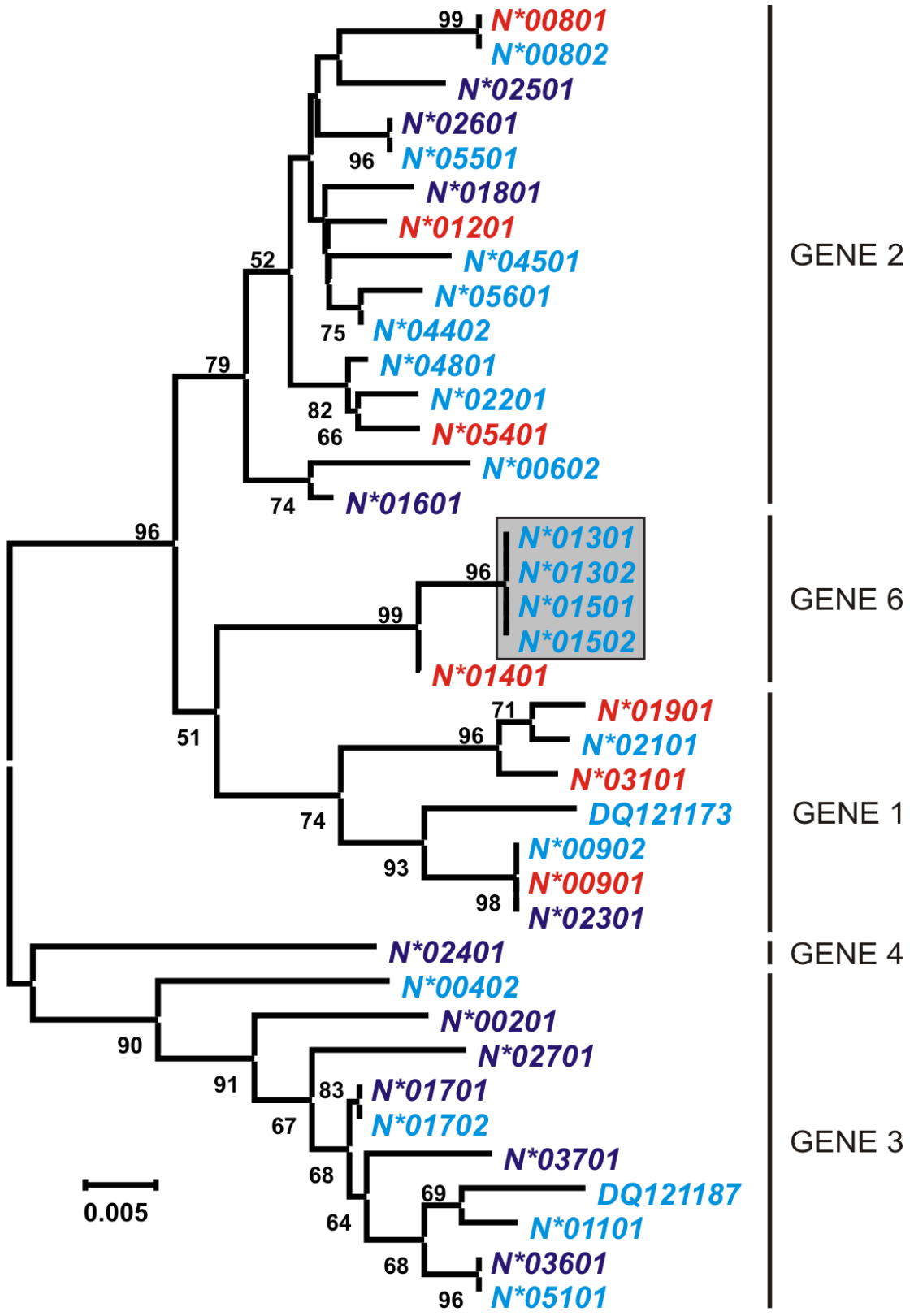


Figure 6.14. Phylogenetic analysis of the alpha 3, cytoplasmic and transmembrane domains of MHC class I allele sequences detected in the Canadian Holstein bulls 2006 sample (red) and the British Friesian bulls (blue). Alleles detected in both populations are purple. The gene 6 alleles exclusive to the British Friesians are boxed in grey. Node support values less than 50 % have been removed.

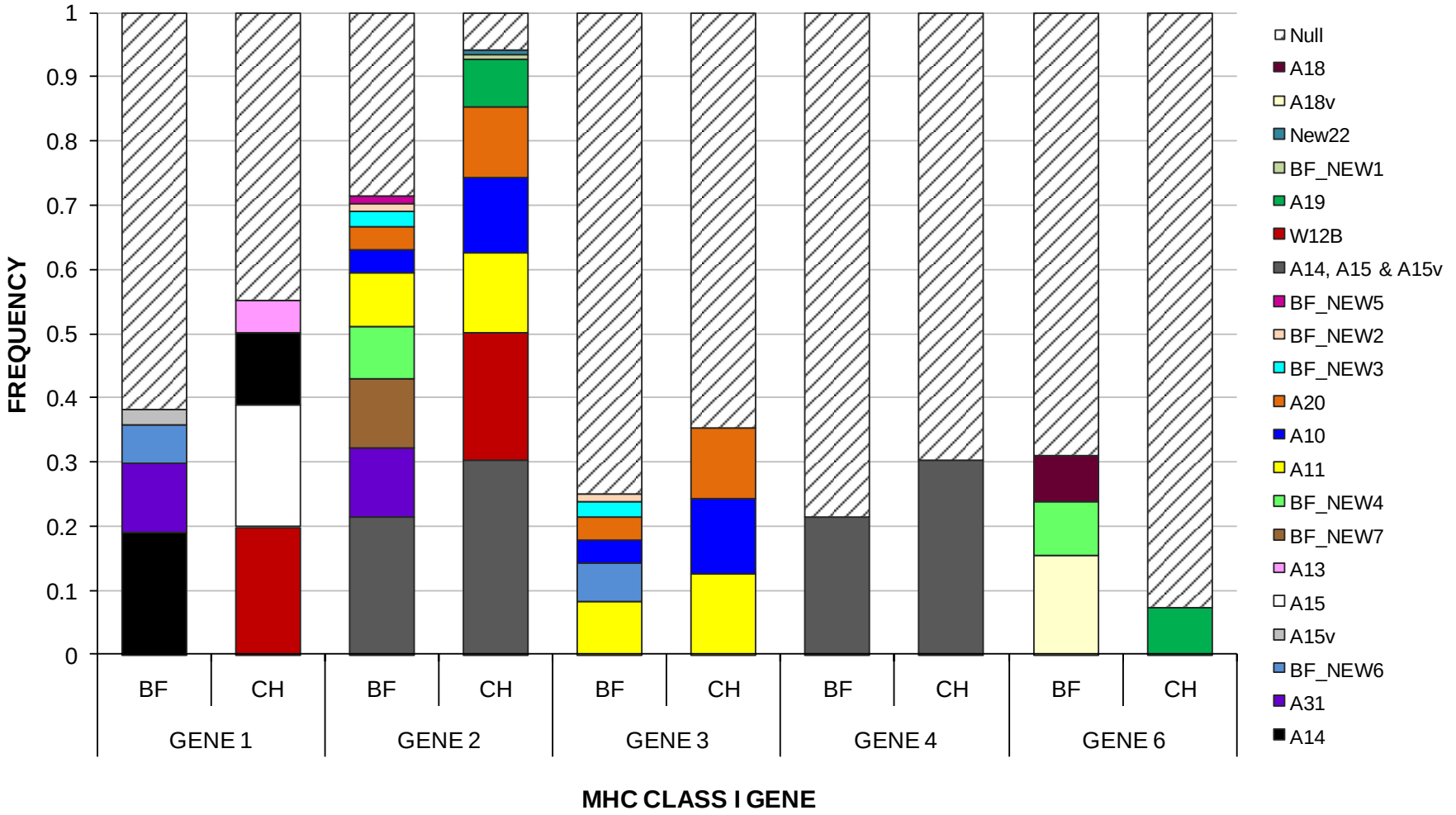


Figure 6.15. Comparison of allele frequencies for each class I gene between the British Friesian (BF) and Canadian Holstein (CH) bulls. Alleles expressed on the same haplotype are shaded in the same colour as denoted in the legend.

6.3.5 Analysis of class I diversity in British Friesian cows

6.3.5.1. MHC class I haplotype frequencies in British Friesian cows

The British Friesian cow sample (n=25) contained 11 class I haplotypes (see Table 6.8). This sample represents a limited number of bloodlines i.e. these 25 cows are descended from seven sires and includes two maternal half-sisters, and two pairs of mother and daughter. The class I haplotype frequencies ranged from 0.02 to 0.26, with the A31 haplotype the most common in this sample. All except BF_NEW1 were found in the British Friesian bulls. However, BF_NEW1 had been previously reported in the Canadian Holstein bull population in which it had been assigned the name New24.

HAPLOTYPE	FREQUENCY
A31	0.260
A14	0.200
A11	0.120
A18v	0.100
BF_NEW2	0.080
BF_NEW3	0.080
BF_NEW1/New24	0.060
A10	0.040
BF_NEW4	0.020
A18	0.020
New5	0.020

Table 6.8. MHC class I haplotype frequencies in the British Friesian cow population.

6.3.5.2. MHC class I allele frequencies in British Friesian cows

Allele frequencies in the British Friesian cow population ranged from 0.02 to 0.26 (see Table 6.9). The alleles found at the highest frequency in the British Friesian cow sample were *N*02101* and *N*02201,* both of which are expressed on the common A31 haplotype. Although this is the only sample population in which *N*02401* and *N*02501* are not the most common alleles detected, they are found at the second highest frequency of 0.200. Again, *N*02401* was the only gene 4 allele to be detected. The frequency of having a null allele was the lowest for gene 2 and the highest for gene 6, despite the presence of three alleles at this locus.

GENE	ALLELE	ASSOCIATED HAPLOTYPE	FREQUENCY
1	*N*02101*	A31	0.260
	*N*02301*	A14	0.200
	Null		**0.540**
2	*N*01201*	A10	0.040
	*N*01801*	A11	0.120
	N*02201	**A31**	**0.260**
	*N*02501*	A14	0.200
	*N*04402*	BF_NEW3	0.080
	*N*04501*	BF_NEW4	0.020
	*N*05401*	BF_NEW1	0.060
	*N*05501*	BF_NEW2	0.080
	Null		0.140
3	*N*00201*	A10	0.040
	*N*01701*	A11	0.120
	*N*01702*	BF_NEW2	0.080
	*N*05101*	BF_NEW3	0.080
	Null		**0.680**
4	*N*02401*	A14	0.200
	Null		**0.800**
6	*N*01301*	A18	0.020
	*N*01302*	A18v	0.100
	*N*01501*	BF_NEW4	0.020
	Null		**0.860**

Table 6.9. MHC class I allele frequencies recorded at each gene in the British Friesian cow population. The highest frequency detected at each gene is highlighted in bold.

6.4. Discussion

Broadening the assessment of class I diversity to encompass the British Friesian population has considerably increased the amount of MHC class I sequence and haplotype data available for the bovine MHC (see Table 6.10). Data collected in this study includes five novel class I allele sequences, confirmation of six allele sequences previously reported in Charolais cross cattle samples (Babiuk et al., 2007), the full characterisation of six new class I haplotypes from cDNA and the partial characterisation of a further three class I haplotypes from gDNA. All new full length allele sequences together with those that had been previously published were submitted to IPD and assigned an official name. Haplotypes detected in the British Friesian populations provided evidence of founder effect, gene duplication and recombination acting upon *BoLA* diversity in the Holstein Friesian breed. Comparing data from this study with that from the Canadian Holstein bull population showed each of the British Friesian populations studied contained a greater number of both alleles and haplotypes. However, a direct comparison of allelic diversity revealed no significant difference in the level of class I allele diversity between the two bull populations. Analysis of haplotype frequencies between the British Friesian and Canadian Holstein populations showed that whilst each population contained unique class I haplotypes, the A10, A11, A14 and New5 haplotypes were shared across all three sample sets. Previous examination of haplotype frequencies over time in the Canadian Holstein AI bull population (see Chapter 3) detected the A10, A11 and A14 haplotypes at comparable frequencies over twenty years of selection i.e. these haplotypes were observed at similar frequencies in both the 1986 and 2006 samples. Furthermore, statistical analysis of trait data showed that the A10, A11 and A14 haplotypes were favourably associated with selection traits (see Chapter 4) which may account for the maintenance of these haplotypes across the three sample sets, with A14 the most common of these shared haplotypes.

HAPLOTYPE	GENE 1	GENE 2	GENE 3	GENE 4	GENE 5	GENE 6
A19		*N*01601*				*N*01401*
A17		*N*00602, N*00802*				*N*01502*
A11		*N*01801*	*N*01701*			
A20		*N*02601*	*N*02701*			
A33		*N*00501*	*N*00401*			
RSCA2[a]		*RSCA2.1*[b]	*N*03801*			
A10		*N*01201*	*N*00201*			
A14	*N*02301*	*N*02501*		*N*02401*		
A15	*N*00901*	*N*02501*		*N*02401*		
W12B	*N*01901*	*N*00801*				
A12 (A30)	*N*02001*	*N*00801*				
A31	*N*02101*	*N*02201*				
A10/ KN1O4[c]			*N*00101*		*N*00301*	
A13	*N*03101*					
A18						*N*01301*
A18v						*N*01302*
BF_NEW1[a]		*N*05401*				
BF_NEW2[a]		*N*05501*	*N*01702*			
BF_NEW3[a]		*N*04402*	*N*05101*			
BF_NEW4[a]		*N*04501*				*N*01501*
BF_NEW5[a]		*N*01601*				
New5[a]		*N*05601*	*N*03601, N*03701*			
A15v[a]	*N*00902*	*N*02501*		*N*02401*		
BF_NEW6[a]	*DQ121173*[b]		*DQ121187*[b]			
BF_NEW7[a]		*N*04801*	*N*00402, N*01101*			

Table 6.10. Transcribed classical MHC class I genes in different haplotypes. [a] A serological specificity has not been determined for this haplotype. [b] A formal allele name has not been assigned to this allele. [c] This haplotype was reported and analysed by Bensaid et al. (1991). It is possible additional genes are transcribed. Alleles with variable expression are highlighted in red. Adapted from Birch et al. (2006).

The A14 haplotype is one of three *BoLA* class I haplotypes, along with A15 and A15v, on which the *N*02501* and *N*02401* alleles are expressed. Notably, the *N*02401* allele is the only gene four allele to be detected in *Bos taurus* cattle populations. The maintenance of these three haplotypes at high frequency across each of the populations sampled in this study may be a reflection of the importance of these alleles in the immune response or an association with a selected trait. Our statistical analysis showed that the A14 haplotype was significantly associated with a favourable reduction in somatic cell score (SCS) in the Canadian Holstein bulls (see Chapter 4). The A15 haplotype is absent from both the British Friesian sample sets. Statistical analysis did not show a favourable association between the A15 haplotype and any of the selection traits included in our analysis. The A15v is found only in the British Friesian bull sample and is present at low frequency, 0.022. The selective forces contributing to the maintenance of these three *N*02501* and *N*02401* expressing haplotypes at cumulative high frequencies in the population may be based upon the gene 1 allele or more specifically the alpha 1 region of the gene 1 allele, whilst the expression of three genes on a haplotype may also confer some selective advantage.

Expression of three class I genes is relatively rare in cattle class I haplotypes, with the A14, A15 and A15v haplotypes the only known examples. The A14 haplotype is the only *BoLA* class I haplotype for which a genomic map is available, demonstrating the chromosomal positions of the classical genes 1, 2 and 4 as well as the non-classical gene NC1 (Di Palma et al., 2002). Haplotype composition homology between A14 and the A15 and A15v haplotypes e.g. expression of a gene 1 allele, the classical alleles *N*02401* and *N*02501* and an NC1 allele, suggests the chromosomal locations of genes on the A15 and A15v haplotypes match those mapped on the A14 haplotype. The differing gene 1 allele (*N*00902*) of the A15v haplotype is the apparent product of intra-locus recombination of the gene 1 alleles of the A14 and A15 haplotypes, *N*00901* and *N*02301,* facilitated by homology across the regions flanking the site of recombination giving rise to both a new allele and a new class I haplotype (see Figure 6.16). In addition, evidence of recombination contributing to the formation of a new haplotype was provided by the BF_NEW2 haplotype. The high

degree of similarity of the gene 2 and gene 3 alleles of the BF_NEW2 haplotype to alleles of the A20 and A11 haplotypes respectively suggest these two haplotypes have recombined to give rise to the combination of genes expressed on the BF_NEW2 haplotype. Further evidence of both intra- and inter-locus recombination in the *BoLA* region has been provided on the basis of phylogenetic analysis (Birch et al., 2006; Holmes et al., 2003). Diversity in other mammalian MHC regions, including human and swine, have also been shown to be driven by recombination events. In human, the clustering of sequences into monophyletic groups indicates that inter-locus recombination is a relatively rare event, but intra-locus recombination has been shown to contribute to *HLA* polymorphism, particularly at the *HLA-B* locus (Hughes et al., 1993; Satta et al., 1997). In a study of cross-bred pigs, in which the penetrance of skin melanomas were linked to the *SLA* region, three recombination events located within the class I region were shown to give rise to new haplotypes (Ho et al., 2010).

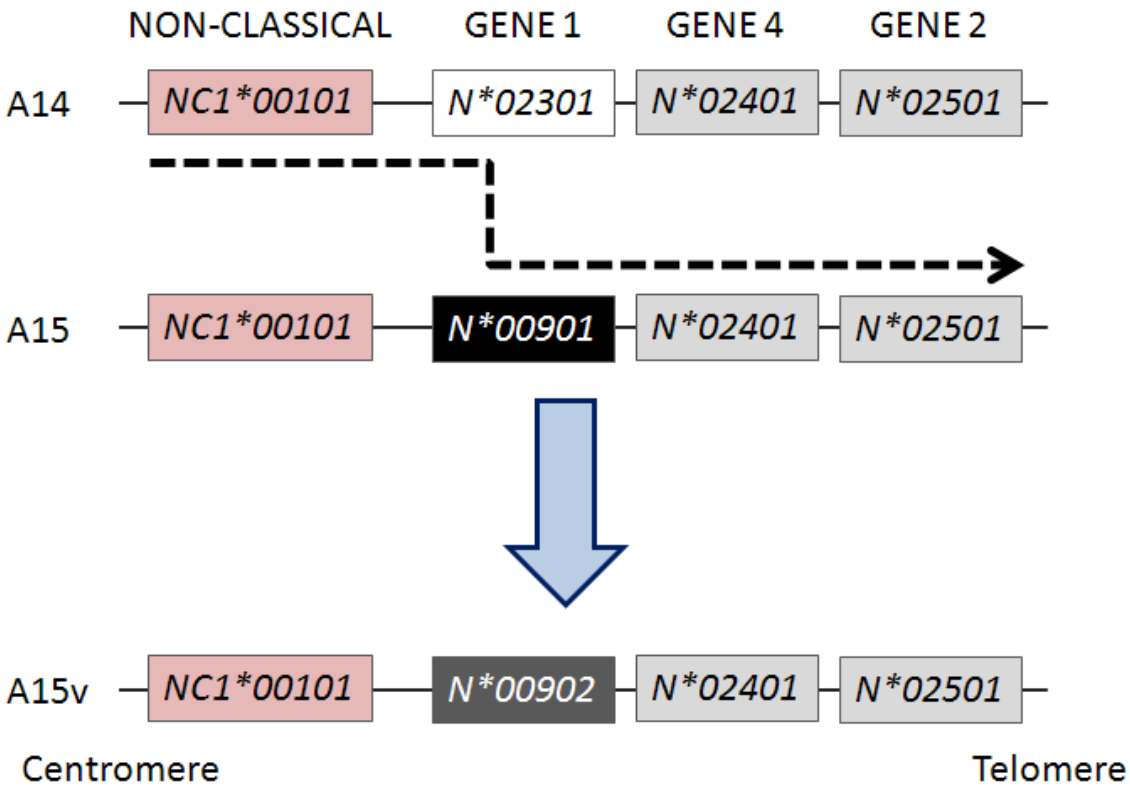


Figure 6.16. Diagrammatic representation of the recombination event which produced the *N*00902* allele and the new A15v haplotype. Gene positions are based upon the mapped A14 haplotype (Di Palma et al., 2002).

Variability in pig class I haplotypes is also generated by the number of loci expressed, with both gene deletion and duplication responsible for these differences in gene number. Haplotype analysis has shown that cattle express between one and three class I genes per haplotype, with gene deletion or silencing believed to be responsible for this variation (Ellis et al., 1999). The exact number of class I loci in cattle has not yet been elucidated due to a limited number of mapping studies and apparent heterozygosity at the MHC region of the published bovine genome sequence. Phylogenetic analyses of all available class I allele alpha 3, cytoplasmic and transmembrane domain sequences has shown clustering of alleles into six weakly defined groups which are believed to broadly correspond with loci (Ellis et al., 2005; Holmes et al., 2003). Gene-specific motifs located in the cytoplasmic and transmembrane domains have been the primary tool for the assignment of *BoLA* class I allele sequences to one of the putative genes, and will remain so until additional haplotypes are mapped. In addition, the combination of class I genes expressed on a given haplotype appear to conform to conserved gene configurations e.g. genes 2 and 3 are often found together whereas genes 1 and 6 are never expressed on the same haplotype which presents the possibility that these are not distinct genes, but in fact divergent lineages of the same locus (Birch et al., 2006; Ellis et al., 2004). These haplotype configurations and phylogenetic data have been integral in assigning new allele sequences to loci and determining haplotype structures in the British Friesian samples. Prior to this study, genes 1 and 3 had never been found on the same haplotype suggesting that these two genes are not distinct but phylogenetic evidence shows no strong relationship between the two groups (Holmes et al., 2003) and the intron sequences are sufficiently divergent to facilitate gene-specific amplification (Birch et al., 2006). The apparent expression of genes 1 and 3 on the BF_NEW6 haplotype provides further support that these are distinct genes, but this can only be confirmed through haplotype mapping i.e. the production of a high resolution sequence map of the class I region in its entirety from an animal carrying this haplotype.

In previous studies of cattle class I haplotypes, it was shown that unrelated animals of the same breed and MHC-type express the same alleles,

demonstrating that bovine class I haplotypes are generally well conserved in cattle. However, class I haplotypes detected in this study, although well conserved, showed differences in the alleles expressed according to the lineage from which the animal was derived. Evidence of founder effect in the Holstein-Friesian breed can be seen in the A15, A15v, BF_NEW5 and New5 haplotypes. The exclusivity of the A15 and A15v haplotypes to the Canadian Holstein and British Friesian bull populations respectively suggests that the effect of the founding populations may be instrumental in this observed segregation. The BF_NEW5 haplotype found in the British Friesian bulls expresses a single gene 2 allele, *N*01601*. This allele is also found in the Canadian Holstein bulls, however in this population it is expressed on the A19 haplotype in combination with the gene 6 allele *N*01401*. However, this allele (*N*01401*) has also been previously reported to be expressed in isolation in a Friesian animal (Ellis et al., 1999). The apparent segregation of these alleles in the Friesian animals and the conserved gene 2 and gene 6 allele combination found in the Holstein animals suggests that these haplotype configurations can be traced back to the founder population stage. However, it is difficult to determine whether the A19 haplotype has segregated to form two single gene haplotypes or whether the A19 haplotype represents the recombination of two single gene haplotypes to form one haplotype expressing two genes.

Similarly, this investigation revealed differences in gene expression on the New5 haplotype depending on the cattle lineage. This haplotype was initially detected in the Canadian Holstein bull sample and was believed to express a single gene 3 allele. However, the availability of cDNA samples from the British Friesian cows showed the presence of another gene 3 allele (*N*03701*) as a result of duplication, as well as a gene 2 allele *N*05601*. Further characterisation of this haplotype in Holstein and Friesian animals showed that the duplicated gene 3 was present in both populations, but that the gene 2 allele (*N*05601*) is absent. It is hypothesised that the loss of the gene 2 allele occurred prior to the segregation of the Holstein and Friesian lineages and that the *N*05601* was absent from the founding population from which the Holstein breed was derived, whereas the duplicated gene 3 alleles were conserved in each population.

Gene duplication is commonly found as a means of driving diversity at the MHC in mammalian species e.g. duplication of the *SLA-1* genes is observed on three swine class I haplotypes with expression of alleles at phylogenetically indistinguishable loci termed '*SLA-1a*' and '*SLA-1b*' (Renard et al., 2001; Smith et al., 2005). Clear evidence of gene duplication is relatively scarce in the *BoLA* class I region, although the separation of class I alleles into two subgroups, based upon cytoplasmic and transmembrane domain sequences, suggests duplication of an ancestral sequence was fundamental in *BoLA* class I evolution (Birch et al., 2006; Holmes et al., 2003). The phylogenetic tree of the alpha 3, transmembrane and cytoplasmic domains of the classical allele sequences detected in the two bull populations in this study corroborates the previous analysis (see Figure 6.14). The characterisation of two allele sequences on the New5 haplotype, apparently assigned to the gene 3 locus, is indicative of gene duplication which, prior to this finding, had only been reported on one other bovine class I haplotype, A17. However, the duplicated gene 2 allele *N*00602* on the A17 haplotype has a deletion at the 3' end causing the removal of the stop codon and extending translation into the 3' UTR, thus casting doubt as to whether this allele is functional (Birch et al., 2006). The duplicated alleles expressed on the New5 haplotype (*N*03601* and *N*03701*) both have an appropriate stop codon in the 3' prime region, indicating that each of these alleles is functional. The detection of gene duplication on the New5 and A17 haplotypes provides evidence that gene duplication contributes to the generation of diversity at the bovine MHC.

Phylogenetic analysis showed that although there were differences in the haplotypes and alleles detected in the British Friesian bulls and Canadian Holstein bulls, no single locus was exclusive to either population. However, the analysis did show that a tightly grouped cluster of gene 6 alleles were exclusive to the British Friesian bulls. The four British Friesian haplotypes expressing a gene 6 allele were A18, A18v, A17 and BF_NEW4. Both the A18 and A18v haplotypes are single gene 6 haplotypes and the alleles differ by a single amino acid change in the alpha 2 region. The BF_NEW4 and A17 haplotypes each express similar gene 6 alleles (*N*01501* and its variant *N*01502*) in combination with a single gene 2 allele and two gene 2 alleles arising through apparent gene

duplication respectively. The four gene 6 alleles *N*01301*, *N*01302*, *N*01501* and *N*01502* dictate that the A18, A18v, A17 and BF_NEW4 haplotypes would all fall under the broad 'A6' serological specificity, which has been previously reported to be immunodominant in *Theileria parva* infections (Morrison et al., 1996). The immunodominance of the 'A6' serotype was also demonstrated by a study investigating the $CD8^+$ T cell response to foot and mouth virus in MHC-typed cattle, with high $CD8^+$ T cell responses shown to be mainly restricted by the *N*01301* allele of the A18 haplotype in both homozygous and heterozygous animals (Guzman et al., 2008).

A crystal structure of the *N*01301* allele has recently been elucidated and shows that a hydrophobic ridge causes bound peptide to bulge out of the peptide binding cleft, potentially giving rise to a unique MHC-TCR interaction contributing to immunodominant peptide presentation (Macdonald et al., 2010). Sequence analysis predicts that the *N*01302*, *N*01501* and *N*01502* alleles also have the hydrophobic ridge within their peptide binding groove and thus the tendency to elicit immunodominant $CD8^+$ T cell responses. These four 'A6' haplotypes are found at a cumulative frequency of 0.304 in the British Friesian bull population suggesting that they may be more immunologically robust than their Canadian Holstein counterparts. However, it could be argued that the high frequency of the A18 and A18v haplotypes in the British Friesian bulls represents a compromise in terms of diversity as they each express a single gene 6 allele in contrast to the majority of class I haplotypes that express at least two genes. However, it is important to note that whilst data from the foot and mouth virus study and the crystal structure provide evidence of immunodominance associated with the 'A6' alleles, this has yet to be correlated with the ability to provide a protective response to pathogenic challenge.

It was initially hypothesised that the levels and pattern of MHC class I diversity found in the British Friesian population would be similar to that seen in the Canadian Holstein bulls (see Chapter 3). This assumption arose because despite segregation into types, all pedigree Holstein-Friesian cattle are derived from founding populations in the Netherlands and are registered on a single herdbook.

Although this study showed that there were no differences in the level of allelic diversity between the two populations, differences in the pattern of MHC class I diversity at both the haplotype and allele level between the two types of cattle under the classification of a single breed were evident. These differences in the pattern of MHC class I diversity are likely to reflect differences at the founder population stage as these two populations have experienced similar selective pressures. It is important to note that whilst the Holstein and Friesian are considered as a single breed and registered on one herdbook, in terms of MHC diversity these two breed 'types' may be considered too disparate for conserved gene expression under the definition of a particular haplotype name. However, it must also be considered that sequence analysis of both the British Friesian and Canadian Holstein animals demonstrated that for animals carrying the A10, A11, A14 and A20 haplotypes i.e. four of the five 'shared' haplotypes, the alleles expressed were identical irrespective of the animal's lineage. The differences in class I alleles detected in the two populations may well translate to functional differences in terms of pathogen challenge e.g. alleles expressed exclusively in one population may confer the ability to better present antigens from a new pathogen than the other population in which these alleles are not present. However, evidence of such functional differences in immune response according to breed strain/ancestry has not yet been reported.

General discussion and further work

7.1. Aims of this study

In natural populations, balancing selection maintains MHC class I diversity. However, in economically important livestock such as dairy cattle, there are concerns that intensive selection for both production and fitness traits may override natural selection and lead to a loss of MHC diversity. Intense selection has been coupled with the widespread introduction of reproductive technologies, resulting in a small effective population size in the common Holstein-Friesian breed of dairy cattle. It had been hypothesised that these focussed dairy breeding practices may lead to a reduction in MHC diversity and leave cattle populations susceptible to new disease pathogens.

The aim of this study was to assess current levels of MHC class I diversity in the UK dairy cattle population. MHC-typing had been previously carried out using serological methods, but the availability of molecular-based methods has facilitated a more detailed analysis of class I diversity in the current cattle populations. This study focussed primarily on Holstein-Friesian AI bull populations from which the wider UK dairy herd is derived. Data from these populations were used to investigate whether intensive selection was leading to a reduction in class I diversity over time, how selection for specific production and fitness traits was impacting upon MHC diversity, and whether similar selection pressures had given rise to similar patterns and levels of class I diversity in AI bull populations of different lineages. These data were also used in the design and development of a SNP-based assay that can detect common class I haplotypes, enabling breeders to formulate future breeding strategies that maintain and promote MHC diversity in the UK herd. Due to the linkage of class I genes with class II genes, it is assumed that selecting for increased MHC

class I diversity will simultaneously promote diversity at class II loci. The aim would be to increase overall MHC diversity, and not just class I diversity.

7.2. Summary of findings and achievements

Canadian and North American bloodlines represent a globally dominant force in dairy breeding and in the UK accounted for over 40 % of 250,000 pedigree dairy registrations in 2008 alone (Alison Maddrell, personal communication, Holstein UK). In a sample of 72 premium AI Canadian Holstein bulls, a total of eleven class I haplotypes were detected, eight of which had been previously characterised at the molecular level (Birch et al., 2006). Of the eleven class I haplotypes in this population, the W12B and A15 haplotypes were the most common, detected at frequencies of 0.188 and 0.181 respectively. Also common in this population was the A14 haplotype, which expresses the *N*02401* and *N*02501* alleles also present on the A15 haplotype. The level of MHC class I allelic diversity was lower than would be expected for an outbred population of comparable size. However, statistical analysis showed that the level of heterozygosity in this population was significantly higher than expected by Hardy-Weinberg calculations. Furthermore, a retrospective comparison with serological data collected in the 1980s demonstrated that there had not been a significant reduction in the overall levels of MHC class I diversity over a twenty year period of selection.

Simulations of gene flow over this twenty year period demonstrated that the changes in frequency of common haplotypes could not be a result of genetic drift and inbreeding alone, implicating trait-based selection as an influential force in shaping MHC class I diversity in this population. Statistical analysis of phenotypic trait data from 220,000 offspring sired by the Canadian Holstein bulls and the MHC class I haplotype frequency data from this bull sample enabled the dissection of these selective forces. The trait data analysis showed correlations between MHC haplotype and specific selection traits. Examples of these

associations include the favourable positive association between milk yield and the A11 haplotype, a favourable decrease in somatic cell score associated with the A10, A14 and A19 haplotypes, as well as the negative association of the common A15 haplotype with fertility. The underlying form of selection responsible for these associations appears to be heterozygote advantage with the rare haplotypes favoured over the common haplotypes. Although it is unlikely that the MHC alleles/genes involved are the cause of the trait associations, they provide a useful genetic marker. Being able to type a sample for these common haplotypes would enable breeders to make more efficient breeding decisions that mimic the selective forces already in action by favouring the rare haplotypes.

The haplotype and allele frequency data from the Canadian Holstein bull sample was used to design and develop a SNP-based assay for the rapid detection of common MHC class I haplotypes. The customised SNP assay is able to detect alleles of the common A14, A15 and W12B haplotypes and determine hetero- or homozygosity in a gDNA sample. This assay was optimised and applied to samples from the wider UK Holstein-Friesian herd, in which similar frequencies of the common A14, A15 and W12B haplotypes were detected.

In order to more thoroughly assess which other class I haplotypes are common in the UK herd, investigations of class I diversity were expanded to incorporate the British Friesian bull population. This represents a different lineage contributing to the genetic diversity of the UK herd that has undergone similar selection pressures. In a sample of 46 British Friesian bulls a total of sixteen class I haplotypes were detected, of which only eight had been previously characterised at the molecular level (Birch et al., 2006). As well as a greater number of haplotypes, the British Friesian bulls also contained a greater number of class I alleles in comparison to the Canadian Holstein bulls. Despite higher levels of haplotype and allelic diversity, the level of heterozygosity in this population was not significantly different than that expected by Hardy-Weinberg calculations. Most importantly, this population contained a large number of

novel alleles and haplotype combinations which has contributed to our knowledge of the mechanisms driving diversity in the *BoLA* class I region.

7.3. Further lines of enquiry arising from this work

7.3.1. Mechanisms driving *BoLA* class I diversity

Prior to this study, there were approximately 60 class I allele sequences submitted to the IPD database and 16 class I haplotypes had been characterised at the molecular level. Previous phylogenetic analysis of 29 available bovine class I sequences demonstrated the presence of six weakly defined groups, based on gene-specific motifs located in the alpha 3, cytoplasmic and transmembrane domains, which were thought to correspond to loci (Holmes et al., 2003). Of these putative loci, analysis has shown that between one and three class I genes are expressed on a class I haplotype. This investigation into class I diversity in AI bull populations, particularly the British Friesian lineage of the Holstein-Friesian breed, and database mining of recently reported sequences in the literature, has provided a growing repository of bovine class I sequences and class I haplotype definitions. There are now over 80 full length classical and non-classical bovine class I allele sequences submitted to IPD and 22 class I haplotypes have been characterised at the molecular level from cDNA and a further three haplotypes from gDNA. This has contributed to our knowledge of the mechanisms generating diversity in the bovine class I region.

7.3.1.1. Gene duplication

Using the new sequence data available, which equates to approximately 50 additional sequences, it is possible to re-analyse the class I allele sequences phylogenetically to further our knowledge of cattle class I evolution. Due to

selection on the alpha 1 and 2 domains, and frequent recombination throughout intron three, it is more informative to analyse the 3' end of class I allele sequences which are more conserved. The phylogenetic tree of the alpha 3, transmembrane and cytoplasmic domains of 80 classical and non-classical class I allele sequences (see Figure 7.1) indicates that non-classical genes 2, 3 and 4 represent a divergent group, confirming the findings of previous analyses (Ellis et al., 1999; Holmes et al., 2003). It is also apparent that the non-classical gene 1 locus (NC1) and all the classical sequences share a common ancestor. Further phylogenetic analysis of the 3' end of the classical class I genes in isolation shows that, by the process of duplication, the classical genes have diverged into two supergroups, with sequences putatively assigned as genes 1, 2 and 6 comprising one group, and the second formed by the putative genes 3, 4 and 5 (Figure 7.2).

Evidence of gene duplication was also found within individual class I haplotypes e.g. A17 and New5, which each express two alleles from a single gene. The A17 haplotype expresses a single gene 6 allele (*N*01502*) in combination with two gene 2 alleles (*N*00602* and *N*00802*). Of the duplicated gene 2 alleles on the A17 haplotype, the *N*00602* allele has a deletion at the 3' end causing the removal of the stop codon and extending translation into the 3' UTR, thus casting doubt as to whether this allele is functional (Birch et al., 2006). The duplicated alleles expressed on the New5 haplotype (*N*03601* and *N*03701*) both have an appropriate stop codon in the 3' prime region, suggesting that each of these alleles is functional. The assumption that New5 expresses duplicated genes is based only on gene-specific motifs in the cytoplasmic and transmembrane domains of these allele sequences. Further investigation of this haplotype would have to be carried out to confirm that the *N*03601* and *N*03701* alleles represent duplication of gene 3 e.g. analysis of intron sequences. If these alleles are the product of duplication, it would be expected that the intron sequences of these two alleles would show limited sequence divergence. However, intron sequence data from these two alleles is lacking. The gene 3-specific primers, which anneal to locus-specific motifs in the intron, are able to amplify the *N*03601* allele but not the *N*03701* allele, indicating that these sequences may not represent a gene duplication e.g. the *N*03701* appears to lack conserved gene 3-specific

motifs in the intron to which these primers bind. Further analysis of those haplotypes with apparent gene duplications is required to better understand the mechanisms responsible.

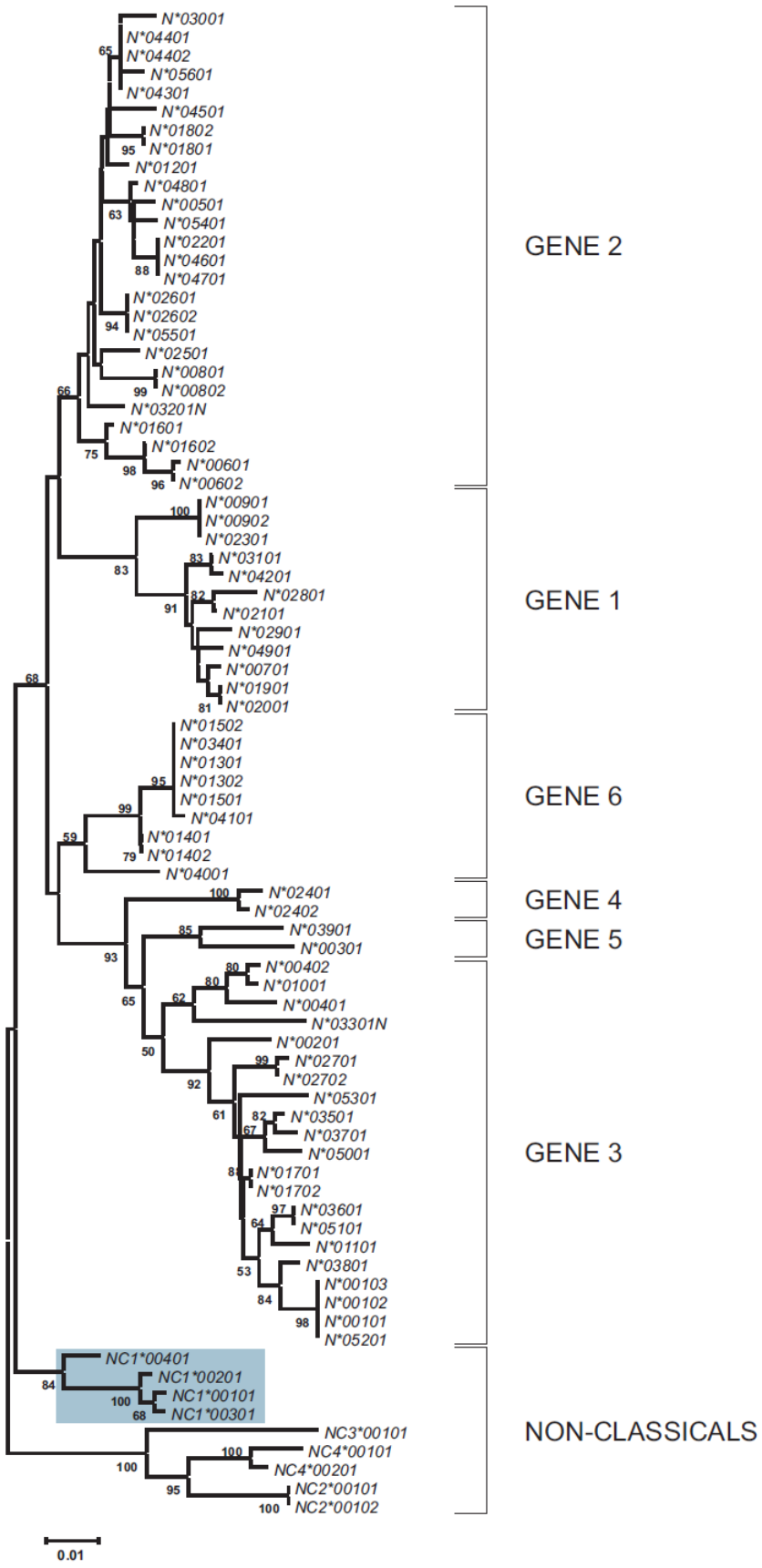


Figure 7.1. Phylogenetic analysis of the alpha 3, transmembrane and cytoplasmic domains of all 80 bovine MHC class I sequences currently submitted to the IPD database. Non-classical gene 1 (NC1) is highlighted in blue. Node support values less than 50 % have been removed.

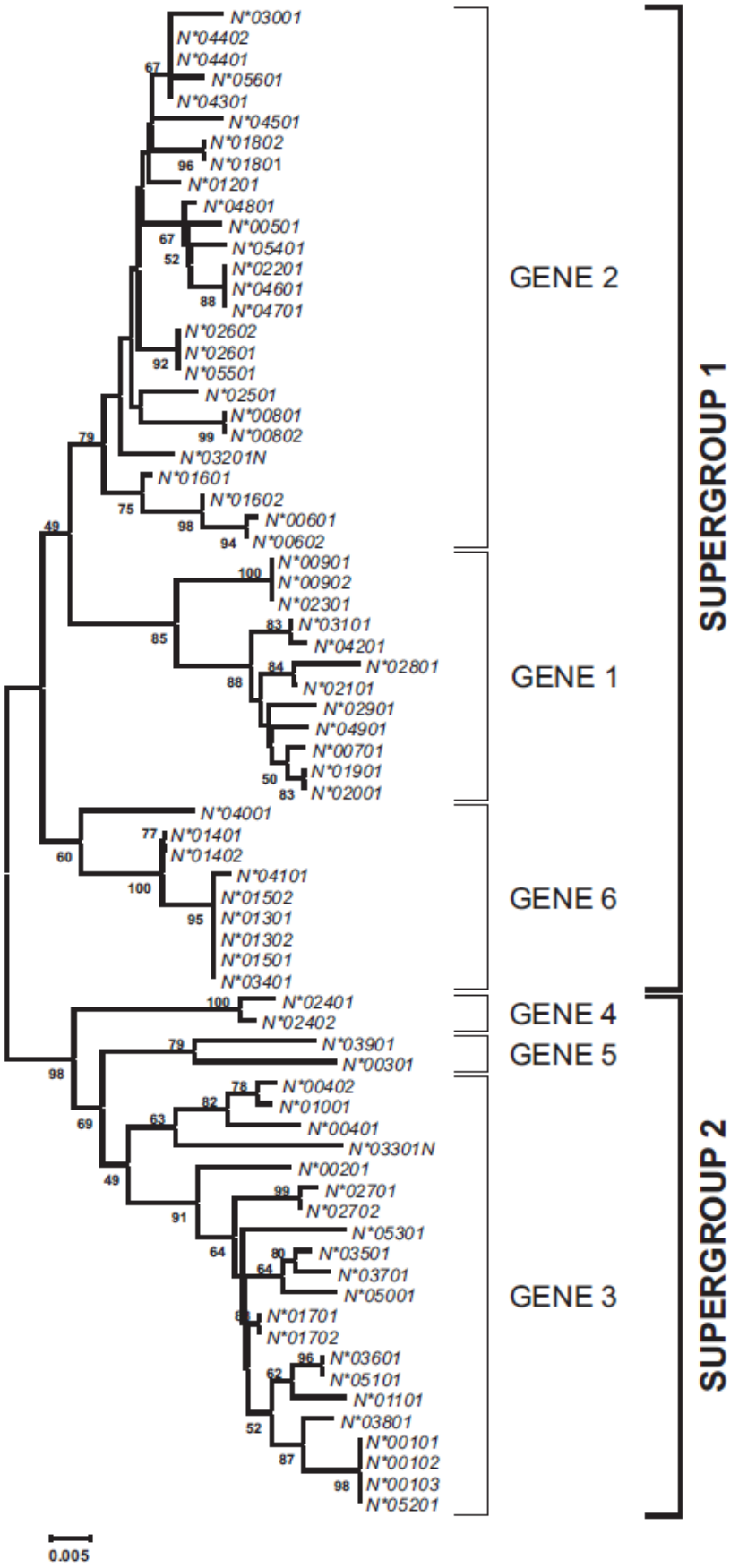


Figure 7.2. Phylogenetic analysis of the alpha 3, transmembrane and cytoplasmic domains of all 71 bovine MHC classical class I sequences currently submitted to the IPD database. Node support values less than 49 % have been removed.

7.3.1.2. Recombination

Besides duplication, the new allele sequences found during this study also provided evidence of both inter-locus and intra-locus recombination in the *BoLA* class I region. The high degree of similarity of the gene 2 and gene 3 alleles of the BF_NEW2 haplotype to alleles of the A20 and A11 haplotypes respectively, suggest inter-locus recombination between these two haplotypes is responsible for the combination of genes expressed on BF_NEW2 (see Chapter 6, Figure 6.2.).

The most clear cut example of intra-locus recombination in the *BoLA* class I region was provided by the *N*00902* allele on the A15v haplotype detected in the British Friesian bulls. The *N*00902* allele appears to have arisen through the recombination of the alpha 1 domain from the *N*02301* allele and the alpha 2 domain from the *N*00901* allele, facilitated by homology of these two alleles across leader, alpha 3, transmembrane and cytoplasmic domains providing a framework for recombination. This kind of recombination event, in which structural repeats and/or sequence homology provide a framework for recombination, is common in both the human MHC and *KIR* region (Hughes et al. 1993; Satta et al., 1997; Traherne et al., 2009).

Whilst it seems that the *N*00902* allele evolved through the recombination of gene 1 alleles on the A14 and A15 haplotypes, we have yet to identify all existing cattle class I alleles and thus there may be an alternate explanation for its evolution. It is also possible that an allele formed by recombination of the alpha 1 domain from *N*00901* and the alpha 2 domain from *N*02301* is also likely to be present within the population. This 'reciprocal' recombination event would give rise to a variant of the *N*02301* allele expressed on what would be considered an A14v haplotype (see Figure 7.3). However, the transverse recombinant of the *N*00901* and *N*02301* alleles has yet to be found in the

population and it is likely that this haplotype would also be found at relatively low frequency within the population.

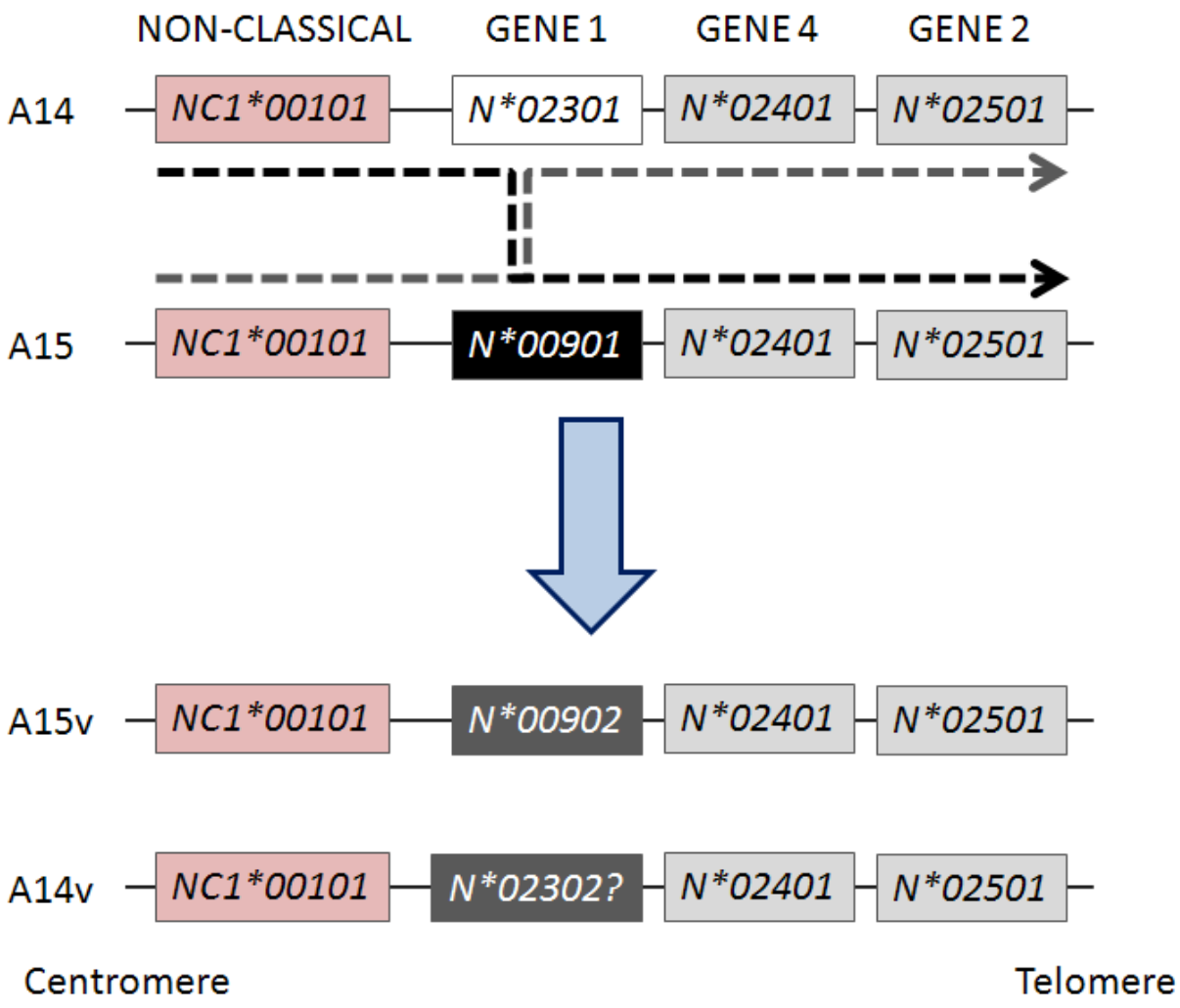


Figure 7.3. Diagrammatic representation of the proposed transverse recombination event between the A14 and A15 haplotypes, giving rise to the A15v haplotype and the speculated A14v haplotype.

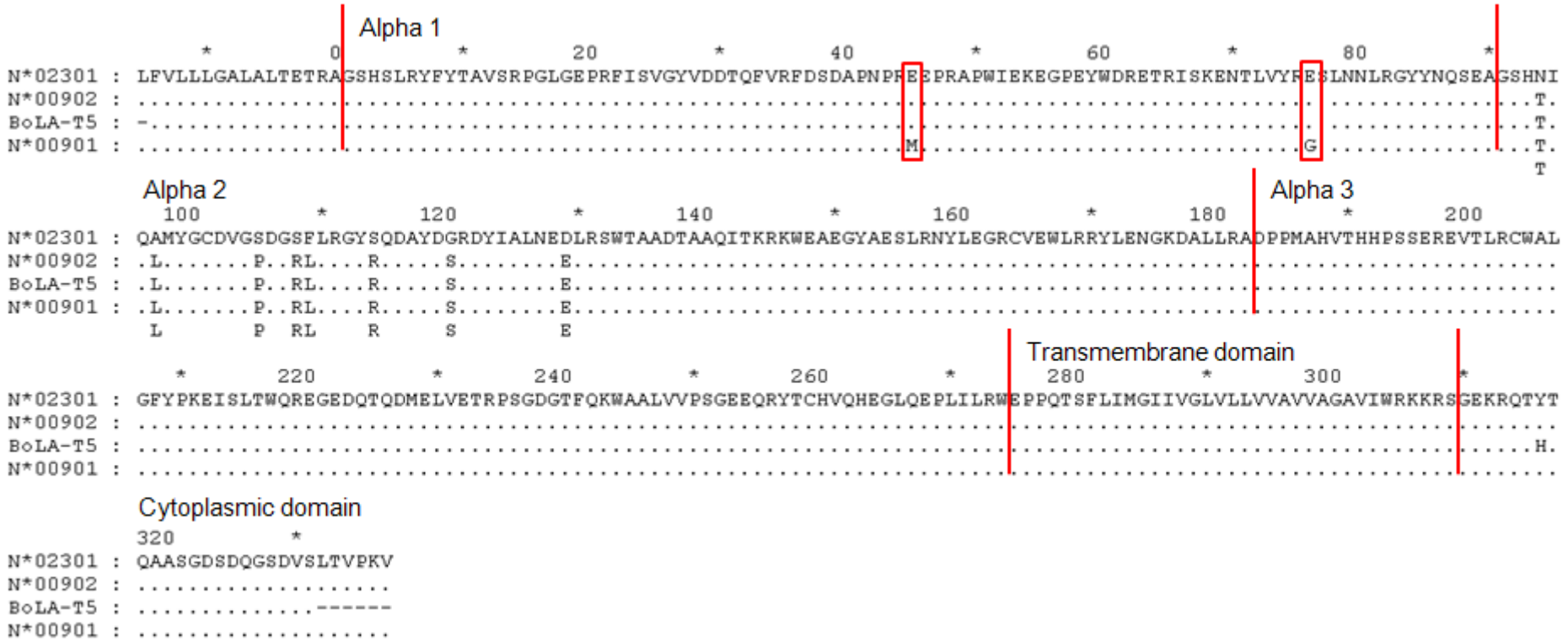


Figure 7.4. Alignment of the *N*02301*, *N*00902*, *N*00901* and BoLA-T5 sequences. Dots indicate identity, dashes indicate gaps/lack of sequence data compared to reference.

Further investigation of the *N*00902* allele sequence has shown that it is identical, apart from one residue in the cytoplasmic domain, to a sequence called BoLA-T5 (Genbank Accession number EU189195.1; see Figure 7.4.). The amino acid change occurs at a well conserved residue and so may be an artefact of PCR or a sequencing error rather than a 'real' change and thus should be considered identical to the *N*00902* allele. Investigations of antigen presentation during *Theileria parva* infections has shown that the *N*00902* allele presents the same Tp5 epitope as the *N*02301* allele on the A14 haplotype (Prof Ivan Morrison, personal communication).

This similarity in antigen presentation is likely to be a result of homology between the *N*02301* and *N*00902* alleles in the alpha 1 domain. Within the alpha 1 domain, the amino acid residues at positions 45 and 76 in the *N*02301* and *N*00902* alleles are both negatively charged (Glutamic acid) but are neutral (Methionine and Glycine) in the *N*00901* allele. However, it seems unusual that the alpha 1 sequence can dictate specificity in this manner given the fact that both the alpha 1 and alpha 2 domains contribute to the specificity of the peptide binding cleft, and that the *N*00902* allele shows greater similarity to *N*00901* than *N*02301*. Looking at the sequence differences in the alpha 2 domain between *N*02301* and both *N*00902* and *N*00901*, the majority of amino acid changes are not that influential i.e. the amino acids substituted do not differ much in terms of their properties. The only major changes in terms of amino acid properties in alpha 2 are seen at positions 105, 108, 114 and 120 i.e. charge changes, and of these only position 114 is categorised as an antigen recognition site. To further this investigation, it would be informative to have functional data for antigen presentation by the gene 1 allele of the A15 haplotype, *N*00901*, as this would provide an interesting comparison and enable a greater understanding of the polymorphic residues directly involved in antigen presentation from this parasite and other pathogens.

7.3.2. Differences between Holstein-Friesian lineages

Examination of the two different Holstein-Friesian lineages showed differences in the patterns of MHC class I haplotype and allelic diversity. This is as expected, as a similar situation arises when examining class I allelic diversity in different ethnicities within human populations (Mori et al., 1997). The most striking difference between these two lineages was observed at the gene 6 locus, with the British Friesian population exhibiting both greater diversity and a greater number of animals expressing an allele at this locus. The high frequency of gene 6 alleles in the British Friesian population was somewhat unexpected given that only a single allele from this locus is expressed in the Canadian Holsteins, at low frequency. Previous studies of haplotypes expressing a gene 6 allele, which fall under the broad 'A6' serological specificity, have been shown to be 'immunodominant' in *Theileria parva* infections. This immunodominance is believed to be attributed to the unique peptide binding conformation of gene 6 alleles, as revealed by the generation of a crystal structure from the gene 6 allele, *N*01301*. These data together with the prevalence of these alleles in the British Friesian population instigates further interest in the gene 6 locus.

Thus far, there are no mapping data showing the position of putative gene 6 in relation to those loci already mapped. For example on the A14 haplotype, 3 classical class I genes (1, 2 and 4) have been mapped, together with four additional class I pseudogenes, but none appear related to gene 6. Of the mapped class I alleles, the closest related sequence to gene 6 is the *N*04001* allele, the alpha 1 and 2 domains of which are related to gene 3, but the 3' end of this allele is similar to that of gene 6 alleles (see section 7.3.1.2.). It is thus difficult to ascertain why gene 6 alleles are now largely lacking in the Canadian Holstein bull population, and why this gene is often expressed in isolation on single gene class I haplotypes.

The combined frequency of single gene haplotypes in the British Friesian lineage, both bull and cow samples, was 0.199. This is markedly higher than the

frequency of single gene haplotypes detected in the Canadian Holstein bull population (0.056). The high combined frequency of single gene haplotypes in the British Friesian is equivalent to the dominance of the W12B or A15 haplotypes, which express two and three class I genes respectively, in the Canadian Holstein bulls. Although it could be speculated that the expression of three classical genes in contrast to a single gene on a class I haplotype will give rise to a more immunologically robust individual, there is no evidence to confirm that this is indeed true.

Examination of antigen presentation by the classical class I genes of the A14 haplotype transfected into COS-7 cells during studies of immunity to *Theileria parva* has shown that *N*02301* is the dominant functional allele, whilst despite being expressed on the cell surface, the *N*02501* allele appears unable to present epitopes from this particular parasite (Prof Ivan Morrison, personal communication). In addition, analysis of levels of MHC class I cell surface expression has been shown to vary both between the genes of class I haplotypes, and between individual haplotypes. For example, expression of the *N*02201* allele on the A31 haplotype is over six times greater than *N*02101*. Also, total MHC class I cell surface expression on PBMCs was one and a half times greater in cells from an A18 (single gene haplotype) homozygote compared with cells from an A14 (three class I genes) homozygote (Smith, 2000).

These differences in both function and expression may indicate that there is dominant expression of one gene and as such the total number of genes expressed on a haplotype may be largely immaterial. This would mirror the situation seen in chicken, in which one class I gene is dominantly expressed per conserved class I haplotype (Kaufman, 1999). With one gene being dominantly expressed on the cell surface, the other genes encoded on multiple gene haplotypes in cattle may serve as ‘reserve’ genes, which only become transcribed when under stressful conditions, in a similar manner to the expression of the *MIC* genes. This would also explain the promiscuity in peptide binding and apparent ‘immunodominance’ of gene 6 alleles which are often found as single gene haplotypes i.e. high levels of expression and diverse antigen

binding properties compensate for the ability to only express a single gene. In essence, a great deal more information regarding the functional differences between class I haplotypes and the different gene configurations expressed on class I haplotypes is required, both in terms of pathogen challenge and interactions with other immune genes i.e. NK receptors. Any differences observed must be taken into account in breeding strategies to ensure that diversity is maintained without any functional detriment to the immune response.

7.3.3. The number of *BoLA* class I loci

The number of MHC class I loci in cattle still remains a fundamental issue in *BoLA* research. A limited number of mapping studies and the derivation of the bovine genome from two consanguineous animals has not provided any concrete data as to the exact number of loci. Using phylogenetic analysis of all the new class I sequence data available, it is possible to make more informed speculations as to the number of *BoLA* class I loci (see Figure 7.1 and 7.2).

The phylogenetic tree of the alpha 3, transmembrane and cytoplasmic domains of all 80 class I sequences available shows the clustering of classical class I allele sequences into six groups, corroborating the findings of Holmes and colleagues (2003). Each of these groups is thought to broadly correspond to a class I locus. However, there appears to be greater overlap between these clusters as a result of more sequence data. If the non-classical sequences are removed from the phylogenetic analysis, a segregation of sequences into two supergroups is apparent (Figure 7.2). Genes 1, 2 and 6 form supergroup 1, and the second supergroup is comprised of genes 3, 4 and 5. The clustering of allele sequences into two supergroups was also observed when analysing the sequences derived from the two AI bull populations (see Chapter 6, Figure 6.14), although gene 5 allele sequences were absent from each of these populations.

The tight clustering of allele sequences from genes 3, 4 and 5 in supergroup 2 may suggest that these sequences actually represent a single locus. Of these three genes, only genes 3 and 5 have been found on the same haplotype. However, the mapping of the gene 3 allele *N*00101* and the gene 5 allele *N*00301* on the A10/KN104 haplotype demonstrate that genes 3 and 5 are indeed physically separate and located within 210 kb of one another (Bensaid et al., 1991). In addition, the transmembrane domains of gene 5 alleles are 36 amino acids in length whereas all other class I alleles are either 35 or 37 amino acids in length (Ellis et al., 1992). As such, it can be assumed that genes 3 and 5 represent separate loci, as previously proposed by Ellis and colleagues (1999). Evidence that gene 4 also represents an independent locus is provided by the weaker clustering of allele sequences from gene 4 with the other two putative genes within supergroup 2. As such, despite tighter clustering within supergroup 2, it appears that each of the three groups represents independent class I loci.

The previous phylogenetic analysis by Holmes and colleagues (2003) also suggested that genes 1 and 6 represent two separate allele lineages of the same locus. This observation appeared to be confirmed when analysing the sequences present in the two AI bull populations, with genes 1 and 6 clustering tightly together (see Chapter 6, Figure 6.14). However, the inclusion of more sequence data in our analysis has shown that genes 1 and 6 do not cluster together but represent sister clades, suggesting that they are independent loci. Indeed, these two genes have yet to be found expressed together on the same haplotype. In addition, there are distinct differences in the cytoplasmic domains of genes 1 and 6, and sufficient differences within the intron to enable discrete amplification of these loci (Birch et al., 2006). Additional phylogenetic analysis of the alpha 1 and 2 domains of all the classical class I sequences (data not shown) shows, once again, genes 1 and 6 clustering together. These data suggest that although genes 1 and 6 represent two independent loci, they are derived from a common ancestor. Any uncertainty as to whether genes 1 and 6 represent two separate loci or two lineages of a single locus, and whether cattle do have six classical class I loci will only be resolved by the physical mapping of class I genes on a range of class I haplotypes.

Although seemingly advantageous to possess a greater number of class I genes, it must be considered that, at the level of the individual, a balance must be struck between the number of MHC loci and the T cell repertoire. The presence of too many class I loci may result in the immoderate elimination of the T cell repertoire during the process of negative selection in the thymus and so reducing the number of TCR able to react to pathogenic antigens. Indeed, in polyploid *Xenopus* species it has been shown that expression of all but one diploid set of MHC genes is suppressed, with the process of 'diploidization' believed to be the result of gene deletion (Flajnik et al., 1999). Mathematical modelling of survival probability versus the number of heterozygous MHC loci has shown that the optimum number of functional genes across both the class I and class II region is six, and survival probability becomes negative when this number exceeds twelve (Takahata, 1995). This type of analysis suggests that six class I loci represent an absolute maximum in cattle, whilst the presence of four loci seems a more favourable situation. However, rhesus macaques provide an exception to this rule and due to rapid expansion of the class I region appear to express at least nine class I loci per haplotype with no apparent detriment to their immune responses (Daza-Vamenta et al., 2004). Similarly, sequencing of a mouse lemur BAC has shown that although the MHC class I region on chromosome 6 only contains pseudogenes, a cluster of six class I genes encoding functional proteins has been mapped to chromosome 26 (Averdam et al., 2009),thus proving it is possible to express more than three class I genes without apparently compromising immune responses.

7.3.4. Variable haplotype configurations

Analysis of cattle class I haplotypes has shown that between one and three putative class I genes are expressed. Mapping of the class I genes on the A14 haplotype in a BAC library showed that four class I genes, *N*02301*, *N*02401*, *N*02501* and *N*04001*, were located within a contig of 400 kb. However, the *N*04001* sequence cannot be detected in cDNA isolated from resting lymphocytes and when transfected into COS-7 cells is unable to present antigen from *Theileria parva*, *Theileria annulata* or bovine herpes virus-1 (Di Palma et al.,

2002; Prof Ivan Morrison, personal communication), thus affirming the analysis that cattle express up to three functional class I genes per haplotype. The class I haplotypes of other artiodactyls have also been shown to express variable number of genes. In sheep, it is predicted that there are a minimum of four classical class I loci, of which three are expressed on ovine class I haplotypes (Miltiadou et al., 2005). Pigs have also been shown to express variable numbers of genes on class I haplotypes (Renard et al., 2001; Tanaka-Matsuda et al., 2009; Ho et al., 2009).

Whilst the majority of cattle class I haplotypes express two functional class I genes, single gene class I haplotypes in cattle are not uncommon. Of the twenty five cattle MHC class I haplotypes characterised at the molecular level, five are known to only express a single gene (see Table 7.1.). Of these five single gene haplotypes, one is detected in both lineages whilst three appear to be exclusive to the British Friesian population and the remaining single gene haplotype is exclusive to the Canadian Holstein bulls.

HAPLOTYPE	GENE	ALLELE EXPRESSED
A13	1	*N*03101*
A18	6	*N*01301*
A18v	6	*N*01302*
BF_NEW1	2	*N*05401*
BF_NEW5	2	*N*01601*

Table 7.1. Single gene haplotypes detected in cattle populations

Although five class I haplotypes expressing a single gene have been found in cattle, the expression of a single class I gene in mammalian species is unusual. The only other mammalian example is the speculative expression of a single gene class I gene in the grey short-tailed opossum. In this marsupial eleven class I genes have been detected, of which seven are known to be transcribed. Further analysis of these seven transcribed loci has led to the speculation that

the opossum expresses only a single classical class I locus, with the remaining six transcribed loci are believed to be non-classical based upon a number of characteristics including unusual alternative splicing, tissue-specific transcription and low levels of polymorphism (Baker et al., 2009).

In a similar manner to the opossum, the single gene haplotype in cattle, A18, was shown to express one classical class I in combination with a number of non-classical allele sequences (Birch et al., 2006). PCR amplification from both cDNA and gDNA derived from an A18 homozygote with two primer pairs designed to amplify partial exon 2 and 3 from all known alleles, followed by cloning and sequencing, demonstrated the presence of seven additional sequences besides the classical gene 6 allele, *N*01301*. Of these seven sequences, two were derived from cDNA templates and thus appear to be transcribed. Of the apparently transcribed sequences, one appeared to be non-classical and the other contained many divergent amino acid residues not seen in either classical or non-classical cattle class I alleles. Two of the remaining five sequences derived from gDNA templates from the A18 homozygote had classical characteristics, whilst the other three sequences had unusual amino acid residues and/or appeared to be non-classical genes. As only exons 2 and 3 were amplified during this study, it is entirely possible that the additional sequences detected on this single gene haplotype represent gene fragments or truncated genes. Birch and colleagues also detected additional sequences, both transcribed and untranscribed, on the A14 haplotype and additional untranscribed sequences on the A31 haplotype. In any case, it can be hypothesised that the presence of non-classical sequences, pseudogenes and gene fragments in the class I region may act as a reservoir of genetic elements with which recombination can occur, thus generating polymorphism (reviewed in Shum et al., 1993). Those additional sequences in which unusual amino acid residues were detected may be representative of alleles undergoing divergent mutations and/or truncations which will ultimately result in the loss of transcription of these genes (Birch et al., 2006).

The mechanism by which the variation in number of expressed genes on class I haplotypes in cattle is created has yet to be definitively determined. In the case of single gene haplotypes, it is unclear as to whether the other five class I loci are deleted completely, or are merely truncated thus rendering them inactive. Identifying the processes involved in creating a single gene haplotype becomes more difficult in the instances of the A18 and A18v haplotypes due to the unknown chromosomal location of gene 6. The study by Birch and colleagues (2006) highlighted the inability to detect the remaining five classical class I loci in either gDNA or cDNA samples from the A18 homozygote and implies that gene deletion is operating to reduce the number of classical class I loci from six to one. Together with gene duplication, gene deletion is known to give rise to the varying numbers of class I genes expressed on pig class I haplotypes (Tanaka-Matsuda et al., 2009) and evidence of each of these processes is apparent in the *BoLA* class I region. However, if gene deletion is responsible for the observed variability in gene number on class I haplotypes, it becomes difficult to explain why the NC1 gene is always present on all haplotypes despite its apparent close proximity to the classical class I genes 1, 2 and 4 (Birch et al., 2008; Di Palma et al., 2002). This may imply that, along with variability in gene number, the organisation of class I loci and possibly the genetic distance over which class I loci are organised on cattle class I haplotypes may also be variable. Evidence of this kind of variability has been provided by the physical mapping of the macaque class I B region of the *Mamu-h1* and *Mamu-h2* haplotypes (Bonhomme et al., 2008), and so a similar situation cannot be ruled out in cattle but could only be confirmed by physical mapping of a greater number of class I haplotypes

With a growing number of next generation sequencing methods becoming available, and a drive towards lowering the associated costs, studies to map the genes located in the *BoLA* class I region for a number of different class I haplotypes are not implausible. It would also be highly informative to extend such studies to the MHC class II region, thus giving information on complete MHC gene haplotypes in cattle. Not only would this shed light on the classical genes expressed in the bovine MHC, but also the pseudogene and gene fragment content which may be acting as a reserve for the generation of new allele sequences in future generations. This would improve our understanding of both

the processes involved in creating diversity and also the variability in gene content. Combining such studies to incorporate additional immune genes i.e. the *KIR* region, would generate complete 'immune gene haplotypes' which would greatly improve knowledge as to the causes of disease resistance and susceptibility to economically important disease e.g. mastitis or foot and mouth. Mapping of the *BoLA* complex will also enable the identification of a greater number of SNP markers across the *BoLA* region which could be refined to form a SNP marker map to be used for the MHC-typing of elite breeding animals to help improve diversity or select for MHC types linked with resistance to certain pathogens.

7.3.5. Forces driving *BoLA* class I diversity in AI bull populations

The evidence presented in this study shows that whilst the processes of genetic drift and inbreeding influence the pattern of MHC class I diversity in the Canadian Holstein bull population, the most influential force involved was that of human imposed selection for agri-economic traits. The panel of traits analysed focused on the traits of highest priority in terms of breeding for production, but the panel was relatively small given that progeny-derived data are available for over 60 individual traits on each published bull proof. The traits on which we focused our analysis were of high heritability and reliability. A number of conformation and type traits have low heritability and would give greater variability in the progeny, thus making this type of association analysis more difficult. In addition, the reliability of the conformation traits is often lower than for production traits due to less recording e.g. not every cow will be classified for type traits but her levels of production are recorded as a matter of course.

The gene substitution model used showed that certain haplotypes were associated with individual traits, and the associations recorded could be either favourable or undesirable. A driving force behind these associations appeared to be heterozygote advantage, with common haplotypes often unfavourably

associated with selection traits. Of interest would be to examine whether similar associations between MHC haplotype and selection traits are detectable in the British Friesian population, and whether heterozygote advantage was instrumental in these associations. The associations observed in the Canadian Holstein bulls were shown to be variable over a twenty year period of selection e.g. the associations found in two studies in the 1980s (Batra et al., 1989; Weigel et al., 1990) were generally not replicated in this study. In addition, the gene substitution effects vary with gene frequency, therefore it is perhaps unlikely that the same associations will be seen in a different breed lineage.

Although the pedigrees of Canadian Holstein and British Friesian derived animals are recorded on the same herdbook, trait data for the British Frieisian population is compared to a different genetic base. As a result, it is difficult to make direct numerical comparisons between the Canadian Holstein and British Friesian populations. In addition, the most informative trait data will be from the country in which the animal was proven e.g. the UK for the British Friesian bulls. Also, the relative numbers of progeny sired by the British Friesian bulls will be lower and as a result, the progeny-derived data will have a lower reliability. However, general trends can be compared between the two populations e.g. the A11 haplotype may be associated with an increase in milk yield in both populations, or the A14 haplotype may be associated with an increase in SCS in the British Friesian bull population, whereas it has been shown to be associated with a decrease in SCS in the Canadian Holstein bulls.

To further this kind of analysis, it would be informative to sample a larger number of animals and to include a number of different Holstein-Friesian lineages e.g. include the Dutch Friesian animals. It would also be of interest to extend the analysis to encompass more of the traits published on bull proofs. The publication of genomic proofs will contribute to selection trait data becoming more rapidly available for young bulls, but when using the gene substitution model, the progeny derived data add greater power to the analysis e.g. genomic proof data is not progeny tested and so cannot account for environmental variability.

7.3.6. The future of the SNP assay

The expansion of diversity studies to incorporate the British Friesian animals has highlighted the need to broaden the SNP assay to include additional common class I haplotypes e.g. A31, A18, and A18v. The shared A10, A11 and A20 haplotypes would also be worthy of inclusion in an extended MHC class I SNP assay panel. Using sequence alignments it was possible to identify SNPs in a number of alleles from these haplotypes, as detailed in Table 7.2.

HAPLOTYPE	GENE	ALLELE	SNP POSITION (CODON NO.)
A10	3	*N*00201*	9, 136
A11	2	*N*01801*	111, 45
	3	*N*01701*	136
A18	6	*N*01301*	45, 97, 105
A18v	6	*N*01302*	45, 97, 105
A20	2	*N*02601*	8, 174
	3	*N*02701*	99, 174
A31	1	*N*02101*	35

Table 7.2. Possible SNP positions (as indicated by codon number) for the identification of alleles expressed on class I haplotypes to be included on the original panel of SNPs.

The acquisition of new class I data dictated that the SNP markers in the A14, A15 and W12B panels had to be checked to confirm they were still unique identifiers. With the inclusion of the new class I sequences, the SNP marker at codon 39 used for the identification of the *N*00801* allele on the W12B haplotype was no longer unique, and would cross-react with the gene 2 allele *N*05401* on the BF_NEW1 haplotype. The possibility of using a SNP marker in codon 19 would have to be re-investigated so that the W12B SNP panel would have two unique markers for the identification of *N*00801*. Due to the way in which MHC genes have evolved, these minor complications are inherent with the use of SNP markers for MHC gene identification i.e. it is not an ideal system. However, this technique can be rapidly adapted to take into account the availability of new sequence data which may render particular SNP markers less informative.

The SNP markers used for the identification of A14 and A15 remained unique. However, the existing panel of SNPs are unable to differentiate between the A15 and A15v haplotypes. This is due to the homology between the *N*00902* allele and the *N*02301* and *N*00901* alleles across the alpha 1 and 2 domains respectively. Sequence alignments show that there are no suitable SNP residues that would enable further differentiation between the *N*00902* and *N*00901* alleles. In cases where the animal carries the A15v haplotype, there could be misinterpretation of the results e.g. in A15v homozygous animals it may appear that the haplotype of that animal is A14/A15 and in A15v heterozygous animals it may appear that the animal is A15 heterozygous. The low frequency of this haplotype in the population should mean that the occurrence of A15v homozygotes is rare and the rates of misidentification are rare. In any case, the expression of the *N*02401* and *N*02501* alleles on each of these three haplotypes means that any breeding scheme should avoid matings that result in A14, A15 and A15v homozygotes, or heterozygotes of these three haplotypes, to avoid fixation of these two alleles at high frequency in the Holstein-Friesian population.

When applying the SNP assay to samples from the wider UK herd, it was found that samples classed as British Friesian were shown to contain the A15

haplotype. The lack of the A15 haplotype in the British Friesian bull sample makes the detection of this haplotype in animals derived from this lineage surprising. Two possibilities are presented as an explanation; the SNP panel used could have misidentified an A15v haplotype as A15, or the breed data recorded with the sample was inaccurate, with each explanation equally plausible.

As with the expansion of the trait association analysis, the future development of the SNP assay would benefit from the acquisition of more MHC class I allele sequence data from a larger sample of animals. This would give a more thorough understanding of allelic diversity in the population and enable the identification of appropriate unique SNP markers with which to type cattle. The use of SNP markers has been applied to studying variation in the extended human MHC to better understand genetic variation responsible for MHC disease associations e.g. diabetes and celiac disease (Bakker et al., 2006; Nejentsev et al., 2007). By increasing the amount of data from cattle populations in terms of MHC class I diversity together with trait associations and also disease incidence it is conceivable that more definitive QTL for production and disease resistance may be identified in and surrounding the bovine MHC. A preliminary assessment of class II sequence alignments showed fewer allele-specific SNP markers and so the examination of the linked class I genes may be just as informative. A more achievable aim may be the use of SNP markers to identify major types at a given class II locus. Certainly, it may be possible to identify MHC supertypes that would aid in the design of vaccination strategies. In addition, the inclusion of other immune genes i.e. NK cell receptor genes, into the SNP panel, in combination with the MHC information, would provide a potentially powerful tool for the selection of cattle that are more immunologically robust when faced with the challenge of emerging pathogens. With the routine application of genomic selection tools in the dairy industry an imminent proposal, the availability of SNP markers that provide information on both the levels of diversity of highly polymorphic immune genes as well as selection traits can only prove beneficial to the dairy industry breeding strategies now and in the future.

Appendices

A1. Primer sequences

GENE	PRIMER NAME	PRIMER SEQUENCE (5'-3')	PRODUCT SIZE (nt)
1	F Group 1 primer4	TGCGAGGGGACCGCCCGA	1350
	R Group 1 primer4	AGGTGAGAACAGGCCTTGAGAA	
2	F2a	GAACRAGCGACCCCGACT	1125
	R2c	CAAGTGGGGCAACTGGTC	
3	F3c	TCGACCGCTTCCATCTCG	1150
	R3e	GAACAGGCCTTGAGAGAC	
6	F6d	TCATTGACCCTCCGCCCA	900
	R6e	GGCGCTGTTYCCACAGGC	

Table A.1. Details of gene-specific primers.

ALLELE	ASSOCIATED HAPLOTYPE	PRIMER NAME	PRIMER SEQEUNCE 5'- 3'	PRIMING SITE
*N*03101*	A13	A12.1F	TCTGAAGTATTTCTACACCGCCGTA	exon 2
		A12.1R	CCTGGATATTGTGAGACCCTGCG	intron 2/exon 3
*N*01901*	W12B	4221.1F	CCCGCTTCATACCGTT	exon 2
		4221.1R	CACGTCGCAGCCGTACATG	exon 3
*N*01201*	A10	A10.2F	GGATCGGGAGACGCGAAACTT	exon 2
		A10.2R	CGTAGGCGTCCTGCCTGTAT	exon 3
*N*02401*	A14 & A15	D18.1F	CCCGCTTCATCACCGTT	exon 2
		D18.1R	ACTGTACCCGCGGAGGAAACT	exon 3
*N*02201*	A31	HD7F	GGCGCCGTGGATGGAGCAA	exon 2
		HD7R	GCACCTGCCGCCTCCCG	exon 3
*N*05601*	New5	*N*05601*F	GCTTCATCGCCGTCGGCTAT	exon 3
		*N*05601*R	TACTGCGTGAACCCGCT	exon 4
*N*03701*	New5	*N*03701*F	CGGGAGGCGGCAGGTGT	exon 2
		*N*03701*R	ACAAACTCCATGTCCTGCA	exon 3

Table A.2. Details of sequence-specific PCR (SSP-PCR) primers.

PRIMER NAME	SEQUENCE 5'-3'	PRIMING SITE
Bov7	GGCTACGTGGACGACACG	Bovine MHC I exon 2
Bov11	CCCTCCAGGTAGTTCCT	Bovine MHC I exon 3
Bov21a	CATGGGGCCGCGAAC	Bovine MHC I exon 1
Bov21g	CATGGGGCCGCGAGC	Bovine MHC I exon 1
Bov3	GGATGAAGCATCACTCAG	3'UTR

Table A.3. Details of generic primers.

PRIMER NAME	SEQUENCE 5'-3'
M13F	GTTTTCCCAGTCACGAC
SP6	ATTTAGGTGACACTATAGAA

Table A.4. Details of primers used for sequencing inserts from pGEM-Teasy vector.

HAPLOTYPE ASSAY	PRIMER NAME	SEQUENCE 5'-3'
A14 & A15	SNP55R	TCTCGCGATCCCAATACTCCGGCCC
	SNP129F	TACGACGGCAGAGATTACATCGCCCTGAACGAGGA
	SNP122R	AGGTCCTCGTTCAGGGCGATGTAAT
W12B	SNP9R_ATAIL	AAAAACCGGGACACGGCGGTG
	SNP39F	GCAGTTCGTGCGGTTCGACAGC

Table A.5. Details of SNP extension primers used for detection of common class I haplotypes.

A2. Experimental details

A2.1. Formula used for calculating the lifetime profit index (LPI) EBVs (referenced in Chapter 1. Table 1.3. and Chapter 4. Table 4.2.)

LPI = Production Component x Emphasis x Factor + Durability Component x Emphasis x Factor + Health & Fertility Component x Emphasis x Factor

Values used for calculating LPI for Holstein breed in January '09 proof run:

COMPONENT	EMPHASIS	FACTOR
Production	51	1.5435
Durability	34	1.5168
Health & Fertility	15	1.4990

Production Component (PROD):

PROD = [WPYx(PY-AvgPY)/SDPY] + [WPDxPD/SDPD] + [WFYx(FY-AvgFY)/SDFY] + [WFDxFD/SDFD]

Where PY = Protein Yield, PD = Protein Deviation, FY = Fat Yield and FD = Fat Deviation, which are standardized using the appropriate averages (Avg) and standard deviations (SD) and then multiplied by their respective relative weight (W), all of which are breed specific as outlined in the following table.

PARAMETER	TRAIT	HOLSTEIN BREED VALUES
EBV Averages	Protein Yield	9.00
	Fat Yield	6.00
EBV Standard Deviations	Protein Yield	25.00
	Protein Deviation	0.12
	Fat Yield	31.00
	Fat Deviation	0.30
Relative Weights Within the Production Component	Protein Yield	5.70
	Protein Deviation	0.30
	Fat Yield	3.80
	Fat Deviation	0.20

Durability Component (DUR):

DUR = [WHL x (HL - 100)/5] + [WMS x MS/5] + [WF&L x F&L/5] + [WDS x DS/5]

Where HL = Herd Life, MS = Mammary System, F&L = Feet and Legs, DS = Dairy Strength and each trait is standardized using the appropriate averages and standard deviations and then multiplied by their respective relative weight (W) as outlined in the following table.

PARAMETER	TRAIT	HOLSTEIN BREED VALUES
Relative Weights Within the Durability Component	Herd Life	2.0
	Mammary System	4.0
	Feet & Legs	3.0
	Dairy Strength	1.0

Health & Fertility Component (H&F):

H&F = [WSCS x -1 x (SCS-3.00)/0.23] + [WUD x UD/5] + [WMSP x (MSP-100)/5] +

[WDF x (DF-100)/5] + [WLP x (LP-100)/5]

Where SCS = Somatic Cell Score, UD = Udder Depth, MSP = Milking Speed, DF = Daughter Fertility and LP = Lactation Persistency. The relative weights for each trait (i.e.: WSCS, WUD, WMSP, WDF and WLP respectively), which are specific to each breed, are provided in the following table.

PARAMETER	TRAIT	HOLSTEIN BREED VALUES
Relative Weights Within the Health & Fertility Component	Somatic Cell Score	2.0
	Udder Depth	1.0
	Milking Speed	0.3
	Daughter Fertility	6.7
	Lactation Persistency	0.0

A2.2. Converting serological specificities to molecular definitions

SEROLOGICAL DEFINITIONS	MOLECULAR DEFINITIONS
w7	A7(w50)
w9A +w9B	A9
w10	A10
w11	A11
w12.1	w12
w13	A13
w8.1(w8)	A14
w8.2(w8)	A15
w16	w16
w6.2(w6)*	A17
w6.1(w6)*	A19
w20A + w20B	A20
CA42B	A21
CC1	A26
w12.2	W12B

As described by Bull et al., (1989), except those marked with an asterisk which, due to discrepancies in local names for sera falling under the w6 supertype, are named differently (Stear et al., 1989).

A2.3. Haplotype frequencies and measures of diversity (data presented in Chapter 3. Section 3.3.7.)

The number of alleles, the heterozygosity and expected heterozygosity (allelic diversity) were determined with the Allele procedure on SAS/Genetics module (SAS Institute, Cary, North Carolina). The procedures have been described previously (Weir, 1996). In brief, as all alleles were inherited codominantly, the number of copies (n_u) of haplotype M_u was determined as $n_u = 2n_{uu} + \Sigma\, n_{uv}$. The haplotype frequencies in the sample were estimated as $p_u = n_u / (2n)$ where n is the number of individuals sampled. These frequencies are unbiased maximum likelihood estimates (MLE) of the population proportions. A total of 10000

permutations were used to approximate the p-value for the Hardy-Weinberg equilibrium test.

The heterozygosity (also known as the observed heterozygosity) is the proportion of heterozygous individuals in the data set. It is 1- Σ P_{uu} summed over all alleles where Puu is the frequency of homozygotes for haplotype u. The allelic diversity, also known as the expected heterozygosity, is the expected proportion of heterozygous individuals in the data set under Hardy-Weinberg equilibrium. It is 1- Σ $(p_u)^2$ summed over all haplotypes where p_u is the frequency of haplotype u.

A permutation version of the exact test for Hardy-Weinberg equilibrium was used (Guo & Thompson 1992). This test is based on the conditional probability of genotype counts, given the counts of alleles under the hypothesis of independence of alleles. Hill diversity measures, derived from Rényi's generalised entropy (Hill 1973; Renyi 1961) were used to compare the diversities of the two samples. This procedure has been used by ecologists to compare different measures of entropy but does not seem to have been used to examine genetic diversity. For ecologists, the Hill diversity measures (Equation 1) give an estimate of the effective number of species (Jost 2006). In our study, 'species' corresponds to haplotype.

$$D_q(p_1 \cdots p_s) = \begin{cases} \left[\sum_{i=1}^{s} p_i^q \right]^{\frac{1}{1-q}} & q \neq 1 \\ \prod_{i=1}^{s} p_i^{-p_i} & q = 1 \end{cases} \qquad \text{Equation 1}$$

where $p_1 \cdots p_s$ are the s non-zero haplotype abundances

The diversity measure, D_q, has a single parameter q that varies from 0 to infinity. It determines the extent to which haplotypes contribute towards overall diversity. In a genetic context, D_0 is the number of haplotypes, D_1 is derived from

the Shannon entropy, and is the sum of the haplotype frequencies multiplied by their logarithms, D_2 corresponds to the square of the haplotype frequencies and thus reflects expected homozygosity and D_∞ is derived from the frequency of the most common haplotype.

To measure the diversity in each sample set (1986 and 2006), the haplotype frequencies were derived directly from the count data ($p_j = n_j / \Sigma n_i$), and the different diversities were calculated from the frequencies. As sample diversity measurements depend heavily on sample size (Sanders 1968), direct comparison of the two samples was conducted by sub- sampling the larger (1986) sample to the size of the smaller (2006) without replacement; confidence intervals were generated by resampling. Extending this method to the level of the population, we calculated the scenarios that maximise and minimise both diversity and the associated confidence intervals for the diversity distributions.

A2.4. Mathematical modelling of genetic drift and inbreeding (data presented in Chapter 3. Section 3.3.8.)

A mathematical model was used to simulate the effects of inbreeding and drift. As 20 years represents approximately four generations of cattle the pedigrees of all bulls sampled in 2006 back to the great-grandparents were used. Haplotypes were randomly assigned to the great grandparents based on the frequencies observed in 1986. Haplotypes were randomly assigned to progeny by mimicking Mendelian inheritance. Each offspring had a 50 % chance of inheriting either of the two haplotypes in each parent. The gene frequencies were then estimated in the 72 bulls sampled in 2006. The algorithm was written in SAS and 5000 replicates were run.

A2.5. Associations of MHC class I haplotypes with production and fitness traits (Chapter 4)

Means and standard deviations for the estimated breeding values (EBV) using these traits were calculated with the univariate procedure on SAS (Table 2). A gene substitution model was used to estimate the average effect of each haplotype, and a separate model was used for each trait EBV. The model has been previously published e.g. Batra et al., 1989. The haplotypes were fitted as covariates. To reduce dependencies among the equations, the most common haplotype (W12B) was set to zero. The most common haplotype was chosen as the reference because the higher frequency means that its effects are estimated more precisely. The mixed procedure of SAS was used in each of these analyses. Rare alleles present in fewer than 5 bulls were included in the model but the results are not presented because their rarity means that the estimates for them are unreliable.

When a large number of comparisons are made, low probabilities may arise by chance alone. We carried out a meta-analysis of probabilities (Fisher 1970) to determine whether the low probabilities were meaningful. The test statistic $-2 \Sigma \ln (p)$ is distributed as chi-square with twice as many degrees of freedom as probabilities tested.

A3. Animal data

A3.1. MHC haplotypes of the Canadian Holstein bull sample (2006)

The 2006 Canadian Holstein bull sample consisted of 72 bulls in total i.e. one sample from which gDNA could not be extracted (Canadian bull number 24), and one duplicated sample (Canadian bull number 23 was a repeat of Canadian bull number 9). Canadian Holstein bulls numbered 50 to 74 were MHC-typed by J Birch in the Compton laboratory (2005).

ANIMAL DATA		
ANIMAL NO.	**HAPLOTYPE 1**	**HAPLOTYPE 2**
1	A11	A19
2	A13	A10
3	A11	W12B
4	A14	New5
5	A15	A10
6	A14	New5
7	A15	A10
8	A15	A15
9	A11	A15
10	A10	W12B
11	W12B	New24
12	W12B	W12B
13	A14	A13
14	A14	A11
15	A11	W12B
16	A20	W12B
17	A11	W12B
18	A11	A15
19	A15	W12B
20	A11	W12B
21	A19	A13
22	A20	A10
23	REPEAT SAMPLE	
24	NO DNA	
25	A20	W12B
26	A10	A14
27	A14	A11
28	A15	A11
29	A14	New5
30	A20	A13
31	A10	New5
32	W12B	A10
33	A20	A11
34	A19	W12B
35	A20	A19
36	W12B	A13
37	W12B	A10
38	A14	A13
39	A14	A15
40	A20	A20
41	A14	W12B
42	A10	W12B
43	A15	A11
44	A11	W12B
45	A15	A19
46	A10	A20

47	A14	A11
48	A10	A11
49	A14	W12B
50	A15	A19
51	W12B	New22
52	A15	A19
53	A15	W12B
54	A10	A14
55	A15	A20
56	A15	A20
57	A15	A20
58	A10	A15
59	A10	A19
60	A15	A15
61	A15	W12B
62	A20	W12B
63	A15	A20
64	A10	A19
65	A14	W12B
66	A10	A19
67	W12B	W12B
68	A14	A20
69	A13	A14
70	A14	A15
71	A14	A15
72	A15	W12B
73	A11	A20
74	A15	A11

A3.1. Production trait proof data for the 2006 Canadian Holstein bull sample

ANIMAL DATA			PRODUCTION EBV *09 JAN				
ANIMAL NO.	HAPLOTYPE 1	HAPLOTYPE 2	MILK (KG)	FAT (KG)	PROTEIN (KG)	NO. OF DAUGHTERS	RELIABILITY
1	A11	A19	1025	45	36	1959	99
2	A13	A10	2003	77	46	1237	98
3	A11	W12B	97	78	18	337	96
4	A14	New5	1332	39	38	2062	99
5	A15	A10	1251	23	27	2012	99
6	A14	New5	840	-5	47	154	94
7	A15	A10	854	51	55	112	94
8	A15	A15	1831	39	58	53	88
9	A11	A15	649	27	22	98	93
10	A10	W12B	1270	31	37	4219	99
11	W12B	New24	1092	58	24	1501	98
12	W12B	W12B	974	48	27	225	96
13	A14	A13	955	27	26	148	95
14	A14	A11	259	6	24	1758	99
15	A11	W12B	1788	84	54	167	93
16	A20	W12B	178	28	5	171	96
17	A11	W12B	110	-1	9	21277	99
18	A11	A15	357	1	25	6052	93
19	A15	W12B	697	17	24	2853	99
20	A11	W12B	1073	12	22	160	95
21	A19	A13	-416	33	-3	14364	99
22	A20	A10	405	2	5	118	94
23	REPEAT SAMPLE						
24	NO DNA						

25	A20	W12B	-699	33	6	88	92
26	A10	A14	1459	13	25	70	80
27	A14	A11	1425	19	40	98	93
28	A15	A11	158	22	19	101	93
29	A14	New5	482	42	15	100	82
30	A20	A13	537	32	39	87	93
31	A10	New5	349	30	9	112	93
32	W12B	A10	318	19	17	94	81
33	A20	A11	449	13	17	120	94
34	A19	W12B	823	43	31	123	94
35	A20	A19	459	-26	13	129	94
36	W12B	A13	-28	8	21	103	93
37	W12B	A10	-40	30	-2	96	81
38	A14	A13	428	1	14	109	81
39	A14	A15	923	11	-1	113	94
40	A20	A20	207	-48	11	179	96
41	A14	W12B	842	22	33	65	79
42	A10	W12B	-145	40	16	104	94
43	A15	A11	233	1	-2	85	91
44	A11	W12B	2364	26	47	102	93
45	A15	A19	943	14	38	72	79
46	A10	A20	749	43	19	96	93
47	A14	A11	766	35	42	88	81
48	A10	A11	942	-4	38	83	80
49	A14	W12B	1582	14	23	249	96
50	A15	A19	-295	14	15	19815	99
51	W12B	New22	471	30	0	33350	99
52	A15	A19	319	-16	8	10360	99
53	A15	W12B	138	10	-2	22112	99

54	A10	A14	507	23	10	3936	99
55	A15	A20	118	50	-6	239	96
56	A15	A20	694	60	24	428	98
57	A15	A20	807	65	27	6669	99
58	A10	A15	1460	42	48	15535	99
59	A10	A19	748	63	38	6683	99
60	A15	A15	433	54	2	112	93
61	A15	W12B	-969	-24	-17	134	94
62	A20	W12B	439	53	25	1619	99
63	A15	A20	544	-8	4	18902	99
64	A10	A19	631	5	25	5523	99
65	A14	W12B	-307	30	21	1908	99
66	A10	A19	-216	46	7	3499	99
67	W12B	W12B	10	-5	-4	546	98
68	A14	A20	239	46	23	1482	98
69	A13	A14	193	-9	18	843	98
70	A14	A15	1223	-1	33	4189	99
71	A14	A15	897	40	21	240	97
72	A15	W12B	-283	29	1	141	95
73	A11	A20	1213	38	39	115	92
74	A15	A11	798	4	31	95	91

A3.2. Functional trait proof data for the 2006 Canadian Holstein bull sample

ANIMAL DATA			FUNCTIONAL EBVs					
ANIMAL NO.	HAPLOTYPE 1	HAPLOTYPE 2	HERD LIFE	RELIABILITY	DAUGHTER FERTILITY	RELIABILITY	SCS	RELIABILITY
1	A11	A19	100	87	97	93	2.89	99
2	A13	A10	102	88	101	91	2.76	98
3	A11	W12B	91	84	100	81	3.18	94
4	A14	New5	103	95	106	96	2.73	99
5	A15	A10	109	88	102	96	3.09	99
6	A14	New5	98	84	98	81	3.18	91
7	A15	A10	106	86	107	79	2.77	90
8	A15	A15	93	75	91	66	3.42	82
9	A11	A15	102	84	104	77	2.92	88
10	A10	W12B	102	91	96	98	3.03	99
11	W12B	New24	106	91	96	89	2.75	98
12	W12B	W12B	100	87	99	83	2.92	93
13	A14	A13	107	85	102	82	2.85	92
14	A14	A11	109	93	102	94	2.67	99
15	A11	W12B	96	80	86	78	3.03	90
16	A20	W12B	98	88	102	81	3.02	93
17	A11	W12B	105	99	101	99	3.42	99
18	A11	A15	104	82	104	63	2.67	92
19	A15	W12B	92	94	91	97	3.69	99
20	A11	W12B	101	89	105	79	3.12	91
21	A19	A13	103	99	96	99	2.9	99
22	A20	A10	101	86	91	79	3.06	90
23	REPEAT SAMPLE							
24	NO DNA							

25	A20	W12B	109	82	108	74	3.07	88
26	A10	A14	95	54	96	45	3.23	70
27	A14	A11	106	77	102	77	2.9	89
28	A15	A11	102	83	102	77	2.86	89
29	A14	New5	99	59	93	47	2.89	74
30	A20	A13	102	73	104	75	2.92	88
31	A10	New5	99	57	97	74	3.17	88
32	W12B	A10	104	54	103	44	2.97	72
33	A20	A11	95	60	94	75	3.14	89
34	A19	W12B	99	56	91	74	3.09	89
35	A20	A19	110	54	103	73	2.68	88
36	W12B	A13	105	63	99	74	3.23	87
37	W12B	A10	91	53	92	44	3.09	72
38	A14	A13	99	53	97	44	3.18	73
39	A14	A15	90	86	97	76	3.35	89
40	A20	A20	102	86	106	81	3.55	92
41	A14	W12B	96	53	96	44	3.08	69
42	A10	W12B	108	83	106	78	3.19	90
43	A15	A11	93	79	97	69	3.43	84
44	A11	W12B	100	83	99	74	2.97	88
45	A15	A19	101	54	100	45	3.05	70
46	A10	A20	106	83	101	76	2.6	88
47	A14	A11	96	59	101	46	3.03	73
48	A10	A11	102	56	98	43	2.74	72
49	A14	W12B	105	90	95	85	3.04	94
50	A15	A19	94	99	94	99	2.88	99
51	W12B	New22	110	99	104	99	3.05	99
52	A15	A19	97	99	93	99	2.78	99

53	A15	W12B	103	99	98	99	3.04	99
54	A10	A14	109	94	100	97	2.73	99
55	A15	A20	95	90	94	84	3	94
56	A15	A20	95	93	95	90	2.87	97
57	A15	A20	90	98	91	98	3.08	99
58	A10	A15	101	98	98	99	3.17	99
59	A10	A19	111	92	100	97	2.48	99
60	A15	A15	91	79	93	74	2.85	88
61	A15	W12B	101	73	101	79	3.23	91
62	A20	W12B	111	97	105	96	2.92	99
63	A15	A20	106	99	100	99	3.19	99
64	A10	A19	104	98	96	98	3.13	99
65	A14	W12B	103	92	108	95	2.89	99
66	A10	A19	103	90	101	96	2.78	99
67	W12B	W12B	109	91	109	90	3.13	97
68	A14	A20	109	86	95	86	2.88	98
69	A13	A14	100	88	95	91	2.86	97
70	A14	A15	106	87	92	96	2.87	99
71	A14	A15	97	90	95	86	3.04	95
72	A15	W12B	103	65	100	78	3.18	91
73	A11	A20	99	55	99	71	3.11	85
74	A15	A11	100	54	101	66	2.9	83

A3.3. Conformation trait proof data for the 2006 Canadian Holstein bull sample

ANIMAL DATA			CONFORMATION TRAITS						
ANIMAL NO.	HAPLOTYPE 1	HAPLOTYPE 2	CONFORMATION	MAMMARY SYSTEM	FEET & LEGS	DAIRY STRENGTH	RUMP	NO. DAUGHTERS	REL %
1	A11	A19	7	7	4	6	3	1907	99
2	A13	A10	4	5	-2	7	1	1170	98
3	A11	W12B	4	3	0	8	4	298	95
4	A14	New5	0	3	-5	-2	-2	1737	99
5	A15	A10	5	8	-3	0	6	1808	99
6	A14	New5	7	5	3	10	0	163	91
7	A15	A10	-1	-2	1	1	2	108	87
8	A15	A15	-1	1	1	3	-6	49	79
9	A11	A15	-3	5	-5	-10	-2	85	86
10	A10	W12B	7	4	9	5	4	3934	99
11	W12B	New24	5	1	7	5	1	1523	98
12	W12B	W12B	4	7	4	0	-3	230	93
13	A14	A13	7	7	-4	8	2	149	90
14	A14	A11	7	13	7	-5	-1	1685	99
15	A11	W12B	2	3	1	-1	4	156	91
16	A20	W12B	9	5	0	11	12	150	91
17	A11	W12B	11	7	5	10	14	19655	99
18	A11	A15	2	5	1	-3	-3	1360	82
19	A15	W12B	3	2	5	0	4	2501	99
20	A11	W12B	2	3	0	1	-1	151	90
21	A19	A13	9	8	3	7	6	13240	99
22	A20	A10	6	6	11	1	-7	111	88
23	REPEAT SAMPLE								
24	NO DNA								

25	A20	W12B	-1	1	-2	-3	2	83	86
26	A10	A14	-1	-2	0	5	2	46	67
27	A14	A11	6	3	4	5	4	97	87
28	A15	A11	0	3	2	0	-11	87	86
29	A14	New5	1	0	1	2	0	49	67
30	A20	A13	3	2	4	2	4	85	86
31	A10	New5	2	5	1	-2	2	98	87
32	W12B	A10	1	0	4	0	2	45	67
33	A20	A11	1	-1	3	2	-3	110	88
34	A19	W12B	1	1	2	2	-2	117	87
35	A20	A19	1	4	-2	-2	-1	123	89
36	W12B	A13	-2	1	-2	-6	1	96	87
37	W12B	A10	9	7	1	7	3	56	68
38	A14	A13	-2	-2	1	-2	1	62	69
39	A14	A15	0	2	-7	2	2	102	88
40	A20	A20	4	4	6	2	-1	153	91
41	A14	W12B	0	-1	0	1	0	46	67
42	A10	W12B	5	6	3	0	3	105	87
43	A15	A11	-3	-3	2	0	-2	67	83
44	A11	W12B	3	6	9	5	-11	104	88
45	A15	A19	-6	-3	-5	-8	3	34	64
46	A10	A20	0	2	3	-5	-1	88	86
47	A14	A11	-2	-4	1	3	1	70	70
48	A10	A11	-4	-4	3	-2	-1	71	70
49	A14	W12B	8	6	-1	10	7	238	93
50	A15	A19	4	5	-4	2	2	17085	99
51	W12B	New22	9	6	7	4	8	29596	99
52	A15	A19	6	1	10	7	6	9287	99
53	A15	W12B	9	7	3	9	8	19997	99

54	A10	A14	6	2	3	6	10	3597	99
55	A15	A20	7	7	1	6	4	204	93
56	A15	A20	4	2	8	4	-3	364	96
57	A15	A20	3	5	0	3	-2	5998	99
58	A10	A15	7	10	8	1	-8	14673	99
59	A10	A19	16	16	14	10	4	7262	99
60	A15	A15	3	0	-2	6	2	89	87
61	A15	W12B	6	3	4	7	1	124	89
62	A20	W12B	5	8	4	-4	4	1438	98
63	A15	A20	11	9	16	6	2	17799	99
64	A10	A19	6	6	3	-1	6	5093	99
65	A14	W12B	5	4	0	2	5	1704	99
66	A10	A19	6	2	8	5	5	3397	99
67	W12B	W12B	7	5	7	1	7	522	97
68	A14	A20	11	9	8	9	5	1952	99
69	A13	A14	6	8	4	3	-1	786	98
70	A14	A15	9	9	4	8	0	4096	99
71	A14	A15	-1	-4	-5	7	5	219	93
72	A15	W12B	2	0	4	2	-1	129	90
73	A11	A20	5	2	8	4	3	98	87
74	A15	A11	4	1	7	4	5	87	86

A3.4 Economic trait proof data for the 2006 Canadian Holstein bull sample

ANIMAL DATA			ECONOMIC	
ANIMAL NO.	HAPLOTYPE 1	HAPLOTYPE 2	LPI	NO. DAUGHTERS
1	A11	A19	1239	1959
2	A13	A10	1584	1237
3	A11	W12B	952	337
4	A14	New5	1076	2062
5	A15	A10	972	2012
6	A14	New5	841	154
7	A15	A10	1637	112
8	A15	A15	812	53
9	A11	A15	574	98
10	A10	W12B	1131	2417
11	W12B	New24	1173	1501
12	W12B	W12B	1107	225
13	A14	A13	1001	148
14	A14	A11	1342	1758
15	A11	W12B	1290	167
16	A20	W12B	480	171
17	A11	W12B	544	21277
18	A11	A15	788	6052
19	A15	W12B	23	2853
20	A11	W12B	579	160
21	A19	A13	622	14364
22	A20	A10	174	118
23	REPEAT			
24	NO DNA			

25	A20	W12B	666	88
26	A10	A14	-21	70
27	A14	A11	1169	98
28	A15	A11	695	101
29	A14	New5	314	100
30	A20	A13	1235	87
31	A10	New5	328	112
32	W12B	A10	597	94
33	A20	A11	-15	120
34	A19	W12B	581	123
35	A20	A19	226	129
36	W12B	A13	262	103
37	W12B	A10	66	96
38	A14	A13	-216	109
39	A14	A15	-649	113
40	A20	A20	-30	179
41	A14	W12B	329	65
42	A10	W12B	1246	104
43	A15	A11	-673	85
44	A11	W12B	1353	102
45	A15	A19	234	72
46	A10	A20	902	96
47	A14	A11	731	88
48	A10	A11	335	83
49	A14	W12B	533	249
50	A15	A19	85	19815
51	W12B	New22	881	33350
52	A15	A19	-38	10360
53	A15	W12B	322	22112

54	A10	A14	649	3936
55	A15	A20	283	239
56	A15	A20	936	428
57	A15	A20	717	6669
58	A10	A15	1689	15535
59	A10	A19	2669	6683
60	A15	A15	-32	112
61	A15	W12B	-370	134
62	A20	W12B	1591	1619
63	A15	A20	797	18902
64	A10	A19	582	5523
65	A14	W12B	1058	1908
66	A10	A19	926	3499
67	W12B	W12B	531	546
68	A14	A20	1429	1482
69	A13	A14	425	843
70	A14	A15	796	4189
71	A14	A15	44	240
72	A15	W12B	351	141
73	A11	A20	1138	115
74	A15	A11	732	95

A3.5. British Friesian bull sample haplotype information

EXPT ANIMAL NAME	MHC CLASS I HAPLOTYPE	
	HAPLOTYPE 1	HAPLOTYPE 2
GEN1	A18v	BF_NEW4
GEN2	A18v	A10
GEN3	A18	BF_NEW6
GEN4	A31	A18v
GEN5	A11	New5
GEN6	A31	A31
GEN7	A15v	A18
GEN8	A31	BF_NEW7
GEN9	A18	BF_NEW7
GEN10	A14	BF_NEW7
GEN11	A11	BF_NEW7
GEN12	A31	A31
GEN13	A10	BF_NEW3
GEN14	A18v	A14
GEN15	A31	A18v
GEN16	A11	A18
GEN17	A18v	A18v
GEN18	A14	A20
GEN19	A11	BF_NEW4
GEN20	BF_NEW6	A18v
GEN21	BF_NEW6	A17
GEN22	BF_NEW7	BF_NEW7
GEN23	BF_NEW4	BF_NEW6
GEN24	BF_NEW4	A18v
GEN25	A18	A14
GEN26	A10	BF_NEW6
GEN27	A31	BF_NEW6
GEN28	A17	BF_NEW3
GEN29	A10	New5
GEN30	A15v	BF_NEW4
GEN31	A11	BF_NEW2
GEN32	A14	A18v
GEN33	A18v	BF_NEW7
GEN34	A14	BF_NEW3
COG1	A14	BF_NEW5
COG2	A18	A18v
COG3	A20	A11
COG4	A14	BF_NEW4
AVONCROFT 1	A14	A20
AVONCROFT 2	A14	A31
BLG 1	A14	A14
BLG 2	A14	A11
BLG 3	A14	A11
BLG 4	BF_NEW4	BF_NEW7
BLG 5	A14	BF_NEW7
BLG 6	A14	A18v

A3.6. British Friesian cow sample haplotype information

HERD NO.	MHC HAPLOTYPE DATA	
	HAPLOTYPE 1	HAPLOTYPE 2
25	A14	A11
202	A31	BF_NEW1
100	BF_NEW3	New5
165	A11	A18v
24	A14	A14
42	A14	A31
44	A14	BF_NEW2
19	A18v	BF_NEW2
75	BF_NEW2	BF_NEW3
32	A31	A11
114	A31	BF_NEW1
15	A31	A31
73	A14	BF_NEW3
189	A10	BF_NEW1
208	A14	A11
23	A18v	BF_NEW3
162	A31	A31
10	A31	BF_NEW2
51	A31	A18v
36	A31	A11
140	A14	A18v
78	A14	A31
61	A14	BF_NEW4
175	A11	A18
52	A10	A31

List of References

Aarestrup, F. U., N. E. Jensen, et al. (1995). Analysis of associations between major histocompatibility complex (BoLA) class I haplotypes and subclinical mastitis of dairy cows. Journal of Dairy Science **78**(8): 1684-1692.

Abi Rached, L., M. F. McDermott, et al. (1999). The MHC big bang. Immunological Reviews **167**: 33-44.

Achilli, A., S. Bonfiglio, et al. (2009). The multifaceted origin of taurine cattle reflected by the mitochondrial genome. PLoS ONE **4**(6): e5753.

Aguilar, A., G. Roemer, et al. (2004). High MHC diversity maintained by balancing selection in an otherwise genetically monomorphic mammal. Proceedings of the National Academy of Sciences of the United States of America **101**(10): 3490-3494.

Aguilar, B., P. Vos, et al. (1997). The role of the major histocompatibility complex in bovine embryo transfer. Theriogenology **47**(1): 111-120.

Alberts, B., Bray, D. et al. (1994). *Molecular biology of the cell*. New York : Garland Publishing Inc. pp1294.

Aldridge, B. M., L. Bowen, et al. (2006). Paucity of class I MHC gene heterogeneity between individuals in the endangered Hawaiian monk seal population. Immunogenetics **58**(2-3): 203-215.

Alfonso, C. & L. Karlsson (2000). Nonclassical MHC class II molecules. Annual Reviews in Immunology **18**: 113-142.

Alfonso, C., G. S. Williams, et al. (2003). Analysis of H2-O influence on antigen presentation by B cells. Journal of Immunology **171**(5): 2331-2337.

Alfonso, C., G. S. Williams, et al. (2003). H2-O influence on antigen presentation in H2-E-expressing mice. European Journal of Immunology **33**(7): 2014-2021.

Al-Murrani, S. W. K., E. J. Glass, et al. (1994). BoLA Class I charge heterogeneity reflects the expression of more than 2 loci. Animal Genetics **25**(3): 165-172.

Amadou, C. (1999). Evolution of the Mhc class I region: the framework hypothesis. Immunogenetics **49**(4): 362-367.

Andersson, L., A. Lunden, et al. (1988). Linkage relationships in the bovine MHC region. High recombination frequency between class II subregions. Immunogenetics **27**(4): 273-280.

Andersson, L. & L. Rask (1988). Characterization of the MHC class II region in cattle - the number of DQ genes varies between haplotypes. Immunogenetics **27**(2): 110-120.

Apanius, V., D. Penn, et al. (1997). The nature of selection on the major histocompatibility complex. Critcal Reviews in Immunology **17**(2): 179-224.

Argüello, J.R. & J.A Madrigal (1999). HLA typing by Reference Strand-Mediated Conformation Analysis (RSCA). Reviews in Immunogenetics **1**(2):209-219.

Averdam, A., B. Petersen, et al. (2009). A novel system of polymorphic NK cell receptors identified in primates. Tissue Antigens **73**(5): 386-386.

Babiuk, S., B. Horseman, et al. (2007). BoLA class I allele diversity and polymorphism in a herd of cattle. Immunogenetics **59**(2): 167-176.

Baker, M. L., S. D. Melman, et al. (2009). Evolution of the opossum major histocompatibility complex: evidence for diverse alternative splice patterns and low polymorphism among class I genes. Immunology **128**(1): e418-e431.

Ballingall, K. T., S. A. Ellis, et al. (2004). The DY genes of the cattle MHC: expression and comparative analysis of an unusual class II MHC gene pair. Immunogenetics **55**(11): 748-755.

Ballingall, K. T., A. Luyai, et al. (1997). Analysis of genetic diversity at the DQA loci in African cattle: Evidence for a BoLA-DQA3 locus. Immunogenetics **46**(3): 237-244.

Ballingall, K. T., N. D. MacHugh, et al. (2001). Transcription of the unique ruminant class II major histocompatibility complex-DYA and DIB genes in dendritic cells. European Journal of Immunology **31**(1): 82-86.

Band, M., J. H. Larson, et al. (1998). A radiation hybrid map of BTA23: identification of a chromosomal rearrangement leading to separation of the cattle MHC class II subregions. Genomics **53**(3): 269-275.

Batra, T. R., A. J. Lee, et al. (1989). Class I alleles of the bovine major histocompatibility system and their association with economic traits. Journal of Dairy Science **72**(8): 2115-2124.

Beauchamp, G. K., K. Yamazaki, et al. (1988). Pre-weaning experience in the control of mating preferences by genes in the major histocompatibility complex of the mouse. Behaviour Genetics **18**(4): 537-547.

Becker, J. W. & G. N. Reeke, Jr. (1985). Three-dimensional structure of beta 2-microglobulin. Proceedings of the National Academy of Sciences of the United States of America **82**(12): 4225-4229.

Beja-Pereira, A., D. Caramelli, et al. (2006). The origin of European cattle: evidence from modern and ancient DNA. Proceedings of the National Academy of Sciences of the United States of America **103**(21): 8113-8118.

Bensaid, A., J. Naessens, et al. (1988). An immunochemical analysis of class I (BoLA) molecules on the surface of bovine cells. Immunogenetics **27**(2): 139-144.

Bernoco, D., H. A. Lewin, et al. (1991). Joint Report of the Fourth International Bovine Lymphocyte Antigen (BoLA) Workshop, East Lansing, Michigan, USA, 25 August 1990. Animal Genetics **22**(6): 477-496.

Berry, D. P., B. L. Harris, et al. (2005). Phenotypic associations between traits other than production and longevity in New Zealand dairy cattle. Journal of Dairy Science **88**(8): 2962-2974.

Birch, J., G. Codner, et al. (2008). Genomic location and characterisation of nonclassical MHC class I genes in cattle. Immunogenetics **60**(5): 267-273.

Birch, J., L. Murphy, et al. (2006). Generation and maintenance of diversity in the cattle MHC class I region. Immunogenetics **58**(8): 670-679.

Birch, J., C. D. J. Sanjuan, et al. (2008). Genomic location and characterisation of MIC genes in cattle. Immunogenetics **60**(8): 477-483.

Birnboim, H. C. & J. Doly (1979). Rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Research **7**(6): 1513-1523.

Bjorkman, P. J., M. A. Saper, et al. (1987). Structure of the human class I histocompatibility antigen, *HLA-A2*. Nature **329**(6139): 506-512.

Blasky, A., J. Karl, et al. (2008). Rapid high-resolution MHC class I genotyping of Chinese rhesus macaques by capillary reference strand-mediated conformational analysis. Immunogenetics **60**(10):575-584.

Bonhomme, M., G. G. M. Doxiadis, et al. (2008). Genomic plasticity of the immune-related Mhc class I B region in macaque species. BMC Genomics **9**:514

Booth, C. J., L. D. Warnick, et al. (2004). Effect of lameness on culling in dairy cows. Journal of Dairy Science **87**(12): 4115-4122.

Borghans, J. A., J. B. Beltman, et al. (2004). MHC polymorphism under host-pathogen coevolution. Immunogenetics **55**(11): 732-739.

Bover, K. H. (1927). Homoisotransplantation van Epidermis bei eineigen Zwillingen. Beitrage zum klinische Chirurgie **141**: 393-453.

Bradley, D. G., D. E. MacHugh, et al. (1996). Mitochondrial diversity and the origins of African and European cattle. Proceedings of the National Academy of Sciences of the United States of America **93**(10): 5131-5135.

Braud, V. M., D. S. J. Allan, et al. (1999). Functions of nonclassical MHC and non-MHC-encoded class I molecules. Current Opinion in Immunology **11**(1): 100-108.

Braud, V. M., D. S. J. Allan, et al. (1998). HLA-E binds to natural killer cell receptors CD94/NKG2A, B and C. Nature **391**(6669): 795-799.

Bray, T. C., L. Chikhi, et al. (2009). The population genetic effects of ancestry and admixture in a subdivided cattle breed. Animal Genetics **40**(4): 393-400.

Brinkmeyer-Langford, C.L., C.P. Childers, et al. (2009). A high resolution RH map of the bovine major histocompatibility complex. BMC Genomics **10**: 182

Callesen, H., T. Liboriussen, et al. (1996). Practical aspects of multiple ovulation embryo transfer in cattle. Animal Reproduction Science **42**(1-4): 215-226.

Cao, K., J. Hollenbach, et al. (2001). Analysis of the frequencies of HLA-A, B, and C alleles and haplotypes in the five major ethnic groups of the United States reveals high levels of diversity in these loci and contrasting distribution patterns in these populations. Human Immunology **62**(9): 1009-1030.

Canadian Dairy Network (2005) Minimum criteria for an official bull proof. (http://www.cdn.ca/articles.php) date accessed 20/04/10

Canadian Dairy Network (2010) Heritability estimates used for genetic evaluation in Canada. (http://www.cdn.ca/articles.php) date accessed 20/04/10

Carrington, M., G. W. Nelson, et al. (1999). HLA and HIV-1: heterozygote advantage and B*35-Cw*04 disadvantage. Science **283**(5408): 1748-1752.

Chen, J. J., J. A. Hollenbach, et al. (1999). Hardy-Weinberg testing for HLA class II (DRB1, DQA1, DQB1, and DPB1) loci in 26 human ethnic groups. Tissue Antigens **54**(6): 533-542.

Clarke, B. & D. R. Kirby (1966). Maintenance of histocompatibility polymorphisms. Nature **211**(5052): 999-1000.

Colonna, M., G. Borsellino, et al. (1993). HLA-C is the inhibitory ligand that determines dominant resistance to lysis by NK1-specific and NK2-specific natural killer cells. Proceedings of the National Academy of Sciences of the United States of America **90**(24): 12000-12004.

Comiskey, M., C. Y. Goldstein, et al. (2003). Evidence that HLA-G is the functional homolog of mouse Qa-2, the Ped gene product. Human Immunology **64**(11): 999-1004.

Crowley, M. P., A. M. Fahrer, et al. (2000). A population of murine gamma-delta T cells that recognize an inducible MHC class Ib molecule. Science **287**(5451): 314-316.

Cunningham, B. & I. Berggard (1975). Partial amino acid sequence of rabbit beta2-microglobulin. Science **187**(4181): 1079-1080.

Cunningham, E. P. (1983). Structure of dairy-cattle breeding in Western Europe and comparisons with North America. Journal of Dairy Science **66**(7): 1579-1587.

Danchin, E., V. Vitiello, et al. (2004). The major histocompatibility complex origin. Immunological Reviews **198**: 216-232.

Dausset, J. & A. Nenna (1952). Présence d`une leuco-agglutinine dans le sérum d`un cas d`agranulocytose chronique. Comptes rendus des séances de la Société de biologie et de ses filiales (Paris) **146**: 1539-1541.

Davies, C. J., J. R. Hill, et al. (2004). Major histocompatibility antigen expression on the bovine placenta: its relationship to abnormal pregnancies and retained placenta. Animal Reproduction Science **82-83**: 267-280.

Davies, C. J., I. Joosten, et al. (1994). Polymorphism of bovine MHC class I genes. Joint report of the Fifth International Bovine Lymphocyte Antigen (BoLA) Workshop, Interlaken, Switzerland, 1 August 1992. European Journal of Immunogenetics **21**(4): 239-258.

Daza-Vamenta, R., G. Glusman, et al. (2004). Genetic divergence of the rhesus macaque major histocompatibility complex. Genome Research **14**(8): 1501-1515.

de Bakker, P. I. W., G. McVean, et al. (2006). A high-resolution HLA and SNP haplotype map for disease association studies in the extended human MHC. Nature Genetics **38**(10): 1166-1172.

De Boer, R. J., J. A. Borghans, et al. (2004). Heterozygote advantage fails to explain the high degree of polymorphism of the MHC. Immunogenetics **55**(11): 725-731.

de Campos Lima, P. O., V. Levitsky, et al. (1997). Epitope-dependent selection of highly restricted or diverse T cell receptor repertoires in response to persistent infection by Epstein-Barr virus. Journal of Experimental Medicine **186**(1): 83-89.

Denniston, C. & J. F. Crow (1990). Alternative fitness models with the same allele frequency dynamics. Genetics **125**(1): 201-205.

Denzin, L. K. & P. Cresswell (1995). HLA-DM induces CLIP dissociation from MHC class II alpha beta dimers and facilitates peptide loading. Cell **82**(1): 155-165.

Denzin, L. K., C. Hammond, et al. (1996). HLA-DM interactions with intermediates in HLA-DR maturation and a role for HLA-DM in stabilizing empty HLA-DR molecules. Journal of Experimental Medicine **184**(6): 2153-2165.

Department for Environment, Food and Rural Affairs (DEFRA; 2008). The Cattle Book 2008. (http://www.defra.gov.uk/foodfarm/farmanimal/diseases/vetsurveillance/documents/cattlebook-2008.pdf) date accessed 02/04/10

Deverson, E. V., H. Wright, et al. (1991). Class-II major histocompatibility complex genes of the sheep. Animal Genetics **22**(3): 211-225.

Di Palma, F., S. D. Archibald, et al. (2002). A BAC contig of approximately 400 kb contains the classical class I major histocompatibility complex (MHC) genes of cattle. European Journal of Immunogenetics **29**(1): 65-68.

Dietz, A. B., J. C. Detilleux, et al. (1997). Genetic association of bovine lymphocyte antigen DRB3 alleles with immunological traits of Holstein cattle. Journal of Dairy Science **80**(2): 400-405.

Ditchkoff, S. S., R. L. Lochmiller, et al. (2001). Major-histocompatibility-complex-associated variation in secondary sexual traits of white-tailed deer (*Odocoileus virginianus*): Evidence for good-genes advertisement. Evolution **55**(3): 616-625.

Doherty, P. C. & R. M. Zinkernagel (1975). A biological role for the major histocompatibility antigens. Lancet **1**(7922): 1406-1409.

Du, H. J., J. S. Zheng, et al. (2010). High MHC DQB Variation and Asymmetric Allelic Distribution in the Endangered Yangtze Finless Porpoise, *Neophocaena phocaenoides asiaeorientalis*. Biochemical Genetics **48**(5-6): 433-449.

Egid, K. & J. L. Brown (1989). The major histocompatibility complex and female mating preferences in mice. Animal Behaviour **38**: 548-550.

Ellis, S. (2004). The cattle major histocompatibility complex: is it unique? Veterinary Immunology and Immunopathology **102**(1-2): 1-8.

Ellis, S. A. & K. T. Ballingall (1999). Cattle MHC: evolution in action? Immunological Reviews **167**: 159-168.

Ellis, S. A., R. E. Bontrop, et al. (2006). ISAG/IUIS-VIC Comparative MHC Nomenclature Committee report, 2005. Immunogenetics **57**(12): 953-958.

Ellis, S. A., K. A. Braem, et al. (1992). Transmembrane and cytoplasmic domain sequences demonstrate at least 2 expressed bovine MHC class-I loci. Immunogenetics **37**(1): 49-56.

Ellis, S. A., E. C. Holmes, et al. (1999). Variation in the number of expressed MHC genes in different cattle class I haplotypes. Immunogenetics **50**(5-6): 319-328.

Ellis, S. A., A. J. Martin, et al. (1995). At least 4 MHC class-I genes are transcribed in the horse- phylogenetic analysis suggests an unusual evolutionary history for the MHC in this species. European Journal of Immunogenetics **22**(3): 249-260.

Ellis, S. A., W. I. Morrison, et al. (2005). Serological and molecular diversity in the cattle MHC class I region. Immunogenetics **57**(8): 601-606.

Ellis, S. A., K. A. Staines, et al. (1996). cDNA sequence of cattle MHC class I genes transcribed in serologically defined haplotypes A18 and A31. Immunogenetics **43**(3): 156-159.

Ellis, S. A., K. A. Staines, et al. (1998). DNA typing for BoLA class I using sequence-specific primers (PCR-SSP). European Journal of Immunogenetics **25**(5): 365-370.

Ennis, P. D., A. P. Jackson, et al. (1988). Molecular cloning of bovine class I MHC cDNA. Journal of Immunology **141**(2): 642-651.

Falconer, D. S. & T. F. C. Mackay (1996). *Introduction to quantitative genetics*. Harlow, Longman Group Limited.

Fan, J.-B., M. S. Chee, et al. (2006). Highly parallel genomic assays. Nature Reviews Genetics **7**(8): 632-644.

Feder, J. N., A. Gnirke, et al. (1996). A novel MHC class I-like gene is mutated in patients with hereditary haemochromatosis. Nature Genetics **13**(4): 399-408.

Fisher, R. A. (1970). Statistical Methods for Research Workers. Edinburgh, Oliver & Boyd.

Flajnik, M. F. (2004). Immunology: another manifestation of GOD. Nature **430**(6996): 157-158.

Flajnik, M. F., Y. Ohta, et al. (1999). Two Ancient Allelic Lineages at the Single Classical Class I Locus in the *Xenopus* MHC. Journal of Immunology **163**(7): 3826-3833.

Flutter, B. & B. Gao (2004). MHC class I antigen presentation--recently trimmed and well presented. Cell Molecular Immunology **1**(1): 22-30.

Food and Agriculture Organisation of the United Nations (FAO), 2000. World Watch List for Domestic Animal Diversity, 3rd edition, edited by Scherf, B.D. (http://www.fao.org/docrep/009/x8750e/x8750e00.HTM) accessed on 01/04/10

Food and agriculture Organisation of the United Nations: Statistics website (FAOSTAT), 2010. (http://faostat.fao.org/site/569/DesktopDefault.aspx?PageID=569#ancor) accessed on 01/04/10

Food and Agriculture Organisation of the United Nations (FAO) Global Food Outlook, 2010 (http://www.fao.org/docrep/012/ak349e/ak349e00.pdf) accessed on 01/04/10.

Foote, R. H. (1999). Development of reproductive biotechnologies in domestic animals from artificial insemination to cloning: a perspective. Cloning **1**(3): 133-142.

Freeman, A. E. & G. L. Lindberg (1993). Challenges to Dairy Cattle Management: Genetic Considerations. Journal of Dairy Science **76**(10): 3143-3159.

Fries, R., R. Hediger, et al. (1986). Tentative chromosomal localization of the bovine major histocompatibility complex by in-situ hybridisation. Animal Genetics **17**(4): 287-294.

Funk, D. A. (2006). Major advances in globalization and consolidation of the artificial insemination industry. Journal of Dairy Science **89**(4): 1362-1368.

Galli, C., R. Duchi, et al. (2003). Bovine embryo technologies. Theriogenology **59**(2): 599-616.

Garber, T. L., A. L. Hughes, et al. (1993). Sequence and evolution of cattle MHC class I cDNAs- concerted evolution in cattle has not taken place. Immunogenetics **38**(1): 11-20.

Garber, T. L., A. L. Hughes, et al. (1994). Evidence for at least three transcribed BoLA class I loci. Immunogenetics **39**(4): 257-265.

Genin, E., C. Ober, et al. (2000). A robust test for assortative mating. European Journal of Human Genetics **8**(2): 119-124.

Geraghty, D. E., J. Pei, et al. (1992). Cloning and physical mapping of the HLA class I region spanning the HLA-E-to-HLA-F interval by using yeast artificial chromosomes. Proceedings of the National Academy of Sciences of the United States of America **89**(7): 2669-2673.

Glass, E. J., R. A. Oliver, et al. (2000). Duplicated DQ haplotypes increase the complexity of restriction element usage in cattle. Journal of Immunology **165**(1): 134-138.

Goldsworthy, M., J. M. Wilkinson, et al. (2000). Sample sequencing of the human major histocompatibility complex class I region identifies further new genes. Immunogenetics **51**(8-9): 615-623.

Gorer, P. A. (1936). The detection of antigenic differences in mouse erythrocytes by the employment of immune sera. British Journal of Experimental Pathology **17**:42-50.

Grisart, B., W. Coppieters et al., (2002). Positional candidate cloning of a QTL in dairy cattle: identification of a missense mutation in the Bovine *DGAT1*

gene with a major effect on milk yield and composition. Genome Research **12**: 222-231

Green, L. E., V. J. Hedges, et al. (2002). The impact of clinical lameness on the milk yield of dairy cows. Journal of Dairy Science **85**(9): 2250-2256.

Groh, V., S. Bahram, et al. (1996). Cell stress-regulated human major histocompatibility complex class I gene expressed in gastrointestinal epithelium. Proceedings of the National Academy of Sciences of the United States of America **93**(22): 12445-12450.

Groh, V., A. Steinle, et al. (1998). Recognition of stress-induced MHC molecules by intestinal epithelial gammadelta T cells. Science **279**(5357): 1737-1740.

Gussow, D., R. S. Rein, et al. (1987). Isolation, expression and the primary structure of *HLA-Cw1* and *HLA-Cw2* genes- evolutionary aspects. Immunogenetics **25**(5): 313-322.

Guzman, E., G. Taylor, et al. (2008). An MHC-restricted CD8(+) T-cell response is induced in cattle by foot-and-mouth disease virus (FMDV) infection and also following vaccination with inactivated FMDV. Journal of General Virology **89**: 667-675.

Hansen, L. B. (1999). Increased inbreeding and relationships of Holsteins - How much further should we go? Proceedings of the Western Canadian Dairy Seminar. Alberta, CA

Hardy, G. (1908). Mendelian proportions in a mixed population. Science **28**(706): 49-50

Hayes, B. J., P. M. Visscher, et al. (2003). Novel multilocus measure of linkage disequilibrium to estimate past effective population size. Genome Research **13**(4): 635-643.

Heath, W. R. & F. R. Carbone (2001). Cross-presentation in viral immunity and self-tolerance. Nature Reviews in Immunology **1**(2): 126-134.

Hedrick, P. W. (1972). Comparison of a frequency dependent model to the over dominant model. Genetics **71**(3 PART 2): S24-S25.

Hedrick, P. W. (1999). Perspective: Highly variable loci and their interpretation in evolution and conservation. Evolution **53**(2): 313-318.

Hedrick, P. W. (2002). Pathogen resistance and genetic variation at MHC loci. Evolution **56**(10): 1902-1908.

Hedrick, P. W., G. Thomson, et al. (1987). Evolutionary genetics and HLA - another classic example. Biological Journal of the Linnean Society **31**(4): 311-331.

Hedrick, P. W. K., T.J (2000). Genetics of complex polymorphisms: parasites and maintenance of the major histocompatibility complex variation. *In* Evolutionary Genetics: From Molecules to Morphology. R. S. K. Singh, C.B. KImbas, eds. (Cambridge, Cambridge University Press) pp204-234.

Hendel, H., S. Caillat-Zucman, et al. (1999). New class I and II HLA alleles strongly associated with opposite patterns of progression to AIDS. Journal of Immunology **162**(11): 6942-6946.

Henderson, M. T. (1949). A consideration of the genetic explanations of heterosis. Agronomy Journal **41**(3): 123-126.

Heringstad, B., G. Klemetsdal, et al. (2000). Selection for mastitis resistance in dairy cattle: a review with focus on the situation in the Nordic countries. Livestock Production Science **64**(2-3): 95-106.

Hess, M., T. Goldammer, et al. (1999). Physical assignment of the bovine MHC class IIa and class IIb genes. Cytogenetics and Cell Genetics **85**(3-4): 244-247.

Heyen, D.W., J.I. Weller et al. (1999) A genome wide scan for QTL influencing milk production and health traits in dairy cattle. Physiological Genomics **1**: 165-175.

Hill, A. V., C. E. Allsopp, et al. (1991). Common west African HLA antigens are associated with protection from severe malaria. Nature **352**(6336): 595-600.

Hines, H. C., F. R. Allaire, et al. (1986). Association of bovine lymphocyte antigens with milk fat percentage differences. Journal of Dairy Science **69**(12): 3148-3150.

Ho, C. S., J. K. Lunney, et al. (2009). Molecular characterization of swine leucocyte antigen class I genes in outbred pig populations. Animal Genetics **40**(4): 468-478.

Ho, C. S., G. W. Martens, et al. (2010). Swine leukocyte antigen (SLA) diversity in Sinclair and Hanford swine. Developmental and Comparative Immunology **34**(3): 250-257.

Hollenbach, J. A., G. Thomson, et al. (2001). HLA diversity, differentiation, and haplotype evolution in Mesoamerican Natives. Human Immunology **62**(4): 378-390.

Holmes, E. C., A. F. Roberts, et al. (2003). Evolution of major histocompatibility complex class I genes in Cetartiodactyls. Immunogenetics **55**(4): 193-202.

Hughes, A. L. (1995). Origin and evolution of HLA class I pseudogenes. Molecular Biology and Evolution **12**(2): 247-258.

Hughes, A. L., M. K. Hughes, et al. (1993). Contrasting roles of interallelic recombination at the HLA-A and HLA-B loci. Genetics **133**(3): 669-680.

Hughes, A. L. & M. Yeager (1998). Natural selection at major histocompatibility complex loci of vertebrates. Annual Reviews in Genetics **32**: 415-435.

Jardetzky, T. S., J. H. Brown, et al. (1996). Crystallographic analysis of endogenous peptides associated with HLA-DR1 suggests a common, polyproline II-like conformation for bound peptides. Proceedings of the National Academy of Sciences of the United States of America **93**(2): 734-738.

Jasiorowski, H., Z. Reklewski, et al. (1983). Testing of different strains of Friesian cattle in Poland I. Milk performance of F1 paternal Friesian strain crosses under intensive feeding conditions. Livestock Production Science **10**(2): 109-122.

Joosten, I., A. J. Teale, et al. (1992). Biochemical evidence of the expression of 2 major histocompatibility complex class I genes on bovine peripheral mononuclear cells. Animal Genetics **23**(2): 113-123.

Juling, K., H. Schwarzenbacher, et al. (2006). A major genetic component of BSE susceptibility. BMC Biology **4**: 33.

Kadarmideen, H. N. (2004). Genetic correlations among body condition score, somatic cell score, milk production, fertility and conformation traits in dairy cows. Animal Science Journal **79**: 191-201.

Kasahara, M. (1999). Genome dynamics of the major histocompatibility complex: insights from genome paralogy. Immunogenetics **50**(3-4): 134-145.

Kasahara, M., M. Hayashi, et al. (1996). Chromosomal localization of the proteasome Z subunit gene reveals an ancient chromosomal duplication involving the major histocompatibility complex. Proceedings of the National Academy of Sciences of the United States of America **93**(17): 9096-9101.

Kasahara, M., T. Suzuki, et al. (2004). On the origins of the adaptive immune system: novel insights from invertebrates and cold-blooded vertebrates. Trends in Immunology **25**(2): 105-111.

Kasahara, M., Y. Watanabe, et al. (2002). A family of MHC class I-like genes located in the vicinity of the mouse leukocyte receptor complex. Proceedings of the National Academy of Sciences of the United States of America **99**(21): 13687-13692.

Katsanis, N., J. Fitzgibbon, et al. (1996). Paralogy mapping: Identification of a region in the human MHC triplicated onto human chromosomes 1 and 9 allows the prediction and isolation of novel PBX and NOTCH loci. Genomics **35**(1): 101-108.

Kaufman, J. (2000). The simple chicken major histocompatibility complex: life and death in the face of pathogens and vaccines. Philosophical Transactions of the Royal Society of London Series B-Biological Sciences **355**(1400): 1077-1084.

Kaufman, J., S. Milne, et al. (1999). The chicken B locus is a minimal essential major histocompatibility complex. Nature **401**(6756): 923-925.

Kelley, J., L. Walter, et al. (2005). Comparative genomics of major histocompatibility complexes. Immunogenetics **56**(10): 683-695.

Kennedy, L.J., R. Ryvar, et al. (2003) Resolution of complex feline leukocyte antigen *DRB* loci by reference strand-mediated conformational analysis (RSCA). Tissue Antigens **62**(4):313-323.

Kennedy, L.J., S. Quarmby, et al. (2005). High-Resolution Characterization of the Canine *DLA-DRB1* Locus Using Reference Strand-Mediated Conformational Analysis. Journal of Heredity **96**(7):836-842.

Kiepiela, P., A. J. Leslie, et al. (2004). Dominant influence of HLA-B in mediating the potential co-evolution of HIV and HLA. Nature **432**(7018): 769-774.

Kimura, M. & T. Ohta (1969). The average number of generations until extinction of an individual mutant gene in a finite population. Genetics **63**(3): 701-709.

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

Knapp, L. A., J. C. Ha, et al. (1996). Parental MHC antigen sharing and pregnancy wastage in captive pigtailed macaques. Journal of Reproductive Immunology **32**(1): 73-88.

Komlos, L., R. Zamir, et al. (1977). Common HLA antigens in couples with repeated abortions. Clinical Immunology and Immunopathology **7**(3): 330-335.

Kristensen, T. N. & A. C. Sorensen (2005). Inbreeding - lessons from animal breeding, evolutionary biology and conservation genetics. Animal Science **80**: 121-133.

Kropshofer, H., A. B. Vogt, et al. (1998). A role for HLA-DO as a co-chaperone of HLA-DM in peptide loading of MHC class II molecules. EMBO Journal **17**(11): 2971-2981.

Kulski, J. K., T. Shiina, et al. (2002). Comparative genomic analysis of the MHC: the evolution of class I duplication blocks, diversity and complexity from shark to man. Immunological Reviews **190**(1): 95-122.

Lebron, J. A., M. J. Bennett, et al. (1998). Crystal structure of the hemochromatosis protein HFE and characterization of its interaction with transferrin receptor. Cell **93**(1): 111-123.

Lee, L. A., J. D. Ferguson, et al. (1989). Effect of disease on days open assessed by survival analysis. Journal of Dairy Science **72**(4): 1020-1026.

Lenz, T.L., C. Eizaguirre, et al. (2009) RSCA genotyping of MHC for high-throughput evolutionary studies in the model organism three-spined stickleback *Gasterosteus aculeatus*. BMC Evolutionary Biology **9**:57.

Little, C. C. & E. E. Tyzzer (1916). Further experimental studies on the inheritance of susceptibility to a transplantable tumour, carcinoma (JWA) of the Japanese waltzing mouse. Journal of Medical Research **33**: 393-453.

Loftus, R. T., D. E. MacHugh, et al. (1994). Evidence for two independent domestications of cattle. Proceedings of the National Academy of Sciences of the United States of America **91**(7): 2757-2761.

Lohuis, M. M. (1995). Potential benefits of bovine embryo manipulation technologies to genetic improvement programs. Theriogenology **43**(1): 51-60.

Macdonald, I. K., M. Harkiolaki, et al. (2010). MHC Class I Bound to an Immunodominant Theileria parva Epitope Demonstrates Unconventional Presentation to T Cell Receptors. PLoS Pathogens **In press**.

Marello, K. L., A. Gallagher, et al. (1995). Expression of multiple DQB genes in Bos indicus cattle. Animal Genetics **26**(5): 345-349.

Matsumura, M., D. H. Fremont, et al. (1992). Emerging principles for the recognition of peptide antigens by MHC class I molecules. Science **257**(5072): 927-934.

Matukumalli, L. K., C. T. Lawley, et al. (2009). Development and Characterization of a High Density SNP Genotyping Assay for Cattle. PLoS ONE **4**(4): e5350.

McMichael, A. (1978). HLA restriction of human cytotoxic T-lymphocytes specific for influenza virus- poor recognition of virus associated with HLA-A2. Journal of Experimental Medicine **148**(6): 1458-1467.

McShane, R. D., D. S. Gallagher, et al. (2001). Physical localization and order of genes in the class I region of the bovine MHC. Animal Genetics **32**(5): 235-239.

Mejdell, C. M., O. Lie, et al. (1994). Association of major histocompatibility complex antigens (BoLA-A) with AI bull progeny test results for mastitis, ketosis and fertility in Norwegian cattle. Animal Genetics **25**(2): 99-104.

Miltiadou, D., K. T. Ballingall, et al. (2005). Haplotype characterization of transcribed ovine major histocompatibility complex (MHC) class I genes. Immunogenetics **57**(7): 499-509.

Moore, K. & W. W. Thatcher (2006). Major advances associated with reproduction in dairy cattle. Journal of Dairy Science **89**(4): 1254-1266.

Mori, M., P. G. Beatty, et al. (1997). HLA gene and haplotype frequencies in the North American population - The National Marrow Donor Program Donor Registry. Transplantation **64**(7): 1017-1027.

Morrison, W. I. (1996). Influence of host and parasite genotypes on immunological control of Theileria parasites. Parasitology **112**: S53-S66.

Murphy, L. (2004) 'Variations in immune response as determined by MHC class I polymorphism in cattle'. PhD thesis, University of Edinburgh.

Nejentsev, S., J. M. M. Howson, et al. (2007). Localization of type 1 diabetes susceptibility to the MHC class I genes HLA-B and HLA-A. Nature **450**(7171): 887-U819.

Nicholas, F. W. (1996). Genetic improvement through reproductive technology. Animal Reproduction Science **42**(1-4): 205-214.

Nielsen, H. M., L. G. Christensen, et al. (2006). Definition of sustainable breeding goals in dairy cattle. Proceedings of the 8th World Congress on Genetics Applied to Livestock Production, Belo Horizonte, Minas Gerais, Brazil, 13-18 August, 2006: 01-36.

Norman, H. D., T. J. Lawlor, et al. (2004). Performance of Holstein clones in the United States. Journal of Dairy Science **87**(3): 729-738.

Ober, C. (1995). Current topic - HLA and reproduction - lessons from studies in the Hutterites. Placenta **16**(7): 569-577.

Ober, C., L. R. Weitkamp, et al. (1997). HLA and mate choice in humans. American Journal of Human Genetics **61**(3): 497-504.

O'Brien, S. J., M. E. Roelke, et al. (1985). Genetic basis for species vulnerability in the cheetah. Science **227**(4693): 1428-1434.

Oddgeirsson, O., S. P. Simpson, et al. (1988). Relationship between the bovine major histocompatibility complex (BoLA), erythrocyte markers and susceptibility to mastitis in Icelandic cattle. Animal Genetics **19**(1): 11-16.

Ohta, T. (1999). Effect of gene conversion on polymorphic patterns at major histocompatibility complex loci. Immunological Reviews **167**: 319-325.

Olsen, K. H., M. Grahn, et al. (2002). Influence of mhc on sibling discrimination in Arctic char, *Salvelinus alpinus (L.*). Journal of Chemical Ecology **28**(4): 783-795.

Olsen, K. H., M. Grahn, et al. (1998). MHC and kin discrimination in juvenile Arctic charr, *Salvelinus alpinus (L.)*. Animal Behaviour **56**: 319-327.

Otting, N., A. J. M. de Vos-Rouweler, et al. (2007). MHC class I A region diversity and polymorphism in macaque species. Immunogenetics **59**(5): 367-375.

Otting, N., C. I. C. Heijmans, et al. (2005). Unparalleled complexity of the MHC class I region in rhesus macaques. Proceedings of the National Academy of Sciences of the United States of America **102**(5): 1626-1631.

Otting, N., C. M. C. Heijmans, et al. (2008). A snapshot of the Mamu-B genes and their allelic repertoire in rhesus macaques of Chinese origin. Immunogenetics **60**(9): 507-514.

Parham, P., E. J. Adams, et al. (1995). The origins of HLA-A,B,C polymorphism. Immunological Reviews **143**: 141-180.

Penn, D. & W. Potts (1998). MHC-disassortative mating preferences reversed by cross-fostering. Proceedings of the Royal Society of London Series B-Biological Sciences **265**(1403): 1299-1306.

Perez-Pardal, L., L. J. Royo, et al. (2010). Multiple paternal origins of domestic cattle revealed by Y-specific interspersed multilocus microsatellites. Heredity **105:** 511–51

Pichowski, J. S., S. A. Ellis, et al. (1996). Sequence of two cattle MHC class I cDNAs associated with BoLA A10 specificity. Immunogenetics **43**(4): 253-254.

Piertney, S. B. & M. K. Oliver (2006). The evolutionary ecology of the major histocompatibility complex. Heredity **96**(1): 7-21.

Potts, W. K., C. J. Manning, et al. (1991). Mating patterns in semi-natural populations of mice influenced by MHC genotype. Nature **352**(6336): 619-621.

Potts, W. K. & E. K. Wakeland (1993). Evolution of MHC genetic diversity - a tale of incest, pestilence and sexual preference. Trends in Genetics **9**(12): 408-412.

Purcell, M. K., A. L. Murray, et al. (2008). Decreased Mortality of Lake Michigan Chinook Salmon after Bacterial Kidney Disease Challenge: Evidence for Pathogen-Driven Selection? Journal of Aquatic Animal Health **20**(4): 225-235.

Rammensee, H. G., T. Friede, et al. (1995). MHC ligands and peptide motifs: first listing. Immunogenetics **41**(4): 178-228.

Renard, C., M. Vaiman, et al. (2001). Sequence of the pig major histocompatibility region containing the classical class I genes. Immunogenetics **53**(6): 490-500.

Rhodes, D. A. & J. Trowsdale (1999). Genetics and molecular genetics of the MHC. Reviews in Immunogenetics **1**(1): 21-31.

Roberts, S. C., L. M. Gosling, et al. (2008). MHC-correlated odour preferences in humans and the use of oral contraceptives. Proceedings of the Royal Society B-Biological Sciences **275**(1652): 2715-2722.

Robinson, J., M. J. Waller, et al. (2005). IPD - the immuno polymorphism database. Nucleic Acids Research **33**: D523-D526.

Rocha, J. L., J. F. Baker, et al. (1992). Statistical associations between restriction fragment length polymorphisms and quantitative traits in beef cattle. Journal of Animal Science. **70**(11): 3360-3370.

Rupp, R., A. Hernandez, et al. (2007). Association of bovine leukocyte antigen (BoLA) DRB3.2 with immune response, mastitis, and production and type traits in Canadian Holsteins. Journal of Dairy Science **90**(2): 1029-1038.

Sambrook, J., E. F. Fritsch, et al. (1989). Molecular cloning a laboratory manual second edition vols. 1, 2 and 3 (New York, Cold Spring Harbor Laboratory Press)

Sanderson, F., M. J. Kleijmeer, et al. (1994). Accumulation of HLA-DM, a regulator of antigen presentation, in MHC class II compartments. Science **266**(5190): 1566-1569.

Satta, Y. & N. Takahata (1997). Intragenic recombination and MHC diversity. Hereditas (Lund) **127**(1-2): 168.

Schacter, B., A. Muir, et al. (1979). HLA-A, B compatibility in parents of offspring with neural tube defects or couples experiencing involuntary fetal wastage. Lancet **313**(8120): 796-799.

Schaeffer, L. R. (2006). Strategy for applying genome-wide selection in dairy cattle. Journal of Animal Breeding and Genetics **123**(4): 218-223.

Sewalem, A., G. J. Kistemaker, et al. (2004). Analysis of the relationship between type traits and functional survival in Canadian Holsteins using a Weibull proportional hazards model. Journal of Dairy Science **87**(11): 3938-3946.

Sharif, S., B. A. Mallard, et al. (1998). Associations of the bovine major histocompatibility complex DRB3 (BoLA-DRB3) alleles with occurrence of

disease and milk somatic cell score in Canadian dairy cattle. Animal Genetics **29**(3): 185-193.

Sher, A. & R. Ahmed (1995). Regulation of innate and adaptive responses to infection: new avenues and angles. Current Opinion in Immunology **7**(4): 471-473.

Shook, G. E. (2006). Major advances in determining appropriate selection goals. Journal of Dairy Science **89**(4): 1349-1361.

Shum, B., D. Avila, et al. (1993). Isolation of a classical MHC class I cDNA from an amphibian. Evidence for only one class I locus in the Xenopus MHC. Journal of Immunology **151**(10): 5376-5386.

Simm, G. (1998). *Genetic improvement of cattle and sheep*. Ipswich, Farming Press.

Skow, L. C., J. E. Womack, et al. (1988). Synteny mapping of the genes for 21-steroid hydroxylase, alpha- a crystallin and class I bovine leukocyte antigen in cattle. DNA-A Journal of Molecular & Cellular Biology **7**(3): 143-149.

Slade, R. W. (1992). Limited MHC polymorphism in the southern elephant seal: implications for MHC evolution and marine mammal population biology. Proceedings of the Royal Society B-Biological Sciences **249**(1325): 163-171.

Sloan, V. S., P. Cameron, et al. (1995). Mediation by HLA-DM of dissociation of peptides from HLA-DR. Nature **375**(6534): 802-806.

Smith, D. M., J. K. Lunney, et al. (2005). Nomenclature for factors of the SLA class-I system, 2004. Tissue Antigens **65**(2): 136-149.

Smith, K. B. (2000). 'Expression of major histocompatibility complex class I genes in cattle'. PhD thesis, School of Animal and Microbial Sciences. Reading, University of Reading.

Smith, M.Z., C.J. Dale, et al. (2005) Analysis of pigtail macaque major histocompatibility complex class I molecules presenting immunodominant simian immunodeficiency virus epitopes. Journal of Virology **79**(2): 684-

695.Snary, D., C. J. Barnstable, et al. (1977). Molecular structure of human histocompatibility antigens - HLA-C series. European Journal of Immunology **7**(8): 580-585.

Snell, G. D. & R. B. Jackson (1958). Histocompatibility genes of the mouse. II. Production and analysis of isogenic resistant lines. Journal of the National Cancer Institute **21**(5): 843-877.

Solache, A., C. L. Morgan, et al. (1999). Identification of three HLA-A*0201-restricted cytotoxic T cell epitopes in the cytomegalovirus protein pp65 that are conserved between eight strains of the virus. Journal of Immunology **163**(10): 5512-5518.

Sordillo, L. M., K. Shafer-Weaver, et al. (1997). Immunobiology of the Mammary Gland. Journal of Dairy Science **80**(8): 1851-1865.

Spooner, R. L., R. A. Oliver, et al. (1979). Analysis of alloantisera against bovine lymphocytes - Joint report of the 1st International bovine lymphocyte antigen (BoLA) workshop. Animal Blood Groups and Biochemical Genetics **10**(2): 63-86.

Stear, M. J., G. T. Innocent, et al. (2005). The evolution and maintenance of polymorphism in the major histocompatibility complex. Veterinary Immunology and Immunopathology **108**(1-2): 53-57.

Stear, M. J., M. J. Newman, et al. (1982). Two closely linked loci and one apparently independent locus code for bovine lymphocyte antigens. Tissue Antigens **20**(4): 289-299.

Stear, M. J., T. S. Pokorny, et al. (1990). Genetic analysis of the antigens defined at the third international BoLA workshop. Journal of Immunogenetics **17**(1-2): 21-28.

Stone, R. T. & N. E. Muggli-cockett (1990). Partial nucleotide sequence of a novel bovine major histocompatibility complex class II beta chain gene, *BoLA-DIB*. Animal Genetics **21**(4): 353-360.

Strandberg, E. & G. E. Shook (1989). Genetic and economic responses to breeding programs that consider mastitis. Journal of Dairy Science **72**(8): 2136-2142.

Strauss, S. (2010). Biotech breeding goes bovine. Nature Biotechnology **28**(6): 540-543.

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics **123**(3): 585-595.

Takahata, N. (1995). MHC diversity and selection. Immunological Reviews **143**: 225-247.

Takahata, N. & M. Nei (1990). Allelic genealogy under overdominant and frequency-dependent sleection and polymorphism of major histocompatibility complex loci. Genetics **124**(4): 967-978.

Takeshima, S., Y. Matsumoto, et al. (2008). Evidence for cattle major histocompatibility complex (BoLA) class II *DQA1* gene heterozygote advantage against clinical mastitis caused by Streptococci and Escherichia species. Tissue Antigens **72**(6): 525-531.

Takeshima, S., N. Saitou, et al. (2003). The diversity of bovine MHC class II *DRB3* genes in Japanese Black, Japanese Shorthorn, Jersey and Holstein cattle in Japan. Gene **316**: 111-118.

Tallmadge, R. L., J. A. Campbell, et al. (2010). Analysis of MHC class I genes across horse MHC haplotypes. Immunogenetics **62**(3): 159-172.

Tamura, K. & M. Nei (1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Molecular Biology and Evolution **10**(3): 512-526.

Tanaka-Matsuda, M., A. Ando, et al. (2009). Difference in number of loci of swine leukocyte antigen classical class I genes among haplotypes. Genomics **93**(3): 261-273.

Tanaka-Takahashi, Y., M. Yasunami, et al. (2007). Reference strand-mediated conformation analysis-based typing of multiple alleles in the rhesus macaque MHC class I *Mamu-A* and *Mamu-B* loci. Electrophoresis **28**(6):918-924.

Teutsch, M. R., J. E. Beever, et al. (1989). Linkage of complement factor B gene to the bovine major histocompatibility complex. Animal Genetics **19**(4): 427.

The Bovine Genome, Sequencing and. Analysis., Consortium, C. G. Elsik, et al. (2009). The Genome Sequence of Taurine Cattle: A Window to Ruminant Biology and Evolution. Science **324**(5926): 522-528.

The Bovine HapMap Consortium, R. A. Gibbs, et al. (2009). Genome-Wide Survey of SNP Variation Uncovers the Genetic Structure of Cattle Breeds. Science **324**(5926): 528-532.

The MHC Sequencing Consortium (1999). Complete sequence and gene map of a human major histocompatibility complex. Nature **401**(6756): 921-923.

Thornhill, R., S. W. Gangestad, et al. (2003). Major histocompatibility complex genes, symmetry, and body scent attractiveness in men and women. Behavioural Ecology **14**(5): 668-678.

Townsend, A. & H. Bodmer (1989). Antigen recognition by class I-restricted T lymphocytes. Annual Reviews in Immunology **7**: 601-624.

Townsend, A., Öhlén, C. et al. (1989b). Association of class I major histocompatibility heavy and light chains induced by viral peptides. Nature **340**(6233): 443-448

Toye, P. G., N. D. Machugh, et al. (1990). Transfection into mouse L-cells of genes encoding 2 serologically and functionally distinct bovine class I MHC molecules from a MHC homozygous animal- evidence for a 2nd class I locus in cattle. Immunology **70**(1): 20-26.

Traherne, J. A., M. Martin, et al. (2010). Mechanisms of copy number variation and hybrid gene formation in the *KIR* immune gene complex. Human Molecular Genetics **19**(5): 737-751.

Trowsdale, J., V. Groves, et al. (1989). Limited MHC polymorphism in whales. Immunogenetics **29**(1): 19-24.

Trowsdale, J. & P. Parham (2004). Mini-review: defense strategies and immunity-related genes. European Journal of Immunology **34**(1): 7-17.

Tu, B., S. J. Mack, et al. (2007). *HLA-A, -B, -C, -DRB1* allele and haplotype frequencies in an African American population. Tissue Antigens **69**(1): 73-85.

Vallejo, R.L., Y. L. Li, et al. (2003) Genetic diversity and background linkage disequilibrium in the North American Holstein Cattle population. Journal of Dairy Science **86**: 4137-4147

Van Bressem, M. F., K. Van Waerebeek, et al. (1999). A review of virus infections of cataceans and the potential impact of morbilliviruses, poxviruses and papillomaviruses on host population dynamics. Diseases of Aquatic Organisms **38**(1): 53-65.

van de Goor, L. H., H. Panneman, et al. (2009). A proposal for standardization in forensic bovine DNA typing: allele nomenclature of 16 cattle-specific short tandem repeat loci. Animal Genetics **40**(5): 630-636.

Van Doormaal, B. J. (1997) Genetic evaluation of dairy cattle in Canada. (http://www.cdn.ca/articles.php) date accessed 20/04/10

Van Doormal, B.J., Miglior, F. et al., (2005) Interbull bulletin 33; Proceedings of the Interbull meeting, June 2-4, 2005, UPPSALA, Sweden: Genetic diversification of the Holstein breed in Canada and Internationally (http://www.interbull.slu.se/bulletins/bulletin33/Van%20Doormal.pdf) date accessed 27/03/10

Vance, R. E., J. R. Kraft, et al. (1998). Mouse CD94/NKG2A is a natural killer cell receptor for the nonclassical major histocompatibility complex (MHC) class I molecule Qa-1(b). Journal of Experimental Medicine **188**(10): 1841-1848.

Vanderpoel, J. J., M. A. M. Groenen, et al. (1990). The nucleotide sequence of the bovine MHC class II alpha genes- DRA, DQA and DYA. Immunogenetics **31**(1): 29-36.

VanRaden, P. M., C. P. v. Tassell, et al. (2009). Genomic data and cooperation result in faster progress. ICAR Technical Series(13): 341-346. *In:* Identification, Breeding, Production, Health and Recording of Farm Animals. Proceedings of the 36th ICAR Biennial Session, Niagara Falls,

USA, 16-20 June, 2008 (proceedings available online at: http://www.icar.org/pages/Publications/technical_series.htm)

Veerkamp, R. F., E. P. C. Koenen, et al. (2001). Genetic correlations among body condition score, yield, and fertility in first-parity cows estimated by random regression models. Journal of Dairy Science **84**(10): 2327-2335.

Von Schantz, T., H. Wittzell, et al. (1997). Mate choice, male condition-dependent ornamentation and MHC in the pheasant. Hereditas **127**(1-2): 133-140.

Von Schantz, T., H. Wittzell, et al. (1996). MHC genotype and male ornamentation: Genetic evidence for the Hamilton-Zuk model. Proceedings of the Royal Society of London Series B-Biological Sciences **263**(1368): 265-271.

Waheed, A., S. Parkkila, et al. (1999). Association of HFE protein with transferrin receptor in crypt enterocytes of human duodenum. Proceedings of the National Academy of Sciences of the United States of America **96**(4): 1579-1584.

Wall, E., S. Brotherstone, et al. (2005). Impact of Nonadditive Genetic Effects in the Estimation of Breeding Values for Fertility and Correlated Traits. Journal of Dairy Science **88**(1): 376-385.

Wall, E., S. Brotherstone, et al. (2003). Genetic evaluation of fertility using direct and correlated traits. Journal of Dairy Science **86**(12): 4093-4102.

Warner, C. M., M. S. Brownell, et al. (1988). Why aren't embryos immunologically rejected by their mothers? Biology of Reproduction **38**(1): 17-29.

Warner, C. M., S. O. Gollnick, et al. (1987). Linkage of the preimplantation-embryo-development (*Ped*) gene to the mouse major histocompatibility complex (MHC). Biology of Reproduction **36**(3): 606-610.

Wedekind, C. & S. Furi (1997). Body odour preferences in men and women: do they aim for specific MHC combinations or simply heterozygosity? Proceedings of the Royal Society of London Series B-Biological Sciences **264**(1387): 1471-1479.

Wedekind, C., T. Seebeck, et al. (1995). MHC-dependent mate preferences in humans. Proceedings of the Royal Society of London Series B-Biological Sciences **260**(1359): 245-249.

Weigel, K. A., A. E. Freeman, et al. (1990). Association of class I bovine lymphocyte antigen complex alleles with health and production traits in dairy cattle. Journal of Dairy Science **73**(9): 2538-2546.

Wiener, P., D. Burton & J.L. Williams (2004) Breed relationships and definition in British cattle: a genetic analysis. Heredity **93**: 597–602.

Wiggans, G. R., T. S. Sonstegard, et al. (2009). Selection of single-nucleotide polymorphisms and quality of genotypes used in genomic evaluation of dairy cattle in the United States and Canada. Journal of Dairy Science **92**(7): 3431-3436.

Winters, M. (2009) Breeding+ Breeding Improvement Programme. (http://www.dairyco.org.uk/library/farming-info-centre/breeding/breedingplus-folder.aspx) date accessed 18/04/10

Yamazaki, K., G. K. Beauchamp, et al. (1988). Familial imprinting determines H-2 selective mating preferences. Science **240**(4857): 1331-1332.

Yamazaki, K., E. A. Boyse, et al. (1976). Control of mating preferences in mice by genes in major histocompatibility complex. Journal of Experimental Medicine **144**(5): 1324-1335.

Yamazaki, K., M. Yamaguchi, et al. (1979). Recognition among mice - evidence from the use of a Y-Maze differentially scented by congenic mice of different major histocompatibility types. Journal of Experimental Medicine **150**(4): 755-760.

Yuhki, N. & S. J. O'Brien (1990). DNA variation of the mammalian major histocompatibility complex reflects genomic diversity and population history. Proceedings of the National Academy of Sciences of the United States of America **87**(2): 836-840.

Zarnecki, A., J. Jamrozik, et al. (1991). Comparison of 10 Friesian strains in Poland for yield traits from 1st three parities. Journal of Dairy Science **74**(7): 2303-2308.

Zarnecki, A. & M. Stolzman (1986). A preliminary comparison of later lactation yields of different Friesian strains in Poland. 3rd World Congress on Genetics applied to Livestock Production, Lincoln, Nebraska, USA, July 16-22, 1986. IX. Breeding programs for dairy and beef cattle, water buffalo, sheep, and goats. 112-114.

Zhang, C., A. Anderson, et al. (1998). Structural principles that govern the peptide-binding motifs of class I MHC molecules. Journal of Molecular Biology **281**(5): 929-947.

Zhou, H., J. G. H. Hickford, et al. (2006). A two-step procedure for extracting genomic DNA from dried blood spots on filter paper for polymerase chain reaction amplification. Analytical Biochemistry **354**(1): 159-161.