Characterization of the ovine Major Histocompatibility Complex (MHC) class I genes

Despoina Miltiadou

A thesis submitted to the University of Edinburgh
for the degree of Doctor of Philosophy
Edinburgh, 2006

DECLARATION

I hereby declare that the work presented in this thesis is the product of my own efforts, and has not been submitted in any previous application for a degree. The work on which it is based is my own except where stated in the text and in the acknowledgement section.

Despoina Miltiadou

To the memory of my mother

Acknowledgments

I would like to express my gratitude to my principal supervisor, Prof. Declan Mckeever, for trusting me this project and ensuring the funding of such a costly research. I would also like to thank him for the useful discussions throughout the duration of the experimental work and the invaluable suggestions during the writing of the thesis.

I am sincerely grateful to my supervisor Dr. Keith Ballingall for his constant support and advice, for his excellent demonstrations of laboratory techniques, as well as for his useful corrections of the present thesis.

I acknowledge the encouragement, support and advice of Dr. George Russell as well as the advice of my external supervisor, Dr. Shirley Ellis, which were invaluable inputs to the present study based on many years of experience on MHC research.

I am indebted to the staff of the Moredun Research Institute and Roslin Institute for their contribution to the present work. In particular, many thanks go to Dr. Mara Rocchi for advice on techniques and for the FACS analysis presented in chapter 7. I also owe special thanks to Dr. Zhong-Wei Chai for providing me with the data from the *Theilleria* expression experiments, as well as for his friendship and excellent collaboration both in the lab and in our shared office. I would also like to thank the staff of the sequencing services of Moredun and Roslin institutes for sequencing large batches of samples for the needs of this study. I am also deeply grateful to the Moredun PhD student's advisor, Dr. John Huntley, for his sincere encouragement and support during my work at Moredun Research Institute.

I acknowledge the partial financial support of the Hellenic Scholarship Foundation (IKY) along with the help and kind assessment on behalf of the IKY of Prof. Emmanouil Rogdakis.

Finally, I extend my gratitude to my husband and love Ouranios Tzamaloukas for his countless support, encouragement and patience, without which I would have never been able to accomplish such a difficult task.

TABLE OF CONTENTS

LIST OF FIGURES

Titles ..... **pages**

Chapter 5

Chapter 6

Chapter 7

Chapter 8

LIST OF TABLES

Abstract

The ovine major histocompatibility complex (MHC) remains poorly characterized compared with those of other livestock species. To develop cellular and molecular tools to support development of vaccines against intracellular pathogens of sheep, a molecular genetic analysis of four distinct ovine MHC haplotypes carried by two heterozygous Blackface rams (501 and 504) was conducted. A total of 17 novel sequences was identified, 11 of which were obtained full length. Initially, using a combination of reverse transcriptase – polymerase chain reaction (RT-PCR), Single Stranded Conformational Polymorphism (SSCP) and sequence analysis, six distinct SSCP patterns were identified from ram 501 and eight from ram 504 each corresponding to a different sequence (chapter 3). SSCP patterns and their corresponding sequences D5, D3, C7, E2, F8 and F12 accounted for 86% of 360 clones analysed from both rams, suggesting that they encode the mainly expressed classical class I molecules, while F7 and G12 represented only 0.56%, indicating they are putative low expressed class I like genes or pseudogenes. Genotyping of 17 offspring from both rams using a novel sequence specific PCR method, allowed the segregation of sequences within the four haplotypes (chapter 4). An additional sequence C7b was identified during genotyping, indicating that at least 6 class I genes are transcribed in a single haplotype. F7 and G12 were shared between all four haplotypes. Full length transcripts for ten of the sequences identified in chapter 3 were obtained by RT-PCR, all having a characteristic MHC gene structure (chapter 5). During full length amplification two additional sequences, N1 and N2, were identified with characteristics consistent with non classical class I loci. Phylogenetic analysis using the identified transcripts and published ovine, bovine and other ruminant class I sequences belonging to the *Bovidae* family indicated that there are at least six ovine MHC class I loci (chapter 6). Sequences N1 and N2 are closer to the non classical bovine MHC class I sequence HD15 than to the remaining ovine class I transcripts. Seven out of eight full length transcripts subcloned into a mammalian expression vector expressed detectable class I cell surface glycoproteins in COS-7 cells (chapter 7). The combination of phylogenetic analysis, haplotype, transcription and expression data suggest that there are at least four distinct polymorphic ovine MHC class I loci, three of which appear to be expressed in a number of combinations in individual haplotypes, a couple of non polymorphic poorly transcribed class I like sequences and at least one additional diverged non classical class I locus (chapter 8). Similarities and differences of the ovine MHC class I region with that of other species are discussed. Using the data generated here, an MHC defined sheep flock, which includes animals homozygous for each of the four MHC haplotypes is currently under development. The MHC defined resource population, along with the transfected cell lines expressing each of the full length ovine MHC class I sequences, comprise tools for immunization and disease association experiments studying the protective immunity to intracellular pathogens of sheep.

Chapter 1. Introduction and Literature Review

1.1 Introduction

1.1.1 The involvement of the Major Histocompatibility Complex (MHC) in immune responses

The immune system of vertebrates has traditionally been divided into innate and adaptive components (Fig. 1.1). Innate or non-specific immunity provides the early lines of defence and induces adaptive immunity (reviewed by Beutler B. 2004; Fearon and Locksley, 1996; Medzhitov and Janeway, 2000). The skin, the epithelial cells of the airways and gastro-intestinal tract and the mucosal layer provide the first line of defence (reviewed by Podolsky D. 1999), while phagocytosis (reviewed by Aderem and Underhill, 1999), assisted by complement (Frank and Fries, 1991), the inflammatory response (Svanborg et al. 1999), antimicrobial substances (reviewed by Zasloff M. 1992) and natural killer (NK) cells comprise the second line of innate defences (reviewed by Biron et al. 1999; French and Yokoyama, 2003; Hamerman et al. 2005).

Non-specific Defence Mechanisms (Innate Immunity)		**Specific Defence Mechanisms (Adaptive Immunity)**
First line of defence	Second line of defence	Third line of defence
• Skin • Mucous membranes • Secretions of skin and mucous membranes	• Phagocytic white blood cells • The inflammatory response • Antimicrobial substances • NK cells	• Lymphocytes • Antibodies

Figure 1.1: An overview of immune responses

The latter effector mechanisms of innate immunity are activated immediately after infection, but they are not specific to the pathogen. Rather, they arise from a limited number (several hundred) of genetically predetermined receptors of the innate immune system. These receptors have evolved by natural selection to recognize structures highly conserved in groups of microrganisms encountered by the immune system during evolution (reviewed by Janeway and Medzhitov, 2002). In contrast, adaptive or specific immunity provides a slower response that is highly specific to a particular pathogen and enables the body to recognize and develop

immunological memory against almost any pathogen even if it has not previously been encountered (Janeway C. 1993).

Protein antigens that are encountered by the immune system for the first time are taken up by an antigen presenting cell (APC). Dendritic cells (DC) are the most efficient APCs (reviewed by Banchereau and Steinman, 1998; Heath et al. 2004; Makala and Nagasawa, 2001), although macrophages (Debrick et al. 1991; Janeway C. 1993; Rock et al. 1993) and possibly B cells (Ke and Kapp, 1996; Parker D. 1993) may also prime an immune response. Immature DC capture antigens at the peripheral tissues by receptor-mediated antigen uptake and macropinocytosis (Sallusto et al. 1995). This is a key event that induces maturation and mobilization of DC, which subsequently migrate to the lymphoid organs (Banchereau and Steinman, 1998). There, they present processed antigenic peptides to distinct subgroups of T lymphocytes, either helper or cytotoxic T cells. The antigenic peptides are displayed to T cells by specialized cell surface glycoproteins, encoded within a genetic region known as the Major Histocompatibility Complex (MHC) (Zinkernagel and Doherty, 1974). In the recognition phase, two other cell surface proteins, CD4 and CD8 (CD denotes 'cluster of differentiation') for helper and cytotoxic T cells respectively act as co-receptors by binding to different classes of MHC molecules (Miceli and Parnes, 1993; Fig. 1.2). $CD4^+$ T helper cells bind to MHC class II molecules, whereas $CD8^+$ T cells bind to class I molecules.

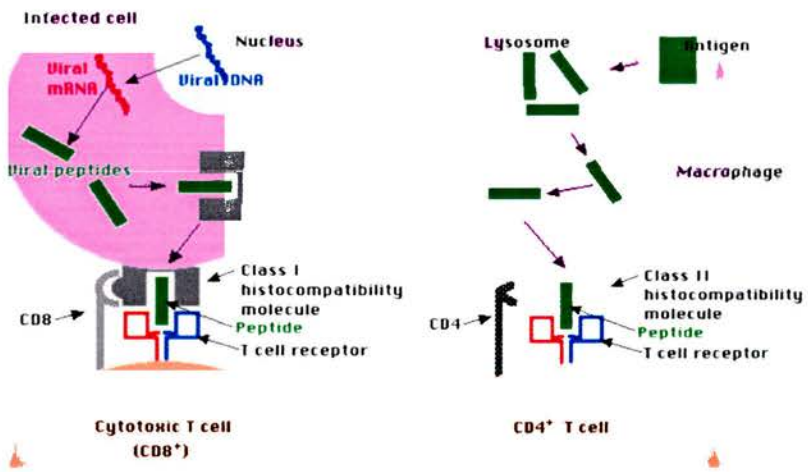


Figure 1.2: Overview of MHC–restricted T cell recognition (Picture obtained from http://www.users.rcn.com).

However, recognition of the antigenic peptide-MHC complex alone is not enough to activate T cells that have not previously encountered an antigen, known as naïve T cells. A second signal provided by innate immunity is required to induce T cell proliferation and differentiation into effector cells, a phenomenon initially called "the two signal theory for self-non self discrimination" (Bretscher and Cohn, 1970; Fig. 1.3). More recently, it was discovered that the second signal is produced by cell surface proteins known as co-stimulators (van der Merwe et al. 1997; Ye et al. 1997). Microbes and soluble factors known as cytokines, which are produced during the innate immune responses, activate APC to express such co-stimulatory molecules (reviewed by Fearon and Locksley, 1996). Two well-characterized co-stimulators are B7.1 (CD80) and B7.2 (CD86), which are recognized by the CD28 protein or the cytotoxic T lymphocyte antigen 4 (CTLA-4) on the T cell membrane (van der Merwe et al. 1997; Ye et al. 1997). The B7.1/B7.2-CD28 interactions stimulate T cells to proliferate and release cytokines, which mediate many of the subsequent immune responses, while the B7.1/7.2-CTLA-4 interactions play an inhibitory role to avoid overstimulation.

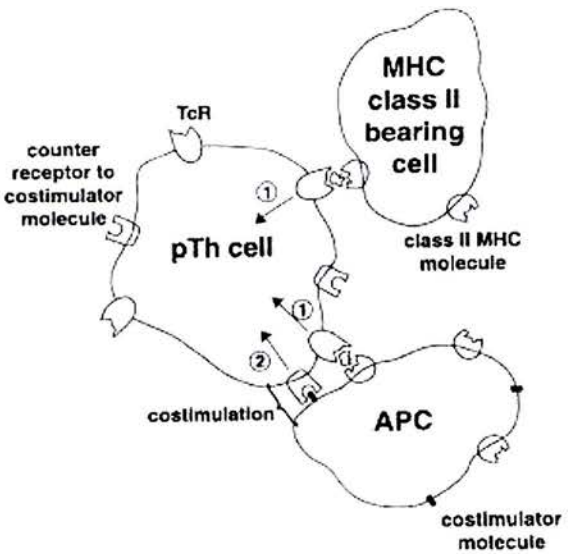


Figure 1.3: Overview of the two signal theory for self-non self discrimination. TcR: T cell receptor, pTh: peripheral T helper cell. Number 1 shows the first signal provided by the interaction of MHC – peptide complex with the T cell receptor, while number 2 shows the second signal provided by the costimulatory molecules (Picture obtained from http://www.usask.ca/medicine/microbio/bretscher/SelectedPublications_ copy.html).

The effector phase of the response includes both innate mechanisms, such as inflammation and macrophage activation, and adaptive mechanisms, such as antibody production and T cell mediated cytolysis, providing the so-called humoral and T cell mediated immunity, respectively (reviewed by Abbas et al. 2000; Janeway et al. 2001). During this phase, effector T cells move from the lymph organs to the site of infection and act after recognition of the peptide-MHC complex on the surface of infected cells or professional APC (DC, macrophage or B cell). This phase needs no co-stimulatory signals, although it has been suggested that costimulation of effector T cell functions by other B7 family members may occur (Chambers C. 2001; Liang and Sha, 2002).

In $CD8^+$ T cell-mediated immune responses, peptides derived from cytosolic antigens of viruses or other intracellular pathogens, tumor proteins (Greenberg P. 1991) or, in the case of allografts, alloantigens are presented by class I MHC molecules, which are expressed on all nucleated cells (Mitchell et al. 1998). This interaction provides the first signal for activation of naïve $CD8^+$ T cells. However, because the B7 molecules that provide the co-stimulatory signal are expressed mainly by APCs, only infected APCs would be expected to initiate cytotoxic responses. When cells other than APCs are infected with an intracellular pathogen, it has been suggested that professional APCs, most probably DCs (den Haan et al. 2001; Heath et al. 2004) and less possibly B cells (Ke and Kapp, 1996) and macrophages (Debrick et al. 1991; Rock et al. 1993), are able to ingest such infected cells during their apoptotic or necrotic phase (reviewed by den Haan and Bevan, 2001; Rock K. 1996), or take up soluble antigens or cell associated proteins from infected cells (Carbone et al. 1998). After that, antigens either move to the cytosol or are presented on the cell surface in association with class I MHC molecules through a non-cytosolic pathway, a process called cross-priming (Bevan M. 1976; reviewed by Heath et al. 2004). It is believed that $CD4^+$ T helper cells may play an important role in cross-priming (Bennett et al. 1997). It was found that APCs may also present the same endocytosed antigen to T helper cells in association with class II MHC molecules. This double recognition of the antigen by both $CD4^+$ and $CD8^+$ cells either causes the release of cytokines, providing the second signal for $CD8^+$ T cell differentiation, or enhances the ability of APCs to express B7 molecules (Bennett et al. 1997; Carbone et al. 1998). However, opponents of this model suggest that cross-priming may occur only in exceptional cases under physiological conditions (Zinkernagel R. 2002). Whichever the case is, activated $CD8^+$ T cells proliferate and differentiate into effector cytotoxic T cells capable of binding and attacking infected or tumor cells, causing cell lysis by apoptosis (reviewed by Harty et al. 2000). Effector $CD8^+$ T cells also

secrete cytokines, mostly interferon γ (IFN-γ), lymphotoxin (LT) and tumor necrosis factor (TNF), to activate phagocytes and induce inflammation (reviewed by Harty et al. 2000).

In helper T cell-dependent immune responses, uptake of soluble antigen by APCs results in presentation of antigenic peptides in association with MHC class II molecules and induction of B7 co-stimulator molecules (reviewed by Fearon and Locksley, 1996; Jenkins et al. 2001). Following recognition of both the co-stimulator and the peptide-MHC complex, T helper cells (Th) express and secrete interleukin 2 (IL-2) – although other factors may be involved (Jenkins et al. 2001). IL-2 functions as a growth and differentiation factor for antigen-specific T cells and promotes their clonal expansion. These interactions stimulate cytokine secretion, the pattern of which varies to some extend between the two well characterized subsets of Th cells (Th1 and Th2) (Mosmann et al. 1986; reviewed by O'Garra and Arai, 2000; Kidd P. 2003). It has also been suggested that the affinity with which the MHC molecules bind certain peptides affects the direction of a T helper response towards Th1 or Th2 (Murray J. 1998).

Although immune responses are not always either Th1 or Th2 exclusively, a general overview is described below. In most cases, when macrophages and dendritic cells present antigens from intracellular pathogens mainly in endosomal compartments, they mainly produce IL-12. This cytokine acts on antigen stimulated Th cells and induces their differentiation into the T helper 1 subset. When effector Th1 cells then encounter their specific antigen displayed on a class II MHC molecule on the surface of a macrophage or a B cell, they secrete cytokines. Secretion of IFN-γ and LT α induces the release of TNF-α by other cells, such as epithelial cells (reviewed by Spellberg and Edwards, 2001). These cytokines in turn activate macrophage microbicidal functions and antibody production, such as IgG1 and IgG2a in human and mice, respectively (Fearon and Locksley, 1996) to eliminate pathogens, such as intracellular bacteria, fungi and mycobacteria (reviewed by Kidd P. 2003; Spellberg and Edwards, 2001).

On the other hand, secretion of IL-4 or IL-13 by B cells and DCs during antigen presentation of extracellular pathogens promotes the differentiation of T cells into the T helper 2 subset (Th2) (reviewed by Kidd P. 2003; Spellberg and Edwards, 2001). When effector Th2 cells recognize their specific antigen-MHC class II complex on the surface of an APC, they secrete IL-4, 5, 6, 10 and 13, which activate mast cells and eosinophils and induce B cell proliferation and differentiation into plasma cells secreting mainly IgE and IgG4 (Fearon and Locksley, 1996) to eliminate large pathogens that cannot be phagocytosed, such as helminths (reviewed by Spellberg and Edwards, 2001).

1.1.1.1 MHC independent humoral immune responses

Apart from protein antigens that elicit T cell-dependent (TD) immune responses, where MHC molecules play an important role, as described, in some cases the immune system develops T cell independent (TI) immune responses against mainly non-protein antigens. Such antigens from parasites, viruses, fungi and bacteria, with the lipopolysaccharide (LPS) one being the most well characterised, may act as mitogens or superantigens, triggering the machinery of polyclonal B lymphocyte responses in type 1 TI immunity (Mond et al. 1995; Reina-San-Martin et al. 2000). Under high concentrations of pathogenic burden, non specific polyclonal B cell proliferation takes place that, however, renders the immune system powerless, may cause immunosuppression and augments autoimmune reactions (Mond et al. 1995; Reina-San-Martin et al. 2000).

In the type 2 TI immune responses, mature B cells can be activated by extensive cross-linking of their antigen specific receptors (reviewed by Mond et al. 1995; Vos et al. 2000). Antibodies bound on the membrane of B lymphocytes serve as specific receptors for almost every kind of extracellular or cell surface antigen, including large polymeric proteins, lipids and phospholipids, carbohydrates, nucleic acids and small chemical substances (Mond et al. 1995). Distinct subsets of B cells, the CD5 B1 and the marzinal zone B cells are mainly involved in this kind of response (reviewed by McHeyzer-Williams M. 2003). Although this type of immune response is regarded as TI, it seems that γδ (Gumperz and Brenner, 2001) and double negative $CD4^-CD8^-$ T cells provide some kind of help (Fairhurst et al. 1998; Gumperz and Brenner, 2001). This T cell help, though, is not MHC restricted. Instead, it has been suggested that an MHC like molecule, CD1, acts as a restriction element for polysaccharide-encapsulated bacterial antigens (Fairhurst et al. 1998) and for lipid mycobacterial antigens (Gumperz and Brenner, 2001). During the TI 2 response, a small number of effector cells are usually generated producing low affinity antibodies, there is limited isotype switching and in some cases immunological memory is not developed (reviewed by Mond et al. 1995).

In conclusion, the vast majority of specific immune responses to protein antigens require the binding of antigenic peptides to a major histocompatibility complex molecule, a co-stimulatory signal provided by the innate immunity and subsequent recognition by T cells. Only foreign peptides that bind to an MHC protein can initiate an immune response, although many peptides that do bind are not immunogenic. Apparently, there are holes in the T cell repertoire for antigen recognition. The development of the T cell repertoire and how the MHC molecules affect that is the subject of the next section.

1.1.2 Developing and maintaining a diverse T cell repertoire

T cell receptors (TCR) are heterodimers consisting of two covalently linked transmembrane polypeptide chains, α and β. Each chain consists of an immunoglobulin (Ig)-like variable (V) domain, an Ig-like constant (C) domain, a transmembrane domain and a short cytoplasmic region. The antigen binding site is situated in the variable domain, where each chain has three hypervariable regions known as complementarity determining regions (CDRs). Each TCR chain is encoded by multiple gene segments that undergo somatic rearrangement during the maturation of T lymphocytes (Davis and Bjorkman, 1988).

The expression of antigen receptor genes is a key event in lymphocyte maturation required for the generation of a diverse T cell repertoire and for maintaining survival of lymphocytes. Functional TCR genes are assembled from V (variable), D (diversity), and J (joining) DNA segments that are separated from one another in the germline and are spliced together during maturation. There are around 20-70 Vα and Vβ segments in mammals and several D and J segments. Two of the CDRs (CDR1 and CDR2) are encoded by the V gene segments, whilst the third CDR (CDR3), which exhibits the greatest variability, is created by the juxtaposition of VJ or VDJ segments for the α and β chains respectively (Davis and Bjorkman, 1988). Therefore, a great number of distinct TCRs are created by all possible different segment combinations and the greatest sequence diversity is concentrated in CDR3.

The process of gene rearrangement occurs in the thymus. In general, thymocyte maturation can be divided into three broad stages based on co-receptor surface expression (Kisielow and VonBoehmer, 1995). The earliest T cell precursors are $CD4^-CD8^-$ double negative (DN) and express only the β TCR chain. These proliferate extensively and become $CD4^+CD8^+$ double positive (DP), which are still immature, but express both α and β TCR chains and undergo either positive selection to proceed to the mature $CD4^+$ or $CD8^+$ single positive (SP) cell stage, or negative selection that inactivates or deletes potentially autoreactive thymocytes. Consequently, although there is a huge germline repertoire, only about 5% of the DPs mature to SPs that will migrate from the thymus to the peripheral organs (Goldrath and Bevan, 1999).

The mature T cell repertoire is established through interactions with self peptide-MHC molecule complexes (reviewed by Goldrath and Bevan, 1999; Huseby et al. 2005; Sebzda et al. 1999; von Boehmer et al. 2003). As the precise pathogen-derived antigens that will be encountered is unpredictable, the immune system uses the diverse array of self peptides bound

to self MHC molecules to select and maintain a T cell repertoire capable of responding promptly to foreign antigens. The model supported by most studies suggests that those DP thymocytes that bind peptide-self MHC with low affinity undergo positive selection, whereas DP cells that express a receptor that interacts with high affinity with MHC molecules in the thymus are either inactivated or destroyed. However, it is more possible that the selection of a T cell repertoire is not mediated by a single strict interaction of one peptide-MHC complex with a TCR (Sebzda et al. 1999). Instead, it is suggested that TCR-peptide-MHC recognition is a degenerate process that involves multiple interactions with a mixture of peptides. TCRs that bind self peptide-self MHC complexes strongly are potentially autoreactive and undergo negative selection (Mathis and Benoist, 2004). That some T cells fail to recognise some foreign antigen-MHC combinations possibly reflects this immunological phenomenon, known as tolerance. Foreign peptide-MHC complexes may escape recognition because they resemble self-MHC complexes for which the appropriate TCRs have been negatively selected (Marx J. 1988). Moreover, 'holes' in the repertoire have been described when T cells are selected in the presence of low amount of thymic peptides during maturation. Thus, the greater the number of peptides in thymus during T cell selection, the more diverse the T cell repertoire becomes (Sebzda et al. 1999).

On the other hand, a variety of models show that 5% to 75% of the positively selected cells are potentially autoreactive and normally undergo clonal deletion. A combination of qualitative and quantitative models for peptide-MHC recognition by the TCR in addition to other TCR interactions with surface/coreceptor molecules explain the selection of the T cell repertoire (reviewed by Carreno et al. 2006; Sebzda et al. 1999). Qualitatively, different peptide-MHC complexes promote positive or negative selection based on their affinity to TCRs, while different concentrations of the same peptide may also promote positive or negative selection. Therefore, those T cells selected by rare peptide-MHC complexes may be regarded as low affinity interactions and initiate positive selection although they are actually self-reactive and must be clonally deleted afterwards. Moreover, the thymus does not contain all the different self peptides found in an organism, because many proteins are only expressed in certain tissues and therefore, it is possible that TCRs capable of recognizing such self peptides could escape negative selection in the thymus and need to be deleted subsequently. Under these circumstances, mechanisms for inducing T cell tolerance in the periphery are crucial. The two signal model for T cell recognition, described in section 1.1, provides peripheral tolerance, since T cells bearing receptors that recognise peptide-MHC complexes, without receiving co-stimulatory signals undergo apoptosis or anergy (reviewed by Chambers C. 2001).

1.1.3 Structural basis of T cell recognition

Antigenic proteins to which T cells respond are degraded into peptides that are bound by MHC molecules and displayed on the cell surface for interaction with the T cell receptor. A feature of T cell responses is that only a few short segments of an antigenic protein, usually containing around 10 to 30 amino acid residues, are recognized. Crystal structures determined for αβ TCR heterodimers, both alone and in complex with the peptide-MHC have shed light on the structural basis of cellular immune recognition (reviewed by Garcia K. 1999).

The TCR is positioned in a roughly diagonal orientation over the face of the peptide-MHC complex to maximise the interaction between the two surfaces (Bjorkman P. 1997; Garboczi et al. 1996; Garcia et al. 1996; 1998; Fig. 1.4). The CDR1 and CDR2 regions of the TCRα chain lie over the N-terminal half of the peptide and the corresponding regions of the TCRβ lie over the C-terminal half. Both CDR3s align near the middle of the peptide. While the CDR1 and CDR2 regions interact largely with MHC residues, the most variable CDR3 regions lie centrally and have most contact with protruding side chains of the peptide (Bjorkman P. 1997; Garcia et al. 1996). Surprisingly, only relatively few contacts are made with peptide side chains that point up toward the TCR. Nevertheless, other peptide side chains are recognized indirectly through subtle conformational changes of MHC residues (Bjorkman P. 1997).

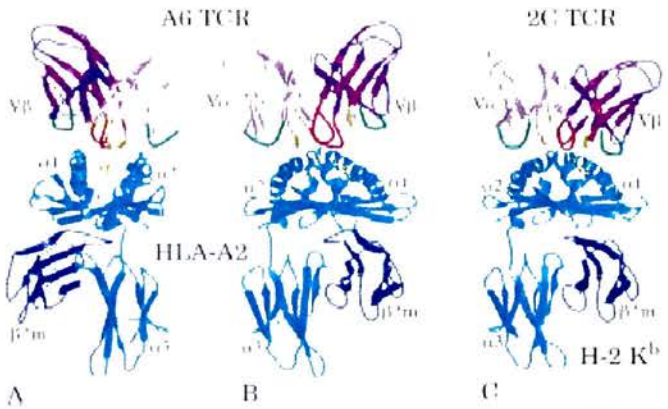


Figure 1.4: Structures of TCR /MHC / Peptide complexes. Only the V domains of the TCR are shown. CDR1, yellow; CDR2, green; CDR3, red. (A and B) A6TCR/HLA-A2/Peptide (Garboczi et al. 1996) (C) 2C/H-2K^b/Peptide (Garcia et al. 1996).

1.2 The Major Histocompatibility Complex (MHC)

1.2.1 Definition and protein structure of MHC molecules

The major histocompatibility complex (MHC) is a genetic region that encodes molecules playing a central role in immune responses. There are three classes of MHC genes (I, II and III) as determined by their structure and function. The class I and II genes are generally categorized as classical or non-classical. Polymorhic and highly expressed genes that encode products presenting antigenic peptides to T cells, and therefore act as restriction elements, are classified as classical. The definition of the non-classical genes is not as clear, because they encode products with various immunological functions and include many genes of unknown function. However, some general characteristics facilitate differentiation of classical and non-classical MHC genes. In contrast to their classical counterparts, non-classical MHC genes are generally expressed at low levels and exhibit limited or no polymorphism (Braud et al. 1999). The class III MHC genes encode non-antigen presenting structurally diverse molecules with various immunological and other functions (reviewed by Lehner et al. 2004; Yu et al. 2000).

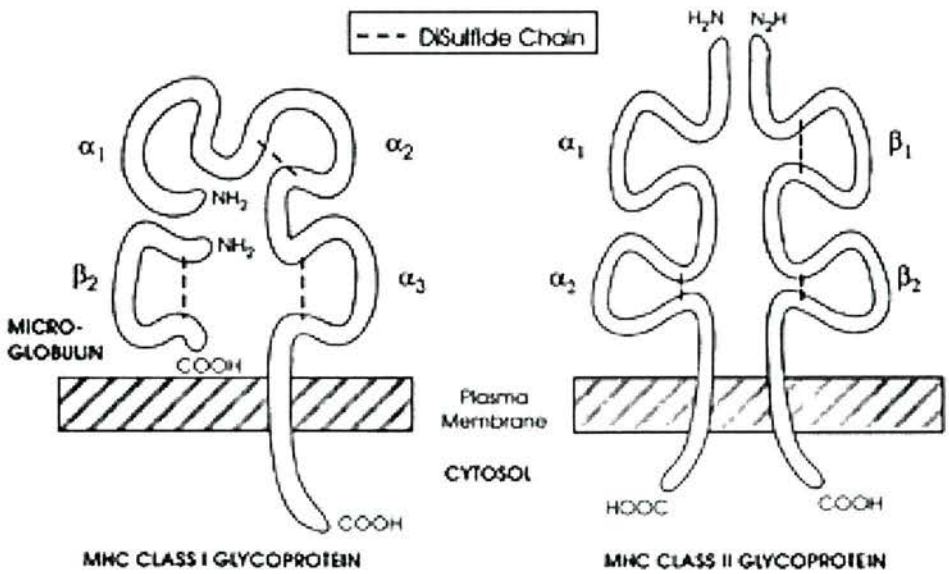


Figure 1.5: Schematic diagram of MHC class I and II glycoproteins (Picture obtained from www.sanbi.ac.za).

MHC class I and II molecules are cell surface glycoproteins that consist of an extracellular peptide-binding domain followed by a pair of immunoglobulin (Ig) like domain. These extracellular components are anchored to the cell by transmembrane and cytoplasmic domains (Fig. 1.5).

Class I molecules consist of two non-covalently linked polypeptide chains: a glycosylated MHC-encoded α chain (heavy chain) and a non MHC-encoded subunit known as β2-microglobulin. The heavy chain consists of three extracellular segments (α1, α2, α3). Two of these (α1 and α2) form the peptide binding groove, whereas the third is an Ig-like domain (Bjorkman et al. 1987a, b; Fig. 1.5, 1.6). The α1 and α2 domains each consist of an antiparallel β sheet with four strands and a long α helical region. The α3 and β2-microglobulin domains are both β sandwich structures composed of two antiparallel β sheets, with four and three β strands respectively (Bjorkman et al. 1987a).

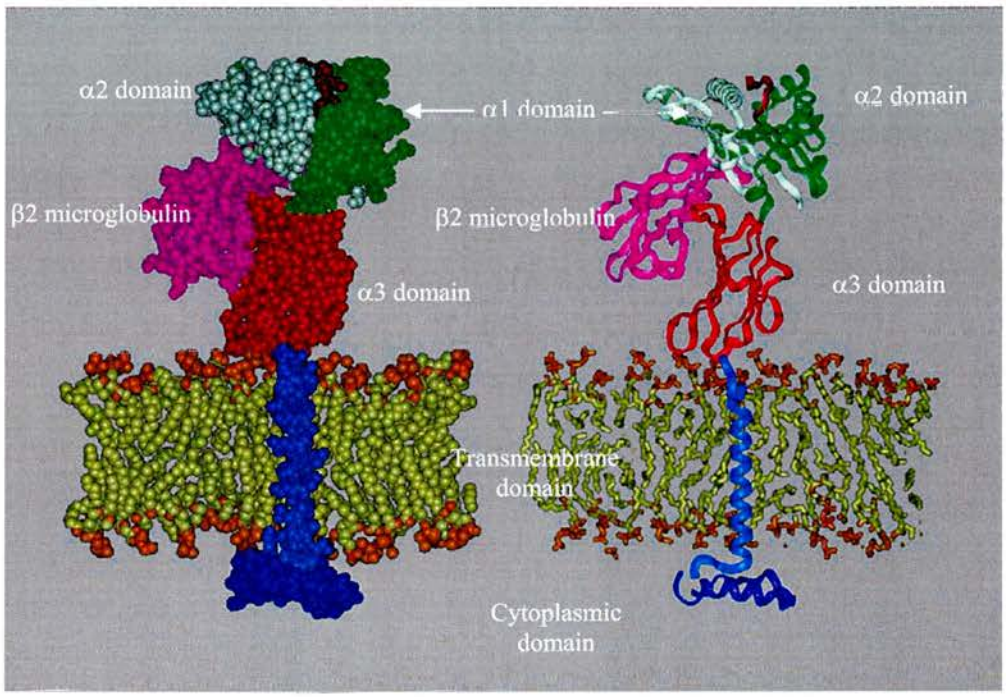


Figure 1.6: Three dimensional structure of an MHC class I molecule (Picture obtained from www.cryst.bbk.ac.uk)

Class II molecules are composed of two non-covalently associated MHC-encoded glycosylated polypeptide chains (α and β). Each chain consists of two extracellular segments

(α1, α2 and β1, β2). The amino terminal α1 and β1 segments form the peptide binding groove, which is composed of an antiparallel β sheet with eight strands and two antiparallel α helical regions, whereas α2 and β2 are Ig-like segments (Brown et al. 1993; Madden D. 1995; Fig. 1.5).

1.2.2 Genomic organization and function

In humans, the MHC complex spans a region of 4Mb on the short arm of chromosome 6 and the general subregion organisation is centromere-class II-class III-class I (http://www.sanger.ac.uk/HGP/Chr6/). The complete nucleotide sequence of the human MHC has been determined (The MHC Sequencing Consortium, 1999), with 224 gene loci identified, 128 of which are believed to be expressed. Although most of the genes are still of unknown function, it was estimated that around 40% have some immunological function. The gene map of all identified human MHC loci is shown in figure 1.7.

The human class I subregion comprises the class Ia subregion, with three genes (B, C, A) encoding the heavy chains of the classical MHC molecules as well as the class Ib subregion containing several other genes encoding non-classical class I like products and molecules of unknown function. It also contains many pseudogenes (Fig. 1.7). Non-classical MHC molecules include HLA-E,-F, -G and -H (or HFE), MICA and MICB. The function of these are quite well characterized, with the exception of HLA-F and G (reviewed by Braud et al. 1999; Shao et al. 2005). The HLA -E molecules are thought to play a regulatory role for NK cell lysis, mainly by inhibiting the action of NK cells on cells not expressing classical MHC molecules that would otherwise be targeted (Braud et al. 1998). HFE is involved in iron metabolism (Gross et al. 1998) and MIC molecules act, mainly in response to heat shock on intestinal epithelia, to activate γδ T cells to kill infected or damaged epithelia (Groh et al. 1998). MICA can also interact with the NK cell receptor NKG2D, which activates NK cells and provides co-stimulation for T cells (Shao et al. 2005). It has also been suggested that HLA -G and MIC expression in thymic epithelia play a role in selecting the T cell repertoire (Braud et al. 1999).

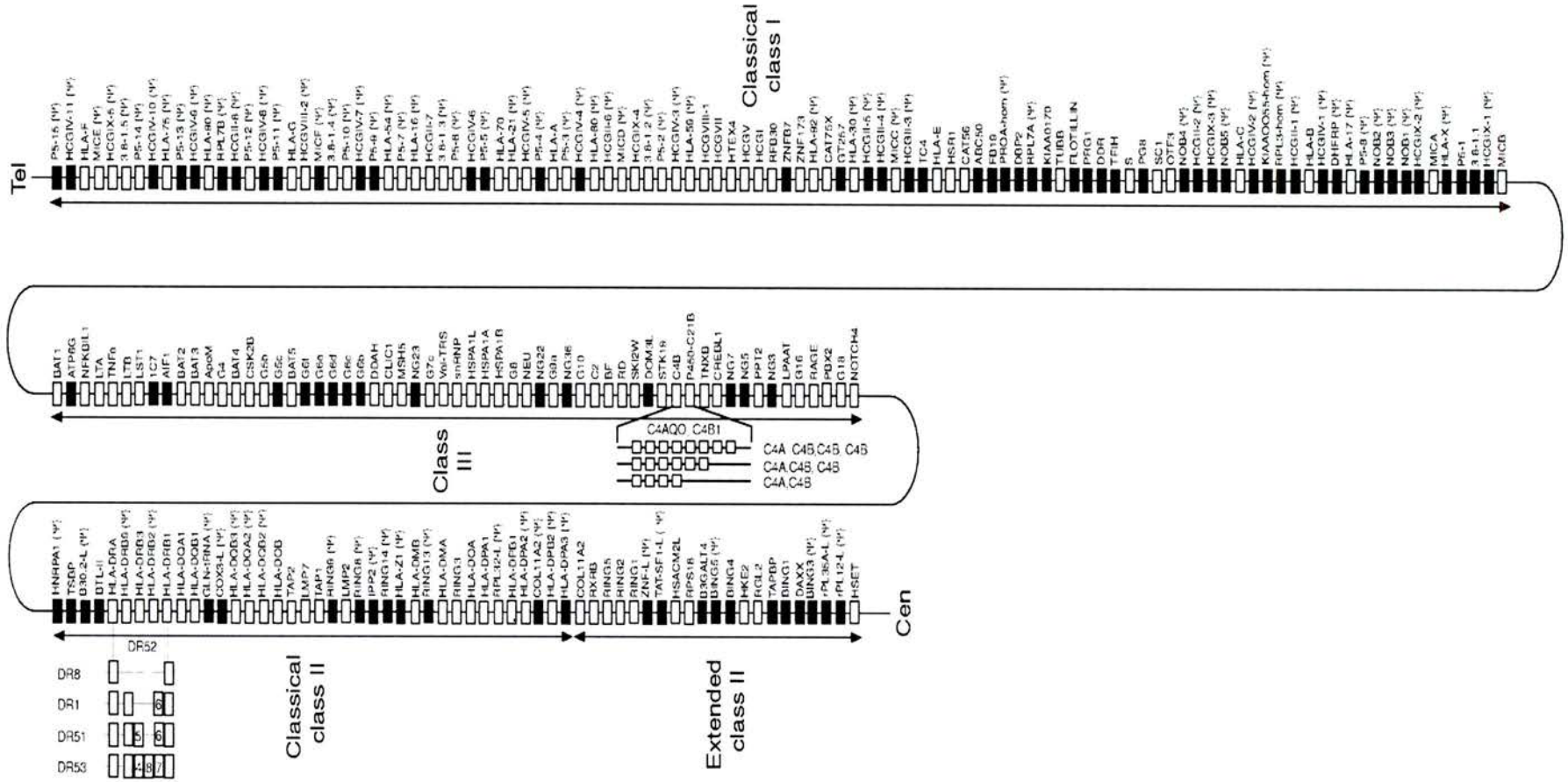


Figure 1.7: The human MHC gene map (Picture obtained from the MHC sequencing consortium, 1999).

The human class II subregion contains three expressed polymorphic gene subclasses (DP, DQ and DR), along with non-polymorphic genes encoding heterodimeric class II molecules (DO and DM). It also contains genes that encode proteins involved in the antigen processing, such as TAP (transporter associated with the antigen processing) genes and LMP (proteasome components), as well as many pseudogenes and a number of other genes, most of which have no known immunological function (Beck and Trowsdale, 1999). The DP, DQ and DR represent the classical restriction elements for T helper cells, whereas the DO and DM genes are non-classical molecules involved in peptide loading of classical class II molecules (reviewed by Alfonso and Karlsson, 2000; Bryant and Ploegh, 2004).

The class III MHC subregion contains 60 and 61 expressed genes in humans and mice, respectively, but the physiological roles of most of these remain to be determined (Xie et al. 2003). It has been suggested that around 1/3 of the proteins encoded in this region have a role in RNA processing (Lehner et al. 2004). Some genes involved in immune responses include complement components, tumour necrosis factor (TNF), several lymphotoxins (LT) and heat shock proteins (HSP).

1.2.3 Antigen processing and presentation pathway

Generation of peptides for presentation to T cells by MHC molecules involves processes that differ for class I and class II proteins.

Class I MHC molecules are expressed by all nucleated cells and initiate cell mediated immune responses by presenting intracellular cytosolic antigens such as virus and tumor proteins to $CD8^+$ T cells. However, as described in section 1.1, cross priming also allows class I molecules to present antigens from endocytosed compartments. These conventional and alternative MHC class I antigen processing and presentation pathways have been reviewed (Gromme and Neefjes, 2002; Heath et al. 2004; Fig. 1.8).

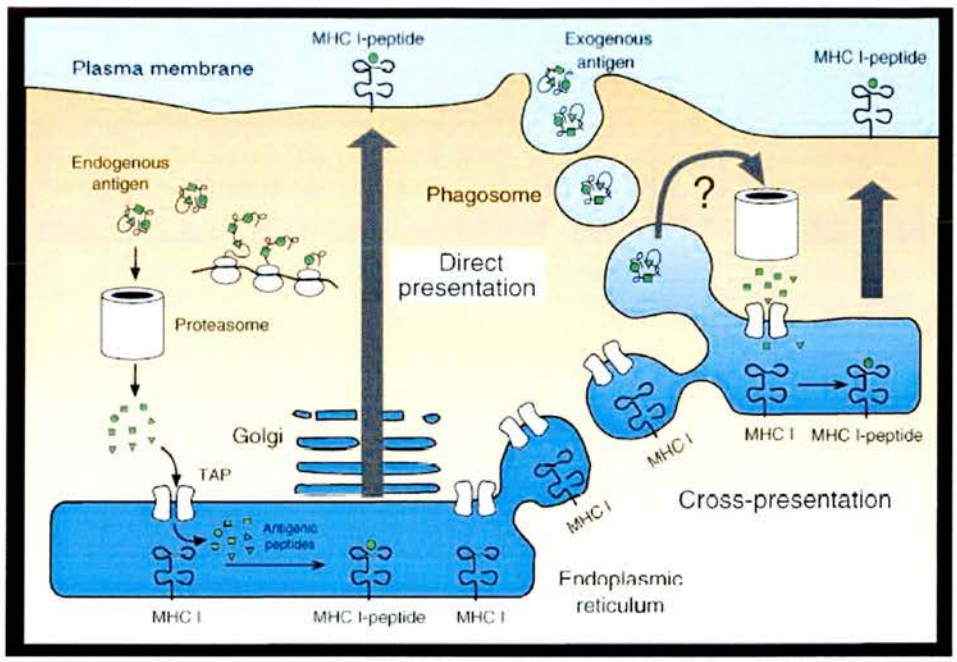


Figure 1.8: Major Histocompatibility Complex class I presentation pathways (Heath et al. 2004).

In the conventional pathway, cytosolic proteins are targeted for degradation by conjugation with ubiquitin. Those proteins with a polyubiquitin tail are degraded by a proteinase complex known as the proteasome. The resulting peptides are then transported to the endoplasmic reticulum (ER) via the TAP protein complex. The number of peptides translocated by TAP is low compared to the number of peptides generated by the proteasome as a result of peptidase activity. This may account for a relatively low efficiency of antigen presentation by class I molecules, especially for low copy antigenic peptides. However, peptidases and proteases in the cytosol and the ER may also contribute to the generation of peptides of appropriate length (9-12 residues, mostly 9mers) for class I binding. Moreover, TAP molecules exhibit limited polymorphism, which affects the set of peptides carried into the ER. There, chaperone-assisted assembly generates a stable heterotrimer containing MHC class I heavy chain, β2-microglobulin and a peptide. After glycosylation in the Golgi apparatus, this complex is transported via the secretory pathway to the cell surface for presentation to $CD8^+$ T cells (Fig. 1.8).

The alternative pathway of antigen processing of exogenous peptides is not yet completely understood (Gromme and Neefjes, 2002; Heath et al. 2004). Endocytosed antigens or proteins from intracellular pathogens such as bacteria gain access to endosomes or

phagosomes. From there, they may be transferred to the cytosol and follow a conventional class I pathway for process and presentation. Transport to the cytosol could result either from phagocytic overload or disruption of the endosomal membrane by toxins produced by the pathogen (Brunt et al. 1990; Mazzaccaro et al. 1996). In the non-cytosolic alternative pathway, phagosytosed proteins are translocated from the phagosome to be degraded by closely-associated proteasomes. Peptides are then transported back to the endosomal or phagolysosomal compartment and loaded onto MHC class I molecules. The class I molecules present in these compartments are either newly synthesized or, more possibly, internalized from cell surface. Following peptide exchange, the new heterotrimers are recycled back to the cell surface (Fig. 1.8).

Class II MHC molecules are expressed only in certain cells of the immune system, such as macrophages, B cells and dendritic cells and they display peptides from external antigens to $CD4^+$ T helper cells. The process involves uptake of antigen into the endocytic compartment and degradation of the proteins into peptides of 10 to 30 residues (usually 12-20), which bind to the class II peptide binding groove, which is more extended than that of class I molecules (Brown et al. 1993). Class II MHC are transported into the ER, where they associate with a protein known as the invariant chain (Ii). Binding of Ii to the class II binding groove prevents binding of peptides in the ER. The class II-Ii complex is then transported to a specialised compartment known as the MHC class II compartment (MIIC), which fuses with endosomal vesicles. The Ii is then cleaved by proteases, leaving only a short fragment called the class II associated invariant chain peptide (CLIP). Finally, a non-classical class II molecule, the DM, catalyses the release of the CLIP fragment and the binding of antigenic peptides to the MHC class II molecules (reviewed by Alfonso and Karlsson, 2000; Bryant and Ploegh, 2004). DM continuously binds and releases weakly bound peptides, affecting the outcome of which peptides are ultimately presented for T cell recognition.

1.2.4 The importance of MHC polymorphism

MHC genes are the most polymorphic genes in the genome of vertebrates and the polymorphic residues are mainly situated in the peptide binding groove (Bjorkman et al. 1987b; Brown et al. 1993), so that different alleles favour both the binding of different peptides (Falk et al. 1991; Rudensky et al. 1991) and the development of distinct T cell repertoires (Messaoudi et al. 2002).

MHC polymorphism is maintained by some form of balancing selection (Hedrick P. 1998). At the population level, such polymorphism ensures that peptides derived from possible pathogens will be recognised by the immune system of at least some individuals and thus, the likelihood of a pathogen evading all individuals is eliminated. Since MHC genes are co-dominantly expressed and most individuals are heterozygous because of the extremely polymorphic nature of MHC genes, two alleles are usually expressed from each locus. Hence, given that there are three classical class I MHC loci in humans, up to six different class I proteins can be expressed on the cell surface (Parham et al. 1995). In the case of class II genes, the total number of distinct class II molecules for a heterozygous individual can range from 10 to 20, because of the possibility of heterologous pairing of α and β chains when multiple DR or DQ loci occur. Taking into consideration that multiple peptides with certain common features can potentially bind to a given MHC molecule, MHC polymorphism and co-dominant expression of multiple loci ensures that the maximum possible number of either pathogen peptides or self peptides can be displayed for T cell recognition during antigen presentation and T cell repertoire development, respectively.

1.2.4.1 MHC polymorphism in disease association studies and vaccine development

Because of the dual function of the MHC molecules in presenting processed antigenic peptides to T cells on the one hand and their role in shaping the T cell repertoire on the other hand, these proteins and their polymorphism has been extensively studied in disease association and vaccine development studies.

1.2.4.1.1 The MHC in disease association studies

An association between a polymorphic MHC gene and an immune response trait or a disease may be due to either a direct effect of that gene or that of a gene linked to it. Thus, both the function and the polymorphic nature of the MHC genes make them likely candidate or marker genes for association studies.

Many disease association studies have been conducted in humans and domesticated animals (reviewed by Dietert R. 1996; Kaufman et al. 1995; Lewin H. 1996; Lewin et al. 1999; Lombard et al. 2006; Schonbach et al. 2002; Takeshima and Aida, 2006; Thorsby E. 1997). In humans, these studies provide useful evolutionary perspectives, showing that specific pathogens

influence the evolution of MHC genes (Lombard et al. 2006). This is apparent from the well established associations between a class I specificity and a class II haplotype with protection from severe malaria (Hill et al. 1991). These alleles are found in higher frequencies in West African populations compared to other racial groups, suggesting that natural selection by malaria has contributed to their elevated frequencies in Africa. Such findings have been exploited in an attempt to design a vaccine against malaria, by identifying the epitopes involved in a protective immune response (Davenport and Hill, 1996; Doolan et al. 1996). Other well established MHC associations in humans include correlations with autoimmune diseases, such as rheumatoid arthritis and type I insulin dependent diabetes (reviewed by Lombard et al. 2006; McDevitt H. 2000; Thorsby E. 1997). The FIMM database for functional molecular immunology includes 52 disease associations (Schonbach et al. 2000; 2002) and a database containing primers for single nucleotide polymorphism (SNP) detection in the MHC area has further facilitated such association studies (Geraghty et al. 2002).

In domesticated animals, MHC association studies were motivated by the potential for improving genetic resistance to infectious diseases. As the relationships between MHC molecules and function were elucidated, however, it became obvious that maintaining heterozygosity is important for herd health. Homozygosity at an MHC locus for an allele or haplotype may provide resistance to a particular disease, but 'holes' in the peptide-binding repertoire for other diseases may exist because of lower diversity of MHC presenting molecules. Thus, creating animals homozygous at an MHC locus is not only difficult to achieve, but also be undesirable. Nevertheless, altering the frequency of resistant alleles or haplotypes for a specific disease is desirable and feasible, particularly for those diseases where resistance is a dominant trait such that heterozygous animals with only one resistant allele are nonetheless resistant (Lewin et al. 1999). To date, Marek's disease in chickens is a well established example of an association between MHC and resistance (Briles et al. 1977). In ruminants, a number of studies support associations between bovine MHC and resistance to Bovine Leukemia Virus (BLV) and mastitis (reviewed by Lewin H. 1996; Lewin et al. 1999; Takeshima and Aida, 2006), while associations of ovine MHC with scrapie (reviewed by Dietert R. 1996) and parasitic infections have been suggested (reviewed by Stear and Wakelin, 1998).

The precise mechanism whereby MHC genotype confers resistance or susceptibility to particular pathogens are unclear, but may relate to the specificity of peptide binding. Animals with different sets of MHC genes preferentially bind and present particular immunogenic

peptides to T cells. Those T cells may be involved in the pathogenesis of the disease or in a protective immune response (Thorsby E. 1997).

Moreover, it is likely that animals with different MHC alleles and endogenous peptides become tolerant to different self antigens and develop distinct T cell repertoires. A direct link has been suggested between MHC polymorphism and defense against pathogens, associated with increased T cell repertoire diversity and T cell avidity (Messaoudi et al. 2002). These authors used co-isogenic mouse strains differing only in the alleles of one locus and showed that differences in resistance to herpes virus was attributable to different T cell repertoires against an immunodominant peptide. The susceptible mice had developed T cells with TCRs exhibiting lower overall avidity for the peptide. It seems, therefore, that different MHC alleles favour the development of distinct T cell repertoires, which affects the magnitude and specificity of some of the variation in immune responses against pathogens. Consequently, MHC polymorphism may account for resistance or susceptibility to pathogens.

1.2.4.1.2 MHC polymorphism in vaccine development

One approach to vaccine development involves vaccination with peptides corresponding to short segments of a pathogen, instead of using the whole organism. This approach is very useful, because it may provide safe alternatives to the conventional live attenuated or inactivated vaccines. Moreover, an intact pathogenic protein may incorporate elements evolved by the pathogen to escape the immune response, whereas immunization with subunit peptides can bypass this problem (Cunha-Neto E. 1999). Research on human subunit vaccines has resulted in the design and development of very successful vaccines, such as that targeting hepatitis B virus (Koff R. 2002), while research on veterinary subunit vaccines has shown that conjugated peptides derived from several animal viruses provide protective immune responses (reviewed by Shams H. 2005).

The immunogenicity of a protein for an individual depends largely both on its ability to bind to MHC molecules and on the ability of the T cells to recognise the resulting peptide-MHC complexes (Doolan et al. 1996; Sette and Fikes, 2003; chapter 1.1). Although parameters such as the role of costimulatory molecules, the concentration of the antigenic protein and the steps involved in the antigen processing and presentation may also influence the immunogenicity of a protein, a critical requirement for an immunodominant peptide is a structure that permits it to

form a complex with an MHC molecule that can be recognized by a T cell (Corradin and Demotz, 1997; Doolan et al. 1996; Sette and Fikes, 2003; chapter 1.1).

Individual MHC molecules can bind a range of peptides that share common features, known as peptide binding motifs. Such motifs have been established for a number of MHC alleles by sequencing naturally processed peptides eluted from their binding grooves (Falk et al. 1991; reviewed by Davenport and Hill, 1996; Sette and Fikes, 2003). Some amino acid residues of a peptide are crucial for efficient binding to MHC molecules and are known as anchor residues. These residues bind to polymorphic amino acids in the binding groove and as a result, especially for class I molecules, related amino acids occur at particular positions in all peptides eluted from individual MHC specificities. A database containing short peptides known to bind MHC class I or II molecules from several organisms has been created (MHCPEP; Brusic et al. 1994, 1996, 1997 and 1998) and more recently, the EPIMHC database was established to address the lack of updates of MHCPEP since 1998 (Reche et al. 2005). On the basis of these developments and information from crystallographic structures of MHC proteins, algorithms have been developed that allow prediction of high affinity peptides from pathogens (reviewed by Cunha-Neto E. 1999).

A successful vaccine should include epitopes that elicit protective immune responses in individuals carrying any of the MHC alleles present in the target population (Cunha-Neto E. 1999). Accommodation of this principle towards development of subunit vaccines for humans has followed two main strategies based on classical or reverse immunogenetics (reviewed by Davenport and Hill, 1996). The classical immunogenetic approach relies on the identification of candidate immunogenic pathogen proteins, demonstration of their protective capacity and establishment whether MHC polymorphism affects the magnitude and specificity of the immune response. In contrast, the reverse immunogenetic strategy starts with an association of a particular MHC allele or haplotype with disease resistance in a given population. This is followed by identification of peptide binding motifs from self peptides eluted from that allele and comparison with peptides from known pathogen proteins (Lauemoller et al. 2001). This allows synthesis of candidate pathogen peptides, which can be screened for the capacity to bind the MHC alleles in question in MHC stabilization assays using TAP-deficient cell lines (Anderson et al. 1993), and in functional assays for T cell recognition such as lysis of peptide-pulsed target cells by specific cytotoxic T lymphocytes.

1.3 The ovine Major Histocompatibility Complex (MHC)

1.3.1 Nomenclature and genomic organization

The ovine MHC system is known as Ovine Lymphocyte Antigens (OLA), in accordance with HLA in humans and BoLA in cattle, because it was initially defined using alloantisera (Millot P. 1978). The name Ovar, which stands for *Ovis aries*, has been also proposed in an effort towards systematic nomenclature of the MHCs in all vertebrates (Klein et al. 1990).

The ovine MHC system is situated on chromosome 20 (Hediger et al. 1991; Mahdy et al. 1989) and spans a region of at least 2.6Mb (Liu et al. 2006). Most mammals for which genome maps have been constructed, including chimpanzees, old world monkeys, pigs, cattle, dogs and cats, appear to have an overall MHC structure similar to those of human and mouse (Kulski et al. 2002). Only very recently, a BAC-based physical map of the ovine MHC complex was constructed, which shows that its general genomic organization is similar to that of the human and mouse MHC, with centromere-class II-class III-class I-telomere orientation (Liu et al. 2006). This was also supported by a linkage map based on microsatellites (Crawford et al. 1990), where the use of three bovine markers showed that the order and the approximate genetic distance between the markers was conserved between cattle and sheep (Dietert R. 1996).

Although the overall genomic organization is similar among different organisms, some notable differences occur in some species. In contrast to humans, there are at least six expressed class I loci in cattle and BoLA class I haplotypes differ both in the number and composition of expressed class I genes (Ellis and Ballingall, 1999; Ellis et al. 1999; Ellis S. 2004). Haplotype variation may also be a feature of ovine class I genes since a number of serological studies have suggested at least two expressed loci (Cullen et al. 1982; Garrido et al. 1995; Jugo and Vicario, 2001; Jugo et al. 2002; Stear and Spooner, 1981; Stear et al. 1996), whereas others propose at least three (Grossberger et al. 1990; Millot P. 1984). In addition, evidence based on biochemistry, serology and one dimensional isoelectric focusing (1D-IEF) suggests that IEF patterns could be explained by different numbers of genes expressed in different haplotypes (Jugo et al. 2002). However, unlike HLA class I alleles, which are readily distinguished into three loci by inter-locus and intra-locus pairwise sequence comparisons (Parham et al. 1995), sheep and cattle class I alleles cannot be allocated into different loci on the basis of sequence differences (Ellis et al. 1999 and personal communication). Identification of the exact number of expressed bovine and ovine class I loci is therefore difficult.

A striking difference in genomic organization between the HLA and the BoLA and OLA complexes is found in the class II region. BoLA class II genes are separated into two subregions (IIa and IIb) by a distance of around 17cM (Andersson et al. 1988; Hess et al. 1999), with the class IIb cluster located proximal to the centromere (Jarrell et al. 1995; van Eijk et al. 1993). Based on a radiation hybrid map, the separation of the IIa-IIb loci has been attributed to a large chromosomal inversion (Band et al. 1998). Subregions IIb and IIa in sheep are separated by a large genomic insertion of approximately 25cM (Maddox et al. 2001), with the novel ruminant DYA and DIB genes (Wright et al. 1994), which are located in the BoLA class IIb region, also localized in the class IIb subregion of sheep (Liu et al. 2006). In an additional structural deviation from HLA, both BoLA and OLA lack a DP locus (Deverson et al. 1991; Scott et al. 1987).

The ovine MHC class II region contains an estimated 7 different alpha and 24 different beta genes, as revealed by screening of cosmid libraries from two unrelated heterozygous sheep with human and murine class II probes (Deverson et al. 1991). In particular, Southern blot hybridization with HLA class II probes revealed a single DRA and multiple copies of DRB, DQA and DQB genes in the ovine MHC class IIa subregion (Scott et al. 1987). It seems that there are at least three DRB genes; a study using monoclonal antibodies specific for sheep MHC class II molecules suggested that two distinct DRB products are expressed (Dutia et al. 1994) and an additional DRB pseudogene has been described (Scott et al. 1991b). Co-expression of ovine DRA and DRB genes in mouse L cells has been achieved (Ballingall et al. 1992).

The DQ area in sheep seems to be more complicated. Scott et al. (1987a, b) provide evidence for at least two DQA- and three DQB-like loci. From these, four partial genomic clones representing two DQA (DQA1 and DQA2) and one DQB locus were sequenced (Scott et al. 1991a, b). A map subsequently constructed for the ovine MHC-DQ subregion, using overlapping cosmid clones, showed that it spans an area of 130kb, with DQA1, DQB1, DQA2 and DQB2 loci in the order mentioned (Wright and Ballingall, 1994). Expression of MHC glycoproteins on the surface of L cell transfectants was accomplished only for the genes designated DQ1 (Ballingall et al. 1992; Deverson et al. 1991; Wright and Ballingall, 1994). Up to five bovine DQA and B loci have been proposed and cattle haplotypes contain either single or duplicated DQ genes, with various combinations of those loci (Gelhaus et al. 1999a, b). This also seems to be the case for ovine MHC DQ genes, as revealed by more recent studies (Escayg et al. 1996; Hickford et al. 2000; Snibson et al. 1998). While the majority (82-89%) of the sheep

examined were found to carry one DQA1 and one DQA2 gene, 11-18% had no DQA1 locus (Escayg et al. 1996; Snibson et al. 1998). DQA1 null haplotypes were subsequently found to contain a dublicated DQA2 region (Hickford et al. 2000).

The ovine MHC class IIb subregion contains single copies of DNA, DOB (Scott et al. 1987; Liu et al. 2006), DYA and DIB genes (Wright et al. 1994; Liu et al. 2006) and the DM loci (liu et al. 2006). In cattle, LMP2, LMP7 and TAP2 genes have been also shown to be located in the BoLA class IIb subregion (Hess et al. 1999), but such data are not available for sheep.

1.3.2 Structure and function of ovine MHC molecules

Using monoclonal antibodies detecting monomorphic ovine MHC determinants and immunoprecipitation techniques, a class I heavy chain of 44kD in association with a β2-microglobulin of 12kDa has been identified (Gogolin-Ewes et al. 1985). For class II, an α chain of 30-32kDa and a β chain of 24-26 kDa have been described (Puri et al. 1985). Although minor structural differences between human and ovine MHC molecules have been described (Puri et al. 1987a, b), the overall structure of ovine MHC glycoproteins is similar to their human counterparts and to those described in other species (Gogolin-Ewes et al. 1985; Puri et al. 1985).

Despite the lack of specific studies showing that ovine MHC class I and II molecules act as restriction elements in sheep, it is assumed that, as in other species, ovine T lymphocyte responses are MHC restricted. Moreover, the tissue distributions of both class I and class II ovine MHC glycoproteins were found to be the same as those of HLA antigens (Gogolin-Ewes et al. 1985; Hopkins et al. 1986; Puri et al. 1985). In addition, MHC molecules are expressed within 35-40 days during development of the foetal sheep thymus and all T cell subsets found in human thymus are also present in sheep (reviewed by Mackay C. 1988). These findings indicate a role for the ovine MHC proteins in thymic T cell development similar to that in man and mouse.

1.3.3 Ovine MHC polymorphism

1.3.3.1 Serology

The sheep MHC system was first identified using serological reagents (Ford C. 1975; Millot P. 1978) and polymorphism of ovine MHC class I molecules has been studied mainly using

serology (reviewed by Spooner et al. 1988). Up to 31 provisional specificities controlled by co-dominant alleles from at least two loci have been described. Although a formal workshop for ovine MHC antigens has yet to be held, exchange of alloantisera among European laboratories resulted in the establishment and nomenclature of four specificities (OLA w1 to w4) (Cullen et al. 1985). Evidence that sheep lymphocyte antigens are encoded by ovine MHC class I genes was obtained by a study showing that OLA specificities co-segregated with genes that hybridized to cDNA probes for human MHC class I (Chardon et al. 1985).

A comparison between European (P) and Australian (SY) specificities showed that w1 to w3 were closely correlated (Outteridge et al. 1988) and four specificities were found in Spanish sheep, the two of which (CO1.1 and CO2.1) showed excellent correlation with alloantisera from Scotland (CA2 and CA5) (Garrido et al. 1995). Subsequently, alloantisera were produced for two Spanish breeds (Laxta and Rasa) and segregations of some of these ovine MHC antigens with MHC class II DRB1 SSCP patterns indicated that they were encoded within the MHC (Jugo and Vicario, 2001). In the latter study, two of the defined specificities (Bi2 and Bi5) corresponded to two specificities (CO3.1 and CO1.1) found by Garrido et al. (1995). More recently, over 600 alloantisera from 516 parous sheep from the Laxta and Rasa Spanish breeds were analyzed in an attempt to produce class I typing reagents (Jugo et al. 2004). Comparison of immunization with either whole blood or leucocytes revealed that a single injection of the dam with whole blood from a lamb is sufficient to produce alloantisera with high utility for class I antigen typing (Jugo et al. 2004), corroborating the observation of Nesse and Larsen (1987) in goats.

In contrast to class I typing, where serology has been the principal means of characterizing the system, serological typing for class II is limited, even for BoLA, despite its more advanced level of characterisation. This can be attributed both to the difficulty of developing serological reagents for class II antigens and advances in the development of more powerful DNA-based methods (Lewin et al. 1996).

1.3.3.2 Monoclonal antibodies, immunochemical and biochemical studies

Monoclonal antibodies against ovine MHC class I antigens were first produced by Gogolin-Ewens et al. (1985). Later, immunoprecipitation of class I products from a homozygous sheep, using those and an additional antibody, revealed three different class I molecules (Puri et al. 1987a). However, because these antibodies recognized monomorphic class I determinants, they

could not be used to study class I polymorphism. More recently, MHC class I polymorphism was studied biochemically in 65 Laxta sheep, by immunoprecipitation with monoclonal antibodies raised against human MHC class I molecules followed by 1D-IEF analyses (Jugo et al. 2002). Eighteen distinct IEF types were defined and comparison with serological specificities showed that some well-defined specificities corresponded to only single patterns, while three specificities segregated with more than one IEF pattern. Some IEF patterns did not correspond to any of the serological specificities. These findings suggest that at this stage of serological advancements, IEF typing detects a higher level of polymorphism than serology.

Monoclonal antibodies against ovine MHC class II antigens were first produced by Puri et al. (1985) and were used to study the tissue distribution of class II molecules. Later, a rat monoclonal antibody against sheep MHC class II molecules was used to study the distribution of these proteins in peripheral and efferent lymph (Hopkins et al. 1986). Subsequent studies characterized ovine MHC class II molecules immunochemically (Puri et al. 1987b) and provided evidence for four distinct class II subpopulations and allelic variation in α and β polypeptides (Puri and Brandon, 1987; Puri et al. 1987a). More recently, monoclonal antibodies specific for α and β class II chains (Dutia et al. 1990) and those that distinguish DR or DQ molecules (Ballingall et al. 1995; Dutia et al. 1993; 1994) have been produced, but they have not been applied in the study of MHC class II polymorphism.

1.3.3.3 DNA studies

Despite the characterization of sheep MHC class I polymorphism by serology, very few studies have been conducted at the DNA level. A study by Chardon et al. (1985) was the first to examine ovine MHC class I polymorphism by restriction fragment length polymorphism (RFLP), using human class I cDNA probes, and demonstrated the segregation of four haplotypes into five offspring of a single family. Segregation of four distinct polymorphic bands corresponded exactly to the inheritance of ovine MHC haplotypes from the ram and the ewe used. Comparison with serological typing showed that the RFLP patterns corresponded to the serological haplotypes.

However, as with other species, the occurrence of many pseudogenes and genes of undefined function in the MHC class I region make such genomic studies complicated. It is therefore prudent to restrict analysis to the cDNA level, especially during the early stages of characterization. At the outset of this study, only a single full length cDNA sequence (accession

number: M34676; Grossberger et al. 1990) was available in public databases, along with 10 partial cDNA sequences (M34672-M34675, U030092-U030094, AY188824-AY188826) and 13 exon 3 fragments (U25998-U26010). All were derived from only four research groups, only one of which had published their results in an international journal (Grossberger et al. 1990). These authors isolated thirteen class I cDNA clones from a sheep thymus library. Five different sequences were found, requiring the expression of at least three loci. One of the sequences was represented six times among the clones and, since no amplification took place during the construction of the cDNA library, this might have represented the major expressed product.

Due to the small number of ovine MHC class I sequences known to date and the limited number of family studies conducted, the only available PCR-based typing method for these genes to date is based on a microsatellite inside the MHC class I region (Groth and Wetherall, 1994; Gruszczynska et al. 2002). However, correspondence of the microsatellite alleles with MHC class I alleles is not known.

Polymorphism of ovine MHC class II has been studied to a much greater extent at the DNA level than class I genes. This has been accomplished by RFLP-Southern hybridization analyses (Grain et al. 1993; Escayg et al. 1996), Single Strand Conformational Polymorphism (SSCP) (van Oorschot et al. 1994; Snibson et al. 1998; Kostia et al. 1998; Jugo and Vicario, 2000; Tkacikova et al. 2005; Arrieta-Aguirre et al. 2006), Reference Strand Conformational Polymorphism (RSCP) (Feichtlbauer-Huber et al. 2000), PCR-RFLP (Konnai et al. 2003) and sequencing of cloned PCR or RT-PCR products, or clones from cDNA libraries (Fabb et al. 1993; van Oorschot et al. 1994; Wright et al. 1994; 1995; Snibson et al. 1998; Kostia et al. 1998; Feichtlbauer-Huber et al. 2000; Jugo and Vicario, 2000; Konnai et al. 2003; Herrmann et al. 2005).

1.3.3.4 Importance of ovine MHC polymorphism-balancing selection

As with other species, it has been suggested that sheep MHC polymorphism is maintained by balancing selection (Gutierrez-Espeleta et al. 2001; Paterson S. 1998; Paterson et al. 1998). An excess of non-synonymous over synonymous substitutions has been observed, especially in the peptide binding region (Gutierrez-Espeleta et al. 2001; Paterson S. 1998; Schwaiger et al. 1993). In addition, heterozygosity in the antigen binding sites (ABS) is over five-fold greater than elsewhere (Gutierrez-Espeleta et al. 2001). These findings imply strongly that the extent of variation is associated with function and therefore, that new allelic variants are favoured,

leading to an increase in MHC diversity. Furthermore, comparison of allelic sequences from six different Bovid species (Bighorn, domestic sheep, goat, bison, cow and musk-ox) showed that the MHC polymorphism is maintained by balancing selection over long periods of time, because many allelic lineages of one species are more similar to alleles of other species, than to other alleles of the same species (Gutierrez-Espeleta et al. 2001).

1.3.4 Ovine MHC disease association and immunization studies

Associations of sheep MHC specificities or alleles with disease resistance or susceptibility have been suggested for sheep nematode infections, including *Trichostrogylous colubriformis* (reviewed by Hohenhaus and Outteridge, 1995), *Ostertagia (Teladorsagia) circumcincta* (Buitkamp et al. 1996; Schwaiger et al. 1995; Stear et al. 1996), *Haemonchus contortus* (reviewed by Hohenhaus and Outteridge, 1995), as well as for scrapie (reviewed by Spooner et al. 1988), footrot (Escayg et al. 1997) and the Bovine Leukemia Virus (BLV)-induced ovine lymphoma (Nagaoka et al. 1999; Konnai et al. 2003). These studies can be divided into two main categories: Those in which natural infection occurred and comparisons of allele frequencies were made between afflicted and non-afflicted groups, and those where the pathogen was introduced experimentally by inoculation or immunisation. In the latter case, allele frequencies were compared between susceptible and resistant groups or between groups that produced high or low immune responses.

Caution is required in interpretation of such associations, because of various difficulties associated with the analyses (reviewed by Spooner et al. 1988). Problems in managed populations arise from sire effects and selective mating based on desirable traits, which affect the frequency of certain MHC alleles in the population. Moreover, due to the large number of available MHC alleles and the high level of heterozygosity, large numbers of animals are needed for statistically valid analyses. It is difficult for all, or even the majority of alleles in a population to be represented by a statistically adequate number of animals. For these reasons, it has been suggested to study susceptibility or resistance within sire families and lineages in relation to segregation of an MHC haplotype. The use of large numbers of sheep and confirmation of associations by several independent studies is also considered necessary to provide a vigorous assessment.

Notwithstanding the above considerations, a possible association of ovine MHC and resistance to gastrointestinal nematodes has been observed in Scottish Blackface sheep

(Buitkamp et al. 1996; Dukkipati et al. 2006; Schwaiger et al. 1995; Stear et al. 1996). These authors found that an ovine MHC DRB1 (G2) allele and an OLA class I specificity (G13br) were associated with reduced faecal egg counts (FEC) in sheep infected naturally with *O. circumcincta*. It was suggested that the DRB1 gene was more closely linked to the trait, and is therefore possibly the actual resistance locus (Stear et al. 1996). Chi-square analysis of typing data for both class I and II loci suggested that G2 and G13br are closely linked, a fact that further supports the association.

Subsequently, DRB1 alleles from the Texel and Suffolk sheep breed were also associated with FEC as a means of measuring resistance/susceptibility to gastrointestinal nematodes (Sayers et al. 2005). These authors suggested that Texels are more resistant than Suffolk because of a different allelic profile at the DRB1 locus. A recent microarray analysis provided further evidence for an association of ovine MHC class II genes and resistance or susceptibility to nematodes (Diez-Tascon et al. 2005). Among eight other genes, DRA, DQA1 and DQB1 were differentially expressed and, in particular, upregulated, in resistant sheep compared to susceptible animals.

The most striking association found to date for sheep MHC class I antigens is between SY1 (divided into SY1a and b and corresponds to European P3), an OLA class I specificity found in Australian sheep, and resistance to *T. colubriformis* (reviewed by Hohenhaus and Outteridge, 1995). This finding has been supported by several studies using different flocks (Outteridge et al. 1985; 1986; Windon and Dineen, 1984), including a family study, where sires and ewes were selected on the basis of the OLA specificity and bred to produce mainly homozygous lambs (Outteridge et al. 1988). Subsequently, the association of SY1 with low FEC for the nematode was also confirmed in a New Zealand breed (Douch and Outteridge, 1989). At the farm level, a test screening rams for SY phenotype has been established in Australia for use in combination with measurement of blood eosinophil levels (high levels indicate resistance) in both rams and ewes, in order to select for parasite resistance (Hohenhaus et al. 1996).

Identification of ovine MHC-disease associations can therefore inform the design of appropriate breeding schemes to increase the overall resistance of a sheep flock to a pathogen. Such studies are also driven by the fact that MHC is a source of variation in responsiveness to immunisation with potential vaccines (Hohenhaus and Outteridge, 1995). However, although T

cell antigen screening of peptides eluted from isolated human MHC molecules have lead to identification of candidate vaccine antigens (section 1.2.4.1.2), analogous studies have yet to be conducted for the ovine MHC. Such studies are constrained by the lack of molecular characterization of the ovine MHC class I region and lack of reagents necessary for analysis of class I MHC restricted T cell responses.

1.4 Aims and objectives of the project

The role of classical MHC molecules in initiation and regulation of immune responses and in the development of the T cell repertoire is well established. The function and extremely polymorphic nature of MHC genes therefore render them as candidate or marker loci for disease resistance with potential for use in selective breeding programmes. Characterisation of the MHC system in a species of interest is a prerequisite for the establishment of typing methods for screening populations for alleles that confer resistance to pathogens.

In the context of vaccine development, the pivotal role played by MHC products in determining the specificity of the immune response provides opportunities for identification of candidate antigens through the reverse genetic approaches described above. Application of these approaches is again dependent on precise information on MHC structure and polymorphism and the availability of cloned genes for expression in mammalian cells.

However, since very few studies have been conducted on ovine MHC class I genes to date (Chardon et al. 1985; Grossberger et al. 1990), this region is poorly characterized. Only limited sequence information is available and the number of expressed class I loci is disputable. In addition, sequence or PCR-based typing methods are unavailable and functional studies using sheep class I molecules to identify potential subunit vaccines are not possible. This study set out to address these gaps by characterising the expression of MHC class I genes in sheep and producing a set of cloned class I genes derived from individual haplotypes for expression *in vitro*.

The study focused on a flock of Blackface sheep. The Blackface breed is the most numerous breed in Britain, accounting for over three million ewes. This represents 16% of the British pure-bred ewe flock and the vast majority are found in Scotland (http://www.scottish-blackface.co.uk/). Two Blackface rams were selected from the flock for laboratory analysis. The project was structured in three sections:

1. Sequence and haplotype characterization of the transcribed ovine MHC class I alleles carried by the rams

This section aimed to identify all ovine MHC class I alleles transcribed by each ram, using single strand conformational polymorphism and sequencing (chapter 3). Based on

the identified partial ovine MHC class I sequences, it was aimed to establish a sequence-specific PCR-based typing system for these alleles (chapter 4). It was aimed to use this system to investigate how the parental alleles segregate among the progeny, and so allocate alleles to haplotypes (chapter 4). Such information would enable generation of sheep defined for a given MHC haplotype for future use in functional studies.

2. Sequencing and analysis of full length ovine MHC class I transcripts

The aim of this section was to obtain full length sequence data for the class I genes transcribed by the rams, in order to gain insight on their functionality and classical or non-classical nature (chapter 5). Sequence data could then be used to perform phylogenetic analysis of the identified sequences in this study and the ruminant class I sequences available in databases (chapter 6). This phylogenetic analysis would allow an evaluation of the number of classical and putative non-classical class I loci transcribed in sheep.

3. Generation of transfectants expressing OLA class I alleles carried by the rams

This section aimed to confirm the capacity of transcripts identified in the study to give rise to appropriately folded molecules on the cell surface. This was approached through generation of mammalian expression constructs for individual transcripts and analysis of transiently transfected L cells for surface expression using monoclonal antibodies and flow cytometry (chapter 7).

Chapter 2. General Materials and Methods

2.1 Animals

Twenty four Blackface sheep from a resource flock established at the Moredun Research Institute were used in this study. Specifically, 2 rams, 5 ewes and 17 lambs were used, the pedigree of which is presented in Table 2.1.

Ram	Ewe	Lamb	Ram	Ewe	Lamb
501	894N	448	504	888N	457
501	306N*	449	504	926N)	529
501	877N*	461	504	926N	530
501	822N*	486	504	922N	573
501	219N	489	504	596N	596
501	740N	518	504	1098N*	609
501	329N	555	504	265N*	610
501	808N	637	504	286N	673
501	197N	680			

Table 2.1: Pedigree of animals used in this study. * Ewes used in this study.

2.2 Chemicals, solutions and plastic ware

All chemicals were of analytical or molecular biology grade from BDH (BDH Laboratory Supplies, Poole, UK), Sigma (Sigma-Aldrich Co. Ltd., Poole, UK), Difco (BD Diagnostic systems, USA) and Bioline (Bioline, London, UK).

All stock solutions and buffers were prepared with MilliQ ion-exchange purified water (Purelab ultra 2, ELGA Process Water, Derbyshire, UK), sterilized DEPC treated water (Appendix 1.1) for RNA based procedures and sterilized distilled water for DNA based procedures. All laboratory plastic used in nucleic acid preparation and analysis was sterile, disposable and DNase/RNase free. Materials were sterilized by autoclaving at 121°C for 15 min.

2.3 Preparation of nucleic acids

2.3.1 Preparation of messenger RNA

Immediately after preparation of peripheral blood mononuclear cells (PBMC; section 2.9.2), poly-adenylated messenger RNA (mRNA) was extracted using the Dynabeads mRNA direct kit (DYNAL ASA, Oslo, Norway), according to the manufacturer's instructions. The kit relies on the base pairing between the poly A tail of messenger RNA and the oligo dT sequence bound to the Dynabeads surface. Dynabeads are uniform, magnetic polymer particles, which rapidly and efficiently bind the target RNA which is subsequently purified by binding to a magnetic particle concentrator (MPC).

Briefly, $4x10^6$ PBMC were resuspended in lysis/binding buffer and the lysate was passed through a 21 gauge needle three times to shear the DNA. The sheared lysate was then added to resuspended Dynabeads Oligo (dT)25. Following binding the mixture was applied to the MPC and the supernatant was removed. The dynabeads were then washed three times and 0.2-0.6μg purified mRNA were eluted in a final volume of 20μl 10mM Tris-Cl (Appendix 1.1).

2.3.2 Preparation of total cellular RNA

Total cellular RNA was extracted from PBMC using the Qiagen RNeasy mini kit (Qiagen Ltd, West Sussex, UK), according to manufacturer's instructions. PBMC were stored at -80°C in the RNA stabilization reagent RLT (Qiagen Ltd, West Sussex, UK). This reagent contains guanidine isothiocyanate (GITC) that is a strong denaturant causing simultaneous disruption of cells and inactivation of ribonucleases. Total RNA is bound to a silica gel membrane at a high salt concentration and contaminants are removed by washing. Any contaminating genomic DNA is removed by digestion with DNAase I and RNA is subsequently eluted under low salt concentration.

Briefly, 1×10^7 PBMC in 600μl RLT were thawed and incubated at 37°C for 10 min. The lysate was then passed 6 times through a 20-gauge needle fitted to an RNase-free syringe. One volume of 70% ethanol was added to the homogenized lysate and mixed well by pipetting. The sample was applied to an RNeasy mini column and centrifuged at 8000g for 15 sec. DNA was digested on column using 30U DNAase I. Columns were washed with ethanol containing buffer (RPE) and finally, total RNA was eluted in 30μl RNase-free water (Appendix 1.1).

2.3.3 Preparation of plasmid DNA

Plasmid DNA was recovered from 5ml of Luria–Bertani (LB) medium (Appendix 1.2) inoculated with the bacterial colony of interest and incubated overnight at 37°C under shaking. The QIAprep spin miniprep kit (Qiagen Ltd, West Sussex, UK) was used for plasmid recovery according to manufacturer instructions. In the QIAprep system, bacteria are lysed under alkaline conditions and the lysate is neutralized and adjusted to high salt binding conditions. The alkaline conditions ensure that only the plasmid DNA is released and denatured, while rapid addition of acidic buffer precipitates cell debris where chromosomal DNA is entraped. After high speed centrifugation, the supernatant is applied to a silica gel membrane column that selectively binds DNA in the presence of high salt. The plasmid DNA is then eluted under low salt concentration.

Briefly, pelleted bacteria were resuspended, lysed for 5 min and neutralized. Cell debris with entrapped chromosomes were precipitated by centrifugation at 10000 g for 10 min in a benchtop microcentrifuge and the supernatant was applied to a silica-gel membrane. After washes and removal of ethanol, the plasmid DNA was eluted in 30µl 10mM Tris-Cl (Appendix 1.1).

2.4 Polymerase chain reactions

2.4.1 Amplification of DNA fragments by polymerase chain reaction (PCR)

PCR, initially developed by Saiki et al. 1988, is a primer-directed in vitro reaction for enzymatic synthesis of a specific DNA fragment. It is based on the annealing and subsequent extension of two oligonucleotide primers that flank the target DNA sequence. Repeated cycles of denaturation, primer annealing and extension result in the exponential accumulation of a specific fragment as the products created in one cycle serve as template in subsequent cycles (Erlich H. 1999). The particular conditions and primers for amplification of ovine MHC class I genes are described in each chapter separately. All primers were designed using the programs PrimerSelect (DNASTAR, Inc. USA) and Primer3 (Rozen and Skaletsky, 2000; http://frodo.wi.mit.edu/primer3/primer3_code.html.). All PCR reactions were carried out in a 96 well Hybaid MBS 0.2S thermal cycler (Thermo Hybaid, Basingstoke, UK).

2.4.2 cDNA synthesis by reverse transcription (RT)-PCR

The mRNA (section 2.3.1) or total RNA (section 2.3.2) was converted into cDNA immediately

after isolation, using the Superscript first-strand synthesis system for RT-PCR (Invitrogen Ltd, Paisley, UK), according to the manufacturer's instructions. First strand cDNA was synthesized using oligo (dT) as the primer and catalysed by Superscript II RNase H- reverse transcriptase (RT). This enzyme has been engineered to eliminate RNase H activity thereby increasing the yield of full length cDNA transcripts.

Up to 0.3μg of mRNA or 5μg of total RNA were mixed with 10mM dNTPs and 0.5μg oligo$(dT)_{12-18}$ in a final volume of 10μl and, incubated at 65°C for 5 min and then placed on ice. The RT reaction mixture (9μl containing 10xRT buffer, 100mM $MgCl_2$, 0.2 M DTT and 40U RNase inhibitor) was mixed with the primer/RNA mixture and incubated at 42°C for 2 min. Fifty U (1μl) of Superscript II RT were added and reverse transcription was conducted at 42°C for 50 min. The reactions were terminated at 70°C for 15 min and chilled on ice. Finally, the RNA was degraded by 2U (1μl) of RNase H at 37°C for 20 min.

2.4.3 Rapid amplification of cDNA ends (RACE)

A 3'-RACE technique was used to determine the unknown nucleotide sequence between a known internal sequence and the poly(A) mRNA tail. Initially, mRNA is converted into cDNA using reverse transcriptase and an oligo-$d(T)_n$ adapter primer. The adapter primer and a gene specific primer are then used for PCR amplification of the cDNA of interest. To reduce non specific amplification a second nested gene specific primer was used with the adapter primer (Frohman and Martin, 1989).

In this study mRNA was converted into cDNA, using the Superscript first-strand synthesis system for RT-PCR (Invitrogen Ltd, Paisley, UK), as described in section 2.4.2, with a slight modification. The oligo-dT primer containing a 36bp adapter incorporated at its 3' prime end was used instead of the conventional Oligo$(dT)_{18}$. The primer sequences and PCR conditions for 3'-RACE of ovine MHC class I genes are described in detail in chapter 5.

2.4.4 Sequence specific primer (SSP)-PCR

This PCR based method for amplification of specific sequences or alleles at polymorphic loci relies on the principle that completely matched primers amplify more efficiently than mismatched primers especially at the 3 prime end. The development and application of the SSP-PCR technique for the amplification of ovine MHC class I genes is described in detail in chapter 4.

2.5 Analysis of nucleic acids

2.5.1 Spectophotometric quantification of nucleic acid

Nucleic acid concentrations were routinely estimated using a Beckman DU-650 spectrophotometer (BECKMAN, USA). The optical density (OD) at 260 was used to measure the nucleic acid concentration. An OD of one equates to an RNA concentration of 40μg/ml and a DNA concentration of 50μg/ml. The OD_{260}/OD_{280} absorbance ratio gives an estimate of the purity of the sample. A ratio greater than 1.8:1 indicates that the sample is relatively free from contamination (Sambrook et al. 1989).

2.5.2 Restriction endonuclease digestion of DNA

Restriction endonucleases from Roche (Roche Molecular Biochemicals, Mannheim, Germany), Promega (Promega Corporation, Madison, USA) and New England Biolabs (New England Biolabs, Beverly, USA) were used according to manufacturer instructions. Restriction endonucleases are degradative enzymes found in bacteria that recognize and cleave DNA at sites with a specific nucleotide sequence (Sambrook et al. 1989).

Generally, 1-2μg of DNA was digested with 5 units of a restriction enzyme in a total reaction volume of 20μl containing the appropriate enzyme buffer provided by the manufacturer. Incubation times varied from 30 min to 2 hours.

2.5.3 Agarose gel electrophoresis

Agarose, a linear polymer extracted from seaweed forms a gel-like matrix that is commonly used to separate DNA or RNA fragments by size when an electric field is applied (Sambrook et al. 1989). Ethidium bromide is used to stain the DNA fragments which can be visualised under ultraviolet light (Sambrook et al. 1989).

0.7, 1.2 and 2% horizontal agarose gels were prepared by dissolving 0.7, 1.2 and 2g of agarose (Bio-Rad Laboratories, Hercules, USA), respectively, in 100ml of 1X TBE buffer (Appendix 1.1). The agarose was dissolved using a microwave and 1μg/ml ethidium bromide was added prior to pouring into a horizontal gel tray with an appropriate sized comb. Once set, the comb is removed, thus forming wells into which the sample was added. Samples were loaded with an appropriate volume of 6x bromophenol blue loading buffer (Appendix 1.1) in a final volume between 6 and 18μl. Elecrophoresis was conducted under 1X TBE buffer

containing 1μg/ml ethidium bromide under constant voltage (80-150V) depending on the size of the gel.

2.5.4 Isolation of DNA fragments from agarose gels

PCR products of the appropriate size were purified from 0.7% agarose gels (Thermo Hybaid, Basingstoke, UK), using the QIAquick gel extraction kit (Qiagen Ltd, West Sussex, UK). In the QIAquick system, a silica gel membrane column binds DNA in the presence of high salt while the contaminants pass through. Impurities are washed away and the pure DNA is eluted under low salt concentration.

Briefly, a gel slice containing the appropriate band was excised and melted at 50°C for 10 min in high salt buffer. The solution was applied to a QIAquick column and centrifuged for 1 min. After washes the DNA was recovered in 30μl of 10mM Tris-HCl buffer.

2.5.5 Quantification of DNA by comparison with known concentrations of marker DNA

The concentration of DNA fragments was estimated on 2% agarose gels, either by using different concentrations (ranging from 10ng-500ng) of the Promega 100bp DNA ladder (Promega Corporation, Madison, USA), or by using 1, 5 and 10μl of the quantification DNA marker Hyperladder IV (Bioline Ltd, London, UK).

2.5.6 Analysis of plasmid DNA

To excise cloned DNA fragments from plasmid vectors in order to confirm the presence and size of the insert, 5μl of plasmid DNA was digested using the restriction enzymes *Eco* RI (pGEM-T Easy and pCR4Blunt-TOPO) or *Bam* HI and *Apa* I (pcDNA3.1/V5-His-TOPO). The digested DNA was separated by agarose gel electrophoresis using a 2% agarose gel (section 2.5.3) and the fragments visualized under ultraviolet light.

Digestions with *Eco* RI were performed at 37°C for 30 min in enzyme buffer H (Promega Corporation, Madison, USA), while digestions with *Bam* HI and *Apa* I were performed at 30°C for 30 min and then at 37°C for another 30 min in enzyme buffer A (Roche Molecular Biochemicals, Mannheim, Germany).

2.5.7 Ethanol precipitation of DNA

Plasmid DNA was purified and concentrated using ethanol precipitation. Three M sodium acetate (Appendix 1.1) was added to the DNA solution to a final concentration of 0.3M along with 2.5 volumes of absolute ethanol (Fisher Scientific Ltd, Leicestershire, UK) in a conical based sterile microcentrifuge tube. Following incubation at -20°C for 30 min samples were centrifuged at 10000g for 5 min in a benchtop microcentrifuge and the supernatant removed. The precipitated DNA was washed with 70% ethanol to remove excess acetate salts and centrifuged for 1 minute at 10000g. Any remaining ethanol was removed by pipette and the DNA was dried for 5 min at room temperature before being resuspended in the appropriate volume of sterile distilled water to a final concentration of 0.5μg/μl.

2.5.8 Single stranded conformational polymorphism (SSCP) analysis

SSCP analysis was used to genotype cloned ovine MHC class I exon 2 and 3 fragments. SSCP analysis is optimal for DNA fragments amplified by PCR of less than 400bp. The PCR product is denatured and rapidly cooled so that each strand folds into a unique conformation that is determined by its sequence. The folded fragments are then separated using non denaturing acrylamide gels (Hayashi and Yandell, 1993) and visualized by silver staining.

Ten μl of each PCR product were mixed with 5μl formamide denaturing solution (Appendix 1.1), denatured at 94°C for 8 min and immediately cooled on ice for 5 min. Ten μl were loaded onto a 6% polyacrylamide gel (Appendix 1.1) without glycerol. Electrophoresis was carried out using a Hoefer SE 600 vertical gel electrophoresis unit (Amersham Biosciences, Buckinghamshire, UK) with 1XTBE running buffer (Appendix 1.1) for 16h at 14mA at a constant temperature of 4°C and then for an additional 1h at 50mA.

Single strand DNA fragments were detected by silver staining (Ainsworth et al. 1991). Silver stains because of the reduction of silver ions complexed with DNA under reducing conditions provided by formaldehyde. The bands were fixed with a solution of 40% ethanol and 10% glacial acetic acid (GAA) for 1 hour, washed twice with 10% ethanol for 15 min and then three times with distilled water for 15 min. The gel was stained with 0.2% silver nitrate for 30 min and then washed twice with distilled water for 30 sec. The samples were developed with a solution of 2.5% sodium carbonate and 0.1% formaldehyde and the developing was terminated with 1% GAA.

2.5.9 Restriction fragment length polymorphism (RFLP) analysis

Restriction fragment length polymorphism is a method for discriminating between polymorphic sequences using the patterns of DNA fragments produced following digestion with restriction enzymes. Specifically, 10µl of each PCR product were digested for 2h at 37°C with *Alu* I or *Bam* HI in enzyme buffer B (Roche Molecular Biochemicals, Mannheim, Germany), or with each of the enzymes *Bam* HI and *Rsa* I in enzyme buffer C (Promega Corporation, Madison, USA).

For *Alu* I and *Bst* YI combined digestions, 10µl of each PCR product were initially digested for 2h at 37°C with restriction enzyme *Alu* I and then for 1h at 60°C with *Bst* YI in enzyme buffer B (New England Biolabs, Beverly, USA).

The digested fragments were separated using 2% agarose gels or 6% polyacrylamide gels (Appendix 1.1) stained with ethidium bromide and visualised under UV light.

2.5.10 Polyacrylamide electrophoresis

Acrylamide/bisacrylamide solution (29:1 ratio) polymerizes in the presence of free radicals supplied by ammonium persulphate and the catalyst tetramethylethaldiamine (TEMED). Polyacrylamide has much greater resolving power than most forms of agarose, since it can discriminate DNA fragments differing in length by a single nucleotide (Sambrook et al. 1989). Polyacrylamide gels of 5% were used for sequencing, and 6% for RFLP and SSCP analysis (Appendix 1.1).

2.6 Cloning into plasmid vectors

2.6.1 Vectors

The pGEM-T Easy vector (Promega Corporation, Madison, USA) and the pCR4Blunt-TOPO vector (Invitrogen Ltd, Paisley, UK) were used for cloning PCR products prior to sequencing (Appendix 1.3). Mammalian expression constructs were made by cloning into the pcDNA3.1/V5-His-TOPO plasmid vector (Appendix 1.3).

The pGEM-T Easy plasmid is a 3015bp vector designed to facilitate cloning of PCR products. The vector contains the lac operator and the lacZ gene for blue/white screening (section 2.7.3), the phage f1 origin of replication and the ampicillin resistance gene for the selection of transformed colonies. The vector also contains the T7 and SP6 RNA polymerase

promoter and transcription initiation sites and the M13 forward and reverse sequencing primer binding sites.

The pCR4Blunt-TOPO plasmid is a 3957bp vector designed to facilitate cloning and sequencing of blunt end PCR products. The vector contains the lac promoter and the ccdB lethal gene fused to the C-terminus of the LacZa fragment. Ligation of a blunt-end PCR product disrupts expression of the LacZa-ccdB gene permitting growth of only positive recombinants, because cells that contain the non-recombinant vector are killed upon plating due to the expression of the lethal ccdB gene (Bernard and Couturier, 1992; Bernard et al. 1993; 1994). The vector also contains the puC origin of replication, the ampicillin and kanamycin promoters and resistance genes for selection of transformed cells and the T3, T7, M13 forward and M13 reverse sequencing priming sites.

The pcDNA3.1/V5-His-TOPO plasmid is a 5.5kb expression vector designed to facilitate rapid directional cloning of blunt-end PCR products for expression in mammalian cells. The vector contains the human cytomegalovirus (CMV) enhancer/promoter for high level constitutive expression of the cloned gene in a wide range of mammalian cells (Boshart et al. 1985; Nelson et al. 1987) and the SV40 early polyadenylation signal for efficient termination of the transcription and polyadenylation of the transcript (Maxwell et al. 1989). The vector also contains the V5 epitope, a 14 amino acid epitope derived from the P and V proteins of the paramyxovirus SV5 (Shouthern et al. 1991), and the C-terminal polyhistidine tag (Lindner et al. 1997) for detection and purification of the recombinant protein. The T7 and BGH priming sites of the vector can be used for sequencing and the pUC origin allows high copy number replication in *E. coli*. Other elements of the vector include the bla promoter and amplicillin resistance gene for selection in E. coli and the promoter origin of replication and early polyadenylation signal along with the neomycin resistance gene for selection of stable transfectants in mammalian cells (Shouthern and Berg, 1982).

2.6.2 TA cloning

TA cloning vectors contain 3'-T overhangs at the insertion site that anneal with 3'-A overhangs of A-tailed PCR products, produced with Taq polymerase without proofreading activity. T4 DNA ligase then creates a diphosphoesteric bond between the vector and the insert.

Up to 7µl of PCR products (150-300ng) extracted from agarose gels were incubated at 70°C for 30min with 5U Taq polymerase and 0.2mM dATP in a final volume of 10µl. The A-

tailed PCR products were then ligated into the pGEM T Easy vector according to the manufacturer instructions. Briefly, one μl of the A tailed product (15-30ng) was mixed with 1μl (50ng) of the vector, 5μl of 2x rapid ligation buffer and 1μl (3U) T4 DNA ligase in a final volume of 10μl and incubated overnight at 4°C. The reaction was then placed on ice prior to transformation of competent *E. coli*.

2.6.3 Blunt end cloning

Blunt end cloning involves the direct insertion of blunt-end PCR products into a plasmid vector using the enzyme topoisomerase I from the *Vaccinia* virus. Topoisomerase I binds to double stranded DNA at specific sites and cleaves the phosphodiester backbone after 5'-CCCTT in one strand of the vector (Shuman et al. 1991). The energy from the broken phosphodiester bond is used in the formation of a covalent bond between the 3' phosphate of the cleaved strand and a tyrosyl residue of topoisomerase I. The phosphotyrosyl bond between the DNA and the enzyme is subsequently attacked by the 5'-hydroxyl of the original cleaved strand, reversing the reaction and releasing topoisomerase (Shuman K. 1994). This way, a PCR product can be quickly cloned into a vector without post PCR procedures and ligase.

PCR products were cloned into the pCR4Blunt-TOPO or the pcDNA3.1/V5-His-TOPO vector according to the manufacturer's instructions. Up to 4μl of PCR products were mixed with 10ng of the vector and 1μl of salt solution (1.2 M NaCl, 0.06 M $MgCl_2$) in a final volume of 6μl and incubated for 30 min at room temperature. The reaction was then placed on ice prior to transformation of competent *E. coli*.

2.7 Transformation and analysis of transformed E. coli

2.7.1 Competent cells and growth media

For the purposes of this study, high efficient JM109 (Promega Corporation, Madison, USA) and TOP10 (Invitrogen Ltd, Paisley, UK) chemically competent *E. coli* were used to select and expand recombinant plasmid vestors. Bacteria were cultured in Luria–Bertani (LB) medium (Appendix 1.2) or on LB agar plates (Appendix 1.2) supplemented with 50μg/ml ampicillin (Sigma-Aldrich Co. Ltd., Poole, UK).

2.7.2 Transformation

Bacterial transformation is the process where small pieces of extracellular DNA are taken up and incorporated into the genetic material of a living competent bacterium (Klug and Cummings, 2000).

Two μl of each cloned PCR product was added into 50μl of just thawed competent cells, mixed gently and placed on ice for 20 min. The cells were then heat-shocked for 50 sec at 42°C and immediately placed on ice for 2 min. Room temperature SOC medium (950μl or 250μl; Appendix 1.2) was added into the JM109 or the TOP10 competent cells, respectively, and the mixture was incubated for 1.5 hours at 37°C with shaking.

2.7.3 Selection and growth of antibiotic resistant colonies

The transformed cells were spread out onto LB agar plates supplemented with 50μg/ml ampicillin (Sigma-Aldrich Co. Ltd., Poole, UK), 100mM isopropyl-β-D-thiogalactopyranoside (IPTG) (Bioline, London, UK) and 40mg/ml 5-bromo-4chloro-3-indolyl-b-D-galactosidase (X-Gal, Bioline, London, UK) which allows selection for white and blue colonies depending on the presence or absence of a cloned insert, respectively.

The blue/white screening is based on interruption of the expression of the *Lac Z* gene, because of the insertion of the cloned DNA fragment. The *Lac Z* gene encodes the β-galactosidase enzyme, which converts lactose into glucose and galactose. This gene is induced when the medium contains lactose. Isopropylthiogalactoside (IPTG) is a chemical analogue of lactose and behaves as a natural inducer, but does not serve as substrate. Instead, a sugar called X-Gal is hydrolysed by β-galactosidase and produces a blue colour. Therefore, cells (cultured on LB agar plates with ampicillin/IPTG/X-Gal) without an insert are induced by IPTG to express the *Lac Z* gene and produce blue colonies, whereas cells containing an insert interrupt the expression of *Lac Z* and produce white colonies (Klug and Cummings, 2000).

2.7.4 Selection of transformed colonies by PCR

White colonies were picked and processed for colony PCR as described by Hawker and Billadello (1993). Briefly, each colony was picked into 100μl sterile water, heated at 98°C for 5 min and cooled immediately. Twenty μl of the lysate were mixed with 5μl PCR reaction mixture (50mM KCl, 10mM Tris-HCl, 1.5mM $MgCl_2$, 0,1% TritonX-100, 200μM of each

dNTP, 200nM of each of the primers and 1 unit Taq polymerase (Promega Corporation, Madison, USA). The thermal cycling profile comprised an initial denaturation step of 5 min at 94°C, followed by 30 cycles of 1 min at 94°C, 2 min at 55°C and 4 min at 72°C.

2.7.5 Preparation of glycerol stocks

Bacterial clones carrying the DNA fragment of interest were stored as glycerol stocks. Each colony was picked and cultured overnight under shaking in 5ml LB medium (Appendix 1.2) with 50μg/ml ampicillin. One ml from the culture was centrifuged at 10000 g for 5 min in a benchtop microcentrifuge. The pelleted cells were then resuspended in 0.5ml of 75% glycerol in LB.

2.8 Nucleotide sequence analysis

2.8.1 Sequencing reactions

The basis of nucleotide sequencing is a PCR based extension of a primer annealed to the DNA template up to the point of termination of the reaction by incorporation of a dideoxy nucleotide (ddNTP) (Sanger et al. 1977). A ddNTP lacks the hydroxide group necessary for generation of the phosphodiesteric bond by the DNA polymerase. In the reaction mixture, dNTPs and ddNTPs are used in a ratio so that all the range of different DNA fragments of up to around 800bp length are generated. Those fragments are then separated on a polyacrylamide gel.

In this study nucleotide sequencing was performed, using a 377 ABI PRISM DNA sequencer (Perkin Elmer, USA). Plasmid template DNA (250-500ng) was sequenced with 3.2pmols of primers and reagents from the ABI-PRISM big dye sequencing kit. This system uses the dideoxy chain termination method with four color fluorescent primers and the DNA polymerase FS. This enzyme is a variant of *Thermus aquaticus* (Taq) DNA polymerase that makes it less discriminating against ddNTPs leading to a more even pick intensity pattern. In addition, it contains a mutation that eliminates the 5'-3' exonuclease activity. Four separate reactions are performed each one terminated at a different base and primed by a different colour dye primer. Ovine MHC class I genes inserted in the pGEM-T Easy vector were sequenced with the T7 and SP6 primers. Class I genes inserted in the pCR4Blunt-TOPO vector were sequenced with primers M13frw and M13rev and those inserted in pcDNA3.1D/V5-His-TOPO vector were primed by T7 and BGH. The sequence of each of these primers is shown in table 2.2.

Forward Primers		Reverse Primers	
Name	*5' Sequence 3'*	*Name*	*5' Sequence 3'*
T7	TAATACGACTCACTATAGGG	SP6	GATTTAGGTGACACTATAG
M13frw	GTAAAACGACGGCCAG	M13rev	CAGGAAACAGCTATGACC
T7	TAATACGACTCACTATAGGG	BGH	TAGAAGGCACAGTCGAGG

Table 2.2: Primers used for sequencing

PCR cycling conditions used to generate the DNA fragments for sequence analysis were 30 cycles of 10 sec at 96°C, 5 sec at 50°C and 4 min at 60°C and conducted on the Applied Biosystems 9700 thermal cycler. The amplified products were precipitated with 1µl 3M sodium acetate and 25µl 95% ethanol. The DNA was resuspended in 2µl of sequence loading dye (Appendix 1.1) and loaded into 5% polyacrylamide gels containing 6M Urea (Appendix 1.1).

2.8.2 Sequence analysis

The ABI PRISM® 377 DNA Sequencer automatically analyzes DNA molecules labeled with multiple fluorescent dyes. After samples are loaded onto the system's vertical gel, they undergo electrophoresis, laser detection, and computer analysis.

Base calling from the raw sequencer output was performed using the sequencer's data collection and sequence analysis software (Applied Biosystems, USA). The traces of sequences obtained from both forward and reverse primers were visually checked, assembled and edited if necessary, using the SeqMan program (DNASTAR, Inc. USA).

2.8.3 Generation of multiple sequence alignments

A multiple alignment algorithm generates a guide tree from a distance matrix. This matrix contains pairwise distances calculated for all combinations of pairs of sequences used. Each pairwise alignment is given a score that arises from positive values for identities and negative values for mismatches and gaps. The sequences are then progressively aligned according to the branching order of the guide tree (Thomson et al. 1994). Pairwise distances used for the generation of the guide tree are not corrected for multiple substitutions.

In this thesis, multiple alignments were generated using the Clustal W (Thomson et al. 1994) or the Clustal V method (Higgins and Sharp, 1989) in the DNAStar software program

MegAlign (DNASTAR, Inc. USA). Clustal W is an advanced version of Clustal V, designed to generate more accurate alignments from highly diverged sequences. Both methods do not assume that sequences should be connected by descent. Although MHC class I sequences analyzed were closely related, Clustal W produced better results compared to Clustal V, in cases where the lengths of available and obtained sequences varied significantly.

2.8.4 Estimation of sequence distances

When two DNA sequences have derived from a common ancestral sequence, the descendant sequences gradually diverge by nucleotide substitution. However, the number of different nucleotides present between two sequences underestimates the actual distances, because it does not take into account backward and parallel substitutions (Nei and Kumar, 2000). Therefore several models that calculate sequence distances by correcting for multiple substitutions have been developed (Jukes and Cantor, 1969; Kimura M. 1980; Tajima and Nei, 1984; Tamura K. 1992; Tamura and Nei, 1993).

It has been observed that transitions and transversion rates differ during evolution and therefore this difference should be taken into account when calculating sequence distances. Transitions are nucleotides changes between purines (A→G or G→A) or pyrimidines (C→T or T→C), whereas tranversions are changes between a purine and a pyrimidine. In this study, nucleotide sequence distances were calculated by correcting with the Kimura two parameter method for multiple substitutions (Kimura M. 1980). This model assumes that the rate of transitional substitution per site per year (α) is different from that of transvertional substitutions (2β).

Apart from the different transition and transversion rates, other models that calculate sequence distances take into consideration the GC content bias because nucleotide frequencies are rarely equal to each other (Tajima and Nei, 1984; Tamura K. 1992; Tamura and Nei, 1993). However, if sequences are closely related and their estimated p distances (number of different nucleotides/total number of nucleotides) are lower than 0.25, as in the case of the analyzed MHC class I sequences in the current study, such models do not give significantly different results. In such cases a simple model or calculation of p distances is preferred especially for construction of phylogenetic trees, because it has a smaller variance (Nei and Kumar, 2000).

In protein coding genes, the first, second and third codon position have different substitution rates. In particular, it has been suggested that the rate variation at different

nucleotide sites approximately follows the gamma distribution and sequence distances are better estimated by the Kimura model for gamma distances (Jin and Nei, 1990). Nevertheless, the effect of variation of each codon is important only when p distances are greater than 0.2, because gamma distances have a larger variance (Nei and Kimura, 2000). Since all pairwise p distances for the analyzed sequences in this study were <0.2, genetic distances were calculated by the Kimura two parameter model, without corrections for gamma distribution.

Aminoacid sequence distances were calculated by the Poisson correction distance (Nei and Kumar, 2000). This model calculates amino acid distances based on the assumption that the rate of amino acid substitution per year at a particular amino acid site is the same for all sites. Practically, this estimation is good unless p (number of amino acid differences/total number of amino acids) is higher than 0.3. Since all amino acid pairwise p distances for the sequences analyzed in this study were smaller or very close to 0.3, the Poisson correction (PC) distance was a good estimate to use.

All distances were calculated using the Molecular Evolutionary Genetics Analysis software (Kumar et al. 2001; MEGA2.1 Arizona State University, Tempe, Arizona, USA).

2.8.5 Phylogenetic analysis

Phylogenetic trees were constructed using the Neighbor Joining (NJ) method developed by Saitou and Nei (1987). NJ is a simplified minimum evolution method, where instead of calculating the sums of all branch lengths for all possible topologies to choose the smallest, the sums are calculated at each stage of taxon clustering. The NJ method sequentially identifies neighbor pairs that minimize the total length of the tree. Genetic distances calculated for the construction of phylogenetic trees were calculated either by the Kimura two parameter method for full length sequences or by p distances for exons 4 through 8 that exhibited lower variability (section 2.8.4).

2.9 Cell culture

2.9.1 Aseptic technique

All cell culture treatments were performed under aseptic conditions in a UniMat-BS class II microbiological safety cabinet (Medical Air Technology LTD, Manchester, UK) and all plasticware, solutions and media used were sterile.

2.9.2 Preparation of peripheral blood mononuclear cells (PBMC)

Preparation of PBMC was conducted using a standard Ficoll-Hypaque (Pharmacia, Uppsala, Sweden) separation method (Boyum A. 1968). Ficoll-Hypaque is a sterile density gradient medium for purifying lymphocytes in high yield and purity from small or large volumes of human peripheral blood using a simple and rapid centrifugation procedure. This method was slightly modified for sheep PBMC.

Twenty ml of peripheral blood were collected in vacutainers with 10U/ml sodium heparin and centrifuged at 1000g for 15 min in a Sorvall RT7 plus table top refrigerated centrifuge at 4°C (GMI, Inc. Minnesota, USA). The buffy coat was recovered and suspended in 20ml phosphate buffer saline (PBS; Appendix 1.4) supplemented with 10U/ml sodium heparin and the mixture was layered over 20ml Ficoll-Hypaque (2 sets of 10ml) and centrifuged at 650g for 25 min at room temperature. The transparent interphase band containing the mononuclear cells was recovered and washed three times with 10ml of PBS-heparin. Each wash was followed by a centrifugation at 350g for 10 min. All centrifugations were conducted at room temperature. Cells were counted using an Olympus CK40-F200 light microscope (Olympus optical ltd, Japan), using a haemocytometer (section 2.9.4) and the volume was then adjusted to 1 or 2×10^6 cells/ml with PBS. Cells were either used immediately or stored at -70°C. For storage, the cells were centrifuged at 350g for 10 min and the PBS was removed. PBMC were then lysed in 600µl RNA stabilization Reagent RNAlater (RLT; Qiagen Ltd, West Sussex, UK) and kept at -70°C until used.

2.9.3 Cell counting

Mammmalian cells were counted using an improved Neubauer haemocytometer (Hawksley, London, UK), after mixture with an equal volume of 0.05% nigrosine which allows discrimination between live and dead cells (BDH Laboratory Supplies, Poole, UK). The haemocytometer is a glass microscope slide with a 5x5 grid engraved into it. When a glass coverslip is positioned, a predetermined volume lies over the grid. This allows the number of cells within the defined volume to be determined using a microscope. By taking any dilution factor into consideration, the number of cells per ml can be determined.

2.9.4 COS-7 cells

The monkey fibroblast cell line COS-7, (provided by Dr. K. Ballingall, Moredun Research

Institute, Edinburgh, UK) was used for transient expression of ovine MHC class I genes. These kidney fibroblast cells originated from the African green monkey *Cercopithecus aethiops* by transformation with an origin defective mutant of simian virus SV40. COS cells carry a stably integrated portion of the SV40 genome that enables it to produce, the wild-type T protein (Gluzman Y. 1981). This protein triggers extrachromosomal replication of the virus by binding to its origin of replication. The plasmid vector pcDNA3.1/V5-His-TOPO (Appendix 1.2) carries the SV40 origin of replication allowing it to be replicated to high copy number after transfection into COS cells.

COS-7 cells were cultured in complete Iscove's Modified Dulbecco's Medium (IMDM; Sigma-Aldrich Co. Ltd., Poole, UK; Appendix 1.4), supplemented with 10% heat-inactivated foetal bovine serum (FBS; Egg Technologies, Brunswick, UK), 100U/ml penicillin, 100µg/ml streptomycin and 1.46% glutamine (Appendix 1.4). Cells were maintained either in T25 or T75 cm^2 tissue culture flasks (costar, Corning Incorporated, Corning, USA) at $37^{O}C$ under 5% CO_2 in a humidified incubator (Jouan incubator, Thermo Electron Corporation, Witchford, UK).

2.9.5 COS cell passage

COS-7 cells were passaged using a 1:5 split ratio just prior to confluence. Since COS-7 cells are adherent, after removal of the medium, the cells were washed three times with 10ml PBS (Appendix 1.4) pre-warmed to 37°C. The PBS was removed and 2.5ml of room temperature trypsin/versene solution (Appendix 1.4) was added. The flasks were incubated for 2 min at 37°C and the adherent cells removed by a sharp tap followed by the addition of 3.5ml of complete IMDM (Appendix 1.4). Two ml from the flask were then transferred into a new flask containing 8ml of complete IMDM medium (Appendix 1.4) pre-warmed at 37°C.

2.9.6 Retrieval of COS-7 cells from liquid nitrogen storage

A vial of COS-7 cells was removed from liquid nitrogen storage and defrosted in a waterbath at 37°C until just thawed. The cells were then transferred to a centrifuge tube and 10ml of chilled complete medium was added and mixed gently. After centrifugation for 5min at 350g at 4°C in a Sorvall RT7 plus table top refrigerated centrifuge (GMI, Inc. Minnesota, USA), the supernatant was removed and the cells were resuspended in 10ml complete IMDM medium warmed to $37^{O}C$ and transferred to a 75cm^2 sterile tissue culture flask (costar, Corning Incorporated, Corning, USA).

2.9.7 Cell cryopreservation

For long term storage COS cells were frozen under liquid nitrogen. Cells were removed from the tissue culture flasks as described in section 2.9.5, centrifuged at 350g for 10 min at 4°C and resuspended in 0.5ml FBS at a concentration of $0.5x10^6$ml. Half ml of 20% DMSO in FBS (Appendix 1.4) was added dropwise and the mixture was transferred to a 1.5ml cryopreservation tube. The tube was then placed in a freezing container insulated with isopropanol and stored at -70°C for 24 hours prior to long term storage in liquid nitrogen.

2.10 Gene transfer and expression analysis

2.10.1 Transfection

Ovine MHC class I genes cloned into the pcDNA3.1/V5-His-TOPO plasmid vector were transferred into COS-7 cells by transfection using the diethylaminoethyl (DEAE) -Dextran method (Sambrook et al. 1989). Dextran is a carbohydrate polymer and DEAE is a positively charged chemical group attached to the polymer. The DNA binds to the positively charged DEAE-dextran via its negatively charged phosphate groups and the polymer particles stick to the cell surface. The DNA/Dextran complex is taken up by the cell, and although some DNA is likely to be destroyed in the cytoplasm, a percentage enters the nucleus. DEAE-Dextran is the transfection method of choice for transient expression in COS cells, because it requires only small amounts of DNA and the toxicity of the polymer is not an issue, as cells are not kept for long periods (Sambrook et al. 1989).

A semiconfluent T75cm^2 flask of COS-7 cells grown in complete IMDM medium (Appendix 1.4) was divided into two T75cm^2 flasks (section 2.9.8), the day prior to transfection. On the day of transfection, the COS-7 cells were removed from the flasks using trypsin/versene solution, as described in section 2.9.8. Cells were pelleted by centrifugation at 350g for 5 min at room temperature in a final volume of 25ml complete IMDM medium prewarmed at 37°C. The pelleted cells were counted and resuspended in 15ml Tris-RPMI medium (Sigma-Aldrich Co. Ltd., Poole, UK; Appendix 1.4) at a concentration of $5x10^5$ cells per ml.

Seventy five µl of 1mg/ml DEAE-Dextran (Sigma-Aldrich Co. Ltd., Poole, UK) was added to 1ml of Tris-RPMI and the Dextran/Tris-RPMI mixture was then added to 10µg of plasmid DNA dropwise with constant agitation. The DNA/Dextran/RPMI mixture was mixed with the 1ml of cells and incubated at 37°C for 2-4 hours with occasional inversion. The cells were then washed with 10ml Tris-RPMI, resuspended in 2ml complete IMDM medium

prewarmed to 37°C, plated out into one well of a 6 well plate and incubated for 48-72h at 37°C. After incubation, the cells were removed from the wells and analyzed for MHC class I surface expression by indirect immunoflorescence (section 2.10.2).

2.10.2 Indirect immunofluorescence

An indirect immunofluorescent technique was used to detect ovine MHC class I molecules expressed on the surfaces of transfected cells. Ovine MHC class I molecules expressed on the surface of COS-7 cells were detected using the murine monoclonal antibody IL-A88, which recognizes a common monomorphic determinant found on bovine and ovine MHC class I molecules (Toye et al. 1990; Chai Z. 2005). Monoclonal antibody IL-A88 binding was detected using a second stage, goat anti-mouse immunoglobulin conjugated to the fluorochrome Alexa fluor 488 (goat-anti-mouse IgG; Alexa fluor 488, Molecular probes). Fluorescent labeling of transfected COS cells was analysed by flow cytometry.

Following transfection, the medium from each well of the 6 well plate described in section 2.10.1 was decanted and the cells were removed using trypsin versene solution, as described in section 2.9.8, Cells were transferred to 15ml tubes and washed in 6ml RPMI medium containing 10% FBS to inactivate the trypsin (Sigma-Aldrich Co. Ltd., Poole, UK). The cells were centrifuged at 350g for 5 min at 4°C, the supernatant removed and the pellet washed with FACS buffer (Appendix 1.4). The cells were then resuspenced in 200µl FACS buffer and transferred to a single well of a 96 well rounded bottomed plate and centrifuged at 600g for 30 sec. The pellet was washed twice with 200µl FACS buffer and resuspended in 50µl FACS buffer. The cells were then labeled for 40 min at 4°C with 50µl of IL-A88 antibody diluted 1:500 in FACS solution (Appendix 1.4). Cells were washed 3 times with FACS buffer, resuspended in 50µl FACS buffer and labeled for 30 min at 4°C with 50µl of a 1:500 dilution of goat anti-mouse IgG (Appendix 1.4). After a further three washes, cells were transferred into FACS analysis tubes in a final volume of 400µl FACS buffer.

2.10.3 Flow cytometry

Cells expressing ovine MHC class I molecules were identified by flow cytometry using a FACSCalibur (Becton Dickinson, Rockville, USA). The flow cytometer utilizes an argon-ion laser beam, through which a single cell suspension is directed. As cells pass through the beam, the scattered light is detected and converted electronically into a number of parameters for each cell. These include the cell size arising from the forward light scatter (FSC), the intracellular

granularity from the side or 90°C scatter and the fluorescence emitted by the fluorochrome attached to the monoclonal antibody. The information accumulated by the instrument is analyzed using the software, the BD FACSCalibur data management system, and the data are displayed graphically in the form of histograms using the cell quest software.

Chapter 3. Identification of transcribed ovine MHC class I genes

3.1 Introduction

In most mammals, the Major Histocompatibility Complex (MHC) consists of three linked gene clusters containing class I, II and III genes (Kulski et al. 2002). Class I genes are classified into classical (class Ia) and non classical (class Ib) depending on their function, expression and level of polymorphism (Braud et al. 1999). Classical MHC class I genes are extremely polymorphic and encode highly expressed cell surface glycoproteins that initiate immune responses by presenting processed antigenic peptides to $CD8^+$ T cells (Harty et al. 2000). These molecules also present self peptides to T lymphocytes during thymus development, thus determining an individual's T cell repertoire (Sebzda et al. 1999). The allelic polymorphism associated with classical MHC class I genes is located principally in exons 2 and 3. These exons code for the peptide binding region (PBR) (Brown et al. 1993) where the repertoire of bound peptides is determined (Bjorkman et al. 1987). MHC polymorphism has been shown to be associated with a range of responses to infectious disease (Hill et al. 1991; Messaoudi et al. 2002) and vaccination (Eckels D. 2000), which may be explained by subtle changes in the repertoire of peptides presented to T cells by different MHC molecules.

Ovine MHC class I allelic and haplotype diversity has been studied largely by serology (Millot P., 1978; Stear and Spooner, 1981; Cullen et al. 1985; Outteridge et al. 1988; Garrido et al. 1995; Jugo and Vicario, 2001; Jugo et al. 2002; Jugo et al. 2004). However, isoelectric focusing (IEF) studies indicated that serology underestimated the degree of diversity in sheep (Jugo et al. 2002). Antigens not previously defined by serology were identified by IEF and three serological specificities were subdivided by biochemical typing. DNA studies are expected to reveal even greater complexity. At the outset of this study, only limited information on the ovine MHC class I region was available at the DNA level (Chardon et al. 1985; Grossberger et al. 1990) with only a single full length cDNA sequence (accession number: M34676; Grossberger et al. 1990), 10 partial cDNA sequences (accn No M34672-M34675, U030092-U030094 AY188824-AY188826) and 13 exon 3 fragments (U25998 to U26010) in public databases.

The work described in this chapter aimed to identify the number of ovine MHC class I sequences transcribed in two Blackface rams and estimate their level of transcription using reverse transcription-polymerase chain reaction (RT-PCR), cloning, polymerase chain reaction-

single stranded conformational polymorphism (PCR-SSCP) and sequence analysis. The work represents the first step towards the establishment of an MHC defined resource sheep flock incorporating animals of defined MHC haplotypes for functional and disease association studies.

3.2 Materials and methods

3.2.1 Animals and preparation of nucleic acid

Peripheral blood samples were obtained from two Blackface rams, 501 and 504, as described in section 2.9.2. Immediately after preparation of peripheral blood mononuclear cell (PBMC), poly-adenylated mRNA was extracted and cDNA was synthesized as described in sections 2.3.1 and 2.4.2, respectively.

3.2.2 PCR amplification

3.2.2.1 PCR primers

Amplification of the greater part of exons 2 and 3 of ovine MHC class I genes, which encodes the peptide binding region, was performed using three different pairs of primers (A and Ar, B and Br, C and Cr, Fig. 3.1, Table 3.1). Primer A is the previously described Bov7 (Ellis et al. 1999); the remaining primers were designed on the basis of all known class I sequences for sheep and cattle, taking into account the HLA and DLA class I sequences (Fig. 3.2). Specifically, primers B and C were designed in the same area as primer A, because it is well conserved in different species (Fig. 3.2). Primers B and C were based on the same area targeted by primer A, but were used instead of A, because they provided a better match (3' end complementarity and Tm) with their corresponding reverse primers. The primer Ar is as described for Bov11 (Pichowski et al. 1996) but with a degenerate nucleotide (C/T) at the third base from the 3'-end to suit the known sheep sequences better. The primer Br is generally conserved among different species (including human, non human primates, pig, cattle, sheep and mice), with over 600 identical class I sequences found by BLAST search (Karlin and Altschul, 1990; 1993), but it does not include all of exon 3. Primer Cr lies in a slightly more polymorphic area, but amplifies the entire exon 3 (Fig. 3.1, 3.2).

Forward Primers		**Reverse Primers**	
Name	***5' Sequence 3'***	***Name***	***5' Sequence 3'***
A (Bov7)	GGCTACGTGGACGACACG	Ar	CCCTCCAGGTAGTTYCT
B	GCTACGTGGACGACACGC	Br	AGCGCAGGTCCTCGTTC
C	CGGCTACGTGGACGACAC	Cr	ATGGGTCACATGTGYCTTTG

Table 3.1: Primers used in this chapter. Nucleotide ambiguity codes recommended by IUPAC-IUB Biochemical Nomenclature Commission are used in degenerate primers.

```
        -exon 1--->
1   CCCATGACCAGAGGATTGCGAGTAATGGGGCCGCGAACCCTCCTGTTGCTGCTCTCGGGAGTCCTGGTCCTGACCGAGAT   80
              -exon 2--->
81  CCGGGCGGGCCCCACTCCATGAGGTATTTCAGCACCGCCGTGTCCCGCGCCGGCGCCGGGGAGCCCCGGTACCTGGAAG   160
      -primer A,B,C---->
161 TCGGCTACGTGGACGACACGCAGTTCGTGCGGTTCGACAGCGACGCCCCGGATCCGAAGATGGAGCAGAGGGAGCCGTGG   240
241 ATGAAGCAGGTGGGGCCGGAGTATTGGGATCGGAACACGCGAAATCCCAAGGGCAACGCACAGACTTTCCGAGTGGGCCT   320
                                        -exon 3--->
321 GACCATCCTGCGCGGCTACTACAACCAGAGCGAGACCGGGTCTCACACCTGGCAGTGTATGTACGGCTGCGACGTGGGGC   400
                                                          <------primer Br
401 CGGACGGGCGTCTCCTCCGCGGGTTCATGCAGTTCGGCTACGACGGCAGAGATTACATCGCCCTGAACGAGGACCTGCGC   480
-                                                                  <----
481 TCCTGGACCGCGGCGGACACGGCGGCTCAGGTCACCCAGCGCAAGTGGGAGAAGGAAGGTGCGGCGGACCACTACAGGAA   560
    --primer Ar-                                                          -exon 4
561 CTACGTGGAGGGCACGTGCGTGGAGTGCGTGCGCAGATACCTGGAGATCGGGAAGGAACAGCTGCAGCGAGCAGACCCTC   640
    <----------primer Cr
641 CAAAGGCACATGTGACCCATCACCCCATCTCTGGCCATGATGTCACCCTGAGGTGCTGGGCCCTGGGCTTCTACCCTGAG   720
721 GAGATCTCACTGACCTGGCAGCGCAATGGGGAGGACCAGTTGCAGGACATGGAGCTTGTGGAGACTAGGCCTTCAGGGGA   800
801 TGGAACCTTCCAGAAGTGGGCGGCCCTTGGTGGTGCTTCTGGAGAGGAGCAGAGATACACGTGCCATGTGCAGCATGAGG   880
881 GGCTTCAGGAGCCCCTCACCCTGAGATGGG                                                     910
```

Figure 3.1: Ovine class I MHC sequence (M34676, Grossberger et al. 1990) showing exons 1 to 4 and the position of primers A and Ar, B and Br and C and Cr. Primers A and Ar are in *italics*, B and Br *underlined* and C and Cr in *bold*. The first three bases of each exon are in *bold*.

Species	Ref	Class I	No	primers *A*, B, C *CGGCTACGTGGACGACACGC*
Human	A	HLA-A	202	G----------
		HLA-B	412	----------S-
		HLA-Cw	107	----------
Cattle	B	BoLA*	27	Y----------R--WY--
Sheep	C	OLA	4	----------
Dog	D	DLA-88	48	S----------R---

Species	Ref	Class I	No	primer *Ar* *AGRAACTACCTGGAGGG*	No	primer Br GAACGAGGACCTGCGCT	No	primer Cr CAAAGRCACATGTGACCCAT
Human	A	HLA-A	201	--AGY----------	201	---M----------	87	-C--RA-R---A----Y--C
		HLA-B	416	--ARC----------	416	----------M---	162	-----A----Y--------C
		HLA-Cw	107	--AGC----------	107	---M-----M-------	55	-----A----C--------C
Cattle	B	BoLA	27	----W----S-B-----	26	----S----S-------	27	---W-R--------R-B-RW
Sheep	C	OLA	16	--R-W----B-K-----	16	----------	5	----------W
Dog	D	DLA-88			47	-R----------		

Figure 3.2: Strategy for design of primers for amplification of ovine MHC class I cDNAs. Forward primer direction is from 5' to 3', whereas the complement sequence of each reverse primer is shown. Degenerate primer nucleotides and polymorphic sites on the alignments are shown using the appropriate IUPAC nomenclature code (Table 1; Appendix). *Ref* stands for references obtained from A: http://sdmc.krdl.org.sg:8080/fimm/, B: http://www.projects.roslin.ac.uk/bola/ except from the KN104 African cattle sequence, C: published ovine MHC class I sequences (see section 3.1), D: http://www.ebi.ac.uk/ipd/mhc/dla/index.html. *No* represents the number of sequences aligned and *dashes* represent the nucleotides that are identical to the primer sequence. Polymorphic sites in *red* and *blue* show that the different nucleotide from the primer sequence is represented by only one or two sequences, respectively. Primers A and Ar are in *italics*, B and Br *underlined* and C and Cr in *bold*.

3.2.2.2 PCR optimization

Optimization of the PCR reaction conditions was conducted with primers A and Ar using annealing temperatures of 50°C and 55°C, with annealing times of 30 sec and 1 min and 1, 2, 4 and 8µl of cDNA template. These PCR reactions were performed for either 30 or 35 cycles as otherwise described in section 3.2.2.3. PCR optimization with primers B-Br and C-Cr was not necessary since the optimized protocol used for primers A and Ar proved appropriate.

3.2.2.3 Optimized PCR protocol

PCR reactions were carried out in a final volume of 50µl containing 48 or 49µl PCR master mix (50mM KCl, 10mM Tris-HCl, 1.5mM $MgCl_2$, 0.1% TritonX-100, 200µM of each dNTP, 200nM of each primer and 1 unit *Taq* polymerase (Promega Corporation, Madison, USA)) and 2 or 1µl cDNA. The thermal cycling profile comprised initial denaturation for 5 min at 94°C; 35 cycles of 1 min at 94°C, 1 min at 50°C and 1 min at 68°C, and final extension for 7 min at 68°C.

Messenger RNA incubated as otherwise described in section 2.4.2, but without reverse transcriptase was amplified with all primer pairs described in section 3.2.2.1, as did controls ensuring that the PCR products were not a result of genomic DNA contamination.

3.2.3 Cloning of PCR products and transformation

PCR products of the appropriate size were excised from 0.7% agarose gels, as described in chapter 2.5.4 and cloned into the pGEM T Easy vector (Promega Corporation, Madison, USA, see Appendix 1), as described in section 2.6.2. The cloned PCR products were then used to transform Promega JM109 high efficiency competent cells, as described in section 2.7.2. The transformed cells were selected on the basis of antibiotic resistance and beta galactosidase deficiency as described in section 2.7.3.

3.2.4 Selection of transformed cells by PCR

White colonies were picked and processed for colony PCR with primers B and Br, as described in section 2.7.4.

3.2.5 SSCP analyses

The colony PCR products of the clones containing ovine MHC class I exons 2 and 3 were screened to identify different sequences, using Single Strand Conformational Polymorphism (SSCP) analysis as described by Ainsworth et al. (1991), with slight modifications, described in section 2.5.8. Although the ovine MHC class I genes were amplified from cDNA with three different primer pairs (section 3.2.2.1), SSCP analysis was performed on amplification fragments of the cloned genes produced with the same primer pair (B-Br) to ensure the comparatability of the single stranded patterns. In total, 180 clones for each animal (60 clones per primer pair) were screened by SSCP.

3.2.6 Sequence analysis

One to 10 replicates of clones with different SSCP patterns were picked for sequencing. Plasmid DNA was recovered and analyzed as described in section 2.3.3 and 2.5.6. Sequencing reactions and analysis were performed, as described in section 2.8. The sequenced clones of interest were stored in glycerol stocks (section 2.7.5).

3.3 Results

3.3.1 PCR products

3.3.1.1 PCR optimization

A range of annealing temperatures, template concentrations and cycle numbers were tested to optimize the PCR cycling conditions for the chosen primers. No PCR products were obtained from reactions with 30 cycles and an annealing temperature of 55°C; decreasing the annealing temperature to 50°C resulted in only limited amplification. However, the yield of the PCR products at 50°C annealing temperature was improved by increasing the number of cycles to 35 (Fig. 3.3). Increasing the annealing time from 30 sec to 1 min further increased the yield (Fig. 3.4). In contrast, increasing the amount of cDNA used in PCR reactions above 4µl decreased the yield. In addition, there was a possible reduction of specificity as indicated by smearing (Fig. 3.4).

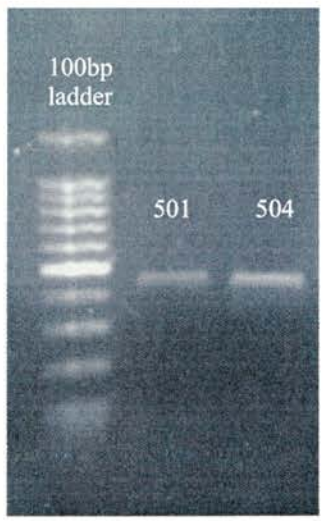


Figure 3.3: Agarose gel showing the products obtained following 35 cycles of PCR using primers A and Ar, 1µl of cDNA template from rams 501 and 504 and an annealing temperature of 50°C. The 100bp DNA ladder contains fragments from 100bp to 1000bp and an additional band at 1500bp.

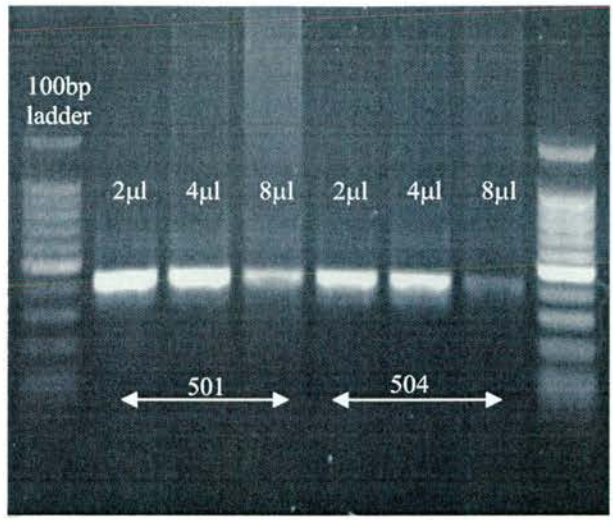


Figure 3.4: Agarose gel showing PCR products amplified from 2, 4 and 8µl of cDNA template derived from rams 501 and 504.

3.3.1.2 PCR products from optimized protocol

Using the optimized PCR protocol described in section 3.2.2.3, a specific band of the appropriate size (Table 3.2) was obtained from cDNA samples prepared from both rams using all three primer pairs (Fig. 3.5 and 3.6). The reverse transcription (RT) control shows that the amplified products were not a result of genomic DNA contamination. Prior to cloning, PCR products were gel purified to remove any high or low molecular weight contaminants. Representative gel purified PCR products are presented in figures 3.5 and 3.6.

Primer pair	PCR product size
A-Ar	410bp
B-Br	318bp
C-Cr	499bp

Table 3.2: PCR product sizes amplified by each primer pair.

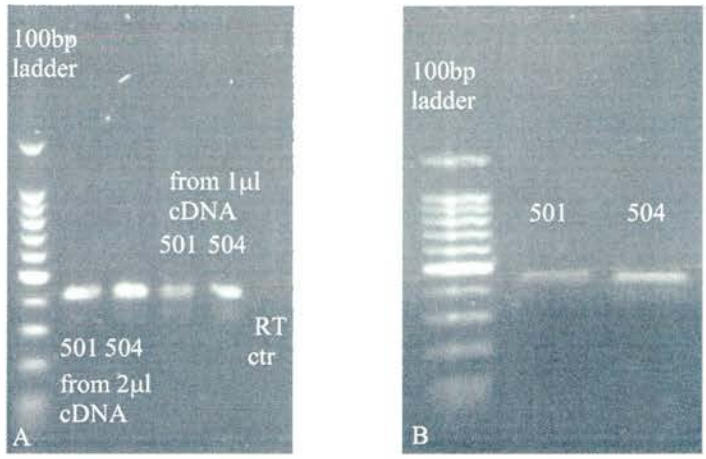


Figure 3.5: Agarose gel showing PCR products amplified using primers A and Ar from either 1 or 2µl of cDNA derived from rams 501 and 504, before (A) and after gel purification (B). The RT ctr controls for genomic DNA amplification.

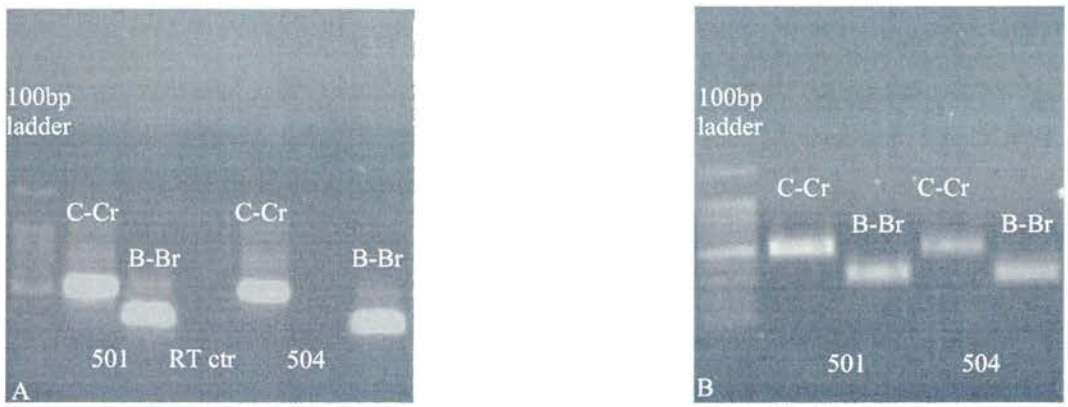


Figure 3.6: Agarose gel showing PCR products amplified from cDNA prepared from rams 501 and 504 using primers B-Br and C-Cr before (A) and after gel purification (B). The RT ctr controls for genomic DNA amplification.

3.3.2 Cloning and selection of recombinants by PCR

Purified PCR products from cDNA derived from both rams were amplified using all three sets of primers and cloned into the pGEM T-easy vector.

Approximately 200 white colonies were screened by colony PCR using primers B and Br (Fig. 3.7). For each ram cDNA, 60 positive clones obtained from each primer pair were chosen for SSCP analysis to identify candidates for sequencing.

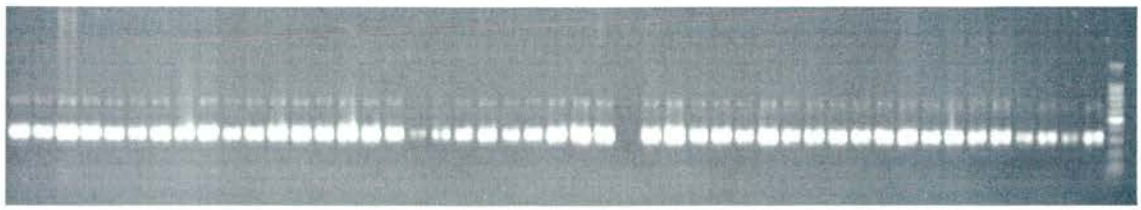


Figure 3.7: Agarose electrophoresis of colony PCR products of clones containing MHC class I exons 2 and 3 products.

3.3.3 SSCP analysis

SSCP analysis of A-Ar amplicons using the B-Br primers revealed two distinct patterns for each ram, whereas B-Br and C-Cr amplicons yielded four additional patterns for ram 501 and six additional patterns for ram 504 (Fig. 3.8). The number of clones arising from each primer pair that corresponds to individual SSCP patterns for the two rams is shown in table 3.3. These were derived by comparison of SSCP patterns on the same polyacrylamide gel and subsequent sequencing of at least one sample from each pattern.

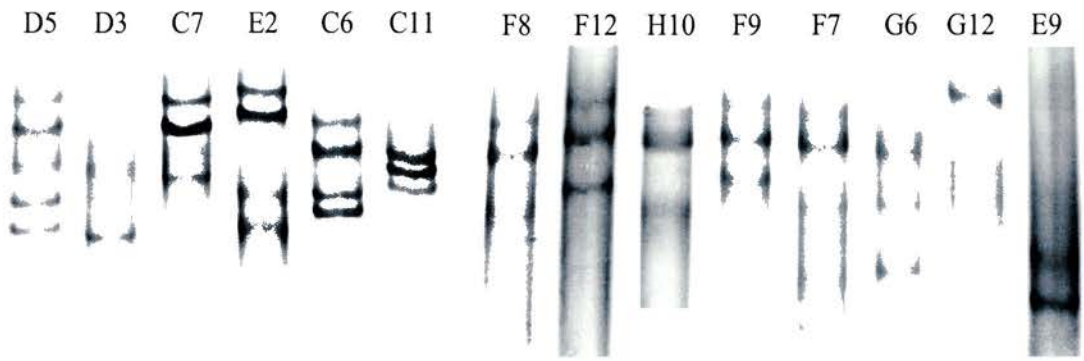


Figure 3.8: SSCP patterns obtained for ram 501 (first six lanes) and ram 504 (eight last lanes).

501 SSCP patterns	D5	D3	C7	E2	C11	C6	Hybrids
A-Ar	30	27	0	0	0	0	3
B-Br	14	18	7	11	4	1	5
C-Cr	18	17	12	7	1	2	3
total	62	62	19	18	5	3	11

504 SSCP patterns	F8	F12	H10	F9	F7	G6	G12	E9	Hybrids
A-Ar	32	25	0	0	0	0	0	0	3
B-Br	20	33	0	1	1	0	0	3	2
C-Cr	27	12	8	9	0	2	1	0	1
total	79	70	8	10	1	2	1	3	6

Table 3.3: Number of clones identified for each SSCP pattern per primer pair used for cloning. Hybrids were clones having a different SSCP pattern, but sequencing showed that they consisted of a combination of two other alleles. The names of the patterns correspond to the names of the clones sequenced (Fig. 3.9 and 3.10).

3.3.4 Sequencing of different clones identified by SSCP

Sequencing confirmed that the cloned products were ovine MHC class I genes. Up to ten replicates corresponding to each SSCP pattern were sequenced and identical sequences were associated with each pattern. Six distinct class I sequences were obtained for ram 501 and 8 for ram 504 for part of exons 2 and 3 (Fig. 3.9 and 3.10). An alignment of their predicted amino acid sequences is shown in figure 3.11. All sequences identified were new, except for part of the D5 sequence, which is identical to the database entry with accession number U26002. However, the length of this entry is only around 200bp and therefore it is possible that it is not identical to D5 throughout its length.

Some primer pairs did not detect all sequences identified from both rams. In particular, the primer pair A-Ar amplified only sequences D5 and D3 from ram 501 and F8 and F12 from ram 504, whereas each primer pair B-Br and C-Cr also amplified four additional sequences from 501 and three or four from 504, respectively (Table 3.3). Furthermore, although all six 501 class I sequences were identified by both primer pairs B-Br and C-Cr, some 504 sequences were not picked by primers B-Br, but were identified by C-Cr and vice versa (Table 3.3).

Additionally, marked variation was observed in the efficiency with which individual sequences were amplified by each primer pair. Each of the sequences D5, D3, F8 and F12 were represented in more than 60 clones and sequences C7, E2, H10 and F9 were found in around 10-20 clones each. The remaining sequences were represented in a low number of clones (1-5).

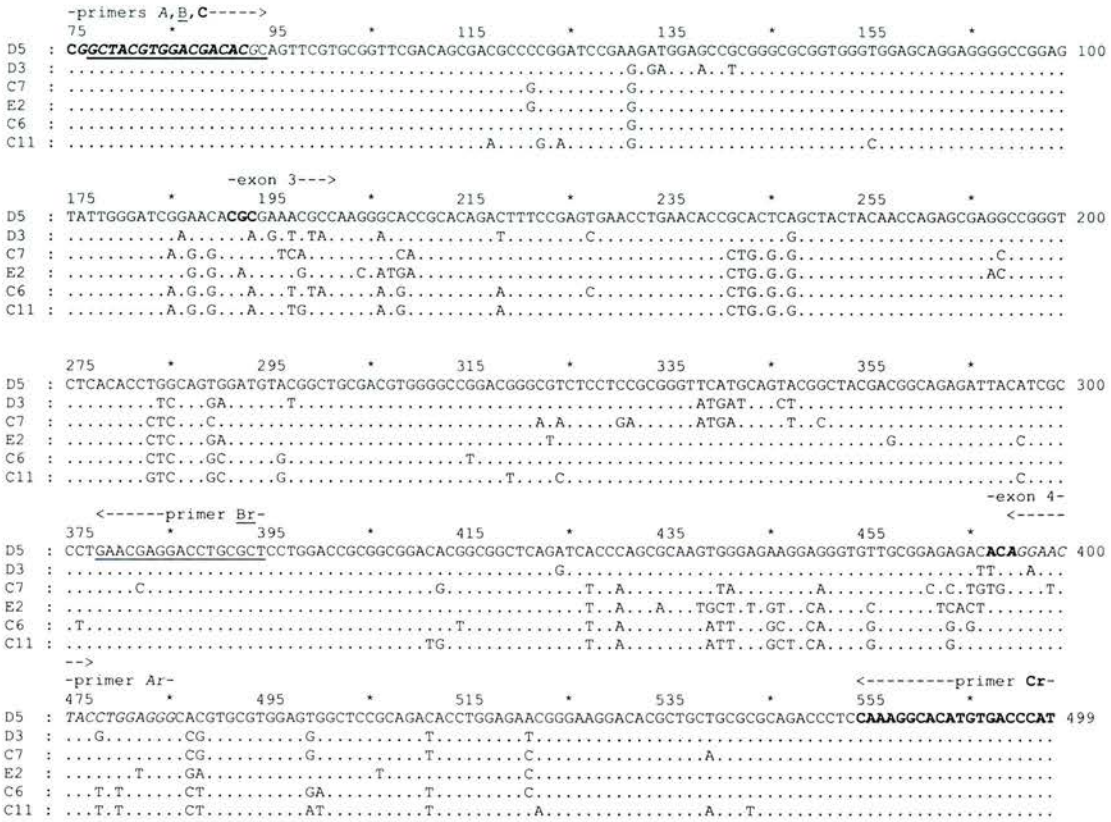


Figure 3.9: Alignment of ovine MHC class I sequences obtained with primer pairs A-Ar, B-Br and C-Cr for ram 501. Primers A and Ar are in *italics*, B and Br *underlined* and C and Cr in *bold*. The first three nucleotides of each exon are in *bold*. Identities are indicated by *dots* (.). The numbering of nucleotides is based on the full length class I sequence M34676 (Grossberger et al. 1990).

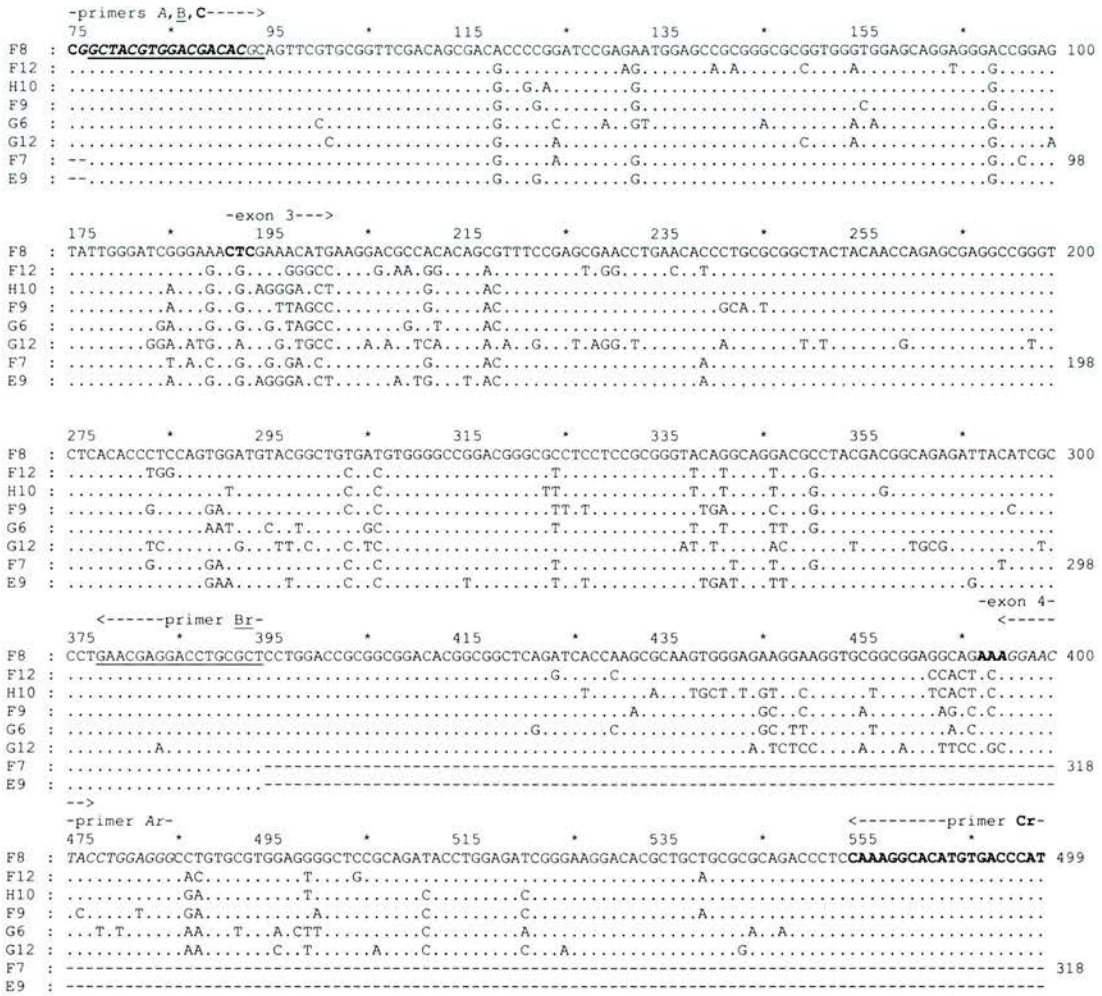


Figure 3.10: Alignment of ovine MHC class I sequences obtained with primer pairs A-Ar, B-Br and C-Cr for ram 504. Primers A and Ar are in *italics*, B and Br *underlined* and C and Cr in *bold*. The first three nucleotides of each exon are in *bold*. Identities and gaps are indicated by *dots* (.) and *dashes* (-), respectively. The numbering of nucleotides is based on the full length class I sequence M34676 (Grossberger et al. 1990).

```
                 Alpha 1                                                          |Alpha 2
                 *   30        40        50    *** *****************  *** *     9|0   * * *100      110 * *   120       100
501_D5   : GYVDDTQFVRFDSDAPDPKMEPRARWVEQEGPEYWDRNTRNAKGTAQTFRVNLNTALSYYNQSEA|GSHTWQWMYGCDVGPDGRLLRGFMQYGYDGRDYIA : 100
501_D3   : ...................E....................IY.D...I..A......G.......|......E................D.L......... : 100
501_C7   : ...............A....................QE..S...H..........LRG......P|....L.R.............S..E..A........ : 100
501_E2   : ...............A.....................E..K.NDD..........LRG......P|....L.E..........L..........E...... : 100
501_C6   : ....................................QE..IY.DA..N..A....LRG.......|....L.A.C.......................... : 100
501_C11  : .............E.R..........A.........QE..M..DA..N.......LRG.......|....V.A.C.......W.................. : 100
504_F8   : ..............T......................E...M.DAT.R..A....LRG.......|....L..................R.DA........ : 100
504_F12  : .....................Q..P....V.......E..R...NG.....G.TILRG.......|................................... : 100
504_H10  : ...............A....................QE.QGT.DA.....A....LRG.......|....L.C..........L..........E...... : 100
504_F9   : ...............R..........A.........QE..L..DA.....A......G.......|....V.E...........F....E.H......... : 100
504_G6   : ................H......T............EE..I..DGI....A....LRG.......|....L.N.HC.G....................... : 100
504_G12  : ........A...............P..........EEM.....KAQ.RL.SG...MRG.......|V...S...A.V...........IW.TA...A...S : 100
504_F7   : ....................................L...RT.DA.....A...NLRG.......|....V.E................S........... : 100
504_E9   : ...............R....................QE.QGT.D..L...A...NLRG.......|....L.E................D..A....E... : 100

                      130       140* *** **** ****** *** *** * *     170       180        200
501_D5   : LNEDLRSWTAADTAAQITQRKWEKEGVAERHRNYLEGTCVEWLRRHLENGKDTLLRADPPKTHVTH : 166
501_D3   : ..............................F......R...G...Y..I............A.... : 166
501_C7   : ............A.....K..L......A.V.I....R...G...Y..T.....Q........... : 166
501_E2   : ..................KHNVVVA.A.DHY....V.E.....L....T................. : 166
501_C6   : ..................K..I.AA.G..G....F..L...E...Y..T................. : 166
501_C11  : ............A.....K..I.AA.G..G....F..L...M...Y..K.....Q........... : 166
504_F8   : ..................K.......A..AE......L...G...Y..I................. : 166
504_F12  : ..........................A.DHY..............Y..I.....Q........... : 166
504_H10  : ..................KHNAVVA.A.DHY......E..........T................. : 166
504_F9   : ..................K...AA.E......S.V.E...E......T.....Q........... : 166
504_G6   : ...............R.......AL.A..EQ...F..K...NL.............T......... : 166
504_G12  : ..................K...IS.E..FQ......K.....H....T................. : 166
504_F7   : ......------------------------------------------------------------ : 106
504_E9   : ......------------------------------------------------------------ : 106
```

Figure 3.11: Alignment of predicted aminoacid ovine MHC class I sequences for rams 501 and 504. Identities and gaps are indicated by *dots* (.) and *dashes* (-), respectively. Asterisks (*) indicate MHC class I residues that bind antigenic peptides or interact with T cell receptors (Bjorkman et al. 1987). The numbering of aminoacids is based on the full length class I sequence M34676 (Grossberger et al. 1990).

	D5	D3	C7	E2	C6	C11	F8	F12	H10	F9	F7	G6	G12	E9
D5		[0.013]	[0.016]	[0.015]	[0.015]	[0.016]	[0.015]	[0.012]	[0.015]	[0.014]	[0.018]	[0.017]	[0.022]	[0.022]
D3	0.078		[0.016]	[0.019]	[0.016]	[0.018]	[0.016]	[0.017]	[0.018]	[0.015]	[0.019]	[0.020]	[0.025]	[0.020]
C7	0.108	0.120		[0.016]	[0.016]	[0.016]	[0.016]	[0.016]	[0.016]	[0.016]	[0.020]	[0.019]	[0.024]	[0.018]
E2	0.106	0.148	0.112		[0.015]	[0.016]	[0.016]	[0.016]	[0.010]	[0.016]	[0.018]	[0.018]	[0.023]	[0.021]
C6	0.101	0.110	0.108	0.096		[0.010]	[0.015]	[0.017]	[0.014]	[0.014]	[0.016]	[0.016]	[0.023]	[0.018]
C11	0.115	0.144	0.120	0.110	0.052		[0.017]	[0.018]	[0.016]	[0.015]	[0.017]	[0.017]	[0.023]	[0.021]
F8	0.094	0.113	0.108	0.119	0.096	0.122		[0.016]	[0.016]	[0.016]	[0.018]	[0.018]	[0.022]	[0.022]
F12	0.072	0.129	0.119	0.110	0.124	0.134	0.108		[0.016]	[0.018]	[0.020]	[0.018]	[0.022]	[0.023]
H10	0.103	0.141	0.110	[illegible]	0.087	0.115	0.108	0.110		[0.016]	[0.015]	[0.017]	[0.023]	[0.016]
F9	0.092	0.101	0.110	0.108	0.092	0.096	0.115	0.134	0.108		[0.017]	[0.018]	[0.023]	[0.017]
F7	0.091	0.098	0.108	0.090	0.076	0.083	0.091	0.116	0.069	0.087		[0.019]	[0.030]	[0.017]
G6	0.127	0.159	0.159	0.139	0.115	0.129	0.134	0.144	0.122	0.132	0.105		[0.024]	[0.022]
G12	0.189	0.235	0.223	0.210	0.207	0.218	0.187	0.194	0.202	0.202	0.227	0.216		[0.032]
E9	0.135	0.109	0.098	0.124	0.098	0.120	0.128	0.146	0.076	0.087	0.080	0.127	0.246	

Table 3.4: Nucleotide sequence distances and variances calculated by the Kimura 2 parameter model, using MEGA 2.1. Sequence distances are presented below the diagonal, whereas their variances are shown above the diagonal. The lowest distance is highlighted in *green* and the largest distance in *yellow*.

	D5	D3	C7	E2	C6	C11	F8	F12	H10	F9	F7	G6	G12	E9
D5		[0.029]	[0.037]	[0.039]	[0.035]	[0.039]	[0.033]	[0.031]	[0.037]	[0.035]	[0.041]	[0.039]	[0.047]	[0.049]
D3	0.128		[0.038]	[0.044]	[0.034]	[0.042]	[0.030]	[0.038]	[0.041]	[0.036]	[0.037]	[0.041]	[0.050]	[0.041]
C7	0.207	0.214		[0.040]	[0.037]	[0.039]	[0.035]	[0.042]	[0.039]	[0.039]	[0.044]	[0.047]	[0.052]	[0.040]
E2	0.222	0.276	0.237		[0.037]	[0.041]	[0.040]	[0.042]	[0.023]	[0.039]	[0.043]	[0.046]	[0.054]	[0.046]
C6	0.185	0.178	0.207	0.207		[0.024]	[0.031]	[0.040]	[0.031]	[0.033]	[0.035]	[0.037]	[0.048]	[0.040]
C11	0.229	0.260	0.229	0.244	0.088		[0.039]	[0.042]	[0.039]	[0.033]	[0.041]	[0.043]	[0.052]	[0.049]
F8	0.163	0.142	0.185	0.237	0.149	0.222		[0.036]	[0.036]	[0.036]	[0.038]	[0.039]	[0.044]	[0.040]
F12	0.149	0.214	0.252	0.252	0.237	0.260	0.192		[0.040]	[0.044]	[0.049]	[0.044]	[0.049]	[0.056]
H10	0.207	0.244	0.222	[illegible]	0.149	0.222	0.192	0.237		[0.037]	[0.037]	[0.042]	[0.051]	[0.035]
F9	0.185	0.192	0.222	0.222	0.163	0.170	0.199	0.276	0.207		[0.038]	[0.043]	[0.049]	[0.041]
F7	0.165	0.132	0.188	0.177	0.121	0.165	0.143	0.223	0.132	0.143		[0.047]	[0.063]	[0.037]
G6	0.229	0.244	0.308	0.300	0.207	0.268	0.229	0.284	0.260	0.268	0.211		[0.048]	[0.050]
G12	0.308	0.350	0.376	0.393	0.324	0.367	0.284	0.333	0.358	0.333	0.350	0.324		[0.070]
E9	0.223	0.165	0.154	0.200	0.154	0.223	0.154	0.285	0.121	0.165	0.132	0.235	0.420	

Table 3.5: Amino acid distances and variances calculated by the Poisson correction, using MEGA 2.1. Amino acid distances are presented below the diagonal, whereas their variances are shown above the diagonal. The lowest distance is highlighted in *green* and the largest distance in *yellow*.

3.3.5 Sequence analysis

One hundred and seventy eight polymorphic sites were identified within the 499 nucleotides (35.7%) of the largest amplified fragment (Fig. 3.9 and 3.10). This translates into variation at 76 of 166 amino acid sites (45.8%) (Fig. 3.11). When genetic distances were estimated, sequences were between 4.6% to 23.5% different at the nucleotide level (Table 3.4) and 8.2% to 39.3% different at the aminoacid level (Table 3.5), with sequences G12 and G6 showing the greatest divergence from the other sequences (Tables 3.4 and 3.5).

3.4 Discussion

In this chapter, the polymorphic and polygenic nature of the ovine MHC class I region was studied by SSCP and sequence analysis of the peptide binding region (PBR). Six distinct patterns were identified from ram 501 and eight were obtained from ram 504, each corresponding to a single sequence. Comparison of the 14 ovine MHC class I sequences identified in the work described in this chapter with those previously published in public databases showed that all were novel.

3.4.1 Genotyping method for identification of ovine MHC class I transcripts

SSCP and sequence analysis of exons 2 and 3 was chosen as an efficient and cost effective means of identifying class I sequences. SSCP was used as a sensitive low cost genotyping method of screening large numbers of clones prior to sequencing. This method allows the detection of single base polymorphisms in short DNA fragments due to the mobility of single stranded DNA molecules, without the need for prior sequence knowledge (Orita et al. 1989). For that reason, SSCP was a useful tool for the identification of new ovine MHC class II genes (van Oorshot et al. 1994; Kostia et al. 1998; Snibson et al. 1998; Jugo and Vicario, 2000). As there is limited sequence information available for ovine MHC class I genes, SSCP was also regarded as the most appropriate genotyping system for the purposes of this study.

SSCP analysis can distinguish single stranded DNA fragments under non-denaturing conditions because they fold into unique conformations determined by their primary sequence. Formation of such secondary structures depends on both the length of the sequence and physical factors such as temperature and ionic strength (Hayashi and Yandell, 1993). For these reasons, all cloned PCR products in this study were amplified with the same primer pair (B-Br) to produce fragments of the same length for SSCP analysis. In addition, the temperature and chemical composition of the buffer were kept constant for all polyacrylamide gel electrophoresis (PAGE). Therefore, the SSCP conditions described here ensured that each sequence produced a unique SSCP migration pattern that was reproducible under exactly the same conditions. However, the method was sensitive to slight variations and as a result, identical sequences sometimes appeared slightly different in separate gels. For this reason, the analysis was restricted to the identification of different patterns within individual gels. Moreover, to ensure that SSCP was detecting all the polymorphism, replicates with the same SSCP pattern were sequenced. In each case all replicates were identical.

Sequencing of exons 2 and 3 of ovine MHC class I transcripts was used to identify the sequences transcribed by the rams, because the majority of variability in MHC class I genes is found within these exons (Trowsdale and Parham, 2004). Sequence analysis of exons 2 and 3 can discriminate between the great majority of HLA class I sequences (Parham et al. 1995; http://www.ebi.ac.uk/imgt/hla/align.html). Moreover, bovine MHC class I sequences available in the databases can be discriminated by sequencing of the fragment amplified by primers C and Cr, used in this study. In sheep, most of the diversity previously described in MHC class I molecules is situated between residues 60 and 80 of the α1 domain and between residues 150 and 170 of the α2 domain (Grossberger et al. 1990) and both of these regions were included in the amplicons generated.

Screening a cDNA library is an alternative approach for the identification of transcribed class I sequences. In a recent study of the horse MHC, class I genes were isolated both from a cDNA library and by reverse transcription PCR (Chung et al. 2003). The second approach isolated only three out of the four sequences identified from the library. However, failure to detect the remaining allele is likely to have arisen from the small number (25-35) of RT-PCR clones that were screened. The much larger number of clones screened in the present study (180 per animal) greatly increases the likelihood of detecting MHC class I genes transcribed at lower levels. Furthermore, failure to detect all transcribed sequences from an animal could also result from cDNA library screening with a probe even under low stringency conditions. Using SSCP and sequence analysis of the PBR, a number of undetected sequences were found to be transcribed by cattle MHC haplotypes previously characterized by cDNA library screening (McHugh N., CTVM, University of Edinburgh, Edinburgh, UK, personal communication).

3.4.2 Number and transcription levels of ovine MHC class I loci

A number of serological studies, to date, investigating segregation of alloantisera markers in sheep families, have identified up to three class I loci (Stear and Spooner, 1981; Cullen et al. 1982; Millot P. 1984; Grossberger et al. 1990; Garrido et al. 1995). Millot P. (1978) suggested that there are two closely linked sheep MHC loci (OLA-A and OLA-B), but in a later study, identified a third linked serologically distinct locus (Millot P. 1984). Other researchers suggested that the distribution of antigens in the population and their segregation is consistent with the assumption that two linked loci are present (Stear and Spooner, 1981; Cullen et al. 1982). However, Cullen et al. (1982) speculated that their serological typing data could be explained by the expression of four loci. Molecular studies of the ovine MHC class I region are restricted to one publication, where a sheep thymus cDNA library was screened with a human

HLA-B7 cDNA probe. The authors identified 5 different sequences, which is consistent with the transcription of at least three loci (Grossberger et al. 1990).

The number of transcripts identified from rams 501 and 504 implies haplotypes with at least three and four MHC class I loci, respectively. To maximize the likelihood of identifying all the transcribed class I sequences, three different primer pairs designed in the most conserved areas possible were used and a large number of clones (180 per animal) were screened by SSCP. However, the possibility of missing transcribed sequences due to polymorphism at the 3' prime ends of the primers cannot be totally excluded. It is also possible that additional class I genes were not identified because they are transcribed at very low levels in PBMCs. In that case, the number of ovine MHC class I loci transcribed by those rams may be even higher.

If the primers used to amplify the class I transcripts were conserved, the number of replicates identified for each sequence should reflect to some extent the abundance of the corresponding mRNA. With the exception of primer Ar, it was subsequently determined that all primers were designed in well conserved areas (chapter 5; Fig. 5.12). Therefore, a preferential amplification of specific sequences with primer pairs B-Br and C-Cr is unlikely. On that basis, sequences D5, D3, C7 and E2 for ram 501 and F8 and F12 for ram 504 were transcribed at much higher levels compared to the remaining sequences, because they were represented by a larger number of clones (Table 3.3). These sequences possibly encode highly expressed classical class I products, whereas the remaining sequences may encode poorly expressed classical or non-classical class I genes. Similarly, in humans alleles of the HLA-C locus are expressed at significantly lower levels compared to HLA-A and HLA-B (Snary et al. 1977; McCutcheon et al. 1995). Moreover, a number of classical bovine class I loci expressed at levels significantly lower than usual have been described (reviewed by Ellis S. 2004).

It is unlikely that all identified class I sequences, especially for ram 504, are functionally expressed classical class I genes. It has been suggested that a small number of highly polymorphic classical class I loci is preferred in vertebrates, because in the presence of a many MHC molecules during the development of the T cell repertoire, depletion of a large number of potentially self reactive T cell receptors is possible. Therefore, a reduced T cell repertoire would be generated that is unfavourable for disease resistance (Kappler et al. 1987). Indeed, the extensive variability in genetic distances between the obtained exon 2 and 3 fragments (Tables 3.5 and 3.6) suggests that although some sequences represent alleles at individual loci, others represent diverged putative non classical loci that have incorporated a large number of possibly non functional mutations (sequences G12 and G6). Moreover, the

identification of partial MHC class I transcripts does not mean that each is functional. A number of transcribed pseudogenes with deletions or early stop codons have been identified within the bovine MHC class I region (Ellis S., IAH, Compton, UK, personal communication). Therefore, additional full length sequence information combined with transfection and expression studies are required to resolve this issue.

A phylogenetic analysis of these sequences may provide some information on the relationships between transcripts and the number of loci involved. However, the topology of trees constructed using the highly polymorphic exon 2 and 3 sequences were unreliable possibly due to the influence of positive selection in the regions coding for the PBR (Hughes and Nei, 1988). In contrast, substitutions that occur in more conserved regions of MHC class I genes are more possibly locus specific. Therefore, phylogenetic analysis using full length or exons 4 through 8 sequence data are expected to yield information about the relationships between the ovine MHC class I loci.

In summary, the results outlined in this chapter suggest that a combination of RT-PCR, SSCP and nucleotide sequence analysis of the ovine MHC class I peptide binding region was a useful tool for identifying novel transcripts. Six distinct SSCP patterns, each corresponding to a unique nucleotide sequence were identified from ram 501 and eight from ram 504, implying the transcription of at least three class I loci by ram 501 and four by ram 504. The abundance of the transcripts and the genetic distances among the sequences suggest that at least one locus for ram 504 and two for ram 501 are highly expressed classical class I genes, while the remaining sequences may represent weakly expressed classical or non classical loci or pseudogenes.

Chapter 4. Assigning the ovine MHC class I transcripts to haplotypes

4.1 Introduction

In the previous chapter, six ovine MHC class I sequences transcribed by ram 501 and 8 transcribed by ram 504 were identified. However, the combination of genes inherited together on a single chromosome comprises a haplotype and characterization of MHC haplotypes is desirable for precise analysis of immune responses and disease association studies.

Traditionally, MHC class I haplotypes have been characterized using serology. However, during the last decade, due to the development of the polymerase chain reaction (PCR) technique, DNA based typing methods have begun to replace serological techniques. DNA methods were initially applied to type the human MHC, known as Human Leukocyte Antigens (HLA). Such methods were first developed for class II typing and more recently they have been used to determine class I HLA alleles (reviewed in Middleton D. 1999). DNA typing revealed that the extent of polymorphism was greater than that detected by serology and proved invaluable in cases where serological typing was difficult due to poor cell viability or expression.

Among the available DNA typing techniques, sequence specific PCR has been widely used in clinical histocompatibility laboratories (Middleton D. 1999). A comprehensive genotyping system which characterizes HLA-A, -B, -C, -DRB1, -DRB3, -DRB4, -DRB5 and DQB1 haplotypes, using multiplex sequence specific amplification has been developed (Bunce et al. 1995). Following the advances developed to type human samples for clinical transplantations, sequence specific amplification methods were developed to type MHC genes in other non human mammalian species. MHC class I genes from Rhesus macaques (*Macaca mulatta*) (Lobashevsky and Thomas, 2000), woodchuck (*Marmota monax*) (Zhou et al. 2003) and cattle (Ellis et al. 1998) have been typed by this method.

However, ovine class I haplotypes have been characterized only by serology. Due to the small number of studies at the DNA level (Chardon et al. 1985; Grossberger et al. 1990) and the limited sequence data available, PCR based typing methods for ovine class I genes have yet to be developed. Therefore, disease association (reviewed by Dietert R. 1996; Hohenhaus and Outteridge, 1995; Stear and Spooner, 1981) and immunization studies on sheep (Outteridge et

al. 1988), where class I antigens are used as markers, are mainly based on serological typing to date.

The work described in this chapter aims to determine which class I genes are inherited together on individual haplotypes carried by rams 501 and 504, by genotyping cohorts of their progeny. This required the development of a sequence specific genotyping system to trace the inheritance of each class I sequence in the progeny of rams 501 and 504. Haplotype characterization of ovine MHC class I genes, availability of MHC defined animals and development of reliable and quick genotyping methods will facilitate immunization and disease association studies.

4.2 Materials and methods

4.2.1 Animals and preparation of nucleic acid

Peripheral blood samples were obtained from rams 501 and 504 and from 9 and 8 of their respective progeny (Table 2.1). Blood was also obtained from 5 of the dams of these progeny as detailed in Table 4.1. Peripheral blood mononuclear cells (PBMC) were prepared from each blood sample as described in section 2.9.2 and total RNA extracted and cDNA synthesized as described in sections 2.3.2 and 2.4.2, respectively. The numbers of PBMC, total RNA concentrations and OD_{260}/OD_{280} ratios for the extracted RNA are shown in Table 4.1.

Animals	No	PBMCs	RNA (μg/ml)	OD_{260}/OD_{280}
rams	501	$1.03x10^7$	12.15	1.79
	504	$0.93x10^7$	10.67	1.89
ewes	265N	$1.27 x10^7$	7.89	1.90
	306N	$1.12 x10^7$	9.86	1.93
	877N	$1.17 x10^7$	8.64	1.92
	822N	$1.35 x10^7$	9.45	1.88
	1098N	$1.86 x10^7$	10.28	1.94
lambs	448	$1.00 x10^7$	5.16	1.95
	449	$1.13 x10^7$	6.30	1.98
	461	$0.96 x10^7$	5.63	1.90
	486	$0.68 x10^7$	7.27	1.89
	489	$0.72 x10^7$	11.38	1.92
	518	$1.27 x10^7$	11.57	1.90
	555	$1.02 x10^7$	12.81	1.90
	637	$1.04 x10^7$	10.43	1.86
	680	$0.50 x10^7$	12.28	1.96
	457	$0.46 x10^7$	8.34	1.93
	529	$1.15 x10^7$	11.9	1.94
	530	$0.88 x10^7$	9.78	1.95
	573	$0.72 x10^7$	9.46	1.96
	596	$0.68 x10^7$	10.98	1.90
	609	$1.20 x10^7$	15.33	1.95
	610	$0.83 x10^7$	6.24	1.93
	673	$0.71 x10^7$	6.33	1.95

Table 4.1: PBMC measurements, RNA concentrations and OD_{260}/OD_{280} ratios for the extracted nucleic acid from animals used in this chapter.

4.2.2 Development of the class I sequence specific PCR amplification method

4.2.2.1 Primer design

For the purpose of establishing a sequence specific amplification method to genotype the progeny of rams 501 and 504, sequence specific primers (SSPs; Fig. 4.1, Table 4.2) were designed using the sequences described in chapter 3 (Fig. 3.9 and 3.10). SSPs were designed on the basis that at least one specificity-dependent nucleotide is located on the terminal 3'-base of either forward, reverse or ideally both primers (Fig. 4.1).

SSPs were designed to generate PCR products of different sizes for each sequence identified in each ram, with the potential of developing a multiplex PCR. In addition, SSPs were designed to have high and compatible melting temperatures (59°C to 63°C for 501 sequences and 58°C to 63°C for 504 sequences, as determined using Primer3 software; section 2.4.1), allowing stringent PCR conditions to be used. The vast majority of SSPs exhibited low end self complementarity and low 3' end stability.

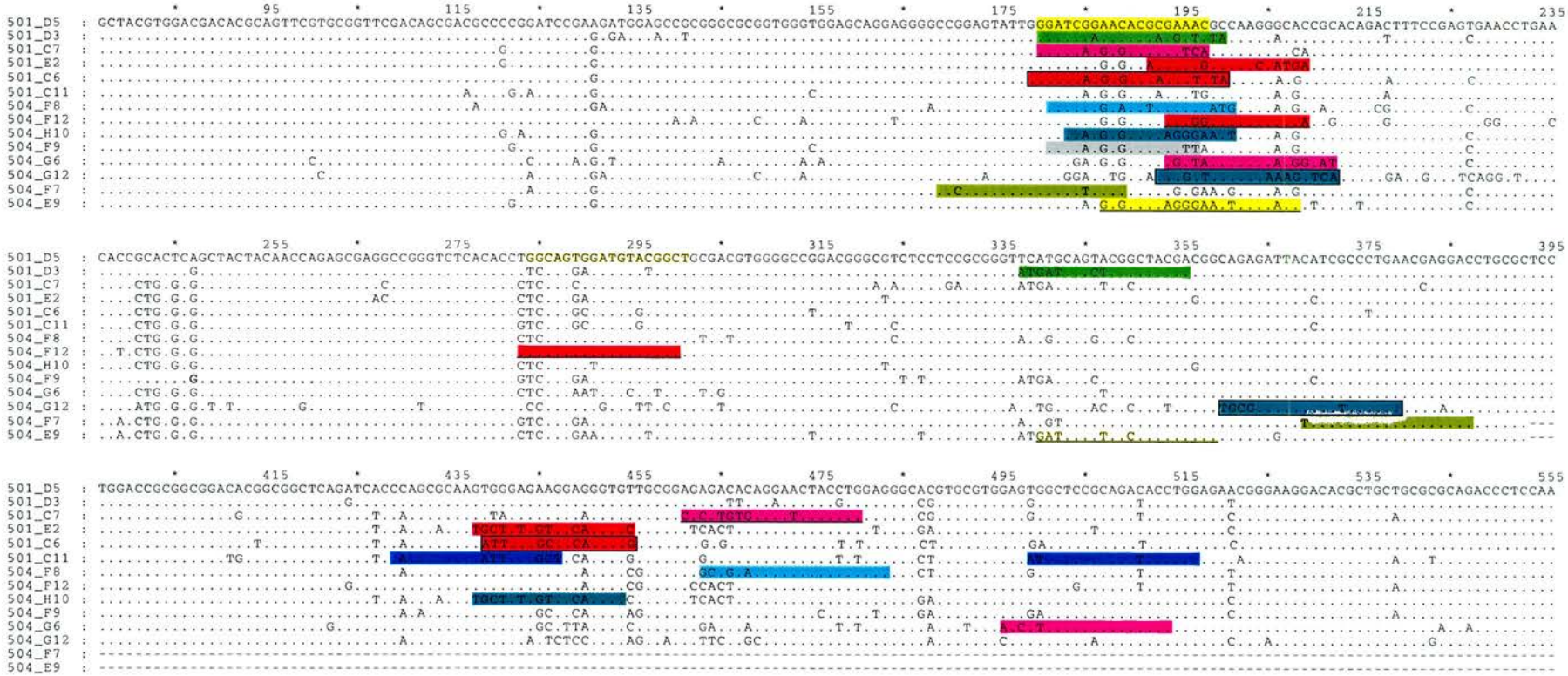


Figure 4.1: Alignment of ovine MHC class I sequences obtained for rams 501 and 504 showing the position of all SSPs . Both the forward and the reverse primer that specifically amplify each sequence are shown with the same colour. The numbering of nucleotides is based on the full length class I sequence M34676 (Grossberger et al. 1990).

Forward Primers		Reverse Primers		
Name	*5' Sequence 3'*	*Name*	*5' Sequence 3'*	*PCR product length*
D5frw	GATCGGAACACGCGAAAC	D5rev	AGCCGTACATCCACTGCC	121bp
D3frw	GGATCGAAACACGAGGATCTA	D3rev	TCGTAGCCGAGCTGATCAT	177bp
C7frw	GGATCAGGAGACGCGATCA	C7rev	CAGGTAGATCCTCACACGCG	301bp
E2frw	ACGAAAGGCCAACGATGA	E2rev	GCACCTGCCACCACAACA	264bp
C6frw	GGGATCAGGAGACGAGAATCTA	C6rev	CACCTGCCGCCTCAAT	276bp
C11frw	CAAGCGCAAGATTGAGGCT	C11rev	CCAGGTATCTGCGGAGCAT	89bp
F8frw	GATCGGGAAACTCGAAACAT	F8rev	CTCCAGGTAGTTCCTTTCTGC	303bp
F12frw	GAAGGGCCAAGGGCAA	F12rev	GCCGTACATCCACTGCCA	107bp
H10frw	TCAGGAGACGCAGGGAACT	H10rev	CACCTGCCACCACAGCA	272bp
F9frw	GGATCAGGAGACGCGATT	F9rev	TGGTTGTAGTAGCCGAGTGC	81bp
G6frw	GGATAGCCAAGGACGGCAT	G6rev	GGTGTCTGCGGAGCAAGTT	321bp
G12frw	CGAGATGCCAAGAAAGCTCA	G12rev	TCAGGGAGATGTAATCCGCA	187bp
F7frw	CCCGAGTATTGGGATCTGAAC	F7rev	GTCCTCGTTCAGGGCGATA	219bp
E9frw	GAGACGCAGGGAACTAAGGACA	E9rev	CCGTCGTAGGCGAACTGATC	173bp

Table 4.2: Sequence specific primers and length of the expected PCR products.

4.2.2.2 PCR optimization using plasmid template DNA

The sequence-specific amplification method was initially optimized using plasmid DNA, containing each of the ovine MHC class I sequences identified from rams 501 and 504. Plasmid DNA was prepared as described in section 2.3.3 and resuspended at a concentration of 100ng/µl. To determine the optimum plasmid template concentration, the plasmid containing the D5 gene was diluted 100, 500 and 1000 times and 1µl from each dilution was mixed with 49µl PCR master mix (50mM KCl, 10mM Tris-HCl, 1.5mM $MgCl_2$, 0,1% TritonX-100, 200µM of each dNTP, 200nM of each primer (D5frw and D5rev) and 1 unit Taq polymerase (Promega Corporation, Madison, USA). Amplification was carried out using the touchdown PCR cycling profile 1 described below.

Thermal cycling profile 1:

- Initial denaturation for 1 min at 94°C;
- 5 cycles of 1 min at 94°C, 1 min at 65°C and 2 min at 72°C;
- 5 cycles of 1 min at 94°C, 1 min at 60°C and 2 min at 72°C;
- 20 cycles of 1 min at 94°C, 1 min at 55°C and 2 min at 72°C and
- final extension for 7 min at 72°C.

4.2.2.3 PCR optimization using cDNA template derived from rams

PCR reactions were carried out in a final volume of 50µl containing 49µl PCR master mix (50mM KCl, 10mM Tris-HCl, 1.5mM $MgCl_2$, 0,1% TritonX-100, 200µM of each dNTP, 200nM of each primer and 1 unit Taq polymerase (Promega Corporation, Madison, USA)), 0.5µl cDNA (around 100ng) and thermal cycling profiles 1 through 5 to identify the optimum amplification conditions for each primer set. In addition, the relative efficiency of Roche Taq DNA polymerase (Roche Diagnostics GmbH, Penzberg, Germany) was compared with the Promega polymerase at a range of different $MgCl_2$ concentrations (1.5mM, 2mM, 2.5mM and 5mM), using the thermal cycling profile 2.

Thermal cycling profile 2:

- Initial denaturation for 1 min at 94°C;
- 5 cycles of 1 min at 94°C, 1 min at 65°C and 2 min at 72°C;
- 25 cycles of 1 min at 94°C, 1 min at 60°C and 2 min at 72°C;
- final extension for 10 min at 72°C.

Thermal cycling profile 3:

- Initial denaturation for 1 min at 94°C;
- 5 cycles of 1 min at 94°C, 1 min at 70°C and 2 min at 72°C;
- 5 cycles of 1 min at 94°C, 1 min at 65°C and 2 min at 72°C and
- 20 cycles of 1 min at 94°C, 1 min at 60°C and 2 min at 72°C.

Thermal cycling profile 4, used only for 501 alleles:

- Initial denaturation for 5 min at 94°C;
- 30 cycles of 1 min at 94°C, 1 min at 65°C and 2 min at 72°C and
- final extension for 10 min at 72°C.

Thermal cycling profile 5, used only for 501 alleles:

- Initial denaturation for 2 min at 94°C;
- 30 cycles of 30 sec at 94°C, 30 sec at 65°C and 30 sec at 72°C and
- final extension for 5 min at 72°C.

4.2.2.4 Optimized PCR protocol for genotyping lambs and ewes

PCR reactions were carried out in 50µl final volume containing 49µl PCR master mix (50mM KCl, 10mM Tris-HCl, 2mM $MgCl_2$, 200µM of each dNTP, 200nM of each allele specific primer and 2 units Taq polymerase (Roche Diagnostics GmbH, Penzberg, Germany) and 0.5µl cDNA (around 100ng). The thermal cycling profile 2 was used for amplification of all sequences, with the exception of alleles C11 (ram 501), F12, F7, G12, F9 and E9 (ram 504), which were amplified for 35 cycles instead of 30.

A total of 17 lambs and 5 ewes were genotyped, using the optimized sequence specific amplification method. The typing of each lamb was repeated at least twice for all sequences and at least three times for those sequences found to be present in the majority of lambs (C7, E2, C6, F12, F7 and G12).

4.2.3 Cloning of PCR products excised from agarose gels

Sequence specific PCR products were excised from 0.7% agarose gels (section 2.5.4) and cloned into the pGEM T Easy vector (section 2.6.2). Each cloned PCR product was used to transform Promega JM109 high efficiency competent cells (section 2.7.2) cultured as described in section 2.7.3.

4.2.4 Selection of transformed colonies by PCR

White and blue colonies were picked and processed for colony PCR with each of the sequence specific primer pairs presented in table 4.2, as otherwise described in section 2.7.4.

4.2.5 Sequence analysis

To confirm the identity of the inserts plasmid DNA was prepared and sequenced as described in section 2.8. This data allowed the generation of a phylogenetic tree based on a Clustal W alignment showing the predicted relationship between sequences (section 2.8.3).

4.3 Results

4.3.1 Preliminary experiments

4.3.1.1 Assessing the quality of RNA

The quality of RNA preparations were assessed by spectrophotometry, as described in section 4.2.1. The ratios of optical density at 260 and 280 nm were between 1.8 and 1.9, (Table 4.1) indicating only limited protein contamination. In addition, PCR amplification of full length MHC class I transcripts from cDNA prepared from these samples indicates that the mRNA preparations were not degraded by contaminating nucleases (Fig. 4.2).

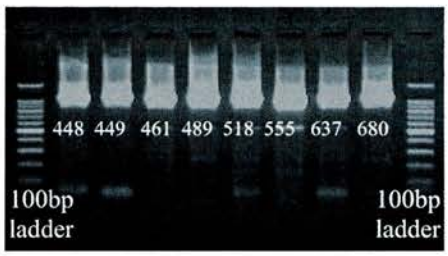


Figure 4.2: Agarose gel showing full length PCR products amplified from lamb cDNA with primers Cfl-Dr1 (chapter 5). The 100bp DNA ladder contains fragments from 100bp to 1000bp and an additional band at 1500bp.

4.3.2 Development and optimization of the class I SSP-PCR genotyping method

4.3.2.1 PCR optimization using previously defined plasmid DNA

4.3.2.1.1 Use of appropriate plasmid DNA concentration

The optimum plasmid template concentration was identified by sequence specific PCR of 100X, 500X and 100X dilutions of a 100ng/µl solution of the D5 plasmid. Using the protocol described in section 4.2.2.2, a band of slightly higher intensity was obtained from the 500 X dilution (Fig. 4.3) and this dilution of plasmid template was subsequently used throughout the optimization process.

Figure 4.3: Agarose gel showing sequence specific PCR amplification of the D5 sequence from x100, x500 and x1000 dilutions of the D5 plasmid

4.3.2.1.2 Confirmation of SSP-PCR specificity

Plasmid DNA containing each of the MHC class I sequence described in the previous chapter was amplified with each of the different primer pairs (Table 4.2) using the optimized conditions selected from the experiments detailed above. This confirmed that only the targeted sequence could be amplified by each primer pair (Fig. 4.4 and 4.5). The sizes of all PCR products were as expected.

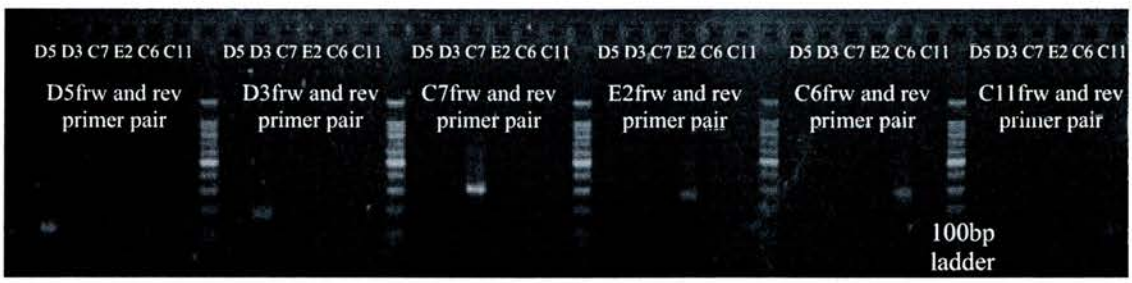


Figure 4.4: Agarose gel showing the specificity of each SSP pair. Amplification was restricted only to the SSP-associated plasmid DNA template. The identity of each plasmid DNA template is shown below the wells.

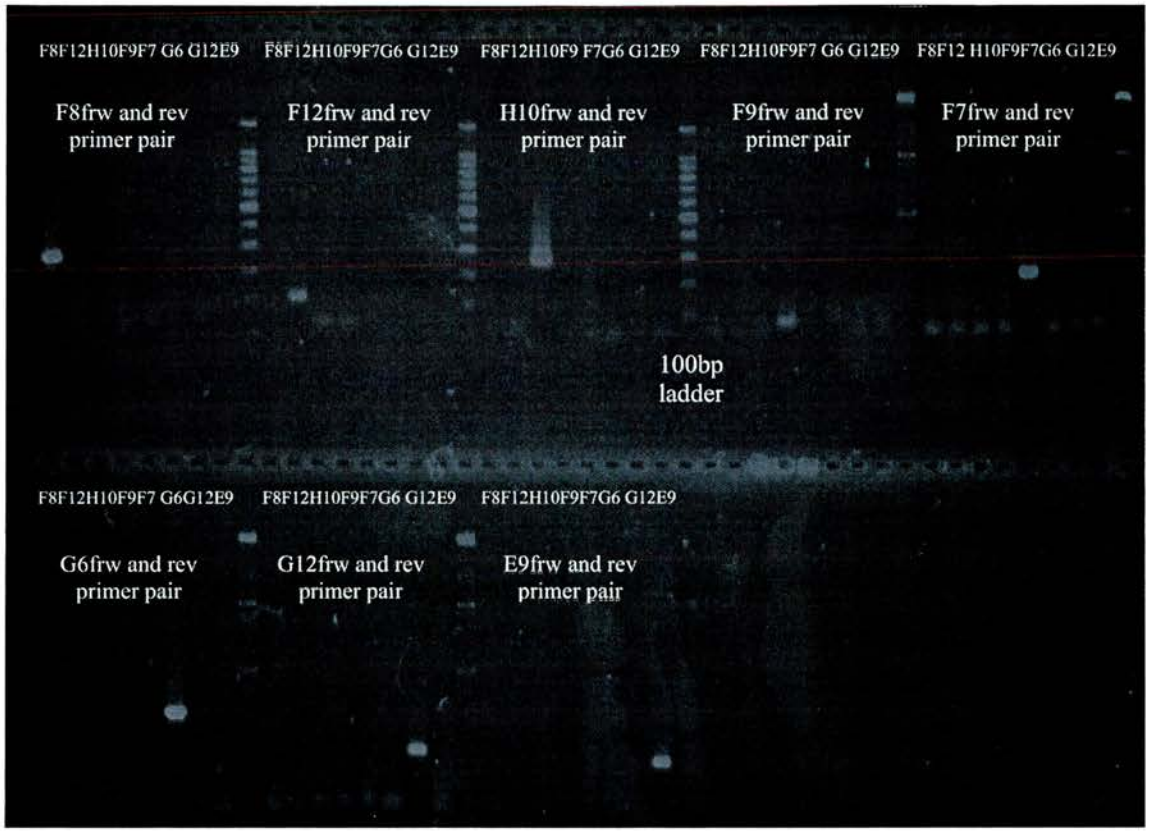


Figure 4.5: Agarose gel showing the specificity of each SSP pair. Amplification was restricted only to the SSP-associated plasmid DNA template. The identity of each plasmid DNA template is shown below the wells.

4.3.2.2 PCR optimization using cDNA prepared from rams 501 and 504

Once the specificities of the sequence specific primers were validated using plasmid templates a range of thermal cycling conditions were tested in order to optimize amplification using cDNA as the template.

4.3.2.2.1 Selection of a PCR cycling profile

With the exception of E2, all 501 and 504 sequences amplified using cycling profiles 1 to 4 (Fig. 4.6 to 4.10). However, when cycling profile 1 was used, non specific bands were evident using C6 SSP and the majority of the 504 SSP pairs (Fig. 4.6). By increasing the annealing temperature of the last 20 cycles from 55°C to 60°C (profile 2), the intensity of the contaminating bands was reduced, but not totally eliminated (Fig. 4.7). Increasing the annealing temperature of the first five cycles from 65°C to 70°C, (profile 3) did not improve the sensitivity

and reduced the efficiency of amplification of F12, H10 and G12 sequences (Fig. 4.8). A non touchdown protocol (profile 4) was tested but this also failed to reduce the intensity of the C6 contaminating bands (Fig. 4.9). Reducing the amplification time from 1 min to 30 sec (profile 5) resulted in a failure to amplify most of the 504 sequences (Fig. 4.10). Therefore, the use of cycling profile 2 provided the best possible results by reducing the intensity of the contaminating bands without deteriorating the efficiency of amplification of the required products. For that reason, cycling profile 2 was further used for optimization with appropriate $MgCl_2$ titration.

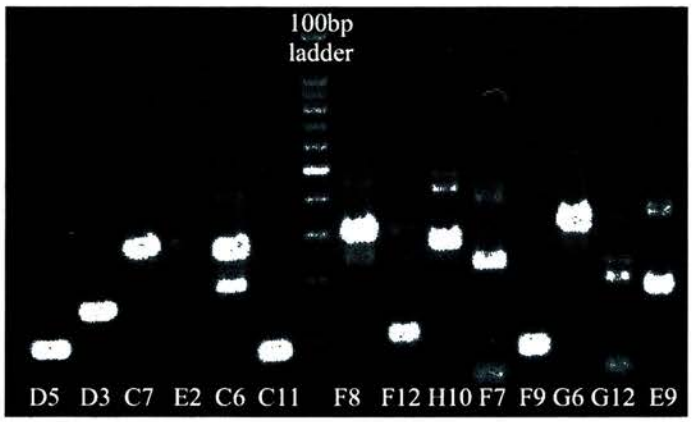


Figure 4.6: Agarose gel showing sequence specific PCR products amplified from cDNA prepared from rams 501 and 504, using the thermal cycling profile 1.

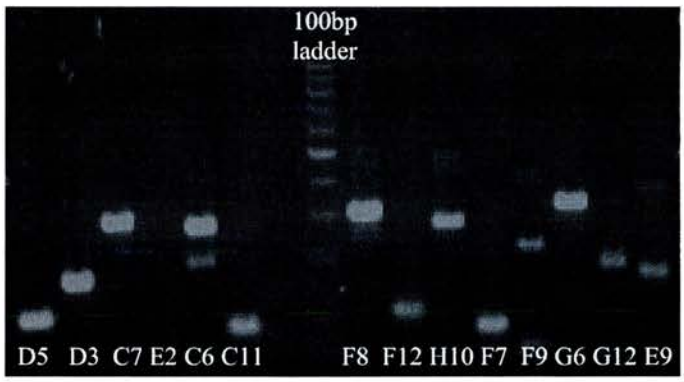


Figure 4.7: Agarose gel showing sequence specific PCR products amplified from cDNA prepared from rams 501 and 504, using the thermal cycling profile 2.

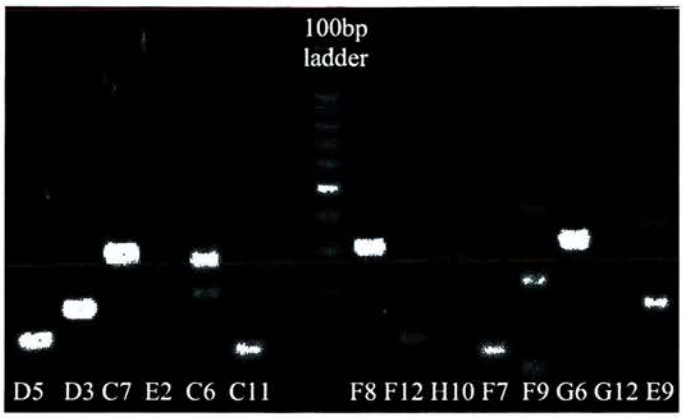


Figure 4.8: Agarose gel showing sequence specific PCR products amplified from cDNA prepared from rams 501 and 504, using the thermal cycling profile 3.

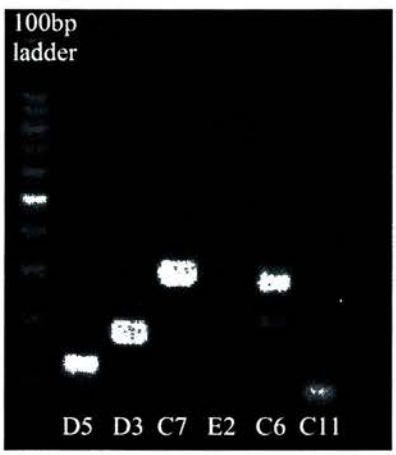


Figure 4.9: Agarose gel showing sequence specific PCR products amplified from cDNA prepared from ram 501, using the thermal cycling profile 4.

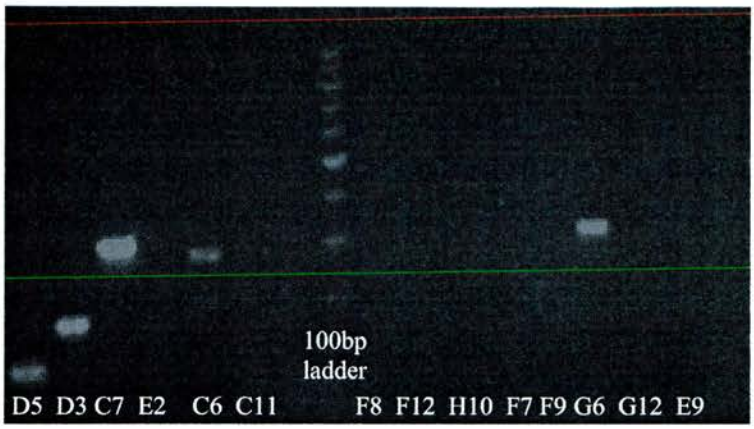


Figure 4.10: Agarose gel showing sequence specific PCR products amplified from cDNA prepared from rams 501 and 504, using the thermal cycling profile 5.

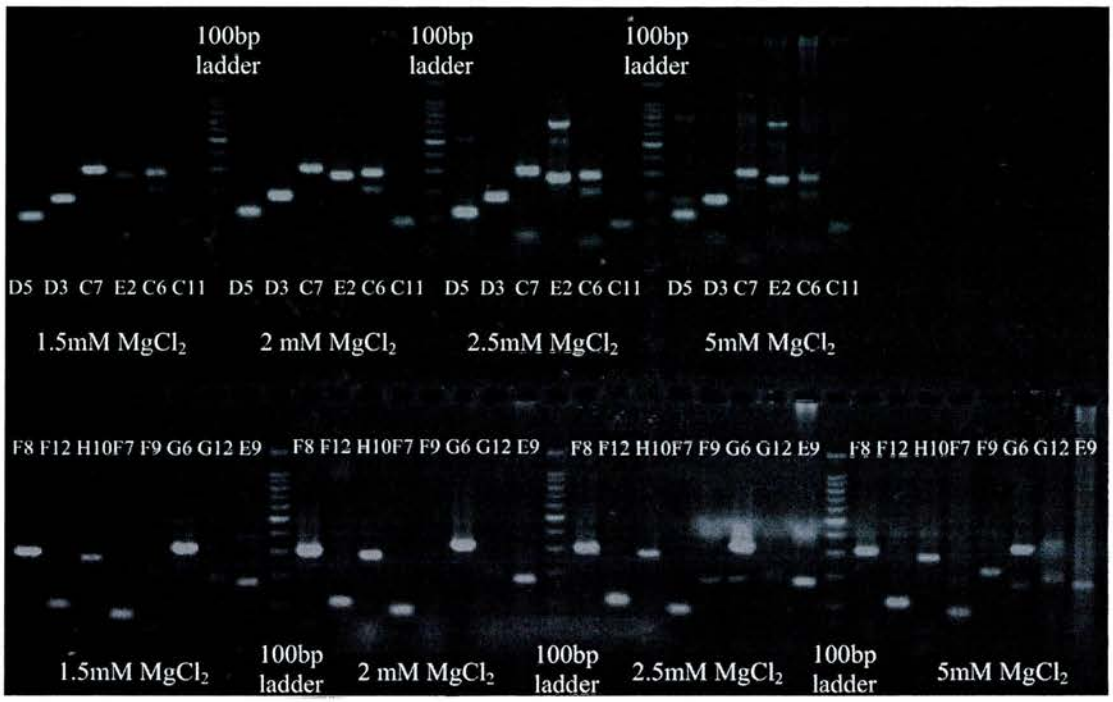


Figure 4.11: Agarose gel showing sequence specific PCR products amplified from cDNA prepared from rams 501 and 504, using 1.5, 2, 2.5 and 5mM $MgCl_2$ concentrations.

4.3.2.2.2 Selection of $MgCl_2$ concentration.

Using cycling profile 2, a range of $MgCl_2$ concentrations were tested to further optimize the SSP amplification method (Figure 4.11). Better amplification in terms of both concentration and specificity were obtained with 2mM $MgCl_2$. Lower $MgCl_2$ concentration produced bands of

lower intensity, while higher concentrations decreased specificity, as indicated by secondary bands and smearing. Sequences C11, F12, F7, G12, F9 and E9 required amplification with 2mM $MgCl_2$, using Roche Taq polymerase and the thermal cycling profile 2 but for 35 cycles (Fig. 4.12).

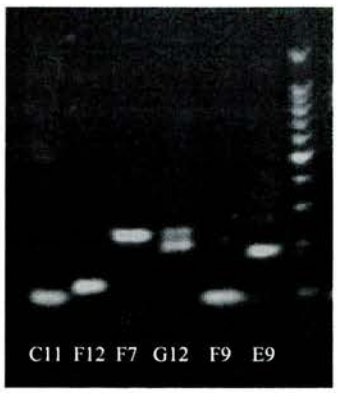


Figure 4.12: Agarose gel showing sequence specific PCR products amplified from cDNA prepared from rams 501 and 504. Sequences C11, F12, F7, G12, F9 and E9 were amplified, using the thermal cycling profile 2, but for 35 cycles

4.3.2.2.3 Shared sequences among animals

cDNA prepared from ram 501 was amplified with primers designed for sequences detected from ram 504 and vice versa. Sequences F7 and G12 that had been detected in single clones according to SSCP and sequence analysis of exons 2 and 3 in chapter 3 (Table 3.3), were shared between rams (Fig. 4.13), as they were detected in both 501 and 504 and in all their offspring (Fig. 4.13 and 4.14).

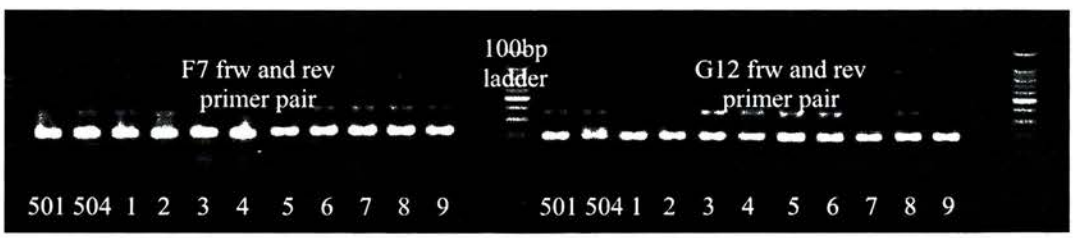


Figure 4.13: Agarose gel showing specific PCR products for sequences F7 and G12 amplified from cDNA prepared from rams 501 and 504 and from nine 501 offspring. The names of the lambs are presented as showed in the table 4.3 (section 4.3.4).

4.3.3 Genotyping of lambs and ewes

The 17 progeny of rams 501 and 504 were genotyped to determine from which parental MHC haplotype a given gene was derived using the optimized sequence specific amplification method described in section 4.2.2.4. The results are shown in Figures 4.13 and 4.14 and summarized in Table 4.3.

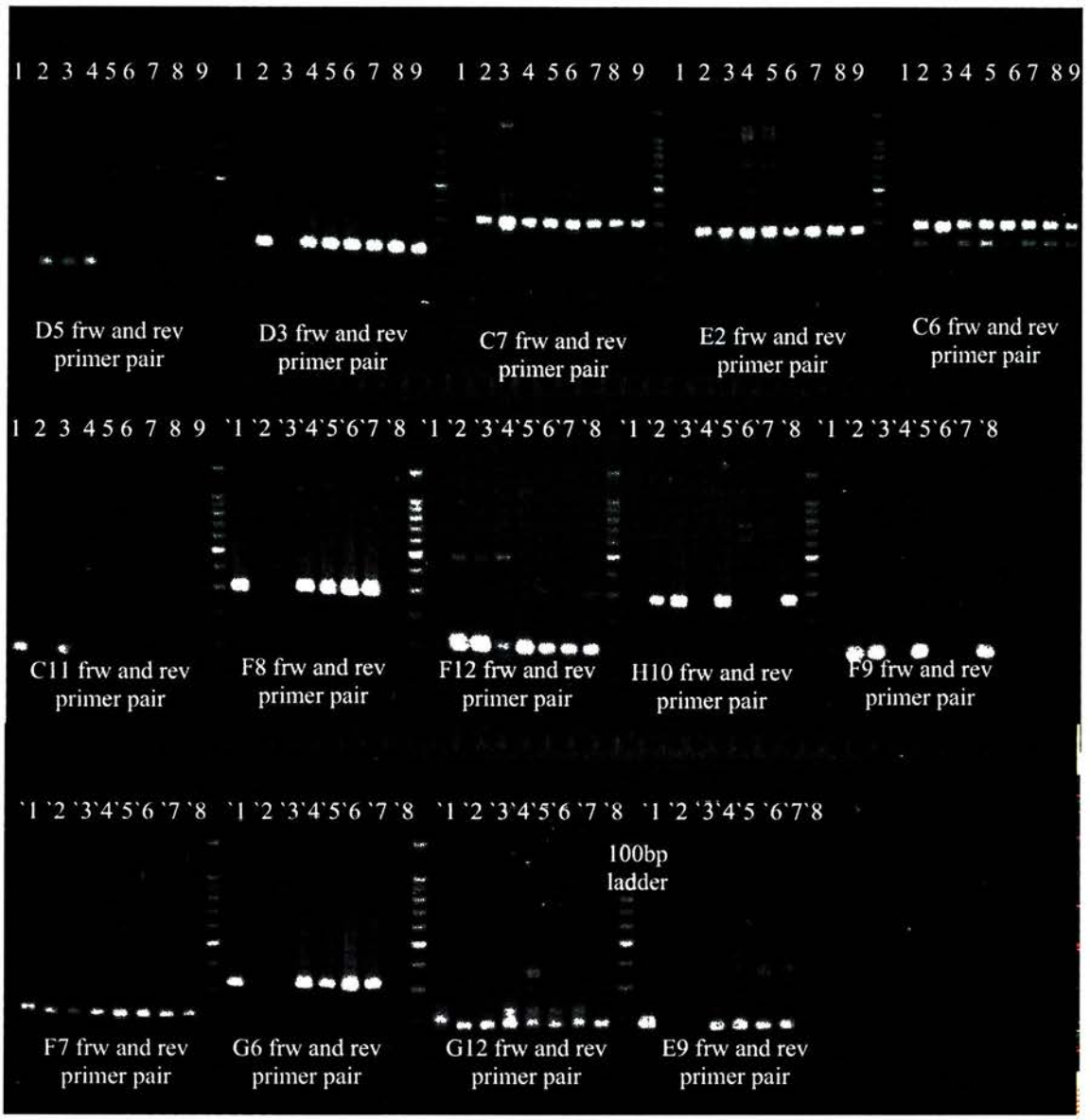


Figure 4.14: Agarose gel showing specific PCR products amplified from cDNA prepared from nine 501 and eight 504 offspring for each sequence specific primer pair. The names of the lambs are presented as shown in the table 4.3 (section 4.3.4).

Seven progeny of 501 (449, 486, 489, 518, 555, 637, and 680) provided sequences D3, E2 and C6, while two (448 and 461) provided sequences D5, C11 and C7. Four offspring of 504 (457, 573, 609 and 610) provided sequences F8, G6 and E9, while three (529, 530 and 673) provided sequences F12, H10 and F9. One lamb (596) sired by 504 yielded all six paternal sequences. The identity of each SSP-PCR product was confirmed by sequencing. SSP-PCR analysis of all the progeny of rams 501 and 504 consistently amplified the F7 and G12 sequences (Fig. 4.13 and 4.14).

4.3.4 Sequence analysis

Sequence specific PCR products from lambs 448, 449, 461, 486, 489, 457, 596, 609, 610, 673 and ewes 265N, 306N, 822N, 877N and 1098N were cloned and sequenced to confirm their identity. All sequence specific PCR products amplified from lambs' cDNA were identical to those that had been identified from rams except from the maternally inherited and newly identified ones.

Some SSP pairs clearly amplified maternally inherited class I sequences, as these differed by at least 4bp from those obtained from rams 501 and 504 (underlined sequences in table 4.3). An alignment of the maternally inherited sequences with those previously described is presented in figure 4.15. In particular, the D5 primer pair amplified sequences D5b1 and D5b2 from lambs 449 and 486. Genotyping of the mothers of 449 and 486 (ewes 306N and 822N) with the D5 specific primer pair followed by cloning and sequence analysis of the amplified products confirmed that D5b1 and D5b2 were maternally inherited (Fig. 4.15). Similar analysis of ewe 877N confirmed that sequences E2b and C6b were maternally inherited to lamb 461 (Fig. 4.15). In addition, SSPs designed for the F12 sequence amplified sequences F12b1 and F12b2 from lambs 609 and 610. These sequences differed from F12 by at least 4bp (Fig. 4.15) and genotyping of ewes1098N and 265N verified that they were maternally inherited.

Genotyping of lambs derived from ram 501 with primers designed for sequence C7 revealed a previously unidentified sequence inherited from 501. This sequence, C7b, differed from C7 by 25 nucleotides (Fig. 4.15) and always segregated with alleles D3, C6 and E2 (Table 4.3). The sequence C7b was not confirmed in ram 501, because the C7 sequence was amplified consistently. However, it is unlikely to be maternally inherited as it was identified in seven lambs all sired by ram 501, when mated with seven different ewes.

501 lambs	Sequences	504 lambs	Sequences
1: 448	D5, C11, C7, F7, G12	`1: 457	F8, F7, G6, E9, G12
2: 449	D5b1, D3, **C7b**, E2, C6, F7, G12	`2: 529	F12, H10, F9, F7, G12
3: 461	D5, C11, C7, E2b, C6b, F7, G12	`3: 530	F12, H10, F9, F7, G12
4: 486	D5b2, D3, **C7b**, E2, C6, F7, G12	`4: 573	F8, F7, G6, E9, G12
5: 489	D3, **C7b**, E2, C6, F7, G12	`5: 596	F12, H10, F9, F7, G12 F8, F7, G6, E9, G12
6: 518	D3, **C7b**, E2, C6, F7, G12	`6: 609	F8, F7, G6, E9, G12, F12b1
7: 555	D3, **C7b**, E2, C6, F7, G12	`7: 610	F8, F7, G6, E9, G12, F12b2
8: 637	D3, **C7b**, E2, C6, F7, G12	`8: 673	F12, H10, F9, F7, G12
9: 680	D3, **C7b**, E2, C6, F7, G12		

Table 4.3: Summary of the genotyping results shown in figures 4.13 and 4.14 by taking into consideration the sequencing results. *Underlined* sequences are maternally inherited. Sequence C7b shaded *in gray* is a newly identified class I sequence derived from ram 501.

An alignment of the predicted amino acid sequences described in chapter 3 with the newly identified 501 sequence C7b and the sequences identified from the ewes (Fig. 4.16) and a phylogenetic tree (Fig. 4.17) show the similarities among the sequences. Based on the available sequence data, the C7b sequence and each maternally inherited sequence amplified by SSP-PCR were phylogenetically closest to the individual sequences targeted by the SSPs, except for sequences D5b1 and D5b2, which were more similar to F12 (Fig. 4.17).

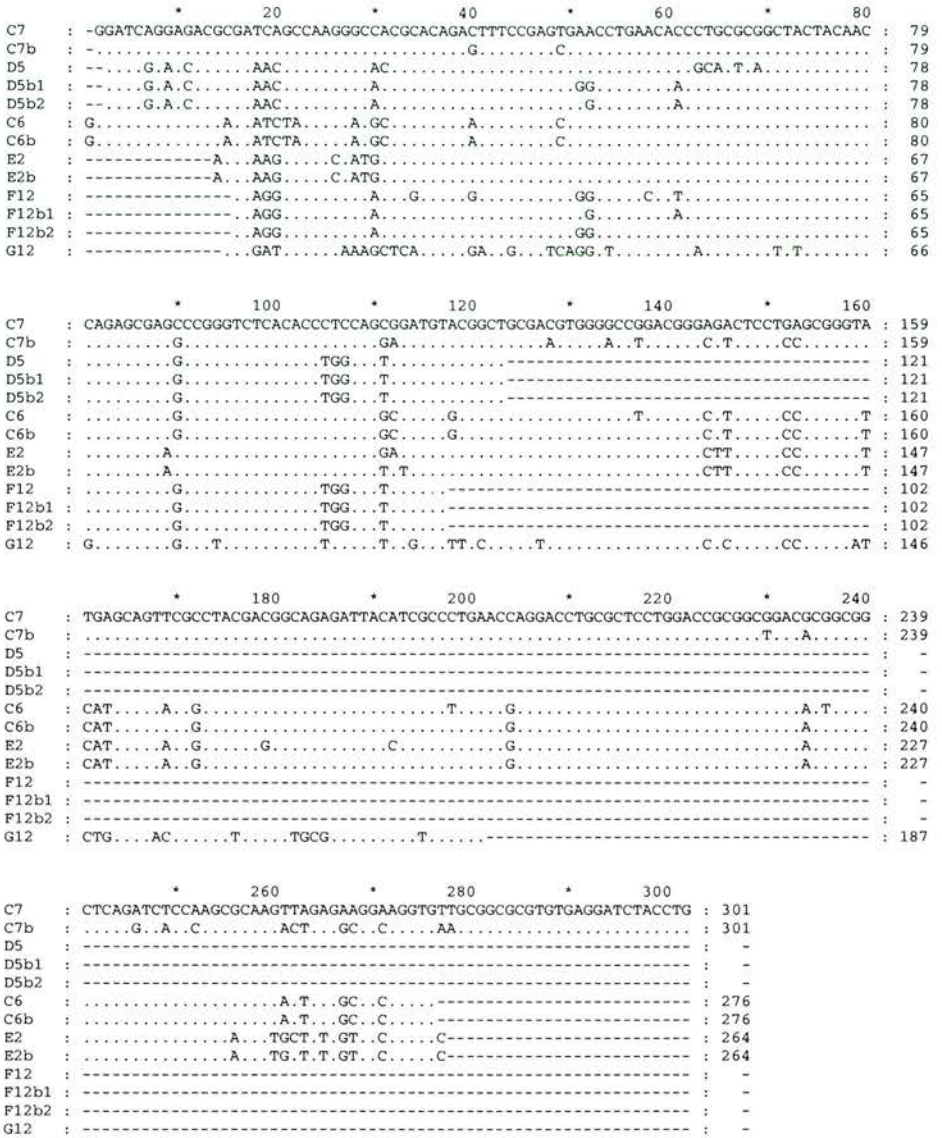


Figure 4.15: Alignment of new ovine MHC class I sequences identified on lambs or ewes cDNA. These sequences were amplified with sequence specific primers designed for class I sequences C7, D5, C6, E2, F12 and G12 previously identified from rams 501 and 504 in chapter 3. Identities and gaps are indicated by *dots* (.) and *dashes* (-), respectively.

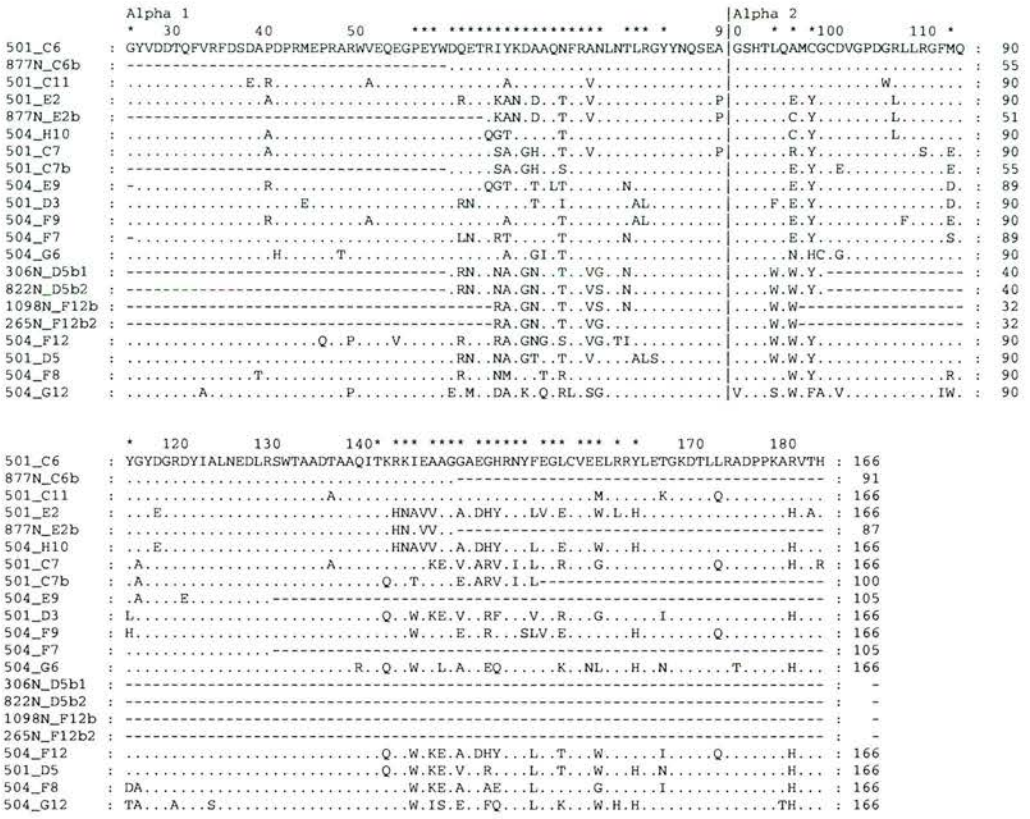


Figure 4.16: Alignment of predicted amino acid ovine MHC class I sequences identified on rams, lambs and ewes. The order of the sequences is based on the phylogenetic tree presented in figure 4.17. *Vertical lines* (|) indicate domain boundaries. Identities and gaps or absence of sequence are indicated by *dots* (.) and *dashes* (-), respectively. *Asterisks* (*) indicate MHC class I residues predicted to bind antigenic peptides or interact with T cell receptors in humans (Bjorkman et al. 1987a).

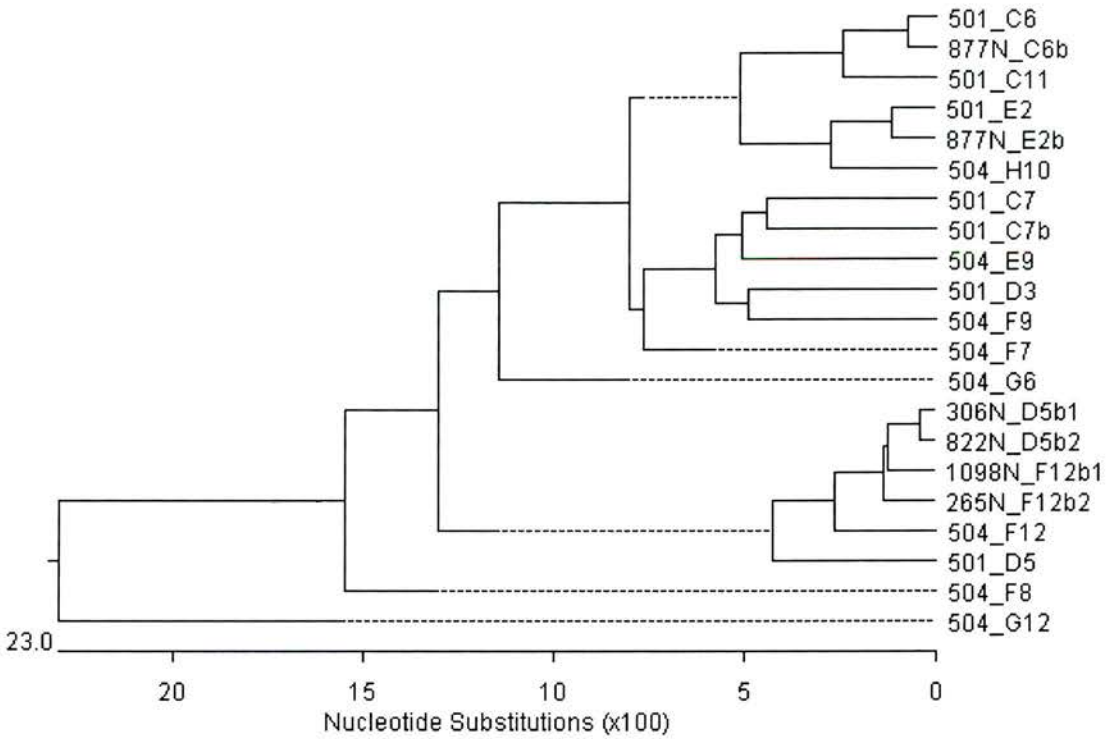


Figure 4.17: A phylogenetic tree of ovine MHC class I sequences identified on rams, lambs and ewes, based on clustal W alignments, visualizes the similarities among the identified sequences from the rams and ewes. The length of each branch is indicated by a *continuous line*.

4.3.5 Ovine MHC class I haplotypes

Based on the genotyping data presented in figures 4.13 and 4.14 and in table 4.3, figure 4.18 shows the 4 ovine MHC class I haplotypes carried by rams 501 and 504.

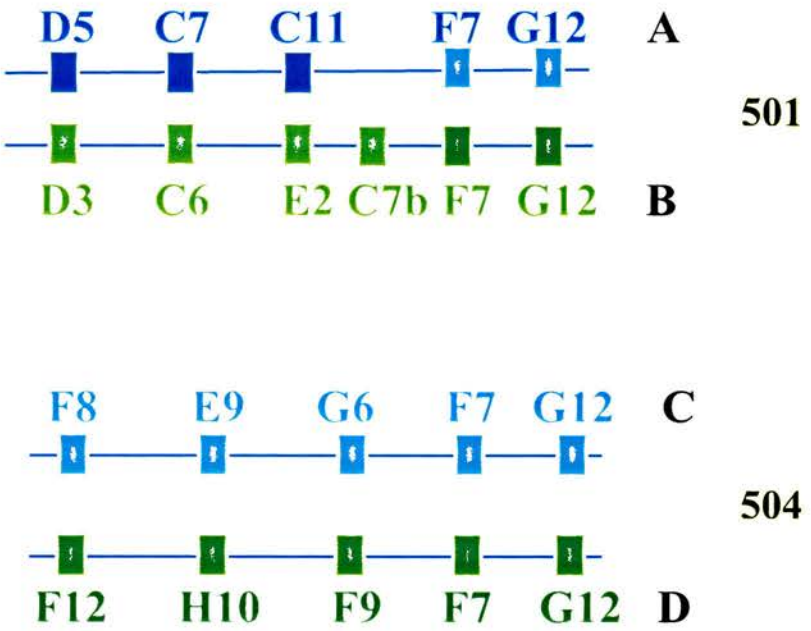


Figure 4.18: The MHC class I haplotypes carried by rams 501 and 504 as determined by sequence specific PCR genotyping. Alleles aligned do not necessarily correspond to loci.

4.4 Discussion

In this chapter each of the 14 class I genes previously identified from rams 501 and 504 were each assigned to individual MHC class I haplotypes using a novel SSP-PCR genotyping method.

4.4.1 Optimization of the SSP-PCR genotyping method

DNA-PCR-based methods can be broadly classified into three main categories (reviewed by Middleton D. 1999; Erlich et al. 1999). (i) those which generate a product containing internally located polymorphisms which can be detected by a second technique, such as restriction fragment length polymorphism (RFLP), specific oligonucleotide probing (SOP) or sequencing, (ii) those in which the polymorphism is identified as part of the PCR process, such as the sequence specific PCR and (iii) conformational analysis in which the polymorphism generates specific conformational changes in PCR products, such as heteroduplex analysis (HA), single strand conformational polymorphism (SSCP) and denaturing gradient gel electrophoresis (DGGE). Among those methods, sequence specific PCR is the method of choice if prior sequence data are available and the number of samples analyzed is not very large, as in the case of this study. The advantages of the sequence specific amplification genotyping include accuracy and rapidity and the ease with which it can be performed once optimized.

Prior to genotyping the progeny of rams 501 and 504, the SSP-PCR genotyping method was optimized and validated for each class I gene. Initially, plasmid DNA was used as the template to test the specificity of each primer pair. SSP pairs were designed on the principle that a mismatch at the 3′ terminal end is sufficient to prevent priming of non-specific reactions, when *Taq* DNA polymerase lacking 3′ to 5′ exonuclease activity is used (Newton et al. 1989). By testing a range of cycling conditions, template concentrations, $MgCl_2$ concentrations and *Taq* polymerases, the amplification of only the target sequence was confirmed for each primer pair (Fig. 4.4 and 4.5).

However, because plasmid DNA is usually super coiled and each construct contains only one sequence, further optimization of the SSP-PCR system using cDNA template was necessary. Touchdown PCR protocols in which the annealing temperature is gradually decreased, so that initial priming occurs only at sequences with perfect homology to the primers (McPherson et al. 1995) were used during the optimization process. As expected, specificity was improved by increasing the annealing temperature of the first cycles (Fig. 4.6 to 4.9).

Decreasing the incubation time during both annealing and extension would also be expected to improve specificity, due to reduced mis-priming (Elrich H. 1999). However, when shorter amplification times were used (Fig. 4.10), the yield of most PCR products was significantly reduced. Therefore, a balance was reached between specificity and yield.

Some sequences required a higher number of cycles than others for amplification. For most sequences, 30 cycles produced PCR products that appeared as intense bands on agarose gels. However, the C11, F9 and G12 sequences each required amplification for 35 cycles while a greater yield was obtained for F12, F7 and E9 using 35 cycles. This is likely to be due to a combination of different primer annealing kinetic parameters and lower levels of mRNA transcripts as discussed in chapter 3 (Table 3.3).

4.4.2 Genotyping of lambs and ewes

A total of nine offspring of ram 501 and eight offspring of ram 504 were genotyped by SSP-PCR. This allowed assignment of the 14 MHC class I sequences previously identified from the rams into four haplotypes (Fig. 4.18).

Although the SSP were each designed to amplify a specific class I sequence, previously unidentified sequences, some of which were maternally inherited, were also amplified using some SSP pairs (Table 4.3; underlined and shaded sequences). Specifically, the novel D5b1, D5b2, E2b, F12b1 and F12b2 sequences all differed by at least 4 nucleotides from their respective target counterparts D5, E2, and F12 (Fig. 4.15). As the PCR-SSP typing system was set up using the limited number of ovine MHC class I sequences identified from rams 501 and 504 and 8 published sequences, amplification of novel maternally inherited sequences was not unexpected. As the number of available ovine MHC class I sequences from different breeds rises, design of additional SSP will be possible, which may reduce cross-reactivity.

The amplification of a C7 product from all the offspring of ram 501 was attributed either to it being homozygous, or to cross-priming of a previously unidentified paternal sequence. Sequencing of C7 SSP-PCR products revealed a new sequence, denoted C7b, in 7 of the 9 offspring of ram 501. These 7 animals each inherited the 501B haplotype, while C7b was not found in two animals that inherited the 501A haplotype. The prior failure to identify the C7b sequence in ram 501 may be a result of polymorphism at the 3' end of the original primers (Newton et al. 1989) or could arise from very low transcript numbers and a failure to identify a clone by SSCP. It may also be due to an age related expression profile. The possibility of

genomic DNA contamination was excluded by digestion with DNAase I during RNA preparation in addition to the inclusion of appropriate controls for non specific amplification from RNA prior to reverse transcription.

4.4.3 Number of ovine MHC class I loci

A number of genetic studies in livestock and other species have revealed significantly greater genetic complexity and allelic diversity in MHC class I than initially predicted by serology (reviewed by Chardon et al. 1999; Ellis S. 2004; Parham et al. 1995). Prior to molecular genetic analyses, it was believed that the majority of MHC class I molecules in cattle were encoded by a single or two highly polymorphic loci (reviewed by Ellis and Ballingall, 1999). However, molecular genetic studies have since revealed evidence for at least six classical bovine class I loci, which occur in a number of combinations, and none of them is consistently transcribed (reviewed by Ellis S. 2004). In sheep, a serological and a cDNA study suggested that the ovine MHC contains at least three MHC class I loci (Grossberger et al. 1990; Millot P. 1984). Other researchers have suggested that the distribution and segregation of class I antigens in the population is consistent with at least two linked class I loci (Cullen et al. 1982; Garrido et al. 1995; Jugo and Vicario, 2001; Jugo et al. 2002; Stear and Spooner, 1981).

In chapter 3, 14 novel ovine MHC class I transcripts were identified from two heterozygous rams. Sequences D5, D3, C7 and E2 for ram 501 and F8 and F12 for ram 504 possibly encode highly expressed classical class I products, whereas the remaining sequences may encode poorly expressed classical or non-classical class I genes, as suggested by their transcription levels. In the current chapter, additional evidence was provided for the non classical nature of the weakly transcribed sequences F7 and G12. These sequences were identified from all the offspring of 501 and 504, suggesting that both rams are homozygous for these loci. Non classical MHC genes generally exhibit limited or no polymorphism, significantly lower expression and in most cases restricted distribution (reviewed by Braud et al. 1999). It is therefore likely that genes F7 and G12 represent such non-classical loci, as they are present in four haplotypes and are almost certainly transcribed at a significantly lower level in peripheral blood, based on the limited number of clones initially identified (2 out of 360; Table 3.3; chapter 3). However, since the full length coding regions of the F7 and G12 genes remain to be determined, it is possible that they represent poorly transcribed pseudogenes. A number of transcribed pseudogenes with deletions or early stop codons have been found within the cattle MHC class I region (Ellis S., IAH, Compton, UK, personal communication).

By genotyping the progeny of rams 501 and 504, the 14 transcripts identified in chapter 3 along with the novel C7b found here were assigned to individual haplotypes. Six transcribed class I loci, two of which are putative non classical were identified in one 501 haplotype (Fig. 4.18; 501 haplotype B), while five transcripts were found in the other (Fig. 4.18; 501 haplotype A) and in both 504 haplotypes (Fig. 4.18; 504 haplotypes C and D), together with the same putative non classical class I genes, suggesting a greater complexity for the ovine class I region than earlier predicted by serology. For the first time, ovine MHC class I haplotypes were characterized at the cDNA level and a PCR genotyping system for rapid screening of class I sequences was established.

Chapter 5. Analysis of full length ovine MHC class I transcripts

5.1 Introduction

The previous chapters describe the identification of 15 partial transcripts from two rams and the assignment of those transcripts into four haplotypes. The work described in this chapter aims to extend these analyses by molecular cloning and sequencing of the entire coding region of those genes.

Molecular cloning and sequencing of transcribed full length Major Histocompatibility Complex (MHC) class I genes in mouse (Evans et al. 1982; Steinmetz et al. 1981a,b) and man (Geraghty et al. 1987; Malissen et al. 1982) has revealed their fine structure. Both classical and non classical MHC class I genes exhibit an eight exon format as follows:

- 5′-untranslated region (5′-UTR): This region varies between 20 and 30bp in the area between the TATA box and the translation codon.
- Exon 1: The first exon encodes the leader peptide in the immature protein and varies from 21 to 29 residues.
- Exons 2 and 3: These exons encode the extracellular α1 and α2 domains that encode the peptide binding region and consist of 90 and 92 residues, respectively. These domains belong to the immunoglobulin superfamily and their overall structure is well conserved among different species.
- Exon 4: This exon consists of 92 residues and encodes the α3 extracellular domain where β2 microglobulin binds.
- Exon 5: This exon encodes the transmembrane region of the protein and exhibits the greatest variability in length among different class I loci within species (Crew et al. 1991).
- Exons 6, 7 and 8: These exons encode the cytoplasmic domain of the protein and may vary in length as well. Exon 8 usually includes the stop codon. However, in HLA-B genes, the termination codon is found in exon 7 and therefore B locus products lack the residues encoded by exon 8 (Gussow et al. 1987). Alternative splicing of those exons

further contributes to the variation of the length of that region in mouse (Lew et al. 1987), human (Krangel S. 1986), cattle (Ellis et al. 1992) and sheep (Grossberger et al. 1990). Variation in length may also be caused by in-frame termination codons that give rise to truncated proteins (Steinmetz et al. 1981b). This is quite common for non classical class I genes (Lew et al. 1987; Ulker et al. 1990).

- 3′-untranslated region (3′-UTR): This region can vary considerably in length and has lower nucleotide similarity between different class I genes, probably because it lacks the structural constrain of coding regions.

At the initiation of this project, only one full length (including the predicted start and stop codons) ovine MHC class I sequence was present in Genbank (Accn. No M34676), which shows greatest similarity to an allele at the HLA-B locus (Grossberger et al. 1990). Its overall fine structure appears similar to that described for functional human and murine class I genes. This and other partial ovine MHC class I sequences appear most similar to their cattle counterparts and in common with BoLA sequences, they contain transmembrane domains of variable length (Grossberger et al. 1990). Transmembrane domain length and sequence motifs within exons 5-8 have provided locus-specific markers in some species (Holmes and Ellis 1999; Parham et al. 1989). In addition, phylogenetic analysis of the α3, transmembrane and cytoplasmic domains is more likely to cluster alleles at distinct loci (Hugh and Nei, 1988) than similar analysis of exons 2 and 3. Therefore, data derived from that region would be useful for phylogenetic analysis of ovine MHC class I sequences. Furthermore, full length coding sequence data for those genes might provide some insight on their functionality and classical or non classical nature.

5.2 Materials and methods

5.2.1 Animals and preparation of nucleic acid

Peripheral blood samples were obtained from rams 501 and 504, as described in section 2.9.2. Immediately after preparation of PBMC, poly-adenylated mRNA was extracted and cDNA synthesized as described in sections 2.3.1 and 2.4.2, respectively.

5.2.2 Amplification of full length class I transcripts

Amplification of full length class I transcripts was carried out using three different forward (Afl, Bfl and Cfl, Table 5.1, Fig. 5.1) and two reverse primers (Dr1 and Dr2, Table 5.1, Fig. 5.1). To obtain sequence data for the 5-prime and 3-prime ends of the various ovine MHC class I transcripts, additional amplifications using the forward primers Afl and Cfl with the internal reverse primer Br (Fig 5.1, Table 5.1) and the internal forward primer Br' (Table 5.1, Fig 5.1) with the reverse primer Dr1 were carried out. Primers Bfl and Dr1 are identical to the Bov21 and Bov3 primers described by Pichowski et al. (1996). The remaining primers were designed using previously described ovine, bovine, human and canine class I sequences (Fig 5.2). Primer Afl was designed using the single full length ovine MHC class I sequence (accession number: M34676) available on the outset of this study. The primer Cfl was based on a region generally conserved between ovine and bovine class I sequences (Fig. 4.2). Primer Dr2 was designed on the basis of 3' RACE data described in section 5.3.5.

PCR reactions were carried out in a final volume of 50µl containing 48 or 49 µl PCR master mix (60 mM Tris-SO_4, 18mM ammonium sulphate, 200 µM of each dNTP, 200 nM of each primer and 1U high fidelity Platinum Taq DNA polymerase (Invitrogen Ltd, Paisley, UK)) and 1 or 2 µl cDNA (≈150 ng). The cycling profile described by Parham et al. (1995) was used for amplification of full length class I transcripts to reduce formation of recombinant alleles. This comprised an initial denaturation for 2 min at 94°C, 25-30 cycles of 1 min at 94°C, 30 sec at 50°C and 2 min at 68°C, and a final extension for 7 min at 68°C. When primer Dr2 was used the annealing time was reduced to 10 seconds. The reactions were carried out in a 96 well Hybaid MBS 0.2S thermal cycler (Thermo Hybaid, Basingstoke, UK).

```
          -Primers Afl, Bfl, Cfl--------------->
          -exon 1--->
1     CCCATGACCAGAGGATTGCGAGTAATGGGGCCGCGAACCCTCCTGTTGCTGCTCTCGGGAGTCCTGGTCCTGACCGAGAT    80
              -exon 2--->
81    CCGGGCGGGCCCCCACTCCATGAGGTATTTCAGCACCGCCGTGTCCCGCGCCGGCGCCGGGGAGCCCCGGTACCTGGAAG   160
161   TCGGCTACGTGGACGACACGCAGTTCGTGCGGTTCGACAGCGACGCCCCGGATCCGAAGATGGAGCAGAGGGAGCCGTGG   240
241   ATGAAGCAGGTGGGGCCGGAGTATTGGGATCGGAACACGCGAAATCCCAAGGGCAACGCACAGACTTTCCGAGTGGGCCT   320
                                          -exon 3--->
321   GACCATCCTGCGCGGCTACTACAACCAGAGCGAGACCGGGTCTCACACCTGGCAGTGTATGTACGGCTGCGACGTGGGGC   400
                                                                  <------primer Br
                                                                  ---primer Br'---
401   CGGACGGGCGTCTCCTCCGCGGGTTCATGCAGTTCGGCTACGACGGCAGAGATTACATCGCCCTGAACGAGGACCTGCGC   480
      -
      ->
481   TCCTGGACCGCGGCGGACACGGCGGCTCAGGTCACCCAGCGCAAGTGGGAGAAGGAAGGTGCGGCGGACCACTACAGGAA   560
                                                                          -exon 4
561   CTACGTGGAGGGCACGTGCGTGGAGTGCGTGCGCAGATACCTGGAGATCGGGAAGGAACAGCTGCAGCGAGCAGACCCTC   640
      --->
641   CAAAGGCACATGTGACCCATCACCCCATCTCTGGCCATGATGTCACCCTGAGGTGCTGGGCCCTGGGCTTCTACCCTGAG   720
721   GAGATCTCACTGACCTGGCAGCGCAATGGGGAGGACCAGTTGCAGGACATGGAGCTTGTGGAGACTAGGCCTTCAGGGGA   800
801   TGGAACCTTCCAGAAGTGGGCGGCCCTTGGTGGTGCTTCTGGAGAGGAGCAGAGATACACGTGCCATGTGCAGCATGAGG   880
                                    -exon 5--->
881   GGCTTCAGGAGCCCCTCACCCTGAGATGGGAACCTCCTCAGACCTCCTTCCTCACTTCCTCAATGGGCATCATTGTTGGC   960
                                                                  -exon 6--->
961   CTGGTTCTCCTCGTCATGGTGGCTGTGGTGGCTGCAGCTGTGATCTGGAGGAAGAAGTGCTCAGGTGAAAAAAGAGGGAC  1040
                                                                        <--Dr1, Dr2
                          -exon 7--->                          exon8-3UTR--->
1041  CTATACCCAGGCTTCAAGCAATGACAGTGCCCAGGGTTCTGATGTGTCTCTCACGGTTCATAAAGTGTGAGACAGTGATG  1120
                                                                        CTG
      primers-
1121  CTGCATCCCGCTATGTGCCATCAGATCCCCGGACCCCTCTTTCTGAAGCTGCATCTGCACGTGTCTGTGCTCCTAGTAGC  1200
1201  ATAACGTGAGGAGTTGGGGAGACCGTTCACCCCTGCCCACCGCGCCCCCTCCTGCCCTGACCTGTGTTCTCCTCCCTGAT  1280
1281  CCACTGTCCTGTTCCAGCAGCAGACAGGGCTGGGCCGTCTCCATCCCTGTCTTTGCTTCGTATGCACTGAGTAATGATGT  1360
1361  CTTATTTCCTTATTGAAAATAAAATCTGTATGTATG                                               1397
```

Figure 5.1: Ovine class I MHC sequence (M34676, Grossberger et al. 1990) showing the position and sequence of the forward primers Afl, Bfl and Cfl in *italics*, *underlined* and in *bold*, respectively. Reverse primers Dr1 and Dr2 are in *bold* and *bold italics*, respectively. The first three bases of each exon are in bold and the stop codon is underlined.

Species	Ref	Class I	*No*	primers *Afl*, Cfl **ATG***ACCAGAGGATTGCGRGTA***ATGGGGCCGCGAACCSTC**
Human	A	HLA-A	83	A--GCC--C----C---C------C--
		HLA-B	168	A---KG--C-Y--CR--C-----SS--
		HLA-Cw	53	A----G--C----C---C-R-R--C--
Cattle	B	BoLA*	10	A----**R**--Y---**R**----**R**------C--
Sheep	C	OLA	1	----------A----------C--
Dog	D	DLA-88	3	-GAR-SKRG-YKR--R-T--SS--R--C--

Figure 5.2: Strategy for design of the Afl and Cfl forward primers based on an alignment of consensus human, bovine, ovine and canine class I sequences obtained from A: http://sdmc.krdl.org.sg:8080/fimm/, B: http://www.projects.roslin.ac.uk/bola/, C: M34672 to M34676, U26000 to U26010 and U03092 to U03094, and D: http://www.ebi.ac.uk/ipd/mhc/dla/index.html. The primer Afl is in *italics* and Cfl is in *bold*. *No* represents the number of sequences aligned and *dashes* indicate nucleotides that are identical to the primer sequence. Polymorphic sites in *red* and *blue* show that the different nucleotide from the primer sequence is represented by only one or two sequences, respectively.

Forward Primers		**Reverse Primers**	
Name	*5' Sequence 3'*	*Name*	*5' Sequence 3'*
Afl	ATGACCAGAGGATTGCGRGT	Dr1 =Bov3	GGATGAAGCATCACTCAG
Bfl =Bov21	ATGCGAGTCATGGGGCC	Dr2	GGATGAAGCATCACTGTC
Cfl	ATGGGGCCGCGAACCSTC	Br	AGCGCAGGTCCTCGTTC
Bf	GAACGAGGACCTGCGCT		

Table 5.1: Nucleotide sequence of oligonucleotide primers used for the amplification of full length class I transcripts.

5.2.3 3' Rapid Amplification of cDNA Ends (RACE)

The 3-prime ends of the D5 and D3 transcripts were amplified by 3'-RACE (section 2.4.3). mRNA was converted into cDNA, using the Superscript first-strand synthesis system for RT-PCR (Invitrogen Ltd, Paisley, UK), as described in section 2.4.2, but using an oligo-dT primer including a 36bp adapter incorporated in its 3'-end. In the initial amplification step, primers D5D3frw and adapter 1 were used. D5D3frw was designed to amplify both D5 and D3 sequences and adapter 1 annealed to the adapter of the oligo-dT primer used for cDNA synthesis (Table 5.2, Fig. 5.3). PCR reactions were carried out in 50 µl final volume containing 49 µl PCR master mix (60 mM Tris-SO_4, 18 mM ammonium sulphate, 200 µM of each dNTP, 200 nM of each primer and 1U high fidelity Platinum *Taq* DNA polymerase (Invitrogen Ltd, Paisley, UK)) and 1 µl cDNA (≈150 ng). The thermal cycling profile was: initial denaturation for 2 min at 94°C, 20 cycles of 1 min at 94°C, 20 sec at 65°C and 2 min at 68°C, and final extension for 10 min at 68°C.

The second amplification step was conducted with 2 µl from the first amplification step and 48 µl PCR master mix (as above), with the forward allele specific primers D5frw or D3frw and the reverse primer adapter 2, which has 12bp overlap with primer adapter 1 (Table 5.2, Fig. 5.3). The thermal cycling profile was: initial denaturation for 1 min at 94°C, 15 cycles of 1 min at 94°C, 20 sec at 60°C and 2 min at 68°C, and final extension for 10 min at 68°C.

Forward Primers		**Reverse Primers**	
Name	***5' Sequence 3'***	***Name***	***5' Sequence 3'***
D5D3frw	CCGAGTGAACCTGAACACCGCA	OligodTadapter	GGCTGTCAACGATACGCTACGTAACGG
			CATGACAGTGTTTTTTTTTTTTTTTT
D5frw	AGGAGGGTGTTGCGGAGAGACA	Adapter 1	CGCTACGTAACGGCATGACAGTG
D3frw	AAGGAGGGTGTTGCGGAGAGATT	Adapter 2	GGCTGTCAACGATACGCTACGTAA

Table 5.2: 3'-RACE primers

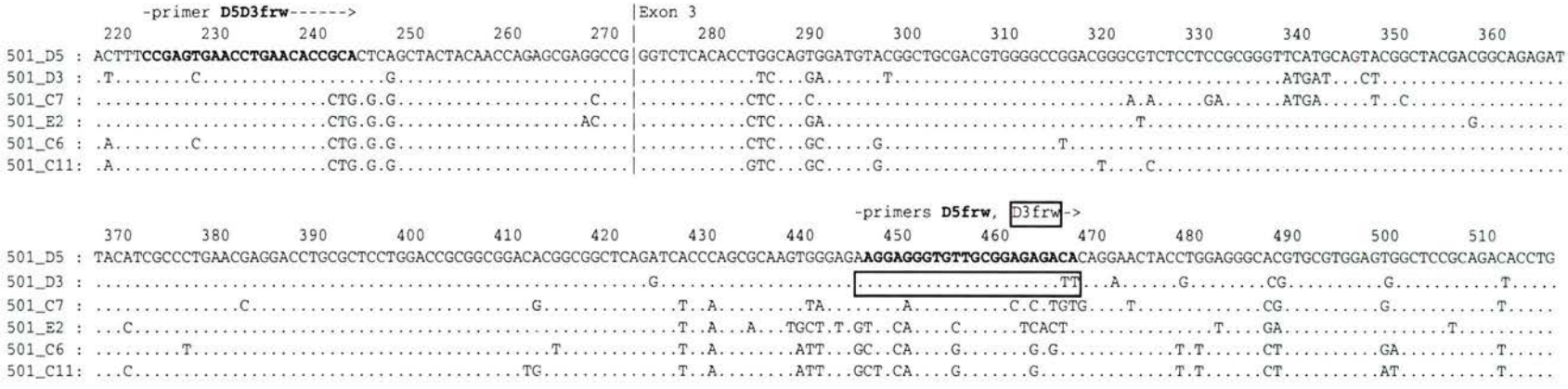


Figure 5.3: Alignment of the ovine MHC class I sequences obtained for ram 501 showing the position of the primers designed for sequence specific RACE amplification of the D5 and D3 sequences. Identities are indicated by *dots*. The numbering of nucleotides is based on the full length class I sequence M34676 (Grossberger et al. 1990).

5.2.4 Cloning of PCR products and transformation

Full length, partial and 3'-end PCR products generated as described in sections 5.2.2 and 5.2.3 were purified from 0.7% agarose gels, as described in section 2.5.4 and cloned into the pGEM T Easy vector or the pCR4Blunt TOPO vector (section 2.6.1), as described in section 2.6.2 and 2.6.3, respectively. Cloned PCR products were then used to transform Promega JM109 or TOP10 competent cells as described in section 2.7.2 and cultured for selection of white colonies, as described in section 2.7.3.

5.2.5 Selection of transformed colonies by PCR

White colonies were picked and processed for colony PCR with primers B and Br, as described in section 2.7.4. Clones containing full length inserts and inserts obtained with primers Cfl and Br were selected by colony PCR, whereas cloned 3'-RACE PCR products and inserts obtained with primers Br' and Dr1 were sequenced without prior PCR screening.

5.2.6 RFLP analyses

Colony PCR products of clones containing full length ovine MHC class I sequences and amplification products with primers Cfl and Br were screened by restriction fragment length polymorphism (RFLP), as a means of distinguishing the different sequences prior to the costly task of sequencing. RFLP genotyping was based on information derived from SSCP and sequence analyses of exons 2 and 3 (Chapter 3), as shown in figure 5.4. Approximately 260 clones per ram were screened by RFLP. Digestions were conducted, as described in section 2.5.9. The digested DNA fragments were separated in 2% agarose gels (section 2.5.3), except where better resolution was required, 6% polyacrylamide gels (section 2.5.10) were used.

5.2.7 Sequence analysis

Up to 5 replicates of clones with the same RFLP pattern were picked for sequencing. Plasmid DNA was recovered and analyzed as described in section 2.3.3 and 2.5.6. Sequencing reactions and analysis were performed as described in section 2.8. The sequenced clones that matched the consensus sequence from at least three clones, where available, were stored in glycerol stocks (section 2.7.5).

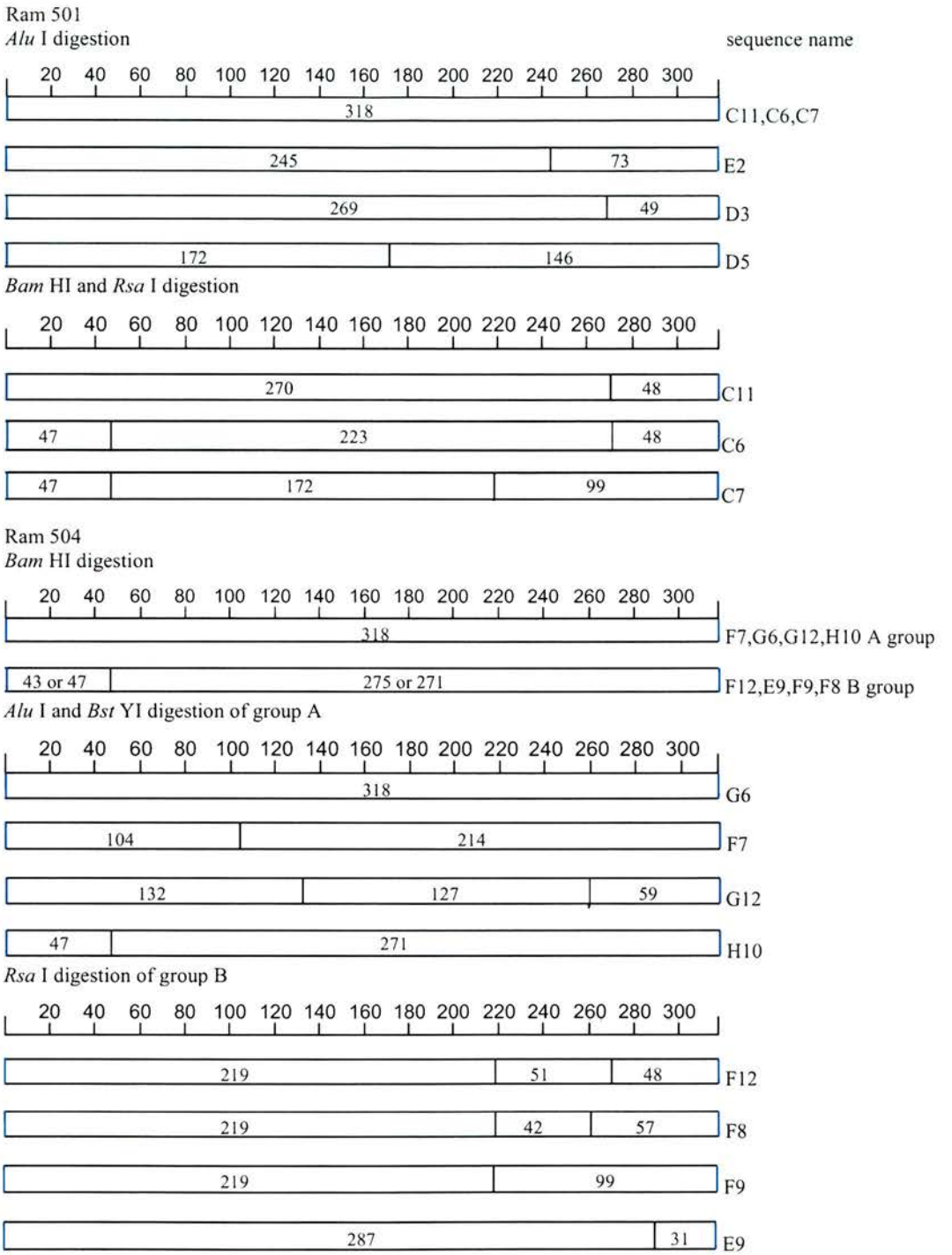


Figure 5.4: Diagramatic representation of the PCR-RFLP patterns associated with the ovine MHC class I exons 2 and 3 sequences obtained from rams 501 and 504. Sizes are indicated in base pairs.

5.3 Results

5.3.1 Amplification of full length class I transcripts

Using the primer pairs shown in table 5.3, PCR products of the expected size were obtained from cDNA template prepared from rams 501 and 504, (Fig. 5.5 and 5.6). The reverse transcription (RT) control indicates that the products were not a result of genomic DNA contamination.

Primer pair	PCR product size
Afl-Dr1	1135bp
Bfl-Dr1	1123bp
Cfl-Dr1	1098bp or 1114bp
Cfl-Br	457bp
Br-Dr1	674bp or 657bp
Cfl-Dr2	1098 or 1104

Table 5.3: Expected size of the PCR products obtained from each primer pair.

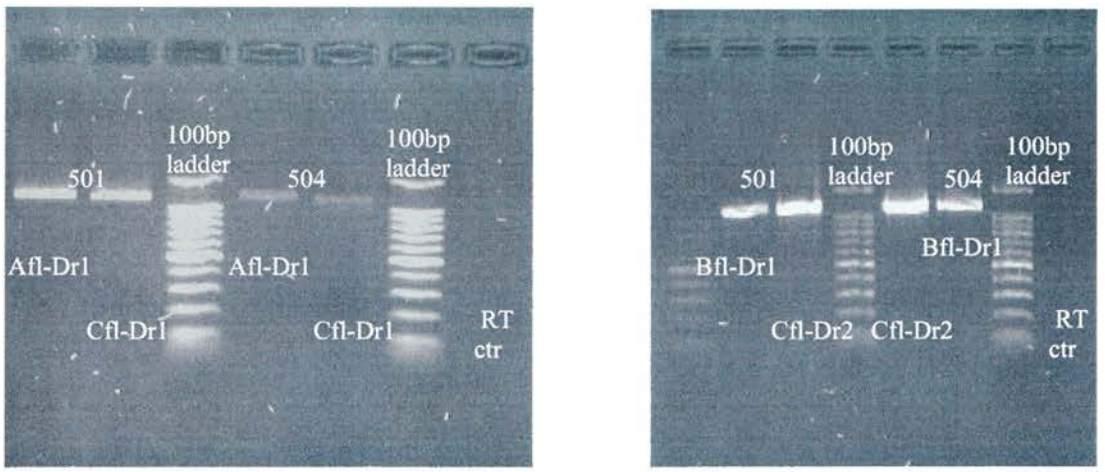


Figure 5.5: Agarose gel showing PCR products of the expected size amplified from cDNA prepared from rams 501 and 504 with primer pairs Afl-Dr1 and Cfl-Dr1 (left), Bfl-Dr1 and Cfl-Dr2 (right). The 100bp DNA ladder contains fragments from 100bp to 1000bp and an additional band at 1500bp. The RT ctr controls for genomic DNA amplification.

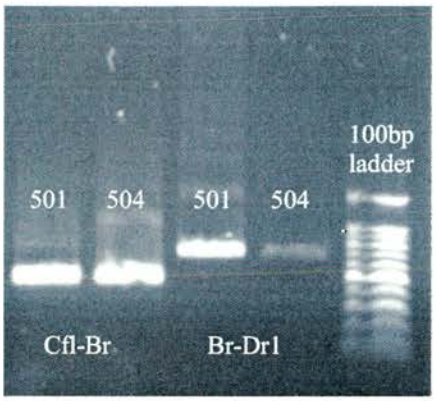


Figure 5.6: Agarose gel showing PCR products amplified from cDNA prepared from rams 501 and 504 with primer pairs Cfl-Br and Br-Drl.

5.3.2 3'-RACE PCR products

Since the reverse primer Drl failed to amplify sequences D5 and D3 (section 5.3.6), sequence data for the 3' untranslated region of D5 and D3 were obtained by 3'-RACE. Using the fully nested amplification protocol described in section 5.2.3, an amplification product of around 850bp was obtained for each sequence (Fig. 5.7).

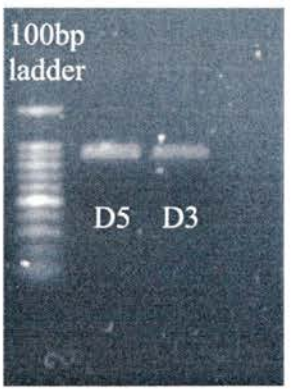


Figure 5.7: Agarose gel showing second step 3'-RACE PCR products.

5.3.3 Cloning and selection of recombinants by PCR

PCR products amplified from cDNA derived from both rams with primer pairs presented in tables 5.1 and 5.2 were gel purified and ligated into bacterial vectors. Following bacterial transformation approximately 500 white colonies were screened by colony PCR with primers B and Br (Fig. 5.8). Sixty positive clones amplified with each of the primer pairs Afl-Drl, Bfl-

Dr1, Cfl-Dr1 and Cfl-Br and 20 clones amplified with primer pair Cfl-Dr2 were chosen for PCR-RFLP analysis to identify candidates for sequencing.

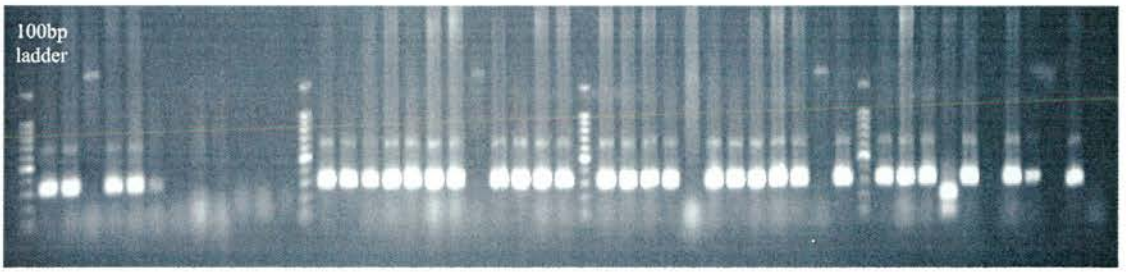


Figure 5.8: Agarose gel showing the results of colony PCR on clones containing full length inserts.

5.3.4 PCR-RFLP analysis

A PCR-RFLP method was developed to distinguish between the class I sequences carried by the rams, prior to sequencing (section 5.2.6). Clones containing full length ovine MHC class I sequences and the 457bp product amplified with primers Cfl and Br were genotyped by RFLP conducted on colony PCR products with primers B and Br. Separation of the restriction fragments on agarose gels distinguished between all the sequences described in figure 5.4, except F12 and F8, which required the increased resolution provided by polyacrylamide gels. The PCR-RFLP genotyping results are presented in table 5.4. Figures 5.9 and 5.10 show the range of different RFLP patterns associated with the sequences identified in chapter 3.

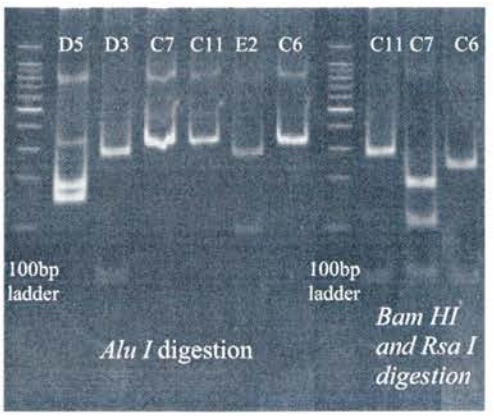


Figure 5.9: Polyacrylamide gel showing the range of PCR-RFLP patterns associated with the class I transcripts amplified from ram 501.

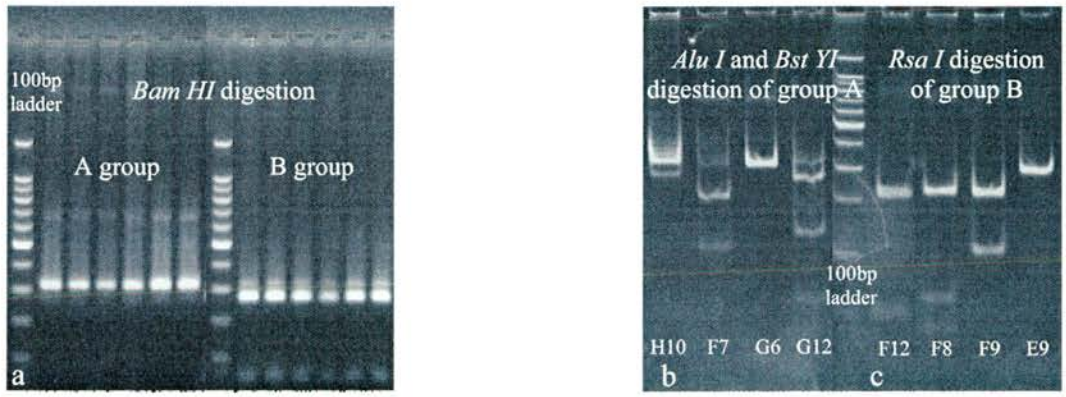


Figure 5.10: Agarose (a) and polyacrylamide (b,c) gel showing PCR-RFLP patterns associated with the class I transcripts amplified from ram 504. Digestions with restriction enzymes and groups A and B are described in figure 5.4.

5.3.5 Sequencing of 3'-RACE PCR products

Sequencing of 3'-RACE products demonstrated that the Dr1 primer differed by 3bp at the 3' prime end from the D5 and D3 sequences due to a 19bp deletion (Fig. 5.11 and 5.12). This deletion was also found in the F8 and F12 sequences from animal 504. The Dr2 primer used to amplify full length D5 and D3 sequences (section 5.3.6) was based on these data (Fig. 5.11).

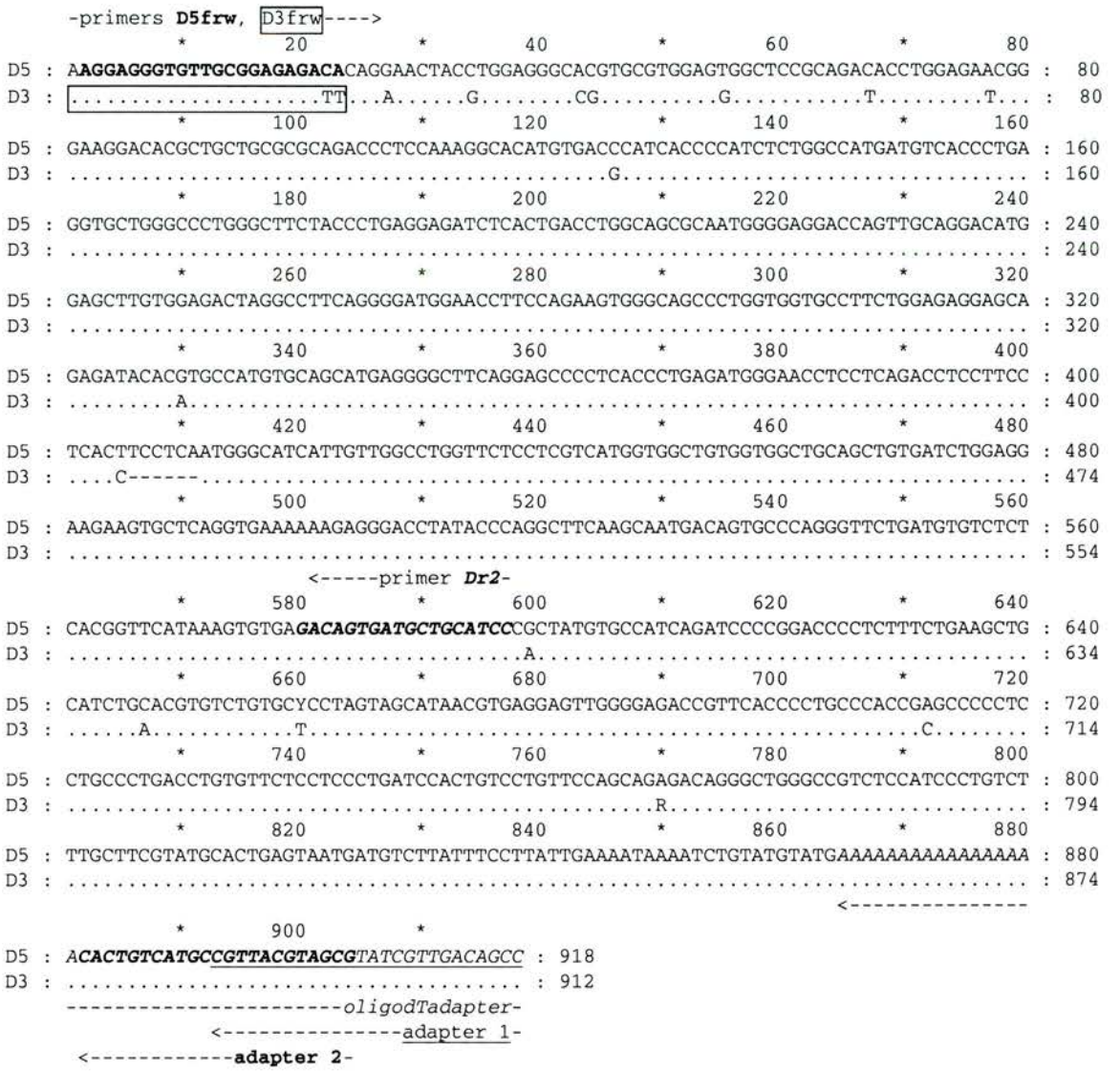


Figure 5.11: Alignment of D5 and D3 sequences obtained from 3'-RACE PCR products. Forward primers D5 and D3 are in *bold* and *boxed*, respectively, while reverse primers oligoadapter, adapter 1 and adapter 2 are in *italics*, *underlined* and in *bold*, respectively. Identities and gaps are indicated by *dots* (.) and *dashes* (-), respectively.

5.3.6 Sequencing of clones identified by PCR-RFLP

Full length sequence data were obtained from at least three clones representing each of the D5, D3, C7, C11 and C6 transcripts from ram 501 and for F8, F12, H10, F9 and G6 transcripts from ram 504 (Table 5.4, Fig. 5.12), Initially, PCR-RFLP analysis of clones containing full length inserts amplified using the forward primers Afl, Bfl and Cfl and the reverse primer Dr1 identified a number of clones as D5, D3, F8 or F12. However, most of these clones contained hybrid sequences with C7 or H10 (Table 5.4). In contrast, PCR products amplified using the primer pair Cfl-Br were either D5 or D3 for 501 and F8 or F12 for 504 (Table 5.4). On the other hand, the primer pair Br-Dr1 did not amplify any of these sequences (Table 5.4). This finding suggested that the reverse primer Dr1 was not efficient in amplifying full length sequences of D5, D3, F8 and F12. Moreover, sequencing data from single clones of F8 and F12 suggested that due to a 19bp deletion present in these alleles, the Dr1 primer had a 3bp mismatch with these sequences in its 3'-end. This finding was also confirmed by 3'-RACE for sequences D5 and D3 (section 5.3.5). For these reasons, the Dr2 primer was designed and full length PCR products obtained with the primer pair Cfl-Dr2 were cloned and sequenced. All these PCR products were either D5 or D3 for ram 501 and F8 or F12 for ram 504 (Table 5.4). Two novel sequences (N1 and N2) were also obtained that had not been previously identified by SSCP and sequence analyses in chapter 3.

The aligned nucleotide and predicted amino acid sequences are shown in figures 5.12 and 5.13.

501	D5		D3		C7		E2		C11		C6	
primers	RFLP	Seq	RFLP	Seq	RFLP	Seq	RFLP	Seq	RFLP	Seq	RFLP	Seq
Afl-Dr1	0		0		57	0[3]	0	0	3	3	0	0
Bfl-Dr1	0		0		53	10	0	0	3	3	4	4
Cfl-Dr1	4	0 D5	7	0 D3	49	0[3]	0	0	0	0	0	0
		2 dif[4]		7 hyb[2]								
		2 hyb[2]										
Cfl-Br	26	5	34	3	0	0	0	0	0	0	0	0
Br-Dr1[4]		0		0		9		1		1		1
Cfl-Dr2	23	4	17	4	0	0	0	0	0	0	0	0

504	F8		F12		H10		F9		F7		G6		G12		E9	
Primers	RFLP	Seq	RFLP	Seq	RFLP	Seq	RFLP	Seq	RFLP	Seq	RFLP	Seq	RFLP	Seq	RFLP	Seq
Afl-Dr1	0	0	0	0	60	3	0	0	0	0	0	0	0	0	0	0
Bfl-Dr1	0	0	0	0	44	8	12	10	0	0	4	4	0	0	0	0
Cfl-Dr1	1	1	1	1	56	1[3]	0	0	0	0	0	0	2	new[5]	0	0
Cfl-Br	22	4	30	5	6	0[3]	2	0[3]	0	0	0	0	0	0	0	0
Br-Dr1[4]		0		0		0		2 new[5]		0		8		0		0
Cfl-Dr2	21	4	19	4	0	0	0	0	0	0	0	0	0	0	0	0

[1]Two non ovine MHC class I genes
[2] Hybrids: recombinant sequences made of two alleles
[3] Only few or none were sequenced because we already had enough identical sequences obtained by other primers
[4] PCR-RFLP was not conducted
[5]Two new sequences were identified for 504 (2 full length clones and 4 Br-Dr clones)

Table 5.4: Number of clones identified by PCR-RFLP in relation to each primer pair used and the number of clones sequenced representing each pattern.

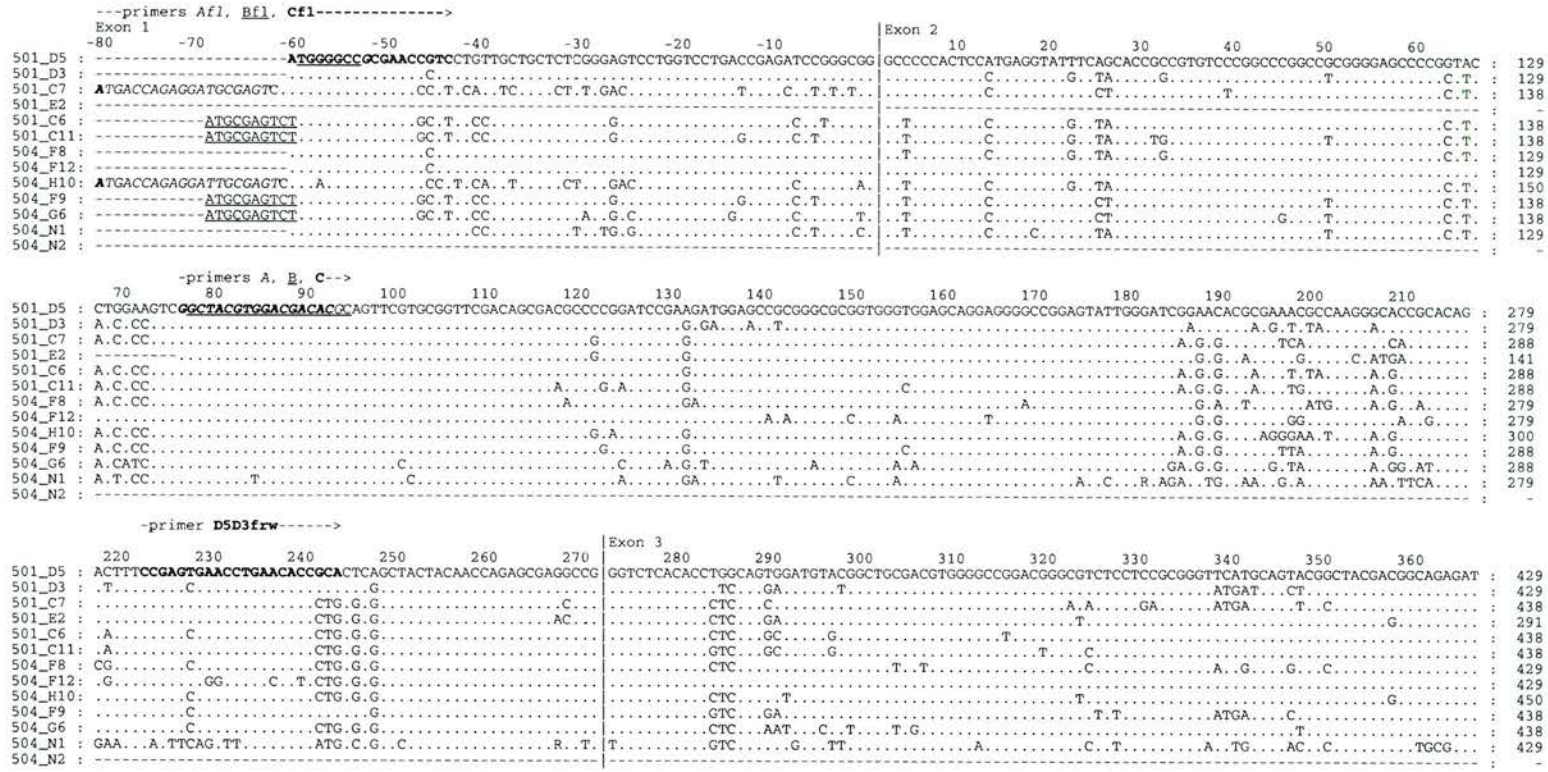


Figure 5.12 continued on next page

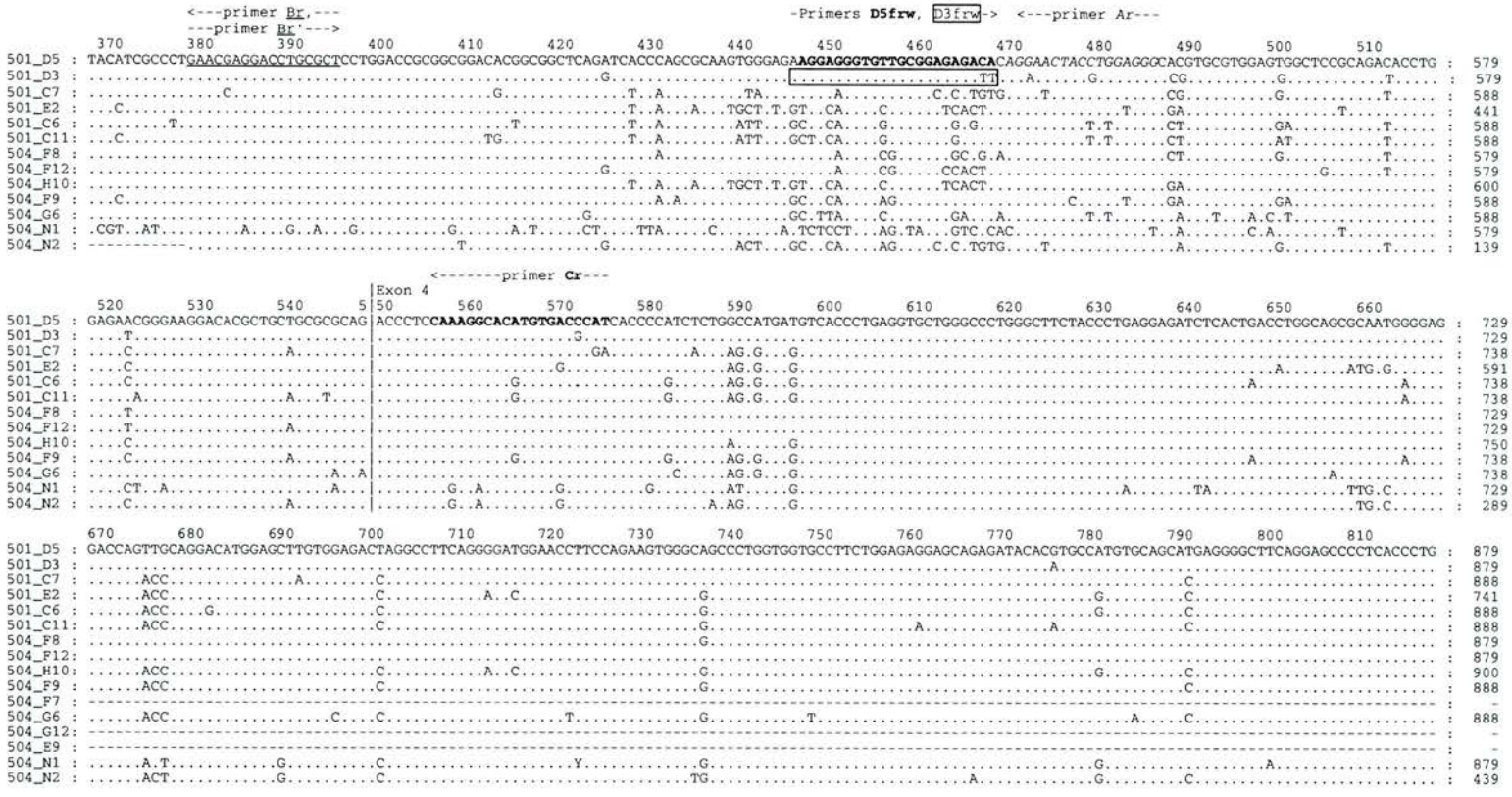


Figure 5.12 continued on next page

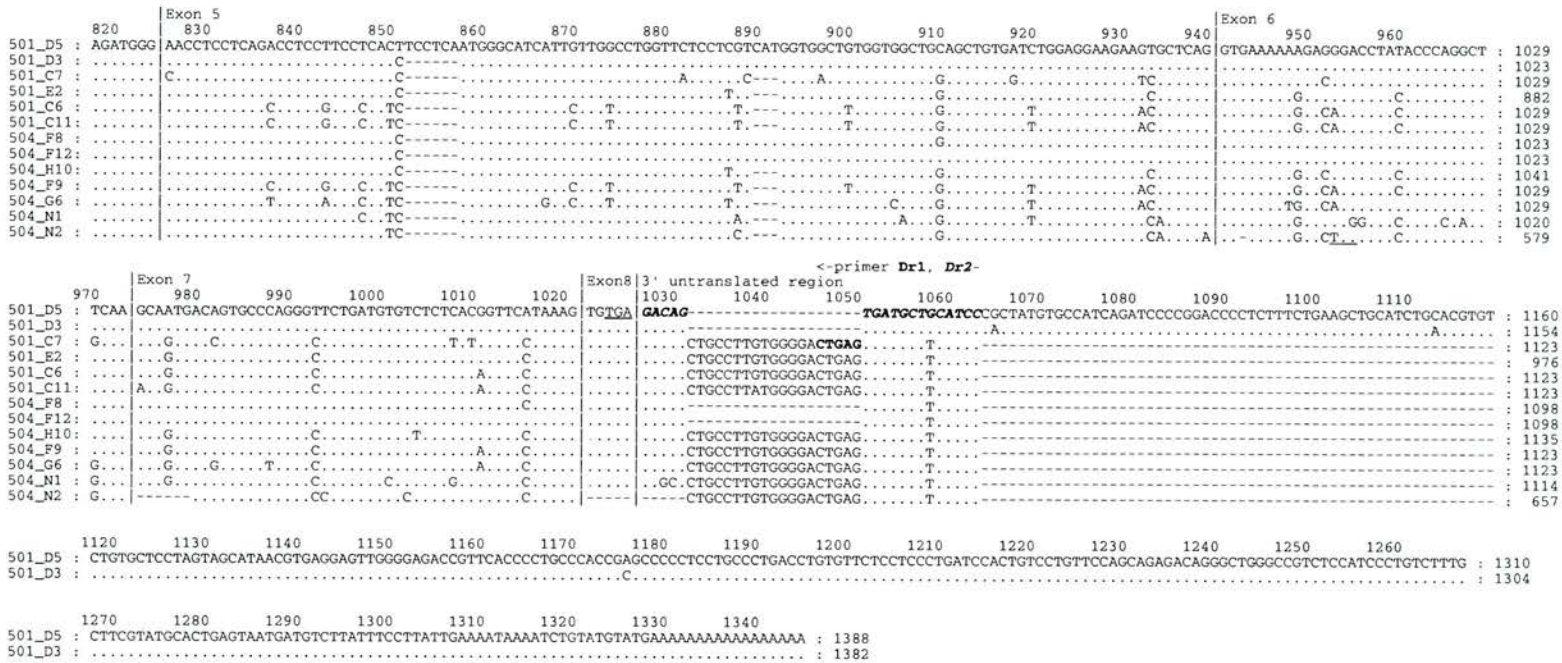


Figure 5.12: Alignment of nucleotide sequences obtained for animals 501 and 504. Names of primers and their position are highlighted. *Vertical lines* indicate exon boundaries. Identities and gaps are indicated by *dots* (.) and *dashes* (-), respectively.

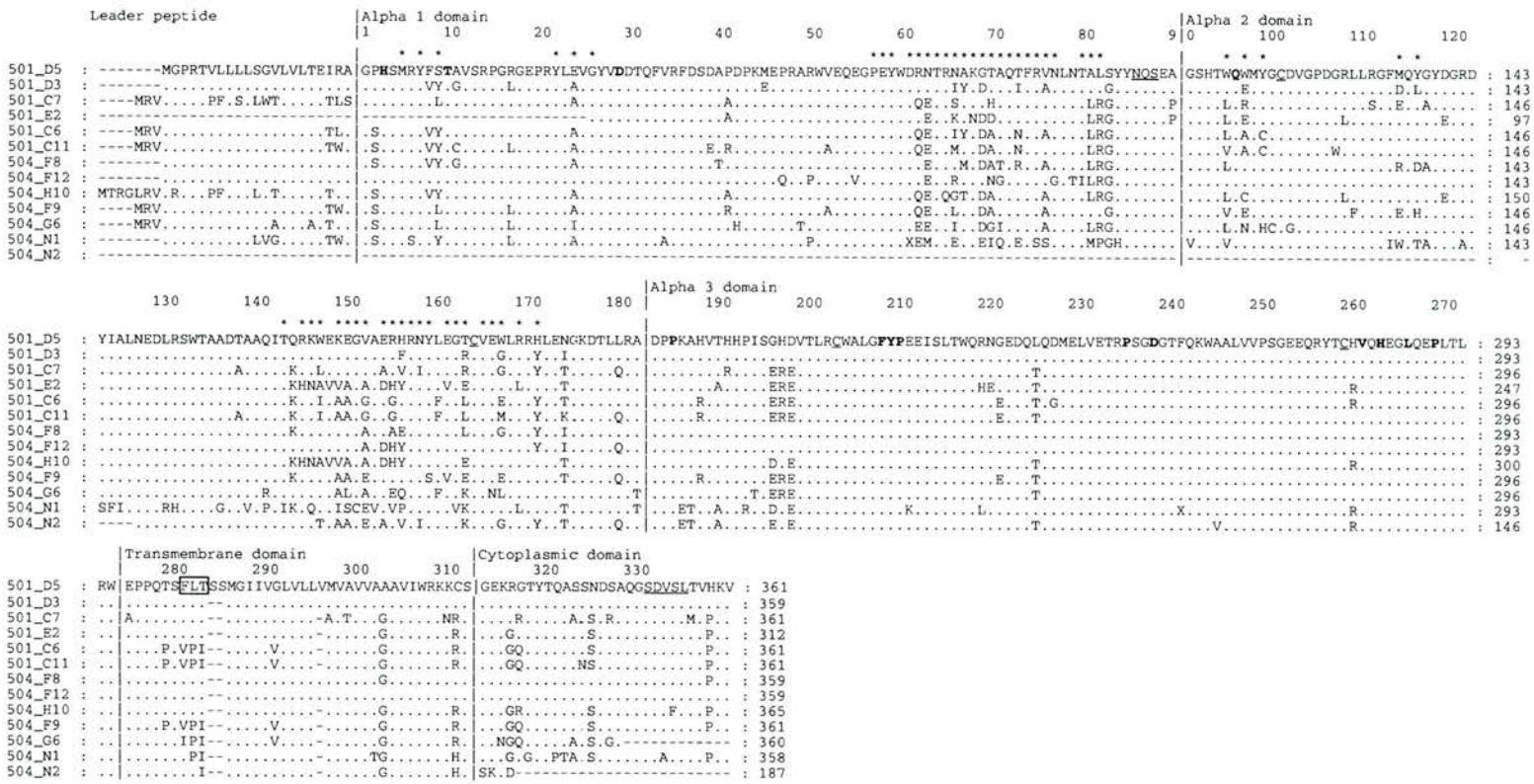


Figure 5.13: Alignment of the predicted amino acid sequences of the class I transcripts obtained for animals 501 and 504. *Vertical lines* (|) indicate domain boundaries. Identities and gaps are indicated by *dots* (.) and *dashes* (-), respectively. *Asterisks* (*) indicate MHC class I residues predicted to bind antigenic peptides or interact with T cell receptors in humans (Bjorkman et al. 1987a). A conserved site for N-linked glycosylation at positions 86-88, a conserved site for phosphorylation at positions 330-334 and sites for intradomain disulfide bond formation (cysteines 101, 164, 203 and 259) are all *underlined*. Other conserved residues among vertebrates are in *bold*. A classical MHC class I identification motif is *boxed*.

	D5	D3	C7	E2	C6	C11	F8	F12	H10	F9	G6	N1	N2
D5		0.084	0.198	0.157	0.181	0.209	0.106	0.069	0.171	0.178	0.209	0.298	0.176
D3	0.051		0.198	0.184	0.161	0.195	0.075	0.118	0.174	0.168	0.205	0.306	0.164
C7	0.108	0.109		0.172	0.198	0.219	0.178	0.219	0.168	0.202	0.244	0.345	0.181
E2	0.087	0.110	0.091		0.153	0.176	0.157	0.172	0.063	0.161	0.211	0.293	0.176
C6	0.101	0.097	0.102	0.080		0.061	0.135	0.202	0.132	0.094	0.151	0.298	0.193
C11	0.112	0.110	0.114	0.089	0.035		0.171	0.219	0.178	0.088	0.174	0.314	0.199
F8	0.059	0.057	0.100	0.089	0.084	0.097		0.115	0.138	0.165	0.191	0.279	0.164
F12	0.034	0.071	0.111	0.088	0.110	0.119	0.063		0.185	0.215	0.229	0.321	0.164
H10	0.094	0.103	0.086	0.035	0.074	0.093	0.081	0.095		0.171	0.205	0.291	0.164
F9	0.096	0.094	0.102	0.086	0.050	0.052	0.095	0.113	0.089		0.161	0.291	0.181
G6	0.122	0.131	0.128	0.108	0.087	0.095	0.114	0.128	0.103	0.089		0.306	0.217
N1	0.179	0.186	0.200	0.173	0.174	0.180	0.165	0.185	0.170	0.169	0.187		0.231
N2	0.101	0.101	0.088	0.081	0.092	0.103	0.090	0.094	0.079	0.088	0.122	0.127	

Table 5.5: Estimates of genetic distances between sequences calculated by the Kimura 2 parameter model, incorporating the Poisson correction. Nucleotide distances are presented below the diagonal, whereas amino acid distances are shown above the diagonal. The lowest and the largest distance are highlighted in *green* and *yellow*, respectively.

All sequences presented in figure 5.12 had not been reported in databases before the onset of the current study, as indicated by both an alignment with the previously published ovine class I sequences and a blast search. Another alignment of the sequences with published human, cattle, dog, murine and ovine class I sequences suggested that they consist of 8 exons that transcribe the leader peptide (LP) and 5 protein domains. The length of the LP was uncertain because the forward primers used to amplify those sequences annealed at the beginning of the LP. As with other species, α1, α2 and α3 domains were 90, 92 and 92 amino acids long, respectively. The transmembrane (TM) domain varied in length between the different sequences, with most being 35 residues. The D5 sequence had the longest TM domain at 38 residues, while those of D3, F8 and F12 sequences were intermediate in length at 36 residues (Fig. 5.13). With the exception of sequences G6 and N2, which were truncated by in-frame stop codons, the cytoplasmic domains were 28 amino acids long (Fig. 5.13).

Three hundred and twenty six of the 1086 nucleotides (30%) of the full length coding region were diverse (Fig. 5.12). This translates into variation at 151 of the 361 amino acids (41.8%) (Fig. 5.13). Estimates of genetic distances between sequences ranged between 3.5% and 20% at the nucleotide level (Table 5.5) and 6.3% to 34.5% at the amino acid level (Table 5.5), with sequences N1 followed by G6 exhibiting the greatest divergence from the other sequences, especially within exons 2 and 3 (Table 5.5, Fig. 5.12 and 5.13).

5.4 Discussion

In this chapter, 11 full length and 2 partial MHC class I transcripts were cloned and sequenced from the MHC heterozygous rams 501 and 504. The entire coding region was obtained from all the highly represented transcripts identified by SSCP in chapter 3 (D5, D3, C7, F8, F12, H10, F9), with the exception of E2, which is missing the first 150bp. Full length sequence was also obtained for the C11, C6 and G6 transcripts, which were initially identified in only small numbers of clones by SSCP. Using the methods described here, full length sequence was not obtained from the F7, G12 and E9 transcripts initially identified in less than 5 clones analysed by SSCP or for the C7b transcript identified in chapter 4. However, two previously unidentified transcripts, N1 and N2, were obtained, raising the number of distinct class I transcripts identified from the two rams to 17. A search of the nucleotide databases revealed that all 17 transcripts were novel.

The genetic distances between the transcripts and their predicted amino acid sequences was determined and shown to range from 0.035 to 0.20 at the nucleotide and 0.063 to 0.345 at the amino acid level (Table 5.5). This reflects the range of classical and non classical class I loci represented. A similar observation of high numbers of class I loci has also been made for the horse MHC and authors suggested that there are four putative non classical horse MHC class I loci (Ellis et al. 1995; Holmes et al. 1999).

Failure to obtain sequence data for the entire coding region of the genes F7, G12, E9 and C7b is likely to reflect the limited abundance of these transcripts. This is supported by the findings presented in chapter 3, where no more than 3 clones were obtained for F7, G12, E9 and the C7b transcript was not identified in the 180 clones analysed. Alternatively, failure to amplify full length sequences may result from unresolved diversity in the primer annealing sites or from a variety of factors involved in a multiplex PCR reaction, including competition between different amplification products for primers and reagents during amplification. Ideally, RACE technology should be applied to provide sequence data for the 3' and 5' ends of all the class I genes described here. This would provide sequence covering of all primer binding sites and could therefore clarify whether some transcripts were not identified in chapter 3 due to primer mismatches.

The majority of the full length class I transcripts described here (Fig. 5.13) have a typical MHC class I exon structure, with nothing to suggest that they are not capable of translating functional class I molecules. By comparison with sequences available from human

(Geraghty et al. 1987; Malissen et al. 1982; http://sdmc.krdl.org.sg:8080/fimm/), murine (Evans et al. 1982; Steinmetz et al. 1981a) and bovine (Bensaid et al. 1991; Brown et al. 1989; http://www.projects.roslin.ac.uk/bola/) class I genes, each ovine class I transcript appears to be composed of eight exons that translate the leader, α1, α2 and α3 extracellular domains, a transmembrane and a cytoplasmic domain. As expected, most of the nucleotide diversity is situated in the α1 and α2 domains encoding amino acids that bind antigenic peptides or interact with the T cell receptor (Bjorkman et al. 1987a; Fig. 5.13). All the sequences include features important for functional antigen presenting molecules, including a conserved site for N-linked glycosylation at positions 86-88 and cysteine sites for intradomain disulfide bond formation at positions 101, 164, 203 and 259 (Bjorkman et al. 1987a). With the exception of N1, N2 and G6, all sequences contain a conserved site for phosphorylation at positions 330-334 (Garber et al. 1994; underlined in Fig. 5.13). These sequences also include conserved residues found in all class I molecules from mammals, birds, reptiles, amphibians and fish (Grossberger and Parham, 1992; in bold in Fig. 5.13) which are thought to be essential for maintaining the architecture of antigen presenting molecules.

In contrast, the N1, N2 and G6 transcripts include features associated with non classical class I molecules. In particular, sequences G6 and N2 each contain an early stop codon situated in the region encoding the cytoplasmic domain. This results in truncated proteins deprived of the conserved phosphorylation site at positions 330-334 (Guild and Strominger, 1984). The sheep class I sequence with accession number U03094 contains an identical cytoplasmic domain to G6 and differs from G6 by only 3bp, suggesting that these sequences may be alleles of the same locus. Furthermore, sequence G6 was identified in only 2 out of 180 clones analyzed by SSCP, suggesting that it is transcribed at low levels in PBMC. A truncated cytoplasmic domain and low levels of transcription are features of non classical class I genes (Braud et al. 1999; Lew et al. 1987; Ulker et al. 1990), suggesting that G6 and N2 are candidates for non classical loci.

Similarly, the N1 transcript, which was also not identified during the SSCP analysis described in chapter 3 is a candidate for a non classical locus. Since the entire coding region of N1 was obtained, it is clear that polymorphism within the primer binding sites was not the reason for the earlier failure to identify this transcript (Fig. 5.12). It is likely that N1 is transcribed at very low levels in peripheral blood cells and for that reason it was not detected in chapter 3. Furthermore, sequence N1 is the most divergent sequence (Table 5.5) and differs in areas otherwise conserved in all other ovine MHC class I sequences (Fig. 5.12, 5.13). N1 may represent a non classical class I like gene that has accumulated a large number of mutations associated with a novel function (Hugh and Nei, 1989). Indeed, two putatively non classical

MHC class I genes identified in cattle were found to be relatively more divergent than classical bovine MHC class I genes (Holmes et al. 1999; 2003).

Additional evidence for the non classical nature of sequences N1 and G6 comes from nucleotide composition characteristics. The lowest third position G+C content of exons 2 and 3 is observed in sequence N1 (78.3%) and the second lowest is that of sequence G6 (85.6%), whilst the corresponding content for all other sequences described here is over 86.2%. In almost every case, as genes become non classical or pseudogenes, the third position G+C content in exons 2 and 3 tends to be reduced in comparison with that in classical genes and that reduction is greater in genes that are highly divergent from classical genes (Hugh and Nei, 1989). Therefore, N1 may represent an evolutionary ancient non-classical gene, whereas G6 may represent a non-classical gene that evolved from duplication of a classical gene more recently.

It has been suggested that exons 4 to 8 contain locus defining motifs in other species (Cadavid et al. 1997; Garber at al. 1994, Holmes and Ellis, 1999; Parham et al. 1989). The class I sequences presented in this chapter could be divided into 3 distinct groups based on the length of the transmembrane domains suggesting that they represent alleles at a minimum of three distinct loci if there no genetic interlocus recombination. The differences in length between ovine MHC molecules may have a biological significance in terms of function. When the haplotype data from chapter 4 and phylogenetic analysis in chapter 6 are combined, a more accurate picture of the number of ovine MHC class I loci, and the extent of allelic diversity and haplotype variability will emerge.

In summary, 9 full length ovine MHC class I sequences encoding molecules with characteristics associated with functional antigen presenting proteins and 2 full length and one partial class I sequence with characteristics consistent with non classical class I loci or pseudogenes were obtained through the work described in this chapter. Additional evidence regarding the classical or non-classical nature of these genes could be obtained through a study of their polymorphism, cellular distribution, cellular location and evolution.

Chapter 6. Phylogenetic analysis of ovine MHC class I sequences

6.1 Introduction

The previous chapters describe the identification of 11 full-length and 6 partial MHC class I transcripts from two rams and the assignment of those transcripts to four haplotypes. In chapter 5, some putative locus specific characteristics were identified in exons 4 to 8, which encode the more conserved α3, transmembrane and cytoplasmic domains. However, assignment of the transcripts to loci could not be accomplished by sequencing alone.

In some species, MHC class I sequences group into discrete, well supported groups that correspond to loci, whereas in others such correlation is more problematic. Class I sequences from human, pig, dog and non-human primates such as gorilla and chimpanzee contain locus-specific features and can therefore be assigned to certain loci merely by sequencing (IPD MHC database; http://www.ebi.ac.uk/ipd/mhc/dla/index.html). In contrast, identification of distinct equine class I loci was initially problematic and required phylogenetic analysis of class I sequences not only from horse, but also from related species (Holmes and Ellis, 1999). In addition, similar analysis of bovine class I sequences with closely and distantly related species suggested that there are at least from four to six classical class I loci (Holmes et al. 2003). However, although phylogenetic groups were identified in this analysis, bovine MHC class I loci could not be clearly defined.

The number of class I loci in the ovine MHC complex is also not clear. Moreover, the limited sequence data available has not allowed a phylogenetic analysis of ovine MHC class I sequences. The work described in this chapter aimed to construct phylogenetic trees using the identified transcripts as well as ovine and other ruminant class I sequences available in the literature. The analysis was expected to facilitate determination of the number of ovine class I loci, examine the phylogenetic relationships among class I sequences within and between species and provide more insight into the classical or non-classical nature of the transcribed sequences from the rams.

6.2 Materials and Methods

6.2.1 Sequences

Phylogenetic analysis of ovine class I sequences was performed with eleven full-length and two partial sequences identified in the previous chapters, ten published sequences obtained from GenBank and four unpublished full-length sequences (Blacklaws B., University of Cambridge, Cambridge, UK; Table 6.1). The partial ovine class I sequences E9, C7b, F7 and G12 were not included in the analysis, because they are restricted to the highly variable exons 2 and 3 (chapters 3 and 4).

Published class I sequences from the Bovidae species *Bos taurus* (domestic cattle), *Bos indicus* (African Boran cattle), *Syncerus caffer nanus* (dwarf forest African buffalo), *Bison bonasus* (bison), *Bubalus bubalis* (water buffalo), *Bos gaurus* (gaur), *Bos grunniens* (yak), *Tragelaphus speckii* (sitatunga), *Tragelaphus strepsiceros* (greater kudu), *Tragelaphus angasii* (nyala) and *Sus scrofa* (domestic pig) were also used for phylogenetic analysis (Table 6.1).

6.2.2 Phylogenetic analysis

Nucleotide sequences were aligned using the MegAlign program (DNASTAR, Madison WI), as described in section 2.8.3. Neighbour-Joining phylogenetic trees (Saitou and Nei, 1987; section 2.8.5) were constructed, using either genetic distances of the entire coding regions calculated by the Kimura two parameter method (Kimura M. 1980) or p distances of exons 4 through 8 (section 2.8.4). The swine sequences AF014001, AF014002 and AF014006 were used as an outgroup to root the trees. Phylogenetic analysis was performed using MEGA2.1 software (Kumar et al. 2001).

Species	Symbol of species on trees	Sequence name on trees	Exons	Accession No
Ovis aries	OLA	501A_D5	1-8 or 4-8	AJ874673
Ovis aries	OLA	501B_D3	1-8 or 4-8	AJ874674
Ovis aries	OLA	501A_C7	1-8 or 4-8	AJ874675
Ovis aries	OLA	501B_E2	2-8 or 4-8	AJ874676
Ovis aries	OLA	501B_C6	1-8 or 4-8	AJ874677
Ovis aries	OLA	501A_C11	1-8 or 4-8	AJ874678
Ovis aries	OLA	504C_F8	1-8 or 4-8	AJ874679
Ovis aries	OLA	504D_F12	1-8 or 4-8	AJ874680
Ovis aries	OLA	504D_H10	1-8 or 4-8	AJ874681
Ovis aries	OLA	504D_F9	1-8 or 4-8	AJ874682
Ovis aries	OLA	504C_G6	1-8 or 4-8	AJ874683
Ovis aries	OLA	N1	1-8 or 4-8	unpublished
Ovis aries	OLA	N2	4-8	unpublished
Ovis aries	OLA	M34673	4-8	M34673
Ovis aries	OLA	M34674	4-8	M34674
Ovis aries	OLA	M34675	4-8	M34674
Ovis aries	OLA	M34676	1-8 or 4-8	M34676
Ovis aries	OLA	U03092	1-8 or 4-8	U03092

Ovis aries	OLA	U03093	1-8 or 4-8	U03093
Ovis aries	OLA	U03094	1-8 or 4-8	U03094
Ovis aries	OLA	AY188824	4-8	AY188824
Ovis aries	OLA	AY188825	4-8	AY188825
Ovis aries	OLA	AY188826	4-8	AY188826
Ovis aries	OLA	Cambridge 1	1-8 or 4-8	unpublished
Ovis aries	OLA	Cambridge 2	1-8 or 4-8	unpublished
Ovis aries	OLA	Cambridge 3	1-8 or 4-8	unpublished
Ovis aries	OLA	Cambridge 4	1-8 or 4-8	unpublished
Bos taurus	BoLA	3349.1	4-8	AJ010862
Bos taurus	BoLA	4221.1	4-8	AJ010865
Bos taurus	BoLA	A10	4-8	M69206
Bos taurus	BoLA	BL3-6	4-8	M21044
Bos taurus	BoLA	BL3-7	4-8	M21043
Bos taurus	BoLA	BSA	4-8	L02832
Bos taurus	BoLA	BSC	4-8	L02833
Bos taurus	BoLA	BSF	4-8	L02834
Bos taurus	BoLA	BSN	4-8	L02835
Bos taurus	BoLA	BSX	4-8	U01187

Bos taurus	BoLA	D18.1	4-8	Y09205
Bos taurus	BoLA	D18.2	4-8	Y09206
Bos taurus	BoLA	D18.3	4-8	Y09207
Bos taurus	BoLA	D18.4	4-8	Y09208
Bos taurus	BoLA	D18.5	4-8	AJ010867
Bos taurus	BoLA	HD1	4-8	X80933
Bos taurus	BoLA	HD6	4-8	X80934
Bos taurus	BoLA	HD7	4-8	X80935
Bos taurus	BoLA	HD15	4-8	X80936
Bos taurus	BoLA	JSP.1	4-8	X92870
Bos taurus	BoLA	KN104	4-8	M69204
Bos taurus	BoLA	MAN1	4-8	AJ010863
Bos taurus	BoLA	MAN2	4-8	AJ010861
Bos taurus	BoLA	MAN3	4-8	AJ010864
Bos taurus	BoLA	MAN8	4-8	AJ010866
Bos taurus	BoLA	pBoLA1	4-8	M24090
Bos taurus	BoLA	pBoLA4	4-8	X97647, X97648, X97649
Bos indicus	Boran	M69204	4-8	M69204

Bos indicus	Boran	M69206	4-8	M69206
Bos indicus	Boran	AY188799	4-8	AY188799
Bos indicus	Boran	AY188801	4-8	AY188801
Bos indicus	Boran	AY188802	4-8	AY188802
Bos indicus	Boran	AY188803	4-8	AY188803
Bos indicus	Boran	AY188804	4-8	AY188804
Bos indicus	Boran	AY188805	4-8	AY188805
Bos indicus	Boran	AY188806	4-8	AY188806
Syncerus caffer nanus	Fbuff	AY188810	4-8	AY188810
Syncerus caffer nanus	Fbuff	AY188811	4-8	AY188811
Bison bonasus	Bison	AY188808	4-8	AY188808
Bison bonasus	Bison	AY188809	4-8	AY188809
Bubalus bubalis	Wbuff	AY188812	4-8	AY188812
Bubalus bubalis	Wbuff	AY188813	4-8	AY188813
Bos gaurus	Gaur	AY188814	4-8	AY188814
Bos gaurus	Gaur	AY188815	4-8	AY188815
Bos grunniens	Yak	AY188827	4-8	AY188827
Bos grunniens	Yak	AY188828	4-8	AY188828
Tragelaphus speckii	Sit	AY188816	4-8	AY188816

Tragelaphus speckii	Sit	AY188817	4-8	AY188817
Tragelaphus speckii	Sit	AY188818	4-8	AY188818
Tragelaphus speckii	Sit	AY188819	4-8	AY188819
Tragelaphus speckii	Sit	AY188820	4-8	AY188820
Tragelaphus speckii	Sit	AY188821	4-8	AY188821
Tragelaphus speckii	Sit	AY188822	4-8	AY188822
Tragelaphus strepsiceros	Kudu	AY188823	4-8	AY188823
Tragelaphus angasii	Nyala	AY188829	4-8	AY188829
Tragelaphus angasii	Nyala	AY188830	4-8	AY188830
Tragelaphus angasii	Nyala	AY188831	4-8	AY188831
Sus scrofus	SLA	AF014001	4-8	AF014001
Sus scrofus	SLA	AF014002	4-8	AF014002
Sus scrofus	SLA	AF014006	4-8	AF014006

Table 6.1: Sequences used for phylogenetic analysis. Names of sequences identified in this study are presented as haplotype ram name (chapter 4), where available, followed by sequence name or as published names for bovine sequences, or as accession numbers for the remaining sequences.

6.3 Results

6.3.1 Sequence alignments

The entire coding regions or exons 4 through 8 of transcripts identified in this study, along with previously published and four unpublished ovine MHC class I sequences (Table 6.1) were aligned prior to construction of phylogenetic trees (Appendix 2.1). In addition, an alignment of exons 4 through 8 of all available ovine MHC class I sequences with published bovine or other bovid sequences reported in GenBank was generated and is presented in Appendix 2.1. An alignment of the predicted amino acid sequences of available ovine MHC class I sequences is presented in figure 6.1.

From a phylogenetic prespective, an interesting feature of the aligned amino acid sequences was the observed variation in length of the transmembrane (TM) domain. Sequences were differentiated into four groups on this basis, with the majority encoding 35 residues (red box, Fig. 6.1). Sequences D5, M34676, Cambridge 1, 2 and 4 had the longest TM domains at 38 residues (blue box, Fig. 6.1), while the D3, F8, F12 and Cambridge 3 sequences were intermediate in length at 36 residues (green box, Fig. 6.1). With the exception of AY188825, sequences with TM 35 residues in length were aligned with the group of sequences with the longest TM by including gaps at position 284-285 and 295 (Fig. 6.1), whereas sequence AY188825 incorporates a gap at position 307 instead of 295. Sequence U03094, which encodes a TM region with 38 amino acids, contains a duplication (VVAVVA) at positions 301-306 that could be attributed to a sequencing error, as it is present in no other ovine or bovid class I sequence. For this reason, U03094 was also grouped with the sequences encoding TM regions of 35 residues (red box, Fig. 6.1).

Variation in sequence length was also observed due to in-frame stop codons in the cytoplasmic domain (CT) of G6, U03093 and N2 (yellow boxes, Fig. 6.1). These sequences encode truncated CT domains with 12, 12 and 24 fewer residues respectively than the CT region of the remaining sequences, which were 28 amino acids long.

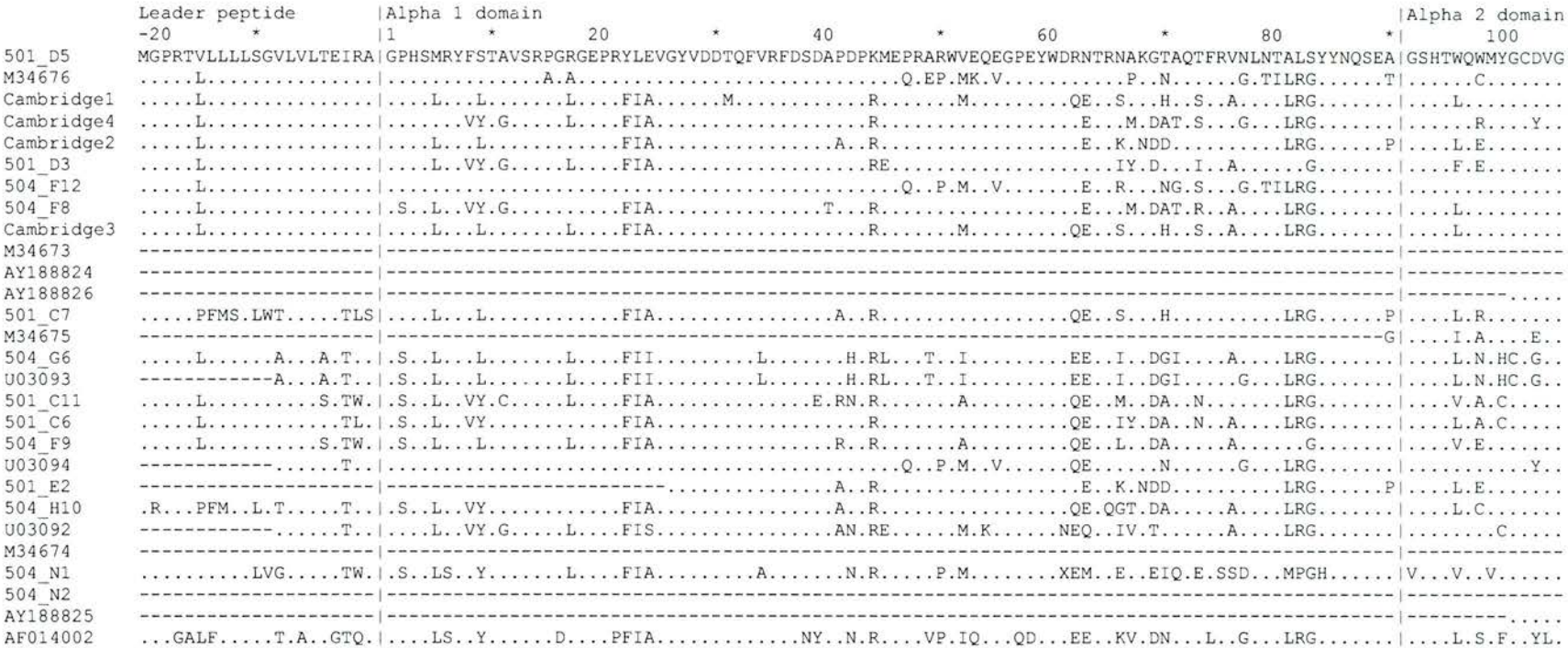


Figure 6.1 continued in next page

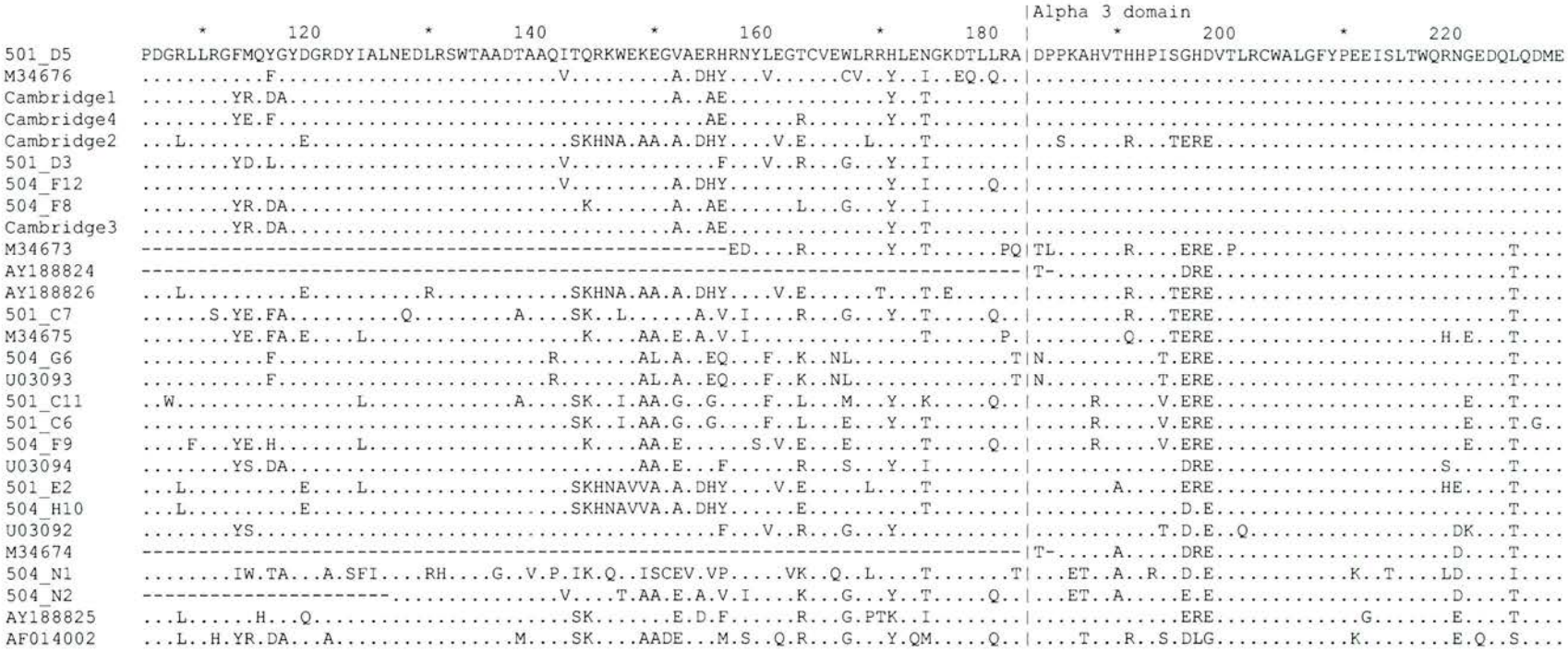


Figure 6.1 continued in next page

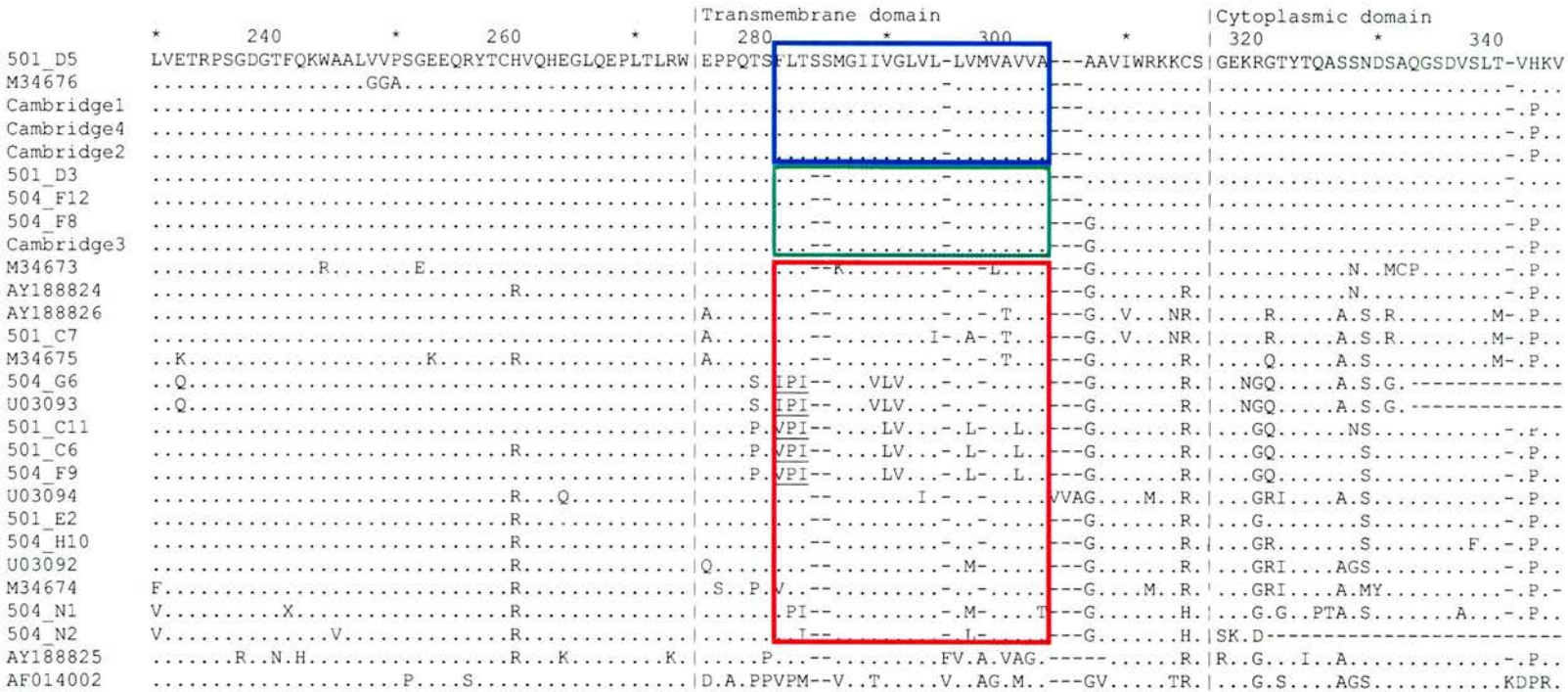


Figure 6.1: Alignment of the predicted amino acid sequences of transcripts identified in this study along with previously published and four unpublished ovine MHC class I sequences. *Vertical lines* (|) indicate domain boundaries. *Identities* and *gaps* are indicated by dots (.) and dashes (-), respectively. Sequences with transmembrane domains at 38, 36 and 35 residues are boxed in blue, green and red, respectively. Truncated parts of cytoplasmic domains due to in frame stop codons are boxed in yellow. The putative non classical discrimination motifs VPI and IPI are underlined.

6.3.2 Phylogenetic analysis

The relationships between the ovine class I sequences were visualized by constructing Neighbour-Joining (NJ) phylogenetic trees using, where available, the entire coding regions (Fig. 6.2), or exons 4 through 8 (Fig. 6.3), encoding the α3, transmembrane and cytoplasmic domains. Although swine sequences were expected to root these trees, sequence N1 appears the most divergent overall in the tree presented in Fig. 6.2. In contrast, when the influence of positive selection on tree topology was excluded by confining the analysis to exons 4 to 8, the swine class I sequence appears as the clear outgroup rooting the tree, as shown in Fig. 6.3. For this reason, phylogenetic trees of available ovine and published bovine (Fig. 6.4) or of all available bovid MHC class I sequences (Fig. 6.5) were constructed using only exons 4 though 8.

Although firm phylogenetic groups that clearly correspond to different ovine class I loci were not defined, some clusters of closely related sequences that could correspond to one or two related loci were apparent. In particular, with the exception of N1, the full length sequences grouped into four distinct clusters (Fig. 6.2), while phylogenetic analysis using exons 4 through 8 distinguished the sequences into five clusters (Fig. 6.3). Clusters I and IV of the phylogenetic tree presented in Fig. 6.2 were similar to those shown in Fig. 6.3. With the exception of the Cambridge 2 sequence, clusters I and IV contained exactly the same sequences in both trees, thus strengthening their designation as putative distinct loci. The Cambridge 2 sequence located to cluster II in figure 6.2 and segregated with the sequences of cluster I in figure 6.3, suggesting that its exons 4 to 8 were more similar to the remaining Cambridge sequences.

Restriction of the analysis to exons 4 through 8 (Fig. 6.3) rearranged the topology of sequences in clusters II and III observed in Fig. 6.2. C7 segregated with the partial sequences M34675 and AY188826 in cluster IIi, whereas E2 and H10 segregated with U03094 in cluster IIii, suggesting that these sequences may represent at least two additional loci (Fig. 6.3). Sequences M34674, U03092, N1 and N2 grouped in cluster V in Figure 6.3, while the partial class I sequences M34673 and AY188824 also segregated together, but distinct from clusters I, IIi, IIii, IV and V, suggesting that these may represent alleles at a sixth locus. The sequence AY188825 appeared relatively deep on the phylogenetic tree and was quite divergent from the other sequences. It may therefore represent a non-classical locus. However, no firm assertions can be made for this sequence, or for clusters IIii, III and V, as the confidence values attached to the corresponding nodes were low.

Further evidence regarding the number of ovine class I loci was obtained by constructing phylogenetic trees of class I sequences from related species. In particular, trees

were constructed using exons 4 through 8 of ovine and bovine (Fig. 6.4) or of all available bovid MHC class I sequences (Fig. 6.5). The identity and topology of sequences grouped with high confidence values in clusters I, IIi and IV in the tree presented in figure 6.3 were almost identical to those appearing in figures 6.4 and 6.5, further supporting their designation as distinct loci. The E2, H10 and U03094 sequences also consistently segregated together in cluster IIii, suggesting that they may represent sequences at a fourth locus. In addition, although the partial sequences M34673 and AY188824 (cluster III; Fig.6.3) appeared between clusters I and IIii in all trees, their relatively long branches in figure 6.3 and their topology as independent sequences in figures 6.4 and 6.5 suggest that these sequences are relatively divergent from the previously described loci and may therefore represent alleles of at least one additional locus. Moreover, sequences N1 and N2 were consistently located deep in the trees (cluster V; Fig. 6.4 and 6.5) together with the non classical bovine class I sequence HD15, and may therefore represent alleles of one or two distant non-classical loci. Similarly, M34674, U03092 and AY188825 show no clear relationship to any ovine cluster and may thus represent distant class I like genes.

The relationships of ovine with bovine or with other bovid MHC class I sequences can be visualized in the trees presented in figures 6.4 and 6.5. Bovine and other bovid class I sequences segregated in three distinct large groups of sequences designated as groups a-c. The relative position of these groups to the clusters containing ovine class I sequences was uncertain, as the topology for the groups was not supported by high confidence values. The identity and topology of bovine class I sequences within each group, however, was almost identical in both trees (Fig. 6.4 and 6.5). Some bovine class I sequences were closely related to those of other bovid species, for example boran cattle, and bison and nyala (asterisks (*), Fig. 6.5), suggesting that these sequences may represent alleles from orthologous loci. In contrast, with the exception of N1 and N2, ovine class I sequences did not segregate with those of other bovid species (Fig. 6.5). Sequences N1 and N2 segregated with HD15 and AY18829, the putative non-classical genes of cattle and nyala (Holmes et al. 2003), whereas the remaining ovine class I clusters contained MHC sequences exclusively from sheep. However, cluster IIii and independent sequences M34674, U03092 and AY188825 seemed to be more closely related to class I sequences from other bovid species than to the remaining sheep sequences, although this topology was weakly supported by boostrap values (Fig. 6.4 and 6.5).

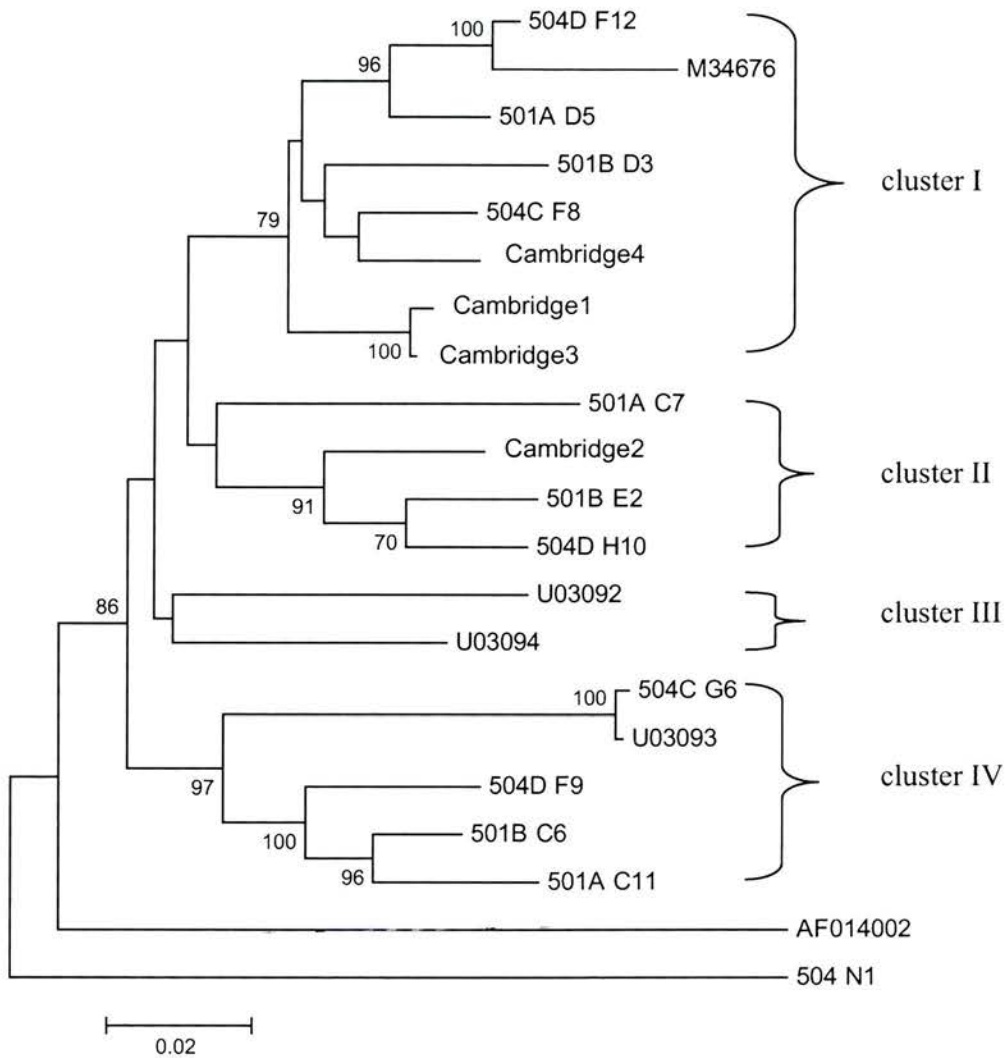


Figure 6.2: Neighbor-Joining phylogenetic tree showing the relationships between the entire coding regions of transcripts identified in this study, along with previously published and four unpublished ovine MHC class I sequences. Sequence names are presented as shown in Table 6.1. The tree is rooted by the swine MHC class I sequence AF14002. Bootstrap values exceeding 70% representing the confidence attached to each node are shown.

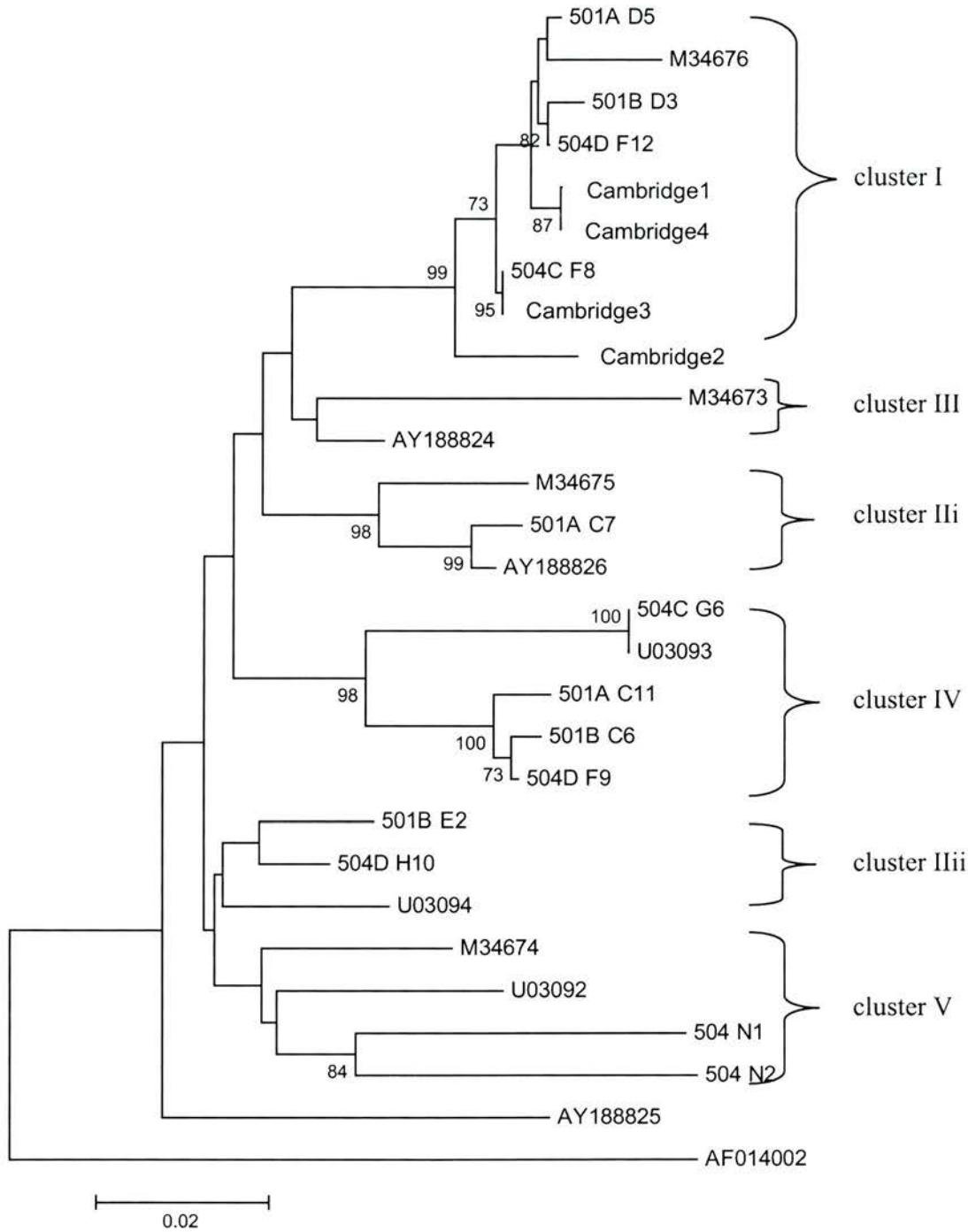
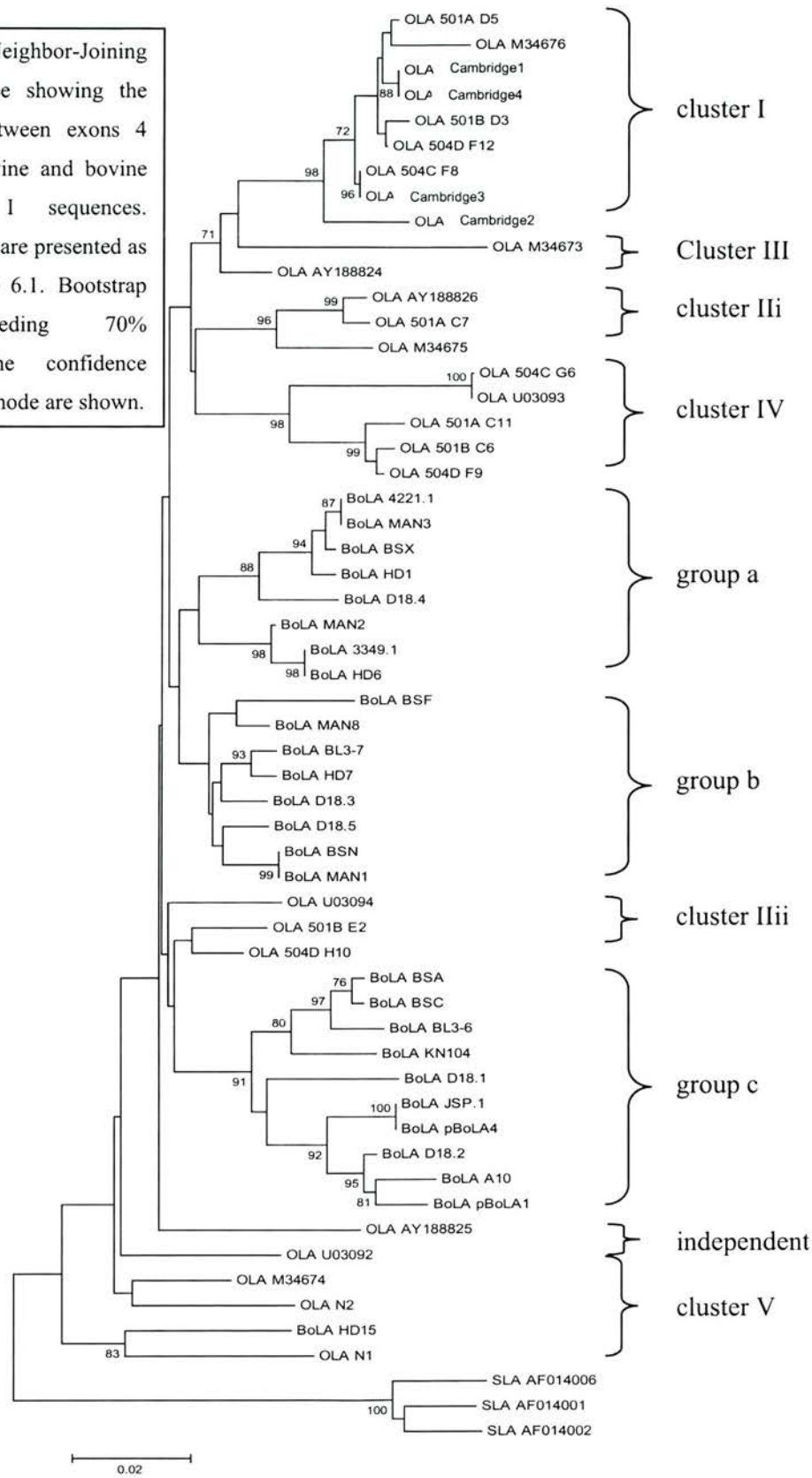


Figure 6.3: Neighbor-Joining phylogenetic tree showing the relationships between exons 4 through 8 of transcripts identified in this study, along with previously published and four unpublished ovine MHC class I sequence . Sequence names are presented as shown in Table 6.1. The trees are rooted by the swine MHC class I sequence AF014002. Bootstrap values exceeding 70% representing the confidence attached to each node are shown.

Figure 6.4: Neighbor-Joining phylogenetic tree showing the relationships between exons 4 through 8 of ovine and bovine MHC class I sequences. Sequence names are presented as shown in Table 6.1. Bootstrap values exceeding 70% representing the confidence attached to each node are shown.



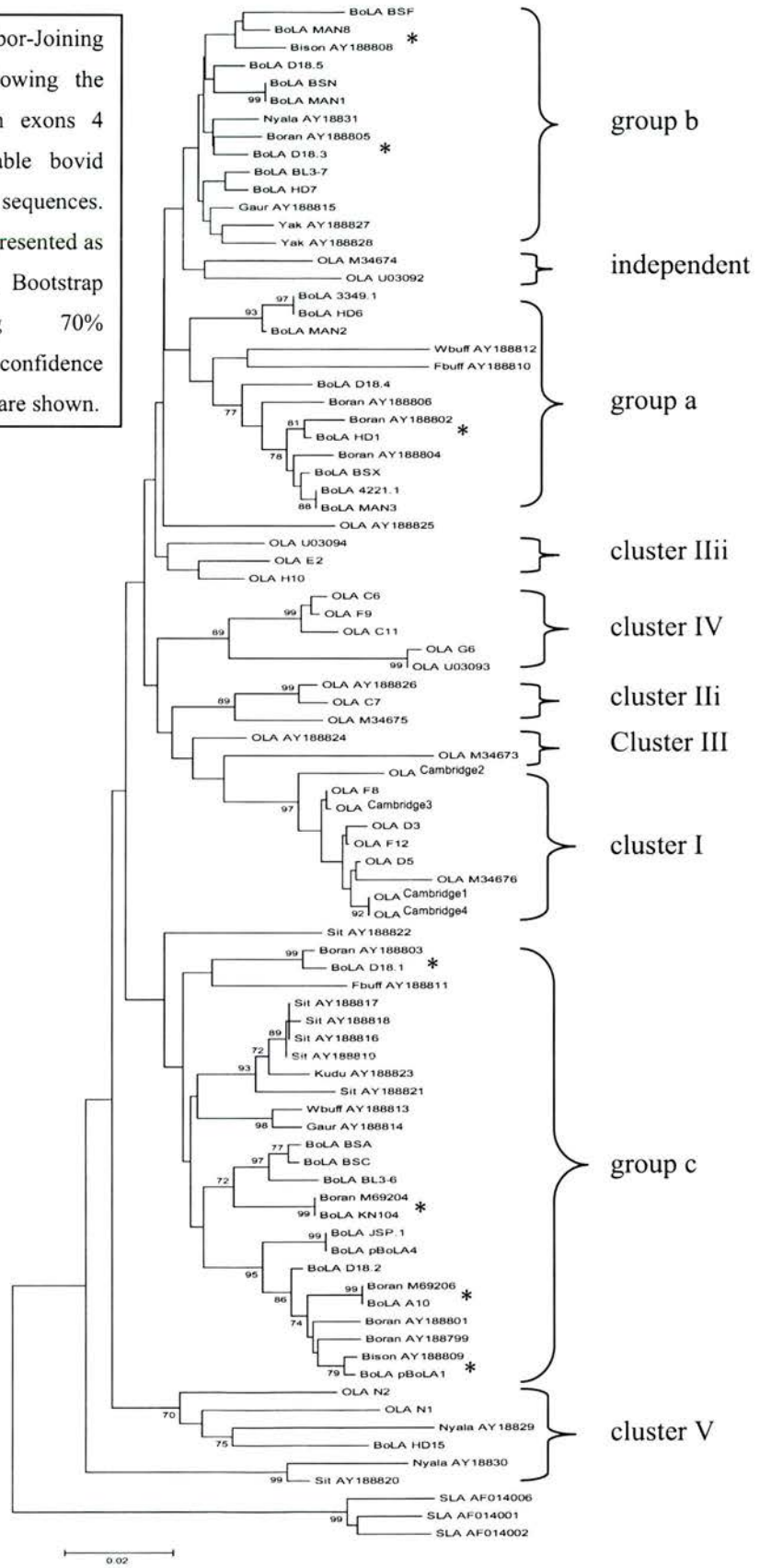


Figure 6.5: Neighbor-Joining phylogenetic tree showing the relationships between exons 4 through 8 of available bovid MHC class I sequences. Sequence names are presented as shown in Table 6.1. Bootstrap values exceeding 70% representing the confidence attached to each node are shown.

6.4 Discussion

Sheep belong to a major branch of mammals that diverged from mice and humans 60 million years ago (Gingerich D. 1977). Based on mammalian phylogeny, the domestic sheep, *Ovis aries*, are ungulates belonging to the Artiodactyla, an order closely related to Perissodactyla within the Laurasiatheria branch of mammals (Murphy et al. 2001). The Artiodactyla order includes species with even numbers of toes, such as pigs, camels, ruminants and hippopotamuses, whereas Perissodactyla include species with odd numbers of toes, such as horses and rhinoceroses (Arnason et al. 2002). The ruminants constitute the largest group of ungulates with over 190 species and are traditionally divided into six families: Antilocapridae, Bovidae, Cervidae, Giraffidae, Moschidae and Tragulidae. Sheep belong to the Caprinae subfamily of the Bovidae family, which also includes the domestic artiodactyl species *Bos taurus* (cow) and *Capra hircus* (goat) (Gatesy et al. 1997).

Given that MHC class I alleles of orthologous loci from different species are sometimes phylogenetically closer to alleles from distinct loci of the same species (Boyson et al. 1996; Gutierrez-Espeleta et al, 2001), phylogenetic analysis of class I sequences from related species could provide information on the relationships and functional significance of ovine class I loci. A study of the evolution of MHC class I genes in Artiodactyls revealed no species-specific groups of sequences, with the exception of giraffe sequences, which form a species-specific cluster situated deep in the tree (Holmes et al. 2003). The authors suggested that diversification of MHC lineages in Artiodactyls occurred after the divergence of giraffe, but before the appearance of the subfamily Bovinae. For that reason, in the work described in this chapter, the ovine class I sequences identified from the rams were compared with published ruminant class I sequences derived from the family Bovidae.

This analysis revealed three distinct well supported clusters of sequences (I, IIi, and IV; Figs. 6.3, 6.4), along with three additional clusters that, although supported by low confidence values, include sequences that consistently segregate together in the phylogenetic trees, (clusters IIii, III and V; Figs. 6.3, 6.4) and a number of branches with one or two sequences distantly related to the clusters (Figs. 6.2-6.4). The tree topology supports the haplotype assignment of individual sequences from rams 501 and 504 (chapter 4), with each sequence within a haplotype being associated with a different cluster (Figs. 6.2-6.4).

In the majority of cases, the segregation of the ovine class I sequences is also in concordance with their estimated level of transcription (chapter 3). The higher frequency

sequences D5, D3, F8 and F12 group in cluster I (Figs. 6.2-6.5) with the also highly transcribed M34676 (Grossberger et al. 1990), suggesting that this cluster contains alleles of a highly expressed classical class I locus. Similarly, the lower frequency sequences C6, C11, F9 and G6 group in cluster IV (Fig. 6.2-6.5), which could be a poorly expressed non-classical locus. These observations are consistent with correspondence of these clusters to individual loci.

When additional sequences become available, however, clusters I and IV may be further divided. In this regard, cluster I sequences D5, M34676, and Cambridge 1, 2 and 4 encode TM domains that are two residues longer than the remaining sequences in the cluster (Figs. 6.1; 6.3 and 6.4). In line with the observation that TM length tends to be conserved within bovine class I loci (Garber et al. 1994), these sequences may represent alleles at a distinct locus. However, the Cambridge 2 sequence seems to be quite divergent from the other sequences in the cluster and does not segregate in the tree with those sequences also having TM domains at 38 residues. The ovine class I sequences designated as Cambridge 1-4 were obtained by PCR and were not confirmed by sequencing of at least three identical copies (Blacklaws B., University of Cambridge, Cambridge, UK, personal communication). Given the large number of hybrid class I sequences observed in the present study, and the observation that the Cambridge 2 sequence contains features from both clusters I and II, it is possible that it is also a PCR artifact. The well supported segregation of the entire coding region of Cambridge 2 with sequences in cluster II (Fig. 6.2), while exons 4 through 8 group with sequences in cluster I, also with high confidence values (Fig. 6.3), further supports the contention that this sequence is a hybrid of at least two other ovine class I sequences. However, it could also be an allele produced by inter-locus recombination as has been observed for bovine class I sequences (Holmes et al. 2003).

Based on phylogeny and the length of the C domain, sequences in cluster IV may also be subdivided further into a classical and a non-classical locus (Fig. 6.1-6.5). Phylogenetic analysis clusters G6 and U03094 very tightly in cluster IV and separates them from the remaining sequences in the cluster by a long branch (Fig. 6.2-6.5). Furthermore, both of these transcripts contain an early stop codon within the cytoplasmic domain (yellow box, Fig. 6.1) and evidence from chapters 3 and 5 strongly suggests that they are alleles of a non-classical locus. In contrast, although transcription levels (chapter 3) and discrimination motifs (VPI or IPI at positions 281-283, underlined in Fig. 6.1; Holmes et al. 2003) suggest that transcripts C6, C11 and F9 could be also non-classical, haplotype assignment in chapter 4 and phylogenetic analysis presented herein indicate that they represent polymorphic alleles at a single locus, which questions their non-classical nature. A number of classical bovine class I loci expressed at levels significantly lower than usual have been described, although this was not a locus

specific phenomenon (reviewed by Ellis S. 2004) and, in humans, alleles of the HLA-C locus exhibit less polymorphism (Falk and Schendel, 1997) and are expressed at significantly lower levels than those at HLA-A and HLA-B (Snary et al. 1977; McCutcheon et al. 1995).

When the analysis was extended to all Bovidae MHC class I sequences, none of the ovine class I clusters described above included sequences from orthologous loci of other bovid species. Although some sheep sequences were more related to those of cattle, they were separated by quite long branches and it is unlikely that they represent alleles from orthologous loci. This finding could be attributed to the use of sequences from bovid species largely belonging to the earliest diverging Bovinae subfamily (Matthee and Davis, 2001). For example, the analysis included no class I sequences from the Caprinae subfamily other than those derived from sheep, due to unavailability of such data in public databases. Therefore, a more accurate estimate of the number of ovine MHC class I loci may be achieved in the future by greatly expanding the number of class I sequences with additional ovine and other Caprinae subfamily sequences. However, experience from phylogenetic analysis of MHC class I sequences from Artiodactyls suggests that assignment of alleles to loci in ruminants is not as straightforward as in humans, due to inter-locus recombination (Holmes et al. 2003). It is therefore possible that proper ovine class I allele designations will ultimately be confirmed only by physical mapping.

Since it is likely that cluster I contains sequences from two highly expressed classical loci, while cluster IV may be divided into a classical and a non classical locus, the transcripts identified in this study may represent alleles of seven or eight loci, of which at least three are classical. Given that some published sequences failed to cluster with our transcripts, the number of ovine MHC class I loci may exceed eight. These results suggest a complexity similar to that found in the bovine MHC class I region, where a combination of phylogenetic analysis (Holmes et al. 2003), mapping (Bensaid et al. 1991; DiPalma et al. 2002), expression studies (Ellis et al. 1996; 1999) and identification of locus-specific motifs from introns has provided evidence for at least six expressed loci (Ellis S. 2004). However, the results obtained to date for the ovine MHC class I region are based only on transcribed genes, some of which may not be expressed. The work described in the following chapter aimed to determine if those transcripts are capable of translation into functionally folded molecules on the surface of mammalian cells.

Chapter 7. Transfection of ovine MHC class I transcripts

7.1 Introduction

In the previous chapters, 6 partial and 11 full length ovine MHC class I transcripts were identified and assigned to four haplotypes. Their exons 4 to 8, where available, were grouped into five distinct clusters, suggesting that excluding the partial transcripts they represent alleles of at least five class I loci. However, the evidence for transcription of these genes does not confirm that their products are expressed at the cell surface as functional molecules.

The process whereby genes are artificially transferred into mammalian cells is known as transfection and can provide essential information regarding the expression and function of a gene. It is particularly useful for the study of eukaryotic proteins that are not expressed efficiently or folded correctly in prokaryotic systems, as well as for those that require posttranslational modifications and targeting to the cell surface (Sambrook et al. 1989), such as MHC glycoproteins.

Transfection of cloned genes encoding major histocompatibility molecules has been used to identify functional class I molecules in many species including human, pig, cattle and mouse (Goodenow et al. 1982; Lemonnier et al. 1982; Sawhney et al. 1995; Singer et al. 1982). Although transfected L cell lines expressing ovine MHC class II heterodimers have been generated (Ballingall et al. 1992; Ballingall et al. 1995; Wright and Ballingall, 1994), transfection and expression of ovine MHC class I molecules has not been described previously. Therefore, to date, the number of expressed ovine MHC class I loci has only been estimated by serology.

The work described in this chapter sought to determine if the highly transcribed class I sequences D5, D3 and C7 from ram 501 and F8 and F12 from ram 504, which are strong candidates for genes encoding classical antigen presenting molecules, can be expressed on the surface of mammalian cells. In addition, since three full-length genes (D5, C7 and C11) were available for the 501 haplotype A and another three (F12, H10 and F9) for the 504 haplotype D (chapter 4; Fig. 4.18), the poorly transcribed sequences C11, H10 and F9 were also transfected into mammalian cells to identify the number of sequences that can be expressed from a single haplotype and whether there is variation in the number of putatively expressed genes between haplotypes. The partial transcripts F7 and G12 identified in both haplotypes were strong

candidates for non-classical class I genes (chapter 4) and were not analyzed further by expression analysis.

7.2 Materials and methods

7.2.1 PCR amplification

The full length ovine MHC class I sequences D5, D3, C7 and C11 from ram 501 and F8, F12, H10 and F9 from ram 504 were amplified from the pGEM-T Easy vector and sub-cloned into the pcDNA3.1/V5-His-TOPO plasmid vector to generate mammalian expression constructs.

Each plasmid DNA was prepared as described in section 2.3.3 and amplified with one forward and two reverse sequence-specific primers (Table 7.1; Fig. 7.1). All forward primers annealed at the first exon of each sequence and contained the kozak sequence CACC (Kozak S. 1987; 1990) and one initiation codon for efficient expression in mammalian cells. All reverse primers annealed at the 5' end of each sequence just before the 3' untranslated region, because sequences within that area have been shown to destabilize transcribed mRNAs from mammalian expression vectors (Sambrook et al. 1989). The reverse primers denoted with A contained the codon TTT instead of the stop codon at the 5'-end. In this way, the amino acid lysine is translated and the open reading frame of the class I sequence is in frame with the C-terminal tag of the pcDNA3.1/V5-His-TOPO vector, which can be detected by the anti-His antibody (Lindner et al. 1997). In contrast, the reverse primers denoted with B contained a stop codon and an additional base (G) at the 5'-end so that the translation of the protein is terminated at that point and the remaining sequence of the vector is out of frame.

PCR reactions were carried out in 50µl final volume containing 49 µl PCR master mix (60 mM Tris-SO_4, 18mM ammonium sulphate, 200 µM of each dNTP, 200 nM of each primer and 1U high fidelity Platinum *Taq* DNA polymerase (Invitrogen Ltd, Paisley, UK)) and 1µl (≈50 pg) plasmid DNA. The cycling profile was 25 cycles of 1 min at 94°C, 20 sec at 60°C and 2 min at 68°C, and one cycle of 1 min at 94°C, 20 sec at 60°C and 10 min at 68°C.

	Forward Primers		Reverse Primers		
Target sequence	*5' Sequence 3'*	*A 5' Sequence 3'*	*B 5' Sequence 3'*		
D5	CACCATGGGGCCGCGAACCGTC	TTTCACTTTATGAACCGTGA GAGACAC	GTCACACTTTATGAACC GTGAGAGACAC		
D3	CACCATGGGGCCGCGAACCCTC	TTTCACTTTATGAACCGTGA GAGACAC	GTCACACTTTATGAACC GTGAGAGACAC		
C7	CACCATGGGGCCGCGAACCCC	TTTCACTTTAGGAACCATAA GAGACAC	GTCACACTTTAGGAACC ATAAGAGACAC		
C11	CACCATGGGGCCGCGAAC	TTTCACTTTAGGAACTGTGA GAGACAC	GTCACACTTTAGGAACT GTGAGAGACAC		
F8	CACCATGGGGCCGCGAACCCTC	TTTCACTTTAGGAACCGTGA GAGACAC	GTCACACTTTAGGAACC GTGAGAGACAC		
F12	CACCATGGGGCCGCGAACCCTC	TTTCACTTTATGAACCGTGA GAGACAC	GTCACACTTTATGAACC GTGAGAGACAC		
H10	CACCATGGGGCCGCGAACCCC	TTTCACTTTAGGAACCGTGA GAAACAC	GTCACACTTTAGGAACC GTGAGAAACAC		
F9	CACCATGGGGCCGCGAAC	TTTCACTTTAGGAACTGTGA GAGACAC	GTCACACTTTAGGAACT GTGAGAGACAC		

Table 7.1: Primers used for amplification of full length ovine MHC class I sequences prior to sub-cloning into the pcDNA3.1/V5-His-TOPO vector.

7.2.2 Cloning of PCR products and transformation

PCR products of the appropriate size were excised from 0.7% agarose gels, as described in section 2.5.4 and cloned into the pcDNA3.1/V5-His-TOPO plasmid vector (Invitrogen Ltd, Paisley, UK, Appendix 1.3), as described in section 2.6.3. The cloned PCR products were then transformed into TOP10 competent cells, as described in section 2.7.2 and cultured for selection of transformed cells (sections 2.7.3 and 2.7.4).

7.2.3 Sequence analysis

For each ovine MHC class I insert, 3-5 replicates of clones were picked for sequencing. Plasmid DNA was recovered and analyzed as described in sections 2.3.3 and 2.5.6. Sequencing reactions and analysis were performed as described in section 2.8. The sequences were compared with the consensus of each ovine MHC class I sequence identified in chapter 5 and sequenced clones that contained no *Taq* polymerase errors were stored in glycerol stocks (section 2.7.5).

7.2.4 Preparation of plasmid DNA for transfection

Mammalian expression plasmids containing each of the full length ovine MHC class I sequence were purified as described in section 2.3.3 and the nucleic acid concentrations were estimated (sections 2.5.1, 2.5.5 and 2.5.6) and adjusted to 0.5μg/μl after ethanol precipitation (section 2.5.7).

7.2.5 Preparation of COS-7 cells for transfection

A vial of frozen COS-7 cells was retrieved from liquid nitrogen as described in section 2.9.6 and cultured for 36 hours (section 2.9.4). The cells were passaged using a 1:5 ratio just prior to confluence (section 2.9.5) 24h before transfection.

7.2.6 Transfection and expression analysis

Aliquots ($5x10^5$) of COS-7 cells were transfected with 10μg of pcDNA3.1/V5-His-TOPO plasmid vector containing each of the full length ovine MHC class I sequences D5, D3, C7 or C11 from ram 501 and F8, F12, H10 or F9 from ram 504 (section 2.10.1). Transfected cells were analysed after 48-72h culture by indirect immunofluorescence and flow cytometry, as described in sections 2.10.2 and 2.10.3, respectively. Photographic images of the COS-7 cells expressing class I genes were obtained under an Olympus CKX41 inverted microscope (Olympus UK Ltd., London, UK).

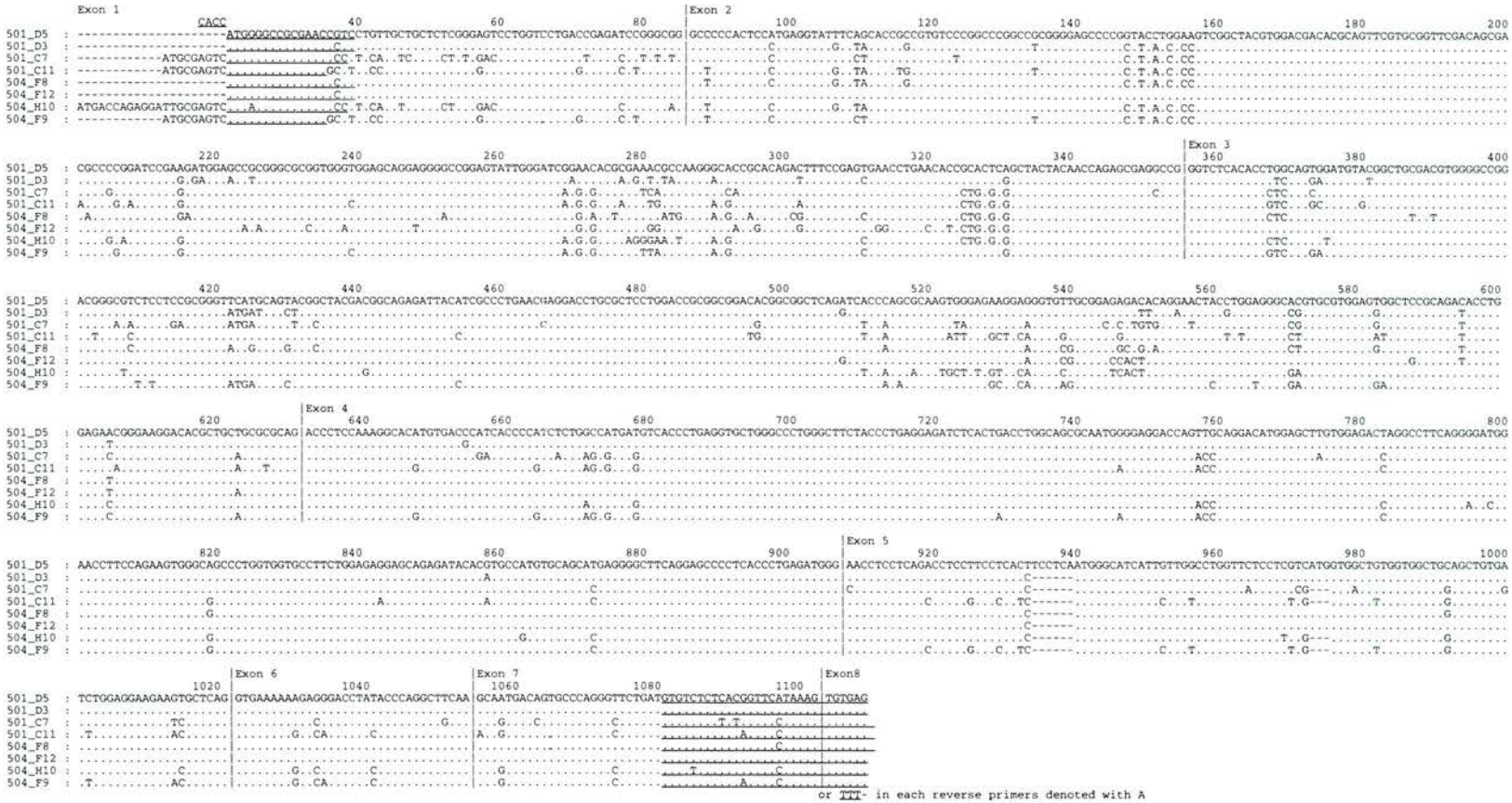


Figure 7.1: An alignment of full length ovine MHC class I transcripts with the sequence specific primers (*underlined)* used for amplification prior to subcloning into mammalian expression constructs.

7.3 Results

7.3.1 PCR products

The full length ovine MHC class I sequences D5, D3, C7 and C11 from ram 501 and F8, F12, H10 and F9 from ram 504 were amplified from their respective pGEM-T Easy vector constructs (chapter 5) using the PCR protocol described in section 7.2.1. A specific band of the expected size (Table 7.2) was obtained from each plasmid DNA using the sequence-specific primer pairs presented in Table 7.1 (Fig. 7.2).

Target sequence	PCR product size with the reverse primers denoted with A (bp)	PCR product size with the reverse primers denoted with B (bp)
D5	1087	1088
D3	1081	1082
C7	1078	1079
C11	1078	1079
F8	1081	1082
F12	1081	1082
H10	1078	1079
F9	1078	1079

Table 7.2: PCR product sizes amplified by each primer pair.

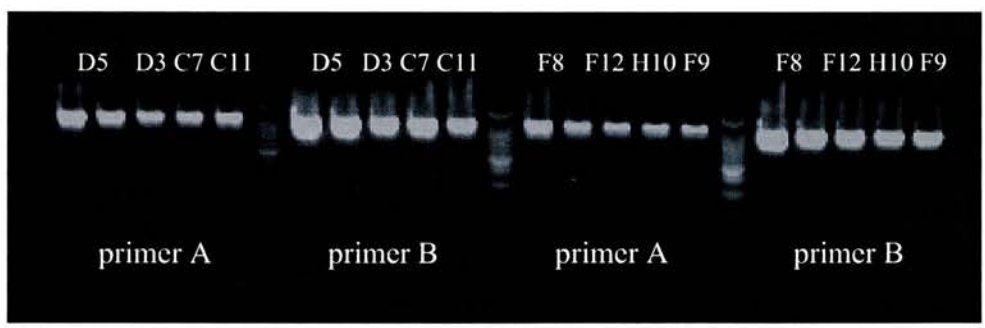


Figure 7.2: PCR products amplified from plasmid DNA containing each of the full length ovine MHC class I sequences D5, D3, C7 and C11 from ram 501 and F8, F12, H10 and F9 from ram 504. For each target sequence two reverse primers (A and B) were used. Five and 10 µl of PCR product amplified from D5 and F8 was loaded.

7.3.2 Subcloning and selection of recombinants by PCR

PCR products amplified from plasmid DNA containing each of the full-length ovine MHC class I transcripts D5, D3, C7, C11, F8, F12, H10 and F9 were purified and subcloned into the

pcDNA3.1/V5-His-TOPO plasmid vector to generate mammalian expression constructs. Approximately 200 white colonies, around 10-15 for each of the 16 PCR products (Fig. 7.2) were screened by colony PCR using primers B and Br and 30-80% were positive (Fig. 7.3). Three to 5 positive clones obtained from each sub-cloned amplification product were selected for sequencing.

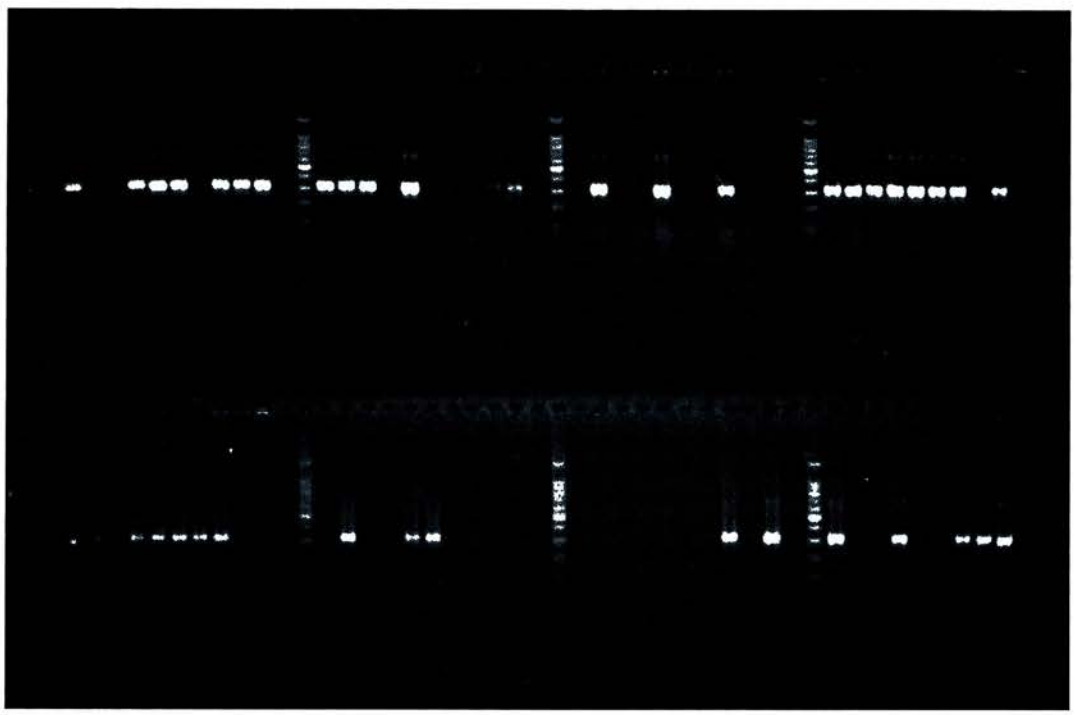


Figure 7.3: Agarose electrophoresis of colony PCR products of clones containing full length ovine MHC class I sequences.

7.3.3 Sequencing

Sequencing identified at least one clone for each PCR product with no *Taq* polymerase errors and a 5' to 3' orientation appropriate for expression analysis.

7.3.4 Expression of ovine MHC class I genes

Ovine MHC class I genes cloned in the pcDNA3.1/V5-His-TOPO plasmid vector and transfected in COS-7 cells were analysed using a flow cytometer after labeling with IL-A88, a monoclonal antibody that recognizes a monomorphic determinant of the bovine MHC class I heavy chain (Toye et al. 1990) and cross-reacts with the ovine product (Chai Z. 2005). IL-A88 surface staining of cells transfected with ovine MHC class I sequences D5, D3, C7 and C11

from ram 501 and F8, F12 and F9 from ram 504 confirmed that the genes can give rise to conformationally intact molecules on the surface of COS cells, whereas the H10 molecule from ram 504 was not detected by IL-A88. Figures 7.4 through 7.11 show transfected cells under a fluorescent microscope after immunostaining (a), the corresponding field under light microscope (b) and the flow cytometry diagram (c) for each class I sequence that was sub cloned in-frame (A) or out-of-frame with the histidine tag (B). Both procedures established successful transfectants. Figure 7.12 shows a positive bovine MHC class I control (A) and a negative control with no DNA during transfection (B).

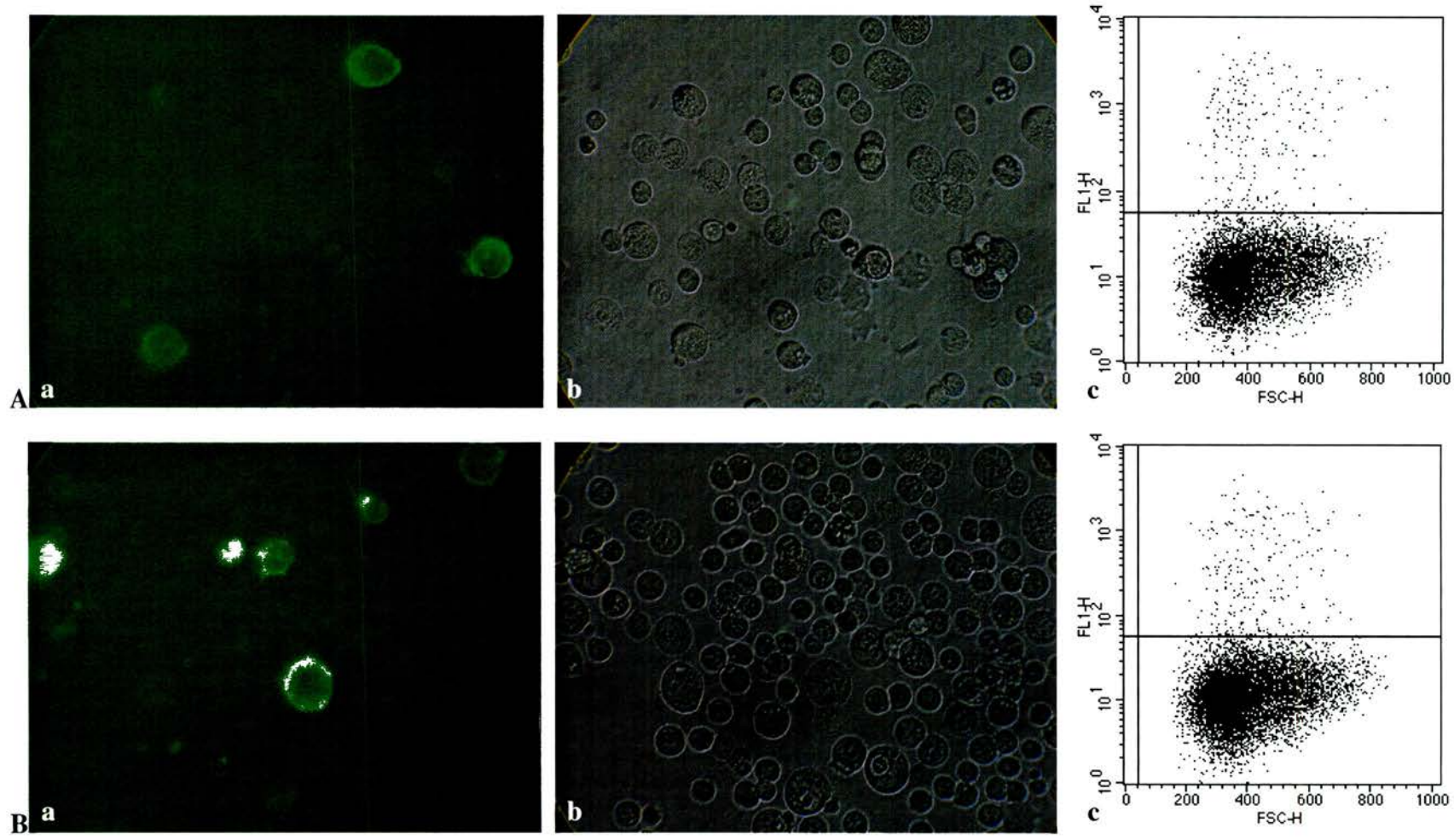


Figure 7.4: Transfected cells with the ovine MHC class I sequence D5 under a fluorescent microscope (a), the corresponding field under light microscope (b) and the flow cytometry diagram (c) for each class I sequence that was subcloned in frame (A) or out of frame with the histidine tag (B).

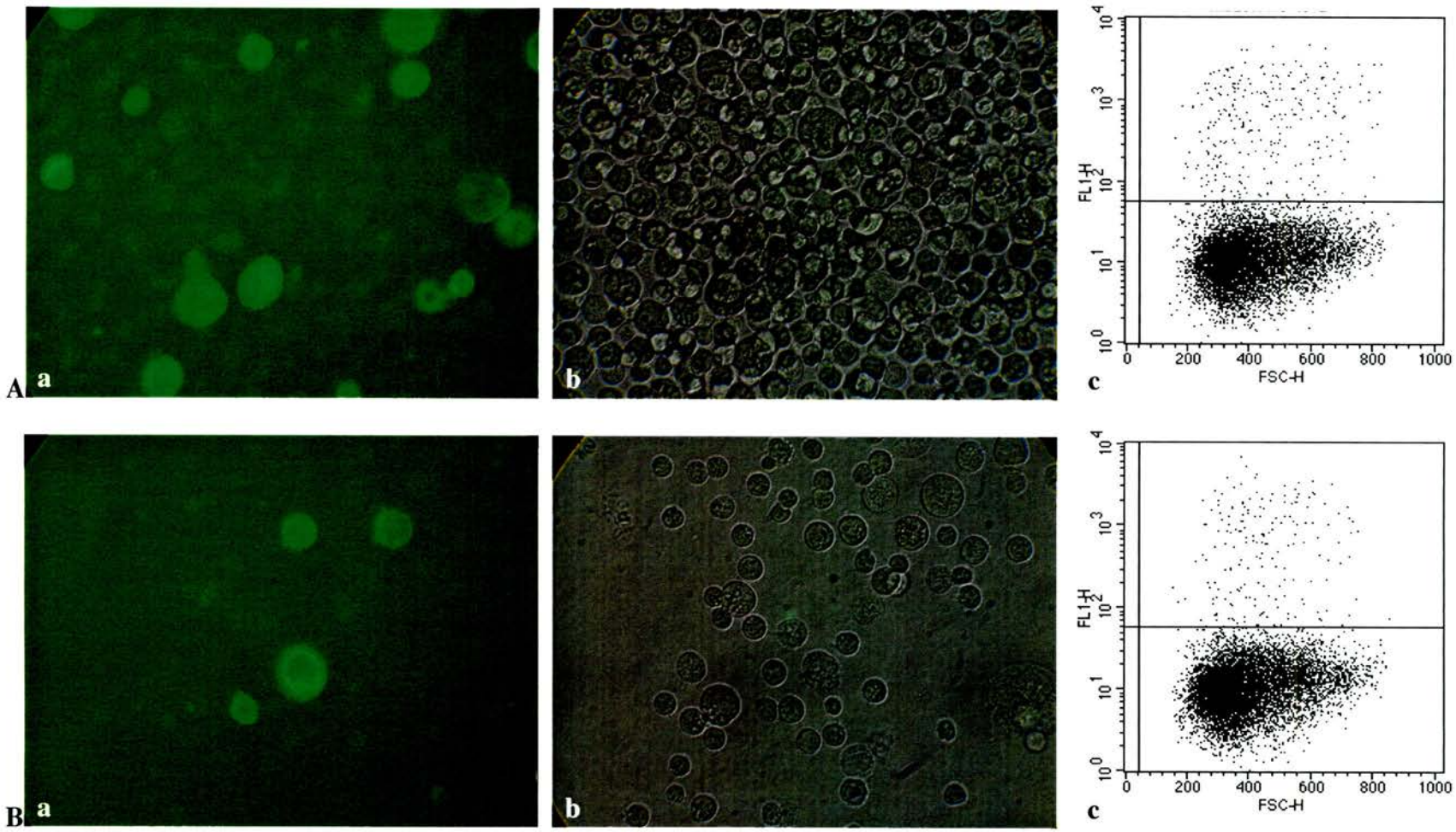


Figure 7.5: Transfected cells with the ovine MHC class I sequence D3 under a fluorescent microscope (a), the corresponding field under light microscope (b) and the flow cytometry diagram (c) for each class I sequence that was subcloned in frame (A) or out of frame with the histidine tag (B).

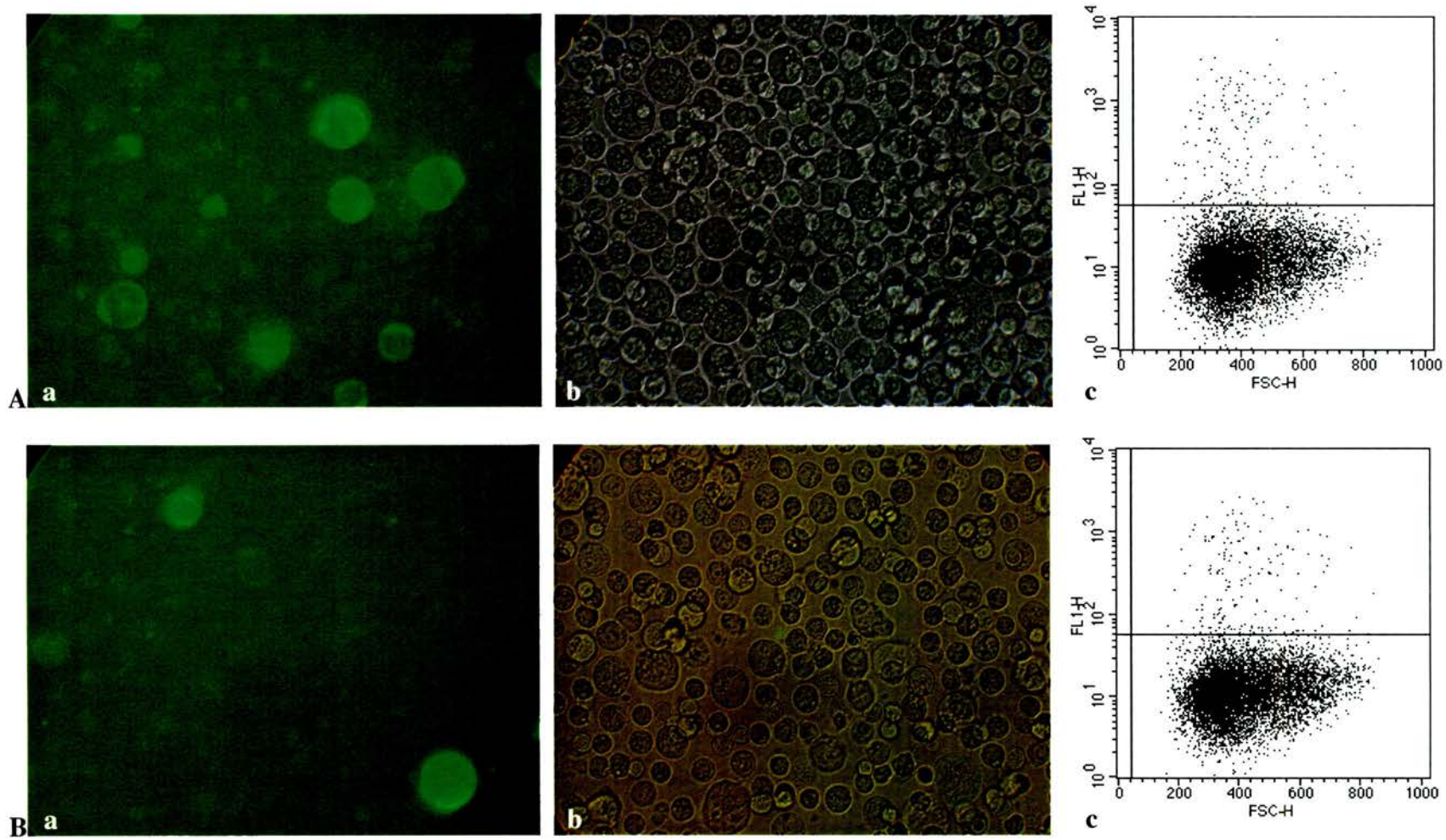


Figure 7.6: Transfected cells with the ovine MHC class I sequence C7 under a fluorescent microscope (a), the corresponding field under light microscope (b) and the flow cytometry diagram (c) for each class I sequence that was subcloned in frame (A) or out of frame with the histidine tag (B).

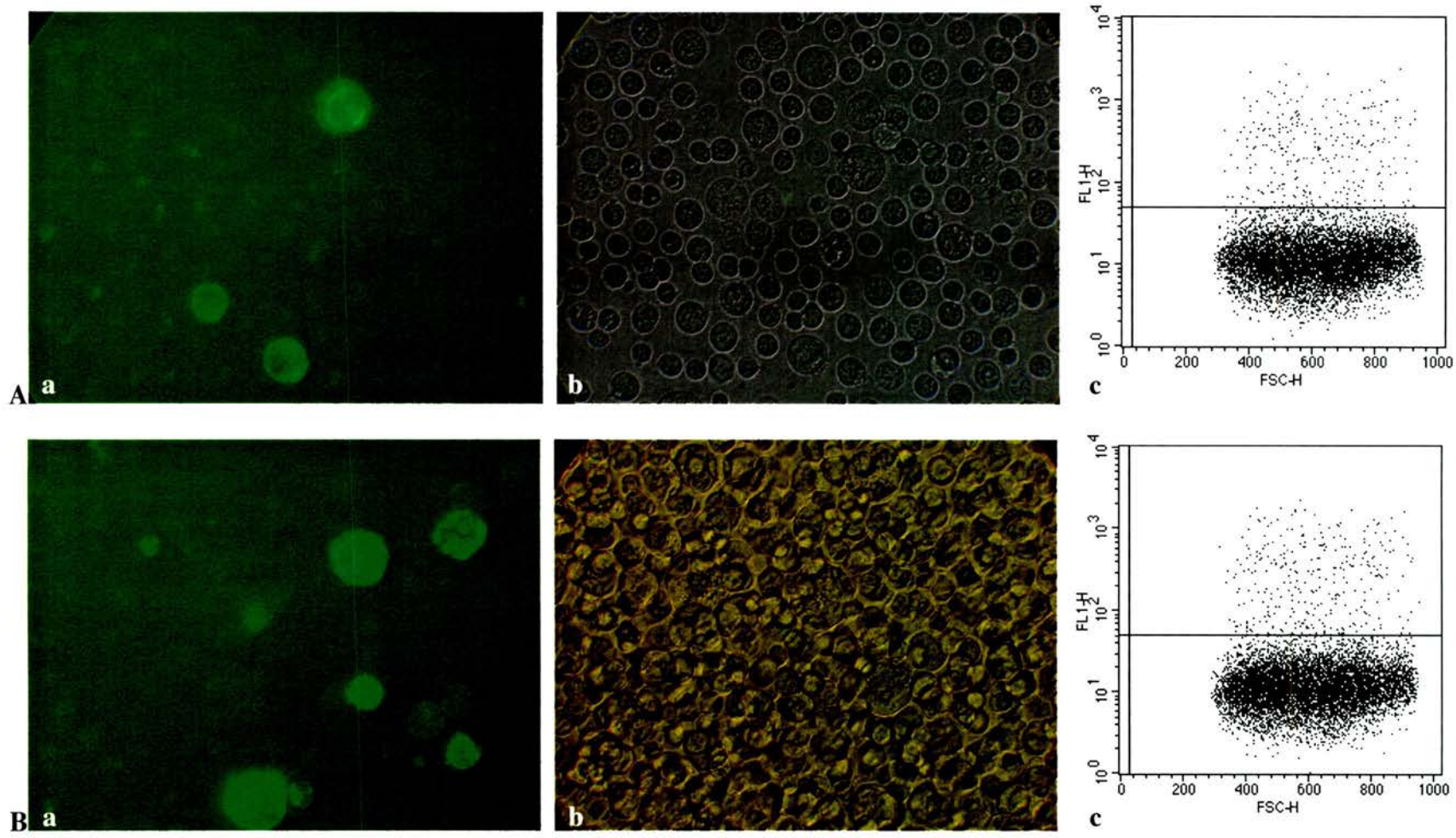


Figure 7.7: Transfected cells with the ovine MHC class I sequence C11 under a fluorescent microscope (a), the corresponding field under light microscope (b) and the flow cytometry diagram (c) for each class I sequence that was subcloned in frame (A) or out of frame with the histidine tag (B).

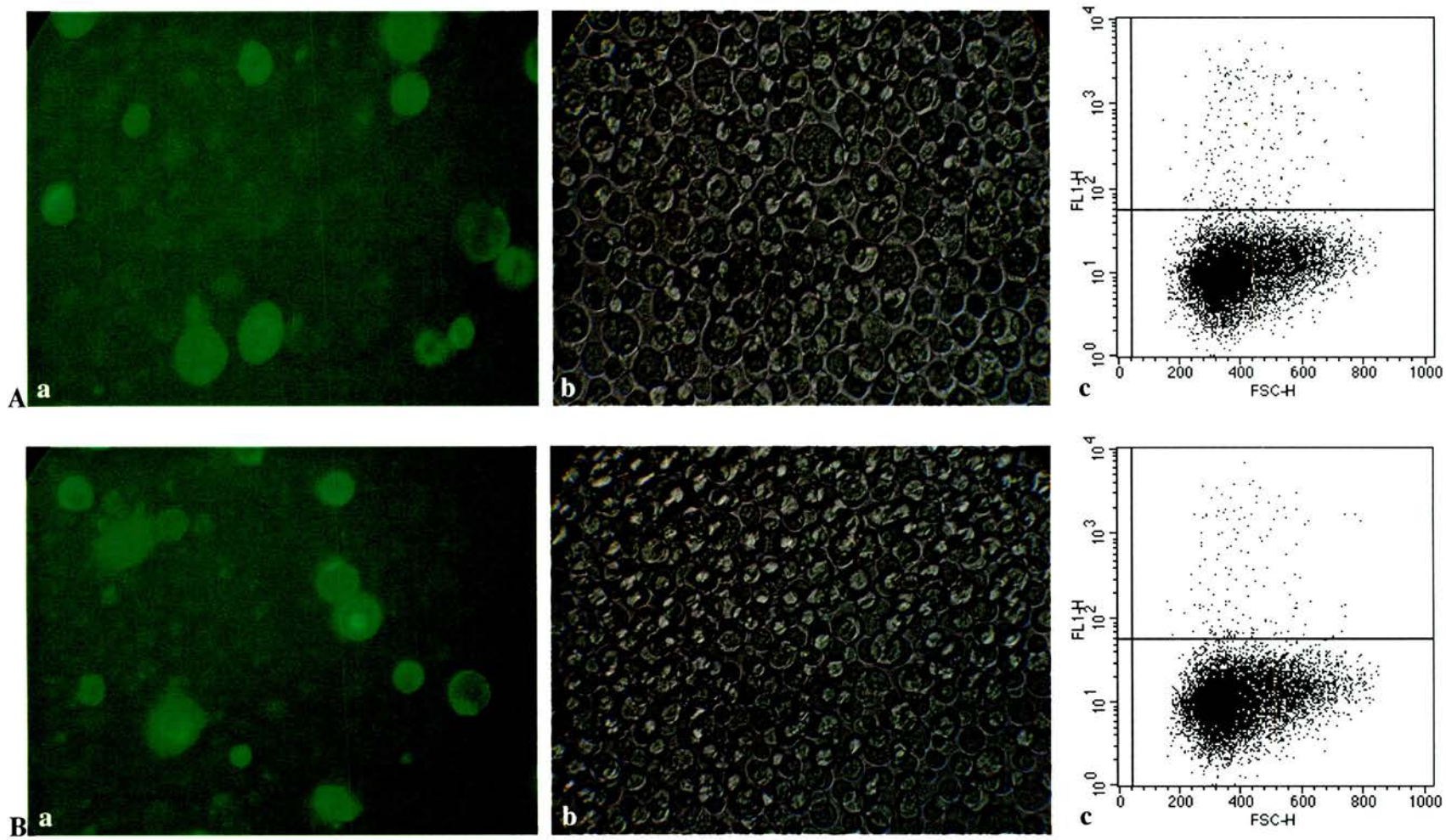


Figure 7.8: Transfected cells with the ovine MHC class I sequence F8 under a fluorescent microscope (a), the corresponding field under light microscope (b) and the flow cytometry diagram (c) for each class I sequence that was subcloned in frame (A) or out of frame with the histidine tag (B).

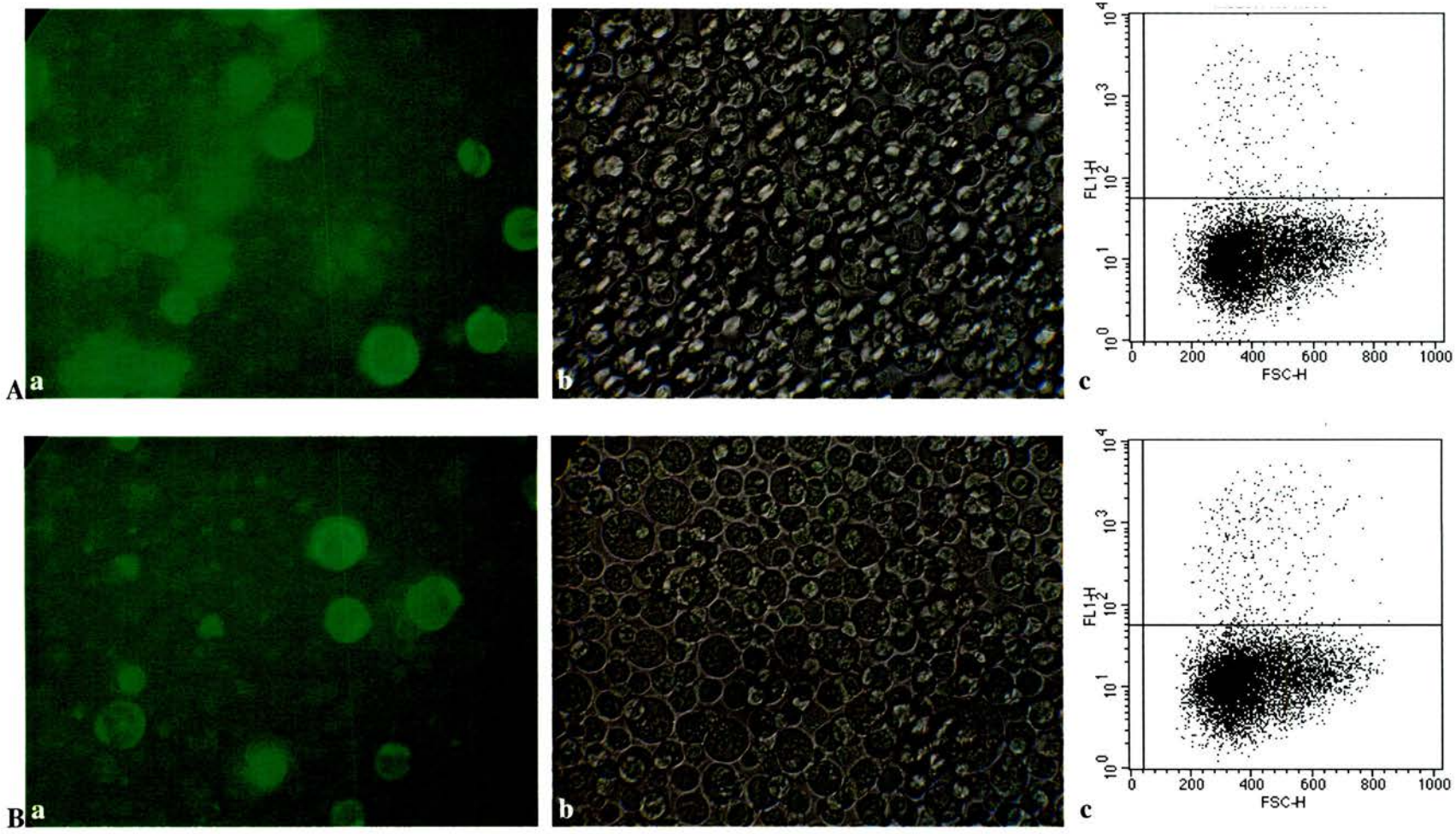


Figure 7.9: Transfected cells with the ovine MHC class I sequence F12 under a fluorescent microscope (a), the corresponding field under light microscope (b) and the flow cytometry diagram (c) for each class I sequence that was subcloned in frame (A) or out of frame with the histidine tag (B).

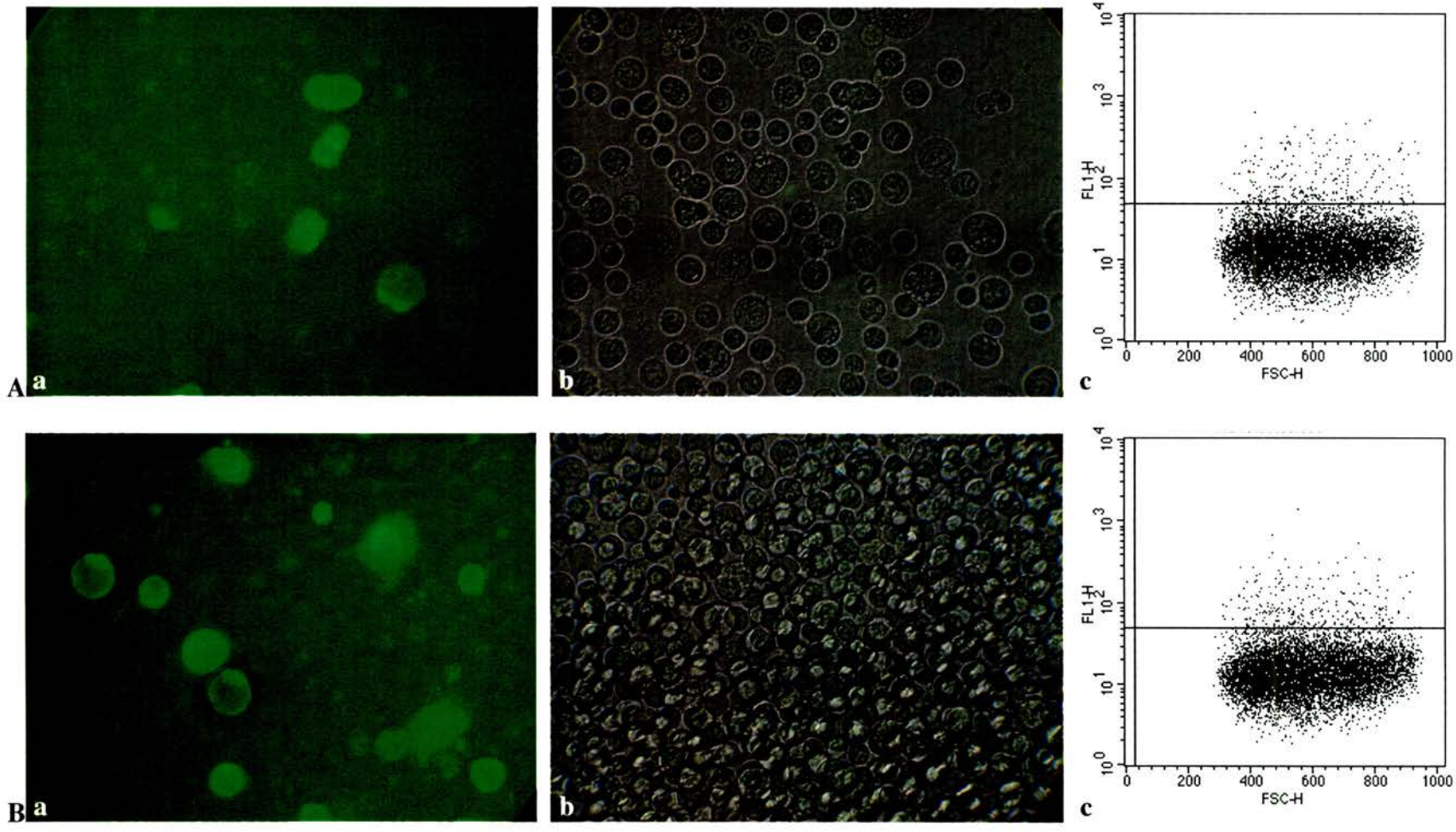


Figure 7.10: Transfected cells with the ovine MHC class I sequence F9 under a fluorescent microscope (a), the corresponding field under light microscope (b) and the flow cytometry diagram (c) for each class I sequence that was subcloned in frame (A) or out of frame with the histidine tag (B).

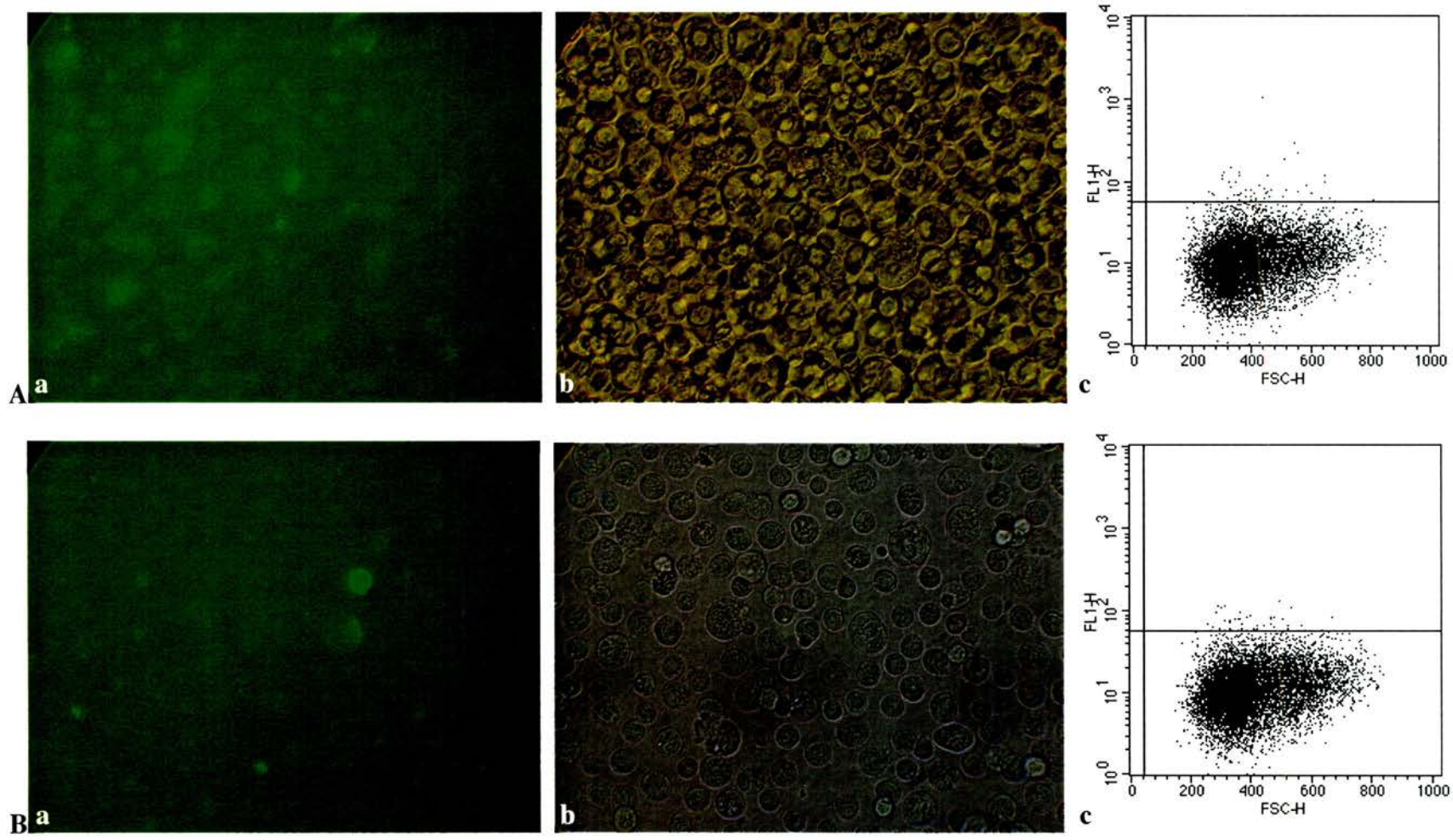


Figure 7.11: Transfected cells with the ovine MHC class I sequence H10 under a fluorescent microscope (a), the corresponding field under light microscope (b) and the flow cytometry diagram (c) for each class I sequence that was subcloned in frame (A) or out of frame with the histidine tag (B).

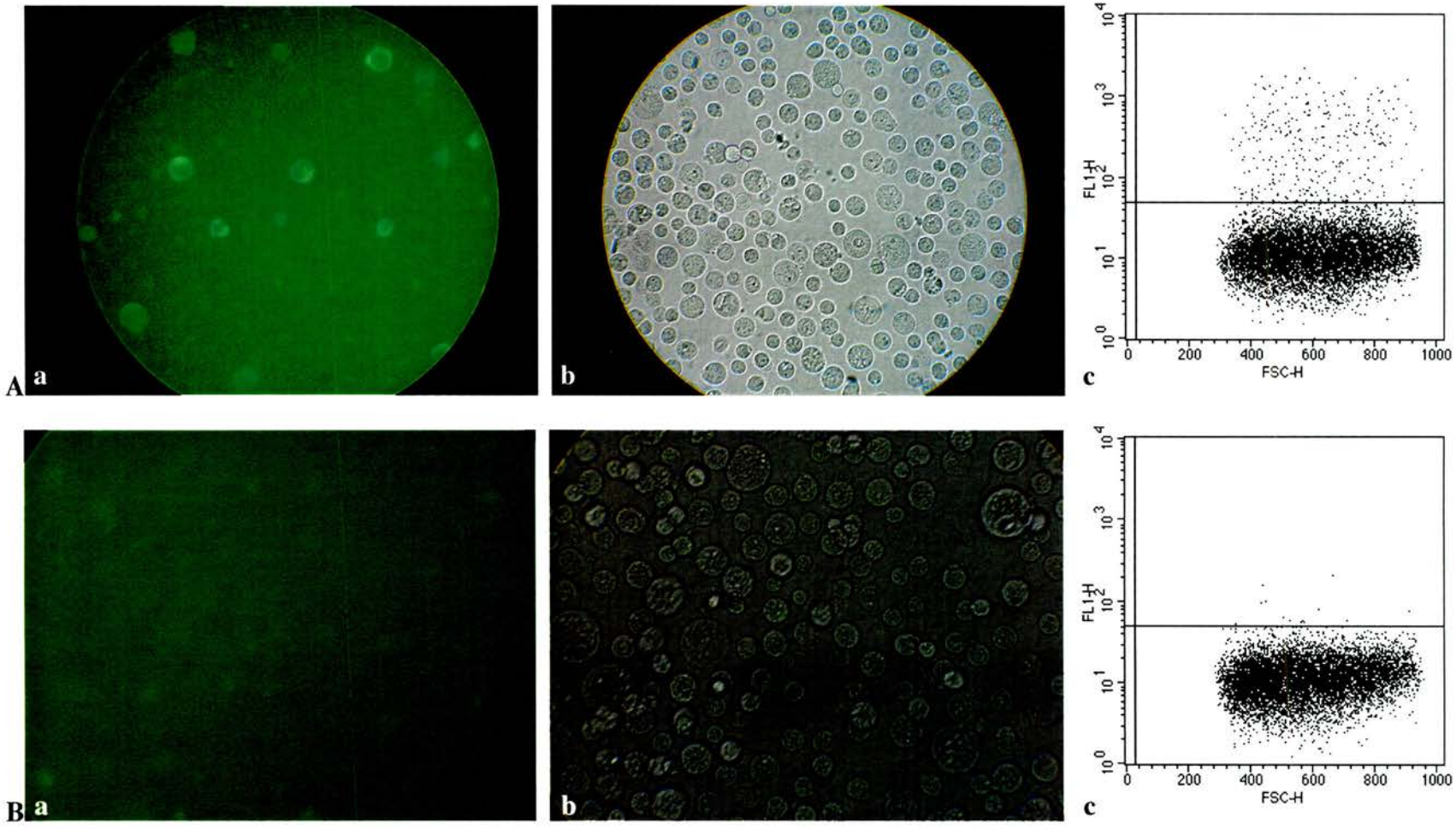


Figure 7.12: (a) The corresponding field under light microscope (b) and the flow cytometry diagram (c) for a positive bovine MHC class I control (A) and a negative control with no DNA during transfection (B).

7.4 Discussion

7.4.1 Development of mammalian expression constructs

This chapter describes the use of 8 full-length ovine MHC class I transcripts to generate mammalian expression constructs. In general, such constructs are based on viruses or plasmids and contain the gene of interest along with elements necessary for efficient expression of that gene in mammalian cells. A plasmid vector was used in the present study, because it has the advantage of containing prokaryotic, eukaryotic and viral functional components (reviewed in Colosimo et al. 2000). The prokaryotic elements allow bacterial propagation and maintenance of the vector, while the eukaryotic and viral elements allow efficient expression of the gene in mammalian cells.

Mammalian expression systems can be stable or transient, depending on whether the gene is expressed permanently or for only a short period (reviewed in Colosimo et al. 2000). Transient expression systems are very useful for studying the expression of a gene quickly, within 12 to 72 hours after transfection. In contrast, stable transfectants require a 3-4 week period of cell culture with selectable markers, but can be stored and used on a long term basis. For the purposes of this chapter, transient transfectants were generated to allow rapid evaluation of whether the class I molecules could be expressed on the surface of COS-7 cells.

7.4.2 Expression of ovine MHC class I transcripts

Although three classical class I loci, HLA-A, -B and –C, are expressed from all human MHC haplotypes, such consistency appears to be the exception rather than the rule for other species. For example, the number of class I genes varies in different inbred strains of mouse (reviewed by Trowsdale J. 1995) and variable numbers of MHC class I genes have been reported in different serologically defined swine haplotypes (Ruohonen-Lehto et al. 1998). Similarly, although two polymorphic genes are ubiquitously expressed in equids, at least four putative non-classical genes may also be expressed at low levels (Ellis et al. 1995; Holmes and Ellis, 1999). In cattle, there is evidence for at least six classical MHC class I loci (reviewed by Ellis S. 2004), which occur in a number of combinations on different haplotypes (Ellis et al. 1999).

Haplotype variation may also be a feature of ovine class I genes. A number of serological studies have suggested two expressed loci (Cullen et al. 1982; Garrido et al. 1995; Jugo and Vicario 2001; Jugo et al. 2002; Stear and Spooner 1981), whereas others propose three (Grossberger et al. 1990; Millot P. 1984). In addition, evidence based on isoelectric focusing

(IEF) suggests that IEF patterns could be explained by different numbers of genes expressed on different haplotypes (Jugo et al. 2002).

In this study, the expression of ovine MHC class I molecules was evaluated on the surface of COS-7 cells using the IL-A88 antibody, which has been used successfully for identification of bovine class I molecules with a wide range of specificities (Dr Shirley Ellis, IAH, Compton, UK, personal communication) and cross-reacts with the ovine product (Chai Z. 2005). Of the 8 ovine class I transcripts that were transfected and analyzed, 7 (D5, D3, C7, C11, F8, F12, F9) were demonstrated to give rise to surface products identified by IL-A88. The remaining construct, H10, failed to result in a detectable product. As transcripts D5, C7 and C11 are derived from the 501 haplotype A and sequences F12, H10 and F9 locate to the 504 haplotype D (chapter 4; Fig. 4.18), these expression analyses indicate that at least three and two class I loci, respectively, are capable of translating ovine MHC class I molecules on the surface of mammalian cells from these haplotypes (Fig. 7.13).

In a collaborative experiment conducted with Dr. Zhongwei Chai to confirm the expression of the transcribed class I genes in vivo, leukocytes derived from lambs homozygous for each of the 501 and 504 haplotypes were transformed by infection with the intracellular parasite *Theileria lestoquardi* (Hooshmandrad P. 1993), immunoprecipitated after lysis with IL-A88 and separated by SDS-polyacrylimide gel electrophoresis. MHC class I heavy chains were then excised from the gels, digested with trypsin and analyzed by MALDI-TOF mass spectrometry. Peptide mass profiles were then screened using the Mascot software against protein sequence databases constructed for each of the four haplotypes (Zhongwei Chai, MRI, Edinburgh, UK, unpublished data). The resulting data corroborate the observations of the present study that sequences D5, C7 and C11 are expressed from 501 haplotype A, as are sequences D3, E2, C6 from 501 haplotype B (Fig. 7.13). In addition, sequences F8 and G6 from 504 haplotype C and F12, F9 and H10 from 504 haplotype D were expressed by transformed homozygous lymphocytes, although it is not clear if the products reach the cell surface. Therefore, it seems that, as with other species, the number of expressed class I loci varies between different ovine haplotypes.

The mass spectrometry expression findings are in agreement with the available transfection results with the exception of the H10 sequence. The H10 molecule was not detected on the surface of transfected COS-7 cells, whereas it was expressed in *Theileria*-infected cell lines. Since the same antibody was used in both systems, it is unlikely that it failed to recognize a phylogenetically diverse H10 molecule in transfected cells. It is possible that a molecular

cloning problem prevented H10 from expression on transfected cells, or incorrectly folded H10 molecules were expressed. Detection of the epitope encoded by the histidine or the V5 tag, on the surface of COS cells, and repetition of the cloning experiment will possibly resolve this issue.

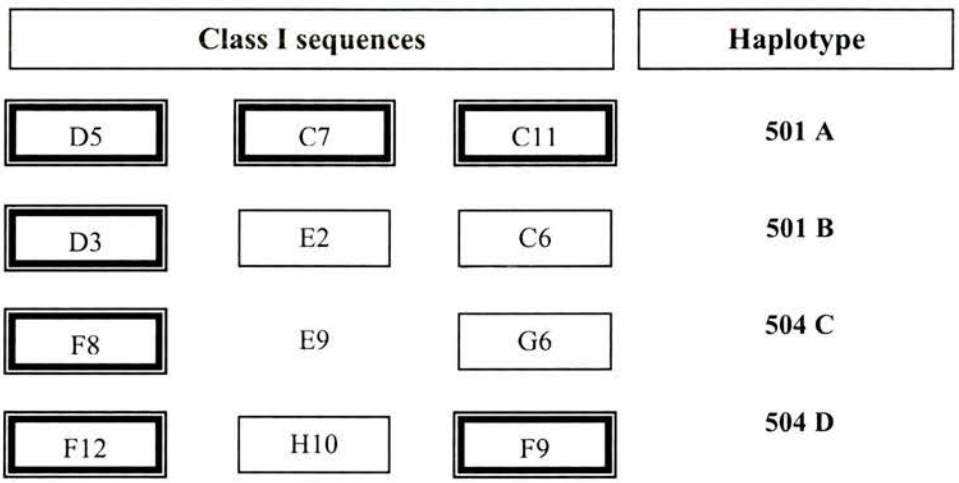


Figure 7.13: A diagrammatic representation of the ovine class I sequences transcribed and expressed by four different haplotypes from rams 501 and 504. Sequences *boxed in a thick line* were expressed on the surface of COS-7 cells transfected with each sequence, while sequences *boxed in a thin line* were expressed by *Theileria*-infected cells from lambs homozygous for the relevant 501 and 504 haplotypes. The partial E9 transcript was not used in the analysis in this chapter.

7.4.3 Applications of mammalian expression constructs

The mammalian expression constructs generated can be used for transfection of a wide range of mammalian cells. Stable transfectants expressing D5, D3, C7, F12 and F8 molecules have already been generated (Keith Ballingall, MRI, Edinburgh, UK, unpublished data). Such transfected cell lines will be valuable tools for producing new serological reagents to MHC molecules, evaluating previous serological specificities and as markers in immunoprecipitation studies with cells from the rams and their offspring. A range of monoclonal antibodies for class I molecules from rams 501 and 504 is currently under development with the use of the mentioned stable transfectants (Keith Ballingall, MRI, Edinburgh, UK, unpublished data). Such transfectants are also useful for studying the role of individual MHC molecules in the restriction of cytotoxic T cell responses of foreign antigens. Using such tools, it has been suggested that $CD8^+$ T cell responses to the intracellular parasite of cattle *Theileria parva* can be restricted by

only one bovine MHC class I allele in a haplotype, a phenomenon known as immunodominance (Taracha et al. 1995).

Cytotoxic T cell responses against intracellular pathogens are restricted by MHC class I molecules on the basis of the presentation of short peptides (Zinkernagel and Doherty, 1974). Identification of potential cytotoxic T cell epitopes capable of inducing protective immune responses is therefore valuable in vaccine development. Transfected cell lines expressing ovine MHC class I molecules infected with an intracellular pathogen or co-transfected with genes expressing putative antigenic peptides from the pathogen will be useful tools for T cell epitope identification. As with studies carried out on MHC of human and cattle (Crawford et al. 2004; Doolan et al. 1996; Gaddum et al. 1996), sequence-specific peptide motifs could be identified for peptides bound to a number of ovine class I alleles.

8. General discussion

The number of MHC class I loci varies between different species or even between individuals of the same species and class I genes can be organized and expressed in a variety of ways (reviewed by Chardon et al. 2000; Ellis S. 2004; Kelley et al. 2005; Trowsdale J. 1995). Including pseudogenes, there are over 100 class I genes in the human extended class I region (The MHC sequencing consortium, 1999) and between 40 and 60 class I genes in rodents (Kulski et al. 2002), whereas there are only 21 in pigs (Chardon et al. 2000) and two in chicken (Kaufman et al. 1999). Since the bovine MHC class I region has yet to be sequenced and relatively few mapping studies have been carried out (Di Palma et al. 2002; Fries et al. 1986; McShane et al. 2001), the total number of class I genes in cattle remains unknown. It is estimated, however, that the bovine class I region is larger than that described for the pig and that it contains more class I genes (Ellis S. 2004). More recently, 24 genes have been mapped in the ovine class I region (Liu et al. 2006) and the authors suggest that the gene organization is similar to that of the swine MHC.

However, among the many genes identified in the class I region of different genomes, only one to three class I loci normally encode molecules involved in antigen presentation. There are three expressed classical class I loci in humans, gorillas and chimpanzees (A, B and C; IPD MHC database; http://www.ebi.ac.uk/ipd/mhc /dla/index.html), between one and three in pigs (SLA -1, -2 and -3; Renard et al. 2001), up to four in rats (RT1-A1, -A2, -A3 and –C/M; Walter et al. 2000), one in dogs (DLA-88; Kennedy et al. 1998) and one dominantly expressed class I gene in each chicken haplotype (B-F class I; Kaufman and Wallny, 1996). Although serological studies had suggested that the majority of MHC class I molecules in cattle were encoded by a single highly polymorphic locus (Ellis and Ballingall, 1999), molecular genetic studies have later provided evidence for at least six classical MHC class I loci in cattle, up to three of which are expressed in individual haplotypes (Ellis S. 2004). At the outset of this study, no molecular data were available on the number of expressed loci in the class I region of sheep, although serological evidence was consistent with the expression of two or three ovine class I loci (Cullen et al. 1982; Garrido et al. 1995; Jugo and Vicario, 2001; Jugo et al. 2002; Millot P. 1984; Stear and Spooner, 1981).

A total of 17 novel ovine MHC class I transcripts were identified from two heterozygous rams during the course of the present study (chapters 3, 4 and 5; Fig. 8.1). With

the exception of two (N1 and N2), these transcripts were assigned to haplotypes by genotyping progeny of the rams (chapter 4). Five class I sequences were identified in each of haplotypes 501A, 504C and 504D, while six class I transcripts were assigned to the 501B haplotype, suggesting the transcription of 5-6 ovine MHC class I loci from the haplotypes (Fig. 8.1). Subsequent evidence for transcription of sequences N1 and N2 by ram 504 (chapter 5) suggests that one or both of 504 haplotypes C and D contain at least one additional class I sequence, which raises the number of transcribed class I loci from these haplotypes to at least six. Phylogenetic analysis of ovine class I sequences with those of other bovid species also supported the existence of over six ovine MHC class I loci (chapter 6). The combination of phylogenetic analysis and haplotype data suggest that, excluding the partial transcripts, the class I sequences identified in the present study are grouped into five of these loci (Figs. 6.3 and 8.1). Four of these appear to be polymorphic (chapter 6), while the fifth locus contains highly divergent sequences bearing features associated with non classical class I genes (chapters 5 and 6). The partial transcripts E9, C7b, F7 and G12 were not included in the phylogenetic analysis, due to unavailability of sequence data for exons 4 though 8.

Despite the occurrence of multiple ovine class I loci, there is evidence for expression of no more than three class I loci from each haplotype (Fig. 8.1). Sequences D5, D3, C7, E2, F8 and F12, which accounted for 86% of the 360 clones analyzed by SSCP (chapter 3), were expressed in *Theileria*-infected cells from lambs homozygous for the relevant 501 and 504 haplotypes and, with the exception of E2, which was not transfected, were capable of expressing class I molecules on the surface of COS cells (chapter 7). Combined with phylogenetic analysis, these expression data suggest that D5, D3, F8 and F12 represent alleles of a single dominantly expressed classical ovine MHC class I locus, while C7 and E2 appear to be alleles at two additional polymorphic expressed class I loci (chapters 6 and 7).

Sequences C11, C6, G6, F9 and H10, which occurred in 8.61% of the clones screened by SSCP (chapter 3) were also found to be expressed in *Theileria*-infected cells (chapter 7). In addition, C11 and F9 transfected in COS cells were capable of expressing class I molecules on the cell surface (chapter 7). Sequence H10 consistently clustered with E2 in the phylogenetic analyses, as did the C11, C6, G6 and F9 sequences. However, the G6 sequence was only distantly associated with the remaining sequences in the cluster (chapter 6) and contained features strongly associated with non-classical genes (chapters 3 and 5). Therefore, G6 was considered to be a poorly expressed non-classical class I gene. However, the status of C11, C6

and F9 requires clarification in functional studies, because they are expressed at levels similar to those of the non classical class I genes, but appear nonetheless to be polymorphic, which is a characteristic found mainly in classical genes (chapters 5, 6 and 7).

Despite the lack of expression data for the partial transcripts E9, C7b, F7 and G12, as well as the N1 and N2 sequences transcribed by ram 504, F7 and G12 remain strong candidates for non-classical class I genes or pseudogenes; since they are shared between the four haplotypes and are almost certainly transcribed at a significantly lower level than the other transcripts in peripheral blood, based on the limited number of clones initially identified (0.56%) (chapters 3 and 4). In addition, sequencing features of N1 and N2 (chapter 5) and their localization deep in the phylogenetic trees, together with the putative non classical class I genes from cattle and nyala and close to the swine sequences (chapter 6) strongly suggest that these sequences also represent non-classical class I genes. A class I gene very similar to the swine AF014001 sequence has recently been identified in a BAC clone used for the construction of a physical map of the ovine class I region (Liu et al. 2006).

It therefore seems that haplotypes 501A and B each contain two highly expressed polymorphic classical class I loci along with a third locus expressed at low levels that requires further characterisation by functional analysis. Each also contains a pair of putative non-polymorphic class I like genes. These findings agree with previous serological and cDNA data suggesting the expression of two or three classical ovine MHC class I genes (Cullen et al. 1982; Garrido et al. 1995; Grossberger et al. 1990; Jugo and Vicario, 2001; Jugo et al. 2002; Millot P. 1984; Stear and Spooner, 1981). Two classical genes were also identified in the class I contig used for the construction of a BAC-based physical map of the ovine MHC complex (Liu et al. 2006). In contrast, haplotype 504C includes only a single highly expressed classical class I locus, together with the same putative non-classical genes, because G6 possibly represents a non classical locus and the E9 gene is predicted to be transcribed at only very low levels. Similar haplotypes with a single classical class I locus have been described for chicken (Kaufman et al. 1999), cattle (Ellis S. 2004), pig (Chardon et al. 2000) and inbred lines of rat (Joly et al. 1996). The 504D haplotype also contains only one putative highly expressed classical class I locus along with one additional classical locus expressed at lower levels and a poorly expressed locus with characteristics associated with neither classical nor non-classical genes. The pair of invariant genes transcribed in all haplotypes is also present. Two highly divergent class I-like genes are also transcribed by ram 504, although they have not been assigned to haplotypes.

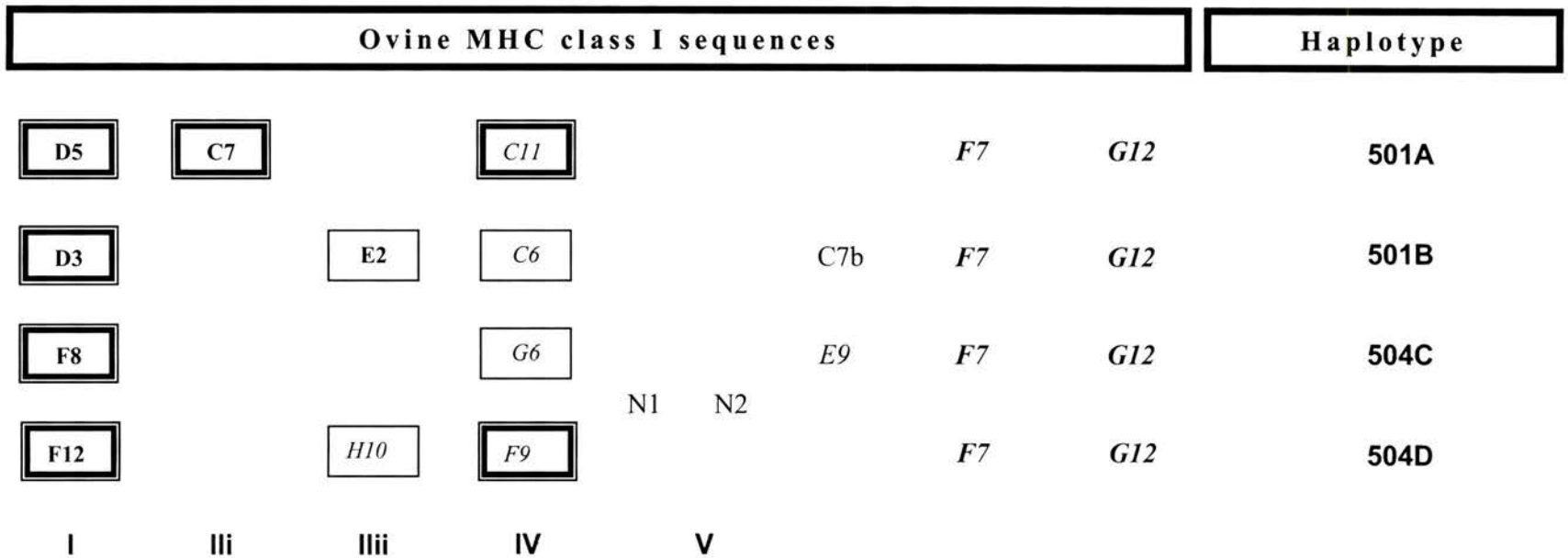


Figure 8.1: A diagrammatic representation of the ovine MHC class I sequences transcribed and expressed by four different haplotypes from rams 501 and 504. N1 and N2 were not assigned to individual haplotypes. Sequences in *bold, italics* and *bold italics* were highly transcribed, poorly transcribed and extremely poorly transcribed, as they were represented in 86%, 8.61% and 0.56% of the 360 clones analysed by SSCP. Sequences *boxed in a thick line* were expressed on the surface of COS-7 cells transfected with each sequence, while sequences *boxed in a thin line* were expressed by *Theileria*-infected cells from lambs homozygous for the relevant 501and 504 haplotypes. The expression of partial sequences C7b, E9, F7 and G12 was not explored due to lack of full length transcripts for transfection and enough sequence data for a robust Mascot analysis. The arrangement of the class I sequences reflects the clustering of the alleles according to the phylogenetic trees presented in Figs. 6.3, 6.4 and 6.5. The assignment of partial transcripts was not determined due to lack of sequence data for exons 4 through 8.

The combination of phylogenetic analysis, haplotype, transcription and expression data are therefore consistent with the existence of at least four distinct polymorphic ovine MHC class I loci, of which three are expressed in varying combinations and intensities in individual haplotypes, along with a pair of poorly transcribed non-polymorphic class I like loci and at least one additional divergent non classical class I locus (Fig. 8.1). Although there are no expression data for the five class I cDNA sequences available in public databases (M34673, M34674, U03092, AY188824, AY188825), these represent alleles of at least three additional loci to those identified in this study (chapter 6; Fig. 6.5). If these transcripts represent alleles of classical class I genes and there is no genetic interlocus recombination, the ovine class I region may be similar to that observed for cattle, with at least six classical loci of which up to three are expressed in a single haplotype (Ellis S. 2004). Alternatively, if sequences of cluster IV and the transcripts identified by other research groups are non classical, then the ovine MHC class I region may resemble that of swine, with three classical class I loci expressed in various combinations in different haplotypes and a large number of transcribed class I-like genes (Chardon et al. 2000). However, most of the latter genes are either truncated or contain stop codons that result in premature termination of the transcription (Chardon et al. 2000). In contrast, with the exception of G6, the above transcripts have a classical eight exon format (chapter 6). Similar observations of high numbers of putative non classical class I loci exhibiting limited polymorphism have been made for the equine MHC (Holmes and Ellis, 1999).

The degree of plasticity observed in the MHC class I region is indeed amazing (reviewed by Kelley et al. 2005; Kumanovics et al. 2003; Trowsdale J. 1995). It ranges from two genes in the chicken genome (Kaufman et al. 1999) to thousands in the genome of the African pigmy mouse *Nannomys setulosus* (Delarbre et al. 1992). Remarkable diversity has also been revealed within the class I region of individual species. Haplotype variation of class I genes is a feature observed in many species including primates (Cadavid et al. 1999), rodents (Joly et al. 1996), horse (Holmes and Ellis, 1999), pig (Chardon et al. 2000) and cattle, with the latter possibly exhibiting the greatest variability (Ellis S. 2004). The results obtained in the current study show that composition of class I loci also varies between different sheep haplotypes. Different numbers and organization of class I genes are believed to arise from rapid evolution of this region of mammalian genomes (Kelley et al. 2005). After divergence from a common ancestor, the MHC changes rapidly through gene duplications, deletions and mutations in a process known as birth and death evolution (Piontkivska and Nei, 2003). Such rapid diversification is believed to be pathogen-driven (Bernatchez and Landry, 2003; Hurt et al.

2004). Determination of whether ruminants exhibit an unusually high level of class I haplotype variability will require further studies not only in sheep and cattle, but also in other ruminant species. Such studies will provide insights into the evolution of vertebrate MHC genes.

Future Studies

Additional work is required to characterize further the four haplotypes transcribed by rams 501 and 504. Full length sequence data for transcripts E9, C7b, F7 and G12 would provide indications as to whether these incorporate an eight exon structure characteristic of functional class I molecules or, more possibly, mutations in conserved residues, deletions or stop codons indicative of diverse function or pseudogene status. Sequence data for the entire coding region of these partial transcripts and the N2 sequence would also allow phylogenetic analyses and transfection experiments to evaluate expression. Further experiments are also required to assign the N1 and N2 sequences to haplotypes using appropriately designed sequence-specific primers. Amplification of the putative non-classical class I genes from other sheep would confirm whether they are indeed non-polymorphic or exhibit limited polymorphism.

The haplotype data generated in this study have facilitated the establishment of a resource population of MHC-defined and MHC-homozygous sheep. These data provide an initial template for more extensive characterization of ovine MHC haplotypes and allelic diversity in different breeds and populations. This resource flock will also provide genetic material for comparative genomic analysis and disease association studies. Identification of alleles offering protection to infections could be integrated in appropriate breeding schemes for the sustainable control of sheep diseases.

A range of cellular and molecular tools to undertake in-depth analyses of $CD8^+$ T cell responses to infection by intracellular pathogens was generated and is still under development. These tools will be used for studying the role of specific ovine MHC loci in determining the specificity of an immune response and for T cell epitope identification. By determining allele-specific motifs of peptides eluted by ovine class I alleles, it will be possible to identify protein fragments important for the induction of $CD8^+$ T cell responses and peptides which stimulate protective responses against intracellular pathogens of sheep.

References

Abbas KA, Lichman HA, Pober SJ (2000) 'Cellular and Molecular Immunology.' (W.B. Saunders Company: London)

Aderem A, Underhill DM (1999) Mechanisms of phagocytosis in macrophages. *Annual Review of Immunology* **17**: 593-623.

Ainsworth PJ, Surh LC, Coulter-Mackie MB (1991) Diagnostic single strand conformational polymorphism, (SSCP): a simplified non-radioisotopic method as applied to a Tay-Sachs B1 variant. *Nucleic Acids Research.* **2**: 405-406.

Alfonso C, Karlsson L (2000) Nonclassical MHC class II molecules. *Annual Review of Immunology* **18**: 113-128.

Anderson KS, Alexander J, Wei M, Cresswell P (1993) Intracellular-transport of Class-I MHC molecules in antigen-processing mutant-cell lines. *Journal of Immunology* **151** (7): 3407-3419.

Andersson L, Lunden A, Sigurdardottir S, Davies CJ, Rask L (1988) Linkage relationships in the bovine MHC Region - High recombination frequency between Class-II subregions. *Immunogenetics* **27** (4): 273-280.

Arnason U, Adegoke JA, Bodin K, Born EW, Esa YB, Gullberg A, Nilsson M, Short RV, Xu XF, Janke A (2002) Mammalian mitogenomic relationships and the root of the eutherian tree. *Proceedings of the National Academy of Sciences of the United States of America* **99** (12): 8151-8156.

Arrieta-Aguirre I, Garcia-Etxebarria K, Jugo BM (2006) Optimization of the MHC Ovar-DRB1 gene typing. *Tissue Antigens* **67** (3): 222-228.

Ballingall KT, Wright H, Redmond J, Dutia BM, Hopkins J, Lang J, Deverson EV, Howard JC, Puri N, Haig D (1992) Expression and characterization of Ovine Major Histocompatibility Complex Class-II (OLA-DR) genes. *Animal Genetics* **23** (4): 347-359.

Ballingall KT, Dutia BM, Hopkins J, Wright H (1995) Analysis of the fine specificities of sheep Major Histocompatibility Complex class II-specific monoclonal- antibodies using mouse L-cell transfectants. *Animal Genetics* **26** (2): 79-84.

Banchereau J, Steinman RM (1998) Dendritic cells and the control of immunity. *Nature* **392** (6673): 245-252.

Band M, Larson JH, Womack JE, Lewin HA (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.

Beck S, Trowsdale J (1999) Sequence organisation of the class II region of the human MHC. *Immunological Reviews* **167**: 201-210.

Bennett SRM, Carbone FR, Karamalis F, Miller JFAP, Heath WR (1997) Induction of a CD8(+) cytotoxic T lymphocyte response by cross- priming requires cognate CD4(+) T cell help. *Journal of Experimental Medicine* **186** (1): 65-70.

Bensaid A, Kaushal A, Baldwin CL, Clevers H, Young JR, Kemp SJ, MacHugh ND, Toye PG, Teale AJ (1991) Identification of expressed bovine class-I MHC genes at 2 loci and demonstration of physical linkage. *Immunogenetics* **33** (4): 247-254.

Bernard P, Couturier M (1992) Cell killing by the F-plasmid Ccdb protein involves poisoning of DNA-topoisomerase-Ii complexes. *Journal of Molecular Biology* **226** (3): 735-745.

Bernard P, Kezdy KE, Vanmelderen L, Steyaert J, Wyns L, Pato ML, Higgins PN, Couturier M (1993) The F-plasmid ccdb protein induces efficient ATP-dependent DNA cleavage by gyrase. *Journal of Molecular Biology* **234** (3): 534-541.

Bernard P, Gabant P, Bahassi EM, Couturier M (1994) Positive-selection vectors using the F-plasmid ccdb killer gene. *Gene* **148** (1): 71-74.

Bernatchez L, Landry C (2003) MHC studies in nonmodel vertebrates: what have we learned about natural selection in 15 years? *Journal of Evolutionary Biology* **16** (3): 363-377.

Beutler B (2004) Innate immunity: an overview. *Molecular Immunology* **40** (12): 845-859.

Bevan MJ (1976) Cross-priming for a secondary cytotoxic response to minor H-antigens with H-2 congenic cells which do not cross-react in cytotoxic assay. *Journal of Experimental Medicine* **143** (5): 1283-1288.

Biron CA, Nguyen KB, Pien GC, Cousens LP, Salazar-Mather TP (1999) Natural killer cells in antiviral defense: Function and regulation by innate cytokines. *Annual Review of Immunology* **17**: 189-220.

Bjorkman PJ, Saper MA, Samraoui B, Bennett WS, Strominger JL, Wiley DC (1987a) The foreign antigen-binding site and T-cell recognition regions of Class-I Histocompatibility antigens. *Nature* **329** (6139): 512-518.

Bjorkman PJ, Saper MA, Samraoui B, Bennett WS, Strominger JL, Wiley DC (1987b) Structure of the human class-I Histocompatibility antigen, HLA- A2. *Nature* **329** (6139): 506-512.

Bjorkman PJ (1997) MHC restriction in three dimensions: A view of T cell receptor/ligand interactions. *Cell* **89** (2): 167-170.

Boshart M, Weber F, Jahn G, Dorschhasler K, Fleckenstein B, Schaffner W (1985) A very strong enhancer is located upstream of an immediate early gene of human cytomegalo-virus. *Cell* **41** (2): 521-530.

Boyson JE, Shufflebotham C, Cadavid LF, Urvater JA, Knapp LA, Hughes AL, Watkins DI (1996) The MHC class I genes of the rhesus monkey - Different evolutionary histories of MHC class I and II genes in primates. *Journal of Immunology* **156** (12): 4656-4665.

Boyum A (1968) Isolation of mononuclear cells and granulocytes from human blood. Isolation of monuclear cells by one centrifugation, and of granulocytes by combining centrifugation and sedimentation at 1 g. *Scandinavian Journal of Clinical and Laboratory Investigation* **97**: 77-89.

Braud VM, Allan DSJ, O'Callaghan CA, Soderstrom K, D'Andrea A, Ogg GS, Lazetic S, Young NT, Bell JI, Phillips JH, Lanier LL, McMichael AJ (1998) HLA-E binds to natural killer cell receptors CD94/NKG2A, B and C. *Nature* **391** (6669): 795-799.

Braud VM, Allan DSJ, McMichael AJ (1999) Functions of nonclassical MHC and non-MHC-encoded class I molecules. *Current Opinion in Immunology* **11** (1): 100-108.

Bretscher P, Cohn M (1970) A theory of self-nonself discrimination. *Science* **169** (3950): 1042-1049.

Briles WE, Stone HA, Cole RK (1977) Mareks-disease - effects of B-Histocompatibility alloalleles in resistant and susceptible chicken lines. *Science* **195** (4274): 193-195.

Brown JH, Jardetzky TS, Gorga JC, Stern LJ, Urban RG, Strominger JL, Wiley DC (1993) 3-Dimensional structure of the human Class-II Histocompatibility antigen HLA-DR1. *Nature* **364** (6432): 33-39.

Brown P, Spooner RL, Clark AJ (1989) Cloning and characterization of a BoLA class-I cDNA clone. *Immunogenetics* **29** (1): 58-60.

Brunt LM, Portnoy DA, Unanue ER (1990) Presentation of *Listeria monocytogenes* to $CD8^+$ T-cells requires secretion of hemolysin and intracellular bacterial-growth. *Journal of Immunology* **145** (11): 3540-3546.

Brusic V, Rudy G, Harrison LC (1994) MHCPEP - A database of MHC-binding peptides. *Nucleic Acids Research* **22** (17): 3663-3665.

Brusic V, Rudy G, Kyne AP, Harrison LC (1996) MHCPEP - A database of MHC-binding peptides: Update 1995. *Nucleic Acids Research* **24** (1): 242-244.

Brusic V, Rudy G, Kyne AP, Harrison LC (1997) MHCPEP, a database of MHC-binding peptides: Update 1996. *Nucleic Acids Research* **25** (1): 269-271.

Brusic V, Rudy G, Harrison LC (1998) MHCPEP, a database of MHC-binding peptides: update 1997. *Nucleic Acids Research* **26** (1): 368-371.

Bryant P, Ploegh H (2004) Class II MHC peptide loading by the professionals. *Current Opinion in Immunology* **16** (1): 96-102.

Buitkamp J, Filmether P, Stear MJ, Epplen JT (1996) Class I and class II major histocompatibility complex alleles are associated with faecal egg counts following natural, predominantly *Ostertagia circumcincta* infection. *Parasitology Research* **82** (8): 693-696.

Bunce M, Fanning GC, Welsh KI (1995) Comprehensive, serologically equivalent DNA typing for HLA-B by PCR using sequence-specific primers (PCR-SSP). *Tissue Antigens* **45** (2): 81-90.

Cadavid LF, Shufflebotham C, Ruiz FJ, Yeager M, Hughes AL, Watkins DI (1997) Evolutionary instability of the major histocompatibility complex class I loci in New World primates. *Proceedings of the National Academy of Sciences of the United States of America* **94** (26): 14536-14541.

Carbone FR, Kurts C, Bennett SRM, Miller JFAP, Heath WR (1998) Cross-presentation: A general mechanism for CTL immunity and tolerance. *Immunology Today* **19** (8): 368-373.

Carreno LJ, Gonzalez PA, Kalergis AM (2006) Modulation of T cell function by TCR/pMHC binding kinetics. *Immunobiology* **211** (1-2): 47-64.

Chai Z (2005) 'Identification and characterization of secreted proteins of Theileria lestoquardi schizonts. Thesis (Ph.D.) University of Edinburgh.'

Chambers CA (2001) The expanding world of co-stimulation: the two-signal model revisited. *Trends in Immunology* **22** (4): 217-223.

Chardon P, Kirszenbaum M, Cullen PR, Geffrotin C, Auffray C, Strominger JL, Cohen D, Vaiman M (1985) Analysis of the sheep MHC using HLA class I, II, and C4 cDNA probes. *Immunogenetics* **22** (4): 349-358.

Chardon P, Renard C, Vaiman M (1999) The major histocompatibility complex in swine. *Immunological Reviews* **167**: 179-192.

Chardon P, Renard C, Gaillard CR, Vaiman M (2000) The porcine Major Histocompatibility Complex and related paralogous regions: a review. *Genetics Selection Evolution* **32** (2): 109-128.

Chung C, Leib SR, Fraser DG, Ellis SA, McGuire TC (2003) Novel classical MHC class I alleles identified in horses by sequencing clones of reverse transcription-PCR products. *European Journal of Immunogenetics* **30** (6): 387-396.

Colosimo A, Goncz KK, Holmes AR, Kunzelmann K, Novelli G, Malone RW, Bennett MJ, Gruenert DC (2000) Transfer and expression of foreign genes in mammalian cells. *Biotechniques* **29** (2): 314-331.

Corradin G, Demotz S (1997) Peptide-MHC complexes assembled following multiple pathways: An opportunity for the design of vaccines and therapeutic molecules. *Human Immunology* **54** (2): 137-147.

Crawford AM, Buchanan FC, Swarbrick PA (1990) Ovine dinucleotide repeat polymorphism at the MAF18 locus. *Animal Genetics* **21** (4): 433-434.

Crawford F, Huseby E, White J, Marrack P, Kappler JW (2004) Mimotopes for alloreactive and conventional T cells in a peptide-MHC display library. *Plos Biology* **2** (4): 523-533.

Crew MD, Filipowsky ME, Neshat MS, Smith GS, Walford RL (1991) Transmembrane domain length variation in the evolution of Major Histocompatibility Complex class-I genes. *Proceedings of the National Academy of Sciences of the United States of America* **88** (11): 4666-4670.

Cullen PR, Bunch C, Brownlie J, Morris PJ (1982) Sheep lymphocyte antigens: a preliminary study. *Animal Blood Groups & Biochemical Genetics* **13** (3): 149-159.

Cullen PR, Millot P, Nguyen TC (1985) Sheep Histocompatibility antigens - A population-level comparison between lymphocyte antigens previously defined in France, England and Scotland, and sheep red-cell groups. *Animal Blood Groups and Biochemical Genetics* **16** (1): 19-34.

Cunha-Neto E (1999) MHC-restricted antigen presentation and recognition: constraints on gene, recombinant and peptide vaccines in humans. *Brazilian Journal of Medical and Biological Research* **32** (2): 199-205.

Davenport MP, Hill AVS (1996) Reverse immunogenetics: From HLA-disease associations to vaccine candidates. *Molecular Medicine Today* **2** (1): 38-45.

Davis MM, Bjorkman PJ (1988) T-cell antigen receptor genes and T-cell recognition. *Nature* **334** (6181): 395-402.

Debrick JE, Campbell PA, Staerz UD (1991) Macrophages as accessory cells for class-I MHC-restricted immune-responses. *Journal of Immunology* **147** (9): 2846-2851.

Delarbre C, Jaulin C, Kourilsky P, Gachelin G (1992) Evolution of the Major Histocompatibility Complex - A 100-fold amplification of MHC class-I genes in the african pygmy mouse *Nannomys-setulosus*. *Immunogenetics* **37** (1): 29-38.

den Haan JMM, Bevan MJ (2001) Antigen presentation to CD8(+) T cells: cross-priming in infectious diseases. *Current Opinion in Immunology* **13** (4): 437-441.

Deverson EV, Wright H, Watson S, Ballingall K, Huskisson N, Diamond AG, Howard JC (1991) Class-II Major Histocompatibility Complex genes of the sheep. *Animal Genetics* **22** (3): 211-225.

Di Palma F, Archibald SD, Young JR, Ellis SA (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.

Dietert R.R. (1996) The ovine major histocompatibility complex. In 'The major histocompatibility complex region of domestic animals'. (Eds Schook L.B. and Lamont S.J.) pp. 65-98. (CRC Press: London)

Diez-Tascon C, Keane OM, Wilson T, Zadissa A, Hyndman DL, Baird DB, Mcewan JC, Crawford AM (2005) Microarray analysis of selection lines from outbred populations to identify genes involved with nematode parasite resistance in sheep. *Physiological Genomics* **21** (1): 59-69.

Doolan DL, Wizel B, Hoffman SL (1996) Class I HLA-restricted cytotoxic T lymphocyte responses against malaria - Elucidation on the basis of HLA peptide binding motifs. *Immunologic Research* **15** (4): 280-305.

Douch PG, Outteridge PM (1989) The relationship between ovine lymphocyte antigens and parasitological and production parameters in Romney sheep. *International Journal for Parasitology*. (1): 35-41.

Dukkipati VSR, Blair HT, Garrick DJ, Murray A (2006) Ovar-MHC-Ovine Major Histocompatibility Complex: Role in genetic resistance to diseases. New *Zealand Veterinary Journal* **54** (4): 153-160.

Dutia BM, Hopkins J, Allington MP, Bujdoso R, Mcconnell I (1990) Characterization of monoclonal-antibodies specific for alpha- chains and beta-chains of sheep MHC class-II. *Immunology* **70** (1): 27-32.

Dutia BM, Mcconnell I, Bird K, Keating P, Hopkins J (1993) Patterns of Major Histocompatibility Complex class-II expression on T-cell subsets in different immunological compartments .1. Expression on resting T-cells. *European Journal of Immunology* **23** (11): 2882-2888.

Dutia BM, Mcconnell I, Ballingall KT, Keating P, Hopkins J (1994) Evidence for the expression of two distinct MHC class-II DR-beta- like molecules in the sheep. *Animal Genetics* **25** (4): 235-241.

Eckels DD (2000) MHC: function and implication on vaccine development. *Vox Sanguinis* **78** (Suppl 2): 265-267.

Ellis SA, Braem KA, Morrison WI (1992) Transmembrane and cytoplasmic domain sequences demonstrate at least two expressed bovine MHC class I loci. *Immunogenetics* **37** (1): 49-56.

Ellis SA, Martin AJ, Holmes EC, Morrison WI (1995) At least four 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 SA, Staines KA, Morrison WI (1996) cDNA sequence of cattle MHC class I genes transcribed in serologically defined haplotypes A18 and A31. *Immunogenetics* **43** (3): 156-159.

Ellis SA, Staines KA, Stear MJ, Hensen EJ, Morrison WI (1998) DNA typing for BoLA class I using sequence-specific primers (PCR-SSP). *European Journal of Immunogenetics* **25** (5): 365-370.

Ellis SA, Ballingall KT (1999) Cattle MHC: evolution in action? *Immunological Reviews* **167**: 159-168.

Ellis SA, Holmes EC, Staines KA, Smith KB, Stear MJ, McKeever DJ, MacHugh ND, Morrison WI (1999) Variation in the number of expressed MHC genes in different cattle class I haplotypes. *Immunogenetics* **50** (5-6): 319-328.

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

Erlich HA (1999) Principles and applications of the polymerase chain reaction. *Reviews in Immunogenetics* **1**: 127-134.

Escayg AP, Hickford JG, Montgomery GW, Dodds KG, Bullock DW (1996) Polymorphism at the ovine major histocompatibility complex class II loci. *Animal Genetics* **27** (5): 305-312.

Escayg AP, Hickford JGH, Bullock DW (1997) Association between alleles of the ovine major histocompatibility complex and resistance to footrot. *Research in Veterinary Science* **63** (3): 283-287.

Evans GA, Margulies DH, Cameriniotero RD, Ozato K, Seidman JG (1982) Structure and expression of a mouse Major Histocompatibility antigen gene, H-2LD. *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences* **79** (6): 1994-1998.

Fabb SA, Maddox JF, Gogolin-Ewens KJ, Baker L, Wu MJ, Brandon MR (1993) Isolation, characterization and evolution of ovine major histocompatibility complex class II DRA and DQA genes. *Animal Genetics* **24** (4): 249-255.

Fairhurst RM, Wang CX, Sieling PA, Modlin RL, Braun J (1998) CD1-restricted T cells and resistance to polysaccharide- encapsulated bacteria. *Immunology Today* **19** (6): 257-259.

Falk CS, Schendel DJ (1997) HLA-C revisited - Ten years of change. *Immunologic Research* **16** (2): 203-214.

Falk K, Rotzschke O, Stevanovic S, Jung G, Rammensee HG (1991) Allele-specific motifs revealed by sequencing of self-peptides e luted from MHC molecules. *Nature* **351** (6324): 290-296.

Fearon DT, Locksley RM (1996) Elements of immunity - The instructive role of innate immunity in the acquired immune response. *Science* **272** (5258): 50-54.

Feichtlbauer-Huber P, Stear MJ, Fries R, Buitkamp J (2000) Reference-strand-mediated conformation analysis of MHC alleles: a new method for high-resolution typing of the Ovar-DQB genes. *Immunogenetics* **51** (1): 65-68.

Ford C.H.J (1975) Genetic studies of sheep leucocyte antigens. *Journal of Immunogenetics* **2**: 31-40.

Frank MM, Fries LF (1991) The role of complement in inflammation and phagocytosis. *Immunology Today* **12** (9): 322-326.

French AR, Yokoyama WM (2003) Natural killer cells and viral infections. *Current Opinion in Immunology* **15** (1): 45-51.

Fries R, Hediger R, Stranzinger G (1986) Tentative chromosomal localization of the bovine Major Histocompatibility Complex by in situ hybridization. *Animal Genetics* **17** (4): 287-294.

Frohman MA, Martin GR (1989) Cut, paste, and save - new approaches to altering specific genes in mice. *Cell* **56** (2): 145-147.

Gaddum RM, Willis AC, Ellis SA (1996) Peptide motifs from three cattle MHC (BoLA) class I antigens. *Immunogenetics* **43** (4): 238-239.

Garber TL, Hughes AL, Watkins DI, Templeton JW (1994) Evidence for at least 3 transcribed BoLA class-I loci. *Immunogenetics* **39** (4): 257-265.

Garboczi DN, Ghosh P, Utz U, Fan QR, Biddison WE, Wiley DC (1996) Structure of the complex between human T-cell receptor, viral peptide and HLA-A2. *Nature* **384** (6605): 134-141.

Garcia KC, Degano M, Stanfield RL, Brunmark A, Jackson MR, Peterson PA, Teyton L, Wilson IA (1996) An alpha beta T cell receptor structure at 2.5 angstrom and its orientation in the TCR-MHC complex. *Science* **274** (5285): 209-219.

Garcia KC, Teyton L, Wilson LA (1999) Structural basis of T cell recognition. *Annual Review of Immunology* **17**: 369-397.

Garrido JJ, de Andres DF, Pintado CO, Llanes D, Stear MJ (1995) Serologically defined lymphocyte alloantigens in Spanish sheep. *Experimental & Clinical Immunogenetics* **12** (4): 268-271.

Gatesy J, Amato G, Vrba E, Schaller G, DeSalle R (1997) A cladistic analysis of mitochondrial ribosomal DNA from the bovidae. *Molecular Phylogenetics and Evolution* **7** (3): 303-319.

Gelhaus A, Forster B, Horstmann RD (1999a) Evidence for an additional cattle DQB locus. *Immunogenetics* **49** (10): 879-885.

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

Geraghty DE, Koller BH, Orr HT (1987) A human Major Histocompatibility Complex class-I gene that encodes a protein with a shortened cytoplasmic segment. *Proceedings of the National Academy of Sciences of the United States of America* **84** (24): 9145-9149.

Geraghty DE, Daza R, Williams LM, Vu Q, Ishitani A (2002) Genetics of the immune response: identifying immune variation within the MHC and throughout the genome. *Immunological Reviews* **190** (1): 69-85.

Gingerich PD (1977) Paleontology, phylogeny, and classification of mammals. *American Zoologist* **17** (4): 872.

Gluzman Y (1981) SV40-transformed simian cells support the replication of early SV40 mutants. *Cell* **23** (1): 175-182.

Gogolin-Ewens KJ, Mackay CR, Mercer WR, Brandon MR (1985) Sheep lymphocyte antigens (OLA). I. Major histocompatibility complex class I molecules. *Immunology* **56** (4): 717-723.

Goldrath AW, Bevan MJ (1999) Selecting and maintaining a diverse T-cell repertoire. *Nature* **402** (6759): 255-262.

Goodenow RS, Mcmillan M, Nicolson M, Sher BT, Eakle K, Davidson N, Hood L (1982) Identification of the class-I genes of the mouse Major Histocompatibility Complex by DNA-mediated gene-transfer. *Nature* **300** (5889): 231-237.

Grain F, Nain MC, Labonne MP, Lantier F, Lechopier P, Gebuhrer L, Asso J, Maddox J, Betuel H (1993) Restriction-fragment-length-polymorphism of DQB and DRB Class- II genes of the ovine Major Histocompatibility Complex. *Animal Genetics* **24** (5): 377-384.

Greenberg PD (1991) Adoptive T-cell therapy of tumors - mechanisms operative in the recognition and elimination of tumor-cells. *Advances in Immunology* **49**: 281-355.

Groh V, Steinle A, Bauer S, Spies T (1998) Recognition of stress-induced MHC molecules by intestinal epithelial gamma delta T cells. *Science* **279** (5357): 1737-1740.

Gromme M, Neefjes J (2002) Antigen degradation or presentation by MHC class I molecules via classical and non-classical pathways. *Molecular Immunology* **39** (3-4): 181-202.

Gross CN, Irrinki A, Feder JN, Enns CA (1998) Co-trafficking of HFE, a nonclassical major histocompatibility complex class I protein, with the transferrin receptor implies a role in intracellular iron regulation. *Journal of Biological Chemistry* **273** (34): 22068-22074.

Grossberger D, Hein W, Marcuz A (1990) Class I major histocompatibility complex cDNA clones from sheep thymus: alternative splicing could make a long cytoplasmic tail. *Immunogenetics* **32** (2): 77-87.

Grossberger D, Parham P (1992) Reptilian class-I Major Histocompatibility Complex genes reveal conserved elements in class-I structure. *Immunogenetics* **36** (3): 166-174.

Groth DM, Wetherall JD (1994) Dinucleotide repeat polymorphism within the ovine Major Histocompatibility Complex class-I region. *Animal Genetics* **25** (1): 61.

Gruszsynska J, Charon K, Swiderek W, Sawera M (2002) Microsatellite polymorphism in locus OMHC1 (MHC Class I) in Polish Health sheep and Polish Lowland Sheep (Zelazna variety). *Journal of Applied Genetics* **43** (2): 217-222.

Guild BC, Strominger JL (1984) Human and murine class I MHC antigens share conserved serine-335, the site of HLA phosphorylation in vivo. *Journal of Biological Chemistry* **259** (14): 9235-9240.

Gumperz JE, Brenner MB (2001) CD1-specific T cells in microbial immunity. *Current Opinion in Immunology* **13** (4): 471-478.

Gussow D, Rein RS, Meijer I, Dehoog W, Seemann GHA, Hochstenbach FM, Ploegh HL (1987) Isolation, expression, and the primary structure of HLA-Cw1 and HLA-Cw2 genes - Evolutionary aspects. *Immunogenetics* **25** (5): 313-322.

Gutierrez-Espeleta GA, Hedrick PW, Kalinowski ST, Garrigan D, Boyce WM (2001) Is the decline of desert bighorn sheep from infectious disease the result of low MHC variation? *Heredity* **86**: 439-450.

Hamerman JA, Ogasawara K, Lanier LL (2005) NK cells in innate immunity. *Current Opinion in Immunology* **17** (1): 29-35.

Harty JT, Tvinnereim AR, White DW (2000) CD8(+) T cell effector mechanisms in resistance to infection. *Annual Review of Immunology* **18**: 275-308.

Hawker KJ, Billadello JJ (1993) A multiplex polymerase chain reaction colony miniprep. *Biotechniques* **14** (5): 764-765.

Hayashi K, Yandell DW (1993) How sensitive is PCR-SSCP. *Human Mutation* **2** (5): 338-346.

Heath WR, Belz GT, Behrens GMN, Smith CM, Forehan SP, Parish IA, Davey GM, Wilson NS, Carbone FR, Villadangos JA (2004) Cross-presentation, dendritic cell subsets, and the generation of immunity to cellular antigens. *Immunological Reviews* **199** (1): 9-26.

Hediger R, Ansari HA, Stranzinger GF (1991) Chromosome-banding and gene localizations support extensive conservation of chromosome structure between cattle and sheep. *Cytogenetics and Cell Genetics* **57** (2-3): 127-134.

Hedrick PW (1998) Balancing selection and MHC. *Genetica* **104** (3): 207-214.

Herrmann LM, Brown WC, Lewis GS, Knowles DP (2005) Identification and phylogenetic analysis of 15 MHC class II DRB1 beta 1 expressed alleles in a ewe-lamb flock. *Immunogenetics* **57** (11): 855-863.

Hess M, Goldammer T, Gelhaus A, Ried K, Rappold G, Eggen A, Bishop MD, Schwerin M, Horstmann RD (1999) Physical assignment of the bovine MHC class IIa and class IIb genes. *Cytogenetics and Cell Genetics* **85** (3-4): 244-247.

Hickford JGH, Ridgway HJ, Escayg AP (2000) Evolution of the ovine MHC DQA region. *Animal Genetics* **31** (3): 200-205.

Higgins DG, Sharp PM (1989) Fast and sensitive multiple sequence alignments on a microcomputer. *Computer Applications in the Biosciences* **5** (2): 151-153.

Hill AVS, Allsopp CEM, Kwiatkowski D, Anstey NM, Twumasi P, Rowe PA, Bennett S, Brewster D, McMichael AJ, Greenwood BM (1991) Common west african HLA antigens are associated with protection from severe malaria. *Nature* **352** (6336): 595-600.

Hohenhaus MA, Outteridge PM (1995) The immunogenetics of resistance to Trichostrongylus colubriformis and Haemonchus contortus parasites in sheep. *British Veterinary Journal* **151** (2): 119-140.

Hohenhaus MA, East IJ, Eisemann CH, Pearson LD, Douch PG, Green RS, Outteridge PM (1995) Variation in immune responsiveness of sheep to the antigens of intestinal nematodes and blowfly larvae. *International Journal for Parasitology* **25** (5): 629-636.

Holmes EC, Ellis SA (1999) Evolutionary history of MHC class I genes in the mammalian order Perissodactyla. *Journal of Molecular Evolution* **49** (3): 316-324.

Holmes EC, Roberts AF, Staines KA, Ellis SA (2003) Evolution of major histocompatibility complex class I genes in Cetartiodactyls. *Immunogenetics* **55** (4): 193-202.

Hooshmandrad P, Magnusson U, Uggla A (1993) Some characteristics of ovine lymphoid-cells infected in vivo by *Theileria hirci* . *Parasitology Research* **79** (3): 195-199.

Hopkins J, Dutia BM, Mcconnell I (1986) Monoclonal-antibodies to sheep lymphocytes .1. Identification of MHC class-II molecules on lymphoid-tissue and changes in the level of class-II expression on lymph-borne cells following antigen stimulation in vivo. *Immunology* **59** (3): 433-438.

Hughes AL, Nei M (1988) Pattern of nucleotide substitution at Major Histocompatibility Complex class-I loci reveals overdominant selection. *Nature* **335** (6186): 167-170.

Hughes AL, Nei M (1989) Evolution of the Major Histocompatibility Complex - Independent origin of nonclassical class-I genes in different groups of mammals. *Molecular Biology and Evolution* **6** (6): 559-579.

Hurt P, Walter L, Sudbrak R, Klages S, Muller I, Shiina T, Inoko H, Lehrach H, Gunther E, Reinhardt R, Himmelbauer H (2004) The genomic sequence and comparative analysis of the rat major histocompatibility complex. *Genome Research* **14** (4): 631-639.

Huseby ES, White J, Crawford F, Vass T, Becker D, Pinilla C, Marrack P, Kappler JW (2005) How the T cell repertoire becomes peptide and MHC specific. *Cell* **122** (2): 247-260.

Janeway CA (1993) How the immune-system recognizes invaders. *Scientific American* **269** (3): 72-79.

Janeway CA, Travers P, Walport M, Shlomchik M (2001) 'Immunology; The Immune System in Health and Disease.' (Garland Publishing: New York)

Janeway CA, Medzhitov R (2002) Innate immune recognition. *Annual Review of Immunology* **20**: 197-216.

Jarrell VL, Lewin HA, Da Y, Wheeler MB (1995) Gene-centromere mapping of bovine DYA, DRB3, and PRL using secondary oocytes and first polar bodies - evidence for 4- strand double crossovers between DYA and DRB3. *Genomics* **27** (1): 33-39.

Jenkins MK, Khoruts A, Ingulli E, Mueller DL, McSorley SJ, Reinhardt RL, Itano A, Pape KA (2001) In vivo activation of antigen-specific CD4 T cells. *Annual Review of Immunology* **19**: 23-45.

Jin L, Nei M (1990) Limitations of the evolutionary parsimony method of phylogenetic analysis. *Molecular Biology and Evolution* **7** (1): 82-102.

Joly E, Leong L, Coadwell WJ, Clarkson C, Butcher GW (1996) The rat MHC haplotype RT1(c) expresses two classical class I molecules. *Journal of Immunology* **157** (4): 1551-1558.

Jugo BM, Vicario A (2000) Single-strand conformational polymorphism and sequence polymorphism of MHC-DRB in Latxa and Karrantzar sheep: implications for Caprinae phylogeny. *Immunogenetics* **51** (11): 887-897.

Jugo BM, Vicario A (2001) Lymphocyte antigens in sheep: linkage to the MHC class II DRB1 gene. *European Journal of Immunogenetics* **28** (4): 451-458.

Jugo BM, Joosten I, Grosfeld-Stulemeyer M, Amorena B, Hensen EJ (2002) Immunoprecipitation and isoelectric focusing of sheep MHC class I antigens reveal higher complexity than serology. *European Journal of Immunogenetics* **29** (5): 391-399.

Jugo BM, Rodriguez JA, Lopez-Otsoa F, Mazon LI, Vicario A (2004) Production and evaluation of alloantibodies against sheep MHC class I antigens. *Comparative Immunology, Microbiology & Infectious Diseases* **27** (2): 105-115.

Jukes TH, Cantor CR (1969) Evolution of protein molecules. In 'Mamalian protein metabolism'. (Ed. HN Munro) pp. 21-132. (Academic: New York)

Kappler JW, Roehm N, Marrack P (1987) T cell tolerance by clonal elimination in the thymus. *Cell* **49** (2): 273-280.

Karlin S, Altschul SF (1990) Methods for assessing the statistical significance of molecular sequence features by using general scoring schemes. *Proceedings of the National Academy of Sciences of the United States of America* **87** (6): 2264-2268.

Karlin S, Altschul SF (1993) Applications and statistics for multiple high-scoring segments in molecular sequences. *Proceedings of the National Academy of Sciences of the United States of America* **90** (12): 5873-5877.

Kaufman J, Volk H, Wallny HJ (1995) A minimal-essential-MHC and an unrecognized-MHC - 2 extremes in selection for polymorphism. *Immunological Reviews* **143**: 63-88.

Kaufman J, Wallny HJ (1996) Chicken MHC molecules, disease resistance and the evolutionary origin of birds. *Immunology and Developmental Biology of the Chicken* **212**: 129-141.

Kaufman J, Milne S, Gobel TWF, Walker BA, Jacob JP, Auffray C, Zoorob R, Beck S (1999) The chicken B locus is a minimal essential major histocompatibility complex. *Nature* **401** (6756): 923-925.

Ke Y, Kapp JA (1996) Receptor-mediated uptake delivers exogenous protein antigens into the MHC class I processing pathway in B cells. *Faseb Journal* **10** (6): 1749.

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

Kennedy LJ, Altet L, Angles JM, Barnes A, Carter SD, Francino O, Gerlach JA, Happ GM, Ollier WER, Polvi A, Thomson W, Wagner JL (1999) Nomenclature for factors of the dog major histocompatibility system (DLA), 1998. First report of the ISAG DLA Nomenclature Committee. *Tissue Antigens* **54** (3): 312-321.

Kidd P (2003) Th/Th2 Balance:The hypothesis, its limitations, and implications for health and disease. *Alternative Medicine Review* **8** (3): 223-243.

Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution* **16** (2): 111-120.

Kisielow P, von Boehmer H (1995) Development and selection of T-cells - facts and puzzles. *Advances in Immunology* **58**: 87-209.

Klein J, Bontrop RE, Dawkins RL, Erlich HA, Gyllensten UB, Heise ER, Jones PP, Parham P, Wakeland EK, Watkins DI (1990) Nomenclature for the major histocompatibility complexes of different species: a proposal. *Immunogenetics* **31** (4): 217-9.

Klug WS, Cummings MR (2000) 'Concepts of Genetics.' (Prentice Hall: New Jersey)

Koff RS (2002) Immunogenicity of hepatitis B vaccines: implications of immune memory. *Vaccine* **20** (31-32): 3695-3701.

Konnai S, Nagaoka Y, Takeshima S, Onuma M, Aida Y (2003) Sequences and diversity of 17 new Ovar-DRB1 alleles from three breeds of sheep. *European Journal of Immunogenetics* **30** (4): 275-282.

Kostia S, Kantanen J, Kolkkala M, Varvio SL (1998) Applicability of SSCP analysis for MHC genotyping: fingerprinting of Ovar-DRB1 exon 2 alleles from Finnish and Russian breeds. *Animal Genetics* **29** (6): 453-455.

Kozak M (1987) An analysis of 5'-noncoding sequences from 699 vertebrate messenger-RNAs. *Nucleic Acids Research* **15** (20): 8125-8148.

Kozak M (1990) Downstream secondary structure facilitates recognition of initiator codons by eukaryotic ribosomes. *Proceedings of the National Academy of Sciences of the United States of America* **87** (21): 8301-8305.

Krangel MS (1986) Secretion of HLA-A and HLA-B antigens via an alternative RNA splicing pathway. *Journal of Experimental Medicine* **163** (5): 1173-1190.

Kulski JK, Shiina T, Anzai T, Kohara S, Inoko H (2002) Comparative genomic analysis of the MHC: the evolution of class I duplication blocks, diversity and complexity from shark to man. *Immunological Reviews* **190** (1): 95-122.

Kumanovics A, Takada T, Lindahl KF (2003) Genomic organization of the mammalian Mhc. *Annual Review of Immunology* **21**: 629-657.

Kumar S, Tamura K, Jakobsen IB, Nei M (2001) MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* **17** (12): 1244-1245.

Lauemoller SL, Holm A, Hilden J, Brunak S, Nissen MH, Stryhn A, Pedersen LO, Buus S (2001) Quantitative predictions of peptide binding to MHC class I molecules using specificity matrices and anchor-stratified calibrations. *Tissue Antigens* **57** (5): 405-414.

Lehner B, Semple JI, Brown SE, Counsell D, Campbell RD, Sanderson CM (2004) Analysis of a high-throughput yeast two-hybrid system and its use to predict the function of intracellular proteins encoded within the human MHC class III region. *Genomics* **83** (1): 153-167.

Lemonnier FA, Malissen M, Golstein P, Lebouteiller P, Rebai N, Damotte M, Birnbaum D, Caillol D, Trucy J, Jordan BR (1982) Expression of human class-I Histocompatibility antigens at the surface of DNA-transformed mouse L-cells. *Immunogenetics* **16** (4): 355-361.

Lew AM, Mccluskey J, Maloy WL, Margulies DH, Coligan JE (1987) Multiple class-I molecules generated from single genes by alternative splicing of pre-messenger-RNAs. *Immunologic Research* **6** (1-2): 117-132.

Lewin HA (1996) Genetic organisation, polymorphism, and function of the bovine major histocompatibility complex. In 'The Major Histocompatibility Complex Region of Dostic Animals'. (Eds Schook L.B. and Lamont S.J.) pp. 65-98. (CRC Press: New York)

Lewin HA, Russell GC, Glass EJ (1999) Comparative organization and function of the major histocompatibility complex of domesticated cattle. *Immunological Reviews* **167**: 145-58.

Liang L, Sha WC (2002) The right place at the right time: novel B7 family members regulate effector T cell responses. *Current Opinion in Immunology* **14** (3): 384-390.

Lindner P, Bauer K, Kremmer E, Krebber C, Honegger A, Klinger B, Mocikat R, Pluckthun A (1997) Specific detection of His-tagged proteins with recombinant anti-His tag scFv-phosphatase or scFv-phage fusions. *Biotechniques* **22** (1): 140-149.

Liu HB, Liu K, Wang JF, Ma RLZ (2006) A BAC clone-based physical map of ovine major histocompatibility complex. *Genomics* **88** (1): 88-95.

Lobashevsky AL, Thomas JM (2000) Six Mamu-A locus alleles defined by a polymerase chain reaction sequence specific primer method. *Human Immunology* **61** (10): 1013-1020.

Lombard Z, Brune AE, Hoal EG, Babb C, van Helden PD, Epplen JT, Bornman L (2006) HLA class II disease associations in southern Africa. *Tissue Antigens* **67** (2): 97-110.

Mackay C (1988) Sheep leukocyte molecules: a review of their distribution, structure and possible function. *Veterinary Immunology & Immunopathology* (1): 1-20.

Madden DR (1995) The 3-dimensional structure of peptide-MHC complexes. *Annual Review of Immunology* **13**: 587-622.

Maddox JF, Davies KP, Crawford AM, Hulme DJ, Vaiman D, Cribiu EP et al. (2006) An Enhanced Linkage Map of the Sheep Genome Comprising More Than 1000 Loci. *Genome Research* **11**: 1275-1289.

Mahdy EA, Makinen A, Chowdhary BP, Andersson L, Gustavsson I (1989) Chromosomal localization of the ovine Major Histocompatibility Complex (OLA) by in situ hybridization. *Hereditas* **111** (1): 87-90.

Makala LHC, Nagasawa H (2002) Dendritic cells: A specialized complex system of antigen presenting cells. *Journal of Veterinary Medical Science* **64** (3): 181-193.

Malissen M, Malissen B, Jordan BR (1982) Exon intron organization and complete nucleotide-sequence of an HLA gene. *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences* **79** (3): 893-897.

Marx JL (1988) What T-cells see and how they see it. *Science* **242** (4880): 863-865.

Mathis D, Benoist C (2004) Back to central tolerance. *Immunity* **20** (5): 509-516.

Matthee CA, Davis SK (2001) Molecular insights into the evolution of the family Bovidae: A nuclear DNA perspective. *Molecular Biology and Evolution* **18** (7): 1220-1230.

Maxwell IH, Harrison GS, Wood WM, Maxwell F (1989) A DNA cassette containing a trimerized SV40 polyadenylation signal which efficiently blocks spurious plasmid-initiated transcription. *Biotechniques* **7** (3): 276-280.

Mazzaccaro RJ, Gedde M, Jensen ER, van Santen HM, Ploegh HL, Rock KL, Bloom BR (1996) Major histocompatibility class I presentation of soluble antigen facilitated by Mycobacterium tuberculosis infection. *Proceedings of the National Academy of Sciences of the United States of America* **93** (21): 11786-11791.

Mccutcheon JA, Gumperz J, Smith KD, Lutz CT, Parham P (1995) Low HLA-C expression at cell-surfaces correlates with increased turnover of heavy-chain messenger-RNA. *Journal of Experimental Medicine* **181** (6): 2085-2095.

McDevitt HO (2000) Discovering the role of the major histocompatibility complex in the immune response. *Annual Review of Immunology* **18**: 1-17.

McHeyzer-Williams MG (2003) B cells as effectors. *Current Opinion in Immunology* **15** (3): 354-361.

McPherson MJ, Hames BD, Taylor GR (1995) 'PCR 2 A Practical Aproach.' (Oxford University Press: Oxford)

McShane RD, Gallagher DS, Newkirk H, Taylor JF, Burzlaff JD, Davis SK, Skow LC (2001) Physical localization and order of genes in the class I region of the bovine MHC. *Animal Genetics* **32** (5): 235-239.

Medzhitov R, Janeway CA (1998) An ancient system of host defense. *Current Opinion in Immunology* **10** (1): 12-15.

Medzhitov R, Janeway CJ (2000) Advances in immunology: Innate immunity. *New England Journal of Medicine* **343** (5): 338-344.

Messaoudi I, Guevara-Patino JA, Dyall R, LeMaoult J, Nikolich-Zugich J (2002) Direct link between MHC polymorphism, T cell avidity, and diversity in immune defense. *Science* **298** (5599): 1797-1800.

Miceli MC, Parnes JR (1993) Role of CD4 and CD8 in T-cell activation and differentiation. *Advances in Immunology* **53**: 59-122.

Middleton D (1999) History of DNA typing for the humane MHC. *Reviews in Immunogenetics* **1**: 135-156.

Millot P (1978) The major histocompatibility complex of sheep (OLA) and two minor loci. *Animal Blood Groups & Biochemical Genetics* **9** (2): 115-21.

Millot P (1984) The OLA major histocompatibility complex of sheep. Study of six new factors and evidence of a third locus of the complex: OLA-C. *Experimental & Clinical Immunogenetics* **1** (1): 31-42.

Mitchell DA, Nair SK, Gilboa E (1998) Dendritic cell macrophage precursors capture exogenous antigen for MHC class I presentation by dendritic cells . *European Journal of Immunology* **28** (11): 3891.

Mond JJ, Lees A, Snapper CM (1995) T-cell-independent antigens type-2. *Annual Review of Immunology* **13**: 655-692.

Mosmann TR, Cherwinski H, Bond MW, Giedlin MA, Coffman RL (1986) 2 types of murine helper T-cell clone .1. Definition according to profiles of lymphokine activities and secreted proteins. *Journal of Immunology* **136** (7): 2348-2357.

Murphy WJ, Eizirik E, Johnson WE, Zhang YP, Ryderk OA, O'Brien SJ (2001) Molecular phylogenetics and the origins of placental mammals. *Nature* **409** (6820): 614-618.

Murray JS (1998) How the MHC selects Th1/Th2 immunity. *Immunology Today* **19** (4): 157-163.

Nagaoka Y, Kabeya H, Onuma M, Kasai N, Okada K, Aida Y (1999) Ovine MHC class II DRB1 alleles associated with resistance or susceptibility to development of bovine leukemia virus-induced ovine lymphoma. *Cancer Research* **59** (4): 975-981.

Nei M, Kumar S (2000) 'Molecular evolution and phylogenetics.' (Oxford University Press: Oxford)

Nelson JA, Reynoldskohler C, Smith BA (1987) Negative and positive regulation by a short segment in the 5'-flanking region of the human Cytomegalovirus Major immediate-early gene. *Molecular and Cellular Biology* **7** (11): 4125-4129.

Nesse LL, Larsen HJ (1987) Lymphocyte antigens in Norwegian goats - serological and genetic-studies. *Animal Genetics* **18** (3): 261-268.

Newton CR, Graham A, Heptinstall LE, Powell SJ, Summers C, Kalsheker N, Smith JC, Markham AF (1989) Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). *Nucleic Acids Research* **17** (7): 2503-2516.

O'Garra A, Arai N (2000) The molecular basis of T helper 1 and T helper 2 cell differentiation. *Trends in Cell Biology* **10** (12): 542-550.

Orita M, Iwahana H, Kanazawa H, Hayashi K, Sekiya T (1989) Detection of polymorphisms of human DNA by gel-electrophoresis as single-strand conformation polymorphisms. *Proceedings of the National Academy of Sciences of the United States of America* **86** (8): 2766-2770.

Outteridge PM, Windon RG, Dineen JK (1985) An association between a lymphocyte antigen in sheep and the response to vaccination against the parasite Trichostrongylus colubriformis. *International Journal for Parasitology* **15** (2): 121-127.

Outteridge PM, Windon RG, Dineen JK, Smith EF (1986) The relationship between ovine lymphocyte antigens and faecal egg count of sheep selected for responsiveness to vaccination against Trichostrongylus colubriformis. *International Journal for Parasitology* **16** (4): 369-374.

Outteridge PM, Windon RG, Dineen JK (1988) An ovine lymphocyte antigen marker for acquired resistance to Trichostrongylus colubriformis. *International Journal for Parasitology* **18** (6): 853-858.

Parham P, Lawlor DA, Lomen CE, Ennis PD (1989) Diversity and diversification of HLA-A,B,C alleles. *Journal of Immunology* **142** (11): 3937-3950.

Parham P, Adams EJ, Arnett KL (1995) The origins of HLA-A,B,C polymorphism. *Immunological Reviews* **143**: 141-180.

Parker DC (1993) T-cell - dependent B-cell activation. *Annual Review of Immunology* **11**: 331-360.

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

Paterson S, Wilson K, Pemberton JM (1998) Major histocompatibility complex variation associated with juvenile survival and parasite resistance in a large unmanaged ungulate population (Ovis aries L.). *Proceedings of the National Academy of Sciences of the United States of America* **95** (7): 3714-3719.

Pichowski JS, Ellis SA, Morrison WI (1996) Sequence of two cattle MHC class I cDNAs associated with BoLA A10 specificity. *Immunogenetics* **43** (4): 253-4.

Piontkivska H, Nei M (2003) Birth-and-death evolution in primate MHC class I genes: Divergence time estimates. *Molecular Biology and Evolution* **20** (4): 601-609.

Podolsky DK (1999) Mucosal Immunity and Inflammation V. Innate mechanisms of mucosal defense and repair: the best offense is a good defense. *American Journal of Physiology-Gastrointestinal and Liver Physiology* **277** (3): 495-499.

Puri NK, Mackay CR, Brandon MR (1985) Sheep lymphocyte antigens (OLA). II. Major histocompatibility complex class II molecules. *Immunology* **56** (4): 725-33.

Puri NK, Gogolin-Ewens KJ, Brandon MR (1987a) Monoclonal antibodies to sheep MHC class I and class II molecules: biochemical characterization of three class I gene products and four distinct subpopulations of class II molecules. *Veterinary Immunology & Immunopathology* **15** (1-2): 59-86.

Puri NK, Gorrell MD, Brandon MR (1987b) Sheep MHC class-II molecules .1. Immunochemical characterization. *Immunology* **62** (4): 567-573.

Puri NK, Brandon MR (1987) Sheep MHC class II molecules. II. Identification and characterization of four distinct subsets of sheep MHC class II molecules. *Immunology* **62** (4): 575-580.

Reche PA, Zhang H, Glutting JP, Reinherz EL (2005) EPIMHC: a curated database of MHC-binding peptides for customized computational vaccinology. *Bioinformatics* **21** (9): 2140-2141.

Reina-San-Martin B, Cosson A, Minoprio P (2000) Lymphocyte polyclonal activation: A pitfall for vaccine design against infectious agents. *Parasitology Today* **16** (2): 62-67.

Renard C, Vaiman M, Chiannilkulchai N, Cattolico L, Robert C, Chardon P (2001) Sequence of the pig major histocompatibility region containing the classical class I genes. *Immunogenetics* **53** (6): 490-500.

Rock KL, Rothstein L, Gamble S, Fleischacker C (1993) Characterization of antigen-presenting cells that present exogenous antigens in association with class-I MHC molecules. *Journal of Immunology* **150** (2): 438-446.

Rock KL (1996) A new foreign policy: MHC class I molecules monitor the outside world. *Immunology Today* **17** (3): 131-137.

Rozen S, Skaletsky HJ (2000) Primer3 on the WWW for general users and for biologist programmers. In 'Bioinformatics Methods and Protocols: Methods in Molecular Biology'. (Eds S Krawetz and S Misener) pp. 365-386. (Humana Press: New Jersey)

Rudensky AY, Prestonhurlburt P, Hong SC, Barlow A, Janeway CA (1991) Sequence-analysis of peptides bound to MHC class-II molecules. *Nature* **353** (6345): 622-627.

Ruohonen-Lehto MK, Renard C, Rothschild MF, Edfors-Lilja I, Kristensen B, Gustafsson U, Larson RG, Varvio SL (1998) Variable number of pig MHC class I genes in different serologically defined haplotypes identified by a 3 '-untranslated region probe. *Animal Genetics* **29** (3): 178-184.

Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, Erlich HA (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA-polymerase. *Science* **239** (4839): 487-491.

Saitou N, Nei M (1987) The Neighbor-Joining Method - A New Method for Reconstructing Phylogenetic Trees. *Molecular Biology and Evolution* **4** (4): 406-425.

Sallusto F, Cella M, Danieli C, Lanzavecchia A (1995) Dendritic cells use macropinocytosis and the mannose receptor to concentrate macromolecules in the Major Histocompatibility Complex class-Ii compartment - down-regulation by cytokines and bacterial products. *Journal of Experimental Medicine* **182** (2): 389-400.

Sambrook J, Fritsch FE, Maniatis T (1989) 'Molecular Cloning. A Laboratory Manual.' (Cold Spring Harbor Laboratory Press: New York)

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America* **74** (12): 5463-5467.

Sawhney SMS, Hasima NN, Glass EJ, Almurrani SWK, Nichani AK, Spooner RL, Williams JL, Russell GC (1995) Transfection, expression, and DNA-sequence of a gene encoding a BoLA-A11 antigen. *Immunogenetics* **41** (4): 246-250.

Sayers G, Good B, Hanrahan JP, Ryan M, Angles JM, Sweeney T (2005) Major Histocompatibility Complex DRB1 gene: its role in nematode resistance in Suffolk and Texel sheep breeds. *Parasitology* **131**: 403-409.

Schonbach C, Koh JLY, Sheng X, Wong L, Brusic V (2000) FIMM, a database of functional molecular immunology. *Nucleic Acids Research* **28** (1): 222-224.

Schonbach C, Koh JLY, Flower DR, Wong LS, Brusic V (2002) FIMM, a database of functional molecular immunology: update 2002. *Nucleic Acids Research* **30** (1): 226-229.

Schwaiger FW, Wayers E, Epplen C, Brun J, Crawford RA, Epplen JT (1993) The paradox pf MHC-DBR exon/intron evolution: α-helix and β-sheet encoding regions diverge while hypervariable intronic simple repeats coevolve with β-sheet codons

Schwaiger FW, Gostomski D, Stear MJ, Duncan JL, McKellar QA, Epplen JT, Buitkamp J (1995) An ovine major histocompatibility complex DRB1 allele is associated with low faecal egg counts following natural, predominantly *Ostertagia circumcincta* infection. *International Journal for Parasitology* **25** (7): 815-22.

Scott PC, Choi CL, Brandon MR (1987) Genetic organization of the ovine-MHC class-II region. *Immunogenetics* **25** (2): 116-122.

Scott PC, Gogolinewens KJ, Adams TE, Brandon MR (1991a) Nucleotide-sequence, polymorphism, and evolution of ovine MHC class-II DQA genes. *Immunogenetics* **34** (2): 69-79.

Scott PC, Maddox JF, Gogolinewens KJ, Brandon MR (1991b) The nucleotide-sequence and evolution of ovine MHC class-II b- genes - DQB and DRB. *Immunogenetics* **34** (2): 80-87.

Sebzda E, Mariathasan S, Ohteki T, Jones R, Bachmann MF, Ohashi PS (1999) Selection of the T cell repertoire. *Annual Review of Immunology* **17**: 829-874.

Sette A, Fikes J (2003) Epitope-based vaccines: an update on epitope identification, vaccine design and delivery. *Current Opinion in Immunology* **15** (4): 461-470.

Shams H (2005) Recent developments in veterinary vaccinology. *Veterinary Journal* **170** (3): 289-299.

Shao L, Kamalu O, Mayer L (2005) Non-classical MHC class I molecules on intestinal epithelial cells: mediators of mucosal crosstalk. *Immunological Reviews* **206**: 160-176.

Shuman S (1991) Recombination mediated by vaccinia virus-DNA topoisomerase-I in *Escherichia coli* is sequence specific. *Proceedings of the National Academy of Sciences of the United States of America* **88** (22): 10104-10108.

Shuman S (1994) Novel-approach to molecular-cloning and polynucleotide synthesis using vaccinia DNA topoisomerase. *Journal of Biological Chemistry* **269** (51): 32678-32684.

Singer DS, Cameriniotero RD, Satz ML, Osborne B, Sachs D, Rudikoff S (1982) Characterization of a porcine genomic clone encoding a Major Histocompatibility antigen - expression in mouse L-cells. *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences* **79** (5): 1403-1407.

Snary D, Barnstable CJ, Bodmer WF, Crumpton MJ (1977) Molecular-structure of human Histocompatibility antigens - HLA-C series. *European Journal of Immunology* **7** (8): 580-585.

Snibson KJ, Maddox JF, Fabb SA, Brandon MR (1998) Allelic variation of ovine MHC class II DQA1 and DQA2 genes. *Animal Genetics* **29** (5): 356-362.

Southern PJ, Berg P (1982) Transformation of mammalian cells to antibiotic resistance with a bacterial gene under control of the SV40 early region promoter. *Journal of Molecular Applied Genetics* **1** (4): 327-337.

Spellberg B, Edwards JE (2001) Type 1 type 2 immunity in infectious diseases. *Clinical Infectious Diseases* **32** (1): 76-102.

Spooner RL, Teale AJ, Cullen P (1988) The MHC of cattle and sheep. *Progress in Veterinary Microbiology & Immunology* **4**: 88-107.

Stear MJ, Spooner RL (1981) Lymphocyte antigens in sheep. *Animal Blood Groups & Biochemical Genetics* **12** (4): 265-276.

Stear MJ, Bairden K, Bishop SC, Buitkamp J, Epplen JT, Gostomski D, McKellar QA, Schwaiger FW, Wallace DS (1996) An ovine lymphocyte antigen is associated with reduced faecal egg counts in four-month-old lambs following natural, predominantly *Ostertagia circumcincta* infection. *International Journal for Parasitology* **26** (4): 423-428.

Stear MJ, Wakelin D (1998) Genetic resistance to parasitic infection. *Revue Scientifique et Technique de l Office International des Epizooties* **17** (1): 143-153.

Steinmetz M, Frelinger JG, Fisher D, Hunkapiller T, Pereira D, Weissman SM, Uehara H, Nathenson S, Hood L (1981a) 3 cDNA clones encoding mouse transplantation antigens - homology to immunoglobulin genes. *Cell* **24** (1): 125-134.

Steinmetz M, Moore KW, Frelinger JG, Sher BT, Shen FW, Boyse EA, Hood L (1981b) A pseudogene homologous to mouse transplantation antigens - transplantation antigens are encoded by 8 exons that correlate with protein domains. *Cell* **25** (3): 683-692.

Svanborg C, Godaly G, Hedlund M (1999) Cytokine responses during mucosal infections: role in disease pathogenesis and host defence. *Current Opinion in Microbiology* **2** (1): 99-105.

Tajima F, Nei M (1984) Estimation of evolutionary distance between nucleotide-sequences. *Molecular Biology and Evolution* **1** (3): 269-285.

Takeshima SN, Aida Y (2006) Structure, function and disease susceptibility of the bovine major histocompatibility complex. *Animal Science Journal* **77** (2): 138-150.

Tamura K (1992) Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G+C-content biases. *Molecular Biology and Evolution* **9** (4): 678-687.

Tamura K, Nei M (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.

Taracha ELN, Goddeeris BM, Teale AJ, Kemp SJ, Morrison WI (1995) Parasite strain specificity of bovine cytotoxic T-cell responses to *Theileria-parva* is determined primarily by immunodominance. *Journal of Immunology* **155** (10): 4854-4860.

The MHC sequencing consortium, Beck S, Geraghty D, Inoko H, Rowen L, Aguado B, Bahram S, Campbell RD, Forbes SA, Guillaudeux T, Hood L, Horton R, Janer M, Jasoni C, Madan A, Milne S, Neville M, Oka A, Qin S, Ribas-Despuig G, Rogers J, Shiina T, Spies T, Tamiya G, Tashiro H, Trowsdale J, Vu Q, Williams L, Yamazaki M (1999) Complete sequence and gene map of a human major histocompatibility complex. *Nature* **401** (6756): 921-923.

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* **22** (22): 4673-4680.

Thorsby E (1997) Invited anniversary review: HLA associated diseases. *Human Immunology* **53** (1): 1-11.

Tkacikova L, Bhide MR, Mikula I (2005) Asymmetric PCR-SSCP: a useful tool for detection of OLA-DRB1 (MHC class II) gene polymorphism in Slovak improved Valachian sheep. *Acta Veterinaria Brno* **74** (2): 275-+.

Toye PG, MacHugh ND, Bensaid AM, Alberti S, Teale AJ, Morrison WI (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.

Trowsdale J (1995) Both Man and Bird and Beast - Comparative Organization of Mhc Genes. *Immunogenetics* **41** (1): 1-17.

Trowsdale J, Parham P (2004) Mini-review: defense strategies and immunity-related genes. *European Journal of Immunology* **34** (1): 7-17.

Ulker N, Lewis KD, Hood LE, Stroynowski I (1990) Activated T-cells transcribe an alternatively spliced messenger-RNA encoding a soluble form of Qa-2 antigen. *Embo Journal* **9** (12): 3839-3847.

van der Merwe PA, Bodian DL, Daenke S, Linsley P, Davis SJ (1997) CD80 (B7-1) binds both CD28 and CTLA-4 with a low affinity and very fast kinetics. *Journal of Experimental Medicine* **185** (3): 393-403.

van Eijk MJT, Russ I, Lewin HA (1993) Order of bovine DRB3, DYA, and PRL determined by sperm typing. *Mammalian Genome* **4** (2): 113-118.

van Oorschot RA, Maddox JF, Adams LJ, Fabb SA (1994) Characterization and evolution of ovine MHC class II DQB sequence polymorphism. *Animal Genetics* **25** (6): 417-24.

von Boehmer H, Aifantis I, Gounari F, Azogui O, Haughn L, Apostolou I, Jaeckel E, Grassi F, Klein L (2003) Thymic selection revisited: how essential is it? *Immunological Reviews* **191** (1): 62-78.

Vos Q, Lees A, Wu ZQ, Snapper CM, Mond JJ (2000) B-cell activation by T-cell-independent type 2 antigens as an integral part of the humoral immune response to pathogenic microorganisms. *Immunological Reviews* **176**: 154-170.

Walter L, Gunther E (2000) Physical mapping and evolution of the centromeric class I gene-containing region of the rat MHC. *Immunogenetics* **51** (10): 829-837.

Windon R.G., Dineen J.K. (1984) Parasitological and immunological completence of lambs selected for high and low responsiveness to caccination with irradiated *Trichostrogylus colubriformis* larvae. In 'Immunogenetics Approaches to the control of endoparasites with particular reference to sheep'. (Eds JK Dineen and PM Outerridge) pp. 13-28. (CSIRO and Australian Wool Corporation: Melbourn)

Wright H, Ballingall KT, Redmond J (1994) The DY sub-region of the sheep MHC contains an A/B gene pair. *Immunogenetics* **40** (3): 230-234.

Wright H, Ballingall KT (1994) Mapping and characterization of the DQ subregion of the ovine MHC. *Animal Genetics* **25** (4): 243-249.

Wright H, Redmond J, Wright F, Ballingall KT (1995) The nucleotide-sequence of the sheep MHC class-II DNA gene. *Immunogenetics* **41** (2-3): 131-133.

Xie T, Rowen L, Aguado B, Ahearn ME, Madan A, Qin SZ, Campbell RD, Hood L (2003) Analysis of the gene-dense major histocompatibility complex class III region and its comparison to mouse. *Genome Research* **13** (12): 2621-2636.

Ye G, Barrera C, Fan XJ, Gourley WK, Crowe SE, Ernst PB, Reyes VE (1997) Expression of B7-1 and B7-2 costimulatory molecules by human gastric epithelial cells - Potential role in CD4(+) T cell activation during Helicobacter pylori infection. *Journal of Clinical Investigation* **99** (7): 1628-1636.

Yu CY, Yang ZY, Blanchong CA, Miller W (2000) The human and mouse MHC class III region: a parade of 21 genes at the centromeric segment. *Immunology Today* **21** (7): 320-328.

Zasloff M (1992) Antibiotic peptides as mediators of innate immunity. *Current Opinion in Immunology* **4** (1): 3-7.

Zhou JH, Ferencik S, Rebmann V, Yang DL, Lu M, Roggendorf M, Grosse-Wilde H (2003) Molecular genetic and biochemical analysis of Woodchuck (Marmota monax) MHC class I polymorphism. *Tissue Antigens* **61** (3): 240-248.

Zinkernagel RM, Doherty PC (1974) Restriction of in-vitro T cell-mediated cytotoxicity in lymphocytic choriomeningitis within a syngeneic or semiallogeneic system. *Nature* **248** (5450): 701-702.

Zinkernagel RM (2002) On cross-priming of MHC class I-specific CTL: rule or exception? *European Journal of Immunology* **32** (9): 2385-2392.

Appendix 1. Materials and Methods

Appendix 1.1: Solutions and gels

Bromophenol blue

Mix

1. 50-80% glycerol
2. 1XTE
3. 0.05% bromophenol blue

TE

Mix

1. 5ml 10mM Tris-Cl from 0.5M Tris-Cl
2. 1ml 1mM EDTA from 25mM EDTA

in 1lt distilled water.

0.5M Tris-HCl

Dissolve

1. 60.5gr Tris-Cl

in 1lt distilled water and adjust the pH to 8 with HCl.

25mM EDTA

Dissolve

1. 93.05g $Na_2EDTAX2H_2O$

in 1 lt distilled water and adjust the pH to 8 with NaOH

DEPC treated water

RNAase free water was produced by treating distilled water with 0.1% diethyl pyrocarbonate (DEPC). Following incubation at 37°C for at least 12 hours, the DEPC treated water was then autoclaved at 121°C for 15 min.

Formamide denaturing solution

Mix

1. 95% formamide
2. 25mM EDTA (as above)
3. 0.05% bromophenol blue
4. 0.05% xylene cyanol

and keep at 4°C.

Polyacrylamide gels

6% polyacrylamide gels for SSCP

1. 16ml 30% acryl/bisacrylamide
2. 20ml 5XTBE
3. 44ml distilled water
4. 320ml of 25% APS
5. 52.8ml TEMED

6% polyacrylamide gels for RFLP

1. 20ml 30% acryl/bisacrylamide
2. 10ml 10XTBE
3. 68ml distilled water
4. 2ml of 3% APS solution
5. 100ml TEMED

5% polyacrylamide gels containing 6M Urea for sequencing

1. 18g Urea
2. 5.68ml 30% acryl/bisacrylamide
3. 10ml 5XTBE
4. up to 50ml distilled water
5. 0.07g resin beans
6. 250µl 10% APS
7. 35µl TEMED

In all preparations APS and TEMED were added after all other ingredients were mixed.

Sequence loading dye

5:1 formamide:blue dye (25mM EDTA, 50mg/ml blue dextran, pH=8.0)

3M Sodium acetate

Dissolve

1. 408.2g sodium acetate

in 1lt distilled water and adjust the pH to 5.2.

10XTBE

Dissolve

1. 108g trisbase
2. 55g boric acid
3. 9.3g EDTA

in 1lt distilled water and adjust the pH to 8,3-8.7 at 50°C with HCl.

This stock solution was dissolved 1:10 to make 1XTBE working solution.

Appendix 1.2: Bacterial Grown Media

LB medium

Dissolve

1. 10g Bacto-tryptone
2. 5g Bacto-yeast extract
3. 5g NaCl

in 1lt distilled water and adjust the pH to 7.0 with NaOH

LB agar plates with ampicillin/IPTG/X-Gal

1g agar was added to 1lt LB medium and autoclaved. When the medium reached temperature of around 50°C, 100μg/ml ampicillin, 0.5mM IPTG and 80mg/ml X-Gal were added.

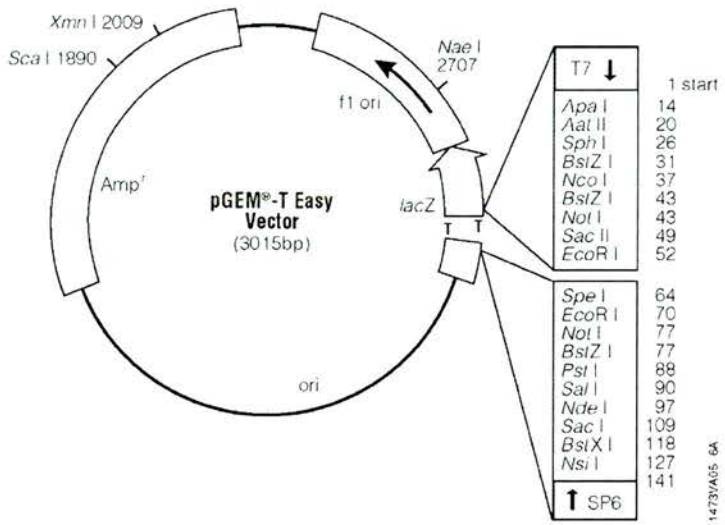
SOC medium

Dissolve

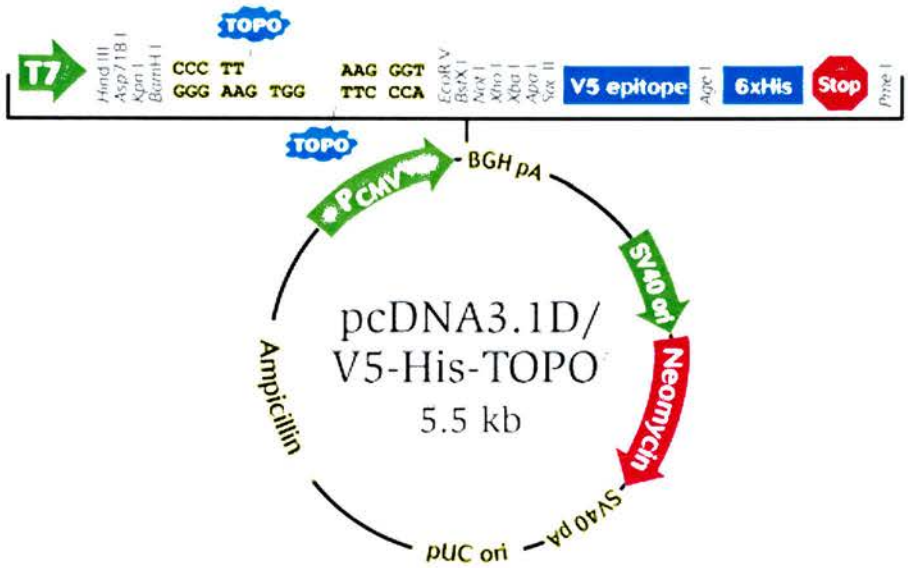
1. 2g Bacto-tryptone
2. 0.5g Bacto-yeast extract
3. 1ml 1M NaCl
4. 0.25ml 1M KCl
5. 1ml 2M Mg^{2+} stock, filter sterilized (20.33g $MgCl_2$ $6H_2O$ and 24.65g $MgSO_4$ $7H_2O$)
6. 1ml 2M glucose, filter sterilized

in 97ml distilled water and autoclave (all except the Mg^{2+} stock and the glucose). Then, add the Mg^{2+} stock and the glucose and bring the total volume to 100ml with sterile distilled water. Adjust the pH to 7.0.

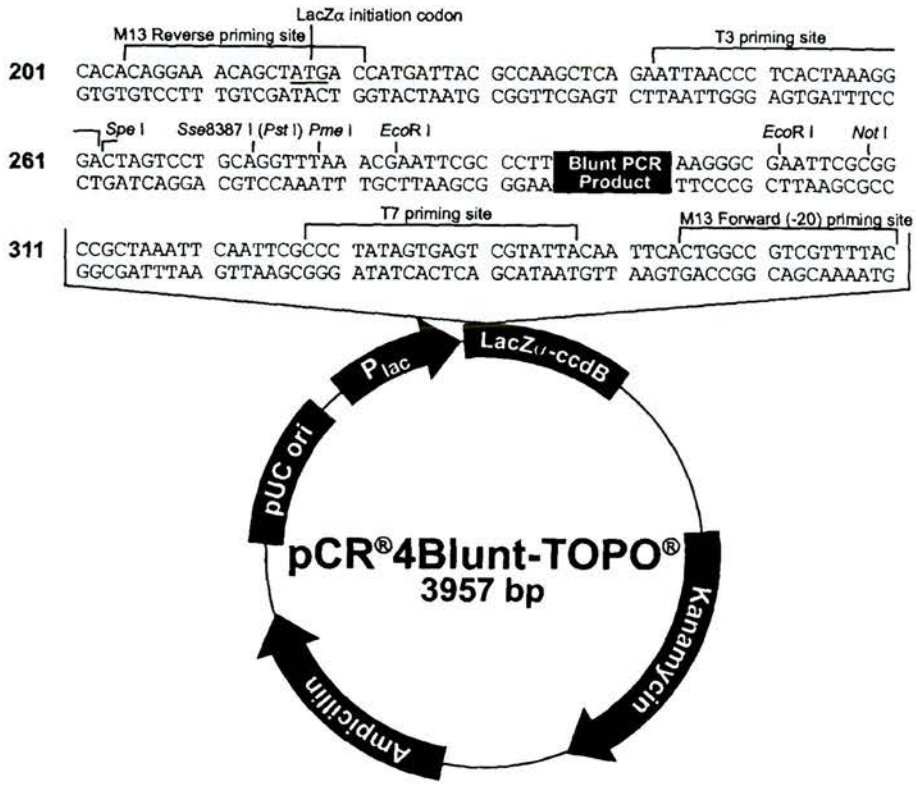
Appendix 1.3: Vectors



pGEM-T Easy Vector



pcDNA3.1D/V5-His-TOPO Vector



pCR4Blunt TOPO Vector

Appendix 1.4: Cell Culture Media

Complete IMDM medium

Mix

1. 500ml IMDM
2. 10ml Penicillin/Streptomycin solution
3. 10ml L-glutamine
4. 50ml heat inactivated foetal bovine serum (FBS)

Penicillin/Streptomycin solution

Dissolve

1. 1.2g benzypenicillin sodium and
2. 2g streptomycin sulphate

in 200ml distilled sterile water and filter under aseptic conditions, using a 0.2µm sterile membrane filter.

Keep at -20°C in 20ml aliquots.

L-glutamine

Dissolve

1. 7.3g L-glutamine

in 500ml distilled sterile water and filter under aseptic conditions, using a 0.2µm sterile membrane filter.

Keep at -20°C in 20ml aliquots.

Heat inactivated foetal bovine serum

Heat a 500ml bottle of FBS at 57°C for 30min.

Keep at -20°C in 20ml aliquots.

Tris-RPMI

Add

1. 25ml 1M Tris

in 500ml RPMI

1M Tris

Dissolve

1. 121.1g Tris base

in 1lt distilled water and adjust the pH to 7.2-7.4 with HCl.

Trypsin/versene solution

Mix

1. 4ml trypsin solution and
2. 16ml versene solution

under aseptic conditions and keep at 4°C for 2-3 weeks

Trypsin sollution

Dissolve

1. 4g NaCl
2. 0.38g KCl
3. 0.05g Na_2HPO_4
4. 0.5g Glucose
5. 1.5g Tris methylamine

in 350ml and adjust the pH to 7.6-7.8 in room temperature.

Then add

1. 148ml phenol/trypsin solution

and filter under aseptic conditions, using a 0.2μm sterile membrane filter.

Keep at -20°C in 20ml aliquots. Once defrosted, keep at 4°C for one month.

Phenol/trypsin solution

Dissolve

1. 7.5 ml 1% phenol red solution and
2. 12.5g Trypsin

in 1480ml distilled sterile water

Versene solution

Dissolve

1. 4.8g NaCl
2. 0.12g KCl
3. 0.69g $NaHPO_4$
4. 0.12g KH_2PO_4
5. 0.12g versene

6. 0.9ml 1% phenol red solution

in 600ml distilled sterile water and adjust the pH to 7.1-7.3.

Filter under aseptic conditions with a 0.2μm sterile filter membrane.

Keep at 4°C in 16ml aliquots.

PBS

Dissolve

1. 80g NaCl
2. 2g KCl
3. 11.5g Na_2HPO_4
4. 20g KH_2PO_4

in 1lt distilled water and autoclave at 121°C for 15 min

FACS buffer

Dissolve

1. 2% FBS
2. 0.02% sodium azide

in PBS and keep in fridge.

ILA 88 antibody

Dissolve

1. 200μl of ILA-88 1% stock solution

in 1000μl FACS buffer.

Goat anti-mouse IgG antibody

Dissolve

1. 20μl of 2mg/ml goat anti-mouse IgG

in 1000μl FACS buffer.

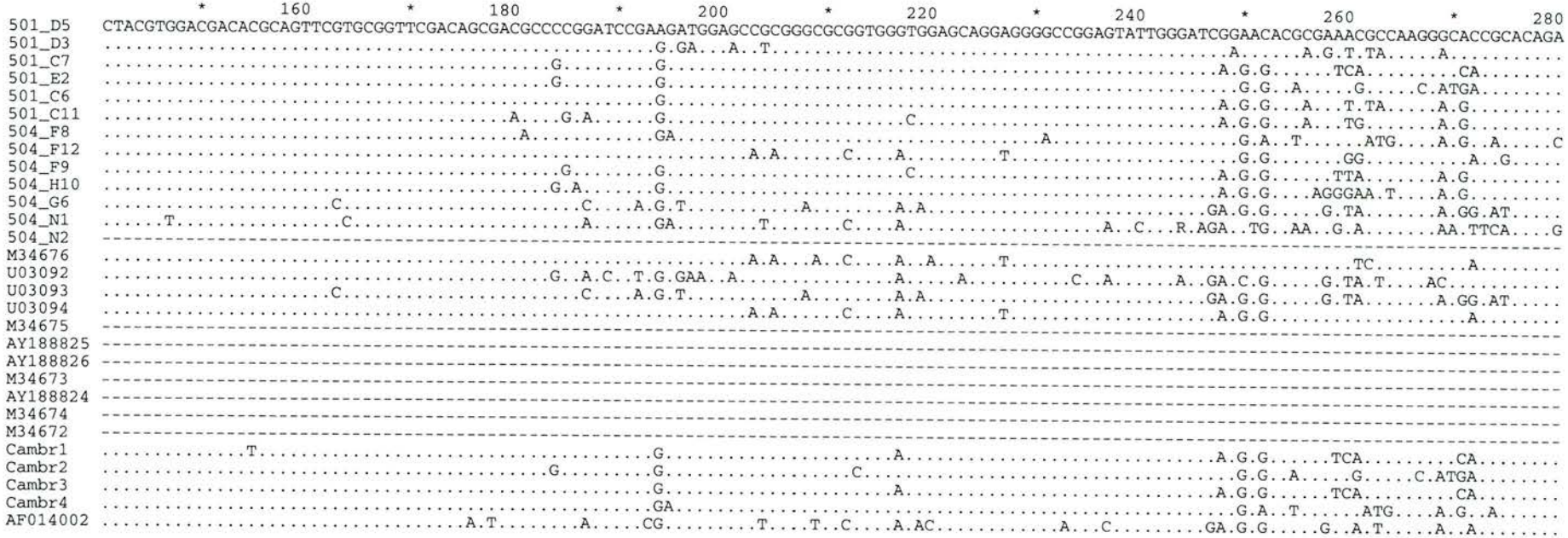
Appendix 2. Results

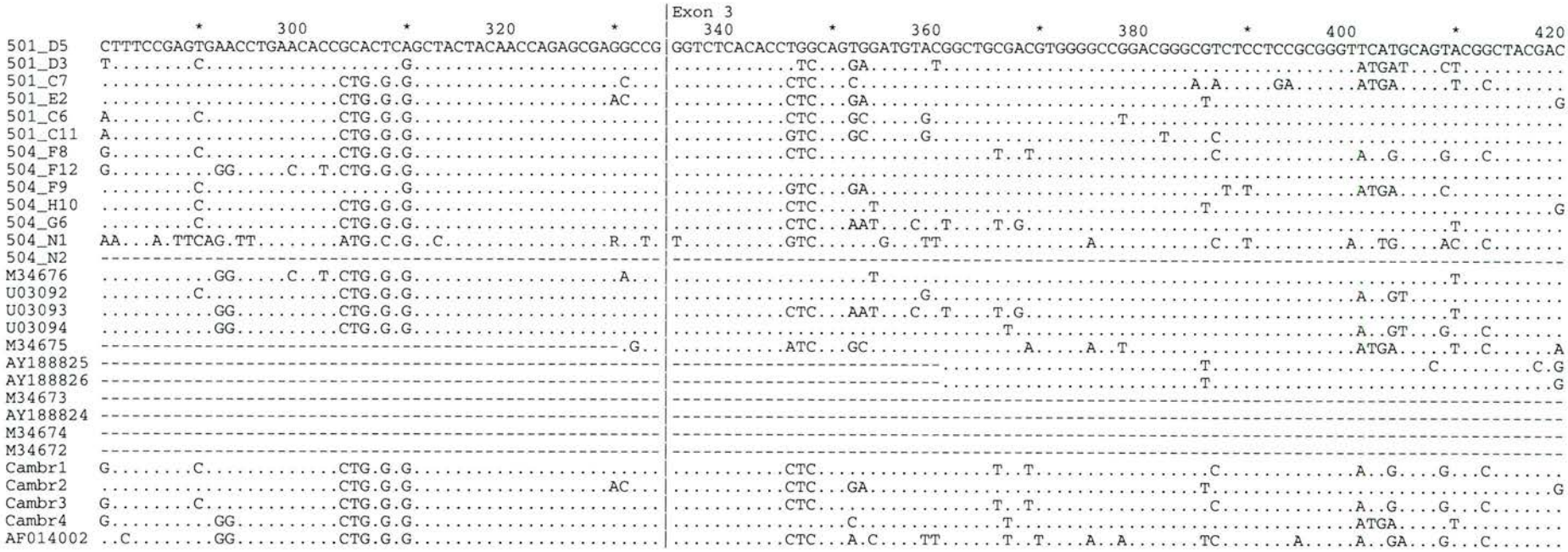
Appendix 2.1: Sequence alignments used for phylogenetic analysis.

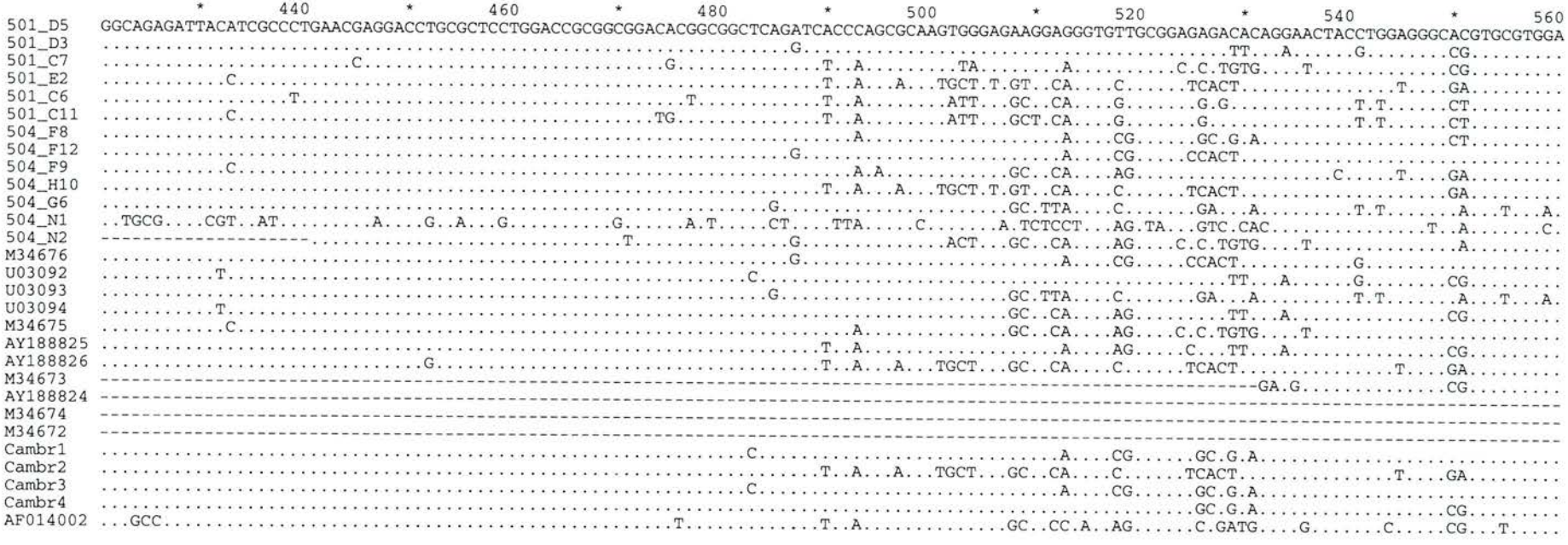
Sequence alignment of ovine MHC class I sequences

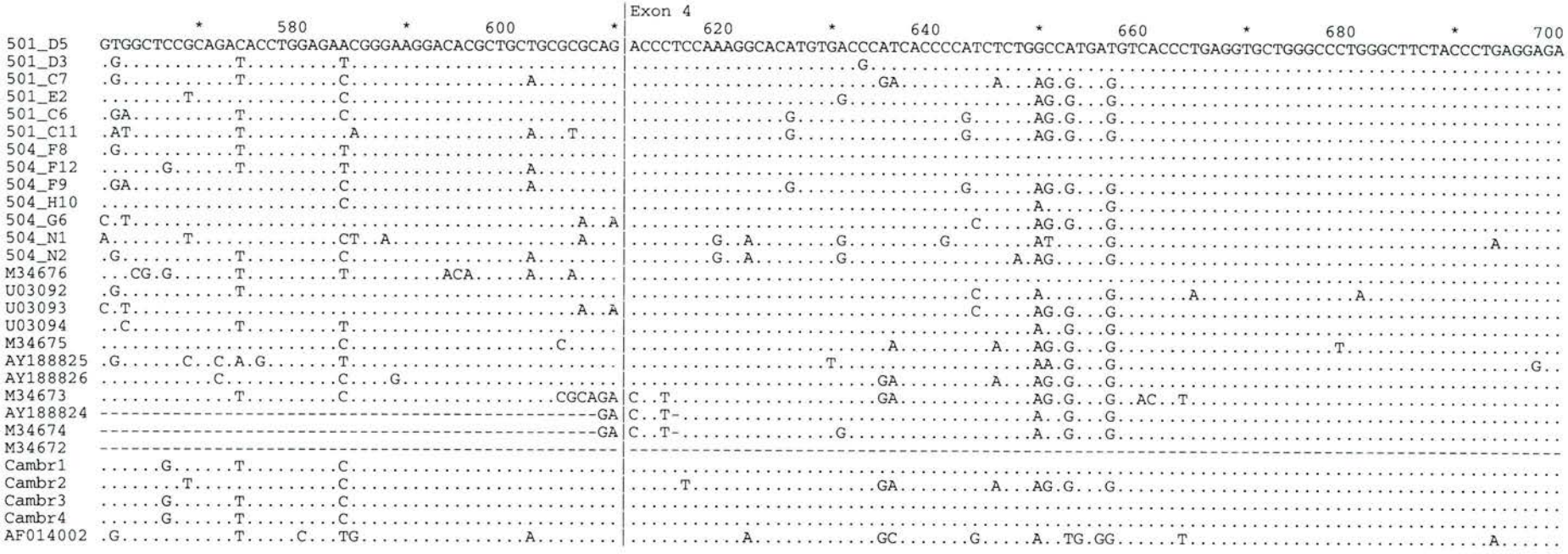
Vertical lines (|) indicate domain boundaries. *Identities* and *gaps* are indicated by dot (.) and dashes (-), respectively. Cambr stands for Cambridge.

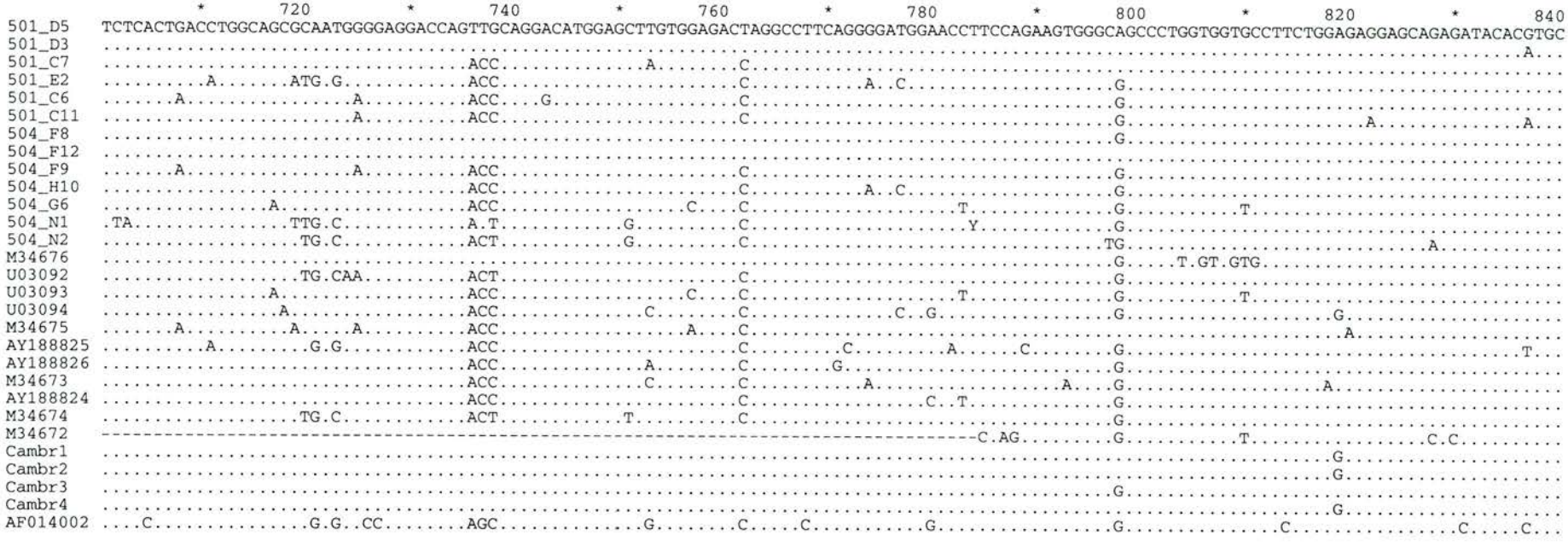
```
          Exon 1                                                          |Exon 2
                   *         20        *         40        *         60   |     *        80         *        100       *        120       *      140
501_D5    ATGGGGCCGCGAACCGTCCTGTTGCTGCTCTCGGGAGTCCTGGTCCTGACCGAGATCCGGGCGG|GCCCCACTCCATGAGGTATTTCAGCACCGCCGTGTCCCGGCCCGGCCGCGGGGAGCCCCGGTACCTGGAAGTCGG
501_D3    ...............C................................................|...........C........G..TA.....G.................T...........C.T.A.C.CC.....
501_C7    ...............CC.T.CA..TC....CT.T.GAC............T....C..T.T.T.|...........C...........CT...........T.......................C.T.A.C.CC.....
501_E2    ----------------------------------------------------------------|-------------------------------------------------------------------------..
501_C6    ..............GC.T..CC.............G...................C..T.....|..T........C........G..TA...................................C.T.A.C.CC.....
501_C11   ..............GC.T..CC.............G.............G.....C.T......|..T........C........G..TA....TG.................T...........C.T.A.C.CC.....
504_F8    ...............C................................................|..T........C........G..TA.....G.............................C.T.A.C.CC.....
504_F12   ...............C................................................|...........................................................................
504_F9    ..............GC.T..CC.............G.............G.....C.T......|..T........C...........CT.......................T...........C.T.A.C.CC.....
504_H10   ...A...........CC.T.CA..T.....CT...GAC.................C......A.|..T........C........G..TA...................................C.T.A.C.CC.....
504_G6    ..............GC.T..CC..........A..G.C..........G......C......T.|..T........C...........CT...................G...T...........C.T.A.CATC.....
504_N1    ....................CC.........T..TG.G.................C.T....C.|..T........C....C......TA.......................T...........C.T.A.T.CC.....
504_N2    ----------------------------------------------------------------|---------------------------------------------------------------------------
M34676    ...............C................................................|.......................................CG.....GC...........................
U03092    -----------------------------------G...................C........|...........C........G..TA.....G.............G...T...........C.T.A.CTCT.....
U03093    -----------------------------------G.C..........G......C......T.|..T........C...........CT...................G...T...........C.T.A.CATC.....
U03094    -----------------------------------....T...............C......T.|...........................................................................
M34675    ----------------------------------------------------------------|...........................................................................
AY188825  ----------------------------------------------------------------|---------------------------------------------------------------------------
AY188826  ----------------------------------------------------------------|---------------------------------------------------------------------------
M34673    ----------------------------------------------------------------|---------------------------------------------------------------------------
AY188824  ----------------------------------------------------------------|---------------------------------------------------------------------------
M34674    ----------------------------------------------------------------|---------------------------------------------------------------------------
M34672    ----------------------------------------------------------------|---------------------------------------------------------------------------
Cambr1.   ...............C................................................|...........C...........CT.......................T...........C.T.A.C.CC.....
Cambr2..  ...............C................................................|...........C...........CT...................................C.T.A.C.CC.....
Cambr3    ...............C................................................|...........C...........CT.......................T...........C.T.A.C.CC.....
Cambr4    ...............C................................................|....................G..TA.....G.................T...........C.T.A.C.CC.....
AF014002  ........TG..G..C..T.CC.......G.....GAC.....C........G..C..A.....|.T.........C....C......TA...................A..............C..T.A.C.CC.....
```

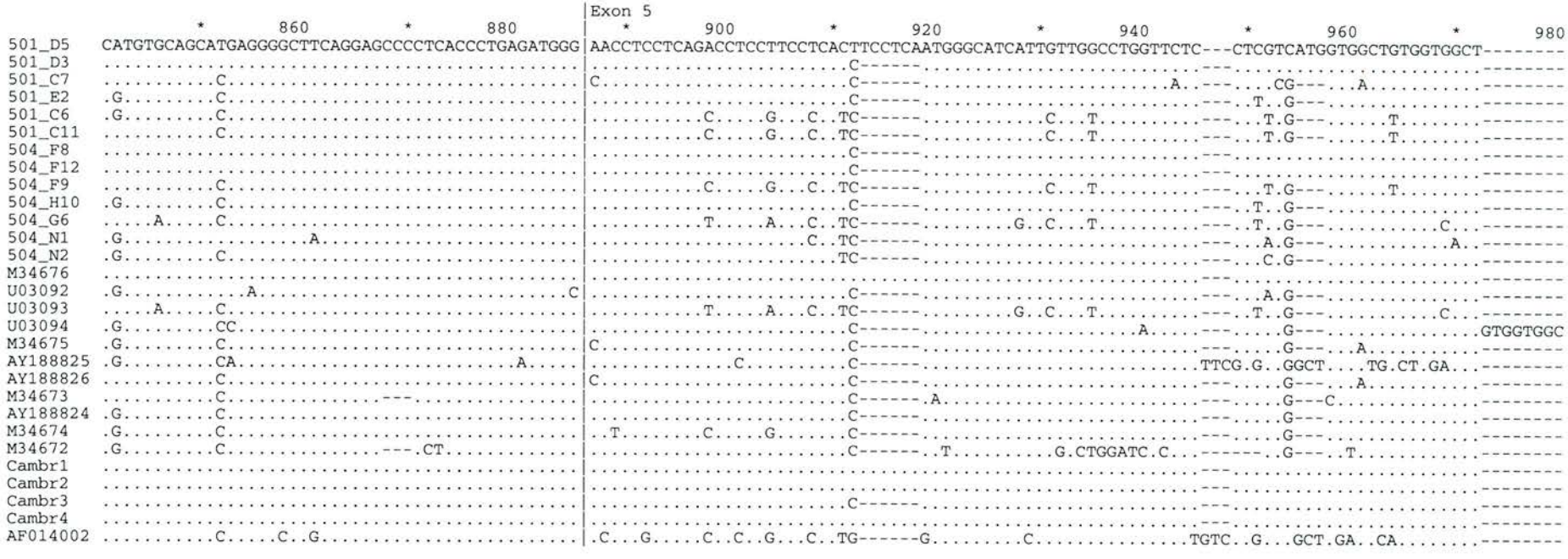

160
180
200
220
240
260
280
501_D5 CTACGTGGACGACACGCAGTTCGTGCGGTTCGACAGCGACGCCCCGGATCCGAAGATGGAGCCGCGGGCGCGGTGGGTGGAGCAGGAGGGGCCGGAGTATTGGGATCGGAACACGCGAAACGCCAAGGGCACCGCACAGA
501_D3
501_C7
501_E2
501_C6
501_C11
504_F8
504_F12
504_F9
504_H10
504_G6
504_N1
504_N2
M34676
U03092
U03093
U03094
M34675
AY188825
AY188826
M34673
AY188824
M34674
M34672
Cambr1
Cambr2
Cambr3
Cambr4
AF014002

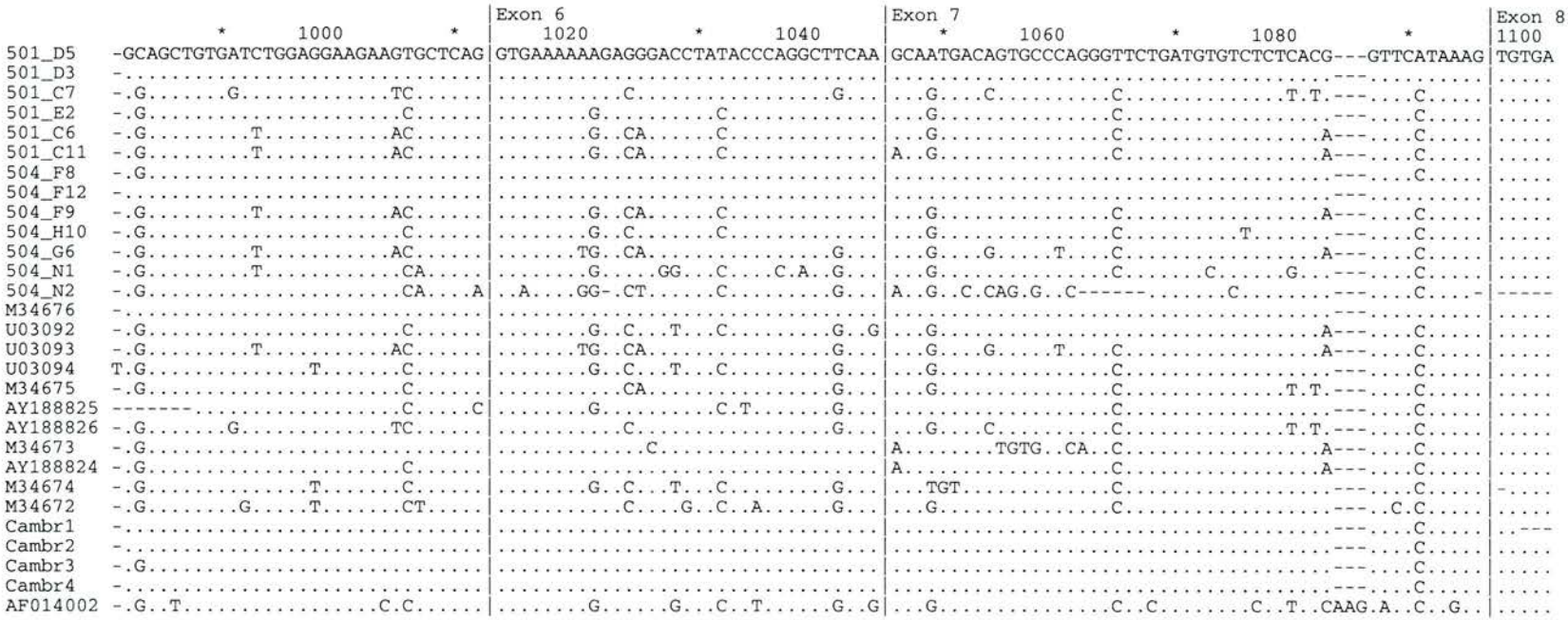

Exon 3
300
320
340
360
380
400
420
501_D5
501_D3
501_C7
501_E2
501_C6
501_C11
504_F8
504_F12
504_F9
504_H10
504_G6
504_N1
504_N2
M34676
U03092
U03093
U03094
M34675
AY188825
AY188826
M34673
AY188824
M34674
M34672
Cambr1
Cambr2
Cambr3
Cambr4
AF014002


* 440 * 460 * 480 * 500 * 520 * 540 * 560
501_D5 GGCAGAGATTACATCGCCCTGAACGAGGACCTGCGCTCCTGGACCGCGGCGGACACGGCGGCTCAGATCACCCAGCGCAAGTGGGAGAAGGAGGGTGTTGCGGAGAGACACAGGAACTACCTGGAGGGCACGTGCGTGGA
501_D3
501_C7
501_E2
501_C6
501_C11
504_F8
504_F12
504_F9
504_H10
504_G6
504_N1
504_N2
M34676
U03092
U03093
U03094
M34675
AY188825
AY188826
M34673
AY188824
M34674
M34672
Cambr1
Cambr2
Cambr3
Cambr4
AF014002


Exon 4
501_D5 GTGGCTCCGCAGACACCTGGAGAACGGGAAGGACACGCTGCTGCGCGCAG ACCCTCCAAAGGCACATGTGACCCATCACCCCATCTCTGGCCATGATGTCACCCTGAGGTGCTGGGCCCTGGGCTTCTACCCTGAGGAGA
501_D3
501_C7
501_E2
501_C6
501_C11
504_F8
504_F12
504_F9
504_H10
504_G6
504_N1
504_N2
M34676
U03092
U03093
U03094
M34675
AY188825
AY188826
M34673
AY188824
M34674
M34672
Cambr1
Cambr2
Cambr3
Cambr4
AF014002


720
740
760
780
800
820
840
501_D5 TCTCACTGACCTGGCAGCGCAATGGGGAGGACCAGTTGCAGGACATGGAGCTTGTGGAGACTAGGCCTTCAGGGGATGGAACCTTCCAGAAGTGGGCAGCCCTGGTGGTGCCTTCTGGAGAGGAGCAGAGATACACGTGC
501_D3
501_C7
501_E2
501_C6
501_C11
504_F8
504_F12
504_F9
504_H10
504_G6
504_N1
504_N2
M34676
U03092
U03093
U03094
M34675
AY188825
AY188826
M34673
AY188824
M34674
M34672
Cambr1
Cambr2
Cambr3
Cambr4
AF014002

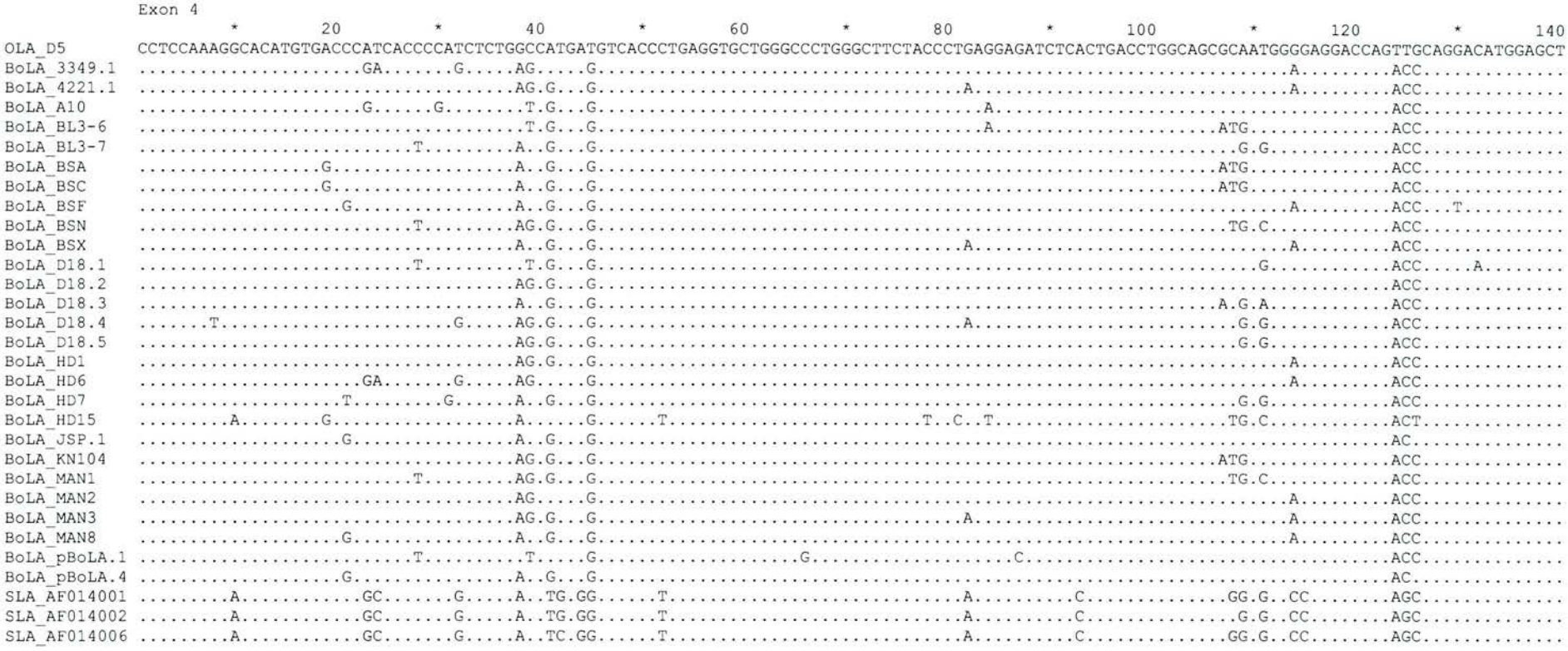

Exon 5
860
880
900
920
940
960
980
501_D5
501_D3
501_C7
501_E2
501_C6
501_C11
504_F8
504_F12
504_F9
504_H10
504_G6
504_N1
504_N2
M34676
U03092
U03093
U03094
M34675
AY188825
AY188826
M34673
AY188824
M34674
M34672
Cambr1
Cambr2
Cambr3
Cambr4
AF014002

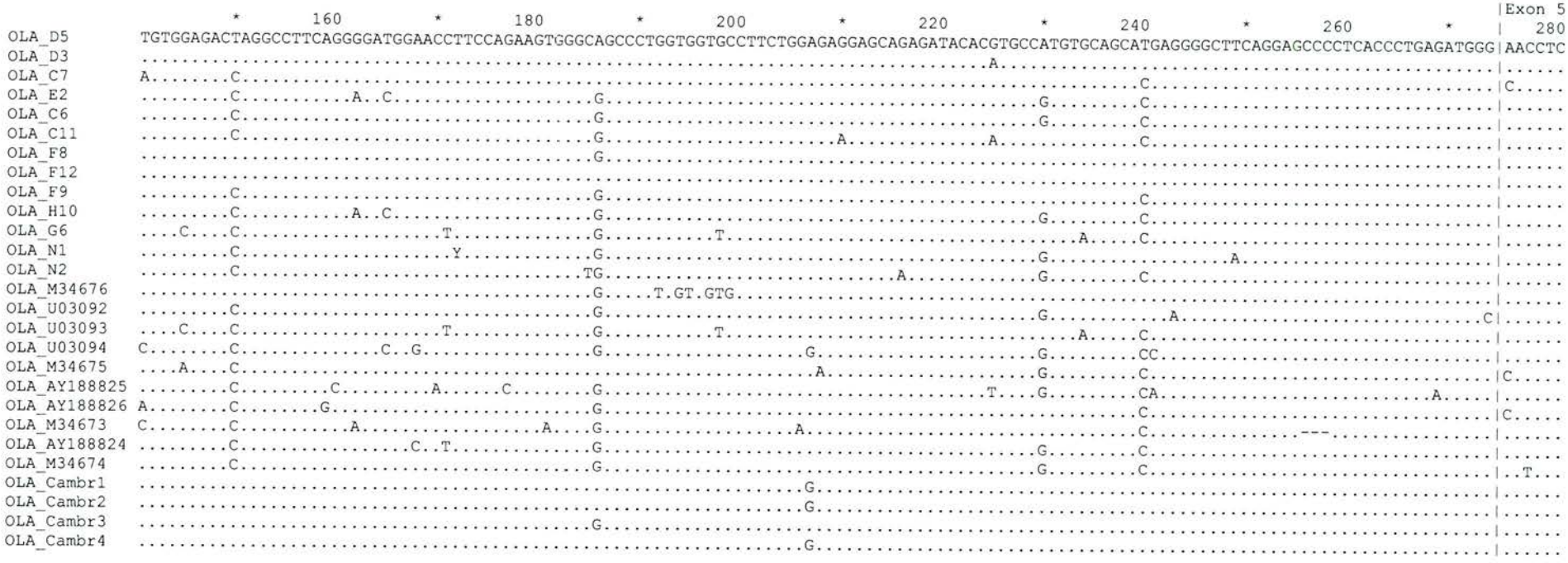

Exon 6
Exon 7
Exon 8
1000
1020
1040
1060
1080
1100
501_D5 -GCAGCTGTGATCTGGAGGAAGAAGTGCTCAG GTGAAAAAAGAGGGACCTATACCCAGGCTTCAA GCAATGACAGTGCCCAGGGTTCTGATGTGTCTCTCACG---GTTCATAAAG TGTGA
501_D3
501_C7
501_E2
501_C6
501_C11
504_F8
504_F12
504_F9
504_H10
504_G6
504_N1
504_N2
M34676
U03092
U03093
U03094
M34675
AY188825
AY188826
M34673
AY188824
M34674
M34672
Cambr1
Cambr2
Cambr3
Cambr4
AF014002

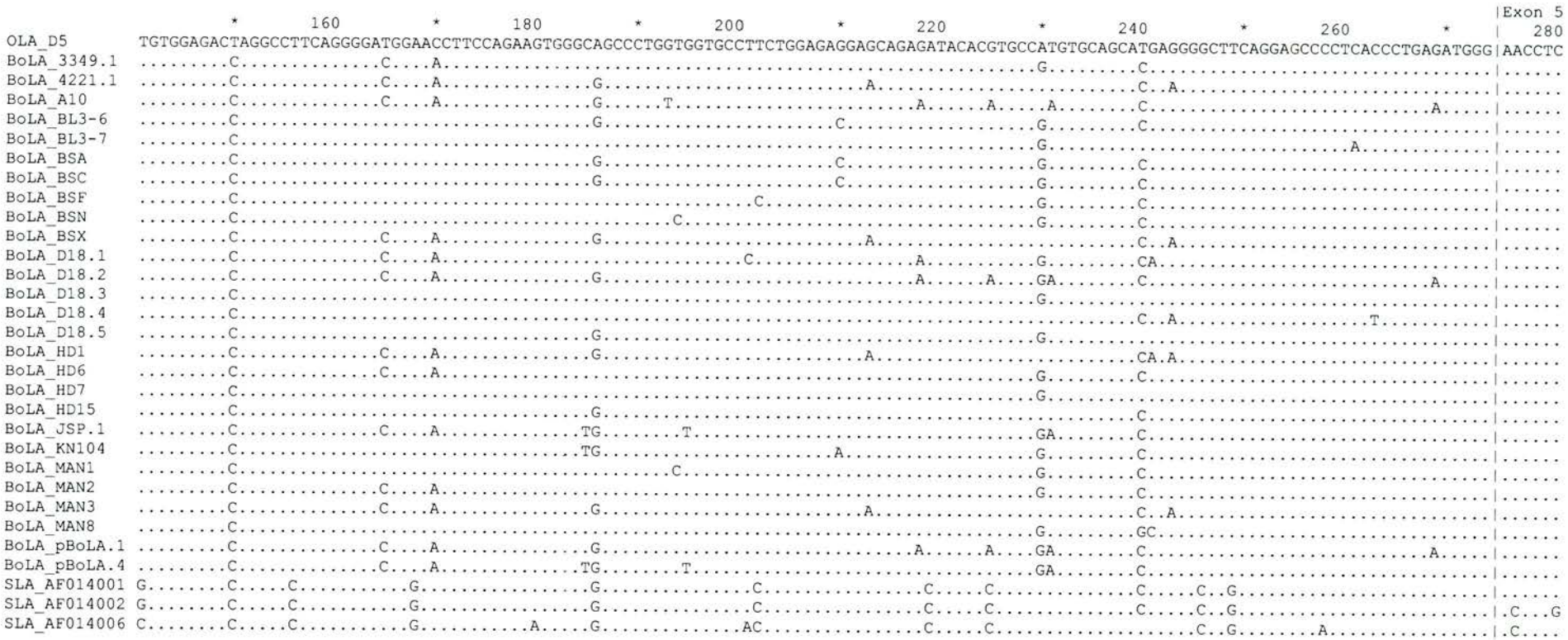
Sequence alignment of ovine with bovine MHC class I sequences

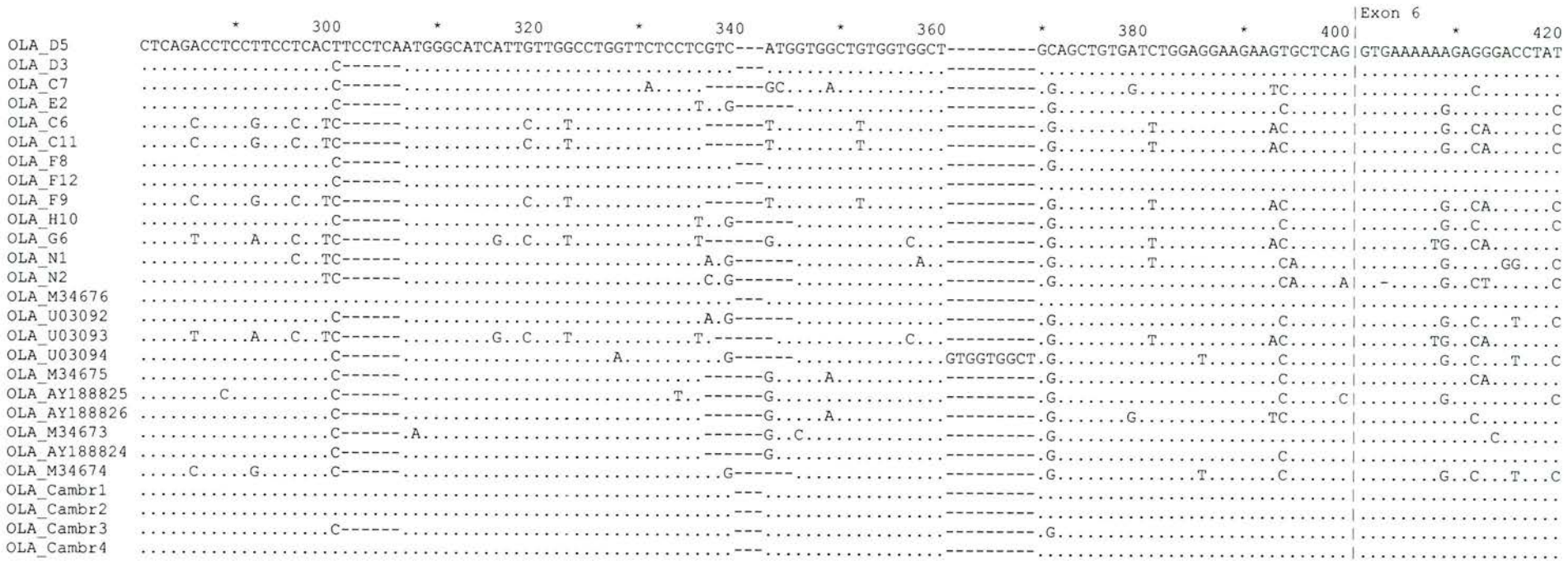
Vertical lines (|) indicate domain boundaries. *Identities* and *gaps* are indicated by dot (.) and dashes (-), respectively. Cambr stands for Cambridge.

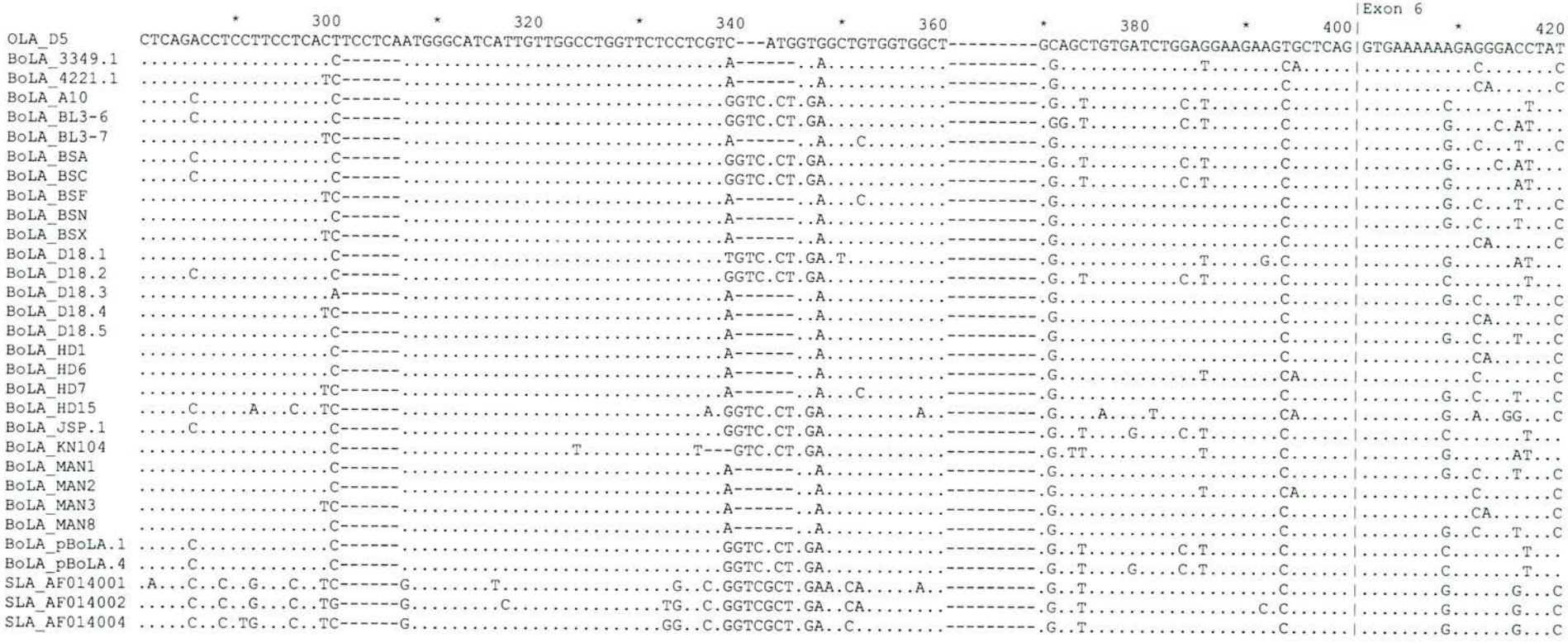
```
            Exon 4
                  *        20        *        40        *        60        *        80        *       100        *       120        *       140
OLA_D5      CCTCCAAAGGCACATGTGACCCATCACCCCATCTCTGGCCATGATGTCACCCTGAGGTGCTGGGCCCTGGGCTTCTACCCTGAGGAGATCTCACTGACCTGGCAGCGCAATGGGGAGGACCAGTTGCAGGACATGGAGCT
OLA_D3      ....................G.....................................................................................................................
OLA_C7      ....................GA.........A...AG.G...G.....................................................................................ACC.............
OLA_E2      ..................G..................AG.G...G.....................................................A.......ATG.G............ACC.............
OLA_C6      ............G..............G......AG.G...G.....................................................A..............A..........ACC....G.........
OLA_C11     ............G..............G......AG.G...G.....................................................A..........ACC.............
OLA_F8      ....................................................................................................................................
OLA_F12     ....................................................................................................................................
OLA_F9      ............G..............G......AG.G...G.....................................................A..............A..........ACC.............
OLA_H10     ..................................A......G.....................................................................ACC.............
OLA_G6      ............................C.....AG.G...G.....................................................A..................ACC.............
OLA_N1      ......G..A........G.........G........AT.....G.....................................A.......TA..............TTG.C............A.T............G.
OLA_N2      ......G..A........G..............A.AG.....G.....................................................TG.C............ACT............G.
OLA_M34676  ....................................................................................................................................
OLA_U03092  ............................C.....A......G.......A..............A.....................................TG.CAA..........ACT.............
OLA_U03093  ............................C.....AG.G...G.....................................................A..................ACC.............
OLA_U03094  ..................................A..G...G.....................................................A..................ACC.............
OLA_M34675  ......................A.........A...AG.G...G.....................T.....................A..........A.....A..........ACC.............
OLA_AY188825 ................T..................AA.G...G.....................................G............A.........G.G............ACC.............
OLA_AY188826 ......................GA.........A...AG.G...G.....................................................................ACC.............
OLA_M34673  ......................GA.............AG.G...G..AC..T.....................................................................ACC.............
OLA_AY188824 ..................................A..G...G.....................................................................ACC.............
OLA_M34674  ..................G..................A..G...G.....................................................TG.C............ACT............T.
OLA_Cambr1  ....................................................................................................................................
OLA_Cambr2 ...T..................GA.........A...AG.G...G.....................................................................................
OLA_Cambr3 ....................................................................................................................................
OLA_Cambr4 ....................................................................................................................................
```

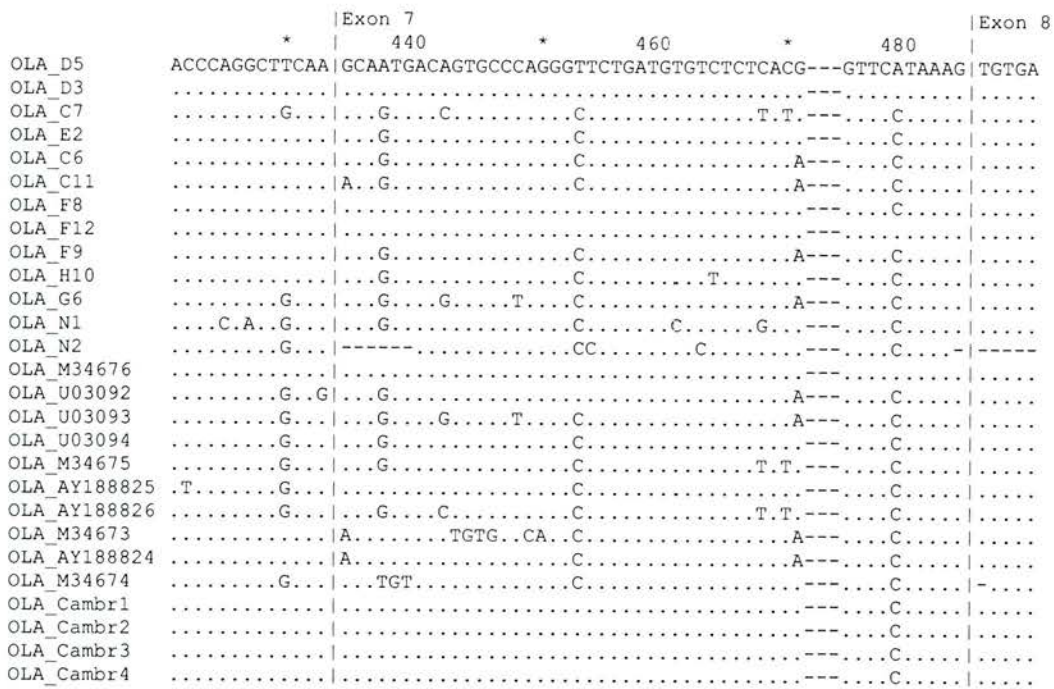

Exon 4
OLA_D5 CCTCCAAAGGCACATGTGACCCATCACCCCATCTCTGGCCATGATGTCACCCTGAGGTGCTGGGCCCTGGGCTTCTACCCTGAGGAGATCTCACTGACCTGGCAGCGCAATGGGGAGGACCAGTTGCAGGACATGGAGCT
BoLA_3349.1
BoLA_4221.1
BoLA_A10
BoLA_BL3-6
BoLA_BL3-7
BoLA_BSA
BoLA_BSC
BoLA_BSF
BoLA_BSN
BoLA_BSX
BoLA_D18.1
BoLA_D18.2
BoLA_D18.3
BoLA_D18.4
BoLA_D18.5
BoLA_HD1
BoLA_HD6
BoLA_HD7
BoLA_HD15
BoLA_JSP.1
BoLA_KN104
BoLA_MAN1
BoLA_MAN2
BoLA_MAN3
BoLA_MAN8
BoLA_pBoLA.1
BoLA_pBoLA.4
SLA_AF014001
SLA_AF014002
SLA_AF014006


Exon 5
160
180
200
220
240
260
280
OLA_D5 TGTGGAGACTAGGCCTTCAGGGGATGGAACCTTCCAGAAGTGGGCAGCCCTGGTGGTGCCTTCTGGAGAGGAGCAGAGATACACGTGCCATGTGCAGCATGAGGGGCTTCAGGAGCCCCTCACCCTGAGATGGGAACCTC
OLA_D3
OLA_C7
OLA_E2
OLA_C6
OLA_C11
OLA_F8
OLA_F12
OLA_F9
OLA_H10
OLA_G6
OLA_N1
OLA_N2
OLA_M34676
OLA_U03092
OLA_U03093
OLA_U03094
OLA_M34675
OLA_AY188825
OLA_AY188826
OLA_M34673
OLA_AY188824
OLA_M34674
OLA_Cambr1
OLA_Cambr2
OLA_Cambr3
OLA_Cambr4


Exon 5
OLA_D5 TGTGGAGACTAGGCCTTCAGGGGATGGAACCTTCCAGAAGTGGGCAGCCCTGGTGGTGCCTTCTGGAGAGGAGCAGAGATACACGTGCCATGTGCAGCATGAGGGGCTTCAGGAGCCCCTCACCCTGAGATGGG AACCTC
BoLA_3349.1
BoLA_4221.1
BoLA_A10
BoLA_BL3-6
BoLA_BL3-7
BoLA_BSA
BoLA_BSC
BoLA_BSF
BoLA_BSN
BoLA_BSX
BoLA_D18.1
BoLA_D18.2
BoLA_D18.3
BoLA_D18.4
BoLA_D18.5
BoLA_HD1
BoLA_HD6
BoLA_HD7
BoLA_HD15
BoLA_JSP.1
BoLA_KN104
BoLA_MAN1
BoLA_MAN2
BoLA_MAN3
BoLA_MAN8
BoLA_pBoLA.1
BoLA_pBoLA.4
SLA_AF014001
SLA_AF014002
SLA_AF014006


Exon 6
OLA_D5
OLA_D3
OLA_C7
OLA_E2
OLA_C6
OLA_C11
OLA_F8
OLA_F12
OLA_F9
OLA_H10
OLA_G6
OLA_N1
OLA_N2
OLA_M34676
OLA_U03092
OLA_U03093
OLA_U03094
OLA_M34675
OLA_AY188825
OLA_AY188826
OLA_M34673
OLA_AY188824
OLA_M34674
OLA_Cambr1
OLA_Cambr2
OLA_Cambr3
OLA_Cambr4

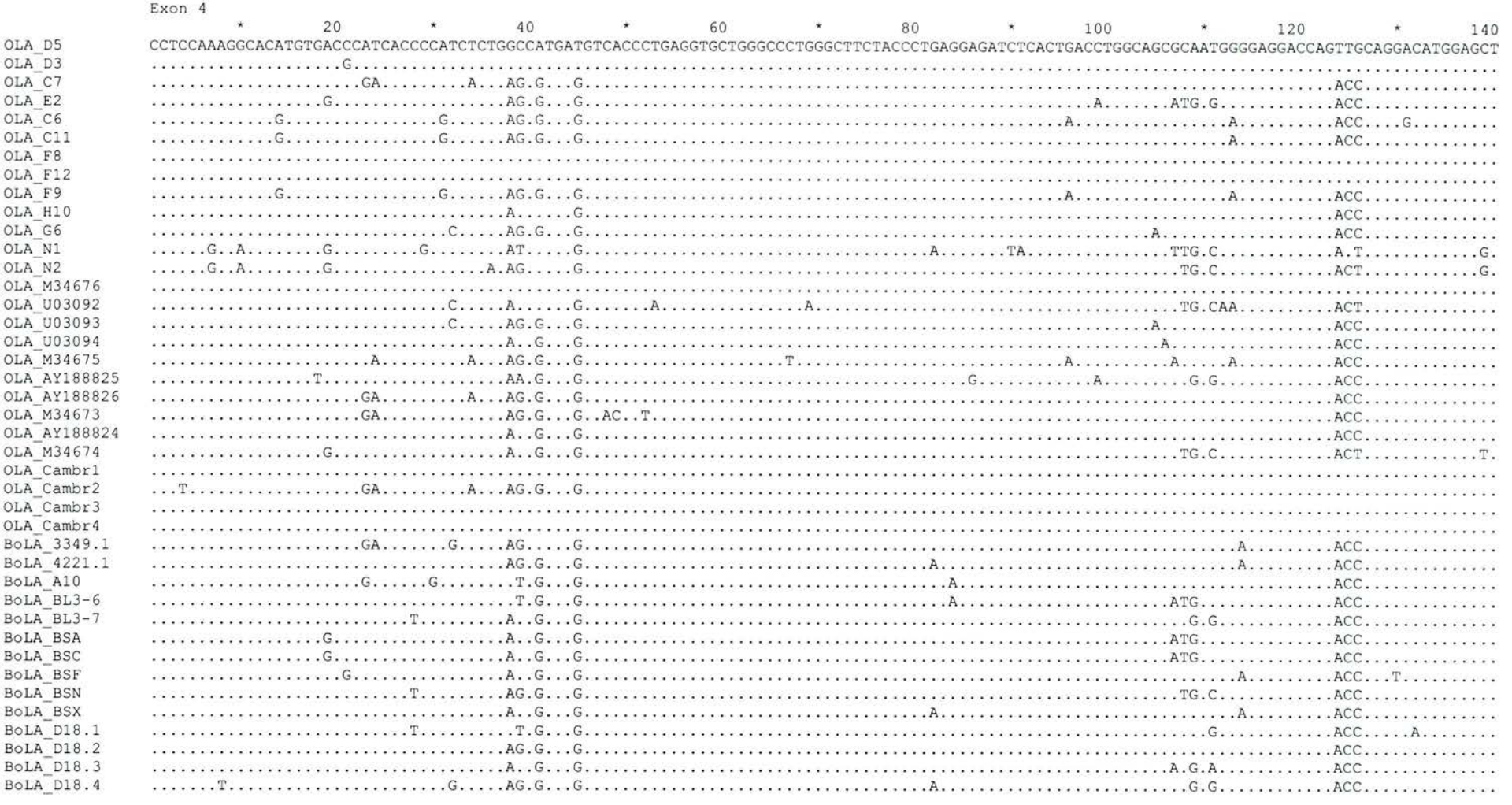

Exon 6
OLA_D5
BoLA_3349.1
BoLA_4221.1
BoLA_A10
BoLA_BL3-6
BoLA_BL3-7
BoLA_BSA
BoLA_BSC
BoLA_BSF
BoLA_BSN
BoLA_BSX
BoLA_D18.1
BoLA_D18.2
BoLA_D18.3
BoLA_D18.4
BoLA_D18.5
BoLA_HD1
BoLA_HD6
BoLA_HD7
BoLA_HD15
BoLA_JSP.1
BoLA_KN104
BoLA_MAN1
BoLA_MAN2
BoLA_MAN3
BoLA_MAN8
BoLA_pBoLA.1
BoLA_pBoLA.4
SLA_AF014001
SLA_AF014002
SLA_AF014004

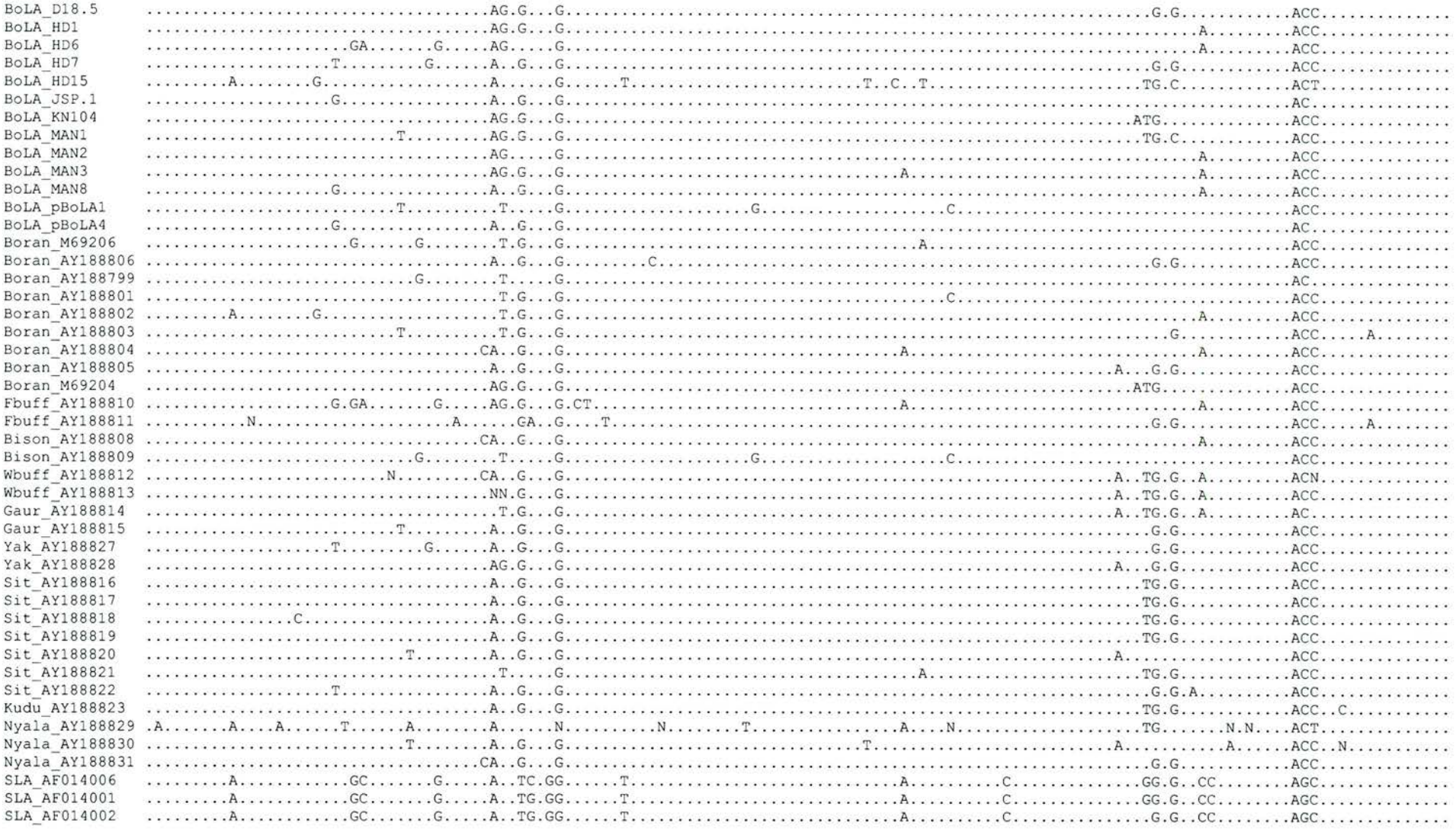

Exon 7
Exon 8
440
460
480
OLA_D5
OLA_D3
OLA_C7
OLA_E2
OLA_C6
OLA_C11
OLA_F8
OLA_F12
OLA_F9
OLA_H10
OLA_G6
OLA_N1
OLA_N2
OLA_M34676
OLA_U03092
OLA_U03093
OLA_U03094
OLA_M34675
OLA_AY188825
OLA_AY188826
OLA_M34673
OLA_AY188824
OLA_M34674
OLA_Cambr1
OLA_Cambr2
OLA_Cambr3
OLA_Cambr4

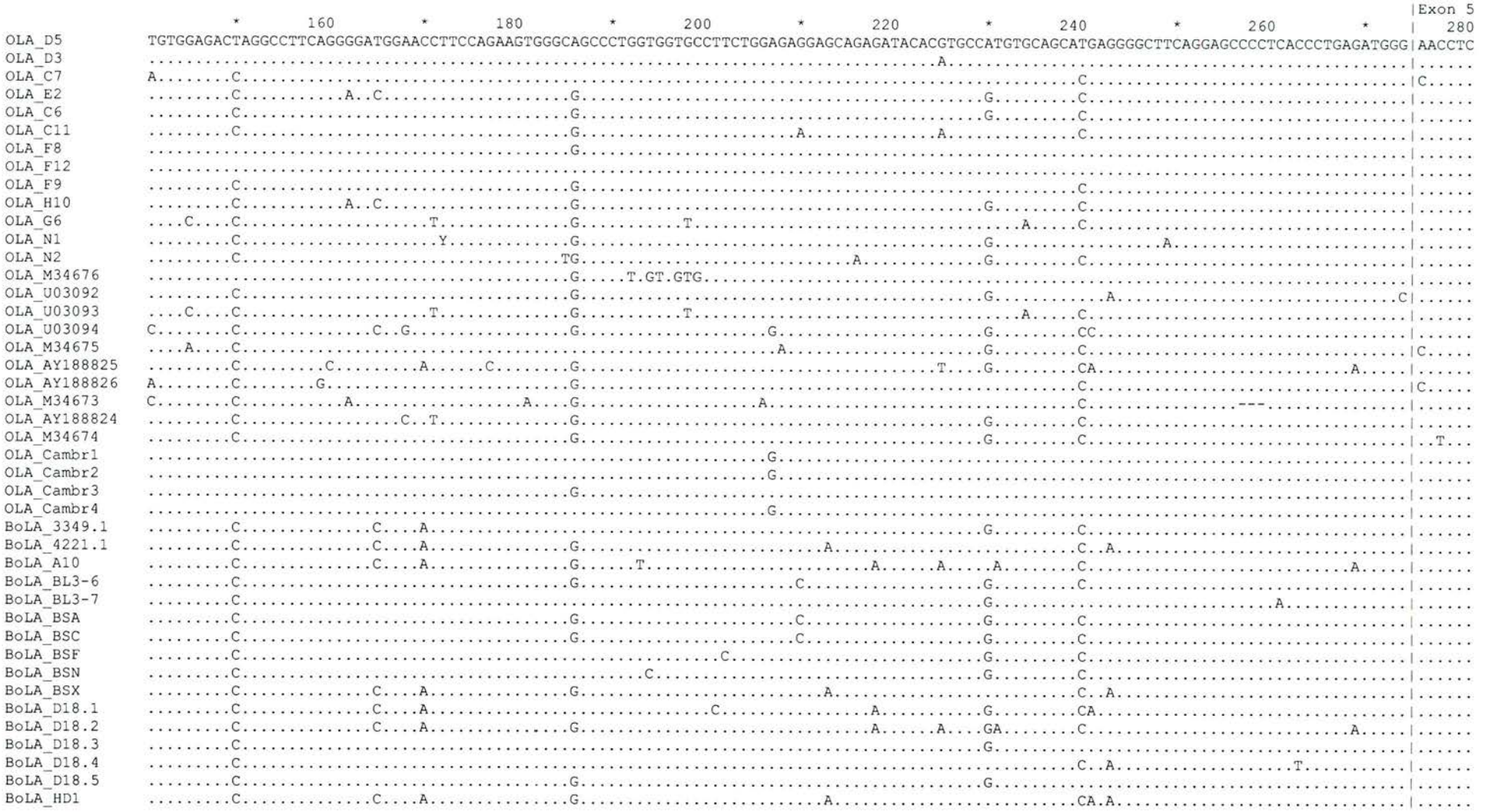
```
                        |Exon 7                                                                 |Exon 8
                    *   |    440         *       460         *       480    |
OLA_D5          ACCCAGGCTTCAA|GCAATGACAGTGCCCAGGGTTCTGATGTGTCTCTCACG---GTTCATAAAG|TGTGA
BoLA_3349.1     .........G...|..................C..................---....C.....|.....
BoLA_4221.1     .........G...|.TGG........A......C..................---....C.....|.T...
BoLA_A10        .T..........|...G..............C..................---....C.....|.....
BoLA_BL3-6      .T..........|...G..............C..................---....C.....|.....
BoLA_BL3-7      .........G...|...G..............C..................---....C.....|.....
BoLA_BSA        .T.........G|...G..............C..................---....C.....|.....
BoLA_BSC        .T.........G|.G.G..............C..................---....C.....|.....
BoLA_BSF        .........G...|...G.......A.......C..................A---....C.....|CTGTC
BoLA_BSN        .........G...|...G.......A.......C..................---....C.....|.T...
BoLA_BSX        .........G...|.TGG........A......C..................---....C.....|.....
BoLA_D18.1      .T..........|...G..............C..................---....C.....|.....
BoLA_D18.2      .T..........|...G..............C..................---....C.....|.....
BoLA_D18.3      .........G...|...G..............C..................---....C.....|.T...
BoLA_D18.4      .........G...|.TGG........A......C..................---....C.....|.....
BoLA_D18.5      .........G...|...G.......A.......C......A..........A---....C.....|.T...
BoLA_HD1        .........G...|.TGG........A......C..................---....C.....|.T...
BoLA_HD6        .........G...|...................C..................---....C.....|.....
BoLA_HD7        .........G...|...G..............C..................---....C.....|.....
BoLA_HD15       ......A..G...|...G.T......A......C................T.---....C.....|.....
BoLA_JSP.1      .T..........|...G..............C..................---....C.....|.....
BoLA_KN104      .T..........|...G..C........G...C..................---....C.....|.....
BoLA_MAN1       .........G...|...G.......A.......C..................---....C.....|.T...
BoLA_MAN2       .........G...|...................C..................---....C.....|.....
BoLA_MAN3       .........G...|.TGG........A......C..................---....C.....|.T...
BoLA_MAN8       .........G...|...G..............C..................---....C.....|.T...
BoLA_pBoLA.1    .T..........|...G..............C..................---....C.....|.....
BoLA_pBoLA.4    .T..........|...G..............C..................---....C.....|.....
SLA_AF014001    ..T......G..G|.G.G..............C..C........C..T..CAAG.A..C..G..|.....
SLA_AF014002    ..T......G..G|...G..............C..C........C..T..CAAG.A..C..G..|.....
SLA_AF014006    ..T......G..G|...G..............C..C........C..T..CAAG.A..C..G..|.....
```

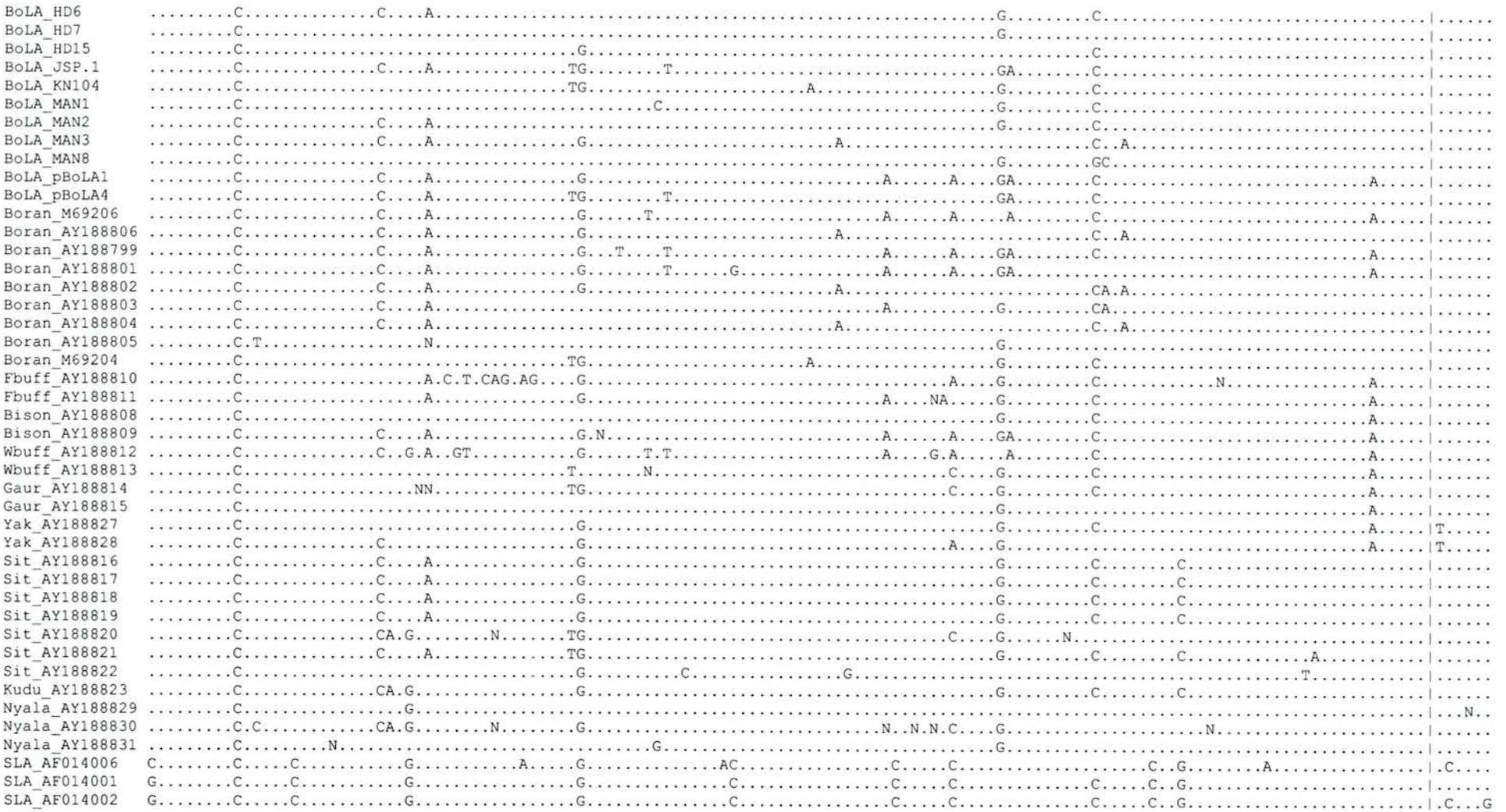
Sequence alignment of bovidae MHC class I sequences

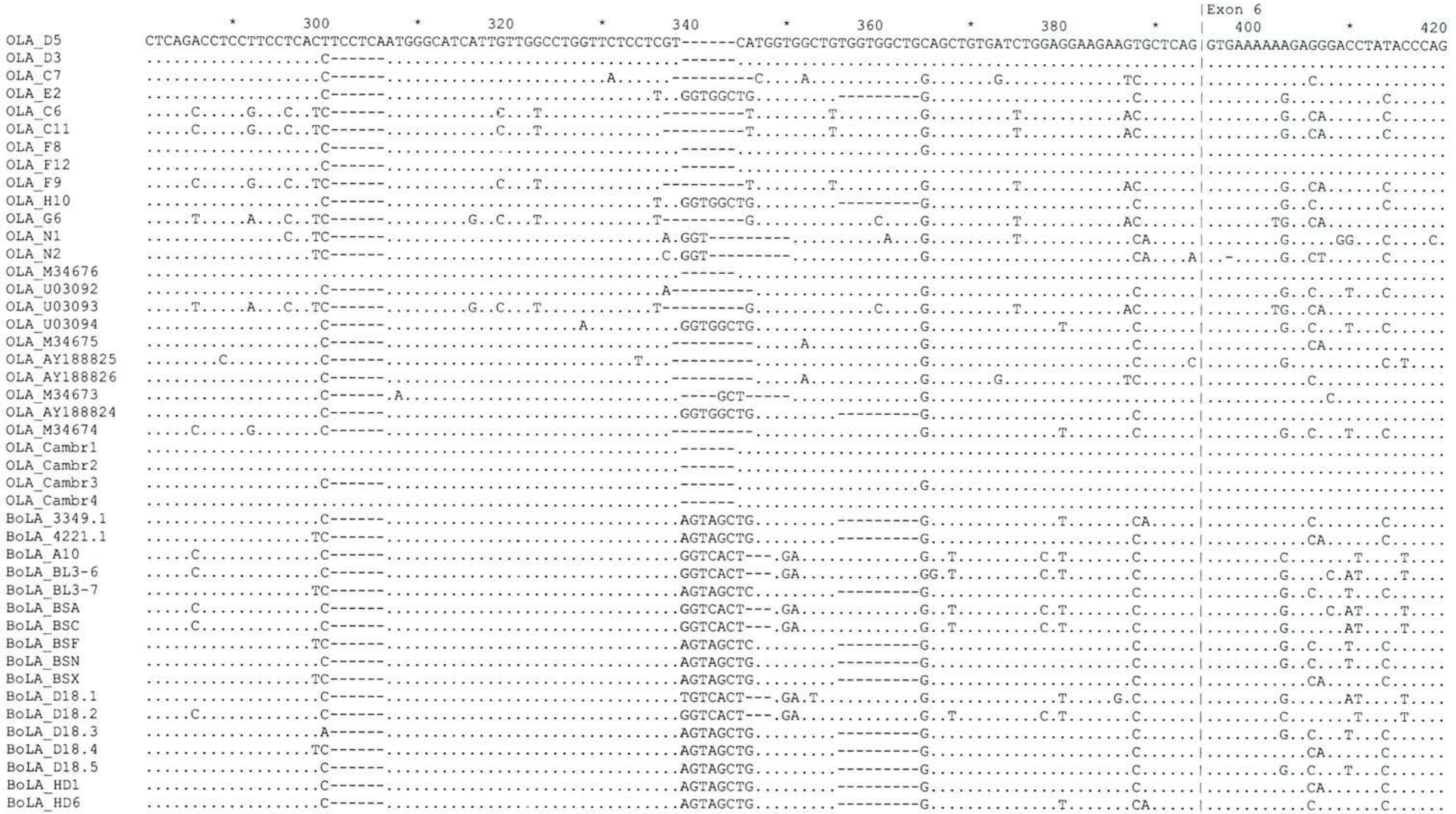
Vertical lines (|) indicate domain boundaries. *Identities* and *gaps* are indicated by dot (.) and dashes (-), respectively. Cambr stands for Cambridge.

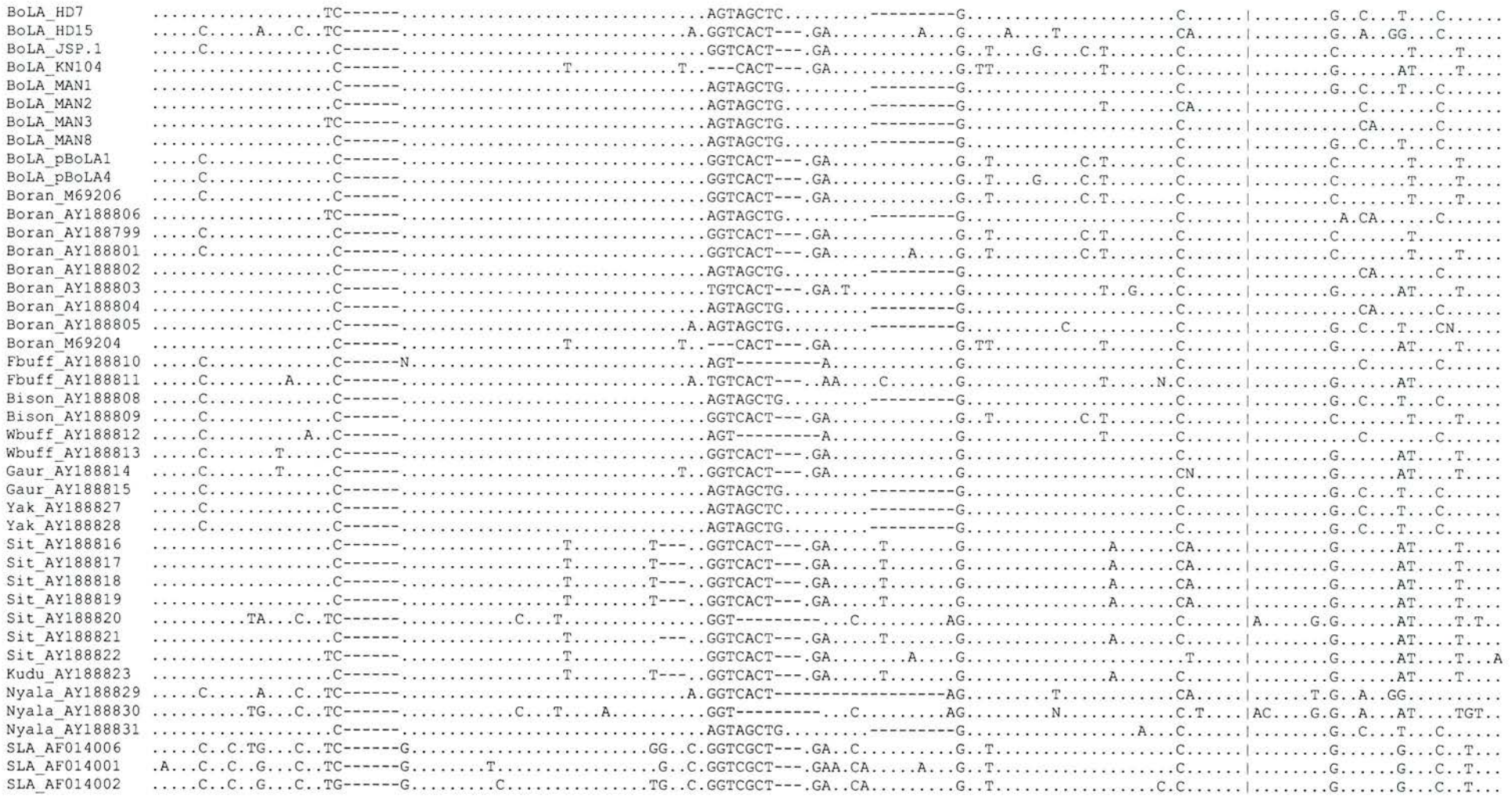



BoLA_D18.5
BoLA_HD1
BoLA_HD6
BoLA_HD7
BoLA_HD15
BoLA_JSP.1
BoLA_KN104
BoLA_MAN1
BoLA_MAN2
BoLA_MAN3
BoLA_MAN8
BoLA_pBoLA1
BoLA_pBoLA4
Boran_M69206
Boran_AY188806
Boran_AY188799
Boran_AY188801
Boran_AY188802
Boran_AY188803
Boran_AY188804
Boran_AY188805
Boran_M69204
Fbuff_AY188810
Fbuff_AY188811
Bison_AY188808
Bison_AY188809
Wbuff_AY188812
Wbuff_AY188813
Gaur_AY188814
Gaur_AY188815
Yak_AY188827
Yak_AY188828
Sit_AY188816
Sit_AY188817
Sit_AY188818
Sit_AY188819
Sit_AY188820
Sit_AY188821
Sit_AY188822
Kudu_AY188823
Nyala_AY188829
Nyala_AY188830
Nyala_AY188831
SLA_AF014006
SLA_AF014001
SLA_AF014002


Exon 5
160
180
200
220
240
260
280
OLA_D5 TGTGGAGACTAGGCCTTCAGGGGATGGAACCTTCCAGAAGTGGGCAGCCCTGGTGGTGCCTTCTGGAGAGGAGCAGAGATACACGTGCCATGTGCAGCATGAGGGGCTTCAGGAGCCCCTCACCCTGAGATGGG|AACCTC
OLA_D3
OLA_C7
OLA_E2
OLA_C6
OLA_C11
OLA_F8
OLA_F12
OLA_F9
OLA_H10
OLA_G6
OLA_N1
OLA_N2
OLA_M34676
OLA_U03092
OLA_U03093
OLA_U03094
OLA_M34675
OLA_AY188825
OLA_AY188826
OLA_M34673
OLA_AY188824
OLA_M34674
OLA_Cambr1
OLA_Cambr2
OLA_Cambr3
OLA_Cambr4
BoLA_3349.1
BoLA_4221.1
BoLA_A10
BoLA_BL3-6
BoLA_BL3-7
BoLA_BSA
BoLA_BSC
BoLA_BSF
BoLA_BSN
BoLA_BSX
BoLA_D18.1
BoLA_D18.2
BoLA_D18.3
BoLA_D18.4
BoLA_D18.5
BoLA_HD1


BoLA_HD6
BoLA_HD7
BoLA_HD15
BoLA_JSP.1
BoLA_KN104
BoLA_MAN1
BoLA_MAN2
BoLA_MAN3
BoLA_MAN8
BoLA_pBoLA1
BoLA_pBoLA4
Boran_M69206
Boran_AY188806
Boran_AY188799
Boran_AY188801
Boran_AY188802
Boran_AY188803
Boran_AY188804
Boran_AY188805
Boran_M69204
Fbuff_AY188810
Fbuff_AY188811
Bison_AY188808
Bison_AY188809
Wbuff_AY188812
Wbuff_AY188813
Gaur_AY188814
Gaur_AY188815
Yak_AY188827
Yak_AY188828
Sit_AY188816
Sit_AY188817
Sit_AY188818
Sit_AY188819
Sit_AY188820
Sit_AY188821
Sit_AY188822
Kudu_AY188823
Nyala_AY188829
Nyala_AY188830
Nyala_AY188831
SLA_AF014006
SLA_AF014001
SLA_AF014002


Exon 6
300
320
340
360
380
400
420
OLA_D5
OLA_D3
OLA_C7
OLA_E2
OLA_C6
OLA_C11
OLA_F8
OLA_F12
OLA_F9
OLA_H10
OLA_G6
OLA_N1
OLA_N2
OLA_M34676
OLA_U03092
OLA_U03093
OLA_U03094
OLA_M34675
OLA_AY188825
OLA_AY188826
OLA_M34673
OLA_AY188824
OLA_M34674
OLA_Cambr1
OLA_Cambr2
OLA_Cambr3
OLA_Cambr4
BoLA_3349.1
BoLA_4221.1
BoLA_A10
BoLA_BL3-6
BoLA_BL3-7
BoLA_BSA
BoLA_BSC
BoLA_BSF
BoLA_BSN
BoLA_BSX
BoLA_D18.1
BoLA_D18.2
BoLA_D18.3
BoLA_D18.4
BoLA_D18.5
BoLA_HD1
BoLA_HD6


BoLA_HD7
BoLA_HD15
BoLA_JSP.1
BoLA_KN104
BoLA_MAN1
BoLA_MAN2
BoLA_MAN3
BoLA_MAN8
BoLA_pBoLA1
BoLA_pBoLA4
Boran_M69206
Boran_AY188806
Boran_AY188799
Boran_AY188801
Boran_AY188802
Boran_AY188803
Boran_AY188804
Boran_AY188805
Boran_M69204
Fbuff_AY188810
Fbuff_AY188811
Bison_AY188808
Bison_AY188809
Wbuff_AY188812
Wbuff_AY188813
Gaur_AY188814
Gaur_AY188815
Yak_AY188827
Yak_AY188828
Sit_AY188816
Sit_AY188817
Sit_AY188818
Sit_AY188819
Sit_AY188820
Sit_AY188821
Sit_AY188822
Kudu_AY188823
Nyala_AY188829
Nyala_AY188830
Nyala_AY188831
SLA_AF014006
SLA_AF014001
SLA_AF014002

```
                    |Exon 7                                                  |Exon 8
                    |  *       440         *       460         *       4|80
OLA_D5         GCTTCAA|GCAATGACAGTGCCCAGGGTTCTGATGTGTCTCTCACG---GTTCATAAAG|TGTGA
OLA_D3         .......|.....................................---..........|.....
OLA_C7         ...G...|...G....C..........C.............T.T.---....C.....|.....
OLA_E2         .......|...G..............C...............---....C.....|.....
OLA_C6         .......|...G..............C.............A---....C.....|.....
OLA_C11        .......|A..G..............C.............A---....C.....|.....
OLA_F8         .......|.....................................---....C.....|.....
OLA_F12        .......|.....................................---..........|.....
OLA_F9         .......|...G..............C.............A---....C.....|.....
OLA_H10        .......|...G..............C..........T....---....C.....|.....
OLA_G6         ...G...|...G....G.....T....C.............A---....C.....|.....
OLA_N1         A..G...|...G..............C.......C......G...---....C.....|.....
OLA_N2         ...G..-|-----A............CC........C.....---....C....-|-----
OLA_M34676     .......|.....................................---..........|.....
OLA_U03092     ...G..G|...G.............................A---....C.....|.....
OLA_U03093     ...G...|...G....G.....T....C.............A---....C.....|.....
OLA_U03094     ...G...|...G..............C...............---....C.....|.....
OLA_M34675     ...G...|...G..............C.............T.T.---....C.....|.....
OLA_AY188825   ...G...|...................C...............---....C.....|.....
OLA_AY188826   ...G...|...G....C..........C.............T.T.---....C.....|.....
OLA_M34673     .......|A........TGTG..CA..C.............A---....C.....|.....
OLA_AY188824   .......|A.................C.............A---....C.....|.....
OLA_M34674     ...G...|...TGT............C...............---....C.....|-....
OLA_Cambr1     .......|.....................................---....C.....|.....
OLA_Cambr2     .......|.....................................---....C.....|.....
OLA_Cambr3     .......|.....................................---....C.....|.....
OLA_Cambr4     .......|.....................................---....C.....|.....
BoLA_3349.1    ...G...|...................C...............---....C.....|.....
BoLA_4221.1    ...G...|.TGG........A......C...............---....C.....|.T...
BoLA_A10       .......|...G..............C...............---....C.....|.....
BoLA_BL3-6     .......|...G..............C...............---....C.....|.....
BoLA_BL3-7     ...G...|...G..............C...............---....C.....|.....
BoLA_BSA       ......G|...G..............C...............---....C.....|.....
BoLA_BSC       ......G|.G.G..............C...............---....C.....|.....
BoLA_BSF       ...G...|...G.......A......C...............A---....C.....|CTGTC
BoLA_BSN       ...G...|...G.......A.......C...............---....C.....|.T...
BoLA_BSX       ...G...|.TGG........A......C...............---....C.....|.....
BoLA_D18.1     .......|...G..............C...............---....C.....|.....
BoLA_D18.2     .......|...G..............C...............---....C.....|.....
BoLA_D18.3     ...G...|...G..............C...............---....C.....|.T...
BoLA_D18.4     ...G...|.TGG........A......C...............---....C.....|.....
BoLA_D18.5     ...G...|...G.......A.......C......A.........A---....C.....|.T...
BoLA_HD1       ...G...|.TGG........A......C...............---....C.....|.T...
```

```
BoLA_HD6        ...G...|.................C.................---....C.....|.....
BoLA_HD7        ...G...|...G.............C.................---....C.....|.....
BoLA_HD15      A..G...|...G.T......A......C...............T.---....C.....|.....
BoLA_JSP.1      .......|...G.............C.................---....C.....|.....
BoLA_KN104      .......|...G..C........G...C.................---....C.....|.....
BoLA_MAN1       ...G...|...G.......A.......C.................---....C.....|.T...
BoLA_MAN2       ...G...|.................C.................---....C.....|.....
BoLA_MAN3       ...G...|.TGG........A......C.................---....C.....|.T...
BoLA_MAN8       ...G...|...G.............C.................---....C.....|.T...
BoLA_pBoLA1     .......|...G.............C.................---....C.....|.....
BoLA_pBoLA4     .......|...G.............C.................---....C.....|.....
Boran_M69206    .......|...G.............C.................---....C.....|.....
Boran_AY188806 ...G...|.TGG........A......C...............A---....C.....|.....
Boran_AY188799  .......|...G.............C.................---....C.....|.....
Boran_AY188801  .......|...G.............C.................---....C.....|.....
Boran_AY188802  ...G...|.TGG........A......C.................---....C.....|.T...
Boran_AY188803  .......|...G.............C.................---....C.....|.....
Boran_AY188804  ...G...|.TGG........A......C..............G..---....C.....|.T...
Boran_AY188805  ...G...|...G.............C.................---....C.....|.T...
Boran_M69204    .......|...G..C........G...C.................---....C.....|.....
Fbuff_AY188810  ...G...|.TGG........A......C.................---....C.....|.....
Fbuff_AY188811  .......|...G.............C.................---A...C.....|.....
Bison_AY188808  ...G...|...G.......A.......C...............A---....C.....|.T...
Bison_AY188809  .......|...G.............C.................---....C.....|.....
Wbuff_AY188812  ...G...|CTGG........A......C.................---....C.....|.....
Wbuff_AY188813  ......G|...G........G......C..............G.---....C.....|.....
Gaur_AY188814   ......N|.N.G.............C.................---....C.....|.....
Gaur_AY188815   ...G...|...G.............C.................---....C.....|.....
Yak_AY188827    ...G...|...G.............C.................---....C.....|.....
Yak_AY188828    ...G...|...G.............C.................---....C.....|.T...
Sit_AY188816    .......|...G.............C.................---....C.....|.T...
Sit_AY188817    .......|...G.............C.................---....C.....|.T...
Sit_AY188818    .......|...G.............C.................---....C.....|.T...
Sit_AY188819    .......|...G.............C.................---....C.....|.----
Sit_AY188820    ...C...|.G.G.............C...............A---....C.....|.....
Sit_AY188821    .......|...G.............C.................---A...C.....|.T...
Sit_AY188822    .......|...G.............C.................---....C.....|.....
Kudu_AY188823   .......|...G.............C.................---....C.....|.T...
Nyala_AY188829 .T.G...|...G........A.T...AC.A...A...........---....C.....|.A...
Nyala_AY188830 ...C...|.G.G.............................N..A---....C.....|.A...
Nyala_AY188831 ...G...|.................C.................---....C.....|.T...
SLA_AF014006    ...G..G|...G.............C..C........C..T..CAA..A..C..G..|.....
SLA_AF014001    ...G..G|.G.G.............C..C........C..T..CAA..A..C..G..|.....
SLA_AF014002    ...G..G|...G.............C..C........C..T..CAA..A..C..G..|.....
```

Appendix 3. Publications from the thesis

Publications in journals

***Miltiadou D.**, Ballingall K.T., Ellis S.A., Russell G.C. and McKeever D.J. (2005) Haplotype characterization of transcribed ovine major histocompatibility complex (MHC) class I genes. Immunogenetics, 57(7):499-509.*

Announcements in conferences

***Miltiadou D.**, Ballingall K.T., Ellis S.A. and McKeever D.J. (2005). Establishment of an MHC defined sheep resource flock for studying cytotoxic T cell responses to intracellular pathogens. (Proceedings of the 6th International Sheep Veterinary Congress, Crete, Greece)*

*Zhongwei C., Ballingall K.T., **Miltiadou D.**, Newlands G., MacHugh N., McKeever D.J. (2005) Proteomic analysis of ruminant Major Histocompatibility Complex (MHC) class I molecules. (Proceedings of the 1st European Farm Animal for Functional Genomics conference, Edinburgh, UK)*

*Ballingall K.T., **Miltiadou D.**, Ellis S.A., McKeever D.J. (2005) Generation and genetic characterization of MHC homozygous sheep. (Proceedings of the 1st European Farm Animal for Functional Genomics conference, Edinburgh, UK)*

***Miltiadou D.**, Ballingall K.T., Ellis S.A. and McKeever D.J. (2004). Haplotype characterization of sheep Major Histocompatibility Complex class I genes. Proceedings of the 7th International Veterinary Immunological Symposium (IVIS), Quebec, Canada*

***Miltiadou D.**, Ballingall K.T., Ellis S.A. and McKeever D.J. (2004). Characterization of transcribed Ovine Lymphocyte Antigen (OLA) class I genes by Single Strand Conformational Polymorphism (SSCP) and sequence analysis. Proceedings of the British Society for Animal Science (BSAS), York, U.K.*

Nucleotide sequences submitted in public databases

Novel nucleotide sequences submitted in DDBJ/EMBL/GenBank were given the accession numbers: AJ874673, AJ874674, AJ874675, AJ874676, AJ874677, AJ874678, AJ874679, AJ874680, AJ874681, AJ874682, AJ874683, AJ874684, AJ874685, AJ874686, AJ874687.

mmunogenetics (2005) 57: 499–509
OI 10.1007/s00251-005-0008-y

ORIGINAL PAPER

Despoina Miltiadou · Keith T. Ballingall ·
Shirley A. Ellis · George C. Russell · Declan J. McKeever

Haplotype characterization of transcribed ovine major histocompatibility complex (MHC) class I genes

Received: 14 January 2005 / Accepted: 23 May 2005 / Published online: 19 July 2005

Abstract The ovine major histocompatibility complex (MHC) remains poorly characterized compared with those of other livestock species. Molecular genetic analysis of the bovine MHC has revealed considerable haplotype and allelic diversity that earlier serological analysis had not detected. To develop cellular and molecular tools to support development of vaccines against intracellular pathogens of sheep, we have undertaken a molecular genetic analysis of four distinct ovine MHC haplotypes carried by two heterozygous Blackface rams. We have identified 12 novel class I transcripts and used a class I sequence-specific genotyping system to assign each of these transcripts to individual haplotypes. Using a combination of phylogenetic analysis, haplotype and transcript expression data, we identified at least four distinct polymorphic class I MHC loci, three of which appear together in a number of combinations in individual haplotypes. The haplotypes were further characterized at the highly polymorphic *Ovar-DRB1* locus, allowing selection of the progeny of the two founder rams for the establishment of an MHC-defined resource population.

Keywords Ovine · Class I · Major histocompatibility complex (MHC) · Haplotype · Single-strand conformational polymorphism (SSCP) · Sequence-specific primer-polymerase chain reaction (SSP-PCR)

Nucleotide sequence data reported here are available in the DDBJ/EMBL/GenBank databases under the accession numbers: AJ874673, AJ874674, AJ874675, AJ874676, AJ874677, AJ874678, AJ874679, AJ874680, AJ874681, AJ874682, AJ874683, AJ874684.

D. Miltiadou (✉) · K. T. Ballingall · G. C. Russell ·
D. J. McKeever (✉)
Moredun Research Institute,
Pentlands Science Park, Bush Loan, Penicuik,
Midlothian, EH26 0PZ, UK
e-mail: Declan.Mckeever@mri.sari.ac.uk
Tel.: +44-131-4455111
Fax: +44-131-4456235

S. A. Ellis
Institute for Animal Health,
Compton, RG20 7NN, UK

Introduction

The highly polymorphic major histocompatibility complex (MHC) class I genes encode glycoproteins expressed at the cell surface in association with peptides derived from intracellular pathogens or from self proteins. Infection is detected by $CD8^+$ T lymphocytes which recognise the combination of self MHC and foreign peptide (reviewed by Harty et al. 2000). The majority of polymorphic sites within classical MHC class I genes encode amino acids that line the peptide binding groove (Bjorkman et al. 1987). The repertoire of peptides presented to T cells in an individual is dependent on the complement of class I molecules inherited on each haplotype. This diversity in the repertoire of peptides presented to T cells can influence the quality of the immune response to infection. As a consequence, polymorphic MHC loci are often used as genetic markers in studies of disease resistance or susceptibility and in cases where variations are observed in responses to infection (Hill et al. 1991; Messaoudi et al. 2002) or vaccination (Eckels 2000).

Three polymorphic classical class I loci, HLA-A, -B and -C, are present on all human MHC haplotypes (reviewed by Trowsdale 1995). Such consistency appears to be the exception rather than the norm for other species. For example, the number of class I genes varies in different inbred strains of mouse (reviewed by Trowsdale 1995) and variable numbers of MHC class I genes have been reported in different serologically defined swine haplotypes (Ruohonen-Lehto et al. 1998). Similarly, although two polymorphic genes are ubiquitously expressed in equids, at least four putative non-classical genes may also be expressed at low levels (Ellis et al. 1995; Holmes and Ellis 1999). In cattle, there is evidence for at least six classical MHC class I loci (reviewed by Ellis 2004), which occur in a number of combinations on different haplotypes (Ellis et al. 1999).

Haplotype variation may also be a feature of ovine class I genes. A number of serological studies have suggested two expressed loci (Cullen et al. 1982; Garrido et al. 1995; Jugo and Vicario 2001; Jugo et al. 2002; Stear and Spooner 1981), whereas others propose three (Grossberger et al.

1990; Millot 1984). In addition, evidence based on isoelectric focusing (IEF) suggests that serology has underestimated the level of ovine MHC diversity (Jugo et al. 2002). Antigens not previously defined by serology were identified by IEF and three serological specificities were subdivided by biochemical typing. Only a single molecular study has been carried out on the MHC class I region of sheep (Grossberger et al. 1990). This study identified five different sequences from a cDNA library, which is consistent with transcription of at least three loci.

Development of vaccines against intracellular pathogens of sheep is currently constrained by the lack of reagents necessary for analysis of class I MHC restricted T-cell responses. To address this deficiency, we have embarked on the establishment of a resource population of MHC-defined sheep and a corresponding array of molecular and cellular reagents. Given the lack of information on ovine MHC haplotype diversity, we have undertaken an extensive molecular characterization of the haplotypes carried by two heterozygous rams used to found the population. We first identified the class I genes transcribed by each ram, using polymerase chain reaction-single strand conformational polymorphism (PCR-SSCP) and sequence analysis. Based on these data, we developed a sequence-specific primer–polymerase chain reaction (SSP–PCR) system for genotyping a cohort of progeny from each ram to allow the allocation of transcripts to individual haplotypes. By supplementing these data with genotyping of the major transcribed and highly polymorphic MHC class II *DRB* locus of each haplotype, we provide the first detailed description of four ovine MHC haplotypes.

Materials and methods

Animals, RNA extraction and cDNA synthesis

Rams 501 and 504 and their progeny were selected from a Scottish Blackface flock maintained at the Moredun Research Institute (Table 1). Peripheral blood mononuclear cells (PBMC) were prepared using a standard Ficoll–Hypaque (Pharmacia, Uppsala, Sweden) separation method (Boyum 1968). Poly-adenylated mRNA was extracted from PBMC obtained from the rams using the Dynabeads mRNA direct kit (Dynal Oslo, Norway), according to the manufacturer's instructions. Total RNA was prepared from 1×10^7 PBMC obtained from all progeny, using the Qiagen RNeasy mini kit (Qiagen Ltd., West Sussex, UK). First strand cDNA was prepared using the Superscript first-strand synthesis system (Invitrogen Ltd., Paisley, UK).

Nomenclature

The genetic region containing the ovine MHC is termed the ovine leukocyte antigen (OLA) region. However, in order to discriminate between the five species within the genus *Ovis*, MHC genes derived from the domestic sheep *Ovis aries* are prefixed with *Ovar* (Klein et al. 1990).

Amplification of ovine MHC class I genes

Amplification of partial ovine MHC class I transcripts prior to SSCP analysis

PCR primers were designed to amplify part of exons 2 and 3 of ovine MHC class I genes. Target sequences were located in conserved areas of published bovine (http://www.ebi.ac.uk/ipd/mhc/bola/index.html) and ovine class I sequences (accession nos. M34672 to M34676, U25998 to U26010, U030092 to U030094 and AY188824 to AY18 8826). Three pairs of primers (A and Ar, B and Br, C and Cr; Table 2) were used to maximize the likelihood of amplifying transcribed class I genes. Primer A had previously been used to amplify cattle class I genes (Bov7; Ellis et al. 1999). Primers B and C were designed in the same area as primer A, but provided reduced 3′ end complementarity and closer Tm with their corresponding reverse primers. The primer Ar is identical to Bov11 (Pichowski et al. 1996) but was designed with a degenerate nucleotide (C/T) at the third base from the 3′-end to complement the

Table 1 Pedigree of animals used in this study and class I and class II genotyping results

Animal	Class I sequences	Class II *DRB* PCR-RFLP	Animal	Class I sequences	Class II *DRB* PCR-RFLP
Ram 501	D3, D5, C6, C7, C11, E2	AAA ABB	Ram 504	F8, F12, E9, F9, G6, H10	BCA CAA
Progeny 501			Progeny 504		
Lamb 448	D5, C7, C11	AAA	Lamb 457	F8, G6, E9	BCA
Lamb 449	D3, E2, C6	ABB	Lamb 529	F12, H10, F9	CAA
Lamb 461	D5, C7, C11	AAA	Lamb 530	F12, H10, F9	CAA
Lamb 486	D3, E2, C6	ABB	Lamb 573	F8, G6, E9	BCA
Lamb 489	D3, E2, C6	ABB	Lamb 596	F8, F12, E9, F9, G6, H10	BCA CAA
Lamb 518	D3, E2, C6	ABB	Lamb 609	F8, G6, E9	BCA
Lamb 555	D3, E2, C6	ABB	Lamb 610	F8, G6, E9	BCA
Lamb 637	D3, E2, C6	ABB	Lamb 673	F12, H10, F9	CAA
Lamb 680	D3, E2, C6	ABB			

PCR-RFLP polymerase chain reaction-restriction fragment length polymorphism

Table 2 Oligo nucleotide primers used for amplification of ovine MHC class I sequences

Name	5′ Forward primer 3′	Location	Name	5′ Reverse primer 3′	Location	Amplicon size (bp)
A[a]	GGCTACGTGGACGACACG	Exon 2	Ar[a]	CCCTCCAGGTAGTTYCT	Exon 3	410
B[a]	GCTACGTGGACGACACGC	Exon 2	Br[a]	AGCGCAGGTCCTCGTTC	Exon 3	318
C[a]	CGGCTACGTGGACGACAC	Exon 2	Cr[a]	ATGGGTCACATGTGYCTTTG	Exon 4	499
Afl[b]	ATGACCAGAGGATTGCGRGT	Exon 1	Dr1[b]	GGATGAAGCATCACTCAG	3UTR	1,135
Bfl[b]	ATGCGAGTCATGGGGCC	Exon 1	Dr1[b]	GGATGAAGCATCACTCAG	3UTR	1,123
Cfl[b]	ATGGGGCCGCGAACCSTC	Exon 1	Dr2[b]	GGATGAAGCATCACTGTC	3UTR	1,098 or 1,104
D5f[c]	GATCGGAACACGCGAAAC	Exons 2, 3	D5r[c]	AGCCGTACATCCACTGCC	Exon 3	121
D3f[c]	GGATCGAAACACGAGGATCTA	Exons 2, 3	D3r[c]	TCGTAGCCGAGCTGATCAT	Exon 3	177
C7f[c]	GGATCAGGAGACGCGATCA	Exons 2, 3	C7r[c]	CAGGTAGATCCTCACACGCG	Exons 3, 4	301
E2f[c]	ACGAAAGGCCAACGATGA	Exons 2, 3	E2r[c]	GCACCTGCCACCACAACA	Exon 3	264
C6f[c]	GGGATCAGGAGACGAGAATCTA	Exons 2, 3	C6r[c]	CACCTGCCGCCTCAAT	Exon 3	276
C11f[c]	CAAGCGCAAGATTGAGGCT	Exon 3	C11r[c]	CCAGGTATCTGCGGAGCAT	Exon 4	89
F8f[c]	GATCGGGAAACTCGAAACAT	Exons 2, 3	F8r[c]	CTCCAGGTAGTTCCTTTCTGC	Exons 3, 4	303
F12f[c]	GAAGGGCCAAGGGCAA	Exon 3	F12r[c]	GCCGTACATCCACTGCCA	Exon 3	107
H10f[c]	TCAGGAGACGCAGGGAACT	Exons 2, 3	H10r[c]	CACCTGCCACCACAGCA	Exon 3	272
F9f[c]	GGATCAGGAGACGCGATT	Exons 2, 3	F9r[c]	TGGTTGTAGTAGCCGAGTGC	Exon 3	81
G6f[c]	GGATAGCCAAGGACGGCAT	Exon 3	G6r[c]	GGTGTCTGCGGAGCAAGTT	Exon 4	321
E9f[c]	GAGACGCAGGGAACTAAGGACA	Exons 2, 3	E9r[c]	CCGTCGTAGGCGAACTGATC	Exon 3	173

3UTR 3 untranslated region
[a]Oligonucleotide primers for amplification of parts of exons 2 and 3
[b]Oligonucleotide primers for amplification of the entire coding region
[c]Sequence specific primers

available diversity in ovine class I sequences. The primer Br is generally conserved among different species, but does not amplify all of exon 3. For this reason, primer Cr was designed to allow amplification of the entire exon 3. PCR was carried out in a final volume of 50 µl containing 50 mM KCl, 10 mM Tris–HCl, 1.5 mM $MgCl_2$, 0.1% TritonX-100, 200 µM of each dNTP, 200 nM of each primer, 2 µl cDNA and 1 unit *Taq* polymerase (Promega, Madison, WI). The thermal cycling profile included an initial denaturation for 5 min at 94°C; 30 cycles of 1 min at 94°C, 1 min at 50°C and 1 min at 68°C, and a final extension for 7 min at 68°C. The reactions were carried out in a 96-well Hybaid MBS 0.2S thermal cycler (Thermo Hybaid, Basingstoke, UK).

Amplification of full-length ovine MHC class I transcripts

Amplification of full-length transcripts was carried out by using three different forward (Afl, Bfl and Cfl; Table 2) and two reverse primers (Dr1 and Dr2; Table 2). Primers Bfl and Dr1 are identical to Bov21 and Bov3 described by Pichowski et al. (1996). The remaining primers were designed using available ovine and bovine class I sequences. PCR was carried out in a final volume of 50 µl containing 49 µl PCR master mix (60 mM Tris–SO_4, 18 mM $(NH_4)_2SO_4$, 200 µM of each dNTP, 200 nM of each primer and 1 U high-fidelity Platinum *Taq* DNA polymerase; Invitrogen) and 1 µl cDNA (≈150 ng). The cycling profile for amplification of full-length transcripts was as described by Parham et al. (1995) to reduce formation of recombinant alleles: initial denaturation for 2 min at 94°C, 25–30 cycles of 1 min at 94°C, 30 s at 50°C and 2 min at 68°C, and final extension for 7 min at 68°C. The annealing time was reduced to 10 s when Dr2 was used as a reverse primer. The reactions were carried out in a 96-well Hybaid MBS 0.2S thermal cycler (Thermo Hybaid).

SSCP and sequence analysis of ovine MHC class I genes

Exon 2 and 3 PCR products were purified from 0.7% agarose gels using the QIAquick gel extraction kit (Qiagen) and cloned into the pGEM-T-Easy vector (Promega). Colony PCR using primers B and Br was used to identify 180 positive clones per ram (60 clones per primer pair). These were each screened by SSCP according to the method described by Ainsworth et al. (1991) with slight modifications. Briefly, 10 µl of each colony PCR product was mixed with 5 µl formamide denaturing solution. The mixtures were denatured at 94°C for 8 min and immediately cooled on ice for 5 min. Ten microliters was loaded onto a 6% polyacrylamide gel without glycerol. Electrophoresis was carried out in a Hoefer SE 600 vertical gel electrophoresis unit (Amersham Biosciences, Buckinghamshire, UK) for 16 h at 14 mA and then for 1 h at 50 mA at a constant temperature of 4°C. Single-stranded DNA fragments were detected by silver staining (Ainsworth et al. 1991). Nucleotide sequence was obtained from up to ten replicates for each SSCP pattern.

PCR products of the entire coding region were also cloned as described above and sequencing data was ob-

tained from at least three replicates. Sequencing was performed using a 377 ABI Prism DNA sequencer (Perkin Elmer, USA), or a 3730XL DNA Analyser (Applied Biosystems, USA).

Genotyping by SSP–PCR

Progeny of rams 501 and 504 were genotyped to determine from which parental haplotypes individual genes were derived. Sequence specific primers were designed for specific amplification of each of the sequences presented in Fig. 1 on the basis that at least the 3′ terminal nucleotide of either forward or reverse primer (or ideally both) was allele-specific. Where possible, ovine class I MHC gene sequences available in public databases were considered in the primer design to avoid cross-priming with published sequences.

PCR reactions were carried out in a final volume of 50 µl containing 50 mM KCl, 10 mM Tris–HCl, 2 mM $MgCl_2$,

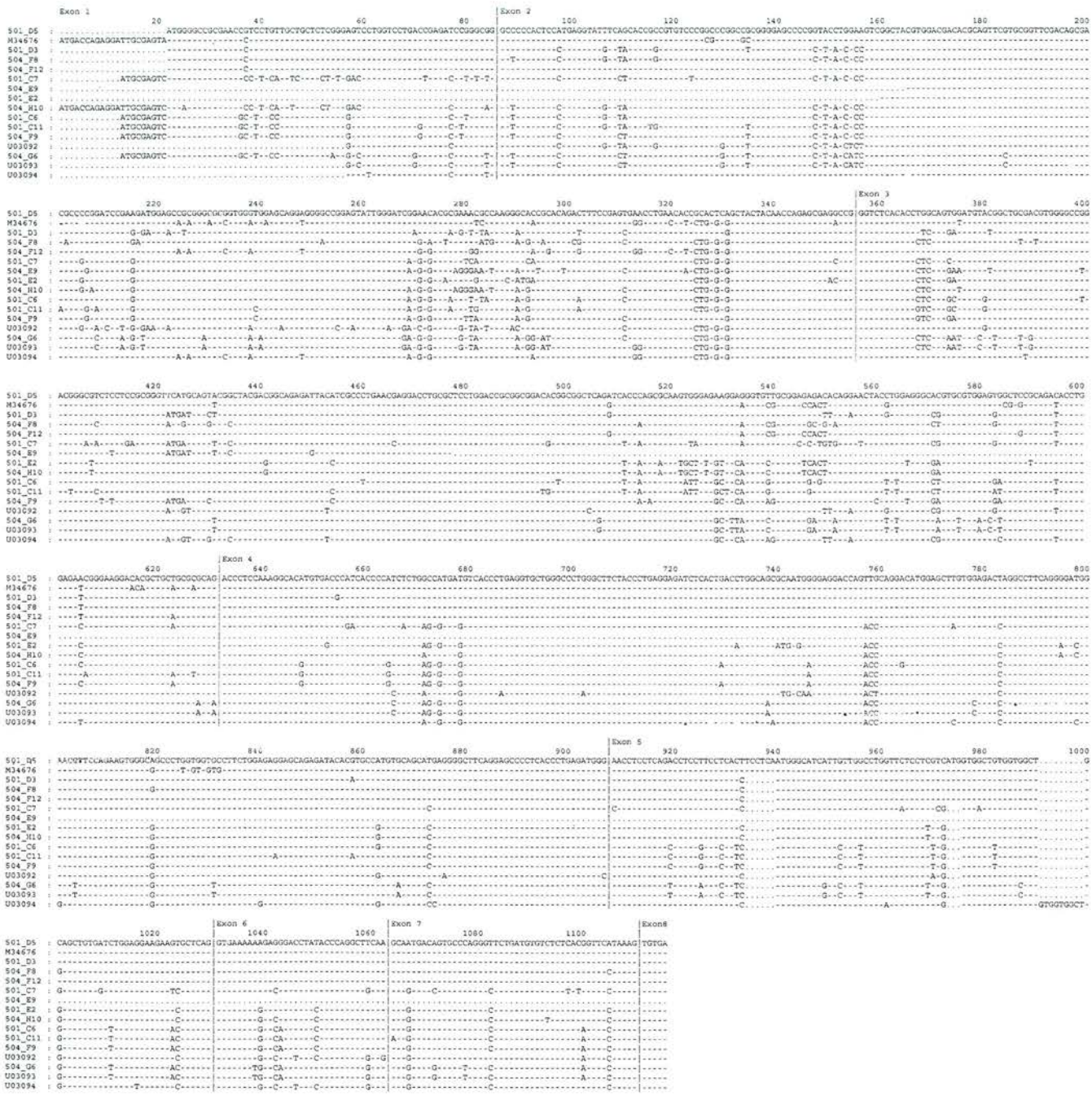


Fig. 1 Alignment of ovine class I transcripts identified from rams 501 and 504 in comparison with the full-length ovine class I sequence M34676 (Grossberger et al. 1990) and the nearly full length sequences U03092 through U03094. *Dashes* indicate identities and *dots* represent gaps or absence of sequence

200 μM of each dNTP, 200 nM of each sequence specific primer, 2 units *Taq* polymerase (Roche Diagnostics Penzberg, Germany) and 0.5 μl cDNA (approximately 100 ng). The thermal cycling profile included an initial denaturation for 1 min at 94°C; 5 cycles of 1 min at 94°C, 1 min at 65°C and 2 min at 72°C followed by 25 cycles of 1 min at 94°C, 1 min at 60°C and 2 min at 72°C (with the exception of sequences C11, F9 and E9, which were amplified for 30 cycles) and a final extension for 10 min at 72°C. The identity of the SSP–PCR products was confirmed by cloning and sequence analysis.

Ovine MHC class II *DRB* genotyping

Exon 2 of the highly polymorphic *Ovar-DRB1* locus was genotyped by PCR-restriction fragment length polymorphism (RFLP). Genomic DNA was prepared from whole blood, using the Easy-DNA kit (Invitrogen). Amplification was performed with primers Ov30 (5′-atcccctctctgcagca catttc-3′) and Ov31 (5′-cccgcgctcacctcgccgctg-3′) in 50 μl PCR buffer containing 2 mM $MgCl_2$, 200 μM of each dNTP, 500 nM of each primer and 1 unit of *Taq* polymerase (Promega). The thermal cycling profile included initial denaturation for 2 min at 94°C; 32 cycles of 30 s at 94°C, 30 s at 52°C and 30 s at 72°C and a final extension for 4 min at 72°C. PCR products of 296 bp were cloned and sequenced as described above. The sequence of the four *Ovar-DRB* alleles obtained from rams 501 and 504 allowed selection of restriction enzymes *Rsa*I, *Mbo*II and *Sac*I to produce characteristic PCR-RFLP patterns following separation on 8% polyacrylamide gels. This PCR-RFLP approach allowed rapid genotyping of the progeny of rams 501 and 504.

Sequence and phylogenetic analysis

Nucleotide sequence was assembled, edited, and aligned using the SeqMan and MegAlign programs (DNASTAR, Madison, WI). Neighbour-joining phylogenetic trees (Saitou and Nei 1987) were constructed, using either genetic distances of the entire coding regions calculated by the Kimura two-parameter method (Kimura 1980) or p distances of exons 4 through 8. The swine sequence AF014002 was used as an outgroup to root the trees. The phylogenetic analysis was performed using MEGA2.1 software (Kumar et al. 2001).

Results

SSCP and sequence analysis of ovine MHC class I genes

PCR amplicons obtained from rams 501 and 504 were cloned and screened by SSCP and those clones showing distinct patterns were selected for sequence analysis. Six different SSCP patterns were identified from 180 clones analysed from ram 501 and eight from the same number of clones analysed from ram 504. SSCP patterns for sequences D5, D3, C7 and E2 accounted for 89% of all clones analysed from ram 501, while C6 and C11 represented only 4%. Similarly, SSCP patterns for sequences F8 and F12 constituted 83% of the total clones identified from ram 504, while the remaining six patterns accounted for only 14%. The remainder comprised PCR artefacts as determined by sequence analysis.

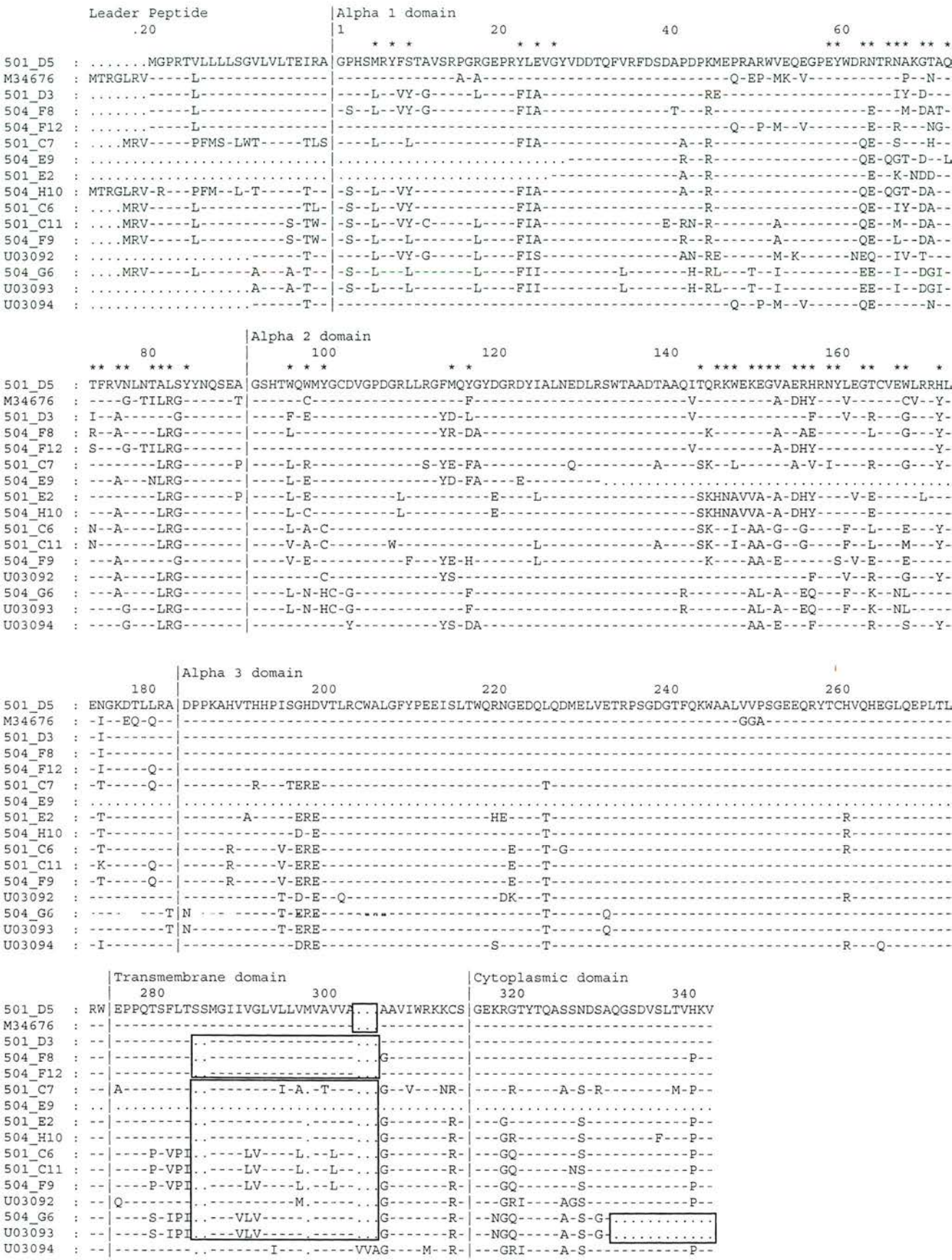
Although some caution is required in the extrapolation of PCR and cloning frequencies to expression levels, the use of three different sets of PCR primers on the most conserved areas possible, thereby amplifying fragments of similar lengths, minimised preferential primer amplification and variation in cloning efficiencies. Cloning frequencies should therefore reflect transcription levels.

Sequence analysis confirmed that the amplified products were ovine MHC class I transcripts, with six distinct sequences being identified from each ram (Fig. 1). Because two of the SSCP patterns identified from ram 504 were found in only single clones and could not be validated, their corresponding sequences are not presented. The remaining class I sequences were each identified in at least three replicate clones. Sequence data for the entire coding region was obtained for ten of these. All sequences were novel and were submitted to the EMBL database with the accession numbers shown in parenthesis: D5 (AJ874673), D3 (AJ 874674), C7 (AJ874675), E2 (AJ874676), C6 (AJ874677), C11 (AJ874678), F8 (AJ874679), F12 (AJ874680), H10 (AJ874681), F9 (AJ874682), G6 (AJ874683), E9 (AJ874684).

Sequence data for only part of exons 2 and 3 were obtained for the E9 sequence. This may simply reflect low frequency of the transcript—only three replicates were obtained in 180 clones analysed by SSCP. Alternatively, failure to amplify full-length sequence for E9 could have arisen through poor primer binding due to unresolved diversity in the primer annealing sites. The latter possibility is supported by the observation that the incomplete 5′ sequence for the E2 gene resulted from failure of the forward primer binding (data not shown).

An alignment of the 12 nucleotide sequences, together with the single published full-length ovine MHC class I sequence (Grossberger et al. 1990; Fig. 1) and three near full-length sequences (U030092–U030094), shows the patchwork nature of nucleotide substitutions characteristic of MHC loci in other species. As expected, most of the nucleotide diversity was situated in regions encoding amino acids of the class I polypeptide that bind antigenic peptides or interact with the T-cell receptor (Fig. 2).

The predicted amino acid sequences of the class I transcripts comprised a leader peptide and α1, α2, α3, cytoplasmic (C) and transmembrane (TM) domains. The TM domains varied in length, with most being 35 residues in length. The D5 sequence had the longest TM, at 38 residues, while those of D3, F8 and F12 were intermediate in length at 36 residues. With the exception of the G6 sequence, which was truncated by an in-frame stop codon, all C domains were 28 residues in length (Fig. 2).


Leader Peptide
Alpha 1 domain
Alpha 2 domain
Alpha 3 domain
Transmembrane domain
Cytoplasmic domain
501_D5
M34676
501_D3
504_F8
504_F12
501_C7
504_E9
501_E2
504_H10
501_C6
501_C11
504_F9
U03092
504_G6
U03093
U03094

◀ **Fig. 2** Alignment of the predicted amino acid sequences derived from the ovine MHC class I transcripts identified from rams 501 and 504 in comparison with the full-length ovine class I sequence M34676 and the nearly full length sequences U03092 through U03094. *Dashes* indicate identities and *dots* gaps or absence of sequence. *Asterisks* indicate MHC class I residues predicted to bind antigenic peptides or interact with T-cell receptors in humans (Bjorkman et al. 1987). Similarities among sequences caused by deletions in the transmembrane domain or by early stop codons in the cytoplasmic domain are indicated by *boxes*

Genotyping the progeny of rams 501 and 504 by SSP-PCR

Validation of the method

SSP pairs (Table 2) were designed in order to assign each of the class I sequences to individual MHC haplotypes. Each SSP pair was selected on the basis that it amplified a product only from the clone that it was designed to amplify. SSP specificities were further validated on cDNA from rams 501 and 504. In addition, SSP pairs designed to amplify class I sequences from ram 501 were also tested on cDNA prepared from ram 504 and vice versa. None of the sequences previously identified in at least three clones was shared between animals. As cross reaction with previously unidentified sequences was considered possible, sequencing of the SSP–PCR products was performed to confirm the identity of the amplicons. In cases where it was suspected that SSPs cross-reacted with maternally inherited class I sequences, the ewes were genotyped. In all such cases, the class I sequences of the ewes differed by at least 4 bp from the ovine MHC sequences of the rams (data not shown), confirming that the amplified product was maternally inherited.

Genotyping results

Seventeen lambs sired by rams 501 and 504 were genotyped by SSP–PCR (Table 1; Fig. 3). Seven progeny of 501 (449, 486, 489, 518, 555, 637, and 680) provided sequences D3, E2 and C6, while two (448 and 461) provided sequences D5, C11 and C7. Four offspring of 504 (457, 573, 609 and 610) provided sequences F8, G6 and E9, while three (529, 530 and 673) provided sequences F12, H10 and F9. One lamb (596) sired by 504 yielded all six paternal sequences, indicating that both the ram and the ewe shared an MHC haplotype.

Phylogenetic analyses

The relationships between ovine class I sequences were visualised by constructing phylogenetic trees using, where available, the entire coding regions (Fig. 4a), or exons 4 through 8 (Fig. 4b). With the exception of U030094, the full-length sequences group into three distinct clusters (Fig. 4a). The tree topology supports the haplotype assignment of individual sequences from rams 501 and 504, with each sequence within a haplotype associated with a different cluster. In the majority of cases, the segregation of the sequences is in concordance with their estimated level of transcription. The higher-frequency sequences D5, D3, F8 and F12 group in cluster I (Fig. 4a,b), while the lower-frequency sequences C6, C11, F9 and G6 group in cluster III (Fig. 4a,b). These observations are consistent with correspondence of these clusters to individual loci. In contrast, cluster II contains sequences with a range of transcription levels (Fig. 4a). For example, C7 and E2 were identified in approximately twice as many clones as H10, while E9 was identified in only three clones.

By excluding the influence of positive selection on tree topology, phylogenetic analysis using only exons 4 to 8 provides a more accurate assignment of sequences to loci (Holmes et al. 2003). Clusters I and III of the phylogenetic tree presented in Fig. 4b are almost identical to those shown in Fig. 4a, thus strengthening their designation as distinct loci. However, this analysis subdivided the sequences in cluster II, with C7 segregating with cluster IIa, and E2 and H10 segregating with cluster IIb, suggesting that these sequences may represent two additional loci. The partial class I sequences M34673 and AY188824 also grouped together, but separately from clusters I, IIa, IIb and III, and may therefore represent alleles at a fifth locus. The sequence AY188825 located relatively deep on the phylogenetic tree is quite divergent from the other sequences and may therefore represent a distant non-classical locus. Sequences U03094 and U03095 grouped in cluster IIb in Fig. 4b, suggesting that their exons 4 through 8 are more similar to the H10 and E2 transcripts. However, no firm assertions can be made for cluster IIb, as the confidence values were low. A more accurate identification of the

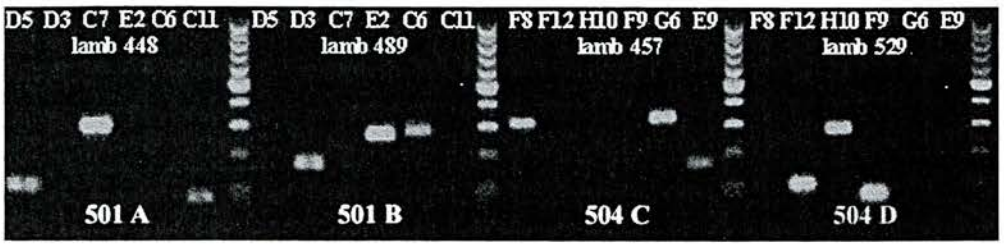


Fig. 3 Sequence-specific primer-polymerase chain reaction analysis of the progeny of rams 501 and 504. From left to right, lamb 448 inherited 501 haplotype A (*lanes 1–6*), lamb 489 inherited 501 haplotype B (*lanes 8–13*), lamb 457 inherited 504 haplotype C (*lanes 15–20*) and lamb 529 inherited 504 haplotype D (*lanes 22–27*). A 100-bp DNA ladder was loaded in *lanes 7, 14, 21 and 28*

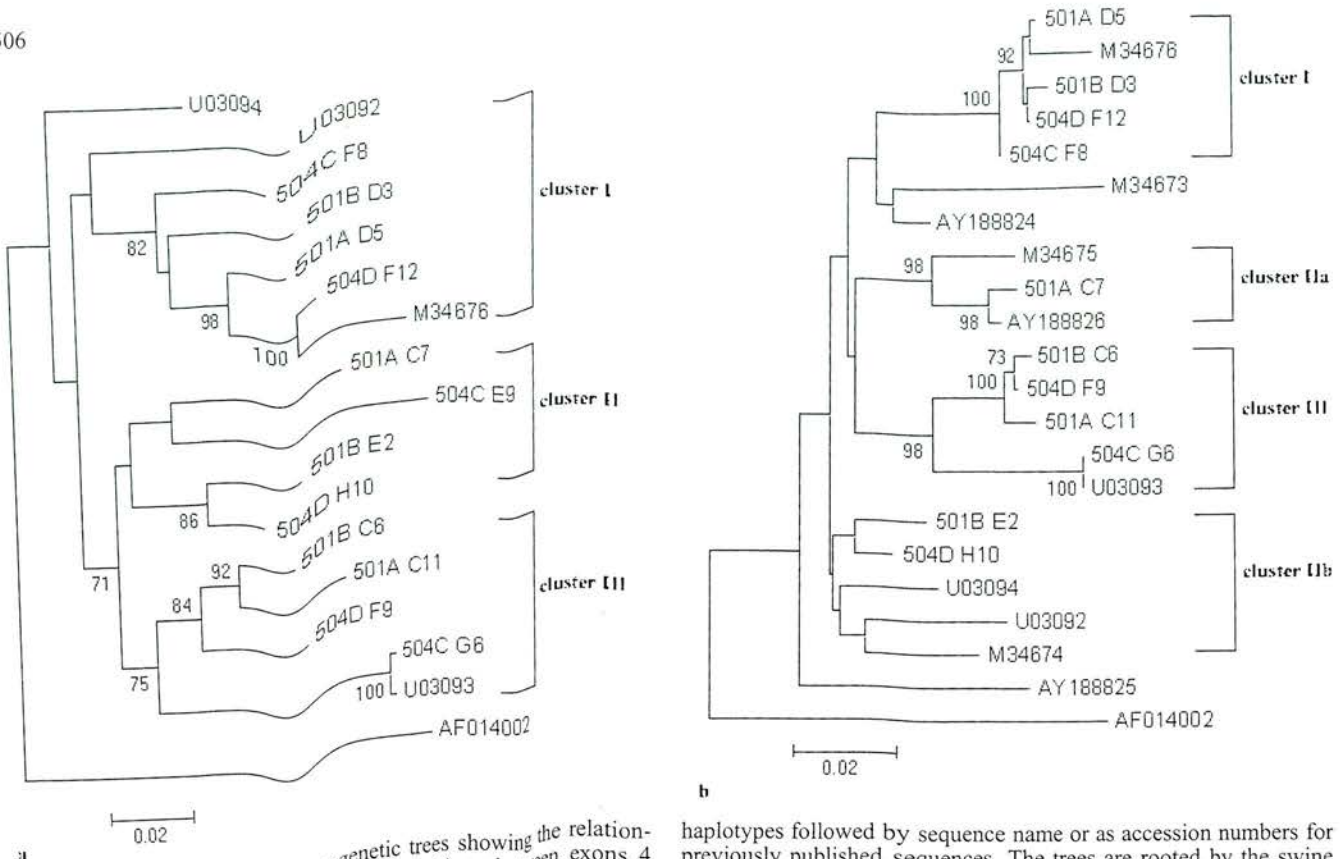


Fig. 4 Neighbour-joining phylogenetic trees showing the relationships between the entire coding regions (**a**) or between exons 4 through 8 (**b**) of previously published ovine MHC class I transcripts and those described here. Sequence names are presented as ram haplotypes followed by sequence name or as accession numbers for previously published sequences. The trees are rooted by the swine MHC class I sequence AF014002. Bootstrap values exceeding 70% representing the confidence attached to each node are shown

number of transcribed ovine MHC class I loci will be achieved only by greatly expanding the number of class I sequences derived from a range of additional ovine MHC haplotypes.

Class II genotyping and the extended MHC haplotypes

We wished to extend the characterization of the four MHC haplotypes to include a gene within the class II region of the ovine MHC to provide a simple and rapid means of genotyping progeny of the two rams using genomic DNA. Exon 2 of the highly polymorphic *Ovar-DRB1* locus was amplified from genomic DNA prepared from rams 501 and 504. Two *Ovar-DRB1* alleles were identified from 501 that correspond to previously described alleles *Ovar-DRB1*0101* (Konnai et al. 2003) and *Ovar-DRB14* (Schwaiger et al. 1994). Two alleles were also identified from 504, which correspond to *Ovar-DRB33* (Schwaiger et al. 1994) and *Ovar-DRB1*0141* (Konnai et al. 2003). Digestion of the PCR products with restriction enzymes *Rsa*I, *Mbo*II and *Sac*I (Table 3) generated allele-specific patterns that allowed discrimination between all four alleles in the progeny of rams 501 and 504. The genotyping of the progeny is shown in Table 1, and the four extended ovine MHC haplotypes are represented in Fig. 5.

Table 3 PCR-RFLP genotyping of the *Ovar-DRB1* alleles from rams 501 and 504

Allele	*Rsa*I fragment size (bp)	*Mbo*II fragment size (bp)	*Sac*I fragment size (bp)	PCR-RFLP type
501 Allele 1	296 (A)	179, 117 (A)	296 (A)	AAA
501 Allele 2	296 (A)	179, 73, 44 (B)	202, 94 (B)	ABB
504 Allele 1	141, 62, 54, 39 (B)	90, 89, 73, 44 (C)	296 (A)	BCA
504 Allele 2	111, 62, 54, 39, 30 (C)	179, 117 (A)	296 (A)	CAA

Ram 501 allele 1 corresponds to *Ovar-DRB1*0101* (Konnai et al. 2003). Ram 501 allele 2 corresponds to *Ovar-DRB14* (Schwaiger et al. 1994). Ram 504 allele 1 corresponds to *Ovar-DRB33* (Schwaiger et al. 1994). Ram 504 allele 2 corresponds to *Ovar-DRB1*0141* (Konnai et al. 2003)

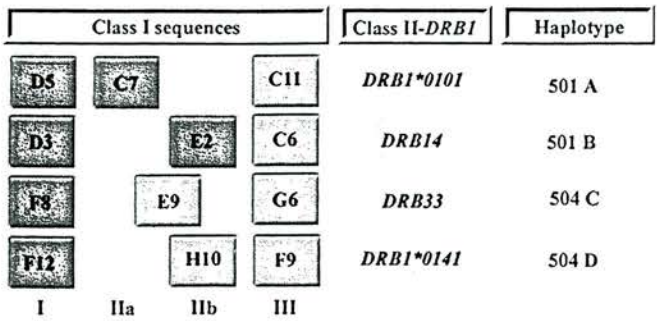


Fig. 5 A diagrammatical representation of the four ovine MHC haplotypes identified from rams 501 and 504 showing the class I sequences and the associated class II *DRB1* PCR-RFLP pattern. The arrangement of the class I sequences reflects the clustering of the alleles according to the phylogenetic trees presented in Fig. 3. The assignment of E9 within cluster IIa or IIb was not determined due to lack of sequence data for exons 4 through 8. The sequences in *dark grey boxes* were represented in 86% of the 360 clones analysed by SSCP

Discussion

A number of genetic studies in livestock and other species have revealed significantly greater genetic complexity and allelic diversity in the MHC class I region than initially predicted by serology (reviewed by Chardon et al. 1999; Ellis 2004; Otting et al. 2005; Parham et al. 1995). Serological studies of the segregation of allotypic determinants in families is consistent with two linked ovine class I loci (Cullen et al. 1982; Garrido et al. 1995; Jugo and Vicario 2001; Jugo et al. 2002; Stear and Spooner 1981). Other researchers have suggested that the ovine MHC contains at least three class I loci (Grossberger et al. 1990; Millot 1984). Prior to molecular genetic analyses, it was believed that the majority of MHC class I molecules in cattle were encoded by a single highly polymorphic locus (reviewed by Ellis and Ballingall 1999). However, molecular genetic studies have since revealed evidence for at least six classical bovine class I loci, which occur in a number of combinations, and none of which is consistently transcribed (reviewed by Ellis 2004).

In this study, we identified 12 novel ovine MHC class I transcripts from two heterozygous rams. By genotyping their progeny, we were able to assign these transcripts to individual haplotypes. Three class I sequences were identified in each haplotype, suggesting the transcription of at least three ovine MHC class I loci. Sequences D5, D3, C7, E2, F8 and F12, which accounted for 86% of the 360 clones analysed, are likely to represent highly expressed polymorphic classical class I loci. The remaining sequences, which occurred in much smaller numbers of clones (9%) screened by SSCP, may represent a mixture of poorly expressed classical or non-classical loci. Interestingly, whereas two highly transcribed genes were identified in each 501 haplotype (Fig. 5), only one was found in each 504 haplotype, indicating that variation may occur among haplotypes in the number of expressed classical class I loci.

The combination of haplotype and phylogenetic analysis described here suggests that these transcripts were products of at least four loci, three of which appear together in a number of combinations in individual haplotypes (Figs. 4b, 5). When additional sequences become available, clusters I and III may also be further divided. In this regard, cluster I sequences D5 and M34676 encode TM domains that are two residues longer than the remaining sequences in the cluster (Fig. 2). In line with the observation that TM length tends to be conserved within bovine class I loci (Garber et al. 1994), these sequences may represent alleles at a distinct locus.

Sequences in cluster III may also be subdivided further based on the length of the C domain (Fig. 2). The G6 sequence contains an early stop codon within the cytoplasmic domain that removes the conserved phosphorylation site at positions 330–334 (Guild and Strominger 1984). The previously described class I sequence U03094 contains an identical C domain and differs from G6 by only 3 bp, suggesting that these sequences may be alleles at a single locus that exhibits limited polymorphism. Phylogenetic analysis clusters G6 and U03094 very close together and separates them from the remaining sequences in cluster III by a long branch. Furthermore, G6 was identified in only 1% of the 180 clones analysed by SSCP, suggesting that it is weakly transcribed. Since non-classical class I sequences often encode truncated C domains (Lew et al. 1987; Ulker et al. 1990), exhibit limited or no polymorphism, and are often expressed at low levels or in a tissue-specific manner (Braud et al. 1999), G6 and U03094 are strong candidates for alleles at a non-classical locus. Additionally, G6 exhibited the lowest third position G+C content of exons 2 and 3 when compared with the remaining sequences (data not shown), a characteristic associated with non-classical class I sequences in many species (Hughes and Nei 1989).

All sequences in cluster III are transcribed at low levels and encode the amino acids VPI or IPI at positions 281–283, discrimination motifs previously suggested to be characteristic of non-classical bovine class I genes (Holmes et al. 2003). However, haplotype and phylogenetic analysis strongly indicate that transcripts C6, C11 and F9 represent polymorphic alleles at a single locus, which raises questions about their non-classical nature. A number of classical bovine class I loci expressed at levels significantly lower than usual have been described, although this was not a locus-specific phenomenon (reviewed by Ellis 2004). In humans, alleles of the HLA-C locus exhibit less polymorphism (Falk and Schendel 1997) and are expressed at significantly lower levels compared to HLA-A and HLA-B (Snary et al. 1977; McCutcheon et al. 1995). These genes are involved in regulation of NK cell activity more than in antigen presentation (Falk and Schendel 1997). Whether sequences in cluster III encode proteins with similar functions to HLA-C molecules needs to be determined by functional studies.

It therefore seems that at least two putative classical class I loci are highly expressed in each of 501 haplotypes A and B, along with a third locus expressed at low levels, which requires further characterisation by functional analysis. In contrast, 504 haplotype C includes only a single highly expressed classical class I locus, since G6 possibly rep-

resents a non-classical locus and the E9 gene is predicted to be transcribed at very low levels. Similar haplotypes with single classical loci have been described for cattle, chicken and inbred lines of rat (reviewed by Ellis 2004). The 504 haplotype D also contains one putative highly expressed classical class I locus and two additional loci with characteristics not clearly associated with either classical or non-classical genes.

Since it is likely that cluster I contains sequences from two classical loci, whereas cluster III may be divided into a classical and a non-classical locus, the transcripts identified in this study may represent alleles of six loci, of which at least three are classical. Given that some published sequences failed to cluster with our transcripts, the number of ovine MHC class I loci may exceed six. Confirmation of this possibility could arise from a more comprehensive phylogenetic analysis using a much larger sample of ovine class I sequences from well-defined haplotypes. Experience from phylogenetic analysis of MHC class I sequences from Cetartiodactyls, however, suggests that assignment of alleles to loci in ruminants is not as straightforward as in humans, due to inter-locus recombination (Holmes et al. 2003). It is therefore possible that proper ovine class I allele designations will ultimately be confirmed only by physical mapping.

In summary, the molecular genetic analysis presented here identified 12 novel MHC class I transcripts from two heterozygous rams. A class I sequence-specific genotyping system allowed the assignment of these transcripts to four haplotypes and phylogenetic analysis of the more conserved exons 4–8 grouped the transcripts into four clusters. The combination of phylogenetic analysis, haplotype data and transcription levels suggest that the number and composition of ovine MHC loci vary between haplotypes. By including information on the highly polymorphic class II MHC *Ovar-DRB1* locus, we have defined for the first time four extended ovine MHC haplotypes. These provide an initial template for more extensive characterization of ovine MHC haplotypes and allelic diversity in different breeds and populations. These haplotype data have facilitated the establishment of a resource population of MHC-defined and MHC-homozygous sheep. This resource flock will provide genetic material for comparative genomic analysis, as well as a range of cellular and molecular tools to undertake in-depth analyses of $CD8^+$ T-cell responses to infection by intracellular pathogens.

Acknowledgements The support of the sequencing services at the Moredun and Roslin Institutes is gratefully acknowledged, along with the expert assistance of the clinical support staff at the Moredun Institute.

References

Ainsworth PJ, Surh LC, Coulter-Mackie MB (1991) Diagnostic single strand conformational polymorphism (SSCP): a simplified non-radioisotopic method as applied to a Tay–Sachs B1 variant. Nucleic Acids Res 2:405–406

Bjorkman PJ, Saper MA, Samraoui B, Bennett WS, Strominger JL, Wiley DC (1987) The foreign antigen binding site and T cell recognition regions of class I histocompatibility antigens. Nature 329:512–518

Boyum A (1968) Isolation of mononuclear cells and granulocytes from human blood. Isolation of mononuclear cells by one centrifugation, and of granulocytes by combining centrifugation and sedimentation at 1 g. Scand J Clin Lab Invest 97:77–89

Braud VM, Allan DS, McMichael AJ (1999) Functions of non-classical MHC and non-MHC-encoded class I molecules. Curr Opin Immunol 11:100–108

Chardon P, Renard C, Vaiman M (1999) The major histocompatibility complex in swine. Immunol Rev 167:179–192

Cullen PR, Bunch C, Brownlie J, Morris PJ (1982) Sheep lymphocyte antigens: a preliminary study. Anim Blood Groups Biochem Genet 13:149–159

Eckels DD (2000) MHC: function and implication on vaccine development. Vox Sang 78:265–267

Ellis S (2004) The cattle major histocompatibility complex: is it unique? Vet Immunol Immunopathol 102:1–8

Ellis SA, Ballingall KT (1999) Cattle MHC: evolution in action? Immunol Rev 167:159–168

Ellis SA, Martin AJ, Holmes EC, Morrison WI (1995) At least four MHC class I genes are transcribed in the horse: phylogenetic analysis suggests an unusual evolutionary history for the MHC in this species. Eur J Immunogenet 22:249–260

Ellis SA, Holmes EC, Staines KA, Smith KB, Stear MJ, McKeever DJ, MacHugh ND, Morrison WI (1999) Variation in the number of expressed MHC genes in different cattle class I haplotypes. Immunogenetics 50:319–328

Falk CS, Schendel DJ (1997) HLA-C revisited: ten years of change. Immunol Res 16:203–214

Garber TL, Hughes AL, Watkins DI, Templeton JW (1994) Evidence for at least 3 transcribed BoLA class-I loci. Immunogenetics 39:257–265

Garrido JJ, de Andres DF, Pintado CO, Llanes D, Stear MJ (1995) Serologically defined lymphocyte alloantigens in Spanish sheep. Exp Clin Immunogenet 12:268–271

Grossberger D, Hein W, Marcuz A (1990) Class I major histocompatibility complex cDNA clones from sheep thymus: alternative splicing could make a long cytoplasmic tail. Immunogenetics 32:77–87

Guild BC, Strominger JL (1984) Human and murine class I MHC antigens share conserved serine-335, the site of HLA phosphorylation in vivo. J Biol Chem 259:9235–9240

Harty JT, Tvinnereim AR, White DW (2000) CD8+ T cell effector mechanisms in resistance to infection. Annu Rev Immunol 18:275–308

Hill AV, Allsopp CE, Kwiatkowski D, Anstey NM, Twumasi P, Rowe PA, Bennett S, Brewster D, McMichael AJ, Greenwood BM (1991) Common West African HLA antigens are associated with protection from severe malaria. Nature 352:595–600

Holmes EC, Ellis SA (1999) Evolutionary history of MHC class I genes in the mammalian order Perissodactyla. J Mol Evol 49: 316–324

Holmes EC, Roberts AF, Staines KA, Ellis SA (2003) Evolution of major histocompatibility complex class I genes in Cetartiodactyls. Immunogenetics 55:193–202

Hughes AL, Nei M (1989) Evolution of the major histocompatibility complex-independent origin of nonclassical class-I genes in different groups of mammals. Mol Biol Evol 6:559–579

Jugo BM, Vicario A (2001) Lymphocyte antigens in sheep: linkage to the MHC class II DRB1 gene. Eur J Immunogenet 28:451–458

Jugo BM, Joosten I, Grosfeld-Stulemeyer M, Amorena B, Hensen EJ (2002) Immunoprecipitation and isoelectric focusing of sheep MHC class I antigens reveal higher complexity than serology. Eur J Immunogenet 29:391–399

Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16:111–120

Klein J, Bontrop RE, Dawkins RL, Erlich HA, Gyllensten UB, Heise ER, Jones PP, Parham P, Wakeland EK, Watkins DI (1990) Nomenclature for the major histocompatibility complexes of different species: a proposal. Immunogenetics 31:217–219

Konnai S, Nagaoka Y, Takeshima S, Onuma M, Aida Y (2003) Sequences and diversity of 17 new Ovar-DRB1 alleles from three breeds of sheep. Eur J Immunogenet 30:275–282

Kumar S, Tamura K, Jakobsen IB, Nei M (2001) MEGA2: molecular evolutionary genetics analysis software. Bioinformatics 17: 1244–1245

Lew AM, Mccluskey J, Maloy WL, Margulies DH, Coligan JE (1987) Multiple class-I molecules generated from single genes by alternative splicing of pre-messenger-RNAs. Immunol Res 6:117–132

McCutcheon JA, Gumperz J, Smith KD, Lutz CT, Parham P (1995) Low HLA-C expression at cell-surfaces correlates with increased turnover of heavy-chain messenger-RNA. J Exp Med 181:2085–2095

Messaoudi I, Guevara Patino JA, Dyall R, LeMaoult J, Nikolich-Zugich J (2002) Direct link between mhc polymorphism, T cell avidity, and diversity in immune defense. Science 298:1797–1800

Millot P (1984) The OLA major histocompatibility complex of sheep. Study of six new factors and evidence of a third locus of the complex: OLA-C. Exp Clin Immunogenet 1:31–42

Otting N, Heijmans CIC, Noort RC, de Groott NG, Doxiadis GGM, van Rood JJ, Watkins DI, Bontrop RE (2005) Unparalleled complexity of the MHC class I region in rhesus macaques. Proc Natl Acad Sci U S A 102:1626–1631

Parham P, Adams EJ, Arnett KL (1995) The origins of HLA-A, B, C polymorphism. Immunol Rev 143:141–180

Pichowski JS, Ellis SA, Morrison WI (1996) Sequence of two cattle MHC class I cDNAs associated with BoLA A10 specificity. Immunogenetics 43:253–254

Ruohonen-Lehto MK, Renard C, Rothschild MF, Edfors-Lilja I, Kristensen B, Gustafsson U, Larson RG, Varvio SL (1998) Variable number of pig MHC class I genes in different serologically defined haplotypes identified by a 3′-untranslated region probe. Anim Genet 29:178–184

Saitou N, Nei M (1987) The neighbour-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Schwaiger FW, Weyers E, Buitkamp J, Ede AJ, Crawford A, Epplen JT (1994) Interdependent MHC-DRB exon-plus-intron evolution in artiodactyls. Mol Biol Evol 11:239–249

Snary D, Barnstable CJ, Bodmer WF, Crumpton MJ (1977) Molecular structure of human histocompatibility antigens: the HLA-C series. Eur J Immunol 8:580–585

Stear MJ, Spooner RL (1981) Lymphocyte antigens in sheep. Anim Blood Groups Biochem Genet 12:265–276

Trowsdale J (1995) "Both man and bird and beast": comparative organization of MHC genes. Immunogenetics 41:1–17

Ulker N, Lewis KD, Hood LE, Stroynowski I (1990) Activated T-cells transcribe an alternatively spliced messenger-RNA encoding a soluble form of QA-2 antigen. EMBO J 9:3839–3847