MOLECULAR IDENTIFICATION OF THE BOVINE MAJOR HISTOCOMPATIBILITY COMPLEX I AND II GENES OF A *BOS TAURUS AFRICANUS* (SANGA) CATTLE BREED OF ZAMBIA

BY

ISAAC KOMBE SILWAMBA

A dissertation submitted to the University of Zambia in fulfilment of the requirements for the award of the degree of Master of Science in Veterinary Epidemiology

THE UNIVERSITY OF ZAMBIA

LUSAKA

2019

COPYRIGHTS

DECLARATION

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

Signature : _____

Print name : _____

Date : _____

APPROVAL

The University of Zambia approves this dissertation of Isaac Kombe Silwamba as fulfilling the requirements for the award of the degree of Master of Science in Veterinary Epidemiology.

Internal Examiner (1)	**Signature**	**Date**
Internal Examiner (2)	**Signature**	**Date**
External Examiner	**Signature**	**Date**
Chairperson, Board of Examiners	**Signature**	**Date**
Supervisor	**Signature**	**Date**

ABSTRACT

The Major Histocompatibility Complex (MHC) region contains many genes that are key regulators of both innate and adaptive immunity and is the most polymorphic locus in the mammalian genome. Consequently, the characterisation of the repertoire of MHC genes is critical to understanding the variation pivotal in determining the nature of immune responses. Currently, our knowledge of the bovine MHCI and II repertoire is limited, with only the Holstein-Friesian breed having been studied intensively. Traditional methods of MHC genotyping are of low resolution and laborious, however, next-generation sequencing (NGS) technologies have enabled high throughput and much higher resolution MHC typing in a number of species. In this study alleles of the BoLA class I and BoLA class II (DRB3, DQA and DQB) loci, from 347 Tonga cattle whole blood from four discrete sampling locations were sequenced using the Illumina MiSeq NGS platform. Analysis of the sequence data was performed with Bioinformatic scripting/tools (FastQC, Flash, Fastx, BLAST, and Perl). During the course of this study, a total of 137 MHC class I haplotypes were identified, of which 37 were novel. Together these novel haplotypes included 13 novel alleles and sequences matching 88 known alleles. Furthermore, 67 novel MHC class II haplotypes were defined by co-occurrence of the same DRB3, DQA and DQB alleles. Together these MHCII haplotypes included 108 novel alleles and sequences matching 121 known alleles. The results of this study have dramatically expanded the known bovine MHC repertoire of African cattle. This is the first report presenting a detailed study of the allelic and haplotype distribution of Bovine MHCI and II genes in a *Bos taurus africanus* cattle breeds. These results will provide a wealth of information for rationalising the development of improved and novel vaccines against infectious diseases of livestock, help in developing improved Livestock breeding strategies and will be a major contributor to upgrading the current Immuno polymorphism database - MHC database of known bovine MHC alleles.

Key words: Major histocompatibility complex; Tonga cattle; next generation sequencing; Bioinformatics

DEDICATION

I dedicate this work to the almighty God, my parents Lilian and Luckford Silwamba and my siblings Vivian (The late), Aubrey, McClemens, Graham, Bertha, Norah, Nathan (The late), Nathan, Abednego and Audrey. To all my friends, my fiancé Twaambo and anybody who has helped me become what I am today and delights in my success.

ACKNOWLEDGEMENTS

The dissertation marks a significant milestone in my life, and it would not have come to pass without those that have helped, nourished and supported me throughout the journey. I would like to first acknowledge my co-mentor, Dr. Timothy Connelley, who took me under his wings to learn about Major Histocompatibility complex, read my numerous revisions and helped, make some sense out of the confusion. Also thanks to my main supervisor, Dr. Martin Simuunza, and co-supervisors, Prof. John. B. Muma, Dr. Geoffrey Mainda and Prof. King Nalubamba, who offered guidance and support. Thanks to the University of Zambia, International Livestock Research Institute (ILRI), and University of Edinburgh, Roslin Institute for providing the necessary facilities. Thanks to the Centre for Tropical Livestock Genetics and Health (CTLGH) for awarding me, the financial support for this project and to CAPAZOMANINTECO for supporting payment of my tuition. Finally, thanks to my wonderful parents and siblings, lab mates, my playfellows Drs Goliath Zulu and James Chanda and numerous friends who endured this long process with me, always offering support and love.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ABBREVIATIONS AND ACRONYMS

AMV	:	Avian Myeloblastosis Virus
APC	:	Antigen Presenting Cells
BAC	:	Bacterial Artificial Chromosome
BoLA	:	Bovine Leukocyte Antigen
CD	:	Cluster of Differentiation
cDNA	:	Complementary DNA
cM	:	Centimorgans
DNA	:	Deoxyribonucleic Acid
dNTP	:	Deoxynucleotide Triphospate
EDTA	:	Ethylenediaminetetraacetic Acid
FAO	:	Food and Agriculture Organisation
GAPDH	:	Glyceraldehyde 3-Phosphate Dehydrogenase
HLA	:	Human Leukocyte Antigen
IPD	:	Immuno Polymorphism Database
KB	:	Kilobases
MB	:	Megabases
MHC	:	Major Histocompatibility Complex
MIC	:	MHC class I chain related
NC	:	Non classical
NGS	:	Next Generation Sequencing
NKC	:	Natural Killer Cells
PACRA	:	Patents and Companies Registration Agency
PBS	:	Phosphate Buffered Saline
PCR	:	Polymerase Chain Reaction
RBC	:	Red Blood Cell
RNA	:	Ribonucleic Acid
TAP	:	Transporter associated with antigen processing
TCRs	:	T cell receptors:
UK	:	United Kingdom
WBC	:	White Blood Cells

CHAPTER ONE

INTRODUCTION

1.1 Background

Domesticated around 10500 years ago (Decker *et al.*, 2014), cattle are currently raised throughout the world for a range of purposes including the provision of milk, meat, hides and draft power. At present, the overall domestic cattle populace is at 999.8 million (World cattle inventory, 2017). The Food and Agriculture Organization (FAO) projections predict that demand for livestock products is going to increase by 70% by 2050 to feed a populace anticipated to reach 9.6 billion (FAO, 2000). Thus, research to promote livestock production of economically important species, such as cattle, is critical to sustaining adequate food provision in the future. One aspect of this is to focus on disease control strategies that aim to generate improved vaccines against major cattle pathogens that can be used in the relevant target populations. To achieve this, characterising the immunogenetic variation that may assist in vaccine development and understanding variability in vaccine responses will be critical, with the Major Histocompatibility Complex (MHC) genes an obvious target for analysis.

The MHC locus is a genetic region found in all mammalian species and it contains many genes crucial in antigen presentation. The classical MHCI and MHCII genes are transcribed to form MHC molecules present on the cell surfaces. The main function of these MHC molecules is to bind pathogen-derived peptides and present them to appropriate T-cells. Classical MHCI molecules are present on all nucleated cells and have a peptide-binding groove formed by the extra-cellular α1 and α2 domains in which short peptide fragments (9–11 amino acids) bind. The combined peptide-MHCI structure forms the ligand recognised by αβ T cell receptors (TRs) expressed by antigen-specific CD8+ T cells and by killer cell immunoglobulin-like (and other) receptors expressed by natural killer (NK) cells (Guethlein *et al.*, 2015). Consequently, MHCI genes play key roles in regulating both innate and adaptive immune responses. On the other hand, MHC class II molecules are exclusively found on specialised antigen-presenting cells (APCs) and present antigens to CD4+ T cells. The α_1 and β_1 domains form the MHC class II binding groove and is capable of loading antigens of 13 amino acids in length. An important feature of both classical MHCI and MHCII genes is their high level of polymorphism within the exons encoding the peptide-binding groove. For instance, in humans, over 12,000 unique classical MHCI alleles of the three classical MHCI genes (HLA-A, HLA-B, HLA-C) have been

described. This polymorphism ensures that at the population level, MHCI and MHCII genes have the capacity to effectively present a diverse range of peptides from any potential pathogen. As a consequence of their critical role in determining the pool of peptides from pathogens that have the potential to serve as epitopes for T cells, knowledge of the diversity of classical MHCI and MHC II genes is fundamental to understanding T cell immunity and vaccine development.

Much of the groundwork in major histocompatibility complex (MHC) research was carried out in humans and mice, where the number of classical MHC class I genes present is consistent across all haplotypes (Parham *et al.,* 1995). However, recent studies of the MHC have demonstrated greater genetic complexity than previously predicted (Ellis 2004; Miltiadou *et al.* 2005; Otting *et al.,* 2005). This complexity involves variable gene content in different haplotypes as well as allelic diversity.

The past traditional methods of MHC genotyping, using low resolution and laborious procedures, have greatly hindered a more comprehensive analysis of the MHC repertoire of cattle breeds other than the Holstein-Friesian. However, rapid and relatively cheap novel Next Generation Sequencing (NGS) techniques that have enabled large-scale gene-typing of bovine MHC alleles from genetically diverse populations have been developed (Vasoya *et al.,* 2016). This is an opportunity for more focused research in understanding the diversity of MHC genes in various cattle breeds including those that are indigenous to Zambia.

1.2 Statement of the Problem

Livestock plays a vital role in food security globally(Mahmood *et al*., 2014). However, the efficiency and viability of livestock production is constrained by a multitude of problems and of these diseases are a major factor (FAO, 2016).

Prevention through vaccination currently represents the best course of action to combat these diseases. (Rueckert, 2012). Consequently, vaccines are still required for a large number of diseases that cause livestock morbidity and mortality.

A key obstacle to the rational design of novel vaccines is the selection epitopes to be included in such vaccines. The MHC is a key factor in determining epitope selection for targeting T-cell responses and,therefore, knowledge of MHC diversity is fundamental in efforts to rationalise development of vaccines that aim to elicit T-cell responses (Rueckert, 2012). Based on evidence from other species it is assumed that the bovine MHC will be characterised by a high level of polymorphism with the polymorphism directly influencing epitope selection. However,

the problem is that, the current knowledge of MHC diversity in cattle is severely limited and for Zambian cattle populations there is currently no data on MHC diversity.

As the MHC molecules are a primary determinant of which peptides are presented to T-cells, our currently restricted understanding of the MHC repertoire in cattle forms a major impediment to the rational selection of epitopes that should be included in new vaccines. Accordingly, this study was undertaken to fill this fundamental knowledge gap and it aimed to apply high-throughput NGS typing methods to study MHCI and II repertoires in Zambian cattle populations. As Tonga cattle is resilient and well adapted to local climatic conditions. An understanding of MHC in Tonga cattle is of key importance for cattle conservation and disease control both of which are crucial for income generation and improved livelihood for the farmers.

1.3 Study Justification

Data on allele sequences that code for MHC molecules is publically available through the curated Immuno Polymorphism Database (IPD - http://www.ebi.ac.uk/ipd). In humans, over 12,000 unique classical MHCI alleles have now been described, whereas for cattle the number of sequences available is <100, and of these >60% of the submitted sequences are derived from, or known to be expressed in Holstein–Friesian cattle, the predominant dairy breed in Europe and North America (Birch *et al.*, 2006; Brown *et al.*, 1989; Davies et *al.*, 2006; Ellis *et al.* 1999; Ellis *et al.*,2005; Ellis *et al.*, 1996; Ellis *et al.*, 1998; Holmes *et al.*, 2003). In contrast, data on MHC allele sequences from the other major lineages of domestic cattle - *Bos indicus* and African *B. taurus* - which are extensively reared in tropical regions, is still scarce (Robinson *et al.*, 2005). Therefore, the repertoire of MHC diversity in cattle is very poorly defined. As MHC is a critical regulator of innate (e.g. NK cell) and cell-mediated (i.e. T-cell mediated) immunity, the lack of knowledge is a severe impediment to research into the development of vaccines that aim at eliciting protective T-cell responses against *Theileria parva*, foot-and-mouth-disease virus, and *Mycobacterium bovis*. The lack of data on major lineages of domestic cattle - *Bos indicus* and African *B. taurus* was largely due to the lack of refined tools to enable the analysis of large and diverse cattle populations. However, recent technological developments that have led to development of rapid and relatively cheap novel techniques provide an opportunity to generate data on other cattle breeds, including animals from the *Bos indicus* and African *Bos taurus* lineages.

The complex origins of the global cattle population and its highly outbred nature make comparison of the MHC allele repertoires of these domestic cattle lineages of evolutionary, as well as of immunological interest. It is clear that breeding strategies, climate instability, emerging pathogens and unpredictable movement of disease vectors may cause MHC diversity to reduce and this could have a detrimental impact on the ability of animals to mount appropriate T cell responses (Ellis., 2005). Recent genetic analysis implies that despite the European and African taurine lineages being derived from the same domestication event, substantial admixture of the ancestors of African *B. taurus* with wild African Aurochs has led to significant genetic divergence (Decker *et al.,* 2014) and therefore data from European lineages cannot be extrapolated to African ones which need to considered separately. Furthermore, knowledge of the MHC diversity at the genomic, haplotype and allele level will help to understand and facilitate the manipulation of immune responses against pathogens in economically important species, such as cattle. Therefore, further studies generating larger datasets, incorporating more breeds (including Zebu and Sanga animals), are needed to more comprehensively evaluate the diversity of MHC genes in the global cattle population.

1.4 Significance of the study

The data generated by this study and the equivalent studies being undertaken in other countries will provide a wealth of information on bovine MHC diversity that will be a major contributor to upgrading the current IPD-MHC database of known bovine MHC alleles. In addition, since the importance of cattle to the global food security is likely to increase, it is crucial that we understand complex genetic systems such as the MHC that underpin their ability to survive in the face of unpredictable pathogen exposure. Finally, an improved understanding of MHC diversity in modern day cattle can be used to assess the ability of these animal populations to mount effective immune responses after infection or vaccination and for application in vaccinology.

1.5 Research Questions

What MHC class I repertoires does the Tonga cattle breed of Zambia have?

What MHC class II repertoires does the Tonga cattle breed of Zambia have?

1.6 Objectives

1.6.1 General objective

The general objective of this study was to perform molecular identification of the Major Histocompatibility Complex class I and class II genes of the Tonga cattle breed of Zambia using Next Generation Sequencing (NGS) and bioinformatics.

1.6.2. Specific Objective

To identify the MHC class I and II repertoires of Tonga cattle in Zambia using NGS and bioinformatics.

1.7 Organisation of the dissertation

The dissertation is organised in five chapters. The first chapter provides the background to the study and highlights the problem statement, justification, study questions and its objectives. Chapter Two gives a review of the available literature on MHC classes and also identifies some knowledge gaps. The third chapter explains the materials used in the study and describes the different methods employed in realising the study findings. This chapter also describes how the data was analysed and ethical considerations observed in undertaking the study as well as study limitations. Chapter Four provides a detailed description of the study findings while Chapter Five provides a discussion of the study findings according to the study objectives; draws conclusions from the findings and makes recommendations.

CHAPTER TWO

LITERATURE REVIEW

2.1 Major Histocompatibility Complex

A large number of genes associated with antigen presentation are situated in the major histocompatibility (MHC) locus. In cattle, the MHC, also known as Bovine leucocyte antigen (BoLA), is located on bands q13-23 of cattle chromosome 23 (Fries *et al.*, 1986) while in humans the MHC, also known as Human Leukocyte Antigen (HLA), is located on chromosome 6p and is known to cover 4 Mb. As in humans and mice, BoLA comprises three gene classes; class I, class II and class III. Amongst these genes are those that encode 'MHC' molecules that flag disease by exhibiting pathogen-derived peptides to receptors on T lymphocytes (Ellis, 2004). Figure 1 below illustrates the essential genomic arrangement of the human MHC on chromosome 6.

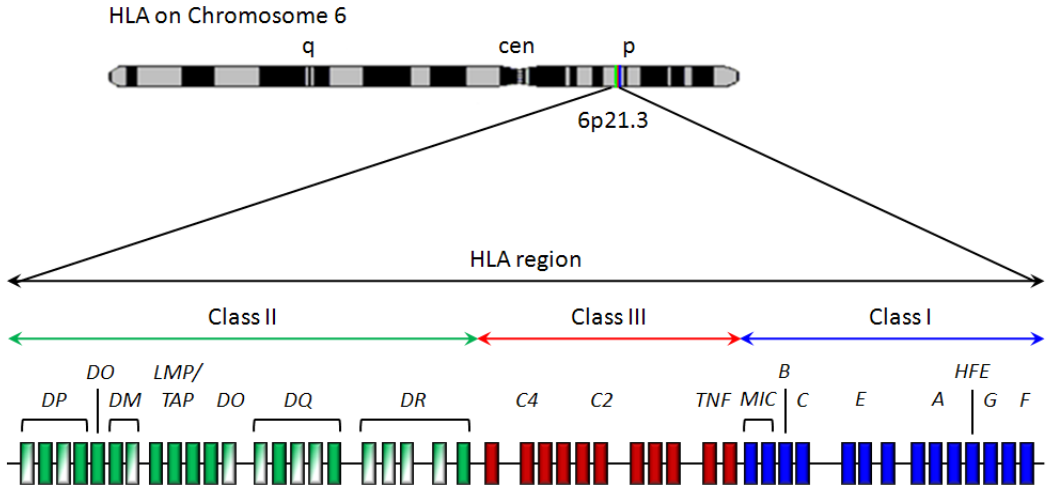


Figure 1 Fundamental genomic organisation of the human MHC on chromosome 6 (adapted from the MHC Sequencing Consortium, 1999).

2.1.1 Genomic Organisation of the MHC Region

In comparison to human MHC, the organisation of the BoLA genes is poorly characterised. In spite of the availability of the bovine genome sequence, the organisation and precise number of bovine classical I loci is still uncertain as the data used for the genome assembly was derived from two related animals, and the sequence contains just three classical class I genes (The

Bovine Genome Sequencing and Analysis Consortium, 2009). Restricted mapping studies have so far reviewed that no less than 9 genes are located inside an area of 400 kb (Di Palma *et al.*, 2002; http://www.hgsc.bcm.tmc.edu/projects/bovine/). A further 10 genes, which are probably going to be pseudogenes or gene fragments, are also found within this region (Di Palma *et al.*, 2002; http://www.hgsc.bcm.tmc.edu/projects/bovine/). Current evidence suggests the presence of 6 separate loci that encode functional MHCI molecules, but the loci content of haplotypes differs.

Currently, the organisation of these loci is undetermined except for A14. Separate studies have demonstrated that genes 1, 2 and 4 on the A14 haplotype (Di Palma *et al.*, 2002) and genes 3 and 5 on the A10/KN104 haplotype (Bensaid *et al.*, 1991) are situated inside 210 kb of each other. Notwithstanding, the situation of genes 1, 2 and 4 in connection to that of genes 3 and 5 remains undetermined as does the relative location of gene 6. Mapping of the MHC class I genes in an A14 BAC library demonstrated the classical class I genes were organised gene 1, gene 4 and gene 2 arranged from the centromere to the telomere (Hammond *et al.*, 2012). An anchor gene, ZNF173 and a gene called 'Z', now known as N*04001, were situated at the telomeric end of this 400 kb bunch. The BAC library mapping study additionally demonstrated that the bovine non-classical gene NC1-N*00101, formally called gene X, is found centromerically adjoining gene 1 (Di Palma *et al.*, 2002). The bovine genome sequence likewise demonstrates that a NC1 gene is arranged close to the classical class I genes (Birch *et al.*, 2008a).

Be that as it may, there is an absence of overlap between the distributed bovine genome sequences and the BAC mapping study because of the DNA sources from which the sequences were inferred i.e. the animals do not convey a similar haplotype. Three extra non-classical class I genes, together with a conceivable four MIC genes have been found roughly 500 kb centromeric to the classical class I group (Birch et al., 2008a; Birch *et al.*, 2008b). The bovine class I locus is located near the telomere whereas class II is located centromerically to class III (Sharif *et al.*, 1998) and the size of the MHC class I region is varied across species. For example, humans and rodents have the biggest MHC class I region (Amadou et al., 1999; Shiina *et al.*, 1999) and the pig the smallest (Chardon *et al.*, 1999). Information from mapping studies and bovine genome sequences has shown that the cattle class I region is bigger than that of the pig, and has a greater number of genes (Ellis *et al.*, 2005). Figure 2 below shows the genetic organisation of the bovine MHC in comparison to HLA and a simplistic interpretation

of a selection of genes encoded on both HLA and BoLA.

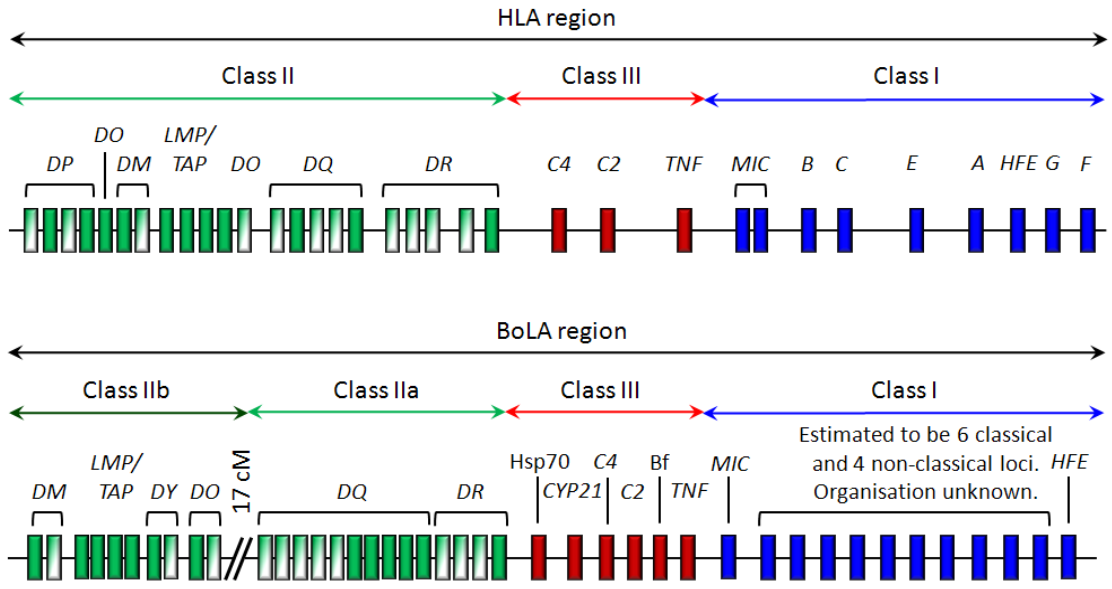


Figure 2. Diagrammatic representation comparing the genetic organisation of the bovine MHC (*BoLA)* on *Bos taurus* chromosome 23 with *HLA*. (Adapted from a thesis by J Codner, Institute of Animal Health, Compton and University of Glasgow, UK)

2.1.2 MHC class I

The classical MHC class I genes code for MHCI molecules that are present on all nucleated cells. MHC class I classification was confirmed in 2004 by the MHC nomenclature committee and depends on the classification utilized for HLA (Ellis *et al.*, 2006). HLA allele names comprise of 5-11 digits and are assigned by the amino acid arrangement, with the initial 3 digits demonstrating allele 'group', the following two showing coding changes, the following two denoting non-coding changes and the last four digits, if required, demonstrating changes inside the promoter or introns. As the task of assigning alleles was previously troublesome in cattle (Ellis, 2004; Davies *et al.*, 2006), the majority of alleles were prefixed 'N' to demonstrate that they couldn't be allocated to a complete locus, and they were numbered in a single series. This is no longer the case and alleles have now been allocated to putative loci. This allocation is based on phylogenetic analyses which has the ability to cluster classical MHCI sequences into six groups that are considered at present to represent the products of six independent loci (Ellis *et al.*, 1999; Ellis *et al.*, 2005; Holmes *et al.* 2003), however, the support for these groups is not consistently strong. Alleles contrasting by less than four amino acids in alpha 1 and 2 and close

to four over the rest of the sequences are allotted to the same 'group'. Those alleles recognized as non-classical alleles are prefixed by 'NC' and the non-classical locus number preceding the 'N'.

The set of bovine MHC genes co-inherited as a group are referred to as a haplotype (a set of genetic determinants located on a single chromosome) and the MHC genes from the two chromosomes constitute the MHC genotype of an individual (Warner *et al.,* 1988). In humans, each MHC haplotype contains three classical MHC class I genes. Detailed investigations of cattle MHC haplotypes have demonstrated that between 1 and 4 classical MHCI loci are expressed per haplotype and such expression of a variable number of MHCI genes has been observed in different species (Ellis *et al.,* 1999). Notably certain combinations of MHCI loci appeared to be preferentially co-expressed e.g. genes 1 and 2 and genes 2 and 3 are often expressed together. Gene 1 is never expressed in combination with either gene 3 or 6, which may suggest that these genes are not distinct. However phylogenetic evidence shows no distinct relationship between these alleles at these putative loci (Holmes *et al.,* 2003) and the intron sequences are sufficiently divergent to facilitate gene-specific amplification (Birch *et al.,* 2006). Gene deletion or gene 'silencing' are accepted to be the main reasons for the variation in the number of genes expressed per haplotype (Ellis *et al.,* 2005). The different permutations of alleles expressed on human class I haplotypes has to date not been observed in cattle where the allele content of defined haplotypes appears to be largely fixed. However, this may reflect the comparatively small number of haplotypes that have been characterised in cattle. Investigation of the non-classical genes on six molecularly characterized MHCI haplotypes showed that the most oligomorphic non-classical locus NC1 is present on all haplotypes. Non-classical genes NC2 and NC4 were likewise recognized in all haplotypes, except for the A18 haplotype.

Previously, only sixteen MHC class I haplotypes (Table 1) were identified at the molecular level. However, recent application of novel next generation sequencing technologies have reviewed some additional 62 bovine MHC class I haplotypes (Vasoya *et al* 2016). To date, roughly <100 cattle MHC class I sequences have been submitted to the IPD database (http://www.ebi.ac.uk/ipd/mhc/bola/index.html; Robinson et al., 2005) of which more than 60% were obtained from, or are known to be expressed in, Holstein-Friesian cattle. In contrast, data on MHC allele sequences from the other major lineages of domestic cattle - *Bos indicus*

and African *B. taurus* - which are extensively reared in tropical regions, is severely limited (Robinson *et al.*, 2005).

Table 1. Transcribed bovine classical MHC class I genes in different haplotypes.

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

Source: Adapted from Birch *et al*., 2006. [a] A serological specificity has not been determined for this haplotype. [b] A formal allele name has not been assigned to this allele. [c] This haplotype was reported and analysed by Bensaid *et al*. (1991).

2.1.3 MHC class II

The MHC class II genes code for classical MHCII molecules that are expressed on specialised antigen-presenting cells (APCs). The class I and class II region are firmly connected within the bovine MHC. The high level of homology of cattle class II with other mammalian class II regions has facilitated characterisation of this BoLA region. Southern blotting with human class II probes provided evidence of cattle orthologues of DO, DR and DQ genes. (Andersson *et al*., 1988). Utilizing the recognized DO, DR and DQ genes together with two extra loci, a family segregation analysis of linkage disequilibrium in the class II was completed. This analysis showed the presence of a substantial molecular distance isolating the DQ and DR genes from the DO genes (Andersson *et al.*, 1988). Additionally, mapping empowered the assignment of class II loci to either the class IIa or class IIb regions, which appeared to be separated by a

physical distance of 17 cM (Hess *et al.,* 1999). Table 2 below presents the genes encoded on the BoLA class II region. The DR region, situated in the class IIa region, contains the monomorphic DRA gene and no less than three DRB genes. Of the DRB genes, just DRB3 is effectively transcribed, expressed at high levels and polymorphic. With more than 120 alleles currently identified, the DRB3 locus is the most polymorphic of the class II loci. This high level of polymorphism and ease of characterisation via PCR and sequence analysis have meant that the relationship between DRB3 and disease resistance/susceptibility has been intensively examined e.g. the DRB3*1502 allele has been associated with lower somatic cell count, a marker of mastitis in Holstein cattle (Sharif *et al.,* 1998). Notwithstanding, it is vital to take note of that in spite of various papers connecting certain DRB3 alleles to mastitis resistance/susceptibility there has been an absence of consistency in these examinations (Dietz *et al.,* 1997; Rupp *et al.,* 2007). Every haplotype expresses one DR gene pair, DRA and DRB3. Arranged centromerically from the DR region in the class IIa region are the DQ genes. Five DQA genes of differing levels of polymorphism and five similarly polymorphic DQB genes are known. The number and combination of DQ genes on any single haplotype remains poorly described. However, to date each haplotype appears to contain either 1 or 2 DQA or 1 or 2 DQB expressed loci, with the duplicated DQ structure seen in around half of the common class II haplotypes (Glass *et al.,* 2000). Gene duplication combined with the variety of the DQA and DQB genes, and also the possibility of intra- and inter-haplotype DQA/DQB pairing can generate an extensive variety of DQ products expressed at the cell surface. The class IIb region encodes the LMP and TAP loci alongside the non-classicals DOB, DOA, DIB, DYA, DMA and DMB. These genes have indicated limited homology with human class II genes, and the DYA and DIB genes have no known homologues in human or mice (Stone and Muggli-Cockett., 1990; Vanderpoel *et al.,* 1990). Both DYA and DIB are transcribed at low levels in cattle dendritic cells (DCs) and together these genes have appeared to code for class II molecules of one of a kind (Ballingall *et al.,* 2001; Ballingall *et al.,* 2004). The close relationship of DIB and DYA has led to the proposal that DIB be renamed DYB. The restricted expression of this gene pair and non-polymorphism in the binding cleft of this transcribed particle classifies these genes as non-classical.

Table 2. Details of BoLA class II genes and molecules. * The *DQ* genes are duplicated on some class II haplotypes

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

Source: (Adapted from a thesis by J Codner, Compton, University of Glasgow, UK)

2.1.4 MHC class III

The MHC class III genes are very conserved (Kulski *et al.*, 2002). A couple of investigations of the class III regions have indicated complement proteins Bf and C4, heat shock protein-70 (HSP70), a 21 hydroxylase and tumour necrosis factor-alpha (TNF-α) (Andersson et al., 1988; McShane *et al.*, 2001; Skow *et al.*, 1988; Teutsch *et al.*, 1989). Unlike MHCI and MHCII, the class III genes do not encode proteins that present peptides to T-cells.

2.1.5 Structure and functional pathways of MHC class I molecules

The bovine MHC class I molecules are heterodimers consisting of a heavy alpha-chain (45 kDa) and a light chain, β_2 microglobulin (12 kDa). The heavy chain comprises a cytoplasmic domain, transmembrane domain and three extracellular domains, α_1, α_2, and α_3, with which the β_2 microglobulin chain is covalently associated (Flutter & Gao, 2004). The α_3 and β_2 microglobulin are highly conserved and they consist of two anti-parallel β-sheets which are joined by disulphide bonds (Becker & Reeke, 1985; Bjorkman *et al.,* 1987; Cunningham *et al.,* 1975). A key feature of MHCI molecules is the peptide binding cleft. It is formed by the hypervariable α_1 and α_2 domains with support from the α_3 and β_2 microglobulin domains. The tertiary structures of the α1 and α2 domains are similar, each consisting of four anti-parallel β-strands with a long α-helical region at the C-terminus. The eight β-strands of the α_1 and α_2 domains form one large β-sheet, constituting the floor of the peptide binding cleft, and the two α-helices form the peripheral boundaries of the cleft in which peptides are bound (Zhang *et al.,* 1998). MHCI molecules have a peptide-binding site that is closed on both ends, and they preferentially bind peptides of eight to ten amino acids, which are able to fit within this closed binding site (Bjorkman *et al.,* 1987; Jardetzky *et al.,* 1994; Rammensee *et al.,* 1995). The peptides bound by MHC class I molecules are derived from both self and foreign peptides originating from the cytosol (Figures 3 and 4). These peptides are generated by the proteasome and are transported into the endoplasmic reticulum (ER) lumen via two transporters associated with antigen processing proteins. Once inside the ER, the processed peptides bind with, and stabilise the folding of MHC class I subunits with the help of chaperone proteins before transport to the cell membrane (Townsend *et al.,* 1989b). The binding of peptides stabilises MHC class I molecules, although it has been demonstrated that at low temperatures, empty MHC class I molecules are able to assemble and are expressed on the cell membrane, but these molecules are not stable at physiological temperatures. Within the peptide binding cleft of the MHC class I molecule are a series of structural pockets, unique to each MHC allele as a result of polymorphism in the α_1 and α_2 domains, which accommodate peptide side chains, facilitate peptide binding and dictate antigen specificity (Matsumura *et al.,* 1992). On expression at the cell surface, T cell receptors (TCR) expressed by CD8+ T-cells recognise cell membrane-bound MHC class I molecules that present the cognate peptides. Once the CD8+ TCR has bound to peptide-MHC ligands composed of peptides from 'altered-self' or foreign pathogens the presenting cell may be lysed. This interaction also initiates proliferation and differentiation of cytotoxic T cells and the establishment of memory T helper cells (Sher *et al.,* 1995).

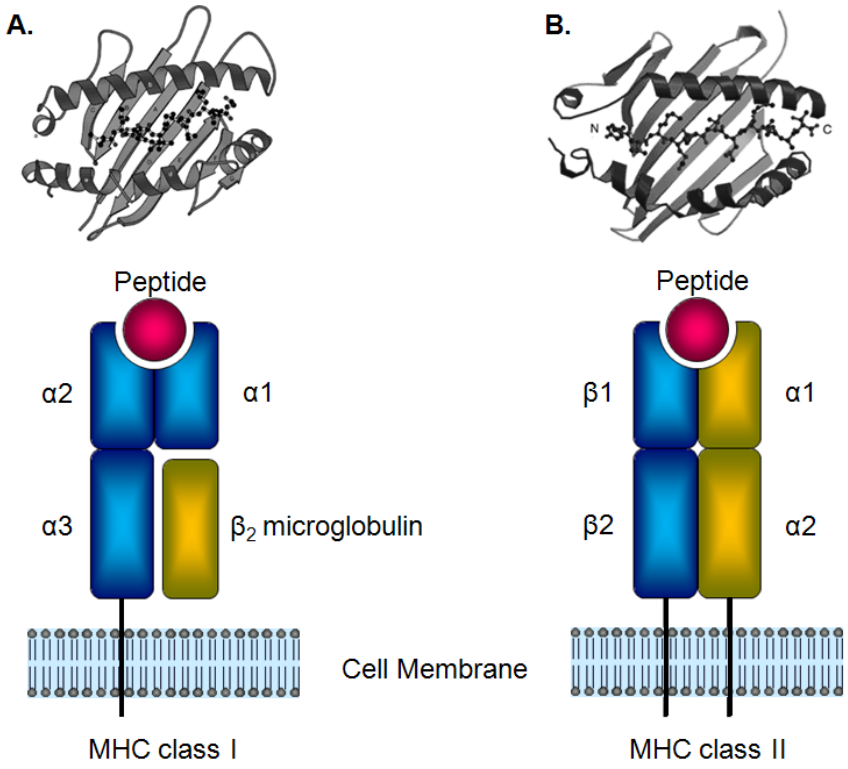


Figure 3. Schematic diagrams of the peptide binding cleft with peptide and overall structure of cell membrane bound **A.** MHC class I and **B.** MHC class II molecules. . *(Adapted from a thesis by J Codner, Institute of Animal Health, Compton and University of Glasgow, UK)*

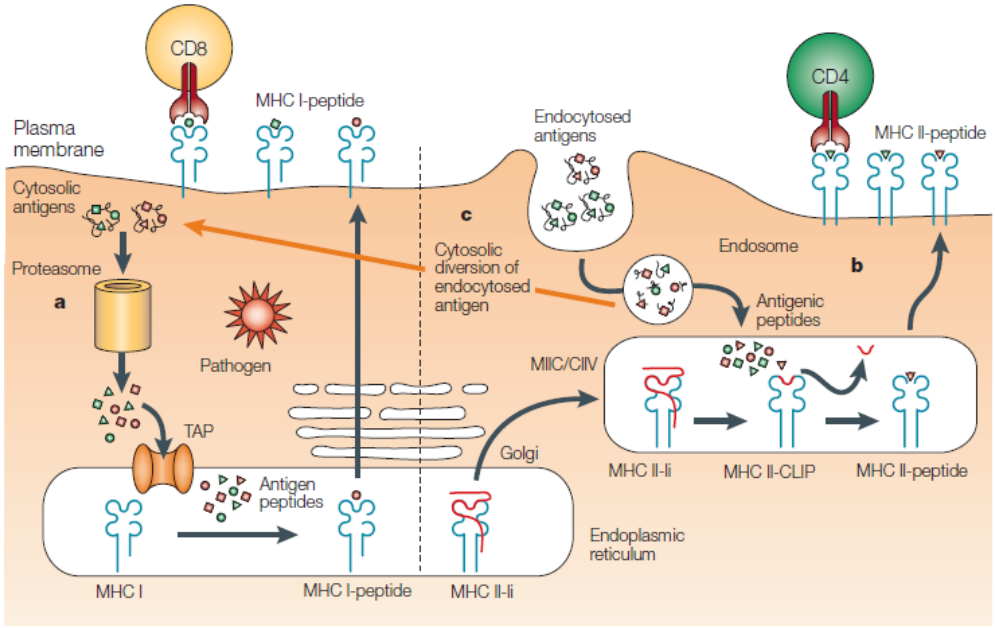


Figure 4. Schematic diagram of antigen presentation pathways by **A.** MHC class I and **B.** class II molecules (Heath & Carbone, 2001).

2.1.6 Structure and functional pathways of MHC class II molecules

MHC class II molecules are made up of two transmembrane glycoprotein chains, a 30-34 kDa alpha-chain and a 26-29 kDa beta-chain, which are joined by a disulphide bond. Each of these chains comprise a cytoplasmic domain, a transmembrane domain and two extracellular domains; α_1 and α_2 on the alpha-chain, and β_1 and β_2 on the beta-chain. Similar to MHCI molecules, the domains proximal to the cell surface, α_2 and β_2, have immunoglobulin-like structures, and the α_1 and β_1 domains are similar to the MHCI α_1 and α_2 domains. The interaction of the α_1 and β_1 domains forms the MHC II binding cleft, comprising eight anti-parallel β-sheets bounded by an α-helix from each contributing domain, in a similar manner to class I molecules. MHCII molecules have a peptide-binding site that is open at both ends, and they preferentially bind peptides of at least 13 amino acids (Chicz *et al.*, 1992; Brown *et al.*, 1993; Jardetzky *et al.*, 1994). The invariant residues that secure the carboxy- and amino-termini of the peptide in the peptide binding cleft of MHCI molecules (closed end) are absent in MHCII molecules enabling the presentation of antigens of at least 13 amino acids in length. As with MHCI, peptide binding within the MHCII binding cleft is facilitated by the insertion of antigenic side chains into structural pockets and hydrogen bonding along the length of the peptide; such that polymorphic residues located within the structural pockets of the peptide binding cleft dictate peptide specificity (Jardetzky *et al.*, 1996). The pathway of peptide presentation for MHC class II molecules is different to that of class I. Following translocation into the ER lumen, MHC class II subunits are bound to an invariant chain (MHC II-Ii) molecule which prevents antigen binding by preventing folding and/or physically interfering with binding. The invariant chain also directs the MHC class II molecule to low-pH endosomal compartments. Once fused with an acidic endosome the invariant chain molecule is degraded by proteases, leaving a short peptide called CLIP. In the presence of foreign peptides, CLIP is released and the antigen loaded into the peptide binding cleft by chaperones. The MHC class II-antigen complex is displayed on the extracellular surface of the cell for detection by CD4+ T cell receptors.

2.1.7 Non-classical MHC molecules

It is well known that both the class I and class II regions of the MHC encode non-classical MHC molecules. Albeit structurally comparable, non-classical MHC molecules are differentiated from their classical partners by restricted polymorphism and cell surface expression. MHC class I non-classical genes in humans include HLA-E, -F, and –G and in

mice the non-classical genes were named Qa and TL. HLA-E molecules in human, and Qa molecules in mice interact with the NK cell receptors, NKG2A and NKG2C and modulate NK cell function (Braud *et al.*, 1998; Vance *et al.*, 1998). HLA-G is expressed at the maternal-fetal interface in humans and is widely accepted to play a role in the immune tolerance of pregnancy (reviewed in Braud *et al.*, 1999). While HLA-E and -G have been widely examined, the role of HLA-F is yet to be elucidated. Likewise encoded inside the class I region are the MIC class I chain related loci and theHFE gene. The expression of MIC proteins on the cell surface without attaching to β2 microglobulin and autonomously of TAP shows these proteins are not dependent on peptide binding for cell surface translocation (Groh *et al.*, 1996). MIC is heat stress inducible, recognized by T cells which express gamma-delta T cell receptors and is widely accepted to be fundamental in maintenance of sound gut epithelium in humans (Groh *et al.*, 1998). Genes encoding non-classical class II molecules have additionally been accounted for. In people, the HLA-DM and HLA-DO genes encode proteins which are structurally similar to classical class II molecules, but these two molecules are limited in their expression. HLA-DM is held intracellularly inside the MHC class II compartments (Sanderson *et al.*, 1994). Here, HLA-DM catalyzes the evacuation of CLIP and the resulting loading of peptides to the class II molecules (Denzin and Cresswell, 1995). It has additionally been reported that HLA-DM stabilises the unfilled class II particle before peptide loading and alters the peptide to permit enduring peptides to be introduced at the cell surface (Denzin *et al.*, 1996; Sloan *et al.*, 1995). H2-M is likewise associated with the processing of antigens for CD4+ T cells in mice (reviewed in Alfonso and Karlsson, 2000).

2.1.8 Peptide binding Motifs for MHC class I and II molecules

MHC molecules are characterised by selective peptide binding and transportation. The peptide-binding motif, which simply infers selectivity of peptide binding of each MHC molecule, is dictated by the structure of the peptide-binding site of the MHC molecule. Detailed study of the amino acid sequences of peptides that are bound by individual MHC molecules as well as the crystal structures of MHC class I and class II alleles provides insights into the peptide selectivity of these molecules (Bjorkman *et al.*, 1987; Brown *et al.*, 1993; Rammensee *et al.*, 1993). Peptide-binding motifs have proved useful in the identification of peptide epitopes that are recognized by both class I and class II restricted T cells (Hammer et al., 1992, 1993; Parker *et al.*, 1992, 1994, 1995; Tsuchida *et al.*, 1994). Class I molecules usually bind peptides with two distinctive dominant anchor residues that bind in separate pockets in the peptide-binding site of the MHC molecule (Rötzschke *et al.*,1990; Rammensee *et al.*, 1993). These dominant

anchor residues are often limited to a single amino acid or a set of chemically similar amino acids. In almost every class I molecule, one dominant anchor is located at the C terminus of the peptide and is commonly abbreviated as P*omega* (P9 in nonamer peptides). Most frequently, the second dominant anchor residue is found at P2 of the peptide but it can be found at variable positions. Peptides that have only dominant anchor residues in common can differ in their binding affinity for the same MHC molecule by four orders of magnitude (Ruppert *et al.,* 1993; Parker *et al.,* 1994). Significant positive or negative effects on peptide binding can occur if amino acids are present at other positions in the peptide. Non-dominant amino acids with positive effects are termed auxiliary anchor positions. The peptide-binding motifs of class II MHC molecules are independent of the peptide termini and instead concentrate on the middle core of the peptide sequence (Hammer *et al.,* 1992, 1993; Vogt *et al.,* 1994; Wucherpfennig *et al.,* 1994; Southwood *et al.,* 1998). Due to the open-endedness of class II molecules, the bound peptide can potentially bind in a number of "registers" depending on the relative location of its middle or "core" amino acids.

2.1.9 Polymorphism and generation of MHC diversity

Genes encoding MHC molecules are among the most polymorphic genes found in gnasthostomes (jawed vertebrates), with some loci recording heterozygosity values as high as 80% (Hughes and Yeager, 1998). There are presently over 12,000 unique classical MHCI alleles identified in humans and more alleles are probably yet to be identified. It has been shown that MHC polymorphism in humans is caused by point mutations and intra-locus recombination and this variation can be mapped to particular regions of the peptide binding clefts of MHC class I and class II molecules (Trowsdale and Parham 2004). Both inter-locus recombination and gene duplication has been reported to happen in the BoLA class I region. Comparison of sequences on gene 1 and 4 of the A14 haplotype demonstrated that genes were different apart from homology all through exon 3. The relatedness of these two genes to each other, and nine different class I genes, proposes recombination accounts for the recognized sequence homology (Birch et al., 2006; Di Palma *et al.,* 2002). In support of gene duplication in the cattle class I region is the expression of two gene 2 alleles on the A17 haplotype. This blend of strategies has all the hallmarks of working inside the cattle class I region to maintain and produce diversity and thus guaranteeing that cattle populations can respond to a scope of various pathogens. Polygeny, polymorphism and heterozygosity at MHC loci afford gigantic

variation, enabling the presentation of a diverse range of antigenic peptides (Clarke and Kirby, 1966; Doherty and Zinkernagel, 1975).

The investigation of DRB polymorphism is valuable as this region is available in the antigen-presenting site and variability in this area might be associated with the variability in immune responsiveness of individuals to specific pathogens. The significant contrasts in allelic frequencies amongst European and African cattle breeds (Mikko and Andersson, 1995) suggests that selection, genetic drift, and population bottlenecks have been influential in determining the collection of cattle MHC class I and class II alleles. An expansive number of studies did to ponder the genetic variability existing at the DRB3.2 loci in different cattle breeds spread out everywhere throughout the world have detailed the event of a high level of polymorphism at this locus and that the allele spectrum and gene frequencies profile of various breeds at this locus shift from each other.Two primary sorts of mechanisms are proposed to drive the high diversity at MHC loci: disease based and reproductive mechanisms, although these two components are obviously not fundamentally unrelated (Ballingal et al., 2004). Likewise, studies have been completed toward a path to investigate if allelic diversity at the class II DRB3 locus plays a part in influencing responsiveness to vaccination. The BoLA class II DRB3∗2703 and DRB3∗1501 alleles are associated with levels of immunity against *Theileria parva* challenge following vaccination with sporozoite p67 antigen (Ballingal et al., 2004)

2.1.10 MHC and its association with disease resistance or susceptibility

The MHC genes are especially intriguing in light of the fact that they are related to genetic resistance and susceptibility to a wide cluster of diseases. Comprehension of the mechanisms that detail genetic variation in resistance and susceptibility may be exceptionally profitable in the designing of new strategies for disease control. The description of MHC association with disease in cattle is extremely expansive. Hence, just a couple of examples will be described here (Table 3). Polymorphism in BoLA-DRB3 is firmly associated with resistance against bovine leukaemia infection (Xu e*t al.*, 1993). Resistance and susceptibility to BLV has been mapped to particular areas of the ß-chain of the DR molecule. Dermatophilosis is another disease that has been mapped to particular amino acid mortifs of the BoLA DR molecule. Maillard *et al.* (2003) observed that a one of a kind BoLA class II haplotype made up of one DRB3 exon2 allele and one DQB allele exceedingly relates with the susceptibility character to bovine Dermatophilosis. Mastitis is another disease with a relationship to MHC polymorphism and resistance (Lewin ., 1996; Weigel *et al.*, 1990;Ballingal *et al.*, 2011) contemplated the

relationship between MHC diversity and the development of bovine neonatal pancytopenia in Holstein Dairy cattle. The molecular investigation and fine mapping of disease associations will most likely assume a focal part in cattle genetics and veterinary medicine for a long time to come. The expanding resistance of pathogens to anti-microbial underlines the significance of understanding the molecular genetic bases underlying resistance to infectious disease and vaccine development.

Table 3: Association of the BoLA-DRB3.2 alleles with different diseases in different breeds of cattle

Disease	Breed	BoLA allele	Type of association (Susceptibility/Resistance)
Dermatophilosis	Brahman	BoLA-A8, BoLA-DRB3 "EIAY"	Higher resistance
Severe mastitis	Canadian Holstein	BoLA-DRB3.2*23	Higher susceptibility
Retained placenta	Canadian Holstein	BoLA-DRB3.2*03	Lower susceptibility
Subclinical mastitis	Iranian Holstein	BoLA-DRB3.2*08	Higher susceptibility
Tick infestation by Amblomma americanum	*Bos taurus* x *Bos indicus*	DRB3*4401	Resistance
Clinical mastitis	Canadian Holstein	DRB3.2*03, DRB3.2*11	Lower susceptibility
Clinical mastitis	Canadian Holstein	DRB3.2*08	Higher susceptibility
Clinical mastitis	Norwegian Red	DRB3.2*13,*18,*22 and *27	Lower susceptibility
Mastitis	Japanese Holstein	DRB3*0101, *1501	Susceptibility
Mastitis	Japanese Holstein	DRB3*1101, *1401, *1201	Resistance
Lymphosarcoma and persistent lymphocytosis caused by Bovine Leukaemia virus	Holstein	BoLA DRB3*0902	Resistance
Leukaemia	Russian Black Pied	BoLA-DRB3.2*11, *23 and *28	Resistance

Source: Review article by Jyotsna Behl et al., 2012

2.1.11 Use of peptide-binding Motif to predict interaction with MHC class I

Binding to a specific MHC class I or class II allele is a prerequisite for a peptide to be available for T cell recognition. Understanding this interaction at the molecular level is, therefore,of great interest not only from a structural point of view (Garboczi *et al.,* 1996; Garcia *et al.,* 1996), but even more for designing vaccines and identifying potential T cell epitopes. For example, in humans, known association of autoimmune diseases with particular HLA class I or class II repertoires like HLA-DRB1*1501 with multiple sclerosis (Vogt *et al.,* 1994) and HLA-DRB1*0401 with rheumatoid arthritis (Nepom and Erlich, 1991; Hammer et al., 1995) and the knowledge of allele specific MHC anchor motifs have been used to predict T cell epitopes within candidate antigens. Since immunodominance of peptides within a given protein often correlates with their binding affinity, and this in turn with the presence of the MHC binding motif, this approach is reasonable. Further, Class I–associated peptides are usually eight or ten amino acids long. Within this sequence, the positioning of the MHC anchor amino acids is crucial and of relatively greater importance than the anchors in MHC class II molecules. However, the entire peptide backbone and amino acid side-chain interactions, their size, hydrophobicity, and charge all contribute to the MHC-peptide interactions and thus ultimately to binding affinity. As mentioned above, the peptide-binding groove of MHC class II molecules is open at either end and, consequently, MHC class II molecules have been shown to accommodate significantly longer peptides of at least 13 amino acids, with much longer ones being eluted from MHC class II alleles (Chicz *et al.,* 1993; Vogt *et al.,* 1994). However, in humans, peptide lengths down to five and three amino acids, respectively, have been described to be recognized by HLA class I– or class II–restricted T cells (Reddehase *et al.,* 1989; Hemmer *et al.,* 2000). Unlike class I–binding peptides, the peptide-binding motif of MHC class II molecules can be shifted to the N- or C-terminal end, repetitive or multiple motifs may result in binding in different registers, and one peptide may contain binding motifs for a number of different MHC class II alleles, and thus appear to be a promiscuous binder (Sette and Sidney, 1998).

2.1.12 Techniques for MHC typing

Bovine MHC typing has customarily been accomplished through serological testing and additionally utilization of MHC-allele specific PCR or a blend of these strategies (Ellis *et al.* 2005; Ellis *et al.* 1998). Notwithstanding, incomplete understanding of MHCI and MHCII genes implies that it is not conceivable to confirm the outright specificity of MHCI and II

allele/haplotype specific antibodies and PCR profiles and that for animals expressing uncharacterised MHCI and II genes, the right now accessible reagents are of restricted use. Thus, haplotype identification, has ultimately relied upon MHC gene sequence analysis utilizing conventional Sanger sequencing, which is expensive and labour intensive when used on an extensive scale. This is particularly valid for inadequately characterized cattle populations, for which sequencing of different sub-clones of PCR products acquired from every animal utilizing pan MHCI and II primers is required. Of late, the utilization of Next Generation Sequencing (NGS) innovations has changed HLA research and diagnostic typing as reviewed in a few recent articles (Cereb *et al.*, 2015; De Santis *et al.*, 2013; Erlich 2012; Gabriel *et al.*, 2014; Hosomichi *et al.*, 2015). Exon sequencing is the most widely used application of NGS in HLA typing (Gabriel *et al.*, 2014), which for MHCI molecules is typically centred around exon 2 and exon 3, which encode the α1 and α2 domains forming the peptide-binding groove (Lange *et al.*, 2014; Cereb *et al.*, 2015). Outside of the human field, NGS has additionally been applied to the identify of MHC repertoire in various mammalian and feathered species (Dudley *et al.*, 2014; Heimbruch *et al.*, 2015; Oomen *et al.*, 2013; Promerova *et al.*, 2012; Sepil *et al.*, 2012). In this study, a rapid and cost effective a high-throughput NGS bovine MHCI genotyping protocol developed by Vasoya *et al*, (2016) was utilized.

2.2 Domestic cattle

Domestic cattle belong to the subfamily *Bovinae* and genus *Bos*. It is composed of two principal types, taurine (*Bos taurus*) and zebu (*Bos indicus*). The taurine type is partitioned into two lineages: African (*Bos taurus africanus*) and European (*Bos taurus european*). Both the *Bos taurus* and *Bos indicus* cattle, which nearly incorporate all present day cattle are relatives of wild aurochs scientifically known as *Bos primigenius*. The presence of aurochs is dated back to 12000 BC and are accepted to have lived all through Europe, North Africa, and Asia up until 1627 when the last animal died in Poland (Vuure, 2003).

2.2.1 Origins of today's cattle breeds

The most broadly held view is that cattle originated from a solitary domestication event. In any case, this view was disproved by Loftus *et al.*, (1994) who investigated mitochondrial DNA (mtDNA) sequences from representative taurine and zebu breeds recommending that cattle started from two separate domestication events, presumably of genetically distinct populations of aurochs. The European taurine is accepted to have started in the Fertile Crescent between

8800-8300 BC and for the Zebu in the indicus valley roughly 1500 years later (Helmer et al., 2005). The African taurine came from Fertile Crescent and inter-mixed with African aurochs. In spite of the fact that the time of origin of the African taurine is the subject of broad debate, a few authors date the first African taurine cattle to 1600 BC, while alternative confirmation demonstrates bones found in East Africa dating to 7000 years ago. Sanga cattle (a cross breed between indicus and taurus cattle) are found in Sub-Saharan Africa. They are recognized by having little cervico-thoracic humps rather than the high thoracic humps which describe the Zebu. The Zebu is believed to have first arrived in sub-Saharan Africa in the vicinity of 700 and 1500 BC and was introduced to the Horn of Africa around 1000 (Marshal and Fiona, 1989). This introduction of Zebu cattle largely influenced African cattle genetics by hybridizing the indigenous humpless cattle with Zebu cattle (Gibbs *et al.,* 2009), particularly in eastern Africa. This has been shown in an investigation by (Decker *et al.,* 2014) who reported that Africa cattle in western Africa have from 0% to 19.9% Zebu ancestry with an average score of 3.3%. He additionally reported that moving from west to east and from south to central Africa, the percent of Zebu ancestry increases from 22.7% to 74.1% with an average score of 56.9%. The second factor to a great extent accepted to underlay dissimilarity of African taurine from *Bos taurus taurus* is high level of wild African auroch (Linseele, 2004; Stock and Gifford-Gonzalez, 2013) introgression getting as much as 26% of their family line from admixture with wild African auroch. Along these lines the African taurine has been accounted as the most diverged of the taurine cattle populations (Decker *et al.,* 2014). As the relocation of man proceeded southwards, new breeds and types developed with some overwhelmingly Zebu, for example, the Boran, Masai, Sokoto and other Sanga composites (Afrikaner, Nguni and Tuli), demonstrating genetic markers unique to both *B. taurus* and *B. indicus* (Meyer, 1984).

Notwithstanding the migratory routes already described from the Fertile Crescent to Europe (Decker *et al.,* 2014), exportations to China, Southeast Asia, Africa, Iberian Peninsula, Mediterranean Europe and America have also been reported. Subsequent to these underlying exportations, there have been endless exportations and importations of cattle around the world. At the point when domestic cattle were available and new germplasm was imported, the introduced cattle were often crossed with the local cattle bringing about an admixed population. Human development hugely added to the conveyance of cattle world over and like other domestic species, adjustment to different natural conditions, selective breeding practices and worldwide relocation of sires have incredibly affected evolution and genetic diversity of cattle (Marsan *et al.,* 2010). The unique position of Africa has meant it has served as cross-roads for

each of the 3 major lineages of cattle. Domestication, exportation, admixture, and breed development have immensely affected the variety show within and between cattle breeds. In Asia, Africa, North and South America, cattle reproducers have crossbred *Bos t. taurus* and *Bos t. indicus* cattle to create hybrids well suited for their local environment and endemic production systems (Marsan *et al.*, 2010).

2.2.2 Cattle breeds in Zambia

In Zambia, there is a significant lack of empirical data on cattle population and its spatial distribution. Initiatives are currently being developed to conduct a livestock census that would provide a comprehensive database for cattle. To this end, surveys on crop farming have been used to gather some data on cattle population. According to Ministry of Fisheries and Livestock (MFL, 2016), the nationwide cattle population was estimated at 4,984,909 with Southern, Eastern and Western Provinces having the highest cattle population. Indigenous cattle breeds in Zambia are classified into three major breeds according to geographical location and physical appearances. The Tonga predominantly found in Southern Province. The Barotse predominantly found in Western province and Angoni predominantly found in Eastern province. They are rather classified by geographical location and named by areas in which they are found. There is still open deliberation on the fourth breed, the Baila of the Kafue flood plains in Southern Province, whether it is a different breed or only a cross between the neighbouring Tonga and Barotse breeds. Of late, the population of these indigenous cattle has significantly declined due to farmers' inclination on exotic breeds that are seen to be more productive. Farmers in rural areas trust that they could profit through cattle raising by just moving from low-producing indigenous breeds to exotic breeds. This frequently has not worked extremely well for most small scale farmers as most exotic breeds fail to flourish in their new surroundings. In this way, for small scale farmers, keeping of indigenous cattle rather than exotic animals might be invaluable as they are adapted to local conditions and will give some level of production even under harsh conditions, for example, restricted supply of feed and water, and/or high disease exposure.

The geographic distribution of indigenous cattle breeds in Zambia

Figures 5 below show the provincial breed distribution of cattle indigenous to Zambia. The geographical areas are represented by the coloured triangles in the legend.

Figure 5. Provincial breed distribution of indigenous cattle in Zambia

2.2.3 Tonga Breed

The Tonga breed is found in Southern Province especially on the Zambezi Plateau and valley areas surrounding Lake Kariba. The Tonga breed has its origin from the Sanga. For the most part, Tonga breed have a small body frame with medium sized horns that tend to spread outwards from the head (Figure 6). The breed is multi-coloured going from unadulterated black to white, albeit black have a tendency to dominate in many herds. The breed is generally kept for meat despite the fact that the Tonga people will infrequently butcher an animal for food or sell unless there is a funeral service or wedding in the family. In view of its little body stature, the Tonga breed is not very much adjusted as a work animal although almost all farmsteads have a set of working animals that are utilized for cultivation and transportation of produce to and from the market. The breed is likewise utilized as part of cultural and social festivals. Because of its versatility to local conditions, the breed is generally crossbred with exotic breeds to generate crossbred beef and milk animals.

Fig 6. The Tonga cattle breed has a small body framework with medium sized horns and mostly used a work animal (Photo: Dr. J. Simbaya)

2.2.4 Angoni Breed

Angoni cattle are generally found in Eastern Province. The Angoni cattle breed belongs to the Shorthorn Zebu and was brought into Zambia by the Ngoni tribe when they crossed the Zambezi river from the South around 1835 (Dagris, 2007). The breed is additionally found in neighbouring countries such as Malawi, Tanzania and Mozambique. The Angoni breed has a compact body that is for the most part bigger than the Tonga breed, but smaller than the Barotse breed. They differ extensively in colour going from black to roan-brown with a blend of all colours in between (Figure 7). They have generally shorter horns when compared with other cattle breeds. The females have characteristic large udders that make the breed perfect for crossbreeding with exotic breeds for dairying. The large size of males and the compact body make them perfect for draft power both for land cultivation and transportation of goods.. The Angoni breed is also kept for cultural and social activities for example, the Ncwala ceremony and consequently the animals are associated with social status and a symbol of wealth. The cattle are also used as a source of direct income with sales used to meet immediate financial needs such as a funeral , wedding or school expenses for children. Finally, the animals can also be used as a source of protein for consumption by the herding communities.

Fig 7. The Angoni with a compact body has short horns (Photo: Dr. I.K. Silwamba)

2.2.5 Barotse breed

The Barotse breed is a group of cattle found exclusively in Western Province and encompassing surroundings of the Kalahari and Namib Deserts in neighbouring Angola, Namibia and Botswana. The breed originated from Sanga cattle and in Zambia, Lozi people living on the Zambezi flood plain mainly rear it. The principle distinguishing feature of Barotse cattle is the huge, but moderately thin body structure with long horns that spread and bend in reverse (Figure 8). The breed is regularly multi-coloured although usually black with white or brown patches on the sides and under the abdomen. In spite of the large body frame, it is uncommon to see a fat, well rounded Barotse animal, which might be credited to the conditions the animals live in. The animals are either swimming in flood waters amid the rainy season and are faced with water-borne disease challenges. In the dry season they are normally grazing on harvested fields and there are feed deficiencies. Barotse cattle are mainly kept for meat and milk production. The breed is additionally utilized for land cultivating,draft power and manure.

Fig 8. The Barotse has moderately thin body structure with long horns that spread and bend in reverse (Photo: Dr. I.K. Silwamba)

2.2.6 Productive and performance characteristics of cattle breeds in Zambia.

The productive and reproductive performance parameters of indigenous cattle breed in Zambia as established in the early 1970s and late 1980s are shown in the table below. The Tonga cattle (Table 4) breed poses the highest calf birth weight in comparison to the Angoni and the Barotse. Being the focus of this research, better understanding of MHC and conservation of this breed therefore may be very beneficial to farmers who rear these cattle.

Table 4: The productive and reproductive performance parameters of indigenous cattle breed in Zambia

Cattle Breed	Angoni	Tonga	Barotse
Calving percentage (%)	82.5	74.4	78.1
Birth weight (kg)	22.9	25.7	19.8
Weaning weight (kg)	147.3	140.8	167.0
Calf mortality (%)	2.7	4.6	5.3
Weight at 18 months (kg)	207.7	200.0	235.0
Weight at 3 years (kg)	283.3	210.3	255.3
Dress weight (kg)	182.4	145.7	185.5
Milk yield/lactation (kg)	990	850	1160

CHAPTER THREE

MATERIALS AND METHODS

3.1 Study areas

This study was carried out in Mazabuka and Namwala districts (see Figure 9 below) of Southern Province of Zambia and were selected purposively based on availability of a high numbers of Tonga cattle in these areas. The distance between the two sampling sites is 200.5 Kilometres.

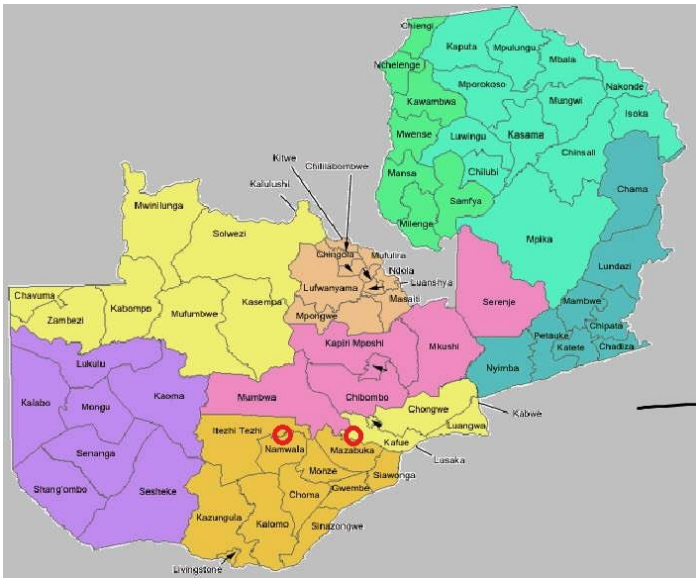


Figure 9. Cattle sampling sites indicated by the red circles. Source: Map world.com

3.2 Sample size

In this study, 347 Tonga cattle were sampled. Conventional formulas for sample size determination were not used in this study because the MHC allele frequencies for the Tonga cattle breed were not known. In addition, the allele frequencies vary in each individual animal and according to breed type. Therefore, no previous study could be used to estimate the allele frequency. However, based on the analysis of sample size used in a previous study by Vasoya *et al*, 2016 it was assumed that a random sample of 347 cattle would be sufficient.

3.3 Sampling strategy

The epidemiological unit of interest in this study was an individual animal, since *breed* is an attribute of an individual animal. Purposive sampling was employed in selecting the province,

districts and veterinary camps and even cattle herds. This was because the interest was a particular cattle breed and not all farmers keep indigenous cattle. To select individual animals from the herds, systematic sampling was used, where by animals were placed in crush pen and systematically sampled. The selected animals were marked and then restrained for blood collection. In situations where there was no crush pen casting down was used as a method of restraint during blood sampling.

3.3.1 Inclusion criteria

Only the Tonga cattle breed was included in this study regardless of sex and age. Identification of Tonga cattle breed within the herds was based on putative phenotypic characteristics, which included a small body frame with medium sized horns that tend to spread outwards from the head and multi-coloured going from unadulterated black to white.

3.3.2 Exclusion criteria

The sick and pregnant animals were not included for humane reasons. This was because most animals required casting down as a method of restraint during blood sampling, more especially that, animal handling facilities were very poor in the target sampling sites.

3.4 Sampling and Laboratory analysis

3.4.1 Blood sampling and extraction of peripheral blood mononuclear cells

A total of 347 blood samples were collected by jugular venipuncture into 10 ml EDTA vacutainer tubes (BD Biosciences, Oxford, UK) and peripheral blood mononuclear cells **(**PBMCs) extracted using ammonium chloride lysis method (Vasoya *et al.*, 2016). Briefly, blood was centrifuged for 10 minutes at 1000xg (2700rpm) using a Kokusan, H – 103N Series swing-bucket centrifuge. After that, the middle layer i.e. buffy coat of the centrifuged blood (Plasma, buffy coat layer and red blood cells) was harvested using a 1 ml Gilson pipette. The harvested buffy coat was then transferred into a 15ml labelled falcon tube before filling it with RBC lysis buffer (0.144 M ammonium chloride/0.0175 M Tris pH 7.4) to ensure a ratio of 1:4. The mixture of buffy coat and lysis buffer was then incubated at room temperature for 5-10 minutes prior to centrifugation for 5 minutes at 300xg (1500 rpm). The supernatant containing the lysed RBC was discarded by pouring off, leaving behind a white pellet (WBC) at the bottom of the tube. The pellet was re-suspended by flicking the tube and 15 ml of Phosphate

buffered saline (PBS) was added. Further, the mixture was centrifuged for 5 minutes at 300xg (1500 rpm) and the supernatant was discarded by pouring off and pellet re-suspended by flicking tube as above. Finally, 1 ml of tri-reagent (Sigma-Aldrich) was added to each re-suspended cell pellet and an aliquot of 0.5ml was transferred to each of 2 labelled Eppendorf tubes while pipetting repeatedly in order to lyse the cells. The tubes were incubated at room temperature for 5 minutes, after which they were stored at – 20 °C until needed for total RNA extraction.

3.4.2 Isolation of Total RNA from PBMCs using the tri-reagent protocol

The white blood cell extracts prepared above were removed from -20°C storage, thawed and 200ul of chloroform added to the tube and vortexed. The tubes were then incubated at room temperature for 2-15 minutes before centrifuging them in a pre-chilled (4 °C) centrifuge (Tomy mix – 207 micro centrifuge) at 12000xg for 15 minutes. The centrifugation resulted in formation of 3 layers of the tube contents – bottom red organic containing protein, a pale coloured interphase containing DNA and the top clear aqueous phase containing RNA. The top clear aqueous phase was pipetted out using a 200µl Gilson pipette and transferred into a clean labelled 1.5 ml eppendorf tube and to it 500µl of isopropanol was added. The mixture was mixed by inverting the tubes and then allowed to stand at room temperature (RT) for 5 – 10 minutes. After incubation at RT, the tubes were then centrifuged at 12000xg for 10 minutes at 4°C. The supernatant was pipetted out leaving behind a white pellet (RNA) on the side of the tube. The Pellet wash up was performed by adding 1 ml of 75% ethanol into the tube and centrifuging at 7500xg for 5 minutes at 4 °C. After that, ethanol was removed by using multiple pipettes as follows - first with a 1ml, then 200µl, and finally with a 10µl Gilson pipette to speed up drying of the pellet. Thereafter, the pellet was air dried until it became almost transparent taking care not to over dry it. Finally, the pellet was re-suspended in 25µl nuclease free water and later stored briefly until needed for use (first strand cDNA synthesis). The concentration of RNA was determined using a nanodrop (Tomy Q5000).

3.4.3 First strand cDNA synthesis

The first strand cDNA was synthesised using a reverse transcription Kit (GOscript kit - Promega A5001) according to manufactures recommendations, but with a few modifications. Briefly, 5µl of RNAase free water was pipetted into the microcentrifuge tubes to which 5ul of RNA sample was added. The tubes were then incubated in a thermocycler (Veriti 96 well

thermocycler) at 70 °C for 10 minutes. After that, the tubes were removed from the thermocycler and centrifuged briefly, and thereafter a mastermix with the following components per sample (4µl 25mM Magnesium Chloride ($MgCl_2$); 2µl 10X Reverse Transcription Buffer; 2µl 10mM dNTP mix; 0.5µl Recombinant RNAsin Ribonuclease inhibitor; 0.6µl AMV reverse transcriptase; and 1µl Oligo $(dT)_{15}$ Primer) were added. The cycling conditions for the reactions were 42 °C for 60 minutes, 95°C for 5 minutes and 0°C for 10 minutes.

3.4.4 GAPDH Polymerase Chain Reaction

The *GAPDH* (Glyceraldehyde-3-Phosphate Dehydrogenase) gene was used as the target gene for quality checks of the synthesized cDNA (Vasoya *et al.*, 2016). GAPDH was targeted because it is a house keeping gene and is always expressed. In brief, the thermo-cycler was pre-heated up to 95°C for 30 minutes. A master mix comprising of (4µl 5X Buffer; 0.6µl Dimethyl sulfoxide (DMSO); 0.4µl dNTP; 1µl Forward primer; 1µl Reverse primer; 11.8µl water; and 0.2µl Phusion enzyme) per sample was prepared adding the Phusion enzyme last and then pipetted out into the micro centrifuge tubes. Thereafter, 1ul of the cDNA template synthesized above was quickly added to the micro centrifuge tubes and the mixture spun down briefly. Finally, the tubes were placed into the thermocycler set to the following cycling conditions; 95 °C for 2 minutes, (98°C for 10s, 63 °C for 30 s, 72 °C for 1 minute) X 30 cycles and final stage at 72 °C for 10 minutes. Gel electrophoresis of the PCR products was performed to check the quality of the cDNA and the gel red stained PCR products were viewed using a Biodoc-it (UVP) Imaging system equipment.

3.4.5 Gene-specific PCR and sequencing

A series of gene specific 3′ (for) and 5′ (rev) pan-MHCI (Annex IV) and MHC II (Annex V) primer incorporating Illumina adaptors and multiplex identifier tags (MID) were used in this study (Vasoya *et al.*, 2016). cDNA from individual animals was subjected to PCR amplification in two separate reactions using either the For1/Rev2 or the For3/Rev1 primer pairs for the MHC I and three separate reactions for the DRB3, DQA, DQB MHC II genes. A unique combination of MID tags for the primers was used to allow subsequent de-multiplexing of the sequence data. PCRs were conducted using the Phusion High-Fidelity PCR kit (New England Bio Labs, Hitchin, UK) with 20 µl reactions composed of Phusion HF amplification buffer, 0.2 mM dNTPs, 3 % DMSO, Nuclease free water, 2 U Phusion Hot Start DNA polymerase, 2.5 pmol

of 3′ and 5′ primers, and 5 µl cDNA. Cycling conditions were 98 °C for 2 minutes for hot start, 98°C for 30s, (98 °C for 10 s, at 61°C for 30 s, 72 °C for 45 s) X 30 cycles and final stage at 72 °C for 10 min. Following amplification, 10 µl of PCR products from each sample were pooled, concentrated using a Qiagen PCR purification kit (500ul concentrated to 100 ul) and run on a 1.5 % agarose gel. Amplicons of the appropriate size were extracted and first purified using the Qiagen Gel extraction kit (Qiagen, Manchester, UK) and then with Agencourt AMPure XP Beads (Beckman Coulter, High Wycombe, UK) at a v/v ratio of 1:1 beads to PCR products. PCR product quantification using 260/280 nm absorption readings was obtained from a Nanodrop spectrophotometer (Wilmington, DE, USA). The PCR products were submitted to Edinburgh Genomics for sequencing on an Illumina MiSeq v3. The raw reads from the sequencing were then obtained for analysis.

3.5 Data Analysis

A bioinformatics pipeline developed by Vasoya *et al.*, (2016) was used to analyse sequence data generated in this study and corrected for various artefacts introduced during amplification/sequencing procedures. Briefly, this pipeline included (i) a cut-off threshold to exclude variants (unique sequences) that constituted <0.2% of reads in the sample that passed quality control filtering. (ii) an algorithm to detect and remove variants resulting from chimaera formation between more frequent variants within the same sample; (iii) removal of variants that had 1 or 2 bp differences from a variant that was present in the same sample but were present at <30 or <50- fold its frequency respectively, on the basis that these likely represented PCR/sequencing error; and (iv) removal of amplicons that were ± 9 bp the anticipated product size (as these likely represented splice variants). In addition, the pipeline automatically identified previously defined bovine MHCI haplotypes when all of the composite alleles have been identified within a sample. However, manual annotation was used to construct novel MHCI and MHCII haplotypes that were formed from either novel alleles or new combinations of known and/ or novel alleles. All MHCII haplotypes were defined by co-occurrence of the same DRB3, DQA and DQB alleles. To be considered a novel haplotype the combination of alleles was required to be present in at least two individuals. Analysis of MHCI and MHCII haplotype sharing was performed using the Bioinformatics & Evolutionary Genomics tool (http://bioinformatics.psb.ugent.be/webtools/Venn/) while haplotype frequency distributions were generated using Microsoft excel. Phylogenetic analysis of nucleotide sequences was

performed with CLC main workbench version 8.0.1 using K-mer based approach (Construction method; neighbour joining, k-mer length; 15, Distance measure; Mahalanobis).

.

3.6 Nomenclature

Novel MHCI alleles and haplotypes identified in Zambia were given a number prefixed by 'za' (e.g. za 1.1 identifies the first novel allele identified in the MiSeq run and zaHP.1.1 for the first haplotype). Similarly, novel MHCII (DRB3, DQA, DQB) alleles and haplotypes were given a number prefixed by 'za' (e.g. za1.DRB3.1:28.27 identifies the first novel allele identified in the first MiSeq run and MHCII.HP.za.1 for the first haplotype). Different combinations of these prefixes were employed as necessary. Previously defined bovine MHCI alleles and haplotypes were prefixed by a two letter code defining a country in which they were first characterised (e.g. ke (Kenya), cm (Cameroon), gb (great Britain), it (Italy), ug (Uganda) and br (Brazil). BoLA-DRB3, DQA and DQB alleles in this study followed the nomenclature decided at the Prague, Czech Republic BoLA workshop. This nomenclature follows the human nomenclature for HLA alleles. For instance, DRB3*0101 denotes an allele at the DRB3 locus. The allele designation is composed of four digits where the first two digits indicate which major allelic type the allele belongs to and the last two digits represent the subtype within a major type; alleles which differ by fewer than five amino acid substitutions are considered subtypes of the same major allelic type.

3.7 Ethical approval and Nagoya protocol clearance

Ethical clearance was sought from and granted (ANNEX I) by Excellence in Research Ethics and Science (ERES) Ref. No 2018-Jan-009 before commencement of the study. Permission to export genetic material (cDNA) to the UK for analysis was sought and granted (ANNEX II) by Patents and Company Registration Authority (PACRA) through the Ministry of Lands and Natural Resources. The district Veterinary officers were informed of the study and consent (ANNEX III) was sought from cattle owners before commencement of the study. Upon arrival at the cattle sampling site, the objectives of the study were explained to the veterinary camp officers and the cattle owners and permission was sought to collect blood samples from the animals. The farmers were informed that their participation in this research was completely voluntary and that they have the right to choose whether or not to participate in the study. Further, the farmers were assured that withdraw from the study will not affect their rights as cattle owners and that they will remain entitled to the same veterinary services provision as other livestock

owners. Only qualified veterinary personnel were allowed to collect the blood samples by jugular venipuncture so as to ensure that the procedure was done with minimal discomfort to the animals. During sample collection, the animals were not denied access to clean water and food.

3.8 Study limitations

This study did not determine the functional importance of the haplotypes and alleles identified other than simply cataloguing polymorphisms. Furthermore, it was not be able to state with certainty that all of the haplotypes and alleles present in the animals studied have been successfully identified. Also, it can not be stated with certainty that all cattle studied are purely Tonga cattle breed due to lack of characterisation of the breed. However, efforts were made to select animals based on the phenotypic attributes of Tonga cattle.

CHAPTER FOUR

RESULTS

4.1 Sample characteristics

Table 5 below shows the structure of the sample-set used in this study. A total of 347 Tonga cattle were in included of which, 25.6% **(n = 89)** were male and 74.4% **(n = 258)** were female. These cattle were from 4 herds in different locations of Namwala and Mazabuka districts of Southern province of Zambia. For each of the 3 herds from Katantila, Muchila and Ndema Veterinary camps of Namwala district 119, 99 and 90 animals respectively (25.9-34.3% of the sample-set) were sampled, whilst from Mazabuka research station the number of animals was substantially smaller (n = 39, 11.3%). All cattle herds were >50 km apart and managed under traditional farming system conditions. The majority of animals (51.9%) were within the productive age group 1-5 years and a large proportion (72.6%) of animals were born within the herd increasing the likelihood of gene localisation. No twinning was recorded. Twining data is important when analysing homozygous haplotypes within the dataset.

Table 5: Structure of cattle sample-set analysed in this study (n = 347)

Variable	**n (%)**
Sex	
Male	89 (25.6)
Female	258 (74.4)
Age group (yrs)	
< 1	36 (10.4)
1 – 5	180 (51.9)
> 5	124 (35.7)
Unknown	7 (2.0)
Herd status	
Born within herd	252 (72.6)
Brought in	95 (27.4)
Twin	
Yes	0 (0.0)
No	347 (100)
Sampling sites (cohorts)	
Mazabuka research station	39 (11.2)
Katantila	119 (34.3)
Muchila	99 (28.5)
Ndema	90 (25.9)

4.2 Analysis of MHCI and MHCII haplotype and allele repertoires

In this study alleles of the BoLA class I and BoLA class II (DRB3, DQA and DQB) loci, from 347 Tonga cattle from four discrete sampling locations (Table 5) were sequenced using the Illumina MiSeq NGS platform. Bioinformatic analysis of the sequence data identified a total of 137 MHCI haplotypes (Table 6), of which 37 were novel. Together these novel haplotypes included 13 novel alleles and sequences matching 88 known alleles. For 255 animals two haplotypes were defined, for 34 animals only one haplotype and no additional alleles were identified (suggesting homozygosity), whilst for 12 animals one haplotype plus a unique combination of alleles were identified (suggesting heterozygosity with an as yet undefined haplotype; definition of a haplotype requires its identification in a minimum of two individuals). These twelve undefined haplotypes included 20 known and 12 putative novel alleles. In 25 animals, three to four haplotypes were defined (possibly due to unrecognised/unrecorded twinning) and in the remaining 21 animals, no sequencing data was generated. GAPDH analysis suggests these samples were negative due to absence of viable cDNA (cDNA upon which gene specific PCR can be performed). Two haplotypes cmHP.137 and keHP3.6 were the most dominant (with haplotype frequencies within the population >5%) while the rest were detected at frequencies that ranged from 0.2 – 4.2%. Likewise, the most dominant MHCI alleles 3*00402, 2*02501, 4*02401 (Table 6) were present at frequency >2.5%, while the rest were seen at moderately low frequency that ranged from 0.3 to 1.4%.

A total of 67 novel MHCII haplotypes were identified (Table 7 – MHCII haplotypes have been defined by co-occurrence of the same DRB3, DQA and DQB alleles). Together these MHCII haplotypes included 108 novel alleles and sequences matching 121 known alleles. Two haplotypes, MHCII.HP.za.5 and MHCII.HP.za.2 were the most dominant (with frequencies within the population > 6%) while the rest were detected at frequencies that ranged from 0.2 – 4.9%. Likewise, the most dominant DRB3 alleles was BoLA-DRB3*2201:38.99 present at frequency 11.9%, while the rest were seen at frequency between 1.5 to 7.5%. The most dominant DQA allele was za1.DQA.1:59.79 present at frequency 20.3%, while the rest were seen at frequencies between 1.4 to 6.6%. The most dominant DQB allele was BoLA-DQB*0302:51.99 present at frequency of 26.1%, while the rest were seen at frequency between 1.1 to 17.0%.

Table 6: MHCI haplotypes and alleles identified in this study. For each haplotype (column 1), the constituent alleles are defined (columns 2-6). Novel haplotypes are highlighted in green and novel alleles in blue.

Haplotypes [n = 137]	Allele 1	Allele 2	Allele 3	Allele 4	Allele 5	# of times each haplotype was seen
zaHP1.brHP1.36	3*07301	3*00402	cm1.16	ke3.9		11
zaHP1.brHP1.13	br1.13	2*05701	cm1.101(For3Rev1)	3*03301N		2
zaHP1.BF8	3*03601	3*03701	2*05601	gb1.6		7
zaHP1.9	br1.36	cm1.12				2
zaHP1.8	cm1.18	ke2.3				6
zaHP1.7	ke1.8	2*04801	3*00402			2
zaHP1.6	br1.89	5*07201				8
zaHP1.5	br1.1	gb1.7	za1.69	2*00601		2
zaHP1.4	za1.31	za1.19	br1.68(For1Rev2)			6
zaHP1.38	ke1.6	ke1.4	gb1.10	6*04001		2
zaHP1.35	5*07201					1
zaHP1.34	ke1.8	cm1.88	3*00402	cm1.46		3
zaHP1.33	za2.7	4*02401	2*02501			2
zaHP1.32	cm1.15	cm1.5	cm1.19	br1.18		2
zaHP1.31	ke2.4	cm1.35				2
zaHP1.30	ke1.16	ke1.14				4
zaHP1.3	ke2.12	br1.90	cm1.75			7
zaHP1.29	za2.1	ke2.3	ke1.26			9
zaHP1.25	za1.123	3*00402	cm1.46			1
zaHP1.23	za1.98	cm1.29				3
zaHP1.22	2*04701	cm1.59	gb1.7			4
zaHP1.21	ke1.24	4*02401	2*02501			3
zaHP1.20	3*01001	za1.35				3
zaHP1.2	za1.3	cm1.33				10
zaHP1.19	za1.37	za1.19	br1.68(For1Rev2)			2
zaHP1.18	2*03202	3*06801				4
zaHP1.16	za1.66	br1.28				5
zaHP1.15	ug1.27	cm1.33				10
zaHP1.14*	za1.28	2*05701	3*03301N			1
zaHP1.14	za1.28	2*05701	3*03301N	cm1.101(For3Rev1)		3
zaHP1.13	ug1.16	2*04701	3*01001			3
zaHP1.12	cm1.62	cm1.66				2
zaHP1.11	ug1.29	2*02501	6*04001			2
zaHP1.10	za1.45	cm1.66				2
zaHP1.1*	br1.10	ke1.4				7
zaHP1.1	br1.10	ke1.4	ke3.3			27
za1.cmHP1.5*	cm1.30	cm1.35	3*00402			4
ugHP1.8	cm1.29	ug1.11	cm1.41			2
ugHP1.7	ug1.14	2*04801	gb1.7			3
ugHP1.23*	ke2.5	cm1.4				1
ugHP1.23	cm1.16	ke2.5	cm1.4			20
ugHP1.20	br1.1	ug1.15	ke1.28			1
ugHP1.2	ug1.18	1*07401				10

Table 6 Continued

ugHP1.19	2*04701	ug1.16	cm1.5			5
ugHP1.12	ug1.22	ug1.23				3
ugHP1.1	ug1.2	ug1.3				1
keHP3.6	br2.1	cm1.86				**34**
keHP3.4	3*01101	gb1.7	2*04801			2
keHP3.3	ke3.6	cm1.29	cm1.41			4
keHP3.11	cm1.39	ke3.4				2
keHP3.1	ke1.1	ke3.1	3*00402			1
keHP2.2	ke2.4	2*04801				3
keHP1.7	ke1.9	ke1.11	ke1.12			4
keHP1.6	ke1.8	3*00402	ke1.3			9
keHP1.5	ke1.7	3*00102	cm1.34			3
keHP1.3**	ke1.2	4*02401				2
keHP1.3*	ke1.2	4*02401	2*02501			1
keHP1.3	ke1.2	4*02401	2*02501	6*04001		4
keHP1.2	ke1.1	cm1.8	3*00402	cm1.10(For3Rev1)		2
keHP1.18	ke1.18	cm1.33				1
keHP1.16*	ke1.22	cm1.80	cm1.84	cm1.96(For3Rev1)		1
keHP1.16	ke1.22	cm1.80	cm1.84	cm1.96(For3Rev1)	cm1.99(For1Rev2)	4
keHP1.15	1*02801	4*02401				1
keHP1.13*	ke1.19	3*00402				1
keHP1.13	ke1.19	3*00402	ke1.3			4
itHP1.9	it1.11	2*06001	cm1.28			7
itHP1.4	it1.3	2*00801				1
itHP1.11	2*01601	6*01401				1
H2	5*03901					18
gbHP1.A10	3*00201	2*01201	gb1.3	gb1.8		1
gbHP1.3***	2*01602	gb1.15				3
gbHP1.3**	2*01602	gb1.15	gb1.17			3
gbHP1.3*	2*01602	gb1.15	gb1.14(For1Rev2)			2
gbHP1.3	2*01602	gb1.15	gb1.17	gb1.14(For1Rev2)		14
gbHP1.1	gb1.11	4*02401				8
cmHP1.7	cm1.13	cm1.6(For3Rev1)				12
cmHP1.5	cm1.5	cm1.30	cm1.35			5
cmHP1.41*	cm1.62	cm1.66	2*01601			6
cmHP1.41	cm1.62	cm1.66	2*01601	cm1.101(For3Rev1)		1
cmHP1.40*	cm1.39					3
cmHP1.40	cm1.39	cm1.98(For3Rev1)				1
cmHP1.4*	cm1.5	cm1.14	cm1.15			1
cmHP1.4	cm1.5	cm1.14	cm1.15	cm1.19		2
cmHP1.37	cm1.40	3*05901				**36**
cmHP1.26*	cm1.53	cm1.56				1
cmHP1.26	cm1.53	cm1.56	cm1.57			1
cmHP1.23	3*00102	cm1.34	cm1.37			4
cmHP1.21	cm1.48	cm1.41	cm1.29			1

Table 6 Continued

cmHP1.2	cm1.3	3*03701	cm1.51			4
cmHP1.18	cm1.44	cm1.28	2*06001			3
cmHP1.17*	cm1.25	cm1.26				2
cmHP1.17	cm1.25	cm1.26	cm1.22	cm1.21(For3Rev1)		5
cmHP1.12	cm1.17	2*04701	2*03201N			1
cmHP1.11	cm1.16	cm1.68				1
cmHP1.10	cm1.4	cm1.16	cm1.27			1
cmHP1.1	cm1.1	cm1.2				11
brHP1.8* **	br1.11	3*06501				1
brHP1.8*	br1.11	3*06501	gb1.17			1
brHP1.7	br1.16	3*03701	cm1.3			11
brHP1.60	2*01601	br1.5				1
brHP1.6	br1.4	2*04801				4
brHP1.59	cm1.23	br1.3				3
brHP1.55	1*02101	br1.74				8
brHP1.54*	br1.82	3*00402				2
brHP1.54	br1.82	3*00402	ke1.3			6
brHP1.53	br1.79	cm1.33	cm1.46			13
brHP1.47	3*03501	3*00402	br1.57	br1.89		2
brHP1.45	br1.54	cm1.23	3*00402	ke1.28		8
brHP1.43	cm1.49	cm1.3	cm1.81			2
brHP1.42*	br1.49	br1.51				1
brHP1.42	br1.49	br1.51	br1.65(For3Rev1)	cm1.101(For3Rev1)		1
brHP1.4*	br1.1	br1.5				4
brHP1.4	br1.1	br1.5	ke1.28			15
brHP1.37	3*03701	br1.48				2
brHP1.33	br1.41	br1.35				9
brHP1.32	br1.42	br1.9				2
brHP1.31	3*00402	3*01001				11
brHP1.3	br1.2	br1.6	2*00601			15
brHP1.25	cm1.91	cm1.88	cm1.46			2
brHP1.23	br1.27	br1.24	ke1.5	2*03201N		1
brHP1.22*	br1.25	cm1.14				1
brHP1.22	br1.25	cm1.14	cm1.19			2
brHP1.20	br1.26	2*00801				1
brHP1.18	br1.18					1
brHP1.16	br1.21	cm1.33				1
brHP1.14*	cm1.73	br1.17	br1.31(For3Rev1)			1
brHP1.1	br1.28	br1.29				2
BF5	2*01601					4
A20v*	3*02702					7
A15v*	1*00902	4*02401	2*02501			1
A15v	1*00902	4*02401	2*02501	6*04001		4
A15	1*00901	4*02401	2*02501			7
A14	1*02301	4*02401	2*02501	6*04001		9
A12(w12B)*	1*01901					2
A12(w12B)	1*01901	2*00801				7
A12(A30)	1*02001	2*00801				1
A11	3*01701	3*03301N	2*01801			2

Table 7: MHCII haplotypes and alleles identified in this study. For each haplotype (column 1), the constituent DRB3, DQA and DQB alleles are defined (columns 2-6). Novel haplotypes are highlighted in green and novel alleles in blue.

Haplotype [n = 67]	DRB3 Allele	DQA Allele 1	DQA Allele 2	DQB Allele 1	DQB Allele 2	n
MHCII.HP.za.1	BoLA-DRB3*0201:35.94	BoLA-DQA*0203:43.50		BoLA-DQB*0201:60.08		16
MHCII.HP.za.2	BoLA-DRB3*2801:85.93	za1.DQA.1:46.74	BoLA-DQA*2603:40.12	BoLA-DQB*1702:54.91	BoLA-DQB*0601:25.76	**42**
MHCII.HP.za.3	za1.DRB3.2:28.27	BoLA-DQA*0203:21.88		BoLA-DQB*3602:51.68		27
MHCII.HP.za.4	za1.DRB3.9:39.00	BoLA-DQA*0301:20.68		BoLA-DQB*1805:47.15		9
MHCII.HP.za.5	za1.DRB3.1:41.28	za1.DQA.2:33.96		BoLA-DQB*0402:30.46		**61**
MHCII.HP.za.6	za1.DRB3.3:40.72	za1.DQA.1:59.79	za1.DQA.8:9.08	BoLA-DQB*0901:22.94	BoLA-DQB*4101:9.12	15
MHCII.HP.za.7	BoLA-DRB3*1801:40.49	BoLA-DQA*2301:50.85		BoLA-DQB*0302:23.73	za1.DQB.2:16.80	20
MHCII.HP.za.8	BoLA-DRB3*03021:35.75	za1.DQA.3:32.30		BoLA-DQB*0302:51.99		23
MHCII.HP.za.9	BoLA-DRB3*2601:19.33	BoLA-DQA*2603:21.19	za1.DQA.17:5.58	BoLA-DQB*0601:12.36	BoLA-DQB*4101:9.04	4
MHCII.HP.za.10	za1.DRB3.13:38.43	za1.DQA.24:23.43		BoLA-DQB*0402:44.99		4
MHCII.HP.za.11	BoLA-DRB3*0901:79.54	za1.DQA.1:45.61	za1.DQA.17:11.61	BoLA-DQB*4101:13.39	BoLA-DQB*0901:2.43	2
MHCII.HP.za.12	BoLA-DRB3*2201:36.30	za1.DQA.1:26.81	za1.DQA.14:6.89	BoLA-DQB*0502:12.39	BoLA-DQB*0701:4.64	14
MHCII.HP.za.13	BoLA-DRB3*2201:38.99	za1.DQA.23:80.69		za1.DQB.6:27.66		8
MHCII.HP.za.14	BoLA-DRB3*25011:44.69	za1.DQA.11:47.81	za1.DQA.5:17.09	BoLA-DQB*2001:21.51	BoLA-DQB*3201:4.70	4
MHCII.HP.za.15	BoLA-DRB3*2101:32.17	BoLA-DQA*0301:53.11		BoLA-DQB*1501:35.97		18
MHCII.HP.za.16	BoLA-DRB3*2101:43.71	BoLA-DQA*0301:59.19		za1.DQB.4:37.08		19
MHCII.HP.za.17	BoLA-DRB3*1201:42.42	za1.DQA.1:44.19	za1.DQA.9:11.60	BoLA-DQB*1005:14.30	BoLA-DQB*0901:12.38	9
MHCII.HP.za.18	za1.DRB3.4:25.97	za1.DQA.1:30.61	za1.DQA.4:2.99	za1.DQB.3:13.87	BoLA-DQB*1001:8.54	30
MHCII.HP.za.19	BoLA-DRB3*3101:39.78	za1.DQA.12:85.36	za1.DQA.7:25.35	BoLA-DQB*2401:55.37	BoLA-DQB*3001:10.77	11
MHCII.HP.za.20	BoLA-DRB3*3101:26.08	BoLA-DQA*2602:54.38	za1.DQA.7:30.30	BoLA-DQB*1301:18.35	BoLA-DQB*3001:9.58	9

Table 7 Continued

MHCII.HP.za.21	newza2.DRB3.1:37.81	za1.DQA.3:38.80		BoLA-DQB*0302:42.30		21
MHCII.HP.za.22	za1.DRB3.10:60.93	BoLA-DQA*27011:65.24		za1.DQB.12:20.90		8
MHCII.HP.za.23	za1.DRB3.7:42.68	BoLA-DQA*2602:46.73	za1.DQA.19:27.35	BoLA-DQB*3001:17.88	BoLA-DQB*1602:13.60	13
MHCII.HP.za.24	newza2.DRB3.5:40.93	za1.DQA.1:48.20	za1.DQA.9:12.03	BoLA-DQB*1005:4.71		10
MHCII.HP.za.25	newza2.DRB3.6:42.08	za1.DQA.5:43.62		BoLA-DQB*1802:44.68		9
MHCII.HP.za.26	za1.DRB3.5:30.32			BoLA-DQB*2501:25.01		25
MHCII.HP.za.27	za1.DRB3.6:40.03	za1.DQA.3:21.48		BoLA-DQB*2502:12.50		16
MHCII.HP.za.28	newza2.DRB3.2:35.28	za1.DQA.1:31.19		BoLA-DQB*2601:27.45		11
MHCII.HP.za.29	za1.DRB3.18:52.74			BoLA-DQB*2501:9.50		6
MHCII.HP.za.30	za1.DRB3.14:45.61			BoLA-DQB*0101:38.58		2
MHCII.HP.za.31	za1.DRB3.14:61.82			BoLA-DQB*1805:11.10		3
MHCII.HP.za.32	BoLA-DRB3*4401:31.75			BoLA-DQB*2701:20.97		1
MHCII.HP.za.33	BoLA-DRB3*14011:28.29			za1.DQB.13:20.65		5
MHCII.HP.za.34	BoLA-DRB3*2301:40.08	BoLA-DQA*0102:50.48				2
MHCII.HP.za.35	BoLA-DRB3*1002:31.44	za1.DQA.13:33.62	BoLA-DQA*0401:17.95	BoLA-DQB*1809:30.72	za1.DQB.5:20.11	8
MHCII.HP.za.36	za1.DRB3.8:39.55	za1.DQA.20:54.26		BoLA-DQB*0401:39.84		3
MHCII.HP.za.37	za1.DRB3.8:32.72	za1.DQA.11:47.05	za1.DQA.5:19.78	BoLA-DQB*2001:39.31	BoLA-DQB*3201:2.46	5
MHCII.HP.za.38	BoLA-DRB3*2703:33.08	BoLA-DQA*0203:88.23		BoLA-DQB*0101:22.17		3
MHCII.HP.za.39	BoLA-DRB3*0701:26.98	za1.DQA.21:39.45		za1.DQB.11:24.46	za1.DQB.10:15.90	5
MHCII.HP.za.40	BoLA-DRB3*1501:29.15	za1.DQA.7:50.32	za1.DQA.22:16.08	BoLA-DQB*1301:29.72	BoLA-DQB*10021:7.63	4
MHCII.HP.za.41	BoLA-DRB3*2802:57.57	za1.DQA.5:45.84		BoLA-DQB*1802:37.26		5
MHCII.HP.za.42	za1.DRB3.15:40.20	za1.DQA.5:29.35		BoLA-DQB*1901:46.85		12
MHCII.HP.za.43	BoLA-DRB3*1302:52.92	BoLA-DQA*27011:22.92		za1.DQB.6:27.36		2
MHCII.HP.za.44	BoLA-DRB3*1001:46.85	za1.DQA.7:23.77	za1.DQA.4:7.48	BoLA-DQB*1301:15.00	BoLA-DQB*1001:6.79	2

Table 7 Continued

MHCII.HP.za.45	BoLA-DRB3*0902:46.84	BoLA-DQA*0301:87.97		BoLA-DQB*3302:53.84		3
MHCII.HP.za.46	za1.DRB3.3:41.24	za1.DQA.15:45.96	za1.DQA.10:43.56	BoLA-DQB*1702:23.89	BoLA-DQB*10021:18.03	5
MHCII.HP.za.47	za1.DRB3.3:24.53			BoLA-DQB*1805:27.37	BoLA-DQB*2001:22.31	3
MHCII.HP.za.48	za1.DRB3.11:47.42	za1.DQA.18:47.32		za1.DQB.14:48.10		4
MHCII.HP.za.49	za1.DRB3.12:38.33	za1.DQA.3:40.58		BoLA-DQB*0302:38.89		5
MHCII.HP.za.50	BoLA-DRB3*2601:33.91	za1.DQA.1:44.01	za1.DQA.10:17.99	BoLA-DQB*0901:14.72	BoLA-DQB*1009:11.07	1
MHCII.HP.za.51	BoLA-DRB3*3601:46.96	za1.DQA.5:13.83		BoLA-DQB*1802:42.38		2
MHCII.HP.za.52	za1.DRB3.16:88.28	BoLA-DQA*0103:44.00		BoLA-DQB*2501:83.55		2
MHCII.HP.za.53	newza2.DRB3.9:28.84	za1.DQA.21:31.42		za1.DQB.11:33.84		5
MHCII.HP.za.54	newza2.DRB3.12:36.60	za1.DQA.7:73.83	za1.DQA.22:16.65	BoLA-DQB*10021:14.73		3
MHCII.HP.za.55	newza2.DRB3.4:42.03			BoLA-DQB*0302:11.32	za1.DQB.10:5.35	5
MHCII.HP.za.56	newza2.DRB3.3:25.79			za1.DQB.13:26.31		12
MHCII.HP.za.57	newza2.DRB3.13:40.86	za1.DQA.5:10.32		BoLA-DQB*1901:64.66		4
MHCII.HP.za.58	newza2.DRB3.14:41.61	BoLA-DQA*0801:34.50		BoLA-DQB*0801:41.19		3
MHCII.HP.za.59	newza2.DRB3.7:20.12	za1.DQA.11:45.38		BoLA-DQB*2501:7.01		7
MHCII.HP.za.60	newza2.DRB3.10:40.53	BoLA-DQA*0203:85.44		BoLA-DQB*0201:60.25		4
MHCII.HP.za.61	newza2.DRB3.15:39.59			BoLA-DQB*2701:38.29		2
MHCII.HP.za.62	za1.DRB3.24:26.73			BoLA-DQB*0101:49.23		3
MHCII.HP.za.63	za1.DRB3.19:51.29	BoLA-DQA*25012:47.29	za1.DQA.1:22.10	BoLA-DQB*1702:29.78	BoLA-DQB*0601:13.84	2
MHCII.HP.za.64	BoLA-DRB3*2201:27.06	za1.DQA.4:7.78		BoLA-DQB*1001:18.37		4
MHCII.HP.za.65	za1.DRB3.23:35.28			BoLA-DQB*2301:19.34		2
MHCII.HP.za.66	newza2.DRB3.18:34.97			newza2.DQB.1:59.08		2
MHCII.HP.za.67	newza2.DRB3.11:49.28		za1.DQA.5:42.04	BoLA-DQB*1802:33.52		3

4.3 Analysis of distribution of MHCI and MHCII haplotypes among cohorts

A total of 83, 80, 64, and 20 MHCI and 46, 42, 35 and 16 MHCII haplotypes were identified in Muchila, Katantila, Ndema and Mazabuka, respectively (Table 8). Due to some sharing of haplotypes between the cohorts the total number of MHCI and MHCII haplotypes identified was 137 and 67 respectively.

Table 8: Number of unique MHCI and II haplotypes identified in each cohort

Cohort	Number of MHCI haplotypes	Number of MHCII haplotypes
Katantila	80	42
Mazabuka	64	16
Muchila	83	46
Ndema	20	35
Total number of haplotypes	**137**	**67**

The results of this study show a 5.8% **(n = 8)** and 9.0% **(n = 6)** sharing of MHCI and MHCII haplotypes among the four animal cohorts studied (Figures 10 and 11 and Tables 9 and 10). More detailed analysis of haplotype sharing shows that only a limited number of MHCI haplotypes (n=8) were observed in all 4 cohorts. Although there were a large number of haplotypes that were observed in multiple cohorts (n=69), nearly the same number of haplotypes were observed in only one cohort (n=66). This shows a surprising level of differentiation between the cohorts with nearly half of the MHCI haplotypes defined being segregated to individual herds.

Figures 12 and 13 display the frequency distribution of MHCI and MHCII Haplotypes. In the Muchila and Ndema cohorts, keHP3.6 was the most commonly observed haplotype, but was rare in Mazabuka and Katantila cohorts, where the most commonly seen haplotypes were zaHP1.1 and cmHP.137 respectively. Of the haplotypes detected in the both Muchila and Katantila cohorts, MHCII.HP.za.5 was the most common, but was moderately rare in Mazabuka and Ndema cohorts, where most frequently observed haplotypes were MHCII.HP.za.2 and MHCII.HP.za.21, respectively.

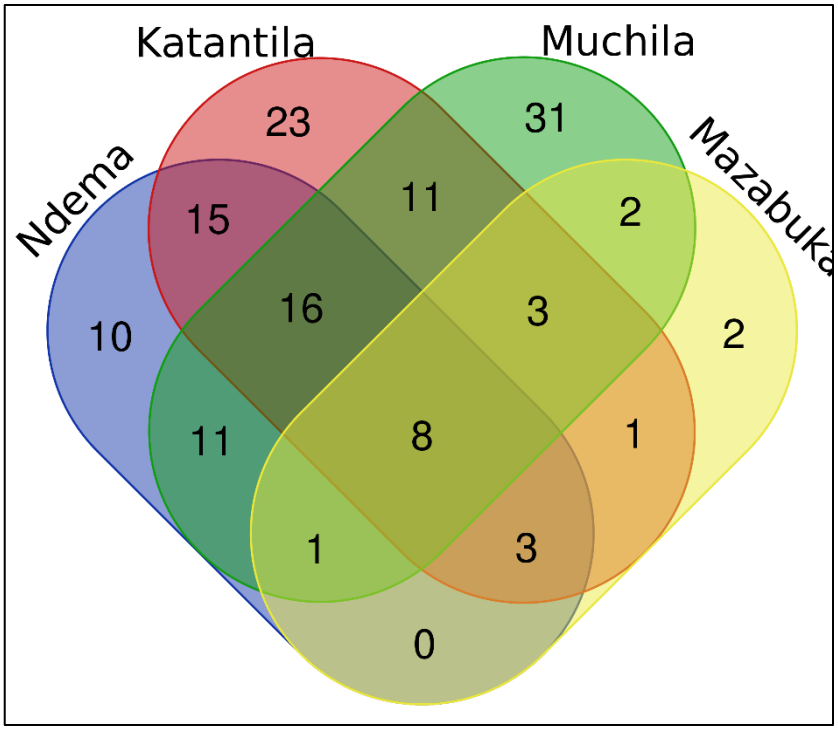


Fig 10. Analysis of MHCI haplotype sharing. The figure was prepared using the Bioinformatics & Evolutionary Genomics tool(http://bioinformatics.psb.ugent.be/webtools/Venn/). Cohorts are coloured and details are as described in Table 9 below.

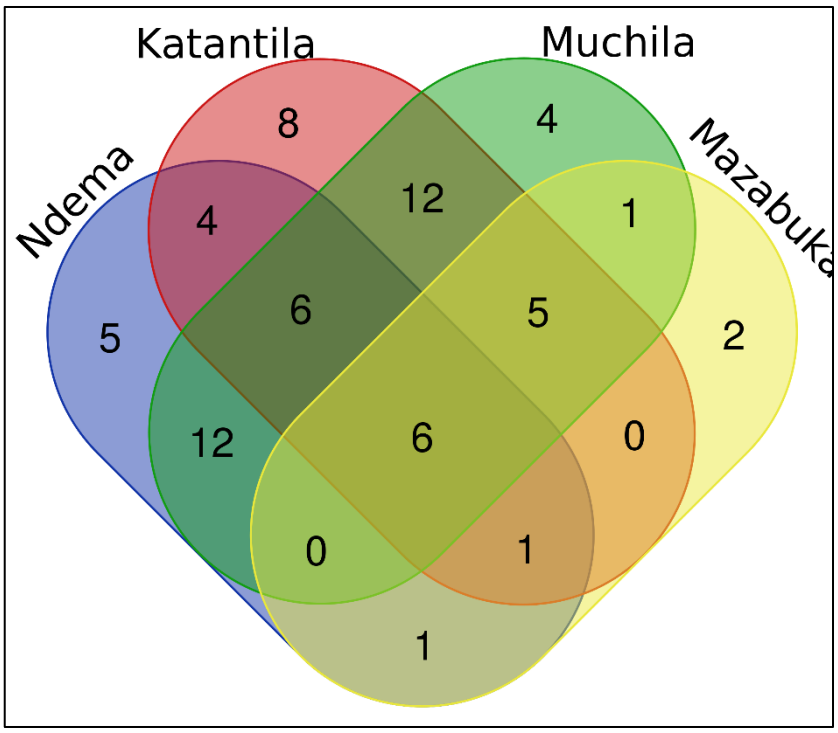


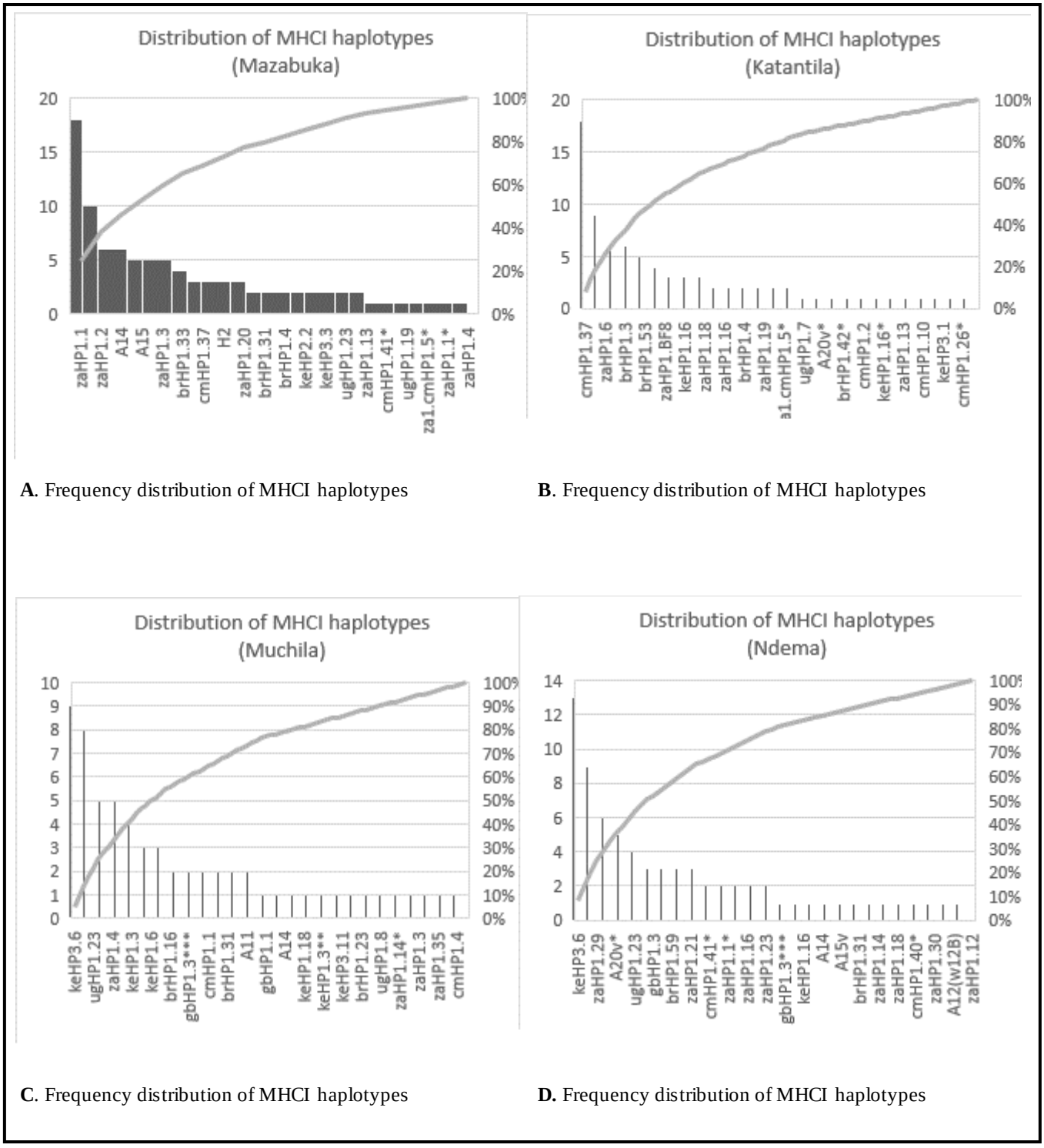
Fig 11. Analysis of MHCII haplotype sharing. The figure was prepared using the Bioinformatics & Evolutionary Genomics tool(http://bioinformatics.psb.ugent.be/webtools/Venn/). Cohorts are coloured and details are as described in Table 10 below.

Table 9: Analysis of MHCI sharing in the four study cohorts

Cohort(s)	# of unique haplotypes	Haplotypes names
Katantila Mazabuka Muchila Ndema	8	A14 brHP1.4 brHP1.31 H2 brHP1.33 cmHP1.37 zaHP1.1 ugHP1.23
Katantila Muchila Ndema	16	zaHP1.15 cmHP1.7 zaHP1.BF8 cmHP1.2 keHP3.6 zaHP1.brHP1.36 zaHP1.8 brHP1.7 gbHP1.3 A15v A12(w12B) brHP1.45 brHP1.53 gbHP1.1 zaHP1.16 A20v*
Katantila Mazabuka Ndema	3	cmHP1.41* zaHP1.1* keHP3.3
Mazabuka Muchila Ndema	1	zaHP1.3
Katantila Mazabuka Muchila	3	A15 za1.cmHP1.5* ugHP1.19
Katantila Ndema	15	brHP1.3 zaHP1.12 zaHP1.18 cmHP1.23 brHP1.55 keHP1.16 keHP1.7 zaHP1.14 brHP1.54 itHP1.9 zaHP1.22 zaHP1.brHP1.13 zaHP1.23 brHP1.47 ugHP1.7
Muchila Ndema	11	gbHP1.3* keHP3.11 gbHP1.3*** keHP1.2 cmHP1.40* A12(w12B)* cmHP1.17 keHP1.3** zaHP1.34 zaHP1.30 brHP1.37
Katantila Muchila	11	ugHP1.8 keHP1.6 cmHP1.4 keHP1.5 ugHP1.2 brHP1.25 brHP1.6 cmHP1.5 cmHP1.18 cmHP1.1 keHP1.13
Katantila Mazabuka	1	zaHP1.13
Mazabuka Muchila	2	zaHP1.4 keHP2.2
Ndema	10	brHP1.14* cmHP1.4* zaHP1.29 zaHP1.31 cmHP1.17* cmHP1.40 zaHP1.21 cmHP1.41 brHP1.59 brHP1.54*
Katantila	23	zaHP1.5 zaHP1.7 brHP1.18 brHP1.43 keHP3.1 cmHP1.12 gbHP1.A10 brHP1.1 itHP1.4 keHP3.4 cmHP1.11 zaHP1.6 brHP1.32 brHP1.42* keHP1.16* zaHP1.11 zaHP1.32 zaHP1.19 cmHP1.26* ugHP1.1 brHP1.42 cmHP1.10 keHP1.15
Muchila	31	brHP1.8* BF5 A15v* brHP1.20 zaHP1.35 zaHP1.33 A11 ugHP1.20 brHP1.23 cmHP1.21 keHP1.3* gbHP1.3** brHP1.60 A12(A30) zaHP1.25 keHP1.13* brHP1.16 brHP1.22* keHP1.3 ugHP1.23* cmHP1.26 itHP1.11 zaHP1.14* keHP1.18 brHP1.8*** brHP1.22 zaHP1.10 zaHP1.9 brHP1.4* ugHP1.12 zaHP1.38
Mazabuka	2	zaHP1.2 zaHP1.20

Table 10: Analysis of MHCII sharing in the four study cohorts

Cohort(s)	# of unique haplotypes	Haplotypes names
Katantila Mazabuka Muchila Ndema	6	MHCII.HP.za.4 MHCII.HP.za.18 MHCII.HP.za.5 MHCII.HP.za.3 MHCII.HP.za.2 MHCII.HP.za.6
Katantila Muchila Ndema	6	MHCII.HP.za.12 MHCII.HP.za.42 MHCII.HP.za.27 MHCII.HP.za.26 MHCII.HP.za.23 MHCII.HP.za.16
Katantila Mazabuka Ndema	1	MHCII.HP.za.15
Katantila Mazabuka Muchila	5	MHCII.HP.za.39 MHCII.HP.za.8 MHCII.HP.za.1 MHCII.HP.za.7 MHCII.HP.za.14
Katantila Ndema	4	MHCII.HP.za.46 MHCII.HP.za.37 MHCII.HP.za.29 MHCII.HP.za.49
Muchila Ndema	12	MHCII.HP.za.25 MHCII.HP.za.64 MHCII.HP.za.65 MHCII.HP.za.55 MHCII.HP.za.57 MHCII.HP.za.28 MHCII.HP.za.24 MHCII.HP.za.67 MHCII.HP.za.21 MHCII.HP.za.56 MHCII.HP.za.66 MHCII.HP.za.60
Mazabuka Ndema	1	MHCII.HP.za.10
Katantila Muchila	12	MHCII.HP.za.22 MHCII.HP.za.30 MHCII.HP.za.41 MHCII.HP.za.19 MHCII.HP.za.43 MHCII.HP.za.13 MHCII.HP.za.40 MHCII.HP.za.17 MHCII.HP.za.33 MHCII.HP.za.44 MHCII.HP.za.31 MHCII.HP.za.62
Mazabuka Muchila	1	MHCII.HP.za.63
Ndema	5	MHCII.HP.za.53 MHCII.HP.za.20 MHCII.HP.za.58 MHCII.HP.za.54 MHCII.HP.za.59
Katantila	8	MHCII.HP.za.48 MHCII.HP.za.45 MHCII.HP.za.36 MHCII.HP.za.50 MHCII.HP.za.35 MHCII.HP.za.34 MHCII.HP.za.11 MHCII.HP.za.32
Muchila	4	MHCII.HP.za.61 MHCII.HP.za.52 MHCII.HP.za.51 MHCII.HP.za.47
Mazabuka	2	MHCII.HP.za.38 MHCII.HP.za.9



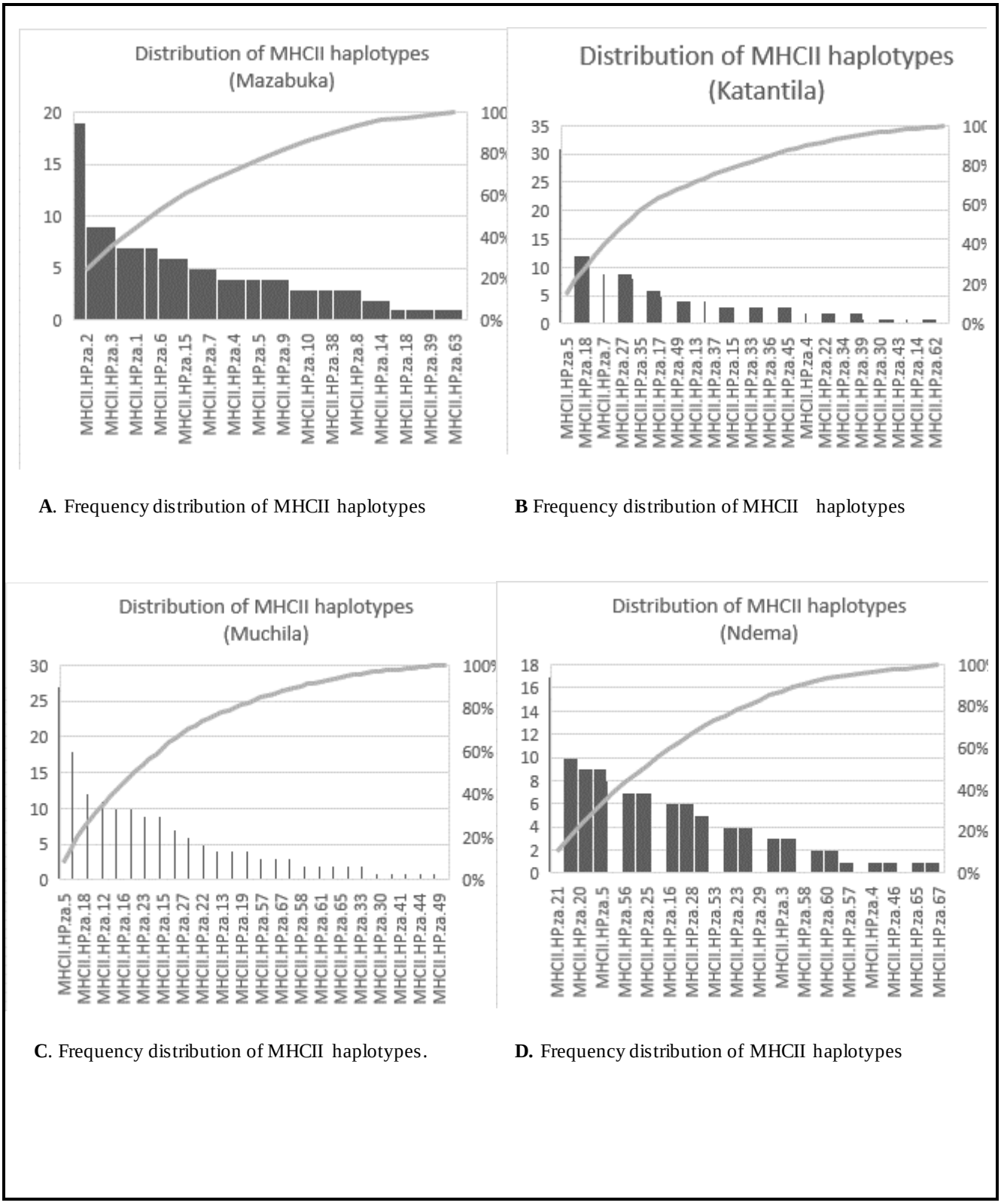
A. Frequency distribution of MHCI haplotypes

B. Frequency distribution of MHCI haplotypes

C. Frequency distribution of MHCI haplotypes

D. Frequency distribution of MHCI haplotypes

Figure 12. The cumulative and density figures of frequency distribution of MHCI haplotypes.



A. Frequency distribution of MHCII haplotypes

B Frequency distribution of MHCII haplotypes

C. Frequency distribution of MHCII haplotypes.

D. Frequency distribution of MHCII haplotypes

Figure 13. The cumulative and density figures of frequency distribution of MHCII haplotypes.

4.4 Phylogenetic analysis of bovine MHCI and MHCII genes

Phylogenetic analysis of the nucleotide sequences generated in this study and available from the IPD-MHC database shows that sequences from the IPD-MHC database (predominantly European taurine breeds), and the Zambian cohort (Tonga) are extensively intercalated, identifying a high level of phylogenetic relatedness (Figures 14 -17).

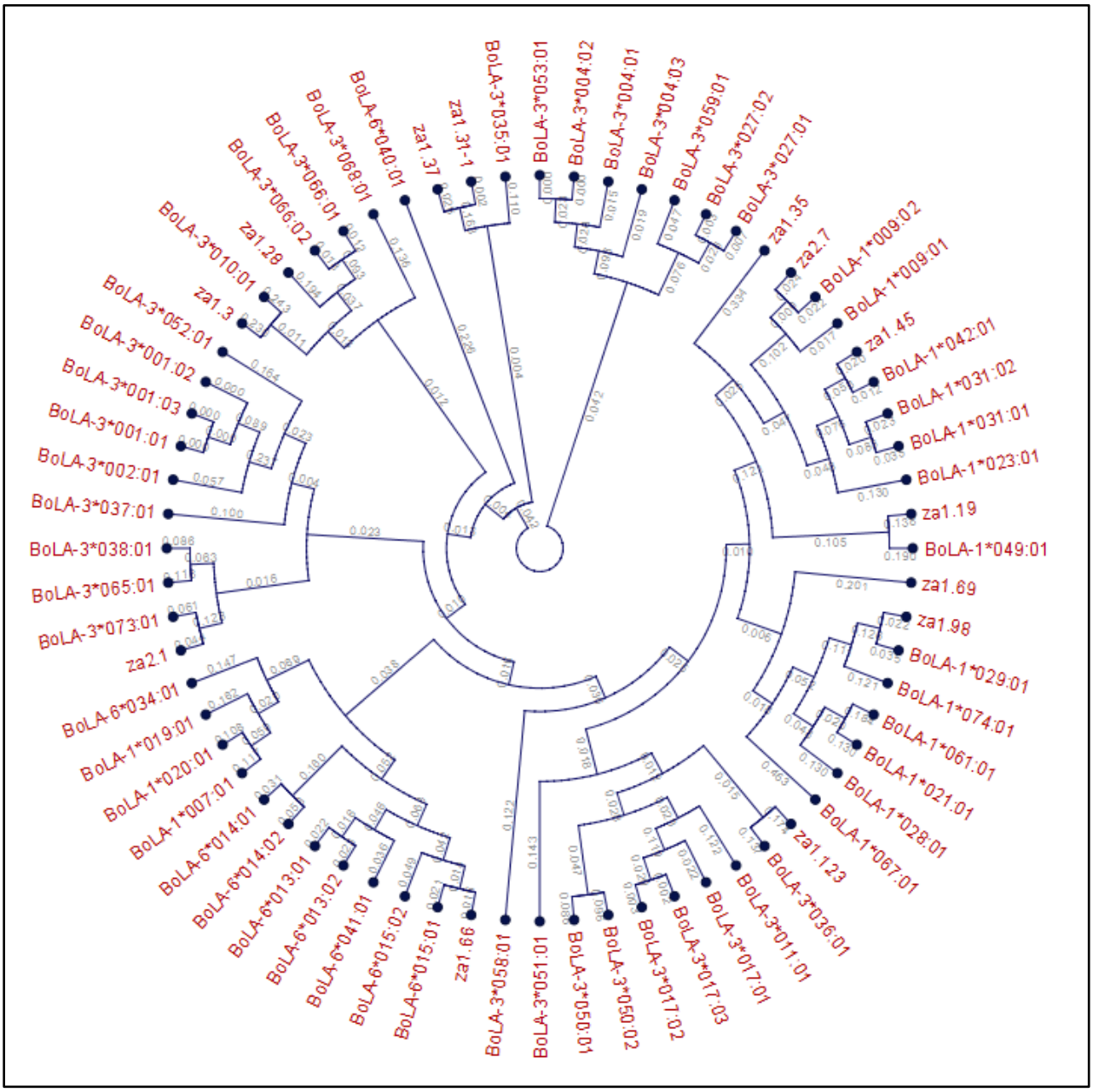


Figure 14. Phylogenetic analysis of bovine MHCI sequences. Sequence labels starting with za in the legends represent novel bovine MHCI alleles identified in this study.

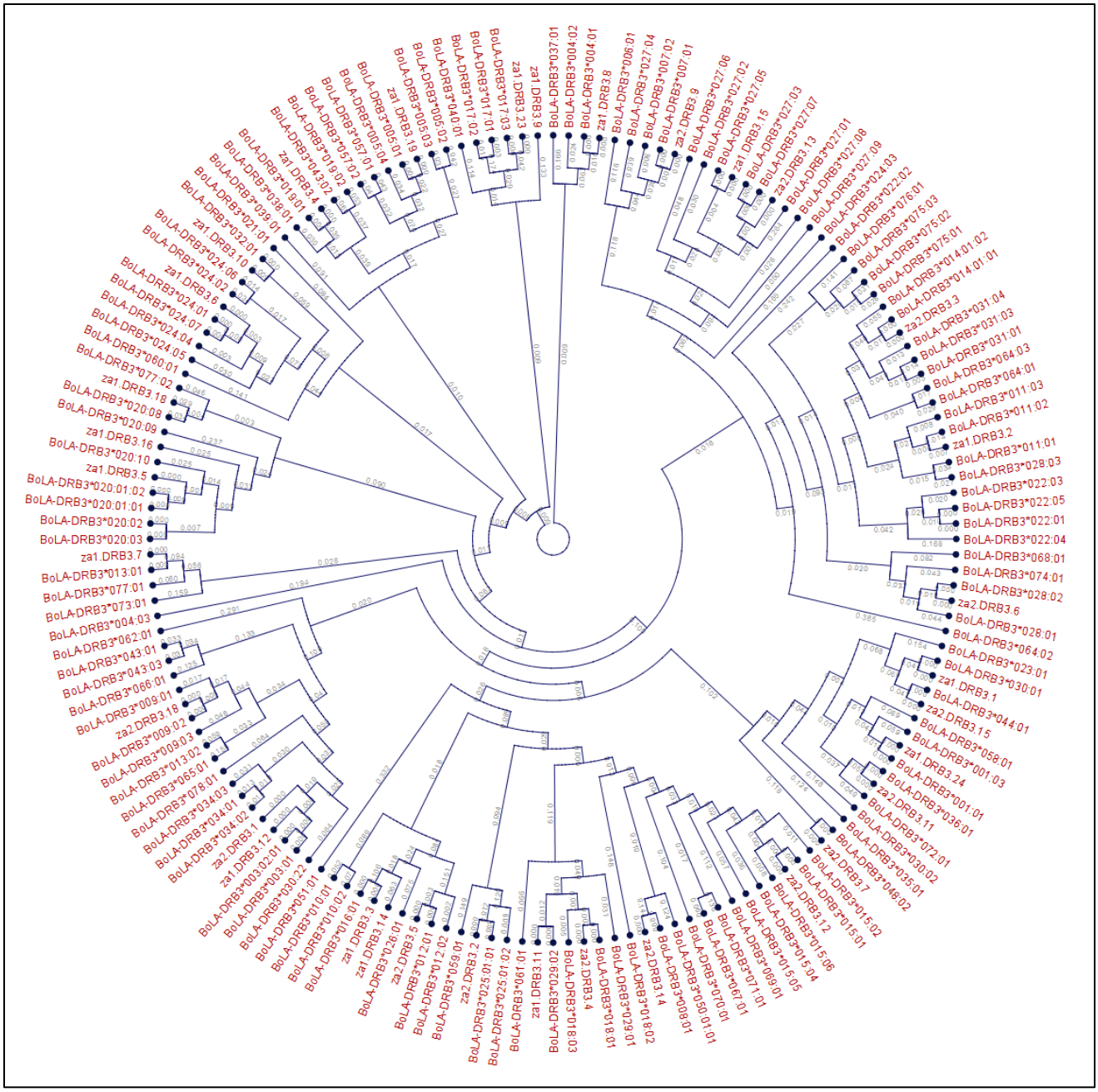


Figure 15. Phylogenetic analysis of bovine DRB3 sequences. Sequence labels starting with za in the legends represent novel bovine MHCII alleles identified in this study.

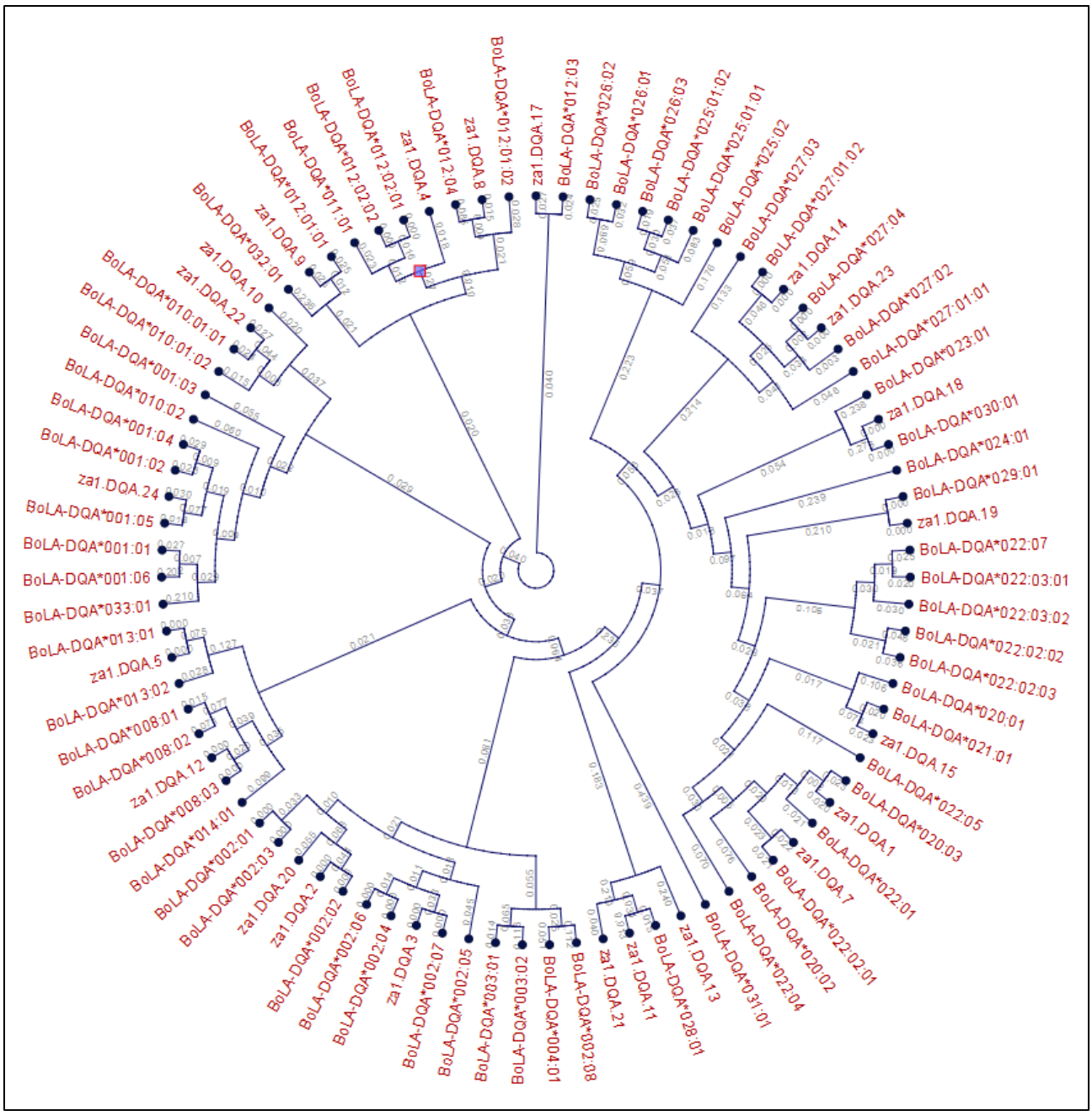


Figure 16. Phylogenetic analysis of bovine DQA sequences.

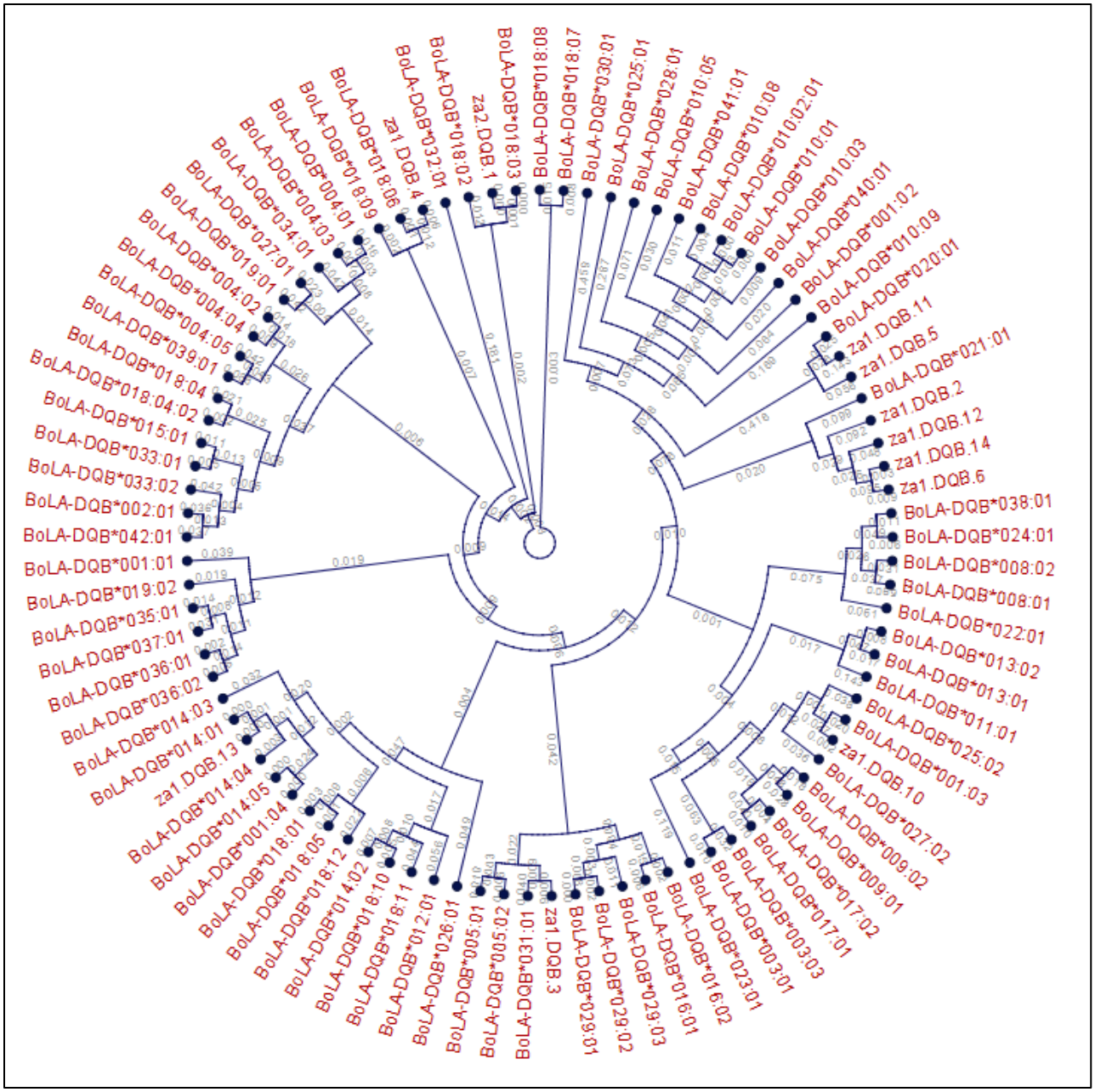


Figure 17. Phylogenetic analysis of DQB sequences.

CHAPTER FIVE

DISCUSSION AND CONCLUSION

5.1 Discussion

Analysis of the sequence data generated in this study identified a total of 137 MHCI haplotypes, of which 37 were novel. Together these novel haplotypes included 13 novel alleles and sequences matching 88 known alleles. Furthermore, 67 novel MHCII haplotypes were defined by co-occurrence of the same DRB3, DQA and DQB alleles. Together these MHCII haplotypes included 108 novel alleles and sequences matching 121 known alleles.

Previously, Vasoya *et al* (2016) examined sequences from a sample of 292 animals representing Holstein Friesian, Boran and Cameroonian cattle breeds using next generation sequencing (NGS) technologies. They reported over 140 novel MHCI alleles and defined 62 novel haplotypes and of these, six haplotypes [A11, A14, A15, A15v, A12 (w12B) and A20 (v2)] were identical to the haplotypes, reported in this study. Overall, more than 360 and nearly 230 unique MHCI and MHCII alleles, respectively were characterised in this study. With these large allele numbers found, in a sample of only (n = 347) cattle, there is possibly a very extensive diversity of MHCI and MHCII repertoires in Tonga cattle. Combining the data generated in this study, and that from Vasoya et al (2016), it appears that MHC diversity of African cattle is higher than that found in European cattle. This would be in agreement with Mikko and Andersson (1995) who compared in detail the diversity of MHC in both African and European cattle. This conjecture is in principal due to a larger effective population size and extensive breeding practices in African cattle. The effective population size in European cattle breeds is often drastically reduced by the extensive use of a limited number of sires (Mikko and Andersson., 1995). Most authorities would agree that high levels of diversity is advantageous (Parham and Ohta, 1996; Hughes and Yeager, 1998). This is because of the important immunological role played by the MHC, which is largely dependent on maintenance of high levels of polymorphism in these genes at the population level (Parham and Ohta, 1996). In humans for example, there are currently more than 12000 validated class I alleles encoded at three loci, HLA-A, -B, -C (http://www.ebi.ac.uk/imgt/hla). To elaborate more, it is well documented that MHC polymorphism is maintained by some form of balancing selection (Hedrick and Thomson 1983; Hughes and Nei 1988;Takahata and Nei 1990). The presence of

extensive MHC diversity in cattle (Schook and Lamont 1995) gives a clue as regards the selection forces promoting MHC diversity. The most favoured hypothesis is that selection at MHC loci is driven by the interaction with pathogens due to the central role of the MHC molecules in immune recognition of pathogens (Bodmer., 1972; Klein., 1986). An alternative hypothesis is that MHC mating preferences have evolved as a general mechanism to avoid inbreeding and that this is the major mechanism maintaining MHC diversity (Potts and Wakeland 1993). Inbreeding may cause MHC diversity to reduce, and have a detrimental impact on the ability of cattle to mount appropriate T cell responses that are needed. In this present time of climate instability, emerging pathogens and unpredictable movement of disease vectors, the importance rearing livestock that are disease resistant and able to respond appropriately to vaccines is paramount. This is especially more important in low and middle income countries (LMICS) including Zambia were livestock plays a major role for income generation and alleviating rural poverty.

Another aspect addressed was the sharing of MHCI and haplotypes sequences among the four cohorts of animals studied. More detailed analysis of haplotype sharing showed that only a limited number of MHCI haplotypes were observed in all four cohorts. Although there were a large number of haplotypes that were observed in multiple cohorts, nearly the same number of haplotypes were observed in only one cohort. This showed a surprising level of separation between the cohorts with nearly half of the MHCI haplotypes defined been segregated to individual herds. This lack of overlap may be explained by the undoubtedly incomplete sampling of the MHCI diversity within the cohorts.

A striking feature of data generated in this study was the presence of MHCI alleles 3*00402, 2*02501 and 4*02401 which go to form A14-like haplotypes common in Holstein Friesians. However, with available data, it is impossible to speculate as to whether such allele similarities are due genetic flow. Nonetheless, it is well known that Holstein Friesian cattle perform poorly in African tropical conditions and that this is partly due to poorresistance to disease. It would be of interest to determine the extent, if any, to which similarities at MHC loci contribute to variations in adaptation to the tropical setting and its flora of pathogens.

Phylogenetic and sequence analysis of the nucleotide sequences generated in this study and currently available from the IPD-MHC database showed that sequences from the IPD-MHC database (predominantly European taurine breeds), and the Zambian cohort (Tonga) are

extensively intercalated. The extensive phylogenetic intermingling of novel sequences from the Tonga cohorts with those in IPD-MHC database identify a high level of phylogenetic relatedness and may further suggests that the exon 2/3 sequences in all cattle are derived from a common set of ancestral MHCI and MHCII alleles.

It has been suggested that some BoLA class II alleles are strongly associated with certain pathogens (Takeshima and Aida, 2006). BoLA-DRB3*1001 is a high responder for IgG1 and IgG2 and appear to be strongly linked to high responses to the foot-and-mouth disease vaccine (Baxter et al., 2009). BoLA class II DRB3∗2703 and DRB3∗1501 alleles have been associated with levels of immunity against *Theileria parva* challenge following vaccination with sporozoite p67 antigen (Ballingal et al., 2004). BoLA-DRB3*0902 is firmly associated with resistance to bovine leukaemia infection (Xu et al., 1993). Interestingly, all these alleles were identified in this study. Identification of the distribution of BoLA class II genes that protect against pathogens may lead to breeding strategies that produce more resistant livestock in the future.

As Tonga cattle is resilient and well adapted to local climatic conditions, studies such as this one will provide data, which is essential for cattle conservation and disease control. Consequently, this will contribute to income generation and improved livelihood for the farmers who are involved in rearing Tonga cattle in Southern province of Zambia. . However, there is still uncertainty as to whether all cattle studied are purely Tonga due to lack of characterisation data.

5.2 Conclusion

This genetic data set represents one of the most comprehensive analyses of bovine MHCI and MHCII (DRB3, DQA and DQB) genes and is the first detailed record of the allelic and haplotype distribution in a *Bos taurus africanus* (sanga) cattle breed in Africa.

The results of this study have dramatically expanded the known bovine MHC repertoire in African cattle, for which there was limited available data. The findings derived from this study suggest a possible high diversity of MHCI and MHCII repertoires and a definite low haplotype sharing among the Tonga cattle herds of Zambia.

Furthermore, BoLA class II alleles that are strongly associated with resistance to certain pathogens are prevalent in Tonga cattle. Equally important, the currently available MHC sequences (predominantly from European taurine breeds) extensively intercalate with Tonga cattle demonstrating a high level of phylogenetic relatedness.

With this work, the Tonga cattle becomes the first African cattle breed, to have both MHCI and MHCII loci haplotyped. This combined with the large number of alleles identified, provides an excellent fundamental resource for use in bovine immunology and immunogenetics. The data generated by this study will provide a wealth of information for rationalising the development of improved and novel vaccines against infectious diseases of livestock and will be a major contributor to upgrading the current IPD-MHC database of known bovine MHC alleles.

5.3 Recommendations

Data on bovine MHCI and MHCII of Tonga cattle has now been documented. However, it cannot be stated with certainty that all MHC haplotypes and alleles have been observed due to incomplete sampling. Therefore, further studies generating larger datasets incorporating more breeds (including Angoni and Barotse), will be needed to more comprehensively evaluate the diversity of MHC genes of indigenous cattle breeds in the Zambia. Further studies can also consider first characterising the Tonga cattle to ensure purity of the breed other than reliance on phenotypic traits. Furthermore, since BoLA class II alleles that are strongly associated with resistance to certain disease pathogens are prevalent in Tonga cattle. It would be better to research further to what extent this is, and if possible, use this genetic data to breed animals that are resistant to such disease pathogens.

REFERENCES

AnderssonL., Lunden A., Sigurdardottir S., Christopher. J., Davies and Rask. L., (1988). Linkage relationships in the bovine MHC region. High recombination frequency between class II sub regions. Immunogenetics 27(4): 273-280.

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

Andersson L., Rask L., (1988). Characterization of the MHC class II region in cattle. Immunogenetics 27(2): 110-120.

Ajmone-Marsan, P., Garcia, J. F., Lenstra, J. A. (2010). On the origin of cattle: How aurochs became cattle and colonized the world. Evolutionary Anthropology: Issues, News, and Reviews, 19(4), 148–157.

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

Band, M., Larson J. H., Womack E.J., Lewin A., (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.

Becker, J. W. and G. N. Reeke, Jr. (1985). Three-dimensional structure of beta 2-microglobulin. Proc Natl Acad Sci U S A 82(12): 4225

Birch J, Murphy L, MacHugh ND, Ellis SA (2006) Generation and maintenance of diversity in the cattle MHC class I region. Immunogenetics 58:670–679

Ballingall, K. T., Ellis, S. A., MacHugh, N. D., Archibald, S. D., & McKeever, D. J. (2004). The DY genes of the cattle MHC: expression and comparative analysis of an unusual class II MHC gene pair. Immunogenetics 55(11): 748-755.

Bodmer, W. E (1972) The evolutionary significance of the HLA system. Nature 237: 139-145.

Behl, J. D., Verma, N. K., Tyagi, N., Mishra, P., Behl, R., & Joshi, B. K. (2012). The Major Histocompatibility Complex in Bovines: A Review. ISRN Veterinary Science, 2012, 1–12. doi:10.5402/2012/872710

Baxter, R., Craigmile, S.C., Haley, C., Douglas, A.J., Williams, J.L., Glass, E.J., (2009) BoLA-DR peptide binding pockets are fundamental for foot-and-mouth disease virus vaccine design in cattle. Vaccine 28, 28–37.

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

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 33:247–254

Ballingall, K. T., MacHugh, N. D., Taracha, E. L. N., Mertens, B., & McKeever, D. J. (2001). Transcription of the unique ruminant class II major histocompatibility complex-DYA and DIB genes in dendritic cells. European Journal of Immunology 31(1): 82-86.

Ballingall, K. T., MacHugh, N. D., Taracha, E. L. N., Mertens, B., & McKeever, D. J. (2011). Transcription of the unique ruminant class II major histocompatibility complex-DYA and DIB genes in dendritic cells. European Journal of Immunology 31(1): 82-86.

Braud, V. M., Allan, D. S. J., O'Callaghan, C. A., Söderström, K., D'Andrea, A., Ogg, G. S., McMichael, A. J. (1998) HLA-E binds to natural killer cell receptors CD94/NKG2A, B and C. Nature 391(6669): 795-799.

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

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

Brown, P., Spooner, R.L., Clark, A.J. (1989) Cloning and characterization of a BoLA class I cDNA clone. Immunogenetics, 29, 58.

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

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

Codner, G. F., Birch, J., Hammond, J. A., Ellis, S. A. (2012). Constraints on haplotype structure and variable gene frequencies suggest a functional hierarchy within cattle MHC class I. Immunogenetics, 64(6), 435–445.

Chicz, R.M., Urban, R.G., Gorga, J.C., Vignali, D.A.A., Lane, W.S., and Strominger, J.L. 1993. Specificity and promiscuity among naturally processed peptides bound to HLA-DR alleles. J. Exp. Med. 178:27-47.

Chardon, P., Renard, C., Vaiman, M. (1999). The major histocompatibility complex in swine. Immunological Reviews167 179.

Carrington M, Nelson GW, Martin MP, Kissner T, Vlahov D, Goedert JJ, Kaslow R, Buchbinder S, Hoots K, O'Brien SJ (1999). HLA and HIV-1: heterozygote advantage and B*35-Cw*04 disadvantage. Science 283(5408): 1748-1752.

Chicz, R.M., Urban, R.G., Lane, W.S., Gorga, J.C., Stern, L.J., Vignali, D.A., Strominger, J.L. (1992). Predominantly naturally processed peptides bound to HLA-DR1 are derived from MHC-related molecules and are heterogeneous in size. Nature 358:764-768.

CerebN, KimHR,Ryu J,YangSY (2015) Advances in DNAsequencing technologies for high resolution HLA typing. Hum Immunol 76: 923–927

Dudley DM, Karl JA, Creager HM, Bohn PS,Wiseman RW, O'Connor DH, (2014) Full-length novel MHC class I allele discovery by next- generation sequencing: two platforms are better than one. Immunogenetics 66:15–24

Decker JE, Pires JC, Conant GC, McKay SD, Heaton MP., (2009) Resolving the evolution of extant and extinct ruminants with high-throughput phylogenomics. Proc Natl Acad Sci U S A 106: 18644–18649.

Doherty, P. C., Zinkernagel. R.M., (1975). A biological role for the major histocompatibility antigens. Lancet 1(7922): 1406-1409.
Dagris (2007), domestic animal genetic resources information system (edited by S. kempo, Y. Mamo, B. Astrat and T. Dessie), International Livestock research institute, Addis Ababa, Ethiopia.

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

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

Davies CJ, Eldridge JA, Fisher PJ, Schlafer DH (2006) Evidence for expression of both classical and non-classical major histocompatibility class I genes in bovine trophoblast cells. Am J Reprod Immunol 55:188–200.

Decker JE, McKay SD, Rolf MM, Kim J, Molina Alcala A, Sonstegard TS, Hanotte O, Gotherstrom A, Seabury CM, Praharani L, Babar ME, Correia de Almeida Regitano L, Yildiz MA, Heaton MP, Liu WS, Lei CZ, Reecy JM, Saif-Ur-Rehman M, Schnabel RD, Taylor JF (2014) Worldwide patterns of ancestry, divergence, and admixture in domesticated cattle. PLoS Genet 10: e1004254

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

Dietz, A. B., Detilleux, J. C., Freeman, A. E., Kelley, D. H., Stabel, J. R., Kehrli, M. E. (1997), Genetic association of bovine lymphocyte antigen DRB3 alleles with immunological traits of Holstein cattle. Journal of Dairy Science 80(2): 400-405.

Ellis, S. (2004) 'The cattle major histocompatibility complex: Is it unique?', Veterinary Immunology and Immunopathology, 102(1–2), pp. 1–8

Ellis SA, Staines KA, Morrison WI., (1996) cDNA sequence of cattle MHC class I genes transcribed in serologically defined haplotypes A18 and A31. Immunogenetics 43:156–159

Ellis, S. A., Morrison, W. I., MacHugh, N. D., Birch, J., Burrells, A., Stear, M. J. (2005). Serological and molecular diversity in the cattle MHC class I region. Immunogenetics 57(8): 601-606.

Ellis, S. A., K. A. Staines, 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, 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

Ellis SA, Bontrop RE, Antczak DF, Ballingall KT, Davies CJ, Kaufman J, Kennedy L, Robinson J, Smith DJ, Stear MJ, Stet RJ, Waller M, Walter L, Marsh SGE (2006)

ISAG/IUIS-VIC Comparative MHC Nomenclature Committee report, 2005. Immunogenetics 57:953–958
Flutter, B., Gao B., (2004). MHC class I antigen presentation--recently trimmed and well presented. Cell Molecular Immunology (1): 22-30.

Hammond, J. A., Marsh, S. G. E., Robinson, J., Davies, C. J., Stear, M. J., Ellis, S. A. (2012). Cattle MHC nomenclature: is it possible to assign sequences to discrete class I genes? Immunogenetics, 64(6), 475–480.

Fries, R., R. Hediger, Stranzinger., G., (1986). Tentative chromosomal localization of the bovine major histocompatibility complex by in-situ hybridisation. Animal Genetics 17(4): 287-294.
Food and Agriculture Organization of the United Nations. (2000). World watch list for domestic animal diversity. 3rd Edition. FAO, Rome, Italy.

Food and Agriculture Organization of the United Nations. (2016). Economic analysis of animal Rome, Italy.

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

Groh. V., Bahram S, Bauer S, Herman A, Beauchamp M, Spies T (1996). Cell stress-regulated human major histocompatibility complex class I gene expressed in gastrointestinal epithelium. Proceedings of the National Academy of Sciences of the United States of America 93(22): 12445-12450.

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

Garboczi, D.N., Ghosh, P., Utz, U., Fan, Q.R., Biddison, W.E., and Wiley, D.C. (1996). Structure of the complex between human T-cell receptor, viral peptide and HLA-A2. Nature 384:134-141.

Gibbs RA, Taylor JF, Van Tassell CP, Barendse W, Eversole KA., (2009) Genome-wide survey of SNP variation uncovers the genetic structure of cattlebreeds. Science (80-) 324: 528–532.

Garcia, K.C., Degano, M., Stanfield, R.L., Brunmark, A., Jackson, M.R., Peterson, P.A., Teyton, L., and Wilson, I.A. (1996). An T cell receptor structure at 2.5Å and its orientation in the TCRMHC complex. Science 274:209-219.

Guethlein LA, Abi-Rached L, Hammond JA, Parham P (2015) The expanded cattle KIR genes are orthologous to the conserved single-copy KIR3DX1 gene of primates. Immunogenetics 59:517–522

Hammer, J., Gallazzi, F., Bono, E., Karr, R.W., Guenot, J., Valsasnini, P., Nagy, Z.A., and Sini- gaglia, F. (1995) Peptide binding specificity of HLA-DR4 molecules: Correlation with rheuma- toid arthritis association. J. Exp. Med. 181:1847- 1855.

Hemmer, B., Kondo, T., Gran, B., Pinilla, C., Cortese, I., Pascal, J., Tzou, A., McFarland, H.F., Houghton, R., and Martin, R. (2000). Minimal peptide length requirements for CD4+ T cell clones—implications for molecular mimicry and T cell survival. Int. Immunol. 12:375-383.

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

Holmes EC, Roberts AFC, Staines KA, Ellis SA (2003) Evolution of MHC class I genes in Cetartiodactyls. Immunogenetics 55:193–202

Helmer, D., Gourichon, L., Monchot, H., Peters, J. & Saña Segui M. (2005), Identifying early domestic cattle from Pre-Pottery Neolithic sites on the Middle Euphrates using sexual dimorphism. In: Vigne, J-D., Helmer, D., Peters, J. (Eds). First steps of animal domestication. New archaeozoological approaches. London: Oxbow Books: p.86-95.

Hughes, A. L. and Nei, M. (1988) Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. Nature 335: 167-170.

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

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

Hammer, J., Takacs, B., Sinigaglia, F. 1992 Identification of a motif for HLA-DR1 binding peptides using M13 display libraries. J. Exp. Med. 176:1007-1013.

Hammer, J., Valsasnini, P., Tolba, K., Bolin, D., Higelin, J., Takacs, B., and Sinigaglia, F. (1993) Promiscuous and allele-specific anchors in HLA-DR-binding peptides. Cell 74:197-203.

Hendel H, Caillat-Zucman S, Lebuanec H, Carrington M, O'Brien S, Andrieu JM, Schächter F, Zagury D, Rappaport J, Winkler C, Nelson GW, Zagury JF.New class I and II HLA alleles strongly associated with opposite patterns of progression to AIDS. Journal of Immunology 162(11): 6942-6946.

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

Heimbruch KE., Karl JA,Wiseman RW., DudleyDM, Johnson Z, Kaur A, O'Connor DH (2015) Novel MHC class I full-length allele and haplotype characterization in sooty mangabeys. Immunogenetics 67:437–445

Hammond J. A., Steven G. E., Christopher J., Michael J., Shirley A. Ellis., (2012) Cattle MHC nomenclature: is it possible to assign sequences to discrete class I genes? Immunogenetics (2012) 64:475–480

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

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

Immuno Polymorphism Database (IPD - http://www.ebi.ac.uk/ipd).
Jardetzky, T.S., Brown, J.H., Gorga, J.C., Stern, L.J., Urban, R.G., Chi, Y.I., Stauffacher, C., Strominger, J.L., Wiley, D.C. (1994) Three-dimensional structure of a human class II histocompatibility molecule complexed with superantigen. Nature 368:711-718.

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

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

Birch J., Codner. G., Guzman. E., Ellis SA., (2008). (2008). Genomic location and characterisation of MIC genes in cattle. Immunogenetics 60(8): 477-483.

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

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

Lewin H.A. (1996). - Genetic organization, polymorphism, and function of the bovine major histocompatibility complex. In The major histocompatibility complex region of domestic animal species (L.B. Schook & S.J. Lamont, eds). CRC Series in Comparative Immunology. CRC Press, Boca Raton, Florida, Chapter 4, 65-98

Linseele, V., Van Neer, W., Thys, S., Phillipps, R., Cappers, R., Wendrich, W., & Holdaway, S. (2014). New Archaeozoological Data from the Fayum "Neolithic" with a Critical Assessment of the Evidence for Early Stock Keeping in Egypt. PLoS ONE, 9(10), e108517.

Loftus, R. T., MacHugh, D. E., Bradley, D. G., Sharp, P. M., Cunningham, P. (1994). Evidence for two independent domestications of cattle. Proceedings of the National Academy of Sciences of the United States of America 91(7): 2757-2761.

Lange V, Bohme I, Hofmann J, Lang K, Sauter J, Schone B, Paul P, Albrecht V, Andreas JM, Baier DM, Nething J, Ehninger U, Schwarzelt C, Pingel J, Ehninger G, Schmidt AH (2014) Cost- efficient high-throughput HLA typing byMiSeq amplicon sequenc- ing. BMC Genomics 15:63

Maillard, J. B., Chantal D., Thevenon I., Sidi S., Stachurski I., (2003). Selection assisted by a BoLA-DR/DQ haplotype against susceptibility to bovine dermatophilosis. Genetics, selection, evolution : GSE. 35 Suppl 1. S193-200. 10.1051/gse:2003027.

Marsan A., Garcia P., Lenstra J. F., (2010). On the origin of cattle: how aurochs became domestic and colonized the world. Evolutionary Anthropology. 19. 148-157.

Mikko. S., Andersson. L., (1995) Extensive MHC class II DRB3 diversity in African and European cattle Immunogenetics 42:408-413 © Springer-Verlag.

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

Mahmood H.Z., Hussain, I., Iftikhar, S., Khan M., Nisa F., (2014), Role of Livestock in Food Security: An Ascertainment from Punjab Pakistan, International Journal of Academic Research in Business and Social Sciences Vol. 4, No. 8 ISSN: 2222-6990-458

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

Marshal. F., (1989), Rethinking the Role of Bos indicus in Sub-Saharan Africa, Current Anthropology. 30 (2) JSTOR 2743556.

Meyer. E.H.H., (1984), Chromosomal and biochemical makers of cattle breeds in Southern Africa, Proc 2nd world congress on beef and cattle breeding in RSA. 328 – 339.

Miltiadou D, Ballingall KT, Ellis SA, Russell GC, McKeever DJ (2005) Haplotype characterisation of transcribed ovine MHC class I genes. Immunogenetics 57:499–509

Nepom, G.G. and Erlich, H. (1991). MHC class-II molecules and autoimmunity. Annu. Rev. Immunol. 9:493-526.

Oomen R.A., Gillett RM, Kyle CJ (2013) Comparison of 454 pyrose- quencing methods for characterizing the major histocompatibility complex of nonmodel species and the advantages of ultra deep cov- erage.Mol Ecol Resour 13:103–116

Otting N, Heijmans CMC, Noort RC, de Groot NG, Doxiadis GGM, van Rood JJ, Watkins DI, Bontrop RE (2005) Unparalleled complexity of the MHC class I region in rhesus macaques. ProcNatl 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

Promerova M, Babik W, Bryja J, Albrecht T, Stuglik M, Radwan J., (2012) Evaluation of two approaches to genotyping major histocompatibil- ity complex class I in a passerine-CE-SSCP and 454 pyrosequenc- ing. Mol Ecol Resour 12:285–292

Parham P, Ohta T., (1996) Population biology of antigen presentation by MHC class I molecules. Science 272:67–74

Parker, K.C., Bednarek, M.A., Hull, L.K., Utz, U., Cunningham, B., Zweerink, H.J., Biddison, W.E., Coligan J.E. (1992). Sequence motifs important for peptide binding to the human MHC class I molecule, HLA-A2. J. Immunol. 149:3580-3587.

Parker, K.C., Bednarek, M.A., Coligan, J.E. (1994). A Scheme for ranking potential HLA-A2 binding peptides based on independent binding of individual peptide side-chains. J. Immunol. 152:163-175.

Parker, K.C., Shields, M., DiBrino, M., Brooks, A., Coligan, J.E. (1995). Peptide binding to MHC class I molecules: Implications for antigenic peptide prediction. Immunol. Res. 14:34-57.

Potts, W. K., Wakeland E.K., (1993). Evolution of MHC genetic diversity - a tale of incest, pestilence and sexual preference. Trends in Genetics 9(12): 408-412.

Robinson J, Waller MJ, Stoehr P, Marsh SGE (2005) IPD—the Immuno Polymorphism Database. Nucleic Acids Res 33:523–526

Rammensee, H.-G., Friede, T., Stevanović, S. (1995). MHC ligands and peptide motifs: first listing. Immunogenetics 41(4): 178-228.

Reddehase, M.J., Rothbard, J.B., Koszinowski, U.H. (1989). A pentapeptide as minimal antigenic determinant for MHC class I–restricted T lym- phocytes. Nature 337:651-653.

Rötzschke, O., Falk, K., Deres, K., Schild, H., Norda, M., Metzger, J., Jung, G., and Rammensee, H.-G. (1990). Isolation and analysis of naturally processed viral peptides as recognized by cytotoxic T cells. Nature 348:252-254.

Ruppert, J., Sidney, J., Celis, E., Kubo, R., Grey, H., and Sette, A. (1993). Prominent roles of secondary anchor residues in peptide binding to HLA-A2.1molecules. Cell 74:929-937.

Rueckert C, Guzmán CA (2012) Vaccines: From Empirical Development to Rational Design. PLoS Pathog 8(11): e1003001. https://doi.org/10.1371/journal.ppat.1003001

Rupp, R., Hernandez, A., Mallard, B. A. (2007). Association of bovine leukocyte antigen (BoLA) DRB3.2 with immune response, mastitis, and production and type traits in Canadian Holsteins. Journal of Dairy Science 90(2): 1029-1038.

Sharif S., Mallard B.A., Wilkie B.N., Sargeant J.M., Scott H.M., Dekkers J.C., Leslie K.E. (1998) Associations of the bovine major histocompatibility complex DRB3 (BoLA-DRB3)

alleles with occurrence of disease and milk somatic cell score in Canadian dairy cattle. Animal Genetics 29, 185–93.

Sher, A., Ahmed. R., (1995), Regulation of innate and adaptive responses to infection: new avenues and angles. Current Opinion in Immunology 7(4): 471-473.

Skow, L. C., Womack J.E, (1988). Synteny mapping of the genes for 21-steroid hydroxylase, alpha- a crystallin and class I bovine leukocyte antigen in cattle. DNA-A Journal of Molecular & Cellular Biology 7(3): 143-149.

Sharif, S., Mallard, B. A., Wilkie, B. N., Sargeant, J. M., Scott, H. M., Dekkers, J. C. M., Leslie, K. E. (1998) Associations of the bovine major histocompatibility complex DRB3 (BoLA-DRB3) alleles with occurrence of disease and milk somatic cell score in Canadian dairy cattle. Animal Genetics 29(3): 185-193.

Stone, R. T., Muggli-cockett NE., (1990). Partial nucleotide sequence of a novel bovine major histocompatibility complex class II beta chain gene, BoLA-DIB. Animal Genetics 21(4): 353-360.
Sanderson, F., Kleijmeer, M., Kelly, A., Verwoerd, D., Tulp, A., Neefjes, J. Trowsdale, J. (1994). Accumulation of HLA-DM, a regulator of antigen presentation, in MHC class II compartments. Science 266(5190): 1566-1569.

Sloan, V. S., Cameron, P., Porter, G., Gammon, M., Amaya, M., Mellins, E., Zaller, D. M. (1995) Mediation by HLA-DM of dissociation of peptides from HLA-DR. Nature 375(6534): 802-806.

Shiina, T., Tamiya, G., Oka, A., Takishima, N., Yamagata, T., Kikkawa, E., Inoko, H. (1999). Molecular dynamics of MHC genesis unraveled by sequence analysis of the 1,796,938-bp HLA class I region. Proceedings of the National Academy of Sciences, 96(23), 13282–13287.

Stock F, Gifford-Gonzalez D., (2013), Genetics and African cattle domestication. Afr Archaeol Rev. ;30:51–72.

Stear, M. J., Innocent, G. T., Buitkamp J., (2005) The evolution and maintenance of polymorphism in the major histocompatibility complex. Veterinry Immunology and Immunopathology 108(1-2): 53-57.

Southwood, S., Sidney, J., Kondo, A., del Guercio, M.F., Appella, E., Hoffman, S., Kubo, R.T., Chesnut, R.W., Grey, H.M., Sette, A. (1998). Several common HLA-DR types share largely overlapping peptide binding repertoires. J. Immunol. 160:3363-3373.

Schook, L. B. and Lamont, S. J. (1995), The Major Histocompatibility Complex Region of Domestic Animals, CRC Press, Boca Raton, In press

Sepil I,MoghadamHK,Huchard E, SheldonBC(2012) Characterization and 454 pyrosequencing ofmajor histocompatibility complex class I genes in the great tit reveal complexity in a passerine system.BMC Evol Biol 12:68

Sette, A. and Sidney J. (1998) HLA supertypes and supermotifs: A functional perspective on HLA polymorphism. Curr. Opin. Immunol. 10:478- 482.

Snell, G. D., Jackson RB., (1958). Histocompatibility genes of the mouse. II. Production and analysis of isogenic resistant lines. J Natl Cancer Inst 21(5): 843-877.

Takeshima, S., N. Saitou., (2003). The diversity of bovine MHC class II DRB3 genes in Japanese Black, Japanese Shorthorn, Jersey and Holstein cattle in Japan. Gene 316: 111-118.

Teutsch, M. R., BeeverJ.E, (1989). Linkage of complement factor B gene to the bovine major histocompatibility complex. Animal Genetics 19(4): 427.

The MHC Sequencing Consortium (1999). Complete sequence and gene map of a human major histocompatibility complex. Nature 401(6756): 921-923.

The Bovine Genome, Sequencing and. Analysis., Consortium, (2009). The Genome Sequence of Taurine Cattle: A Window to Ruminant Biology and Evolution. Science 324(5926): 522-528.

Trowsdale, J., Parham P., (2004). Mini-review: defense strategies and immunity-related genes. European Journal of Immunology 34(1): 7-17.

Takeshima, S.N., Aida, Y., (2006). Structure, function and disease susceptibility of the major histocompatibility complex. Anim. Sci. J. 77, 138–150

Tsuchida, T., Parker, K.C., Turner, R.V., McFarland, H.F., Coligan, J.E., and Biddison, W.E. (1994). Autoreactive CD8+ T-cell responses to human myelin protein-derived peptides. Proc. Natl. Acad. Sci. U.S.A. 91:10859-10863.

Takahata, N., Nei M., (1990). Allelic genealogy under overdominant and frequency-dependent sleection and polymorphism of major histocompatibility complex loci. Genetics 124(4): 967-978.

Townsend, A., Öhlén, C. (1989b). Association of class I major histocompatibility heavy and light chains induced by viral peptides. Nature 340(6233): 443-448

Vasoya D., Law A., Motta., P., Yu M., Muwonge A., Cook., E., Li X., Bryson K., MacCallam A., Sitt T., Toye P., Bronsvoort B., Watson M., Morrison WI., Connelley T., (2016), Rapid identification of bovine MHCI haplotypes in genetically divergent cattle populations using next-generation sequencing, Immunogenetics 68:765–781

Van der Poel, J. J., Groenen, M. A. M., Dijkhof, R. J. M., Ruyter, D., Giphart, M. J. (1990). The nucleotide sequence of the bovine MHC class II alpha genes- DRA, DQA and DYA. Immunogenetics 31(1): 29-36.

Vogt, A.B., Kropshofer, H., Kalbacher, H., Kalbus, M., Rammensee, H.-G., Coligan, J.E., Martin,R. (1994). Ligand motifs of HLA-DRB5*0101 and DRB1*1501 molecules delineated from self-peptides. J. Immunol. 153:1665-1673.

Vance, R. E., Kraft, J. R., Altman, J. D., Jensen, P. E., Raulet, D. H. (1998). Mouse CD94/NKG2A is a natural killer cell receptor for the nonclassical major histocompatibility complex (MHC) class I molecule Qa-1(b). J Exp Med 188(10): 1841-1848.

Vuure V, C.T. (2005). Retracing the Aurochs: History, Morphology and Ecology of an Extinct Wild Ox. Pensoft Publishers. Sofia-Moscow

Weigel, K. A., Freeman, A. E., Kehrli, M. E., Stear, M. J., Kelley, D. H. (1990) Association of class I bovine lymphocyte antigen complex alleles with health and production traits in dairy cattle. Journal of Dairy Science 73(9): 2538-2546.

World Cattle Inventory vs. Human Population (2017) http://farmcentric.com//
Wucherpfennig, K.W., Sette, A., Southwood, S., Oseroff, C., Matsui, M., Strominger, J.L., Hafler, D.A. (1994). Structural requirements for binding of an immunodominant myelin basic protein peptide to DR2 isotypes and for its recognition by human T cell clones. J. Exp. Med.179:279-290.

Warner, C. M., Brownell, M. S., Ewoldsen, M. A. (1988). Why aren't embryos immunologically rejected by their mothers? Biology of Reproduction 38(1): 17-29.

Xu, A., van Eijk, M. J. T., Park, C., Lewin, H. (1993) Polymorphism in BoLA-DRB3 exon 2 correlates with resistance to persistent lymphocytosis caused by bovine leukaemia virus. J ImmunoI 151: 6977-6985.

Zhang, C., A. Anderson, (1998). Structural principles that govern the peptide-binding motifs of class I MHC molecules. Journal of Molecular Biology 281(5): 929-947.

APPENDICES

APPENDEX I: RESEARCH ETHICAL APPROVAL

ERES CONVERGE

33 Joseph Mwilwa Road
Rhodes Park, Lusaka
Tel: +260 955 155 633
+260 955 155 634
Cell: +260 966 765 503
Email: eresconverge@yahoo.co.uk

I.R.B. No. 00005948
F.W.A. No. 00011697

13th March, 2018

Ref. No. 2018-Jan-009

The Principal Investigator
Dr. Isaac Silwamba
The University of Zambia
School of Veterinary Medicine
Dept. of Disease Control
P.O. Box 32379,
LUSAKA.

Dear Dr. Silwamba,

RE: MOLECULAR CHARACTERIZATION OF THE BOVINE MAJOR HISTOCOMPATIBILITY COMPLEX I AND II GENES OF BOS INDICUS (ZEBU) AND BOS TAURUS AFRICANUS (SANGA) CATTLE BREEDS OF ZAMBIA.

Reference is made to your corrections dated 28th February, 2018. The IRB resolved to approve this study and your participation as Principal Investigator for a period of one year.

Review Type	**Ordinary**	Approval No. **2018-Jan-009**
Approval and Expiry Date	Approval Date: 20th February, 2018	Expiry Date: 12th March, 2019
Protocol Version and Date	Version - Nil	12th March, 2019
Information Sheet, Consent Forms and Dates	• English.	12th March, 2019
Consent form ID and Date	Version - Nil	12th March, 2019
Recruitment Materials	Nil	12th March, 2019
Other Study Documents Questionnaires,	n/a	12th March, 2019
Number of participants approved for study	-	12th March, 2019

Specific conditions will apply to this approval. As Principal Investigator it is your responsibility to ensure that the contents of this letter are adhered to. If these are not adhered to, the approval may be suspended. Should the study be suspended, study sponsors and other regulatory authorities will be informed.

Conditions of Approval

- No participant may be involved in any study procedure prior to the study approval or after the expiration date.
- All unanticipated or Serious Adverse Events (SAEs) must be reported to the IRB within 5 days.
- All protocol modifications must be IRB approved prior to implementation unless they are intended to reduce risk (but must still be reported for approval). Modifications will include any change of investigator/s or site address.
- All protocol deviations must be reported to the IRB within 5 working days.
- All recruitment materials must be approved by the IRB prior to being used.
- Principal investigators are responsible for initiating Continuing Review proceedings. Documents must be received by the IRB at least 30 days before the expiry date. This is for the purpose of facilitating the review process. Any documents received less than 30 days before expiry will be labelled "late submissions" and will incur a penalty.
- Every 6 (six) months a progress report form supplied by ERES IRB must be filled in and submitted to us.
- ERES Converge IRB does not "stamp" approval letters, consent forms or study documents unless requested for in writing. This is because the approval letter clearly indicates the documents approved by the IRB as well as other elements and conditions of approval.
- Every 6 (six) months a progress report form supplied by ERES IRB must be filled in and submitted to us. Late submission of these will attract a penalty.

Should you have any questions regarding anything indicated in this letter, please do not hesitate to get in touch with us at the above indicated address.

On behalf of ERES Converge IRB, we would like to wish you all the success as you carry out your study.

Yours faithfully,
ERES CONVERGE IRB

Prof. E. Munalula-Nkandu
BSc (Hons), MSc, MA Bioethics, PgD R/Ethics, PhD
CHAIRPERSON

APPENDIX II: NAGOYA PROTOCOL CLEARANCE

PACRA

Patents and Companies Registration Agency

Head Office: PACRA House, Haile Selassie Avenue, Longacres, P.O. Box 32020, Lusaka, Zambia
Tel: (+260) 211 255 127, (+260) 211 255 151 Fax: (+260) 211 255 426
E-mail: pro@pacra.org.zm Website: www.pacra.org.zm

1st March, 2018

Dean – School of Veterinary Medicine,
University of Zambia,
P.O. Box 32379,
Lusaka.

Dear Sir,

RE: REQUEST FOR CLEARANCE ON THE NAGOYA PROTOCOL FOR THE RESEARCH PROJECT ENTITLED "MOLECULAR CHARACTERISATION OF THE BOVINE MAJOR HISTOCOMPATIBILITY COMPLEX I AND II (MHC I AND II) OF ZEBU AND SANGA CATTLE BREEDS OF ZAMBIA"

I write with respect to the above captioned matter and in response to your letter dated 6th February, 2018 addressed to the Permanent Secretary, Ministry of Lands and Natural Resources.

I note that the request came by way of a letter, undoubtedly because of the absence of Regulations enabling the implementation of the Protection of Traditional Knowledge, Genetic Resources and Expressions of Folklore Act No. 16 of 2016.

The absence of regulations notwithstanding, my advice on the way forward is that the use or access to traditional knowledge, genetic resources or expression of folklore for educational, research and experimental purposes is exempted from the provisions of the Act.

The research outcomes should be communicated to the Agency, Ministry of Lands and Natural resources and to the relevant communities where the samples will be collected.

I wish the student well in his research project.

Yours faithfully,

Anthony Bwembya
REGISTRAR & CEO

All correspondence should be addressed to the Registrar

APPENDIX III: INFORMATION AND CONSENT FORM FOR THIS STUDY

THE UNIVERSITY OF ZAMBIA

SCHOOL OF VETERINARY MEDICINE

INFORMATION & CONSENT FORM FOR BOVINE MAJOR HISTOCOMAPATIBILITY (MHC) RESEARCH

Project Title: **Molecular characterisation of the Bovine Major Histocompatibility Complex I and II genes of *Bos indicus* (Zebu) and *African Bos taurus* (Sanga) cattle of Zambia.**

SECTION A: INFORMATION SHEET

What is this study about?

This is a research project being conducted by Dr. Isaac Silwamba (MSc. Student), Dr. Martin Simuunza (Principal supervisor) and Prof John B. Muma (Co-supervisor) of the University of Zambia, School of Veterinary Medicine.

As the owner or duly authorized agent for the owner, you have been asked to have your cattle herd participate in a research study that will be evaluating diversity of the Major Histocompatibility Complex II and II of Zebu and Sanga cattle breeds of Zambia. This research study has been approved by the University of Zambia, School of Veterinary Medicine. Your informed consent is required prior to this use. The participation of your cattle herd is completely voluntary. Before agreeing to take part in the study, we request you to read this form, or we will go through it with you. This form describes the purpose of the study, the risks, benefit and your alternatives to participating in the study.

The major histocompatibility complex (MHC) is one of the most diverse genetic regions found in mammals. The genes contained in this region play a crucial role in immunological responses. The important immunological role played by the MHC is largely dependent on polymorphism (presence of genetic variation within a population). This polymorphism ensures that at the population level, MHC genes have the capacity to effectively present a diverse range of peptides from any potential pathogen. Therefore, an improved understanding of MHC diversity

is critical in understanding the nature of immune responses of to various disease pathogens, disease susceptibility or resistance and for application in related areas of applied research such as vaccinology.

Your district has been selected for sampling because of the large numbers of indigenous cattle.

What will I be asked to do if I agree to participate?

When you volunteer to participate in this study, you will be requested to bring your cattle herd to a central place for blood sampling. About 10ml of whole blood will be taken from selected animals within your herd for this study. The blood will be processed and stored frozen. No other samples will be required from your animal for this study. The processed sample from your animal will be used only for the purposes explained to you in this form. It will be kept for the entire period of the study and as long as it is still available. However, should our findings necessitate further investigations outside the present plan, we shall contact you once again to seek permission to use the sample or even undertake another sampling.

Would my participation in this study be kept confidential?

We will do our best to keep your personal information and other information pertaining to your animal and cattle herd confidential. To help protect your confidentiality, your name will not appear with the information pertaining to results of our tests. Only authorized research staff will have access to information bearing your name.

What are the risks of this research?

There is slight discomfort that your animals may feel when the blood sample is taken from them. But this is a normal routine process used in veterinary practice. The research team will do everything possible to minimize any possible discomfort or harm to your animal. Blood collection will be done by registered veterinary personnel.

What are the benefits of this research?

By allowing your cattle herd to participate in this study, you provide the opportunity to gain more understanding of the Major Histocompatibity complex (MHC) of cattle native to Zambia. This may provide avenues for improved disease control strategies or Vaccine production. Participation of your cattle herd will therefore go a long way towards reducing the suffering and death of your animal populations from various diseases.

Do I have to be in this research and may I stop participating at any time?

Your participation in this research is completely voluntary. You have the right to choose whether or not to participate in the study. This will not affect your rights as a cattle owner. You will remain entitled to the same veterinary services as other livestock owners.

.

Incentives/compensation

No incentives or compensation will be paid to you for participating in this study.

Sharing the data and the results

In this project, we are working together with researchers from Roslin institute in the UK. The processed samples obtained from your animals will be sent to researchers abroad, because they have got the machines needed to work on them. We will take the genetic material from your animal's blood and read it (almost like we read a book) and create information about MHC alleles and haplotypes that is specific to individual animals within your cattle herd. We will use this information together with the other information about your animal that will be collected. It is now common practice that genetic information is shared with researchers around the world. The benefit is that many researchers can use the same information for different research projects. We would also like to share the genetic information of your cattle with other researchers for other projects after we finish our study. If you participate in this study, you will also need to agree to share your cattle's genetic information for other research in the future.

We will publish the results of this study in journals and might present it in seminars.

What if I have questions?

This research is being conducted by Dr. Isaac Silwamba (MSc. Student) Dr. Martin Simuunza (Principal supervisor) and Prof John B. Muma (Co-supervisor) of the University of Zambia, School of Veterinary Medicine. Should you have any questions regarding this study and your rights as a research participant or if you wish to report any problems you have experienced related to the study, please contact:

1. Dr. Isaac Silwamba (MSc. Student)
   Department of Disease Control,
   School of Veterinary Medicine,
   University of Zambia,
   PO Box 32379,
   Lusaka.
   Mobile: 0977 628318

2. Dr. Martin Simuunza (Principal Supervisor)
   Department of Disease Control,
   School of Veterinary Medicine,
   University of Zambia,
   PO Box 32379,
   Lusaka.
   Tel: +260-211-293727

3. Prof John Bwalya Muma (Co-supervisor)
   Department of Disease control,
   School of Veterinary Medicine,
   University of Zambia,
   P.O BOX 32379,
   LUSAKA.
   Tel: +260-211-293727

SECTION B: CONSENT FORM

I have read and understood the information that has been presented to me both in vernacular and English languages. I have had all my questions answered to my satisfaction. I have been asked to participate in the above study by presenting my cattle herd for blood sampling and given free consent by signing this form. My consent to participate is voluntary and I may withdraw from the study at any time. I am further aware that the information I disseminate will be treated in confidence and I will not be personally identified. I also understand that some risk always exists when animal handling and animal procedures are performed.

(a) Signature cattle owner/duly authorized owner and date _____

(b) Signature of Researcher and date _____

(c) Signature of witness and date _____

APPENDEX IV: PRIMER SEQUENCES AND THEIR ANNEALING TEMPERATURES

A - MHCI - Miseq		Annealing temperature	Product size - no adaptors	Product size - adaptors (+132bp)
TCMHCfor1	GTY GGC TAY GTG GAC GAC	61	419	551
TCMHCrev 2	GGC CCT CSA SGT AGT TCC T			
TCMHCfor3	GGG CYV GAG TAT TGG GA	61	358	490
TCMHCrev 1	CTC CAG GTR TCT GSG GAG C			
B - MHCII - Miseq				
Bovine DRB3for	TAG TGA TGC TGA TGV TGC TG	61	364	496
Bovine DRB3 rev	GGY TGR GTC TTT GCA GGA TA			
Bovine DQAfor3	STG GRR GTG AAG ACA TYG TG	61	301	433
Bovine DQArev2	GAY TTG GRA AAC ACA GYC AC			
Bovine DQBfor1	GGR CYG AGG GCA GAG ACT	61	319	451
Bovine DQBrev2	GGR GAG ATG GTC ACT GTA GG			

(Vasoya *et al.*, 2016 supplementary data)

APPENDIX V: PRIMERS FOR UNIVERSAL BOVINE MHCI ALLELE AMPLIFICATION

Sequences of the sets of TCMHCfor1 and TCMHCfor3 primers with the Illumina D501-D508 adaptors and TCMHCrev1 and TCMHCrev2 primers with the Illumina D701-712 used for PCR amplification. The bovine MHCI-specific portion of the primers, the sequences of the MID tags and the Illumina adaptor/ligation sequences are shown in green, red and black script respectively.

Primer name	
TCMHCfor1-D501	AAT GAT ACG GCG ACC ACC GAG ATC TAC AC TATAGCCT ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT GTY GGC TAY GTG GAC GAC
TCMHCfor1-D502	AAT GAT ACG GCG ACC ACC GAG ATC TAC AC ATAGAGGC ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT GTY GGC TAY GTG GAC GAC
TCMHCfor1-D503	AAT GAT ACG GCG ACC ACC GAG ATC TAC AC CCTATCCT ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT GTY GGC TAY GTG GAC GAC
TCMHCfor1-D504	AAT GAT ACG GCG ACC ACC GAG ATC TAC AC GGCTCTGA ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT GTY GGC TAY GTG GAC GAC
TCMHCfor1-D505	AAT GAT ACG GCG ACC ACC GAG ATC TAC AC AGGCGAAG ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT GTY GGC TAY GTG GAC GAC
TCMHCfor1-D506	AAT GAT ACG GCG ACC ACC GAG ATC TAC AC TAATCTTA ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT GTY GGC TAY GTG GAC GAC
TCMHCfor1-D507	AAT GAT ACG GCG ACC ACC GAG ATC TAC AC CAGGACGT ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT GTY GGC TAY GTG GAC GAC
TCMHCfor1-D508	AAT GAT ACG GCG ACC ACC GAG ATC TAC AC GTACTGAC ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT GTY GGC TAY GTG GAC GAC
TCMHCfor3-D501	AAT GAT ACG GCG ACC ACC GAG ATC TAC AC TATAGCCT ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT GGG CYV GAG TAT TGG GA
TCMHCfor3-D502	AAT GAT ACG GCG ACC ACC GAG ATC TAC AC ATAGAGGC ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT GGG CYV GAG TAT TGG GA
TCMHCfor3-D503	AAT GAT ACG GCG ACC ACC GAG ATC TAC AC CCTATCCT ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT GGG CYV GAG TAT TGG GA
TCMHCfor3-D504	AAT GAT ACG GCG ACC ACC GAG ATC TAC AC GGCTCTGA ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT GGG CYV GAG TAT TGG GA
TCMHCfor3-D505	AAT GAT ACG GCG ACC ACC GAG ATC TAC AC AGGCGAAG ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT GGG CYV GAG TAT TGG GA
TCMHCfor3-D506	AAT GAT ACG GCG ACC ACC GAG ATC TAC AC TAATCTTA ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT GGG CYV GAG TAT TGG GA
TCMHCfor3-D507	AAT GAT ACG GCG ACC ACC GAG ATC TAC AC CAGGACGT ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT GGG CYV GAG TAT TGG GA
TCMHCfor3-D508	AAT GAT ACG GCG ACC ACC GAG ATC TAC AC GTACTGAC ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT GGG CYV GAG TAT TGG GA
TCMHCrev1-D701	CAA GCA GAA GAC GGC ATA CGA GAT CGAGTAAT GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T CTC CAG GTR TCT GSG GAG C
TCMHCrev1-D702	CAA GCA GAA GAC GGC ATA CGA GAT TCTCCGGA GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T CTC CAG GTR TCT GSG GAG C
TCMHCrev1-D703	CAA GCA GAA GAC GGC ATA CGA GAT AATGAGCG GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T CTC CAG GTR TCT GSG GAG C
TCMHCrev1-D704	CAA GCA GAA GAC GGC ATA CGA GAT GGAATCTC GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T CTC CAG GTR TCT GSG GAG C

```
TCMHCrev1-D705          CAA GCA GAA GAC GGC ATA CGA GAT TTCTGAAT GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T CTC CAG GTR TCT GSG GAG C
TCMHCrev1-D706          CAA GCA GAA GAC GGC ATA CGA GAT ACGAATTC GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T CTC CAG GTR TCT GSG GAG C
TCMHCrev1-D707          CAA GCA GAA GAC GGC ATA CGA GAT AGCTTCAG GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T CTC CAG GTR TCT GSG GAG C
TCMHCrev1-D708          CAA GCA GAA GAC GGC ATA CGA GAT GCGCATTA GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T CTC CAG GTR TCT GSG GAG C
TCMHCrev1-D709          CAA GCA GAA GAC GGC ATA CGA GAT CATAGCCG GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T CTC CAG GTR TCT GSG GAG C
TCMHCrev1-D710          CAA GCA GAA GAC GGC ATA CGA GAT TTCGCGGA GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T CTC CAG GTR TCT GSG GAG C
TCMHCrev1-D711          CAA GCA GAA GAC GGC ATA CGA GAT GCGCGAGA GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T CTC CAG GTR TCT GSG GAG C
TCMHCrev1-D712          CAA GCA GAA GAC GGC ATA CGA GAT CTATCGCT GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T CTC CAG GTR TCT GSG GAG C

TCMHCrev2-D701          CAA GCA GAA GAC GGC ATA CGA GAT CGAGTAAT GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T GGC CCT CSA SGT AGT TCC T
TCMHCrev2-D702          CAA GCA GAA GAC GGC ATA CGA GAT TCTCCGGA GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T GGC CCT CSA SGT AGT TCC T
TCMHCrev2-D703          CAA GCA GAA GAC GGC ATA CGA GAT AATGAGCG GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T GGC CCT CSA SGT AGT TCC T
TCMHCrev2-D704          CAA GCA GAA GAC GGC ATA CGA GAT GGAATCTC GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T GGC CCT CSA SGT AGT TCC T
TCMHCrev2-D705          CAA GCA GAA GAC GGC ATA CGA GAT TTCTGAAT GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T GGC CCT CSA SGT AGT TCC T
TCMHCrev2-D706          CAA GCA GAA GAC GGC ATA CGA GAT ACGAATTC GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T GGC CCT CSA SGT AGT TCC T
TCMHCrev2-D707          CAA GCA GAA GAC GGC ATA CGA GAT AGCTTCAG GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T GGC CCT CSA SGT AGT TCC T
TCMHCrev2-D708          CAA GCA GAA GAC GGC ATA CGA GAT GCGCATTA GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T GGC CCT CSA SGT AGT TCC T
TCMHCrev2-D709          CAA GCA GAA GAC GGC ATA CGA GAT CATAGCCG GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T GGC CCT CSA SGT AGT TCC T
TCMHCrev2-D710          CAA GCA GAA GAC GGC ATA CGA GAT TTCGCGGA GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T GGC CCT CSA SGT AGT TCC T
TCMHCrev2-D711          CAA GCA GAA GAC GGC ATA CGA GAT GCGCGAGA GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T GGC CCT CSA SGT AGT TCC T
TCMHCrev2-D712          CAA GCA GAA GAC GGC ATA CGA GAT CTATCGCT GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T GGC CCT CSA SGT AGT TCC T
```

(Vasoya *et al.*, 2016 supplementary data)

APPENDIX: VI: OUTPUT FROM BIOINFORMATICS ANALYSIS

The output of the analysis is given as an excel workbook containing 4 spreadsheets and a set of fasta files.

Spreadsheet 1 – 'Stats'

Provides a comprehensive statistical overview of the sequencing parameters and quantitative details of the MHC analysis pipeline for each sample. (E.g. MHCI below)

Column	Parameter	Notes
A	Well	The same MID tags are used for For1/Rev2 and For3/Rev1 primers and therefore PCR products bearing the same MID tags are clustered together in the same well.
B	Breed	Breed of animal – Tonga - X (no sample)
C	ID	Animal Identification
D	Total pairs	Number of paired reads sequenced for this well (combining For1/Rev2 and For3/Rev1)
E	Overlapping pairs	Number of paired reads sequenced for this well (combining For1/Rev2 and For3/Rev1) that overlap to form a contiguous sequence
F	For1Rev2	Number of reads in well that have the For1Rev2 primers
G	For3Rev1	Number of reads in well that have the For3Rev1 primers
I/Z	Total For1/Rev2 or For3/Rev1 reads	
J/AA	Total For1/Rev2 or For3/Rev1 variants	
K/AB	Threshold	Number of reads required for a variant to pass the 0.2% cut-off threshold
L/AC	Variants removed by cut-off	Number of variants that fail to reach threshold
M/AD	Reads removed by cut-off	Number of reads removed by application of threshold
N/AE	Singletons	Number of variants removed by threshold that were represented by a single read
O/AF	Variants above threshold	Number of variants above threshold

P/AG	Reads above threshold	Number of reads that form dataset to be entered into MHCI analysis pipeline
Q/AH	Chiamaeras	Number of chiamaeras identified
R/AI	Length >+/-9bp difference	Number of variants with length that is >+/-9bp different from that anticipated
S/AJ	1/2bp variant	Number of variants that are 1 or 2bp different from variants in the sample that are present at >30x or >50x frequency (variants potentially arising from PCR/sequencing errors)
T/AK	Known classical MHCI	Number of variants matching known classical MHCI alleles
U/AL	Known non-classical MHCI	Number of variants matching known non-classical MHCI alleles
V/AM	Unknown	Novel variants that have passed all parameters in pipeline and have the anticipated length– putative novel MHCI alleles
W/AN	Length <9bp difference	Novel variants that have passed all parameters in pipeline but have an unanticipated length - putative novel MHCI alleles.
X/AO	Corrected number of variants	Variants above threshold minus numbers removed as chimaeras, having lengths >+/-bp difference or 1/2bp variant.

Spreadsheet 2 and 3 – 'For1Rev2' and 'For3Rev1'

Provide a description of the data obtained for each individual sample for the For1/Rev2 and For3/Rev2 reactions respectively. In these spread sheets the following data for each sample is given:

Column	Parameter	Notes
A	Well	
B	Breed	Breed of animal – Tonga - X (no sample)
C	Identification	Animal Identification
D/E/F	Identified haplotypes	Designated when the combination of alleles that form a known haplotype are identified in a sample.
G/H/I	Alleles assigned to haplotypes	Lists each known MHCI allele in the identified haplotype and the percentage of the total overlapping reads for that sample which are represented by each of these alleles.*
J/K	Alleles unassigned to identified haplotypes	Lists each known MHCI allele that is not part of an identified haplotype in the sample and the percentage of the total overlapping reads for that sample which are represented by each of these alleles.
L/M/N/O	Unknown	Lists each variant of the anticipated length for the PCR which is not a known MHCI allele and the percentage of the total overlapping reads for that sample which are represented by each of these variants. In the For1Rev2 and For3Rev1 spreadsheet these variants are designated as 'Foo' and 'Bar' respectively, with consecutive numerical identifiers give to each variant.
P/Q	Unknown > +/- 9bp difference	Lists each variant differing from the anticipated PCR product length by <9bp which is not a known MHCI allele and the percentage of the total overlapping reads for that sample which are represented by each of these variants. In the For1Rev2 and For3Rev1 spreadsheet these variants are designated as 'FooSZ' and 'BarSZ' respectively, with consecutive numerical identifiers give to each variant.
R/S	Known non-classical	Lists each known non-classical MHCI allele and the percentage of the total overlapping reads for that sample which are represented by each of these alleles.

*when a sequence could match to multiple alleles each potential allele is shown unless the allele can be deduced from the co-expressed alleles.

Spreadsheet 4 – 'Foo Bar'

Provides 2 sets of information regarding the Foo and Bar sequences.

1 – A list of Foo and Bar sequences identified from the For1Rev2 and For3Rev1 data that match and are therefore considered to represent products of the same allele.

2 – A summary of BLAST analysis of each Foo and Bar sequence against the database of known bovine MHCI alleles. For each Foo/Bar sequence the known MHCI allele to which it showed the highest BLAST value is shown as is the % identity, length of sequence matched and the e-value.**

The current chimaera detection algorithm may not detect chiamaeras if one parent sequence is different from the canonical product size. Consequently any unknown variants identified ***in samples which contain sequences known to generate products of different lengths (e.g. 3*03301N) and at a lower read frequency than such sequences* have to be manually checked for chimaera. Such unknown variants are identified and highlighted in the accompanying sample excel spreadsheet.