Aus dem Institut für Parasitologie und Tropenveterinärmedizin
des Fachbereichs Veterinärmedizin
der Freien Universität Berlin

Sizing up the host and parasite genotype considerations relevant to the choice of candidate subunit vaccine antigens intended to render cattle immune to *Theileria parva*

Inaugural-Dissertation
zur Erlangung des Grades eines
PhD of Biomedical Sciences
an der Freien Universität Berlin

vorgelegt von
Isaiah Otieno Obara
Immunologist aus Nairobi, Kenia,
MSc. Immunology, B.Ed. Science (First-class honours)

Berlin 2017
Journal-Nr.:3945

Gedruckt mit Genehmigung des Fachbereichs Veterinärmedizin
der Freien Universität Berlin

Dekan:	Univ.-Prof. Dr. Jürgen Zentek
Erster Gutachter:	Prof. Dr. Peter-Henning Clausen
Zweiter Gutachter:	Prof. Dr. Jabbar Ahmed
Dritter Gutachter:	Prof. Dr. Richard Bishop

Deskriptoren (nach CAB-Thesaurus):
Cattle; Ankole; Holstein; *Theileria parva*; genotypes; vaccination; vaccines; antigens; East Coast Fever; Africa

Tag der Promotion: 04.05.2017

Bibliografische Information der *Deutschen Nationalbibliothek*
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.ddb.de> abrufbar.

ISBN: 978-3-86387-811-5
Zugl.: Berlin, Freie Univ., Diss., 2017
Dissertation, Freie Universität Berlin
D 188

Choriner Str. 85 - 10119 Berlin
verlag@menschundbuch.de – www.menschundbuch.de

Table of contents

List of abbreviations

AIC	Akaike Information Criterion
BEB	Bayes Empirical Bayes
BLAST	Basic Local Alignment Search Tool
BoLA	Bovine Leucocyte Antigen
bp	Base Pair
cDNA	Complementary Deoxyribonucleic Acid
CD4	Cluster of Differentiation 4
CD8	Cluster of Differentiation 8
CO_2	Carbon Dioxide
CTL	Cytotoxic T lymphocytes
dN/dS, ω	Ratio of Non-Synonymous to Synonymous Evolutionary Changes
DNA	Deoxyribonucleic Acid
ECF	East Coast fever
EcoRI	*Escherichia coli* RI
ELISA	Enzyme-Linked Immunosorbent Assay
ELISpot	Enzyme-Linked ImmunoSpot
FAO	Food and Agriculture Organization
FLEXBAR	Flexible Barcode and Adapter Removal
HLA	Human Leukocyte Antigen
IACUC	Institutional Animal Care and Use Committee
IFN-γ	Interferon gamma
IPD	Immuno-Polymorphism Database
ITM	Infection and Treatment
LB-Medium	Lysogeny Broth
LnL	Maximized Log-Likelihood
M1	Nearly Neutral Model
M2	Positive Selection Model
M7	Null B Model
M8	Positive Selection Plus B Model
mAb	Monoclonal antibodies
MAFFT	Multiple Alignment using Fast Fourier Transform
MHC	Major Histocompatibility Complex
MID	Multiplex Identifiers
MSP	Merozoite Surface Protein
NCBI	National Center for Biotechnology Information
NJ	Neighbor-Joining
PAML	Phylogenetic Analysis by Maximum Likelihood
PAUP	Phylogenetic Analysis Using Parsimony
PBMC	Peripheral Blood Mononuclear Cell
PCR	Polymerase Chain Reaction
PIM	Polymorphic Immunodominant Molecule
PP	Percent Positivity
PSCPL	Positional Scanning Combinatorial Peptide Library
RNA	Ribonucleic Acid
RPMI	Roswell Park Memorial Institute
SFU	Spot-Forming Units
SQL	Structured Query Language

TCR	T cell Receptor
TIM+I+G	Transition Model, Proportion of Invariable Sites and Gamma Distribution
TVM+I+G	Transversion Model, Proportion of Invariable Sites and Gamma Distribution
UV	Ultraviolet
VNTR	Variable Number Tandem Repeats

Preface

The disease caused by the apicomplexan protozoan parasite *Theileria parva* is typically lethal in cattle mostly owned by resource poor small holder farmers, particularly pastoralists, in large regions of eastern, central, and southern Africa. In order to develop an effective subunit vaccine against *T. parva,* it is important to bear in mind what such a candidate might be expected to achieve. Firstly, a subunit vaccine must confer protection in genetically diverse out-bred cattle populations. Secondly, the vaccine should render cattle immune to challenge by antigenically diverse parasite populations in the field. The latter requirement must take into account that a wildlife reservoir of infection exists in cape buffalo (*Syncerus caffer*) which is known to harbor *T. parva* parasites of much higher antigenic diversity than those maintained in cattle, and will frequently be present in co-grazed pastures, particularly in pastoralist areas. The principal aim of this thesis is to elicit information about host and parasite genotypes that contribute to immune responses to *T. parva* and would therefore have practical application in the identification of antigens for the development of subunit vaccines suitable for field use in ECF endemic areas.

The thesis is structured into chapters. The first chapter is intended to justify the search for vaccine-based solutions for control of the disease caused by *T. parva* by highlighting the extent of the ECF problem based on distribution of the disease and economic appraisals. Moreover, the chapter also highlights the parasite's development in the tick vector, and the mechanism of invasion and clonal expansion in the mammalian host target cells relevant to vaccination.

The second chapter addresses the limitations of measures presently available to control the disease caused by *T. parva*, as exemplified by increasing acaricide resistance, the practical difficulties of early diagnosis for use of chemotherapeutics, and the logistical constraints to production and delivery of the

live infection and treatment vaccination. In addition to presenting the case for subunit vaccination, the chapter reviews the information available on the efficacy of current subunit vaccine candidates under experimental and field conditions and identifies data gaps amenable to experimental investigation.

The third chapter leverages the progress in 'reverse vaccinology' to generate estimates of the relevance of the extent of sequence diversity present within the bovine host immune surveillance genes as a constraint to the development of effective cellular immunity in response to subunit vaccination against *T. parva* infection. Within this broad theme, the main areas of focus included: (i) the utilization of next-generation sequencing to identify novel allelic variants of class I MHC genes in distinct cattle breeds residing in the ECF endemic areas, (ii) *in silico* characterization of the predicted functionality of cattle class I MHC molecules and (iii) use of *ex vivo* techniques to infer the extent of functional divergence among the class I MHC alleles expressed by the African native (Ankole) and exotic (Holstein) *Bos taurus* cattle.

The fourth chapter is intended to provide a more detailed picture of the possible impacts of heterologous parasite challenge on the protection afforded to cattle immunized using a defined *T. parva* sporozoite surface antigen recognized by neutralizing antibodies. A key epidemiological consideration in this chapter is the fact that the disease in cattle is often perpetuated through spread from a wildlife reservoir. The primary analysis in this chapter involved examining allelic variation, specifically length polymorphisms and amino acid diversity in the closely juxtaposed B cell epitopes mapped to the sporozoite surface antigen targeted by neutralizing antibodies. In addition, the chapter also seeks to shed light on the mechanisms that underpin molecular evolution of the gene encoding the major sporozoite surface protein.

The final chapter attempts to synthesize data on the influence of polymorphisms and selective pressures on bovine adaptive immune response to *T. parva,* using the findings from the result sections described herein, supplemented with information from published research. The chapter reaches the conclusion that the constraints to recombinant vaccine development for ECF control imposed by cattle class I MHC diversity and the dynamic nature of parasite populations in the field may not be insurmountable and makes two key recommendations.

1.0 Chapter 1: Introduction

1.1 The extent of the East Coast Fever problem: geographical scale and economic impact

East Coast Fever (ECF), caused by the apicomplexan protozoan parasite *Theileria parva* (*T. parva*), is one of the most devastating animal health problems throughout a wide geographical region in eastern, central, and southern Africa ranging from Sudan to South Africa in the north and south respectively (Norval et al. 1992). At present, an estimated 28 million cattle in the region are at risk of the disease, with the annual number of cattle deaths approximated at 1 million. These deaths occur within three weeks of infection with the major cause of pathology until recently considered to be the dissemination of parasitized lymphoblasts into lymphoid and non-lymphoid tissues (Irvin and Morrison 1987). However, it has recently been discovered that tissue invasion by a subset of macrophages may also be important in inducing pulmonary edema in the lungs (Fry et al. 2016). Lymphoproliferation of parasitized cells is often exacerbated by non-specific cytolytic events provoked by infection. The three-host tick *Rhipicephalus appendiculatus,* the major vector of the parasite is widely distributed geographically (Norval et al. 1992, Minjauw and McLeod, 2003). Theileriosis can also be caused by tick transmission from the wildlife reservoir, the African buffalo (*Syncerus caffer*), that can infect cattle populations that share grazing pastures. The latter results in a different, but nonetheless highly pathogenic clinical syndrome characterized by low schizont parasitosis and low piroplasm parasitaemia, termed corridor disease (Neitz et al. 1955).

Most loss is incurred by pastoralists and other resource-poor farmers. An initial economic appraisal placed the annual burden of ECF at approximately US$ 189 million (Mukhebi, 1992). This estimate was derived from a number of measures including cattle mortality, reduced productivity, cost of acaricides for tick control, and in a broader context the fact that the disease limits the introduction into Africa of the more productive European breeds. This has recently been increased to US$ 300 million (McLeod and Kristjanson, 1999).

1.2 Development of *T. parva* parasites and the disease process

1.2.1 Haploid mammalian life cycle stages

The development of inflammatory reactions at the attachment site of feeding *Rhipicephalus appendiculatus* ticks coincides with the inoculation of tick-derived sporozoites into the mammalian host. Subsequent to their injection, the sporozoites rapidly invade bovine lymphocytes (Fawcett et al. 1982). As opposed to the more typical development of apicomplexan parasites within membrane-bound parasitophorous vacuoles, the contents of the rhoptries and microspheres of *T. parva* sporozoites have been implicated in the dissolution of the enveloping host endosomal cell membrane after invasion (Andrews et al. 1991). Within the cytosol of the leukocyte, further sporozoite development results into multinucleate schizonts, a stage which is associated with cancer-like transformation of the infected cell (Dobbelaere et al. 1999). The parasite's multiplication strategy at this stage involves associating with the host cell mitotic spindle apparatus during cell division ensuring that the infection is inherited by daughter cells (Hulliger et al. 1964).

This clonal expansion results in a distinctive pattern of pathology involving dissemination of parasitized cells via lymph and blood throughout the lymphoid system and to other tissues, notably the lungs and gut mucosa (Fry et al. 2016). Within the lymphoid tissues, the outcome is extensive depletion of lymphocytes attributable both to the lysis of infected and uninfected cells and invasion of bone marrow with proliferating lymphoblasts. Within the non-lymphoid systems, organ dysfunction in diseased animals provoked by infection is a well-recognized pathological feature. Further differentiation is possible within approximately two weeks of infection, during which a proportion of the schizonts differentiate into merozoites which are released following rupture of the infected lymphocytes (Shaw et al. 1995). The merozoites invade red blood cells to give rise to the tick infective intraerythrocytic piroplasm stage (Mehlhorn et al. 1984).

1.2.2 Transient diploid life cycle in the tick gut

Feeding ticks ingest piroplasm-infected RBCs which lyse in the tick gut liberating the piroplams and commencing gametogenesis (Mehlhorn et al, 1984). Morphologically distinct gametes fuse to form a diploid zygote which takes residence in the cells of the gut epithelium. The results of subsequent development are motile kinetes capable of migrating through the haemocoel to the salivary glands and invading the acinar cells (Walker et al, 1990). The life cycle is completed by the development of a syncytial sporont, from kinete precussors, which differentiate into a sporoblast (Fawcett et al, 1985). Stimuli received when a tick attaches to a mammalian host have been shown to be essential for the formation of multiple sporozoites from the sporoblast and their subsequent release into the acinar duct in preparation for inoculation. The life cycle stages of *T. parva* in both the mammalian host and the tick vector are shown in Figure 1.

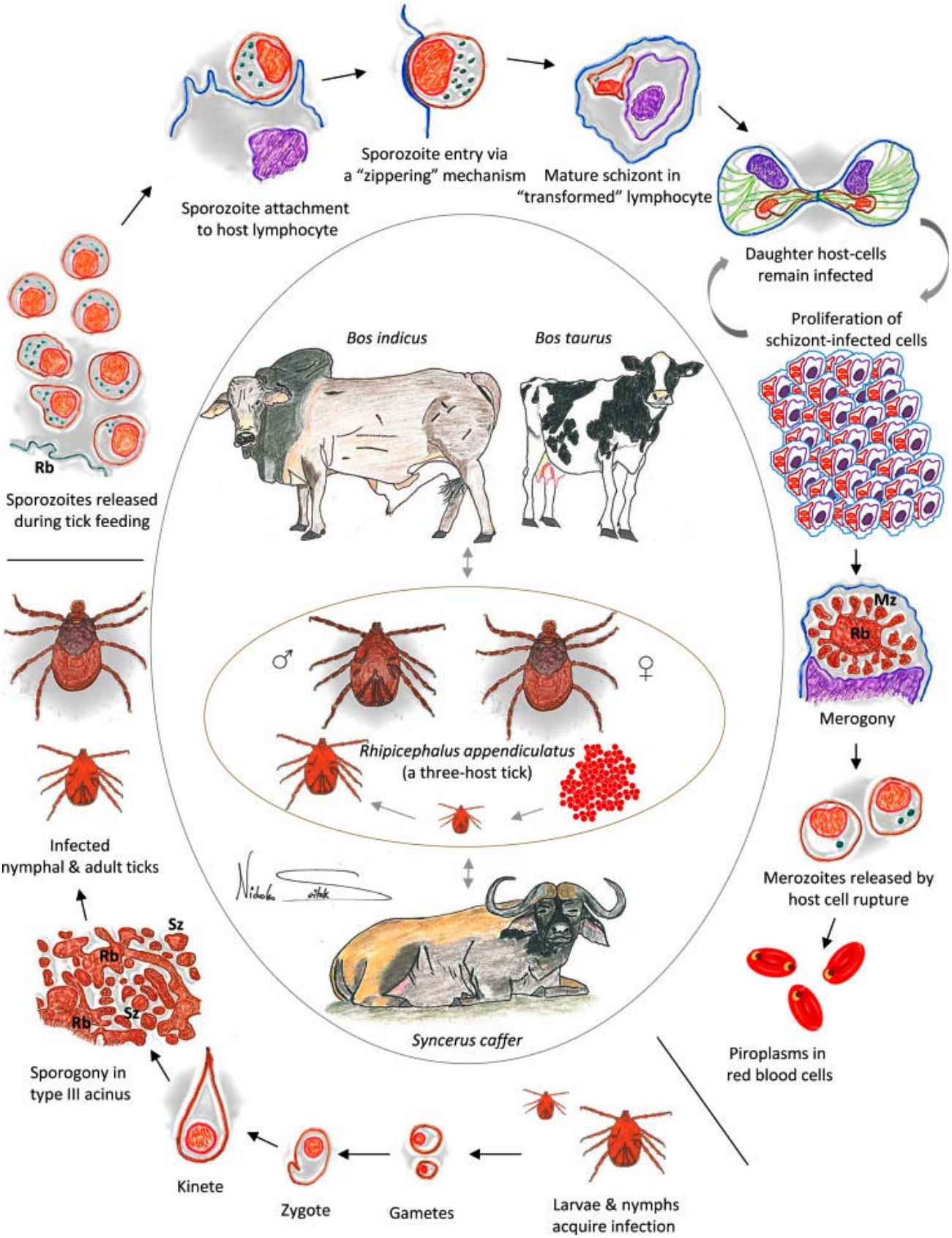


Fig 1 Obligate differentiation steps of *T. parva* in both the mammalian host and the tick vector Reproduced, with permission, from V. Nene et al. / Ticks and Tick-borne Diseases 7 (2016) 549–564

1.3 References

Andrews NW, Webster P (1991) Phagolysosomal escape by intracellular pathogens. Parasitol Today. 7(12):335-40

Dobbelaere D, Heussler V (1999) Transformation of leukocytes by *Theileria parva* and *T. annulata*. Annu Rev Microbiol. 53:1-42

Fawcett DW, Doxsey S, Stagg DA, Young AS (1982) The entry of sporozoites of *Theileria parva* into bovine lymphocytes in vitro. Electron microscopic observations. Eur J Cell Biol. 27(1):10-21

Fawcett DW, Young AS, Leitch BL (1985) Sporogony in *Theileria* (Apicomplexa: Piroplasmida). A comparative ultrastructural study. J Submicr Cyt. 17:299–314

Fry LM, Schneider DA, Frevert CW, Nelson DD, Morrison WI, Knowles DP (2016) East Coast Fever caused by *Theileria parva* is characterized by macrophage activation associated with vasculitis and respiratory failure. PLoS One. 19;11(5):e0156004

Hulliger L, Wilde JKH, Brown CGD, Turner L (1964) Mode of multiplication of *Theileria* in cultures of bovine lymphocytic cells. Nature. 203:728-30

Irvin AD, MorrisonWI (1987) Immunopathology, immunology and immunoprophylaxis of *Theileria* infections. In Immune Responses in Parasitic Infections: Immunology, Immunopathology, and Immunoprophylaxis, Volume III: Protozoa, ed. EJL Soulsby, pp. 223–74. Florida: CRC Press Inc

McLeod A, Kristjanson R (1999) Impact of ticks and associated diseases on cattle in Asia, Australia and Africa. ILRI and eSYS report to ACIAR. Internatiional Livestock Research Institute, Naurobi, Kenya

Mehlhorn H, Schein E (1984) The piroplasms: life cycle and sexual stages. Adv Parasitol. 23:37–103

Minjauw B, McLeod A (2003) Tick-borne diseases and poverty. The impact of ticks and tickborne diseases on the livelihood of small-scale and marginal livestock owners in India and eastern and southern Africa. Research report, DFID Animal Health Programme, Centre for Tropical Veterinary Medicine, University of Edinburgh, UK

Mukhebi AW (1992) Economic impact of theileriosis and its control in Africa. In The Epidemiology of Theileriosis in Africa Edited by: Norval RA, Perry BD, Young AS. London: Academic Press; 379-403

Neitz WO, Canham AS, Kluge EB (1955) Corridor disease: A fatal form of bovine theileriosis encountered in Zululand. Journal of the South African Veterinary Medical Association 26, 79–87

Norval RAI, Perry BD, Young AS (1992) The epidemiology of *Theileria* in Africa. Academic, London

Shaw MK, Tilney LG (1995) The entry of *Theileria parva* merozoites into bovine erythrocytes occurs by a process similar to sporozoite invasion of lymphocytes. Parasitology. 111:455–461

Walker AR (1990) Parasite adaptations in the transmission of *Theileria* by ticks. Trop Anim Health Prod. 22(1):23-33

2.0 Chapter 2: Literature review - Research milestones with impact on ECF control

2.1 ECF control: limitations of available control options

2.1.1 Acaricides and chemotherapy

Throughout the endemic areas, control of ECF has for the most part relied on the application of acaricides to limit infestation with the tick, *R. appendiculatus.* The key issue is the extent to which acaricides are sustainable. Acaricides cannot be continually applied with effectiveness in the long term due to practical limitations imposed by: (i) increasing incidence of acaricide resistance, with each new product released having a shorter effective commercial lifespan, (ii) environmental pollution concerns including the possibility of pesticide residues ending up in the food chain (George et al. 2008), and (iii) the need to apply acaricides regularly over prolonged periods to be effective and the lack of public sector support for the finance required for extended applications, as exemplified by the high numbers of non-operational dipping facilities, particularly in pastoral systems. As regards drug therapy, the cost of available theilericidal compounds makes them too expensive for the majority of resource poor small-holder farmers. Additional constraints to chemotherapy derive from the need to dispense drugs at the early stages of clinical disease leading to the requirement for early diagnosis, which is difficult using serology. This is due to the kinetics of the antibody response, which typically results in a scenario in which animals may die, before sero-diagnosis can be confirmed.

2.1.2 Immunization with live sporozoites and a long acting formulation of tetracycline

For the reasons outlined above, it is not surprising that priority has been given to pursuing vaccination as a more sustainable control tool. Efforts to date have focused on field use of a live vaccination regime termed infection and treatment (ITM). To confer broad protection against multiple *T. parva* stocks in the field, ITM is administered by subcutaneous injection of a composite of three live *T. parva* stocks as sporozoite stabilates combined with concurrent oxytetracycline prophylaxis (Radley et al. 1975a). Adoption of this vaccination strategy was made possible by research that permitted harvesting and

cryopreservation of *T. parva* sporozoites as homogenized whole infected ticks that are stored in liquid nitrogen (Patel et al. 2016).

Given the documented diverse genetic and antigenic composition of parasite and vector populations in the field, the efficacy of the ITM vaccine is a product of complex interaction between genetic diversity of the immunizing stocks and field tick challenge. At present, the standard formulation incorporates three selected parasite isolates (Serengeti- transformed, Kiambu 5, and Muguga), termed the 'Muguga cocktail', that was originally shown to confer protection against experimental challenge with a range of heterologous isolates (Radley et al. 1975b). This cocktail has been extensively tested in the field, mainly in Tanzania (Di Giulio et al. 2009). Additionally, experience from different regions within ECF endemic areas of Africa has demonstrated that it is effective, in some situations where the parasite populations may be antigenically more homogeneous, to use local stocks for ITM immunization (Geysen et al. 1999). This is exemplified by the use of the Katete and Boleni stocks in North East Zambia and Zimbabwe respectively.

ITM immunization has been widely deployed in pastoralist regions in Tanzania and this has stimulated demand in pastoralist regions in Southern Kenya. However there are constraints that may limit wider scale out of this control method (Morrison and McKeever, 2006). Challenges include the complexity of vaccine production requiring passage of sporozoite batches through feeding large numbers of nymphal ticks on infected cattle, a procedure that carries with it the risk of contamination with other bovine pathogens. This implies that quality control is problematic and often involves extensive cattle inoculation tests to evaluate efficacy as well as determining the optimum dose in titration studies (Patel et al. 2016). To induce effective immunity, the vaccine needs to remain viable during storage as a cryopreserved stabilate and cattle must be inoculated shortly after thawing. The cold chain requirement

for the storage and delivery of the vaccine means that liquid nitrogen must constantly be available, although linking into artificial insemination programs mitigates this constraint in some areas. The requirement for the simultaneous administration of oxytetracycline results in a high price of between 8–12 US dollars per calf for a dose of the vaccine in pastoralist systems where there are economies of scale.

Other concerns about ITM deployment include problems with consumer and regulatory acceptance stemming from the perception that inoculation with live parasites carries the potential risk of transmission of 'foreign' parasites into areas previously free of them. In practice, although this perception of transfer of parasites into field ticks and subsequently un-vaccinated cattle has been demonstrated to be correct (Oura et al. 2007), it does not pose a serious problem due to the diversity of *T. parva* in the field and occurrence of sexual recombination in the tick (Henson et al. 2012). The constraints discussed above have provided impetus for the development of subunit vaccines that are designed to mimic the protection afforded to cattle by live immunization.

2.2 The efficacy of subunit vaccine candidates under experimental and field conditions: data gaps and areas of uncertainty

In studies of bovine immune responses to the *T. parva*, extensive research has shown that experimental subunit vaccination can be achieved by targeting two life-cycle stages. The sporozoite, the first of two stages, is injected into cattle by infected ticks and is exposed to humoral immune responses directed against the surface antigens. The considerations for designing a subunit vaccine which targets the intracellular schizont are different from those targeting the sporozoite stage. The most striking of these differences stems from the fact that schizonts replicate within, and in synchrony with, the infected cells inducing cancer-like immortalization of bovine lymphocytes. It thus follows that cell-mediated, as

opposed to humoral immunity is essential for resolution of schizont parasitosis. Two experimental anti-pathogen subunit vaccines at different stages of development are available for ECF. One based on schizont antigens that are the target of $CD8^+$ T-cell responses in cattle immunized using infection and treatment (Graham et al. 2006). The second is based on the sporozoite stage (Musoke at al. 1992) that was consistently effective in inducing 70% protection in the laboratory, but achieved limited efficacy under field tick challenge (Musoke at al. 2005). The research milestones that have underpinned the identification and testing of the protective activity of these candidate subunit vaccines, as well as the necessary further work to achieve a vaccine with adequate efficacy and coverage in an outbred cattle population are outlined below.

2.2.1 Peptides derived from schizont antigens complexed with cattle class I MHC induce protective $CD8^+$ T-cell responses

The basis of natural immunity induced by delivery of sporozoites from *T. parva*-infected ticks in the field is less certain. However, information available from cattle that are immune as a result of ITM vaccination demonstrates the possibility of inducing long-term protective immunity artificially. A number of findings from research devoted to exploring the requirements for the induction of this protective immunity have reached the conclusion that class I MHC restricted cytotoxic T cells (CTLs) specific for schizont-infected lymphocytes are the main effectors (McKeever et al. 1994). The most compelling data in support of this conclusion was the demonstration that adoptive transfer of $CD8^+$ enriched T cells between *T. parva* immune and its naive identical twin calve can protect against challenge (McKeever et al. 1994). By contrast, transfer of schizont specific antibodies from immune to naïve cattle fails to confer protection from challenge. These experimental observations, coupled with strong correlational evidence of links between strain specificity of $CD8^+$ T-cell line with the outcome of cross-protection experiments (Taracha et al. 1995), indicate that antigens recognized by *T. parva*-specific $CD8^+$ T-cells represent promising candidates for the development of subunit vaccines.

Through analysis of the *T. parva* genome, it is now clear that the majority of the 4000 annotated protein-encoding genes in the *T. parva* genome are transcribed in the intralymphocytic schizont stage (Gardner et al, 2005, Bishop et al. 2005), although there has been little proteomics work to confirm this. The main methodological challenge to the design of subunit vaccine has been how to systematically identify candidate vaccine antigens within this large predicted schizont proteome. Studies to date have generally taken one of two approaches to the discovery of immunogenic targets. The more labor intensive of these approaches relies on the use of $CD8^+$ T-cells from *T. parva*-immune cattle to screen antigen-presenting cells co-transfected with class I MHC heavy chain cDNAs and parasite cDNA libraries prepared from purified schizonts (Graham et al. 2006). Candidates are initially identified on the basis of IFN-γ release as assessed using ELISpot, and subsequently validated using cytolytic assays based on measurement of chromium release from *T. parva* infected bovine lymphocytes that are recognized by $CD8^+$ T-cell lines. The parasite specific $CD8^+$ T-cells used in these studies recognized multiple distinct antigens distributed across the genome and these have been designated Tp1–Tp8 (Graham et al. 2006). Only six of these eight antigens were ultimately validated as candidates, but the nomenclature is based on the initial screens. A vaccination trial was subsequently performed to investigate the efficacies of a number of these antigens. These experiments involved immunization of 24 cattle and reported a degree of immunity to needle challenge with *T. parva* sporozoite stabilates in 30% of the vaccinates (Graham et al. 2006). Additionally there was a degree of correlation of the level of protection with induction of cytolytic responses, although most antigens were positive according to ELISpot (Graham et al. 2006). Subsequently, similar methodologies have been used to identify additional *T. parva* candidate antigens (MacHugh et al. 2009). Although the basis for the observed level of protection is not fully understood, studies of the immune responses induced following vaccination suggests that cattle of different class I MHC genotypes tend to recognize different *T. parva* antigens (Graham et al, 2008).

The second approach follows directly from the findings of the first and is intended to expedite vaccine candidate identification by *in silico* characterization of the functionality of cattle class I MHC molecules (reviewed by Nene et al. 2012). Termed 'reverse vaccinology', the validity of this approach will be demonstrated by concordance between $CD8^+$ T-cell epitopes predicted by machine learning algorithms and those identified by screening expressed parasite cDNA with specific $CD8^+$ T-cell lines (Hansen et al. 2014, Nene et al. 2012). Antigens identified using such predictions can be further assessed for their suitability for inclusion in subunit vaccines based on recognition by cytokine-secreting or tetramer-positive T cell populations (Nene et al. 2012, Svitek et al. 2014, Hansen et al. 2014). The ultimate test will then be to evaluate candidate antigens *in vivo* in cattle using defined delivery systems, such as adenovirus priming and modified vaccinia virus (MVA) boosting.

Several concerns relating to the schizont antigen vaccine development approach are presented by the complexity of the expressed cattle class I MHC loci and require attention to ensure that any predicted candidate vaccine antigens offer sufficient coverage and efficacy (Ellis, 2004). For example, in contrast to humans that have a fixed complement of three classical loci, six loci are available for transcription in cattle, MHC gene content varies between haplotypes and most importantly currently available data, albeit limited, strongly suggests that different cattle breeds often carry distinct alleles (Robinson et al. 2013). Based on the information documented in the FAO cattle database http://dad.fao.org/, 211 cattle 'breeds' are resident in the ECF endemic regions in eastern, central and southern Africa. Whereas there has been extensive work carried out in European *Bos taurus* cattle (Holstein/Friesian) to investigate the extent of class I MHC allelic polymorphisms, the expressed class I MHC sequences from African taurine 'breeds' have not yet been analyzed in detail. In the absence of such data, it is impossible to make wide predictions using reverse vaccinology approaches regarding candidate epitopes/antigens that may have widespread efficacy in outbred African cattle populations. Consequently, at present there

are no credible estimates of the number of antigens/epitopes that may need to be incorporated in a recombinant antigen cocktail vaccine that will be effective in the control of ECF in African native, as well as introduced cattle populations in the field.

2.2.2 An experimental recombinant anti-sporozoite vaccine induces antibody based protection

Analysis of serological responses in cattle following experimental as well as natural *T. parva* challenge demonstrated that antibodies directed against sporozoite surface epitopes are capable of neutralizing infectivity of sporozoites for bovine lymphocytes and can be protective (Dobbelaere et al. 1984). Studies using a panel of neutralizing monoclonal antibodies (mAb) have demonstrated that a 67 kilodalton (p67) antigen expressed by *T. parva* as a major component of the sporozoite surface, but not demonstrated to be expressed as a protein in either schizont or piroplasm stages, represents a major target of neutralizing antibodies. It has been demonstrated that vaccination with *E. coli* expressed recombinant p67 invariably protects approximately 70% of immunized cattle from the development of severe disease following needle challenge with *T. parva* sporozoite stabilates, regardless of the recombinant expression system used. Following this successful experimental proof of principle, a field trial was carried out to monitor the effects of vaccination with p67 on situations where cattle were exposed to infective ticks. The results confirmed that immunization with p67 provides some protection to cattle against *T. parva* infection in the field although the efficacy was low at approximately 45 % (Musoke et al. 1992, Bishop et al 2003, Musoke et al. 2005). The reasons for the lower efficacy in field situations are not yet clearly understood, but any heterogeneity in the field isolates of *T. parva* could make such an outcome particularly likely. In addition, a tick challenge is different to a stabilate challenge in the route of sporozoite entry and association of a complex cocktail of immunomodulatory and anti-inflammatory tick-derived molecules (reviewed by Nuttall and Labuda 2004).

One scenario that could have a profound effect on p67 vaccine efficacy in cattle is if *T. parva* challenge is from buffalo-associated ticks. This is because in comparative genomics studies, it has become apparent that the gene encoding p67 is conserved in all cattle-derived populations of the parasite, whilst a limited degree of polymorphism is present in buffalo-derived isolates (Nene et al. 1996; Nene et al. 1999; Musoke et al. 2005). Consequently, one concern that deserves attention to permit deployment of the p67 vaccine in areas where buffalo and cattle co-graze with any expectation of effectiveness is whether *T. parva* parasites that originate from buffalo, and capable of causing severe disease in co-grazing cattle, contain diverse or invariant p67 genotypes. Importantly, it would be useful to obtain information on the extent of B cell epitope polymorphism of the gene encoding p67 among field isolates originating from cattle exposed to buffalo associated ticks. In additional to documenting *T. parva* p67 polymorphisms, there are clearly further lines of investigation, relating to the mechanisms that generate and maintain the diversity, that need to be pursued to fully explore the potential of this candidate vaccine in situations where the wildlife reservoir of infection persists in the environment.

2.3 Study objectives

The principal aim of the research described in this thesis is to elicit information about host and parasite genotypes that determine immune responses to *T. parva* induced in infected cattle. Increased knowledge of these factors would have practical applications in terms of the choice of antigens for the development of subunit vaccines suitable for filed use throughout the ECF endemic areas. There is little doubt that class I MHC restricted CTLs specific for schizont-infected lymphocytes are one important category of effectors induced by ITM live immunization for control of *T. parva*. However, it still remains unclear whether the allelic diversity at the expressed classical class I MHC loci is consistent between cattle breeds or whether the repertoire is broader than indicated by current databases and publications. Consequently there is insufficient information to predict the likely influence of MHC genotypes on the antigenic specificity of $CD8^+$ T-cell responses to *T. parva* particularly in African cattle. This paucity of host genotype data defines the overarching goal of the first part of this thesis whose specific objectives included:

1) Generating the first detailed catalogue of expressed class I MHC sequences from an African taurine 'breed' resident in the ECF endemic areas. It is important to note that in addition to being outbred, breeds in the strict sense exemplified by Holstein/Friesian are rare in Africa because most cattle represent an admixture of alleles derived from taurine and indicine lineages (Machugh et al. 1997). This objective encompassed the following:

   a) Generating amplicons corresponding to the peptide binding regions of the expressed class I MHC sequences from an African taurine (Ankole cattle) using primers based on the most conserved sequences present in all publicly available cattle class I sequence databases which are additionally conserved between cattle and the relatively distantly related European Bison (Babik et al. 2012).

b) High throughput, deep coverage bidirectional sequencing of barcoded, clonally amplified and multiplexed class I MHC peptide binding region-derived amplicons on the Roche 454 platform.

c) Discriminating between true variants and artifactual 454 pyrosequences and discarding reads categorized as mis-called variants by successive filtering algorithms. This is particularly relevant given that errors inherent to amplification and ultra-deep pyrosequencing mean that the potential for some reads to be artifactual is high.

2) Molecular characterization of allelic variants at the class I MHC locus of an exotic European domesticated *Bos taurus* (Holstein) present in sub-saharan Africa by amplification of the peptide binding region, plasmid sub-cloning of amplicons and chain termination sequencing.

3) Establishing the extent of the functional divergence between class I MHC 'haplotypes' expressed in the African native (Ankole) and exotic (Holstein) *Bos taurus* resident in sub-Saharan Africa in the context of assessing the influence of MHC genotypes on the antigenic specificity of $CD8^+$ T-cell responses to *T. parva.* To achieve this, the following sub-objectives can be identified:

a) Exploiting the large amount of currently available peptide-MHC binding data in machine learning computational analysis that incorporates the complexities of peptide-MHC interaction to reliably predict the peptide-binding repertoires of class I MHC alleles from the two 'breeds'.

b) Quantifying the predicted binding specificity distance between the expressed class I MHC sequences from the African native and introduced *Bos taurus* cattle and incorporating the distance scores into a phylogenic clustering algorithm to investigate the extent to which the repertoire of the predicted class I MHC peptide-binding motifs may overlap.

c) Serodiagnosis of *T. parva* in a field population of African taurines (Ankole) and application of an IFNγ release assay to evaluate the ability of PBMCs and purified $CD8^+$ T-cells from ELISA seropositive Ankole cattle to recognize currently identified candidate *T. parva* $CD8^+$ T-cell target antigens. It is worthy of note that the candidate antigens were identified by using immune bovine $CD8^+$ T-cell lines from both Holstein-Friesian, European *Bos taurus* and Boran *B. indicus* cattle to screen expressed *T. parva* cDNA libraries.

4) Codon-by-codon analysis of breed differentiated class I MHC allelic sequences to clarify what processes may underpin selection primarily among amino acid positions that determine specificity for antigenic peptide binding.

A second concern associated with *T. parva* subunit vaccine development that deserves attention relates to the fact that stocks of the parasite that are transmissible between cattle by ticks are not homogeneous at many loci (Henson et al. 2012). This is exacerbated by the fact that a wildlife reservoir of infection in African Cape buffalo (*Syncerus caffer*) is known to harbour parasites of much higher antigenic diversity than those maintained in cattle often co-grazes with cattle. The stock heterogeneity is exemplified by the gene encoding p67 which is conserved in all cattle-derived populations of the parasite, although a limited degree of polymorphism is present in buffalo-derived isolates. It remains unclear whether *T. parva* parasites originating from buffalo, with diverse p67 genotypes, can infect and cause severe disease in co-grazing cattle, in which case, immunisation with cattle derived recombinant p67 may result in susceptibility to challenge with antigenically dissimilar (heterologous) isolates. To provide clarity on these parasite genotype issues, the following objectives were set:

1) Characterization of heterogeneity in the p67 gene among field isolates of *T. parva* originating from cattle that were part of a field trial of ITM vaccines designed to evaluate protection afforded to immunized animals that received *T. parva* challenge from buffalo-associated ticks. The objectives primarily involved the following:

   a) Examining allelic variation, principally length polymorphism due to insertions and deletions (indels), in the central region of the p67 gene. This is particularly relevant since available data suggest that the gene encoding p67 is invariant among *T. parva* parasites that are sustainably transmitted between cattle by ticks and has a unique deletion not present in the vast majority of genotypes present in buffalo.

   b) Assessing the nature and extent of variation in the closely juxtaposed B cell epitopes that have been mapped to the central region of the p67 protein among the field isolates.

2) Shedding light on the mechanisms that underpin molecular evolution of the gene encoding p67, the sporozoite surface protein that is the major target of neutralizing antibody activity and has induced protection as a recombinant antigen. The in depth sub-objectives that are required to deliver this objective are outlined below:

   a) A preliminary reconstruction of the phylogenetic relationships between the p67 sequences using a neighbor-joining (NJ) tree under the Jukes and Cantor model.

   b) Selection of best-fit nucleotide substitution models for *T. parva* p67 evolution by calculating the likelihood scores of different models with varying sets of substitution rate parameters that are capable of accommodating rate heterogeneity between sites

c) Exploration of maximum likelihood inferences of positive selection on the basis of a codon-by-codon statistical analysis to determine the ratio of non-synonymous to synonymous evolutionary changes across the p67 gene sequence dataset

d) Judging the possible influence of immunological selection in contributing to the evolution of the p67 gene on the basis of whether the region of the gene encoding the B-cell epitopes has undergone positive selection for sequence diversity.

2.4 References

Babik W, Kawalko A, Wojcik JM, Radwan J (2012) Low major histocompatibility complex class 1 (MHC I) variation in the European bison (*Bison bonasus*). J Hered 103:349–359

Bishop R, Nene V, Staeyert J, Rowlands J, Nyanjui J, Osaso J, Morzaria S, Musoke A (2003) Immunity to East Coast fever in cattle induced by a polypeptide fragment of the major surface coat protein of *Theileria parva* sporozoites. Vaccine. 21:1205–1212

Bishop R, Shah T, Pelle R, Hoyle D, Pearson T, Haines L, Brass A, Hulme H, Graham SP, Taracha EL, Kanga S, Lu C, Hass B, Wortman J, White O, Gardner MJ, Nene V, de Villiers EP (2005) Analysis of the transcriptome of the protozoan *Theileria parva* using MPSS reveals that the majority of genes are transcriptionally active in the schizont stage. Nucleic Acids Res. 33(17):5503-11

Di Giulio G, Lynen G, Morzaria S, Oura C, Bishop R (2009) Live immunization against East Coast fever--current status. Trends Parasitol. 25(2):85-92

Dobbelaere, DA, Spooner PR, Barry WC, Irvin AD (1984) Monoclonal antibody neutralizes the sporozoite stage of different *Theileria parva* stocks. Parasite Immunol. 6(4):361-70

Ellis S (2004) The cattle major histocompatibility complex: is it unique? Vet Immunol Immunopathol. 102(1–2):1–8

Gardner MJ, Bishop R, Shah T, et al. (2005) Genome sequence of a bovine pathogen that transforms lymphocytes. Science. 309:134–137

George J.E, Pound JM, Davey RB (2008) Acaricides for controlling tick on cattle and the problem of acaricide resistance. Ticks: Biology, Diseases and Control (ed. by A.S. Bowman & P. Nuttall), pp. 408–423. Cambridge University Press, Cambridge, MA

Geysen, D. et al. (1999) Molecular epidemiology of *Theileria parva* in the field. Trop. Med. Int. Health 4, A21–A27

Graham SP, Pellé R, Yamage M, Mwangi DM, Honda Y, et al. (2008) Characterization of the fine specificity of bovine $CD8^+$ T-cell responses to defined antigens from the protozoan parasite *Theileria parva*. Infect. Immun. 76:685–94

Graham SP, Pellé R, Honda Y, Mwangi DM, Tonukari NJ, Yamage M, Glew EJ, de Villiers EP, Shah T, Bishop R, Abuya E, Awino E, Gachanja J, Luyai AE, Mbwika F, Muthiani AM, Ndegwa DM, Njahira M, Nyanjui JK, Onono FO, Osaso J, Saya RM, Wildmann C, Fraser CM, Maudlin I, Gardner MJ, Morzaria SP, Loosmore S, Gilbert SC, Audonnet JC, van der Bruggen P, Nene V, Taracha EL (2006) *Theileria parva* candidate vaccine antigens recognized by immune bovine cytotoxic T lymphocytes. Proc Natl Acad Sci U S A 103(9):3286–3291

Hansen AM, Rasmussen M, Svitek N, Harndahl M, Golde WT, Barlow J, Nene V, Buus S, Nielsen M (2014) Characterization of binding specificities of bovine leucocyte class I molecules: impacts for rational epitope discovery. Immunogenetics. 66(12):705–718

Henson S, Bishop RP, Morzaria S, Spooner PR, Pelle R, Poveda L, Ebeling M, Küng E, Certa U, Daubenberger CA, Qi W (2012) High-resolution genotyping and mapping of recombination and gene conversion in the protozoan *Theileria parva* using whole genome sequencing. BMC Genomics. 23; 13:503

MacHugh DE, Shriver MD, Loftus RT, Cunningham P, Bradley DG (1997) Microsatellite DNA variation and the evolution, domestication and phylogeography of taurine and zebu cattle (*Bos taurus* and *Bos indicus*). Genetics 146(3):1071–1086

MacHugh ND, Connelley T, Graham SP, Pelle R, Formisano P, Taracha EL, Ellis SA, McKeever DJ, Burrells A, Morrison WI (2009) CD8+ T-cell responses to *Theileria parva* are preferentially directed to a single dominant antigen: Implications for parasite strain-specific immunity. Eur J Immunol 39(9):2459-69

McKeever DJ, Morrison WI (1994) Immunity to a parasite that transforms T lymphocytes. Curr Opin Immunol. 6:564–567

Morrison WI, McKeever DJ (2006) Current status of vaccine development against *Theileria* parasites. Parasitology 133(Suppl.):S169–87

Musoke A, Morzaria S, Nkonge C, Jones E, Nene V (1992) A recombinant sporozoite surface antigen of *Theileria parva* induces protection in cattle. Proc Natl Acad Sci U S A. 89:514–518

Musoke A, Rowlands J, Nene V, Nyanjui J, Katende J, Spooner P, Mwaura S, Odongo D, Nkonge C, Mbogo S, Bishop R, Morzaria S (2005) Subunit vaccine based on the p67 major surface protein of *Theileria parva* sporozoites reduces severity of infection derived from field tick challenge. Vaccine. 23:3084–3095

Nene V, Musoke A, Gobright E, Morzaria S (1996) Conservation of the sporozoite p67 vaccine antigen in cattle-derived *Theileria parva* stocks with different cross-immunity profiles. Infect Immun. 64:2056–2061

Nene V, Gobright E, Bishop R, Morzaria S, Musoke A (1999) Linear peptide specificity of bovine antibody responses to p67 of *Theileria parva* and sequence diversity of sporozoite-neutralizing epitopes: implications for a vaccine. Infect Immun. 67:1261–1266

Nene V, Svitek N, Toye P, Golde WT, Barlow J, Harndahl M, Buus S, Nielsen M (2012) Designing bovine T cell vaccines via reverse immunology. Ticks Tick Borne Dis. 3(3):188–192

Nuttall P A and Labuda M (2008) Saliva-assisted transmission of tick-borne pathogens. Ticks: Biology, Diseases and Control (ed. by A.S. Bowman & P. Nuttall), pp. 408–423. Cambridge University Press, Cambridge, MA

Oura CA, Bishop R, Asiimwe BB, Spooner P, Lubega GW, Tait A (2007) *Theileria parva* live vaccination: parasite transmission, persistence and heterologous challenge in the field. Parasitology. 134(Pt 9):1205-13

Patel E, Mwaura S, Kiara H, Morzaria S, Peters A, Toye P (2016) Production and dose determination of the Infection and Treatment Method (ITM) Muguga cocktail vaccine used to control East Coast fever in cattle. Ticks Tick Borne Dis. 7(2):306-14

Radley DE, Brown CGD, Burridge MJ, Cunningham MP, Kirimi IM, Purnell RE, Young AS (1975a) East Coast Fever. 1. Chemoprophylactic immunisation of cattle against *Theileria parva* (Muguga) and five Theileria strains. Vet. Parasitol. 1, 35–41

Radley DE, Brown CGD, Cunningham MP, Kimber CD, Musisi FL, Payne RC, Purnell RE, Stagg DA., Young AS (1975b) East Coast fever: 3. Chemoprophylactic immunization of cattle using oxytetracycline and a combination of Theilerial strains. Vet. Parasitol. 1, 51–60

Robinson J, Halliwell JA, McWilliam H, Lopez R, Marsh SG (2013) IPD—the immuno polymorphism database. Nucleic Acids Res. 41(Database issue): D1234–D1240

Svitek N, Hansen AM, Steinaa L, Saya R, Awino E, Nielsen M, Buus S, Nene V (2014) Use of 'one-pot, mix-and-read' peptide-MHC class I tetramers and predictive algorithms to improve detection of cytotoxic T lymphocyte responses in cattle. Vet Res. 45:50

Taracha EL, Goddeeris BM, Morzaria SP, Morrison WI (1995) Parasite strain specificity of precursor cytotoxic T cells in individual animals correlates with cross-protection in cattle challenged with *Theileria parva*. Infect Immun. 63(4):1258-62

3.0 Chapter 3: Publication I

This chapter considers the extent to which recombinant subunit vaccine development for ECF control based on induction of protective $CD8^+$ T-cell responses may be constrained by sequence diversity present within the class I MHC loci of cattle. Given that the expressed class I MHC sequences of African taurine cattle are yet to be analyzed in detail, this chapter reports on an initial study that utilized amplicon-based pyrosequencing next generation sequencing (NGS) techniques combined sophisticated processing algorithms to obtain reliable genotypes from Ankole cattle.

The 454 GS FLX platform run yielded a total of 172,681 reads (over 56 million bases) from 17 Ankole cattle and after discarding reads categorized as being of insufficient quality by an initial set of filters, 966 variants in exon 2 and 1642 variants in exon 3 were retained. To discriminate true allelic variants from artifacts, especially where several functionally distinct alleles may differ by a single nucleotide, the study adapted an allele calling workflow initially used for filtering some 86,153 reads from a wild rodent with high copy number variation in the MHC class II DRB locus (Sommer et al. 2013). In the end, only 27 reads were classed as legitimate sequence variants. The authenticity of these 'true' variants was corroborated by class I MHC cDNA Sanger sequencing data. Of the 27 unique Ankole class I MHC transcripts, only five (18.5%) shared 100% nucleotide identity with genes in the existing catalog of known cattle class I MHC sequences, while 18 allelic variants (67%) were novel. The remaining variants in the Ankole class I MHC catalog shared considerable sequence identity with either published non classical class I sequences or null alleles suffixed 'N' based on IPD nomenclature conventions to denote that they are transcribed but contains features that are inconsistent with expression as functional proteins.

For comparison with the African taurine class I MHC transcripts, plasmid sub-clones containing a 632 base pair amplicon spanning exons 2 to 4 of the class I MHC alpha chain from European Holstein cattle

resident in sub-Saharan Africa were also sequenced. As anticipated, the Holstein class I MHC data set was dominated by sequences displaying 100% nucleotide identity with allelic variants present in the existing cattle MHC repository. However, the analysis also indicated that novel Holstein class I variants will continue to be discovered with increased sampling as exemplified by the discovery of two new sequences in this study, one of which exhibited only 93.64% sequence similarity to any cattle class I MHC molecule currently present in the databases.

The extent of the functional divergence between Ankole and Holstein class I MHC sequences was assessed using *in silico* peptide binding analysis for functional clustering of class I MHC sequences which provides a more useful and readily interpretable classification of likely peptide binding specificities. The data clearly indicates likely functional differences between Ankole and Holstein class I MHC and strongly suggests that a number of different antigens/epitopes will need to be incorporated in a recombinant antigen cocktail vaccine to provide broad coverage. Furthermore, recognition of currently identified candidate 'Tp' $CD8^+$ T-cell target antigens was assessed with an *ex vivo* gamma interferon ELISpot assay using peripheral blood monocytes (PBMC) and purified $CD8^+$ T-cells from Ankole cattle that were *T. parva* positive according to an indirect ELISA based on a recombinant version of the PIM antigen. It may be significant that the vaccine candidate antigens were originally identified using $CD8^+$ T-cell lines from Holstein (European taurine) cattle to screen expressed *T. parva* cDNA libraries. To summarize, PBMC from each of the eleven seropositive Ankole did not recognize any of the six 'Tp' antigens. This includes the Tp1 and Tp2 antigens, which are currently the main focus of ECF recombinant vaccine trials involving an international consortium of laboratories. The findings from this study confirm the need for development of a multivalent vaccine that can potentially overcome the issue of cattle MHC diversity in the field in Africa.

The nucleotide sequences of all the novel alleles described in this study have been submitted to a public database and can be accessed from the European nucleotide archive with the accessions LN8886781-LN8886805.

Full details of the scientific findings of this chapter have been subjected to peer review and published as an original paper in Immunogenetics. In compliance with the copyright transfer statement to Springer-Verlag Berlin Heidelberg, an author-created version of the published article appears overleaf. The final publication, together with the supplementary material published electronically is available at http://link.springer.com/article/10.1007%2Fs00251-016-0902-5.

Sequence diversity between class I MHC loci of African native and introduced *Bos taurus* cattle in *Theileria parva* endemic regions: *In silico* peptide binding prediction identifies distinct functional clusters

Obara Isaiah[1]*, Nielsen Morten[3], Jeschek Marie[2], Nijhof Ard[1], Mazzoni Camila[2], Svitek Nicholas[4], Steinaa Lucilla[4], Awino Elias[4], Olds Cassandra[4], Jabbar Ahmed[1], Clausen Peter-Henning[1], Bishop Richard P. [4]

1. Institute for Parasitology and Tropical Veterinary Medicine, Freie Universität Berlin, Robert-von-Ostertag-Str. 7-13, D 14163 Berlin, Germany
2. Berlin Center for Genomics in Biodiversity Research (BeGenDiv), Königin-Luise-Str. 6-8, D 14195 Berlin, Germany
3. Center for Biological Sequence Analysis (CBS), Department of Systems Biology, The Technical University of Denmark, Building 208, DK-2800 Lyngby, Denmark
4. International Livestock Research Institute (ILRI), P.O. Box 30709, Nairobi, 00100, Kenya

*Corresponding author

Obara Isaiah, Email: iobara@zedat.fu-berlin.de

This is an author-created version of a peer-reviewed journal article published as an original paper in Immunogenetics. The final publication is available at http://link.springer.com/article/10.1007%2Fs00251-016-0902-5

DOI: 10.1007/s00251-016-0902-5, May 2016, Volume 68, Issue 5, pp 339-352

Abstract

There is strong evidence that the immunity induced by live vaccination for control of the protozoan parasite *Theileria parva* is mediated by class I MHC-restricted $CD8^+$ T-cells directed against the schizont stage of the parasite that infects bovine lymphocytes. The functional competency of class I MHC genes is dependent on the presence of codons specifying certain critical amino acid residues that line the peptide binding groove. Compared with European *Bos taurus* in which class I MHC allelic polymorphisms have been examined extensively, published data on class I MHC transcripts in African taurines in *T. parva* endemic areas is very limited. We utilized the multiplexing capabilities of 454 pyrosequencing to make an initial assessment of class I MHC allelic diversity in a population of Ankole cattle. We also typed a population of exotic Holstein cattle from an African ranch for class I MHC and investigated the extent, if any, that their peptide-binding motifs overlapped with those of Ankole cattle. We report the identification of 18 novel allelic sequences in Ankole cattle and provide evidence of positive selection for sequence diversity, including in residues that predominantly interact with peptides. *In silico* functional analysis resulted in peptide binding specificities that were largely distinct between the two breeds. We also demonstrate that $CD8^+$ T-cells derived from Ankole cattle that are seropositive for *T. parva* do not recognize vaccine candidate antigens originally identified in Holstein and Boran (*Bos indicus*) cattle breeds.

Keywords

Theileria parva, $CD8^+$ T-cell epitopes, class I MHC, Ankole cattle, pyrosequencing

Introduction

The protozoan parasite *Theileria parva* (*T. parva*) is responsible for East Coast fever (ECF), an acute lymphoproliferative disease of cattle in eastern, central and southern Africa (Norval et al. 1992). Transmitted by the tick *Rhipicephalus appendiculatus*, *T. parva* sporozoites are infective to mammalian hosts and these differentiate rapidly into intracellular schizonts that immortalize bovine lymphocytes. The Infection and treatment (ITM) procedure has been used as an approach to live vaccination for control of *T. parva* for over 40 years (Radley et al. 1975). Despite a high efficacy against field challenge, there are a number of constraints to widespread deployment of ITM. The major ones are the technical and ethical issues relating to production and subsequent delivery of live sporozoites using a cold chain combined with the perception that *T. parva* parasites that are foreign to an area may be introduced (reviewed by Di Giulio et al. 2009).

Although monoclonal antibodies against a polymorphic *T. parva* sporozoite surface molecule (Nene et al. 1999; Obara et al. 2015) have been shown to neutralize infectivity *in vitro* (Musoke et al. 1982), at present there is no direct evidence that ITM induces significant levels of antibody against the sporozoite stage. In contrast, available data indicate that one important mechanism of protection afforded by ITM is attributable to $CD8^+$ T-cells specific for the intra-lymphocytic schizont stage of *T. parva*. Earlier time-course studies showed an association between the kinetics of class I MHC restricted $CD8^+$ T-cell cytotoxicity and remission of infection in immune animals (Emery et al. 1981; Eugui et al. 1981). The most compelling evidence for the role of $CD8^+$ T-cells as major mediators of immunity to *T. parva* came from a study in which transfer of $CD8^+$ T-cell enriched lymphocyte populations from an immune animal to its naive chimeric twin conferred protection (McKeever et al. 1994). These observations led to the screening of *T. parva* schizont cDNA libraries with $CD8^+$ T-cells from Holstein and Boran cattle immunized by ITM, with the aim of defining the antigenic specificity of the protective response (Graham et. al 2006). It was evident from the findings of the antigen screens that animals of

different class I MHC genotypes recognized distinct antigens (Graham et. al 2008). A more recent study utilizing a positional scanning combinatorial peptide library (PSCPL) driven analysis corroborated the findings from the antigen screening experiments (Hansen et al. 2014). Taken together with the evidence that different cattle breeds often carry distinct class I MHC alleles (Robinson et al. 2013), the fact that recognition of *T. parva* $CD8^+$ T-cell target antigens is class I MHC restricted raises the issue of the likely impact of MHC diversity on the coverage of epitope-based vaccines.

The number of 'breeds' so far documented from the ECF endemic region, based on the information contained within the FAO breed database (http://dad.fao.org/) is two hundred an eleven in the eleven countries where the disease currently occurs. However such breeds cannot be directly equated to defined European breeds, such as the Holstein, since although ancestral African cattle populations typically have taurine mitochondrial haplotypes, introgression of allelles into the nuclear genome from *Bos indicus* has occurred at high frequency (Machugh, 1997; Anderung et al. 2007). Although potentially serious challenges are presented by the complexity and diversity of expressed bovine MHC loci in African cattle, there are currently significant on-going efforts to design vaccines based on $CD8^+$ -T cells targeting *T. parva*-infected lymphocytes using reverse vaccinology approaches (Nene et al. 2012; Hansen et al. 2014; Svitek et al. 2014; Graham et al. 2007). These are based on training neural network software (Hoof et al. 2009) using the $CD8^+$ T-cell target antigens and epitopes described above (Graham et al. 2006; 2008), with the predictions validated using tetramers designed to be recognized by $CD8^+$ T-cells designed from *T. parva* antigen-peptide complexes (Svitek et al. 2014).

Available data on cattle class I MHC diversity demonstrates that majority of the variability among allelic sequences reside in amino acid positions that determine specificity for peptide binding (Robinson et al. 2013). A listing of cattle MHC alleles and a description of the nomenclature convention is provided in the Immuno-Polymorphism database (IPD;

http://www.ebi.ac.uk/ipd/mhc/bola/). By comparison to humans that have a fixed complement of three classical class I loci (HLA-A, HLA-B and HLA-C), mapping and phylogenetic data (Holmes et al. 2003; Ellis 2004; Di Palma et al. 2002; Birch et al. 2006) support the grouping of cattle class I MHC sequences into six classical loci (genes 1–6). Antigen binding and presentation capacity has been demonstrated for products of all the six loci (Codner et al. 2012). A further level of complexity in cattle class I MHC results from variations in gene content between haplotypes (Ellis et al. 1999; Ellis 2004). Despite recent advances in assay throughput capabilities (Harndahl et al. 2009), the required scale imposes practical limitations that make such experimental determination of the ligand selectivity of all possible haplotypes impractical. However, the artificial neural network based MHC binding predictor, *NetMHCpan* (Hoof et al. 2009; Nielsen et al. 2007) represents a significant advance in the possibility to provide accurate predictions of peptide-class I MHC interactions, even across different species. Although trained predominantly on peptide binding data covering human and nonhuman primates, the versatility of *NetMHCpan* has been demonstrated in the analysis of the binding specificities of porcine (Pedersen et al. 2011) and bovine (Svitek et al. 2014) class I MHC molecules. Most recently, PSCPL analysis of peptide binding for bovine class I MHC molecules known to restrict *T. parva* candidate vaccine antigens provided data that proved useful for refining the predictive performance of *NetMHCpan* (Hansen et al. 2014). Furthermore, procedures exist for validation of the natural presentation and immunogenicity of the predicted epitopes based on selection of cytokine-secreting or tetramer-positive T cell populations (Nene et al. 2012; Svitek et al. 2014; Hansen et al. 2014). Similar approaches have recently been applied for prediction of bovine class I MHC restricted epitopes in foot and mouth disease (Pandya et al. 2015).

A key information gap constraining application of the expedited epitope identification procedures in strategies for *T. parva* control is the paucity of class I MHC data for cattle populations in the endemic regions of Eastern Africa. A major challenge to typing cattle for class I MHC is presented by a

complex genomic organization that necessitates simultaneous genotyping of co-amplifying loci (reviewed by Ellis 2004). In the current study, we utilized the depth and throughput of pyrosequencing, coupled with the multiplexing capabilities of the 454 GS FLX platform to resolve variations at the structurally important peptide-binding regions of class I MHC transcripts in a population of Ankole cattle, an African taurine with demonstrated tolerance to infections with *T. parva* (Paling et al. 1991). For comparison, we also characterized a European Holstein population, introduced to East Africa, for the sequence of the class I MHC alleles present. To gain perspective into differential ligand selectivity of the alleles described in the present study, we derived functionally relevant allele groupings based on sequence features defining peptide binding grooves using *MHCcluster* (Thomsen et al. 2013). Application of this software provides evidence for major specificity groups, known as 'supertypes' in cattle (Pandya et al. 2015), as previously described for humans by Sette and Sidney (1999), suggesting that the constraint of class I MHC diversity for recombinant vaccine development for ECF control may not be insurmountable. In the study described herein, the use of the *MHCcluster* software provided evidence for minimal functional overlap between Ankole and Holstein peptide binding specificity, suggesting a requirement for additional epitopes in ECF subunit vaccines designed to induce $CD8^+$ T-cell based protection.

Materials and Methods

Taurine samples, RNA isolation and cDNA synthesis

Peripheral blood lymphocytes from seventeen African taurines (Ankole) were used for preparation of pyrosequencing libraries, fourteen of which were from Ol Pejeta - a ranch that integrates game conservancy with cattle production in central Kenya (www.olpejetaconservancy.org/). The other three animals, originally from a farm in Uganda, were part of a herd used for genetic studies at ILRI. For comparison, seventeen Holstein cattle, which were part of a vaccine trial that combined tick salivary gland and a *T. parva* sporozoite antigen at ILRI (R. Bishop unpublished), were also typed by plasmid sub-cloning of class I MHC amplicons and chain termination sequencing. Ficoll-based density centrifugation was used for separation of peripheral blood mononuclear cells (PBMCs) from cattle blood drawn into heparinized vacutainers. Cellular RNA was extracted from PBMCs using Trizol (Invitrogen), quantified by measuring Abs_{260} with a nanodrop UV spectrophotometer (Inqaba biotech, SA) and resolved by 1.2% agarose gel electrophoresis to determine integrity. Total RNA was used as template for first-strand cDNA synthesis using a Superscript III Reverse Transcriptase kit (Invitrogen).

Ankole class I MHC amplicon library construction and pyrosequencing

We generated PCR amplicons spanning 410 base pairs (bp) between conserved sequences in exons two and three of class I MHC sequences from Ankole cattle using the Phusion High-Fidelity PCR kit (Thermo scientific). For each sample, we introduced a unique ten-nucleotide multiplex identifiers (MID) index to both the sense strand primer: Bov 7 (5'-GGCTACGTGGACGACACG-3') and the anti-sense strand primer: Bov 11 (5'-CCCTCCAGGTAGTTCCT-3'). Thermal conditions and PCR mix composition were as previously described (Birch et al. 2006). Following MinElute PCR purification (Qiagen), the amplicons were quantified using a Nanodrop UV spectrophotometer (Inqaba biotech) and normalized to equimolar concentrations. Adaptor ligation, library amplification by emulsion PCR, and

bidirectional pyrosequencing of amplicons as one pool in 1/16 region of a 70 x 75 picotiter plate were performed at ILRI using GS FLX Titanium chemistry (454 Life Sciences).

Variant calling: putative alleles, artefacts and chimeras

Since errors inherent to pyrosequencing mean that the potential for some proportion of the reads to be artifactual cannot be excluded, we subjected the reads passing the standard GS FLX base-calling criterion to a multilevel artifact filtering workflow. Foremost, reads that failed the following initial filtering criteria were considered as putative artifacts and excluded from further analysis: (i) sequence length of at least 196 base pairs which could allow reads to be assigned to either exon 2 or 3 ; (ii) presence of complete MID sequences to allow reads to be disaggregated into the indexed individuals; (iii) absence of base pair mismatches with the PCR primers sequences; (iv) reads that could be assigned to either exon 2 or 3 on the basis of exonic nucleotide sequence alignment to the published reference sequence, BoLA-3*00101; (v) a minimum phred scaled quality score value of 20 in more than 95% of bases; and (vi) absence of ambiguous bases ('N' calls). Whilst the demultiplexing steps were accomplished using the flexible barcode and adapter removal tool (FLEXBAR; Dodt et al. 2012), we used the FASTX command line tools, FASTQ/A Trimmer and FASTQ Quality Filter (http://hannonlab.cshl.edu/fastx_toolkit), for length trimming and quality filtering respectively. Exon sequences were identified by similarity searches using the NCBI nucleotide BLAST program (Altschul et al. 1990).

The second set of analyses was intended to evaluate variant authenticity by detecting chimeric reads and base miscalls based on intra and inter-amplicon comparisons as described in Sommer et al. (2013). Briefly, identical reads from each MID indexed sample were collapsed into variants, using the FASTQ/A Collapser tool (http://hannonlab.cshl.edu/fastx_toolkit) and singletons (variants represented by single reads in an amplicon) were discarded. The variants, ordered by frequency, were aligned using

MAFFT version 7.122 using the L-INS-i option with minor manual adjustments (Katoh et al. 2013). The collection of aligned reads from indexed animals was stored in a SQLite database for downstream analysis. SQL queries and custom Biopython scripts were written to classify variants as putative alleles if they comprised the most frequent cluster and did not match any of the artefact definitions. The remaining variants were designated as putative alleles, 'unclassified' variants or artifactual sequences based on inter and intra-animal comparisons and frequency relative to artifacts within the same amplicon as described elsewhere (Sommer et al. 2013). Artifacts were excluded from downstream analysis. Finally we used a custom Biopython script to rank, separately for each animal, all the potential combinations of the exon 2 and exon 3 variants classified as putative alleles according to their abundance in the MID demultiplexed reads from an individual. Since the exon sequences were already known at this step, we relied on partial identifying sequences to allow the inclusion of MID demultiplexed reads that were too short to comprise the complete sequences of exon two and three in this step. We further evaluated the reliability of the genotypes obtained by amplicon pyrosequencing by comparing them to partial or full-length class I MHC cDNA Sanger sequencing datasets for the three Ankole cattle from the ILRI herd.

Generation of cattle class I cDNA-PCR clone libraries and Sanger sequencing

We used enzyme *Pfu* DNA polymerase (Thermo scientific) to generate an amplicon that spans 632 bp between conserved sequences in exons 2 and 4 of class I MHC sequences from the seventeen Holsteins and three Ankole cattle. The inclusion of the three Ankole samples, already genotyped by pyrosequencing, was intended to highlight the reliability of the pyrosequencing reads. We used cDNA templates and the highly conserved primer pairs, Bov 7 (5'-GGCTACGTGGACGACACG-3') and Bov 14 (5'-GAAGGCCTGGTCTCCACAA-3'), for amplification as previously described (Birch et al. 2006). For the three Ankole cattle, full length class I MHC genes were also amplified from cDNA as described in Birch et al. (2006). The amplicons were resolved on 1.2% ethidium bromide stained

agarose gels, excised and purified with DNA gel extraction spin columns (Qiagen). PCR products were sub-cloned into the pGEM T-Easy vector (Promega) and subsequently transformed into competent JM 109 *Escherichia coli* cells. From each animal, 10 positive (blue/white selection) clones were picked and grown overnight in 5 ml of LB-Medium. Plasmid DNA was purified according to QlAprep Miniprep Kit Protocol (QIAGEN) and insert presence confirmed by *EcoRI* digestion or by colony screening PCR. Sanger sequencing of sub-cloned products was undertaken bidirectionally using T7 and SP6 primers. Reads from each clone, assembled into unidirectional contigs using CLCgenomics version 8.0.2 (http://www.clcbio.com/products/clc-genomics-workbench/), were subjected to the same artefact and chimera filter as described for the Ankole data, but with a modification of the preprocessing criterion to retain singletons after collapsing if they were carried by more than one animal. This relaxation in stringency reflects the differences in throughput between pyrosequencing and Sanger sequencing.

In silico clustering of cattle class I MHC sequences based on predicted ligand selectivity

We investigated overlaps, if any, in the predicted ligand selectivities of the Ankole and Holstein class I MHC products based on *NetMHCpan* 2.8 predictions and used *MHCcluster* 2.0 (Thomsen et al. 2013) to generate a specificity heat map and a functional tree constructed using 100 bootstrap calculations and 50,000 peptides. The predictions are based on sequence features defining the peptide binding grooves of the class I MHC products.

$CD8^+$ T-cell specificities in the peripheral circulation of *T. parva* infected Ankole cattle

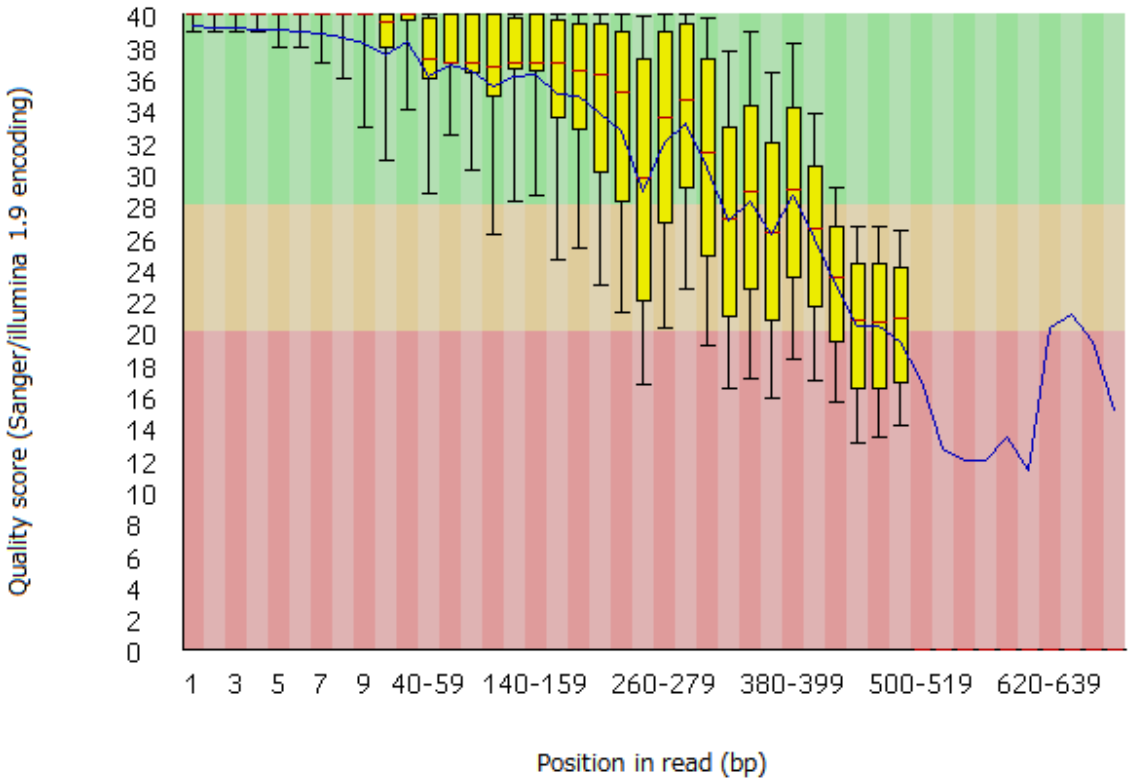
We assayed 40 field Ankole serum samples for antibodies to *T. parva* by enzyme-linked immunosorbent assay (ELISA) using the recombinant polymorphic immunodominant molecule (PIM) as described previously (Katende et al. 1998). IFN-γ responses generated following stimulation of PBMCs, as well as $CD8^+$ T-cell fractions, from serologically positive animals with complete sets of

overlapping peptides representing *T. parva* candidate vaccine antigens were assayed in an *ex vivo* enzyme-linked immunospot (ELISpot) assay as described previously (Steinaa et al. 2012). Briefly, sterile nitrocellulose-lined 96-well microplates (Millipore, Billerica, MA, USA) were coated overnight at 4°C with anti-bovine IFN-γ antibody (Serotec, Oxford, UK, cat.no. MCA1783) and blocked with RPMI containing 10% heat-inactivated FBS for 2 hours at 37 °C. The cryopreserved PBMCs were plated out in the coated wells in triplicate at 2×10^5 cells/well in 50 µl and an equal volume of complete sets of overlapping peptide pools derived from the *T. parva* vaccine candidate antigens added at 1 µg/ml final concentration (Graham et al. 2006). Negative (no peptide) control triplicate wells contained culture in supplemented RPMI medium. Cytotoxic T lymphocytes (CTLs) from BV115 - an animal with demonstrated reactivity to Tp1, were used as positive assay control in triplicate wells. The cells were incubated for 36 hours at 37 °C, 5% CO_2. Spots were visualized by incubation with primary rabbit polyclonal anti-bovine IFN-γ antibody (Sigma–Aldrich, St. Louis, MO, USA), secondary AP-conjugated monoclonal anti-rabbit antibody (Sigma–Aldrich) and Fast substrate solution (Sigma–Aldrich). The results are expressed as the mean spot-forming units (SFU)/10^6 PBMCs.

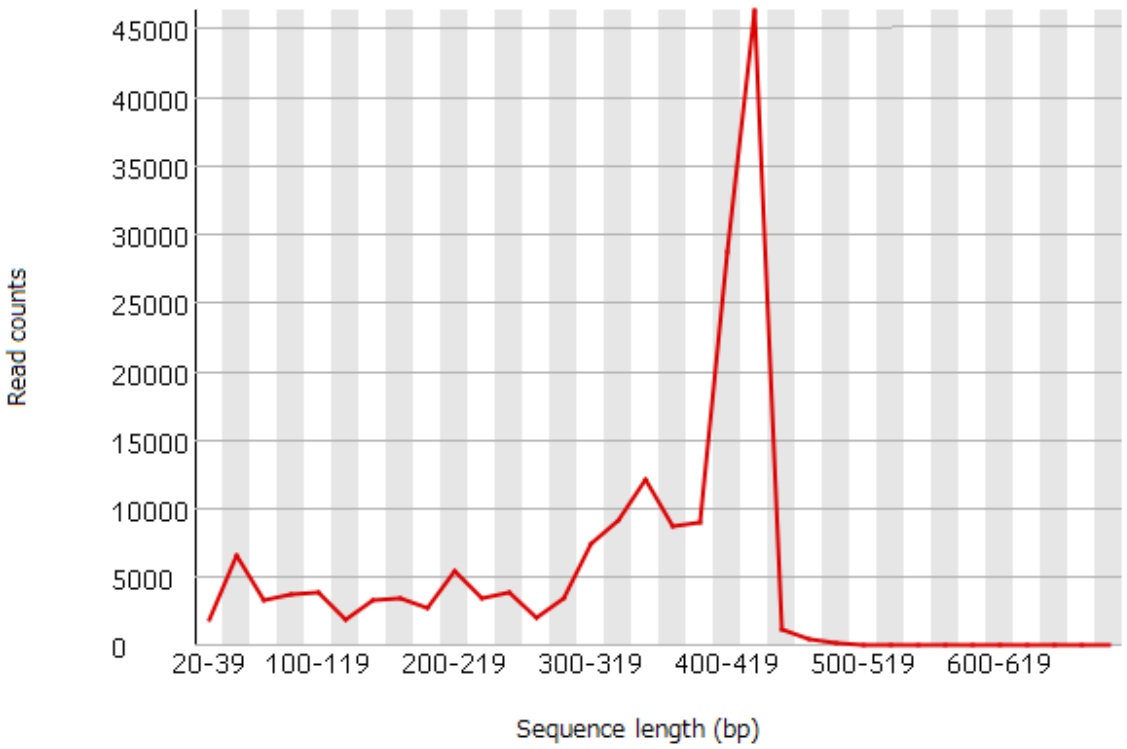
Results

A high throughput assessment of class I MHC diversity from a population of Ankole cattle

We sequenced normalized, GS FLX adaptor flanked, 410 base pair amplicon libraries that extended exons 2 and 3 of cattle class I MHC from seventeen Ankole cattle on a Roche 454 instrument. A total of 172,681 reads (over 56 million bases) passed the standard GS FLX base-calling filters. Majority of the bases that passed the GS FLX base-calling filters had a quality score value of more than 20 on the Phred scale, indicating a base call accuracy of more than 95% (Supplementary figure 1). Furthermore, the modal sequence length was consistent with the expected product size (Supplementary figure 2).



Supplementary figure 1 Distribution of quality scores generated using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/), across all bases that passed the GS FLX base-calling filters. Center lines (red) show the medians; the box limits (yellow) indicate the 25th and 75th percentiles; the upper and lower whiskers extend to the 10th and 90th percentiles respectively and the blue line represents the mean quality. The green, orange and red background partitions of the y axis correspond to very good quality calls, calls of reasonable quality and poor quality calls respectively. n= 56,528,008 bases



Supplementary figure 2 Length distribution over all sequences that passed the GS FLX base-calling quality filters. The plot was generated using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/).

Reads disaggregated on the basis of unique MID barcodes and further partitioned into individual exons were used as input for two sets of filters intended to detect and eliminate errors known to influence the reliability of pyrosequencing calls. The initial data quality check was restricted to filtering: (i) reads bearing mismatches to amplification primers; (ii) reads with ambiguous bases ('N' calls); (iii) reads with intra-primer sequence lengths shorter than 196 base pairs; and (iv) reads containing more than 5% low quality bases. Candidate variants that passed these initial filters were preprocessed for artifact filtering by collapsing identical reads, while maintaining read counts, and discarding singletons. The artifact filtering criterion excluded data from reads with bases that could be associated with chimerism and/or miscalls on the basis of intra and inter-amplicon comparisons as described in Sommer et al.

(2013). The distribution of read counts over all MID indexed samples after successive rounds of filtering, normalized to sum to 1 (divided by the MID demultiplexed read counts) and partitioned into exons 2 and 3 is shown in Fig. 1.

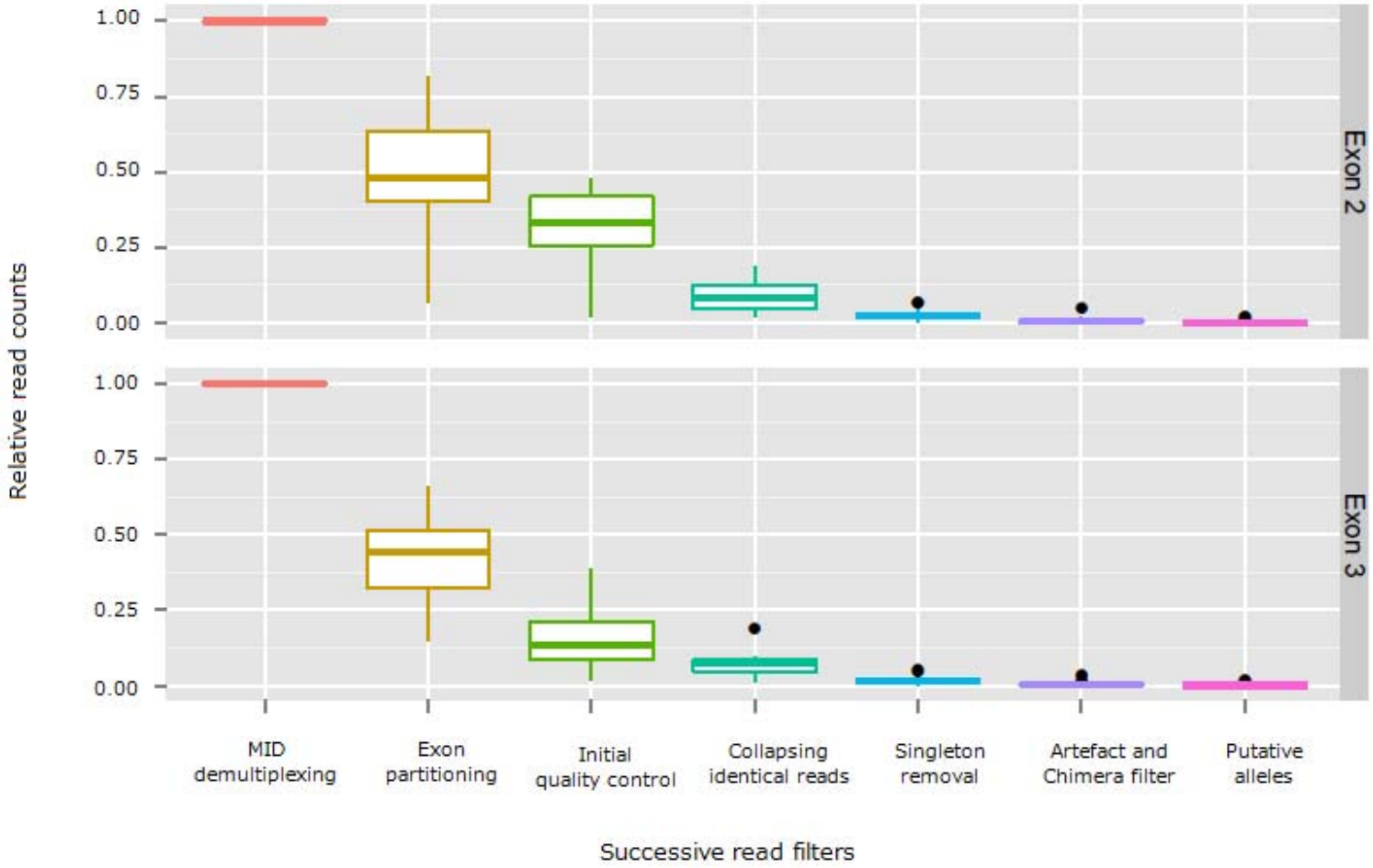


Fig. 1 Distribution of relative read counts over all MID indexed samples after successive rounds of filtering generated using ggplot in the R software. Counts are normalized to the initial number of reads after MID demultiplexing. Center lines show the medians; box limits indicate the 25th and 75th percentiles; whiskers extend 1.5 times the interquartile range from the 25th and 75th percentiles, outliers are represented by black dots. n = 17 Ankole cattle.

After removal of PCR and sequencing associated artifacts, 27 putative alleles were obtained from a total of 966 exon 2 and 1642 exon 3 variants across the 17 Ankole samples (Table 1). The exonic variants classed as 'putative alleles', as well as the 'unclassified variants' which were carried by each of the 17 Ankole animals is shown in supplementary table 1. The latter includes a small number of variants whose classification as either putative allele or artefacts was inconclusive. It should be emphasized that this is a minimum estimate of alleles for these 17 animals and the stringent filtering process may have resulted in some alleles being missed. It also possible that the primers did not amplify all class I MHC loci from the Ankole, although they were designed to be as representative as possible of MHC diversity in cattle (Birch et al. 2006) and are also conserved in European Bison (*Bison bonasus*) (Babik et al. 2012). BLAST searches of the IPD and NCBI nucleotide databases conducted with the nucleotide sequences of the 27 unique Ankole class I MHC transcripts identified only five variants sharing 100% identity with published cattle classical class I MHC sequences. With one exception, these previously reported sequences have been assigned to gene 2 and none has been reported in Holstein cattle. Similarly, two variants matched published non classical class I MHC sequences with 100% nucleotide identity. In addition to the characterized non classical class I sequences, our analysis identified two putatively novel variants that share more than 99% nucleotide identity with the published non classical allele BoLA-NC2*00102 (Table 1). Furthermore, two of the sequences, BoLA-Ank_12 and 16, share in excess of 98% nucleotide identity to BoLA-2*03201N, a null allele suffixed 'N' based on IPD nomenclature conventions to denote that it is transcribed but contains features that are inconsistent with expression as a functional protein.

Considerable variation was observed among the remainder of the sequences with respect to the percent nucleotide identity to previously described cattle classical MHC I alleles. Interestingly, with a few exceptions, these putatively novel classical class I variants share identities with published sequences known to be expressed in breeds other than the Holstein. The diversity per site, excluding variants

shown to share nucleotide identity with non-classical class I alleles was p = 0.09436(or 9.436%) ±0.00385. The nucleotide sequences designated as putatively novel in this study have been submitted to the European nucleotide archive (LN8886781- LN8886805). In addition to the stepwise variant validation procedure, a further consideration involved in evaluating the reliability of the genotypes obtained by pyrosequencing related to comparing them to the genotypes obtained by cloning and Sanger sequencing for the three Ankole cattle from the ILRI herd. Amongst the partial (632 bp) as well as the full-length Sanger sequencing datasets, five distinct alleles were identified, namely BoLA-Ank_04, 06, 08, 09 and 25. These matched the sequences retained from the amplicon pyrosequencing data by the cascade of filters used for variant calling.

Table 1 Sequence diversity present within the second and third exons of class I MHC genes in Ankole cattle

Working allele name	Read counts			IPD BLASTn search		GenBank Accession	Breed known to express allele
	Exon 2	Exon 3	Individuals carrying allele	Allele Name	Percent identity		
BoLA-Ank_02	30	23	2	BoLA-2*00501	100	LN886782	Hereford
BoLA-Ank_05	150	54	2	BoLA-2*06001	100	LN886785	Angus cross
BoLA-Ank_06	1550	1199	3	BoLA-2*04701	100	LN886786	Charolais cross
BoLA-Ank_08	133	70	2	BoLA-2*04801	100	LN886787	Charolais cross
BoLA-Ank_15	376	281	2	BoLA-3*05301	100	LN886794	Charolais cross
BoLA-Ank_01	72	37	1	BoLA-1*04901	93.17	LN886781	Charolais cross
BoLA-Ank_19	434	1137	2	BoLA-1*06101	97.06	LN886798	Angus cross
BoLA-Ank_21	315	245	1	BoLA-1*02101	96.57	LN886799	Holstein
BoLA-Ank_03	49	37	1	BoLA-2*04801	94.85	LN886783	Charolais cross
BoLA-Ank_11	2290	2139	3	BoLA-2*06901	99.51	LN886790	Angus cross
BoLA-Ank_12	10	5	1	BoLA-2*03201N	98.28	LN886791	Holstein
BoLA-Ank_13	73	68	1	BoLA-2*07101	94.36	LN886792	Angus cross
BoLA-Ank_16	11	6062	2	BoLA-2*03201N	98.04	LN886795	Holstein
BoLA-Ank_17	3204	3457	2	BoLA-2*04701	98.53	LN886796	Charolais cross
BoLA-Ank_18	939	1007	2	BoLA-2*03001	93.87	LN886797	Boran
BoLA-Ank_27	11	15	2	BoLA-2*01602	94.12	LN886803	Charolais cross
BoLA-Ank_04	2014	1707	2	BoLA-3*05002	98.28	LN886784	Angus cross
BoLA-Ank_09	180	28	2	BoLA-3*01101	97.06	LN886788	Charolais cross
BoLA-Ank_10	80	11	1	BoLA-3*01101	96.32	LN886789	Charolais cross
BoLA-Ank_14	30	25	1	BoLA-3*05001	97.06	LN886793	Charolais cross
BoLA-Ank_26	857	76	2	BoLA-3*00403	95.34	LN886802	
BoLA-Ank_23	3024	424	1	BoLA-3*05301	97.3	LN886800	Charolais cross
BoLA-Ank_25	81	78	2	BoLA-6*01501	96.57	LN886801	Holstein
BoLA-Ank_07	83	20	3	BoLA-NC1*00301	100	DQ140370	Holstein
BoLA-Ank_24	111	32	1	BoLA-NC1*00601	100	JN792886	Angus cross
BoLA-Ank_20	760	271	2	BoLA-NC2*00102	99.75		Holstein
BoLA-Ank_22	104	87	1	BoLA-NC2*00102	99.51		Holstein

Supplementary table 1 Frequency distribution of putative alleles and unclassified variants across 17 Ankole individuals

(A) Exon 2 variants

Animal	Putative alleles																							Unclassified variants
	#1	#2	#3	#9	#778	#898	#146	#147	#148	#36	#37	#167	#168	#169	#170	#171	#10	#201	#81	#85	#92	#39	#111	#42	
Ankole-1	72	49	27	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
Ankole-2	0	0	0	0	0	0	0	0	0	125	97	0	0	0	0	0	0	0	0	0	0	0	0	0	4
Ankole-3	0	86	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	106	0	0	0	0	0	1
Ankole-4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	74	0	0	0	80	0	4
Ankole-5	0	13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Ankole-6	0	0	0	7	0	0	34	22	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Ankole-7	0	0	0	26	0	0	0	59	46	25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
Ankole-8	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0
Ankole-9	0	0	0	0	0	0	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
Ankole-10	0	0	0	0	0	0	19	0	0	0	0	10	4	2	0	0	0	0	0	0	0	0	0	0	0
Ankole-11	0	47	0	13	0	0	0	0	0	0	0	0	0	0	73	56	10	0	0	30	0	0	0	0	0
Ankole-12	0	0	0	0	0	0	0	0	0	0	0	3818	719	0	0	0	0	0	0	0	0	1473	0	0	10
Ankole-13	0	0	0	0	0	0	2262	0	0	0	1917	0	0	0	0	0	336	0	0	0	0	1484	0	0	1
Ankole-14	0	0	0	0	0	0	0	0	0	0	0	3125	647	437	0	0	0	315	0	0	387	1731	0	0	8
Ankole-15	0	0	0	0	3024	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Ankole-16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	852	0	0	0	0	0	0	0	0
Ankole-17	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	4	0

(B) Exon 3 variants

Animal	Putative alleles																								Unclassifiec variants
	#1	#2	#1542	#13	#16	#17	#18	#21	#42	#43	#44	#47	#69	#70	#972	#120	#86	#1529	#88	#91	#1372	#1530	#1528	#121	#122	
Ankole-1	37	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Ankole-2	0	0	0	0	53	40	24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5
Ankole-3	0	0	0	8	0	0	0	0	50	20	14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
Ankole-4	0	0	0	0	0	0	0	0	0	8	0	0	12	11	0	0	0	0	0	0	0	0	0	0	0	1
Ankole-5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Ankole-6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Ankole-7	0	0	0	0	0	14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
Ankole-8	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0
Ankole-9	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0
Ankole-10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	5	0	0	0	0	0	0	1
Ankole-11	0	0	0	0	0	0	0	12	20	0	25	22	0	0	0	0	0	0	0	68	0	0	0	0	0	1
Ankole-12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1735	0	0	3573	0	0	0	0	635	568	6
Ankole-13	0	0	0	0	1654	0	1149	269	0	0	0	0	0	0	0	0	2128	0	0	0	0	0	0	0	0	2
Ankole-14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1722	0	0	2489	0	0	0	0	372	569	5
Ankole-15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	424	0	0	0	0	2
Ankole-16	0	0	0	0	0	0	26	0	0	0	0	0	0	34	53	0	0	69	0	0	0	60	76	0	0	0
Ankole-17	0	0	6	0	0	0	0	0	0	0	0	0	0	0	7	0	0	7	0	0	0	0	0	0	0	2

Holstein class I MHC amplification, plasmid sub-cloning and Sanger sequencing

We characterized the class I MHC transcript profiles of 17 Holstein cattle by sequencing of plasmid sub-clones containing a 632 base pair amplicon spanning exons 2 to 4 of the class I MHC alpha chain. After testing for incorporation of the correct insert by PCR and EcoRI digestion, 10 positive plasmid sub-clones from each animal were Sanger sequenced using T7 and SP6 primers. We subjected the Sanger sequences to similar artifact and chimera filters as described for the Ankole data, but retained singletons if they were carried by more than one animal.

Our analysis indicated the existence of 23 distinct class I MHC sequences in the Holstein population genotyped in this study (Table 2). As anticipated, the call set was dominated by sequences displaying 100% nucleotide identity to allelic variants present in the definitive cattle MHC repository, the IPD, available at www.ebi.ac.uk/ipd/mhc/bola. The skew in the distribution of these known gene products over the class I MHC classical loci, characterized by a prominence for products of genes 1-3, is consistent with published analyses (Codner et al. 2012). In the absence of an identical match in the IPD database, the remaining sequences within the Holstein class I MHC dataset were conservatively designated as putatively novel if their existence was independently supported by plasmid sub-clone sequences from at least two animals. Two variants, working allele names BoLA-Hol_05 and BoLA-Hol_10 (Table 2), met this criterion. Interestingly BoLA Hol_10 exhibited only 93.64% sequence similarity to any cattle class I MHC molecule currently in the databases, suggesting that despite the relatively limited diversity within the Holstein breed and intensive class I MHC sequence analysis in recent years, the BoLA database is not yet complete. The two novel Holstein sequences have been submitted to the European nucleotide archive (LN8886781- LN8886805).

Table 2 Exonic sequence diversity present within the Holstein class I MHC genes

	Allele frequency		IPD BLASTn search			
Working allele name	No. of plasmid sub-clones	Individuals carrying allele	Allele Name	Percent identity	GenBank Accession	Breed known to express allele
BoLA-Hol_02	3	1	BoLA-1*00902	100	DQ121149	Charolais cross
BoLA-Hol_19	2	2	BoLA-1*01901	100	DQ140362	Holstein
BoLA-Hol_21	6	1	BoLA-1*02301	100	Y09208	Holstein
BoLA-Hol_07	2	2	BoLA-2*01201	100	DQ304655	Holstein
BoLA-Hol_09	1	1	BoLA-2*01802	100	DQ140361	Holstein
BoLA-Hol_11	4	2	BoLA-2*00802	100	L02835	Holstein
BoLA-Hol_04	19	4	BoLA-3*00201	100	X92870	Holstein
BoLA-Hol_08	3	1	BoLA-3*06601	100	JN792876	Angus cross
BoLA-Hol_14	3	1	BoLA-3*05301	100	DQ121175	Charolais cross
BoLA-Hol_15	3	2	BoLA-3*01701	100	Y09206	Holstein
BoLA-Hol_17	4	2	BoLA-3*00403	100	JN792860.1	Holstein
BoLA-Hol_20	3	1	BoLA-3*01101	100	M24090	Holstein
BoLA-Hol_12	3	1	BoLA-6*01402	100	DQ140358	Holstein
BoLA-Hol_05	5	3	BoLA-2*02201	98.04	LN886804	
BoLA-Hol_10	3	2	BoLA-2*00501	93.64	LN886805	
BoLA-Hol_06	4	1	BoLA-3*06501	97.56		
BoLA-Hol_22	1	1	BoLA-1*01901	99.76		
BoLA-Hol_16	4	4	BoLA-NC1*00301	100	DQ140370	Holstein
BoLA-Hol_18	4	3	BoLA-NC1*00501	100	JN792885	Angus cross
BoLA-Hol_01	3	3	BoLA-NC2*00102	100	DQ140372	Holstein
BoLA-Hol_03	6	5	BoLA-NC3*00101	100	DQ140373	Holstein
BoLA-Hol_23	1	1	BoLA-NC1*00701	98.58		
BoLA-Hol_13	1	1	BoLA-NC3*00101	99.27		

Departure from neutrality at the class I MHC locus in *Bos taurus* cattle

We used neighbor-joining (NJ) trees, calculated separately for the Ankole and Holstein class I MHC datasets according to the Jukes and Cantor (JC) algorithm, as the basis for assessing the relative fits of 56 distinct models of nucleotide substitution. For each model, likelihood scores were calculated in PAUP (Swofford, 2002) and the best-fit model for each dataset was selected using the Akaike Information Criterion. We found evolution of both datasets to be consistent with the model TVM+I+G, which assumes variation in base frequencies and transversion rates, as well as presence of invariable sites and rate heterogeneity among sites.

Likelihood settings from the best fit nucleotide substitution model were used in a codon-by-codon analysis to determine the ratio of nonsynonymous to synonymous evolutionary changes (dN/dS, ω) across each dataset. Specifically, we made maximum likelihood inferences of positive selection based on fitting two sets of codon evolution models and evaluating model fit using the AIC approach - a function of the logarithm of the likelihood (lnL), and the number of estimated parameters (K) for each model. The first codon evolution model set comprised the nearly neutral model (M1) and the positive selection model (M2). M1 allows for two site classes: p_0, proportion of codons undergoing purifying selection ($0<\omega_0<1$) and p_1, proportion of codons undergoing neutral evolution ($\omega_1=1$), while M2 extends M1 by additionally allowing for p_2, proportion of codons undergoing positive selection ($\omega_2>1$). The second model set assessed ω based on a β distribution indicated by shape parameters p and q. We implemented both the null β model (M7) which allows for $0<\omega_0<1$ and the positive selection plus β model M8, an extension of M7 to include a proportion of codons undergoing positive selection ($\omega_1>1$). The results of the analysis are shown in Table 3.

Table 3 The Akaike information criterion (AIC) evaluation of the goodness of fit for different models of codon evolution.

	Model fit				Estimated ω			Site classes			β distribution	
Breed	Model	LnL	K	AIC	*ω 0*	*ω 1*	*ω 2*	*p 0*	*p 1*	*p 2*	*p*	*q*
Ankole	M1	-2306.153228	46	4704.306456	0.06109	1		0.72527	0.27473			
	M2	-2232.131755	48	4560.26351	0.0605	1	7.2843	0.61741	0.28724	0.09536		
	M7	-2309.319323	46	4710.638646							0.14694	0.32814
	M8	-2233.607141	48	4563.214282			7.00833			0.09676	0.14597	0.28104
Holstein	M1	-2021.65299	34	4111.30598	0.04266	1		0.67663	0.32337			
	M2	-1976.487153	36	4024.974306	0.04347	1	6.5286	0.63003	0.30727	0.0627		
	M7	-2025.182837	34	4118.365674							0.11398	0.22844
	M8	-1979.071619	36	4030.143238			5.99023			0.07126	0.10506	0.20035

LnL, maximized log likelihoods; K, number of estimated parameters. The three site classes: purifying selection, neutral selection and positive selection are denoted by ω0, ω1 and ω2 respectively, while p0, p1 and p2 indicate the proportion of codons belonging to each of the three site categories respectively. β distribution shape parameters are denoted p and q.

For both datasets, M2 provided a better fit than M1 indicating that evolution at the class I MHC genes in cattle is driven primarily by positive selection. Independent support for this conclusion came from the models implementing the β distribution, with M8 providing better fit to both datasets than M7. Since the p_2 parameter for M8 is not substantially different from that of M2, unless explicitly stated, the positive selection remarks hereafter refer to estimates obtained under M2. For the Ankole and Holstein datasets, 9.536 % and 6.27 % of the codons carried historical signatures of positive selection respectively. Furthermore, ω values of 7.2843 and 6.5286 for Ankole and Holstein class I MHC datasets, respectively were substantially higher than 1. Similar estimates were obtained with M8 (Table 3). For codons identified as carrying signatures of positive selection, Bayes Empirical Bayes (BEB) analysis was used to assess their posterior probability of falling into this site class. For both datasets, the positively selected codons, their corresponding posterior probabilities, and their location relative to sites known to be critical for antigen recognition by cattle class I MHC (Hansen et al. 2014) are shown in Fig. 2.

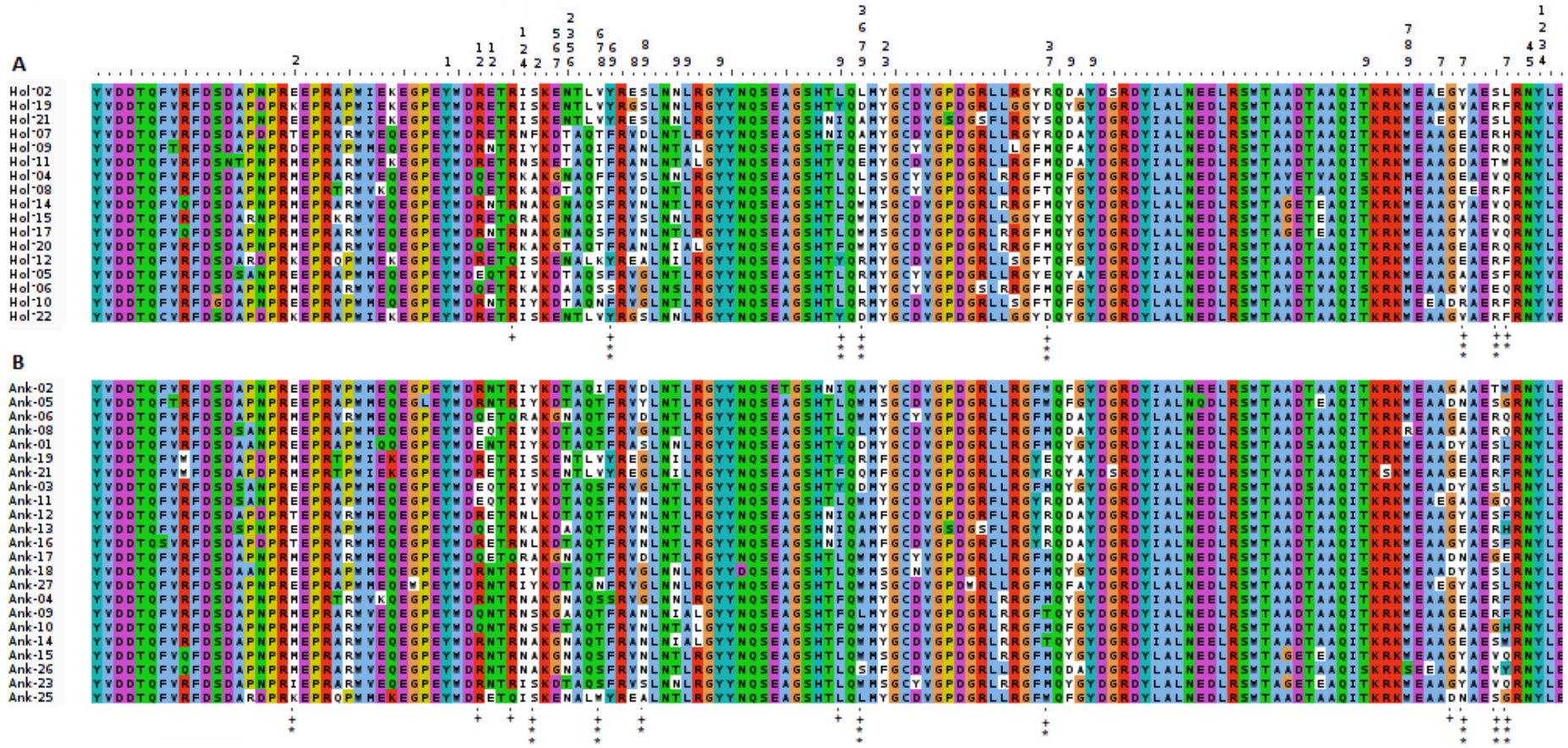


Fig. 2 Polymorphisms in amino acid sequence of the class I MHC peptide-binding regions (α_1 and α_2 domains) expressed in Holstein (A) and Ankole (B) cattle. In the top panel, 1-9 represent the class I MHC peptide ligand positions numbered from the amino to the carboxyl peptide termini, and mark the class I MHC residues that comprise the antigen binding groove (Hansen et al, 2014). Residue columns known to have the most influence on cattle class I MHC binding specificity (F pocket) are marked 9, while those with evidence of positive selection as revealed by Bayes Empirical Bayes analysis are indicated with a plus symbol (+). Single and double asterisks denote 95% and 99% posterior probabilities of falling into the positive selection site class respectively.

In silico analysis of peptide-binding motifs for Ankole and Holstein class I MHC proteins

Using the sequences described in the current study as input for *in silico* analysis, we sought to compare the predicted peptide binding motifs of the Ankole and Holstein class I MHC transcripts as well as investigate the extent to which the motifs overlap. Overlaps in the *NetMHCpan* predicted peptide-binding motifs, conveyed in a specificity heat-map and a functional tree, were assessed using *MHCcluster* (Fig. 3). A prominent feature of the results of these ligand selectivity analyses was the short specificity distances among the Ankole class I MHC sequences, indicated by their juxtaposition on the heat-map. Our results suggest considerable dissimilarity in the peptide binding preferences between the Ankole and Holstein class I MHC products, as indicated by the specificity distances between them on the heat-map. The differences in the predicted peptide binding preferences are further depicted as clusters in the functional tree, while the functional repertoire of Ankole class I MHC is exemplified by sequence logo representations of predicted binding motifs (Fig 4). Based on the height of the stack of amino acid residues, and as indicated by the cattle class I MHC allelic variants restricting known *T. parva* $CD8^+$ T-cell antigens (Hansen et al. 2014), scrutiny of the predicted specificity pattern for all the Ankole sequences revealed that besides the C- terminal residues, side chains at position 2 had the second highest relative contribution to peptide binding (Fig 4). Additional variations were predicted at other motif positions possibly resulting in additional functional divergence among the Ankole class I MHC products.

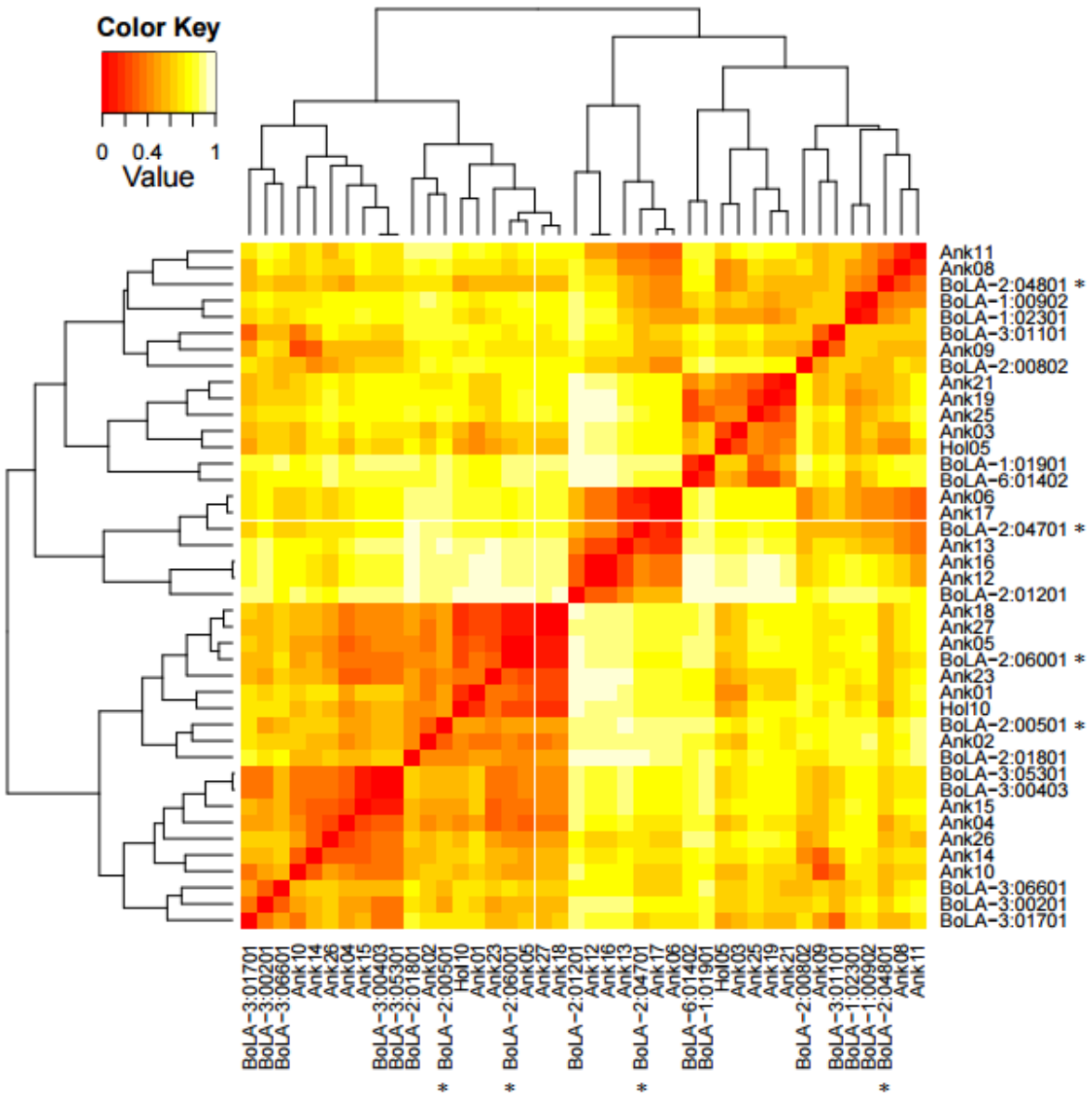

Color Key
0
0.4
1
Value
Ank11
Ank08
BoLA-2:04801 *
BoLA-1:00902
BoLA-1:02301
BoLA-3:01101
Ank09
BoLA-2:00802
Ank21
Ank19
Ank25
Ank03
Hol05
BoLA-1:01901
BoLA-6:01402
Ank06
Ank17
BoLA-2:04701 *
Ank13
Ank16
Ank12
BoLA-2:01201
Ank18
Ank27
Ank05
BoLA-2:06001 *
Ank23
Ank01
Hol10
BoLA-2:00501 *
Ank02
BoLA-2:01801
BoLA-3:05301
BoLA-3:00403
Ank15
Ank04
Ank26
Ank14
Ank10
BoLA-3:06601
BoLA-3:00201
BoLA-3:01701
BoLA-3:01701
BoLA-3:00201
BoLA-3:06601
Ank10
Ank14
Ank26
Ank04
Ank15
BoLA-3:00403
BoLA-3:05301
BoLA-2:01801
Ank02
BoLA-2:00501 *
Hol10
Ank01
Ank23
BoLA-2:06001 *
Ank05
Ank27
Ank18
BoLA-2:01201
Ank12
Ank16
Ank13
BoLA-2:04701 *
Ank17
Ank06
BoLA-6:01402
BoLA-1:01901
Hol05
Ank03
Ank25
Ank19
Ank21
BoLA-2:00802
Ank09
BoLA-3:01101
BoLA-1:02301
BoLA-1:00902
BoLA-2:04801 *
Ank08
Ank11

Fig. 3 Specificity heat-map visualization of the functional relationships between Ankole and Holstein class I MHC transcripts inferred in *MHCcluster*. Ankole alleles are either prefixed "ANK" if they are novel or marked with an asterisk (*) if they have been described previously. Similarly, novel Holstein sequences are prefixed "Hol", while the previous described ones are prefixed "BoLA" without the asterisk marks. Inset is the legend showing the color key used to create the heat-map and the corresponding MHC specificity distance value ranges.

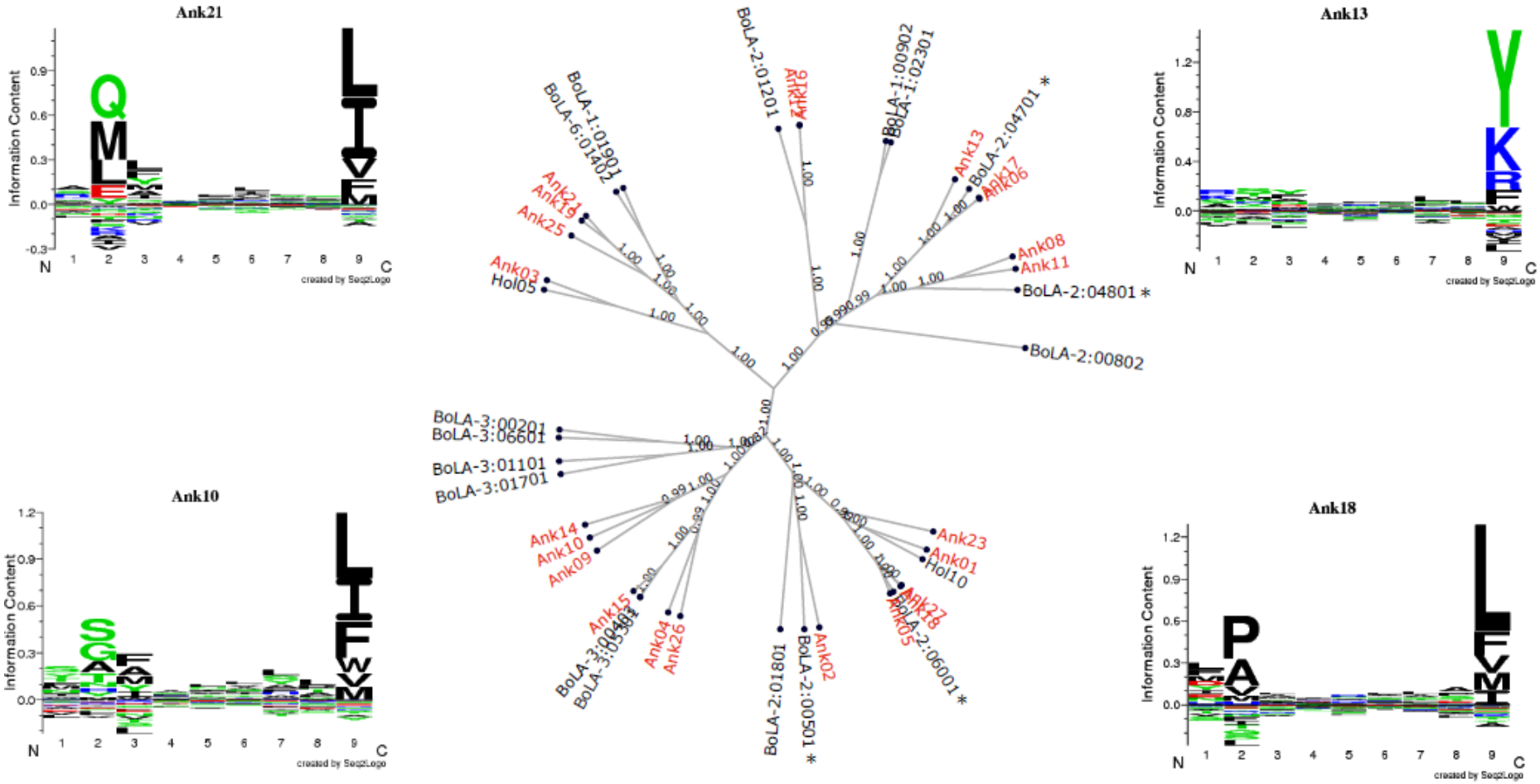


Fig. 4 Comparisons of predicted peptide-binding motifs for Ankole and Holstein class I MHC products. The *MHCcluster 2.0* procedure leverages the predictions generated by *NetMHCpan 2.8* to group the class I MHC molecules into a number of functional clusters. On the MHC functional distance tree, Ankole class I MHC transcripts are coloured red if they are novel or marked with an asterisk (*) if they have

been described previously. Novel Holstein sequences are prefixed “Hol”, while previous described ones are prefixed “BoLA” without an asterisk and are all coloured black. On the representative predicted binding motif sequence logos, Enriched amino acids (AA) are shown on the positive *y*-axis and depleted (under-represented) amino acids on the negative *y*-axis. The height of the stack of AA residues represents the level of conservation at a particular position, while the relative residue height denotes its frequency. Polar AA are depicted in green, basic AA in blue, acidic AA in red and hydrophobic AA in black.

Recognition of *T. parva* candidate vaccine antigens in seropositive Ankole cattle

Studies using immune bovine $CD8^+$ T-cell lines from Holstein-Friesian cattle and the Boran breed of *Bos indicus* to screen expressed parasite cDNA libraries have identified *T. parva* candidate vaccine antigens, designated Tp (Graham et al. 2006). Using these antigens as stimuli, we assessed their recognition in *T. parva* seropositive Ankole cattle by measuring cytokine (IFN-γ) responses. As a pre-requisite, we first determined the titers of *T. parva* specific antibodies in Ankole sera by quantitative antibody-detection ELISA based on the PIM antigen (Katende et al. 1998). The result of this analysis, expressed as a percentage of the strong positive control serum, hereafter referred to as percent positivity (PP), is shown in Fig. 5. Eleven of the 40 Ankole from the field screened were found to be antibody positive for *T. parva.*

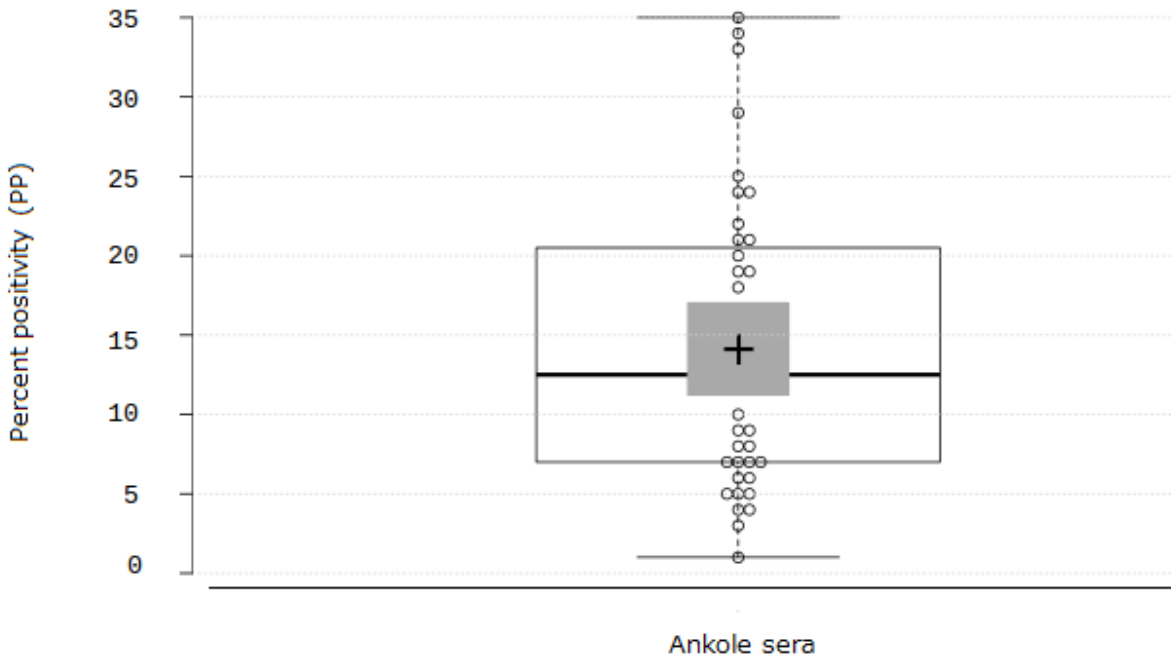


Fig. 5 Sero-prevalence of *T. parva* in field Ankole cattle. The center line show the medians; box limits indicate the 25th and 75th percentiles as determined by R software; whiskers extend 1.5 times the interquartile range from the 25th and 75th percentiles; the cross represents the sample mean; the grey bar indicate 95% confidence intervals of the mean; data points are plotted as open circles. Results are given as a percentage of the strong positive control serum (percent positivity, PP). Threshold PP = 20 and n = 40 Ankole cattle.

Detection and enumeration of IFN-γ secreting cells in the peripheral circulation of *T. parva* seropositive Ankole cattle was assessed *ex vivo* by ELISpot upon stimulation of PBMCs from the eleven seropositive cattle, with overlapping peptide pools derived from each of the defined *T. parva* candidate vaccine antigens (Graham et al. 2006). The breadth and magnitude of this cytokine response for each of the eleven positive animals are shown in Fig. 6 as spot-forming units (SFU) per million PBMCs.

The no peptide control mean numbers of IFN-γ SFU per 1.0 x 10^6 PBMCs were generally markedly less than those obtained from $CD8^+$ T-cell lines (derived from animal BV115) with known Tp1 antigenic specificity that served as a positive control for the assay. The PBMCs from animals Ankole-2 and Ankole-17, were positive in repeat experiments for peptide pools derived from antigens Tp4 and Tp5, respectively (see Fig. 6). However, when the *ex vivo* IFN-γ ELISpot assays were repeated using $CD8^+$ T-cell fractions purified from the PBMCs, the responses from all the Tp antigens were markedly lower than the BV115 positive control, and none attained the empirical threshold of three times mean SFU in Tp stimulated wells (Janetzki et al. 2015) compared to negative control wells (data not shown).

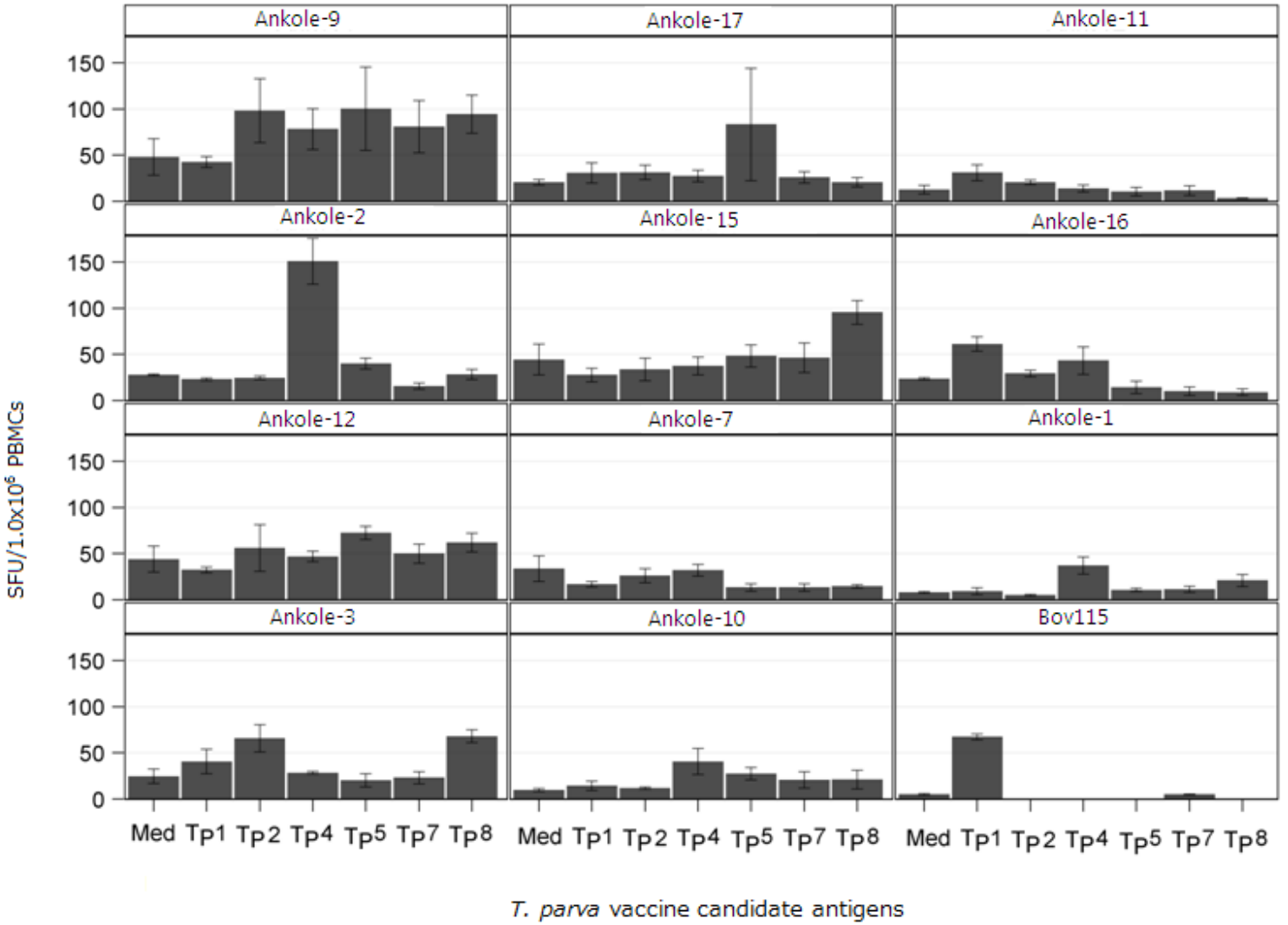


Fig. 6 Cellular immune recognition of *T. parva* vaccine candidate antigens by seropositive Ankole cattle as determined by an *ex vivo* whole PBMC IFN-γ ELISpot assay. Results are expressed as mean numbers of IFN-γ spot-forming units (SFU) per $1.0x10^6$ PBMCs. Assay functionality controls comprised of: (i) the non-specific mitogen con A; (ii) background (no peptide) medium control; and (iii) CTLs from BV115 - an animal with demonstrated reactivity to Tp1 but not Tp7. Pooled overlapping peptides that span the full length of the defined *T. parva* vaccine candidate antigens were used at a final concentration of 1 µg/ ml in three replicate wells. n = 11 *T. parva* seropositive Ankole cattle.

Discussion

Whilst the expressed bovine class I MHC gene repertoire has been examined extensively in European taurines in a number of studies, published analyses of class I MHC transcripts in African cattle is limited (IPD; www.ebi.ac.uk/ipd/mhc/bola). Consequently, there is still uncertainty as to the likely impacts of class I MHC diversity on the coverage of *T. parva* $CD8^+$ T-cell epitope based vaccines in endemic regions. The current study utilized cDNA from a population of African taurines (Ankole) as templates for the amplification of the domains of the class I MHC heavy chain that interact with antigenic peptides, in order to provide an initial assessment of the levels of class I MHC diversity among African taurine cattle. The amplicon libraries were typed by pyrosequencing using the Roche 454 GS-FLX chemistry, an approach that has proved useful in a wide range of situations, notably cataloguing of the complement of class I MHC transcripts in organisms with high levels of complexity at the class I MHC locus (O'Leary et al. 2009).

Whilst the catalogue of Holstein class I MHC transcripts obtained in the current study was in general agreement with published data (www.ebi.ac.uk/ipd/mhc/bola), the results of our evaluation of the Ankole transcripts pointed to minimal overlap in the ligand selectivity of the class I MHC products between the two breeds. Foremost, our data suggest a high level of inter-taurine sequence divergence, with the great majority of the Ankole class I MHC allelic sequences described in the present study being designated as putatively novel (Table 1). Notably, the nucleotide BLAST searches of the cattle section of the IPD revealed that in general, the Ankole class I MHC allelic variants shared higher nucleotide identities with sequences that are not known to be expressed in the extensively studied Holstein breed. In contrast, using the Holstein allelic sequences described herein as the query sequences, BLASTN similarity searches showed that with few exceptions, one of which was dramatically distinct in sequence, the African Holstein sequences shared 100% sequence similarity to already published cattle class I MHC sequences (Table 2)

The results of the codon-based statistical tests that we performed separately on the Ankole and Holstein class I MHC datasets, revealed patterns of evolution that markedly exceeded those predicted by neutrality. Positive selection of amino acid diversity was evident from the highly contrasting levels of nonsynonymous versus synonymous substitutions and the list of codons identified as carrying signatures of positive selection included those that code for residues that line the functionally important F pockets. Taken together, the evidence that the polymorphisms are generated and maintained by positive selection for amino acid diversity, and the concordance between the location of positively selected sites and known antigen binding sites are consistent with potential differences in the spectrum of *T. parva* epitopes recognized by $CD8^+$ T-cells from the two breeds. This prediction is consistent with the results of a previous study that demonstrated that only one of the eight *T. parva* $CD8^+$ T-cell target antigens, designated Tp2, was recognized by T cells from animals of different MHC backgrounds (Graham et al. 2008). Similarly, a separate study utilizing experimental peptide-binding data generated by PSCPL analysis coupled with a high-throughput peptide–class I MHC dissociation assay (Harndahl et al. 2009) found limited evidence for cross binding of known *T. parva* epitopes to different MHC genotypes (Hansen et al. 2014). In addition to the role of class I MHC diversity in the context of recognition by $CD8^+$ T-cells, there is data strongly suggesting that other selective forces also contribute to diversity of class I MHC loci in cattle and significant associations have been demonstrated between MHC haplotypes and production and fertility traits (Codner et al. 2011).

Despite the marked differences in exonic sequence diversity within the peptide binding regions of class I MHC allelic sequences expressed in Ankole and Holstein cattle, allelic polymorphism *per se* are not necessarily sufficient to infer functional differences. A variety of algorithms have been used for prediction of peptide binding to class I MHC and a tool called *MHCcluster* that functionally clusters MHC molecules based on their predicted binding specificity proved particularly useful when applied to our analysis of Ankole and Holstein class I sequences. The classification of allelic sequences based on

their predicted specificity distances, as indicated by different intensities of the colors in a heat-map (see Fig. 3), resulted in the juxtaposition of the Ankole alleles, while the Holstein alleles formed separate functional specificity clusters. Although a comprehensive definition of cattle class I MHC 'supertypes' is beyond the scope of the current study, it may be reasonable to conclude that the observed differences in the predicted ligand selectivity might result in differential peptide recognition by $CD8^+$ T-cells from the two breeds.

More directly relevant to ECF vaccine development is the question of whether the dissimilarities seen in *in silico* class I MHC ligand predictions are sustained in *ex vivo* tests assaying differential cytokine responses to defined *T. parva* antigen between the two taurine lineages. To address this, we employed pools of peptides derived from *T. parva* vaccine candidate antigens selected by immunoscreening of *T. parva* cDNA libraries using Holstein and Boran cattle immunized by ITM with the *T. parva* Muguga stock (Graham et al. 2008). These peptides were utilized to stimulate PBMCs and also purified $CD8^+$ T-cells isolated from *T. parva* seropositive Ankole cattle and assayed by gamma interferon ELISpot. Two of the eleven seropositive Ankole cattle were reproducibly positive for antigens Tp4 and Tp5, respectively. However, cytokine responses could not be demonstrated using purified Ankole $CD8^+$ T-cells. It seems possible that the Tp4 and Tp5 ELISpot responses may have been the result of $CD4^+$ T-cell responses. Whether polymorphisms within the target antigens in the Ankole population or the time point of infection in the field Ankole cattle contributed to the negative outcome of the IFN-γ ELISpot assay using purified $CD8^+$ T-cells remains to be demonstrated. However the failure of purified $CD8^+$ T-cells from the Ankole cattle sampled in this study to recognize antigens previously shown to contain $CD8^+$ T-cell epitopes recognized in the context of other bovine haplotypes, is worthy of further investigation.

Our findings, highlighting inter-taurine differences in class I MHC diversity and functional capability, are also broadly consistent with, and extend, findings from studies that examined recognition of the same *T. parva* candidate vaccine antigens by $CD8^+$ T-cells derived from Zebu (*Bos indicus*) cattle (Akoolo et al. 2008). With the exception of the Tp2 antigen, this study showed that the known *T. parva* $CD8^+$ T-cell target candidate vaccine antigens were not recognized by $CD8^+$ T-cells from any of the animals studied. With respect to Tp2, it concluded that this *T. parva* candidate antigen constituted only a minor $CD8^+$ T-cell specificity in the single animal that appeared to be responding and importantly, noted that the target Tp2 epitope was distinct from those previously identified using $CD8^+$ T-cells derived from *Bos taurus* cattle.

The data that we report herein reveals a surprising degree of functional divergence within the class I MHC of two different but relatively closely related taurine lineages and emphasizes the value of the *MHCcluster* software tool relative to simple comparison of sequences. Given these findings, it will be important to apply similar approaches, more comprehensively to the Ankole breed and also to a wider range of African cattle populations. This will contribute to understanding whether class I 'supertypes' can be defined for African cattle and the number of antigens/epitopes that may need to be incorporated in a recombinant antigen cocktail vaccine that is effective for control of ECF in the field.

Acknowledgments

We are grateful to Dr. Joerg Jores of ILRI for reviewing the manuscript. We also wish to thank Giles Prettejohn of the Ol Pejeta Conservancy in Kenya whose efforts were instrumental to the completion of this study. The study was funded by the DFG German-African Cooperation Projects in Infectiology: "Molecular epidemiology network for promotion and support of delivery of life vaccines against *Theileria parva* and *Theileria annulata* infection in Eastern and Northern Africa (SE 862/2-1)". Some of the work described in this paper was also supported by the CGIAR Consortium research project CRP3.7.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol. 215(3):403-10

Akoolo L, Pellé R, Saya R, Awino E, Nyanjui J, Taracha EL, Kanyari P, Mwangi DM, Graham SP (2008) Evaluation of the recognition of *Theileria parva* vaccine candidate antigens by cytotoxic T lymphocytes from Zebu cattle. Vet Immunol Immunopathol. 121(3-4):216-21

Anderung C, Hellborg L, Seddon J, Hanotte O, Götherström A (2007) Investigation of X- and Y-specific single nucleotide polymorphisms in taurine (Bos taurus) and indicine (Bos indicus) cattle. Anim Genet. 38(6):595-600

Babik W, Kawalko A, Wojcik JM, Radwan J. (2012) Low major histocompatibility complex class 1 (MHC I) variation in the European Bison (Bison bonasus). Journal of Heredity 103: 349-359

Birch J, Murphy L, MacHugh ND, Ellis SA (2006) Generation and maintenance of diversity in the cattle MHC class I region. Immunogenetics. 58(8):670-9

Codner GF, Birch J, Hammond JA, Ellis SA (2012) Constraints on haplotype structure and variable gene frequencies suggest a functional hierarchy within cattle MHC class I. Immunogenetics. 64(6):435-45

Codner GF, Stear MJ, Reeve R, Mathews, L, Ellis SA (2011) Selective forces shaping diversity in the class I region of the major histocompatibility complex in dairy cattle. Animal Genetics 43: 239-249

Di Giulio G, Lynen G, Morzaria S, Oura C, Bishop R (2009) Live immunization against East Coast fever--current status. Trends Parasitol. 25(2):85-92

Di Palma F, Archibald SD, Young JR, Ellis SA (2002) A BAC contig of approximately 400 kb contains the classical class I major histocompatibility complex (MHC) genes of cattle. Eur J Immunogenet. 29(1):65-8

Dodt M, Roehr J, Ahmed R, Dieterich C (2012) FLEXBAR—flexible barcode and adapter processing for next-generation sequencing platforms. Biology. 1(3):895–905

Ellis SA, Holmes EC, Staines KA, Smith KB, Stear MJ, McKeever DJ, MacHugh ND, Morrison WI (1999) Variation in the number of expressed MHC genes in different cattle class I haplotypes. Immunogenetics. 50(5-6):319-28

Ellis S (2004) The cattle major histocompatibility complex: is it unique? Vet Immunol Immunopathol. 102(1-2):1-8

Emery DL, Eugui EM, Nelson RT, Tenywa T (1981) Cell-mediated immune responses to *Theileria parva* (East Coast fever) during immunization and lethal infections in cattle. Immunology. 43(2):323-36

Eugui EM, Emery DL (1981) Genetically restricted cell-mediated cytotoxicity in cattle immune to *Theileria parva*. Nature. 290(5803):251-4

Graham SP, Pellé R, Honda Y, Mwangi DM, Tonukari NJ, Yamage M, Glew EJ, de Villiers EP, Shah T, Bishop R, Abuya E, Awino E, Gachanja J, Luyai AE, Mbwika F, Muthiani AM, Ndegwa DM, Njahira M, Nyanjui JK, Onono FO, Osaso J, Saya RM, Wildmann C, Fraser CM, Maudlin I, Gardner MJ, Morzaria SP, Loosmore S, Gilbert SC, Audonnet JC, van der Bruggen P, Nene V, Taracha EL (2006) *Theileria parva* candidate vaccine antigens recognized by immune bovine cytotoxic T lymphocytes. Proc Natl Acad Sci U S A. 103(9):3286-91

Graham S, Honda Y, Pelle r, Mwangi DM, Glew EJ, DeVilliers EP, Shah T, Bishop R, van der Bruggen P, Nene V, Taracha ELN (2007) A novel strategy for the identification of antigens that are recognised by Bovine MHC class I restricted cytotoxic T cells in a protozoan infection using reverse vaccinology. Immunome Res. 3: 2

Graham SP, Pellé R, Yamage M, Mwangi DM, Honda Y, Mwakubambanya RS, de Villiers EP, Abuya E, Awino E, Gachanja J, Mbwika F, Muthiani AM, Muriuki C, Nyanjui JK, Onono FO, Osaso J, Riitho V, Saya RM, Ellis SA, McKeever DJ, MacHugh ND, Gilbert SC, Audonnet JC, Morrison WI, van der Bruggen P, Taracha EL (2008) Characterization of the fine specificity of bovine CD8 T-cell responses to defined antigens from the protozoan parasite *Theileria parva*. Infect Immun. 76(2):685-94

Hansen AM, Rasmussen M, Svitek N, Harndahl M, Golde WT, Barlow J, Nene V, Buus S, Nielsen M (2014) Characterization of binding specificities of bovine leucocyte class I molecules: impacts for rational epitope discovery. Immunogenetics. 66(12):705–18

Harndahl M, Justesen S, Lamberth K, Roder G, Nielsen M, Buus S (2009) Peptide binding to HLA class I molecules: homogenous, high-throughput screening, and affinity assays. J Biomol Screen. 14:173–180

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

Hoof I, Peters B, Sidney J, Pedersen LE, Sette A, Lund O, Buus S, Nielsen M (2009) NetMHCpan, a method for MHC class I binding prediction beyond humans. Immunogenetics. 61(1):1-13

Janetzki S, Price L, Schroeder H, Britten CM, Welters MJ, Hoos A (2015) Guidelines for the automated evaluation of Elispot assays. Nat Protoc. 10(7):1098-115

Katende J, Morzaria S, Toye P, Skilton R, Nene V, Nkonge C, Musoke A (1998) An enzyme-linked immunosorbent assay for detection of *Theileria parva* antibodies in cattle using a recombinant polymorphic immunodominant molecule. Parasitol Res. 84(5):408-16

Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 30(4):772-80

Obara I, Ulrike S, Musoke T, Spooner PR, Jabbar A, Odongo D, Kemp S, Silva JC, Bishop RP (2015) Molecular evolution of a central region containing B cell epitopes in the gene encoding the p67 sporozoite antigen within a field population of *Theileria parva*. Parasitol Res. 114(5):1729-37

O'Leary CE, Wiseman RW, Karl JA, Bimber BN, Lank SM, Tuscher JJ, O'Connor DH (2009) Identification of novel MHC class I sequences in pig-tailed macaques by amplicon pyrosequencing and full-length cDNA cloning and sequencing. Immunogenetics. 61(10):689-701

MacHugh DE, Shriver MD, Loftus RT, Cunningham P, Bradley DG (1997) Microsatellite DNA variation and the evolution, domestication and phylogeography of taurine and zebu cattle (Bos taurus and Bos indicus). Genetics. 146(3):1071-86

McKeever DJ, Taracha EL, Innes EL, MacHugh ND, Awino E, Goddeeris BM, Morrison WI (1994) Adoptive transfer of immunity to *Theileria parva* in the $CD8^+$ fraction of responding efferent lymph. Proc Natl Acad Sci U S A. 91(5):1959-63

Musoke AJ, Nantulya VM, Buscher G, Masake RA, Otim B (1982) Bovine immune response to *Theileria parva*: neutralizing antibodies to sporozoites. *Immunology* 45:663–68

Nene V, Svitek N, Toye P, Golde WT, Barlow J, Harndahl M, Buus S, Nielsen M (2012) Designing bovine T cell vaccines via reverse immunology. Ticks Tick Borne Dis. 3(3):188-92

Nene V, Gobright E, Bishop R, Morzaria S, Musoke A (1999) Linear peptide specificity of bovine antibody responses to p67 of *Theileria parva* and sequence diversity of sporozoite-neutralizing epitopes: implications for a vaccine. Infect Immun 67:1261–1266

Nielsen M, Lundegaard C, Blicher T, Lamberth K, Harndahl M, Justesen S, Røder G, Peters B, Sette A, Lund O, Buus S (2007) NetMHCpan, a method for quantitative predictions of peptide binding to any HLA-A and -B locus protein of known sequence. PLoS One. 2(8):e796

Norval RAI, Perry BD, Young AS (1992) The epidemiology of Theileria in Africa. London: Academic

Paling RW, Mpangala C, Luttikhuizen B, Sibomana G (1991) Exposure of Ankole and crossbred cattle to Theileriosis in Rwanda. Trop Anim Health Prod. 23(4):203-14

Pandya M, Rasmussen M, Hansen A, Nielsen M, Buus S, Golde W, Barlow J (2015) A modern approach for epitope prediction: identification of foot-and-mouth disease virus peptides binding bovine leukocyte antigen (BoLA) class I molecules. Immunogenetics. Epub ahead of print

Pedersen LE, Harndahl M, Rasmussen M, Lamberth K, Golde WT, Lund O, Nielsen M, Buus S (2011) Porcine major histocompatibility complex (MHC) class I molecules and analysis of their peptide-binding specificities. Immunogenetics. 63(12):821-34

Radley DE, Brown CGD, Burridge MJ, Cunningham MP, Kirimi IM, Purnell RE, Young AS (1975) Chemoprophylactic immunisation of cattle against *Theileria parva* (Muguga) and five Theileria strains. Vet Parasitol. 1:35–41

Robinson J, Halliwell JA, McWilliam H, Lopez R, Marsh SG (2013) IPD--the Immuno Polymorphism Database. Nucleic Acids Res. 41(Database issue):D1234-40

Sette A, Sidney J (1999) Nine major HLA class I supertypes account for the vast preponderance of HLA-A and -B polymorphism. Immunogenetics. 50(3-4):201-12

Sommer S, Courtiol A, Mazzoni CJ (2013) MHC genotyping of non-model organisms using next-generation sequencing: a new methodology to deal with artefacts and allelic dropout. BMC Genomics. 14:542

Steinaa L, Saya R, Awino E, Toye P (2012) Cytotoxic T lymphocytes from cattle immunized against *Theileria parva* exhibit pronounced cross-reactivity among different strain-specific epitopes of the Tp1 antigen. Vet Immunol Immunopathol. 145(3-4):571-81

Svitek N, Hansen AM, Steinaa L, Saya R, Awino E, Nielsen M, Buus S, Nene V (2014) Use of 'one-pot, mix-and-read' peptide-MHC class I tetramers and predictive algorithms to improve detection of cytotoxic T lymphocyte responses in cattle. Vet Res. 45:50

Swofford DL (2002) PAUP: phylogenetic analysis using parsimony and other methods. Version 4. Sinauer Associates, Sunderland

Thomsen M, Lundegaard C, Buus S, Lund O, Nielsen M (2013) *MHCcluster*, a method for functional clustering of MHC molecules. Immunogenetics.

4.0 Chapter 4: Publication II

This chapter evaluates the hypothesis that induction of widely effective humoral immunity in cattle against the infective stage of *T. parva* may be constrained by the dynamic nature of parasite populations in the field particularly where the cape buffalo wildlife reservoir of *T. parva* co-grazes. The chapter presents the findings from a study that characterized heterogeneity in the p67 gene among field isolates of *T. parva* originating from cattle that were part of a field trial of ITM vaccines designed to evaluate protection afforded to immunized animals that received *T. parva* challenge from buffalo-associated ticks.

Ten discrete p67 allelic sequences were identified from the samples examined, with an overall level of DNA polymorphism of 19.6%. This contrasts strongly with the complete conservation of nucleotides at the p67 locus, including introns and the third base in codons, in cattle-transmissible parasites. Allelic sequences varied considerably in size principally due to the 129 and 174 base pair deletions in the central region of the p67 gene. An alignment of predicted allelic amino acid sequences corresponding to two closely juxtaposed epitopes in the central region of the p67 protein revealed both in-frame deletions and nonsynonymous substitutions resulting in a truncated protein and relatively low overall amino acid conservation respectively. In sum, these findings suggest that a p67 based vaccine suitable for field use should incorporate the polymorphic epitopes to enhance the protection conferred by the current conserved p67 vaccine antigen.

The findings from this study also offer insights into the evolution of the *T. parva* p67 locus. In particular, the chapter presents data indicating that (i) the majority of nucleotide substitutions at the p67 locus results from transitions (mutations that conserve the nucleotide base chemistry) rather than transversions (mutations that substitute purines for pyrimidines, or vice-versa), (ii) the codon substitution model that captures the proportion of codons undergoing positive selection provides a

better fit to the p67 data than the one based on neutral evolution indicating an evolutionary patterns that is consistent with the positive selection.

The chapter concludes that *T. parva* parasites originating from buffalo, with diverse p67 genotypes, can infect and cause severe disease in co-grazing cattle and therefore careful consideration will be required in the design of p67 based vaccines in areas where buffaloes are present. The findings also highlight the fact that buffalo- derived *T. parva* can sometimes break through the immunity induced by ITM immunization and induce severe clinical reactions, in areas with a high tick challenge. The sequences described in this report are available from the European Nucleotide Archive at http://www.ebi.ac.uk/ena/data/view/LK054504-LK054513.

The scientific findings from the analyses conducted in this chapter have been subjected to independent peer review and published as an original open access paper in Parasitology Research. The final publication is available at http://link.springer.com/article/10.1007%2Fs00436-015-4358-6. An author-created version of the published article appears overleaf.

Molecular evolution of a central region containing B-cell epitopes in the gene encoding the p67 sporozoite antigen within a field population of *Theileria parva*

Obara Isaiah [1*], Ulrike Seitzer[2+], Musoke Tony[1], Spooner Paul R [1], Jabbar Ahmed [2], Odongo David [1], Kemp Stephen[1], Joana C. Silva[3], Richard P. Bishop [1]

(1) International Livestock Research Institute (ILRI), P.O. Box 30709 Nairobi 00100, Kenya

(2) Veterinary Infection Biology and Immunology Research Center Borstel, Parkallee 22, 23845 Borstel, Germany

(3) Institute for Genome Sciences and Department of Microbiology and Immunology, University of Maryland School of Medicine, Baltimore21201, Maryland, United States of America

[+] Deceased

*Corresponding author, E-mail: i.obara@cgiar.org , Tel: +254 724 379 843

This is an author-created version of a peer-reviewed journal article published as an original open access paper in Parasitology Research. The final publication is available at http://link.springer.com/article/10.1007%2Fs00436-015-4358-6

DOI: 10.1007/s00436-015-4358-6, May 2015, Volume 114, Issue 5, pp 1729-1737

Abstract

Protective immunity induced by the infective sporozoite stage of *Theileria parva* indicates a potential role for antibodies directed against conserved serologically reactive regions of the major sporozoite surface antigen *p67* in vaccination to control the parasite. We have examined the allelic variation and determined the extent of B-cell epitope polymorphism of the gene encoding *p67* among field isolates originating from cattle exposed to infected ticks in the Marula area of the rift valley in central Kenya where the African cape buffalo (*Syncerus caffer*) and cattle co-graze. In the first of two closely juxtaposed epitope sequences in the central region of the p67 protein, an in-frame deletion of a 7 amino acid segment results in a truncation that was observed in parasites derived from cattle that co-grazed with buffalo. In contrast, the variation in the second epitope was primarily due to non-synonymous substitutions, resulting in relatively low overall amino acid conservation in this segment of the protein. We also observed polymorphism in the region of the protein adjacent to the two defined epitopes but this was not sufficient to provide statistically significant evidence for positive selection. The data indicates that B-cell epitopes previously identified within the *p67* gene are polymorphic within the Marula field isolates. Given the complete sequence identity of the *p67* gene in all previously characterised *T. parva* isolates that are transmissible between cattle by ticks, the diversity observed in *p67* from the Marula isolates in combination with the clinical reaction of the infected cattle is consistent with them originating from ticks that had acquired *T. parva* from buffalo.

Introduction

Theileria parva is an apicomplexan protozoan parasite of cattle that is transmitted by the tick *Rhipicephalus appendiculatus*, and is the causal agent of East Coast Fever, a frequently fatal disease of cattle in eastern and southern Africa (Norval et al. 1992). The tick inoculates T. *parva* sporozoites that are infective to the mammalian host and these differentiate rapidly into an intracellular schizont that immortalizes bovine lymphocytes, subsequently *T.parva* differentiates into a piroplasm stage that is infective to erythrocytes. Transmission to cattle can originate from ticks that have fed on African Cape buffalo (S*yncerus caffer*), the major wildlife reservoir of *T. parva*, in which case there is a different clinical syndrome, known as 'corridor disease' involving low levels of schizont parasitosis and piroplasm parasitaemia. Transmission by ticks that have previously fed on other cattle results in typical East Coast Fever symptoms with higher levels of parasites. Previous studies using both variable number tandem repeats (VNTRs) and antigen genes have revealed genetic differences between parasites transmissible among cattle and those found in buffalo (Oura et al. 2011; Pelle et al. 2011), suggesting that the two sets of parasites represent distinct populations.

Theileria parva has been the subject of more than 40 years of research aimed at development of a recombinant vaccine using antigens derived from both the sporozoite and schizont stages (reviewed in Morrison 2009; McKeever et al. 1999). The closest approach to an effective anti-sporozoite vaccine to date has used a recombinant version of p67, the major sporozoite surface antigen of *T. parva*. The p67 protein induces a consistent level of 70% efficacy against either heterologous or homologous needle challenge using sporozoite stabilates in the laboratory (Musoke et al. 1992; Bishop et al. 2003) and reduced severe disease by approximately 45% in field trials where cattle were exposed to infective ticks (Musoke et al. 2005). Five regions of the p67 protein sequence that are reactive with anti-sporozoite monoclonal antibodies and also correlate with the results of an in vitro parasite-neutralizing

assay have been identified using PepScan analysis (Nene et al. 1999). The central region of the *p67* gene includes two closely linked epitopes that are the target of host B cell responses, and whose sequences in the *T. parva* Muguga reference stock are ^{169} TKEEVPPADLSDQVP ^{183} and ^{209} LQPGKTS ^{215}. These are subsequently referred to as epitopes 1 and 2, respectively.

While the *p67* gene is variable among buffalo-derived isolates, the predicted p67 protein appears to be invariant in cattle-derived stocks of *T. parva* that are transmissible between cattle by ticks (Nene et al. 1996; Nene et al. 1999; Musoke et al. 2005). A method of live vaccination known as infection and treatment (ITM) was developed approximately 40 years ago (Radley et al. 1975**)** and there is evidence that the protection induced by ITM is attributable primarily to class I MHC-restricted $CD8^+$ cytotoxic T cells (McKeever et al. 1994). However, this cytotoxic T cell response is strain-specific and strongly dependent on the bovine class I MHC phenotype of the host (Taracha et al. 1995). This may constrain development of a broadly cross-protective recombinant vaccine that mimics the cellular responses induced by ITM and highlights the potential importance of the conservation of the p67 antigen in cattle-derived *T. parva*. Extensive sequence divergence in the central region of the *p67* gene in buffalo derived parasites from South Africa has recently been described (Sibeko et al. 2010). However, the p67 vaccine has not yet been tested in areas where the parasite challenge is mainly from *T. parva* originating from buffalo. We focus in the study described herein on in-depth analysis of p67 B-cell epitope polymorphism in cattle-infective isolates from a specific geographical locality in central Kenya where buffalo and cattle co-graze. In addition, we examine whether any codons show signatures of positive selection in the central region of the *p67* gene.

Materials and Methods

Parasite isolates and genomic DNA extraction

Genomic DNA preparations were made from eighteen cryopreserved pellets of 10^7 *T. parva* schizont-infected lymphocyte cultures initially isolated from cattle that co-grazed with the African Cape buffalo, using the DNeasy® Tissue Kit (Qiagen, Germany) according to the manufacturer's instructions. The cattle were part of a field trial of ITM vaccines performed in the year 2000 to explore protection afforded to immunized animals that received a *T. parva* challenge from buffalo-associated ticks at Marula farm, central Kenya (Pelle et al. 2011). The trial was carried out in strict accordance with the recommendations in the standard operating procedures of the ILRI's Institutional Animal Care and Use Committee (IACUC). Fifty three of the 113 tick-exposed cattle developed clinical disease and died; mortality was observed in 40 immunized animals and 13 of the control cattle. Most animals exhibited clinical and parasitological features typical of those induced by buffalo-derived *T. parva* (Norval et al. 1992) with a low schizont parasitosis and low, or in some cases, no piroplasm parasitaemia. As in table 1, among the eighteen cattle from which schizont-infected lymphocyte cultures were generated from lymph node biopsies were non-immunised controls, as well as those immunised with one of the following stabilates: *T. parva* Marikebuni stabilate 3014 (Morzaria et al. 1995); *T. parva* Marikebuni stabilate 316 (Payne 1999) and *T. parva* composite trivalent Muguga cocktail stabilate FAO 1 (Morzaria et al. 1999). All cattle were monitored daily from day seventeen after exposure and the clinical reactions are summarized in table 1.

Table 1 Classification of *T. parva* p67 alleles based on indels and B cell epitope sequence variation

T.parva strain	Animal No.	Vaccine stabilate	p67 sequence Accession	Antibody epitope		Deletion (bp)		Allele type
				Epitope 1	Epitope 2	129	174	
Muguga			XM_758212.1	TKEEVPPADLSDQVP	LQPGKTS	+	-	
KNP2			AF079177.1	.....	.....	+	-	1
Marula	N16(SRD)	316	LK054513	.....	.....	+	-	
Marula	N32(FD)	FAO 1						
Buffalo			U40703.1	.....	.....	-	-	
Marula	N4 (NR)	FAO 1	LK054504	.....	.P.....	-	-	
Marula	N17(MR)	FAO 1	LK054505	.....	.....	-	-	2
Marula	N25(NR)	316	LK054506	.....	.....	-	-	
Marula	N3 (SR)	FAO 1						
Marula	N31(NR)	FAO 1	LK054507	.....	.P.....	-	-	
KNPW8_35			JX442249.1	.....-------	.KN.R.D	-	+	
Wel24_2-5			JX442248.1	.....-------	.KN.R.D	-	+	
Marula	N7 (MR)	3014	LK054508	.....-------	.KN.R.D	-	+	
Marula	N8 (SRD)	316	LK054509	.....-------	.KN.R.D	-	+	
Marula	N1 (NR)	3014						
Marula	N19(MR)	Control						
Marula	N27(NR)	316						
Marula	N10(NR)	Control	LK054510	.....-------	.KN.R.D	-	+	3
Marula	N12(MR)	Control						
Marula	N26(NR)	Control						
Marula	N28(NR)	FAO 1						
Marula	N11(NR)	316	LK054511	.....-------	.KN.R.D	-	+	
Marula	N33(SRD)	FAO 1	LK054512	.....-------	.KN.R.D	-	+	
Ita8-13			JX442245.1	....I..K-------	.KN.R.D	-	+	
Mar1_49			JX442244.1	....I..K-------	.KN.R.D	-	+	
HIP5_1-16			JX442246.1	....I..K-------	.KN.R.D	-	+	
KNP102-9			JX442247.1	....I..K-------	.KN.R.D	-	+	

Clinical reactions of cattle exposed to tick challenge at Marula farm are in parenthesis and abbreviated as follows: SRD – severe reaction and died; SR – severe reaction; MR – mild reaction; NR – non-reactor; FD – found dead. Curly braces enclose animals from which the same sequence type was obtained, '.' epitope positions which have a conserved residue; '-' gapped positions in the antibody epitopes.

Amplification and sequencing of the gene encoding *T. parva* p67

The amplification of the gene encoding p67 was performed using genomic DNA templates with primer pair IL 613 3 (ACAAACACAATCCCAAGTTC) and IL 792 2 (CCTTTACTACGTTGGCG), designed to amplify a 900 base pair (bp) internal fragment containing part of exon 1 and part of exon 2 separated by the 29-bp intron sequence (Nene et al. 1996). These conserved primers amplify multiple *p67* alleles (Nene et al. 1996; Sibeko et al. 2010). The polymerase chain reaction (PCR) reagents and cycling parameters were as described previously (Nene et al. 1996). The PCR products were purified after fractionation through 1.2% agarose gels, using the QIAQuick gel extraction kit (Qiagen). The amplicons were sequenced bi-directionally using IL 613 3 forward and IL 792 2 reverse primers using Sanger dideoxy technology. Sequences were assembled and edited using the CLC Genomics Workbench version 6.0.

p67 linear epitope sequence alignments, variant calling and sequence logo creation

Basic local alignment search tool (BLAST) version 2.2.29+ (Altschul et al. 1990) search against a local database containing a repertoire of *T. parva* isolates in East and South Africa assigned the Marula query nucleotide sequences based on alignment scores to one of the four previously identified *T. parva p67* allele categories. The 13 sequences derived from *T. parva p67* antigen gene included in the local database were retrieved using the BLAST utility of the NCBI nucleotide database and had GenBank accessions: XM_758212.1, U40703.1, AF079177.1, JX442251.1, JX442249.1, JX442247.1, JX442246.1, JX442245.1, AF079176.1, JX442248.1, JX442244.1, JX442250.1 and AF079175.1.The sequences included in the local database were used as references in a multiple sequence alignment with *p67* sequences from isolates genotyped in this study covering the polymorphic central region of the gene. The alignment was generated with MAFFT version 7.122 using the L-INS-i option (Katoh et al. 2002), with minor subsequent manual adjustments. The program *seq2logo* version 1.2 (Thomsen et al.

2012) was used to create a Kullback-Leibler sequence logo by calculating the information content as a function of the observed and background probability of residues at each position in the alignment region spanning the mapped targets of host B-cell responses. In addition, estimates of *p67* gene polymorphism, π, calculated as the average number of nucleotide differences per site, were generated with DnaSP version 5 (Librado et al.2009).

Selection of best-fit nucleotide substitution models for *T. parva* p67 evolution

The phylogeny of the *T. parva* p67 antigen gene sequences was inferred from an optimal subset of the alignment generated as described above. This alignment subset was determined by the number of residues in gap-free columns by the program MaxAlign version 1.1 (Gouveia et al. 2007). We performed a preliminary reconstruction of the phylogenetic relationships between sequences using a neighbor-joining (NJ) tree (Saitou et al. 1987) under the Jukes and Cantor (JC) model. The NJ-JC tree served as the basis for calculating the likelihood scores for 56 different nucleotide substitution models with varying sets of substitution rate parameters capable of accommodating rate heterogeneity between sites. The best fitting model for the *p67* data was selected based on the Akaike Information Criterion ($AIC = -2\ln L + 2K$), a function of the maximized log-likelihood (lnL) and the number of estimated parameters (K) for a model (Wagenmakers et al.2004)**,** using the program Modeltest version 3.06 (Posada et al. 1998)**.**

Assessing model fits by using Akaike information criterion (AIC)

The phylogeny of the *T. parva* p67 antigen gene sequences was estimated with the maximum-likelihood (ML) criterion and the parameters from the best-fit nucleotide substitution model, using

PAUP version 4 (Swofford 2002). Estimations of d_S and d_N were performed using the codeml program from the PAML package version 4 (Yang 2007). These estimates were obtained under two different models of evolution: the nearly neutral (M1) and the positive selection (M2) models. M1 assumes two categories of sites, one evolving under purifying selection ($0 < d_N/d_S < 1$) and the other neutral ($d_N/d_S = 1$). Model M2 adds to these an additional category of positively selected sites ($1 < d_N/d_S$). Both models were fitted by allowing different nucleotide frequencies for each codon position. The Akaike information criterion was used to seek the fitted model where the information loss based on the expected relative Kullback-Leibler distance as defined by AIC differences (Δ_i) was minimal.

Results

p67 gene variants among *T. parva* isolates from Marula ranch (central Kenya)

We sequenced the central region of the *T. parva p67* gene from a collection of cultured schizont-infected lymphocytes recently isolated from cattle which had been exposed to tick challenge within Marula farm where the African buffalo co-grazed (Pelle et al. 2011). We generated an alignment of the amplified internal fragment of the *T. parva p67* gene, which contains 900 base pairs in the reference *T. parva* Muguga isolate. The alignment contains 27 sequences, nine of which are available in GenBank and 18 of which are derived from the Marula isolates (Fig. 1). There were 10 different p67 sequences represented among the total of 18 from Marula, each of which was derived from a schizont culture isolate from a different animal (Table 1). None of the ten sequences was the same as the p67 sequences of the *T. parva* stocks within the immunizing stabilates, all of which are identical to one another (Nene et al. 1996, M. Norling and J. Silva in preparation).The Marula sequences described in this report are available from the European Nucleotide Archive at http://www.ebi.ac.uk/ena/data/view/LK054504-LK054513 and animals from which the same sequence type was obtained are shown in table 1. Among the alleles sequenced from the Marula isolates, the overall DNA polymorphism, π, within and outside the mapped B-cell epitopes in the amplified central fragment of the *p67* gene was 0.19627 (19.6%). All sequences were divided into three allele types, 1 through 3, defined by the presence or absence of two indels (Table 1). The indel set comprised both a 129 bp deletion (allele type 1) and an additional 174 bp deletion (allele type 3). Fifty seven percent of the sequences were within *p67* allele type 3 based on the indels and only 10% were within allele size type 1. Approximately one third of all sequences in the Marula isolate collection were assigned to *p67* allele type 2 on the basis of sequence similarity to allele type 1 combined with the absence of the 129 bp deletion (Table 1).

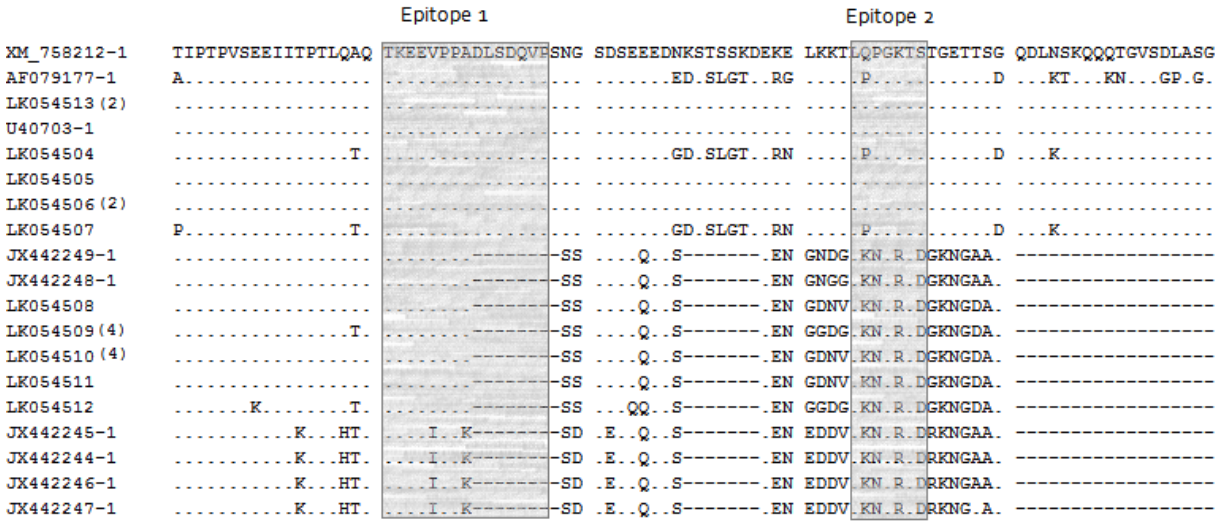


Fig. 1 Multiple deduced amino acid sequence alignment of a section of the amplified internal fragment of the *T. parva p67* gene generated using the multiple sequence alignment program, MAFFT version 7.122 (Katoh et al. 2002) '.' Amino acid (aa) positions which have a conserved residue;'-' gapped aa positions; number of Marula isolates carrying an allele is indicated in parentheses. The epitopes recognised by anti-sporozoite monoclonal antibodies are boxed.

Sequence polymorphism between antibody epitopes within the *p67* gene for Marula isolates

We determined the degree of sequence variation within epitopes 1 and 2 of the *T. parva p67* gene in the Marula field isolates. The predicted protein sequences were compared with the previously characterized p67 protein sequences derived from East and South African isolates (Sibeko et al. 2010). Within the region encoding the mapped epitopes, the high level of nucleotide sequence polymorphism translated into coding changes that resulted in amino acid mutations relative to the *T. parva* Muguga reference allele (XM_758212.1), a cattle-type *p67* sequence (Nene et al. 1996). These variants are shown in Table 1. Among the sequences assigned to allele type 1 and 2, epitope 1 was remarkably conserved, perfectly matching the reference sequence. However, an in-frame deletion of a 7 amino acid segment was observed in a majority of the Marula sequences, all of which were assigned to allele type 3. This truncated epitope sequence resulting from the deletion appears to be a unique feature of the *p67* indel-defined allele type 3. In contrast, epitope 2 sequences were less homogeneous among the

sequences assigned to allele types 1 and 2 due to a nucleotide substitution that conferred a predicted amino acid change, with the consensus glutamine, ^{210}Q, being replaced by proline, P. Additionally, there were non-synonymous mutations in epitope 2 relative to the reference sequence in all of the isolates assigned to *p67* allele type 3.

Position specific enrichment and depletion of residues within the p67 linear epitopes

Information content, a function of the observed and background probability for a particular column in the multiple sequence alignment was calculated for each position spanning the mapped epitopes to assess their relative contributions to epitope polymorphisms and visualized in a Kullback-Leibler logo. As can be seen in figure 2, seq2logo (Thomsen et al. 2012) captures the extent of B-cell epitope polymorphism in *T.parva p67* among field isolates by taking advantage of pseudo count estimates and sequence weighting to deal with data redundancy and the low number of observations respectively. The amino acids enriched at each epitope position are shown on the positive Y axis and the corresponding depleted amino acids are on the negative Y axis.

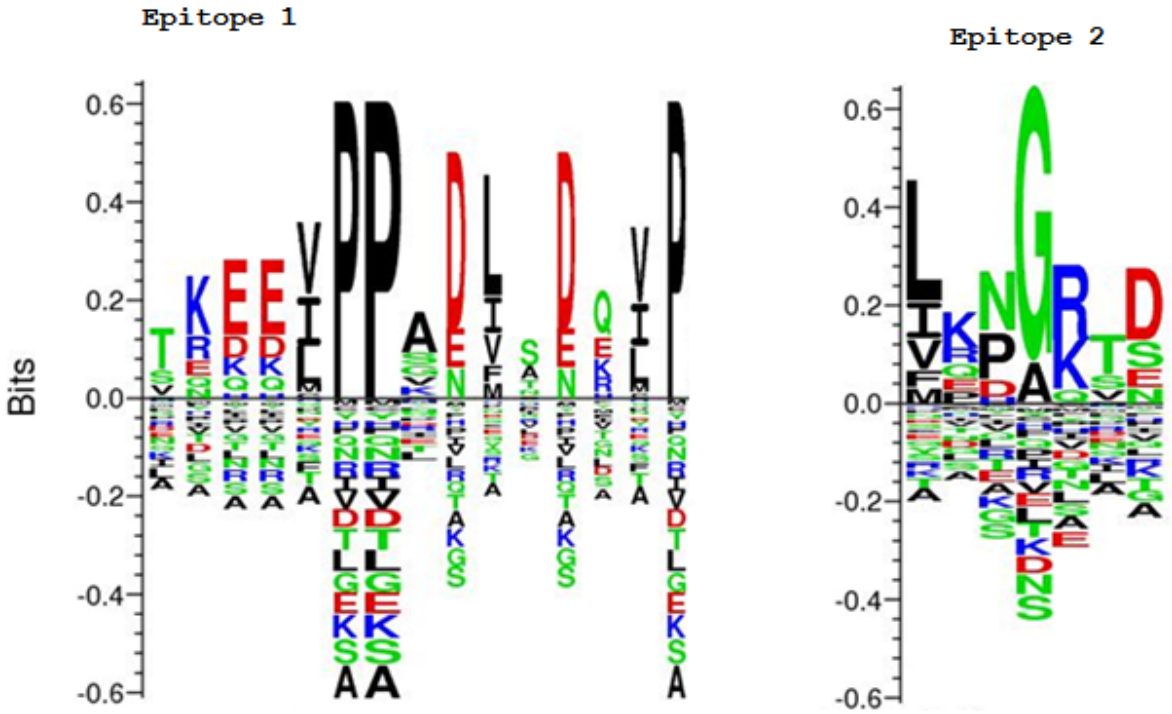


Fig. 2 Kullback-Leibler logo for sequences corresponding to B cell epitopes 1 and 2 for sporozoite neutralizing antibodies on the *T. parva* p67 based on a multiple alignment of the predicted amino acid sequences within the central region. Enriched amino acids (aa) are shown on the positive *y*-axis and depleted (under-represented) amino acids on the negative *y*-axis. The height of the column of aa residues represents the level of conservation at a particular position, while the relative residue height denotes its frequency. Polar aa are depicted in green, basic aa in blue, acidic aa in red, and hydrophobic aa in black.

Analysis of Marula *p67* sequences revealed an excess of nucleotide substitutions resulting from transitions relative to transversions.

To investigate the nature of diversity in the sequenced *p67* alleles, the model of nucleotide substitution that represents the best fit to the data, to be used for phylogenetic inference, was established by statistical testing. This was assessed by the Akaike information criterion (AIC), a function of the maximized log-likelihood (lnL) and the number of estimated parameters (K) for a model. The *p67* dataset was best explained by the TIM+I+G model of evolution. This model assumes variable base

frequencies, variable transition rates, two transversion rates, a proportion of invariable sites (I) and a gamma distribution with shape parameter (α). With the corresponding base frequencies taken into consideration, maximum-likelihood estimates of substitution rates revealed more transitions than transversions. In particular, the A-to-G transition occurred most frequently, 65% of the sites were invariable (*p-inv*) and the gamma shape parameter (α) was less than 1, the latter indicating a relatively slow rate of evolution at the *p67* locus (Table 2).

Table2 AIC model selection and maximum-Likelihood estimates from best-fit model (TIM+I+G)

AIC nucleotide substitution model selection

Model	-lnL	K	AIC	delta	weight	cumWeight
TIM+I+G	3243.4773	8	6502.9546	0.0000	0.2086	0.2086
TrN+I+G	3244.5530	7	6503.1060	0.1514	0.1934	0.4021
K81uf+I+G	3245.0901	7	6504.1802	1.2256	0.1130	0.5151
HKY+I+G	3246.1763	6	6504.3525	1.3979	0.1037	0.6188
TIM+G	3245.1899	7	6504.3799	1.4253	0.1023	0.7211
TrN+G	3246.3713	6	6504.7427	1.7881	0.0853	0.8065
K81uf+G	3246.7419	6	6505.4839	2.5293	0.0589	0.8654
HKY+G	3247.9426	5	6505.8853	2.9307	0.0482	0.9136
TIM	3313.7852	6	6639.5703	136.6157	4.50e-31	1.0000
K80	3358.1204	1	6718.2407	215.2861	0.00e+00	1.0000
JC	3368.9949	0	6737.9897	235.0352	0.00e+00	1.0000

Maximum-Likelihood estimates from best-fit model (TIM+I+G)

Base frequencies

freqA = 0.3548
freqC = 0.2131
freqG = 0.2104
freqT = 0.2216

Among-site rate variation

Proportion of invariable sites (I) = 0.3332

Variable sites (G)
Gamma distribution shape parameter = 0.6479

Substitution model

Rate matrix

R(a) [A-C] = 1.0000 R(b) [A-G] = 2.5858 R(c) [A-T] = 1.3319
R(d) [C-G] = 1.3319 R(e) [C-T] = 1.7247 R(f) [G-T] = 1.0000

The Akaike information criterion (AIC), a function of the maximized log-likelihood (lnL) and the number of estimated parameters (K) used to select the best fitting model. Delta denotes the AIC differences. Also shown are the Maximum-Likelihood estimates of base frequencies, nucleotide substitution rates, and rate heterogeneity parameters (proportion of invariable sites and shape parameter of the gamma distribution).

Spatial clustering of positively selected codons in the *p67* gene

We made maximum likelihood inferences of positive selection based on the AIC approach to compare the fit of models M1, the quasi-neutral model, which allows for two site classes ($d_N/d_S = 1$, $0< d_N/d_S < 1$) relative to a positive selection model, M2, with three site classes ($d_N/d_S = 1$, $0< d_N/d_S < 1$ and $d_N/d_S > 1$). As shown in table 3, fitting the codon substitution model that captures the proportion of codons for which $d_N/d_S > 1$ within the *p67* gene provided a better fit than M1 (minimum AIC value). In the best fit model (M2), about 53% of the codons are strongly conserved (average $d_N/d_S = 0.11037$), 35% are consistent with neutral evolution ($d_N/d_S = 1.0$) and the other 10% of codons exhibit d_N/d_S values suggesting evolution under positive selection (average $d_N/d_S = 2.95491$). Bayesian posterior probabilities calculated using Bayes empirical Bayes (BEB) (Yang et al. 2005) in M2 revealed the codons with signatures of positive selection (Table 4). However, in the case of most codons, the standard error interval includes $d_N/d_S = 1$, and the posterior probability of any codon evolving with $d_N/d_S > 1$ is never equal to, or larger than, 95% and therefore not significant at this level.

Table 3 Parameters and Akaike information criterion (AIC) scores of the *p67* gene under codon site models.

Model	K	lnL	AIC_c	Δ_i	$d_N/d_S < 1$	$d_N/d_S = 1$	$1 < d_N/d_S$
M1	49	1260	2618	2.0	p: 0.54325 ω: 0.09055	p: 0.45675 ω: 1.00000	
M2	51	1257	2616	0.0	p: 0.53538 ω: 0.11037	p: 0.35855 ω: 1.00000	p: 0.10607 ω: 2.95491

The Akaike information criterion (AIC), a function of the maximized log-likelihood (lnL) and the number of estimated parameters (K) used to select the best fitting model. ω is the Akaike weight and p gives the proportion of codon sites belonging to each of the d_N/d_S ratio classes, Δ_i denotes AIC differences.

Table 4 Posterior probabilities (Pr) and estimated d_N/d_S ratios (post mean +- SE for d_N/d_S) calculated using Bayes empirical Bayes (BEB) analysis for sites likely to be under positive selection on the *T. parva p67* gene identified using the reference stock *T. parva* Muguga.

Positively selective site	Amino acid	Pr (d_N/d_S >1)	post mean +- SE for d_N/d_S
151	T	0.645	2.257 +- 1.363
186	G	0.868	2.815 +- 1.394
194	N	0.686	2.259 +- 1.212
204	E	0.532	1.893 +- 1.094
205	L	0.545	1.907 +- 1.080
208	T	0.772	2.505 +- 1.294
237	L	0.801	2.596 +- 1.323
262	G	0.535	1.871 +- 1.047
277	H	0.745	2.469 +- 1.335
278	Q	0.671	2.247 +- 1.243
280	V	0.654	2.238 +- 1.291

Discussion

We present evidence for *p67* alleles grouped into three distinct indel types within a single *T. parva* population infecting cattle that were part of a field trial of ITM vaccines designed to evaluate protection afforded to immunized animals that received a *T. parva* challenge from buffalo-associated ticks (Pelle et al. 2011). All of these p67 alleles were derived from *T. parva* genotypes, present in the Marula tick population, since although ITM immunization typically induces a persistent carrier state in cattle, both Marikebuni and the Muguga, Serengeti and Kiambu V stocks within the trivalent FAO1 Muguga cocktail, have an identical p67 sequence, that differs from the p67 sequences of any of the Marula isolates. Comparable levels of heterogeneity at the *T. parva p67* locus have recently been reported in South Africa (Sibeko et al. 2010). Classification relative to previously described alleles was based on presence or absence of 129 and 174 base pair indels in the central region of the *p67* gene. We show that parasites with a 174 base pair deletion (allele type 3) were most frequent in the isolates from the cattle under challenge at Marula farm. Whether there is any underlying selective advantage that the deletion might confer is unclear. One possibility is that epitope 1 is truncated and therefore not recognized in parasites containing the 129 bp sequence but with the 174 bp deleted.

The 10 different *T. parva p67* sequences described herein (derived from a total of 18 sequences) were distinct at a minimum of 144 nucleotides excluding sequence reads covering indels.This contrasts strongly with the complete conservation of nucleotides at the *p67* locus, including introns and the third base in codons, in cattle-transmissible parasites (Nene et al. 1996, Musoke et al. 2005). We assessed the nature and extent of variation in the two closely juxtaposed B-cell epitopes that have been mapped to the central polymorphic region of the p67 protein (Nene et al. 1999). Our analysis demonstrated that all allele type 1 sequences containing the 174 bp sequence but lacking the 129 bp sequence were identical to the prototype Muguga (XM_758212.1) at both epitopes analyzed. This was also the case

for the isolates within the indel-defined allele type 2 with the exception of a single variation present in two isolates. The high level of conservation of these epitopes suggests that they are candidates for inclusion in multivalent vaccines with the potential to reduce establishment of infection when challenge occurs with multiple *T. parva* genotypes. Indeed, neutralizing antibodies directed against conserved epitopes have been found to be correlates of protection against infective sporozoites in several species of parasitic protozoans within the phylum apicomplexa including *Plasmodium* (Weedall et al. 2007), *Eimeria* (Wallach et al. 2008) and *Theileria parva* (Musoke et al. 1992).

Although studies with monoclonal antibodies have defined ^{169} TKEEVPPADLSDQVP ^{183}, as a p67 epitope in *T. parva* isolates that are transmissible between cattle by ticks (Nene et al. 1999), our analysis reveal a deletion variant in this region that is common to all type 3 alleles, suggesting that it may not represent a B cell epitope in many buffalo-derived *T. parva* parasites. Interestingly, there is a high level of conservation of the epitope sequence flanking the deletion in the isolates carrying allele type 3. The amino acid sequence of epitope 2 in allele 3 isolates is also conserved but distinct from that of allele types 1 and 2, although not altered in length. Collectively, these findings support the hypothesis that these *p67* regions are potential targets for inclusion in a combinatorial vaccine formulation that includes immunodominant B cell epitopes. Vaccine strategies similar to those being pursued for the polymorphic merozoite surface protein (MSP) 1 malaria protein could also be applicable to *T. parva*. In the case of MSP 1, it has been suggested that focusing on both the conserved C-terminal region (Blackman et al. 1994) in combination with polymorphic subtype-restricted epitopes (Tetteh et al. 2005) to create chimeric constructs could be a viable vaccine development strategy. For *T. parva* p67, a vaccine could also incorporate conserved epitopes located outside the p67 central region, for example in the C-terminal region in which no polymorphism has been reported, within a single chimeric construct.

We used the multiple sequence alignment generated from the *p67* dataset described in table 1, to investigate what modality of selection may have contributed to the evolution of the *p67* gene. Such tests typically assume a particular phylogeny and can be influenced by the nucleotide substitution model selected (Yang et al. 1994). After optimization of the parameters, an AIC analysis indicated that the TIM+I+G model of substitution represented the best fit to the *p67* polymorphism data. AIC is an efficient alternative to likelihood ratio tests (Posada et al. 2004) and has recently been implemented in the identification of the codon-substitution model that best fits sequence polymorphisms in a region of VAR2CSA –a *Plasmodium falciparum* protein involved in placental sequestration in the mammalian host (Dahlbäck et al. 2006). We detected a subset of codons with signatures of positive selection using codon-based likelihood analysis based on the observed d_N/d_S rate ratio (Yang et al. 2000). The parameters defining the codon evolution model used in this study were derived by capturing heterogeneity of d_N/d_S across multiple sites, with positive selection inferred when $d_N/d_S > 1$. Under the codon model that provided the best fit to the data, model M2, about 10% of codons showed potential evidence of positive selection (average $d_N/d_S = 2.95$). However, despite the demonstration of evolutionary patterns consistent with the effects of positive selection, none of the amino-acid sites in gap-free alignment columns in the *p67* gene identified by maximum likelihood analysis had a highly significant (> 95%) posterior probability of being positively selected (Table 4). It is possible that the high d_N/d_S ratio observed for several amino acids results from allelic dynamics and does not accurately reflect the effect of selection (Kryazhimskiy et al. 2008). Interestingly, none of the codons located within defined linear epitopes recognized by monoclonal antibodies show signatures of positive selection. These results are consistent with a recent study that showed no preferential distribution of sites under positive selective pressure within T cell epitopes in candidate vaccine antigens from the pathogenic *T. parva* schizonts (Pelle et al. 2011). Among the 115 cattle exposed to challenge at Marula, from which the 18 isolates analysed were derived, approximately 80 % of those that developed severe disease and died, exhibited features associated with corridor disease, specifically low schizont

parasitosis and piroplasm parasitaemia, irrespective of whether they were isolated from vaccinated or control cattle (R. Bishop and A. Musoke, unpublished data). One conclusion from this study is therefore that *T. parva* parasites originating from buffalo can infect and cause severe disease in co-grazing cattle and have diverse p67 genotypes, unlike the conserved p67 sequence observed in *T. parva* populations that have adapted to transmission between cattle. The data also suggests that buffalo-derived *T. parva* can sometimes 'break-through' the immunity induced by Infection and treatment immunisation and induce severe clinical reactions, in areas with a high tick challenge.

Acknowledgements

We are grateful to Dr Jan Naessens for proof reading the manuscript. The work was supported by funding from the DFG German-African Cooperation Projects in Infectiology: "Molecular epidemiology network for promotion and support of delivery of life vaccines against *Theileria parva* and *Theileria annulata* infection in Eastern and Northern Africa (SE 862/2-1)". This is ILRI publication number 92-9146-369-8.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215: 403–10

Bishop R, Nene V, Staeyert J, Rowlands J, Nyanjui J, Osaso J, Morzaria S, Musoke A (2003) Immunity to East Coast fever in cattle induced by a polypeptide fragment of the major surface coat protein of *Theileria parva* sporozoites. Vaccine 21: 1205–12

Blackman MJ, Scott-Finnigan TJ, Shai S, Holder AA (1994) Antibodies inhibit the protease-mediated processing of a malaria merozoite surface protein. J Exp Med 13:389–93

Dahlbäck M, Rask TS, Andersen PH, Nielsen MA, Ndam NT, Resende M, Turner L, Deloron P, Hviid L, Lund O, Pedersen AG, Theander TG, Salanti A (2006) Epitope mapping and topographic analysis of VAR2CSA DBL3X involved in *P. falciparum* placental sequestration. PLoS Pathog 2(11): e124

Gouveia-Oliveira R, Sackett PW, Pedersen AG (2007) MaxAlign: maximizing usable data in an alignment. BMC Bioinformatics 8: 312

Katoh K, Misawa K, Kuma K, Miyata T (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res 30: 3059–66

Kryazhimskiy S, Plotkin JB (2008) The population genetics of dN/dS. PLoS Genet 4(12): e1000304

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25:1451–2

McKeever DJ, Taracha EL, Innes EL, MacHugh ND, Awino E, Goddeeris BM, Morrison WI (1994) Adoptive transfer of immunity to *Theileria parva* in the $CD8^+$ fraction of responding efferent lymph. Proc Natl Acad Sci USA 91: 1959–63

McKeever DJ, Taracha EL, Morrison WI, Musoke a J, Morzaria SP (1999) Protective immune mechanisms against *Theileria parva*: evolution of vaccine development strategies. Parasitol Today 15:263–7

Morrison WI (2009) Progress towards understanding the immunobiology of *Theileria* parasites. Parasitology 136:1415-26

Morzaria SP, Dolan TT, Norval RA, Bishop RP, Spooner PR (1995) Generation and characterisation of cloned *Theileria parva* parasites. Parasitology 111: 39-49

Morzaria S, Spooner P, Bishop R, Mwaura S (1999) The preparation of a composite stabilate for immunisation against East Coast fever. In: Morzaria, S. and Williamson, S. (eds.) Live vaccines for *Theileria parva*: deployment in eastern, central and southern Africa. International Livestock Research Institute, Kenya. 56-61

Musoke A, Morzaria S, Nkonge C, Jones E, Nene V (1992) A recombinant sporozoite surface antigen of *Theileria parva* induces protection in cattle. Proc Natl Acad Sci USA 89: 514–8

Musoke A, Rowlands J, Nene V, Nyanjui J, Katende J, Spooner P, Mwaura S, Odongo D, Nkonge C, Mbogo S, Bishop R, Morzaria S (2005) Subunit vaccine based on the p67 major surface protein of *Theileria parva* sporozoites reduces severity of infection derived from field tick challenge. Vaccine 23:3084–95

Nene V, Musoke A, Gobright E, Morzaria S (1996) Conservation of the sporozoite p67 vaccine antigen in cattle-derived *Theileria parva* stocks with different cross-immunity profiles. Infect Immun 64:2056-61

Nene V, Gobright E, Bishop R, Morzaria S, Musoke A (1999) Linear peptide specificity of bovine antibody responses to p67 of *Theileria parva* and sequence diversity of sporozoite-neutralizing epitopes: implications for a vaccine. Infect Immun 67:1261-6

Norval RAI, Perry BD, Young AS (1992) The epidemiology of *Theileria* in Africa. London: Academic.

Oura CA, Tait A, Asiimwe B, Lubega GW, Weir W (2011) *Theileria parva* genetic diversity and haemoparasite prevalence in cattle and wildlife in and around lake mburo national park in Uganda. Parasitol Res 108: 1365–74

Payne R (1999) Preparation of stabilates for immunisation against East Coast fever at NVRC, Muguga, Kenya. In: Morzaria, S. and Williamson, S. (eds.). Live vaccines for *Theileria parva*: deployment in eastern, central and southern Africa. International Livestock Research Institute, Kenya: 62-65

Pelle R, Graham SP, Njahira MN, Osaso J, Saya RM, Odongo DO, Toye PG, Spooner PR, Musoke AJ, Mwangi DM, Taracha EL, Morrison WI, Weir W, Silva JC, Bishop RP (2011) Two *Theileria parva* CD8 T cell antigen genes are more variable in buffalo than cattle parasites, but differ in pattern of sequence diversity. PLoS One 6(4): e19015

Posada D, Buckley TR (2004) Model selection and model averaging in phylogenetics: Advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. Syst Biol 53:793–808

Posada D, Crandall KA (1998) MODELTEST: testing the model of DNA substitution. Bioinformatics 14: 817–8

Radley DE, Brown CGD, Burridge MJ, Cunningham MP, Kirimi IM, Purnell RE, Young AS (1975) Chemoprophylactic immunisation of cattle against *Theileria parva* (Muguga) and five *Theileria* strains. Vet Parasitol 1:35–41

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4: 406–25

Sibeko KP, Geysen D, Oosthuizen MC, Matthee CA, Troskie M, Potgieter FT, Coetzer JA, Collins NE (2010) Four p67 alleles identified in South African *Theileria parva* field samples. Vet Parasitol 167:244–54

Swofford DL (2002) PAUP: phylogenetic analysis using parsimony and other methods. Version 4. Sinauer Associates, Sunderland, Massachusetts

Taracha EL, Goddeeris BM, Teale AJ, Kemp SJ, Morrison WI (1995) Parasite strain specificity of bovine cytotoxic T cell responses to *Theileria parva* is determined primarily by immunodominance. J Immunol 155:4854–60

Tetteh KK, Cavanagh DR, Corran P, Musonda R, Mcbride JS, Conway DJ (2005) Extensive antigenic polymorphism within the repeat sequence of the *Plasmodium falciparum* merozoite surface protein 1 block 2 is incorporated in a minimal polyvalent Immunogen. Infect Immun 73:5928-35

Thomsen MC, Nielsen M (2012) Seq2Logo: a method for construction and visualization of amino acid binding motifs and sequence profiles including sequence weighting, pseudo counts and two-sided representation of amino acid enrichment and depletion. Nucleic Acids Res 40:W281–7

Wagenmakers EJ, Farrell S (2004) AIC model selection using Akaike weights. Psychon Bull Rev 11: 192–6

Wallach MG, Ashash U, Michael A, Smith NC (2008) Field application of a subunit vaccine against an enteric protozoan disease. PLoS One 3: e3948

Weedall GD, Preston BM, Thomas AW, Sutherland CJ, Conway DJ (2007) Differential evidence of natural selection on two leading sporozoite stage malaria vaccine candidate antigens. Int J Parasitol 37: 77–85

Yang Z, Bielawski JP (2000) Statistical methods for detecting molecular adaptation. Trends Ecol Evol 15:496–503

Yang Z, Goldman N, Friday A (1994) Comparison of models for nucleotide substitution used in maximum-likelihood phylogenetic estimation. Mol Biol Evol 11:316–24

Yang, Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol 24:1586–91

Yang Z, Wong WS, Nielsen R (2005) Bayes empirical bayes inference of amino acid sites under positive selection. Mol Biol Evol 22:1107–18

5.0 Chapter 5: Overall discussion and recommendations

The research described herein has primarily focused on understanding the potential impacts of variations in the host and parasite genotypes for the development of protective immunity in response to subunit vaccination for control of *T. parva*. It has long been hypthesised based on the protective mechanism demonstrated for the infection and treatment method that a strategy based on induction of protective $CD8^+$ T-cells offers one potential route for development of subunit vaccines against *T. parva* infections. This study has provided strong evidence that implementation of such a vaccine strategy has to overcome significant constraints imposed by class I MHC diversity present among African cattle 'breeds' residing in the ECF endemic regions. Such variability has been recognized previously, but no systematic analyses of class I variability in African cattle appear to have been conducted. This paucity of data is attributable largely to the complexity of cattle MHC genetics (Ellis 2004) and is illustrated by the duration it has taken to generate a reference list of expressed class I MHC alleles carried by the inbred European Holstein cattle (Robinson et al. 2013). In particular, studies cataloguing the repertoires of class I MHC genes in cattle have to address the need for simultaneous genotyping of multiple co-expressed loci that are difficult to analyse by routine PCR amplification and Sanger sequencing approaches. This problem is exacerbated by the fact that the expressed gene complement varies between haplotypes. This complexity is illustrated by the long time it has taken to generate a reference list of expressed class I MHC alleles expressed by the inbred European Holstein cattle populations (Robinson et al. 2013).

It is worthy of note that ‘breeds’ in the strict sense exemplified by European Holstein or Friesian are rare in Africa, because most cattle contain an admixture of alleles derived from taurine and indicine lineages (MacHugh et al. 1997, Anderung et al. 2007). This represents one reason why although the expressed class I MHC sequences of African taurine cattle has not yet been analyzed in detail, it can be expected that allelic polymorphism and haplotype composition will vary significantly between

individuals.

The current study adapted recent advances in MHC typing for use in African cattle populations. In particular, the study utilized the multiplexing capabilities and depth of coverage of the GS FLX 454 pyrosequencing platform (454 Life Sciences). However, high throughput sequencing data require careful interpretation for MHC typing as there is little doubt that some of the sequence variants generated by these approaches are artefactual and this may obscure the detection of 'true' rare variants (Babik et al. 2009). Consequently, the data presented here excluded a specified number of reads classed either as 'putative artifacts' or 'chimeras' following passage through a series of sequence filtering algorithms. The design of the typing also allowed comparisons to be made between the genotypes obtained by amplicon pyrosequencing and those obtained by class I MHC cDNA Sanger sequencing for a subset of the samples analyzed. Overall, the data suggests that use of amplicon pyrosequencing in conjunction with robust read filtering algorithms can achieve the level of accuracy of Sanger sequencing and therefore has the potential for routine application in unraveling the complexities of cattle MHC genetics.

It is worth mentioning that it does not suffice to simply catalog expressed class I MHC sequences from an African native (Ankole) and exotic (Holstein) *Bos taurus* cattle as functional differences between alleles are not readily explicable. In the current study, machine learning and phylogenic clustering algorithms were necessary to reliably deliver a functional classification of the MHC molecules using information drawn solely from the residues that interact with peptides (Thomsen et al. 2013). The most important information revealed by analysis of the data is not that the Ankole class I MHC differ from other *Bos taurus* animals (in this case Holstein) *per se*, but the scale of the predicted class I MHC functional divergence despite the fact that the two populations analyzed are both within the *Bos taurus* lineage and therefore share a relatively recent common ancestor on an evolutionary timescale.

One conclusion supported by all the analyses undertaken in this study is that there is convincing evidence that the overlap between the peptide binding specificities of cattle class I MHC molecules is likely to be largely confined to alleles belonging to the same 'breed'. These differences carry the implication that the repertoire of class I alleles expressed by different cattle breeds will determine the number of antigens/epitopes that may need to be incorporated in a recombinant antigen cocktail vaccine before field use at a scale encompassing multiple cattle populations can be contemplated.

The constraint of class I MHC diversity for recombinant vaccine development for ECF control may seem insurmountable, especially given that based on FAO statistics, 200 phenotypically distinguishable zebu, taurine and indicine-taurine crossbreeds are resident in areas known to sustain endemic *T. parva* infections in eastern, central, and southern Africa (FAO; http://dad.fao.org/). However, grouping class I products based on sequence features defining their peptide binding grooves as implemented in this study and elsewhere (Pandya et al. 2015), provides evidence for major specificity groups, known as 'supertypes' in cattle. Such 'supertypes' have previously been described for humans by Sette and Sidney (1999). Consequently, cellular responses induced by live vaccination can potentially be mimicked by immunization with a restricted set of epitopes covering the specificity defined by class I MHC allele 'supertypes' without compromising coverage of cattle populations with an epitope-based vaccine. For this strategy to be pursued, data on class I MHC genotypes from a wider range of African cattle, especially 'breeds' that comprise pastoralists' herds and those important to smallholder dairy farmers in ECF endemic regions would have to be generated. Such efforts will certainly benefit from the work described in this thesis that has clearly established the feasibility and value of new high throughput sequencing technologies in tackling the complexity of cattle MHC genetics.

The findings from the *in silico* analysis that suggested significant functional divergence between the Ankole and Holstein class I sequences were confirmed by subsequent *ex vivo* data collected using T-cell lines from *T. parva* infected cattle. In particular, recognition of recently identified candidate 'Tp' $CD8^+$ T-cell target antigens was assessed in *ex vivo* ELISpot assays using PBMCs and purified $CD8^+$ T-cells from Ankole cattle that were serologically positive for *T. parva* based on a recombinant version of the *T. parva* PIM antigen. Its worthy of note that *T. parva* $CD8^+$ T-cell target vaccine candidate antigens were originally identified using $CD8^+$ T-cell lines derived primarily from Holstein-Friesian cattle to screen *T. parva* cDNA expression libraries. To summarize, in PBMC derived from each of the eleven seropositive Ankole, there were no animals that were positive by ELISpot following sensitization by peptide pools derived from the Tp antigens. Positivity is defined as a gamma interferon response significantly greater than three times the medium control, which is the normally accepted standard for definition of an ELISpot response. This suggests that none of these eleven animals recognized any of the six Tp antigens. This includes the Tp1 and Tp2 antigens, which are immunodominant in the context of the class I haplotypes so far investigated and are currently the main focus of ongoing ECF recombinant vaccine trials being performed by an international consortium of laboratories.

The BV115 animal included as a positive control for the ELISpot assay was by contrast positive since only the Tp1 peptide pool, which was previously determined to be recognized by this animal, provided significant responses in the ELISpot assay. Additionally, the media control and peptide pools from other antigens, including Tp7 which has previously been shown to be negative using $CD8^+$ T-cells derived from BV115, did not generate a response. Furthermore, experience from relying on Tp specific cell lines (experimentally seeded into total PBMC isolated from the same animals) has suggested a detection level down to approximately 0.03% of PBMCs, which is a relatively low threshold of detection and provides support for the suggestion that this study would have detected T cells specific

for these antigens if they were present in the Ankole cattle.

There are no previous studies that have analysed recognition of candidate $CD8^+$ T-cell target antigens by a population of African taurine cattle kept by pastoralists following immunization by ITM for control of *T. parva* or field challenge. This is important since *T. parva* recombinant vaccine development projects seek to reproduce the T-cell based protection induced by ITM. The results described herein are consistent with the findings from an earlier study that demonstrated failure of T-cell lines from *Bos indicus* cattle to recognize the currently identified $CD8^+$ T-cell target antigens (Akoolo et al. 2008). *Bos indicus* cattle are generally believed to have resulted from an independent domestication from a widely geographically separated population of the now extinct Auroch (*Bos primigenius*) separated from the taurine lineage by several hundred thousand years of evolution. Taken together, these data allow reasonably firm conclusions to be drawn on the need to develop a multivalent vaccine that can potentially overcome the issue of genetic heterogeneities between cattle 'breeds'. For immediate purposes, it is necessary to comprehensively define cattle class I MHC ‘supertypes’ as a basis for selection of antigens for inclusion in a multivalent recombinant vaccine for field use. Combined with an appropriate cocktail of ‘supertpe’ epitopes with broad class I MHC binding specificity, an antigen delivery system that induces long-term T cell-based immunity in cattle, and is scalable and affordable to resource poor farmers will need to be developed. This is not yet commercially available for humans or livestock, although experimental progress is promising in murine model systems. The results are also consistent with the hypothesis that cattle infected with *T. parva* through delivery of sporozoites by *R. appendiculatus* ticks may possibly be immune through other effector mechanisms than induction of CD8+ T-cells.

One other conclusion that has emerged from this study is that for robust humoral immunity against the early stage of the parasite to be achieved, it is desirable to develop vaccines that allow for the dynamic nature of parasite populations in the field. In particular, this study investigated the extent of sequence diversity within the gene encoding p67, the *T. parva* sporozoite surface protein that is a major target of antibodies that can neutralize the establishment of infection and induce protection against subcutaneous needle challenge with homologous and heterologous sporozoite stabilates. To provide a more informed picture of the possible impacts of heterologous parasite challenge on protection afforded to p67 immunised cattle, this study also assessed whether the defined B-cell epitopes within the p67 protein are conserved between parasites isolated from clinically reacting cattle in regions where there is a reservoir of disease in wildlife. The findings demonstrate considerable variation among *T. parva* p67 alleles, with the encoded proteins varying widely in size due to length polymorphism, primarily indels within the central region of the p67 gene. This is in marked contrast to the absolute conservation of the p67 gene among parasites transmissible between cattle by ticks. This data suggests that buffalo-derived *T. parva,* with diverse p67 genotypes; 'break through' the immunity induced by recombinant p67 and induce severe clinical reactions in co-grazing cattle. In this context, it is relevant that there is as yet no field trial of recombinant p67 against a buffalo-derived *T. parva* tick challenge.

Importantly, the polymorphism among alleles of the p67 antigen includes sequence differences in sites encoding epitopes that are the targets of neutralizing antibodies. Only six of the 15 residues in the first of two closely juxtaposed epitope sequences in the central region of the p67 protein were identical among the alleles of the p67 antigen described herein and just three residues were conserved in the second epitope. The 18 p67 sequences contained three variants of the first epitope, differing mainly due to an in-frame deletion of a seven-amino acid (21 base pair) segment resulting in a truncated protein. By contrast, the three variants of the second epitope differed primarily due to coding changes that conferred amino acid substitutions. One conclusion supported by these differences is that there is

evidence to suggest that the p67 genotype selected for vaccination may lack efficacy against field challenge with buffalo-derived *T. parva*.

Although the field isolates of *T. parva* do not have identical p67 sequences, the observation of considerable overlap in the pattern of polymorphisms between antibody epitopes within the buffalo-derived p67 alleles suggests that the constraint of p67 diversity for recombinant vaccine development for ECF control may not be insurmountable. Clearly, functional studies would be required to confirm that neutralizing antibodies differ in their ability to recognize allelic variants of the epitopes and based on the outcome of these experiments, multiple epitopes could be combined in a single construct. The p67 sequence diversity findings if extended to isolates from additional buffalo-derived populations in a more comprehensive analysis highlight the potential for simultaneous immunization with vaccine constructs expressing multiple p67 allelic variants to achieve broad protection in the field. An additional hurdle that will need to be overcome in order to develop an effective anti-sporozoite vaccine will be to demonstrate protection against a *T. parva* infected tick challenge, rather than the needle challenges that have been employed to date. The importance of this is emphasized by the reduced efficacy of p67 in field trials (Musoke et al. 2005) relative to experimental laboratory challenges (Musoke et al. 1992, Bishop et al. 2003).

To summarize, sub-unit vaccine development for control of ECF in cattle is ultimately likely to require a multi-faceted approach which will incorporate a cocktail of antigens derived from the sporozoite and schizont stages in order to achieve a robust immunity. It is therefore not surprising that considerable donor resources are currently being expended on attempts to induce T cell-based vaccines using the schizont antigens used in this study, by an international consortium of laboratories (Nene et al. 2012). In addition, efforts are also underway to optimize the p67 sporozoite vaccine and identify additional antigens to permit deployment in the field. Initial attempts to combine delivery of recombinant p67

with schizont antigens are also ongoing. Ultimately a tick antigen component may also need to be included to enhance protection in the field (Olds et al. 2012). The data presented here helps to provide an evidence-based analysis of the factors that will be important for development of subunit vaccines with adequate efficacy and coverage. The recommendation is that priority should be given to grouping of cattle class I MHC alleles into functional ‘supertypes’ once extensive cattle polymorphism data is collated so as to better inform the feasibility of generating T cell based vaccines with a high population coverage regardless of 'breed' differences. It could also be helpful at this point to re-assess the field performance of the p67 vaccine with the major adjustment being the simultaneous immunization with vaccine constructs expressing multiple p67 allelic variants and/or additional sporozoite antigens.

References

Akoolo L, Pellé R, Saya R, Awino E, Nyanjui J, Taracha EL, Kanyari P, Mwangi DM, Graham SP (2008) Evaluation of the recognition of *Theileria parva* vaccine candidate antigens by cytotoxic T lymphocytes from Zebu cattle. Vet Immunol Immunopathol 121(3–4):216–221

Anderung C, Hellborg L, Seddon J, Hanotte O, Götherström A (2007) Investigation of X- and Y-specific single nucleotide polymorphisms in taurine (*Bos taurus*) and indicine (*Bos indicus*) cattle. Anim Genet 38(6):595–600

Babik W, Taberlet P, Ejsmond MJ, Radwan J (2009) New generation sequencers as a tool for genotyping of highly polymorphic multilocus MHC system. Molecular Ecology Resources, 9, 713–719

Bishop R, Nene V, Staeyert J, Rowlands J, Nyanjui J, Osaso J, Morzaria S, Musoke A (2003) Immunity to East Coast fever in cattle induced by a polypeptide fragment of the major surface coat protein of *Theileria parva* sporozoites. Vaccine. 21:1205–1212

Ellis S (2004) The cattle major histocompatibility complex: is it unique? Vet Immunol Immunopathol 102(1–2):1–8

MacHugh DE, Shriver MD, Loftus RT, Cunningham P, Bradley DG (1997) Microsatellite DNA variation and the evolution, domestication and phylogeography of taurine and zebu cattle (*Bos taurus* and *Bos indicus*). Genetics 146(3):1071–1086

Musoke A, Morzaria S, Nkonge C, Jones E, Nene V (1992) A recombinant sporozoite surface antigen of *Theileria parva* induces protection in cattle. Proc Natl Acad Sci U S A. 89:514–518

Musoke A, Rowlands J, Nene V, Nyanjui J, Katende J, Spooner P, Mwaura S, Odongo D, Nkonge C, Mbogo S, Bishop R, Morzaria S (2005) Subunit vaccine based on the p67 major surface protein of *Theileria parva* sporozoites reduces severity of infection derived from field tick challenge. Vaccine. 23:3084–3095

Nene V, Svitek N, Toye P, Golde WT, Barlow J, Harndahl M, Buus S, Nielsen M (2012) Designing bovine T cell vaccines via reverse immunology. Ticks Tick Borne Dis 3(3):188–192

Olds C, Mwaura S, Crowder D, Odongo D, van Oers M, Owen J, Bishop R, Daubenberger C (2012) Immunization of cattle with Ra86 impedes *Rhipicephalus appendiculatus* nymphal-to-adult molting. Ticks Tick Borne Dis. 3(3):170-8

Pandya M, Rasmussen M, Hansen A, Nielsen M, Buus S, Golde W, Barlow J (2015) A modern approach for epitope prediction: identification of foot-and-mouth disease virus peptides binding bovine leukocyte antigen (BoLA) class I molecules. Immunogenetics 67(11):691–703

Robinson J, Halliwell JA, McWilliam H, Lopez R, Marsh SG (2013) IPD—the immuno polymorphism database. Nucleic Acids Res 41(Database issue):D1234–D1240

Sette A, Sidney J (1999) Nine major HLA class I supertypes account for the vast preponderance of HLA-A and -B polymorphism. Immunogenetics 50(3–4):201–212

Thomsen M, Lundegaard C, Buus S, Lund O, Nielsen M (2013) *MHCcluster*, a method for functional clustering of MHC molecules. Immunogenetics 65(9):655–665

Summary

The dissemination of *Theileria parva* immortalized bovine lymphoblasts into lymphoid and non-lymphoid tissues results in East Coast Fever – a disease that continues to ravage cattle herds owned by resource poor small holder farmers and pastoralists throughout eastern, central, and southern Africa. The shortcomings of available control options are exemplified by failing acaricides and chemotherapeutics, and the technical and operational inadequacies of the live vaccination regime termed infection and treatment method (ITM). These constraints have provided impetus towards the development of subunit vaccines intended to render genetically diverse out-bred cattle populations immune to challenge by antigenically distinct parasites in the field throughout the endemic areas.

Schizont antigens complexed with cattle class I MHC induce $CD8^+$ cytotoxic T-lymphocytes (CTLs) that are the major effectors for immune control of *T. parva*. However, at present, there is insufficient information to predict the constraints to recombinant vaccine development based on induction of protective CTLs that may be imposed by functional divergence within cattle class I MHC genes. The first of the two broad studies described herein, utilized amplicon-based next generation sequencing combined with rigorous read processing algorithms to obtain reliable class I MHC genotypes from a field population of African native *Bos taurus* (Ankole). The study then leveraged progress in 'reverse immunology' to infer the extent of functional difference among the class I MHC alleles expressed by Ankole cattle as well as by exotic (Holstein) cattle. Finally, the study sought to ascertain if the dissimilarities seen in the *in silico* predictions of class I MHC peptide binding specificities between the two taurine lineages could be corroborated by *ex vivo* tests assaying cytokine responses to defined *T. parva* antigens.

The major findings from the first study included: (i) the identification of 18 novel cattle class I MHC allelic sequences in Ankole cattle, (ii) the evidence of positive selection for sequence diversity

including in residues that predominantly interact with peptides in Ankole class I MHC, (iii) the evidence from *in silico* functional analysis of peptide binding specificities that are largely distinct between the two breeds and (iv) the demonstration that $CD8^+$ T-cells derived from Ankole cattle that were seropositive for *T. parva* did not recognize vaccine candidate antigens originally identified in Holstein cattle breeds. This includes the immunodominant Tp1 which is currently the main focus of ongoing ECF recombinant vaccine trials. Taken together, the data clearly demonstrates that overlap between the peptide binding specificities of cattle class I MHC molecules is likely to be largely confined to alleles belonging to the same breed. These differences have the implication that a number of different antigens/epitopes will need to be incorporated in a $CD8^+$ T-cell based recombinant antigen cocktail vaccine to provide broad coverage.

In addition to the cellular responses, a contribution of humoral responses in mediating immunity against the sporozoite stage of the parasite has been documented. The closest approach to an effective anti-sporozoite vaccine to date has used a recombinant version of p67, the major sporozoite surface antigen of *T. parva* that is the target of antibodies that can potentially neutralize the establishment of infection. However, exposure of p67 immunized cattle to field tick challenge has resulted in very limited protection relative to experimental laboratory trials. The second study sought to shed light on the possible impacts of heterologous parasite challenge on the protection afforded to p67 immunized cattle. The concern was that whereas the gene encoding p67 is conserved in all cattle-derived populations of *T. parva*, it remained unclear whether parasites that originate from buffalo, and capable of causing severe disease in co-grazing cattle, contain diverse or invariant p67 genotypes. The primary analysis involved examining allelic variation, principally length polymorphisms and amino acid diversity in the closely juxtaposed B cell epitopes mapped to the central region of the p67 protein. The study also sought to shed light on the mechanisms that underpin molecular evolution of the p67 gene.

The findings included: (i) the identification of ten discrete p67 allelic sequences with an overall DNA polymorphism of 19.6%. This contrasts with the complete conservation of nucleotides at the p67 locus, including introns and the third base in codons, in cattle-transmissible isolates, (ii) the demonstration that the p67 allelic sequences described herein varied widely in size principally due to the 129 and 174 base pair deletions in the central region of the gene, (iii) the identification of an in-frame deletion and nonsynonymous substitutions in the two closely juxtaposed B cell epitopes in the central region of the p67 protein respectively, (iv) the demonstration of an evolutionary pattern within the *T. parva* p67 locus that is consistent with the effects of positive selection. In sum, these findings suggest that a p67 based vaccine suitable for field use, particularly where a wildlife reservoir of infection is present, should incorporate polymorphic epitopes as an improvement to the current conserved p67 vaccine antigen.

The totality of the data presented here provides a more informed picture of the development needs of subunit vaccines satisfying both the efficacy and coverage criteria. The key recommendation is that priority should be given to grouping cattle class I MHC alleles into functional ‘supertypes’, once extensive polymorphism data is collated so as to better inform the feasibility of generating $CD8^+$ T-cell based vaccines with a high population coverage regardless of 'breed' differences. It could also be helpful to re-assess the field performance of the p67 vaccine based on simultaneous immunization with vaccine constructs expressing the allelic variants described herein.

Zusammenfassung

Betrachtung der Genotypen von Wirt und Parasit in Bezug auf die Wahl von Subunit-Impfstoffantigenen zur Immunisierung von Rindern gegen *Theileria parva*

Theileria parva, der Erreger des Ostküstenfiebers, verursacht im östlichen, zentralen und südlichen Afrika eine Erkrankung von großer wirtschaftlicher Bedeutung, wobei die an Ressourcen armen Kleinbauern und Pastoralisten am meisten darunter zu leiden haben. Der Erreger dringt in Lymphozyten der Rinder ein und ruft eine unkontrollierte Teilung und Immortalisierung der infizierten Wirtszellen Zellen hervor. Das Fehlen geeigneter Kontrollmaßnahmen wird durch umweltschädliche Akarizide und Chemotherapeutika sowie durch technische und operative Unzulänglichkeiten der Immunprophylaxe („infection and treatment method - ITM“) verdeutlicht. Diese Beschränkungen haben die Entwicklung von Subunit-Impfstoffen gefördert, um in endemischen Gebieten genetisch unterschiedliche Rinderpopulationen gegen eine Infektion durch antigenetisch verschiedene Feldstämme zu immunisieren.

Schizonten-Antigene, die in Form von Petiden von MHC-Klasse-I-Molekülen infizierter Rinder-Zellen präsentiert werden, induzieren $CD8^+$ zytotoxische T-Lymphozyten (CTLs), die als Haupteffektoren der Immunkontrolle gegen *T. parva* fungieren. Derzeit gibt es allerdings ungenügende Informationen um die Einschränkungen für die Entwicklung eines rekombinanten Impfstoffes basierend auf der Induktion von protektiven CTLs vorherzusagen, die durch die funktionelle Divergenz des Rinder-Klasse I MHC bedingt sind. Um zuverlässige Klasse I MHC Genotypen von einer Feldpopulation des in Afrika beheimateten *Bos taurus* Rindes (Ankole) zu erhalten, benutzte diese Studie Amplikon-basierte Sequenzierungen kombiniert mit rigoros gefilterten Sequenz-Algorithmen. Außerdem, nutzte die Studie den erzielten Fortschritt in der „ reverse Immunologie“, um das Ausmaß der funktionalen

Unterschiede zwischen den Klasse I MHC Allelen, die durch die Ankole und exotische (Holstein-) Rinder exprimiert wurden, zu bestimmen. Schließlich wurde versucht festzustellen, ob die beobachteten Unterschiede in den *in silico*-Vorhersagen der Klasse I MHC Peptid-Bindungsspezifitäten zwischen den beiden taurinen Linien durch die Zytokinbildung durch CTLs auf eine Stimulation mit definierten *T. parva*-Antigenen bestätigt werden können.

Die wichtigsten Ergebnisse dieser Studie waren: (i) die Identifizierung von 18 neuen Klasse I MHC Allelsequenzen bei Ankole Rindern, (ii) der Nachweis für eine positive Selektion für Sequenzvielfalt des Ankole Klasse I MHC, einschließlich der Positionen, die vorwiegend mit Peptiden interagieren, (iii) die Beweise aus der *in silico*-funktionellen Analyse von Peptidbindungsspezifitäten, die sich weitgehend zwischen Ankole und Holstein Rinderrassen unterscheiden, und (iv) dass $CD8^+$ -T-Zellen, die aus Ankole-Rindern stammten, die für *T. parva* seropositiv waren, keine Impfkandidaten-Antigene erkannten, die ursprünglich in Holstein Rindern identifiziert wurden.

Dies schließt das immun-dominante Tp1 Antigen ein, das der Schwerpunkt der laufenden rekombinanten Impfstoff-Studien gegen Ostküstenfieber ist. Zusammengenommen zeigen die Daten eindeutig funktionelle Unterschiede zwischen den in verschiedenen Rinderrassen exprimierten Klasse I MHC Genen und deuten stark darauf hin, dass eine Anzahl von verschiedenen Antigenen / Epitopen in einen rekombinanten Antigen-Cocktail-Impfstoff auf $CD8^+$ -T-Zellen integriert werden müssen, um eine breiten Impfschutz zu erzielen.

Zusätzlich zur Rolle der CTLs wurde auch die Bedeutung der humoralen Immunantwort bei der Schutzvermittlung gegen die Sporozoite des Parasiten analysiert. Der naheliegenste Ansatz für einen wirksamen Anti-Sporozoiten-Impfstoff war bisher die Verwendung einer rekombinanten Version von p67, dem Haupt-Sporozoiten-Oberflächen-Antigen von *T. parva*, dass das Angriffsziel von Antikörpern ist, die möglicherweise die Entstehung einer Infektion neutralisieren können. Jedoch führte in einem Feldversuch eine Zeckenexposition bei mit p67 immunisierten Rindern nur zu einem sehr beschränkten Schutz im Vergleich zu experimentellen Labortests.

Die zweite Studie untersuchte die möglichen Auswirkungen eines heterologen Parasitenchallenges auf den Schutz von mit p67 immunisierten Rindern. Während das für p67 kodierende Gen in allen in Rindern vorkommenden *T. parva* Populationen konserviert ist, ist nicht bekannt, ob die aus dem Büffel stammenden Parasiten, die eine schwere Erkrankung bei Rindern verursachen, verschiedene oder unveränderte p67-Genotypen enthalten. Die primäre Analyse befaßte sich zunächst mit Allele-Variationen, vor allem Längenpolymorphismen und Aminosäurevielfalt in den eng benachbarten B-Zell-Epitopen, die der zentralen Region des p67-Proteins zugeordnet sind. Ein weiteres Ziel der Studie war die molekularen Mechanismen der Evolution des p67-Gens aufzuklären.

Hierbei wurden folgende Ergebnisse erzielt: (i) die Identifizierung von zehn diskreten p67-Allelsequenzen mit einem Gesamt-DNS-Polymorphismus von 19,6%. Dies steht im Gegensatz zu der vollständigen Konservierung von Nukleotiden am p67-Locus, einschließlich Introns und der dritten Base in Codon, in von Rindern übertragbaren Isolaten, (ii) der Nachweis, dass die hier beschriebenen p67-Allelsequenzen hauptsächlich aufgrund der 129- und 174-Basepaar-Deletionen im zentralen Bereich des Gens variierten, (iii) die Identifizierung einer In-Frame-Deletion und Substitutionen, die zu

Aminosäureveränderungen in den zwei eng benachbarten B-Zellepitopen im zentralen Bereich des p67-Proteins führen, (iv) die Demonstration eines Evolutionsmusters innerhalb des *T. parva* p67-Lokus, was im Einklang mit den Effekten der positiven Selektion steht. Zusammenfassend deuten diese Befunde darauf hin, dass ein p67-basierter Impfstoff, der für den Feldgebrauch geeignet ist, insbesondere wenn ein Wildtierreservoir der Infektion vorhanden ist, die polymorphen Epitope als eine Verbesserung des gegenwärtig konservierten p67-Impfstoffantigens enthalten sollte.

Basierend auf diese Daten ist zu empfehlen die Gruppierung von Rinder-Klasse-I-MHC-Allelen in funktionelle "Supertypen" Priorität zu erhalten, sobald umfangreiche Polymorphismus-Daten zusammengetragen sind, um die Herstellung von $CD8^+$ T-Zell-basierten Impfstoffen mit einem breiten Anwendungsbereich zu verbessern, unabhängig von den unterschiedlichen Rinderrassen, zu ermöglichen. Es könnte auch hilfreich sein, die Feldeignung des p67-Impfstoffes, basierend auf der gleichzeitigen Immunisierung mit Impfstoffkonstrukten, die die hier beschriebenen Allelvarianten exprimieren, neu zu bewerten.

List of publications

Peer-reviewed journal articles

Obara I, Nielsen M, Jeschek M, Nijhof A, Mazzoni CJ, Svitek N, Steinaa L, Awino E, Olds C, Jabbar A, Clausen PH, Bishop RP (2016) Sequence diversity between class I MHC loci of African native and introduced *Bos taurus* cattle in *Theileria parva* endemic regions: in silico peptide binding prediction identifies distinct functional clusters. Immunogenetics 68: 339

Obara I, Ulrike S, Musoke T, Spooner PR, Jabbar A, Odongo D, Kemp S, Silva JC, Bishop RP (2015) Molecular evolution of a central region containing B cell epitopes in the gene encoding the p67 sporozoite antigen within a field population of *Theileria parva*. Parasitol Res 114: 1729

Noyes H, Brass A, **Obara I**, Anderson S, Archibald AL, Bradley DG, Fisher P, Freeman A, Gibson J, Gicheru M, Hall L, Hanotte O, Hulme H, McKeever D, Murray C, Oh SJ, Tate C, Smith K, Tapio M, Wambugu J, Williams DJ, Agaba M, Kemp SJ (2011) Genetic and expression analysis of cattle identifies candidate genes in pathways responding to Trypanosoma congolense infection. Proc Natl Acad Sci U S A 31;108(22):9304-9

Published Abstracts

Obara I, Nielsen M, Jeschek M, Nijhof A, Mazzoni CJ, Svitek N, Steinaa L, Awino E, Olds C, Jabbar A, Clausen PH, Bishop RP (2016) Recombinant vaccine development for T. parva control based on induction of protective $CD8^+$ T-cells: constraints imposed by class I MHC diversity. The first Joint International Conference of the Association of Institutions for Tropical Veterinary Medicine (AITVM) and the Society of Tropical Veterinary Medicine (STVM) at the Humboldt Universitaet Berlin, 4 – 8 September 2016, Programme & Abstract band ISBN 978-3-86345-338-1, page 106

Nanteza A, **Obara I,** Lubega G (2016) Neutral and adaptive genetic diversity in field isolates of *Theileria parva* derived from Ankole cattle in southwestern Uganda: Implications for the deployment of the live vaccine in Uganda. The first Joint International Conference of the Association of Institutions for Tropical Veterinary Medicine (AITVM) and the Society of Tropical Veterinary Medicine (STVM) at the Humboldt Universitaet Berlin, 4 – 8 September 2016, Programme & Abstract band ISBN 978-3-86345-338-1, page 105

Obara I, Nijhof A, Karl-HZ, Kurt P,Odongo D, Lubega G, Gwakisa P, Hussein A, Salah L, Amira E, Darghouth M, Shawgi H, Idris A, Agol Malak K, Wani M, Rabei ES, Ulrike S, Ahmed J, Bishop R, Clausen, PH (2015). Cattle class I MHC diversity and $CD8^+$ T-cell responses to *Theileria parva.* 25th International Conference of the World Association for the Advancement of Veterinary Parasitology, Liverpool, United Kingdom, 16-20 August 2015.

Obara I, Nijhof A, Karl-HZ, Kurt P, Odongo D, Lubega G, Gwakisa P, Hussein A, Salah L, Amira E, Darghouth M, Shawgi H, Idris A, Agol Malak K, Wani M, Rabei ES, Ulrike S, Ahmed J, Bishop R, Clausen, PH (2015). Molecular epidemiology network for promotion and support of delivery of live vaccines against *Theileria parva* and *Theileria annulata* infection in Eastern and Northern Africa. Annual meeting of the German Society for Veterinary Medicine (DVG; Working group

Parasitology), Stralsund, Germany, 29 June - July 1 2015, Abstract band ISBN 978-3-86345-261-2, pages 77-78

Obara I, Noyes H, Gicheru M, Agaba M, Kemp S (2011). ARHGAP15 allelic heterogeneity and the pathophysiology of trypanosomiasis associated anaemia. Advances in the Knowledge of Parasite Resistance of Ruminant Hosts and Parasites Course, Embrapa Pecuária Sudeste, São Carlos, São Paulo, Brazil

Obara I, Noyes H, Gicheru M, Agaba M, Kemp S (2010).Genotypes at polymorphisms in the trypanotolerance candidate gene ARHGAP15 and the pathophysiology of trypanosomiasis elicited anaemia. International symposium on animal Genomics for Animal Health, 31 May - 2 June 2010, Paris, France

Obara I, Noyes H, Gicheru M, Agaba M, Kemp S (2009). Characterization of Trypanotolerance Candidate Genes: Determination of the Kinetic Parameters of ARHGAP15 Allelic Variants Stimulated RAC1-GTP Hydrolysis. Plant & Animal Genomes XVII Conference San Diego, CA

Disclosure of contribution to the intellectual content

Prof. Dr. Jabbar S. Ahmed and Prof. Dr. Richard Bishop jointly conceived the study and were both lead applicants on the grant awarded by DFG that supported this study. Prof. Dr. Peter-Henning Clausen was responsible for overseeing the study and made significant contributions to the design, acquisition and interpretation of data. Isaiah Otieno Obara designed the experiments, carried out the molecular biology procedures including generating next generation sequence data and cloning selected genes, carried out the immunological assays, compiled, analysed and interpreted the data and drafted both manuscripts. Dr. Camila Mazzoni and Marie Jeschek contributed to discussions on algorithms for next generation sequence analysis. All co-authors contributed to revising the manuscripts for intellectual content and approved the published articles.

Acknowledgements

I am indebted to a number of individuals and organizations whose efforts were instrumental to the completion of this study. I want to particularly express my gratitude to Prof. Dr. Peter-Henning Clausen and Prof. Dr. Jabbar S. Ahmed, Freie Universität Berlin, who continually provided invaluable support and guidance over the period of study. I also wish to express my appreciation to the International Livestock Research Institute (ILRI), in particular Prof. Dr. Richard Bishop, for the numerous suggestions throughout the design and implementation of this study. I have also enjoyed research discussions with Dr. Ard Nijhof, Freie Universität Berlin, and I wish to acknowledge the contributions that his observations and shared experiences made to this work. Much of the bioinformatics research was dependent upon expertise at the Berlin Center for Genomics in Biodiversity Research (BeGenDiv). I am particularly grateful to Dr. Camila Mazzoni, the head of Bioinformatics, at BeGenDiv who made valuable contributions to this work and to her staff, Marie Jeschek for excellent technical assistance. I am also grateful to scientists at the Center for Biological Sequence Analysis, Technical University of Denmark, in particular Dr. Morten Nielsen, for the training in Immunological Bioinformatics.

The cooperation of the field and administrative staff at Ol Pejeta - a ranch that integrates game conservancy with cattle production in central Kenya, is gratefully acknowledged. The level of support from the Dahlem Research School (DRS) – a strategic center for junior researchers at Freie Universität Berlin, has been outstanding. The death of Prof. Dr. Ulrike Seitzer, a leading International figure in parasitology research who enlightened our scientific discussions at the onset of this study was a sad event.

The work was supported by a grant awarded by the Deutsche Forschungsgemeinschaft (DFG) to the project ’Molecular epidemiology network for promotion and support of delivery of live vaccines against *Theileria parva* and *Theileria annulata* infection in Eastern and Northern Africa’.

Statement of Authorship

Except where reference is made in the text, this thesis contains no material published elsewhere or extracted in whole or in part from any previous work presented by me for another degree or diploma. No other persons work has been used without due acknowledgement in the main text of the thesis. This thesis has not been submitted for the award of any other degree or diploma in any other tertiary institution.

Berlin, 04.05.2017 Isaiah Otieno Obara