THE UNIVERSITY *of* EDINBURGH

Antigenic diversity in *Theileria parva* in vaccine stabilate and African buffalo

Johanneke Dinie Hemmink

PhD

University of Edinburgh

2014

I declare that the work presented in this thesis is my own original work, except where specified, and it does not include work forming part of a thesis presented successfully for a degree in this or another university.

Johanneke Dinie Hemmink

Edinburgh 2014

Abstract

Theileria parva is a tick-borne intracellular protozoan parasite which infects cattle and African buffalo in Eastern and Southern Africa. Cattle may be immunised against *T. parva* by the infection and treatment method (ITM), which involves inoculation with live sporozoites and simultaneous treatment with oxytetracycline. One such ITM vaccine is the Muguga Cocktail, which is composed of a mixture of three parasite stocks: Muguga, Serengeti-transformed and Kiambu 5. Although the vaccine has been used with success in the field in several areas in Eastern Africa, there is evidence that vaccination using cattle-derived parasites does not always provide adequate protection against buffalo-derived *T. parva*. A number of *T. parva* antigens recognised by $CD8^+$ T cells from cattle immunised by ITM have been identified in previous studies. A proportion of these antigens show a high degree of sequence polymorphism and allelic diversity is believed to be much greater in buffalo-derived *T. parva* than in cattle-derived parasites. The present study focussed on the development and application of a deep sequencing technique for characterising genotypically heterogeneous *T. parva* DNA samples. A panel of genes encoding $CD8^+$ T cell antigens was used as the basis of a multi-locus sequence typing system (MLST) built upon Roche 454 amplicon sequencing technology. This system was validated using parasite stocks of known composition and then utilised to investigate genetic and antigenic diversity in vaccine stabilates and samples derived from African buffalo. The MLST profile obtained for the Muguga Cocktail stocks was compared to those of African buffalo in two geographically separated sites and was also compared with micro/mini-satellite DNA profiles of Muguga Cocktail stocks.

The three components of the *T. parva* Muguga Cocktail vaccine were found to have limited genotypic and antigenic diversity using both methods. The composition of vaccine batches produced in a single production run (ILRI0801-ILRI0804) was shown to be relatively consistent. In contrast, the composition of the component stocks was shown to alter following passage through cattle and ticks. The deep multi-locus sequence profile and satellite DNA profile established in this study may be used as a reference for comparison with future vaccine batches. It is suggested that

formulation of a new cocktail vaccine containing three parasite clones selected on the basis of genotypic and antigenic divergence may well provide protection comparable to that obtained with the Muguga Cocktail. The components of such a vaccine could readily be distinguished and the composition of vaccine batches monitored, thus allowing improved quality control and greater consistency of the vaccine.

Genetic and antigenic diversity was found to be very high in parasite populations from African buffalo from the Kruger National Park, South Africa and the Ol Pejeta conservancy, Kenya. The estimated average genetic 'distance' between any two alleles in the Kruger National Park and within the Ol Pejeta conservancy was very similar for all six genes investigated. Many of the identified alleles were 'private' to either the buffalo from Ol Pejeta or the Kruger National Park and many of these alleles were present in several individuals in one location. Principal co-ordinate analysis and phylogenetic investigation of several antigen-encoding loci indicated that extant buffalo parasite populations are geographically sub-structured although some of the underlying diversity may reflect 'ancient' polymorphism in an ancestral population.

A subset of the $CD8^+$ T cell antigens examined exhibited extensive antigenic polymorphism while others were highly conserved at the amino acid level. These conserved genes may represent good candidates for the development of next generation vaccines, as strain specificity may be overcome if protective $CD8^+$ T cell responses could be generated against these conserved antigens. This would enable the use of sub-unit vaccines in areas where cattle co-graze with buffalo.

Theileria sp (buffalo) was identified in cell lines isolated from cattle, indicating that this parasite can transform bovine lymphocytes and may therefore be implicated in pathology in cattle. Phylogenetic analysis of *T. parva* and *T. sp (buffalo)* clones using the 5S subunit ribosomal RNA gene, Tp6, Tp7 and Tp8 showed a clear distinction between the two parasite species. These genes could thus be considered as candidates for an improved diagnostic test for *T. parva* in South Africa.

Acknowledgements

Firstly, I would like to thank my supervisor Ivan Morrison, for all the advice, help and discussions about my work. I also would like to thank all the other members of the group. Especially the coffee breaks in the morning gave me the feeling to be part of a group. The discussions about science, life and random subjects often revived me. I would also like everyone at the University of Glasgow for welcoming me on several occasion and Willie Weir for his great input into my PhD work. The days scheduled of tryps to help me with bioinformatics and population genetics are greatly appreciated. Your helpful and positive encouragement is much appreciated.

I am also grateful I had the opportunity to spend time at ILRI to work on several aspects of my work. It is great to feel so welcome in another country. Thanks everyone on campus and all others I interacted with outside the campus. I especially like to thank the people that helped me in the lab or with the animal work: Elias, Rosemary, Thomas and Stephen. Not to forget Nick for hosting me during one of my visits.

Thanks to all the people I have lived with over the years. It was a pleasure sharing a house with you and to be able to share the ups and downs of life with you. I also would like to thank my family for all their support and beliefs in me. Especially, my dad had an unbelievable faith I would succeed (RIP). I also would like to thank all my friends in the UK, the Netherlands and elsewhere in the world for their friendship and support in sometimes difficult times. Lastly, I would like to thank Dennis for all his support and love for the time we have known each other.

Table of Contents

List of figures

List of Tables

Abreviations

°C	Degrees celsius
µg	micro gram
µl	micro litres
^{111}In	Indium radioactive isotope 111
AMV-RT	Avian Myeloblastoma virus reverse transcriptase
bp	base pair
BSA	Bovine serum albumin
CD	Cluster of differentiation
cDNA	Complementary deoxyribonucleic acid
CO_2	carbon dioxide
CTTBD	Center for Ticks and Tick-borne Diseases
CTVM	Centre for Tropical Veterinary Medicine
dATP	2'-deoxyadenosine triphosphate
dCTP	2' deoxycytidine triphosphate
dGTP	2' -deoxyguanosine triphosphate
DMSO	Dimethylsulphoxide
dNTP	Mixture of dATP, dCTP, dGTP, dTTP
dTTP	2' deoxythymidine triphosphate
EAVRO	East Africa veterinary research organisation
ECF	East coast fever
ELISA	enzyme-linked immunosorbent assay
ELISPOT	enzyme-linked immunosorbent spot
ER	endoplasmic reticulum
ERAAP	endoplasmic reticulum aminopeptidase associated with antigen processing
FACS	Fluorescent-activated cell sorting
FAO	Food and Agriculture Organization

FCS	Foetal calf serum
FITC	fluorescent isothiocyanate
g	grams
GUTS	Ground up tick stabilate
hr	hour
IFAT	Immunofluorescent antibody test
IFN	Interferon
Ig	immunoglobulin
ILRAD	Internationl laboratory for research on animal diseases
ILRI	Internation Livestock Research Institute
ITM	Infection and treatment method
mAbs	Monoclonal Antibodies
MACS	Magnetic- activated cell sorting
MEM	Minimum Essential Media
MHC	Major histocompatibility complex
MHC I	Major histocompatibility complex class I
MHC II	Major histocompatibility complex class II
MID	Multiplex Identifier
min	minutes
ml	mili litres
MLST	Multi-locus sequence typing
mRNA	messenger ribonucleic acid
ms	micro satellite DNA marker
MS	mini satellite DNA marker
ng	nano grams
NGS	Next-generation sequencing
NK	Natural killer
PBMC	peripheral blood mono-nuclear cells
PBS	Phosphate buffered saline
PCA	Principal component analysis

PCR	Polymerase chain reaction
PIM	polymorphic immunodominant molecule
pmol	pico molar
RFLP	Restriction fragment length polymorphism
rh IL-2	recombinant human interleukin two
RLB	Reverse line blot
rpm	revolutions per minute
RPMI	Roswell Park Memorial Institute medium
rRNA	ribosomal RNA
RT	reverse transcriptase
RT-PCR	reverse transcriptase- Polymerase chain reaction
s	seconds
TAE	Tris-acetate-EDTA buffer
TAP	Transporter associated with antigen processing
TCR	T cell receptor
TCR Vβ	T cell receptor beta chain variable gene segment
TCRβ	T cell receptor beta chain
Tp	*Theileria Parva* antigen
TpM	*Theileria parva* (Muguga)- infected cell
x g	force of gravity

Chapter 1: General introduction

1.1 *Theileria parva*

The tick-borne protozoan parasite *Theileria parva* causes disease in cattle in Eastern and Southern Africa. The disease is a major constraint on the development of the livestock industry in affected areas as it is associated with high mortality rates together with economic losses sustained through treatment and prevention measures (Minjauw and McLeod, 2003; Nambota et al., 1994). Both indigenous (*Bos indicus*) and exotic breeds (*Bos taurus*) are susceptible to infection with *T. parva* (Ndungu et al., 2005). However, the exotic taurine breeds, which are in demand for high production, are more susceptible to tick infestation and tend to develop more severe disease (Wambura et al., 1998)

1.1.1 Life-cycle

The main vector transmitting *Theileria parva* is *Rhipicephalus appendiculatus*, but a closely related species *Rhipicephalus zambeziensis* also plays an important role in transmission in Southern Africa. *R. appendiculatus* is a three-host tick and transmission of *T. parva* occurs transstadially from larvae to nymphs or from nymphs to adult. The transmission of *T. parva* from nymph to adults is quantitatively more important (Anonymous, 1981; Blouin et al., 1989; Jongejan et al., 2004; Lawrence et al., 1983; Potgieter et al., 1988). The tick vector infects cattle by inoculating sporozoites while taking a blood meal. The non-motile sporozoites are released several days after the tick commences feeding and rapidly enter lymphocytes by receptor-mediated endocytosis; this is believed to occur at the site of the tick bite (Fawcett et al., 1984; Fawcett et al., 1982b). The sporozoite surface coat is shed during the entry process, resulting in a more close apposition of the parasite and host cell membranes (Shaw, 2003). The *Theileria* parasites then rapidly escape into the cytoplasm of the host cell by dissolution of the surrounding host cell membranes (Shaw et al., 1991). *T. parva* infects lymphocytes of the T- and B-cell lineages, but infected T cells dominate *in vivo* and in *in vitro* cell lines (Baldwin et al., 1988).

Development of the sporozoites into the intracellular multinucleate schizont stage, which occurs over 3-4 days, results in activation and proliferation (transformation) of the infected cells. Division of the parasite is synchronised with division of the host cell, enabling the parasites to remain in an intracellular position while replicating (Shaw, 2003). The schizont-infected cells disseminate throughout the lymphoid system and other tissues, such as the lungs and intestine (Siefert, 1996). Rapid multiplication of parasitised cells is followed by extensive lymphocytolysis, affecting infected and uninfected cells, and profound leukopenia (Morrison et al., 1981). The most important clinical signs of *T. parva* infection are pyrexia, enlarged lymph nodes and pulmonary oedema, which develops rapidly, usually 1-2 days before death. Clinical signs associated with disease occur within two to four weeks of infection, with lung pathology being the primary cause of death (Siefert, 1996).

After about ten days, a proportion of the multinucleated schizonts undergo a differentiation process called merogony, which coincides with reduced proliferation of the host cell, enlargement of the schizonts with increasing numbers of nuclei and eventually production of uni-nucleated merozoites through a budding process from the enlarged schizont (Shaw et al., 1992). The infected cells lyse releasing mature merozoites into the bloodstream where they enter erythrocytes to give rise to piroplasms (Shaw et al., 1995). The piroplasm stage of *T. parva* does not multiply. Animals that recover from infection can continue to carry small numbers of parasites (schizonts and piroplasms) for years (Kariuki et al., 1995; Odongo et al., 2010; Oura et al., 2007b; Oura et al., 2004b; Skilton et al., 2002). The piroplasms in these carrier animals, although often undetectable microscopically, are infective for feeding ticks (Shaw, 2003; Siefert, 1996).

All parasite stages in cattle have a haploid genome. Diploid zygotes are formed in the tick gut through fusion of male and female gametes, which develop from ingested piroplasms (Gauer et al., 1995). Feeding ticks are likely to ingest a mixture of parasite genotypes, as mixed infections are common in cattle under field conditions (Muleya et al., 2012; Oura et al., 2005). This allows sexual recombination during development within the tick vector (Katzer et al., 2006) and can lead to the exchange of alleles of genes encoding polymorphic antigenic determents. The zygotes invade

gut epithelial cells of the tick and develop into motile kinetes, which enter the haemolymph and migrate to the tick's salivary glands. The kinetes enter a specialised cell type (the E cell) within the salivary gland, in which each kinete transforms into a sporoblast and subsequent sporogony is initiated by the commencement of a tick feed (Fawcett et al., 1982a). Sporogony in the salivary gland, which involves several nuclear divisions, results in the production of thousands of sporozoites (Fawcett et al., 1982a; Shaw, 2003).

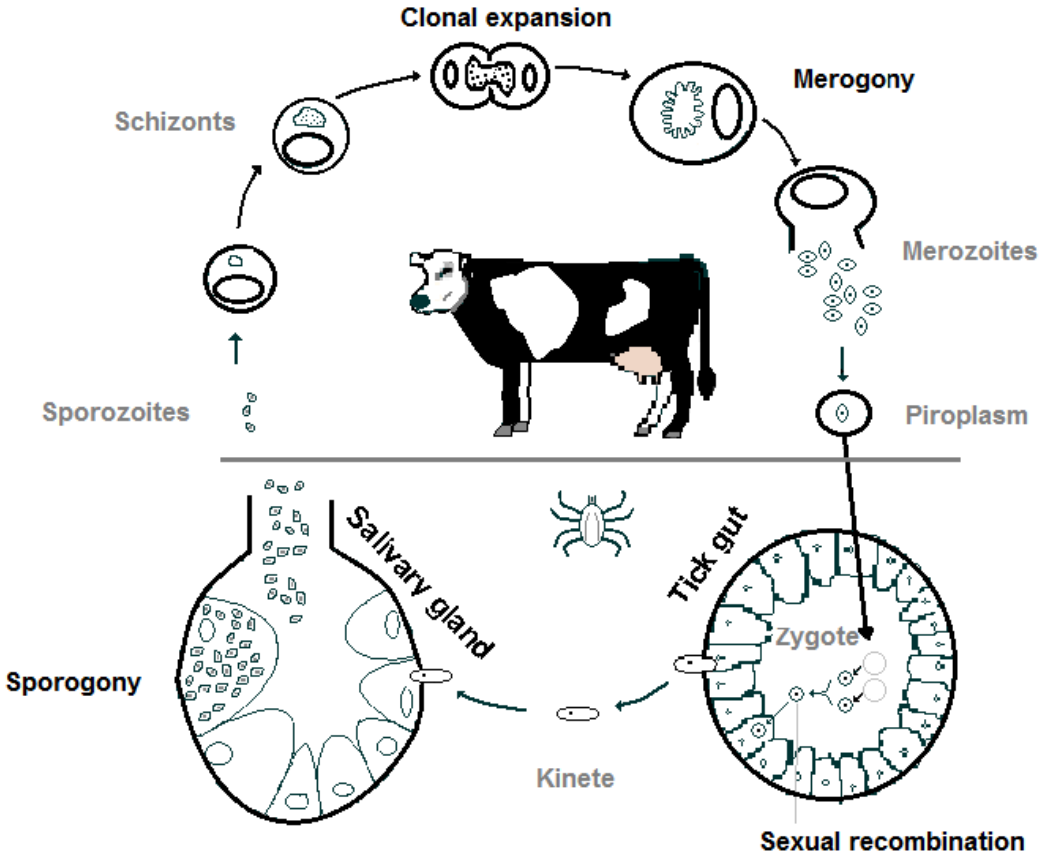

Figure 1.1 Life-cycle of *Theileria parva*

Sporozoites are inoculated into cattle by the tick host and invade monocytes of the T cell and B cell lineages. The parasites develop into schizonts within the cytoplasm of the host cell and induce clonal expansion of the host cell. Merozoites are produced during merogony, which upon release from the host cell invade erythrocytes. Ticks ingest erythrocytes containing piroplasms when feeding on an infected animal. Diploid zygotes are formed, which invade the ticks gut epithelial cells and develop into motile kinetes, which migrate to the tick's salivary glands where sporogony results in the production of thousands of sporozoites.

1.1.2 Clinical signs and nomenclature

The classical presentation of the disease in cattle caused by *Theileria parva* has been named East Coast fever (ECF) and is characterised by anorexia, fever, enlarged lymph nodes containing large number of parasitised cells and breathing difficulties. High mortality rates have been observed in naïve cattle with death ensuing two to four weeks after infection.

African buffalo can carry *T. parva* and may play a role in disease epidemiology in areas where buffalo are present. Cattle infected with buffalo-derived parasites develop severe disease rapidly, despite having relatively low levels of parasites in superficial lymph nodes and blood (Potgieter et al., 1988). In South Africa, disease in cattle caused by buffalo-derived *T. parva* was named Corridor disease, because of its occurrence in a corridor between two game reserves (Norval et al., 1992). Transmission of buffalo-derived parasites from cattle to cattle is limited, due to the low levels of the tick-infective piroplasm stage and the lack of long term carrier infections in cattle (Lawrence, 1979).

While only the buffalo-associated form of disease occurs in South Africa, a milder, seasonal form of disease due to *T. parva* is seen in other regions of Southern Africa (e.g. Zimbabwe). This disease has been named January disease due to its seasonal occurrence after the cold dry season from January to March after which the rainy season commences. The milder nature of the disease and its seasonal occurrence has been proposed to be due to differences in the distribution of tick vectors and their ability to transmit *Theileria* (Norval et al., 1991b; Ochanda et al., 1988). In addition, ticks in Southern Africa can undergo diapause, a predetermined state of low activity and metabolism, which is dependent on changes in hours of daylight and other factors. This aspect of their biology, results in only a single generation of ticks per calendar year (Norval et al., 1991a; Young et al., 1994).

The differences in transmission dynamics and clinical manifestation of theileriosis have led to a variety of nomenclatures being introduced to describe the parasites responsible (Lawrence, 1977, 1979). At the beginning of the twentieth century, three different disease presentations of theileriosis transmitted by *Rhipicephalus*

appendiculatus were recognised and the causative pathogens of East coast fever, corridor disease and January disease were named *T. parva, T. lawrencei* and *T. bovis.* However, after the occasional successful transmission of *T. lawrencei* from cattle to cattle, the morphological similarities between the parasites (Young and Purnell, 1973) and the cross-reactivity of monoclonal antibodies raised against buffalo-derived *T. parva* with cattle derived *T. parva* (Conrad et al., 1989), it was suggested that the three parasites should be considered sub-species of *T. parva*; hence parasites causing the classical cattle disease were named *T. parva parva,* those derived from buffalo *T. parva lawrencei* and those causing January disease *T. parva bovis* (Lawrence, 1979; Uilenberg, 1978). The subsequent finding that these sub-species could not be distinguished genetically based on DNA typing methods then led to the recommendation that all parasites be named *T. parva* and that further descriptive terms may be used: cattle-derived *T. parva* and buffalo-derived *T. parva* (Perry et al., 1993). Cattle-derived *T. parva* includes the previously named *T. parva parva* and *T. parva bovis* (Anonymous, 1988; Norval et al., 1991a; Perry et al., 1993).

1.1.3 Diagnosis

Diagnosis in the field relies on clinical signs and, in some cases, is confirmed by microscopic detection of parasites in leukocytes (the schizont stage) or erythrocytes (the piroplasm stage) in Giemsa-stained smears of lymph node biopsies and/or blood (Minjauw et al., 2003; Siefert, 1996). However, when parasitaemia is low, it can be difficult to detect parasites and differentiation between the different haemoparasites by microscopy is challenging (Norval et al., 1992). Exposure to *T. parva* can also be assessed by measuring antibody responses. Initially an indirect immunofluorescent antibody test (IFAT) was employed using fixed smears of schizont-infected cells as antigen (Goddeeris et al., 1982). More recently this has been replaced with an antibody enzyme-linked immunosorbent assay (ELISA) using a recombinant protein of the polymorphic immunodominant molecule (PIM), a schizont surface protein that induces strong antibody responses (Katende et al., 1998). In South Africa, a real-time PCR assay based on the 18S rRNA gene is used for the detection of *T. parva* in African buffalo and cattle (Sibeko et al., 2008). Other methods developed for diagnostics under experimental settings include a nested PCR assay based on the

p104 gene and a reverse line blot (RLB) assay based on amplification of the 18S rRNA gene using Theileria/Babesia specific primers and hybridisation with species-specific oligonucleotide probes (Bishop et al., 2001; Odongo et al., 2010; Oura et al., 2004a; Tait and Oura, 2004).

1.1.4 Control of *T. parva*

Tick control is an important aspect of the management of tick-borne diseases, including *Theileria parva*. However, frequent application of acaricide is required to prevent transmission of the parasite and this demands organisation and expense. Intensive use of acaricide favours the development of acaricide-resistance (Jongejan and Uilenberg, 2004). Furthermore, there are issues with acaricides in terms of residues in the environment and in livestock products and there is a direct risk to the animal handlers (Morrison and McKeever, 2006; Mugisha et al., 2005). In situations where acaricides are used frequently, the cattle may remain fully susceptible to tick-borne diseases and disruption in the supply of acaricides can lead to devastating outbreaks of disease (Jongejan and Uilenberg, 2004). The use of acaricides is therefore viewed as an unsustainable method for the control of *T. parva* and other tick-borne diseases. A more integrated form of control for ticks and tick-borne disease is advocated, which includes a more strategic use of acaricides in combination with vaccination against ticks and/or tick-borne pathogens (Jongejan et al., 2004; Morrison and McKeever, 2006).

Affected animals can be treated successfully using theilericidal compounds such as parvaquone or buparvaquone, providing treatment is started in the early stages of clinical disease (Muraguri et al., 1999). In the later stages of the disease, several doses of drug may be required to cure the animal and recovery may be protracted. Although drugs can be effective, they are relatively expensive and may not always be available to poor smallholder farmers (Morrison and McKeever, 2006; Norval et al., 1992).

The policy of the South African government, where African buffalo are the source of infection, is focussed on prevention of cattle becoming infected with *Theileria parva* rather than control (Anonymous, 2002). Game reserves and game ranches are fenced

to prevent cattle and buffalo grazing on the same land. Quarantine measures are firmly enforced and these are coupled with strict tick control of cattle in areas bordering the disease-endemic areas, particularly when outbreaks occur. Spread of *T. parva* from African buffalo to cattle is also prevented by prohibiting the translocation of *T. parva*-carrying buffalo into non-endemic areas (Anonymous, 2002; Sibeko et al., 2008). Despite these measures, occasional small outbreaks of disease due to *T. parva* infection are still recorded in cattle (Mbizeni et al., 2013; Thompson et al., 2008; Yusufmia et al., 2010).

Cattle may be immunised against ECF using the 'infection and treatment' method (ITM), which involves infection with live sporozoites in the form of cryopreserved stabilates and simultaneous treatment with long-acting tetracycline (Radley et al., 1975b). Cattle immunised experimentally by this method are solidly immune against challenge with the same (homologous) parasite stock. However, immunity induced by one parasite stock does not always give full protection against challenge with other stocks (Hove et al., 1995; Radley et al., 1975a; Taracha et al., 1995a). To overcome this, the Muguga cocktail vaccine was developed in the 1970s and is composed of three *T. parva* stocks. These stocks were selected on the basis of experimental cross-immunity trials in cattle (Radley et al., 1975b). The Muguga cocktail has been used successfully to vaccinate cattle in the field (discussed further below).

Other *T. parva* stocks have also been used in different local regions in sub-Saharan Africa where *T. parva* is prevalent. The Katete and Chitongo stocks, are used for immunisation in Malawi and Zambia (Geysen et al., 1999; Marcotty et al., 2002). The Marikebuni stock, isolated from the coastal region of Kenya, has been used for immunisation in Kenya (Minami et al., 1983; Morzaria et al., 1987; Wanjohi et al., 2001) and the Boleni stock has been used for vaccination in Zimbabwe (Hove et al., 1995; Latif et al., 2011).

1.2 The Muguga cocktail

1.2.1 Introduction

The Muguga cocktail is the most widely used parasite preparation for vaccination by ITM. The three *T. parva* stocks incorporated in the vaccine are Muguga, Serengeti-transformed and Kiambu 5 (Radley et al., 1975b). The *T. parva* Muguga stock was isolated in the 1950s and was originally maintained at the East African Veterinary Research Organization (EAVRO) by acute passage through ticks and cattle before a stabilate was made in 1970. The Kiambu 5 stock was isolated in the early 1970s by tick pick-up from a cow with clinical East coast fever in the Kiambu district of Kenya (Radley et al., 1975a) The Serengeti-transformed stock was derived from an isolate from a buffalo in the Serengeti National Park, Tanzania by feeding ticks on an African buffalo that was captured from the Serengeti National Park as a calf in the early 1970s. This stock is said to be 'transformed' since following serial passage through cattle, the isolate produced infections characteristic of cattle-derived *T. parva* (Young and Purnell, 1973).

The Muguga cocktail has been used in experimental and field trials with encouraging but variable results. Protection was observed against a range of cattle-derived *T. parva* stabilates from different geographical areas (Radley et al., 1975b). A substantial reduction of the number of deaths due to theileriosis was also seen under field conditions in Kenya and Tanzania. However, fatality was still observed in immunised animals in the different experiments; between 0.2% and 20% of the animals in immunised groups died compared to a death rate of between 20% and 100% in groups of naïve animals (Radley, 1976). In other studies, breakthroughs were observed upon challenge with the Malawian *T. parva* Kasobo stock (Musisi et al., 1996) and the Zambian *T. parva* Katete stock (Geysen, 2000). Broad protection has been reported against field challenge under pastoral conditions in Tanzania where large numbers of animals have been immunised using the Muguga cocktail over the last two decades (Giulio et al., 2009; Martins et al., 2010; Morzaria and Williamson, 1999). However, the protection provided against challenge with buffalo-derived *T. parva* parasites is in most cases disappointing (Lohre, 1978; Radley et al., 1979).

Although the Muguga cocktail has been used with a degree of success in experimental and field conditions, large-scale uptake of the vaccine has been limited by complexities in production, standardisation of the vaccine production process and control of vaccine safety and consistency. Another major constraint to the delivery of vaccine stabilate is the requirement to maintain a liquid nitrogen cold chain for the transport of stabilates to the field. In some regions, there have also been concerns regarding the risk of introducing the vaccine strains into local tick populations should immunised animals become carriers (De Deken et al., 2007; Geysen, 2008; Geysen et al., 1999; Giulio et al., 2009; Kariuki et al., 1995; Oura et al., 2007a). The fear is that the introduced parasites would undergo sexual recombination with local strains, which could potentially create parasites with altered virulence or antigenicity (Morzaria and Williamson, 1999). Nevertheless, efforts have been made to commercially produce and distribute the Muguga cocktail and to promote its widespread use.

1.2.2 Vaccine stabilate production – Muguga Cocktail

The production process for the ITM vaccine (illustrated in Figure 1.2) is intricate and relies on passaging of the three component stocks through cattle and ticks. There are currently no guidelines for assaying the composition of the components and it is possible that stocks may vary at the genotypic level from generation to generation. This clearly has implications for vaccine consistency and quality control. It is important to appreciate the way in which the vaccine is manufactured. Immunising stabilates are derived from working seed stabilates that are in turn derived from master seed stabilates. In other words, a master seed parasite population will go on to complete the life-cycle twice before it forms an actual vaccine stabilate. Importantly, this method includes the potential for two rounds of sexual recombination in the tick. Unfortunately, the production of a live multi-component vaccine such as this is necessarily complex and the methodology used has been developed in order to minimise passage of the parasite stocks and to allow production of large quantities of vaccine stabilates.

Production of the stabilates is carried out as follows (OIE, 2008). Healthy cattle, serologically negative for tick-borne diseases, are infected by subcutaneous injection of the master seed stabilate. During the following parasitaemic phase, laboratory-raised *R. appendiculatus* nymphs are applied to the ears of the cattle in ear bags. The engorged ticks are collected, allowed to moult and then fed on the ears of rabbits for four days to allow maturation of sporozoites in the tick salivary glands (Kimbita and Silayo, 1997; OIE, 2008). The replete ticks are removed from the rabbit ears after four days and the infection rate of dissected salivary glands is determined in a sample of the ticks; the remaining ticks are used to produce the working seed (OIE, 2008) stabilates.

In order to produce a stabilate, batches of 1,000 infected ticks are washed under fast-flowing tap water and the surface of the ticks disinfected by washing with 1% benzalkonium chloride or 70% alcohol and rinsed with distilled water. The ticks are then ground up and deposited in 7.5% glycerol minimum essential media supplemented with bovine serum albumin (MEM/BSA) and the resulting ground up tick stabilates (GUTS) are aliquoted and frozen. Stabilate may be stored in glass vials or plastic artificial insemination straws. Artificial insemination straws have the benefit that colour coding can be used to indicate the identity of the immunising stabilate. Stabilates are frozen in a deep freezer within a container that facilitates a slow rate of freezing. Once frozen, the stabilates are transferred to liquid nitrogen for permanent storage. The same procedure is used to produce the immunising stabilate from working seed stabilates. However, ticks infected with each of the three components are mixed before grinding, to provide batches with equivalent infectious units of each component stock (OIE, 2008; Speybroeck et al., 2008). Normally, the mixture would be formulated to contain approximately ten tick-equivalents per ml, but this can be adjusted if the salivary gland infection rate in a particular tick batch is either very high or very low (OIE, 2008).

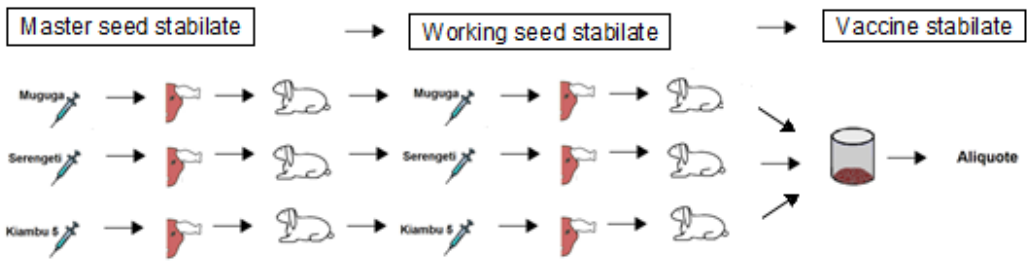

Figure 1.2 Schematic representation of the production process of the Muguga cocktail

Three groups of cattle are inoculated with a single component of the Muguga cocktail (Muguga, Serengeti or Kiambu 5; one component per group) and clean laboratory *Rhipicephalus appendiculatus* nymphs fed on the ears of these cattle. The resulting infected ticks are allowed to moult before being allowed to pre-feed on rabbits for four days in order to allow maturation of sporozoites. These ticks are then used to produce the working seed stabilates for each of the components in the form of cryopreserved ground up tick stabilate (GUTS). This process is then repeated using the working seed stabilates for each of the three component stocks. The resulting infected ticks, containing mature sporozoites, are pooled to provide batches for immunisation with equivalent infectious units of each of the components.

The infectivity of the immunising stabilates is determined by inoculating groups of susceptible cattle with a series of stabilate dilutions in a standard volume of one ml. It is essential to titrate the effective dose, despite the number of infected acini/tick being known prior stabilate production. This is because an unknown proportion of infectivity is lost during the stabilate production process and through freezing and thawing. For the titration, the contents of two to four randomly selected aliquots are mixed and cattle are then inoculated with different dilutions of the stabilate to establish its infectivity and lethality. The aim of this procedure is to determine an immunising dose of stabilate that, in combination with the tetracycline, reproducibly infects all animals but results in a very mild or inapparent clinical reaction. All animals should develop a parasite-specific antibody response after immunisation and be immune to lethal homologous parasite challenge. Once a safe and effective immunising dose is established, it is adopted for vaccination in the field (OIE, 2008).

1.2.3 Challenges in the vaccine production process

The production process is reliant on live animals and ticks and these potentially harbour pathogens that could contaminate the vaccine. By using laboratory-reared rabbits and ticks, this risk is greatly reduced. However, cattle are not sourced from a pathogen-free environment, which means there is always a risk of introducing pathogens during the production process. While a disease-free status cannot be guaranteed, only clinically healthy cattle, serologically negative for tick-borne diseases are used for vaccine production (OIE, 2008).

The dose for immunisation must be calculated on the basis of titration due to an unknown proportion of parasites losing infectivity during the production process. Determining the optimal dose for immunisation is further complicated by the varying susceptibility of cattle breeds. Exotic *Bos taurus* breeds, like the Holstein, develop more severe disease than local *Bos indicus* breeds and their crosses. In a vaccine trial with the Marikebuni stock at a location on the Kenyan coast, a higher proportion of Jersey than *Bos indicus* x *Bos Taurus* cross-bred cattle were observed to be clinical reactors following vaccination (Mutugi et al., 1991). Hence, cattle that are most likely to develop a severe reaction are generally used for titration of immunising stabilates, in order to minimise the chance of clinical reaction following immunisation in the field. The potential risk of using a lower dose of the stabilate is that parasite genotypes present at low frequency in the immunising stabilate, which may contribute to the broad immunity, may not feature at sufficient quantity in every vaccine dose. Furthermore, titration of a mixture of stocks of *T. parva* does not ensure optimal doses of each of the component stocks. However, results from titrating each of the component stocks separately could not necessarily be extrapolated to determine a safe dose of a vaccine mixture (Speybroeck et al., 2008).

Further losses of infectivity are known to occur when the vaccine is distributed in the field and this reduces the effectiveness of each immunising dose. For this reason, it is recommended that thawed stabilate be kept on ice and used within eight hours of de-frosting (Marcotty et al., 2001; Morzaria and Williamson, 1999).

1.3 Characterisation of vaccine stabilates

The Muguga cocktail was originally formulated by selecting parasites based on experimental cross-immunity trials conducted in the 1970s (Radley et al., 1975b). At that time, before molecular techniques were available, the only hope of ensuring consistency among different batches was by adhering to the production protocol. Despite these efforts, the consistency of different vaccine batches over the years is unknown in terms of both numbers of parasites per dose and the genotypic composition of the parasites.

With advances in immunology and molecular biology, several techniques have been used to characterise the three parasite stocks in the Muguga cocktail. Bishop *et al.* (2001) employed three techniques: an indirect fluorescent antibody test (IFAT) with a panel of anti-schizont monoclonal antibodies, southern blotting using four *T. parva* repetitive DNA probes and polymerase chain reaction (PCR)-based assays, detecting polymorphism within four single-copy genes encoding antigens recognised by antibody. Collectively, the data indicated that the Muguga and Serengeti-transformed components of the cocktail are genetically closely related but nevertheless distinct, while Kiambu 5 was highly distinct from the other two components (Bishop et al., 2001).

The availability of the *T. parva* genome sequence (Gardner et al., 2005) facilitated the development of a genome-wide panel of mini (MS) and micro (ms) satellite DNA markers for high-resolution genotyping of parasite populations. These were used for the analysis of the three components of the Muguga cocktail vaccine (Muguga, Serengeti-transformed and Kiambu 5); only four (MS7, MS19, MS25 and MS43) out of 31 markers used exhibited alleles that differed between the Muguga and Serengeti-transformed component stocks. Using similarity-based analysis, it was demonstrated that these two stocks are genetically very closely related (Oura et al., 2004b). An explanation for this could be that the prolonged passage through cattle selected those parasites within the Serengeti-transformed stock that can easily be maintained in cattle with the loss of other genotypes that transmit poorly between cattle. Other researchers have hypothesised that the original *T. parva* Serengeti-

transformed stock has been contaminated with and supplanted by *T. parva* Muguga genotypes over the years (Bishop et al., 2001; Oura et al., 2004b).

Direct analysis of the FAO1 and FAO2 batches of the Muguga Cocktail using six satellite markers provided evidence for at least six alleles at each locus (Oura et al., 2007b). Comparison of Muguga cocktail batches FAO1 and ILRI0804 produced by the Food and Agriculture Organization (FAO) and the International Livestock Research Institute (ILRI) demonstrated considerable commonality for the alleles identified using five satellite DNA markers. However, the ILRI Serengeti-transformed stock had an additional allele for four of the five markers compared to the FAO Serengeti-transformed stock. These additional alleles in the ILRI Serengeti-transformed stock had the same size as the alleles found in the Kiambu stock (for three of the markers) and the Muguga stock (the remaining marker) (Patel et al., 2011). If these are true differences between the Serengeti-transformed stocks, it means that the relative abundance of alleles in vaccine batches produced using different seed stabilates may change.

A more extensive set of satellite DNA markers was used to analyse the composition of the heterogeneous Marikebuni stock (Katzer et al., 2006) with differences in marker profiles found among different passages. A large number of genotypically distinct clones could be found when all the markers were used for analysis. However, for the majority of markers, a single dominant allele with one or two low abundance alleles were identified. Clonal analysis of the most recent passage of Marikebuni indicated that a specific clonotype had selectively expanded within the stock (Katzer et al., 2010). Direct genotyping of stabilates thus provides a reasonable indication of the extent of genetic diversity of the parasites present at relatively high frequency in the vaccine. However, it is possible that a higher resolution genotyping methodology would reveal additional alleles representing minor genotypic components of vaccine stabilates, which may contribute to the broad protection provided by the vaccine. The differences found by Patel *et al.* in the two different seed stabilates of the Serengeti-transformed stock (Patel et al., 2011) highlights the risk that the allelic composition of the components of the Muguga Cocktail may vary from generation to generation.

1.4 Characterisation of *Theileria parva* in the field

The same methods used for the characterisation of *T. parva* populations in vaccine stabilates have been used for the analysis of genotypic diversity of *T. parva* field isolates. Initial characterisation of *T. parva* isolates was performed using monoclonal antibodies raised against three different *T. parva* stocks, namely Muguga, Kiambu 5 and Marikebuni (Bishop et al., 1994; Conrad et al., 1989; Conrad et al., 1987). Isolates from African buffalo appeared to carry mixed *T. parva* populations and there were marked differences in antigenic diversity between isolates from different buffalo and between isolates from the same buffalo taken at different time-points (Conrad et al., 1987).

1.4.1 DNA probes and Restriction Fragment length polymorphism (RFLP)

Other methods used include southern blotting using radio-labeled probes on restriction enzyme digested genomic DNA and RFLP-PCR assays targeting polymorphic antigen-encoding loci (Bishop et al., 1993; Bishop et al., 1994; Conrad et al., 1989; Geysen et al., 1999). Extensive polymorphism was observed among *T. parva* isolates from a variety of geographical areas (Kenya, Tanzania, Zambia, Zimbabwe and Uganda) using repetitive ribosomal and telomeric DNA probes and anti-schizont monoclonal antibodies (mAbs) (Bishop et al., 1993). RFLP-PCR assays using the three antigen-encoding genes p104, p150 and PIM, were used for the analysis of Zambian isolates. The majority of the isolates had a similar hybridisation pattern, which differed from that of 'exotic' stocks such as the Muguga Cocktail components (Muguga, Serengeti-transformed, Kiambu 5) (Geysen et al., 1999). Different PIM PCR-RFLP profiles were observed in different samples from African buffalo from the Kruger National Park, whereas the profiles seen for samples from African buffalo from Hluhluwe-iMfolozi Park, Abalingwe Game Reserve and Ithala Game Reserve were much more similar to one another (Sibeko et al., 2011). RFLP-PCR profiles from Zambian cattle isolates were similar to one another but were distinct from those of the Muguga Cocktail components (Geysen et al., 1999). Relative homogeneity was observed among cattle isolates from Zimbabwe, as

assessed using RFLP analysis with the *T. parva* Tpr repetitive and ribosomal DNA probes and the panel of anti-schizont monoclonal antibodies (Bishop et al., 1994).

1.4.2 Satellite DNA markers

More recently, a higher resolution genotyping method was developed using a genome-wide panel of satellite DNA markers (Asiimwe et al., 2013; Katzer et al., 2011; Katzer et al., 2006; Katzer et al., 2007; Odongo et al., 2006; Oura et al., 2007b; Oura et al., 2004b; Oura et al., 2003; Oura et al., 2011c, d) Across the markers, a range of allele sizes were detected in samples from Zambia, Kenya, Zimbabwe and Uganda suggesting an overall high level of diversity in the parasite population. However, no relationship between genetic similarity and geographical proximity was observed (Oura et al., 2003). A high degree of genetic diversity has been found in *T. parva* parasites in Kenyan cattle from different regions with minimal genetic sub-structuring corresponding to geographical location (Odongo et al., 2006). In contrast, there was evidence for geographical sub-structuring in *T. parva* parasites in cattle in Uganda (Oura et al., 2005).

The *T. parva* parasites found in African buffalo from Queen Elizabeth National Park and Lake Mburo National Park were distinct, indicating there is limited gene flow between the populations (Oura et al., 2011a). A higher mean multiplicity of infection was observed in buffalo from Queen Elizabeth National Park compared to those from Lake Mburo National Park. These values were both much higher than for the multiplicity of infection encountered in cattle-derived isolates in Uganda (Oura et al., 2011a, b). Little genetic differentiation was observed between *T. parva* isolated from cattle within Lake Mburo National Park and cattle outside the park (Oura et al., 2011b)

1.4.3 Sequencing of antigens

Following the identification of $CD8^+$ T cell antigens (Graham et al., 2006), there has been an interest in investigating allelic polymorphism in their encoding-genes. For example, a large number of allelic sequence variants was found for the genes encoding Tp1 and Tp2 in buffalo-derived isolates and isolates from cattle associated with buffalo. This diversity was much more limited in cattle-derived parasites from

buffalo-free areas, both in terms of the number of alleles found and the differences between the different alleles. Allelic polymorphism was evident at the amino acid level in $CD8^+$ T cell epitopes identified within these genes (Pelle et al., 2011).

In addition, sequence analysis has been performed focussing on a number of B cell antigen-encoding genes, namely PIM, P67 and P104 antigen genes, for which there is evidence the proteins can be recognised by antibody responses (Graham et al., 2007b; Musoke et al., 2005; Sibeko et al., 2011; Sibeko et al., 2010; Toye et al., 1995). PCR amplification of the P67 gene of *T. parva* parasites in African buffalo from South Africa resulted in up to four different product sizes in individual animals and sequencing of the PCR products demonstrated further polymorphism (Sibeko et al., 2010).

To summarise, a high degree of diversity of *T. parva* can be found in the field. A range of *T. parva* genotypes have been detected in samples from cattle from different countries in Africa (Bishop et al., 1993; Geysen et al., 1999). Relative homogeneity was observed among cattle isolates from Zimbabwe (Bishop et al., 1994) and among cattle isolates from Zambia (Geysen et al., 1999) whereas isolates from Kenya and Uganda were found to be more heterogeneous (Odongo et al., 2006; Oura et al., 2005; Pelle et al., 2011). There was evidence for geographical sub-structuring in *T. parva* parasites in cattle in Uganda (Oura et al., 2005) although minimal geographical sub-structuring was found in parasites in cattle from different regions in Kenya (Odongo et al., 2006). The multiplicity of infection found in carrier buffalo is much higher than that found in cattle (Conrad et al., 1987b; Oura et al., 2011a; Pelle et al., 2011) and the alleles found in buffalo are more distinct from each other (Conrad et al., 1987b; Oura et al., 2011a; Pelle et al., 2011). The diversity of *T. parva* parasites in individual buffalo can vary from buffalo to buffalo and this may be associated with the sampling location (Oura et al., 2011a; Sibeko et al., 2011; Sibeko et al., 2010).

1.5 Next generation sequencing

Although a number of different methods have been used for the characterisation of the Muguga Cocktail and *T. parva* field isolates, these methods have been hampered by the limited level of resolution they could achieve. Next generation sequencing (NGS) technologies now offer the opportunity to investigate parasite samples in much higher resolution than was previously achievable. These technologies have the major advantage that a much higher level of throughput can be achieved at a greatly reduced cost per nucleotide compared to traditional Sanger sequencing. One million to three billion sequences may be obtained in a single instrument 'run' depending on the platform used. Several platforms are available including Roche 454 sequencing, Illumina Solexa sequencing and ABI Solid sequencing (biosystems, 2010; Illumina, 2012; Kircher and Kelso, 2010; Liu et al., 2012; Margulies et al., 2005; Roche)). Each platform uses a different technical approach and different throughputs, read lengths and error rates are associated with each of them. These parameters change regularly as each platform is being continually developed and improved.

Generally, all the platforms have the same advantages and disadvantages if compared to Sanger sequencing. The main advantage is that a high throughput can be achieved at a greatly reduced cost per nucleotide. However, read lengths are reduced and error rates increased in comparison to Sanger sequencing. The increased error rate can partly be overcome by the increased sequencing depth and bioinformatic analysis of the data. Although a large amount of sequencing data can be obtained in a short period of time, sequence analysis can be challenging and requires knowledge of bioinformatic processing methods and the availability of good computing infrastructure (Kircher and Kelso, 2010; Liu et al., 2012). Each of the platforms has its own strengths: the Illumina platform has the lowest cost per nucleotide sequenced, SOLid has the highest throughput and Roche 454 the longest read lengths. Each also has associated limitations: Illumina has the lowest accuracy in base pair calling, SOLid achieves very short read lengths and Roche 454 has homopolymer errors and single base pair insertions/deletions (see Table 1.1). Hence, the choice of next-generation sequence technology depends on the application for which it is intended (Kircher and Kelso, 2010; Liu et al., 2012).

Table 1.1 Summary of characteristics of three platforms for next generation sequencing

Characteristic	Roche 454 FLX Titanium	Illumina	SOLiDv4
Chemistry	Pyrosequencing	Sequencing-by-synthesis, reversible terminator technology	Sequencing by ligations
Read length	700 bp	MiSeq: 150 bp HiSeq: 100 bp	Up to 50 +35 bp
Error rate	10^{-3} - 10^{-4}	10^{-2} - 10^{-3}	10^{-2} - 10^{-3}
Accuracy	99.9%	98%	99.94%
Type of errors	Homopolymers, single nucleotide insertions/ deletions	Substitution errors	Substitution errors
Cost/million bases	$10	$0.07	$0.13
Reads/run	Up to 1 million (24 hrs)	MiSeq: 3.4 million (27 hrs) HiSeq: 3 billion (11 days)	1.2 billion (12-14 days)
Total output/run	0.7 Gb	MiSeq: 1 Gb HiSeq: 600 Gb	100 Gb
Recently launched improvements	Roche FLX+: read lengths up to 1000bp	Paired-end sequencing, read lengths up to 250 bp	Solid 4hq: Accuracy 99.99%, read lengths up to 75bp
Challenges	Homopolymer errors, relatively low throughput	Short read assembly	Short read assembly, reference sequence required

(biosystems, 2010; Illumina, 2012; Kircher and Kelso, 2010; Liu et al., 2012; Roche)

1.6 Immunity against *T. parva*

The appearance of $CD8^+$ T cells specific for parasitised lymphocytes following immunisation or challenge with *T. parva* coincides with clearance of the infection (Morrison et al., 1981). Further evidence that $CD8^+$ T cells play an important role in protection against *T. parva* was provided by an adoptive transfer experiment performed using identical twin calves. Transfer of $CD8^+$ T cells, derived from the efferent lymph draining the site of challenge with *T. parva* in the immune twin, to the respective naïve twin provided protection against *T. parva* infection (McKeever et al., 1994).

Taracha et al. (1995b) investigated the specificity of $CD8^+$ T cells after immunisation with two immunologically distinct stocks, Muguga and Marikebuni. $CD8^+$ T cell responses in some animals immunised with Muguga recognised both parasites whereas in other animals, only the immunising Muguga parasite was recognised. All animals that developed cross-reactive $CD8^+$ T cell responses were protected against heterologous challenge, whereas the animals that had strain-specific $CD8^+$ T cell responses were completely or partially susceptible to heterologous challenge. It was hypothesised that this variation in strain specificity of the immune response is due to the $CD8^+$ T cell response in different animals being focused on different immunodominant antigens which exhibit allelic variation among parasite strains (MacHugh et al., 2009a; Taracha et al., 1995b).

Parasite-specific $CD4^+$ T cells can be detected in immune animals and some of these $CD4^+$ T cells have cytolytic activity against *T. parva*-infected cells. Evidence from *in vitro* studies suggests that the main role of $CD4^+$ T cells is to provide assistance in the activation of memory $CD8^+$ T cells (Baldwin et al., 1987; Taracha et al., 1997). Although T cell-mediated immunity appears to be the main mechanism for natural protective immunity against *T. parva*, sporozoite-neutralising antibodies can be induced after repeated challenge or vaccination using recombinant p67 sporozoite surface antigen. However, high antibody titres do not necessary correlate with protection upon challenge (Gentschev et al., 1998; Kaba et al., 2004; Musoke et al., 2005).

1.7 Development of 'next-generation' vaccines against *T. parva*

The fact that cattle may develop protective immunity after natural infection or immunisation using the ITM suggests it may, in the future, be feasible to develop a subunit vaccine against *T. parva*. Such a vaccine would have a number of advantages over the current live vaccine, i.e. production would be easier, the cost per-dose would be lower, quality control would be more straightforward, it would be safer to administer and it may not require a liquid nitrogen cold chain for distribution in the field.

Several experiments have been conducted with recombinant p67 sporozoite surface antigen using different delivery constructs for immunisation (Gentschev et al., 1998; Kaba et al., 2004; Musoke et al., 1992; Musoke et al., 2005). Antibodies against p67 have been shown to inhibit parasite entry into the cell and neutralise the infection *in vitro* (Musoke et al., 1992). Although high titres of neutralising antibodies could be induced in all animals through vaccination, only a proportion of the animals were protected against challenge and it has been inferred that this variability is due to other immune mechanisms that were not investigated in the experiments (Bishop et al., 2003; Gentschev et al., 1998; Kaba et al., 2004; Musoke et al., 1992; Musoke et al., 2005).

Over the last decade, efforts have been directed toward identifying *T. parva* antigens that can be recognised by $CD8^+$ T cells in immune animals, as this cell type is known to be important in protective immunity against the parasite. Two different systematic approaches have been used to identify high value candidates. Firstly, *Theileria parva*-specific $CD8^+$ T cell lines were used to screen a schizont cDNA library and secondly, cDNA candidate genes were selected and tested on the basis of a bioinformatic screen of the *T. parva* genome (Graham et al., 2007a; Graham et al., 2006). A number of genes were identified whose products can be recognised by $CD8^+$ T cell responses and, for some of the genes, epitopes within the proteins were identified for cattle of certain major histocompatibility complex class I (MHC I) backgrounds (See Tables 1.2 and 1.3).

The immune responses of cattle immunised with a cocktail of five $CD8^+$ T cell antigens by a heterologous prime-boost regime with recombinant poxvirus vectors was examined (Graham et al., 2006). The animals used for the study had been identified as expressing a MHC I allele capable of presenting defined antigens to $CD8^+$ T cells. Antigen-specific IFN-γ responses were confirmed for the specific MHC I haplotypes, but only partial protection in a few of the animals was seen after challenge with a lethal dose of *T. parva* Muguga stock sporozoite stabilate. Only two of the six animals expressing the 2*01201 class I allele (the restriction element for Tp2) had $CD8^+$ T cell IFN-γ responses against the Tp2 antigen and it was suggested that this could be due to a lack of potency of the vaccination protocols rather than an inability of the animals to respond to the antigens (Graham et al., 2008b). Despite the disappointing results of the vaccine trials, the identified antigens provide an opportunity to investigate specificity of $CD8^+$ T cell responses in more detail.

Table 1.2 *Theileria parva* genes which can be recognised by bovine $CD8^+$ T cells

Gene	Genbank Accession number	Chromosome	Gene annotation	Reference
Tp1	XP_762973	3	Hypothetical	(Graham et al., 2006)
Tp2	XP_765583	1	Hypothetical	(Graham et al., 2007a)
Tp3	XP_766389	1	Hypothetical	
Tp4	XP_763228	3	Ɛ-TCP-1	(Graham et al., 2006)
Tp5	XP_765334	2	eIF-1A	(Graham et al., 2006)
Tp6	XP_765715	1	Prohibitin	
Tp7	XP_764810	2	Hsp90	(Graham et al., 2006)
Tp8	XP_764709	2	Cysteine protease	(Graham et al., 2006)
Tp9	XP_765463	2	Hypothetical	(Aguado-Martinez et al., 2012)
Tp10	XP_764408	4	Coronin	

Table 1.3 Epitopes recognised by bovine CD8$^+$ T cell and their presenting MHC I

Gene	Epitope	Bovine haplotype	Presenting bovine MHC I[3]	Reference
Tp1 $_{214-224}$	VGYPKVKEEML	A18	6*01301	(Graham et al., 2008b)
Tp2 $_{27-37}$	SHEELKKLGML	Animal BW002[2]	6*04001	(Graham et al., 2008b)
Tp2 $_{40-48}$	DGFDRDALF			(Pelle et al., 2011)
Tp2 $_{49-59}$	KSSHGMGKVGK	A10	2*01201	(Graham et al., 2008b)
Tp2 $_{50-59}$	SSHGMGKVGK	A10	2*01201	(Graham et al., 2008b)
Tp2 $_{96-104}$	FAQSLVCVL	Animal BX017[2]	T2c	(Graham et al., 2008b)
Tp2 $_{98-106}$	QSLVCVLMK	A10	2*01201	(Graham et al., 2008b)
Tp2 $_{138-147}$	KTSIPNPCKW			(Pelle et al., 2011)
Tp4 $_{328-338}$	TGASIQTTL	A10-KN104	3*00101	(Graham et al., 2008b)
Tp5 $_{87-95}$	SKADVIAKY	Animal BV050[2]	BoLa T5	(Graham et al., 2008b)
Tp7 $_{206-214}$	EFISFPISL	Animal BW012[2]	BoLa T7	(Graham et al., 2008b)
Tp8 $_{378-388}$	CGAELNHFL	A10-KN104	3*00101	(Graham et al., 2008b)
Tp9 $_{64-73}$	AKFPGMKKS	A14	1*02301	
Tp10 $_{304-318}$[1]	TNNFNNPELIPVL	A10	3*00201	(MacHugh Unpublished)

[1] The minimal epitope length is not yet identified, estimated amino acid composition
[2] Haplotype not defined
[3] Naming of the MHC I molecules was adapted to comply with the new nomenclature as described in the ImmunoPolymorphism Database (IPD) (Codner et al., 2012; Hammond et al., 2012)

1.8 Immunodominance

MacHugh *et al.* (2009) provided direct evidence for immunodominance in $CD8^+$ T cell responses against *T. parva*. In MHC I homozygous animals immunised with *T. parva*, a large component of the $CD8^+$ T cell response was found to be focused on a single dominant antigen. In animals homozygous for the A18 and A10 MHC I haplotypes, over 60% of the responding T cell clones were found to recognise defined epitopes in the Tp1 and Tp2 antigens respectively (MacHugh et al., 2009a). Immunodominance is a common feature in $CD8^+$ T cell responses in viral infections; $CD8^+$ T cell responses are preferentially directed towards one or a few dominant epitopes despite the presence of a T cell repertoire (TCR) with multiple specificities that potentially is capable of responding to a wide range of peptides (Yewdell, 2006; Yewdell and Bennink, 1999).

Immunodominance is the result of a complex combination of factors involved in antigen processing and antigen recognition by $CD8^+$ T cells. The relative importance of the various factors is likely to differ for different combinations of pathogen and host (Yewdell, 2006; Yewdell and Bennink, 1999; Yewdell et al., 2003). Firstly, factors that influence the abundance of the peptide epitope associated with MHC I on priming antigen-presenting cells play a role. This is dependent on abundance of protein prior to antigen processing and the efficiency of peptide processing by the MHC I antigen-presenting pathway (Luciani et al., 2013; York et al., 2006). The stability of peptide MHC I complex (pMHC) is further dependent on the affinity of peptide for the MHC I molecules (Yewdell, 2006). Secondly, immunodominance is influenced by factors that influence $CD8^+$ T cell precursor frequency and the affinity of the T cell receptors in the T cell repertoire for pMHC. Specific memory CD8 T cell will be present at a higher frequency in the memory CD8+ T cell pool after initial exposure. Prior experiences can therefore influence the immunodominance hierarchy (Cole et al., 1997). Thirdly, so called 'cross-competition' can influence immunodominance. In cross-competition, $CD8^+$ T cells of different specificities compete for interaction with antigen presenting cells (APC) upon challenge with a mixture of multiple antigens (such as whole pathogens) (Galea et al., 2012; Kastenmuller et al., 2007; Luciani et al., 2013).

1.8.1 Immunodominance and sequence polymorphism in *Theileria parva*

The effect of antigen polymorphism on MHC I presentation and T cell recognition has been investigated for two parasite-encoded genes, Tp1 and Tp2. Firstly, an epitope was identified ($Tp1_{214-224}$) in Tp1 of the Muguga genome sequence, which can be presented by MHC I allele 6*01301 (Graham et al., 2006). Anchor positions involved in peptide binding to MHC I were defined to be P2 (i.e. the residue at position 2), P5 and the C-terminal residue. However, alanine substitution at P2 for $Tp1_{214-224}$ did not have an effect on MHC I binding, while substitutions at P5 and the C-terminal residue reduced MHC I binding, the latter having the greatest effect(Macdonald et al., 2010).

X-ray crystallography of the pMHC complex (6*01301-$Tp1_{214-224}$) demonstrated that the peptide is presented in a raised conformation. The residue at position nine is predicted to be at the point where the main chain of the peptide dips back in the binding groove, resulting in several interactions which are critical for the shape and stability of the pMHC. Residue changes between P6 and P9 abrogated T cell recognition, with alanine substitutions at P9 being most critical. On the basis of the crystallographic structure, it was predicted that substitution at P9 would disrupt the shape and stability of the complex and thus abrogate TCR recognition (Macdonald et al., 2010). However, no amino acid substitutions are found for the P6 and P9 residues in natural allelic variants of Tp1 (Pelle et al., 2011). Among allelic sequences of the peptide discovered to date, only relatively conserved substitutions of P10 and P11 have been found (Pelle et al., 2011) and peptides but not cells infected with parasites carrying the allelic variants can be recognised by $CD8^+$ T cells *in vitro* (Steinaa et al., 2012); (MacHugh et al., 2009a)

Connelley *et al.* investigated the role of polymorphism in one of the epitopes in Tp2 ($Tp2_{49-59}$) (Connelley et al., 2011; Pelle et al., 2011), which is presented by MHC I allele 2*01201 (Graham et al., 2008b). The binding of variant epitopes was investigated using a series of peptides with alanine substitutions at different positions relative to the reference sequence. This was undertaken using a competitive binding

assay with the $Tp2_{98\text{-}106}$ epitope, which is also presented by the 2*01201 MHC I allele. Serial dilutions of variant peptides were tested for their ability to competitively block the binding of the $Tp2_{98\text{-}106}$ peptide. The MHC I binding capacity of the $Tp2_{49\text{-}59}$ variants was determined by inhibition of cytotoxicity against the $Tp2_{98\text{-}106}$ epitope. The residue in position eleven was found to be critical for MHC I binding and is thus likely to be an anchor residue. All but one of the ten allelic variants from the field with substitutions at P11 were still able to bind to MHC I, but with varying degrees of efficiency (Connelley et al., 2011). Each of the $CD8^+$ T cell clones tested failed to recognise peptide presented with an alanine substitution at P8. In addition, none of the natural $Tp2_{49\text{-}59}$ variants with amino acid substitutions (relative to the reference sequence) at this position were recognised. A proportion of the variants with an alanine substitution at P8 retained MHC I binding, indicating that the loss of killing activity must be the result of reduced T cell recognition rather than MHC I binding (Connelley et al., 2011).

1.9 Aims and objectives

It is clear that allelic polymorphism in *Theileria parva* antigens recognised by $CD8^+$ T cell responses has a profound influence on T cell recognition *in vitro*. A high degree of allelic polymorphism has been observed for some of the antigens that are known to be recognised by $CD8^+$ T cell responses (i.e. Tp1 and Tp2), particularly in buffalo-derived *T. parva*. Cattle immunised with the Muguga Cocktail are not solidly protected against challenge with buffalo-derived *T. parva*. It has been hypothesised that immunodominance of $CD8^+$ T cell responses against polymorphic antigens is responsible for the strain-specificity associated with bovine immunity following ITM immunisation using stabilates with limited antigenic diversity. Challenging genotypes may have different epitope sequences that may prevent presentation by MHC I and/or recognition by $CD8^+$ T cells thus allowing parasites to escape the immune response and cause severe clinical disease.

The overall objectives of this project are to investigate antigenic diversity within the Muguga Cocktail vaccine and to determine how this relates to antigenic diversity in field parasite populations, with particular reference to buffalo-derived *T. parva*. The specific aims of the project were to:

1. Determine the degree of genetic and antigenic diversity of *T. parva* parasites within the Muguga cocktail vaccine

2. Determine the degree of genetic and antigenic diversity between and within buffalo-derived *Theileria parva* in samples from African buffalo from Kenya and South Africa

3. Examine the biological relevance of antigenic diversity by conducting immunisation and challenge experiments using parasites of known antigenic composition

4. Determine whether parasites of the *Theileria* genus other than *T. parva* are present in African buffalo and sentinel animals

Chapter 2: General Materials and Methods

2.1 Cellular techniques

2.1.1 Isolation of peripheral blood mononuclear cells

Peripheral blood mononuclear cells (PBMC) were obtained as described by Goddeeris *et al.* (Goddeeris and Morrison, 1988). Blood was collected from the jugular vein in an equal volume of Alsever's solution. PBMC were separated by density gradient sedimentation over Ficoll-Paque Plus (GE Healthcare, Uppsala, Sweden) in 50 ml Falcon tubes and centrifugation at 900 g for 30 minutes. Cells were harvested from the Ficoll-Paque interface and mixed with an equal volume of Alsever's solution before centrifuging at 450 g for ten minutes. The pellet was washed three times in Alsever's solution by re-suspending and centrifuged for five minutes at 180 g. The final pellet was then re-suspended in standard culture medium (referred to as SCM, see Appendix A) and counted using a haemocytometer.

2.1.2 Magnetic-activated cell sorting (MACS) for cell subset isolation

$CD8^+$ cells and blood monocytes were isolated from PBMC by positive selection using Magnetic-activated cell sorting (MACS) (Miltenyi Biotec). PBMC were obtained as described previously and washed once in phosphate buffered saline (PBS) with 2% foetal calf serum (FCS) (PBS-FCS). $CD14^+$ cells were purified from PBMC using magnetic beads directly conjugated to goat-anti-human $CD14^+$ IgG (Miltenyi Biotec) according to the manufacturer's recommendations. PBMCs were mixed with 10 µl CD14 microbeads/10^7 cells and incubated on ice for 30 minutes. $CD8^+$ T cells were purified from PBMC using a mixture of CD8-specific monoclonal antibodies (IL-A51 and IL-A105). PBMC were pelleted and re-suspended before adding 12.5µl IL-A51/ILA105 (1:500) per 10^6 cells. The cells were incubated on ice for 30 minutes and then washed twice (six to eight minutes at 300 g) in PBS-FCS to remove unbound primary antibody. The pellet was re-suspended in PBS-FCS and 10 µl of anti-mouse IgG microbeads (Miltenyi Biotec) were added per 10^7 cells. The cells were incubated on ice for 30 minutes before transfer to a mini MACS column.

The mini MACS column was mounted on the magnet and flushed with 0.5 ml MACS buffer (99.5 ml PBS-FCS and 0.5 ml 2 mM Edetic Acid (EDTA)). The cells were washed twice in PBS-FCS and re-suspended in 0.5-1 ml MACS buffer. The re-suspended cells were carefully transferred to the mini MACS column and fluid was allowed to pass through the column. The column was washed three times in 0.5 ml MACS buffer. Then 2 ml of SCM was added to the column before removal from the magnet. The cells were eluted in SCM by removing the column from the magnet and inserting the plunger. Flow through was collected in a tube.

2.1.3 Establishment of *T. parva*-infected cell lines

Bovine PBMC were infected with *T. parva* parasites using cryopreserved aliquots of sporozoite stabilate. Straws of stabilates were removed from liquid nitrogen, thawed rapidly and the contents transferred to a 1.5 ml Eppendorf tube, before adding 1 ml of SCM. Particulate tick material was removed by centrifuging at 100 g for five minutes and the supernatant containing the sporozoites collected. PBMC were re-suspended in SCM at 2 x 10^7/ml and 1 ml of stabilate supernatant added and the mixture incubated at 37 °C for one hour. The cells were washed once in SCM, re-suspended at 5 x 10^6/ml and added to the wells of a 24-well plate. Cultures were incubated at 37 °C in 5% CO_2 (Goddeeris and Morrison, 1988).

2.1.4Generation of cloned *Theileria*-infected cell lines

Established *Theileria*-parasitised cell lines were cloned by limiting dilution in conditioned medium (containing 50% SCM and 50% supernatant from established buffalo *T. parva*-infected cell lines) and irradiated bovine PBMC as filler cells. *Theileria*-infected cells were diluted in conditioned medium to a concentration of 15 cells/ml, 3 cells/ml, 1.5 cells/ml and 0.5 cells/ml. 200 µl of cells were distributed into wells of 96-well plates to give a concentration of 3 cells/well, 1 cell/well, 0.3 cells/well and 0.1 cells/well. Irradiated filler cells were added to each well at a concentration of 2 x 10^5 filler cells/well. The plates were then screened microscopically and the frequency of wells giving cell growth was determined. Only cells from plates showing growth in less than 30% of wells were selected as they have a greater than 83% likelihood of being clonal (Goddeeris and Morrison, 1988).

Cells from the selected wells were expanded in several wells of a 96-well plate before transfer to 24-well plates. The cells from four wells in a 24-well plate were then transferred to T25 flasks.

2.1.5 Generation of $CD8^+$ T cell lines

$CD8^+$ T cell lines were generated in a similar manner to that previously described (Goddeeris and Morrison, 1988). Briefly, PBMC were stimulated with autologous irradiated parasitised cells in wells of 48-well plates, each well containing 2×10^6 PBMC (responder cells) and 2×10^5 stimulator cells. All stimulator cells were irradiated by exposure to 60Gy of gamma irradiation (ILRI, Nairobi). Cells were incubated at 37 °C in 5% CO_2 for seven days. Viable cells were harvested, washed and stimulated a second time as described above, but this time using 1×10^6 responder cells/well and 4×10^5 stimulator cells/well. Cells from the second stimulation were harvested after seven days and enriched by positive selection of $CD8^+$ cells by MACS sorting as described previously. Cells were stimulated for a third time by distributing 10^6 responder cells/well with 4×10^5 stimulator cells/well in SCM with 100 units recombinant human (rh) IL-2/ml into wells of 48-well culture plates. Irradiated autologous *T. parva*-infected cell lines were used as stimulators (*T. parva* Muguga stabilate 3308). Plates were incubated at 37 °C in 5% CO_2 for a further seven days. The same protocol was used for further re-stimulations of the cultures.

2.1.6 Phenotyping of cells by flow cytometry

Purity of the $CD8^+$ T cell lines was checked by flow cytometry before assaying for cytotoxicity. Aliquots of 50 µl of cells were distributed in wells of 96-well round-bottom well plates and pelleted by centrifugation at 900 g for 30 seconds. The medium was discarded, the cells re-suspended and 25 µl of the required primary antibody added. Negative controls were incubated with FACS medium instead of primary antibody. The plate was incubated at 4 °C for 30 minutes before washing in FACS medium (see Appendix A) for a total of three washes. The cells were then re-suspended in 25 µl secondary antibody, 1:100 diluted fluorescein isothiocyanate

(FITC)-labelled goat polyvalent anti-mouse-immunoglobulin antibody (Sigma-Aldrich, UK). The cells were incubated at 4 °C for 30 minutes, washed three times as previously described, re-suspended in 200 µl FACS medium and analysed using a BD FACScalibur flow cytometer (Becton Dickinson, USA). The antibodies used for typing are listed in Table 2.1.

Table 2.1 Monoclonal antibodies used for FACS analysis

Monoclonal antibody	Specificity	Cell populations identified
MMIA	CD3	All T cells
IL-A12	CD4	CD4 T cells
IL-A105/IL-A51	CD8	CD8 T cells, some NK cells
CC15 or αT4	WC1	γδ T cells
IL-A24	SIRP-1A	Monocytes
IL-A30	IgM	B cells

2.1.7 Cytotoxicity assays - Chromium release assay

$CD8^+$ T cell lines were tested for specificity using a four-hour Chromium release cytotoxicity assay. 2 x 10^6 target cells were labelled with 10 µl ^{51}Chromium and incubated for one hour at 37 °C in 5% CO_2. The cells were washed three times by adding 8 ml cytotoxicity medium (see Appendix A) and centrifugation at 180 g for five minutes. The pellets were re-suspended in cytotoxicity medium to give a concentration of 10^6 cells/ml.

For each target cell, 3 x 10^5 $CD8^+$ T cells were distributed into 96-well plates together with 5 x10^4 of labelled target cells to give an effector to target ratio of 6:1. Spontaneous release by each target cell was measured in triplicate wells containing 100 µl cytotoxicity medium instead of effector cells. Maximum release was measured in triplicate wells containing 100 µl of 0.2% Tween instead of effector cells.

Plates were centrifuged briefly before incubating at 37 °C in 5% CO_2 for four hours. The plates were centrifuged again and 75 µl supernatant of each well harvested using a multi-pipette and transferred to 96-well lumaplates (Perkin Elmer, USA) which were dried in an oven overnight. The next day the lumaplates were taken from the oven and sealed with Top seal-A (Perkin Elmer, USA). Radioactivity release was measured using a TopCount NXT machine (Perkin Elmer, USA).
The percentage cytotoxicity was calculated with the following formula:

$$\% \text{ cytotoxicity} = 100 \times \frac{(\text{test } ^{51}\text{Chromium release} - \text{spontaneous } ^{51}\text{Chromium release})}{(\text{maximum } ^{51}\text{Chromium release} - \text{spontaneous } ^{51}\text{Chromium release})}$$

2.1.8 Cytotoxicity assays - Indium release assay

$CD8^+$ T cell lines were tested for specificity using a four-hour ^{111}Indium release cytotoxicity assay. Target cells were labelled by incubation with 5 µCi of Indium for 30 minutes at 37°C in a 15 ml falcon tube. Then the cells were washed six times in 8 ml cytotoxicity medium (see Appendix A) and centrifuged at 180 g for five minutes. The pellets were re-suspended in cytotoxicity medium to give a concentration of 10^5 cells/ml. 50 µl of target cell were added to each of the wells to give a concentration of 5 x10^3 target cells/well. $CD8^+$ T cells were added at different concentrations to give effector to target ratios between 20:1 and 0.6:1. A fixed effector to target ratio was used when examining the recognition of peptide using an autologous *T. annulata* line as presenter. Peptides were obtained from JPT Peptide Technologies GmbH, Germany. The peptide concentration was titrated using ten-fold dilutions ranging from 100 ng/ml to 0.01 ng/ml.

Spontaneous and maximum releases were measured for each of the target cells. Spontaneous release was measured in triplicate wells containing 100 µl cytotoxicity medium instead of effector cells. Maximum release was measured in triplicate wells containing 100 µl of 0.2% Tween instead of effector cells. Plates were gently centrifuged (180 g) for one minute and incubated at 37 °C in 5% CO_2 for four hours.

After four hours the plates were centrifuged in order to pellet the cells and 75 µl supernatant of each well was transferred into individual gamma counter tubes and a small amount of bromophenol blue-stained 1.5% agarose added to capture radioactivity in solid form in case of spill. Radioactivity release was measured with a Wallac Wizard 1470 Automatic Gamma Counter and percentage cytotoxicity was calculated using the following formula:

$$\% \text{ cytotoxicity} = 100 \times \frac{(\text{test } ^{111}\text{In release} - \text{spontaneous } ^{111}\text{In release})}{(\text{maximum } ^{111}\text{In release} - \text{spontaneous } ^{111}\text{In release})}$$

2.1.9 Ex vivo enzyme-linked immunosorbent spot assay (ELISPOT)

An *ex vivo* ELISPOT was performed to check whether $CD8^+$ T cells from immunised animals were capable of recognising *T. parva*-infected cells and selected epitopes within the Tp2 protein ($Tp2_{49\text{-}59}$ and $Tp2_{98\text{-}106}$). ELISPOT plates (Millipore, Billerica, MA, USA) were pre-coated overnight with a monoclonal anti-bovine IFN-γ antibody CC302 (Serotec, Oxford, UK). Plates were washed twice with 200 µl of un-supplemented minimum essential media (MEM) before blocking with 200 µl of Roswell Park Memorial Institute medium (RPMI) supplemented with 10% heat-inactivated FCS. Plates were incubated at 37 °C for two hours. The RPMI supplemented with 10% heat-inactivated FCS was flicked off and replaced with cells and incubated in a humidified incubator at 37% with 5% CO_2.

$CD8^+$ T cell and $CD14^+$ cells were isolated from blood as previously described. *Theileria parva* Muguga cell lines were irradiated and left overnight at room temperature and washed twice before use. For presentation of peptide, $CD14^+$ cells were used at 2.5×10^4 cells/well and peptides were added at a concentration of 1 µM. Irradiated and washed *T. parva*-infected cells were used at 2.5×10^4/well. $CD8^+$ T cells were added at 2.5×10^4 cells/well. Controls wells were included that contained only $CD14^+$ cells, only $CD8^+$ T cells or only *Theileria parva* parasitised cells. After incubation, well contents were flicked off and plates were washed four times using 200 µl of distilled water with 0.05% Tween20 (Sigma–Aldrich, USA), followed by four washes using 200 µl PBS (Dulbeco) with 0.05% Tween 20 (PBS-T). Then

50 µl/well of rabbit anti-bovine IFN-γ anti-sera 1:500 (ILRI, Kenya) was added and incubated for one hour at room temperature. The cells were washed four times in 200 µl PBS-T, before adding 50 µl/well of alkaline phosphatase-conjugated mouse-anti-rabbit IgG (1:2,000) diluted in PBS + 0.05% Tween20 + 0.1% BSA (Sigma-Aldrich, USA). Plates were incubated for one hour at room temperature followed by six washes in PBS-T. Plates were developed by adding 50 µl/well of substrate solution to one tablet of Sigma Fast BCIP/NBT substrate (Sigma–Aldrich, USA) in 10 ml distilled water.

2.2 *In vivo* experiment

2.2.1 Ethical approval

Ethical permission for the experiment was obtained from the Institute Animal Care and Use Committee (IACUC) committee at ILRI, Nairobi, Kenya. The experiment was performed as stated in the standard operating procedures for animal experimentation, Sections 11 and 20. These SOPs were prepared initially by Internationl laboratory for research on animal diseases (ILRAD) IACUC on the 2nd of August 1994 and were last revised by ILRI/ IACUC on the 19th March 2004. The only exception in this study was that the end-point determination was left to the discretion of the institute veterinarian. The SOP was developed for studies using cattle-derived *T. parva* for challenge and decisions on treatment or euthanasia of animals were based on an ECF reaction index incorporating parasitological and clinical parameters. However, in previous studies using buffalo-derived *T. parva* parasites, disease progression occurred more rapidly. Therefore, decisions to treat or euthanize animals were taken when deemed necessary, even when this was not yet required on the basis of the ECF index.

2.2.2 Screening of animals for *in vivo* studies

Animals expressing the MHC I haplotype A10 were selected by workers at ILRI. Preliminary selection was performed by immunofluorescence staining of PBMC using a monoclonal antibody (IL-A10) specific for the MHC I haplotype A10. This was further validated first, by demonstrating recognition by a $CD8^+$ T cell line of a peptide known to bind to the 2*01201 A10 class I gene product, following incubation of peptide with PBMC from test animals, and second, by sequencing of the second A10 class I gene 3*00201.

2.2.3 Infection and treatment

Animals were immunised by Infection and Treatment using *T. parva* Muguga (stabilate 3308) and long acting oxytetracycline as previously described (Radley et al., 1975a). 1.0 ml of stabilate was inoculated subcutaneously below and in front of the parotid lymph node using a 1 ml syringe and a 21 G needle. The site was

swabbed with 70% alcohol prior to inoculation. At the same time, animals were injected with 20 mg/kg of long acting oxytetracycline by deep intra-muscular injection. Following immunisation, cattle were examined daily and rectal temperatures were taken and recorded. Lymph node smears were taken when the rectal temperature was greater than 39.5 °C. Cattle were treated with an anti-theilerial drug when developing a severe reaction. Blood samples were taken at weekly intervals to determine serological responses to immunisation.

2.2.4 Challenge

Ten weeks after immunisation the animals were challenged, along with two naïve controls, with buffalo derived stabilate 3081. Animals were challenged with 1 ml of 1:20 dilution of Stabilate 3081 by inoculating subcutaneously below and in front of the right parotid lymph node using a 1 ml syringe and a 21 G needle. Five straws were taken from liquid nitrogen, thawed rapidly and the contents thoroughly mixed before diluting 1:20 using vaccine diluent (see Appendix A) while kept on ice. Time of thawing was noted. Briefly, animals were inspected three times daily and monitored for temperature rise and lymph node enlargement daily. Lymph node and blood smears were taken daily when the body temperature was increased and/or when the lymph nodes were enlarged. Blood samples were taken at weekly intervals to determine serological responses to immunisation.

In addition, blood samples and lymph node aspirates were taken twice weekly for DNA extraction. This frequency was increased to daily when clinical signs were detected until the day of treatment or recovery. Haematology was performed every two days.

2.2.5 Examination of lymph node and blood smears

The site for lymph node biopsy was swabbed with 70% alcohol prior to taking the biopsy. The lymph node was immobilised between the thumb and forefinger of one hand and a 20 G needle on a 2 ml syringe was pushed into the centre of the lymph node with the other hand. The plunger of the syringe was carefully withdrawn until a small volume of lymph was observed in the hub of the needle. The needle was removed, the sample deposited on a microscope slide and a smear made using a drop of blood collected in EDTA.

Lymph node and blood smears were allowed to dry at room temperature then fixed in 100% methanol for three minutes and allowed to air dry. The slides were then stained using 10% Giemsa diluted in staining buffer (see Appendix A) for twenty minutes. The slides were then washed by overflowing the jar containing the stained slides with tap water and pouring off the Giemsa stain. They were then washed a further two times by filling the jar with tap water and pouring the water off before allowing them to dry. The slides were examined under oil immersion using a standard light microscope. The criteria used for scoring the levels of parasites (parasitosis) were as follows: Ma+, schizonts not easy to find; Ma++, one or two schizonts visible per field; Ma+++, more than two schizonts per field.

2.3 Molecular techniques

2.3.1 Processing samples prior to DNA extraction

Some pre-processing of samples was required before DNA extraction. Following thawing of sporozoite stabilate from liquid nitrogen, excess tick debris was removed by centrifuging at 100 g for five minutes. The supernatant containing the sporozoites was transferred to a clean Eppendorf tube. The sporozoites were pelleted at 16,000 g for ten minutes if a smaller volume was required for DNA extraction. *Theileria*-infected cells were harvested from a T25 flask, centrifuged at 900 g for five minutes and DNA extracted from the re-suspended pellet. Lymph node aspirates were put in PBS and the cells were pelleted by centrifuging at 16,000 g for one minute prior to extraction of DNA from the re-suspended pellet. DNA was extracted from 100 µl of whole blood collected in EDTA.

2.3.2 DNA extraction

DNA was extracted from various samples using the Qiagen DNeasy Blood and Tissue Kit, according to the manufacturer's instructions. Briefly, 20 µl of proteinase K was added to the sample and the tubes were mixed in a vortex mixer. Then 200 µl lysis buffer (AL) was added and tubes incubated at 56 °C for ten minutes while shaking. Then 200 µl ethanol was added and tubes mixed thoroughly. The mixture was then transferred to a DNeasy mini spin column and centrifuged for one minute at 6,000 g. The fluid in the collection tube was discarded and 500 µl of washing buffer AW1 was added to the column. The DNeasy mini spin column was centrifuged for two minutes at 6,000 g and the flow-through discarded. 500 µl of washing buffer AW2 was added and the column centrifuged for three minutes at 20,000 g. The DNeasy mini spin column was transferred to a clean Eppendorf tube and between 100 and 200 µl elution buffer (AE) was added to the membrane of the DNeasy mini spin column. The Eppendorf tube and DNeasy mini spin column were centrifuged for two minutes at 6,000 g and DNA within the eluate was collected in the Eppendorf. DNA was stored at -20 °C until further use.

2.3.3 PCR

A number of genes were amplified by PCR in this study. Each PCR reaction was comprised 20 pmol forward primer, 20 pmol reverse primer, 0.5 units BIOTAQ polymerase (Bioline, UK), 2.0 µl 10 x PCR master mix, 50 ng genomic DNA and nuclease-free water, to give a final volume of 20 µl. A G-storm thermal cycler (Genetic Research instrumentation, UK) was used for amplification. The PCR cycle programme was as follows: 95°C for three minutes, 30 cycles of (95 °C for one minute, X °C for one minute, 72 °C for two minutes) and a final extension period of 72 °C for seven minutes. The annealing temperature was adjusted according to the primer pair used and was equal to the lowest melting temperature of the primer pair. PCR amplification of samples for Roche 454 amplicon sequencing, which used fusion primers and DNA polymerase with proofreading capacity, will be discussed in Chapter 3.

2.3.4 Purification of PCR products

PCR products were purified using the Promega Wizard Gel and PCR Clean-up System (Promega Corporation, USA). PCR products were visualised by UV on a 1.5% agarose gel (Bioline, UK) containing SafeView nucleic acid stain (NBS, Biologicals Ltd, UK) in Tris-acetate-EDTA buffer (TAE) and bands of the product of interest were excised. Membrane-binding solution was added at 10 µl per 10 µg of excised product. The agarose was then allowed to melt in a water bath at 60 °C for ten minutes. For direct purification, PCR products were mixed with an equal amount of membrane-binding solution. The mixture was added to a SV minicolumn fitted on a collection tube and incubated for one minute. The SV minicolumn and collection tube were centrifuged at 16,000 g for one minute. The flow-through was discarded and 700 µl of membrane washing solution added to the column and another spin performed at 16,000 g for one minute. The flow-through was discarded and 500 µl of membrane washing solution added to the column. The SV minicolumn and collection tube were centrifuged at 16,000 g for five minutes. After discarding the flow-through the empty SV minicolumn was centrifuged at 16,000 g for one minute. The SV minicolumn was transferred to a new Eppendorf tube and between 30 and 50 µl of

nuclease-free water was added. After another incubation of one minute, the SV minicolumn was centrifuged at 16,000 g for one minute. The SV minicolumn was discarded and the eluate containing purified DNA stored at -20 °C until further use.

2.3.5 Cloning of PCR products

PCR products were cloned into the pGEM-T Easy vector system (Promega Corporation, USA) according to the manufacturer's instructions. The ligation reaction was composed of 5µl 2 x Rapid ligation buffer, 1µl pGEM-T Easy vector, 1µl T4 DNA ligase (3 Weiss units/µl), ~25 ng PCR product and nuclease-free water to give a final volume of 10 µl. The reaction mix was incubated for one hour at room temperature or overnight at 4 °C. The ligated plasmid/PCR products were transformed into JM109 high efficiency competent cells (Promega Corporation, USA). Aliquots of 50 µl JM109 cells were added to 2µl of ligation reaction and incubated for twenty minutes on ice. The cells were heat-shocked for 45 to 50 seconds at exactly 42 °C and immediately returned to ice. 950 µl of SOC medium at room temperature was added and the cells were incubated for 1.5 hours in a shaking incubator at 37 °C and 150 rpm. 100 µl of the transformant was plated on duplicate LB/ampicillin/IPTG/X-Gal plates and incubated overnight at 37 °C. Transformed colonies were identified on the basis of blue/white screening. Plasmid clones were expanded in tubes containing 6 ml of Luria Broth (LB) media containing ampicillin at a concentration of 100 µg/ml. Small quantities of individual clones were added to the LB/ampicillin using a 10 µl pipette tip. The tubes were incubated overnight in a shaking incubator at 37 °C and 150 rpm.

2.3.6 Purification of plasmid clones

Expanded colonies were purified using the PureYield^{TM} Plasmid Miniprep System (Promega Corporation, USA) according to the manufacturer's instructions. Briefly, 600 µl bacterial culture grown in LB medium was transferred to a 1.5 ml Eppendorf tube and 100 µl of cell lysis buffer added. This was then mixed by inverting the tube six times. Then 350µl of cold neutralisation buffer was added and the contents of the tube mixed by inverting until yellow precipitate formed followed by another three

inversions. The tube was centrifuged at maximum speed for three minutes and the supernatant transferred to a PureYield™ minicolumn placed in a collection tube. The PureYield™ Minicolumn/collection tube assembly was centrifuged for 15 seconds at 16,000 g and the flow-through discarded. 200 µl of endotoxin removal wash was then added to the PureYield™ minicolumn and centrifuged for 15 seconds at 16,000 g, following which 400 µl of column wash solution was added to the minicolumn. After a further spin of 30 seconds at 16,000 g, the PureYield™ minicolumn was transferred to a clean 1.5 ml Eppendorf and 30 µl elution buffer was added directly to the PureYield™ minicolumn matrix. The PureYield™ minicolumn/Eppendorf assembly was then centrifuged for 15 seconds at 16,000 g after one minute of incubation. Eluted plasmid DNA was stored at -20 °C until further use.

2.3.7 Bioinformatic sequence analysis

Purified PCR products or plasmid were sequenced using Sanger technology at DBS genomics, University of Durham. Sequencing was carried out using custom primers at a concentration of 3.2 pmol/µl. Resulting sequence traces were inspected and DNA sequences manually corrected and/or trimmed if required. The sequences were then analysed using the MEGA5 software package (Tamura et al., 2011). Sequence alignments were performed using the default setting of MUSCLE. Phylogenetic analysis was performed using MEGA5 (Tamura et al., 2011). Phyogenetic trees were inferred using the maximum likelihood method, parsimony method and/or neighbour joning method. The optimal nucleotide substitution model was identified using Datamonkey (Delport et al., 2010). Phylogenetic trees of the 18S rRNA gene were rooted using the 18S rRNA gene sequence of *Prororectum micans* (M14649), *Sarcocystic muris* (M64244) and *Toxoplasma gondii* (X68523) as previously done by Chaisi *et al.* (Chaisi et al., 2013) and other more closely related species when available. Phylogenetic trees for 5S, Tp6, Tp7 and Tp8 were rooted using sequences obtained from the NCBI database, identified using a BLAST search. Heterozygosity was calculated using the formula: $H_e = 1 - \Sigma Pi^2$ where P is the proportion of allele i. The mean diversity in a parasite population was calculated using a maximum composite likelihood model in MEGA5; this represents the number of base

substitutions per site from averaging over all sequence pairs in the population (Tamura et al., 2004; Tamura et al., 2011). Distance matrices were constructed using MEGA5 (Tamura et al., 2011) by calculating pair-wise distances using the maximum composite likelihood model (Tamura et al., 2004). The rate variation among sites was modelled with a gamma distribution (shape parameter = 2). These distance matrices were exported for principal co-ordinate analyses using the Genalex plug-in for Excel (Peakall and Smouse, 2012). The K-means for the different sub-populations was calculated using the statistical package, R.

2.3.8 Processing of PCR products for 454 sequencing

For the 454 sequencing runs, equimolar quantities of purified PCR product, generated as described in Chapter 3, were pooled. Purity and product size were determined by analysing samples on a Bioagilent DNA1000 chip (Agilent technologies, UK) using the Bioanalyser 2100 (Agilent Technologies) and 2100 expert software (Agilent Technologies) according to the manufacturer's protocol. Preliminary quantification of PCR products was performed by Nanodrop. The PCR products were then diluted in triplicate to a concentration of between 20 and 1000 pg/µl and the DNA concentration determined using a picogreen assay (Invitrogen). The concentration of the PCR products was calculated using the standard curve of the standards in the picogreen assay. The molarity of the PCR products was calculated using the concentration obtained by the picogreen assay and the size obtained from the Bioagilent 100 chip, using the following formula:

$$\text{Molecules/µl} = \frac{(\text{Sample concn in ng/µl}) \times (6.022 \times 10^{23})}{(656.6 \times 10^{9}) \times (\text{amplicon length in bp})}$$

Analysis of the sequences obtained by 454 sequencing will be described in Chapter 3

2.3.9 Satellite DNA typing

PCR amplification was performed using satellite markers and methods as previously described by Oura *et al.* (Oura et al., 2003). The distribution of the satellite markers across the genome is illustrated in Figure 2.1 and their primer sequences are detailed in table 2.2. One primer of each primer pair was labelled at the 5' end with the fluorophore FAM. Each PCR reaction comprised 10 pmol forward primer, 10 pmol reverse primer, 0.5 units BIOTAQ DNA polymerase (Bioline, UK), 1 µl of 10 x custom PCR buffer (Thermo Fischer, UK), 1 µl of stabilate DNA and nuclease-free water to give a final volume of 10 µl. The cycling conditions were: 35 cycles of denaturation at 94 °C for one minute, annealing at 60 °C for one minute and extension at 65 °C for one minute.

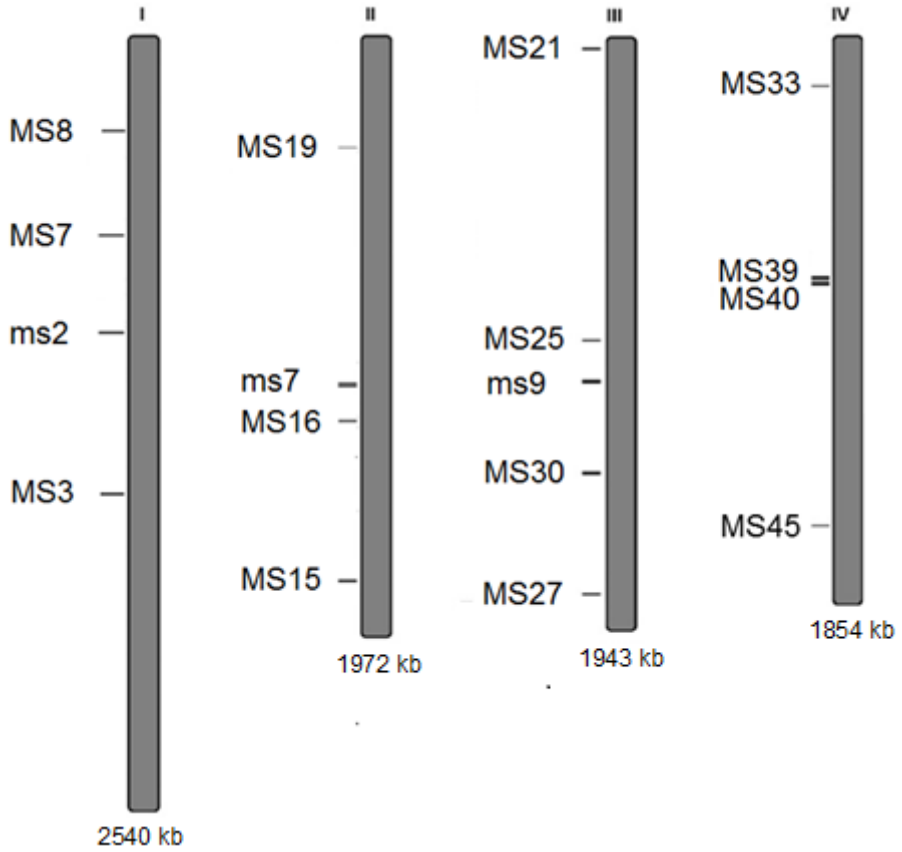

Figure 2.1 Graphical illustration of the satellite loci used in this study
Adapted from Katzer *et al.* 2010. Chromosomes 1-4 are denoted by I-IV. Sizes of individual chromosomes are identified below. The location of mini satellite (MS) and micro satellite (ms) DNA marker is indicated.

The PCR products were analysed using capillary flow electrophoresis with the fluorophore signal detected using a laser. In the first instance, the presence of PCR products was confirmed by running 6 µl of product on a 1.5% agarose gel containing ethidium bromide or SafeView nucleic acid stain (NBS, Biologicals Ltd, UK). PCR products were then diluted in sterile water (between 1:5 and 1:200 depending on the intensity of the band observed on the gel) and 20 µl of each diluted sample was transferred to a 96-well non-skirted PCR plate. The plates were sealed with a sterile silicon mat and sent to the University of Dundee Sequencing Service for genotyping using Genescan^{TM} technology (Applied BioSystems) with GSLIZ500 size standards. The peak size, in base pairs (bp), was calculated using Peak Scanner software (Applied Biosystems) using the default sizing method with the presence of a primer peak and then the data was exported to an Excel spreadsheet. Only peaks in the range of 100 to 500 bp were considered for further analysis. In order to measure the relative abundance of each of the alleles, the area under the curve was calculated for each of peaks separately and for all peaks of a single trace combined. Peaks with an area less than 3% of the total area under the curve were excluded from further analysis. The relative abundance of individual peaks was expressed as a percentage of the total area under the curve for all peaks in a trace. The high-precision peak sizes were inspected and then manually adjusted to integer values to facilitate allele 'calling'.

Table 2.2 Satellite DNA markers used in this study

Satellite marker	Consensus repeat sequence	PCR primers sequence (5'-3')	Chromosome
ms2	tat	FAM-tgaggcagtgtagagcgcataac aaatccgcaacgctattgccgagg	1
MS3	tataccaaat	FAM-ccaccgtaaccctctataccat gacatctccctcaaatcagactc	1
MS7	gtaactataactatgtaaaca	FAM-gttcagtcctatggcaattcag caaacctcttcaaattcactctagg	1
MS8	ttacacagta	FAM-gcctcctcaagcaattcagta ctggttcaaacacatcaaggtaca	1
MS19	ataattaa	FAM-ccagacacctcaaatcccaagta ccacactgccacctaatacaaa	2
ms7	att	FAM-ttaacttatctccctctccctccc acactctcaacaactcactcttcc	2
MS16	actaatatttgttattt	FAM-cctcctccatactactaacctacc cagcgctcagattcacttgtact	2
MS15	aatttaacat	FAM-caagagatcaatggctcctactc ctcttcccctcttttccctctac	2
MS21	atactatt	FAM-ttctacccaacgccactctatgcg tgactcccgtcttatccaaattcg	3
MS25	ttatatagttaagt	FAM-acacacccatcaacgtagtaac caccatcacactcttaaccat	3
ms9	tatac	FAM-ctggttcctcatcttcacacta ctttccagaacctacaatcac	3
MS30	atttggtgagta	FAM-caagaaagtctaaactcgactg tatctctgggcactcagaac	3
MS27	taatcaaattat	FAM-cctgcgatacatttctaatcc gtaataccattcccacctctac	3
MS33	atatagttaatt	FAM-cttctcaaggtaccgtaaacc cctcactactccaatagttcttc	4
MS39	tttacaca	FAM- ccaatcaacatcaactactcc cgaactccaaacgatctaaac	4
MS40	gaattaataaata	FAM-catatcacctcatggtacacac ccagccctaatacacaaatc	4
MS45	tacacatttt	FAM-aacgaggtgactgaggaatac ccgcctacaacaaattacatac	4

MS=Mini satellite DNA marker and ms=micro satellite DNA marker

Chapter 3: Establishing a deep sequencing method to analyse diversity in *Theileria parva* antigens

3.1 Introduction

Cattle may be immunised against *Theileria parva* using the infection and treatment method (ITM). The Muguga cocktail vaccine stabilate has been used for ITM and consists of three parasite stocks (Radley et al., 1975b): Muguga, Serengeti and Kiambu 5. The vaccine has been used under experimental conditions and in the field in East Africa and appears to give good protection against challenge from cattle-derived *Theileria parva* (Dolan, 1989; ILRAD, 1989; Martins et al., 2010; Musisi et al., 1999 ; Radley et al., 1975b; Radley et al., 1979; Uilenberg et al., 1976; Uilenberg et al., 1977). The components of the Muguga cocktail have previously been analysed using a variety of techniques, including parasite-specific monoclonal antibodies, RFLP-PCR of four polymorphic genes and microsatellite markers (Bishop et al., 2001; Oura et al., 2007a; Oura et al., 2004b; Patel et al., 2011). These analyses demonstrated limited diversity in the Muguga cocktail and a high degree of similarity between the Muguga and Serengeti-transformed components (Bishop et al., 2001; Oura et al., 2007a; Oura et al., 2004b; Patel et al., 2011). It is believed that the Muguga cocktail provides broad protection against parasite challenge in the field due to the diversity of parasites in the vaccine. It has been hypothesised that some parasite genotypes may be present at a low frequency and are beneath the detection threshold of currently used methods.

The different components of the Muguga cocktail should be viewed as individual parasite populations with a potential for genotypic diversity within each component part. This diversity may be due to a number of genotypes being present when the parasite stocks were established and this diversity may be compounded through sexual recombination in the tick as the stocks are passaged. Even though discrete parasite genotypes may not be preserved over different passages through the tick host, allelic diversity could be maintained across the genome within each component. However, the influence of genetic drift on a relatively small parasite population may

result in a reduction in allelic diversity over the course of multiple passages. Assuming that an element of allelic diversity is required for the induction of broad protective immunity, for example at key antigen-encoding loci, it is important that this diversity is maintained in every new vaccine batch and that methods are available for their detection.

It has been shown that CD8 T cell responses play an important role in the control of *T. parva* in immune animals after vaccination by infection and treatment (McKeever et al., 1999; McKeever et al., 1994; Morrison et al., 1981). A number of antigen-encoding loci have been identified in the parasite genome, alleles of which have been shown to be recognised by CD8 T cells. These antigens, referred to as Tp1-Tp10, are expressed in the schizont stage of *T. parva* (Graham et al., 2007a; Graham et al., 2006). DNA sequencing has revealed that while a number of these antigens are highly polymorphic in field populations (Pelle et al., 2011), others show little variation among different parasite genotypes. The loci encoding these genes may be utilised as markers that can be used to gain an insight into antigenic diversity in *T. parva*. With the development of high-throughput DNA sequencing technology, it is now possible to undertake high-resolution characterisation of the different components of the Muguga cocktail, targeting genes encoding known CTL antigens to provide a multi-locus sequence typing (MLST) system.

Multi-locus sequence typing was initially developed to analyse genetic diversity in prokaryotic and eukaryotic organisms by applying conventional Sanger sequencing to determine the sequences of multiple genes in a single organism. Applying MLST to a large number of individuals would provide a broad picture of genetic diversity within a population. The advantage of MLST over other genotyping methods is that DNA sequences can provide precise, discrete, reproducible genetic datasets. Importantly, this methodology allows comparison of results between different laboratories (Aanensen and Spratt, 2005; Boers et al., 2012), which may be challenging or impossible with other techniques.

Next generation sequencing strategies offer the opportunity to analyse antigenic diversity at far greater depth than traditional Sanger sequencing, due to the large number of clonal sequences that can be obtained. In the case of heterogeneous samples of *T. parva* (i.e. stabilates, or DNA extracted from an individual animal or tick) this would allow the detection of parasite genotypes present at a low frequency. The Roche 454 sequencing platform has the main advantage, compared to other high-throughput sequencing platforms that relatively long read lengths of 400-500 bp can be achieved (Liu et al., 2012).

The Roche 454 platform uses a pyrosequencing approach and one of the strategies that can be employed using this technology is amplicon sequencing. DNA template samples are PCR amplified using 'fusionprimers', which consists of three components: a sequence-specific primer used for PCR amplification, a multiplex identifier (MID), which is a ten nucleotide barcode used for the identification of samples, and a 454-adaptor which is required for processing of the PCR products for sequencing. The use of these PCR primers is illustrated in Figure 3.1. Following this initial amplification, the 454-adaptor region of single-stranded PCR products anneals to sequencing beads to allow clonal amplification of products using an emulsion-based PCR. Following amplification, individual beads are deposited in wells of a picotitre plate along with the sequencing reagents. The Roche 454 technology is based on a 'sequencing by synthesis' approach, which involves 'flows' of nucleotides being washed over the picotitre plate. When nucleotides are incorporated in the growing DNA molecule, a light burst is emitted and the intensity of this signal correlates with the number of nucleotides incorporated (Margulies et al., 2005). The intensities of the light signals are registered and subsequently translated into nucleotide sequences.

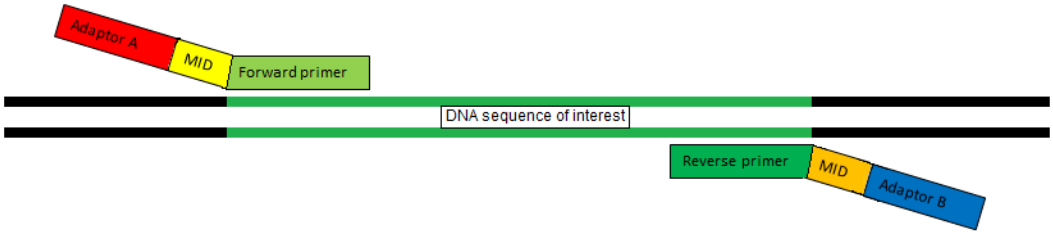

Figure 3.1 Graphical illustration of 454 amplicon sequencing
The fusionprimers used for PCR amplification consist of three components: The sequence-specific primers, Multiplex IDentifier (MID) and sequencing adaptor.

The work presented in this chapter describes the development of a MLST system based on 454 sequencing technology with the capacity to analyse antigenic diversity in *Theileria parva.* This includes the design and validation of PCR primers, optimisation of sample preparation protocols and the development of bioinformatic workflows for downstream sequence analysis. The overall aim was to develop and apply a system that could be used to analyse the composition of parasite vaccine stocks and different vaccine batches and to compare diversity in vaccine stocks with that of field parasite populations

3.2 Material and Methods

A system was developed to PCR amplify segments of genes encoding *T. parva* antigens from a variety of DNA samples and to determine the nucleotide sequence of these amplicons using Roche 454 FLX Titanium technology. A bioinformatics pipeline was developed to allow removal of artefacts from the raw sequencing data and to summarise the results. Development of the system involved the following steps:

- Design of PCR primers
- Selection of parasites for primer validation
- Primer validation and optimisation of PCR conditions
- DNA extraction of sporozoite stabilate and carrier animals
- Sample preparation
- Bioinformatic analysis

These steps are described in detail below.

3.2.1 Design of PCR primers

Primers were designed to amplify 300-500 bp segments of a number of genes known to encode antigens recognised by CD8 T cells (Tp1-Tp10). The amplicons included epitope regions, where epitopes within the antigens had been identified. The genomic reference sequences of the *T. parva* (Muguga) antigen genes and additional sequences of allelic variants provided by R. Pelle (ILRI, Kenya) and C. Sugimoto (Hokkaido University, Japan) were used to facilitate primer design. Available sequences were aligned using the MUSCLE algorithm in the Mega5 software package (Edgar, 2004; Tamura et al., 2011) and primers were designed to match regions of the genes that were conserved among the different *T. parva* parasites. Primer sequences were BLASTed against the NCBI nucleotide database to ensure they only matched the sequence of the gene of interest and would not anneal to other loci in the *T. parva* or the bovine genome. Additional primers were designed when a given primer set failed to amplify all DNA samples in the test panel or non-specific amplification was observed. Redesign of primers was necessary for most genes for a variety of reasons. The most frequent reason for failure to obtain a PCR product was

polymorphism in one or both of the primer binding sites in certain genotypes. Other reasons for re-designing primers included non-specific amplification of DNA from uninfected bovine PBMC and/or from bovine samples infected with other *Theileria* species. On one occasion, a primer was found to anneal to the adaptor part of the fusion primer. In most cases these problems were solved by testing alternative primers.

Due to sequence variation among genotypes, the use of degenerate primers was required for the Tp1, Tp2, Tp4 and Tp10 genes. For Tp2, two different forward primers and two different degenerate reverse primers were used in order to amplify each of the parasite clones in the test panel. The aim was to amplify products between 300 and 500 bp as this is the optimal length for the Roche 454 Titanium platform. For Tp2 it was impossible to design a product smaller than 500bp due to the high degree of polymorphism within the gene. The primers selected for use in subsequent studies are listed in Table 3.1.

It proved impossible to identify primers that would amplify the Tp9 gene from the majority of clones in the test panel. The gene is known to be highly polymorphic in the population and exhibits a variety of insertions and deletions. This, together with the size limitations of 454 amplicon sequencing, made it impossible to design primers that would amplify most alleles and for this reason Tp9 was excluded from the study.

Table 3.1 Primers selected for amplification of *T. parva* antigen genes and the predicted amplicon size

Primer name	Fusion primer sequence *adaptor sequence - MID - sequence specific primer*	Product size (bp) including MIDs and adaptors
Tp1_for_HH2_Ax	CGTATCGCCTCCCTCGCGCCATCAG-XXXXXXXXXX-TGGTGTACAATTTGGTGGG	498
Tp1_rev_HH2_Bx	CTATGCGCCTTGCCAGCCCGCTCAG-XXXXXXXXXX-TAACTTNMCTTCTTGCGAACC	
Tp2_for_PFdeg_Ax	CGTATCGCCTCCCTCGCGCCATCAG-XXXXXXXXXX-GCCAGATTAATHAGYCTTTAC	562
Tp2_for_RP_Ax	CGTATCGCCTCCCTCGCGCCATCAG-XXXXXXXXXX-ATGAAATTGGCCGCCAGATTA	
Tp2_rev_HH3_Bx	CTATGCGCCTTGCCAGCCCGCTCAG-XXXXXXXXXX-AGATTTGTCACTAYCTGTWBYAGG	
Tp2_rev_HH4_Bx	CTATGCGCCTTGCCAGCCCGCTCAG-XXXXXXXXXX-AGATTCGTCCTCAYCTGTWBYAGG	
Tp3_for_HH3_Ax	CGTATCGCCTCCCTCGCGCCATCAG-XXXXXXXXXX-AGCAGATTTCACTCAAGCTGC	477
Tp3_rev_HH2_Bx	CTATGCGCCTTGCCAGCCCGCTCAG-XXXXXXXXXX-TCCCCCAGAACATTAAACGG	
Tp4_for_HH3_Ax	CGTATCGCCTCCCTCGCGCCATCAG-XXXXXXXXXX-GCAACACAATACTTTGCAGG	494
Tp4_rev_HH4_Bx	CTATGCGCCTTGCCAGCCCGCTCAG-XXXXXXXXXX-CCTCAAACACWCCACAAGTTCC	
Tp5_for_HH2_Ax	CGTATCGCCTCCCTCGCGCCATCAG-XXXXXXXXXX-GTATGCTCGGTAATGGCAG	417
Tp5_rev_HH2_Bx	CTATGCGCCTTGCCAGCCCGCTCAG-XXXXXXXXXX-GATTTTGGTCGCTTCAGGC	
Tp6_for_HH2_Ax	CGTATCGCCTCCCTCGCGCCATCAG-XXXXXXXXXX-CGTCCAATAATTTACGATGTGAG	406
Tp6_rev_HH4_Bx	CTATGCGCCTTGCCAGCCCGCTCAG-XXXXXXXXXX-GCTTAAGTGGGTTAAGGAGACA	
Tp7_for_HH1_Ax	CGTATCGCCTCCCTCGCGCCATCAG-XXXXXXXXXX-TGAAGAAGGACGACTCGCAC	362
Tp7_rev_HH2_Bx	CTATGCGCCTTGCCAGCCCGCTCAG-XXXXXXXXXX-TCCTCGTCAGTGACGTCGG	
Tp8_for_HH1_Ax	CGTATCGCCTCCCTCGCGCCATCAG-XXXXXXXXXX-ATCCACAACCAAGTGCCCAG	375
Tp8_rev_HH2_Bx	CTATGCGCCTTGCCAGCCCGCTCAG-XXXXXXXXXX-TGCTATTGCGAGTCAACAGT	
Tp10_for_HH2_Ax	CGTATCGCCTCCCTCGCGCCATCAG-XXXXXXXXXX-GGTCGTCTGACAATAACC	384
Tp10_rev_HH2_Bx	CTATGCGCCTTGCCAGCCCGCTCAG-XXXXXXXXXX-CTAMCATGTAAATCCAGC	

The amplicon size was predicted based on the *T. parva* Muguga genome reference sequence. xxxxxxxxx represents the sequence of the Multiplex Identifier (MID) as can be found in the Roche Technical bulletin (TCB013-2009) 'Amplicon Fusion Primer Design Guidelines for GS FLX Titanium Series Lib-A Chemistry'.

3.2.2 Selection of parasites for primer validation

It was essential to ensure that the primers were capable of amplifying a high proportion of the allelic variants of the genes of interest. To this end, a panel of DNA samples from cloned parasitised cell lines was established in order to test the designed primers. The use of DNA from clonal parasites was considered essential, because failure to amplify alleles containing polymorphism in the primer binding sites could be disguised when attempting to amplify from DNA derived from mixed parasite populations.

Primers were initially tested using a set of ten *T. parva* (Marikebuni) clones of diverse genotypes as defined by satellite DNA typing (Katzer et al., 2006). A further set of cloned cell lines was established by the limiting dilution of five cell lines derived from African buffalo from the Masai Mara and Laikipia districts in Kenya (Conrad et al., 1987b). Further cloned cell lines were established from cell lines isolated from cattle that were used in a vaccine trial on the Marula farm near lake Naivasha, Rift Valley District, Kenya (Pelle et al., 2011; Young et al., 1992). The pastures of the farm were frequented by buffalo thus disease breakthrough was attributed to buffalo-derived *T. parva* challenge. The use of buffalo-derived *T. parva* parasites provided a rigorous test of the ability of the primers to amplify a wide range of parasite allelic variants, because there is evidence that these parasites are genotypically more diverse than cattle-derived *T. parva* parasites (Conrad et al., 1989; Oura et al., 2011b; Pelle et al., 2011). DNA was extracted from all the cloned cell lines using the DNeasy Blood & Tissue Kit (Qiagen) and a number of clones from each isolate were used for primer testing The DNA concentrations were adjusted to a standard concentration of 50 ng/µl. For 7 of the 39 cloned cell lines in the original test panel, PCR amplification failed for most genes. Further investigation demonstrated that these lines were infected with *Theileria sp (buffalo)* rather than *Theileria parva.* All the remaining clonal cell lines in the test panel (see Table 3.2). were subsequently sequenced at the 18S rRNA locus to ensure they were *T. parva* (Gubbels et al., 1999; Oura et al., 2004a). To test primer specificity, DNA preparations derived from uninfected *Rhipicephalus appendiculatus* ticks, uninfected bovine PBMC and samples infected with *Theileria taurotragi, Theileria buffeli, Theileria sp. (buffalo)* and *T. annulata* were included in the test panel. DNA samples from bovine PBMC were derived from a *Bos indicus* animal at ILRI, Kenya and a *Bos taurus* animal at the University of Edinburgh, United Kingdom (see Table 3.3).

Table 3.2 List of cloned *T. parva* cell lines used to validate primers

Location or name	Clone	Country	Origin	Material	Source
Marikebuni	A3	Kenya	bovine	*In vitro* culture	UoE
Marikebuni	A7	Kenya	bovine	*In vitro* culture	UoE
Marikebuni	B12	Kenya	bovine	*In vitro* culture	UoE
Marikebuni	E43	Kenya	bovine	*In vitro* culture	UoE
Marikebuni	F31	Kenya	bovine	*In vitro* culture	UoE
Marikebuni	F44	Kenya	bovine	*In vitro* culture	UoE
Marikebuni	F53	Kenya	bovine	*In vitro* culture	UoE
Marikebuni	I8	Kenya	bovine	*In vitro* culture	UoE
Marikebuni	I38	Kenya	bovine	*In vitro* culture	UoE
Marikebuni	J17	Kenya	bovine	*In vitro* culture	UoE
Marula N33	1	Kenya	bovine	*In vitro* culture	ILRI
Marula N33	3	Kenya	bovine	*In vitro* culture	ILRI
Marula N33	4	Kenya	bovine	*In vitro* culture	ILRI
Marula N33	5	Kenya	bovine	*In vitro* culture	ILRI
Marula N43	1	Kenya	bovine	*In vitro* culture	ILRI
Marula N43	3	Kenya	bovine	*In vitro* culture	ILRI
Marula N43	5	Kenya	bovine	*In vitro* culture	ILRI
Marula N43	6	Kenya	bovine	*In vitro* culture	ILRI
Mara 3	3	Kenya	buffalo	*In vitro* culture	ILRI
Mara 3	6	Kenya	buffalo	*In vitro* culture	ILRI
Mara 3	7	Kenya	buffalo	*In vitro* culture	ILRI
Mara 3	9	Kenya	buffalo	*In vitro* culture	ILRI
Mara 30	2	Kenya	buffalo	*In vitro* culture	ILRI
Mara 30	5	Kenya	buffalo	*In vitro* culture	ILRI
Mara 30	8	Kenya	buffalo	*In vitro* culture	ILRI
Mara 30	11	Kenya	buffalo	*In vitro* culture	ILRI
Mara 42	2	Kenya	buffalo	*In vitro* culture	ILRI
Mara 42	5	Kenya	buffalo	*In vitro* culture	ILRI
Mara 42	8	Kenya	buffalo	*In vitro* culture	ILRI
Mara 42	12	Kenya	buffalo	*In vitro* culture	ILRI
Mara 6998	9	Kenya	buffalo	*In vitro* culture	ILRI
Mara 6998	11	Kenya	buffalo	*In vitro* culture	ILRI

UoE University of Edinburgh, UK; ILRI International Livestock Research Institute, Nairobi, Kenya

Table 3.3List of samples to test primer specificity

Genus and species	Location or name	Country	Origin	Material	Source
T. buffeli	Marula	Kenya	buffalo	blood	ILRI
T. taurotragi	-	Kenya	bovine	culture	ILRI
T. sp (Buffalo)	Buffalo 6834 clone 3	Kenya	buffalo	culture	ILRI
T. annulata	Ankara C9	Turkey	bovine	culture	UoE
R. appendiculatus	Muguga	Kenya	tick	tick	ILRI
Bos indicus	ILRI	Kenya	bovine	PBMC	ILRI
Bos taurus	Edinburgh	UK	bovine	PBMC	UoE

UoE University of Edinburgh, UK; ILRI International Livestock Research Institute, Nairobi, Kenya

3.2.3 Primer validation and optimisation of PCR conditions

PCR conditions were optimised using DNA from a *T. parva* Muguga cell line employing different primer concentrations and different cycling conditions. The optimised conditions were then used to amplify DNA from the test panel of *T. parva* clones. At least three PCR products per primer pair were sequenced to ensure that the amplified products were indeed amplicons of the gene of interest. Following this, fusion primers were used for amplification, as these would be required to facilitate 454 sequencing. The use of different combinations of MIDs allowed the reads to be identified as belonging to specific samples when different templates are sequenced together on the same region of the picotitre plate.

Two different polymerases with proofreading ability were used in an attempt to reduce PCR error: Fast start high fidelity enzyme blend (Roche, Mannheim, Germany) and PFU DNA polymerase (Promega, Madison, USA). The reaction mixtures are detailed in Table 3.4. However, for a number of genes, amplification using these proofreading polymerase was poor and in these cases, a combination of Pfu DNA polymerase and BIOTAQ^{tm} (Bioline, UK) was used at a ratio of 1:1. PCR conditions were optimised by testing a range of temperatures, primer concentrations and cycling conditions. Both the use of fusion primers and the use of proofreading polymerases reduced the efficiency of the PCR amplification and the reaction mix used was adapted to the polymerase used (see Table 3.5).

Table 3.4 Components of PCR reaction mix for sample preparation

Roche FastStart High Fidelity PCR system (cat no. 03553426001 or 03553400001)	
5 µl	FastStart High Fidelity Reaction buffer 10x with 18mM MgCl2
1 µl	dNTPs (10mM)
X*	Forward primer
X*	Reverse primer
2.5 units	Fast Start High fidelity enzyme blend
	water
1 µl	DNA
50 µl	total reaction volume

BIOTAQ + PFU	
5 µl	µl 10x custom PCR Master mix (ThermoScientific SM-005)
X*	Forward primer
X*	Reverse primer
1 unit	BIOTAQ DNA polymerase (Bioline BIO 21040)
1 unit	PFU DNA polymerase (Promega M7741)
	water
1 µl	DNA
50 µl	total reaction volume

X* For primer concentration see table 3.5. Recommended quantity of DNA: 50 ng for parasitised cell lines, DNA equivalent to 5µl of whole blood or tick stabilate.

Table 3.5 Summary of primer conditions for sample preparation for 454 amplicon sequencing

Gene	Primers	Concn (pmol/ 50µl reaction)	PCR reaction	Cycling conditions
Tp1	F CTGGTGTACAATTTGGTGGG R AACTTNMCTTCTTGCGAACC	20 20	Roche	95°C for 3 min 30 cycles (95°C for 30 s, 55°C for 30 s, 72°C for 30 s) 72°C for 7 min
Tp2	F ATGAAATTGGCCGCCAGATTA R1 AGATTTGTCACTAYCTGTWBYAGG R2 AGATTCGTCCTCAYCTGTWBYAGG	10 25 25	BIOTAQ + PFU	95°C for 3 min 35 cycles (95°C for 1 min, 49°C for 1 min, 72°C for 2 min) 72°C for 7 min
	F GCCAGATTAATHAGYCTTTAC R1 AGATTTGTCACTAYCTGTWBYAGG R2 AGATTCGTCCTCAYCTGTWBYAGG	10 25 25	BIOTAQ + PFU	95°C for 3 min 35 cycles (95°C for 1 min, 49°C for 1 min, 72°C for 2 min) 72°C for 7 min
Tp3	F AGCAGATTTCACTCAAGCTGC R TCCCCCAGAACATTAAACGG	20 20	Roche	95°C for 3 min 30 cycles (95°C for 30 s, 54°C for 30 s, 72°C for 30 s) 72 °C for 7 min
Tp4	F GCAACACAATACTTTGCAGG R CCTCAAACACWCCACAAGTTCC	10 10	Roche	95°C for 3 min 30 cycles (95°C for 30 s, 54°C for 30 s, 72°C for 30 s) 72 °C for 7 min
Tp5	F GTATGCTCGGTAATGGCAG R GATTTTGGTCGCTTCAGGC	10 10	Roche	95°C for 5 min 30 cycles (95°C for 30 s, 55°C for 30 s, 72°C for 30 s) 72°C for 7 min
Tp6	F CGTCCAATAATTTACGATGTGAG R GCTTAAGTGGGTTAAGGAGACA	10 10	Roche	95°C for 3 min 30 cycles (95°C for 30 s, 55°C for 30 s, 72°C for 30 s) 72°C for 7 min
Tp7	F TGAAGAAGGACGACTCGCAC R TCCTCGTCAGTGACGTCGG	20 20	Roche	95°C for 3 min 30 cycles (95°C for 30 s, 58°C for 30 s, 72°C for 30 s) 72 °C for 7 min
Tp8	F ATCCACAACCAAGTGCCCAG R TGCTATTGCGAGTCAACAG	10 10	Roche	95°C for 3 min 30 cycles (95°C for 30 s, 54°C for 30 s, 72°C for 30 s) 72°C for 7 min
Tp10	F GGTCGTCTGACAATAACC R CTAMCATGTAAATCCAGC	10 20	BIOTAQ + PFU	95°C for 3 min 35 cycles (95°C for 1 min, 49°C for 1 min, 72°C for 2 min) 72°C for 7 min

Only the sequence-specific regions of the fusion primers are displayed. PCR conditions are indicated with Roche or BIOTAQ + PFU (See table 3.4)

3.2.4 DNA extraction from sporozoite stabilate and carrier animals

Amplification using the sporozoite stabilate derived DNA was poor compared to that using DNA derived from cell lines (see Figure 3.2), which is likely to be due to the relative low abundance of parasites in the *T. parva* sporozoites stabilates.

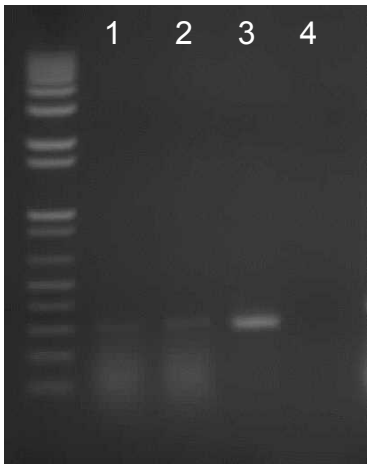

Figure 3.2 PCR amplification using Tp8 specific primers and DNA extracted from stabilate and T. parva infected cell line
1. *T. parva* Muguga stabilate 4230; **2.** *T. parva* Muguga cocktail stabilate ILRI0801; **3.** Positive control: *T. parva* Muguga cell line; **4.** Negative control: H_2O

A number of different protocols for extracting DNA were investigated in an attempt to improve PCR amplification (see Figure 3.3). DNA was extracted from Stabilate 93 (*T. parva* Muguga, produced at CTVM) using a variety of DNA extraction kits according to the manufacturers' instructions: Wizard® Genomic DNA Purification Kit (Promega), DNeasy Blood & Tissue Kit (Qiagen) and ChargeSwitch® gDNA Micro Tissue Kit (Invitrogen™). The efficiency of PCR following DNA extraction with the DNeasy Blood & Tissue Kit was comparable with DNA extraction using ChargeSwitch® gDNA Micro Tissue Kit. The efficiency of PCR was slightly lower with DNA extracted using the Wizard® Genomic DNA Purification Kit. It was therefore decided that DNA extraction would continue to be performed using the DNeasy Blood & Tissue Kit. Stabilate 93 provided a higher yield of PCR product than the Muguga cocktail component stabilates regardless of the method of DNA extraction.

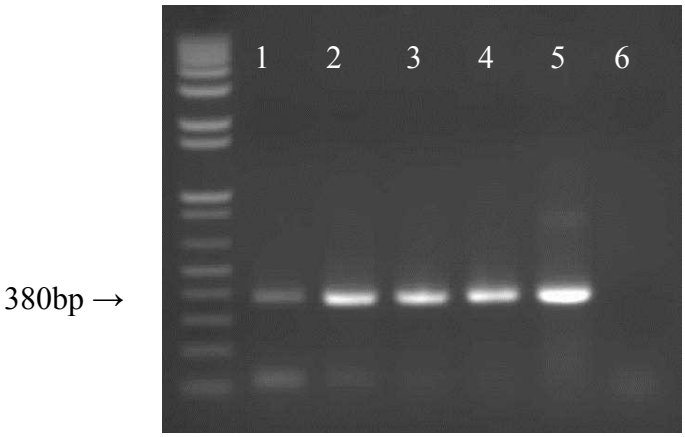

Figure 3.3 PCR amplification using Tp8-specific fusion primers and DNA extracted from stabilate using different methods
1. Promega kit; **2.** Charge switch kit; **3.** Qiagen column, using water for elution; **4.** Qiagen column, using AE buffer for elution; **5.** positive control DNA from *T. parva* Muguga cell line; **6.** Negative control H_2O

A modified protocol for processing tick stabilate material for DNA extraction had been developed at ILRI. This protocol involved an additional centrifugation step to discard particulate tick material present in the stabilates. Thawed stabilate was centrifuged at 500 x g for 5 min and the supernatant, which contained the sporozoites was used for further processing using the Qiagen DNeasy Blood and tissue kit, according to the manufacturer's instructions. DNA extracted using this modified protocol produced a more abundant product following PCR amplification (see Figure 3.4).

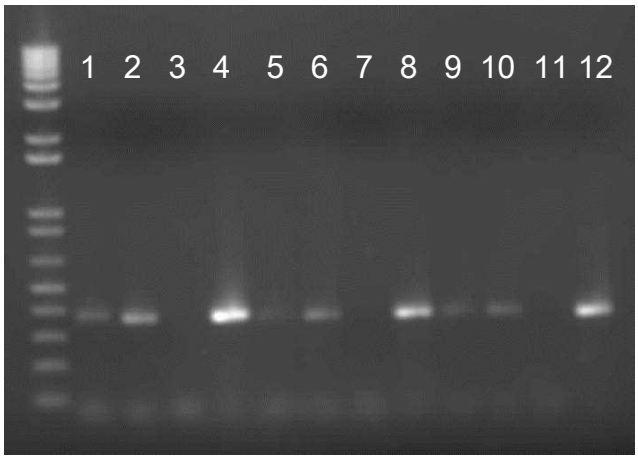

Figure 3.4 PCR amplification of the Tp8 gene before and after removal of excess tick material by centrifugation

DNA was extracted from the Muguga (St4230), Serengeti-transformed (St4229) and Kiambu 5 (St4228) stabilates by using stabilate material directly or by using material obtained after a centrifuge step and removal of particulate tick material (spin). Negative control is water. Postive control is DNA from *T. parva* Muguga cell line.

1. St4230 directly
2. St4230 spin
3. Negative control
4. Postive control
5. St4229 directly
6. St4229 spin
7. Negative control
8. Postive control
9. St4228 directly
10. St4228 spin
11. Negative control
12. Positive control

3.2.5 Sample preparation

Genes of interest were amplified from each sample using a unique combination of MID tags, employing the PCR conditions detailed in Tables 3.4 and 3.5. The PCR products were purified by gel extraction using the Wizard® PCR Preps DNA Purification System (Promega Corporation, USA) according to the manufacturer's instructions. The purity of PCR products was assessed using Agilent DNA1000 chips on a 2100 Bioanalyzer (Agilent Technologies, USA). A more accurate quantification of the PCR product was performed using a picogreen assay (Thermo Fischer Scientific, USA) before equimolar quantities of PCR products were pooled. Products were pooled such that amplicons did not differ in size by more than 50 bp in a single pool. In some cases it was necessary to submit several pools of different size to the sequencing centre. The pools of PCR product were submitted to commercial 454 sequencing centres, initially Genepool, University of Edinburgh and later to the Centre for Genomics Research (CGR), University of Liverpool, where amplicon pools of different size were processed in separate emulsion-based PCRs (emPCR). The pools of different sizes of amplicons were themselves pooled following emPCR.

3.2.6 Bioinformatics

3.2.6.1 De-multiplexing sequence reads

A large number of samples may be combined in a single 454 sequencing run providing that individual samples are uniquely identified. Amplicons from different animals or different parasite isolates may therefore be pooled and analysed simultaneously and this is facilitated by the use of primers incorporating unique Multiplex Identifier (MID) tags. The MID tags allow identification of the source of the sequences and different combinations of 5' and 3' MID tags may be used to provide unique identifiers where larger numbers of template samples are being processed. A bioinformatics pipeline was developed to de-multiplex read data from the raw sequencing files into files corresponding to individual PCR reactions. This bioinformatics pipeline is illustrated in Figure 3.5. Briefly, the raw reads (contained within flowgram files) were first segregated on the basis of the MID tag identified at the 5' end of the read using a script provided by Roche, 'sfffile'. The sff files were then converted to FASTA files using another script provided by Roche, 'sffinfo', and further segregation was done by identifying the MID tag at the 3' end of the read using a script provided by Genepool, 'split_by_end_MID_exact_match_only.sh'. The reads in these FASTA files were further partitioned on the basis of the primer sequence identified using another script provided by Genepool, 'split_by_fasta_file.sh' and a reference file containing the primer sequences, e.g. 'hannekesprimers.txt'. The reads were all oriented in the direction they were sequenced, from either the adaptor A-side or the adaptor B-side. This meant that reads with a specific MID tag at the 5' end of the adaptor A-side needed to be differentiated from reads with the same MID tag at the 5' end of the adaptor B-side. Therefore, reads derived from the same sample (i.e. PCR reaction) but sequenced from a different adaptor side were combined in a single file. The names of the reads belonging to an individual sample were extracted using a script provided by Dr Willie Weir, University of Glasgow, 'extract_readnames.pl'. The read names were then used to extract the raw reads from the original flowgram file using the sfffile script provided by Roche.

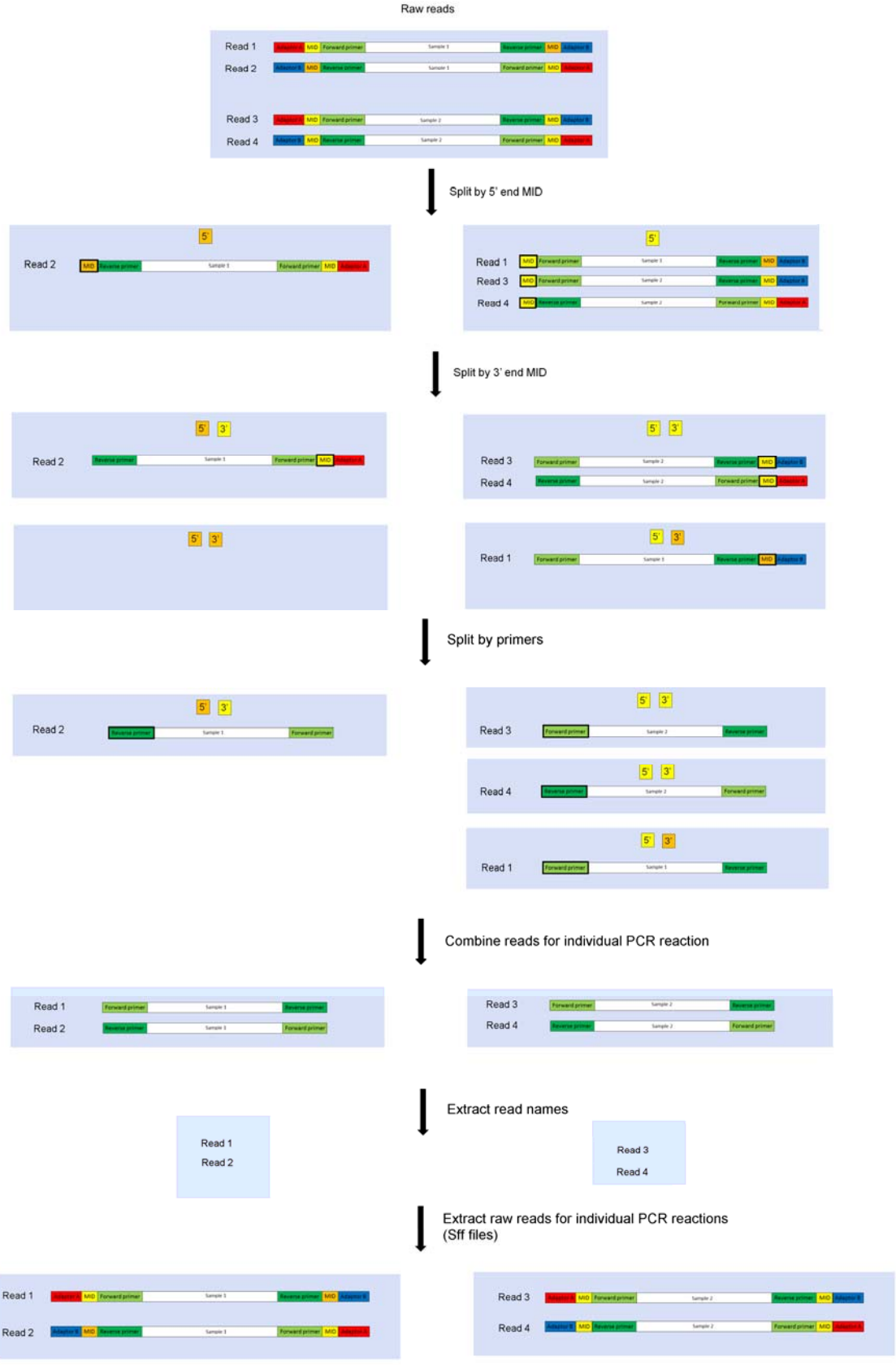
Raw reads
Read 1
Read 2
Read 3
Read 4
Split by 5' end MID
5'
Read 2
Read 1
Read 3
Read 4
Split by 3' end MID
5' 3'
Read 2
Read 3
Read 4
Read 1
Split by primers
Read 2
Read 3
Read 4
Read 1
Combine reads for individual PCR reaction
Read 1
Read 2
Read 3
Read 4
Extract read names
Read 1
Read 2
Read 3
Read 4
Extract raw reads for individual PCR reactions
(Sff files)
Read 1
Read 2
Read 3
Read 4

Figure 3.5 Illustration of bioinformatics pipeline for generating files with reads from individual samples.
The files were generated using the sequences of the adaptor, MID and sequence specific components of the fusion primers (See figure 3.1)The adaptor parts of the fusion primers at the end of the reads are displayed in red for adaptor A and in blue for Adaptor B. The MID sequence in the fusionprimer is displayed in either yellow or orange depending on the MID combination used for amplification of the sample. The sequence specific part of the fusion primer is displayed in green, light green for the forward primer and dark green for the reverse primer.

3.2.6.2 Segregating reads Tp2

Due to the high level of polymorphism in the Tp2 gene, it was impossible to design primers that would generate amplicons of less than 500 bp. Therefore, primers were used which amplified a PCR product of around 560 bp, including fusion primers and adaptors. Although read lengths of up to 600 bp may sometimes be achieved using Roche 454 FLX Titanium chemistry, in practice many reads fall short of this length and this renders them too short to identify both the 5' and 3' MID tag. Therefore, for Tp2 the same unique MID tag was used in the both the 5' and 3' primers for PCR from each sample, allowing assignment of reads to a specific PCR reaction on the basis of the presence of a single primer and MID tag. Since sequencing quality deteriorates toward the end of a read, it was decided to analyse the forward and reverse reads separately and to only use the first 350 nucleotides of sequence data for variant identification. Noise reduction in the sequence dataset was undertaken using the Mothur platform, which is an open-source platform which incorporates a suite of bioinformatic tools that are useful for the analysis of microbial ecology data (see below) (Schloss et al., 2009).

3.2.6.3 Reducing ‘noise’ in the sequencing data

Large amounts of data can be generated using high throughput sequencing platforms. However, due to limitations of the technology combined with the very high number of reads generated, the absolute number of errors is considerable. For the purpose of genome assembly, these errors can largely be eliminated as a consensus sequence may be created when sufficient sequencing depth is achieved. However, these errors present a challenge when using high throughput sequencing to discover novel allelic variants in specific genes. It can be difficult to determine whether minor sequence variations are simply errors introduced during sample processing or whether they

represent genuine allelic variants. Errors can be divided into those introduced during PCR and those which arise during the sequencing process. Mis-incorporation of nucleotides during the polymerisation process of PCR is a common source of error and it is known that different DNA polymerases are associated with different error rates. For most genes Roche high fidelity polymerase was used for amplification, which has six-fold higher accuracy than *Taq* polymerase. Chimaera formation during PCR may also introduce additional false sequence variants. Chimaeras are formed when an incompletely extended DNA strand functions as a primer and anneals to a second DNA strand of different sequence in a subsequent round of the reaction. This results in a product which is a hybrid of two or more target sequences. Lastly, a bias toward particular variants may be introduced during the sample preparation process and PCR and this may be related to DNA extraction methodology and primer specificity. While not a source of sequence error, as such, the proportion of different amplicon species in the PCR product may not accurately reflect the composition of the template material (Schloss et al., 2011).

For Roche 454 pyrosequencing, the error rate per nucleotide base introduced during the sequencing process is comparable with that of conventional Sanger sequencing (Liu et al., 2012) and is lower than that of other sequencing platforms such as Illumina and Solid v4. 454 pyrosequencing utilised the sequencing by synthesis process where 'flows' of nucleotides are washed over the picotitre plate. When nucleotides are incorporated in the growing DNA molecule, a light signal is emitted and the intensity of this light signal correlates with the number of nucleotides incorporated (Margulies et al., 2005). This correlation is less precise when there is sequential incorporation of a high number of the same nucleotide and this frequently results in mis-calling the number of single nucleotides incorporated. This is a well-recognised phenomenon and is termed a 'homopolymer error'. In addition, spurious single base-pair insertions and deletions are known to occur. A bioinformatics pipeline was developed to account for these types of error in the raw sequencing data.

The most common application for 454 FLX Titanium amplicon sequencing is the analysis of 16S subunit rRNA gene sequences in microbial ecology studies. A frequently used platform for the processing of 16S RNA deep sequencing data is Mothur. The platform, which incorporates a suite of bioinformatic tools that are useful for the analysis of microbial ecology data, is continuously improved and updated (Schloss et al., 2009). The algorithms utilised by the software have been developed to either correct errors in the reads or by entirely removing erroneous reads. Quince et al. wrote the 'Pyronoise' and 'SeqNoise' algorithms (Quince et al., 2009; Quince et al., 2011). Pyronoise operates by clustering the raw flowgram data and aims to remove errors introduced during the sequencing process. The distribution of observed light intensities is used to define distances between the flowgrams within a cluster and the probability that a flowgram was generated by a given sequence is calculated. The most likely sequence is then calculated using an expectation-maximisation algorithm (Quince et al., 2009). Although the algorithm was written for the analysis of 16S rRNA data, it can be used for other loci that have been subjected to pyrosequencing (Quince et al., 2009). The algorithm was re-implemented in a more accessible and robust format in the Mothur software package as 'shhh.flows' (Schloss et al., 2011). This package employs a modified single-linkage clustering algorithm, which clusters sequences within a specified distance of each other. It assumes that sequences that are more abundant are correct (Huse et al., 2010; Schloss et al., 2011). A second noise-removal step utilises a program called 'shhh.seqs', which is a re-implementation of the Seqnoise algorithm written by Quince *et al.* (Quince et al., 2011). This algorithm deals with single base pair errors by clustering sequences with a set maximum difference between sequences. The nucleotide transition probabilities are calculated and used to define a distance between sequences that reflects the probability that a given read could have been generated from a true sequence, given PCR error. This was done using a mixture model to cluster sequences; all sequences within a cluster are considered to correspond to a true sequence about which observed noisy reads are distributed.

Lastly, the mothur platform incorporates several chimera detection algorithms (Edgar et al., 2011; Haas et al., 2011; Quince et al., 2011). Perseus was written for the use of detecting chimeras in the absence of a reference database, the algorithm is trained using a set of sequences similar to the sequences of interest (Quince et al., 2011). Uchime and Perseus outperform Chimera slayer when used without a reference database (Schloss and Westcott, 2011). All three programs were used in order to minimise the chance of false variant calling due to the presence of chimera.

A shell script, 'process_sfffiles.sh', was written by Dr Willie Weir to automate the analysis of multiple flowgram files using Mothur. This script requires an input directory with flowgram files of individual PCR reactions and subjects the sequence data in each of these files to a series of bioinformatics steps. The minimum number or minimum percentage of variant reads to be considered a genuine allelic variant can be set. The sequence data representing the output of the bioinformatic pipeline could then be manually inspected using a MUSCLE alignment using MEGA5 (Edgar, 2004; Tamura et al., 2011). Further manual correction of homopolymer errors was performed as deemed appropriate. This was achieved by adding or deleting single nucleotides when the length of the homopolymer tract differed from that of the most common allele(s); not correcting such errors would have lead to frame shifts and thus a radical change to the translated sequence and the appearance of spurious stop codons. In the case of single base pair insertions and deletions, reads were discarded if this resulted in non-coding sequence. A number of the genes have introns (Tp4, Tp5 and Tp10) and this issue was taken into consideration when translating to amino acid sequence.

3.3 Results

3.3.1 Performance of bioinformatic pipeline on clonal samples

In order to gain an insight into the efficiency of the bioinformatics pipeline in accounting for noise in the sequence dataset, amplicons from a cloned parasite were sequenced. This cloned parasite was obtained by *in vitro* cloning of a *T. parva* infected cell line. The reads for individual PCR reactions were obtained by identification of one MID tag and one primer and this was possible as the sample was amplified using the same MID tag in both the forward and reverse primers. The files were then processed using Mothur. A summary of the file for the Mothur pipeline is shown in Table 3.6. For the majority of genes, the bioinformatics pipeline was capable of removing the noise and a single, genuine, allele was determined following execution of the pipeline. However, for some of the genes (Tp3, Tp5 and Tp7) there were still two sequences present, meaning that the bioinformatics pipeline had not fully resolved all the PCR and/or sequencing errors that occurred during processing.

The type of error varied between different genes (See figure 3.6-3.8). For Tp3, single base-pair insertions were found for the control sample and for other samples sequenced in the same run as the control. However, the same type of error was not encountered for Tp3 for other samples in different sequencing runs. When translating the nucleotide sequence to amino acid sequence, insertions lead to frame shifts and premature stop codons. These erroneous reads could be filtered out by discarding reads in which single base pair insertions and deletions lead to incomplete coding sequence.

For Tp5, differences in the length of homopolymer were found, despite the application of algorithms for homopolymer correction in the bioinformatics pipeline. These homopolymer errors, which also resulted in altered open reading frames, were corrected manually by removing or adding single nucleotides to adjust the length to that of the most common allele. These errors were not reproduced in samples sequenced on other sequencing runs.

For Tp7, three single base-pair substitutions were found, which were difficult to identify as errors when manually inspecting the resulting sequencing reads. The reads with errors were present at a low frequency (7/2196 = 0.32%). In light of this result, the script was adapted to allow a minimum percentage of reads before a variant was deemed genuine. Thus, assuming that a 0.32% error rate is representative, a cut-off of 0.4% would remove the reads with potential errors but still allow detection of allelic variants at a frequency as low as 1 in 250. The errors were not reproduced in a repeat of sequencing Tp7 product using the same clone in a different sequencing run (see Table 3.6). Gilles *et al.* estimated that a minimum of five sequences are required to be able to effectively correct errors (Gilles et al., 2011) and therefore a minimum of five reads was applied in addition to the minimum of 0.4%.

Table 3.6 Summary of bioinformatics processing of sequences obtained for the control clone

Filename	Initial no. Reads	No. of unique sequences after de-noising			No. of chimaeras detected			Sensitivity (%)				Variants after chimaera removal & above thresholds		'Total Good reads	Hz	Counts
		1st	2nd	CLUS	S	P	U	100	200	500	1000	>min no. reads	>min %			
control.Tp1.sff	334	87	11	11	0	0	0	94.94	76.9	44.5	25.4	1	1	294	0	294
Control_Forward. Tp2.sff	718	203	26	26	0	0	0	99.82	95.7	71.44	46.13	1	1	625	0	625
Control_Reverse. Tp2.sff	858	481	114	112	0	0	0	99.91	96.78	74.92	49.7	1	1	690	0	690
control.Tp3.sff	3159	443	45	45	0	0	0	100	100	99.7	94.6	2	2	2921	0.014	2899 22
control.Tp4.sff	1206	241	36	36	0	0	0	99.99	99.6	88.8	66.3	1	1	1087	0	1087
control.Tp5.sff	979	121	22	22	0	0	0	99.99	99.0	83.7	59.9	2	2	912	0.228	792 120
control.Tp6.sff	1787	136	5	5	0	0	0	100	99.9	96.7	81.6	1	1	1701	0	1701
control.Tp7.sff	2246	141	16	16	0	0	0	100	100	98.8	89.1	2	2	2196	0.007	2188 8
control.Tp7II.sff	4353	131	3	2	0	0	0	100	100	99.98	98.7	1	1	4327	0	4327
control.Tp8.sff	2505	160	20	19	0	0	0	100	100	99.3	91.3	1	1	2444	0	2444
control.Tp10.sff	1577	154	16	16	0	0	0	100	99.9	95.2	78.1	1	1	1517	0	1517

The number of reads at the different stages of the bioinformatics pipeline are shown together with the results before manual editing. 1st, 2nd and CLUS refer to the stages of noise removal. Chimaera detection was performed using three different algorithms: S = Chimaera Slayer P = Perseus U = Uchime. Sensitivity was calculated as the probability of identifying a minor allele present at 1:100, 1:200, 1:500 and 1:1000. '>min no. reads' = the number of variants found after removal of variants with less than 3 reads; '>min%' = the number of variants found after removal of variants which are present at a frequency of less than 0.3% of the total good reads; 'Total good reads' = the number of reads maintained after filtering of reads in the bioinformatics pipeline; Hz= heterozygosity; Counts = the number of reads for each of the alleles detected.

```
Control Tp3_1_2899      CTGTTAATTCCATGAGTAATGCTGTTAGTGCCATGAACAATACTGTTGGTCCAGC---AA
Control Tp3_2_22        CTGTTAATTCCATGAGTAATGCTGTTAGTGCCATGAACAATACTGTTGGTCCAGC---AA

Control Tp3_1_2899      ATCCCATGGCTACTGTTGGTACTATGAACTCCTTTACTGGAACGCCTGGTGTACAGGATA
Control Tp3_2_22        ATCCCATGGCTACTGTTGGTACTATGAACTCCTTTACTGGAACGCCTGGTGTACAGGATA

Control Tp3_1_2899      CTTTTTCTCAGACACCGCCTGTTAATGTTCAAGACACCTCTACCCAAGAGAACAGTCTTG
Control Tp3_2_22        CTTTTTCTCAGACACCGCCTGTTAATGTTCAAGACACCTCTACCCAAGAGAACAGTCTTG

Control Tp3_1_2899      ACAACCTAAATCTCCTCTTAGATCCTTCGTTAGCAAAGATATCTCAAGCTGATAGTCACA
Control Tp3_2_22        ACAACCTAAATCTCCTCTTAGATCCTTCGTTAGCAAAGATATCTCAAGCTGATAGTCACA


Control Tp3_1_2899      TAAAAGAAAGCATGGTTAGTTTTATTTTTATTAATTTGTTC-AGGAAAAAGCTGTACACA
Control Tp3_2_22        TAAAAGAAAGCATGGTTAGTTTTATTTTTATTAATTTGTTCAAGGAAAAAGCTGTACACA

Control Tp3_1_2899      G-CCTTAAAAAGGTCTT-GGAGGGGCTAACCAACCTTGCGACTCTGT-CTAAAA-GTAGG
Control Tp3_2_22        GACCTTAAAAAGGTCTTGGGAGGGGCTAACCAACCTTGCGACTCTGTACTAAAAGGTAGG

Control Tp3_1_2899      GATACAG
Control Tp3_2_22        GATACAG
```

Figure 3.6 Alignment of Tp3 sequences after bioinformatic processing of sequences obtained from the control clone

```
Control Tp5_1_792       ACTTGAAGCCTACTG-TTTTGACGGCACTAAACGTCTTTGCCATATTAGGTATCGCTACT
Control Tp5_2_120       ACTTGAAGCCTACTGTTTTTGACGGCACTAAACGTCTTTGCCATATTAGGTATCGCTACT

Control Tp5_1_792       ACGCTAATTCTACCATTCACCTAATATCTTTAACTAGTATTAGTTTTACACTATTACGCT
Control Tp5_2_120       ACGCTAATTCTACCATTCACCTAATATCTTTAACTAGTATTAGTTTTACACTATTACGCT

Control Tp5_1_792       AATAAATTAGTTATACATATGGTTAACAATAATTT-AAAAATGGATTAGAGGAAAGATGA
Control Tp5_2_120       AATAAATTAGTTATACATATGGTTAACAATAATTTAAAAAATGGATTAGAGGAAAGATGA

Control Tp5_1_792       GGAAGCGAGTTTGGGTAAATGCTGGTGATATTATTTTGGTATCTCTTAGAGATTTCCAGG
Control Tp5_2_120       GGAAGCGAGTTTGGGTAAATGCTGGTGATATTATTTTGGTATCTCTTAGAGATTTCCAGG

Control Tp5_1_792       ACAGCAAGGCTGACGTGATCGCAAAGTACACTGCTGAGGAGGCTCGTACTCTGAAGGCTT
Control Tp5_2_120       ACAGCAAGGCTGACGTGATCGCAAAGTACACTGCTGAGGAGGCTCGTACTCTGAAGGCTT

Control Tp5_1_792       ACGGCGAGTT
Control Tp5_2_120       ACGGCGAGTT
```

Figure 3.7 Alignment of Tp5 sequences after bioinformatic processing of sequences obtained from the control clone

```
Control Tp7_1_2188      GAGCCGCTCAAAAGAGGAACGAGACTAATACTGCACTTGAAGGAGGACCAAACTGAGTAC
Control Tp7_2_8         GAGCCGCTCAAAAGAGGAACTAGACTAATACTGCACTTGAAGGAGGACCAAACTGAGTAC

Control Tp7_1_2188      CTTGAGGAGAGAAGGCTGAAAGAGCTTGTTAAGAAGCACAGCGAGTTCATTTCATTCCCA
Control Tp7_2_8         CTTGAGGAGAGAAGGCTGAAAGAGCTTGTTAAGAAGCACAGCGAGTTCATTTCATTCCCA

Control Tp7_1_2188      ATCTCGCTCTCAGTAGAGAAGACCCAGGAGACCGAGGTCACTGACGACGAGGCAGAGCTA
Control Tp7_2_8         ATCTCGCTCTCAGTAGAGAAGACCCAGGAGACCGAGGTCACTGACGACGAGGCAGAGCTA

Control Tp7_1_2188      GACGAGGACAAGAAGCCCGAGGAGGAAAAGGCCAAGGACGATAAGGTGGAGGACGTTACT
Control Tp7_2_8         GACGAGGACAAGAAGCCCGAGGAGGAGAAGCCCAAGGACGATAAGGTGGAGGACGTTACT

Control Tp7_1_2188      GACGAGAAAGTGA
Control Tp7_2_8         GACGAGAAAGTGA
```

Figure 3.8 Alignment of Tp7 sequences after bioinformatic processing of sequences obtained from the control clone

In summary, analysis of sequences from the cloned parasite revealed a low incidence of errors that were not identified and removed by the bioinformatics pipeline. However, it proved possible to remove any remaining errors by manually inspecting and adjusting suspect read data and by adjusting the minimum number (or percentage) of reads that is required before a variant is deemed genuine.

3.3.2 Performance of bioinformatic pipeline on heterogeneous samples

The PCR amplification of the cloned parasite as described in the previous section was unlikely to result in chimaeric sequences. Moreover, sequencing errors would theoretically be easier to detect given that the one 'genuine' variant is present at a very high frequency and therefore, it was important to further test the bioinformatics pipeline. A second sequencing run was performed and this included amplicons of Tp1, Tp3, Tp5, Tp7 and Tp8 from two samples; Kiambu 5 a component of the Muguga cocktail vaccine and SC06, a sample derived from a carrier buffalo from the Kruger National Park. It was anticipated that Kiambu 5 would show limited diversity whereas the sample from the carrier buffalo would display more extensive diversity, thus providing examples of the different types of data that will be analysed in subsequent studies.

3.3.2.1 De-multiplexing reads on the basis of MID tags

The samples described above were amplified using the same MID tag at the 5' and 3' end of the amplicon. The sequence reads were segregated first by finding the 5' MID tag and subsequently by finding the 3' MID tag. The percentages of reads that had MID tags on both ends in forward and reverse reads are shown in Table 3.7; 86% of the sequences in this sequencing run were full-length. Thus, if both MIDs were required to assign reads to a specific sample (e.g. when relying on combinations of different 5' and 3' MIDs), only 86% of the reads could be used for further analysis.

Table 3.7 Summary of number of reads after partitioning the raw reads by 5' MID only or by 5' MID and 3' MID subsequently

	5' MID	5' MID and 3' MID	% reads with both MIDs
Total	37,530	32,342	86%
5' MID1	19,442	17,119	88%
5' MID2	18,088	15,223	84%

3.3.2.2 De-multiplexing reads on the basis of primer sequence

To obtain the reads corresponding to individual PCR reactions the sequence data required to be further partitioned using the primer sequences. This was done using a script provided by Stephen Bridgett, Genepool, University of Edinburgh. In some cases, the sequence quality decreases towards the end of the read and permitting mismatches with the primer sequence can be used to maximise the number of sequence reads obtained for each PCR reaction. Reads were processed using four different methods in an attempt to establish the optimal parameters for bioinformatic processing (see Table 3.8) to produce files that can be used for further processing using Mothur (Tables 3.9 and 3.10). Two methods, which required the identification of MIDs at both the 5' and 3' end of the reads, reduced the number of reads that could be used for final analysis in Mothur. Methods which involved allowing mismatches in the primer region increased the number and percentage of identified reads for the different PCR reactions. This increase was more pronounced for Tp1 and Tp3. These genes corresponded to the two longest amplicons of the five genes amplified during this analysis; amplicon sizes for Tp1 were between 510 and 560 nucleotides and for Tp3 were approximately 485 nucleotides.

Table 3.8 Summary of the characteristics of the four read groups tested

Read group	Number of MIDs	Number of primers	Mismatches in primer(s)
1	1	1	0
2	1	2	2
3	2	1	0
4	2	2	2

Reads were processed in four different ways: 1 The reads were assigned to a sample by the identification of one Multiplex Identifier (MID) and the identification of one primer (either forward or reverse), which was identical to the primer used for PCR (no-mismatches). 2) Reads were assigned by the identification one MID and the identification of both the reverse and forward primer used for PCR amplification. Up to two mismatches were allowed to identify the primer sequences. 3) Reads were assigned to a sample by the identification of both MIDs (at adaptor A and adaptor B) and one primer (either forward or reverse) 4) Reads were assigned to a sample by the identification of both MIDs and both primers, allowing up to two mismatches for the identification of primers.

Tables 3.9 and 3.10 show a summary of the progress of the different steps of the Mothur bioinformatics pipeline, including the number of unique variants remaining after the different steps and the number of chimaeric variants detected. It also details the number of variants present that exceed the minimum threshold number of reads (n = 5) and the number of variants present above a minimum frequency value (0.4%).

Gilles *et al.* estimated that a minimum of five sequences are required to be able to effectively correct errors (Gilles et al., 2011). Based on this and the analysis of sequences obtained for the control clone a minimum threshold of 5 reads was set together with a minimum threshold of 0.4% of the total number of good quality reads before a variant was deemed genuine. These thresholds led to the rejection of a considerable number of variants. However, for the buffalo sample, the number of variants was still high even when these thresholds were applied (see Table 3.10). The method used for de-multiplexing reads had little impact on the number of variants found. Therefore, it was decided to use method 4 for de-multiplexing reads as identification of both MIDs and both primers was the most stringent filtering of the four methods assessed.

Table 3.9 Summary of the file generated by processing reads from Kiambu 5

Gene	Group	Initial no. of reads	No. of unique sequences after de-noising			No. of chimaeras detected			Sensitivity		Variants after chimaera removal & above thresholds		Total good reads	Hz
			1st	2nd	Pre Clus	S	P	U	100	200	>min no. reads	>min %		
Tp1	1	890	135	12	12	0	0	0	100	98.3	2	2	820	0.015
	2	823	112	11	11	0	0	0	100	98.0	2	2	776	0.013
	3	803	108	12	12	0	0	0	99.9	97.8	2	2	757	0.016
	4	762	96	12	12	0	0	0	99.9	97.4	2	2	729	0.016
Tp3	1	2760	672	75	74	7	1	0	100	99.9	14	2	2238	0.117
	2	2473	594	79	79	6	1	0	100	99.9	11	2	2062	0.108
	3	2372	604	79	78	6	1	0	100	99.9	12	4	1934	0.113
	4	2195	522	78	78	7	1	0	100	99.9	11	2	1848	0.122
Tp5	1	5477	732	38	38	0	0	0	100	100	4	1	5079	0
	2	4933	602	38	38	0	0	0	100	100	5	1	4714	0
	3	4886	616	37	37	0	0	0	100	100	2	1	4591	0
	4	4636	527	38	38	0	0	0	100	100	2	1	4464	0
Tp7	1	5800	420	16	15	0	0	0	100	100	4	1	5551	0
	2	5275	323	16	15	0	0	0	100	100	4	1	5169	0
	3	5309	347	18	17	0	0	0	100	100	3	1	5115	0
	4	4983	287	17	16	0	0	0	100	100	3	1	4890	0
Tp8	1	4062	614	50	50	0	0	0	100	100	6	1	3809	0
	2	3633	512	51	51	0	0	0	100	100	5	1	3496	0
	3	3709	544	47	47	0	0	0	100	100	4	1	3514	0
	4	3458	464	47	47	0	0	0	100	100	3	1	3353	0

The number of reads at the different stages of the bioinformatics pipeline are shown together with the results before manual editing. 1st, 2nd and CLUS refer to stages of noise removal. Chimaera detection was performed using three different algorithms: S = Chimaera Slayer P = Perseus U = Uchime. Sensitivity was calculated as the probability of identifying minor allele variants present at 1:100 and 1:200. '>min no.reads' = the number of variants found after removal of variants with less than 5 reads. '>min%' = the number of variants found after removal of variants which are present at a frequency of less than 0.4% of the total good reads. 'Total good reads'= the number of reads maintained after filtering of reads in the bioinformatics pipeline Hz= heterozygosity

Table 3.10 Summary of the file generated by processing reads from SC06

Gene	Group	Initial no. of reads	No. of unique sequences after de-noising			No. of chimaeras detected			Sensitivity		Variants after chimaera removal & above thresholds		Total good reads	Hz
			1st	2nd	Pre Clus	S	P	U	100	200	>min no. reads	>min %		
Tp1	1	4395	1811	195	195	2	60	0	100	100	22	16	2519	0.89
	2	3799	1633	207	206	0	53	0	100	100	20	15	2236	0.89
	3	3240	1511	200	199	0	49	0	100	100	18	15	1888	0.88
	4	3058	1383	203	202	0	50	0	100	100	19	15	1846	0.88
Tp3	1	939	335	86	86	1	0	0	100	97.4	20	17	732	0.89
	2	842	278	88	88	1	0	0	99.9	96.9	23	17	687	0.89
	3	801	297	84	84	1	0	0	99.8	96.2	22	18	636	0.89
	4	746	254	84	84	1	0	0	99.8	95.7	21	17	621	0.89
Tp5	1	3167	743	133	129	0	0	0	100	100	30	20	2759	0.91
	2	2846	618	133	127	0	0	0	100	100	28	20	2587	0.91
	3	2715	624	129	124	0	0	0	100	100	25	20	2435	0.90
	4	2579	546	129	123	0	0	0	100	100	24	20	2382	0.90
Tp7	1	5639	879	87	76	0	0	0	100	100	25	13	5181	0.81
	2	5072	693	98	87	0	0	0	100	100	25	13	4811	0.81
	3	5044	735	93	83	0	0	0	100	100	25	14	4703	0.83
	4	4742	609	94	84	1	0	0	100	100	25	14	4527	0.83
Tp8	1	3682	565	70	68	0	0	0	100	100	7	5	3415	0.25
	2	3340	470	74	71	0	0	0	100	100	6	4	3179	0.26
	3	3367	514	77	75	0	0	0	100	100	8	5	3146	0.26
	4	3188	452	76	73	0	0	0	100	100	8	5	3038	0.31

The number of reads at the different stages of the bioinformatics pipeline are shown together with the results before manual editing. 1st, 2nd and LCUS refer to stages of noise removal. Chimaera detection was performed using three different algorithms: S = Chimaera Slayer P = Perseus U = Uchime. Sensitivity was calculated as the probability of identifying minor allele variants present at 1:100 and 1:200. '>min no.reads' = the number of variants found after removal of variants with less than 5 reads. '>min%' = the number of variants found after removal of variants which are present at a frequency of less than 0.4% of the total good reads. 'Total good reads'= the number of reads maintained after filtering of reads in the bioinformatics pipeline Hz= heterozygosity

The number of acceptable reads after the bioinformatics process differed per gene and sample (Table 3.11). For Tp1, the percentage of reads available after the bioinformatics processing was dramatically reduced for the buffalo sample, mainly due to the detection of chimaeric sequences. The percentage of reads available after bioinformatics processing was relatively high in both samples for Tp5, Tp7 and Tp8.

Table 3.11 Percentage of reads in which both MIDs and both primers could be identified

Gene	Total reads	Kiambu 5 acceptable reads	%	Total reads	SC06 acceptable reads	%
Tp1	890	729	82	4,395	1,846	42
Tp3	2,760	1,867	68	939	629	67
Tp5	5,477	4,464	82	3,167	2,391	75
Tp7	5,800	4,890	84	5,639	4,527	80
Tp8	4,062	3,353	83	3,682	3,038	83

3.3.2.3 Assessment of chimaera detection by sequencing a mixture of two cloned parasites

For Tp1, the number of chimaeras detected was high in the buffalo sample and only a single chimaera was detected for Tp3. In contrast, for Kiambu 5, no chimaeras were detected for Tp1 and several were identified for Tp3. The Tp1 chimaeras were detected using Perseus and the Tp3 chimaeras using Chimaera Slayer. The different programs use different algorithms for detecting chimaeras. For the genes used in this study, the optimal chimaera detection algorithm appears to differ among genes. The number of chimaeras detected also appears to be influenced by the number of nucleotide changes between reads and the total number of variants present in a sample. However, the precise genetic composition of the Kiambu 5 and SC06 samples used in this experiment is unknown. To increase the likelihood chimaeras will be detected, it was decided that all three chimaera detection algorithms would be applied to every dataset. For some of the genes no chimaeras were detected. This could be because chimaeras were not present or because chimaeras were not detected. These genes are relatively conserved and the differences between the two different sequences might be too small to be marked as chimaeras using the chimaera detection algorithms. The formation and detection of chimaeras was further tested for Tp1, Tp3 and Tp5 using an artificial mix consisting of equal DNA concentrations (as determined by Nanodrop) of two cloned parasites. The two clones were Marula N43-5 and Marula N33-4. To provide accurate baseline data for these two clones, the sequences of the antigen genes were determined by Sanger sequencing. The results of the Mothur pipeline are summarised in Table 3.12.

Perseus detects four chimaeras in the Tp1 gene, removing true variant as well as true chimaeras. This means there is a risk that true variants could also be removed from other samples. No chimaeras were detected in the reads obtained for Tp5. The two most frequent reads correspond to the sequences identified using conventional Sanger sequencing. The sequences of the two clones differed in 6 nucleotides (see figure 3.9). The other four sequences appear to be chimaeras of these two reads. This demonstrates that none of the three algorithms is capable of detecting the chimaeras in this dataset. No chimaeras were detected in the reads for Tp3 either. Interestingly, chimaeras were detected in both the reads obtained for SC06 (see Table 3.10) and Kiambu 5 (See Table 3.9). This clearly indicates that the efficiency of the algorithms depends on the dataset being analysed. Chimaera detection is presently sub-optimal, despite several different algorithms being available and this issue needs to be taken into account when interpreting results

Table 3.12 Summary of Mothur pipeline of sequencing of an equal mixture of two clones

Gene	Initial no. of reads	No. of unique sequences after de-noising			No. of chimaeras detected			Variants after chimaera removal & above thresholds		Total good reads	Read counts
		1st	2nd	Clus	S	P	U	>min no. reads	>min %		
Tp1	1520	285	13	13	0	4	0	1	1	1272	1272
Tp3	780	211	21	19	0	0	0	5	5	319	176, 91, 23, 14, 8
Tp5	1145	190	26	25	0	0	0	9	7	1007	382, 334, 152, 119, 8, 6, 6

The number of reads at the different stages of the bioinformatics pipeline are shown together with the results before manual editing. 1st, 2nd and CLUS refer to stages of noise removal. Chimaera detection was performed using three different algorithms: S = Chimaera Slayer P = Perseus U = Uchime. '>min no.reads' = the number of variants found after removal of variants with less than 5 reads. '>min%' = the number of variants found after removal of variants which are present at a frequency of less than 0.4% of the total good reads. 'Total good reads'= the number of reads maintained after filtering of reads in the bioinformatics pipeline 'Read counts'= the number of reads for each of the alleles detected.

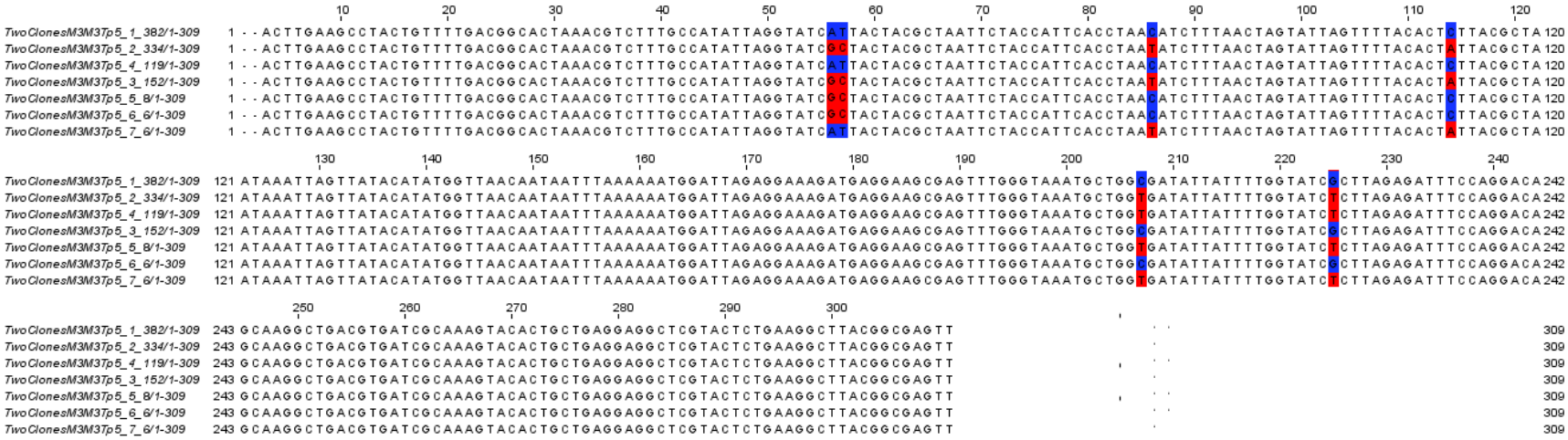

Figure 3.9 Alignment of nucleotide sequences of Tp5 amplicons generated from a mixture of two clones
Nucleotides that differed between different identified alleles are highlighted.

3.4 Discussion

A multi-locus sequence typing system was developed for the analysis of genetic diversity among *Theileria parva* antigen-encoding genes using Roche 454 amplicon sequencing technology. A total of nine genes known to be recognised by $CD8^+$ T cell responses in immune animals were selected for the typing system. During development of the system, PCR amplicons were generated from a vaccine stabilate and blood from a carrier African buffalo. These were then successfully subjected to 454 sequencing and the derived sequences were analysed using a custom bioinformatics pipeline.

3.4.1 Generation of amplicons

Extensive primer validation was necessary during the development of the system and each candidate primer set was tested on a panel of *T. parva* clones. This was essential in order to ensure that the PCR primers were able to amplify the majority of, if not all, allelic variants of *T. parva*. The clones were derived from parasitised cell lines isolated from buffalo or from cattle undergoing clinical theileriosis after infection with buffalo-derived *T. parva.* The test panel used in this study was an ideal resource for the validation of primers, as buffalo-derived *T. parva* is known to be more genotypically diverse than the cattle-derived parasite. For the majority of genes, several primer combinations required testing before a primer combination could be found that would amplify every clone in the test panel. For the primer validation process, the DNA concentration was standardised to 50 ng/µl. However, the amount of amplicon yielded varied among clones, as judged by the intensity of bands on the agarose gel. While this suggests that the efficiency of PCR may not equivalent for different template sequences, it is unknown whether the host/parasite DNA ratio was equivalent for all clones and whether better amplification was associated with a higher parasite load. Nevertheless, it is appreciated that primer specificity and sensitivity may potentially be an issue and that 'imperfect' template sequences may be associated with lower yield. The PCR protocol should therefore be regarded as semi-quantitative, with caveats applied when drawing inferences on the relative abundance of allelic variants.

Amplification from sporozoite stabilates was poor compared to amplification of DNA derived from parasitised cell lines. This may be improved by removal of particulate tick material using an additional centrifugation step before DNA extraction. It was desirable to use DNA extracted directly from sporozoite stabilates, as maintaining parasites *in vitro* is likely to exert a selective pressure on the parasite population under study (Sutherland et al., 1996; Taylor et al., 2005).

Polymerase with proofreading capacity was used in order to minimise the number of errors introduced during PCR amplification and the number of PCR cycles was kept as low as reasonably possible. Unfortunately, amplification using fusion primers together with proofreading polymerase was inefficient and for most samples only small quantities of template DNA were available. The number of cycles required to obtain sufficient PCR products for all the genes was relatively high and varied between 30 and 35 cycles. Chimaeras were detected for Tp3 in both Kiambu 5 and SC06 and for Tp1 in SC06. The number of sequences available for further analysis was greatly reduced for Tp1 from SC06, with only 42% of raw reads passing quality control. Ideally, PCR conditions should be optimised to reduce the number of chimaeras when sufficient DNA is available for optimisation.

3.4.2 Bioinformatics pipeline

A bioinformatics pipeline was established for the processing of the sequences obtained and to identify and account for PCR and sequencing errors. Reads were assigned to a sample following identification of both MID tags and both primers. The next step in the pipeline was to remove 'noise' from the dataset, using algorithms to either correct errors *in situ* or to discard potentially erroneous reads. Sequencing of the control clone demonstrated that the bioinformatics pipeline could not account for all the PCR and/or sequencing errors introduced during sample processing and it was shown that the type of error differed among target genes. This agrees with other studies, which have documented differences in the type of error and error rate when sequencing different gene products (Gilles et al., 2011).

It is clear that 454 amplicon sequencing gives rise to non-random sequencing errors (Niklas et al., 2013; Schloss et al., 2011; Skums et al., 2012). Homopolymer errors

may be found throughout reads, whereas single nucleotide insertions and nucleotide substitutions are more common towards the end of long reads (Gilles et al., 2011). Errors can vary between sequencing centres and even between sequencing runs (Schloss et al., 2011). Gilles *et al.* estimated that a minimum of five sequences are required to be able to effectively correct errors (Gilles et al., 2011).

Homopolymer errors are relatively easy to detect and manually correct, but manual correction is impossible for single nucleotide insertions and substitutions. Based on Gilles estimation that a minimum of five sequences are required for error correction (Gilles et al., 2011) and the error rate observed in sequencing the control clone, a minimum threshold of five reads and 0.4% abundance before a sequence was deemed genuine. It is possible that some genuine variants are rejected using this threshold, thereby underestimating the true number of variants present in a sample. It is advisable to include an internal control in each sequencing run in order to estimate the sequence error rate and the rate of chimaerism in that run (Schloss et al., 2011; Skums et al., 2012) and, ideally, the control should have a level of complexity similar to the other samples in the run (Skums et al., 2012).

3.4.3 Chimaera detection

Three different algorithms were used for chimaera detection in the bioinformatics pipeline to maximise the chance of detecting chimaeras - Chimaera Slayer, Perseus and Uchime (Edgar et al., 2011; Haas et al., 2011; Quince et al., 2011). For the samples analysed in this study, it was found that Perseus and Chimaera Slayer detected chimaeras in the same datasets, while chimaeras were not detected by Uchime. As these algorithms were used successively, the results do not necessarily indicate that Uchime was ineffective and in previous studies, Perseus and Uchime were shown to have a similar efficiency (Edgar et al., 2011; Schloss et al., 2011). Perseus identified a genuine variant as chimaeric in sequences obtained for Tp1 in a mixture of two clones. None of the chimaeric sequences were detected in a mixture of two clones for Tp3 and Tp5. However, chimaeras were detected in Tp3 amplicons using Chimaera Slayer for both Kiambu 5 and SC06. Perseus also detected a chimaera in Tp3 for Kiambu 5. These database-independent methods of chimaera

detection use sequences within the dataset as a reference and thus the reference sequences vary between different samples. It is important to be aware that some genuine sequences may be identified as chimaeric while some chimaeric sequences may escape detection despite the use of three different algorithms.

Chimaera formation is commonly encountered in the field of population genetics and has been increasingly observed when using high throughput sequencing (Haas et al., 2011; Quince et al., 2009; Schloss et al., 2011). It is estimated that individual sequence libraries for environmental bacterial diversity studies can contain over 45% chimaeric sequences (Haas et al., 2011) and such sequences can lead to overestimation of species diversity in habitats when targeting the16S ribosomal RNA gene, although this has been partially overcome with the use of 16S rRNA reference databases. Unfortunately, databases are not available for the allelic variants of the genes used in this study.

Theileria parva undergoes sexual recombination in the tick gut and it is possible that recombination break-points may lie within the loci of interest, with such break-points relating to ancient and/or recent recombination events. Katzer *et al* (2011) performed a genetic cross of two cloned isolates of *T. parva* (Muguga and Marikebuni) and identified 35 recombinant progeny using a panel of 79 satellite DNA markers. The rate of recombination differed across the chromosomes (between 0.12 and 0.34 cm/Kb) and hot and cold spots were identified (Katzer et al., 2011). A high number of cross-over events were observed in two recombinant *T. parva* clones using comparative genomics. The distribution of cross-over events (CO) differed among the four chromosomes and between the two recombinant clones, varying from one cross-over per 0.3 Mb to one cross-over per 1.3 Mb (Henson et al., 2012). Thus, it is possible that some genuine recombinant allelic sequences may be mis-indentified as 'chimaeric', particularly if the putative 'parental' alleles can be identified in the dataset. The effect of such mis-identification is that the multiplicity of infection would be underestimated. However, for the present study, it was decided to implement each of the chimaera detection methods in order to minimise the chance of designating artefactual sequences as real.

3.4.4 Adaptation of amplification and bioinformatics pipeline for Tp2

The amplicons for Tp2 were relatively large and only a small proportion of the reads were sufficiently long to span both MID tags. This necessitated Tp2 being amplified using the same MID tag on both the forward and reverse primer in order to assign reads to a specific sample and this allowed sequences obtained in the forward and reverse direction to be analysed separately. Errors increase toward the ends of long reads and these cannot always be corrected accurately by a bioinformatics pipeline. Schloss *et al*. truncated reads at 720 flows, which corresponds to around 400 bp to remove this noise (Schloss et al., 2011). In the case of Tp2, sequence reads were truncated around 350 flows.

In summary, a MLST typing system was developed for the analysis of the antigenic diversity of *Theileria parva* using Roche 454 amplicon sequencing technology. Amplicons were successfully generated and sequenced from vaccine stabilate and carrier animals. Noise was removed from sequencing reads using a bioinformatic pipeline constructed around the Mothur software package. The pipeline was able, for the most part, to adequately identify and remove 'noise' from the dataset, but further manual editing was required in some cases and chimaeric sequences were not always detected. A threshold of a minimum of five reads and least 0.4% abundance was determined to differentiate between genuine allelic variants and erroneous reads.

Chapter 4: Analysis of the Muguga cocktail vaccine

4.1 Introduction

The Muguga cocktail is the most widely used parasite preparation for immunisation against *Theileria parva* by the 'infection and treatment' method (ITM) and incorporates three parasite stocks: Muguga, Serengeti-transformed and Kiambu 5. Since the initial testing of the components of the Muguga cocktail in the 1970s, each has been passaged a number of times. A consistent protocol for production of different passages of the vaccine components has been strictly followed in an attempt to maintain the consistency of the vaccine. Stabilates used for vaccination are derived from seed stabilates that are in turn derived from master seed stabilates. In other words, a master seed parasite population will go on to complete the life-cycle twice before it forms an actual vaccine stabilate. Reference stabilates are produced alongside the production of the immunising stabilates. Although several molecular typing methods have been used to characterise *T. parva* parasites in the Muguga cocktail (Bishop et al., 2001; Geysen et al., 1999; Oura et al., 2007a; Oura et al., 2004b; Patel et al., 2011), these have provided limited information on the detailed composition of the individual stocks, in particular the presence of minor genotypes. Little detail is available on the level of consistency between different passages of the three components and among different batches of the Muguga cocktail produced from the same passage.

Studies of another *T. parva* stock (Marikebuni), known to be genetically heterogeneous (Katzer et al., 2010), illustrate the necessity to monitor parasite composition. Analyses using a large panel of satellite DNA markers revealed major differences in profile among different passages of this stock. Patel *et al.* used a subset of five satellite DNA markers to compare two batches of the Muguga cocktail vaccine (FAO1 and ILRI0804) and the corresponding reference stabilates. All the reference stabilates were produced using the same seed stabilates (Stabilate 68, 69 and 73) (see Figure 4.1). Differences in allele sizes were found for four out of five satellite loci between the two Serengeti transformed reference stabilates (stabilate

4229 and 4139) produced from the same seed stabilate (stabilate 69) (Patel et al., 2011).

The Global Alliance for Veterinary Medicine (Galvmed) is aiming to facilitate the more widespread use of the Muguga cocktail vaccine, which includes the production of new vaccine batches at the Centre for Ticks and Tick-Borne Diseases (CTTBD) in Malawi. For the production of future vaccine batches of the Muguga cocktail vaccine, the current reference stabilates (Stabilate 4228, Stabilate 4229 and 4230) will be used as the new seed stabilates (see figure 4.1). Hence, future vaccine batches will be derived from a different passage compared to the two most recent large vaccine production rounds, FAO in 1993 and ILRI08 in 2008. This could potentially alter the composition of the parasites in the vaccine and possibly influence the immunity induced.

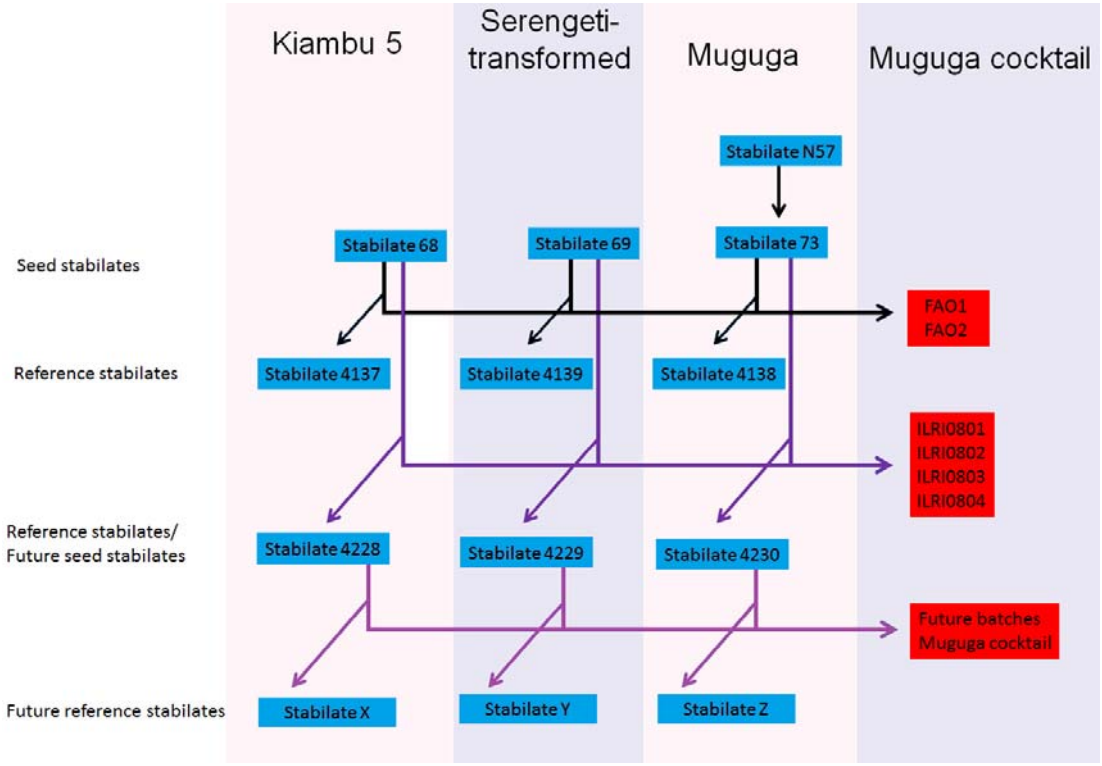

Figure 4.1 Relationship between the different Muguga cocktail seed, reference and vaccine stabilates.
Each arrow represents a single passage through cattle/ticks. Muguga cocktail batches are displayed in the red boxes and were formulated by mixing ticks infected with the Kiambu 5, Serengeti-transformed and Muguga stock after a passage through cattle/ticks

While satellite DNA typing provides a profile of the major components within a parasite stock, it does not allow detection of minor components. The multilocus sequence typing system based on 454 technology developed in this study can be used for high resolution characterisation of parasite populations and thus we now have the capacity to identify minor components. The main aim of the work described in this chapter was to use multilocus sequence typing in combination with satellite DNA typing to determine the extent of parasite diversity in components of the Muguga cocktail, focusing on the masterseed stabilates that will be used for future vaccine production. The profiles generated for the seed stocks may be used for comparison with progeny vaccine stabilates, other vaccine stabilates and field parasite populations.

4.2 Material and Methods

4.2.1 Samples

DNA was extracted from stabilates used for infection and treatment as previously described in Chapter 2. Details of the stabilates can be found in Table 4.1.

Table 4.1 Stabilates used for DNA extraction

Stabilate name	Stabilate number	Parent Stabilate	Source
Muguga	4230	73	ILRI
Muguga	N57	147	CTTBD
Serengeti transformed	4229	69	ILRI
Kiambu 5	4228	68	ILRI
ECF MC ILRI	0801	68,69,73	ILRI
ECF MC ILRI	0802	68,69,73	ILRI
ECF MC ILRI	0803	68,69,73	ILRI
ECF MC ILRI	0804	68,69,73	ILRI
Marikebuni	3014	2 cattle-tick passages	RVC (made at ILRAD)
Marikebuni	128	3014	CTVM, UoE
Katete	KL4		CTTBD
Chitongo	CL20		CTTBD

ILRI= International Livestock Research Institute, Nairobi, Kenya; CTTBD= Centre for ticks and tick borne diseases, Lilongwe, Malawi; RVC= Royal Veterinary college, University of London; ILRAD= Internationl laboratory for research on animal diseases, Nairobi, Kenya; CTVM, UoE= Centre of tropical veterinary medicine, University of Edinburgh

4.2.2 Satellite DNA typing

Satellite DNA typing was performed as described in Chapter 2 on all the stabilates listed in Table 4.1 except for stabilates 57 and 3014, for which a limited amount of DNA was available. Satellite markers were distributed across the genome (See figure 2.1). Allele sizes were determined by capillary flow electrophoresis and detection of the fluorophore signal from a labelled primer in each amplicon. The relative abundance of individual peaks was expressed as percentage of the total area under the curve for all peaks in a trace combined. Peaks with an area of less than 3% of the total area under the curve for all peaks were excluded.

4.2.3 454 amplicon sequencing

Amplicons were generated for the genes Tp1 through to Tp8 and Tp10 for the Muguga (stabilate 4230), Serengeti transformed (stabilate 4229), Kiambu 5 (stabilate 4228) and Marikebuni (stabilate 3014) stocks. The amplicons were generated and pooled as described in Chapter 3 and submitted for sequencing. In addition, amplicons were generated for Marikebuni stabilate 128 for the genes Tp1, Tp5 and Tp10. Reads derived from different sequencing runs were combined when sequencing data were available from more than one sequencing run. Sequencing results were analysed as previously described in Chapter 3. Resulting sequences were aligned using MUSCLE and were visualised using Jalview (Edgar, 2004; Waterhouse et al., 2009).

4.3 Results

4.3.1 Satellite genotyping

4.3.1.1 Comparison of the Muguga cocktail components and batches

Genotypic diversity within batches of the Muguga cocktail and the three components of the Muguga cocktail was assessed using a panel of satellite DNA markers. Product size and relative abundance were determined using capillary electrophoresis (see figure 4.2 for an example of histograms). Products were successfully obtained for eleven of the satellite DNA markers (See figure 4.3). The data obtained was used to calculate a heterozygosity value for each stock (See table 4.2).

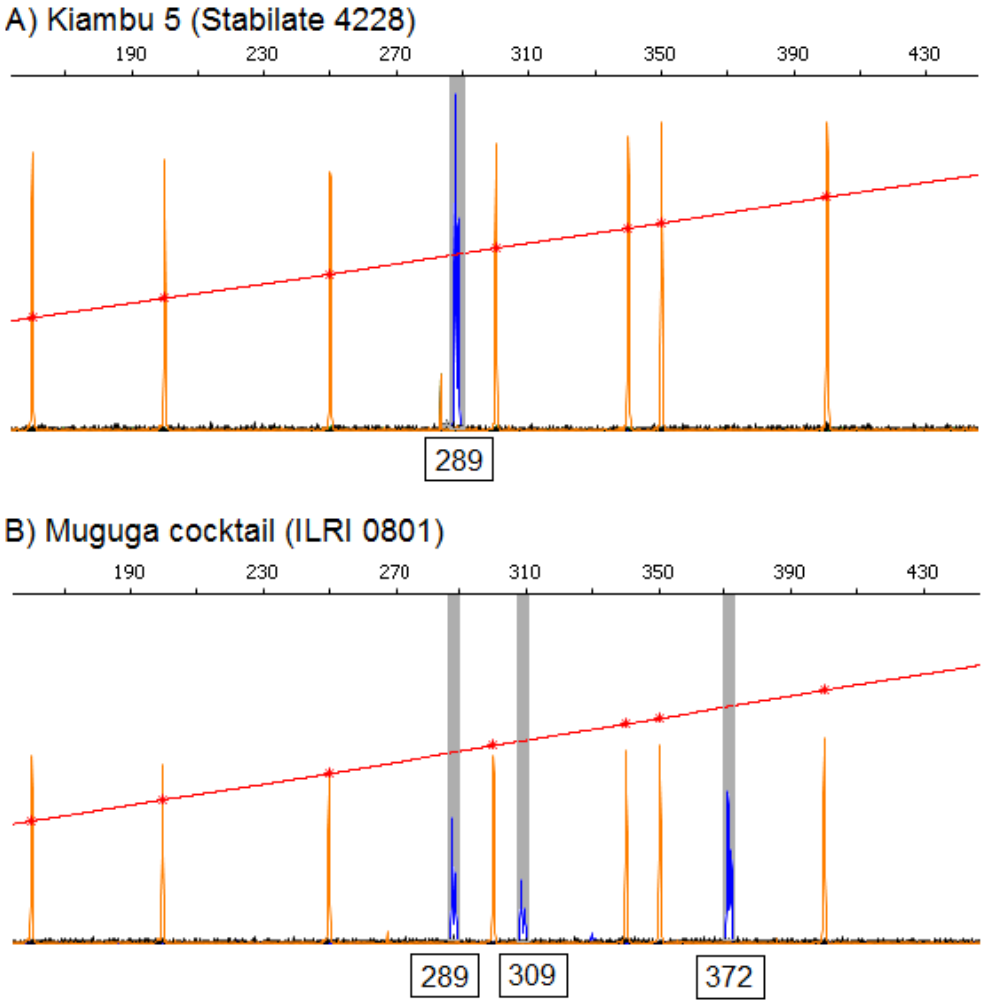

Figure 4.2 Example of representative histograms obtained for MS7 for samples Kiambu 5 and Muguga cocktail. Blue peaks represent sample peaks, the estimated allele size are indicated below the peaks. The allele size was estimated using a standard curve (red line), which was obtained using the peaks obtained for the LIZ500 size standard (peaks in orange)

For each of the satellite markers, between one and four alleles were found in individual samples representing the components of the Muguga cocktail vaccine batches and the reference stabilates. The four batches of the Muguga cocktail had a very similar satellite profile, both in terms of which alleles were found and the relative abundance of each of these alleles. Small differences in relative abundance of alleles between different batches of parasites produced from the same stabilates using similar batches of ticks were observed. This could be a result of stochastic variation among straws and may be related to clumping of sporozoites within the sample. As PCR is not a fully quantitative process, it is possible that minor differences in template composition may be magnified during amplification. In general terms, however, this genotyping methodology is capable of providing highly reproducible results. In a related experiment, different straws representing the same Serengeti –transformed reference stabilate (stabilate 4229) were genotyped by another worker at another institution using the same satellite reagents (University of Glasgow). The results were very similar to those generated in the present study, indicating that the methodology is robust and the results are repeatable (see figure 4.4).

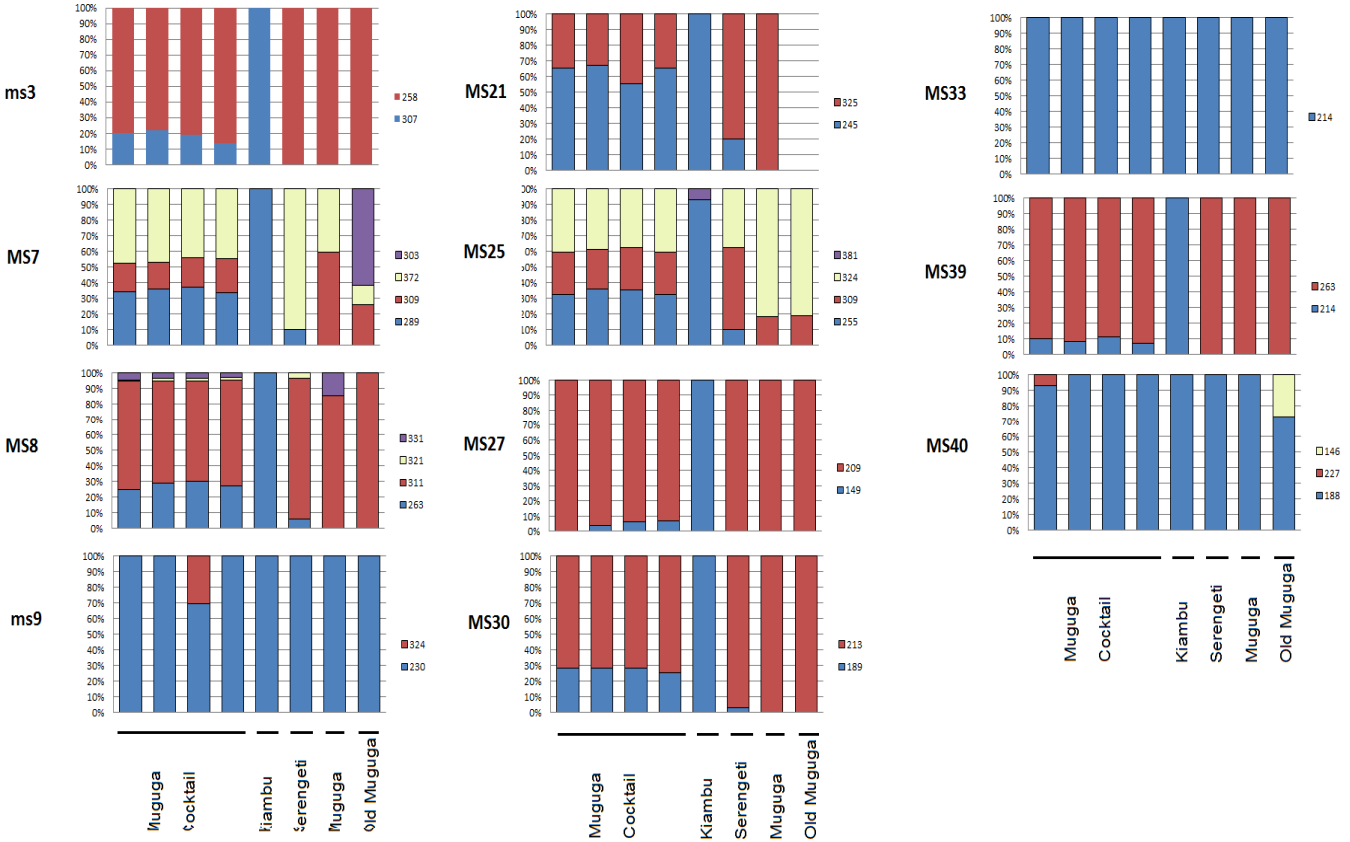

Figure 4.3 Results of satellite DNA typing of four Muguga cocktail vaccine batches and reference stabilates using 11 DNA Satellite markers
Muguga cocktail vaccine batches (ILRI0801-ILRI0804) and reference stabilates Kiambu 5 (Stabilate 4228), Serengeti-transformed (Stabilate 4229) and Muguga (Stabilate 4230) were produced in the same vaccine production run using stabilates 68, 69 and 72 as seed stabilates. In the last column the results are displayed for an old Muguga cocktail stabilate (Stabilate N57) (the result for MS21for stabilate N57 is missing). Different allele sizes are displayed in a different colour and the relative abundance of each of the alleles is displayed in the stacked bars, estimated allele sizes for each of the markers are displayed next to the diagrams. For more information on the stabilates see figure 4.1.

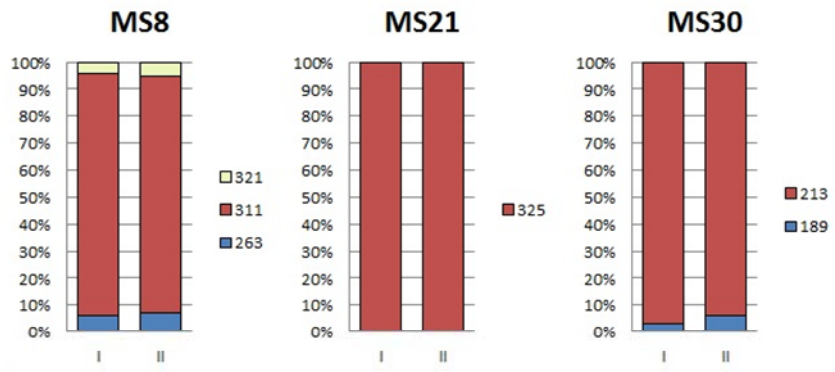

Figure 4.4 Comparison of repeats for three markers for the Serengeti transformed stock. Allele sizes and relative abundance for the Serengeti-transformed stock using three satellite DNA markers (MS8, MS21, MS30) obtained in this study (I) (see figure 4.3) were compared with allele sizes and relative abundance obtained at the University of Glasgow (II).

A few inconsistencies were observed in the typing results obtained in the present study for different batches of the Muguga cocktail. For MS7 applied to ILRI0804, an extra peak corresponding to 214 bp was identified which was not present in the other batches of the Muguga cocktail (ILRI0801-ILRI0803). For MS9 applied to ILRI0803, an extra peak corresponding to 245 bp was found and for MS40 applied to ILRI0801, an extra peak of 227 bp was found. These extra peaks are likely to represent contamination, as the sizes corresponded to allelic sizes of other markers that were processed in the same 96-well plate used for Genescan. For ILRI0801, a minor peak at 149 bp was missing for MS27 compared to the other ILRI08 batches, which may reflect either a failure to detect a low concentration of the PCR product, or a failure for this locus to prime due to low template concentration. The assays for these markers need to be repeated in order to resolve these anomalies. Nevertheless, overall the composition of the four vaccine batches produced in the same vaccine production run (i.e. ILRI0801-ILRI0804) appeared to be consistent. The satellite DNA profiles were similar in allele size and abundance and the allele sizes identified corresponded with those found in the three reference stabilates. However, for marker MS8, by applying an area threshold of 3% for a peak to be considered a genuine allele, allele d would not be identified/accepted in the Serengeti transformed stock, as this allele was only present at 1.7%.

Table 4.2 Heterozygosity in the Muguga cocktail vaccine and reference stabilates

Sample	Average Hz
ILRI0801	0.292
ILRI0802	0.288
ILRI0803	0.338
ILRI0804	0.304
4228	0.012
4229	0.120
4230	0.094
N57	0.129

The heterozygosity of the different batches of the Muguga cocktail and the reference stabilates was calculated for each of the loci. The average heterozygosity is displayed.

Based on satellite typing, the diversity within the reference stabilates and the Muguga cocktail appeared to be limited. The higher heterozygosity observed for ILRI0803 compared to the other vaccine batches is due to the higher heterozygosity for the ms9 locus, which is likely to be caused by a contaminant. The heterozygosity is conserved between the other Muguga cocktail vaccine batches. The heterozygosity for each of the components is low (Hz between 0.012 and 0.120)

Two alleles were detected using MS25 for Kiambu 5; In contrast, only a single allele was found using each of the remaining ten successfully amplifying markers for Kiambu 5. Four of these alleles were unique to the Kiambu 5 stocks. At another four loci, the Kiambu 5 allele was also found as a minor component of the Serengeti-transformed stock. For the remaining two markers, the Kiambu 5 allele was also present in all the components of the Muguga cocktail. A single allele was also found in the Muguga stock for eight loci and two alleles for the remaining three loci. For the Serengeti-transformed stock, a single allele was detected at seven of the eleven loci, two alleles at two loci and three alleles at the remaining two loci.

The Muguga (St 4230) and Serengeti (St 4229) stocks differed from each other at four loci, although one or two relatively abundant alleles were shared at these loci. The Kiambu 5 stock differed from the Muguga stock at eight of the ten loci and from the Serengeti transformed stock at seven of the ten loci.

DNA from Muguga stabilate N57 was available for this study and was genotyped using the satellite markers. Historically, this stabilate was used to produce Muguga stabilate 73, from which the Muguga reference stabilate 4230 was produced (see figure 4.1). Differences between stabilate N57 and 4230 were found across three of the eleven markers. For two of the markers, MS7 and MS40, alleles present at 6-58% in stabilate N57 appear to be lost in stabilate 4230. However, for MS8, stabilate 4230 had an extra allele, at a relative abundance of 15%, compared to stabilate N57. The composition of parasite stocks can thus change over passage.

4.3.1.2 Comparison of the Muguga cocktail with other vaccine stabilates

The same eleven satellite markers were used to investigate the degree of genotypic diversity in other stabilates used for vaccination in the field (see figure 4.5) and the diversity of the stabilates was estimated by calculating the average heterozygosity over all the successfully amplified markers (see Table 4.3). The Marikebuni stock was originally isolated from a collection of ticks from the coastal region in Kenya and has been used in vaccine trials in Kenya. The Katete and Chitongo stocks were isolated in Zambia and have been used for vaccination in Malawi and Zambia (Geysen et al., 1999).

Table 4.3 Average heterozygosity in vaccine stabilates

Sample	Average Hz
ILRI0802	0.288
Katete	0.000
Chitongo	0.032
Marikebuni	0.132

The heterozygosity was highest for the Muguga cocktail (HZ 0.288). The heterozygosity was very low for both Zambian stocks (Katete and Chitongo) (Hz 0,00 and 0,032), while the Marikebuni stock had an intermediate level of heterozygosity (Hz 0.132)

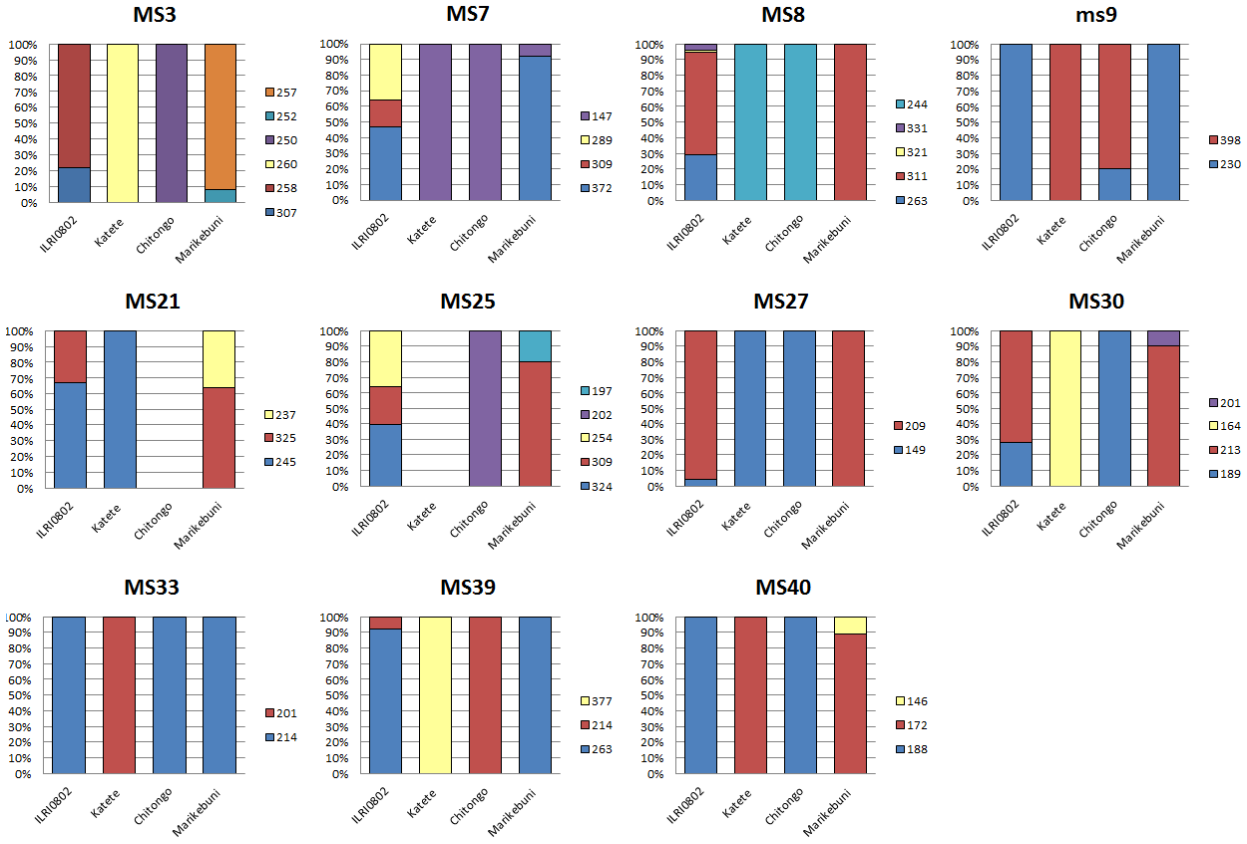

Figure 4.5 Results of DNA satellite typing of four different ITM stabilates using 11 markers
The four stabilates are the Muguga cocktail (ILRI0802), the Katete stock (stabilate KL4), the Chitongo stock (CL20) and the Marikebuni stock (stabilate 128). Data is missing for MS21 for the Chitongo stabilate and MS25 for the Katete stabilate due to shortage of material available for analysis. Different allele sizes are displayed in a different colour and the relative abundance of each of the alleles is displayed in the stacked bars, estimated allele sizes for each of the markers are displayed next to the diagrams.

Very limited diversity was detected within the Zambian stocks (Katete and Chitongo), but differences were found between the two stocks. Identical alleles were detected for three satellite DNA markers (MS7, MS8 and MS27); the Chitongo stock had an additional allele for ms9, besides a shared allele; different alleles were detected for the other five successfully obtained amplified markers (MS3, MS30, MS33, MS39 and MS40). The Katete stock and the Muguga cocktail shared alleles at two of the ten loci (MS21 and MS27), whereas the Chitongo stock and the Muguga cocktail had alleles in common at six of the ten loci (ms9, MS27, MS30, MS33, MS39 and MS40)

The Marikebuni stock was found to be more heterogenous than the Zambian stock, but less so than the Muguga cocktail; two alleles were detected at six of the eleven loci. With the exception of MS3, alleles were shared between the Marikebuni stock and the Muguga cocktail. For five of the loci, a second non-shared allele was found in the Marikebuni stock (MS7, MS21, MS25, MS30, MS40).

4.3.2 'Deep' multilocus sequence typing

4.3.2.1 Muguga cocktail reference stabilates

The satellite genotyping results indicated that genotypic diversity within the Muguga cocktail vaccine batches and reference stabilates is limited. To examine diversity in greater depth, the three stabilates that will be used as seed stabilates for future vaccine batches were subjected to multilocus sequence typing using deep sequencing technology. This offers the facility to generate high resolution allelic profiles which can be used as a reference to compare against future seed stabilates. The aim was to obtain a minimum of 1,000 reads for each of the genes examined, to facilitate detection of allelic variants present at a frequency of 0.5% with a confidence of >99%. A summary of the number of reads obtained for each of the samples/genes and the number of sequences processed at the different points in the bioinformatics pipeline can be found in Table 4.4.

The alleles identified in different components of the Muguga cocktail were compared by aligning resulting DNA sequences using MUSCLE (Edgar, 2004) and a number was assigned to each allele identified. Alignments of the different alleles can be found in Appendix B. The frequencies of the different alleles in the different reference stabilates of the Muguga cocktail are illustrated in Figure 4.6.

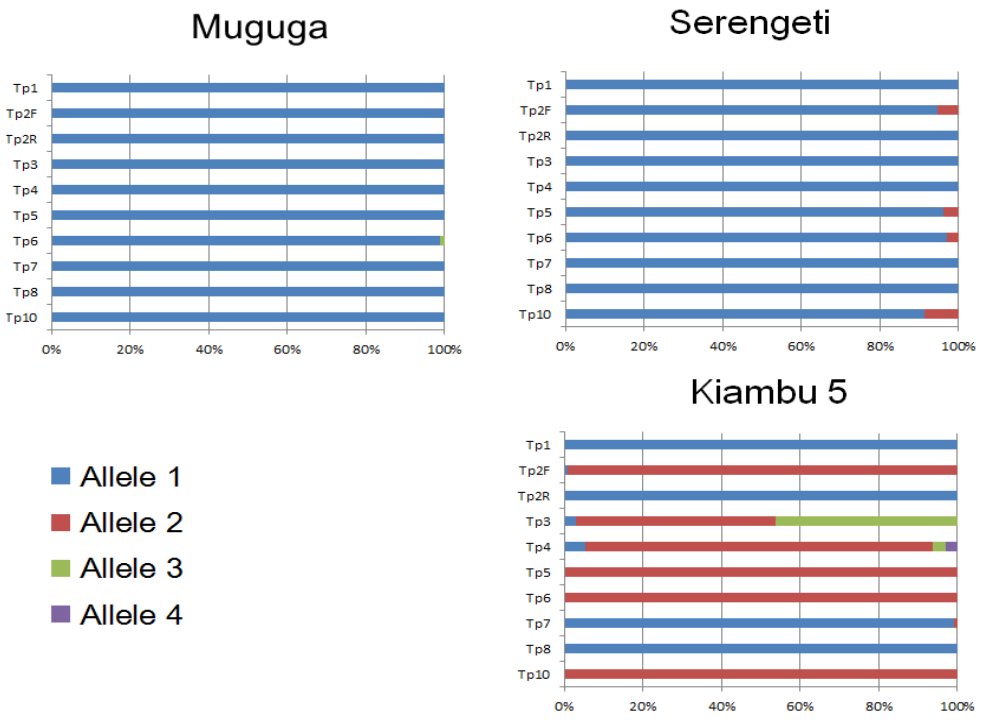

Figure 4.6 Allelic composition of the three components of the Muguga cocktail based on nucleotide sequence of nine *Theileria parva* antigens.

Despite the high sequencing depth achieved, the number of allelic variants identified was low. In the Muguga stock only a single allele was detected at each locus, except for a minor second allele in Tp6 with an abundance of 0.77%. Across the loci, the same alleles found in Muguga dominated the Serengeti-transformed stock, but a second allele was detected for four genes at a frequency of less than 10%. For the Kiambu 5 stock, three genes had a single allele identical to the allele found in the Muguga and Serengeti-transformed stocks. Four genes had a single allele identical to the minor alleles found in the Serengeti-transformed stock. For the remaining genes, between two and four alleles were found. One allele of Tp7, two alleles of Tp3 and three alleles of Tp4 were not represented in either the Muguga or Serengeti-

transformed stocks. In summary, multilocus sequence typing identified between one and four alleles across the nine genes examined within the combined components of the Muguga cocktail, revealing a similar level of diversity to that found using satellite genotyping.

Table 4.4 Summary of the bioinformatic processing of the sequencing reads obtained from the Muguga cocktail

Sample	Gene	Reads from sequencing run	Initial no. of reads	No. of unique sequences after de-noising			No of chimeras detected			Sensitivity		Variants after chimaera removal & above thresholds		Total good reads	Hz	No. of reads obtained for each allele			
				1st	2nd	Clus	S	P	U	100	200	>min no. reads	>min %			1	2	3	4
Kiambu St4228	Tp1	MC_Test_AB	2195	361	22	22	0	0	0	100	99.9	2	1	2039	0	2039			
	Tp2F	MC	906	377	48	48	0	0	0	99.9	97.3	2	2	721	0.01	716	5		
	Tp2R	MC	941	573	153	152	0	0	0	99.9	97.5	3	3	731	0.11	688	36	7	
	Tp3	MC_Test	4976	959	120	120	8	1	0	100	100	6	4	4309	0.53	2179	1979	118	33
	Tp4	MC_AB	1282	213	20	20	1	0	1	100	99.7	4	4	1172	0.21	1039	60	39	34
	Tp5	MC_Test_AB	13310	952	50	50	0	0	0	100	100	2	1	12934	0	12934			
	Tp6	MC	1118	123	12	12	0	0	0	100	99.5	1	1	1055	0	1055			
	Tp7	MC_Test	7575	655	33	32	0	0	0	100	100	3	2	7361	0.02	7304	57		
	Tp8	MC_Test	6453	919	87	87	0	0	0	100	100	1	1	6146	0	6146			
	Tp10	MC_AB	5146	411	13	13	0	0	0	100	100	1	1	5036	0	5036			
Serengeti St4229	Tp1	MC	2182	447	38	38	0	0	0	100	100	1	1	1942	0	1942			
	Tp2F	MC	559	256	56	56	0	0	0	98.3	86.6	3	3	403	0.14	372	21	10	
	Tp2R	MC	817	477	123	123	0	0	0	99.9	95.9	1	1	637	0	637			
	Tp3	MC	1794	277	34	34	0	0	0	100	100	3	2	1626	0.01	1617	9		
	Tp4	MC	1697	370	47	47	0	0	0	100	100	1	1	1493	0	1493			
	Tp5	MC	1954	331	34	33	0	0	0	100	100	2	2	1817	0.08	1746	71		
	Tp6	MC	1259	142	14	14	0	0	0	100	99.7	2	2	1158	0.06	1123	35		
	Tp7	MC	2432	307	25	24	0	0	0	100	100	1	1	2310	0	2310			
	Tp8	MC	3205	388	37	37	0	0	0	100	100	2	2	3012	0.01	2997	15		
	Tp10	MC	256	64	8	8	0	0	0	90.5	68.9	2	2	233	0.16	213	20		
Muguga St4230	Tp1	MC	2394	419	25	25	0	0	0	100	100	1	1	2175	0	2175			
	Tp2F	MC	586	199	41	40	0	0	0	99.1	90.7	2	2	472	0.07	455	17		
	Tp2R	MC	843	393	92	92	0	0	0	99.9	96.9	2	2	693	0.03	682	11		
	Tp3	MC	1394	230	25	25	0	0	0	100	99.8	2	2	1261	0.01	1252	9		
	Tp4	MC	1863	378	26	26	0	0	0	100	100	1	1	1649	0	1649			
	Tp5	MC	1577	264	33	33	0	0	0	100	99.9	1	1	1463	0	1463			
	Tp6	MC	2200	226	9	9	0	0	0	100	100	2	2	2050	0.02	2025	25		
	Tp7	MC_AB	3827	172	11	10	0	0	0	100	100	3	1	3715	0	3715			
	Tp8	MC	2345	330	46	46	1	0	0	100	100	2	2	2206	0.01	2194	12		
	Tp10	MC_AB	7299	523	42	42	0	0	0	100	100	1	1	7091	0	7091			

The number of reads at the different stages of the bioinformatics pipeline are shown together with the results before manual editing. Reads of different sequencing runs were combined when available (MC=Run with samples of the Muguga cocktail, Test=Testrun (see chapter 3), AB=Run with samples from African Buffalo (Chapter 5). 1st, 2nd and CLUS refer to the stages of noise removal. Chimaera detection was performed using three different algorithms: S = Chimaera Slayer P = Perseus U = Uchime. Sensitivity was calculated as the percentage chance of identifying a minor allele present at 1:100 and 1:200. '>min no. reads' = the number of variants found after removal of variants with less than 5 reads; '>min%' = the number of variants found after removal of variants which are present at a frequency of less than 0.4% of the total good reads; 'Total good reads' = the number of reads maintained after filtering of reads in the bioinformatics pipeline; Hz= heterozygosity; The number of reads obtained for each of the alleles detected are displayed.

4.3.2.2 Marikebuni stabilates

A high resolution multilocus sequencing profile was obtained for another *T. parva* stock, Marikebuni stabilate 3014, which has also been used successfully as a single component ITM vaccine, mainly in the coastal region of Kenya. The same genes analysed in the Muguga cocktail components were examined in Marikebuni 3014. Additionally, sequencing was performed on Marikebuni stabilate 128 for four of these genes, Tp1, Tp2, Tp5 and Tp10. Marikebuni 128 was produced at the University of Edinburgh by a single tick passage from an animal infected with stabilate 3014. A summary of the number of reads obtained for each of the samples/genes and the number of sequences processed at the different points in the bioinformatics pipeline can be found in Table 4.5.

Table 4.5 Summary of the bioinformatic processing of the sequencing reads obtained from Marikebuni stabilates 3014 and 128

Sample	Gene	Initial no. of reads	No. of unique sequences after de-noising			No. of chimaeras detected			Sensitivity		Variants after chimaera removal & above threshold		Total good reads	Hz	Number of sequences obtained for alleles				
			1st	2nd	Clus	S	P	U	100	200	>min no. reads	>min %			Allele 1	2	3	4	5
Marikebuni St 3014	Tp1	339	63	4	4	0	0	0	95.8	79.3	2	2	315	0.15	290	25			
	Tp2for	1679	444	25	24	1	1	2	100	100	3	3	1644	0.18	1484	106	54		
	Tp2rev	4416	1193	40	39	0	0	0	100	100	5	4	4372	0.09	4169	127	58	18	
	Tp3	374	45	4	4	0	0	0	97.1	82.9	1	1	352	0.00	352				
	Tp4	543	60	7	7	0	0	0	99.5	93.0	1	1	530	0.00	530				
	Tp5	5518	231	14	13	0	0	0	100	100	7	5	5076	0.11	4771	227	31	24	23
	Tp6	2027	137	4	4	0	0	0	100	100	1	1	1933	0.00	1933				
	Tp7	1910	87	4	3	0	0	0	100	100	2	2	1875	0.05	1826	49			
	Tp8	15949	298	5	5	0	0	0	100	100	2	1	15445	0	15445				
	Tp10	14003	947	53	53	0	0	0	100	100	4	4	13145	0.13	12267	463	216	199	
Marikebuni St 128	Tp1	524	123	21	21	0	0	0	98.7	57.8	3	3	426	0.090315	406	15	5		
	Tp2for	62	29	7	7	0	0	0	13.0	6.8	1	1	14	0	14				
	Tp2rev	418	196	31	29	0	0	0	97.85	84.96	1	1	378	0	378				
	Tp5	1265	92	7	7	0	0	0	100	99.8	2	2	1217	0.10842	1147	70			
	Tp10	897	145	26	25	0	0	0	100	98.3	1	1	813	0	813				

The number of reads at the different stages of the bioinformatics pipeline are shown together with the results before manual editing. Reads of different sequencing runs were combined when available (MC=Run with samples of the Muguga cocktail, Test=Testrun (see chapter 3)AB=Run with samples from African Buffalo (Chapter 5)) 1st, 2nd and Clus refer to the stages of noise removal. Chimaera detection was performed using three different algorithms: S = Chimaera Slayer P = Perseus U = Uchime. Sensitivity was calculated as the percentage chance of identifying a minor allele present at 1:100 and 1:200. '>min no. reads' = the number of variants found after removal of variants with less than 5 reads; '>min%' = the number of variants found after removal of variants which are present at a frequency of less than 0.4% of the total good reads; 'Total good reads' = the number of reads maintained after filtering of reads in the bioinformatics pipeline; Hz= heterozygosity; The number of reads obtained for each of the alleles detected are displayed.

The frequency of the different alleles in the Marikebuni stabilates is illustrated in figure 4.7 (allele colours correspond with the colours used in figure 4.6).

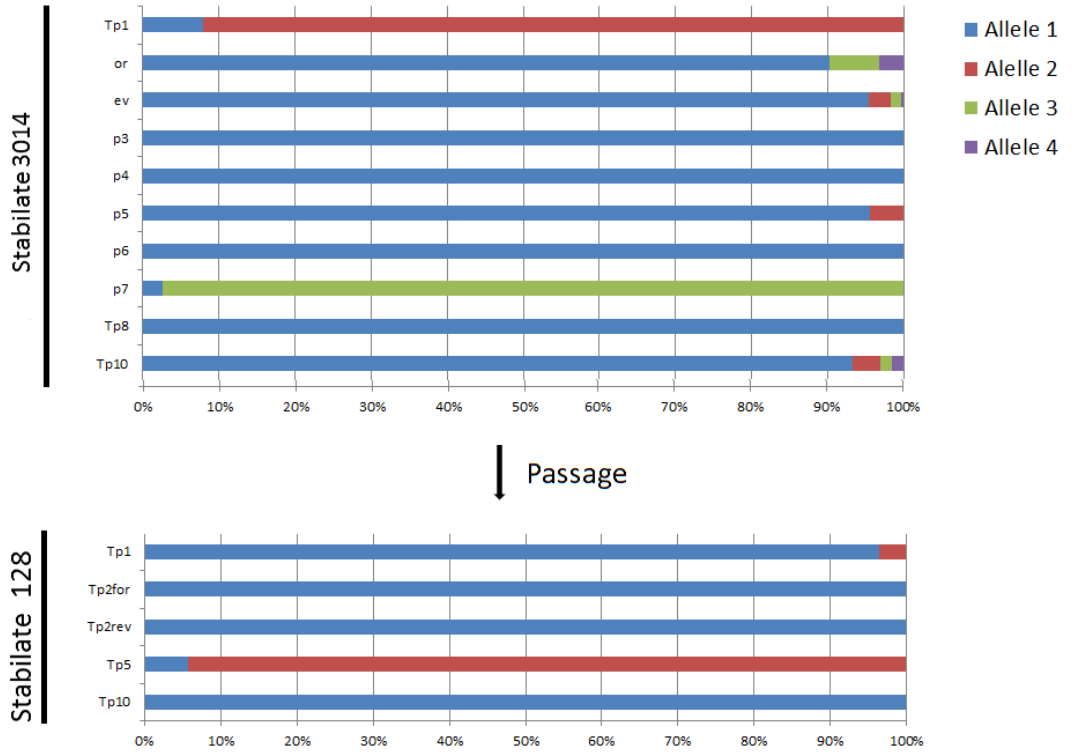

Figure 4.7 Allelic composition of Marikebuni stabilates 3014 and 128 based on nucleotide sequence of nine *Theileria parva* antigens

In Marikebuni stabilate 3014, single alleles were detected for three of the genes (Tp3, Tp4 and Tp6), and these were identical to the alleles found in Muguga and Serengeti. For Tp5 and Tp10, a second allele was detected which was identical to the allele in Kiambu 5. A minor allele of Tp7, which is identical to the allele in Muguga and Serengeti-transformed stocks, was found along with a second major allele, which is unique to the Marikebuni stock. For Tp2for, Tp2rev and Tp10, three or four alleles were found for stabilate 3014. However, only a single allele was detected in sequence reads obtained for these genes in stabilate 128. This single allele was identical to the major allele in stabilate 3014, Muguga and Serengeti. The number of reads obtained for stabilate 128 was considerably lower than that for stabilate 3014 despite aiming for a similar sequencing depth. The reduced sequencing depth coupled with the requirement to identify a minimum of five reads to define an allele

may be responsible for the failure to detect some of the additional alleles.The same two alleles were found for Tp1 and Tp5 in the two Marikebuni stabilates, but the relative proportion of the two alleles differed between the two stabilates. In stabilate 3014, Tp1 allele 1 was represented by 7.9% of the reads and allele 2 by 82.1% of the reads. For stabilate 218, allele 1 was represent by 96.5% of the reads and allele 2 by 3.5%. For Tp5, allele 1 was present at 95.5% and allele 2 at 4.5% in stabilate 3014 compared to 5.8% and 94.2% respectively in stabilate 128. Despite the relatively low number of reads obtained for stabilate 128, the results indicate that tick passage of this *T. parva* stock results in changes in the composition of the parasite population in the Marikebuni stock.

4.3.3 Antigenic diversity in the Muguga cocktail reference stabilates and Marikebuni stabilate

Multilocus sequence typing of the Muguga cocktail and Marikebuni stocks resulted in the identification of between one and four alleles per locus at the nucleotide level. The nucleotide sequences were translated into predicted amino acid sequences in order to investigate allelic diversity at the protein level and this resulted in the identification of between one and three alleles at each locus. For Tp1, T2rev, Tp4, Tp5, Tp6, Tp8 and Tp10, a single allele was found at each locus in the Muguga cocktail. Two alleles that differed by two amino acid residues were found in the Tp2for sequences. These residues, which were found within the $Tp2_{27\text{-}37}$ epitope region (H→D for residue $Tp2_{28}$ and K→N for residue $Tp2_{32}$), were outwith the sequences generated by the Tp2rev primer. Two alleles were found for Tp7, which differed by one amino acid (P →A). Three alleles were found for Tp3, involving differences in seven amino acid residues.

A series of epitopes recognised by $CD8^+$ T cell responses have been identified within the proteins encoded by seven of these genes. A summary of the allelic variants of the epitopes found in the Muguga cocktail components and the two Marikebuni stabilates, based on predicted amino acid sequences, is shown in Table 4.6. For all but one of the genes, only a single allele of the epitopes was found in the Muguga cocktail. A second variant allele was found for the $Tp2_{27\text{-}37}$ epitope in the Serengeti

and Kiambu 5 stocks. The Marikebuni stocks were found to contain between one and three variant epitopes for the genes examined: one allele was found for the Tp2 $_{138-147}$, Tp4 $_{328-338}$, Tp5 $_{87-95}$, Tp7 $_{206-214}$, Tp8 $_{378-388}$ and Tp10$_{304-318}$ epitopes. Two alleles were found for Tp1 $_{214-224}$ and three alleles for Tp2 $_{27-37}$, Tp2 $_{40-48}$, Tp2 $_{49-59}$, Tp2 $_{50-59}$, Tp2 $_{96-104}$ and Tp2 $_{98-106}$. The epitope variants differed by one to six amino acid residues from the reference sequence

Table 4.6 Epitope variants found in vaccine stabilates used for infection and treatment against *T. par va*

A) Muguga cocktail (Muguga stabilate 4230, Serengeti stabilate 4229 and Kiambu 5 stabilate 4230). Amino acid residues that differ from the *T. parva* Muguga reference genome are highlighted in red.

Var.	$Tp1_{214-224}$	$Tp4_{328-338}$	$Tp5_{87-95}$	$Tp10_{304-318}$
V1	VGYPKVKEEML	TGASIQTTL	SKADVIAKY	TNNFNNPELIPVL

Var.	$Tp2_{27-37}$	$Tp2_{40-48}$	$Tp2_{49-59}$	$Tp2_{50-59}$	$Tp2_{96-104}$	$Tp2_{98-106}$	$Tp2_{138-147}$
V1	SHEELKKLGML	DGFDRDALF	KSSHGMGKVGK	SSHGMGKVGK	FAQSLVCVL	QSLVCVLMK	KTSIPNPCKW
V2	S**D**EEL**N**KLGML						

B) Marikebuni (stabilate 3014). Amino acid residues that differ from the *T. parva* Muguga reference genome are highlighted in red.

Var.	$Tp1_{214-224}$	$Tp4_{328-338}$	$Tp5_{87-95}$	$Tp10_{304-318}$
V1	VGYPKVKEEML	TGASIQTTL	SKADVIAKY	TNNFNNPELIPVL
V2	VGYPKVKEE**II**			

Var.	$Tp2_{27-37}$	$Tp2_{40-48}$	$Tp2_{49-59}$	$Tp2_{50-59}$	$Tp2_{96-104}$	$Tp2_{98-106}$	$Tp2_{138-147}$
V1	SHEELKKLGML	DGFDRDALF	KSSHGMGKVGK	SSHGMGKVGK	FAQSLVCVL	QSLVCVLMK	KTSIPNPCKW
V2	S**D**EEL**ET**LGML		**LT**S**KA**M**MT**VGK	**T**S**KA**M**MT**VGK			
V3	S**EA**EL**RKM**GM**I**		**K**SSHGMGK**I**GK	SSHGMGK**I**GK			

4.4 Discussion

The aim of the work in this chapter was to generate a reference profile for the masterseed stocks of the Muguga cocktail using satellite DNA typing and multilocus sequence typing using 454 technology. This profile can be used for comparison with progeny vaccine stabilates, other vaccine stabilates and field parasite populations. Satellite DNA typing has previously been used for characterization of *T. parva* parasites, including the components of the Muguga cocktail (Oura et al., 2004b; Oura et al., 2003; Patel et al., 2011). In this study, satellite DNA typing was used to examine consistency of content of the three reference stabilates and the corresponding Muguga cocktail vaccine batches. In order to obtain higher resolution, a larger panel of satellite DNA markers was used than in the previous study by Patel *et al.* (Patel et al., 2011). The benefits of using satellite typing are that this method is relatively low-cost and rapid and that data analysis is comparatively straightforward. However, the method may fail to detect differences between closely related parasites and its sensitivity is limited in detecting minor alleles present at low abundance. Furthermore, results between labs cannot easily be compared unless the methods used for product size determination and product quantification are carefully standardised (Oura et al., 2004b; Patel et al., 2011).

4.4.1 Comparison of methods for satellite DNA typing

Multiple alleles were identified at a number of satellite markers for the individual components of the Muguga cocktail (Stabilate 4228-4230). Patel *et al.* also detected a second allele with four markers (MS7, MS19, MS25 and MS39) in the ILRI Serengeti reference stabilate (Stabilate 4229) and for two markers (MS7 and MS19) in the ILRI and FAO Muguga reference stabilates (Stabilate 4229 and 4139) (Patel et al., 2011). However, Oura *et al.* detected single alleles for each of the 31 satellite markers used in infected cell lines established *in vitro* by infecting PBMC with stabilates 68,69 and 73 (see figure 4.1) (Oura et al., 2007b). This discrepancy is probably due to the use of cell lines which are frequently dominated by individual genotypic components, which overgrow other components following *in vitro* passage. Alternatively, the discrepancy could be due to the analysis of PCR products using Spreadex gels rather than the capillary flow electrophoresis. Capillary flow

electrophoresis allows detection of low abundance products not visible on agarose gel.

In this study, allele sizes were determined by capillary flow electrophoresis and fluorophore detection, as was done in the study by Patel *et al.* (Patel et al., 2011). Overall, the estimated allele sizes were very similar in each study, but some minor differences in allele size did arise and this is likely due to minor differences in allele calling methodology and the use of different instruments. In the present study, the relative abundance of alleles was determined by calculating the area under the curve in the traces and this was shown to be relatively reproducible between laboratories (Edinburgh and Glasgow). The most accurate method in terms of allele sizing and determining the relative abundance of alleles is therefore capillary flow electrophoresis.

4.4.2 Satellite DNA profile of the Muguga cocktail

The ILRI Muguga and Serengeti-transformed stocks had identical single alleles at seven out of the eleven markers. For the remaining four markers (MS7, MS8, MS25 and MS30), one or two major alleles were found, which were shared between the Muguga and Serengeti-transformed stocks and in some cases an additional allele was found in one of the stocks. Differences between the two stocks were also detected in the study by Patel *et al.*, which utilised four satellite markers (MS7, MS8, MS25 and MS39). Oura *et al.* detected different single alleles for four of the 31 satellite markers used (MS7, MS19, MS25 and MS43) (Oura et al., 2007b). Two of these satellites (MS19 and MS43) were not examined in the current study. The results from the present study indicate that the Kiambu 5 component is genotypically distinct from Muguga and Serengeti. It has a distinct and unique single allele for four of the ten markers that successfully amplified. The single alleles found at four other markers were also found as minor components in the Serengeti-transformed stock. All three components of the Muguga cocktail were identical at the remaining two markers.

The diversity in the Muguga cocktail batches is limited, with between one and four alleles present for each of the eleven satellite DNA markers used and this reinforces the previous finding that diversity within the Muguga cocktail is limited (Oura et al., 2004b; Patel et al., 2011). The satellite DNA profiles were similar for the four batches of the Muguga cocktail vaccine, ILRI 0801-ILRI0804, and the alleles identified corresponded with those found in the three reference stabilates, which were produced simultaneously with the Muguga cocktail vaccine batches. This indicates that there is consistency between vaccine batches produced in the same vaccine production run.

Comparison of two Muguga reference stabilates two passages apart demonstrated differences involving loss or gain of minor alleles at three of the ten loci (Stabilate N57 and Stabilate 4230). Patel *et al.* also demonstrated additional alleles in the ILRI Serengeti reference stock compared to the FAO Serengeti reference stock (Patel et al., 2011). Differences in satellite profiling following tick passage have previously been observed in studies with Marikebuni *T. parva* stocks (Katzer et al., 2006). These findings highlight the potential for changes in the precise composition of parasite populations following tick passage. The components of the Muguga Cocktail are, in effect, small recombining parasite populations based on small founder populations represented by the initial isolates. Consequently, low frequency alleles are at particular risk of drifting out of the populations and particular alleles have becoming fixed at many loci. Current vaccine components therefore have a low, but measurable, level of diversity. It is therefore desirable to produce large quantities of seed stabilates for vaccine production to minimise such variation during the production process. The satellite DNA profiles established in this study can be used for comparison with future progeny reference stabilates.

4.4.3 Deep multilocus sequence typing – quantification of abundance of allelic variants

The deep multilocus sequence typing system was applied to establish a high-resolution profile of the antigenic composition of the current reference stabilates and future seed stabilates. The depth of this analysis was limited by the relatively high error rate inherent in the process, as observed while sequencing control samples (see Chapter 3). A conservative cut-off of 0.4% abundance was chosen to differentiate between true variants and potentially erroneous reads. This still afforded a high of level of resolution in assessing the composition of parasites used for vaccination. Moreover, alleles present beneath this threshold in a population/generation may not always be present in a single dose of vaccine stabilate as the effective vaccination dose is relatively low as discussed in the section below.

4.4.4 Estimation of effective number of sporozoites in a single ITM dose

The Muguga cocktail vaccine stabilate batches produced at ILRI in 2008 contained 275.53 infected salivary gland acini/vaccine component/ ml of stabilate, based on the average number of infected acini per tick and the number of ground-up ticks per ml of stabilate (personal communication Ekta Patel, ILRI). Based on the dimensions of sporozoites and mature infected salivary gland cells, Fawcett *et al.* estimated that each infected acinus contains up to 4 $x10^4$ sporozoites (Fawcett et al., 1982a). It is unlikely that all these sporozoites would have been mature and capable of infecting bovine cells *in vivo* when they were harvested for stabilate production. The number of sporozoites maturating depends on the length of time ticks pre-feed on rabbits; the highest number of mature sporozoites in the salivary gland, before their release into the saliva, is observed after four days of feeding on rabbits (Kimbita et al., 1997). If we assume 90% of the sporozoites are mature and infective when harvested at four days, then 1 ml of undiluted vaccine stabilate can be estimated to contain 1 x 10^7 viable sporozoites (= 275.53 x 4 $x10^4$ x 90%). Major losses in infectivity are predicted to occur at different stages of the vaccine production process. For example, for the production of *Theileria annulata* stabilate it has been estimated that only 24% of sporozoites are recovered after the grinding process (Walker et al., 1983).

Theileria sporozoites are very fragile and considerable further losses of sporozoites occur through the freezing and thawing process. In *T. parva*, there is an estimated infectivity loss of 30- to 100-fold when frozen stabilate is thawed compared to the use of fresh sporozoites (Personal communication Daniel Ngugi, RVC). Extrapolating from a 100-fold loss in infectivity, each ml of undiluted stabilate may be estimated to contain 10^5 viable sporozoites per stock, which will result in 10^3 sporozoites per stock for each vaccination dose after a 100-fold dilution (ie. a total dose of 3,000 sporozoites). However, the proportion of these infective sporozoites that successfully infect bovine cells *in vivo* following inoculation into an animal is unknown. Further loss of infectivity can occur in the field, as it is difficult to maintain a cold chain at all stages of vaccine delivery in the field. An *in vitro* infectivity loss of between 1% and 4% per hour, depending on the parasite stock, was found for stabilates stored in an ice bath (Mbao et al., 2007). A proportion of the animals vaccinated with parasites kept on ice for eight hours or more did not withstand subsequent parasite challenge (Marcotty et al., 2001). It was therefore recommended that stabilate be used within eight hours of thawing (Marcotty et al., 2001; Morzaria et al., 1999). The absolute loss of infectivity in the field has not been quantified, but if we assume a modest loss of 20% takes place, then each dose will contain approximately 800 infective parasites. Based on these assumptions, variant genotypes/alleles detected at 0.4% frequency by methods used in the current study would equate to 3.2 infective parasites per vaccine dose. Given the numerous factors that can potentially result in variation in the dose of parasites delivered to an animal, it is likely that not all animals will receive minor parasite genotypes present at 0.4% in the bulk vaccine. These calculations indicate that, for practical purposes, the level of resolution provided by the deep sequencing procedure used in this study is sufficient to assess the composition of the parasites used for vaccination.

4.4.5 Deep multilocus sequence typing - Muguga cocktail components

The deep multilocus sequence typing system focused on genes encoding $CD8^+$ T cell target antigens. Previous studies of $CD8^+$ T cell responses to *T. parva* have shown that these frequently exhibit parasite strain specificity and analyses of allelic sequences of two of the parasite antigen genes in field isolates of *T. parva* revealed extensive diversity (Pelle et al., 2011; Taracha et al., 1995a). However, the parasites in the Muguga cocktail show relatively limited diversity for each of the loci examined in this study, with each component appearing to contain one predominant allele plus, in some cases, one or more minor components. The differences between the alleles were, in most cases, minimal, with only a few variant nucleotides. Moreover, in five of the genes, the nucleotide variation did not result in any amino acid changes in the protein. In the Muguga cocktail reference stabilates, allelic variation was found in the encoded protein for Tp2, Tp3, and Tp7. This did not result in variation in the epitopes identified in Tp7 and only a single allelic variant was identified in one of the seven epitopes identified in Tp2 (Tp2 $_{27\text{-}37}$). These findings coupled with the results from the DNA satellite typing indicate that diversity within the Muguga cocktail is very limited both in terms of the number of alleles and the differences between the alleles. Despite this limited diversity within the Muguga cocktail, it appears to give broad protection against a range of *T. parva* parasites. This contrasts with the high degree of diversity for Tp1 and Tp2 genes that was documented in field isolates from both cattle and buffalo (30 and 41 alleles for Tp1 and Tp2 respectively). Diversity was especially high in buffalo-derived and buffalo-associated parasite isolates with between 18 and 25 variants found for the CD8 epitopes within the Tp2 gene (Pelle et al., 2011).

4.4.6 Deep multilocus sequence typing - Marikebuni vaccine stabilate

A second parasite population used for vaccinating cattle in Kenya, the Marikebuni stock, also showed limited genotypic diversity based on satellite DNA typing. However, there was a higher degree of diversity in the Marikebuni stock compared to the Muguga cocktail based on the antigens sequenced in this study. Multilocus sequencing identified the presence of two variants of the Tp1 epitope and three variants in six of the epitopes in Tp2 (Tp2 $_{27\text{-}37}$, Tp2 $_{40\text{-}48}$ Tp2 $_{49\text{-}59}$, Tp2 $_{50\text{-}59}$, Tp2 $_{96\text{-}104}$, Tp2 $_{98\text{-}106}$). Differences in satellite profile following tick passage have been observed in previous studies with the Marikebuni stock (Katzer et al., 2006). A marked difference in the allelic composition between passages was detected in the present study by deep sequencing the Tp1 and Tp5 loci. Thus, for the Marikebuni stock, deep multilocus sequence typing gave a higher degree of resolution compared to satellite DNA typing.

4.4.7 Composition of the Muguga cocktail

The Muguga cocktail was formulated in the 1970s based on cross-immunity studies. The inClusion of the Serengeti-transformed component gave better protection than Kiambu 5 and Muguga used as a bivalent vaccine (Radley et al., 1975b). However, the Muguga and Serengeti-transformed stocks appear very similar using multilocus sequence typing and DNA satellite typing, confirming the findings of several previous studies (Bishop et al., 2001; Geysen et al., 1999; Oura et al., 2007b). It has been speculated that the original *T. parva* Serengeti-transformed stock has been contaminated with and supplanted by *T. parva* Muguga genotypes over the years (Bishop et al., 2001; Oura et al., 2004b). Buffalo often demonstrate a high multiplicity of infection with *T. parva* with individual animals harbouring genotypically diverse parasites (Oura et al., 2011a, b; Pelle et al., 2011). While it is likely that passage of a buffalo-derived parasite stock through cattle by tick transmission would result in loss of genetic diversity, it seems unlikely that such a parasite would end up being so genetically similar to the cattle-derived Muguga stock. Unfortunately, stabilate material representing early passages of the Serengeti transformed stock are not available and therefore it is impossible to confirm whether

the composition of current parasite stocks is the same as those tested in the 1970s. Repeating of the cross-immunity experiments with the current parasite stocks would confirm whether the current parasite stocks induce a similar cross-protection profile. If so this might suggest that the minor components of the Serengeti-transformed stock, which are distinct from Muguga, are important in the broad protection provided by the Muguga cocktail.

Even though the Muguga cocktail appears to protect against a range of cattle-derived *T. parva* parasites and provides protection in the field, there are some reports indicating that not all animals are protected against parasites from different areas of Africa (Lynen et al., 1991). As suggested before by Bishop *et al.* and Oura *et al.*, the formulation of the Muguga cocktail should be validated, as better protection might be provided by formulating an alternative cocktail (Bishop et al., 2001; Oura et al., 2007a). If minor components in the parasite stocks contribute to the breadth of immunity, it is important to know whether their presence or absence within a vaccine can explain in variation in the protection elicited. When two passages of the Marikebuni stock were analysed, at certain markers, particular alleles were identified in only one of the passages (This study and (Katzer et al., 2010) and there is some indication that there is variation in composition of components of the Muguga cocktail (This study and (Patel et al., 2011). A more standardised composition could be achieved by using carefully selected, genotypically distinct, *T. parva* clones. The use of clones would make quality control easier as the presence of minor components would not need to be monitored and maintained. The inClusion of clones that are genotypically more distinct may make the cocktail more suitable for deployment over a wider geographical area.

Two local stocks (Katete and Chitongo) are currently used for immunisation against ECF in Malawi and Zambia. The stocks are genotypically quite different from the stocks in the Muguga cocktail (This study; (Geysen et al., 1999; Hayashida et al., 2013). Interestingly, breakthroughs were seen when the Katete stock was used to challenge Muguga cocktail-immunised animals (Geysen, 2000), whereas Muguga cocktail-immunised animals were immune against challenge from the Chitongo stock (Geysen, 2000). The Katete and Chitongos stocks had alleles in common with the

Muguga cocktail for two out of ten and six out of ten DNA satellite markers respectively.

The techniques used in this study may be used for preliminary screening of candidate parasites stocks for inClusion in an improved formulation of a cocktail vaccine. Although it would be relatively easy to identify a number of genotypically distinct parasite clones, it will, unfortunately, be necessary to perform extensive *in vivo* cross-immunity experiments to confirm the utility of the novel components.

Chapter 5: *Theileria parva* in African buffalo

5.1 Introduction

African buffalo are considered to be a reservoir host for *T. parva* and can be infected with high levels of the parasite in the absence of clinical disease (Oura et al., 2011a; Young et al., 1978). In contrast, cattle develop severe clinical disease with a high mortality rate when infected with buffalo-derived *T. parva* (Lawrence, 1979). Previous studies have shown that only a proportion of animals ITM-immunised using cattle-derived *T. parva* stocks, including the Muguga Cocktail, were protected against buffalo-derived *T. parva* (Lohre, 1978; Radley et al., 1975c; Radley et al., 1979). This lack of protection against buffalo-derived parasites is believed to be due to strain specificity in the immunity induced by infection and treatment. In particular, this has been attributed to the use of vaccine stabilates with limited parasite diversity and subsequent challenge with a mixture of strains representing a variety of allelic variants in epitopes recognised by the immune system.

A series of experiments have been conducted using buffalo-derived *T. parva* stabilates for the infection and treatment method of vaccination of cattle. These stabilates were produced by either collecting ticks from the ground of paddocks frequented by buffalo (Dolan, 1984; Latif et al., 2001) or by feeding ticks on infected buffalo (Cunningham et al., 1974; Mutugi et al., 1988a; Mutugi et al., 1988b; Young et al., 1977a). In most cases, vaccination with these stabilates led to the induction of protective immunity against buffalo-derived *T. parva* challenge. Furthermore, cross-protection was noted between buffalo-derived stabilates from different locations in Kenya when cattle were immunised with a single buffalo-derived *T. parva* stabilate (Mutugi et al., 1988a; Mutugi et al., 1988b). These results clearly show that cattle can develop immunity against buffalo-derived *T. parva* and that there is the potential for a single buffalo-derived stabilate to be used effectively in different geographical areas (Mutugi et al., 1988a; Mutugi et al., 1988b).

The multiplicity of infection with *T. parva* in carrier buffalo has been found to be much higher than that in cattle (Conrad et al., 1987b; Oura et al., 2011a; Pelle et al., 2011). Oura *et al.* (2011) found evidence for the presence of at least nine alleles at a single locus in an individual African buffalo, indicating that a minimum of nine

parasite genotypes were present. Marked differences in antigenic diversity have been noted between isolates from different buffalo and between isolates from the same buffalo taken at different time-points (Conrad et al., 1987). Microsatellite-based analysis of *T. parva* maintained in African buffalo in Uganda indicated the presence of distinct parasite populations in Lake Mburo National Park and Queen Elizabeth National Park (Oura et al., 2011a). Variation in PIM PCR-RFLP profiles was observed among samples from buffalo from the Kruger National Park, whereas the profiles from buffalo from Hluhluwe-iMfolozi Park, Abalingwe Game Reserve and Ithala Game Reserve were much more similar to one another (Sibeko et al., 2011). Recent work characterising the parasite has focused on Tp1 and Tp2, two genes known to encode antigens that can be recognised by bovine $CD8^+$ T cell responses. A large number of allelic sequence variants were found for these genes among buffalo-derived isolates and buffalo-associated cattle-derived isolates from Kenya (Pelle et al., 2011).

The main aim of this chapter is to gain an insight into genotypic diversity between and within buffalo-derived *T. parva* populations from two geographically distinct areas in South Africa and Kenya. Genotypic diversity of buffalo-derived *T. parva* will be analysed using a multi-locus sequence typing system based on Roche 454 technology, the development of which is detailed in Chapters 3 and 4. This system has the major advantage that it will allow the identification of multiple allelic variants of the parasite within individual animals. The antigenic diversity found in the samples from buffalo will then be compared to antigenic diversity found in the Muguga Cocktail vaccine.

5.2 Material and methods

5.2.1 Samples from buffalo

DNA from African buffalo was provided by Nicholas Yules, Pirbright Institute and ILRI, Kenya. Genomic DNA was extracted from the whole blood of buffalo from the Kruger National Park (South Africa) and from blood clots of buffalo from the Ol Pejeta conservancy (Kenya). The buffalo were found in different parts of the conservancies and varied in age and sex (see Table 5.1).

Table 5.1 Origin of buffalo samples used in this study

Animal ID	Approx. age	Sex	Location	Park	Country
SC01	18 months	F	Crocodile bridge	Kruger National Park	South Africa
SC02	2 years	M	Crocodile bridge	Kruger National Park	South Africa
SC03	2 years	F	Mnyeleni	Kruger National Park	South Africa
SC04	3 years	F	Mnyeleni	Kruger National Park	South Africa
SC05	3.5 years	M	Sekurakwane	Kruger National Park	South Africa
SC06	3 years	M	Sekurakwane	Kruger National Park	South Africa
301	5 years	M	Sweet waters N 00° 02.752' E 36° 56.165'	Ol Pejeta Conservancy	Kenya
302	2 years	M	Sweet waters N 00° 01.331' E 36° 56.104'	Ol Pejeta Conservancy	Kenya
303	3 years	F	Sweet waters N 00° 01.332' E 36° 56.104'	Ol Pejeta Conservancy	Kenya
304	1 year	M	West of Ewaso Ngiro river N 00° 01.332' E 36° 56.050'	Ol Pejeta Conservancy	Kenya
305	5 years	F	West of Ewaso Ngiro river N 00° 02.225' E 36° 51.224'	Ol Pejeta Conservancy	Kenya
306	> 6 years	M	Sweet waters N 00° 00.799' E 36° 57.797'	Ol Pejeta Conservancy	Kenya
307	6 years	F	Sweet waters N 00° 01.668' E 36° 55.446'	Ol Pejeta Conservancy	Kenya
308	5 years	F	Sweet waters N 00° 01.668' E 36° 55.489'	Ol Pejeta Conservancy	Kenya

5.2.2 Amplicon sequencing using Roche 454 technology

Amplicons were generated of the Tp1, Tp2, Tp4, Tp5, Tp6 and Tp10 genes using the PCR conditions and 'fusionprimers' as described in Chapter 3. The raw sequencing data was processed using the bioinformatics pipeline described in Chapter 3. The resulting sequences were subjected to a number of analyses described in chapter 2; these include: calculation of the number of polymorphic sites in nucleotide and amino acid sequences, heterozygosity, mean distance between variants, principal co-ordinate analysis (PCoA) and construction of phylogenetic trees.

5.3 Results

Deep multi-locus sequence typing was performed on one of the buffalo samples from the Kruger National Park, using all nine *T. parva* loci previously used to characterise the Muguga Cocktail components. Polymorphism was found in each of the genes, with between three and 19 alleles identified at each locus (see Figure 5.1).

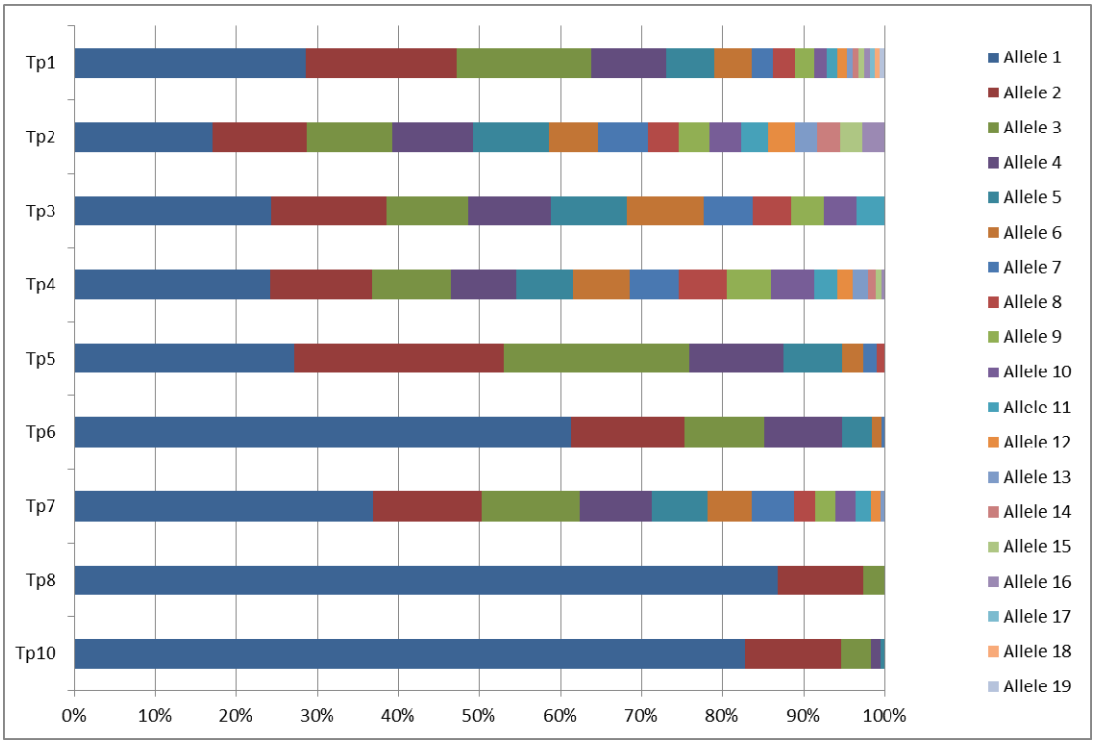

Figure 5.1 Proportion of different alleles found at nine loci representing the parasite population in a single buffalo. The sequences were obtained using DNA from a buffalo from the Kruger national Park (SC06)

Following the analysis of a single buffalo parasite population, multi-locus sequence typing was performed on all the buffalo samples available using a set of six genes, namely Tp1, Tp2, Tp4, Tp5, Tp6 and Tp10. The resulting sequences were subjected to a number of analyses to estimate diversity for the different genes in the different buffalo populations. These include: calculation of the number of polymorphic sites in nucleotide and amino acid sequences, heterozygosity and mean distance between variants; the results can be found in Table 5.2. Further investigation of the relationship between sequences was undertaken using principal co-ordinate analysis and the construction of phylogenetic trees.

Table 5.2 Diversity of buffalo-derived *T. parva* at six loci

Gene	Amplicon					CDS	Protein	
	Polymorphic sites	No. of variants	Mean distance between variants	S.E.M.	Mean Hz	Polymorphic sites	Polymorphic sites	No. of variants
Overall								
Tp1	36/354[1]+ind	76	0.024	0.004	0.829	-	20/118	67
Tp2[2]	171/246	70	0.319	0.032	[3]	-	68/82	62
Tp4	115/384	74	0.073	0.008	0.879	15/120	2/40	3
Tp5	29/306	47	0.018	0.004	0.831	11/186	0/62	1
Tp6	27/288	29	0.017	0.004	0.667	-	1/96	2
Tp10	30/278	41	0.026	0.007	0.284	11/147	0/49	1
Ol Pejeta								
Tp1	23/354[1]+ind	37	0.018	0.004	0.795	-	15/118	31
Tp2[2]	161/246	32	0.311	0.030	[3]	-	65/82	32
Tp4	73/384	31	0.069	0.009	0.855	11/120	2/40	2
Tp5	15/306	15	0.017	0.005	0.797	5/186	0/62	1
Tp6	7/288	9	0.015	0.004	0.531	-	1/96	2
Tp10	23/306	26	0.025	0.007	0.282	7/147	0/49	1
Kruger National Park								
Tp1	19/357[1]+ind	40	0.017	0.004	0.870	-	12/119	36
Tp2[2]	162/246	38	0.348	0.043	[3]	-	64/82	38
Tp4	98/384	46	0.073	0.009	0.912	15/120	2/40	3
Tp5	25/306	33	0.018	0.005	0.887	11/186	0/62	1
Tp6	25/288	25	0.018	0.005	0.850	-	0/96	1
Tp10	22/306	17	0.026	0.007	0.284	7/147	0/49	1

[1] number of polymorphic sites shown does not include residues in insertion/deletion regions (ind); [2] only the first 252 bp of Tp2 forward reads for were used for analysis; [3] heterozygosity could not be calculated for Tp2 as the number of reads obtained for some of the animals was low. The number of variants based on the nucleotide sequence and the mean divergence between sequence variants and the corresponding standard error (S.E.M) and mean heterozygosity (Hz) are shown. The Tp4, Tp5 and Tp10 gene contain introns and the number of polymorphic sites in the coding sequence (CDS) is shown. The number of polymorphic sites in the predicted protein sequence and the number of allelic variants is shown.

Overall, a high number of variant nucleotide sequences was detected for each of the genes in this study combined, ranging from 27 alleles for Tp6 to 76 alleles for Tp1. The number of alleles was highest in Tp1, Tp2 and Tp4, which led to a high number of amino acid variants for Tp1 and Tp2 (Table 5.2). For all the genes, many of the alleles were 'private' to either African buffalo from the Kruger National Park or to African buffalo from the Ol Pejeta conservancy. A limited number of alleles was

shared between African buffalo from the two locations (between 1 and 4 depending on the gene). For Tp5 and Tp10 nucleotide changes in the coding region did not result in amino acid changes. The number of nucleotide changes was proportionally higher in the intron segments of the amplicons representing spliced genes, i.e. Tp4, Tp5 and Tp10 (based on data presented in Table 5.2). Within the different buffalo parasite populations the highest diversity was observed for the Tp2 gene, as expressed by the mean distance between sequences. The distance between sequences within a population was much lower for the other genes (see Figure 5.2). The high mean heterozygosity found for Tp1, Tp4, Tp5 and Tp6 indicates that each buffalo carries multiple parasite genotypes and no one genotype dominates the mixture (Table 5.2).

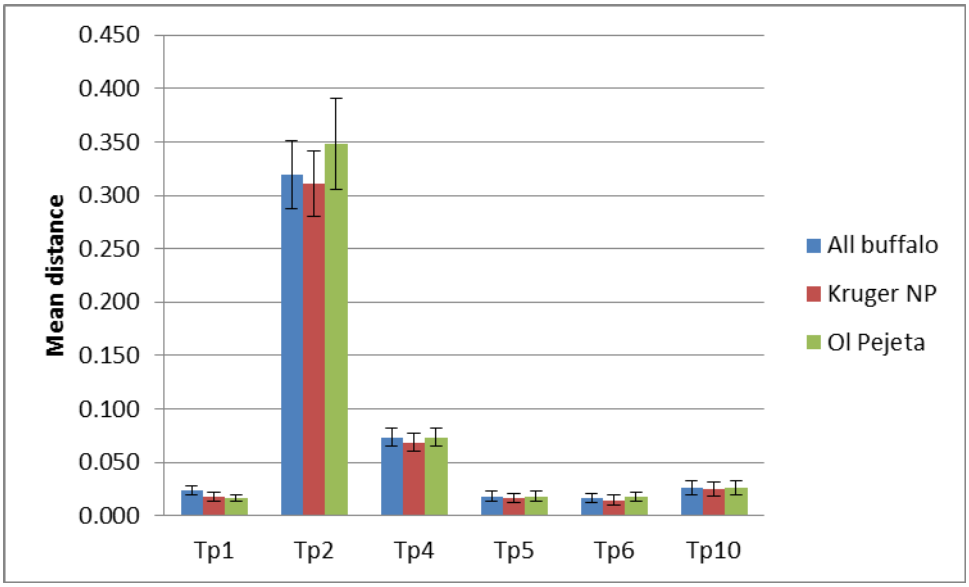

Figure 5.2 Mean distance between sequences obtained for *T. parva* populations from African buffalo from Ol Pejeta and the Kruger National Park

5.3.1 Number of allelic variants in buffalo from Kruger National Park and the Ol Pejeta conservancy

The average number of alleles detected in individual animals was high for all of the genes for African buffalo from both the Kruger National Park and the Ol Pejeta conservancy (see Figure 5.3). Although the total of number of alleles found in African buffalo the Kruger National Park and the Ol Pejeta conservancy was similar, the number of allelic variants for Tp4, Tp5 and Tp10 within an individual African buffalo appears to be higher for African buffalo from the Kruger National Park (see Figure 5.3). This difference cannot be explained by a difference in the number of reads obtained. A similar trend was observed for the Tp10 gene, but unfortunately only five of the eight datasets from the Ol Pejeta buffalo had sufficient reads for further analysis (the other three were excluded from further analysis); substantial variation was found in the remaining datasets. Unfortunately, the number of reads obtained for Tp2 in individual buffalo was too low for analysis of the number of alleles in individual animals and could not be included in the analysis. Nevertheless, it appears that the allelic diversity of *T. parva* is greater in buffalo from the Kruger National Park than from the Ol Pejeta conservancy.

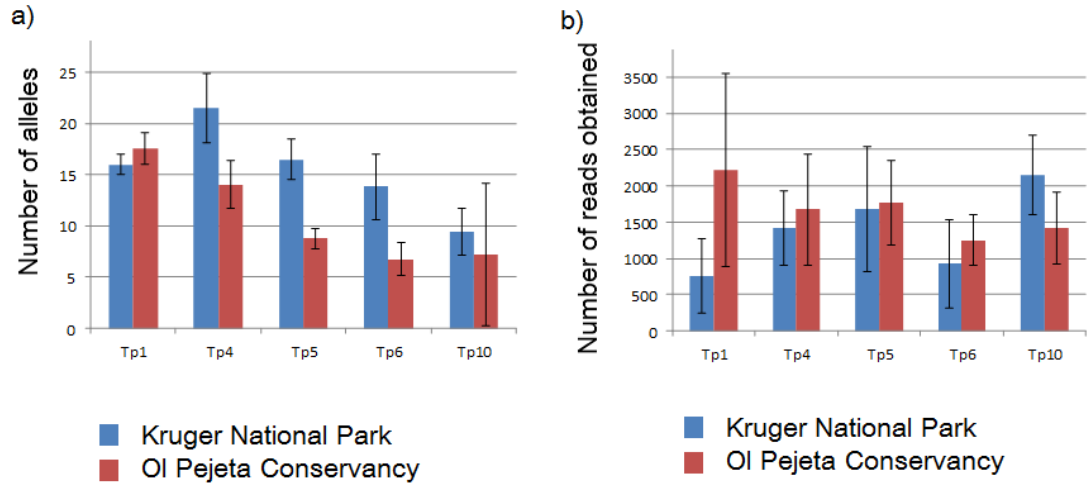

Figure 5.3 Number of alleles found in individual buffalo and read depth for Tp1, Tp4, Tp5, Tp6 and Tp10
a) The average number of allelic variants and standard deviation are shown for buffalo samples from the Kruger National Park and the Ol Pejeta conservancy. Only datasets that contained > 400 reads after bioinformatic processing were used for the analysis. **b)** The average number of reads and standard deviations for the datasets used to make Figure 5.3a.

5.3.2 Comparison of *T. parva* maintained in buffalo from the Kruger National Park and the Ol Pejeta conservancy

The sequences obtained for the different *T. parva* genes from buffalo from the Kruger National Park and the Ol Pejeta conservancy were used for principal co-ordinate analysis. For Tp2, Tp4, Tp5, Tp6 and Tp10, no meaningful sub-clustering was observed using PCoA analysis; no Clustering was observed among sequences from individual animals, nor between sequences derived from African buffalo from the two locations. However, the fact that many 'private' alleles were identified in each location from more than one buffalo suggests that the parasite populations have differentiated, despite the lack of discrimination in the PCoA analysis. As an example, the results for Tp5 are illustrated in Figure 5.4.

A degree of sub-Clustering can be seen for *T. parva* from buffalo from the Ol Pejeta conservancy and the Kruger National Park using sequences obtained for Tp1 and this is illustrated in Figure 5.5. Sequences from each sampling location are largely discriminated by the first axis of the graph, which explains 30.1 % of the variation in the dataset. Individual buffalo were associated with a variety of sequences and clustering was not observed among sequences derived from individual animals. A number of cattle-derived Tp1 sequences (n = 22) from a previous study by Pelle *et al.* (2011) were incorporated into a PCoA (see Figure 5.6). The extensive diversity among the buffalo-derived sequences contrasts with the limited diversity among those derived from cattle. The cattle-derived sequences clustered around the edges of the buffalo-derived sequences obtained.

Similarly, a number of cattle-derived Tp2 sequences (n = 22) from a previous study by Pelle *et al.* (2011) and this study (Muguga Cocktail and Marikebuni vaccine stock, see Chapter 3) were incorporated into a PCoA and this is shown in Figure 5.7. The extensive diversity among the buffalo-derived sequences contrasts with the limited diversity among those derived from cattle. For this gene the cattle-derived sequences also cluster around the edges of the buffalo-derived sequences obtained.

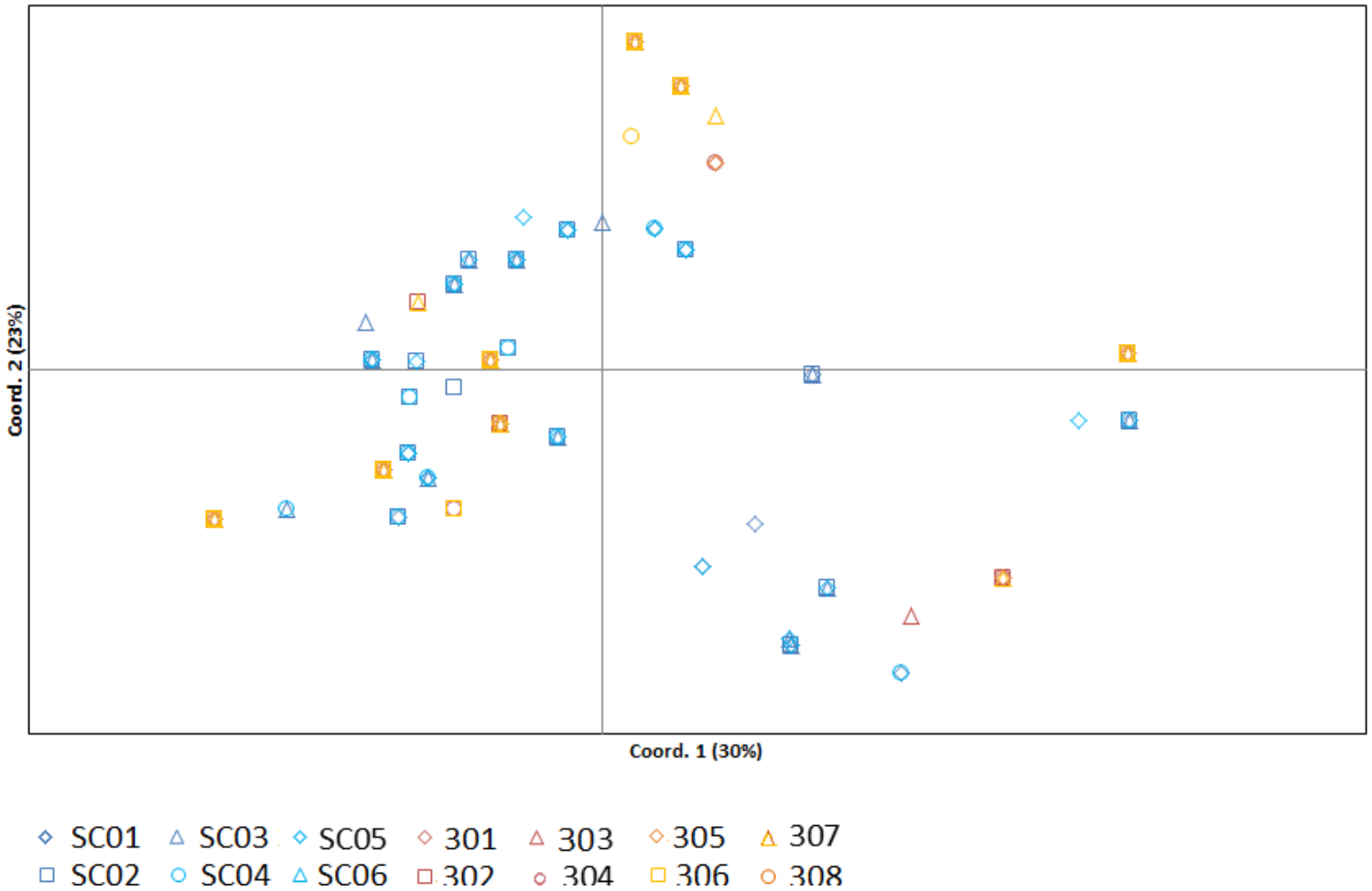

Figure 5.4 PCoA of Tp5 allelic diversity in *T. parva* in African buffalo
Illustration of the relationship between *T. parva* present in buffalo from the Kruger National Park (SC01-SC06) and the Ol Pejeta conservancy (301-308)

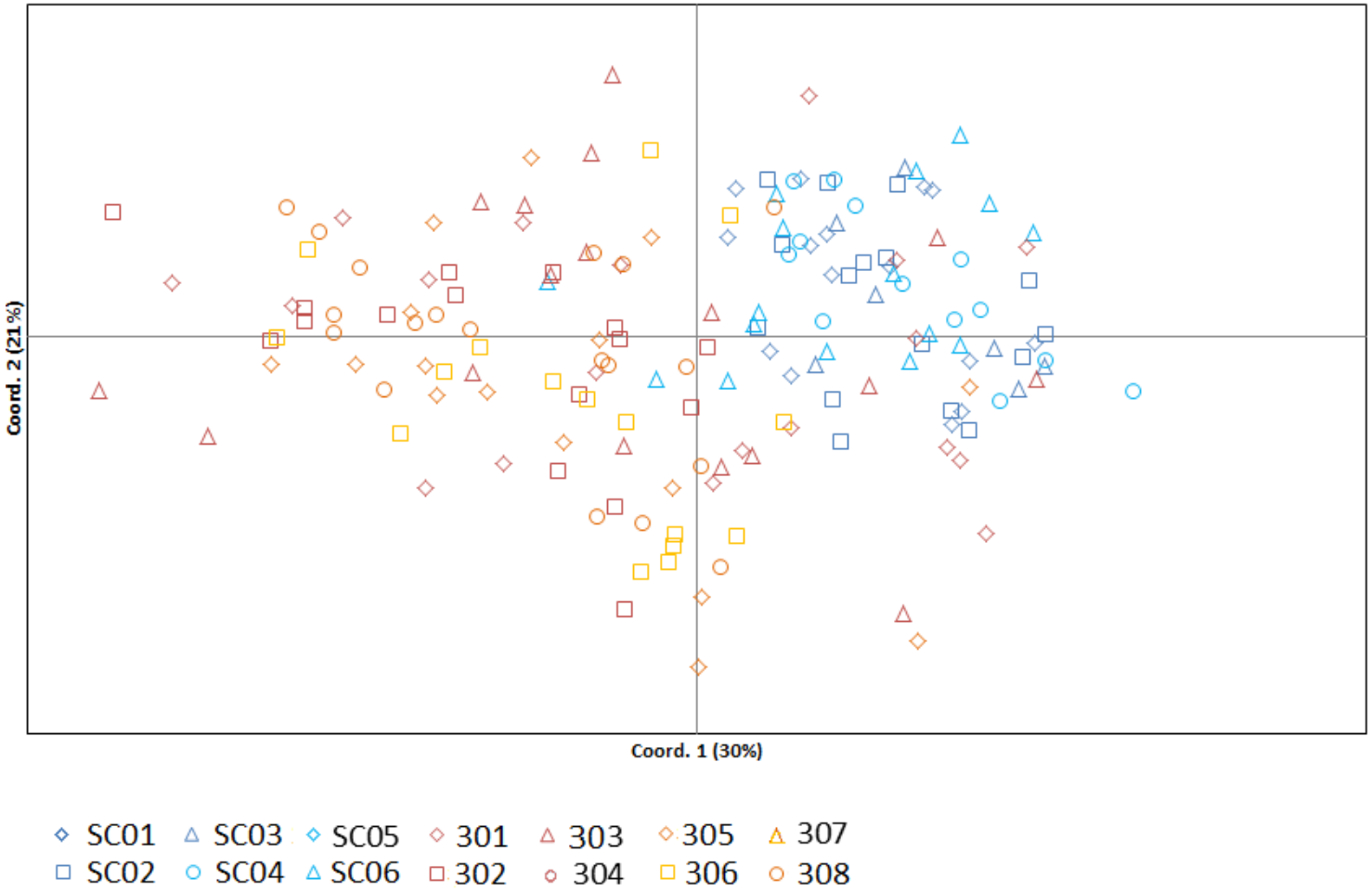

Figure 5.5 PCoA of Tp1 allelic diversity in buffalo
Illustration of the relationship between *T. parva* present in buffalo from the Kruger National Park (SC01-SC06) and the Ol Pejeta conservancy (301-308)

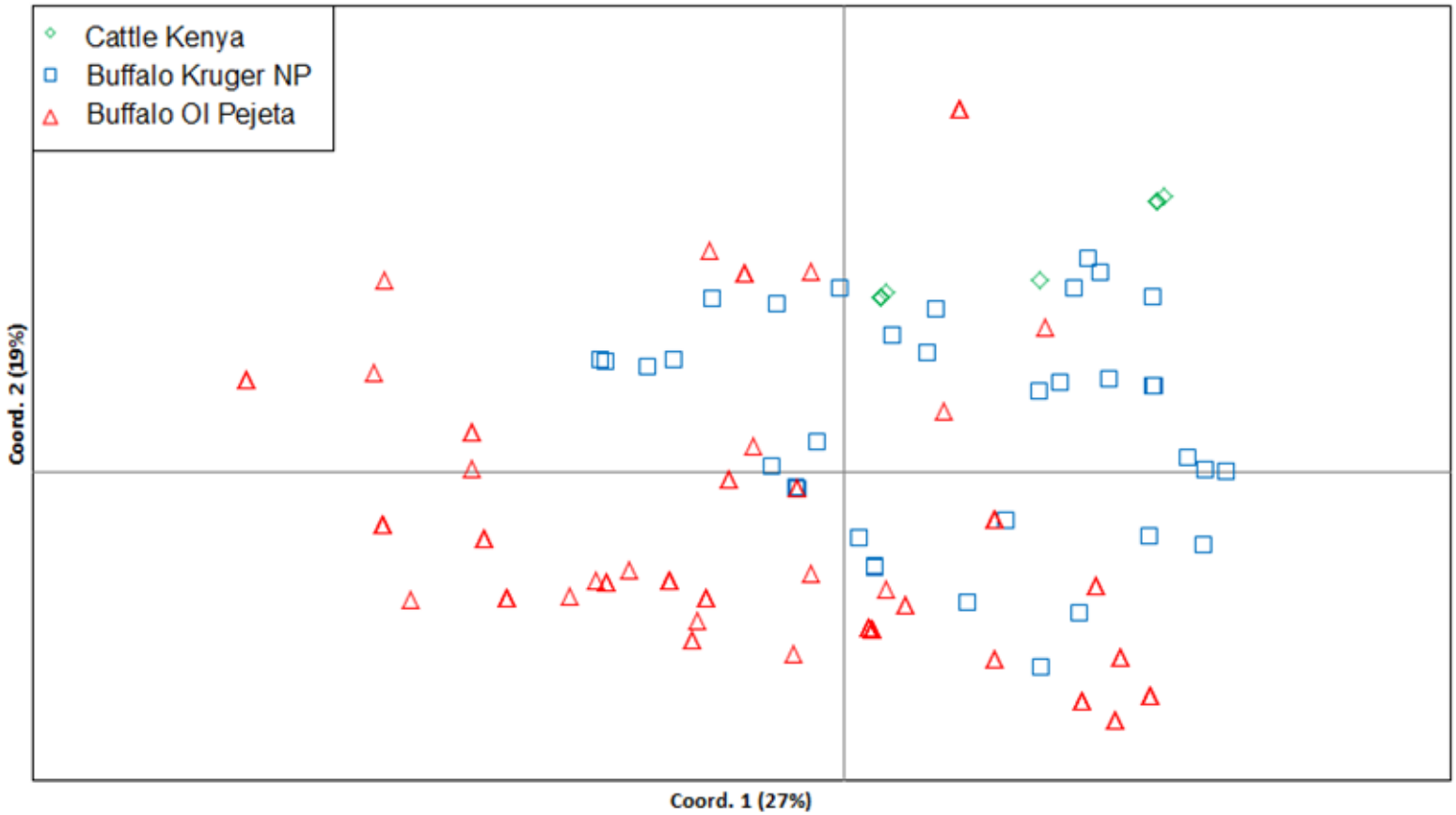

Figure 5.6 PCoA of the Tp1 Allelic diversity in *T. parva* in African buffalo and cattle
This diagram illustrates the relationship between *T. parva* present in African buffalo from the Kruger National Park, the Ol Pejeta conservancy and cattle from different regions in Kenya (Kilifi, Nyairo and Kakuzi) (Pelle et al., 2011)

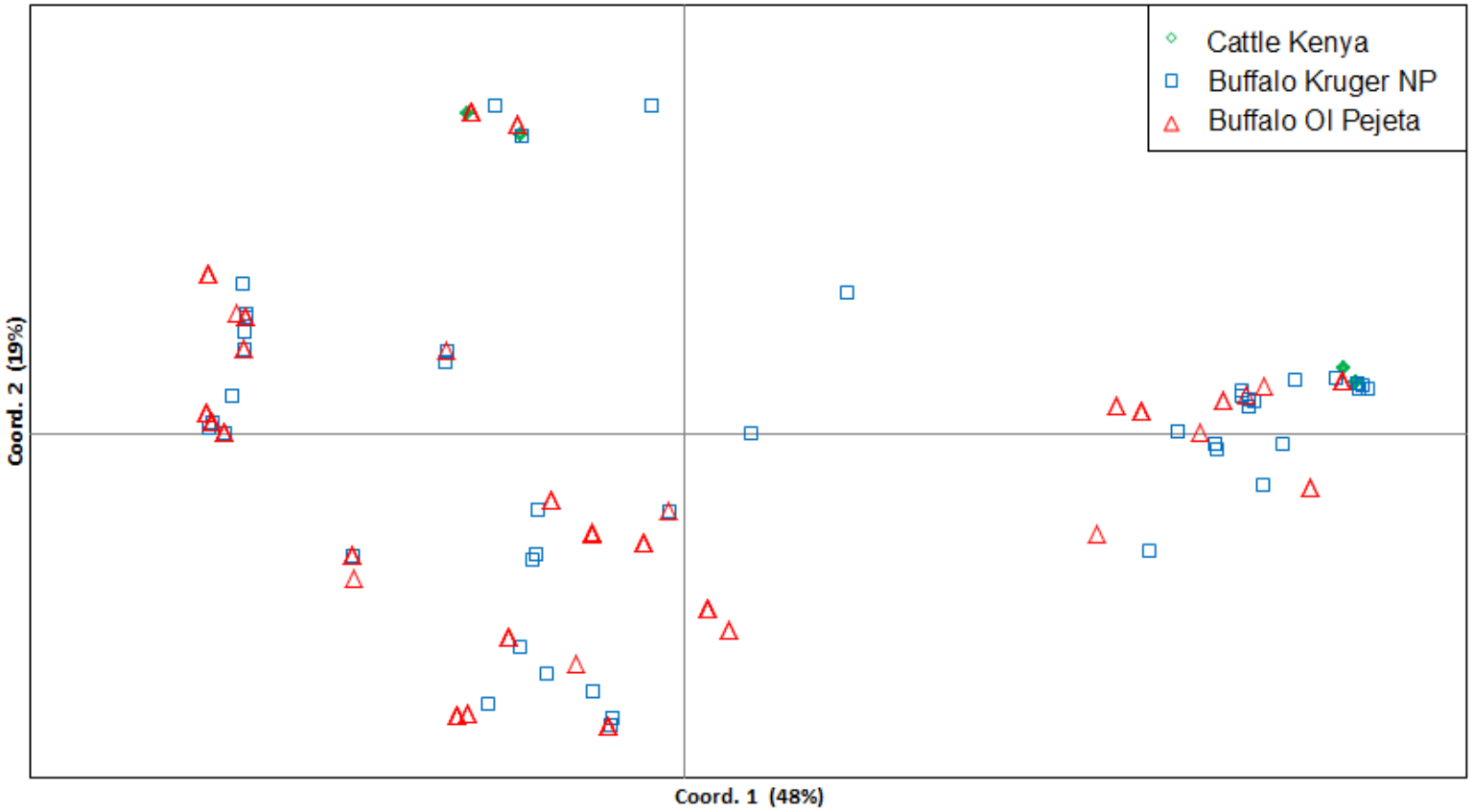

Figure 5.7 PCoA of Tp2 allelic diversity in *T. parva* in African buffalo and cattle

This diagram illustrates the relationship between *T. parva* present in African buffalo from the Kruger National Park, the Ol Pejeta conservancy and cattle from different regions in Kenya (Kilifi, Nyairo and Kakuzi) (Pelle et al., 2011)

Phylogenetic trees were constructed to further investigate the relationship between the different sequences obtained. Two exemplar phylogenetic trees are shown in Figures 5.8 and 5.9. Additional trees can be found in Appendix C. For each of the genes there appears to be a degree of sub-clustering between sequences obtained from African buffalo from the Kruger National Park and the Ol Pejeta conservancy. However, the bootstrap support of the constructed trees was very low, so no firm conclusions can be drawn on the phylogenetic relationship between sequences of the different parasites. The lack of bootstrap support of the constructed trees might be due to lack of true differentiation between the different parasite populations

.

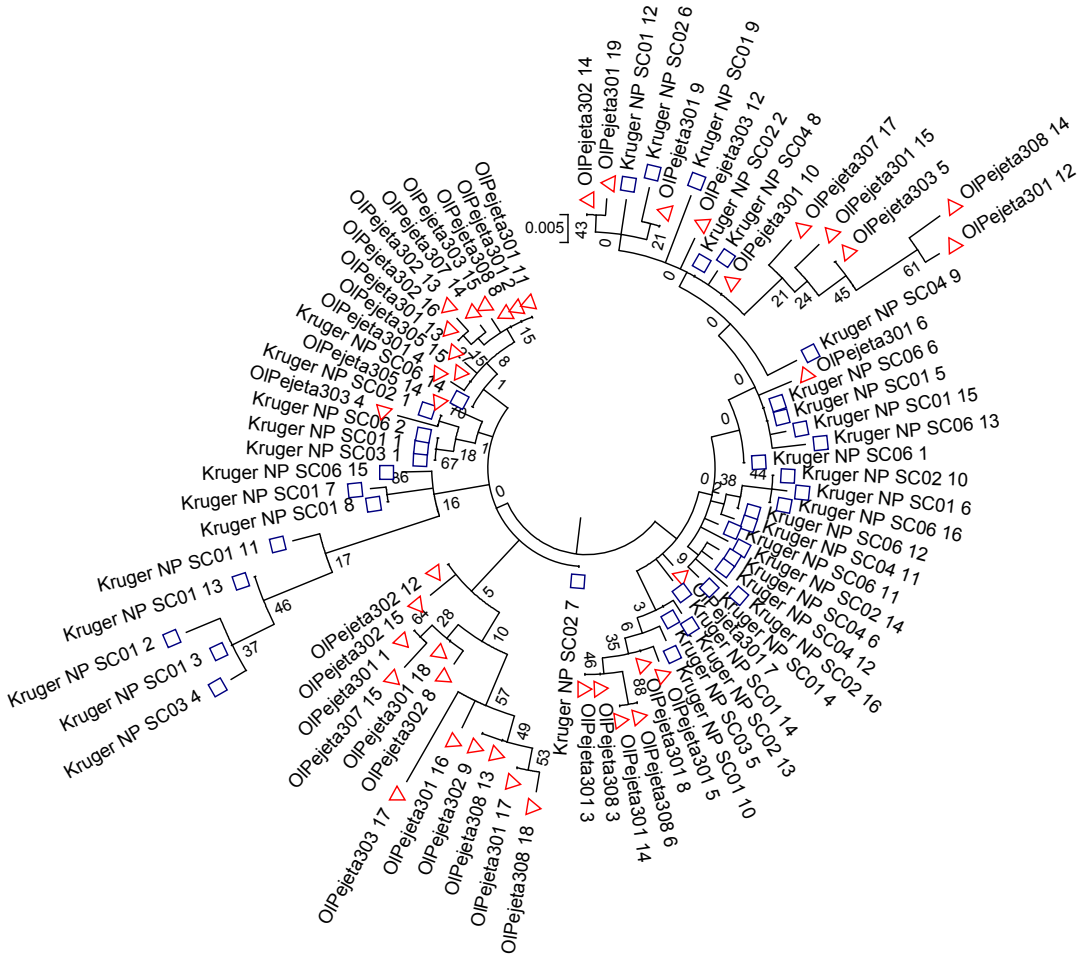

Figure 5.8 Phylogenetic tree constructed using buffalo Tp1 sequences
The tree was constructed using a maximum likelihood method using 1,000 bootstrap replicates. The bootstrap values indicate the degree of support for each of the nodes. Sequences were obtained from buffalo from the Kruger National Park and the Ol Pejeta conservancy, all unique sequences obtained within the Kruger National Park or the Ol Pejeta conservancy are displayed. The location (Ol Pejeta (Δ) or Kruger NP (□)), animal number and allele ranking are shown.

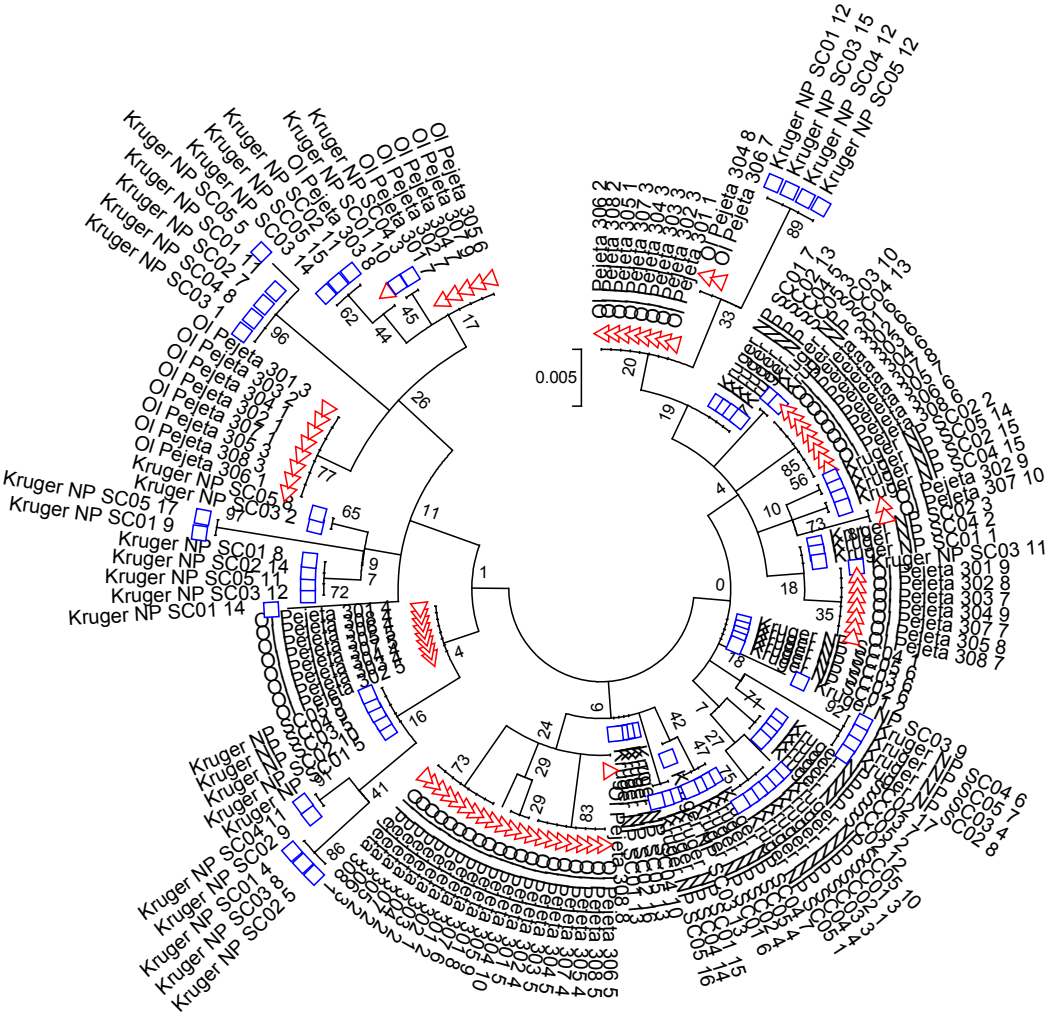

Figure 5.9 Phylogenetic tree constructed using buffalo Tp5 sequences
The tree was constructed using a maximum likelihood method using 1,000 bootstrap replicates. The bootstrap values indicate the degree of support for each of the nodes. Sequences were obtained from buffalo from the Kruger National Park and the Ol Pejeta conservancy. The location (Ol Pejeta (Δ) or Kruger NP (□)), animal number and allele ranking are displayed.

5.3.3 Antigenic diversity in buffalo-derived *T. parva*

A series of epitopes recognised by $CD8^+$ T cell responses have been identified within the proteins encoded by Tp1, Tp2, Tp4, Tp5 and Tp10. A large number of epitope variants was found for the epitopes within the Tp2 gene (see Table 5.3). Between 31 and 42 allelic variants were found at the amino acid level for the epitopes within the forward reads of this gene, some of which were shared between buffalo from the Kruger National Park and the Ol Pejeta conservancy. The alleles of the $Tp2_{49\text{-}59}$ epitope are detailed in Table 5.4. For the other genes, the antigenic diversity was more limited despite the high number of allelic variants at the nucleotide sequence level. Many nucleotide changes were found to be synonymous and did not lead to coding changes. No diversity was found for either $Tp5_{87\text{-}95}$ or $Tp10_{304\text{-}318}$, two alleles were found for $Tp4_{328\text{-}338}$ and three alleles were found for $Tp1_{214\text{-}224}$. The alleles for the $Tp1_{214\text{-}224}$ and $Tp4_{328\text{-}338}$ epitopes include the reference alleles found in the Muguga Cocktail. The two newly discovered variants of the $Tp1_{214\text{-}224}$ epitope have relatively subtle residue changes in the last two residues of the epitope compared to the reference sequence, i.e. $Tp1_{223}$ (M →I) and $Tp1_{224}$ (L→I), and this is shown in Table 5.3. Tp7 and Tp8 were only examined for one sample from a buffalo from the Kruger National Park (see Figure 5.1) and no epitope variants were found for $Tp7_{206\text{-}214}$ and $Tp8_{378\text{-}388}$ in this sample (data not shown)

Table 5.3 Epitope variants found in samples from buffalo from the Kruger National Park and the Ol Pejeta conservancy

Var.	$Tp1_{214-224}$	$Tp2_{27-37}$	$Tp2_{40-48}$	$Tp2_{49-59}$	$Tp4_{328-338}$	$Tp5_{87-95}$	$Tp10_{304-318}$
V1	VGYPKVKEEML	NDDELENLGML	DGFDKEKLF	KSSHGMGKIGR	TGASIQTTL	SKADVIAKY	TNNFNNPELIPVL
V2	VGYPKVKEEMI	NDDELKKLGML	DGFDKELLF	KSSHGMGKVGK	SGASIQTTL		
V3	VGYPKVKEEII	NDDELKRMGMV	DGFDKETLF	KSSHGMGKVGR			
V4		NDDELNKLGML	DGFDRDALF	KSSHGMGMVGR			
V5		NDDELRKLGMV	DGFDREKLF	KSSKAMTTTGK			
V6		NDEELENLGML	DGFDRELLF	KSSKGLTAVGK			
V7		SDDDLKKLGML	DGFDRQRLF	KSSKGMTAVGK			
V8		SDDDLNKLGML	DGFEKERLF	KSSKSMGIVGR			
V9		SDDELDTLGML	DKLDKDKLF	KTSHSMGMIGK			
V10		SDDELETLGML	DNLDKDKLF	KTSKAMTATGR			
V11		SDDELKKLGML	DNLNKERLF	KTSKAMTMTGR			
V12		SDDELKKMGLI	EGFDKEKLF	KTSKAMTSTGR			
V13		SDDELNKLGML	EGFEKEKLF	KTSKGMTEVGK			
V14		SDDELRKLGML	EGLDMEALF	KTSKGMTKVGR			
V15		SDDELRKLGMV	GNFDRELLF	KTSKSMGQVGR			
V16		SDDELSKMGML	GNFNKETLF	LSSHGMGKVGK			
V17		SDEELDTLGML	PDFEKERLF	LTSHGMGEVGK			
V18		SDEELESLGML	PDLDKNRLF	LTSHGMGKIGR			
V19		SDEELETLGML	PDPDKERLF	LTSHGMGKVGR			
V20		SDEELKELGML	PDPDKETLF	LTSHGMGRIGR			
V21		SDEELKKMGML	PDPNKDRLF	LTSKAMGAVGR			
V22		SDEELKKLGML	PDPDKQRLF	LTSKAMMTVGK			
V23		SDKELDTLGML	PDPNKERLF	LTSKAMSTVGK			
V24		SDNELDTLGLL	PDPVKERLF	LTSKAMTHVGK			
V25		SDSELETLGML	PGFDKELLF	LTSKAMTHVGR			
V26		SEAELRKMGMI	PGFDKEVLF	LTSKAMTTVGK			
V27		SEDELKKMGMV	PHPDKERLF	LTSKAMTTVGR			
V28		SEEELRKMGMI	PNPDKEKLF	LTSKCMTEVGR			
V29		SHEELKKLGML	PNPDKEKLF	LTSKGMAAVGR			
V30		TDEELENLGML	QDPDKETLF	LTSKGMTKIGK			
V31		TDEELREMGMI	SNFDRESLF	LTSKGMTKIGR			
V32		TEDELKKLGML	TDPVKERLF	LTSKSMMTVGK			
V33		TEDELKKLGMV		LTSKSMSEVGR			
V34		TEDELKNMGLI		MTSKAMTATGR			
V35		TEEELKKLGMV		RSSHGMGKVGK			
V36		TEEELKRMGML					
V37		TEEELKRMGMV					
V38		TEEELRKIGMV					
V39		TEEELRKLGMA					
V40		TEEELRKLGMV					
V41		TNEELKKLGMV					
V42		TQEELRKLGMV					
V43		TVEELRKMGMV					

Epitope variants highlighted in grey were found in a previous study by Pelle et al (2011).

Table 5.4 Variants of $Tp2_{49-59}$ epitope found in African buffalo from the Kruger National Park, The Ol Pejeta conservancy and in a previous Kenyan study.

Variant	Kruger National Park	Ol Pejeta	Kenyan study[1]
1	KSSHGMGKIGR	KSSHGMGKIGR	KSSHGMGKIGR
2	KSSHGMGKVGK	KSSHGMGKVGK	KSSHGMGKVGK
3	KSSHGMGKVGR	KSSHGMGKVGR	KSSHGMGKVGR
4	KSSHGMGMVGR		
5	KSSKAMTTTGK	KSSKAMTTTGK	KSSKAMTTTGK
6		KSSKGLTAVGK	
7		KSSKGMTAVGK	KSSKGMTAVGK
8	KSSKSMGIVGR	KSSKSMGIVGR	
9		KTSHSMGMIGK	KTSHSMGMIGK
10	KTSKAMTATGR	KTSKAMTATGR	KTSKAMTATGR
11		KTSKAMTMTGR	KTSKAMTMTGR
12	KTSKAMTSTGR		
13	KTSKGMTEVGK	KTSKGMTEVGK	KTSKGMTEVGK
14	KTSKGMTKVGR	KTSKGMTKVGR	KTSKGMTKVGR
15	KTSKSMGQVGR		
16		LSSHGMGKVGK	
17	LTSHGMGEVGK		
18		LTSHGMGKIGR	
19	LTSHGMGKVGR		
20	LTSHGMGRIGR	LTSHGMGRIGR	LTSHGMGRIGR
21	LTSKAMGAVGR		
22	LTSKAMMTVGK	LTSKAMMTVGK	
23	LTSKAMSTVGK	LTSKAMSTVGK	LTSKAMSTVGK
24	LTSKAMTHVGK	LTSKAMTHVGK	
25	LTSKAMTHVGR		
26	LTSKAMTTVGK	LTSKAMTTVGK	LTSKAMTTVGK
27	LTSKAMTTVGR		
28	LTSKCMTEVGR		
29	LTSKGMAAVGR	LTSKGMAAVGR	
30		LTSKGMTKIGK	
31	LTSKGMTKIGR		LTSHGMGKIGR
32	LTSKSMMTVGK	LTSKSMMTVGK	LTSKSMMTVGK
33	LTSKSMSEVGR	LTSKSMSEVGR	LTSKSMSEVGR
34		MTSKAMTATGR	MTSKAMTATGR
35	MTSKAMTMTGR	MTSKAMTMTGR	
36	RSSHGMGKVGK	RSSHGMGKVGK	RSSHGMGKVGK
37			AASHGLGKVGK
38			KSSHGMGEVGK
39			KSSKGMGKVGK
40			KSSKGMTKVGK
41			KTSNGMTKVGK

[1] Pelle et al (2011)

5.4 Discussion

5.4.1 Diversity in *T. parva* in African buffalo from the Kruger National Park and the Ol Pejeta conservancy

The multiplicity of infection was high in both buffalo from the Kruger National Park and the Ol Pejeta conservancy. Using deep multilocus sequence typing of Tp1, Tp2, Tp4, Tp5, Tp6 and Tp10, an average of between 8.8 and 21.5 alleles were found in individual animals for each of the loci for buffalo from the Kruger National Park and between 6.8 and 17.6 alleles for buffalo from the Ol Pejeta conservancy. In a previous study, a maximum of nine alleles was found at a single locus using satellite DNA markers in individual samples from buffalo from national parks in Uganda (Oura et al., 2011a). While it is possible that the buffalo sampled in Uganda simply had a lower multiplicity of infection, it is likely that the lower number of alleles per locus identified in that study was due to the use of a different genotyping technology, both in terms of the locus type and the method of allele-calling. The multi-locus typing system used in this study is capable of detecting single nucleotide differences between alleles and thus provides a very high level of resolution.

It has been hypothesised that *T. parva* was introduced into South Africa at the beginning of the twentieth century through the importation of cattle from East Africa for the purpose of restocking, following the rinderpest epidemic at the end of the 19th century (Mbizeni et al., 2013; Norval et al., 1991a; Norval et al., 1992). It has been argued that the local *T. parva* population was decimated at the end of the 19th century due to the death of considerable numbers of cattle and buffalo during the rinderpest epidemic, thus causing a severe bottleneck of the *T. parva* population. The imported cattle were hypothesised to carry *T. parva* from East Africa, thereby re-introducing East Coast fever to the region with subsequent transmission into African buffalo (Mbizeni et al., 2013). Current evidence indicates that *T. parva* in cattle in East Africa is much less diverse than that in African buffalo (Pelle et al., 2011). Given the high multiplicity of infection and genotypic diversity of *T. parva* within individual South African buffalo revealed in the study, the hypothesis that African buffalo became infected with *T. parva* after introduction from cattle from East Africa appears unlikely.

A simpler explanation is that the remaining South African buffalo population acted as a reservoir for the local *T. parva* population. The size of the buffalo population in the Kruger National Park alone, as measured by the 2009 census, was 27,000. Even if the buffalo population fell to a few thousand following the rinderpest epidemic, the effective size of the parasite population would still have been very large and sufficient to ensure the maintenance of a parasite population with a high level of genotypic diversity. Analysis of the major histocompatibility complex gene, DRB3, in buffalo revealed high allelic diversity for buffalo from four national parks in Africa, including the Kruger National Park. It was hypothesised that following decimation of the buffalo population in South Africa because of the rinderpest epidemic, the population was restored by the rapid re-colonisation of buffalo from unaffected to affected areas (Wenink et al., 1998). It seems likely, therefore, that the re-colonising buffalo would still be infected with the vestiges of the local population and this would re-establish in the country.

Over six loci, the mean distance between alleles sequences was surprisingly similar in each of the three sample sets: (1) buffalo from Ol Pejeta and Kruger National Park combined, (2) buffalo from the Ol Pejeta conservancy and (3) buffalo from the Kruger National Park. Figure 5.2 illustrates the fact there was a good agreement over the three 'populations' and taken together with the relatively small error bars, this data demonstrates converging results, which in turn suggests that each population was sampled with adequate depth to meaningfully calculate this metric. It had been expected that sequences obtained from *T. parva* parasites from buffalo in the Kruger National Park would display less diversity than that those from Eastern Africa buffalo (Mbizeni et al., 2013) for two reasons: firstly, the suggested bottleneck due to the rinderpest as mentioned above and secondly, differences in climatic conditions and tick behaviour (Mbizeni et al., 2013). Ticks have adapted to the harsh climatic conditions during the dry season in South Africa by developing the ability to go into 'diapause', a predetermined state of low activity and metabolism. This state is under neurohormonal control and is initiated, maintained and terminated by changes in daylight length and other factors (Young et al., 1994). As a consequence of this behaviour, *Rhipicephalus appendiculatus* completes only one life-cycle within a

calendar year. Parasites must be maintained long enough in the mammalian host to allow ticks to pick up infection following diapause (Young et al., 1994) and not all *T. parva* genotypes are capable of inducing a long-term carrier state in the cattle host (Skilton et al., 2002). However, buffalo can carry *T. parva* for long periods of time, up to 888 days (Grootenhuis et al., 1987), which will allow parasite diversity to be maintained in the buffalo host over the dry season. When ticks feed on buffalo after their diapause, they will readily become infected with a mixture of parasites. Seasonality of ticks thus does not necessarily hamper the maintenance of parasite diversity in the buffalo host.

5.4.2 Sub-structuring of *T. parva* in Kruger National Park and the Ol Pejeta conservancy

A degree of geographical sub-structuring was observed between the parasites in the Kruger National Park and the Ol Pejeta conservancy. This sub-structuring was most evident for the Tp1 gene, based on both phylogenetic and PCoA analysis. This suggests the parasite populations have differentiated and the simplest explanation is that of genetic drift.

The estimated average genetic 'distance' (as measured by the number of nucleotide substitutions per site from averaging over all sequence pairs) between any two alleles in the Kruger National Park and within the Ol Pejeta conservancy was very similar for all six genes examined, i.e. Tp1, Tp2, Tp4, Tp6 and Tp10. Principal co-ordinate analysis using Tp2, Tp4, Tp5, Tp6 and Tp10 did not reveal any obvious partitioning of the dataset. Phylogenetic analysis also did not provide evidence for the parasite populations in the Kruger national park and the Ol Pejeta conservancy evolving separately. Sequences from the two parks can be found in the same clades. However, many of the identified alleles were 'private' to either the African buffalo from Ol Pejeta or to African buffalo from the Kruger National Park. The fact that several of these 'private' alleles were identified in each location suggests that the parasite populations have differentiated, despite the lack of discrimination in the PCoA analysis. Identical sequences were found in several buffalo from one area, which suggests a degree of homogeneity in local populations, i.e. little evidence of a high rate of novel mutations in the locality. The diversity observed has probably been

around for a long time. This pattern may be consistent with a lot of 'ancient' diversity in combination with present day geographical sub-structuring, which may be at a fine level.

5.4.3 Antigenic diversity in *T. parva* maintained in African buffalo

For buffalo-derived *T. parva*, a moderate to very high degree of allelic polymorphism was observed for the genes targeted for deep multi-locus sequence typing even within single buffalo. This diversity within an individual animal allows feeding ticks to pick up several genotypically distinct parasites that can sexually recombine in the tick to give rise to novel parasite genotypes. A large number of allelic variants of Tp1 and Tp2 have previously been identified in isolates from buffalo (Pelle et al., 2011). In the present study, additional allelic variants were identified and this was anticipated as the population was sampled in greater depth. The diversity in the epitope regions within the Tp1 and Tp2 gene was high in buffalo-derived *T. parva* from both the Kruger National Park and the Ol Pejeta conservancy. Many of the allelic variants of the epitopes in the Tp2 gene were shared between Ol Pejeta and the Kruger National Park, but a number of private alleles were found in each area.

Although many variants of the Tp1 gene were found, the most notable feature of this variation was perhaps the difference in copy number of the proline-rich repeats towards the end of the reads. Only three alleles of the $Tp1_{214\text{-}224}$ epitope were found in both Ol Pejeta and the Kruger National Park. Only one allele was detected for the $Tp5_{87\text{-}95}$ and $Tp10_{304\text{-}318}$ epitopes and two alleles were found for the $Tp4_{328\text{-}338}$ epitope. Most nucleotide changes in these genes did not result in amino acid changes when translated into amino acid sequences. This level of conservation at the amino acid level suggests that the encoding nucleotide sequences are largely under purifying selection and suggest that conservation of these genes is required for parasite fitness. Moreover, the data indicates that these epitopes have not been influenced either by positive diversifying or balancing selection and this in turn suggests they are not under strong immune selection by the African buffalo host.

Animals immunised by infection and treatment using the Muguga Cocktail are not in every case protected against challenge with buffalo-derived *T. parva* parasites. The diversity within the Muguga Cocktail is known to be limited, based on DNA satellite typing (see Figure 4.3) and deep multi-locus sequence typing (see Figure 4.6). Although the allelic variants of the Tp2 gene differed somewhat between the Kruger National Park and the Ol Pejeta conservancy, many allelic variants were detected in each population which were very different from the Muguga alleles of Tp2. Previous work has shown that $CD8^+$ T cell clones derived from animals immunised with *T. parva* Muguga could not recognise peptides with a substitution at P8 of the $Tp2_{49\text{-}59}$ epitope (Connelley et al., 2011). In this study, a substitution with respect to the Muguga allele at P8 was found for many of the $Tp2_{49\text{-}59}$ variants (see Table 5.4) and it is therefore likely they will not be recognised by CD8 T cells induced by the *T. parva* Muguga stock. For the other epitopes in the Tp2 gene, the relationship of allelic polymorphism to immune recognition is unknown. However, it seems likely that the high degree of polymorphism in the epitopes and specificity of $CD8^+$ T cell responses in immunity will prevent or reduce recognition by $CD8^+$ T cells, leading to failure of the Muguga Cocktail to protect against buffalo-derived parasites.

In contrast, epitopes in other genes were relatively conserved among parasite genotypes and should these genes be investigated in the future, it may be concluded that immunisation using a stabilate with a more limited diversity (such as the Muguga Cocktail) can be used for successful vaccination against buffalo-derived *T. parva*. The conservation of some of the antigens recognised by bovine CD8+ T cells is promising for the design of a next-generation vaccine. The use of conserved genes would overcome the issue of genetic diversity between different parasites and would be predicted to lead to broad immunity.

Chapter 6: Investigating the role of antigenic polymorphism in immune protection

6.1 Introduction

Immunity against *Theileria parva* is mediated by a $CD8^+$ T cell response (McKeever et al., 1994) and the strain-specificity of the response has been shown to correlate with the strain-specificity of immunity upon parasite challenge (Taracha et al., 1995a). A number of *T. parva* antigens recognised by $CD8^+$ T cells from cattle immunised by ITM have been identified (Graham et al., 2007a; Graham et al., 2006). It has been shown in animals of defined MHC I haplotypes that antigens recognised by the $CD8^+$ T cell response are determined by the host MHC phenotype and a large component of responding $CD8^+$ T cells (up to 75%) is specific for a single antigen and usually only one or two epitopes (MacHugh et al., 2009a). Such immunodominance is a common feature in $CD8^+$ T cell responses in viral infections; $CD8^+$ T cell responses are preferentially directed towards one or a few dominant epitopes despite the presence of a T cell repertoire with multiple specificities, which is potentially capable of responding to a wide range of peptides (Yewdell, 2006; Yewdell and Bennink, 1999).

Immunodominance results from a complex combination of factors involved in antigen processing and antigen recognition. The relative importance of the different factors is likely to differ for different pathogen/host combinations (Yewdell, 2006; Yewdell and Bennink, 1999; Yewdell et al., 2003). The mechanisms that determine immunodominance in $CD8^+$ T cell responses against *T. parva* are unknown, but the immunodominant nature of these responses has important implications for the design of effective vaccines (MacHugh et al., 2009a). When $CD8^+$ T cell responses are focused on only one or a few antigens, polymorphism in the epitopes can result in failure of $CD8^+$ T cells to recognise heterologous parasite strains, resulting in susceptibility to challenge with such parasites (MacHugh et al., 2009a). At least some of the identified *T. parva* antigens show a high degree of sequence polymorphism (Pelle et al., 2011).

The effect of polymorphism on MHC I presentation and T cell recognition has been investigated for two of the identified epitopes recognised by $CD8^+$ T cell responses

in Holstein cattle. Using specific $CD8^+$ T cell clones and synthetic peptides containing alanine substitutions at each amino acid residue, it was possible to identify MHC I binding residues and residues that affected T cell recognition. Some residue changes had little effect on MHC I binding or T cell recognition, whereas substitution at other residues had a detrimental effect on both types of interaction (Connelley et al., 2011; Macdonald et al., 2010). Animals homozygous for the A10 MHC I haplotype and immunised with *T. parva* Muguga have been shown to mount an immunodominant $CD8^+$ T cell response against epitopes in the Tp2 antigen (MacHugh et al., 2009a). The Tp2 antigen is highly polymorphic; two epitopes, $Tp2_{49\text{-}59}$ and $Tp2_{98\text{-}106}$ both presented by MHC I allele 2*01201, are recognised (Graham et al., 2008a). Studies with single alanine substituted peptides and peptides representing allelic variants of the $Tp2_{49\text{-}59}$ epitope using specific $CD8^+$ T cell clones, demonstrated that the residue at position eleven influences MHC I binding, that the residue eight is essential for T cell recognition and that several other residues have variable effects on T cell recognition (Connelley et al., 2011).

It has been proposed that sequence polymorphism in *T. parva* antigens combined with immunodominance in $CD8^+$ T cell responses allows particular parasites to escape $CD8^+$ T cell responses *in vivo*. To address this hypothesis, the current study sought to determine whether immunisation with a clonal population of *T. parva* that induces $CD8^+$ T cell responses to a dominant polymorphic antigen will result in selective breakthrough of parasite genotypes expressing alternate alleles of this antigen following challenge with an antigenically diverse *T. parva* isolate, i.e. containing the homologous allele and heterologous alleles of the dominant antigen. Animals expressing the MHC I A10 haplotype were used, which (as discussed above) recognises two dominant epitopes in the reference sequence of the Tp2 antigen, $Tp2_{49\text{-}59}$ and $Tp2_{98\text{-}106}$ (Graham et al., 2008a). The initial aim of the study was to identify a challenge stabilate containing allelic variants of $Tp2_{49\text{-}59}$ and $Tp2_{98\text{-}106}$, including the reference allele. The selected stabilate was then used to challenge animals immunised with the Muguga stock and the breakthrough population of parasites in the immunised animals, along with parasites in control animals, were analysed by 454 amplicon sequencing of the Tp2 gene. Figure 6.1 illustrates the experimental design.

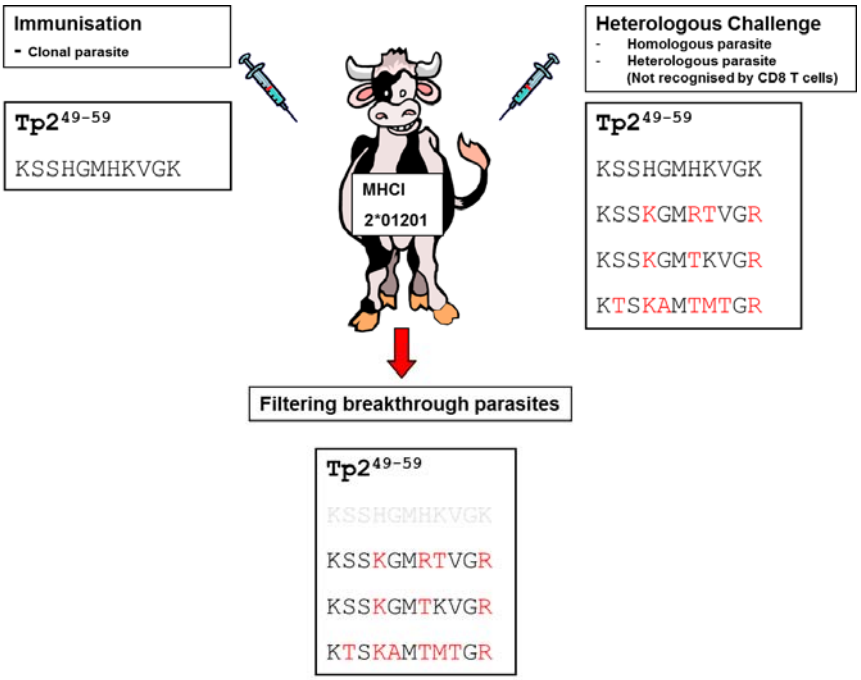

Figure 6.1 Conceptual design of challenge experiment, illustrating the anticipated 'knock-down' of homologous genotypes

Only the Tp2 $_{49\text{-}49}$ epitope is shown for illustration purposes.

6.2 Material and methods

6.2.1 Sequence analysis of the Tp2 gene

The Tp2 gene was PCR amplified from samples of DNA prepared from two stabilates of *T. parva* 3081 and 4110 produced from infected buffalo. The PCR products were cloned into the P-GEM-T vector and plasmid clones were purified using the PureYield^{TM} Plasmid Miniprep System (Promega Corporation, USA) as previously described in Chapter 2. The Tp2 gene was also amplified from nine cloned *T. parva*-infected cell lines derived from a cell line infected with the Marikebuni stock (A3, A7, B12, F31, F44, F53, I8, I38 and J17). The purified plasmids or PCR products were sent to Durham Genomics for sequencing using Sanger technology. Chromatograms were converted into FASTA sequences and alignment was undertaken using MUSCLE in the MEGA5 software package (Edgar, 2004; Tamura et al., 2011). The nucleotide sequences were translated into amino acid sequences and the sequences for the epitopes $Tp2_{49\text{-}59}$ and $Tp2_{98\text{-}106}$ were extracted.

6.2.2 Recognition of Tp2 variants present in *T. parva* Marikebuni

A $CD8^+$ T cell line from animal 011 was tested for specificity with respect to the Tp2 epitope variants found in the *T. parva* Marikebuni stock using a four-hour ^{111}Indium release cytotoxicity assay. $CD8^+$ T cells from this animal have been shown to recognise both epitopes in the reference sequence of theTp2 antigen (MacHugh et al., 2009b). *Theileria parva* Muguga and two *T. parva* Marikebuni lines (B12 and F53) containing epitope variants of $Tp2_{49\text{-}59}$ and $Tp2_{98\text{-}106}$ were used as targets at effector-to-target ratios between 20:1 and 0.6:1 (see Chapter 2). The recognition of peptide was examined using autologous *T. annulata*-infected cells as antigen presenting cells and peptide concentrations ranging from 100 ng/ml to 0.01 ng/ml.

6.2.3 *In vivo* challenge experiment

All experimental work was performed with the help of Thomas Njoroge and Stephen Mwaura, ILRI, Kenya. Seven animals carrying MHC I haplotype A10 were selected as described in Chapter 2 (see Table 6.1 for further information). Animals were transported to the ILRI farm and received acaricide treatment on arrival, three days after arrival and subsequently once a week. Animals were tested for evidence of exposure to *Theileria parva, Theileria mutans, Anaplasma* and *Babesia bigemina* using parasite species-specific ELISAs. Samples for ELISA were taken on the day of arrival, day of immunisation, day of challenge and six weeks after challenge. The ELISAs were performed by Alice Njeri, ILRI, Kenya.

Table 6.1 Animals used for the challenge experiment

Animal number	Breed	Sex	Age	ITM vaccinated
BG040	Friesian/Ayrshire	M	1 year	yes
BG041	Friesian	M	7 months	yes
BG044	Friesian	M	9 months	no
BF080	Ayrshire	M	2.3 years	yes
BF083	Friesian	M	2.5 years	no
BF085	Ayrshire	M	2.5 years	yes
BF086	Friesian	M	2.5 years	yes

Animals were immunised by infection and treatment (ITM) using *T. parva* Muguga stabilate 3308.

Two weeks after arrival at ILRI, five of the animals were immunised by infection and treatment with *T. parva* Muguga Stabilate 3308. Ten weeks after immunisation, these animals, along with two unimmunised controls, were challenged with buffalo-derived *T. parva* stabilate 3081 at a dilution of one in twenty. Animals were monitored closely after challenge. Briefly, animals were inspected three times a day and temperature and lymph node enlargement were monitored daily. Smears of lymph node aspirates and blood were taken daily when body temperature increased and/or when lymph node enlargement was observed. Animals were treated with buparvaquone when they developed severe disease.

6.2.4 Tp2 specificity of CD8+ T cell responses of immunised animals

The specificity of $CD8^+$ T cell responses of immunised animals was investigated using a ^{51}Chromium release cytotoxicity assay and an *ex vivo* ELISPOT assay. $CD8^+$ T cells were cultured by stimulation with irradiated autologous *T. parva*-infected cells as previously described in Chapter 2 and specificity was then determined by assaying cytotoxic activity using autologous *T. parva* cell lines, an MHC I-mismatched *T. parva* cell line and an A10-homozygous cell line (592 *T. parva* Muguga) as target cells. The *ex vivo* ELISPOT assay was performed with the help of Rosemary Saya, ILRI, Kenya. Purified $CD8^+$ cells were used as effector cells and purified $CD14^+$ cells pulsed with peptides as target cells. The peptides used were *T. parva* Muguga $Tp2_{49\text{-}59}$ KSSHGMGKVGK and $Tp2_{98\text{-}106}$ QSLVCVLMK.

6.2.5 Analysis of allelic variants of Tp2 *ex vivo*

DNA was extracted from blood and lymph node samples obtained from the animals undergoing clinical reactions upon challenge, using the DNeasy Blood and Tissue Kit (Qiagen, UK). Amplicons of the Tp2 gene were generated by PCR amplification using fusion primers and resulting PCR products were pooled and submitted for 454 Sequencing (see Chapter 3). Sequence reads were partitioned into those belonging to individual samples after which the bioinformatic pipeline described in Chapter 3 was applied in order to account for 'noise' generated through PCR and the sequencing process.

6.3 Results

6.3.1 Assessment of *T. parva* Marikebuni as the challenge parasite

For the proposed challenge, a stabilate was required that contained the Muguga Tp2 allele in addition to allelic variants of Tp2 that are likely not to be recognised by $CD8^+$ T cells induced by immunisation with Muguga. Ideally, a challenge stabilate would represent a genotypically diverse parasite population and the Tp2 reference allele would be present on a number of different genotypes. This would mean that any immune-selection acting upon other undefined antigen-encoding loci would not result in appreciable selection at the Tp2 locus and confound the planned analysis.

The Marikebuni stock is genotypically diverse as judged by satellite DNA typing of clones derived from CTVM stabilate 72 (Katzer et al., 2010) and has previously also been shown to be antigenically heterogeneous (Goddeeris et al., 1990). Sequencing of the Tp2 gene in nine cloned *T. parva* Marikebuni lines, shown previously by satellite DNA typing to be genotypically distinct (Katzer et al., 2011), demonstrated the presence of two alleles; seven of the clones possessed the reference Muguga allele, while the remaining two clones had a variant allele, which differed from the $Tp2_{49\text{-}59}$ and $Tp2_{98\text{-}106}$ reference sequences at two and three amino acid residues respectively (see Table 6.2).

Table 6.2 Sequences of $Tp2_{49\text{-}59}$ and $Tp2_{98\text{-}106}$ obtained from nine clonal *T. parva* Marikebuni infected cell lines

Cell line	$Tp2_{49\text{-}59}$	$Tp2_{98\text{-}106}$
T. parva Muguga	KSSHGMGKVGK	QSLVCVLMK
T. parva Marikebuni A3, A7, F31, F44, I8, I38, J17	KSSHGMGKVGK	QSLVCVLMK
T. parva Marikebuni B12, F53	KSSHGMGK**I**G**R**	QS**I**VCV**VS**K

An un-cloned $CD8^+$ T cell line available from an immune animal, homozygous for MHC I A10 (Animal 1011) and known to recognise Tp2 (MacHugh et al., 2009a), was used to determine whether the allelic variants of the Tp2 epitopes in these cloned

lines are recognised by Muguga-induced $CD8^+$ cells. The $CD8^+$ T cell line recognised autologous *T. parva* Muguga but not an MHC I-mismatched *T. parva* line. One of the clonal *T. parva* Marikebuni lines (F53) was lysed, but at a lower percentage than the *T. parva* Muguga line. The other *T. parva* Marikebuni line (B12) was not lysed, even at higher effector-to-target ratios (see Figure 6.2). The absence of killing of the B12 line indicates that the Tp2 allelic variant is not recognised efficiently by the Muguga-induced $CD8^+$ T cell response. Differential killing of the two clonal *T. parva* Marikebuni cell lines, which have identical sequence of the Tp2 gene, could be due to recognition of another gene product that is conserved between F53 and Muguga but has a different allele in B12.

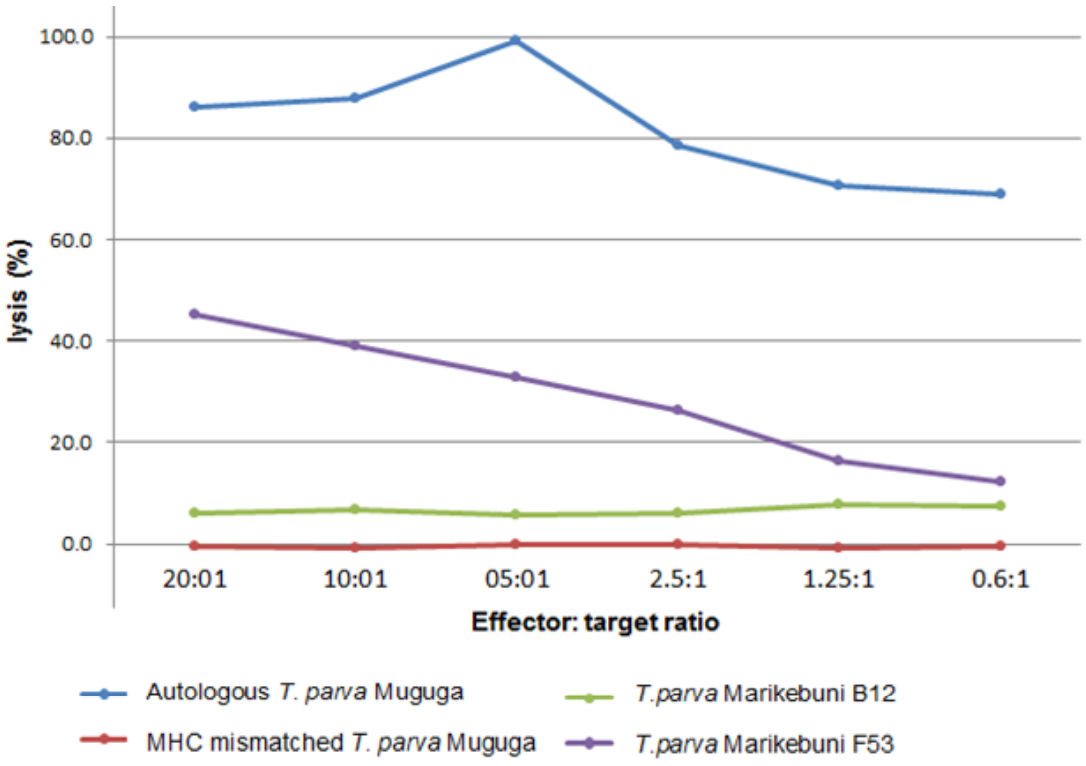

Figure 6.2 Specificity of CD8 T cells from animal 1011 tested against *T. parva*-parasitised cell lines that **carry different allelic variants of the Tp2 gene**
Four cell lines were assayed by Indium release assay: autologous and MHC I mismatched *T. parva* Muguga and two MHC I matched *T. parva* Marikebuni parasitised cell lines (B12 and F53) both of which carry an allelic variant of the Tp2 gene (See Table 6.2).

The uncloned $CD8^+$ T cell line was able to recognise peptides representing both the *T. parva* Muguga and the *T. parva* Marikebuni alleles of the $Tp2_{49\text{-}59}$ epitope, although the latter was recognised less efficiently (see Figure 6.3). This reduced efficiency may account for the apparent lack of recognition of the Marikebuni-infected cells. Nevertheless, in view of the observed cross-reactivity the *T. parva* Marikebuni stock was considered unsuitable for the proposed challenge experiment.

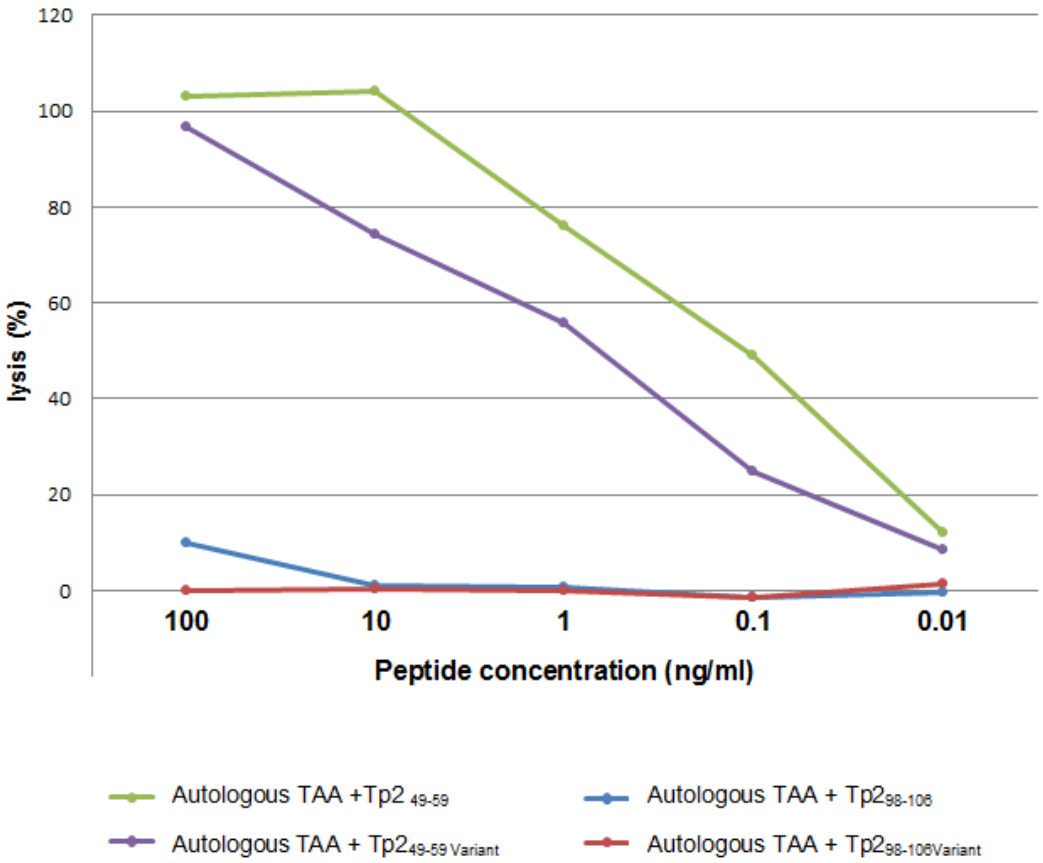

Figure 6.3 Specificity of CD8 T cells from animal 1011 tested against target cells pulsed with different concentrations of Tp2 peptide
Specificity was assayed by Indium relese assay. Peptides were presented by an autologous *T. annulata* Ankara parasitised cell line (TAA). Peptide sequences were reference Tp2 $_{49\text{-}59}$: KSSHGMGVGK, Tp2 $_{49\text{-}59\ Variant}$: KSSHGMGKIGR ,reference Tp2 $_{98\text{-}106}$: QSLVCVLMK and Tp2 $_{98\text{-}106\ variant}$ QSIVCVVSK.

6.3.2 Assessment of buffalo-derived *T. parva* as the challenge parasite

Subsequently, two buffalo stabilates (Stabilate 3081 and 4110) were examined for their potential to be used as a challenge stabilate. Buffalo are known to have a high multiplicity of infection with parasites that are genetically and antigenically diverse and extensive sequence polymorphism has been found in alleles of the Tp2 gene among buffalo-derived *T. parva* (Pelle et al., 2011); data presented in Chapter 4). The Tp2 gene of these two buffalo-derived stabilates was amplified and sequenced after cloning in the P-GEM-T vector. Both stabilates showed a high degree of allelic diversity at the protein level; twelve and nine allelic variants of the Tp2 gene were found in fourteen and eleven sequences obtained by Sanger sequencing, respectively, from stabilates 3081 and 4110; eleven variants of both $CD8^+$ T cell epitopes were found in stabilate 3081, while stabilate 4110 contained nine allelic variants of $Tp2_{49\text{-}59}$ and eleven and eight variants for $Tp2_{98\text{-}106}$ (see Tables 6.3 and 6.4). The *T. parva* Muguga (i.e. reference) allele of $Tp2_{49\text{-}59}$ and $Tp2_{98\text{-}106}$ was found to be present in both stabilates in sequences obtained by Sanger sequencing. A single Muguga allelic sequence was obtained for both stabilates, out of a total of twelve 4110 sequences and fourteen 3081sequences.

The allelic diversity of Tp2 in Stabilate 3081 was also examined using Roche 454 amplicon sequencing. Six allelic variants of $Tp2_{49\text{-}59}$ and twelve allelic variants of $Tp2_{98\text{-}106}$ were found corresponding to 481 and 603 reads respectively (see Figure 6.4). Allelic variants that were found by both Sanger and Roche 454 amplicon sequencing are denoted by an asterisk in Table 6.4. Interestingly, the number of variants detected was lower using Roche 454 sequencing compared to Sanger sequencing and the *T. parva* Muguga allele of $Tp2_{49\text{-}59}$ was not detected using 454 sequencing. This might indicate the *T. parva* Muguga allele is present at a very low frequency and the detection of this allele by Sanger sequencing was fortuitous. This low frequency could be a problem using Roche 454 sequencing, as a minimum of five reads is required before a putative 454 sequence-variant is accepted.

Table 6.3 Sequences for $Tp2_{49-59}$ and $Tp2_{98-106}$ obtained for buffalo stabilate 4110

	$Tp2_{49-59}$	$Tp2_{98-106}$
V1 (1)	KSSHGMGKVGK	FAQSLVCVLMK
V2 (2)	KSSHGMGKIGR	LAQSIVCVVSK
V3 (1)	KSSKGMTAVGK	FAKSIKCVSQH
V4 (1)	KTSKGMTKVGR	FGQSIKCVAQH
V5 (1)	KTSRSMGQVGR	LAASIKCVSQH
V6 (1)	KTSKGMTRVGR	FGQSIKCVAQH
V7 (1)	KTPKGMTKVGR	FGQSIKCVAQH
V8 (1)	LTSKSMSEVGR	FAQSIYCVVKN
V9 (2)	LTSHGMGRIGR	FAASIKCVAQY

Variant sequences of the $Tp2_{49-59}$ and $Tp2_{98-106}$ are displayed, the number of sequences is displayed in parentheses. Residues that differ from $Tp2_{49-59}$ and $Tp2_{98-106}$ of *T. parva* Muguga are displayed in red.

Table 6.4 Sequences for $Tp2_{49-59}$ and $Tp2_{98-106}$ obtained for buffalo stabilate 3081

	$Tp2_{49-59}$	$Tp2_{98-106}$
V1 (1)	KSSHGMGKVGK	QSLVCVLMK*
V2 (1)	KSSHGMGKVGR*	ESIMCVIKK
V3 (1)	KSSKGMTAVGK*	QSIYCVANN
V4 (1)	KSSKGMTAVGK	QSIYCVVKN*
V5 (1)	KTSKGMTKVGR*	QSIKCVAQH
V6 (2)	LTSKAMTTVGK*	QSIKCVAHH*
V7 (1)	LTSKAMTHVGR	QSIYCVAKN
V8 (1)	LTSKAMTHVGK	QSIYCVAKN*
V9 (1)	LTSKAMTTVGK	QSIKCVSQH*
V10 (1)	LTSKAMSTVGK*	QSIYCVANN
V11 (1)	KTSKGMTEVGK	ASIHCVSNK*
V12 (1)	LTSKGMAAVGR	QSIQCVSKY*

Variant sequences of the $Tp2_{49-59}$ and $Tp2_{98-106}$ are displayed, the number of sequences is displayed in parentheses. Residues that differ from $Tp2_{49-59}$ and $Tp2_{98-106}$ of *T. parva* Muguga are displayed in red. Sequences that were also obtained using 454 sequencing are marked * (see Figure 6.4)

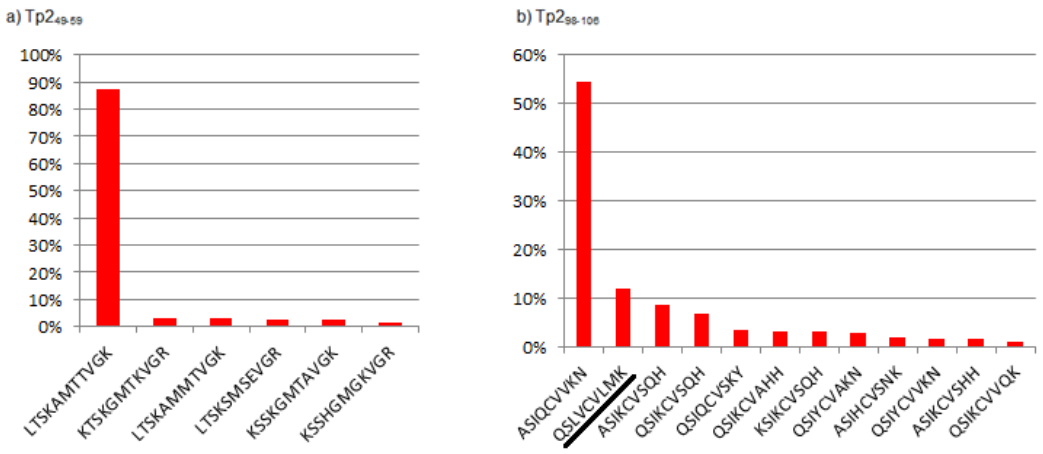

Figure 6.4 Allelic diversity of the Tp2$_{49-59}$ and Tp $_{98-106}$ epitopes found in Stabilate 3081 using Roche 454 amplicon sequencing
The percentage of reads of Tp2$_{49-59}$ allelic variant found in the forward reads (481 reads) or Tp2$_{98-106}$ allelic variants in reverse reads (603 reads) are displayed. The Muguga Tp2$_{49-59}$ allele (KSSHGMGKVGK) was not detected whereas the Muguga Tp2$_{98-106}$ allele (QSLVCVLMK) was detected at ~11% of the reads (underlined).

The infectivity of different sporozoite stabilates can vary greatly. Estimation of infection rates with *T. parva* in ticks used to produce the stabilate provide an indication of the likely infectivity for cattle and *in vitro* titrations provide data on infectivity for blood lymphocytes *in vitro*. However, the only definite way to establish infectivity *in vivo* is to perform *in vivo* titration. Unfortunately, very limited information was available on the *in vivo* infectivity of the two available buffalo stabilates, as titration was performed using only a few animals. It was considered important that the stabilate used for challenge in this study produced sufficiently severe reaction *in vivo* to ensure breakthrough parasites could be isolated *ex vivo*. Ngumi *et al.* used stabilate 3081 for evaluating whether buffalo-derived *T. parva* stabilates could be safely used for immunisation by infection and treatment (Ngumi et al., 1992). Control animals that did not receive treatment were included in this experiment and it was shown that the stabilate was highly infective at a 1/10 dilution and remained infective at a 1/100 dilution (Ngumi et al., 1992). Given this information, it was decided to utilise stabilate 3081 for challenge and a 1/20 dilution of the stabilate was used in an attempt to limit the severity of disease while still having a reasonable probability that each dose contains sufficient parasite numbers to represent all the allelic variants present in the stabilate.

6.3.3 Immunisation and challenge of cattle

Animals were immunised with *T. parva* Muguga (Stabilate 3801) with success assessed by the detection of a serological response. The results showed that animals BG040, BF080 and BF085 developed significant titres of antibody against *T. parva*, indicating successful immunisation. The titre of antibody in animal BF086 was inconClusive. Unfortunately, animal BG041 seroconverted before immunisation and could therefore not be used for the remainder of the study.

Table 6.3 Serological screening for antibodies against *Anaplasma marginale, Babesia bigemina, Theileria mutans* and *Theileria parva*

Animal	Timing	pp			
		A.Marginale	B.bigemina	T.Mutans	T.parva
BG040	Arrival	6	3	7	5
	Pre-immunisation	8	5	18	5
	Pre-Challenge	4	3	10	46
	6 weeks post challenge	7	3	6	97
BG041	Arrival	10	54	7	9
	Pre-immunisation	11	61	14	36
	Pre-Challenge	5	45	35	77
	6 weeks post challenge	5	48	16	99
BG044	Arrival	12	4	10	3
	Pre-immunisation	nd	nd	nd	nd
	Pre-Challenge	5	2	4	5
	6 weeks post challenge	5	3	2	24
BF080	Arrival	17	91	14	10
	Pre-immunisation	33	75	5	16
	Pre-Challenge	5	48	4	31
	6 weeks post challenge	7	47	3	87
BF083	Arrival	97	44	19	3
	Pre-immunisation	nd	nd	nd	nd
	Pre-Challenge	96	28	8	5
	6 weeks post challenge	76	14	3	11
BF085	Arrival	64	74	17	11
	Pre-immunisation	69	78	14	16
	Pre-Challenge	12	72	65	71
	6 weeks post challenge	19	68	37	92
BF086	Arrival	29	33	4	4
	Pre-immunisation	28	27	9	5
	Pre-Challenge	9	22	9	20
	6 weeks post challenge	12	19	5	67

Samples were taken on day of arrival (two weeks pre-immunisation), on the day of immunisation, on the day of challenge (ten weeks after immunisation) and at six weeks post-challenge. Seropositivity was determined using a pathogen-specific ELISA. Values in green are considered seronegative, values in red are considered seropositive and values in black are considered inconClusive.

CD8$^+$ T cell lines were isolated from the immunised animals and stimulated *in vitro* using autologous *T. parva* Muguga cell lines as previously described in Chapter 2. The specificity of the CD8$^+$ T cells was tested using a ^{51}Cr release assay. Unfortunately, only a limited number of targets could be tested using in this assay with relatively low effector-to-target ratios, due to limited proliferation *in vitro.* Low levels of killing against autologous *T. parva* cell lines were found for each of the immunised animals with a maximum killing between 8 and 24% (see Figure 6.5). Low levels of killing were also seen against the A10-homozygous *T. parva* line for animals BG040, BF080 and BF085. Animal BF086 did not show reactivity with the A10 homozygous line and was therefore considered unlikely to recognise Tp2$_{49\text{-}59}$ or Tp2$_{98\text{-}106}$.

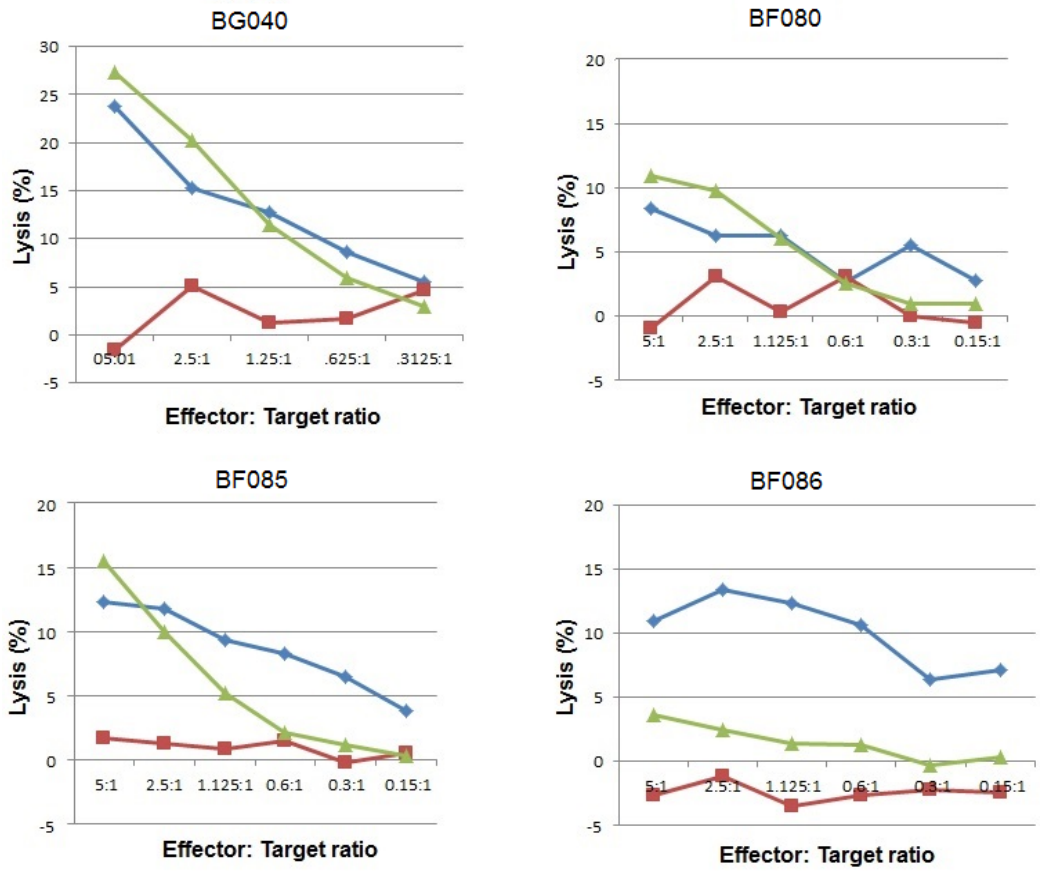

Figure 6.5 MHC I restriction of CD8 T cell lines from *T. parva* Muguga immune animals.
A chromium release assay was used to test the ability of CD8$^+$ T cell lines of *T. parva* immunised animals BG040, BF080, BF085 and BF086, to lyse autologous *T. parva* (-◊-), MHC I A10 *T. parva* Muguga line (592 Tpm A10/A10) (-Δ-) and a MHC I mis-matched *T. parva* Muguga (2229 Tpm A18/A18) (-□-)

The ability of the immunised animals to recognise $Tp2_{49\text{-}59}$ and/or $Tp2_{98\text{-}106}$ was further investigated using an *ex vivo* IFN-γ ELISPOT assay (see Figure 6.6). All immunised animals recognised autologous *T. parva* Muguga cells, indicating that immunisation successfully induced $CD8^+$ responses against *T. parva*. $CD8^+$ cells from both animal BG040 and BF085 were able to recognise $Tp2_{98\text{-}106}$. $CD8^+$ cells from animal BG040 also recognised $Tp2_{49\text{-}59}$. Thus, two animals were suitable for investigating the selective breakthrough of non-*T. parva* Muguga Tp2 variants.

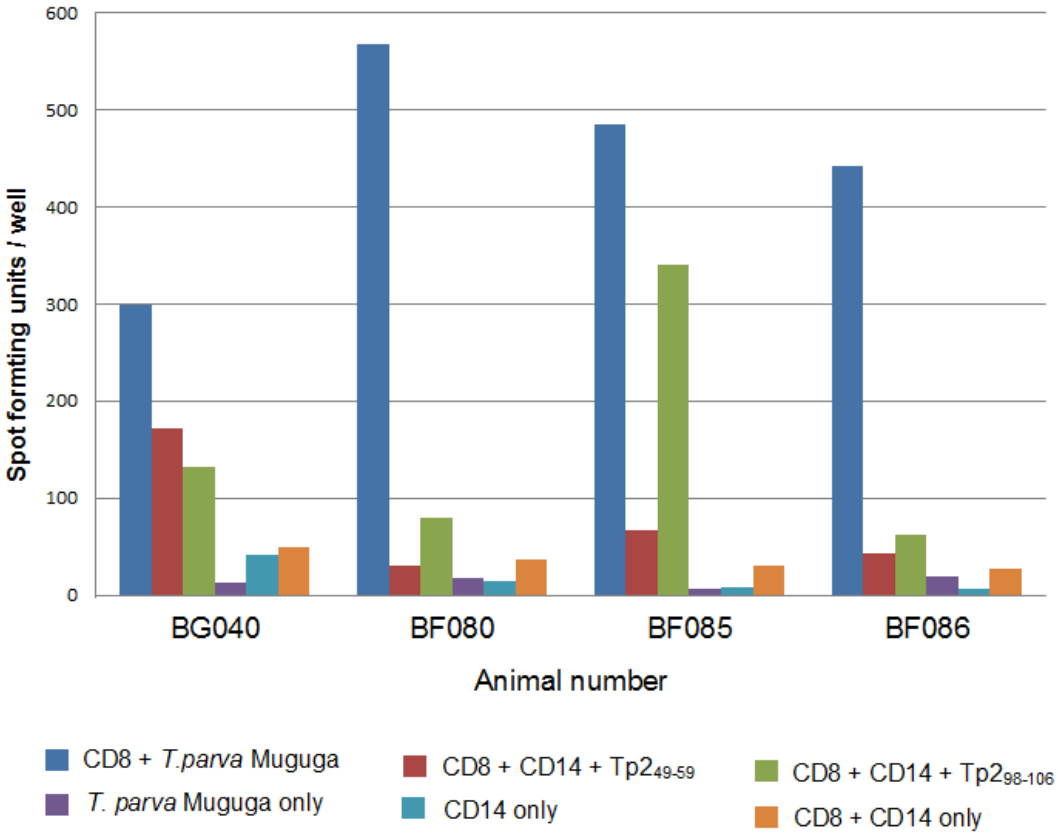

Figure 6.6 Recognition of *T. parva* Muguga and $Tp_{249\text{-}59}$ and $Tp_{298\text{-}106}$ of $CD8^+$ T cells *ex vivo* six weeks after immunisation by infection and treatment
An IFN-γ ELISPOT was used to test the reactivity of isolated $CD8^+$ T cells (2.5×10^4 cells) with autologous *T. parva*-infected cells (2.5×10^4 cells) or autologous monocytes (2.5×10^4 cells) pulsed with either the $Tp2_{49\text{-}59}$ or the $Tp2_{98\text{-}106}$ peptide at a concentration of 1 µM. Autologous *T. parva* Muguga cell line, $CD14^+$ cells and $CD8^+$ cells in combination with $CD14^+$ were used as negative controls.

6.3.4 Reaction of animals upon challenge with buffalo stabilates 3081

Animals were then challenged with 1 ml of a 1:20 dilution of buffalo-derived stabilates 3081. The clinical reaction of each animal was very different despite great effort to ensure homogenisation of the stabilates before challenge. The reactions differed in terms of the time to first detection of schizonts in the lymph nodes, the levels and duration of detectable parasites and the severity of clinical signs (see Table 6.6 and appendix D).

Table 6.6 Clinical parameters for immune and naive animals after challenge with stabilate 3081

Animal number		Parasites detected in lymph node	Fever (>39.5°C)		Day of treatment	Clinical signs
No.	ITM		Day	Max temp		
BG040	I	12,14-16	9,11,16	40.9	16	Nasal discharge, anorexia, increased respiratory rate
BF085	I	NPS	24,25	40	-	-
BF080	I	9,10,17,23	18-20	41.2	-	Slight anorexia
BF086	I	13-16	15-16	40.8	16	Recumbent, anorexia, nasal discharge, increased respiratory rate
BG044	C	7-9,11-15	14,15	41.1	15	Recumbent, anorexia, neurological signs
BF083	C	20,22	19-24	40.5	24	Slight anorexia

I = Animal was immunised by infection and treatment using *T. parva* Muguga stabilate 3308, C = Naïve control. Fever was defined as a body temperature (≥ 39.5°C) and the days on which animals had a fever and the maximum body temperature are displayed. NPS= No parasites seen

The clinical reactions of the two non-immunised animals (BG044 and BF083) were very different. Parasites were detected from day seven for animal BF044 and treatment was required at day 15 after challenge. Interestingly, fever, an early and consistent feature of infections with cattle-derived *T. parva*, only developed a week after the detection of the first parasites, but the animal showed obvious clinical signs before the onset of fever. The other control animal (BF083) showed only a delayed transient mild reaction, with fever between 19 and 24 days after infection and detectable parasites on days 20 and 22.

Among the immunised animals the reactions were also very variable. Schizonts were detected in lymph node smears from animal BG040 twelve days after challenge and this animal required treatment at 16 days after challenge. Animal BF085 only developed fever 24 days after challenge, which disappeared by day 26 without intervention. No parasites could be detected in lymph node and blood smears from this animal, but parasite-positive samples were found using PCR at day 17, 18 and 25. For animal BF080, parasites were intermittently detected from day 9 to day 23. The animal showed only very mild clinical signs and did not require treatment. The first parasites were detected at day 13 after challenge for animal BF086. This animal developed a severe clinical reaction and required treatment at day 16.

Overall, the clinical signs observed after challenge with this stabilate were different from those following the commonly used *T. parva* Muguga stabilates: the lymph node enlargement was limited in all the animals and in some animals fever was detected days after the first parasites were detected in the lymph nodes. The number of parasites found in lymph node smears was generally very low despite, in some instances, severe clinical reactions.

6.3.5 Analysis of allelic variants of Tp2 in challenge animals

The number of samples available for the analysis of Tp2 allelic variants of *T. parva* parasites *ex vivo* was limited due to one animal sero-converting before immunisation and the variability observed in the responses following immunisation and challenge. Only two animals recognised $Tp2_{49\text{-}59}$ or $Tp2_{98\text{-}106}$ and the added variability of the reaction in the control and immunised animals made it difficult to compare parasites in naïve controls and breakthrough parasites in immune animals that responded to Tp2. Therefore, the scale and objectives of the experimental plan were reduced to that of a preliminary study designed to compare the parasites at the beginning of the infection with those at the time of treatment in a single pair of animals. It was anticipated that the Tp2-specific T cell response in the immune animal would alter the composition of Tp2 alleles.

Samples were chosen from one naïve animal and one immune animal that had $CD8^+$ T cells capable of recognising Tp2. *Theileria parva* PCR-positive samples taken at the beginning of the infection (day 8) and on the day of treatment (day 15) were examined from control animal BG044. Samples from the immune animal, BF085 taken on days 18 and 25 following challenge were also selected for analysis. Thus, the time interval between each pair of samples was equivalent for the naïve and the immune animal. The Muguga $Tp2_{49\text{-}59}$ epitope KSSHGMGKVGK was not detected in samples from either the immune or the naïve animal at either time-point (see Figure 6.7). In the immune animal certain alleles in the breakthrough appear to reduce in frequency between the two time points examined (KTSKAMTMTGR and KSSHGMGKVGR), whereas others are boosted (eg. KSSKGMTAVGK). Conversely, the allele frequencies in the naïve animal are more similar between the time points.

The percentage of reads representing the Muguga $Tp2_{98\text{-}106}$ QSLVCVLMK epitope was lower at both time-points in the immunised animal (BF085) compared to the naïve animal (see Figure 6.8). In this animal, the percentage of reads was also lower for a number of other allelic variants, i.e. QSIKCVAHH, ASIHCVSNK, ASIKCVSQH and QSIKCVSQH. The frequency of reads that had a valine (V) in position seven, a lysine (K) in position eight or an asparagine (N) in position nine of the $Tp2_{98\text{-}106}$ epitope, was higher in samples from the immune animal (BF085) compared to the naïve animal (BF044) (see Table 6.7). However, no obvious difference was observed between the pairs of samples from individual animals. Hence, it was impossible to draw any clear conClusions from the results of this experiment.

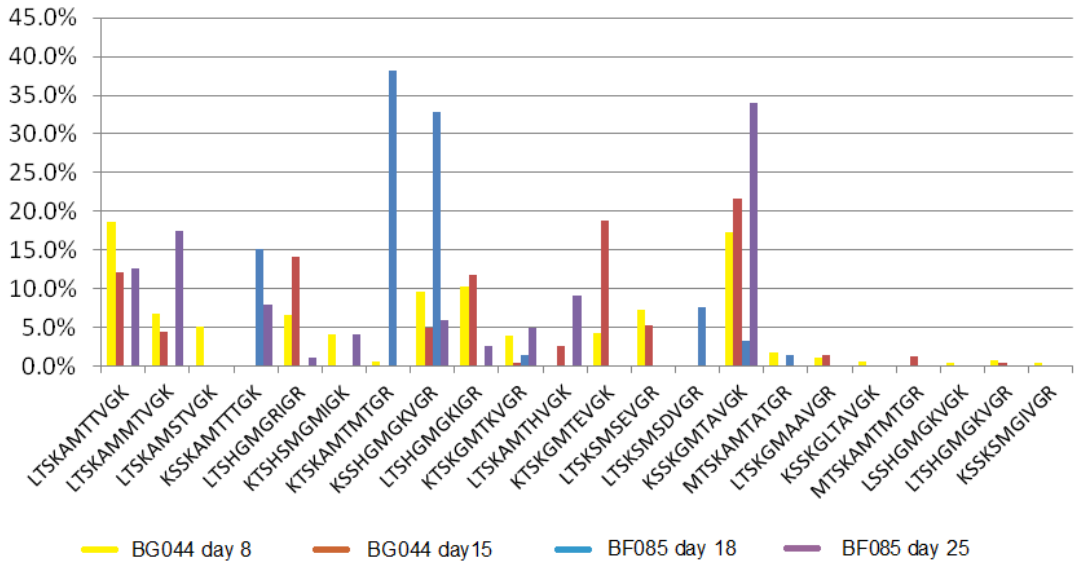

Figure 6.7 Percentage of reads representing different alleles of $Tp2_{49\text{-}59}$ forward reads in parasites isolated *ex vivo*
Reads are shown for one control animal (BG044) and one immune animal (BF085) at the beginning of infection (day 8 and 18 respectively) and at time of treatment (BG044, day 15) or at the peak of the response (BF085, day 25).

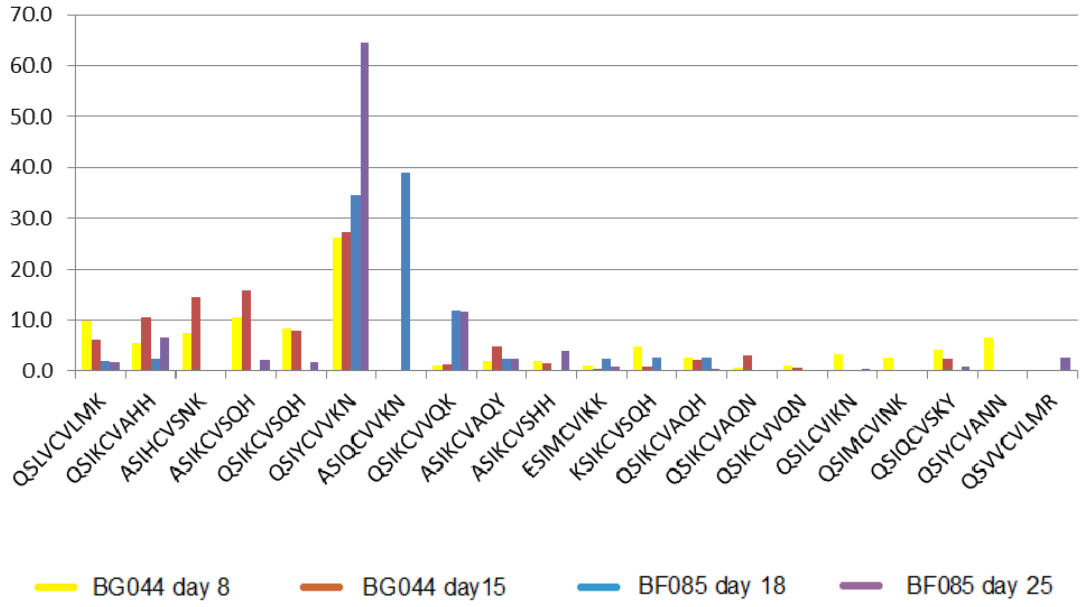

Figure 6.8 Percentage of reads representing different alleles of $Tp2_{98\text{-}106}$ reverse reads of parasites obtained *ex vivo*
Reads are displayed for one control animal (BG044) and one immune animal (BF085) at the beginning of infection (day 8 and 18 respectively) and at time of treatment (BG044, day 15) or at the peak of the response (BF085, day 25).

Table 6.7 Proportion of reads displaying certain amino acid residues of the $Tp2_{98-106}$ epitope in *T. parva* isolated *ex vivo*

Position 1		Q	A	E	K			
	4423	72.2	21.9	1.1	4.8			
	4427	61.7	36.8	0.5	0.9			
	8506	53.6	41.3	2.4	2.8			
	8507	90.8	8.4	0.9	0.0			
Position 2		S						
	4423	100.0						
	4427	100						
	8506	100						
	8507	100						
Position 3		L	V					
	4423	100.0						
	4427	100						
	8506	100						
	8507	97.4	2.6					
Position 4		V	H	K	L	M	Q	Y
	4423	9.7	7.5	38.8	3.4	3.8	4.0	32.8
	4427	6.2	14.7	49.0	0.0	0.5	2.4	27.3
	8506	2.0	0.0	22.2	0.0	2.4	38.9	34.5
	8507	4.5	0.0	28.9	0.5	0.9	0.8	64.4
Position 5		C						
	4423	100.0						
	4427	100						
	8506	100						
	8507	100						
Position 6		V						
	4423	100.0						
	4427	100						
	8506	100						
	8507	100						
Position 7		L	A	I	S	V		
	4423	9.7	17.3	7.2	37.2	28.7		
	4427	6.2	20.6	0.5	43.5	29.2		
	8506	2.0	7.5	2.4	2.8	85.3		
	8507	4.5	9.5	1.4	8.5	76.1		
Position 8		M	N	H	K	Q		
	4423	9.7	16.7	7.4	34.7	31.4		
	4427	6.2	14.7	12.2	30.2	36.8		
	8506	2.0	0.0	2.4	75.8	19.8		
	8507	4.5	0.0	10.3	66.6	18.5		
position 9		K	H	N	T	Y		
	4423	22.2	33.7	38.1	0.0	6.0		
	4427	22.7	39.2	31.0	0.0	7.1		
	8506	16.3	7.9	73.4	0.0	2.4		
	8507	14.5	14.7	65.0	2.6	3.2		

The background of cells was coloured from no colour to dark red to indicate a low to high percentage of reads with a certain amino acid residue in a certain position of the $^{Tp2}_{98-106}$ epitope.

6.4 Discussion

6.4.1 Selection of animals

This study set out to investigate whether $CD8^+$ T cell responses to a dominant *T. parva* antigen (Tp2) coincide with non-recognition of parasites with a different Tp2 genotype and, consequently, the immune 'escape' of such genotypes. Animals expressing a MHC I A10 haplotype, known to recognise epitopes within the dominant Tp2 antigen (Graham et al., 2008b; MacHugh et al., 2009b), were used for this study. Unfortunately, not all immunised animals in this study recognised the Tp2 epitopes. The animals were heterozygous for the MHC I A10 haplotype and the lack of recognition of Tp2 epitopes by some of the animals is likely a reflection of differences in the immunodominance hierarchy among animals of different MHC I backgrounds. During the course of these studies, other researchers in the Morrison Bovine Immunology Group (University of Edinburgh) have also found that only a proportion of the MHC I A10 heterozygous animals have detectable $CD8^+$ T cell responses to Tp2. Moreover, A10-homozygous animals recognise both Tp2 epitopes while A10-heterozygous animals only gave detectable responses to the $Tp2_{98\text{-}106}$ epitope (Xaioying Li and Tim Connelley, unpublished). A similar phenomenon has been observed in immune responses to influenza**;** immune responses to several $H2D^b$- and $H2K^b$-restricted influenza epitopes were diminished in $H2^{bxk}$ F1 versus homozygous mice (Day et al., 2011). Graham *et al.* (2008) examined immune responses of cattle immunised with a cocktail of five $CD8^+$ T cell antigens by a heterologous prime-boost regime with recombinant poxvirus vectors. The animals used for the study had been identified as expressing a MHC I allele capable of presenting defined antigens to $CD8^+$ T cells. Only two of the six animals expressing 2*01201 had $CD8^+$ T cell IFN-γ responses against the Tp2 gene. It was suggested that this could be due to a lack of potency of the vaccination protocols rather than an inability of the animals to respond to the antigens (Graham et al., 2008b). For future studies, it would be informative to use MHC I homozygous animals or a larger number of heterozygous animals in order to obtain sufficient numbers of animals responding to the antigen under examination. Alternatively, another as yet

unidentified polymorphic antigen that induces more consistent reactions in heterozygous animals could be used.

The reaction upon challenge with the buffalo-derived stabilate 3081 was very variable in both immune and naïve animals. There is evidence that the Muguga stock which was used for immunisation contains very little antigenic diversity. It was therefore anticipated that immune animals would only recognise a small proportion of the challenge parasites and that they would suffer severe reactions upon challenge. Differences in reactions in immune animals could occur due to different antigen specificities of $CD8^+$ T cells induced by the immunisation using Muguga. However, this cannot explain the differences in the control animals. One explanation could be that animals used were not truly immunologically naïve, despite the fact that only animals that were serological negative for *T. parva* were used. One animal was excluded from further analysis as it seroconverted before immunisation (BG041). All the animals were transported from the same farm and the presence of one seropositive animal indicates that the animals had a chance of being exposed to *Theileria parva* prior to the start of the experiment. Another explanation is that prior exposure to other tick-borne pathogens such as *Theileria mutans* and *Babesia bigemina* could have induced a degree of cross-immunity in some of the animals.

6.4.2 Challenge dose

Great efforts were made to homogenise the stabilate before inoculation to ensure consistency of the dose and content of parasites administered to each animal. Previously available data indicated that the stabilate was highly pathogenic at a 1/10 dilution and led to fatality in five challenged animals at an average of 16.4 days (Ngumi et al., 1992). The most likely explanation, perhaps, is that variability of reaction is due to a reduction in infectivity of the stabilate compared to when it was used previously. A further two-fold dilution of the dose used in previous studies, which resulted acute disease and rapid death, would not have been expected to result in the substantial reduction in disease severity in a proportion of the animals observed in the present study. Stabilates can lose infectivity during long-term storage and through the freezing and thawing process. Since the stabilates were generated in

the 1980s and it has been more than twenty years since the last infectivity data, it is likely that a degree of infectivity has been lost in the material used in the present study. While a higher dose may have yielded a more homogenous reaction on challenge it would likely also have elicited more severe clinical signs. As the animals were expected to have immunity against only a proportion of the challenge parasites, increased severity of disease may have resulted in an overwhelming parasite load which may in turn have dysregulated the immune response. This may have resulted in failure of the immune response despite the presence of $CD8^+$ T cells capable of recognising a proportion of challenge parasites. Thus, the specific 'knock-down' of genotypes recognised by the $CD8^+$ T cell response would not be observed.

6.4.3 Selective removal of genotypes

A previous study using satellite DNA markers provided evidence that selective removal, or filtering, of certain genotypes occurred following challenge of animals previously immunised with a cloned population of the Marikebuni stock and then challenged with the parental Marikebuni stock. Filtering was not observed following challenge of naïve animals (Katzer et al., 2007). Different parasite genotypes emerged at different time-points after challenge and in immune animals of different MHC I backgrounds (Katzer et al., 2007). The filtering appeared to be focused on a few genomic loci although it could not be related to specific antigens. The present study set out to determine whether the emergence of the different parasite genotypes in challenged animals is related to $CD8^+$ T cell responses directed to immunodominant polymorphic antigens.

Due to the variability discussed above, a limited number of samples were available for analyses, and therefore the study was reduced to a preliminary analysis in a single pair of immune and naïve animals. The Muguga $Tp2_{49\text{-}59}$ epitope KSSHGMGVGK was not detected in either immune or naïve animal and thus no preferential filtering could be seen for this epitope. This is probably a reflection of the low proportion of Muguga $Tp2_{49\text{-}59}$ epitope KSSHGMGVGK in the challenge stabilate and thus a stabilate with a higher proportion of the 'immunising' epitope would be desirable. In the immune animal certain alleles in the breakthrough appear to be reduced in

frequency whereas others were boosted. In contrast, the allele frequencies in the naïve animal were more similar between the time points. The differences in the breakthrough parasite populations at the different timepoint, may be evidence that immunity has disrupted the challenge populations. As animal BF085 did not recognise $Tp2_{49\text{-}59}$ peptide, this could be evidence for strain specificity acting on an additional unidentified antigen.

The Muguga $Tp2_{98\text{-}106}$ epitope was detected in both the immune (BF085) and naïve animal (BG044). The proportion in animal BG044 was similar to that found in the stabilate at both the start of the infection and at the time of treatment. The proportion appeared to be reduced in animal BF085 at both 18 and 25 days after challenge. $CD8^+$ T cells of this animal recognised $Tp2_{98\text{-}106}$ peptide using the ELISPOT assay and thus genotypes with this epitope would be expected to be preferentially removed after memory CD8+ T cell responses become effective Effector function of memory CD8+ T cells can be detected from 8 days after challenge (Graham et al., 2006) and thus filtering due to CD8 + T cell responses might have occurred prior to day 18 after challenge The results indicate that there might be a degree of filtering, but as samples were derived from only a single immune animal and the fact that the Muguga Tp2 allelic variants of $Tp2_{49\text{-}59}$ and $Tp2_{98\text{-}106}$ were only present at a low proportion in the challenge stabilate, no firm conClusions can be drawn from this experiment. In an ideal study, a challenge stabilate consisting of a mixed parasite population containing a lower number of more distinct allelic variants of the target $CD8^+$ T cell antigen would be used. The $CD8^+$ T cell responses induced by immunisation should be focused on an allelic variant that is present in higher abundance in the stabilate than in the current study, thus making it easier to detect 'knock-down' following challenge. No such stabilate is currently available, although one could be synthesised by mixing different cloned parasites and co-transmitting them through ticks as utilised by Katzer *et al.* (2011). This would allow a round of sexual reproduction to take place in the tick and thus the predicted re-assortment of genes encoding different polymorphic $CD8^+$ T cell target antigens (Katzer et al., 2011).

Chapter 7: Haemoparasites in cattle and African buffalo

7.1 Introduction

Buffalo are commonly infected with a range of *Theileria* species, including *Theileria parva*, *Theileria velifera*, *Theileria mutans*, *Theileria buffeli* and *T. sp (buffalo)* (Allsopp et al., 1999; Oura et al., 2011a, b; Young et al., 1978). Grazing of cattle and buffalo in proximity to each other provides the opportunity for transmission of pathogens from buffalo to cattle via ticks. Different *Theileria* species are transmitted by a number of different tick species, but the vector for *T. sp* (buffalo) and *T. buffeli* have not yet been identified (see Table 7.1) (Norval et al., 1992). Severe disease is seen in cattle after infection with buffalo-derived *T. parva* (Norval et al., 1991a). However, the role of other haemoparasites transmitted from buffalo in causing disease in cattle is less clear.

Table 7.1 *Theileria* spp found in cattle and/or African buffalo in Africa and corresponding tick vectors

Parasite species	Tick vector	Mammalian host
T. parva[1,2,3]	*R. appendiculatus,* *R. zambeziensis* *R. duttoni* (other *Rhipicephalus spp)* *(Hyalomma spp.).*	Bovine, African buffalo
T. taurotragi[2]	*R. appendiculatus* *R. zambeziensis* *R. pulchellus*	Cattle, sheep, goats, wild ungulates (e.g. eland, Bushbuck)
T. mutans[2,3]	*A. astrion* *A. cohaerens* *A. gemma* *A. hebraem* *A. lepidum* *A. variegatum*	Cattle, African buffalo, eland, bushbuck
T. velifera[2]	*A. variegatum* *A. hebraem* *A. astrion* *A. lepidum*	Cattle, African buffalo
T. sp (buffalo)	*Unknown*	African buffalo
T. buffeli[4]	*Unknown*	African buffalo, cattle, eland, bushbuck

[1] (Thompson et al., 2008) [2] (Norval et al., 1992) [3](Jongejan et al., 2004)[4] (Ngumi et al., 1994)

Theileria mutans and *Theileria velifera* can infect cattle and are generally considered to be non-pathogenic, although there are some reports of clinical disease associated with *T. mutans* infection in cattle (Irvin et al., 1972; Paling et al., 1981; Snodgrass et al., 1972). Until recently, *Theileria buffeli* and *Theileria sp. (buffalo)* were thought to be exclusive to buffalo and incapable of infecting cattle. However, Oura *et al.* were able to detect *Theileria buffeli* in young calves grazing in Lake Mburo National Park in Uganda using a reverse line blot assay (Oura et al., 2011b). Young *et al.* also found a then unidentified parasite in cattle with clinical theileriosis on the Marula Estate, Nakuru district, Kenya (Young et al., 1992). Only a few macroschizonts were present and large piroplasms were detected in erythrocytes with veils and bars (Young et al., 1992) Later studies demonstrated that *T. sp Marula* was, in fact, *T. buffeli* (Allsopp et al., 1993; Ngumi et al., 1994). *Theileria sp (buffalo)* was not detected in any of the cattle tested in Oura's study, whereas most buffalo examined were found to be infected with this species (Oura et al., 2011b). The cattle examined by Oura et al did not have signs of clinical disease but many had carrier infections with *T. parva*, suggesting that *T.sp (buffalo)* had not established carrier infections in these animals. However, it is possible that an acute infection can establish in cattle, which does not lead to a carrier state. Similarly, in South Africa, *T. sp (buffalo)* was not detected in around 2,500 cattle samples tested with a real-time hybridisation PCR assay during routine diagnostic screening (Mans et al., 2011a).

Evidence for the existence of *Theileria sp (buffalo)* as a distinct species was first reported by Conrad *et al.* (1987) who used monoclonal antibodies and restriction fragment length polymorphism profiles, employing a *T. parva*-specific repetitive DNA probe, to try to differentiate between cattle- and buffalo-derived *Theileria parva*-parasitised cell lines. DNA from sub-clones of the buffalo 6834 cell line hybridised poorly with the IgTPM-23 and IgTpM probes and the buffalo-derived sub-clones showed a somewhat different reaction pattern to that of the cattle-derived ones using monoclonal antibodies (Allsopp et al., 1993; Conrad et al., 1987a; Conrad et al., 1989; Conrad et al., 1987b). Sequencing of the 18S ribosomal subunit from these lines indicated that the cells were infected with a novel species of *Theileria* that was closely related to but distinct from *T. parva* and this was provisionally named

Theileria sp (buffalo) (Allsopp et al., 1993). A cell line infected with the same species has subsequently been isolated from a buffalo in South Africa (Zweygarth et al., 2009). In the present study, cloned cell lines infected with *T. sp (buffalo)* were identified among clones generated from two *Theileria*-parasitised cell lines derived from buffalo. The isolation of *T. sp (buffalo)*-infected cell lines from buffalo indicates that the parasite is able to transform buffalo leukocytes. Whether *T. sp (buffalo)* is also able to transform cattle cells is unknown. Co-infections with *T. parva* and *T. sp (buffalo)* are very common in buffalo in Eastern Africa and Southern Africa (Allsopp et al., 1993; Oura et al., 2011a, b; Pienaar et al., 2011b). These co-infections present a challenge to the diagnosis of *T. parva* in buffalo in South Africa, where a real-time PCR diagnostic test is used. The test is based on the 18S ribosomal RNA gene, the DNA sequence of which is very similar for the two parasites (Sibeko et al., 2008). Pienaar *et al.* examined the effect of mixed *T. sp (buffalo)*/*T. parva* infection on the accuracy of the diagnostic test and estimated that about 10% of *T. parva* infections are misdiagnosed (Pienaar et al., 2011). A further complication is the detection in South African buffalo of additional 18S ribosomal RNA gene sequences that differ by a few nucleotides from *T. sp (buffalo)* (Chaisi et al., 2011; Mans et al., 2011b). This led to questioning whether the two sequence types represented different species of *Theileria*. Mans *et al.* used the 5S ribosomal subunit gene, which shows greater sequence divergence, to successfully differentiate the two parasites (Mans et al., 2011a).

As described in Chapter 3 of this thesis, analyses of a set of cloned buffalo parasitised cell lines with a panel of PCR primers designed to amplify antigen genes of *T. parva* identified several lines that failed to yield PCR products with many of the primers. These cell lines were subsequently shown to be infected with *T. sp (buffalo).*

The aims of the studies described in this chapter were:

- To examine the phylogenetic relationship of *T. sp (buffalo)* and *T. parva* utilising clonal cell lines
- To investigate the presence of *T. sp (buffalo)* in cell lines and samples from cattle with clinical theileriosis after grazing on pastures shared with African buffalo
- To identify *Theileria* species present in African buffalo, which could potentially be transmitted to cattle
- To identify *Theileria* species present in tick-stabilate produced by feeding *R. appendiculatus* on African buffalo

7.2 Material and Methods

7.2.1 Analysis of *T. sp (buffalo)* in *Theileria*-infected cell lines

DNA from 18 uncloned *Theileria*-infected cell lines derived from cattle on the Marula farm in the Rift valley region of Kenya were examined. These DNA samples were provided by David Odongo, ILRI, Kenya. The Marula farm is located close to Naivasha National Park and animals were grazed in an area known to be frequented by buffalo. The cell lines were established from animals that showed clinical signs characteristic of East Coast fever and schizonts were detected in lymph node aspirates.

The 18S subunit was amplified from genomic DNA samples with *Theileria*/*Babesia*-specific primers as described previously (Gubbels et al., 1999; Oura et al., 2004a). 1 µl of PCR product was then used in a semi-nested PCR reaction using a *T. parva* and a *T. sp (buffalo)*-specific reverse primer and visualised following electrophoresis on a 1.5% agarose gel. The presence of *T. sp (buffalo)*-specific sequence was confirmed for some of the cell lines by sequencing of PCR products amplified using the *Theileria*/*Babesia* species-specific primers, which were cloned using the pGEM-T Easy vector system (Promega Corporation, USA) as previously described in Chapter 2. A very small quantity of culture was used directly for PCR amplification using the *T. sp (buffalo)*-specific reverse primer. *T. sp (buffalo)* positive molecular clones were further expanded and purified using the PureYield Plasmid Miniprep System (Promega Corporation, USA). Sequencing of PCR products was performed by Durham Genomics.

Table 7.2 Primers used for amplification of the 18S rRNA gene

Specificity	Name	Sequence 5'-3'
Pan *Theileria*/*Babesia* primers	18s_Forward_Gub	GAGGTAGTGACAAGAAATAACAATA
	18s_reverse_Our2	CTAAGAATTTCACCTCTGACAGT
T. parva-specific reverse primer	18s_ parva_HH1	CTTATTTCGGACGGAGTTCG
T. sp (buffalo)-specific reverse primer	18s_Sp_HH1	GGCTTATTTCAGACGGAGTTTA

7.2.2 454 Sequencing of 18S ribosomal RNA subunit

Samples examined for the presence of Theileria and Babesia species were as follows: (i) Nine DNAs from blood samples from a group of twelve sentinel Boran cattle that developed clinical theileriosis following introduction into an area grazed by African buffalo on the Ol Pejeta conservancy, Laikipia district, Kenya. DNA from these samples was extracted by ILRI staff from whole blood; (ii) DNA derived from blood samples of African buffalo collected in the Ol Pejeta conservancy in Kenya (blood clots) and the Kruger National Park in South Africa (whole blood); (iii) DNA extracted from two previously generated stabilates of sporozoites, which had been produced by feeding *R. appendiculatus* ticks on carrier buffalo that were captured as calves. Stabilate 3081 was prepared from *R. appendiculatus* ticks which fed on a *T. parva* carrier buffalo (7014) captured on Ol Pejeta Ranch, Nanyuki, Kenya. Stabilate 4110 was prepared from *R. appendiculatus* ticks which fed on a *T. parva* carrier buffalo (7752).DNA samples from the sentinel cattle and African buffalo in the Ol Pejeta conservancy and buffalo-derived stabilates were supplied by ILRI, Kenya. DNA from buffalo from the Kruger National Park was provided by Dr Nick Juleff, Pirbright Institute, UK.

The presence of different species of *Theileria* and *Babesia* in samples described above was investigated by sequencing a 375bp segment of the 18S ribosomal RNA gene using Roche 454 amplicon sequencing technology. The 18S ribosomal RNA gene was amplified from samples of genomic DNA using fusion primers that generate a product from all *Theileria* and *Babesia* species, as described in Chapter 3. Fusion primer sequences are detailed in Table 7.3. Further processing of the samples and bioinformatic processing was performed as described in Chapter 3.

Table 7.3 Sequences of fusion primers used for amplification of the 18S rRNA gene for sequencing using Roche 454 technology

Name	Adaptor	MID tag	Sequence
18s_for_gub_Ax	CGTATCGCCTCCCTCGCGCCATCAG	xxxxxxxxxx	GAGGTAGTGACAAGAAATAACAATA
18s_rev_Our2_Bx	CTATGCGCCTTGCCAGCCCGCTCAG	xxxxxxxxxx	CTAAGAATTTCACCTCTGACAGT

7.2.3 PCR amplification for phylogenetic analysis of *T. sp (buffalo)* and *T. parva*

Selected genes of *T. sp (buffalo)* and *T. parva* clones were amplified for phylogenetic analysis. At least four clones were sequenced for each of 4 categories of Theileria parasites: cattle-associated *T. parva*, cattle-derived buffalo-associated *T. parva*, buffalo-derived *T. parva* and buffalo-derived *T. sp (buffalo)* (see Table 7.4). The genes amplified were 5S ribosomal RNA gene (Mans et al., 2011a) and three genes encoding T cell antigens in *T. parva*: Tp6, Tp7 and Tp8 (see Table 7.5). The primers for the amplification of Tp6, Tp7 and Tp8 and their *T.sp (buffalo)* orthologues were designed when trying to design primers that would amplify all/most *Theileria parva* variants (See chapter 3), but were not used for this purpose because the primers also amplified genes from *T. sp (buffalo)* clones. The amplified PCR products were cleaned using the Promega Wizard Gel and PCR Purification Kit (Promega Corporations, USA) and sent to Durham Genomics for Sanger sequencing. Sequence analysis was performed as described in Chapter 2.

Table 7.4 Samples used for the phylogenetic analysis of *T. sp (buffalo)* and *T. parva*

Name/Animal	Clone	Location
Cattle derived *T. parva*		
Muguga	Reference genome	Kenya
Marikebuni	A3, A7, B12, E43, F44, F53, I8, I38, J17	Kenya
Mariakani	St 3231 clone 2/3	Kenya
Boleni	St3230 clone 1:1	Zimbabwe
Uganda	St 3645clone 1/2	Uganda
Cattle associated with African buffalo *T. parva*		
Marula N33	Clone 2, 4, 5, 7	Nakuru, Kenya
Marula N43	Clone 1, 2, 5, 6	Nakuru, Kenya
Buffalo-derived *T .parva*		
Mara 3	Clone 3	Maasai Mara, Kenya
Mara 30	Clone 5, 8, 11	Maasai Mara, Kenya
Mara 42	Clone 2,5	Maasai Mara, Kenya
Buffalo 6998	Clone 2,4,9	Kenya
Buffalo-derived *T. sp (buffalo)*		
Buffalo 6998	Clone 3,8,10	Kenya
Buffalo 6834	Clone 1,5,7,10	Laikipia, Kenya

Table 7.5 List of primers and primer sequences to amplify the 5S ribosomal RNA gene , Tp6, Tp7 and Tp8

Gene	Primer name	Primer sequence
5S*	5S_for_Mans	ATGACAAACACAGAAGTCGCCCT
	5S_rev_Mans	ATTTCATCCTTCTTCTTGATTGCGT
Tp6	Tp6_for_HH2	CGTCCAATAATTTACGATGTGAG
	Tp6_rev_HH3	CTTGTTTAGCCTCTACAGC
Tp7	Tp7_for_HH1	TGAAGAAGGACGACTCGCAC
	Tp7_rev_HH1	TAAGCATTTCCCACTCACGC
Tp8	Tp8_for_HH1	ATCCACAACCAAGTGCCCAG
	Tp8_rev_HH	ACTGCGAAGGAGGTCAATCC

*(Mans et al., 2011a)

7.3 Results

7.3.1 Demonstration of *T. parva* and *T. sp (buffalo)* in cell lines derived from naturally infected cattle

Eighteen parasitised cell lines were obtained from cattle undergoing clinical disease after grazing in an area frequented by buffalo on the Marula estate, Kenya. *Theileria parva* was detected in 15 of the samples and *T. sp (buffalo)* was detected in seven of the samples using a semi-nested PCR with reverse primers specific for either *T. parva* or *Theileria sp (buffalo)* in a second round of amplification (see Figure 7.1).

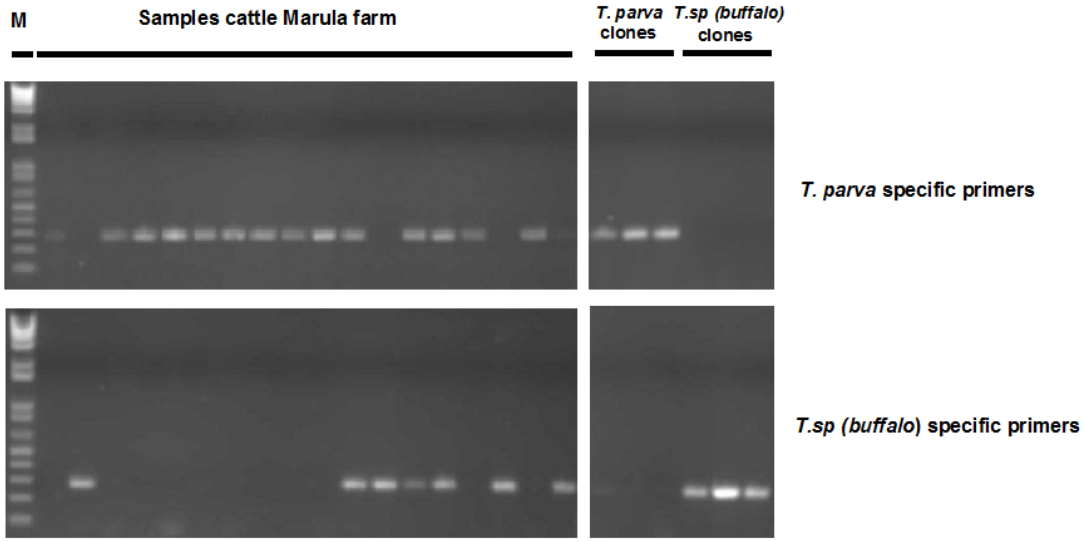

Figure7.1 Results of a semi-nested PCR assay of 18S ribosomal subunit using primers specific for *T. parva* and *T. sp (buffalo)* in the second round of PCR.

The presence of *T. parva* and *T. sp (buffalo)* in samples from 18 cattle from the Marula farm is illustrated. Three *T. parva* and three *T. sp (buffalo)* clones were used as controls to confirm species-specificity. M = molecular weight marker

That the PCR products were derived from *T. sp (buffalo)* was confirmed by sequencing two of the products obtained for the 18S rRNA gene after cloning into the P-GEM-T cloning vector. Sequences obtained for animals N18 and N86 are shown in Figure 7.2.

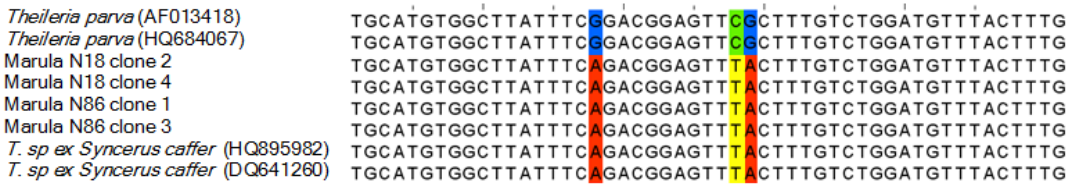

Figure 7.2 Nucleotide alignment of a 50 bp region of the 18S rRNA gene amplified from samples from cattle from the Marula farm
Sequences are aligned with sequences of *Theileria parva (*AF013418 and HQ684067) and *T. sp (buffalo)* (HQ895982 and DQ641260) obtained from the NCBI database. Nucleotides that were not conserved between sequences were highlighted.

7.3.2 Phylogenetic analysis of *T. parva* and *T. sp (buffalo)*

A set of clonal *T. parva* and *T. sp (buffalo)* parasitised cell lines was used for further analysis of the phylogeny of the two parasites. The 5S rRNA gene used in a previous study to analyse samples from South African buffalo was examined (Mans et al., 2011a). The other genes used in the present study were Tp6 (predicted to encode prohibitin), Tp7 (predicted to encode heat shock protein 90) and Tp8 (predicted to encode a cysteine protease). These three antigen genes are relatively conserved among *T. parva* genotypes and do not appear to have been subject to immune selection (see chapter 5). Four groups of cloned parasitised cell lines were used, each group consisting of a minimum of four cloned parasitised cell lines: (1) buffalo-derived *T. parva*, (2) *T. parva* from cattle co-grazed with buffalo (buffalo-associated), (3) *T. parva* from cattle in buffalo-free areas and (4) buffalo-derived *T. sp (buffalo)*.

Phylogenetic trees constructed for all four genes demonstrate that *T. sp (buffalo)* sequences form a cluster distinct from the *T. parva* sequences, with high levels of bootstrap support at the node separating the species. This provides further evidence that *T. sp (buffalo)* might represent a distinct species. No clear demarcation can be found between buffalo-derived *T. parva* and buffalo-associated *T. parva* on phylogenetic trees based on any of the markers. There is a tendency for cattle-derived *T. parva* to cluster separately from the buffalo-derived and buffalo-associated *T. parva*, but the phylogenetic distance is small and much smaller than the distance between *T. parva* and *T. sp (buffalo)* (see Figures 7.3-7.10). Moreover, the Tp6 sequence from one buffalo-derived *T. parva* clusters with the cattle-derived *T. parva* (see Figures 7.7 and 7.8).

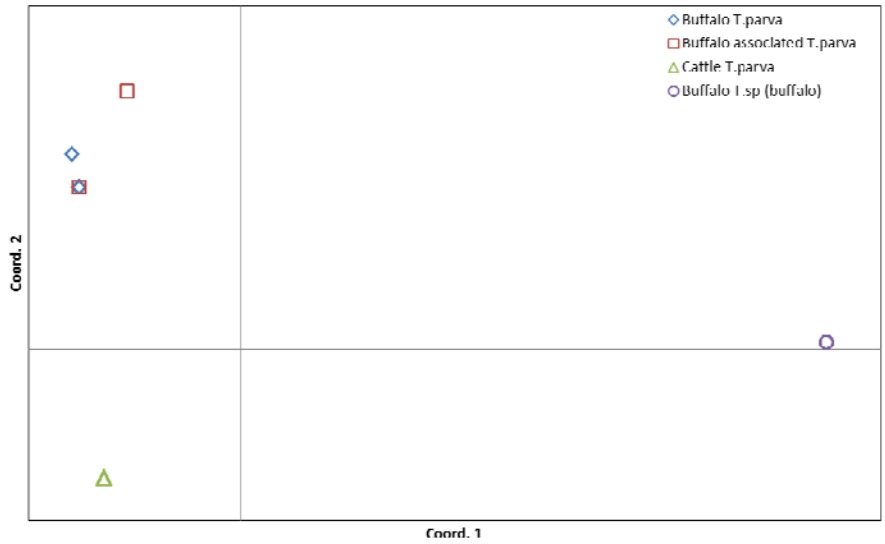

Figure 7.3 Principal component analysis of Tp8 and Tp8 orthologues in *T. sp (buffalo)*. This diagram illustrates the relationship between buffalo-derived (◊), buffalo-associated (□) and cattle-derived *T. parva* (Δ) isolates and buffalo-derived *T. sp (buffalo* (o), suggesting *T. sp (buffalo)* isolates form a distinct population. The datapoints in the PCoA correspond with the samples displayed in the phylogenetic tree in figure7.4

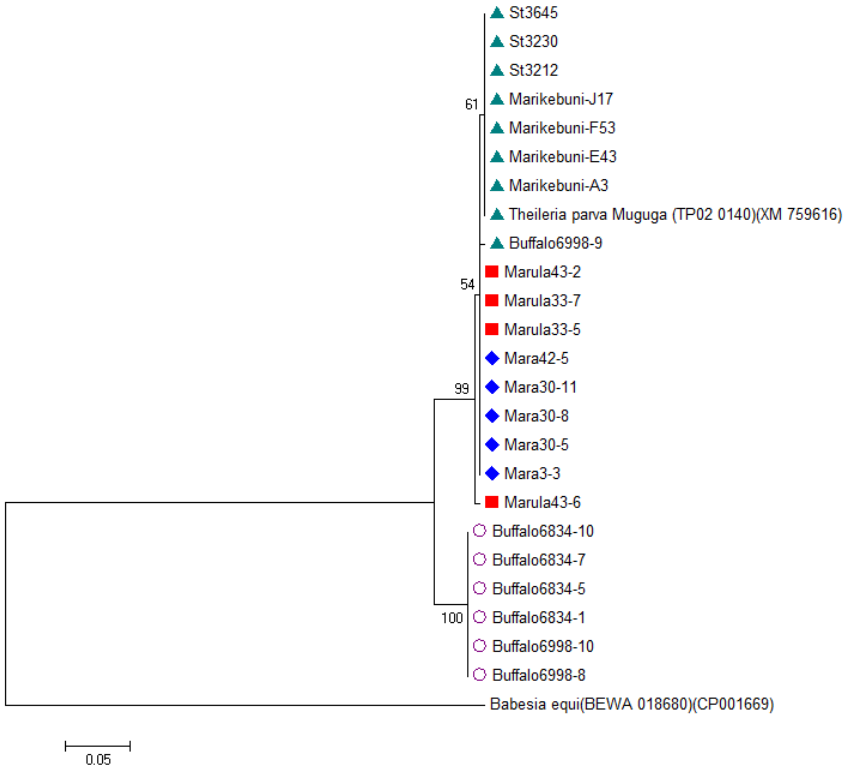

Figure 7.4 Phylogenetic tree showing the relationship between *T. parva* and *T. sp (*buffalo) based on Tp8 sequences. Sequences for Tp8 were used to construct a maximum likelihood tree using 1,000 bootstrap replicates. The percentage of bootstrap values indicating the degree of support for each node are shown. The tree was rooted using an orthologue of Tp8 found in *Babesia equi* (CP001669). The samples used were buffalo-derived (◊), buffalo-associated (□) and cattle-derived *T. parva* (Δ) isolates and buffalo-derived *T. sp (buffalo)* (o).

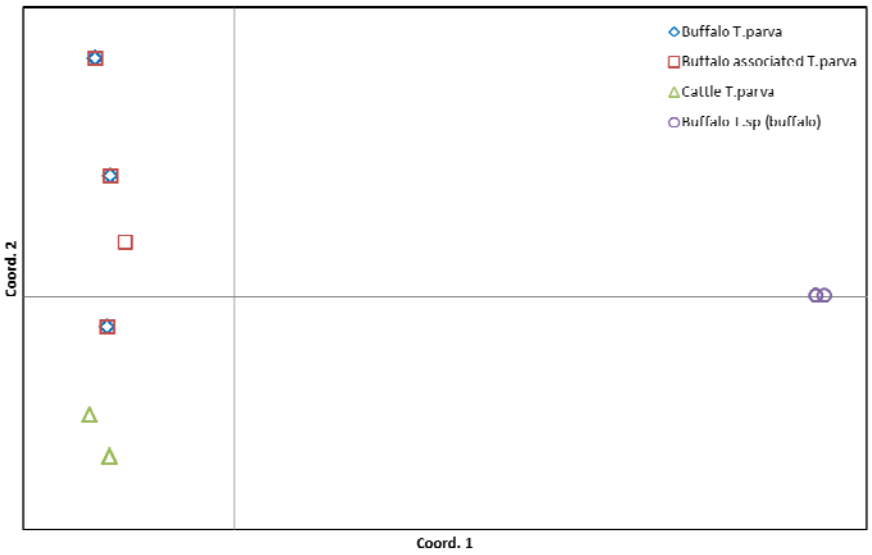

Figure 7.5 Principal component analysis of Tp7 and Tp7 orthologues in *T. sp (buffalo)*
This diagram illustrates the relationship between buffalo-derived (◊), buffalo-associated (□) and cattle-derived *T. parva* (Δ) isolates and buffalo-derived *T. sp (buffalo* (o), suggesting *T. sp (buffalo)* isolates form a distinct population. The datapoints in the PCoA correspond with the samples found in the phylogenetic tree in figure 7.6.

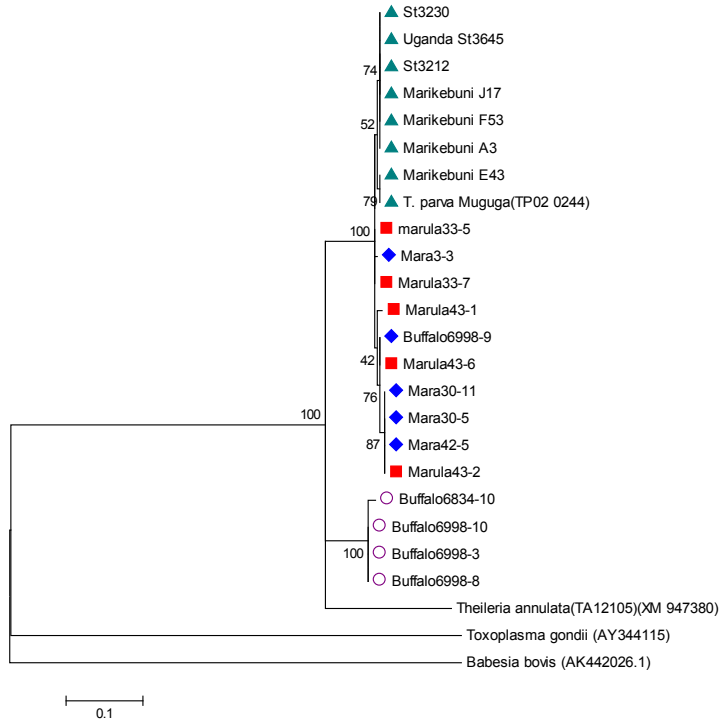

Figure 7.6 Phylogenetic tree showing the relationship between *T. parva* and *T. sp (buffalo)* based on Tp7 sequences
Sequences for Tp7 were used to construct a tree by maximum likelihood analysis using 1,000 bootstrap replicates. The bootstrap values indicating the degree of support for each node are shown. The tree was rooted using the putative Heat shock protein 90 gene sequence for *Toxoplasma gondii* (AY344115), *Babesia bovis* (AK442026*)* and *Theileria annulata* (XM_947380). The samples used were buffalo-derived (◊), buffalo-associated (□) and cattle-derived *T. parva* (Δ) isolates and buffalo-derived *T. sp (buffalo) (*o).

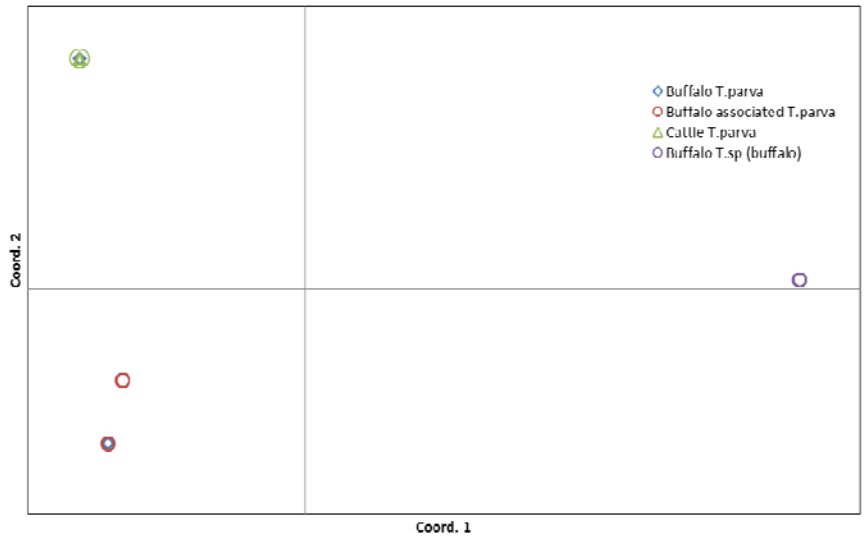

Figure 7.7 Principal component analysis of Tp6 and Tp6 orthologues in *T. sp (buffalo)*. This diagram illustrates the relationship between buffalo-derived (◊), buffalo-associated (□) and cattle-derived *T. parva* (Δ) isolates and buffalo-derived *T. sp (buffalo* (o), suggesting *T. sp (buffalo)* isolates form a distinct population. The datapoints in the PCoA correspond to the samples in the phylogenetic tree in figure 7.8.

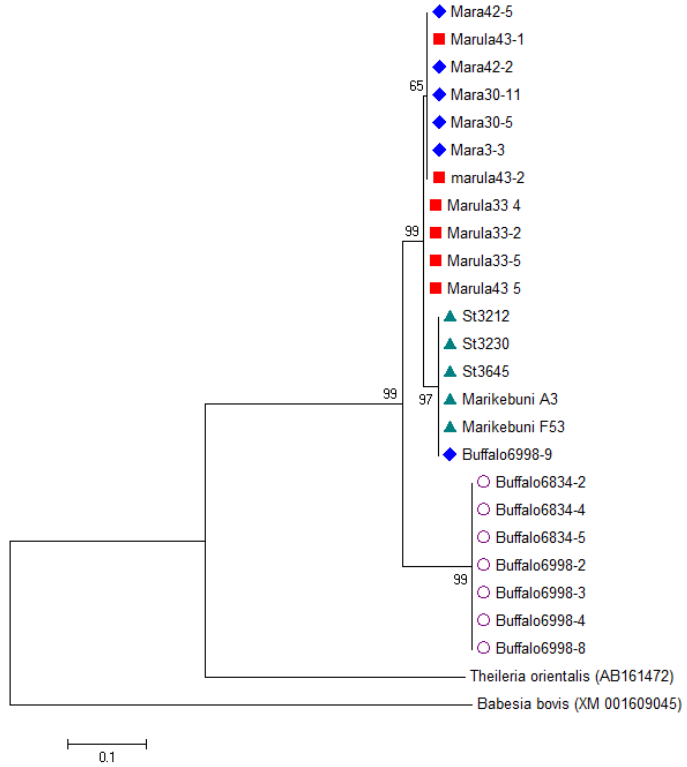

Figure 7.8 Phylogenetic tree showing the relationship between *T. parva* and *T. sp (buffalo)* based on Tp6 sequences of Tp6

The tree was constructed by maximum likelihood analysis using 1,000 bootstrap replicates. The bootstrap values indicating the degree of support for each node are shown. The tree was rooted using the prohibitin gene sequence for *Babesia bovis* (XM001609045*)* and *Theileria orientalis* (AB161472).The samples used were buffalo-derived (◊), buffalo-associated (□) and cattle-derived *T. parva* (Δ) isolates and buffalo-derived *T. sp (buffalo)* (o).

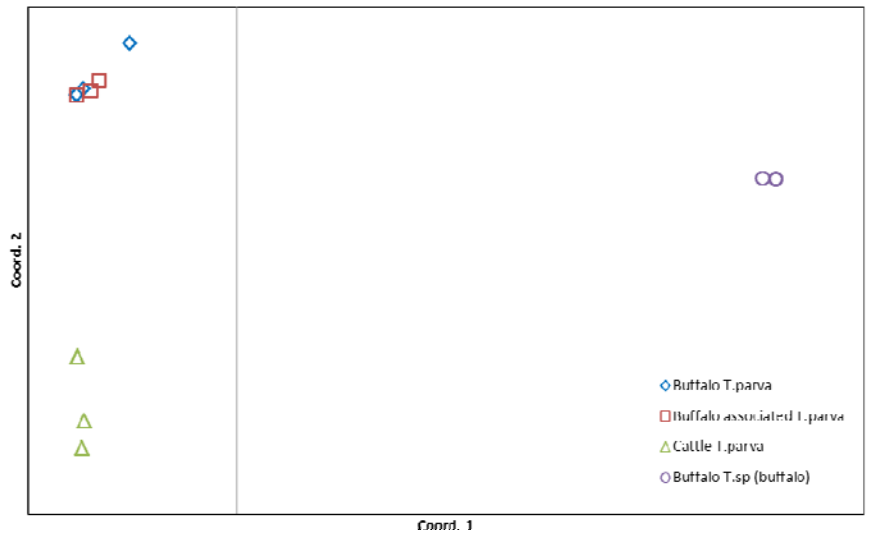

Figure 7.9 Principal component analysis of 5S rRNA gene sequences of *T. parva* and *T. sp (buffalo)*
This diagram illustrates the relationship between buffalo-derived (◊), buffalo-associated (□) and cattle-derived *T. parva* (Δ) isolates and buffalo-derived *T. sp (buffalo* (o), suggesting *T. sp (buffalo)* isolates form a distinct population. The datapoints in the PCoA correspond with the samples found in the phylogenetic tree in figure 7.10.

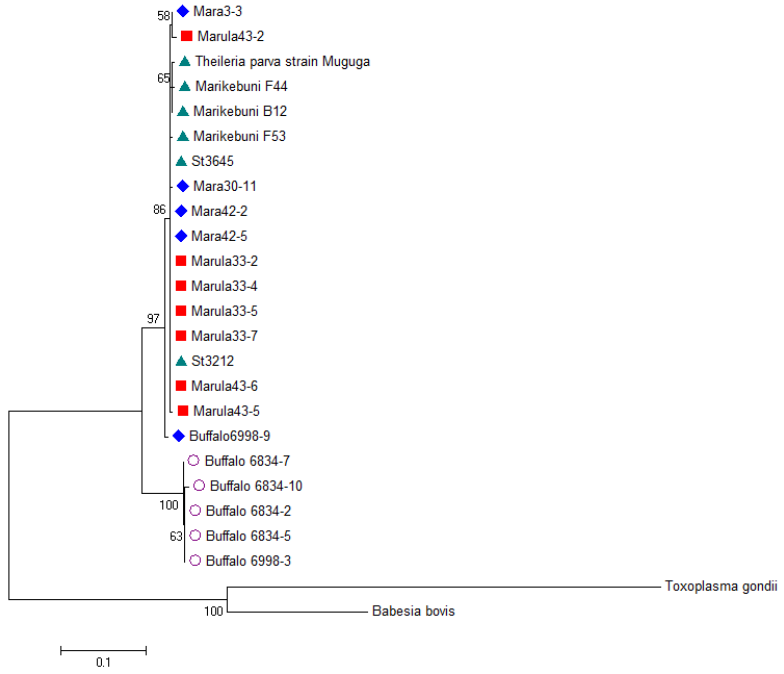

Figure 7.10 Phylogenetic tree showing the relationship between *T. parva* and T. sp (buffalo) based on the 5S rRNA gene
The tree was constructed by maximum likelihood analysis using 1,000 bootstrap replicates. The bootstrap values indicating the degree of support for each node are shown. The tree was rooted using the 5S rRNA gene sequence for *Babesia bovis* (AK441686.1) and *Toxoplasma gondii* (XM_002366722.1). The samples used were buffalo-derived (◊), buffalo-associated (□) and cattle-derived *T. parva* (Δ) isolates and buffalo-derived *T. sp (buffalo* (o).

7.3.3 Investigation of *Theileria* species in sentinel cattle grazed in proximity of African buffalo

Twelve sentinel cattle were introduced into a fenced area within the Ol Pejeta Conservancy and became infected, developing disease from 14 days after introduction. Ten out of twelve animals were detectably infected by day 16 after introduction. Animals were treated within 2-3 days of the first detection of parasites in lymph node smears. 454 sequencing of the 18S rRNA gene was conducted on nine samples obtained immediately prior to treatment and two samples obtained 27 days after treatment. In the samples obtained immediately prior to treatment, *T. parva* was detected in all nine animals and was the only detectable *Theileria* species (see Table 7.6). Co-infection with *Babesia bigemina* and another undefined *Theileria* species, which did not share a high percentage of 18S identity with other known *Theileria* species, was detected in one of the sentinel animals. Samples from this animal (32) and another animal (35) were also examined 27 days after treatment. In both of these samples, *Babesia bigemina* sequences represented the majority of the reads, with the proportion of *T. parva* sequences correspondingly reduced. Sequences representing *Theileria taurotragi* and the undefined *Theileria* species referred to above were found in animal 32, while *T. sp (buffalo)* sequences were detected in animal 35.

Table 7.6 *Theileria* and *Babesia* species detected in sentinel cattle at the Ol Pejeta conservancy

	Animal number	Theileria parva	Theileria velifera	Theileria mutans	T.sp (buffalo)	Theileria buffeli	Theileria seninsis	Babesia bigemia	Theileria taurotragi	Undefined Theileria sp
Ol Pejeta Sentinel cattle	23	895								
	25	2237								
	27	3232								
	28	1402								
	32	1942						84		33
	32*	45						691	31	27
	33	3567								
	34	2577								
	35	12915								
	35*	31			22			1516		
	36	2876								

The numbers in the table represent the number of reads obtained for each of the different *Theileria* species in each of the samples. * Samples were taken 27 days after treatment

7.3.4 *Theileria* parasites in African buffalo

DNA samples obtained from African buffalo on the Ol Pejeta conservancy provided an opportunity to examine the *Theileria* species carried by buffalo, which could potentially be transmitted to cattle. The eight African buffalo examined were all found to be infected with *T. parva*, *T. velifera* and *T. sp (buffalo)* (see Table 7.7). A low number of reads were obtained for *T. mutans* for two of the eight buffalo. One animal had co-infections with both *T. sinensis* and *T. buffeli*. One animal was co-infected with *T. buffeli* only and two animals with *T. sinensis* only.

Samples were also available from African buffalo from the Kruger National Park, South Africa. Five of the six buffalo sampled were found to be co-infected with *T. parva*, *T. velifera*, *T. mutans* and *T. sp (buffalo)* (see Table 7.7). One animal (SC02) only yielded sequences for *T. mutans* and *T. sp (buffalo),* but only a small number of sequence reads were obtained from the sample from this buffalo. Minor sequence variation was observed among the 18S rRNA sequences obtained for all of the Theileria species. These will be discussed further.

Table 7.7 *Theileria* and *Babesia* species found in African buffalo from the Kruger National Park, South Africa and the Ol Pejeta conservancy, Kenya

	Animal number	Theileria parva	Theileria velifera	Theileria mutans	T.sp (buffalo)	Theileria buffeli	Theileria seninsis	Babesia bigemia	Theileria taurotragi	Undefined Theileria sp
Kruger National park South Africa	SC01	27	116	118	89					
	SC02			10	13					
	SC03	6	97	13	357					
	SC04	23	50	23	29					
	SC05	149	370	72	455					
	SC06	59	96	62	242					
Ol Pejeta Conservancy Kenya	301	691	465		691		277			
	302	187	111		246					
	303	463	134		869					
	304	139	99	5	492					
	305	400	173	8	786					
	306	8	11		44	39	25			
	307	300	42		188		244			
	308	56	37		71	542				

The numbers in the table represent the number of reads obtained for each of the different Theileria species in each of the samples

7.3.5 Heterogeneity of *T. parva* and *T. sp (buffalo)* in African buffalo

Previous work examining the diversity of *Theileria* species in buffalo in South Africa, demonstrated the presence of sequence variants among the 18S rRNA gene sequences attributed to *T. parva* and *T. sp (buffalo)*; in each case the sequences were very closely related to, but distinct from both *T. parva* and *T. sp (buffalo)* (Chaisi et al., 2011; Mans et al., 2011). Sequences similar to the 18S rRNA sequences of *T. parva*/*T. sp* (buffalo) were obtained by 454 sequencing in the present study. The sequences were obtained from DNA samples of five African buffalo in the Kruger National Park (represented by 264 reads) and from eight African buffalo in the Ol Pejeta conservancy (2,244 reads). There were only three nucleotide differences between *T. parva* and *T. sp (buffalo*) reference sequences over the 375 bp amplicon examined. Sequence variants were present that differ in one nucleotide from the *T. parva* reference sequence and in two nucleotides from the *T. sp (buffalo)* reference sequence. Due to the limited number of nucleotide changes over the amplicon examined, no clear clustering was seen of the sequences on alignment and phylogenetic analyses of *T. parva* and *T. sp* (buffalo) sequences (see figure 7.11 and appendix E) This is in agreement with previously published data (Chaisi et al., 2011).

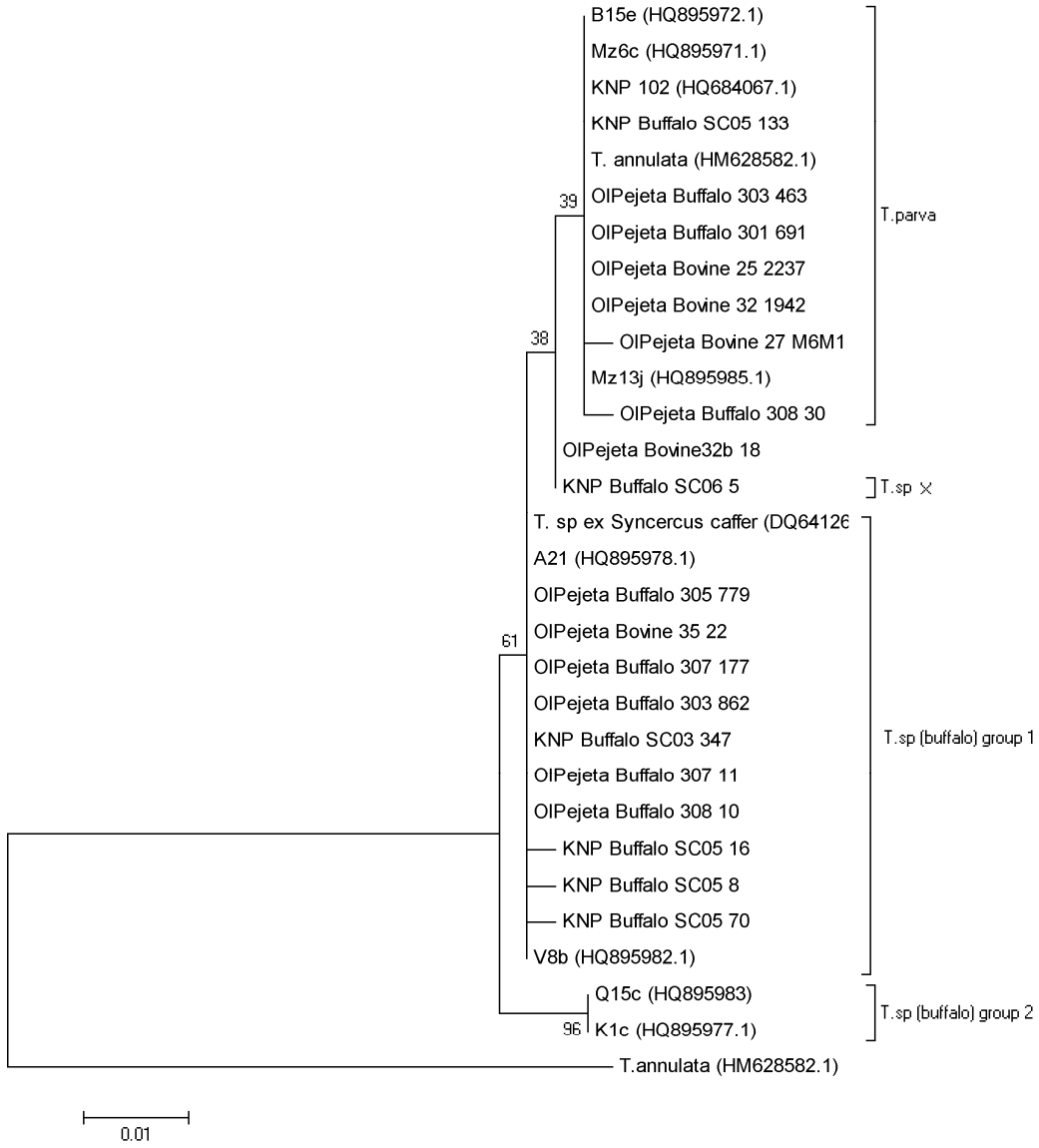

Figure 7.11 Phylogenetic analysis of *T. sp (buffalo)* and *T. parva* sequences obtained from African buffalo

The tree was constructed using a maximum likelihood method using 1,000 bootstrap replicates. The bootstrap values indicating the degree of support for each node are shown. The tree was rooted using the 18S rRNA gene sequence for *T. annulata (*HM628582.1). Samples from African buffalo from the Kruger National Park are labelled KNP Buffalo, followed by the animal number (SC01-SC06), and followed by the number of reads obtained for the allele. Samples from African buffalo from the Ol Pejeta conservancy are labelled Ol Pejeta Buffalo, followed by the animal number (301-308) and the number of reads obtained for the allele.

7.3.6 Heterogeneity in the 18S rRNA gene of *T. velifera*

Twenty-five distinct 18S rRNA sequences similar to *Theileria velifera* were obtained from thirteen buffalo samples, corresponding to 1,801 reads. The BLASTN sequence homology searches indicated that the seventeen sequences obtained from samples from buffalo from the Kruger National Park (corresponding to 729 reads) were identical or similar to previously identified 18S rRNA sequences of *T. velifera* (JN5272701.1- JN5272705) and only differed by a few nucleotides from each other. However, the sequence homology searches indicated that *T. velifera*-like sequences obtained from buffalo on the Ol Pejeta conservancy (corresponding to 1,072 reads) formed a distinct cluster from previously identified 18S rRNA sequences of this organism. In all the samples from African buffalo from the Ol Pejeta conservancy, only one *T. velifera*-like variant was found, which had 6 unique nucleotide changes in the amplified gene segment compared to previously identified *T. velifera*-like variants (see figure 7.12 and Appendix E). Overall, polymorphism was found in 11 nucleotides over the amplicons when looking at all the *T. velifera*-like sequences for African buffalo from both Kenya and South Africa.

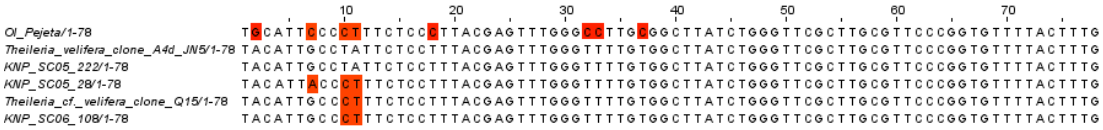

Figure 7.12 Alignment of *T. velifera*-like sequence variants

The alignment was made using a representative sequence obtained in this study from each *T. velifera*-like clade together with *T. velifera*-like sequences from the NCBI database; *T. velifera* (JN572703) and *T. velifera*-like 1 (JN572704)

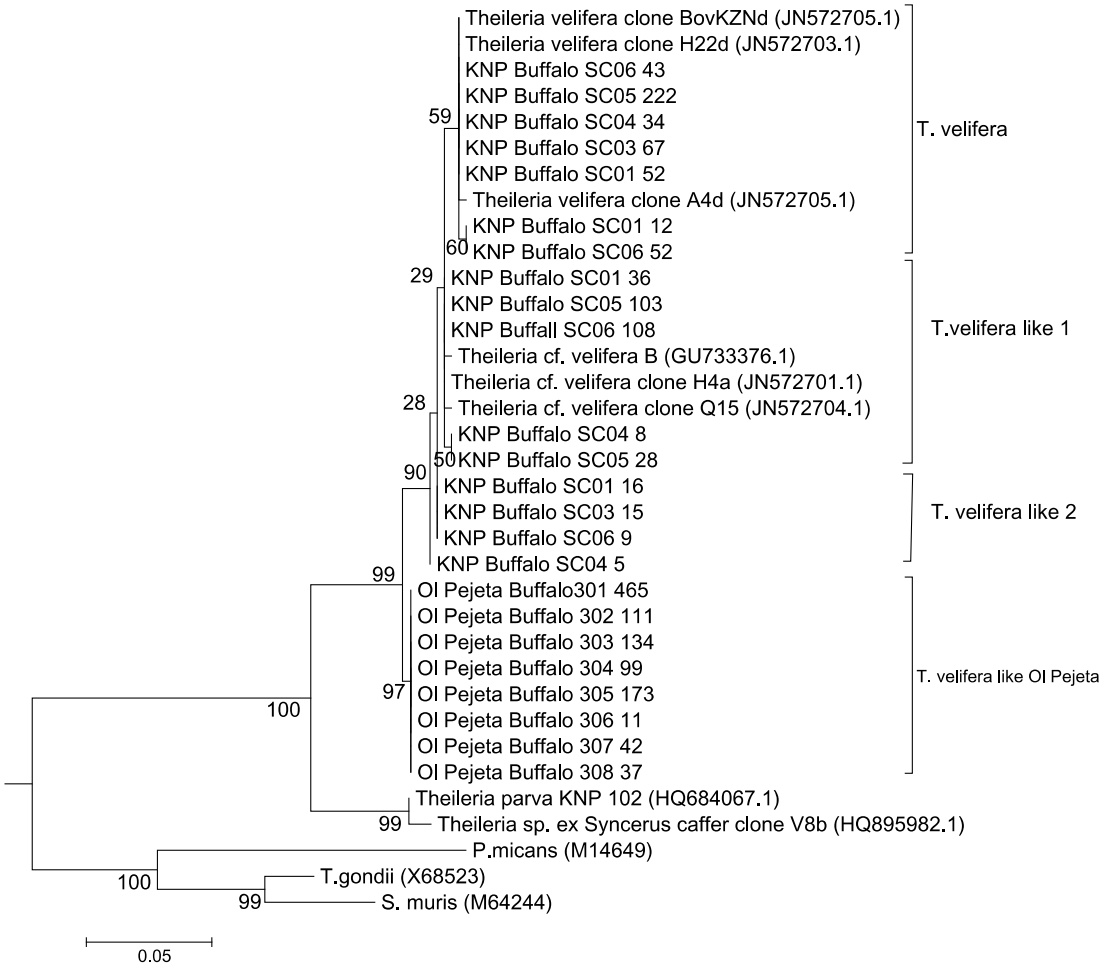

Figure 7.13 Phylogenetic tree showing the relationship between *T. velifera* 18S rRNA gene sequence variants
The tree was constructed using a maximum likelihood method using 1,000 bootstrap replicates. The bootstrap values indicating the degree of support for each cluster are shown. The tree was rooted using rRNA gene sequences for *Prororectum micans* (M14649), *Sarcocystic muris* (M64244), *Toxoplasma gondii* (X68523), *Theileria parva* (HQ684067.1) and *Theileria sp ex Syncerus caffer* (HQ895982.1). Samples from African buffalo from the Kruger National Park are labelled KNP Buffalo, followed by the animal number (SC01-SC06) and the number of reads obtained for the allele. Samples from African buffalo from the Ol Pejeta conservancy are labelled Ol Pejeta Buffalo, followed by the animal number (301-308) and the number of reads obtained for the allele.

7.3.7 Heterogeneity in the 18S rRNA gene of *T. mutans*

Eighteen distinct 18S rRNA sequences (corresponding to a total of 311 reads detected in eight buffalo) similar to *T. mutans* were obtained using BLASTN sequence homology searches. Most of these sequences were identical or similar to previously identified 18S rRNA sequences of *T. mutans* and *T. mutans*-like sequences (JN573694.1, FJ213585.1, FJ213586.1, JN572700.1, JN572696.1, AF078815.1, JN572694.1). For the samples from the Kruger National Park,

T. mutans, T. mutans-like 1 and *T. mutans-like 2* sequences were found as previously described for buffalo samples from South Africa (Chaisi et al., 2013). Unfortunately, only a low number of *T. mutans*-like reads was obtained from two buffalo from the Ol Pejeta conservancy. These were all found to be identical to another variant found previously in South African buffalo (JN572700; *T. mutans*-like X in Figure 7.14 and Appendix E) by (Chaisi et al., 2013).

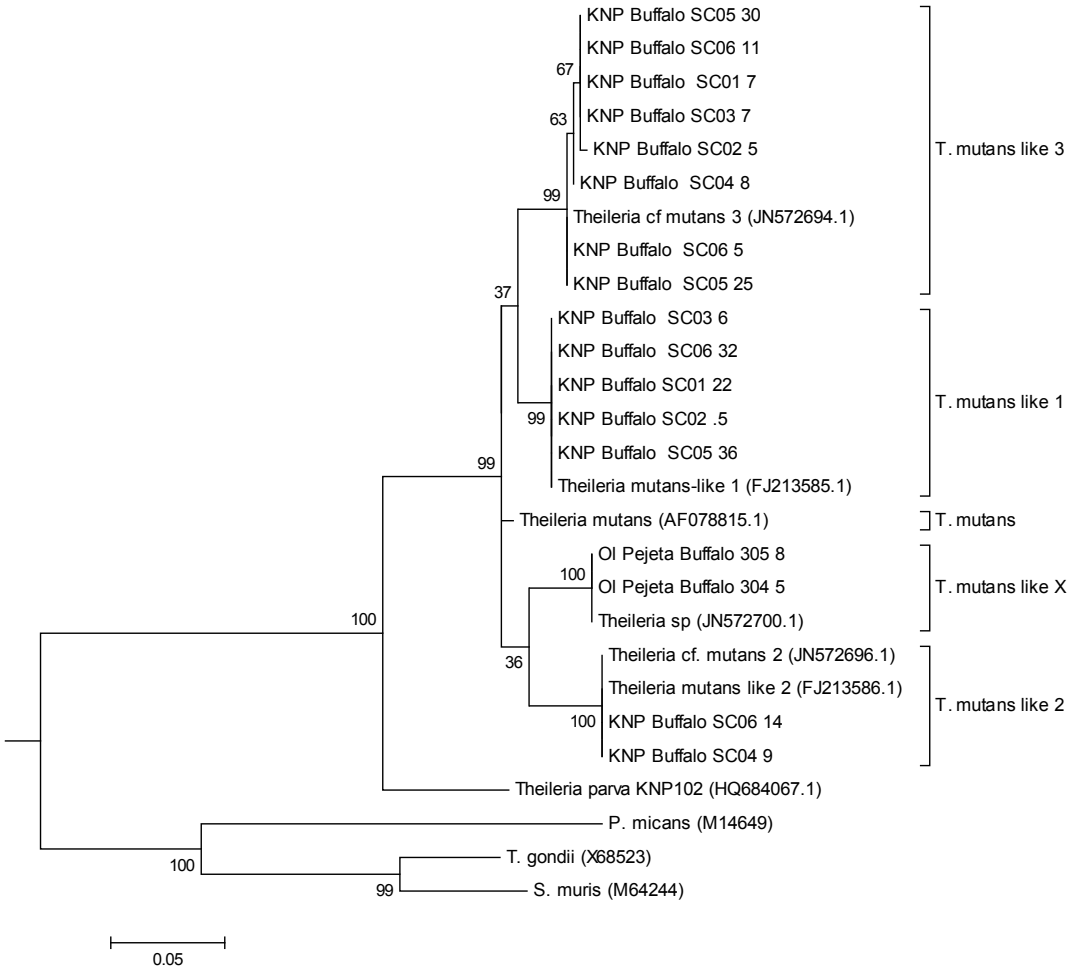

Figure 7.14 Phylogenetic tree showing the relationship between *T. mutans*-like 18S rRNA sequence variants
The tree was constructed using a maximum likelihood method using 1,000 bootstrap replicates. The bootstrap values indicating the degree of support for each node are shown. The tree was rooted using the 18S rRNA gene sequence for *Prororectum micans, Sarcocystic muris, T. gondii* and *T. parva* (HQ684067.1). Samples from African buffalo from the Kruger National Park are labelled KNP Buffalo, followed by the animal number (SC01-SC06), and followed by the number of reads obtained for the allele. Samples from African buffalo from the Ol Pejeta conservancy are labelled Ol Pejeta Buffalo, followed by the animal number (301-308) and the number of reads obtained for the allele.

7.3.8 *Theileria* species in buffalo-derived sporozoite stabilates

To investigate whether *R. appendiculatus* could be a vector for *T. sp (buffalo)*, deep amplicon sequencing of the 18S rRNA gene was performed on DNA prepared from two buffalo-derived *T. parva* stabilates. The stabilates had been produced previously by feeding laboratory *R. appendiculatus* ticks on buffalo that were captured from the field when they were young. Stabilate 3081 was produced by feeding ticks on a buffalo captured at the Laikipia district, Kenya, whereas stabilate 4110 was produced by feeding ticks on a buffalo captured in Kenya. DNA from a buffalo from the Kruger National Park and DNA from a parasitised bovine cell line from an animal on the Marula estate were also sequenced as positive controls. It was shown in a previous sequencing run that the sample from the buffalo from the Kruger National Park contained *T. parva, T. sp (buffalo), T. mutans* and *T. velifera* (see Table 7.8).The cattle cell line sample from the Marula estate contained both *T. parva* and *T. sp (buffalo)* based on the semi-nested PCR assay (see Figure 7.1). Samples from both controls yielded 18S sequences for the expected *Theileria* species. In the sequences obtained from the African buffalo from Kruger National Park (467 reads), *T. parva, T. sp (buffalo), T. mutans* and *T. velifera* were all demonstrated. Only *Theileria parva* could be demonstrated in the two buffalo-derived stabilates (represented by 1,479 and 607 reads in each stabilates) (see Table 7.8). However, for both stabilates, a number of variant sequences of the *T. parva* 18S rRNA gene (111 and 58 sequences for one of the variants for respectively stabilate 3081 and stabilate 4110, and 45 sequences for a second variant in stabilates 4110) were found differing by one nucleotide from *T. parva* sequences and two or three nucleotides from *T. sp* (buffalo) sequences (see Figure 7.15).

Table 7.8 Sequencing of the 18S rRNA gene from two buffalo-derived stabilates

Stabilate name	*T. parva*	*T. sp (buffalo)*	*T. mutans*	*T. velifera*
Buffalo-derived stabilate 3081	1,479	0	0	0
Buffalo-derived stabilate 4110	607	0	0	0
Buffalo SC06	49	203	94	121
Bovine Marula N79	765	345	0	0

The number of reads representing each of the Theileria species is indicated for the samples used; The two buffalo-derived stabilates used were stabilate 3081(Ngumi et al., 1992) and Stabilate 4110; As positive controls a blood sample from buffalo SC06 from the Kruger national Park and a *Theileria* infected cell line from bovine N79 from the Marula farm were used.

Figure 7.15 Alignment of 18S rRNA sequences obtained for buffalo-derived stabilates 3081 and 4110
Representatives of the different sequence varaints obtained for buffalo-derived stabilates 3081 and 4110 were aligned with sequences for *T. parva* (HQ895970) and *T. sp ex Syncercus caffer* (HQ895980). Nucleotide that differ from the T. parva reference genome are displayed.

7.4 Discussion

7.4.1 Detection of *Theileria sp (buffalo)* in parasitised cell lines isolated from cattle

The presence of *T. sp (buffalo)* was demonstrated in seven out of 18 cell lines derived from cattle that had been introduced into an area of the Marula estate known to be grazed by buffalo. The parasite can thus infect and transform bovine lymphocytes and can be maintained *in vitro* in mixed cultures that also contain *Theileria parva.* Further clonal analyses of these lines are required to confirm that the *T. sp (buffalo)*-transformed bovine cells can be maintained *in vitro* in the absence of *T. parva*. However, this seems likely, since cloned *T. sp (buffalo)*-infected cell lines could be established from buffalo-derived cultures. Given the ability of this parasite to transform bovine lymphocytes, a role in pathology in cattle could be expected.

The cell lines were isolated from animals with clinical disease and the presence of both *T. parva* and *T. sp (buffalo)* in cultures suggest that both parasites were abundant in the infected animal, which might imply that *T. sp (buffalo)* contributes to the pathogenesis of disease in cattle. The contribution of *T. sp (buffalo)* either alone or in combination with *T. parva* may have been overlooked in the past because the parasites are morphologically similar and are serologically cross-reactive (Conrad et al., 1987b). There is a need to obtain a *T. sp (buffalo)* sporozoite stabilate to test its pathogenicity in cattle.

The transforming parasites *Theileria parva* and *Theileria annulata* are able to cause severe disease in cattle. Candidate genes potentially involved in transformation have been identified by comparative genomic analysis of *T. orientalis* (a non-transforming *Theileria*), *T. parva* and *T. annulata* (transforming *Theileria* species) (Hayashida et al., 2012). *Theileria sp (buffalo)* provides an additional transforming species that could be added to comparative genome studies to provide further insight into the genes that are important for transformation.

Differences in cell tropism can be found between and within *Theileria* species. *Theileria parva* parasites can generally infect lymphocytes of different lineages, including $CD4^+$ T cells, $CD8^+$ T cells, γδ T cells and B cells (Baldwin et al., 1988). However, a Zambian stock of *T. parva* (Chitongo) reported to be of relatively low virulence could only infect $CD8^+$ T cells (Tindih et al., 2012). Studies of the cell tropism of *T. parva* in relation to disease have indicated that infected T cells are able to cause disease, whereas infected B cells give rise to self-limiting infections (Morrison et al., 1996). *Theileria annulata* infects cells of the monocyte/macrophage and B cell lineages (Spooner et al., 1988) and clinical signs are generally milder than infection with *T. parva.* Since *T. sp (buffalo)* is more closely related to *T. parva* than *T. annulata*, its cell tropism might be expected to be similar to *T. parva*. Future work should attempt to examine the cell tropism of this parasite both in cattle and in buffalo. Cloned infected buffalo cell lines are available for such studies while cloned bovine lines could be generated from the mixed infection lines isolated from Marula cattle.

7.4.2 Phylogenetic analysis of *T. parva* and *T. sp (buffalo)*

Theileria parva and *T. sp (buffalo)* have been shown to be genetically closely related based on analysis of the DNA sequence of the 18S rRNA gene (Chaisi et al., 2011; Mans et al., 2011a). The phylogenetic relationship between the two species was analysed further in the present study using sequences obtained for four other genes: 5S rRNA, Tp6, Tp7 and Tp8. This analysis demonstrated that for all four genes *T. sp (buffalo)* sequences form a distinct Cluster separate from *T. parva,* with high levels of bootstrap support. However, since *T. sp (buffalo)* clonal cell lines were available from only two buffalo and thus the analysis was performed with a limited number of sequences, these findings will need to be confirmed with a larger sample of parasites. If the clear differentiation between the two *Theileria* species is confirmed upon examination of additional samples, these genes could be candidates for an improved differential diagnostic test to distinguish *T. parva* and *T. sp (buffalo)* in African buffalo and cattle (Pienaar et al., 2011; Sibeko et al., 2008).

The phylogenetic analysis of buffalo-derived *T. parva* and buffalo-associated cattle *T. parva* showed that these parasites Cluster together. The sequences from buffalo-derived *T. parva* and buffalo-associated cattle *T. parva* also Clustered together on PCoA. This is in agreement with the results obtained by Pelle *et al.*, where samples from cattle on the Marula estate Clustered with buffalo-derived *T. parva* in principal component analysis based on Tp2 allelic sequences (Pelle et al., 2011). A study of samples collected in Lake Mburo National Park in Uganda employing microsatellite DNA typing found that the *T. parva* genotypes obtained from buffalo and cattle were distinct from each other (Oura et al., 2011b). However, the cattle in the Ugandan study did not have clinical theileriosis and since buffalo-derived parasites do not readily establish a carrier state in cattle (Lawrence, 1979), the cattle parasites examined were likely to represent a more restricted subset of cattle-maintained parasites compared to those examined in the current study. This is consistent with the observation that in the current study, the non-buffalo-associated cattle *T. parva* appear to form a sub-Cluster within the overall *T. parva* population based on allelic sequences at each of the four loci. Pelle *et al.* (2011) also detected a limited subset of alleles in cattle compared to buffalo for Tp1 and Tp2. There is modest bootstrapping support for the separation of the groups and the tree is unstable. The distances between the groups are small, suggesting that limited differentiation has occurred and this supports the concept of a relatively recent adaptation of a sub-population of *T. parva* to the cattle host. The collection of samples used in the current analysis was too small to determine definitively the overall relationship between the three categories of *T. parva*. A large panel of sympatric samples would be required and therefore to begin to address this issue, a number of sentinel calves located in a known buffalo area are the subject of a separate on-going study.

7.4.3 *Theileria* species in blood of cattle grazed in proximity to buffalo

Most of the sentinel animals introduced to the buffalo-grazed area on the Ol Pejeta conservancy only had detectable infection with *T. parva* at the time they received theileriacidal treatment. Oura *et al.* studied haemoparasites in resident cattle grazed in proximity to wildlife in Uganda and, in addition to *T. parva,* found *T. mutans, T. velifera, T. taurotragi, A. marginale* and *E. bovis* (Oura et al., 2011b). However, most of these animals would have had prolonged exposure to the local parasites and had persistent carrier infections. The sentinel animals introduced at the Ol Pejeta conservancy were maintained tick-free in the period before their introduction and as a precautionary measure were also treated with acaricide on the day of arrival in the conservancy. All animals became heavily infested with ticks (>1000 ticks/animal) over the following two weeks, clinical/parasitological evidence of theileriosis was first observed for most animals from 16 days after acaricide treatment, and treatment for *Theileria* was required for these animals by day 19. Hence, the time-frame between the acaricide losing residual effect and anti-*Theileria* treatment was relatively short and might not have allowed the development of infection with other *Theileria* species that were present at lower levels in the ticks and/or were transmitted by a different tick species. In addition, since the animals all had acute infections with *T. parva*, low levels of infections with other parasites might not have been detectable because of the much higher frequency of *T. parva* 18S sequence reads.

The rapid development of disease in the sentinel animals is likely due to a combination of very high levels of exposure to buffalo-derived *T. parva* and the lack of immunity. The large number of ticks found on the cattle might have further stressed the animals making them more susceptible to disease. Since this area of the conservancy is fenced to exclude cattle, ticks will have fed exClusively on wildlife. The levels of infection with *Theileria* parasites in the ticks were not examined in the current study. *Theileria parva* was found in all the samples of African buffalo examined and this persistent parasitaemia, in the absence of clinical signs, likely results in high infection rates in buffalo-fed ticks.

Two of the sentinel animals on the Ol Pejeta conservancy were found to be infected with *Babesia bigemina.* This parasite is transmitted by a number of species of *Rhipicephalus* (*R. decoloratus, R. microplus* and *R. annulatus*) (Jongejan and Uilenberg, 2004). No *Babesia bigemina* was detected in any of the African buffalo examined in this study, indicating that African buffalo are unlikely to be the source of *Babesia bigemina* in the sentinel cattle.

7.4.4 Heterogeneity of *Theileria* species in African buffalo

In previous studies, allelic variation has been identified in the 18S rRNA gene of *T. parva*, *T. sp (buffalo)*, *T. mutans* and *T. velifera* from African buffalo from different National Parks and ranches in South Africa (Chaisi et al., 2013; Chaisi et al., 2011; Mans et al., 2011a). The *Theileria parva* and *T. sp (buffalo)* sequences were closely related, differing at eleven nucleotide positions over the 18S rRNA gene over the length of 1,505 bp. One to three nucleotide differences were detected among allelic variants of *T. parva,* whereas, between one and ten nucleotide differences were detected between alleles of the *T. sp (buffalo)* group (Chaisi et al., 2013; Mans et al., 2011a). Four *T. mutans*-like alleles were found, including a previously obtained sequence variant which was named *T. sp* (Strain MSD) (Chae et al., 1999). The other variants identified in the present study were named *T. mutans*-like 1, *T. mutans*-like 2 and *T. mutans*-like 3. Between 18 and 30 nucleotide differences were observed among different *T. mutans*-like sequences over a region of 1,562 base pairs. For *T. velifera*, between one and seven base pair differences were found between the *T. velifera*-like sequences (Chaisi et al., 2011; Mans et al., 2011a).

In this study, all buffalo were found to be infected with *T. parva, T. sp (buffalo)* and *T. velifera*. The amplicons sequenced were shorter than those in previous studies (~375 base pairs), but over this region the *T. parva* and *T. sp (buffalo)* species possessed a very similar 18S rRNA gene sequence to sequences reported in previous studies (Chaisi et al., 2011; Mans et al., 2011a; Sibeko et al., 2008). *Theileria mutans*, *T. mutans*-like 1, *T. mutans* like-2, *T. mutans*-like 3 sequences were obtained from buffalo from the Kruger National Park and these had been previously found in South Africa buffalo (Chaisi et al., 2013; Mans et al., 2011a). In this study, more

than one *T. mutans*-like variant was found within an individual animal. In all the samples from African buffalo from the Ol Pejeta conservancy, only one *T. velifera*-like variant was found, which had 6 unique nucleotide changes over the size of the amplicons compared to previously identified *T. velifera*-like variants. Polymorphism was found in 11 nucleotides over the amplicons size when looking at all the *T. velifera*-like sequences for African buffalo from the Ol Pejeta conservancy and Kruger National Park combined.

There are no universal criteria for species classification on the basis of variation in the 18S rRNA gene (Norval et al., 1991a) and no data are available on the morphology of the parasites, their possible vectors or their role in clinical disease that would support speciation. Thus, Chaisi *et al.* suggested that the variants of the different *Theileria* species should be classified as variants of the *Theileria* species (Chaisi et al., 2011) until further evidence for speciation is available. Ultimate species definition would require the loss of capability of sexual recombination in the tick vector, if transmitted by the same tick species. However, should parasites be transmitted by separate tick species, this in itself would imply a degree of differentiation has already taken place and there would clearly exist a barrier to sexual recombination. This scenario would facilitate parasite populations evolving in isolation, a driving force for speciation. Further studies of the biology of these parasites coupled with genetic analyses are required to address this question.

The lack of diversity of *Theileria velifera* and *T. mutans* species in African buffalo from the Ol Pejeta conservancy could have a number of reasons. Firstly, the Ol Pejeta conservancy is relatively small and is well fenced. This could have resulted in small founder populations of *Theileria spp* establishing at the conservancy. The effects of genetic drift are more pronounced in small populations and certain variants of a species might be able to become dominant or disappear from a population due to random fluctuations and competition for tick vectors. The Ol Pejeta conservancy has a large wildlife-proof electric fence and the integrity of the fence is checked daily, preventing the introduction of parasites by wildlife outside the conservancy. However, cattle from neighbouring communities are fattened and slaughtered on the conservancy. Thus, the introduction of ticks with new *Theileria* genotypes is likely to

be limited to those brought in with cattle. The animals are generally healthy animals and thus only parasites which can induce a carrier state in cattle will be introduced. Even though ticks in theory could move across the fence, in the absence of movement of mammalian hosts this will be at a low level. Cattle at fenced wildlife-livestock interfaces in southern African were reported not to have *T. parva*, whereas animals in unfenced areas were infected (Caron et al., 2013), indicating that the transmission of *T. parva* and presumably other tick-borne pathogens is limited in the presence of adequate fencing. Secondly, the limited diversity of *Theileria sp* could be possibly due to differences in availability of the different tick vectors at the Ol Pejeta conservancy, due to environmental factors, due to the relatively small area of the conservancy or due to the changes in tick population due to the regular acaricide treatment of all cattle on the conservancy. Around 7,000 head of cattle are present in the conservancy, which receive frequent acaricide treatment (every three to five days depending on the season). The rigorous acaricide treatment regime could greatly reduce the tick burden on the conservancy and impact on the tick populations present in the conservancy and consequently the transmission dynamics of tick-borne pathogens.

7.4.5 Tick vector of *T. sp (buffalo)*

Theileria sp (buffalo) was not detected in two *T. parva* tick stabilates derived previously by feeding clean ticks on naturally infected buffalo. No information is available on whether the buffalo from which the stabilates were derived were infected with *T. sp (buffalo).* However, it is of note that one of the stabilates was prepared from a buffalo captured on the Ol Pejeta site in the mid-1980s (Morzaria et al., 1995). Moreover, each of the buffalo examined in the present study, including the animals on Ol Pejeta, were infected with *T. sp (buffalo)*, as were most buffalo in previously published papers (Mans et al., 2011a; Oura et al., 2011a). In this study, the proportion of reads corresponding to *T. sp (buffalo)* and *T. parva* varied between individual buffalo. However, for the majority of the buffalo from Ol Pejeta, the number of reads obtained for *T. sp (buffalo)* and *T. parva* were roughly equivalent. In contrast, for African buffalo from the Kruger National Park, the number of sequences obtained for *T. sp (buffalo)* was many times greater than the number of sequences

obtained for *T. parva.* Thus, although the results of the current study do not exclude the possibility that *R. appendiculatus* can act as a vector for *T. sp (buffalo)*, they suggest that it is transmitted by another tick species, with differential level of challenge in each area.

Theileria sp. (buffalo) has been found in buffalo in South Africa, Kenya and Uganda and often occurs in combination with *T. parva*. This would indicate that the tick vector (if not *R. appendiculatus* itself) requires climatic conditions similar to *R. appendiculatus*, has a wide geographical distribution, is relatively abundant and frequently feeds on buffalo. Pienaar *et al.* have suggested *Amblyomma hebraeum, Amblyomma marmoreum, Hyalomma truncatum*, *Rhipicephalus appendiculatus, Rhipicephalus evertsi evertsi* or *Rhipicephalus simus* as potential vectors for *T. sp (buffalo)* based on the fact that these tick species were found on African buffalo at the Kruger National Park (Horak et al., 2007; Pienaar et al., 2011). A large number of other tick species are present in both Eastern and Southern Africa (see Table 7.9) (Anderson et al., 2013; Fyumagwa et al., 2011; Horak et al., 2007; Kariuki et al., 2012; Lynen et al., 2007; Pascucci et al., 2011; Walker et al., 2001) and could be potential vectors for *T. sp (buffalo),* in addition to the vectors suggested by Pienaar *et al.* (Pienaar et al., 2011). Further studies to analyse ticks collected from field sites where *T. sp (buffalo)* is known to be present are required to identify the vector(s) of this parasite.

Table 7.9 Tick species found in East and Southern Africa

East Africa	Reference	Southern Africa	Reference
Argas persicus	*W*	*Argas persicus*	*W*
		Argas walkerae	*W*
Amblyomma gemma	*K W L F*		
Amblyomma hebraeum	*L*	*Amblyomma hebraeum*	*A W H*
Amblyomma lepidum	*K W L F*		
Amblyomma variegatum		*Amblyomma variegatum*	*W H P*
Haemaphysalis leachi	*W L*	*Haemaphysalis leachi*	*W*
		Haemaphysalis aciculifer	*H*
Haemaphysalis silacea	*L*	*Haemaphysalis silacea*	*H*
Hyalomma albiparmatum	*K L*		
Hyalomma dromedarii	*W*		*H*
		Hyalomma glabrum	
Hyalomma impeltatum	*W L*		*H W*
Hyalomma marginatum rufipes	*K W L*	*Hyalomma marginatum rufipes*	*H W P*
Hyalomma truncatum	*K W L*	*Hyalomma truncatum*	*W P*
		Hyalomma marginatum turanicum	*W*
		Ixodes pilosusgroup	*H W*
		Ixodes rubicundus	*H W*
		Margaropus winthemi	*H W*
Ornithodoros moubata	*W*	*Ornithodoros moubata*	*W*
Ornithodoros savignyi	*W*	*Ornithodoros savignyi*	*W*
Otobius megnini	*W*	*Otobius megnini*	*W*
Rhipicephalus bequarti	*L*		
		Rhipicephalus capensis	*H*
Rhipicephalu compositus	*F*		
Rhipicephalus decoloratus	*W F L*	*Rhipicephalus decoloratus*	*H W*
		Rhipicephalus duttoni	
Rhipicephalus geigyi	*W*		
Rhipicephalus microplus	*W L*	*Rhipicephalus microplus*	*H W*
Rhipicephalus appendiculatus	*W L*	*Rhipicephalus appendiculatus*	*A H W*
Rhipicephalus evertsi evertsi	*K W L F*	*Rhipicephalus evertsi evertsi*	*A H W P*
		Rhipicephalus evertsi mimeticus	*H*
		Rhipicephalus exophthalmos	*H*
		Rhipicephalus follis	*H*
		Rhipicephalus gertrudae	*H*
		Rhipicephalus glabroscutatum	*H*
Rhipicephalus humeralis	*K*		
Rhipicephalus hurti	*L*		
Rhipicephalus interventus	*L*		
Rhipicephalus kochi	*K L*		
		Rhipicephalus longiceps	*H*
		Rhipicephalus lounsburyi	*H*
Rhipicephalus lunulatus	*W L*	*Rhipicephalus lunulatus*	*H W*
		Rhipicephalus maculatus	*H*
Rhipicephalus muehlensi	*K*	*Rhipicephalus muehlensi*	*H*
Rhipicephalus Muhsamae	*W*		
Rhipicephalus praetextatus	*W L F*		
Rhipicephalus pravus	*K W L*	*Rhipicephalus pravus* group	*H*
Rhipicephalus pulchellus	*K W L F*		
Rhipicephalus sanguineus.	*K W F*	*Rhipicephalus sanguineus*	*W*
Rhipicephalus senegalensis	*W*		
Rhipicephalus simus	*W L*	*Rhipicephalus simus*	*H W*
		Rhipicephalus supertritus	*H*
Rhipicephalus turanicus	*W*	*Rhipicephalus turanicus*	*W*
Rhipicephalus zambeziensis	*W*	*Rhipicephalus zambeziensis*	*W*

A= Anderson et al., 2013 F=Fyumagwa et al., 2011 H=Horak et al., 2007 L= Lynen et al., 2007 W=Walker et al., 2001

Chapter 8: General discussion

The most significant findings of this study were:

- The three components of the *T. parva* Muguga Cocktail vaccine show limited genotypic and antigenic diversity.
- Extensive diversity was found for buffalo-derived *T. parva* both at the population level and within individual animals.
  - Buffalo parasites examined from South Africa and Kenya showed a similarly high level of diversity.
  - A subset of $CD8^+$ T cell antigens examined exhibited antigenic polymorphism.
  - Other $CD8^+$ T cell antigens were highly conserved at the amino acid level.
- *Theileria sp (buffalo)* was identified in cell lines isolated from cattle, indicating that this parasite can transform bovine lymphocytes.

8.1 Limited diversity within the Muguga Cocktail Vaccine

Based on the deep multilocus sequence typing, the diversity of the Muguga Cocktail was found to be limited; each of the components contained a predominant allele with in some cases one or two minor alleles. The Muguga and Serengeti-transformed stocks showed a high level of similarity in antigen gene sequence content and satellite DNA profiles, confirming the findings of several previous studies (Bishop et al., 2001; Geysen et al., 1999; Oura et al., 2007b). Because of this similarity, other workers have speculated that the original *T. parva* Serengeti-transformed stock has been contaminated with and supplanted by *T. parva* Muguga genotypes over the years (Bishop et al., 2001; Oura et al., 2004b). The inClusion of two parasites stocks that are genotypically very similar may not be optimal for obtaining broad immunity. In the present study, minor components found in the Serengeti-transformed stock were not detected in the Muguga stock. These minor components may represent remnants of the parasites present in the original Serengeti-transformed stock. If these

minor components are important for the induction of broad immunity, their presence in different preparations of the vaccine needs to be monitored. Information on the broad protective activity of the Muguga Cocktail is based on experiments conducted in the 1970s (Radley et al., 1975b). Because the original stocks used in these studies are no longer available, it is impossible to determine whether their composition has changed in the intervening period. Hence, there is a case for repeating some of the cross-protection experiments to determine whether the Muguga Cocktail shows broad protection similar to that reported by Radley and colleagues (Radley et al., 1975b; Young et al., 1973).

There is evidence from the current study and previous studies that there is a risk that the composition of vaccine stocks can vary between batches or with tick passage. The composition of vaccine batches produced in the same vaccine production run was very similar, but some variation was observed with tick passage. It is thus important to monitor the consistency of parasite stocks used for vaccination, especially when an additional passage through ticks is involved. The results generated from the current generation of parasite stocks in the present study provide a useful baseline with which future vaccine batches can be compared for consistency of composition.

In this study, the components of the Muguga Cocktail were characterised by deep multilocus sequencing using 454 technology and satellite DNA typing, using a modified protocol in comparison to previous studies. Deep sequencing revealed very limited diversity in the three components of the Muguga Cocktail. This technique is relatively costly and laborious with data analysis being technically challenging and requiring sophisticated bioinformatic support. Satellite DNA typing, although having somewhat lower resolution in detecting low frequency alleles, proved to be a more practical method for characterising the parasites in the Muguga Cocktail stocks. The method used in this study to determine allele sizes and relative abundance of the alleles on satellite DNA typing (capillary flow electrophoresis) was shown to be relatively reproducible between laboratories. However, deep sequencing can provide greater level of resolution when investigating the genetic composition of parasite stocks and remains an immensely useful tool for experimental research. Its resolution

in the current study was limited by the subsequent finding that some of the antigens targeted were highly conserved at the amino acid level and it is clear this approach would be more fruitful if additional polymorphic antigens were analysed.

To overcome the potential inconsistencies in the composition of the Muguga Cocktail, consideration should be given to developing an alternative cocktail vaccine based on the use of clonal parasites. Formulation of a new cocktail vaccine containing three parasite clones selected on the basis of genotypic and antigenic divergence may well provide protection comparable to that obtained with the Muguga Cocktail. The components of such a vaccine could readily be distinguished and the composition of vaccine batches monitored, thus allowing improved quality control and greater consistency of the vaccine.

8.2 Extensive diversity in buffalo-derived *T. parva*

A high level of genetic diversity was detected in *T. parva* in African buffalo both at the population level and within individual animals. Although the average number of nucleotide sequence alleles detected at each locus in individual animals was slightly lower for the Kenyan buffalo in Ol Pejeta compared to the South African animals in the Kruger National Park, the diversity was still very high in both populations. The mean distance between sequences of parasites in the two locations was very similar. No differentiation was observed for sequences from the different locations in PCoA analysis for the majority of the genes. However, a degree of geographical sub-Clustering was observed in the PCoA analysis of the Tp1 gene. A phylogenetic tree was constructed to investigate the evolutionary relationships among Tp1 alleles and this provided evidence that particular lineages have evolved separately in each population. Although population-specific lineages could not be identified among Tp5 sequences, identical alleles were identified within a multiple buffalo hosts from the same population, either the Kruger National Park or Ol Pejeta. Taken together, this suggests the two buffalo populations have differentiated and this is most readily observed in a ‘fast-evolving’ gene such as Tp1.

It has been hypothesised that South African buffalo were infected with *T. parva* at the beginning of the 20th century. This was attributed to the introduction of *T. parva* from cattle imported from Kenya for restocking purposes after the rinderpest pandemic killed the vast majority of South African cattle and buffalo (Mbizeni et al., 2013; Norval et al., 1991b; Norval et al., 1992). Given the high degree of diversity found in individual African buffalo from the Kruger National Park in the current study and the limited antigenic diversity in isolates from cattle in Kenya (Pelle et al., 2011), it seems more likely that the low number of surviving African buffalo in the Kruger National Park would have been sufficient to ensure the maintenance of a diverse parasite population.

8.3 High level of genetic diversity in buffalo-derived *T. parva* at some antigen-encoding loci

A high level of antigenic diversity in the Tp2 gene was found, as previously reported by Pelle *et al.* (2011). A large number of allelic variants of the epitopes in the Tp2 gene were identified, many of which were shared between African buffalo from the Kruger National Park and the Ol Pejeta conservancy. A number of novel variants were identified in addition to those reported by Pelle *et al.* (2011). For Tp1, although the extent of nucleotide polymorphism was lower than for Tp2, it often resulted in changes in amino acid residues and a large number of amino acid sequence variants were detected. Consequently, three alleles of the $Tp1_{214\text{-}224}$ epitope were detected, which were similar to those reported previously by Pelle *et al.* (2011).

The identification of a large number of alleles of the sequenced genes in individual animals indicates that they have acquired infection from multiple tick infection events and that ticks feeding on the animals are likely to become infected with a genotypically mixed parasite population. This will provide the opportunity for frequent sexual recombination between genetically distinct parasites in the tick, which can result in the generation of further heterogeneity in the parasite population.

8.4 Tp4, Tp5, Tp6 and Tp10 antigens are highly conserved in buffalo-derived *T. parva*

The other four genes, namely Tp4, Tp5, Tp6 and Tp10, were highly conserved at the amino acid level despite the diversity observed at the nucleotide level. Only a single allele was found for the epitopes in Tp5 and Tp10 and two alleles for the $Tp4_{328\text{-}338}$ epitope. The $CD8^+$ T cell responses induced by infection and treatment frequently display parasite strain specificity, indicating that the target antigens are polymorphic. This may reflect the greater dominance of the more polymorphic antigens in induction of the $CD8^+$ T cell responses, with responses to the conserved antigens not being of sufficient magnitude to confer protection. Further studies are required to determine the relative contribution of responses to these conserved antigens to the overall $CD8^+$ T cell response and their role in protection.

These conserved genes may represent good candidates for the development of next generation vaccines as no strain specificity in immunity would be expected. However, it is unknown which cattle MHC I genotypes are capable of presenting these epitopes and this may limit their effectiveness when applied to populations of cattle in the field. Nevertheless, if protective $CD8^+$ T cell responses could be generated using these conserved antigens, this would potentially overcome the problem of strain specificity of immunity associated with vaccination using live parasites, and enable the use of sub-unit vaccines in areas where cattle co-graze with African buffalo.

8.5 Implication of buffalo parasite diversity for vaccination

Immunisation of cattle with several cattle-derived parasites stocks, including the Muguga Cocktail, has failed to protect against challenge with buffalo-derived parasites under field and experimental conditions. Although in some cases the mortality was slightly reduced or delayed, the mortality and severity of disease remained high in the majority of cases (Cunningham et al., 1974; Latif et al., 2001; Lohre, 1978; Radley et al., 1975a; Radley et al., 1979; Snodgrass et al., 1972; Young et al., 1973). Despite these disappointing results, the Muguga Cocktail has been used

with success in cattle of pastoralist communities in the Ngoro Ngoro conservancy region in Tanzania, where buffalo are present and are infected with *T. parva* (Di Giulio et al., 2009; Martins et al., 2010). This apparent variation in the efficacy of vaccination raises questions about the *T. parva* diversity in its broadest sense and further studies are required to understand the epidemiology of the parasite in areas where the land is co-grazed by cattle and buffalo. In particular there is a need to understand parasite challenge at the level of the tick vector. The availability of epidemiological data on infections in ticks and the cattle and buffalo mammalian hosts would enable the development of mathematical models of the transmission of *T. parva* in different epidemiological settings.

An important question is - how likely are cattle in different epidemiological situations to encounter an infected tick that has acquired its infection by feeding on a buffalo compared to an infected tick that acquired infection from a cow? In areas with buffalo where no or very limited tick control is employed, it is possible that most ticks feeding on cattle may have fed in their previous stage on cattle. If the chance of an animal encountering a tick that previously fed on a buffalo is low, it can be envisaged that challenge of Muguga Cocktail-vaccinated cattle with infected ticks predominantly fed previously on cattle could boost and broaden the antigenic specificity of immunity and hence provide greater protection against challenge with both cattle-fed and buffalo-fed ticks.

In areas with buffalo where livestock management includes regular acaricide treatment of cattle, the numbers of ticks that complete feeding on cattle and are able to transmit *T. parva* will be greatly reduced. Although this may also result in reduction in overall tick burden, the proportion of infected ticks that have acquired their infections from buffalo will increase. This, coupled with reduced boosting of immunity by challenge with cattle-fed infected ticks, may result in an increased likelihood of vaccine breakthroughs in Muguga Cocktail-vaccinated animals. Vaccination with buffalo-derived parasites could be considered. Previous studies have provided evidence that vaccination of cattle by ITM with buffalo-derived *T. parva* provides protection against challenge with buffalo-derived *T. parva.* (Cunningham et al., 1974; Dolan, 1984; Latif et al., 2001; Mbizeni et al., 2013;

Mutugi et al., 1988a; Mutugi et al., 1988b; Radley et al., 1975c; Young et al., 1977b). However, in view of the requirement of buffalo for tick feeding to produce stabilates of these parasites, the practicality of this approach is questionable.

8.6 Infection of cattle with *Theileria sp. (buffalo)*

Theileria parva, T. sp (buffalo), T. velifera were detected in all African buffalo examined in this study, based on sequencing of the 18S ribosomal RNA gene. *Theileria mutans* and *T. mutans*–like sequences were detected in all buffalo from the Kruger National Park and in some animals from the Ol Pejeta conservancy. In addition, *T. sinensis* and *T. buffeli* were detected in buffalo from the Ol Pejeta conservancy and these parasites have the potential to be transmitted from buffalo to cattle. *Theileria parva* is generally considered the only species of Theileria pathogenic for cattle in Africa. However, a potential role of *T. mutans* in disease in cattle has been described (Snodgrass et al., 1972) and *T. buffeli* has been detected in cattle grazed in proximity of buffalo in Kenya and Uganda (Oura et al., 2011b; Young et al., 1992), although it was not associated with disease.

Of particular interest in the present study was the detection of *T. sp (buffalo)* in *in vitro* cell lines isolated previously from cattle with clinical disease, which had grazed alongside buffalo. This parasite was also detected in blood from one of the sentinel cattle at the Ol Pejeta conservancy. The demonstration of *T. sp (buffalo)* in cattle cell lines indicates that *T. sp (buffalo)* is capable of infecting and transforming bovine cells and may thus be involved in pathology in cattle. It is difficult to differentiate between *T. parva* and *T. sp. (buffalo)* parasites microscopically or serologically and therefore the role of *T. sp (buffalo)* in pathology in cattle may be underestimated. The establishment of *T. sp (buffalo)* sporozoite stabilates will be required to investigate the role of this parasite in disease in cattle. Examination of available buffalo-derived sporozoites stabilates would suggest *Rhipicephalus appendiculatus* is not the tick vector for *T. sp (buffalo)* and therefore further efforts to identify the tick vector of *T. sp (buffalo)* are required.

The 18S rRNA sequences of *T. parva* and *T. sp (buffalo)* are very similar and mixed parasite infection in buffalo is known to affect the accuracy of the diagnostic test for

T. parva (Pienaar et al., 2011; Sibeko et al., 2008). In this study, phylogenetic analysis of sequences obtained for Tp6, Tp7, Tp8 and the 5S rRNA gene and their orthologues in *T. sp (buffalo)* showed a clear distinction between each species. These genes could thus be considered as candidates for an improved diagnostic test for *T. parva* in South Africa.

8.7 Main conclusions

The results of these studies have demonstrated that the three *T. parva* components of the Muguga Cocktail live vaccine (Muguga, Serengeti, Kiambu 5), as well as the combined cocktail, contain very limited genotypic diversity. The vaccine was found to contain a number of minor components present at low levels within some of the individual parasite stocks. This finding has highlighted the risk that such components could be lost during tick passage of these parasites, thus altering the composition of the vaccine and potentially reducing its ability to induce broad protection against field challenge. To overcome this problem, it is proposed that an improved cocktail vaccine should be produced using a mixture of clonal parasites, to ensure consistency of parasite composition and immune protection. A mixture of three parasites carefully selected on the basis of genotypic and antigenic divergence could be used for such a vaccine.

Buffalo-derived *T. parva* genotypes showed a high level of diversity both at the population level and within individual buffalo. In particular, there was extensive antigenic diversity in the epitopes recognised by bovine $CD8^+$ T cells in the Tp1 and Tp2 antigens. However, despite allelic variation at the nucleotide level, the amino acid sequences of the remaining four genes, namely Tp4, Tp5, Tp6 and Tp10, showed a high level of conservation. The strain specificity of $CD8^+$ T cell responses in immunity in animals immunised by ITM indicate that the immune responses are directed predominantly to polymorphic antigens. The identification of antigens that are conserved between parasite isolates offers the potential to explore the possibility of induction of protective immunity using these molecules, in order to obtain protection against all parasite strains of *T. parva* including those derived from buffalo.

Appendix A

Media and solutions

Standard Culture Medium (SCM)

RPMI 1640 medium + 25 mM HEPES + L-glutamine (Gibco, Paisley, UK)

10% (heat-inactivated) foetal bovine serum (Gibco)

100 U/ml penicillin, 100 μg/ml streptomycin, 293 μg/ml L-glutamine (penicillin-streptomycin-glutamine 100 x solution) (Gibco)

5 x 10^{-4} M 2-Mercaptoethanol

Alsever's solution (10 x stock)

D-glucose	205.0g	(113.8 mM)
citric acid	5.5 g	(2.9 mM)
sodium chloride	42.0g	(71.9 mM)
tris-sodium citrate di-hydrate	80.0 g	(27.2 mM)

Made up to a volume of 1 litre with DDW. Made up to 1 x solution by filter sterilising 0.1 litre stock solution using 0.45μm Minisart® single use filter (Sartorious, Germany) and adding 0.9 litre of DDW.

FACS Medium

RPMI 1640 medium + 25 mM HEPES + L-glutamine (Gibco)

5% (heat-inactivated) foetal bovine serum (Gibco)

0.2% sodium azide

Cytotoxicity Medium

RPMI 1640 medium + 25 mM HEPES + L-glutamine (Gibco)

5% (heat-inactivated) foetal bovine serum (Gibco)

Tris-Acetate/EDTA electrophoresis buffer - TAE buffer (50 x stock)

Trizma base	242 g	(2M)
Glacial acetic acid	59.1ml	(2M)
EDTA	100ml of 0.5M solution	(50mM)

pH adjusted to 7.7-8.0 with glacial acetic acid if required. Made up to a volume of 1 litre with DDW.

Buffer for Giemsa staining

potassium di-hydrogen phosphate	3 g
di-sodium hydrogen phosphate	15 g

Made up to a volume of 5 litres with distilled water. pH adjusted to 7.2 using either sodium hydroxide or hydrochloric acid.

Stabilate diluent

Minimum essential media (MEM) with Hanks' salts, L-glutamine and nonessential amino acids (Sigma Aldrich) supplemented with sodium hydrogen carbonate (0.35 g/L)
3.5% of volume BSA – albumin bovine fraction V, pH 7, standard grade (Serva electrophoresis)
7.5% of volume glycerol, pH 7.2
1% of volume Penstrep (150 µg/ml streptomycin and 200 µg/ml penicillin)

Appendix B

Figure 1 Nucleotide alignment of Tp1 within the Muguga Cocktail and the Marikebuni stabilates

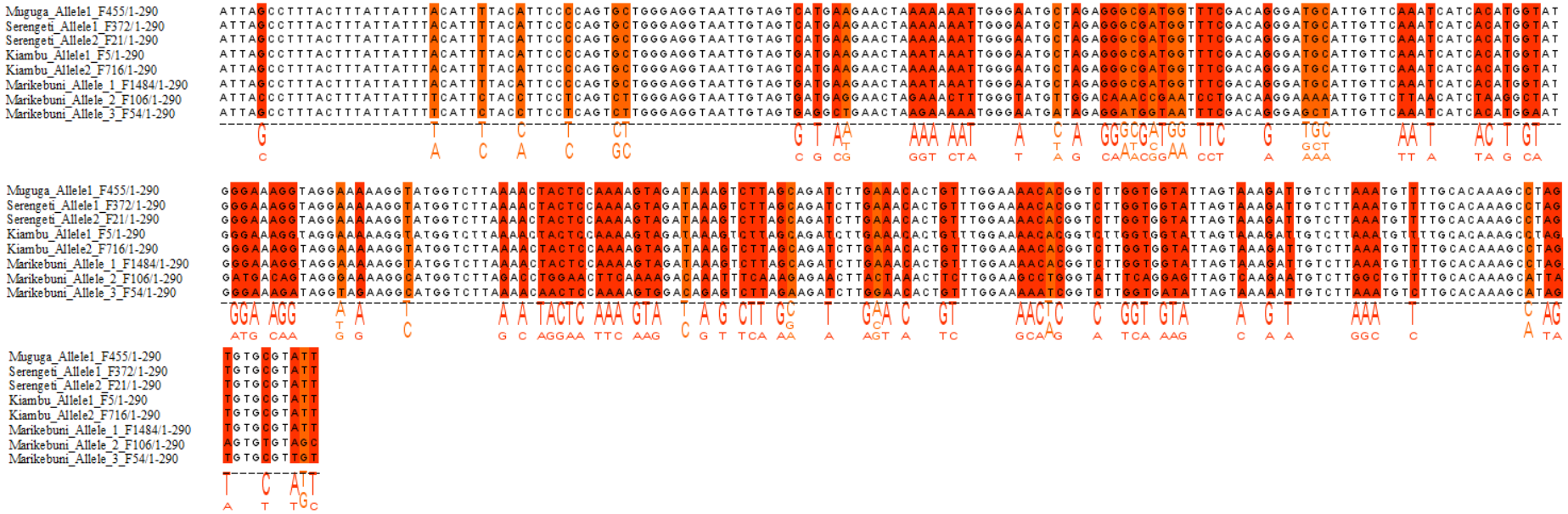

Figure 2 Nucleotide alignment of Tp2 forward reads within the Muguga Cocktail and the Marikebuni stabilates

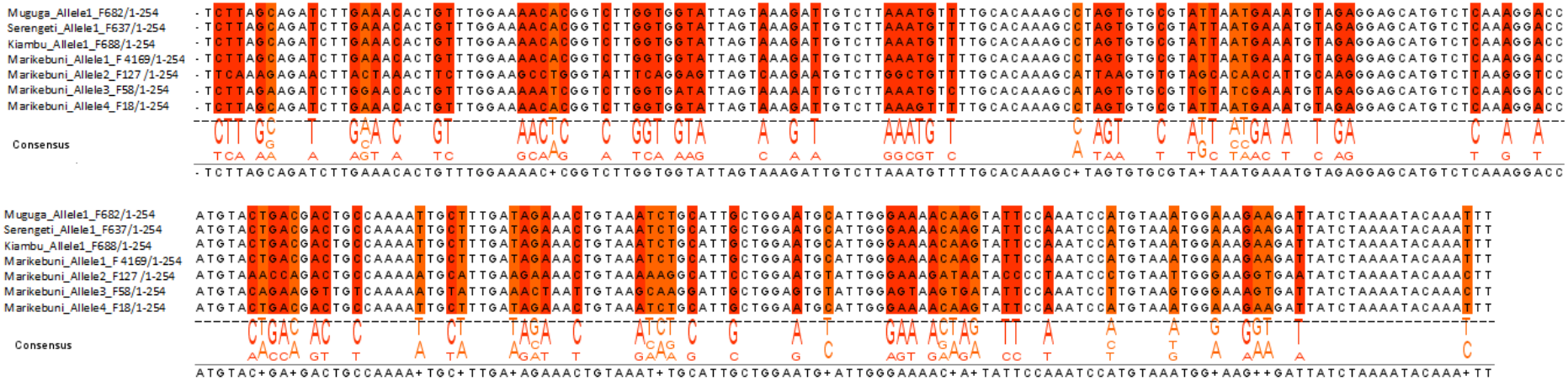

Figure 3 Nucleotide alignment of Tp2 reverse reads within the Muguga Cocktail and the Marikebuni stabilates

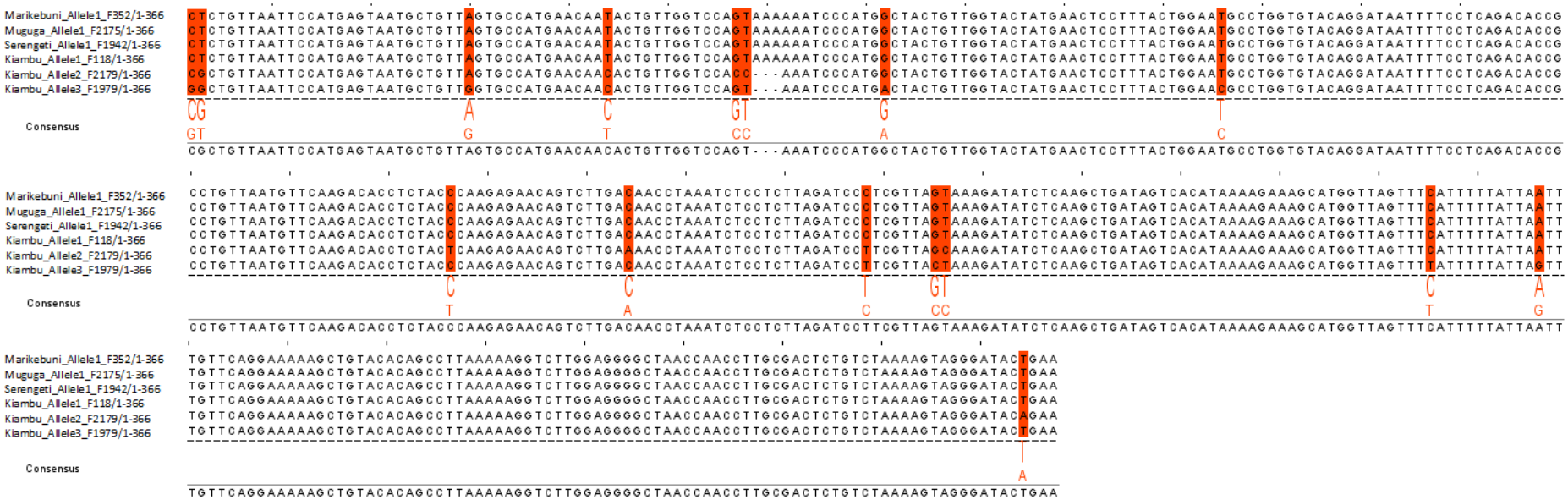

Figure 4 Nucleotide alignment of Tp3 reads within the Muguga Cocktail and the Marikebuni stabilates

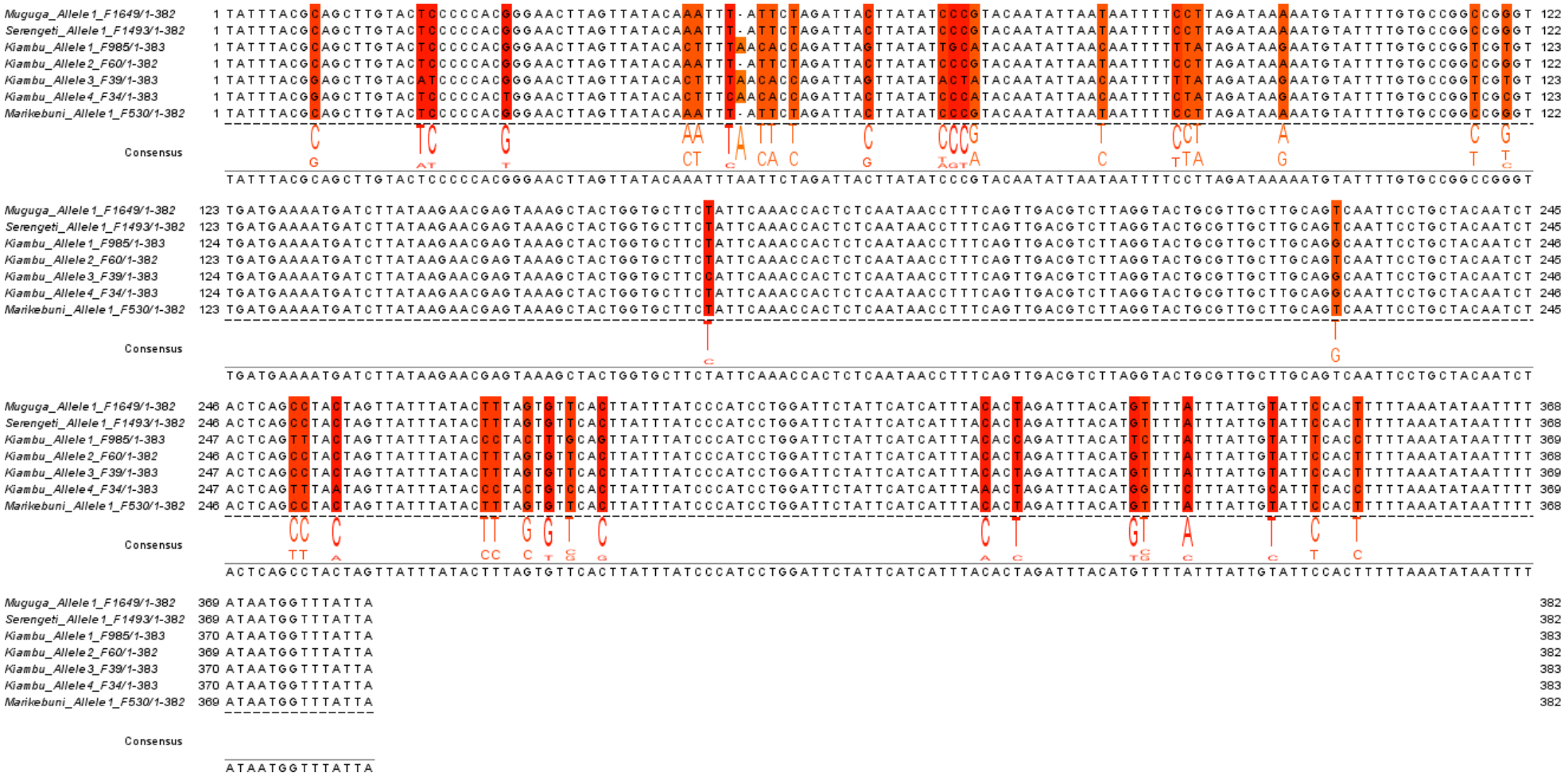

Figure 5 Nucleotide alignment of Tp4 within the Muguga Cocktail and Marikebuni stabilates

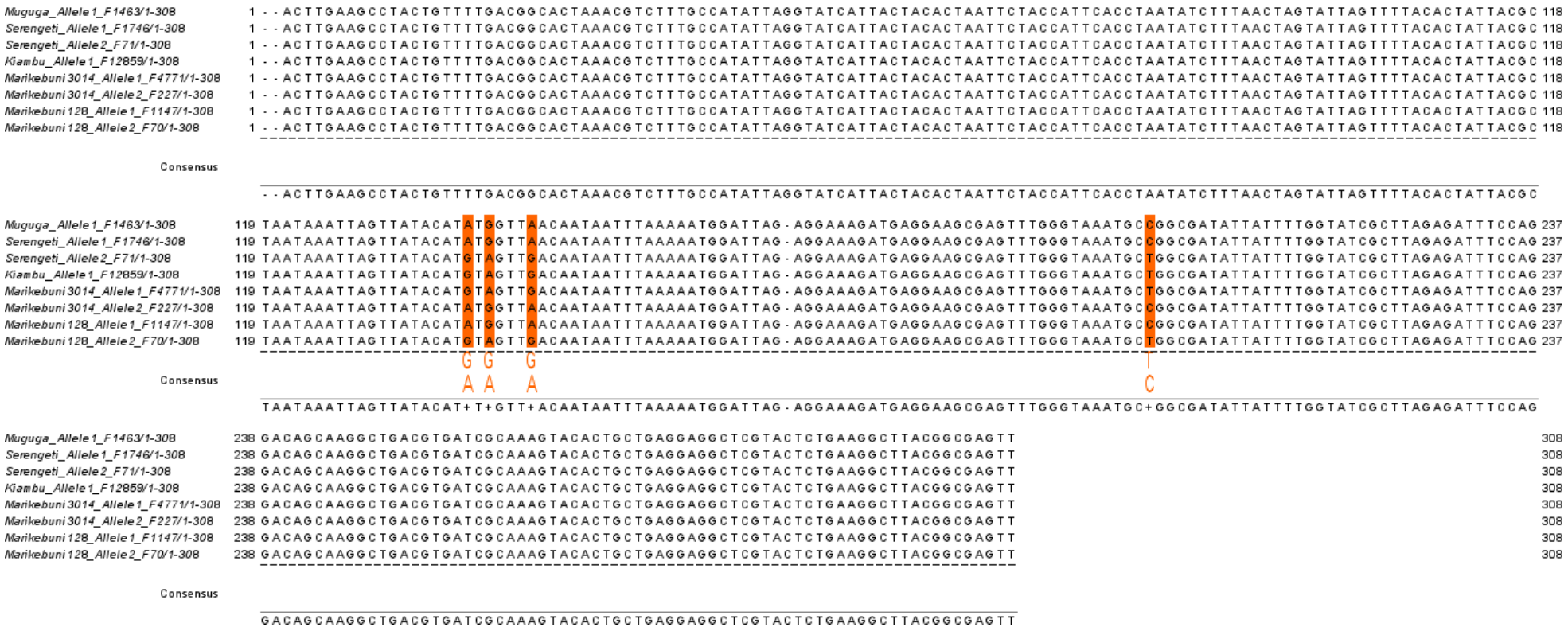

Figure 6 Nucleotide alignment of Tp5 within the Muguga Cocktail and Marikebuni stabilates

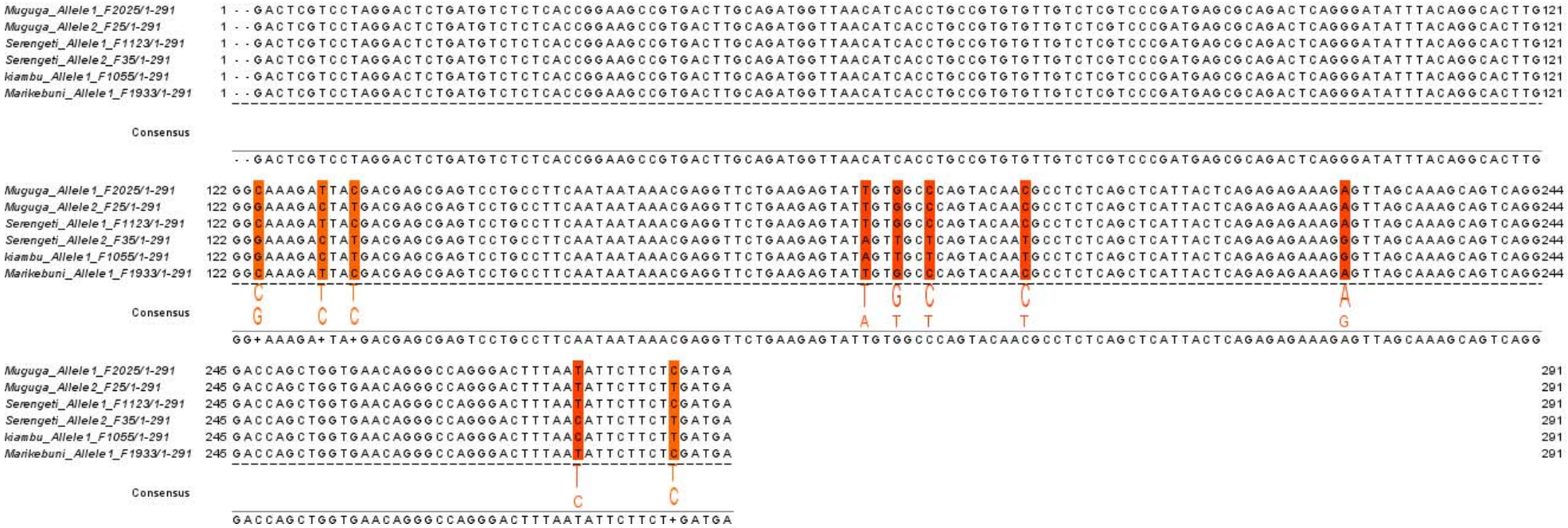

Figure 7 Nucleotide alignment of Tp6 within the Muguga Cocktail and Marikebuni stabilates

Figure 8 Nucleotide alignment of Tp7 within the Muguga Cocktail and Marikebuni stabilates

Figure 9 Nucleotide alignment of Tp8 within the Muguga Cocktail

Figure 10 Nucleotide alignment of Tp10 within the Muguga Cocktail and Marikebuni stabilates

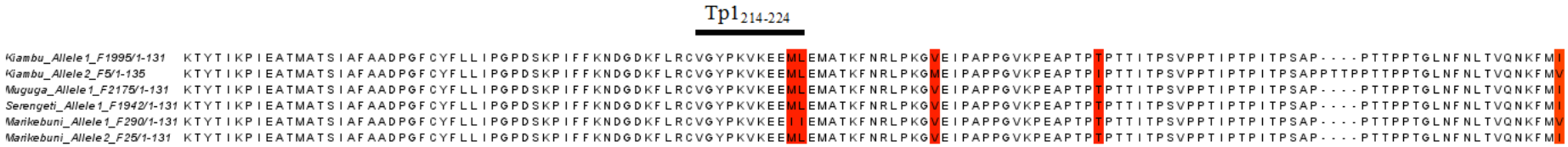

Figure 11 Amino acid alignment of Tp1 within the Muguga Cocktail and Marikebuni stabilates.
The position of the $Tp1_{214-224}$ epitope is indicated above the sequences

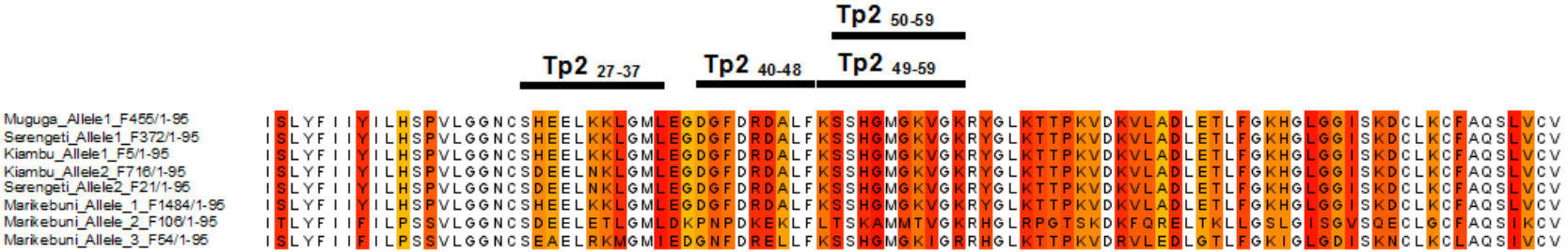

Figure 12 Amino acid alignment of Tp2 forward reads within the Muguga Cocktail and Marikebuni stabilates.
The positions of CD8 T cell epitopes is indicated above the sequences.

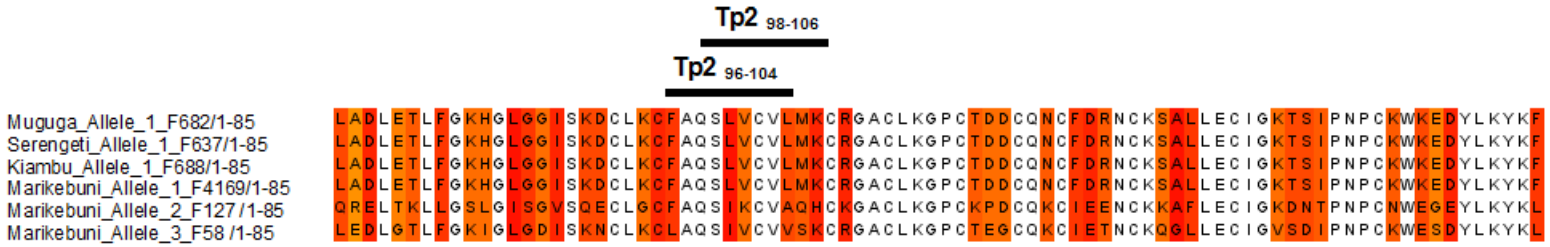

Figure 13 Amino acid alignment of Tp2 reverse reads within the Muguga Cocktail and Marikebuni stabilates.
The positions of CD8 T cell epitopes is indicated above the sequences.

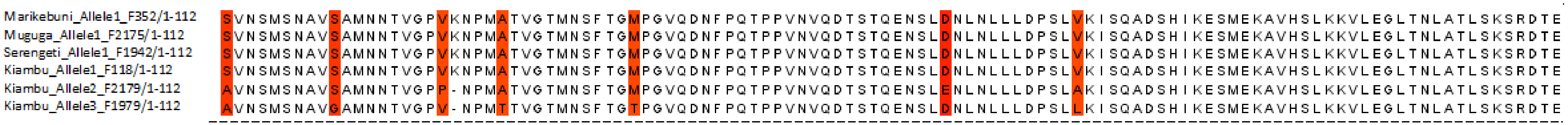

Figure 14 Amino acid alignment of Tp3 within the Muguga Cocktail and Marikebuni stabilates

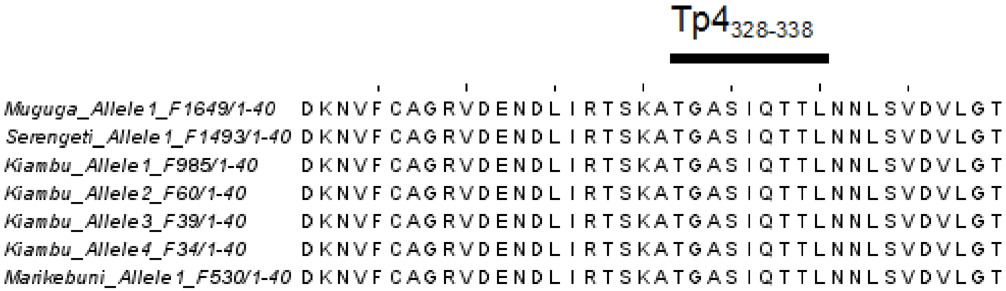

Figure 15 Amino acid alignment of Tp4 within the Muguga Cocktail and Marikebuni stabilates

The position of the $Tp4_{328\text{-}338}$ epitope is indicated above the sequences

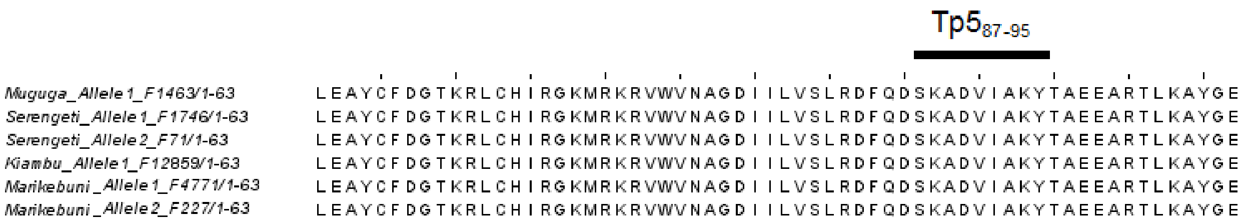

Figure 16 Amino acid alignment of Tp5 within the Muguga Cocktail and Marikebuni stabilates

The position of the $Tp5_{87\text{-}95}$ epitope is indicated above the sequences

Figure 17 Amino acid alignment of Tp6 within the Muguga Cocktail and Marikebuni stabilates

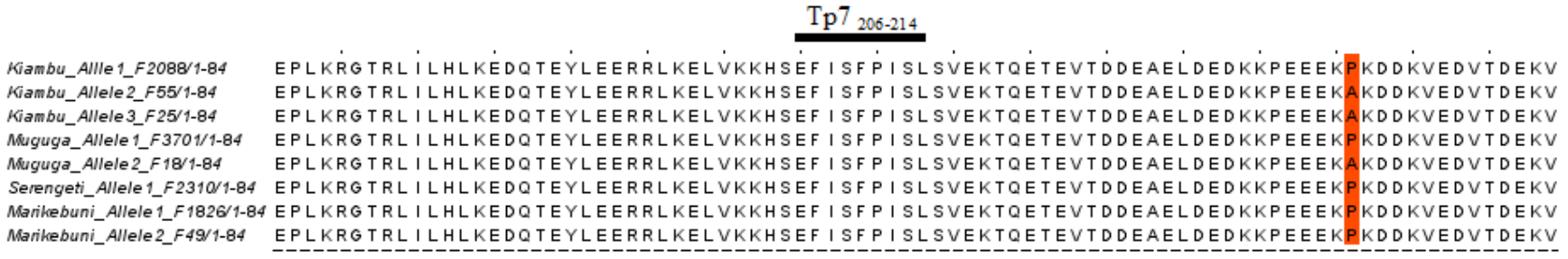

Figure 183 Amino acid alignment of Tp7 within the Muguga Cocktail and Marikebuni stabilates
The position of the $Tp7_{206\text{-}214}$ epitope is indicated above the sequences

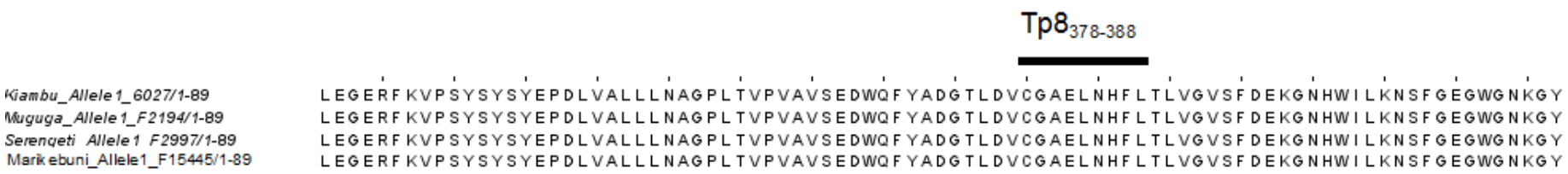

Figure 19 Amino acid alignment of Tp8 within the Muguga Cocktail and Marikebuni stabilates
The position of the $Tp8_{378\text{-}388}$ epitope is indicated above the sequences

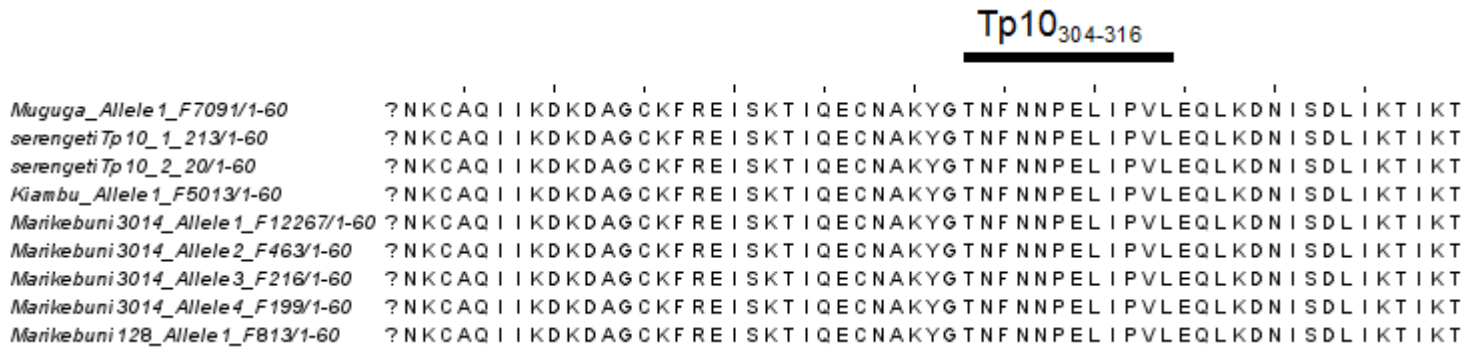

Figure 20 Amino acid alignment of Tp10 within the Muguga Cocktail and Marikebuni stabilates
The position of the $Tp10_{304-316}$ epitope is indicated above the sequences

Appendix C

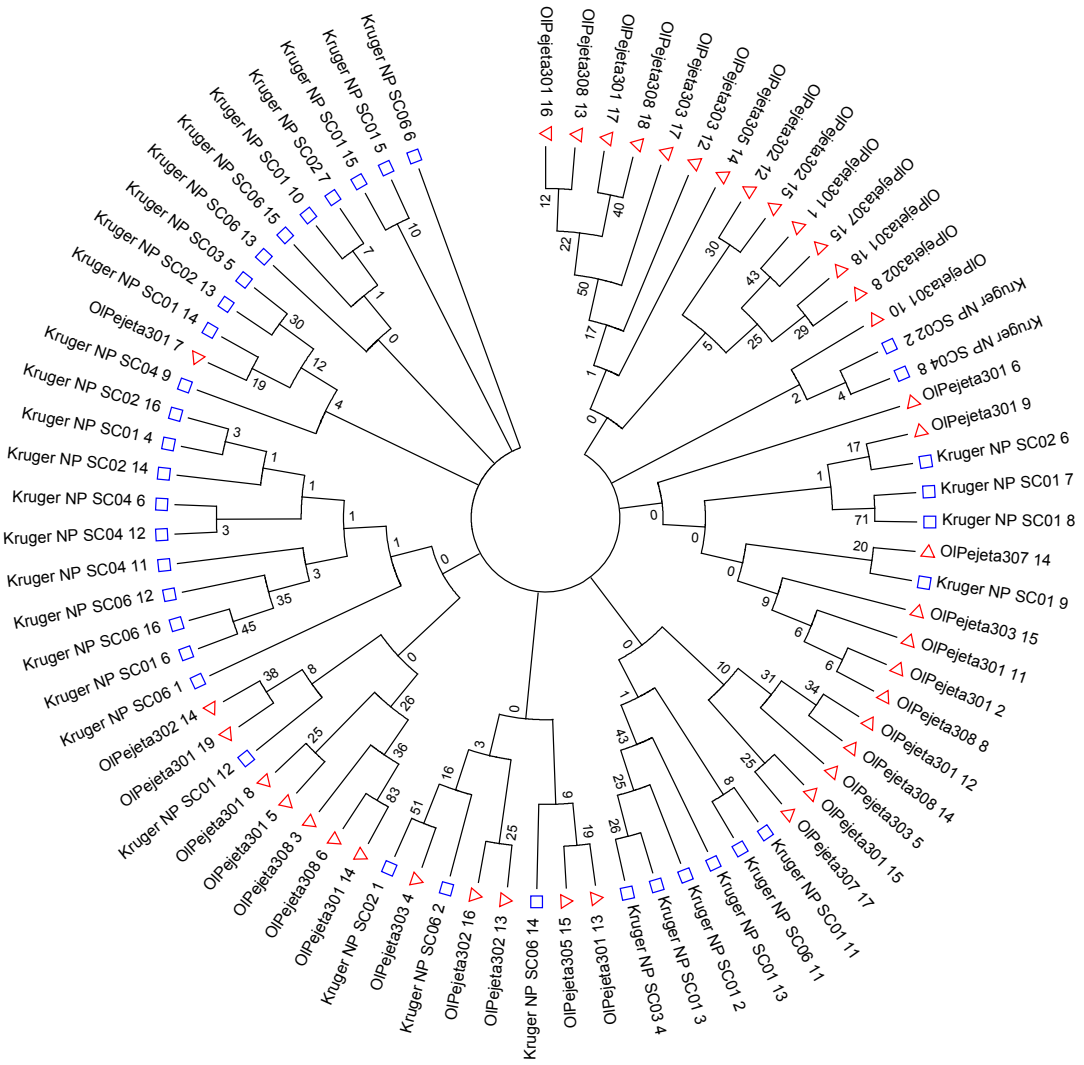

Figure 1 Phylogenetic tree showing the relationship between Tp1 (buffalo) sequences obtained from African buffalo – Maximum Parsimony method.
The phylogenetic relationship was inferred using the Maximum Parsimony method. The bootstrap consensus tree inferred from 1000 replicates is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test is shown next to the branches. There were a total of 312 positions in the final dataset. Samples from African buffalo from the Kruger national park are labelled Kruger NP, followed by the animal number (SC01-SC06), the allele number and the number of sequences obtained for the allele. Samples from the African buffalo from the Ol Pejeta conservancy are labelled Ol Pejeta, followed by the animal number (301-306) and allele number.

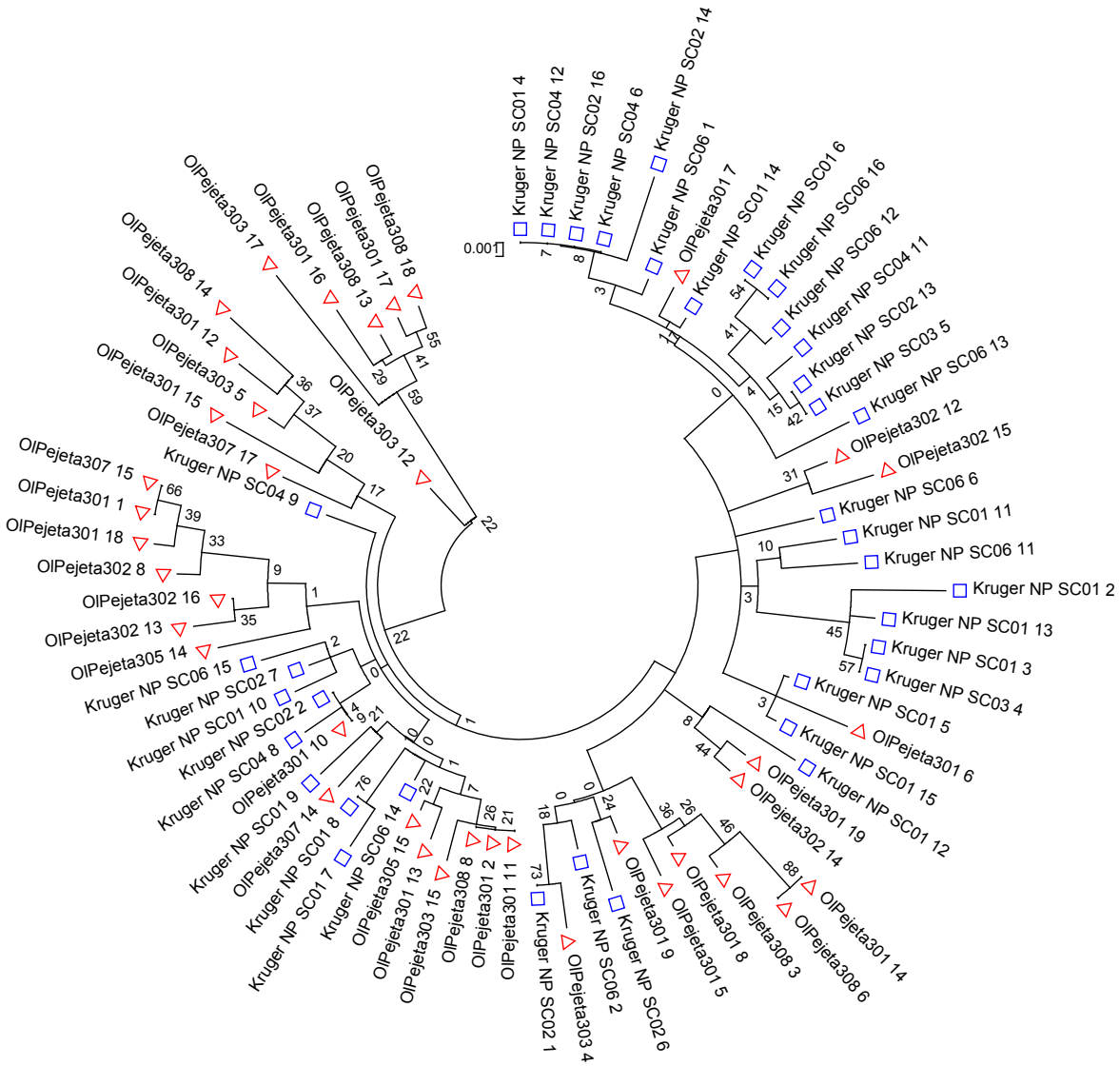

Figure 2 Phylogenetic tree showing the relationship between Tp1 (buffalo) sequences obtained from African buffalo- Neighbor Joining method

The phylogenetic relationship was inferred using the Neighbor-Joining method. The bootstrap consensus tree inferred from 1000 replicates is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test are shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Maximum Composite Likelihood method and are in the units of the number of base substitutions per site. The rate variation among sites was modelled with a gamma distribution (shape parameter = 2).). The analysis involved 76 nucleotide sequences. Samples from African buffalo from the Kruger national park are labelled Kruger NP, followed by the animal number (SC01-SC06), the allele number and the number of sequences obtained for the allele. Samples from the African buffalo from the Ol Pejeta conservancy are labelled Ol Pejeta, followed by the animal number (301-306), allele number and the number of sequences obtained.

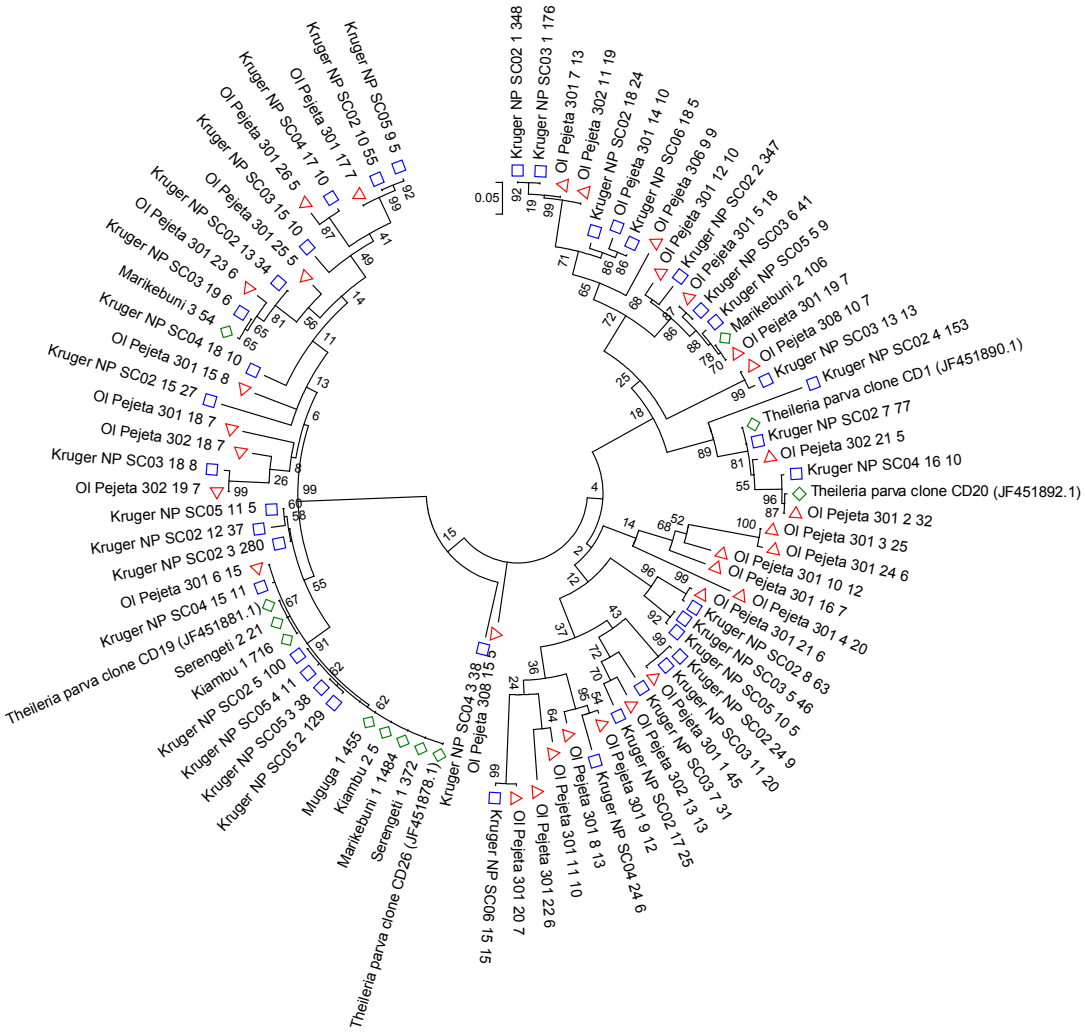

Figure 3 Phylogenetic tree showing the relationship between Tp2 (buffalo) sequences obtained from African buffalo - Maximum likelihood method.
The tree was constructed using a maximum likelihood method using 1,000 bootstrap replicates. The bootstrap values indicate the degree of support for each of the nodes. Sequences were obtained from buffalo from the Kruger National Park and the Ol Pejeta conservancy, all unique sequences obtained within the Kruger National Park or the Ol Pejeta conservancy are displayed. Samples from African buffalo from the Kruger national park are labelled □ Kruger NP, followed by the animal number (SC01-SC06), the allele number and the number of sequences obtained for the allele. Samples from the African buffalo from the Ol Pejeta conservancy are labelled Δ Ol Pejeta, followed by the animal number (301-306), allele number and the number of sequences obtained. Samples from laboratory maintained parasites are labelled ◊, followed by the stabilate name, the allele number and the number of sequences obtained. Sequences from cattle from a previous study are labelled ◊, followed by the name and the Genbank accession number.

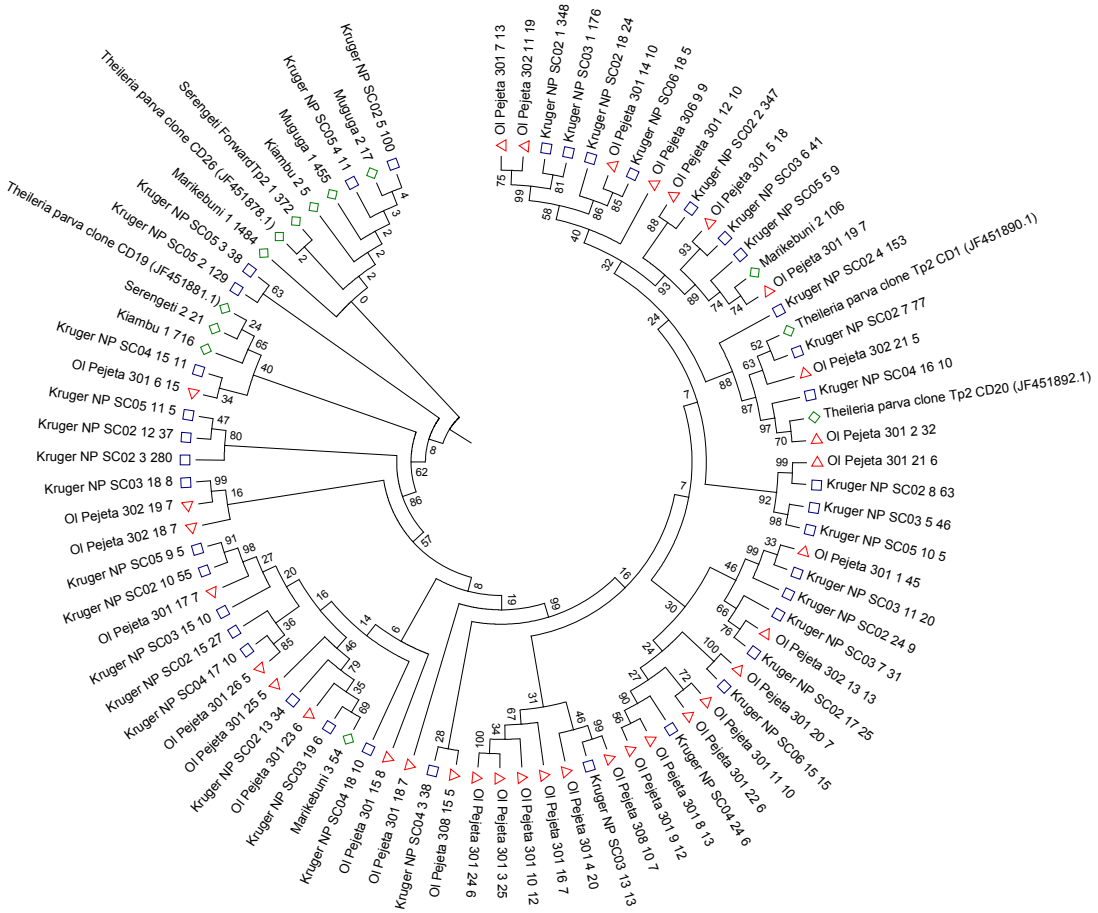

Figure 4 Phylogenetic tree showing the relationship between Tp2 (buffalo) sequences obtained from African buffalo- Maximum Parsimony method.
The phylogenetic relationship was inferred using the Maximum Parsimony method. The bootstrap consensus tree inferred from 1000 replicates is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test is shown next to the branches. The evolutionary distances were computed using the Maximum Composite Likelihood method and are in the units of the number of base substitutions per site. The rate variation among sites was modelled with a gamma distribution (shape parameter = 2). The analysis involved 84 nucleotide sequences. All positions containing gaps and missing data were eliminated. There were a total of 249 positions in the final dataset. Samples from African buffalo from the Kruger national park are labelled □ Kruger NP, followed by the animal number (SC01-SC06), the allele number and the number of sequences obtained for the allele. Samples from the African buffalo from the Ol Pejeta conservancy are labelled Δ Ol Pejeta, followed by the animal number (301-306), allele number and the number of sequences obtained. Samples from laboratory maintained parasites are labelled ◊, followed by the stabilate name, the allele number and the number of sequences obtained. Sequences from cattle from a previous study are labelled ◊, followed by the name and the Genbank accession number.

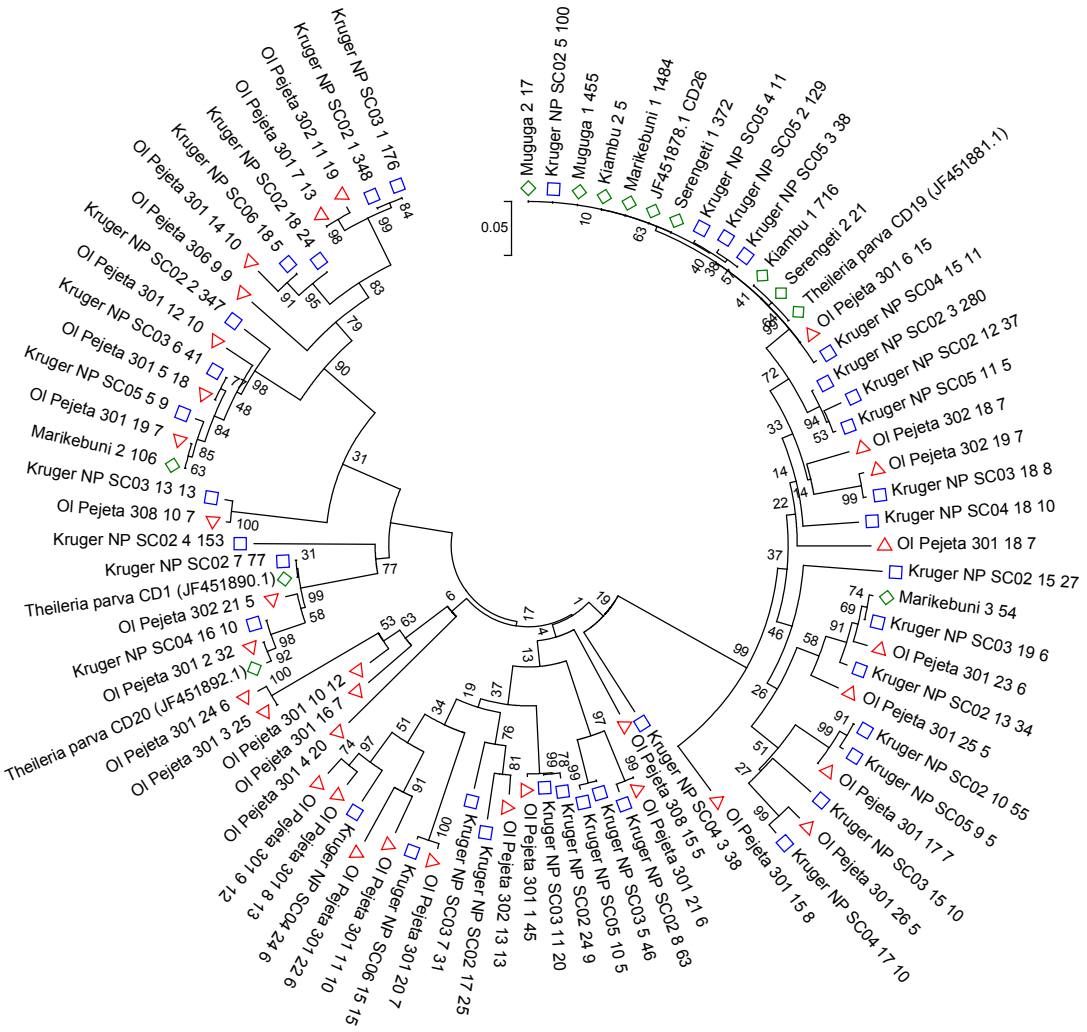

Figure 5 Phylogenetic tree showing the relationship between Tp2 (buffalo) sequences obtained from African buffalo- Neighbor Joining method

The phylogenetic relationship was inferred using the Neighbor-Joining method. The bootstrap consensus tree inferred from 1000 replicates is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test are shown next to the branches. The evolutionary distances were computed using the Maximum Composite Likelihood method and are in the units of the number of base substitutions per site. The rate variation among sites was modelled with a gamma distribution (shape parameter = 2). The analysis involved 84 nucleotide sequences. Samples from African buffalo from the Kruger national park are labelled □ Kruger NP, followed by the animal number (SC01-SC06), the allele number and the number of sequences obtained for the allele. Samples from the African buffalo from the Ol Pejeta conservancy are labelled Δ Ol Pejeta, followed by the animal number (301-306), allele number and the number of sequences obtained. Samples from laboratory maintained parasites are labelled ◊, followed by the stabilate name, the allele number and the number of sequences obtained. Sequences from cattle from a previous study are labelled ◊, followed by the name and the Genbank accession number.

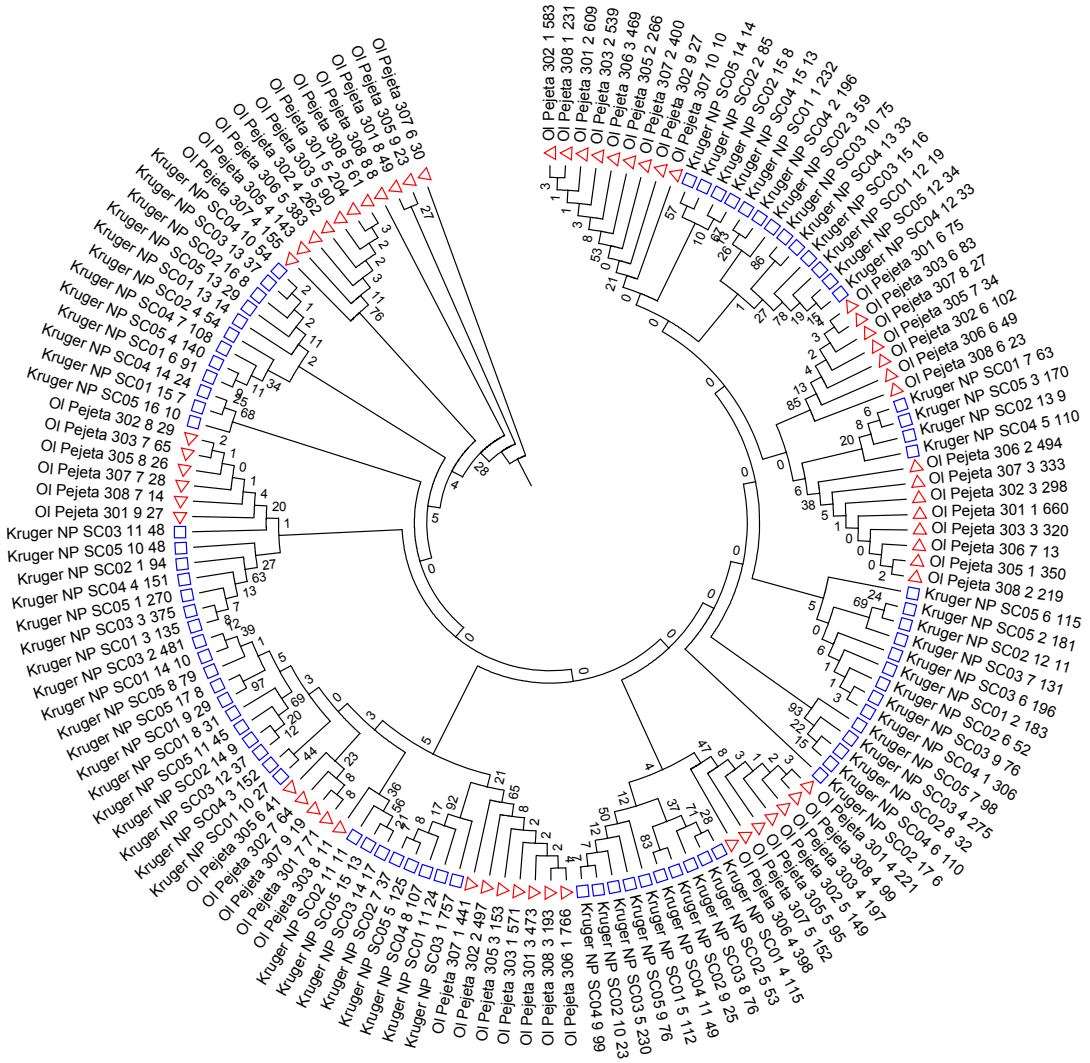

Figure 6 Phylogenetic tree showing the relationship between Tp5 (buffalo) sequences obtained from African buffalo- Maximum Parsimony method.
The phylogenetic relationship was inferred using the Maximum Parsimony method. The bootstrap consensus tree inferred from 1000 replicates is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test is shown next to the branches. There were a total of 309 positions in the final dataset. Samples from African buffalo from the Kruger national park are labelled □ Kruger NP, followed by the animal number (SC01-SC06), the allele number and the number of sequences obtained for the allele. Samples from the African buffalo from the Ol Pejeta conservancy are labelled Δ Ol Pejeta, followed by the animal number (301-306), allele number and the number of sequences obtained.

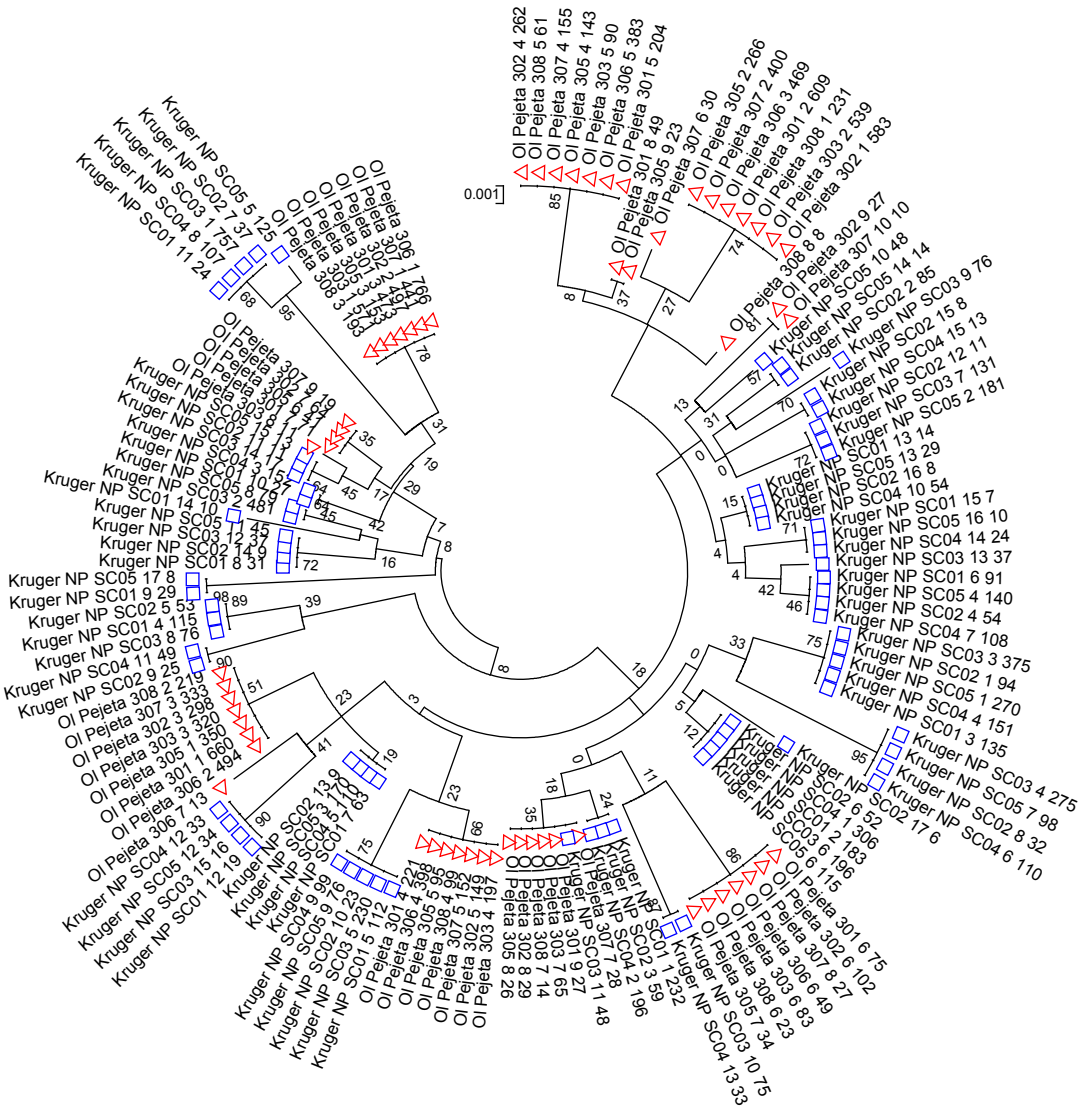

Figure 7 Phylogenetic tree showing the relationship between Tp5 (buffalo) sequences obtained from African buffalo- Neighbor-Joining method.
The phylogenetic relationship was inferred using the Maximum Parsimony method. The bootstrap consensus tree inferred from 1000 replicates is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test is shown next to the branches. The evolutionary distances were computed using the Maximum Composite Likelihood method and are in the units of the number of base substitutions per site. The rate variation among sites was modelled with a gamma distribution (shape parameter = 2). The analysis involved 139 nucleotide sequences. Samples from African buffalo from the Kruger national park are labelled □ Kruger NP, followed by the animal number (SC01-SC06), the allele number and the number of sequences obtained for the allele. Samples from the African buffalo from the Ol Pejeta conservancy are labelled Δ Ol Pejeta, followed by the animal number (301-306), allele number and the number of sequences obtained.

Appendix D

Table 4 Monitoring of parasitological parameters in lymph nodes and blood after challenge with Stabilate 3081

BG040	0	1	2	3	4	5	6	7	8	9	10	11	12
right parotoid ln	-	-	-	-	-	-	-	-	-	-	-	-	-
left parotoid ln	-	-	-	-	-	-	-	H++	H++	H++	H++	H++	MA+
right prescapular ln	-	-	-	-	-	-	-	-	-	-	-	-	-
left prescapular ln	-	-	-	-	-	-	-	-	-	-	-	-	-
blood		-	-	-	-	-	-	-	-	-	NPS	NPS	-

13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
H+	MA+	MA+	MA+	-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
H++	H++	MA+	MA+	-	-	-	-	-	-	-	-	-	-	-
-	NPS	NPS	-	-	NPS	NPS	-	-	NPS	-	-	-	-	-

BG044	0	1	2	3	4	5	6	7	8	9	10	11	12
right parotoid ln	-	-	-	-	-	-	-	-	-	-	-	-	-
left parotoid ln	-	-	-	-	-	-	-	MA+	MA++	MA+	H+	MA+	MA+
right prescapular ln	-	-	-	-	-	-	-	-	-	-	-	-	-
left prescapular ln	-	-	-	-	-	-	-	-	H+	H+	H+	-	-
blood		-	-	-	-	-	-	-	-	-	NPS	NPS	NPS

13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
MA+	MA+	MA++	-	-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
MA++	MA+	MA+	-	-	-	-	-	-	-	-	-	-	-	-
-	NPS	NPS	-	-	NPS	-	-	-	NPS	-	-	-	-	-

BF080	0	1	2	3	4	5	6	7	8	9	10	11	12
right parotoid ln	-	-	-	-	-	-	-	-	-	-	-	-	-
left parotoid ln	-	-	-	-	-	-	-	-	H+	MA+	MA+	H+	H+
right prescapular ln	-	-	-	-	-	-	-	-	-	-	-	-	-
left prescapular ln	-	-	-	-	-	-	-	-	-	-	-	-	-
blood	-	-	-	-	-	-	-	-	-	-	-	-	-

13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
-	-	-	-	-	-	-	-		-	-	-	-	-	-
H+	H++	H+	-	MA+	H+	H+	H+	H+	-	MA+	H+	-	-	-
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	NPS	NPS	NPS	-	-	NPS	-	-	-	-	-

BF083	0	1	2	3	4	5	6	7	8	9	10	11	12
left parotoid ln	-	-	-	-	-	-	-	-	-	H++	H++	H+	H++
right prescapular ln	-	-	-	-	-	-	-	-	-	-	-	-	-
left prescapular ln	-	-	-	-	-	-	-	-	-	-	-	-	-
blood		-	-	-	-	-	-	-	-	-	-	-	-

13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
H++	H++	H++	H+	H++	H++	H+	H+	H+	H+	H+	H+	-	-	-
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-	+	-	-	-	-

BF085	0	1	2	3	4	5	6	7	8	9	10	11	12
right parotoid ln	-	-	-	-	-	-	-	-	-	-	-	-	-
left parotoid ln	-	-	-	-	-	-	-	-	-	-	-	H+	H+
right prescapular ln	-	-	-	-	-	-	-	-	-	-	-	-	-
left prescapular ln	-	-	-	-	-	-	-	-	-	-	-	-	-
blood	-	-	-	-	-	-	-	-	-	-	-	-	-

13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-	H+	H+	-	H+	H+	-	-	H+	H++	H++	H++	-	-	-
-	-	-	-	-	-	-	-	-		-	-	-	-	-
-	-	-	-	-	-	-	-	-		-	-	-	-	-
-	NPS	-	-	-	NPS	-	-	-	NPS	NPS	NPS	-	-	-

BF086	0	1	2	3	4	5	6	7	8	9	10	11	12
right parotoid ln	-	-	-	-	-	-	-	-	-	-	-	-	-
left parotoid ln	-	-	-	-	-	-	-	-	-	-	H+	H+	H++
right prescapular ln	-	-	-	-	-	-	-	-	-	-	-	-	-
left prescapular ln	-	-	-	-	-	-	-	-	-	-	-	-	-
blood		-	-	-	-	-	-	-	-	-	-	-	-

13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
MA+	MA+	MA+	MA++	MA+	MA+	-	H+	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-	MA+	MA+	MA+	-	-	-	-	-	-	-	-	-	-	-
-	NPS	NPS	-	-	NPS	-	-	-	NPS	-	-	-	-	-

Animals were monitored for 27 days after challenge and the day of measurement is shown. The presence of parasites was examined in lymph nodes (ln) and blood. H+ slight hyperplasia of lymphocytes, H++ mild hyperplasia of lymphocytes, Ma+ schizonts not easy to find, Ma++ 1 to 2 schizonts visible per field, Ma +++ > 2 schizonts per field, NPS no parasites seen.

Table 5 Body temperature of animals after challenge with stabilate 3081

Animal	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
BG040	38.2	39.2	38.2	38.2	38.6	39.4	38.3	38.2	39.7	38.7	40.6	38.6	38.6	39.2	39.3
BG044	38.6	38.6	38.4	38.5	38.6	39	39.1	39.3	39.2	39.2	39.1	38.9	39.5	39.6	41.1
BF080	38.5	39.1	38.3	38.2	38.4	38.4	38.1	37.8	38.3	38.5	38.8	38.8	38.5	38.1	38.1
BF083	38.5	39.3	38.3	38.3	37.9	38.8	38.2	38.2	38.3	38.7	38.9	38.7	38.4	38.3	38.1
BF085	38.5	38.9	38.3	38.2	38.4	38.5	38.2	38.4	38.4	38.8	38.8	38.8	38.7	38.4	38.5
BF086	38.2	39.6	38.1	38.1	38.4	38.4	37.9	38.7	38.6	38.5	38.6	38.8	38.7	38	39.6

Animal	16	17	18	19	20	21	22	23	24	25	26	27	28	29
BG040	40.9	40.9	40.4	39.7	39.1	38.7	38.7	38	38.4	38.2	38.3	38.4	38.3	38.1
BG044	41.1	39.4	39.1	38.9	38.6	38.8	38.7	38.2	38.7	38.3	38.7	38.4	38.4	38.4
BF080	38.6	38.9	40.3	41.2	40.2	38.8	38.6	38.6	39.1	39	38.7	38.6	38.7	38.1
BF083	39.1	39.1	38.9	40.1	40.5	40.1	39.5	40.1	40	40.1	39.6	39.1	39	38.6
BF085	39	38.7	38.9	38.7	38.5	38.6	38.6	38.9	39.8	40	38	38.5	38.6	38.4
BF086	41	40.8	39.7	39.6	38.9	38.5	38.3	38.4	38.4	38.4	38.5	38.6	38.7	38.3

The body temperature for each of the animals is shown in °C

Appendix E

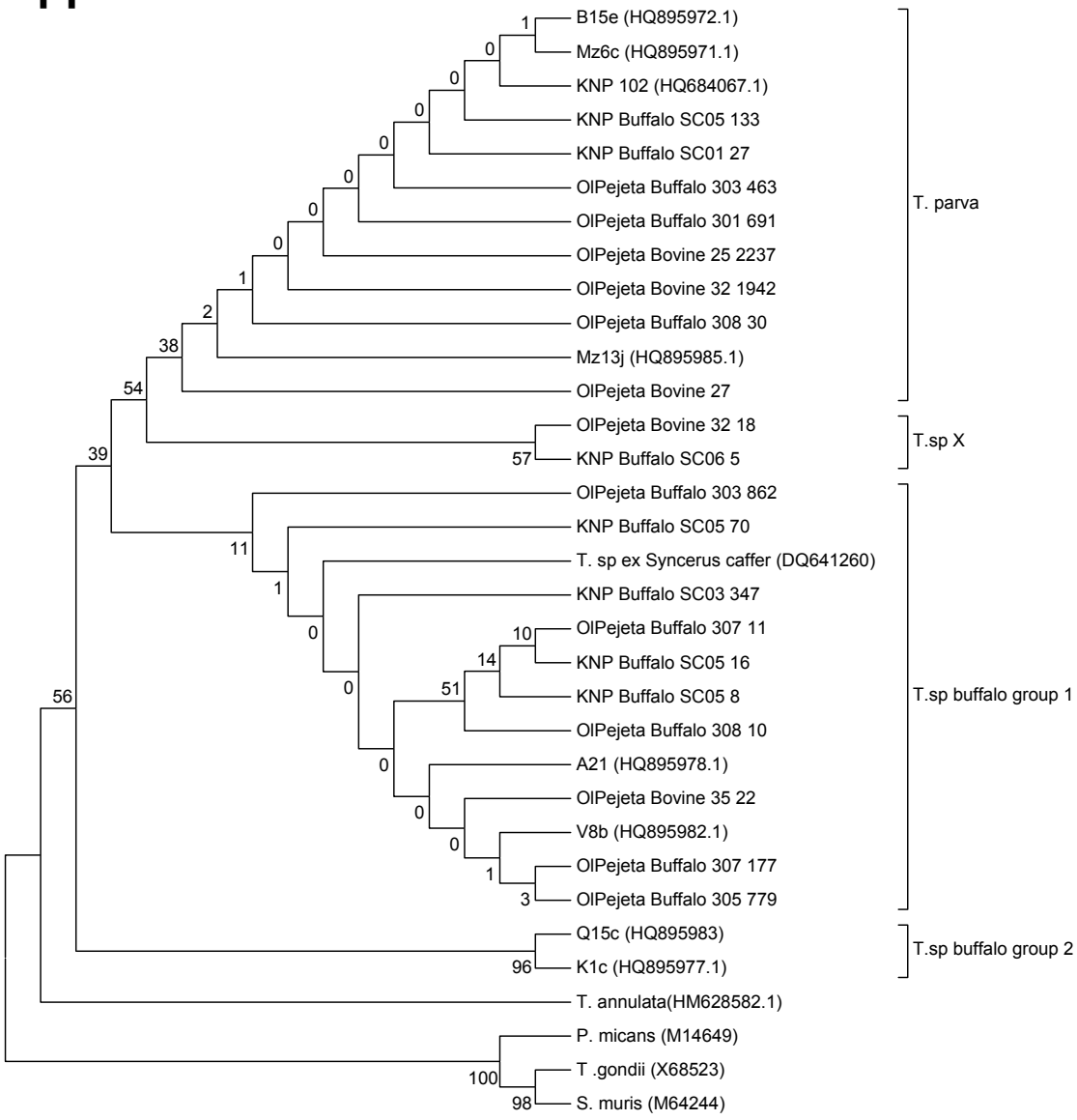

Figure 1 Phylogenetic tree showing the relationship between *T. parva* and *T.sp (buffalo)* sequences obtained from African buffalo.

The phylogenetic relationship was inferred using the Maximum Parsimony method. The bootstrap consensus tree inferred from 1000 replicates is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test is shown next to the branches. The tree was rooted using rRNA gene sequences for *Prororectum micans* (M14649), *Sarcocystic muris* (M64244), *T. gondii* (X68523),*T. parva* (HQ684067.1) and *T. annulata (*HM628582.1) Samples from African buffalo from the Kruger national park are labelled KNP buffalo, followed by the animal number (SC01-SC06), followed by the number of sequences obtained for the allele. Samples from the African buffalo from the Ol Pejeta conservancy are labelled Ol Pejeta buffalo, followed by the animal number (301-306), followed by the number of sequences obtained for the allele. Bovine samples from the Ol Pejeta conservancy are labelled Ol Pejeta bovine, followed by the animal number and followed by the number of sequences obtained for the allele

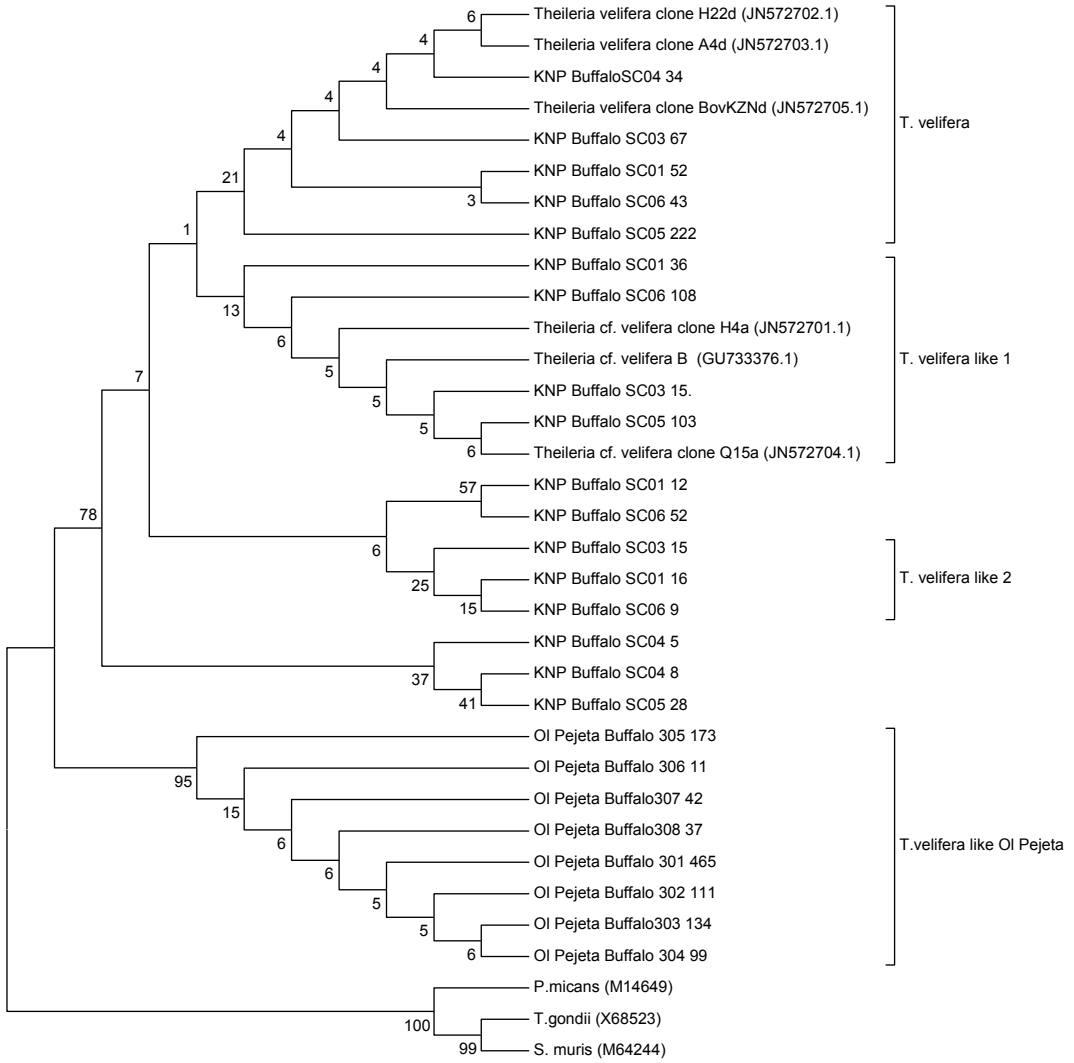

Figure 2 Phylogenetic tree showing the relationship between *T. velifera* like sequences obtained from African buffalo.

The phylogenetic relationship was inferred using the Maximum Parsimony method. The bootstrap consensus tree inferred from 1000 replicates is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test is shown next to the branches. The tree was rooted using rRNA gene sequences for *Prororectum micans* (M14649), *Sarcocystic muris* (M64244), *T. gondii* (X68523). Samples from African buffalo from the Kruger national park are labelled KNP Buffalo, followed by the animal number (SC01-SC06) and the number of sequences obtained for the allele. Samples from the African buffalo from the Ol Pejeta conservancy are labelled Ol Pejeta Buffalo, followed by the animal number (301-306), followed by the number of sequences obtained for the allele.

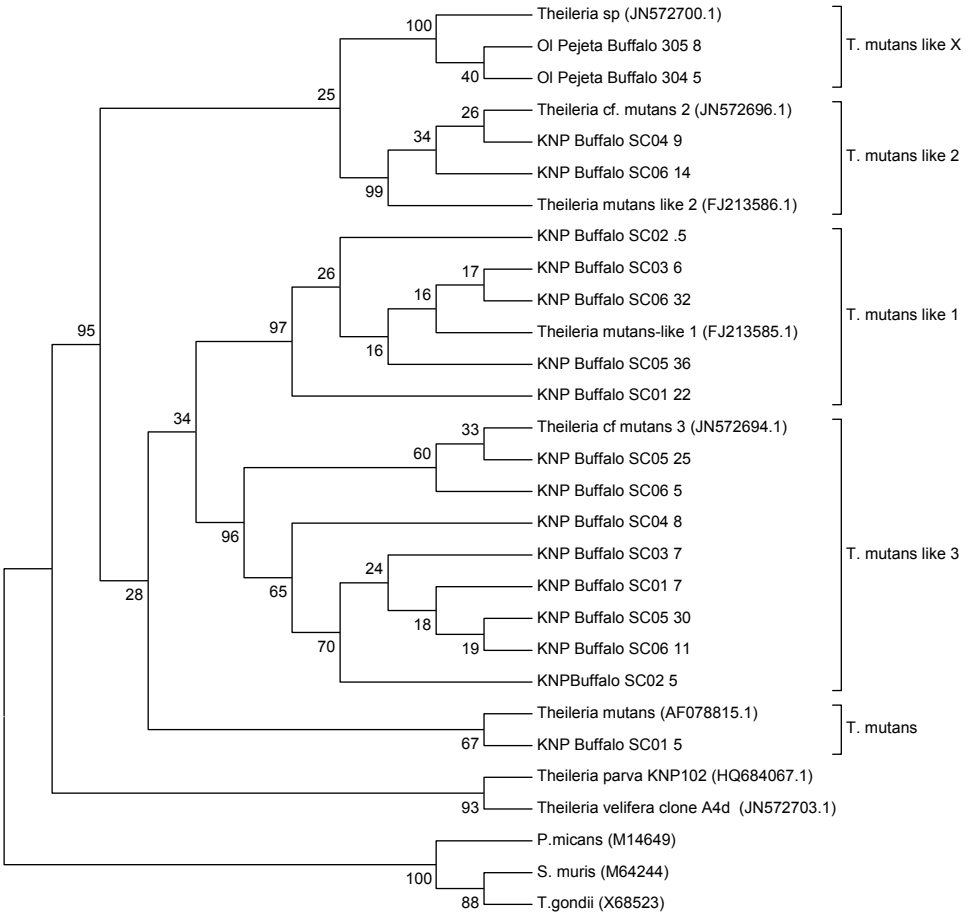

Figure 3 Phylogenetic tree showing the relationship between *T. mutans*-like 18S rRNA sequence variants

The phylogenetic relationship was inferred using the Maximum Parsimony method. The bootstrap consensus tree inferred from 1000 replicates is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test is shown next to the branches. The tree was rooted using rRNA gene sequences for *Prororectum micans* (M14649), *Sarcocystic muris* (M64244), *T. gondii* (X68523) *T. parva* (HQ684067.1), *T.velifera* (JN572703.1) and *T. mutans* (AF078815.1). Samples from African buffalo from the Kruger National Park are labelled KNP Buffalo, followed by the animal number (SC01-SC06) and the number of reads obtained for the allele. Samples from African buffalo from the Ol Pejeta conservancy are labelled Ol Pejeta Buffalo, followed by the animal number (301-308) and the number of reads obtained for the allele.

Appendix F

DNA of cloned *T. parva* infected cell lines used for PCR

Sample Name	Species of isolation	Animal Id	Source	Chapter
Marula N33 Clone 1	Bovine	N33	ILRI	3
Marula N33 Clone 3	Bovine	N33	ILRI	3
Marula N33 Clone 4	Bovine	N33	ILRI	3
Marula N33 Clone 5	Bovine	N33	ILRI	3
Marula N43 Clone 1	Bovine	N43	ILRI	3
Marula N43 Clone 3	Bovine	N43	ILRI	3
Marula N43 Clone 5	Bovine	N43	ILRI	3
Marula N43 Clone 6	Bovine	N43	ILRI	3
Mara 3 Clone 3	Buffalo	Mara 3	ILRI	3
Mara 3 Clone 6	Buffalo	Mara 3	ILRI	3, 6
Mara 3 Clone 7	Buffalo	Mara 3	ILRI	3
Mara 3 Clone 9	Buffalo	Mara 3	ILRI	3
Mara 30 Clone 2	Buffalo	Mara 30	ILRI	3
Mara 30 Clone 5	Buffalo	Mara 30	ILRI	3,6
Mara 30 Clone 8	Buffalo	Mara 30	ILRI	3, 6
Mara 30 Clone 11	Buffalo	Mara 30	ILRI	3,6
Mara 42 Clone 2	Buffalo	Mara 42	ILRI	3,6
Mara 42 Clone 5	Buffalo	Mara 42	ILRI	3,6
Mara 42 Clone 8	Buffalo	Mara 42	ILRI	3
Mara 42 Clone 12	Buffalo	Mara 42	ILRI	3
Mara 6998 Clone 9	Buffalo	Mara 6998	ILRI	3,6
Mara 6998 Clone 11	Buffalo	Mara 6998	ILRI	3
Marikebuni Clone A3	Bovine	592	UoE	3, 6
Marikebuni Clone A7	Bovine	592	UoE	3, 6
Marikebuni Clone B12	Bovine	592	UoE	3, 6
Marikebuni Clone E43	Bovine	592	UoE	3, 6
Marikebuni Clone F31	Bovine	592	UoE	3, 6
Marikebuni Clone F44	Bovine	592	UoE	3, 6
Marikebuni Clone F53	Bovine	592	UoE	3, 6
Marikebuni Clone I8	Bovine	592	UoE	3, 6
Marikebuni Clone I38	Bovine	592	UoE	3, 6
Marikebuni Clone J17	Bovine	592	UoE	3, 6

DNA from blood samples of African buffalo (*Syncerus caffer)* used for PCR

Animal ID	Location	Source	Chapter
SC01	Kruger National Park, South	Pirbright	5,7
SC02	Kruger National Park, South	Pirbright	5,7
SC03	Kruger National Park, South	Pirbright	5,7
SC04	Kruger National Park, South	Pirbright	5,7
SC05	Kruger National Park, South	Pirbright	5,7
SC06	Kruger National Park, South	Pirbright	3,5,7
301	Ol Pejeta Conservancy, Kenya	ILRI	5,7
302	Ol Pejeta Conservancy, Kenya	ILRI	5,7
303	Ol Pejeta Conservancy, Kenya	ILRI	5,7
304	Ol Pejeta Conservancy, Kenya	ILRI	5,7
305	Ol Pejeta Conservancy, Kenya	ILRI	5,7
306	Ol Pejeta Conservancy, Kenya	ILRI	5,7
307	Ol Pejeta Conservancy, Kenya	ILRI	5,7
308	Ol Pejeta Conservancy, Kenya	ILRI	5,7

DNA from lymphnode samples of sentinel cattle after field exposure used for PCR

Animal ID	Location	Source	Chapter
023	Ol Pejeta Conservancy, Kenya	ILRI	7
025	Ol Pejeta Conservancy, Kenya	ILRI	7
027	Ol Pejeta Conservancy, Kenya	ILRI	7
028	Ol Pejeta Conservancy, Kenya	ILRI	7
032	Ol Pejeta Conservancy, Kenya	ILRI	7
033	Ol Pejeta Conservancy, Kenya	ILRI	7
034	Ol Pejeta Conservancy, Kenya	ILRI	7
035	Ol Pejeta Conservancy, Kenya	ILRI	7
036	Ol Pejeta Conservancy, Kenya	ILRI	7

DNA from sporozoite stabilates used for molecular biology

Stabilate name	Stabilate number	Source	Chapter
Muguga	4230	ILRI	4
Muguga	N57	CTTBD	4
Serengeti transformed	4229	ILRI	4
Kiambu 5	4228	ILRI	3,4
ECF MC ILRI	0801	ILRI	4
ECF MC ILRI	0802	ILRI	4
ECF MC ILRI	0803	ILRI	4
ECF MC ILRI	0804	ILRI	4
Marikebuni	3014	RVC (ILRAD)	4
Marikebuni	128	UoE	4
Katete	KL4	CTTBD	4
Chitongo	CL20	CTTBD	4
Marikani	3231 clone 2/3	ILRI	7
Boleni	3230 clone 1/1	ILRI	7
Uganda	3645 clone 1/2	ILRI	7
Buffalo-derived *T. parva*	3081	ILRI	6, 7
Buffalo-derived *T. parva*	4110	ILRI	6, 7

Other DNA's used for PCR

Genus and species	Location or name	Country	Species of isolation	Material	Source	Chapter
T. buffeli	Marula	Kenya	Buffalo	blood	ILRI	3
T. taurotragi	-	Kenya	Bovine	*In vitro* culture	ILRI	3
T. sp (Buffalo)	Buffalo 6834 clone 3	Kenya	Buffalo	*In vitro* culture	ILRI	3
T. annulata	Ankara C9	Turkey	Bovine	*In vitro* culture	UoE	3
R. appendiculatus	Muguga	Kenya	Tick	tick	ILRI	3
Bos indicus	ILRI	Kenya	Bovine	PBMC	ILRI	3
Bos taurus	Edinburgh	UK	Bovine	PBMC	UoE	3

Animals used for *in vivo* experiment

Animal number	Breed	Sex	Age	MHC	Source	Chapter
BG040	Friesian/Ayshire	M	1 year	A10/..	ILRI	6
BG041	Friesian	M	7 months	A10/..	ILRI	6
BG044	Friesian	M	9 months	A10/..	ILRI	6
BF080	Ayshire	M	2.3 years	A10/..	ILRI	6
BF083	Friesian	M	2.5 years	A10/..	ILRI	6
BF085	Ayshire	M	2.5 years	A10/..	ILRI	6
BF086	Friesian	M	2.5 years	A10/..	ILRI	6

In vitro established *T. parva* infected cell lines used for immunological assays

Sample name	Animal ID	MHC I haplotype	Used for	Source	Chapter
011 Muguga	011	A10/A10	Cytotoxicity assay	UoE	6
641 Muguga	641	A18/A18	Cytotoxicity assay	UoE	6
Marikebuni clone B12	592	A10/A10	Cytotoxicity assay	UoE	6
Marikebuni clone F53	592	A10/A10	Cytotoxicity assay	UoE	6
BG040 Muguga	BG040	A10/..	Cytotoxicity assay ELISPOT	ILRI	6
BG041 Muguga	BG041	A10/..	Cytotoxicity assay ELISPOT	ILRI	6
BG044 Muguga	BG044	A10/..	Cytotoxicity assay ELISPOT	ILRI	6
BF080 Muguga	BF080	A10/..	Cytotoxicity assay ELISPOT	ILRI	6
BF083 Muguga	BF083	A10/..	Cytotoxicity assay ELISPOT	ILRI	6
BF085 Muguga	BF085	A10/..	Cytotoxicity assay ELISPOT	ILRI	6
BF086 Muguga	BF086	A10/..	Cytotoxicity assay ELISPOT	ILRI	6

References

Aanensen, D.M., and Spratt, B.G. (2005). The multilocus sequence typing network: mlst.net. Nucleic acids research *33*, W728-733.

Aguado-Martinez, A., MacHugh, N., and Morrison, I. (2012). Identification of a parasite antigen recognised by both CD4 and CD8 T cells specific for lymphoblasts infected with *Theileria parva*. Immunology *137*, 626-627 Abstract P1420.

Allsopp, B.A., Baylis, H.A., Allsopp, M.T., Cavalier-Smith, T., Bishop, R.P., Carrington, D.M., Sohanpal, B., and Spooner, P. (1993). Discrimination between six species of *Theileria* using oligonucleotide probes which detect small subunit ribosomal RNA sequences. Parasitology *107 (Pt 2)*, 157-165.

Allsopp, M.T., Theron, J., Coetzee, M.L., Dunsterville, M.T., and Allsopp, B.A. (1999). The occurrence of Theileria and Cowdria parasites in African buffalo (*Syncerus caffer*) and their associated *Amblyomma hebraeum* ticks. The Onderstepoort journal of veterinary research *66*, 245-249.

Anderson, K., Ezenwa, V.E., and Jolles, A.O. (2013). Tick infestation patterns in free ranging African buffalo (*Syncerus caffer*): Effects of host innate immunity and nice segregration among tick species. International Journal for Parasitology: Parasites and wildlife *2*, 1-9.

Anonymous (1981). The eradication of East Coast fever in South Africa. Journal of the South African Veterinary Association *52*, 71-73.

Anonymous (1988). Nomenclature. In Theileriosis in Eastern, Central and Southern Africa, T.T. Dolan, ed. (Lilongwe, Malawi, The International Laboratory for research on animal diseases), pp. 182-186.

Anonymous (2002). Disease risk management directive for buffalo (*Syncerus caffer*) in South Africa. (2002/09/19).

Asiimwe, B.B., Weir, W., Tait, A., Lubega, G.W., and Oura, C.A. (2013). Haemoparasite infection kinetics and the population structure of *Theileria parva* on a single farm in Uganda. Veterinary parasitology *193*, 8-14.

Baldwin, C.L., Goddeeris, B.M., and Morrison, W.I. (1987). Bovine helper T-cell clones specific for lymphocytes infected with *Theileria parva* (Muguga). Parasite immunology *9*, 499-513.

Baldwin, C.L., Black, S.J., Brown, W.C., Conrad, P.A., Goddeeris, B.M., Kinuthia, S.W., Lalor, P.A., MacHugh, N.D., Morrison, W.I., Morzaria, S.P., and et al. (1988). Bovine T cells, B cells, and null cells are transformed by the protozoan parasite *Theileria parva*. Infection and immunity *56*, 462-467.

Bishop, R.P., Sohanpal, B.K., Allsopp, B.A., Spooner, P.R., Dolan, T.T., and Morzaria, S.P. (1993). Detection of polymorphisms among *Theileria parva* stocks using repetitive, telomeric and ribosomal DNA probes and anti-schizont monoclonal antibodies. Parasitology *107 (Pt 1)*, 19-31.

Bishop, R.P., Spooner, P.R., Kanhai, G.K., Kiarie, J., Latif, A.A., Hove, T., Masaka, S., and Dolan, T.T. (1994). Molecular characterization of Theileria parasites: application to the epidemiology of theileriosis in Zimbabwe. Parasitology *109 (Pt 5)*, 573-581.

Bishop, R., Geysen, D., Spooner, P., Skilton, R., Nene, V., Dolan, T., and Morzaria, S. (2001). Molecular and immunological characterisation of *Theileria parva* stocks which are components of the 'Muguga cocktail' used for vaccination against East Coast fever in cattle. Veterinary parasitology *94*, 227-237.

Bishop, R., Nene, V., Staeyert, J., Rowlands, J., Nyanjui, J., Osaso, J., Morzaria, S., and 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.

Blouin, E.F., and Stoltsz, W.H. (1989). Comparative infection rates of *Theileria parva* lawrencei in salivary glands of *Rhipicephalus appendiculatus* and *Rhipicephalus zambeziensis*. The Onderstepoort journal of veterinary research *56*, 211-213.

Boers, S.A., van der Reijden, W.A., and Jansen, R. (2012). High-throughput multilocus sequence typing: bringing molecular typing to the next level. PloS one *7*, e39630.

Caron, A., Miguel, E., Gomo, C., Makaya, P., Pfukenyi, D.M., Foggin, C., Hove, T., and de Garine-Wichatitsky, M. (2013). Relationship between burden of infection in ungulate populations and wildlife/livestock interfaces. Epidemiology and infection *141*, 1522-1535.

Chae, J.S., Allsopp, B.A., Waghela, S.D., Park, J.H., Kakuda, T., Sugimoto, C., Allsopp, M.T., Wagner, G.G., and Holman, P.J. (1999). A study of the systematics of Theileria spp. based upon small-subunit ribosomal RNA gene sequences. Parasitology research *85*, 877-883.

Chaisi, M.E., Sibeko, K.P., Collins, N.E., Potgieter, F.T., and Oosthuizen, M.C. (2011). Identification of *Theileria parva* and *Theileria sp. (buffalo)* 18S rRNA gene sequence variants in the African Buffalo (*Syncerus caffer*) in southern Africa. Veterinary parasitology *182*, 150-162.

Chaisi, M.E., Collins, N.E., Potgieter, F.T., and Oosthuizen, M.C. (2013). Sequence variation identified in the 18S rRNA gene of *Theileria mutans* and *Theileria velifera* from the African buffalo (*Syncerus caffer*). Veterinary parasitology *191*, 132-137.

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

Cole, G.A., Hogg, T.L., Coppola, M.A., and Woodland, D.L. (1997). Efficient priming of CD8+ memory T cells specific for a subdominant epitope following Sendai virus infection. Journal of immunology *158*, 4301-4309.

Connelley, T.K., MacHugh, N.D., Pelle, R., Weir, W., and Morrison, W.I. (2011). Escape from CD8+ T cell response by natural variants of an immunodominant epitope from *Theileria parva* is predominantly due to loss of TCR recognition. Journal of immunology *187*, 5910-5920.

Conrad, P.A., Iams, K., Brown, W.C., Sohanpal, B., and ole-MoiYoi, O.K. (1987a). DNA probes detect genomic diversity in *Theileria parva* stocks. Molecular and biochemical parasitology *25*, 213-226.

Conrad, P.A., ole-MoiYoi, O.K., Baldwin, C.L., Dolan, T.T., O'Callaghan, C.J., Njamunggeh, R.E., Grootenhuis, J.G., Stagg, D.A., Leitch, B.L., and Young, A.S. (1989). Characterization of buffalo-derived theilerial parasites with monoclonal antibodies and DNA probes. Parasitology *98 Pt 2*, 179-188.

Conrad, P.A., Stagg, D.A., Grootenhuis, J.G., Irvin, A.D., Newson, J., Njamunggeh, R.E., Rossiter, P.B., and Young, A.S. (1987b). Isolation of Theileria parasites from African buffalo (*Syncerus caffer*) and characterization with anti-schizont monoclonal antibodies. Parasitology *94 (Pt 3)*, 413-423.

Cunningham, M.P., Brown, C.G.D., Burridge, M.J., Musoke, A.J., Purnell, R.E., Radley, D.E., and Sempebwa, C. (1974). East Coast Fever - Titration in cattle of suspensions of *Theileria parva* derived from ticks. British Veterinary Journal *130*, 336-345.

Day, E.B., Charlton, K.L., La Gruta, N.L., Doherty, P.C., and Turner, S.J. (2011). Effect of MHC class I diversification on influenza epitope-specific CD8+ T cell precursor frequency and subsequent effector function. Journal of immunology *186*, 6319-6328.

De Deken, R., Martin, V., Saido, A., Madder, M., Brandt, J., and Geysen, D. (2007). An outbreak of East Coast Fever on the Comoros: a consequence of the import of immunised cattle from Tanzania? Veterinary parasitology *143*, 245-253.

Delport, W., Poon, A.F., Frost, S.D., and Kosakovsky Pond, S.L. (2010). Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. Bioinformatics *26*, 2455-2457.

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

Dolan, T.T. (1984). Immunization against Theileriosis on Ol Pejeta ranch, laikipia Kenya. In Immunization against theileriosis in Africa (Nairobi, Kenya).

Dolan, T.T. (1989). Theileriosis in Eastern, Central and Southern Africa, proceedings of a workshop on East Coast fever immunization held in Lilongwe, Malawi 20-22 September 1988 (Nairoby, Kenya: The International Laboratory for Research on Animal Diseases).

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic acids research *32*, 1792-1797.

Edgar, R.C., Haas, B.J., Clemente, J.C., Quince, C., and Knight, R. (2011). UCHIME improves sensitivity and speed of chimera detection. Bioinformatics *27*, 2194-2200.

Fawcett, D., Musoke, A., and Voigt, W. (1984). Interaction of sporozoites of *Theileria parva* with bovine lymphocytes in vitro. I. Early events after invasion. Tissue Cell *16*, 873-884.

Fawcett, D.W., Buscher, G., and Doxsey, S. (1982a). Salivary gland of the tick vector of East Coast fever. III. The ultrastructure of sporogony in *Theileria parva*. Tissue Cell *14*, 183-206.

Fawcett, D.W., Doxsey, S., Stagg, D.A., and Young, A.S. (1982b). The entry of sporozoites of *Theileria parva* into bovine lymphocytes in vitro. Electron microscopic observations. European journal of cell biology *27*, 10-21.

Fyumagwa, R.D., Simmler, P., Meli, M.L., Hoare, R., Hofmann-Lehmann, R., and Lutz, H. (2011). Molecular detection of Anaplasma, Babesia and Theileria species in a diversity of tick species from Ngorongoro Crater, Tanzania. South African journal of wildlife research *41*, 79-86.

Galea, I., Stasakova, J., Dunscombe, M.S., Ottensmeier, C.H., Elliott, T., and Thirdborough, S.M. (2012). CD8+ T-cell cross-competition is governed by peptide-MHC class I stability. European journal of immunology *42*, 256-263.

Gardner, M.J., Bishop, R., Shah, T., de Villiers, E.P., Carlton, J.M., Hall, N., Ren, Q., Paulsen, I.T., Pain, A., Berriman, M., *et al.* (2005). Genome sequence of Theileria parva, a bovine pathogen that transforms lymphocytes. Science *309*, 134-137.

Gauer, M., Mackenstedt, U., Mehlhorn, H., Schein, E., Zapf, F., Njenga, E., Young, A., and Morzaria, S. (1995). DNA measurements and ploidy determination of developmental stages in the life cycles of *Theileria annulata* and *T. parva*. Parasitology research *81*, 565-574.

Gentschev, I., Glaser, I., Goebel, W., McKeever, D.J., Musoke, A., and Heussler, V.T. (1998). Delivery of the p67 sporozoite antigen of *Theileria parva* by using recombinant Salmonella dublin: secretion of the product enhances specific antibody responses in cattle. Infection and immunity *66*, 2060-2064.

Geysen, D. (2000). The application of Molecular Biology techniques to analyse diversity in *Theileria parva* populations in Zambia. In Department of Biology and Biochemistry (London, Brunel University).

Geysen, D. (2008). Live immunisation against *Theileria parva*: spreading the disease? Trends in parasitology *24*, 245-246.

Geysen, D., Bishop, R., Skilton, R., Dolan, T.T., and Morzaria, S. (1999). Molecular epidemiology of *Theileria parva* in the field. Tropical medicine and international health *4*, A21-27.

Gilles, A., Meglecz, E., Pech, N., Ferreira, S., Malausa, T., and Martin, J.F. (2011). Accuracy and quality assessment of 454 GS-FLX Titanium pyrosequencing. BMC genomics *12*, 245.

Giulio, G.D., Lynen, G., Morzaria, S., Oura, C., and Bishop, R. (2009). Live immunization against East Coast fever- current status. Trends in parasitology *25*, 85-92.

Goddeeris, B.M., Katende, J.M., Irvin, A.D., and Chumo, R.S. (1982). Indirect fluorescent antibody test for experimental and epizootiological studies on East coast fever (*Theileria parva* infection in cattle). Evaluation of a cell culture schizont antigen fixed and stored in suspension. Research in veterinary science *33*, 360-365.

Goddeeris, B.M., and Morrison, W.I. (1988). Techniques for the generation, cloning and characterization of bovine cytotoxic T cells specific for the protozoan *Theileria parva*. Journal of Tissue Culture Methods *11*, 101-110.

Graham, S.P., Honda, Y., Pelle, R., Mwangi, D.M., Glew, E.J., de Villiers, E.P., Shah, T., Bishop, R., van der Bruggen, P., Nene, V., and Taracha, E.L. (2007a). 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 research *3*, 2.

Graham, S.P., Pellé, R., Honda, Y., Mwangi, D.M., Tonukari, N.J., Yamage, M., Glew, E.J., Villiers, E.P.d., Shah, T., Bishop, R., *et al.* (2006). *Theileria parva* candidate vaccine antigens recognized by immune bovine cytotoxic T Lymphocytes. Proceedings of the National Academy of Sciences of the United States of America *103*, 3286-3291.

Graham, S.P., Pelle, R., Yamage, M., Mwangi, D.M., Honda, Y., Mwakubambanya, R.S., de Villiers, E.P., Abuya, E., Awino, E., Gachanja, J., *et al.* (2008). Characterization of the fine specificity of bovine CD8 T-cell responses to defined antigens from the protozoan parasite *Theileria parva*. Infection and immunity *76*, 685-694.

Graham, S.P., Saya, R., Awino, E., Ngugi, D., Nyanjui, J.K., Hecker, R., Taracha, E.L., and Nene, V. (2007b). Immunostimulatory CpG oligodeoxynucleotides enhance the induction of bovine CD4+ cytotoxic T-lymphocyte responses against the polymorphic immunodominant molecule of the protozoan parasite *Theileria parva*. Veterinary immunology and immunopathology *115*, 383-389.

Grootenhuis, J.G., Leitch, B.L., Stagg, D.A., Dolan, T.T., and Young, A.S. (1987). Experimental induction of *Theileria parva lawrencei* carrier state in an African buffalo (*Syncerus caffer*). Parasitology *94 (Pt 3)*, 425-431.

Gubbels, J.M., de Vos, A.P., van der Weide, M., Viseras, J., Schouls, L.M., de Vries, E., and Jongejan, F. (1999). Simultaneous detection of bovine Theileria and Babesia species by reverse line blot hybridization. Journal of clinical microbiology *37*, 1782-1789.

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

Haas, B.J., Gevers, D., Earl, A.M., Feldgarden, M., Ward, D.V., Giannoukos, G., Ciulla, D., Tabbaa, D., Highlander, S.K., Sodergren, E., *et al.* (2011). Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. Genome research *21*, 494-504.

Hayashida, K., Abe, T., Weir, W., Nakao, R., Ito, K., Kajino, K., Suzuki, Y., Jongejan, F., Geysen, D., and Sugimoto, C. (2013). Whole-genome sequencing of *Theileria parva* strains provides insight into parasite migration and diversification in the African continent. DNA research *20*, 209-220.

Hayashida, K., Hara, Y., Abe, T., Yamasaki, C., Toyoda, A., Kosuge, T., Suzuki, Y., Sato, Y., Kawashima, S., Katayama, T., *et al.* (2012). Comparative Genome Analysis of Three Eukaryotic Parasites with Differing Abilities To Transform Leukocytes Reveals Key Mediators of Theileria-Induced Leukocyte Transformation. Mbio *3*, (5) e00204-12.

Henson, S., Bishop, R.P., Morzaria, S., Spooner, P.R., Pelle, R., Poveda, L., Ebeling, M., Kung, E., Certa, U., Daubenberger, C.A., and Qi, W. (2012). High-resolution genotyping and mapping of recombination and gene conversion in the protozoan *Theileria parva* using whole genome sequencing. BMC genomics *13*, 503.

Horak, I.G., Golezardy, H., and Uys, A.C. (2007). Ticks associated with the three largest wild ruminant species in southern Africa. The Onderstepoort journal of veterinary research *74*, 231-242.

Hove, T., Musisi, F.L., Kanhai, G.K., Latif, A., Masaka, S., Munatswa, F.C., Pegram, R.G., Kamwendo, S.P., Quiroga, J.C., Mwangondwe, R., and et al. (1995). Challenge of *Theileria parva* (Boleni) immunised cattle with selected east African Theileria stocks. Tropical animal health and production *27*, 202-210.

Huse, S.M., Welch, D.M., Morrison, H.G., and Sogin, M.L. (2010). Ironing out the wrinkles in the rare biosphere through improved OTU clustering. Environmental microbiology *12*, 1889-1898.

Illumina (2012). An introduction to Next-generation sequencing technology. (www.ilumina.com/NGS).

ILRAD (1989). Theileriosis in Eastern, Central and Southern Africa: Proceedings of a Workshop on East Coast Fever Immunization held in Lilongwe, Malawi20-22 September 1988. T.T. Dolan, ed. (Nairobi, Kenya).

Irvin, A.D., Brown, C.G., Burridge, M.J., Cunningham, M.P., Musoke, A.J., Pierce, M.A., Purnell, R.E., and Radley, D.E. (1972). A pathogenic theilerial syndrome of cattle in the Narok District of Kenya. I: Transmission studies. Tropical animal health and production *4*, 220-229.

Jongejan, F., and Uilenberg, G. (2004). The global importance of ticks. Parasitology *129*, S3-14

Jongejan, F., Nene, V., de la Fuente, J., Pain, A., and Willadsen, P. (2007). Advances in the genomics of ticks and tick-borne pathogens. Trends in parasitology *23*, 391-396.

Kaba, S.A., Schaap, D., Roode, E.C., Nene, V., Musoke, A.J., Vlak, J.M., and van Oers, M.M. (2004). Improved immunogenicity of novel baculovirus-derived *Theileria parva* p67 subunit antigens. Veterinary parasitology *121*, 53-64.

Kariuki, D., Young, A., Morzaria, S., Lesan, A., Mining, S., Omwoyo, P., Wafula, J., and Molyneux, D. (1995). *Theileria parva* carrier state in naturally infected and artificially immunised cattle. Tropical animal health and production *27*, 15-25.

Kariuki, E.K., Penzhorn, B.L., and Horak, I.G. (2012). Ticks (*Acari: Ixodidae*) infesting cattle and African buffaloes in the Tsavo conservation area, Kenya. The Onderstepoort journal of veterinary research *79*.

Kastenmuller, W., Gasteiger, G., Gronau, J.H., Baier, R., Ljapoci, R., Busch, D.H., and Drexler, I. (2007). Cross-competition of CD8+ T cells shapes the immunodominance hierarchy during boost vaccination. Journal of experimental medicine *204*, 2187-2198.

Katende, J., Morzaria, S., Toye, P., Skilton, R., Nene, V., Nkonge, C., and Musoke, A. (1998). An enzyme-linked immunosorbent assay for detection of *Theileria parva* antibodies in cattle using a recombinant polymorphic immunodominant molecule. Parasitology research *84*, 408-416.

Katzer, F., Ngugi, D., Oura, C., Bishop, R.P., Taracha, E.L., Walker, A.R., and McKeever, D.J. (2006). Extensive genotypic diversity in a recombining population of the apicomplexan parasite *Theileria parva*. Infection and immunity *74*, 5456-5464.

Katzer, F., Ngugi, D., Schnier, C., Walker, A.R., and McKeever, D.J. (2007). Influence of host immunity on parasite diversity in *Theileria parva*. Infection and immunity *75*, 4909-4916.

Katzer, F., Ngugi, D., Walker, A.R., and McKeever, D.J. (2010). Genotypic diversity, a survival strategy for the apicomplexan parasite *Theileria parva*. Veterinary parasitology *167*, 236-243.

Katzer, F., Lizundia, R., Ngugi, D., Blake, D., and McKeever, D. (2011). Construction of a genetic map for *Theileria parva*: Identification of hotspots of recombination. International journal for parasitology *41*, 669-675.

Kimbita, E.N., and Silayo, R.S. (1997). Use of an in vitro infectivity assay in comparison with histological techniques in the study of *Theileria parva* sporozoite maturation. Veterinary parasitology *70*, 83-97.

Kircher, M., and Kelso, J. (2010). High-throughput DNA sequencing- concepts and limitations. BioEssays : news and reviews in molecular, cellular and developmental biology *32*, 524-536.

Latif, A.A., Hove, T., Kanhai, G.K., and Masaka, S. (2001). Exposure of cattle immunized with different stocks of *Theileria parva* to buffalo-associated Theileria challenge on two game parks in Zimbabwe. The Onderstepoort journal of veterinary research *68*, 197-201.

Latif, and Jongejan, F. (2002). The wide-use of acaricide control of livestock diseases in Africa needs a re-appraisal. In Newsletter intergrated control pathogen trypanosomasis vectors.

Latif, A.A., and Hove, T. (2011). History and critical review of *Theileria parva* (Boleni), the vaccine stock against Zimbabwean cattle theileriosis. Ticks and tick-borne diseases *2*, 163-167.

Lawrence, J.A. (1977). The serological relationship between *Theileria parva* (Muguga) and Theileria lawrence from Rhodesia. The Veterinary record *100*, 470-471.

Lawrence, J.A. (1979). The differential diagnosis of the bovine theilerias of Southern Africa. Journal of the South African Veterinary Association *50*, 311-313.

Lawrence, J.A., Norval, R.A., and Uilenberg, G. (1983). *Rhipicephalus zambeziensis* as a vector of bovine Theileriae. Tropical animal health and production *15*, 39-42.

Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., Lin, D., Lu, L., and Law, M. (2012). Comparison of next-generation sequencing systems. Journal of biomedicine & biotechnology *2012*, 251364.

Lohre (1978). In Tick-borne diseases and their vectors, J.H. Wilde, ed. (Univeristy of Edinburgh).

Luciani, F., Sanders, M.T., Oveissi, S., Pang, K.C., and Chen, W. (2013). Increasing viral dose causes a reversal in CD8+ T cell immunodominance during primary influenza infection due to differences in antigen presentation, T cell avidity, and precursor numbers. Journal of immunology *190*, 36-47.

Lynen, G., Makala, L.H.C., and Pas, W.M., (1991) East Coast fever immunization in Eastern Province, Zambia. In Ticks and Tick-Borne Disease Control (ed. T. T. Dolan) Proceedings of a Joint OAU, FAO and ILRAD Workshop held in Kampala, Uganda, 12–14 September 1991.(ILRAD, Nairobi, Kenya) , pp. 16–21.

Lynen, G., Zeman, P., Bakuname, C., Di Giulio, G., Mtui, P., Sanka, P., and Jongejan, F. (2007). Cattle ticks of the genera *Rhipicephalus* and *Amblyomma* of economic importance in Tanzania: distribution assessed with GIS based on an extensive field survey. Experimental & applied acarology *43*, 303-319.

Macdonald, I.K., Harkiolaki, M., Hunt, L., Connelley, T., Carroll, A.V., MacHugh, N.D., Graham, S.P., Jones, E.Y., Morrison, W.I., Flower, D.R., and Ellis, S.A. (2010). MHC class I bound to an immunodominant *Theileria parva* epitope demonstrates unconventional presentation to T cell receptors. PLoS pathogens *6*, e1001149.

MacHugh, N.D., Connelley, T., Graham, S.P., Pelle, R., Formisano, P., Taracha, E.L., Ellis, S.A., McKeever, D.J., Burrells, A., and Morisson, W.I. (2009). $CD8^+$ T-cell responses to *Theileria parva* are preferentially directed to a single dominant antigen: Implications for parasite strain-specific immunity. European journal of immunology *39*, 2459-2469.

Mans, B.J., Pienaar, R., Latif, A.A., and Potgieter, F.T. (2011a). Diversity in the 18S SSU rRNA V4 hyper-variable region of *Theileria spp.* in Cape buffalo (*Syncerus caffer*) and cattle from southern Africa. Parasitology *38*, 766-779.

Mans, B.J., Pienaar, R., Potgieter, F.T., and Latif, A.A. (2011b). *Theileria parva*, *T. sp. (buffalo)* and *T. sp. (bougasvlei)* 18S variants. Veterinary parasitology *182*, 382-383.

Marcotty, T., Billiouw, M., Chaka, G., Berkvens, D., Losson, B., and Brandt, J. (2001). Immunisation against East Coast fever by the infection and treatment method: evaluation of the use of ice baths for field delivery and appraisal of an acid formulation of long-acting tetracycline. Veterinary parasitology *99*, 175-187.

Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y.J., Chen, Z., *et al.* (2005). Genome sequencing in microfabricated high-density picolitre reactors. Nature *437*, 376-380.

Martins, S.B., Di Giulio, G., Lynen, G., Peters, A., and Rushton, J. (2010). Assessing the impact of East Coast Fever immunisation by the infection and treatment method in Tanzanian pastoralist systems. Preventive veterinary medicine *97*, 175-182.

Mbao, V., Berkvens, D., Dorny, P., Van den Bossche, P., and Marcotty, T. (2007). Comparison of the survival on ice of thawed *Theileria parva* sporozoites of different stocks cryoprotected by glycerol or sucrose. The Onderstepoort journal of veterinary research *74*, 9-15.

Mbizeni, S., Potgieter, F.T., Troskie, C., Mans, B.J., Penzhorn, B.L., and Latif, A.A. (2013). Field and laboratory studies on Corridor disease (*Theileria parva* infection) in cattle population at the livestock/game interface of uPhongolo-Mkuze area, South Africa. Ticks and tick-borne diseases *4*, 227-234.

McKeever, D.J., Taracha, E.L.N., Innes, E.L., MacHugh, N.D., Awinl, E., Goddeeris, B.M., and Morrison, W.I. (1994). Adoptive transfer of immunity to *Theileria parva* int he CD8+ fraction of responding efferent lymph. Proceedings of the National Academy of Sciences of the United States of America *91*, 1959-1963.

McKeever, D.J., Taracha, E.L., Morrison, W.I., Musoke, A.J., and Morzaria, S.P. (1999). Protective immune mechanisms against *Theileria parva*: evolution of vaccine development strategies. Parasitology today *15*, 263-267.

Minjauw, B., and McLeod, A. (2003). Tick-borne diseases and poverty. The impact of ticks and tick-borne diseases on the livelihood of small-scale and marginal livestock owners in India and eastern and southern Africa. In Research report, DFID animal health programme, (Centre for Tropical veterinary medicine, university of Edinburgh, UK).

Morrison, W.I., Buscher, G., Murray, M., Emery, D.L., Masake, R.A., Cook, R.H., and Wells, P.W. (1981). *Theileria parva*: kinetics of infection in the lymphoid system of cattle. Experimental parasitology *52*, 248-260.

Morrison, W.I., MacHugh, N.D., and Lalor, P.A. (1996). Pathogenicity of *Theileria parva* is influenced by the host cell type infected by the parasite. Infection and immunity *64*, 557-562.

Morrison, W.I., and McKeever, D.J. (2006). Current status of vaccine development against Theileria parasites. Parasitology *133*, S169-187.

Morzaria, S.P., Dolan, T.T., Norval, R.A., Bishop, R.P., and Spooner, P.R. (1995). Generation and characterization of cloned *Theileria parva* parasites. Parasitology *111 (Pt 1)*, 39-49.

Morzaria, S., and Williamson, S. (1999). Live Vaccines for *Theileria parva*: Deployment in Eastern, Central and Southern Africa. Proceedings of an FAO, OAU–IBAR and ILRI Workshop held 12 March 1997 at ILRI, Nairobi, Kenya In –, S. Morzaria, and S. Williamson, eds. (ILRI (International Livestock Research Institute) Nairobi, Kenya).

Mugisha, A., McLeod, A., Percy, R., and Kyewalabye, E. (2005). Strategies, effectiveness and rationale of vector-borne disease control in the pastoralist system of south-western Uganda. Tropical animal health and production *37*, 479-489.

Muleya, W., Namangala, B., Simuunza, M., Nakao, R., Inoue, N., Kimura, T., Ito, K., Sugimoto, C., and Sawa, H. (2012). Population genetic analysis and sub-structuring of *Theileria parva* in the northern and eastern parts of Zambia. Parasites & vectors *5*, 255.

Muraguri, G.R., Kiara, H.K., and McHardy, N. (1999). Treatment of East Coast fever: a comparison of parvaquone and buparvaquone. Veterinary parasitology *87*, 25-37.

Musisi, F.L., Quiroga, J.C., Kanhai, G.K., Kamwendo, S.P., Mzoma, F.J., and Njuguna, L.M. (1996). *Theileria parva* (Kasoba): isolation and challenge of cattle recovered from infection with other *Theileria parva* stocks. Revue d'elevage et de medecine veterinaire des pays tropicaux *49*, 42-45.

Musisi, F.L., and Dolan, T.T. (1999). Ticks and Tick-borne Disease Control in Eastern, Central and Southern Africa, 1991-1994. Proceedings of a Joint OAU, FAO and ILRAD Workshop held in Lilongwe, Malawi, 25-28 April 1994. (ILRI ,Nairobi, Kenya), p. 150.

Musoke, A., Morzaria, S., Nkonge, C., Jones, E., and Nene, V. (1992). A recombinant sporozoite surface antigen of *Theileria parva* induces protection in cattle. Proceedings of the National Academy of Sciences of the United States of America *89*, 514-518.

Musoke, A., Rowlands, J., Nene, V., Nyanjui, J., Katende, J., Spooner, P., Mwaura, S., Odongo, D., Nkonge, C., Mbogo, S.*, et al.* (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.

Mutugi, J.J., Young, A.S., Maritim, A.C., Linyonyi, A., Mbogo, S.K., and Leitch, B.L. (1988a). Immunization of cattle using varying infective doses of *Theileria parva* lawrencei sporozoites derived from an African buffalo (*Syncerus caffer*) and treatment with buparvaquone. Parasitology *96 (Pt 2)*, 391-402.

Mutugi, J.J., Young, A.S., Maritim, A.C., Ndungu, S.G., Stagg, D.A., Grootenhuis, J.G., and Leitch, B.L. (1988b). Immunization of cattle against theileriosis using varying doses of *Theileria parva* lawrencei and T. parva parva sporozoites and oxytetracycline treatments. International journal for parasitology *18*, 453-461.

Mutugi, J.J., Ndungu, S.G., Lllnyonyi , A., Maritim, A.C., Mining, S.K., Ngumi, P.N. , Kariuki , D.P., Leitch, B.L. ,D'souza, D., Maloo, S. and Lohr, K.F. (1991). Responses to a vaccine trial for East Coast fever in two cattle herds at the Kenyan coast. Preventive Veterinary Medicine *21*, 215-235.

Nambota, A., Samui, K., Sugimoto, C., Kakuta, T., and Onuma, M. (1994). Theileriosis in Zambia: etiology, epidemiology and control measures. Japanese journal of veterinary research *42*, 1-18.

Ndungu, S.G., Brown, C.G., and Dolan, T.T. (2005). In vivo comparison of susceptibility between Bos indicus and Bos taurus cattle types to *Theileria parva* infection. The Onderstepoort journal of veterinary research *72*, 13-22.

Ngumi, P.N., Young, A.S., Lampard, D., Mining, S.K., Ndungu, S.G., Lesan, A.C., Williamson, S.M., Linyonyi, A., and Kariuki, D.P. (1992). Further evaluation of the use of buparvaquone in the infection and treatment method of immunizing cattle against *Theileria parva* derived from African buffalo (*Syncerus caffer*). Veterinary parasitology *43*, 15-24.

Ngumi, P.N., Lesan, A.C., Williamson, S.M., Awich, J.R., Morzaria, S.P., Dolan, T.T., Shaw, M.K., and Young, A.S. (1994a). Isolation and preliminary characterisation of a previously unidentified Theileria parasite of cattle in Kenya. Research in veterinary science *57*, 1-9.

Ngumi, P.N., Lesan, A.C., Williamson, S.M., Awich, J.R., Morzaria, S.P., Dolan, T.T., Shaw, M.K., and Young, A.S. (1994b). Isolation and Preliminary Characterization of a Previously Unidentified Theileria Parasite of Cattle in Kenya. Research in veterinary science *57*, 1-9.

Niklas, N., Proll, J., Danzer, M., Stabentheiner, S., Hofer, K., and Gabriel, C. (2013). Routine performance and errors of 454 HLA exon sequencing in diagnostics. BMC bioinformatics *14*, 176.

Norval, R.A., Lawrence, J.A., Young, A.S., Perry, B.D., Dolan, T.T., and Scott, J. (1991a). *Theileria parva*: influence of vector, parasite and host relationships on the epidemiology of theileriosis in southern Africa. Parasitology *102 Pt 3*, 347-356.

Norval, R.A.I., Perry, B.D., and Young, A.S. (1992). The epidemiology of theileriosis in Africa (London: academic press limited).

Ochanda, H., Young, A.S., Mutugi, J.J., Mumo, J., and Omwoyo, P.L. (1988). The effect of temperature on the rate of transmission of *Theileria parva* parva infection to cattle by its tick vector, *Rhipicephalus appendiculatus*. Parasitology *97 (Pt 2)*, 239-245.

Odongo, D.O., Oura, C.A., Spooner, P.R., Kiara, H., Mburu, D., Hanotte, O.H., and Bishop, R.P. (2006). Linkage disequilibrium between alleles at highly polymorphic mini- and micro-satellite loci of *Theileria parva* isolated from cattle in three regions of Kenya. International journal for parasitology *36*, 937-946.

Odongo, D.O., Sunter, J.D., Kiara, H.K., Skilton, R.A., and Bishop, R.P. (2010). A nested PCR assay exhibits enhanced sensitivity for detection of *Theileria parva* infections in bovine blood samples from carrier animals. Parasitology research *106*, 357-365.

OIE (2008). OIE Terrestrial Manual 2008. pp. 789-804.

Oura, C.A.L., Odonga, D.O., Lubega, G.W., Spooner, P.R., Tait, A., and Bishop, R.P. (2003). A panel of microsatellite and minisatellite markers for the characterisation of field isolates of *Theileria parva*. International journal for parasitology *33*, 1641-1653.

Oura, C.A., Bishop, R.P., Wampande, E.M., Lubega, G.W., and Tait, A. (2004a). Application of a reverse line blot assay to the study of haemoparasites in cattle in Uganda. International journal for parasitology *34*, 603-613.

Oura, C.A., Bishop, R., Wampanda, E.M., Lubega, G.W., and Tait, A. (2004b). The persistence of component *Theileria parva* stocks in cattle immunized with the 'Muguga cocktail' live vaccine against East Coast fever in Uganda. Parasitology *129*, 27-42.

Oura, C.A., Asiimwe, B.B., Weir, W., Lubega, G.W., and Tait, A. (2005). Population genetic analysis and sub-structuring of *Theileria parva* in Uganda. Molecular and biochemical parasitology *140*, 229-239.

Oura, C.A., Bishop, R., Asiimwe, B.B., Spooner, P., Lubega, G.W., and Tait, A. (2007). *Theileria parva* live vaccination: parasite transmission, persistence and heterologous challenge in the field. Parasitology *134*, 1205-1213.

Oura, C.A., Tait, A., Asiimwe, B., Lubega, G.W., and Weir, W. (2011a). Haemoparasite prevalence and *Theileria parva* strain diversity in Cape buffalo (*Syncerus caffer*) in Uganda. Veterinary parasitology *175*, 212-219.

Oura, C.A., Tait, A., Asiimwe, B., Lubega, G.W., and Weir, W. (2011b). *Theileria parva* genetic diversity and haemoparasite prevalence in cattle and wildlife in and around Lake Mburo National Park in Uganda. Parasitology research *108*, 1365-1374.

Paling, R.W., Grootenhuis, J.G., and Young, A.S. (1981). Isolation of *Theileria mutans* from Kenyan buffalo, and transmission by Amblyomma gemma. Veterinary parasitology *8*, 31-37.

Pascucci, I., Capobianco Dondona, A., Camma, C., Marcacci, M., Di Domenico, M., Lelli, R., Scacchia, M., Jago, M., Khaiseb, S., Hager, A.L., *et al.* (2011). Survey of ixodid ticks and two tick-borne pathogens in African buffaloes, *Syncerus caffer*, from the Caprivi Strip, Namibia. Journal of zoo and wildlife medicine *42*, 634-640.

Patel, E.H., Lubembe, D.M., Gachanja, J., Mwaura, S., Spooner, P., and Toye, P. (2011). Molecular characterization of live *Theileria parva* sporozoite vaccine stabilates reveals extensive genotypic diversity. Veterinary parasitology *179*, 62-68.

Peakall, R., and Smouse, P.E. (2012). GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research--an update. Bioinformatics *28*, 2537-2539.

Pelle, R., Graham, S.P., Njahira, M.N., Osaso, J., Saya, R.M., Odongo, D.O., Toye, P.G., Spooner, P.R., Musoke, A.J., Mwangi, D.M., *et al.* (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*, e19015.

Perry, B.D., and Young, A.S. (1993). The Naming Game - the Changing Fortunes of East-Coast Fever and Theileria-Parva. Veterinary Record *133*, 613-616.

Pienaar, R., Potgieter, F.T., Latif, A.A., Thekisoe, O.M.M., and Mans, B.J. (2011). Mixed Theileria infections in free-ranging buffalo herds: implications for diagnosing *Theileria parva* infections in Cape buffalo (*Syncerus caffer*). Parasitology *138*, 884-895.

Potgieter, F.T., Stoltsz, W.H., Blouin, E.F., and Roos, J.A. (1988). Corridor disease in South Africa: a review of the current status. Journal of the South African Veterinary Association *59*, 155-160.

Quince, C., Lanzen, A., Curtis, T.P., Davenport, R.J., Hall, N., Head, I.M., Read, L.F., and Sloan, W.T. (2009). Accurate determination of microbial diversity from 454 pyrosequencing data. Nature methods *6*, 639-641.

Quince, C., Lanzen, A., Davenport, R.J., and Turnbaugh, P.J. (2011). Removing noise from pyrosequenced amplicons. BMC bioinformatics *12*, 38.

Radley, D.E., Brown, C.G.D., Burridge, M.J., Cunningham, M.P., Kirimi, I.M., Purnell, R.E., and Young, A.S. (1975a). East coast fever: 1. Chemoprophylactic immunization of cattle against *Theileria parva* (Muguga) and five theilerial strains. Veterinary parasitology *1*, 35-41.

Radley, D.E., Brown, C.G.D., Cunningham, M.P., Kimber, C.D., Musisi, F.L., Payne, R.C., Purnell, R.E., Stagg, S.M., and Young, A.S. (1975b). East coast fever: 3. Chemoprophylactic immunization of cattle using oxytetracycline and a combination of theilerial strains. Veterinary parasitology *1*, 51-60.

Radley, D.E., Young, A.S., Brown, C.G.D., Burridge, M.J., Cunningham, M.P., Musisi, F.L., and Purnell, R.E. (1975c). East coast fever: 2. Cross-immunity trials with a Kenya strain of *Theileria lawrencei*. Veterinary parasitology *1*, 43-50.

Radley, D.E. (1976). Chemoprophylactic immunisation against east coast fever. In Tick-borne disease and their vectors, J.K.H. Wilde, ed. (University of Edinburgh, centre for tropical veterinary medicine), pp. 324-329.

Radley, D.E., Young, A.S., Grootenhuis, J.G., Cunningham, M.P., Dolan, T.T., and Morzaria, S.P. (1979). Further studies on the immunization of cattle against *Theileria lawrencei* by infection and chemoprophylaxis. Veterinary parasitology *5*, 117-128.

Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., Lesniewski, R.A., Oakley, B.B., Parks, D.H., Robinson, C.J., *et al.* (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Applied and environmental microbiology *75*, 7537-7541.

Schloss, P.D., Gevers, D., and Westcott, S.L. (2011). Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. PloS one *6*, e27310.

Schloss, P.D., and Westcott, S.L. (2011). Assessing and improving methods used in operational taxonomic unit-based approaches for 16S rRNA gene sequence analysis. Applied and environmental microbiology *77*, 3219-3226.

Shaw, M.K., Tilney, L.G., and Musoke, A.J. (1991). The entry of *Theileria parva* sporozoites into bovine lymphocytes: evidence for MHC class I involvement. The Journal of cell biology *113*, 87-101.

Shaw, M.K., and Tilney, L.G. (1992). How individual cells develop from a syncytium: merogony in *Theileria parva* (Apicomplexa). Journal of cell science *101 (Pt 1)*, 109-123.

Shaw, M.K., and Tilney, L.G. (1995). The entry of *Theileria parva* merozoites into bovine erythrocytes occurs by a process similar to sporozoite invasion of lymphocytes. Parasitology *111 (Pt 4)*, 455-461.

Shaw, M.K. (2003). Cell invasion by Theileria sporozoites. Trends in parasitology *19*, 2-6.

Sibeko, K.P., Oosthuizen, M.C., Collins, N.E., Geysen, D., Rambritch, N.E., Latif, A.A., Groeneveld, H.T., Potgieter, F.T., and Coetzer, J.A. (2008). Development and evaluation of a real-time polymerase chain reaction test for the detection of *Theileria parva* infections in Cape buffalo (*Syncerus caffer*) and cattle. Veterinary parasitology *155*, 37-48.

Sibeko, K.P., Geysen, D., Oosthuizen, M.C., Matthee, C.A., Troskie, M., Potgieter, F.T., Coetzer, J.A., and Collins, N.E. (2010). Four p67 alleles identified in South African *Theileria parva* field samples. Veterinary parasitology *167*, 244-254

Sibeko, K.P., Collins, N.E., Oosthuizen, M.C., Troskie, M., Potgieter, F.T., Coetzer, J.A., and Geysen, D. (2011). Analyses of genes encoding *Theileria parva* p104 and polymorphic immunodominant molecule (PIM) reveal evidence of the presence of cattle-type alleles in the South African T. parva population. Veterinary parasitology *181*, 120-130.

Siefert, H.S.H. (1996). Tropical animal health (Kuwer academic publisher).

Skilton, R.A., Bishop, R.P., Katende, J.M., Mwaura, S., and Morzaria, S.P. (2002). The persistence of *Theileria parva* infection in cattle immunized using two stocks

which differ in their ability to induce a carrier state: analysis using a novel blood spot PCR assay. Parasitology *124*, 265-276.

Skums, P., Dimitrova, Z., Campo, D.S., Vaughan, G., Rossi, L., Forbi, J.C., Yokosawa, J., Zelikovsky, A., and Khudyakov, Y. (2012). Efficient error correction for next-generation sequencing of viral amplicons. BMC bioinformatics *13(Suppementl 10)*, S6.

Snodgrass, D.R., Trees, A.J., Bowyer, W.A., Bergman, J.R., Daft, J., and Wall, A.E. (1972). East Coast Fever: field challenge of cattle immunised against *Theileria parva* (Muguga). Tropical animal health and production *4*, 142-151.

Speybroeck, N., Marcotty, T., Aerts, M., Dolan, T., Williams, B., Lauer, J., Molenberghs, G., Burzykowski, T., Mulumba, M., and Berkvens, K. (2008). Titrating *Theileria parva*: Single stocks against combinations of stocks. Experimental parasitology *118*, 522-530.

Spooner, R.L., Innes, E.A., Glass, E.J., Millar, P., and Brown, C.G. (1988). Bovine mononuclear cell lines transformed by *Theileria parva* or Theileria annulata express different subpopulation markers. Parasite immunology *10*, 619-629.

Steinaa, L., Saya, R., Awino, E., and 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. Veterinary immunology and immunopathology *145*, 571-581.

Sutherland, I.A., Shiels, B.R., Jackson, L., Brown, D.J., Brown, C.G., and Preston, P.M. (1996). Theileria annulata: altered gene expression and clonal selection during continuous in vitro culture. Experimental parasitology *83*, 125-133.

Tait, A., and Oura, C. (2004). Reverse line blotting: a new technique for the sensitive detection of tick borne pathogens. Archives de l'Institut Pasteur de Tunis *81*, 47-50.

Tamura, K., Nei, M., and Kumar, S. (2004). Prospects for inferring very large phylogenies by using the neighbor-joining method. Proceedings of the National Academy of Sciences of the United States of America *101*, 11030-11035.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011). MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood,

Evolutionary Distance, and Maximum Parsimony Methods. Molecular biology and evolution *28*, 2731-2739.

Taracha, E.L., Goddeeris, B.M., Morzaria, S.P., and Morrison, W.I. (1995a). Parasite strain specificity of precursor cytotoxic T cells in individual animals correlates with cross-protection in cattle challenged with *Theileria parva*. Infection and immunology *63*, 1258-1262.

Taracha, E.L., Goddeeris, B.M., Teale, A.J., Kemp, S.J., and Morrison, W.I. (1995b). Parasite strain specificity of bovine cytotoxic T cell responses to *Theileria parva* is determined primarily by immunodominance. Journal of immunology *155*, 4854-4860.

Taracha, E.L., Awino, E., and McKeever, D.J. (1997). Distinct CD4+ T cell helper requirements in *Theileria parva* immune and naive bovine CTL precursors. Journal of immunology *159*, 4539-4545

Taylor, L.H., Matthews, L., Shaw, D.J., and Haydon, D.T. (2005). Competitive suppression in mixed-clone parasite cultures. Biology letters *1*, 108-111.

Thompson, B.E., Latif, A.A., Oosthuizen, M.C., Troskie, M., and Penzhorn, B.L. (2008). Occurrence of *Theileria parva* infection in cattle on a farm in the Ladysmith district, KwaZulu-Natal, South Africa. Journal of the South African Veterinary Association *79*, 31-35.

Tindih, H.S., Geysen, D., Goddeeris, B.M., Awino, E., Dobbelaere, D.A., and Naessens, J. (2012). A *Theileria parva* isolate of low virulence infects a subpopulation of lymphocytes. Infection and immunity *80*, 1267-1273.

Toye, P., Gobright, E., Nyanjui, J., Nene, V., and Bishop, R. (1995). Structure and sequence variation of the genes encoding the polymorphic, immunodominant molecule (PIM), an antigen of *Theileria parva* recognized by inhibitory monoclonal antibodies. Molecular and biochemical parasitology *73*, 165-177.

Uilenberg, G., Schreuder, B.E., Silayo, R.S., and Mpangala, C. (1976). Studies on Theileriidae (Sporozoa) in Tanzania. IV. A field trial on immunization against East Coast fever (*Theileria parva* infection of cattle). Tropenmedizin und Parasitologie *27*, 329-336.

Uilenberg, G., Silayo, R.S., Mpangala, C., Tondeur, W., Tatchell, R.J., and Sanga, H.J. (1977). Studies on Theileriidae (Sporozoa) in Tanzania. X. A large-scale field trial on immunization against cattle Theileriosis. Tropenmedizin und Parasitologie *28*, 499-506.

Uilenberg, G. (1978). Some Recent Research on Theileriidae (Sporozoa) of Cattle. Tropical and geographical medicine *30*, 156-156.

Walker, A.R., and McKellar, S.B. (1983). Observations on the separation of theileria sporozoites from ticks. International journal for parasitology *13*, 313-318.

Walker, A.R., Bouattour, A., Camicas, J.L., Estrada-Peña, A., Horak, I.G., Latif, A.A., Pegram, R.G., and Preston, P.M. (2001). Ticks of Domestic Animals in Africa: a Guide to Identification of Species (Edinburgh, U.K.,).

Wambura, P.N., Gwakisa, P.S., Silayo, R.S., and Rugaimukamu, E.A. (1998). Breed-associated resistance to tick infestation in Bos indicus and their crosses with Bos taurus. Veterinary parasitology *77*, 63-70.

Waterhouse, A.M., Procter, J.B., Martin, D.M., Clamp, M., and Barton, G.J. (2009). Jalview Version 2--a multiple sequence alignment editor and analysis workbench. Bioinformatics *25*, 1189-1191.

Wenink, P.W., Groen, A.F., Roelke-Parker, M.E., and Prins, H.H. (1998). African buffalo maintain high genetic diversity in the major histocompatibility complex in spite of historically known population bottlenecks. Molecular ecology *7*, 1315-1322.

Yewdell, J.W., and Bennink, J.R. (1999). Immunodominance in major histocompatibility complex class I-restricted T lymphocyte responses. . Annual Reviews of Immunology *17*.

Yewdell, J.W., Reits, E., and Neefjes, J. (2003). Making sense of mass destruction: quantitating MHC class I antigen presentation. Nature reviews. Immunology *3*, 952-961.

Yewdell, J.W. (2006). Confronting Complexicity: Real-world immunodominance in antiviral CD8+ T cell responses. Immunity *25*, 533-543.

York, I.A., Brehm, M.A., Zendzian, S., Towne, C.F., and Rock, K.L. (2006). Endoplasmic reticulum aminopeptidase 1 (ERAP1) trims MHC class I-presented

peptides in vivo and plays an important role in immunodominance. Proceedings of the National Academy of Sciences of the United States of America *103*, 9202-9207.

Young, A.S., Brown, C.G., Burridge, M.J., Cunningham, M.P., Kirimi, I.M., and Irvin, A.D. (1973). Observations on the cross-immunity between *Theileria lawrencei* (Serengeti) and *Theileria parva* (Muguga) in cattle. International journal for parasitology *3*, 723-728.

Young, A.S., Brown, C.G., Burridge, M.J., Cunningham, M.P., Payne, R.C., and Purnell, R.E. (1977a). Establishment of an experimental field population of *Theileria lawrencei*-infected ticks maintained by African buffalo (*Syncerus caffer*). The Journal of parasitology *63*, 903-907.

Young, A.S., Brown, C.G.D., Burridge, M.J., Grootenhuis, J.G., Kanhai, G.K., Purnell, R.E., and Stagg, D.A. (1978). The incidence of Theilerial parasites in East African buffalo (*Syncerus caffer*). Tropenmedizin und Parasitologie *29*, 281-288.

Young, A.S., Mutugi, J.J., Kariuki, D.P., Lampard, D., Maritim, A.C., Ngumi, P.N., Linyonyi, A., Leitch, B.L., Ndungu, S.G., Lesan, A.C., and et al. (1992). Immunisation of cattle against theileriosis in Nakuru District of Kenya by infection and treatment and the introduction of unconventional tick control. Veterinary parasitology *42*, 225-240.

Young, A.S., Ochanda, H., Perry, B.D., Morzaria, S.P., Dolan, T.T., Medley, G.F., and Gettinby, G. (1994). The biology of the transmission dynamics of *Theileria parva*, Vol ISVEE 7: Proceedings of the 7th International Symposium on Veterinary Epidemiology and Economics, Nairobi, Kenya (published as Vol 18(2) of The Kenya Veterinarian) (International Symposia on Veterinary Epidemiology and Economics).

Young, A.S., and Purnell, R.E. (1973). Transmission of *Theileria lawrencei* (Serengeti) by the ixodid tick, *Rhipicephalus appendiculatus*. Tropical animal health and production *5*, 146-152.

Young, A.S., Radley, D.E., Cunningham, M.P., Musisi, F.L., Payne, R.C., and Purnell, R.E. (1977b). Exposure of immunized cattle to prolonged natural challenge of *Theileria lawrencei* derived from african buffalo (*Syncerus caffer*). Veterinary parasitology *3*, 283-290.

Yusufmia, S.B., Collins, N.E., Nkuna, R., Troskie, M., Van den Bossche, P., and Penzhorn, B.L. (2010). Occurrence of *Theileria parva* and other haemoprotozoa in cattle at the edge of Hluhluwe-iMfolozi Park, KwaZulu-Natal, South Africa. Journal of the South African Veterinary Association *81*, 45-49.

Zweygarth, E., Koekemoer, O., Josemans, A.I., Rambritch, N., Pienaar, R., Putterill, J., Latif, A., and Potgieter, F.T. (2009). Theileria-infected cell line from an African buffalo (*Syncerus caffer*). Parasitology research *105*, 579-581.