Limited diversity in the CD8+ antigen-coding loci in *Theileria parva* parasites from cattle from southern and eastern Africa

Donald M. Lubembe[a],*, David O. Odongo[b], Fourie Joubert[c], Kgomotso P. Sibeko-Matjila[a]

[a]Vectors and Vector-borne Diseases Research Programme, Department of Veterinary Tropical Diseases, Faculty of Veterinary Science, University of Pretoria, Private Bag X04, Onderstepoort, South Africa
[b]School of Biological Sciences, University of Nairobi, P.O. Box 30197, Nairobi, 00100, Kenya
[c]Centre for Bioinformatics and Computational Biology, Faculty of Natural and Agricultural Sciences, University of Pretoria, Private Bag X20, Hatfield, South Africa

*Corresponding author. Present address at: Department of Veterinary Pathology, Microbiology and Parasitology, Faculty of Veterinary Medicine and Surgery, Egerton University, P.O. Box 536, Egerton, 20115, Kenya. Email: dlubembe@gmail.com

Highlights

•Limited diversity in antigenic loci in both ECF and Corridor disease parasites.
•Among ECF parasites, there are distinctive and common Tp2 epitope variants.
•Tp1 variant (VGYPKVKEEII) is common in both ECF and Corridor disease parasites.
•Tp4, Tp5, Tp7 and Tp8 are conserved in both parasite types.

Abstract

Theileria parva infections in cattle causes huge economic losses in the affected African countries, directly impacting the livelihood of the poor small-holder farmers. The current immunization protocol using live sporozoites in eastern Africa, is among the control measures designed to limit *T. parva* infections in cattle. However, the ability of the immune protection induced by this immunization to protect against field parasites has been compromised by the diversity of the parasite involving the schizont antigen genes. Previous studies have reported on the antigenic diversity of *T. parva* parasites from southern and eastern Africa, however, similar reports on *T. parva* parasites particularly from cattle from southern Africa remains scanty, due to the self-limiting nature of Corridor disease. Thus, we evaluated the diversity of CD8+ T-cell regions of ten schizont antigen genes in *T. parva* parasites associated with Corridor disease and East Coast fever (ECF) from southern and eastern Africa respectively.

Regions of schizont antigen (TpAg) genes containing the CD8+ T-cell epitopes (CTL determinants) were amplified from genomic DNA extracted from blood of *T. parva* positive samples, cloned and sequenced. The results revealed limited diversity between the two parasite groups from cattle from southern and eastern Africa, defying the widely accepted notion that antigen-encoding loci in cattle-derived parasites are conserved, while in buffalo-derived parasites, they are extensively variable. This suggests that only a sub-population of parasites is successfully transmitted from buffalo to cattle, resulting in the limited antigenic

diversity in Corridor disease parasites. Tp4, Tp5, Tp7 and Tp8 showed limited to absence of diversity in both parasite groups, suggesting the need to further investigate their immunogenic properties for consideration as candidates for a subunit vaccine. Distinct and common variants of Tp2 were detected among the ECF parasites from eastern Africa indicating evidence of parasite mixing following immunization. This study provides additional information on the comparative diversity of TpAg genes in buffalo- and cattle-derived *T. parva* parasites from cattle from southern and eastern Africa.

Keywords: *Theileria parva;* schizont antigen genes; diversity; cattle; southern and eastern Africa

1. Introduction

The protozoan parasite *Theileria parva*, transmitted by the brown ear tick, *Rhipicephalus appendiculatus* causes fatal lymphoproliferative disease syndromes in cattle in Africa. The common forms of *T. parva* infections in cattle are East Coast fever (ECF), Corridor disease and January disease present in various African countries (Neitz et al., 1955; Norval et al., 1992; Lawrence et al., 1994a, b; Nene et al., 2016), and in the Comoros Islands (De Deken et al., 2007). This parasites exerts significant economic consequences arising from the costs of tick and disease control, losses in milk and meat production, mortality of cattle, and hinderance to introduction of susceptible but high yielding exotic breeds of cattle in endemic areas (Mukhebi et al., 1992; Gachohi et al., 2012). East Coast fever is estimated to cause an annual loss of USD 300 million arising from an annual mortality of about 1 million cattle (Nene et al., 2016), and although not well documented, Corridor disease and January disease also cause vast economic losses in the affected countries. These economic losses largely affect the small-scale resource poor households whose livelihood depend on cattle production (Minjauw and McLeod, 2003; Randolph et al., 2007; Perry and Grace, 2009). The rain-fed crop production which supplements livestock production in these communities has also been negatively affected by the current climate change (Gornall et al., 2010). Therefore, in order to mitigate the negative impact on the health and economic status of the small-scale farmers, a multidisciplinary approach, including control of tick-transmitted pathogens of cattle like *T. parva*, is paramount.

Several control strategies for *T. parva* infections in cattle targeted at breaking the life cycle of the parasite have been devised. In eastern Africa, such strategies include regular tick control using acaricides (Young et al., 1988), chemotherapy using hydroxynaphthoquinone derivatives; buparvaquone and parvaquone (McHardy et al., 1985; Dolan et al., 1988), and the widely adopted immunization protocol (infection and treatment method) using live *T. parva* sporozoites (Radley et al., 1975; Di Giulio et al., 2009). In South Africa, where ECF was eradicated but Corridor disease still persists, transmission of *T. parva* from buffalo to cattle is limited by regular tick control, and restriction of buffalo-cattle interaction through strict monitoring of movement of buffalo from endemic to non-endemic areas (Mbizeni et al., 2013). Chemotherapy and immunization are prohibited in South Africa due to the risk of development of carrier status which could potentially become the source of infection to other susceptible cattle, resulting in the adaptation of the parasite in cattle as in the case of ECF.

Despite the wide adoption of sundry control methods of *T. parva* in the affected regions of Africa, limitations to these approaches have been reported. The increasing prevalence of acaricide resistance among tick populations (Gachohi et al., 2012; Robbertse et al., 2016; Vudriko et al., 2016), high cost of the acaricides (Minjauw and McLeod, 2003), and

environmental contamination with toxic residues, are major concerns to the use of acaricides. Use of theilericidal drugs and immunization with live sporozoites are not only expensive to the resource-poor smallholder farmers, but significantly promote development of *T. parva* carrier status. In addition, the immune protection induced by the current immunization protocol does not protect against buffalo-derived *T. parva* parasites responsible for Corridor disease (Sitt et al., 2015). Therefore, the need to improve or develop new cost-effective and environment friendly methods of *T. parva* control still remains.

Due to the aforementioned limitations of immunization using live sporozoites, efforts have been made to develop alternative vaccines based on the defined antigens on the sporozoite (Musoke et al., 1984; Nyagwange et al., 2018) and schizont (Gardner et al., 2005; Graham et al., 2006) developmental stages of *T. parva*. This has necessitated studies to understand the biology of the parasite, its genomic and antigenic diversity in the field, as well as the host's immune responses in protection against the parasite. Immunity to *T. parva* infections is mainly cell-mediated involving CD8+ cytotoxic T lymphocytes (CTL) which recognize parasite peptides bound to MHC Class I molecules found on the surface of infected lymphocytes (Bjorkman et al., 1987; McKeever et al., 1994; Taracha et al., 1995; McKeever, 2006). The parasite peptides are encoded by the antigen-encoding genes (designated Tp1-Tp10) found on the lymphocyte-transforming schizont stage of *T. parva* (Gardner et al., 2005; Graham et al., 2006), and their CD8+ T-cell epitopes (CTL determinants) have been defined (Akoolo et al., 2008; Graham et al., 2008; Nene et al., 2012). Tp1 and Tp2 have been demonstrated to dominate CTL responses (MacHugh et al., 2009; Connelley et al., 2011, 2016), whereas the roles of the other genes during an immune response still remains unclear (Morrison et al., 2015).

Analysis of *T. parva* cell lines and field parasites from cattle and buffalo from eastern Africa, as well as the vaccine parasites, revealed high levels of polymorphism in CTL determinants in buffalo-derived than cattle-derived parasites (Pelle et al., 2011; Hemmink et al., 2016; Salih et al., 2017; Sitt et al., 2018; Kerario et al., 2019). A comparative analysis of parasites of buffalo origin from Kenya and South Africa, based on similar antigen-encoding loci, revealed similar extensive diversity (Hemmink et al., 2018). Notwithstanding the extensive analysis of *T. parva* parasites from cattle and buffalo hosts, the extent of antigenic diversity in buffalo-derived *T. parva* parasites from the cattle host is yet to be established. East Coast fever parasites are believed to have originated from buffalo, and adapted to cattle for cattle-to-cattle transmission, while Corridor disease parasites are only transmissible from the buffalo to cattle, and do not establish a carrier status. Thus, the current study sort to extend the analyses, to determine the extent of polymorphism in schizont antigen-encoding loci in *T. parva* parasites associated with clinical and non-clinical cases of Corridor disease and ECF in southern and eastern Africa respectively.

2. Materials and methods

2.1. Ethics statement

This study was approved by the Animal Ethics Committee of the Faculty of Veterinary Science, University of Pretoria (AEC Certificate # V080-16). Additional approvals for collection of blood samples from each site, and import permits for shipping DNA samples to South Africa were obtained from the relevant authorities (Mukolwe et al., 2020).

Table 1. Geographical origin of blood samples collected from cattle, and *T. parva* antigen genes analyzed.

Geographical origin				Date	Number of samples collected[b]	*T. parva*-positive samples selected for PCR[f]	Region of antigen gene targeted for all samples	AT[g] (°C)
Country	**Province/Region**	**Locality**	**Collection site[a]**					
Kenya	Rift Valley Region	Nakuru County	Nakuru Municipality	2017	25[c] (15)	15	Tp1(CD8+ T-cell epitope)	50
Uganda	**Mbarara Region**	Kiruhura District	Nyakashashara & Sanga sub-counties	2017	137 (73)	20	Tp2(Full-length ORF)	65
	Karamoja Region	**Kaabong District**	Sidok & Karenga sub-counties	2017	99 (08)	08	Tp3(antigenic region)	65
		Nakapiripirit District	Nabilatuk sub-county	2017	123 (21)	12	Tp4(CD8+ T-cell epitope)	55
Tanzania	**Tanga Region**	Tanga District	Tanga	2014	20[d] (06)	06	Tp5(CD8+ T-cell epitope	65
	Manyara Region	Simanjiro Districts	Simanjiro plains	2014	20[d] (15)	15	Tp6(antigenic region)	55
South Africa	**KwaZulu-Natal Province**	uMkhanyakude District	Hluhluwe, Dukuduku & Hlabisa	2016	223 (04)	04	Tp7 (CD8+ T-cell epitope)	65
							Tp8 (CD8+ T-cell epitope)	65
	Mpumalanga Province	**Bushbuckridge Municipality (Mnisi)**	Hluvukani	2013	24[d], [e] (24)	24	Tp9 (Full-length ORF)	53
							Tp10 (antigenic region)	50
Total					671 (166)	104		

[a]There was no history of immunization (using Muguga cocktail) of cattle in the collection sites in East Africa as at the time of sample collection, except for Tanzania.
[b]The number in brackets indicates samples that tested positive for *T. parva* on qPCR assay (Sibeko et al., 2008). Except for cattle whose clinical status is provided, all other cattle sampled were non-clinical cases.
[c]10 of these samples were collected from active clinical cases of East Coast fever, and 15 from non-clinical cases.
[d]Samples obtained from previous studies.
[e]14 of these samples were collected from active clinical cases of Corridor disease, and 10 from non-clinical cases from a herd with previous *T. parva* infections. Detail on these samples has been provided previously (Mukolwe et al., 2020).

[f]PCR on all the 10 *T. parva* antigen genes.
[g]AT - annealing temperature.

2.2. Sample collection and detection of *T. parva*

A maximum of 4 mL of whole blood was collected in EDTA vacutainers (Becton Dickinson, Oxford, UK) from cattle from South Africa, Kenya, Tanzania and Uganda, as indicated in Table 1. Two hundred and forty-seven (247) samples were collected from clinical and non-clinical cases of Corridor disease in South Africa, and 424 from ECF cases in eastern Africa. DNA extraction and detection of *T. parva* genomic DNA was performed as described by Mukolwe et al. (2020).

2.3. PCR amplification

A total of 104 *T. parva* positive DNA samples, representing all study sites were selected for PCR amplification of 10 TpAg genes (Tp1−Tp10) (Table 1). At least 50 ng of DNA and 10 pmol of each primer were used in a total reaction volume of 12.5 µL. Primer sequences and the expected sizes (bp) of amplicons from each antigen gene have been provided previously (Hemmink et al., 2016). Amplification conditions were modified according to Phusion Flash High-Fidelity PCR Master Mix protocol as described by Mukolwe et al. (2020), using the annealing temperatures indicated in Table 1. PCR products were analyzed by agarose gel electrophoresis using a 2% agarose containing ethidium bromide in 1X TAE running buffer.

2.4. Ion torrent sequencing of Tp1−Tp8, and Tp10 genes

2.4.1. Preparation of PCR products

The number of samples successfully amplified for each TpAg gene and those selected for cloning and sequencing for each sample group (geographic) is shown in Table 2. Purification of the PCR products from all TpAg genes and cloning was done as previously described (Mukolwe et al., 2020). Recombinant clones were selected as indicated in Table 3, purified and then pooled in equimolar concentrations in two stages (Procedure S1) before sequencing. Consequently, for each sample group; the first pooling resulted in nine pools (except sample groups from Tanzania) (Table 3 and Procedure S1), where each pool represented each TpAg gene, while the second pooling resulted in one pool which included the nine TpAg gene products (Table 3 and Procedure S1). The DNA concentrations of the final seven samples (representing the 7 sample groups) were determined using Qubit 2.0 fluorometer (Invitrogen, Carlsbad USA) (Table 3), and further resolved in 2% agarose gel electrophoresis using 1X TAE as the running buffer before library preparation. The minimum recommended DNA concentration for Ion Torrent sequencing is 10 ng/µL.

2.4.2. Library preparation

Since Ion Torrent sequencing has a sequence read size limit of 400 bp, and the sizes of the amplicons ranged from 450 to 650 bp, a fragmentation stage was included in the library preparation. The samples were fragmented using the Ion Xpress™ Plus Fragment Library kit (ThermoFisher Scientific, USA) according to the Ion Xpress™ Plus gDNA fragment library preparation protocol (Publication number MAN0009847). This was followed by an initial clean-up, ligation of adapters and barcodes (for specific sample groups), and final clean-up according to the same protocol. Subsequently, each of the 7 samples (pools) corresponding to each sample group had a unique barcode. Size selection of the fragmented samples was done using the E-Gel™SizeSelect™ II Agarose Gels, 2% (Invitrogen, Carlsbad USA). The

Table 2. Samples successfully amplified for each TpAg gene and further selected for cloning.

Sample group[a]	Number of samples successfully amplified on targeted regions of TpAg genes[b]										Samples selected for cloning[c]	Samples selected for Tp9 cloning	Clones selected for Sanger sequencing[d]
	Tp1	**Tp2**	**Tp3**	**Tp4**	**Tp5**	**Tp6**	**Tp7**	**Tp8**	**Tp9**	**Tp10**			
Nakuru cattle	12	12	13	12	12	13	11	12	10	13	10	08	10
Mbarara cattle	15	16	16	17	15	17	15	15	14	16	10	10	10
Karamoja cattle	15	13	14	15	13	15	13	15	12	14	10	10	10
Tanga cattle	04	04	ns	ns	ns	ns	ns	ns	04	ns	04	04	06
Simanjiro cattle	11	10	ns	ns	ns	ns	ns	ns	09	ns	10	09	10
uMkhanyakude cattle	00	00	00	00	00	00	00	00	00	00	–	–	–
CD clinical cases (Hluvukani)	10	10	09	09	09	10	09	11	08	10	09	08	10
Non-clinical *T. parva*-positive cases (Hluvukani)	01	01	01	01	01	01	01	01	01	01	01	01	04
Total	68	66	53	54	50	56	49	54	58	54	54	50	60

[a]CD − Corridor disease.
[b]ns − "no samples" were available for analysis due to low quantities of DNA.
[c]Selection was random and based on the quality of PCR amplicons. The number of selected samples represent each antigen gene (except Tp9) for each sample group.
[d]Clones refer to those from Tp9 and were selected based on the quality of colony PCR amplicons. At least one colony was selected from each sample.

Table 3. Samples selected for cloning per gene, and recombinant clones selected for preparing sample pools for Ion Torrent sequencing.

Sample group[a]	Amplicons selected for cloning	1st Pooling[b]		2nd Pooling[c]	
		Samples selected for pooling	1st Pools	2nd Pools	Final Concentration (ng/µL)
Nakuru cattle	10	10−20	9	1	70.8
Mbarara cattle	10	10−20	9	1	68.0
Karamoja cattle	10	10−20	9	1	66.4
Tanga cattle	04	05−10	2	1	35.2
Simanjiro cattle	10	10−20	2	1	90.8
CD clinical cases (Hluvukani)	09	10−15	9	1	80.0
Non-clinical *T. parva*-positive case (Hluvukani)	01	05−10	9	1	66.8
Total	54		49	7	

[a]Due to limited quantity of DNA, the two sample groups from Tanzania were analyzed on Tp1 and Tp2 only.
[b]Samples refer to recombinant clones selected for each TpAg gene. At least one colony was selected from each sample. The number of pools represent the number of TpAg genes in the specific sample groups.
[c]Each pool includes all TpAg genes in the specific sample groups.

libraries were then PCR amplified, cleaned-up and quantified using Qubit 2.0 fluorometer (Invitrogen, Carlsbad USA) following the manufacture's protocol.

2.4.3. Ion Proton™ and ion S5™ systems sequencing

Deep sequencing was done on the Ion Proton™ and Ion S5™ Systems (ThermoFisher Scientific, Carlsbad, USA) at the Central Analytical Facilities, Stellenbosch University, South Africa. On the Ion Proton™ System, template amplification was completed using the Ion PI™ Hi-Q™ Chef Kit (Part No. A27198) (ThermoFisher Scientific, Carlsbad, USA). Subsequent sequencing was done on the Ion Torrent Proton using the Ion PI™ Hi-Q™ Sequencing 200 Solutions (Part No. A26430) with the Ion PI™ Chip Kit v3 (Part No. A26770) (ThermoFisher Scientific, Carlsbad, USA). Template amplification and sequencing was performed as described in the manufacturer's protocol (MAN0010967, Rev B.0). On the Ion S5™ System, template amplification was completed using the Ion 520™ & Ion 530™ Chef-Kit (Part No. A30010) (ThermoFisher Scientific, Carlsbad, USA). Subsequent sequencing was done on the Ion GeneStudio™ S5 system using the Ion S5™ Sequencing Solutions (Part No. A27767) and Ion S5™ Sequencing Reagents (Part No. A27768) with the Ion 530™ Chip Kit (Part No. A27764) (ThermoFisher Scientific, Carlsbad, USA). Template amplification and sequencing was performed as described in the manufacturer's protocol (MAN0010846, Rev D.0).

2.5. Sanger sequencing of Tp9 gene

Approximately 4–10 clones were selected per sample group for Tp9 sequencing (Table 2). Recombinants were confirmed by colony PCR using PCR conditions described earlier, following which amplicons were purified using the QIAquick® PCR Purification Kit (Qiagen, Germany) as described by the manufacturer. Bidirectional sequencing of the purified products was done using pJET1.2 primers on the ABI 3500XL Genetic Analyzer, POP-7™ (ThermoFisher Scientific, USA) at INQABA Biotechnologies, South Africa.

2.6. Sequence data analysis

2.6.1. Analysis of ion torrent sequencing data

A total of ~20 million sequence reads in FASTQ format were obtained from all sample groups with each group contributing ~1.7 million reads. Adapter and barcode sequences, and pJET1.2 primers were trimmed off using FLEXBAR (Dodt et al., 2012) and CutPrimers (Kechin et al., 2017) respectively. The quality of sequence reads was checked using FastQC version 0.11.8 (Andrews, 2014) where ends of sequence reads with a per base sequence quality score <20 were trimmed off using TrimGalore Version 0.5.0 (Andrews, 2014). Trimmed sequences were then mapped to the TpAg genes reference sequences (Table 4) using Bowtie2 version 2.3.4.3 (Langmead and Salzberg, 2012), where Binary Alignment Map (BAM) files were generated for each TpAg gene per sample group. In order to visualize regions of sequence reads with SNPs, the alignment maps were viewed using Integrative Genomics Viewer (IGV) version 2.4.15 (Robinson et al., 2011). The BAM files were then converted to FASTA files using SAMtools (Li et al., 2009), for use in subsequent analyses.

Table 4. TpAg gene reference sequences used in the alignment of respective sequence reads.

Antigen Gene	Reference Sequence (Locus Tag)	Accession number	Antigen annotation[a]	CD8+ T-cell Epitope	ROI[b] (defined by reference nucleotide positions)	Reference
Tp1	***T. parva* Muguga (TP03_0849)**	XM_757880	Hypothetical	Tp1$_{214-224}$VGYPKVKEEML	Incl:779−811	(Graham et al., 2008)
					Excl:1−778;812−1040	
Tp2	***T. parva* Muguga (TP01_0056)**	XM_760490	Hypothetical	Tp2$_{27-37}$SHEELKKLGML	Incl:207−239	(Graham et al., 2008)
					Excl:1−206; 240−245	
				Tp2$_{40-48}$DGFDRDALF	Incl:246−272	(Nene et al., 2012)
				Tp2$_{49-59}$ KSSHGMGKVGK	Incl:273−305	(Graham et al., 2008)
				Tp2$_{50-59}$SSHGMGKVGK	Incl:276−305, Excl:306−413	(Connelley et al., 2016)
				Tp2$_{96-104}$FAQSLVCVL	Incl:414−440	(Graham et al., 2008)
				Tp2$_{98-106}$QSLVCVLMK	Incl:420−446, Excl:447−539	(Graham et al., 2008)
				Tp2$_{138-147}$KTSIPNPCKW	**Incl:540−569**	(Akoolo et al., 2008)
					Excl:570−675	
Tp3	*T. parva* Muguga (TP01_0868)	XM_761296	Hypothetical	Unidentified	Incl:280−356;360−550;590−650, Excl:1−279;357−359;551-589;651−730	(Gardner et al., 2005)
Tp4	***T. parva* Muguga (TP03_0210)**	XM_758135	eta-TCP1	Tp4$_{328-336}$TGASIQTTL	Incl:982−1008	(Graham et al., 2008)
					Excl:1−981;1009−1060	
Tp5	***T. parva* Muguga (TP02_0767)**	XM_760241	eIF-1A	Tp5$_{87-95}$SKADVIAKY	Incl:318−344	(Graham et al., 2008)
					Excl:1−317;345−435	
Tp6	***T. parva* Muguga (TP01_0188)**	XM_760622	Prohibitin	Unidentified	Incl: 220−550	(Gardner et al., 2005)
					Excl: 1−219;551−625	
Tp7	***T. parva* Muguga (TP02_0244)**	XM_759717	Hsp90	Tp7$_{206-214}$EFISFPISL	Incl: 718−744	(Graham et al., 2008)
					Excl: 1−717;745−900	
Tp8	***T. parva* Muguga (TP02_0140)**	XM_759616	Cysteine proteinase	Tp8$_{379-387}$CGAELNHFL	Incl: 1135−1161	(Graham et al., 2008)
					Excl: 1−1134;1162−1300	
Tp9	*T. parva* Muguga (TP02_0895)	XM_760370	Hypothetical	Tp9$_{67-76}$AKFPGMKKSK	N/A	(Nene et al., 2012)
Tp10	*T. parva* Muguga (TP04_0772)	XM_759315	Coronin	Unidentified	Incl: 770−930, Excl: 1−769	(Gardner et al., 2005)

[a]eta-TCP1 (eta subunit of the T-complex protein 1), eIF-1A (translation elongation initiation factor 1A), Hsp90 (heat shock protein 90).
[b]ROI − "Region of Interest" defined by the indicated nucleotide positions on the reference sequence, Incl - ROI of the sequence reads included in the analysis as defined by the indicated nucleotide positions, Excl − region of the sequence reads excluded from the analysis as defined by the indicated nucleotide positions, N/A − not applicable.

For identification of nucleotide and the predicted amino acid sequence variations within the antigenic epitopes, the sequence reads in FASTA format, for each TpAg gene in each sample group, were aligned to the GenBank files (.gb files) of the respective reference sequences (Table 4) using NextGENe version 2.4.2.3 (https://softgenetics.com/NextGENe.php). In the NextGENe analysis setup, the overall matching base percentage was set at 85% and sequence reads within 100–400 bp size range were selected for alignment. In order to align reads that contained indels, a rigorous alignment was performed where the overall mutation score was set at ≥ 12.00, a value equivalent to the statistical 95% confidence, indicating that the mutation call is true. Since Tp1, Tp2, Tp4, Tp5, Tp7 and Tp8 have defined antigenic epitopes (Table 4), these epitope regions were set as the "regions of interest" (ROI) (Table 4) in the mutation report settings. Tp3, Tp6 and Tp10 do not have defined antigenic epitopes and therefore, the known antigenic regions (Hemmink et al., 2016) of the aligned sequence reads were set as ROI (Table 4). The ROI were defined by the first and last nucleotides of each epitope or antigenic region in a text file (Table 4). All other alignment and mutation report settings were used as default on NextGENe or adjusted according to NextGENe user's manual. The mutation reports generated, which reflected the ROI only, were used to tabulate the mutation details in the CD8+ T-cell epitope or antigenic region of each TpAg gene for each sample group. Bar charts were used for comparative analysis of the epitope variants identified in *T. parva* parasites from cattle from eastern and southern Africa.

2.6.2. Analysis of data from Sanger sequencing

Raw nucleotide sequences were confirmed to be Tp9 sequences by sequence similarity analysis using the Basic Local Alignment Search Tool (BLAST) executed on the National Centre for Biotechnology Information (NCBI) platform. Sequence assembly, editing and translation was done using CLC Main Workbench version 8.0 (Qiagen, Germany). Multiple sequence alignment of consensus sequences together with the Tp9 reference sequence (Table 4) was done using the online version 7 of MAFFT (Katoh and Standley, 2013) applying the default parameters (http://mafft.cbrc.jp/alignment/server/). Aligned sequences/sequence matrices were viewed, edited manually and truncated using MEGA version 7 (Kumar et al., 2016). Phylogenetic analysis of predicted protein sequences selected based on the epitope variants was performed using MEGA version 7 (Kumar et al., 2016). Evolutionary history was inferred using Maximum Likelihood, by applying the default substitution model (General Reverse Transcriptase model) in MEGA 7, and the phylogeny tested using 100 bootstrap replicates, where a bootstrap support ≥0.7 (70%) was considered significant.

3. Results

3.1. Tp1−Tp8, and Tp10 epitope variants

The overall sequence coverage for the targeted TpAg genes epitope/antigenic regions from cattle sample groups ranged from 817 to 94,958 reads (Tables S1-S9). Variants were detected in all TpAg gene epitopes/antigenic regions except in two of the Tp2 epitopes (^{40}DGFDRDALF^{48} and ^{96}FAQSLVCVL^{104}), Tp7 and Tp8 epitopes, and the antigenic region of Tp6 (Fig. 1). Most epitope variants were detected in Tp2, where epitope ^{27}SHEELKKLGML^{37} had five variants, ^{138}KTSIPNPCKW^{147} three variants, while ^{49}KSSHGMGKVGK^{59} and ^{50}SSHGMGKVGK^{59} had two variants each (Fig. 1 and Table 5). Tp4 (^{328}TGASIQTTL^{336}) had two variants, whereas Tp1 (^{214}VGYPKVKEEML^{224}), Tp2 (^{98}QSLVCVLMK^{106}) and Tp5 (^{87}SKADVIAKY^{95}) epitopes were the least variable with each having one variant (Fig. 1 and Table 5). Among the TpAg genes with unidentified antigenic

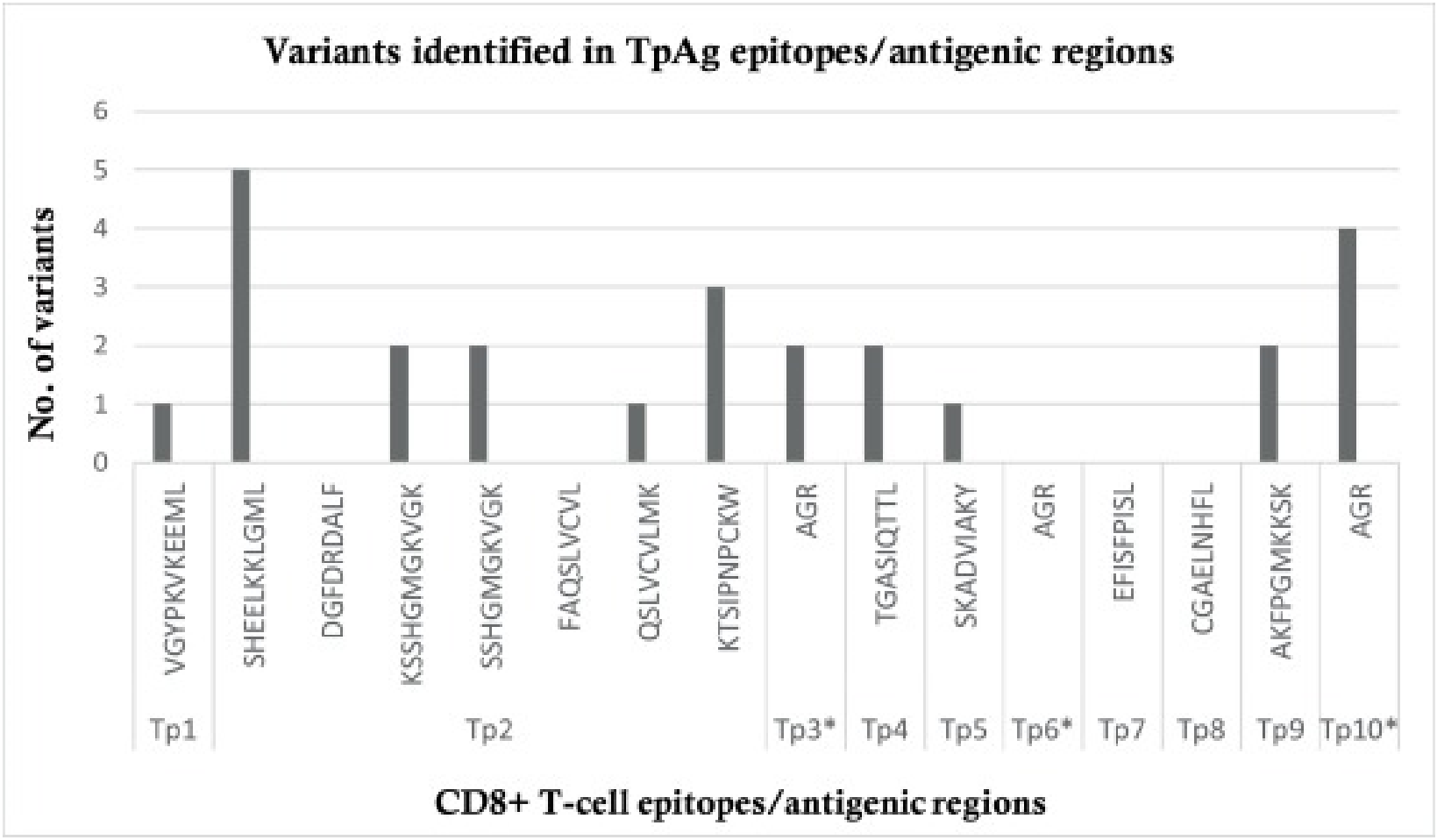

Fig. 1. Variants identified in TpAg gene epitopes/antigenic regions in T. parva parasites associated with Corridor disease and ECF from southern and eastern Africa respectively.

*Antigen genes with unidentified epitopes hence antigenic regions (AGR) were analyzed.

Table 5. CD8 + T-cell epitope variants identified in ECF and Corridor disease parasites from eastern and southern Africa respectively.

Country	Sample group	CD8 + T-cell epitope variants[a]				
		Tp1	**Tp2**	**Tp4**	**Tp5**	**Tp9**
Kenya	**Nakuru cattle**	VGYPKVKEE**II**	S**D**EELKKLGML	TGASIQT**S**L	–	AKFPGMKK**G**K
			SHEELNKLGML			
Uganda	**Mbarara cattle**	VGYPKVKEE**II**	K**P**SIPNPCKW	TG**D**SIQTTL	NKADVIAKY	–
			KTSVPNPCKW			
			KTSIPNPCEW			
	Karamoja cattle	**–**	K**P**SIPNPCKW	–	–	AKFPGMKK**G**K
			KTSVPNPCKW			
			KTSIPNPCEW			
Tanzania	**Tanga cattle**	VGYPKVKEE**II**	S**D**EELKKLGML	NA	NA	AKFPGMKK**G**K
			SHEELNKLGML			
	Simanjiro cattle	**–**	SH**G**ELKKLGML	NA	NA	–
			SDEELKKLGML			
			SDEELKKLGML			
South Africa	**CD clinical cases (Hluvukani)**	VGYPKVKEE**II**	SH**D**ELKKLGML	–	–	**N**KFPGMKK**G**K
			SHEELTKLGML			AKFPGMKK**G**K
			KTSHGMGKVGK			
			KSSHAMGKVGK			
			TSHGMGKVGK			
			SSHAMGKVGK			
			ESLVCVLMK			
	Non-clinical *T. parva*-positive cases (Hluvukani)	–	–	–	–	**N**KFPGMKK**G**K

[a]Bolded and underlined shows the variant amino acid residue(s) in the epitope. Variants of the immunodominant epitopes ($Tp1_{214-224,}$ $Tp2_{49-59}$ and $Tp2_{50-59}$) are highlighted in grey. NA (not analyzed) for the respective sample groups.

epitopes, Tp10 and Tp3 were the most and least variable with four and two non-synonymous substitutions within the antigenic regions, respectively (Fig. 1 and Tables S3 & S9). Tp6 was invariable with all substitutions being synonymous (Fig. 1 and Table S6).

3.2. Tp9 epitope variants

A total of 41 Tp9 sequence reads were obtained after quality and BLAST analysis of 60 consensus sequences obtained from 50 selected samples from cattle (Table 2). The sequence reads were aligned to the previously published *T. parva* Muguga Tp9 sequence (Table 2). Two variants; ^{67}AKFPGMKKGK^{76} and ^{67}NKFPGMKKG K^{76}, were identified in Tp9 epitope (Table 5), one (^{67}AKFPGMKKG K^{76}) from ECF parasites from eastern Africa, and two (^{67}AKFPGMKKG K^{76} and ^{67}NKFPGMKKG K^{76}) from Corridor disease parasites from cattle in South Africa (Table 5 and Fig. 2). Phylogenetic analysis displayed clustering that was congruent to the variants identified, where three main clusters corresponding to the Muguga epitope (^{67}AKFPGMKKSK^{76}) and the two variants (^{67}AKFPGMKKGK^{76}, ^{67}NKFPGMKKGK^{76}) were identified (Fig S1).

3.3. Comparison of variants identified in Corridor disease and ECF parasites

Analysis of variants (n = 19) from *T. parva* parasites associated with Corridor disease and ECF revealed no significant difference in the mean number of variants identified per group (p>0.05, 2-tailed T test). Of the 11 variants detected in each parasite group, majority were from Tp2 (Fig. 2 and Table 6), where variants in the immunodominant epitopes of Tp2 were detected in Corridor disease parasites from South Africa only (Table 5). The only Tp1 variant identified was common for the two parasite groups, and the other variants that also occurred in both groups were Tp2 (S**D**EELKKLGML) and Tp9 (AKFPGMKK**G**K). Variants in Tp4 and Tp5 epitopes were identified in ECF parasites from eastern Africa only (Fig. 2 and Table 6). The conserved epitopes in parasites from both groups were Tp2 (^{40}DGFDRDALF^{48} and ^{96}FAQSLVCVL^{104}), Tp7 (^{206}EFISFPISL^{214}) and Tp8 (^{379}CGAELNHFL^{387}), as well as the Tp6 antigenic region. When Tp2 epitopes variants from ECF parasites from the three east African countries were compared, no variant was common to the three countries. Only Kenya and Tanzania had two common variants, while Tanzania and Uganda had one and three unique variants respectively, and no variant was unique to Kenya (Fig. 3).

Table 6. Summary of the variants exclusive and common to ECF and Corridor disease parasites from southern and eastern Africa respectively.

Group	TpAg gene	Epitope variants	Number (%)
Exclusively in "ECF parasites"	**Tp2**	SHEEL**N**KLGML, SH**G**ELKKLGML, K**P**SIPNPCKW, KTS**V**PNPCKW, KTSIPNPC**E**W	8 (42.1)
	Tp4	TGASIQT**S**L, TG**D**SIQTTL	
	Tp5	**N**KADVIAKY	
Exclusively in "CD parasites"	**Tp2**	SH**D**ELKKLGML, SHEEL**T**KLGML, K**T**SHGMGKVGK, KSSH**A**MGKVGK, **T**SHGMGKVGK, SSH**A**MGKVGK, **E**SLVCVLMK	8 (42.1)
	Tp9	**N**KFPGMKK**G**K	
Common in "ECF and CD parasites"	**Tp1**	VGYPKVKEE**II**	3 (15.8)
	Tp2	S**D**EELKKLGML	
	Tp9	AKFPGMKK**G**K	

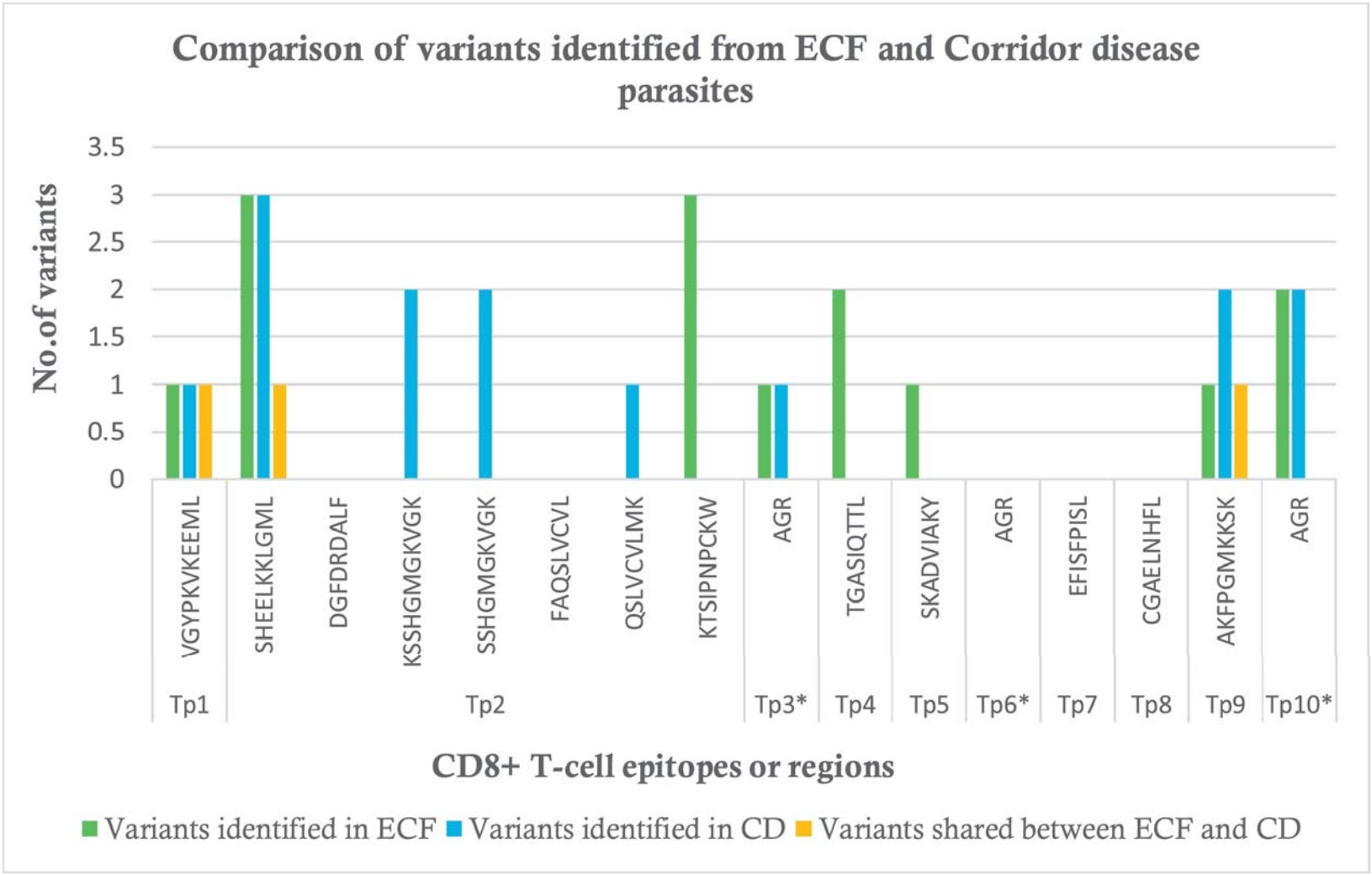

Fig. 2. Comparison of variants identified in TpAg genes epitopes/antigenic regions in *T. parva* parasites associated with ECF and Corridor disease.

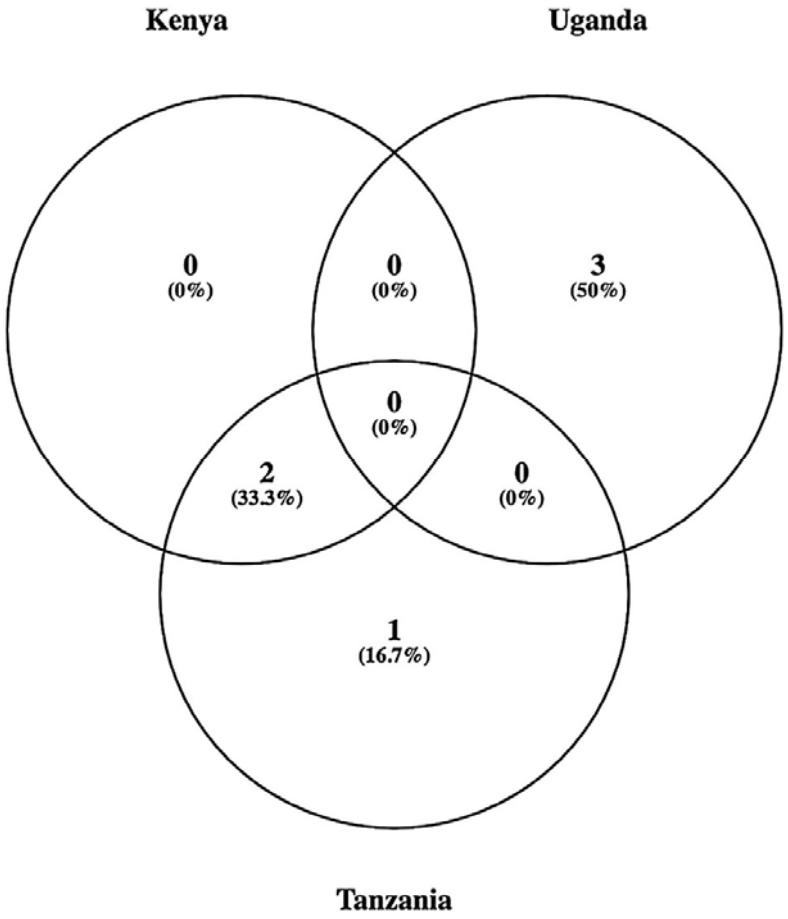

Fig. 3. Comparison of Tp2 epitopes variants from ECF parasites, showing variants common and unique to parasites from Kenya, Tanzania and Uganda.

4. Discussion

Theileria parva is a widespread tick-transmitted protozoal parasite that infects cattle in eastern and southern Africa, causing ECF and Corridor disease. Both disease syndromes impact negatively on the economic status and health of communities that rely on cattle production for their livelihood (Randolph et al., 2007; Perry and Grace, 2009). Control of *T. parva* infections in cattle using the current immunization protocol (Radley et al., 1975) widely deployed in eastern Africa (Di Giulio et al., 2009) has been considerably limited by antigenic diversity in field parasites, especially the buffalo-derived *T. parva*. Hence, efforts have been made to develop an alternative subunit vaccine that could induce immune protection against most field parasites based on the mapped *T. parva* antigens (Musoke and Nene, 1990; Musoke et al., 2005; Graham et al., 2006; Mwangi et al., 2011). In an effort to understand the functional differences in the immune protection induced by the schizont antigen-based vaccines, various studies have investigated possible hypotheses, including but not limited to examining the diversity of CTL determinants in *T. parva* field parasites. However, comparative diversity of the CD8+ antigen-coding loci in *T. parva* parasites associated with clinical and non-clinical cases of ECF and Corridor disease have not been investigated until the current study. Sequence analysis of CTL determinants from 10 schizont antigen genes showed that Tp2 was the most polymorphic, followed by Tp9, Tp4, Tp1 and Tp5 in that order. Tp6, Tp7 and Tp8 were the most conserved in *T. parva* populations from

cattle from eastern and southern Africa. The number of epitope variants in Tp1, Tp2, Tp4, Tp5 and Tp9, occurred at comparable levels in both parasite groups ($p > 0.05$), with Tp1 and Tp9 having variants that were common for both ECF and Corridor disease parasites.

Similar to the previous reports (Pelle et al., 2011; Salih et al., 2017; Kerario et al., 2019), variants in Tp1 epitope involve the last two terminal amino acid residues, with the common variant being the one ending with two isoleucine residues (-**II**). Tp1 epitope has been identified as one of the immunodominant epitopes recognized by CTL in response to *T. parva* infection (MacHugh et al., 2009; Connelley et al., 2011). The only variant identified in the 11-mer Tp1$_{214\text{-}224}$ epitope was common for the two parasite groups, and entailed substitutions at positions 10 and 11 involving two terminal amino acid residues. In this variant, ^{214}VGYPKVKEE**II**^{224}, methionine and leucine were substituted with isoleucine. The three amino acid residues (methionine, leucine and isoleucine) that seem to be commonly involved in mutations in this epitope are non-polar and similar in size and shape, a feature that promotes cross-reactivity of the CTL (Frankild et al., 2008; Steinaa et al., 2012). Substitutions involving amino acid residues at positions 3, 6, 7, 8 and 9 of the epitope are reported to substantially affect CTL killing of infected lymphocytes (Macdonald et al., 2010; Steinaa et al., 2012), whereas position 10 and 11 substitutions fairly reduce CTL killing (Macdonald et al., 2010). Consistent with the previous studies (Pelle et al., 2011; Salih et al., 2017; Hemmink et al., 2018; Kerario et al., 2019), the current study demonstrate that Tp1 epitope is not as polymorphic as Tp2, and that the common mutations occurring in this epitope in both ECF and Corridor disease parasites, mostly involve the two carboxy-terminal amino acid residues. Considering the involvement of the Tp1 epitope in the recognition of *T. parva* infected cells by the host's CTL, and that the variant identified is likely to be cross-reactive to CTL, and is common to both cattle- and buffalo-derived parasites from cattle, reinforces the evidence that Tp1 could be an ideal candidate for a subunit vaccine for both parasite types.

Two variants of Tp9 epitope were identified; two (**N**KFPGMKK**G**K, AKFPGMKK**G**K) from the buffalo-derived *T. parva* parasites from cattle from South Africa, and one (AKFPGMKK**G**K) in parasites from eastern Africa, which was also common for both parasite groups, suggesting that Tp9 gene is less variable in cattle-derived than buffalo-derived parasites from cattle. It has been demonstrated that the Tp9 epitope elicits a weak CTL response, hence it is not a suitable vaccine candidate (Bastos et al., 2019). On the contrary, a full-length recombinant signal peptide of Tp9 has been shown to induce both humoral and cellular response, and with further immunological evaluation, it could be a potential candidate for the *T. parva* subunit vaccine (Bastos et al., 2019). Since the possible effects of the natural variants of Tp9 epitope on CTL recognition have not been investigated, an evaluation of the significance of substitutions of specific epitope residues on CTL recognition is necessary, otherwise, any speculations on the same are unfounded.

Tp4 and Tp5 were conserved in Corridor disease parasites and demonstrated limited diversity among ECF parasites from eastern Africa. Tp6, Tp7, and Tp8 were highly conserved in both parasite types. Although the specific roles of Tp4, Tp5, Tp7 and Tp8 in CD8+ T-cell response have not been defined (Morrison et al., 2015), further investigation of their immunogenic properties would be significant, considering the level of conservation in these loci in both parasite types. Similarly, the roles of three antigens with undefined epitopes (Tp3, Tp6 and Tp10) in immunity against *T. parva* is not known. Therefore, until their roles in *T. parva* immunity are demonstrated, sequence variations on these genes could only suggest possible differences among *T. parva* populations from the two groups investigated.

Among the former four TpAg genes with defined epitopes, non-synonymous amino acid substitutions were only detected in Tp4 and Tp5, with two and one variants respectively, exclusively detected in cattle-derived parasites from eastern Africa. Previous studies on parasites from buffalo in Kenya (Hemmink et al., 2018; Sitt et al., 2018) and South Africa (Hemmink et al., 2018) detected limited variations on these four TpAg genes. It was rather peculiar in the current study that epitope variants of these genes (Tp4 and Tp5) were detected in cattle-derived but not buffalo-derived parasites, as the latter would be expected to have more variants. However, since diversity of TpAg genes is not a universal phenomenon (Sitt et al., 2018), there is a likelihood that some TpAg genes may display limited to no diversity in both buffalo- and cattle-derived *T. parva* parasites from cattle. It could also be possible that these genes are either under negative selection (Hemmink et al., 2018), or much of the evolution of *T. parva* involving these genes has occurred in the wildlife reservoir, and that the evolved parasites are maintained in the wildlife cycle (buffalo/ticks), as well as in cattle.

Seven antigenic epitopes have been identified on the Tp2 gene (Akoolo et al., 2008; Graham et al., 2008; Nene et al., 2012; Connelley et al., 2016), hence the high level of polymorphism in this gene. Only two of the seven Tp2 epitopes seem to be conserved in both *T. parva* parasites types. Notably, four variants of the immunodominant epitopes, $Tp2_{49\text{-}59}$ (K**T**SHGMGKVGK, KSSH**A**MGKVGK) and $Tp2_{50\text{-}59}$ (**T**SHGMGKVGK, SSH**A**MGKVGK) were exclusively identified in Corridor disease parasites, and involved substitutions at positions 2 and 5 in $Tp2_{49\text{-}59}$, and positions 1 and 4 in $Tp2_{50\text{-}59}$. Although the natural variants of these epitopes would be expected to affect the MHC-binding capacity, the magnitude of their effect depends on the number of substitutions and the epitope positions affected (Connelley et al., 2011). Nevertheless, any change that disrupts the distinct epitope ligands for the T-cell receptor (TCR) recognition will result in partial or total escape from immune recognition (Connelley et al., 2011). The CTL response induced by the other Tp2 epitopes has been reported to be minimal (Morrison et al., 2015). Furthermore, it has been demonstrated that substitutions resulting in the loss of lysine residue at position 8 of $Tp2_{49\text{-}59}$, which corresponds to position 7 of $Tp2_{50\text{-}59}$, result in evasion of TCR recognition (Connelley et al., 2011). Although the epitope variants identified in the current study may affect MHC-binding capacity, it is evident that there could be no abrogation of TCR recognition since none of the substitutions involve the residue at position 8. The results of the current study suggest that the immunodominant epitopes of Tp2 are less prone to mutation in cattle-derived parasites, and that the variants identified largely represent a *T. parva* sub-population circulating in the wildlife reservoir and sporadically transmitted to cattle to cause Corridor disease.

Previous studies on the diversity of *T. parva* CD8+ genes have reported that the schizont antigen genes are more variable in parasites of buffalo origin than cattle-derived parasites, and that Tp2 is the most polymorphic (Pelle et al., 2011). However, in the current study, Tp2 showed limited diversity in both ECF and Corridor disease parasites, contrary to the expected high polymorphism in the latter group (buffalo-derived parasites). One possible explanation to this occurrence could be that, only a sub-population of parasites transmitted from the buffalo to cattle can infect and subsequently transform the bovine lymphocytes, and that this presumed selection probably contributes to the observed limited diversity. It would therefore be significant to characterize more parasites from Corridor disease cases from southern and eastern Africa to support or invalidate this hypothesis. Distinct variants among ECF parasites were detected, where parasites from Uganda and Tanzania had unique variants, and parasites from Kenya and Tanzania had two common variants. It is most likely that the wide deployment of the Muguga cocktail vaccine in Tanzania compared to Uganda contributed to

the presence of common variants with parasites from Kenya. The three components of the vaccine (Muguga, Kiambu 5 and Serengeti-transformed) are isolates from Kenya, and therefore chances of parasite mixing with the local *T. parva* strains in Tanzania following vaccination are high.

The TCR of CD8+ cytotoxic T lymphocytes recognize parasite peptides encoded by the epitope regions on the antigen genes and presented by MHC class I haplotype. Three epitopes ($Tp1_{214\text{-}224}$, $Tp2_{49\text{-}59}$ and $Tp2_{50\text{-}59}$) have been shown to be essential and predominate immune response to *T. parva* infections (MacHugh et al., 2009; Connelley et al., 2011, 2016). However, the CTL response to *T. parva* is also affected by the polymorphism of the host's MHC molecules cognate to the Bovine Leukocyte Antigen (BoLA) haplotype (reviewed in Steinaa et al., 2018). For instance, previous studies have demonstrated that CTL from animals homozygous for BoLA-A18 and BoLA-A10 haplotypes predominantly recognize $Tp1_{214\text{-}224}$, and $Tp2_{49\text{-}59}$ & $Tp2_{50\text{-}59}$ epitopes, respectively (MacHugh et al., 2009; Connelley et al., 2016). Although the BoLA types of the animals sampled in this study were not determined, these three epitopes ($Tp1_{214\text{-}224}$, $Tp2_{49\text{-}59}$ and $Tp2_{50\text{-}59}$) important in immune response to *T. parva* infections had variants. All variants of $Tp2_{49\text{-}59}$ and $Tp2_{50\text{-}59}$ were only detected in buffalo-derived parasites from clinical cases of Corridor disease, and the $Tp1_{214\text{-}224}$ variant was common for both parasite groups. Considering the involvement of $Tp1_{214\text{-}224}$, $Tp2_{49\text{-}59}$ and $Tp2_{50\text{-}59}$ in *T. parva* immunity (MacHugh et al., 2009; Connelley et al., 2011, 2016), it could be concluded from the findings of the current study that Tp1 and Tp2 are promising candidates for development *T. parva* subunit vaccine for protection of *T. parva* infections in cattle in the two regions of Africa. Most of the variants of these immunodominant epitopes ($Tp1_{214\text{-}224}$, $Tp2_{49\text{-}59}$ and $Tp2_{50\text{-}59}$) in cattle- and buffalo-derived parasites from cattle do not involve amino acid residues that would lead to abrogation of TCR recognition.

In conclusion, the search for a subunit vaccine of *T. parva* alternative to the current live vaccine, with no risk of carrier state development and affordable to the resource poor small-holder farmers is inevitable. The comparative analysis of the diversity of antigenic epitopes on the immunodominant genes in *T. parva* parasites from cattle from southern and eastern Africa, suggest that Tp1 and Tp2 are promising candidates for a subunit vaccine. Considering the level of conservation of Tp4, Tp5, Tp7 and Tp8 detected in both parasite types, a further investigation in their immunogenic properties is recommended. Successful development of an efficacious and affordable subunit vaccine based on these two genes (Tp1 and Tp2), and the others, will not only improve the welfare of the resource-poor farmers who rely on cattle production for their livelihoods, but also reduce spillover of toxic residues to the environment arising from the excessive use of acaricides.

Funding

This work was supported by the National Research Foundation, South Africa; University of Pretoria, South Africa; and the National Research Fund, Kenya.

CRediT authorship contribution statement

Donald M. Lubembe: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Data curation, Writing - original draft, Writing - review & editing, Visualization. **David O. Odongo:** Conceptualization, Methodology, Resources, Writing - review & editing, Supervision, Project administration, Funding acquisition. **Fourie Joubert:** Formal analysis, Visualization. **Kgomotso P. Sibeko-Matjila:** Conceptualization,

Methodology, Resources, Writing - review & editing, Visualization, Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

None.

Acknowledgements

The authors wish to thank Nicola Collins for providing DNA samples from South Africa, Paul Gwakisa for providing DNA samples from Tanzania, Paul Ayieko for coordinating collection of blood samples from Kenya, Sarah Mwangi (FABI, University of Pretoria) for helping in the initial analysis of NGS data, and the DNA Ion Torrent Sequencing Facility of the University of Pretoria. The National Research Foundation (NRF) is thanked for funding the acquisition of the equipment based at the Central Analytical Facilities of Stellenbosch University (Grant number: UID 88063) that was used in this work.

References

Akoolo, L., Pelle, R., Saya, R., Awino, E., Nyanjui, J., Taracha, E.L., Kanyari, P., Mwangi, D.M., Graham, S.P., 2008. Evaluation of the recognition of *Theileria parva* vaccine candidate antigens by cytotoxic T lymphocytes from Zebu cattle. Vet. Immunol. Immunopathol. 121, 216–221.

Andrews, S., 2014. FastQC A Quality Control Tool for High Throughput Sequence Data.

Bastos, R.G., Franceschi, V., Tebaldi, G., Connelley, T., Morrison, W.I., Knowles, D.P., Donofrio, G., Fry, L.M., 2019. Molecular and antigenic properties of mammalian cell-expressed *Theileria parva* antigen Tp9. Front. Immunol. 10.

Bjorkman, P.J., Saper, M.A., Samraoui, B., Bennett, W.S., Strominger, J.L., Wiley, D.C., 1987. The foreign antigen binding site and T cell recognition regions of class I histocompatibility antigens. Nature 329, 512–518.

Connelley, T.K., MacHugh, N.D., Pelle, R., Weir, W., 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. J. Immunol. 187, 5910–5920.

Connelley, T.K., Li, X., MacHugh, N., Colau, D., Graham, S.P., van der Bruggen, P., Taracha, E.L., Gill, A., Morrison, W.I., 2016. CD8 T-cell responses against the immunodominant *Theileria parva* peptide Tp249-59 are composed of two distinct populations specific for overlapping 11-mer and 10-mer epitopes. Immunology 149, 172–185.

De Deken, R., Martin, V., Saido, A., Madder, M., Brandt, J., Geysen, D., 2007. An outbreak of East Coast Fever on the Comoros: a consequence of the import of immunised cattle from Tanzania? Vet. Parasitol. 143, 245–253.

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

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

Dolan, T.T., Linyonyi, A., McHardy, N., Bond, A.L., Clampitt, R.B., 1988. Chemotherapy of East Coast fever: parvaquone treatment of *Theileria parva* parva at intervals after infection. Res. Vet. Sci. 44, 15–20.

Frankild, S., de Boer, R.J., Lund, O., Nielsen, M., Kesmir, C., 2008. Amino acid similarity accounts for t cell cross-reactivity and for "Holes" in the t cell repertoire. PLoS One 3, e1831.

Gachohi, J., Skilton, R., Hansen, F., Ngumi, P., Kitala, P., 2012. Epidemiology of East Coast fever (*Theileria parva* infection) in Kenya: past, present and the future. Parasit. Vectors 5, 194.

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., Wilson, R.J., Sato, S., Ralph, S.A., Mann, D.J., Xiong, Z., Shallom, S.J., Weidman, J., Jiang, L., Lynn, J., Weaver, B., Shoaibi, A., Domingo, A.R., Wasawo, D., Crabtree, J., Wortman, J.R., Haas, B., Angiuoli, S.V., Creasy, T.H., Lu, C., Suh, B., Silva, J.C., Utterback, T.R., Feldblyum, T.V., Pertea, M., Allen, J., Nierman, W.C., Taracha, E.L., Salzberg, S.L., White, O.R., Fitzhugh, H.A., Morzaria, S., Venter, J.C., Fraser, C.M., Nene, V., 2005. Genome sequence of *Theileria parva*, a bovine pathogen that transforms lymphocytes. Science 309, 134–137.

Gornall, J., Betts, R., Burke, E., Clark, R., Camp, J., Willett, K., Wiltshire, A., 2010. Implications of climate change for agricultural productivity in the early twenty-first century. Philos. Trans. Biol. Sci. 365, 2973–2989.

Graham, S.P., Pelle, R., Honda, Y., Mwangi, D.M., Tonukari, N.J., Yamage, M., Glew, E. J., de Villiers, E.P., Shah, T., Bishop, R., Abuya, E., Awino, E., Gachanja, J., Luyai, A. E., Mbwika, F., Muthiani, A.M., Ndegwa, D.M., Njahira, M., Nyanjui, J.K., Onono, F. O., Osaso, J., Saya, R.M., Wildmann, C., Fraser, C.M., Maudlin, I., Gardner, M.J., Morzaria, S.P., Loosmore, S., Gilbert, S.C., Audonnet, J.C., van der Bruggen, P., Nene, V., Taracha, E.L., 2006. *Theileria parva* candidate vaccine antigens recognized by immune bovine cytotoxic T lymphocytes. Pro. Natl. Acad. Sci. U.S.A. 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., Mbwika, F., Muthiani, A.M., Muriuki, C., Nyanjui, J.K., Onono, F.O., Osaso, J., Riitho, V., Saya, R.M., Ellis, S.A., McKeever, D.J., MacHugh, N.D., Gilbert, S.C., Audonnet, J.C., Morrison, W.I., van der Bruggen, P., Taracha, E.L., 2008. Characterization of the fine specificity of bovine CD8 T-cell responses to defined antigens from the protozoan parasite Theileria parva. Infect. Immun. 76, 685–694.

Hemmink, J.D., Weir, W., MacHugh, N.D., Graham, S.P., Ekta, P., Paxton, E., Shiels, B., Toye, P.G., Morrison, W.I., Pelle, R., 2016. Limited genetic and antigenic diversity within parasite isolates used in a live vaccine against Theileria parva. Int. J. Parasitol. 46, 495–506.

Hemmink, J.D., Sitt, T., Pelle, R., Klerk-Lorist, L.Md., Shiels, B., Toye, P.G., Morrison, W. I., Weir, W., 2018. Ancient diversity and geographical sub-structuring in African buffalo *Theileria parva* populations revealed through metagenetic analysis of antigen-encoding loci. Int. J. Parasitol. 48, 287–296.

Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30, 772–780.

Kechin, A., Boyarskikh, U., Kel, A., Filipenko, M., 2017. CutPrimers: A New Tool for Accurate Cutting of Primers from Reads of Targeted Next Generation Sequencing, Vol. 24.

Kerario, I.I., Chenyambuga, S.W., Mwega, E.D., Rukambile, E., Simulundu, E., Simuunza, M.C., 2019. Diversity of two *Theileria parva* CD8+ antigens in cattle and buffalo-derived parasites in Tanzania. Ticks Tick. Dis. 10, 1003–1017.

Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 33, 1870–1874.

Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359.

Lawrence, J.A., De Vos, A.J., Irvin, A., 1994a. East Coast fever. Infectious Diseases of Livestock, Vol 1. Oxford University Press, London, UK.

Lawrence, J.A., De Vos, A.J., Irvin, A., 1994b. Zimbabwe theileriosis. Infectious Diseases of Livestock, Vol 1. Oxford University Press, London, UK.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 2009. The sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079.

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., Ellis, S.A., 2010. MHC class I bound to an immunodominant *Theileria parva* epitope demonstrates unconventional presentation to t cell receptors. PLoS Pathog. 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., Morrison, W.I., 2009. CD8 T-cell responses to *Theileria parva* are preferentially directed to a single dominant antigen: implications for parasite strain-specific immunity. Eur. J. Immunol. 39, 2459–2469.

Mbizeni, S., Potgieter, F.T., Troskie, C., Mans, B.J., Penzhorn, B.L., 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 Tick. Dis. 4, 227–234.

McHardy, N., Wekesa, L.S., Hudson, A.T., Randall, A.W., 1985. Antitheilerial activity of BW720C (buparvaquone): a comparison with parvaquone. Res. Vet. Sci. 39, 29–33.

McKeever, D.J., 2006. *Theileria parva* and the bovine CTL response: down but not out? Parasite Immunol. 28, 339–345.

McKeever, D.J., Taracha, E.L., Innes, E.L., MacHugh, N.D., Awino, E., Goddeeris, B.M., Morrison, W.I., 1994. Adoptive transfer of immunity to *Theileria parva* in the CD8+ fraction of responding efferent lymph. Proc. Natl. Acad. Sci. U.S.A. 91, 1959–1963.

Minjauw, B., McLeod, A., 2003. Tick-borne Diseases and Poverty. The Impact of Ticks and Tick-borne Diseases on Thelivelihoods of Small-scale and Marginal Livestock Owners in India and Eastern and Southern Africa. Research Report, DFID Animal Health Programme, Centre for Tropical Veterinary Medicine. University of Edinburgh, UK.

Morrison, W.I., Connelley, T., Hemmink, J.D., MacHugh, N.D., 2015. Understanding the basis of parasite strain-restricted immunity to Theileria parva. Annu. Rev. Anim. Biosci. 3, 397–418.

Mukhebi, A.W., Perry, B.D., Kruska, R., 1992. Estimated economics of theileriosis control in Africa. Prev. Vet. Med. 12, 73–85.

Mukolwe, L.D., Odongo, D.O., Byaruhanga, C., Snyman, L.P., Sibeko-Matjila, K.P., 2020. Analysis of p67 allelic sequences reveals a subtype of allele type 1 unique to buffalo-derived *Theileria parva* parasites from southern Africa. PLoS One 15, e0231434.

Musoke, A.J., Nene, V., 1990. Development of recombinant antigen vaccines for the control of theileriosis. Parassitologia 32, 73–85.

Musoke, A.J., Nantulya, V.M., Rurangirwa, F.R., Buscher, G., 1984. Evidence for a common protective antigenic determinant on sporozoites of several *Theileria parva* strains. Immunology 52, 231–238.

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

Mwangi, D.M., Honda, Y., Graham, S.P., Pelle, R., Taracha, E.L., Gachanja, J., Nyanjui, J. K., Bray, J., Palmer, G.H., Brown, W.C., Mwangi, W., 2011. Treatment of cattle with DNA-encoded Flt3L and GM-CSF prior to immunization with *Theileria parva* candidate vaccine antigens induces CD4 and CD8 T cell IFN-γ responses but not CTL responses. Vet. Immunol. Immunopathol. 140, 244–251.

Neitz, W.O., Canham, A.S., Kluge, E.B., 1955. Corridor disease: a fatal form of bovine Theileriosis encountered in Zululand. J. S. Afr. Vet. Assoc. 26, 79–88.

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

Nene, V., Kiara, H., Lacasta, A., Pelle, R., Svitek, N., Steinaa, L., 2016. The biology of *Theileria parva* and control of East Coast fever - current status and future trends. Ticks Tick. Dis. 7, 549–564.

Norval, R.A.I., Perry, B.D., Young, A.S., 1992. The Epidemiology of Theileriosis in Africa. Academic Press Limited, London.

Nyagwange, J., Tijhaar, E., Ternette, N., Mobegi, F., Tretina, K., Silva, J.C., Pelle, R., Nene, V., 2018. Characterization of the *Theileria parva* sporozoite proteome. Int. J. Parasitol. 48, 265–273.

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., Taracha, E.L.N., Morrison, W.I., Weir, W., Silva, J.C., Bishop, R.P., 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., Grace, D., 2009. The impacts of livestock diseases and their control on growth and development processes that are pro-poor. Philos. Trans. R. Soc. Lond. B, Biol. Sci. 364, 2643–2655.

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., Young, A.S., 1975. East coast fever: 3. Chemoprophylactic immunization of cattle using oxytetracycline and a combination of theilerial strains. Vet. Parasitol. 1, 51–60.

Randolph, T.F., Schelling, E., Grace, D., Nicholson, C.F., Leroy, J.L., Cole, D.C., Demment, M.W., Omore, A., Zinsstag, J., Ruel, M., 2007. Invited Review: Role of livestock in human nutrition and health for poverty reduction in developing countries. J. Anim. Sci. 85, 2788–2800.

Robbertse, L., Baron, S., van der Merwe, N.A., Madder, M., Stoltsz, W.H., Maritz-Olivier, C., 2016. Genetic diversity, acaricide resistance status and evolutionary potential of a Rhipicephalus microplus population from a disease-controlled cattle farming area in South Africa. Ticks Tick. Dis. 7, 595–603.

Robinson, J.T., Thorvaldsd´ottir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., Mesirov, J.P., 2011. Integrative genomics viewer. Nat. Biotechnol. 29, 24.

Salih, D.A., Pelle, R., Mwacharo, J.M., Njahira, M.N., Marcellino, W.L., Kiara, H., Malak, A.K., El Hussein, A.R., Bishop, R., Skilton, R.A., 2017. Genes encoding two *Theileria parva* antigens recognized by CD8+ T-cells exhibit sequence diversity in South Sudanese cattle populations but the majority of alleles are similar to the Muguga component of the live vaccine cocktail. PLoS One 12, e0171426.

Sibeko, K.P., Oosthuizen, M.C., Collins, N.E., Geysen, D., Rambritch, N.E., Latif, A.A., Groeneveld, H.T., Potgieter, F.T., Coetzer, J.A.W., 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. Vet. Parasitol. 155, 37–48.

Sitt, T., Poole, E.J., Ndambuki, G., Mwaura, S., Njoroge, T., Omondi, G.P., Mutinda, M., Mathenge, J., Prettejohn, G., Morrison, W.I., Toye, P., 2015. Exposure of vaccinated and naive cattle to natural challenge from buffalo-derived *Theileria parva.* Int. J. Parasitol. Parasites Wildl. 4, 244–251.

Sitt, T., Pelle, R., Chepkwony, M., Morrison, W.I., Toye, P., 2018. *Theileria parva* antigens recognized by CD8+ T cells show varying degrees of diversity in buffalo-derived infected cell lines. Parasitology 145, 1430–1439.

Steinaa, L., Saya, R., Awino, E., Toye, P., 2012. Cytotoxic T lymphocytes from cattle immunized against *Theileria parva* exhibit pronounced cross-reactivity among different strain-specific epitopes of the Tp1 antigen. Vet. Immunol. Immunopathol. 145, 571–581.

Steinaa, L., Svitek, N., Awino, E., Saya, R., Toye, P., 2018. Cytotoxic T lymphocytes from cattle sharing the same MHC class I haplotype and immunized with live *Theileria parva* sporozoites differ in antigenic specificity. BMC Res. Notes 11, 44.

Taracha, E.L., Goddeeris, B.M., Teale, A.J., Kemp, S.J., Morrison, W.I., 1995. Parasite strain specificity of bovine cytotoxic T cell responses to *Theileria parva* is determined primarily by immunodominance. J. Immunol. 155, 4854–4860.

Vudriko, P., Okwee-Acai, J., Tayebwa, D.S., Byaruhanga, J., Kakooza, S., Wampande, E., Omara, R., Muhindo, J.B., Tweyongyere, R., Owiny, D.O., Hatta, T., Tsuji, N., Umemiya-Shirafuji, R., Xuan, X., Kanameda, M., Fujisaki, K., Suzuki, H., 2016. Emergence of multi-acaricide resistant Rhipicephalus ticks and its implication on chemical tick control in Uganda. Parasit. Vectors 9, 4-4.

Young, A.S., Groocock, C.M., Kariuki, D.P., 1988. Integrated control of ticks and tick-borne diseases of cattle in Africa. Parasitology 96 (Pt 2), 403–432.