Ticks and Tick-borne Diseases 10 (2019) 1003–1017

Contents lists available at ScienceDirect

Ticks and Tick-borne Diseases

journal homepage: www.elsevier.com/locate/ttbdis

Original article

# Diversity of two *Theileria parva* CD8+ antigens in cattle and buffalo-derived parasites in Tanzania

Isack I. Kerario[a,b,*], Sebastian W. Chenyambuga[a], Elisa D. Mwega[c], Elpidius Rukambile[d], Edgar Simulundu[b], Martin C. Simuunza[b]

[a] Department of Animal, Aquaculture and Range Sciences, College of Agriculture, Sokoine University of Agriculture P.O. Box 3004, Morogoro, Tanzania
[b] Department of Disease Control, School of Veterinary Medicine, University of Zambia, P.O. Box 32379, Lusaka, Zambia
[c] College of Veterinary Medicine and Biomedical Sciences, Sokoine University of Agriculture (SUA), P.O. Box 3019, Morogoro, Tanzania
[d] Tanzania Veterinary Laboratory Agency, P.O. Box 9254, Dar es Salaam, Tanzania

ARTICLE INFO

*Keywords:*
*Theileria parva*
East Coast fever
Muguga vaccine cocktail
Nucleotide polymorphism
Tajima's D, Fu'sFs
Tanzania

ABSTRACT

*Theileria parva* is a tick-transmitted protozoan parasite that causes a disease called East Coast fever (ECF) in cattle. This important tick borne-disease (TBD) causes significant economic losses in cattle in many sub-Saharan countries, including Tanzania. Cattle immunization using Muguga cocktail has been recommended as an effective method for controlling ECF in pastoral farming systems in Tanzania. However, immunity provided through immunization is partially strain-specific. Therefore, the control of ECF in Tanzania is still a challenge due to inadequate epidemiological information. This study was conducted to assess genetic diversity of Tp1 and Tp2 genes from *T. parva* isolates that are recognized by CD8 + T-cells in cattle and buffalo. The Tp1 and Tp2 genes are currently under evaluation as candidates for inclusion in a subunit vaccine. A total of 130 blood samples collected from cattle which do not interact with buffalo (98), cattle co-grazing with buffalo (19) and buffalo (13) in Mara, Mbeya, Morogoro, Tanga, and Coast regions in Tanzania were used in this study. Genomic DNA was extracted from the blood samples, Tp1 and Tp2 genes were amplified using nested PCR and the PCR products were purified and sequenced. The partial sequencing of the Tp1 and Tp2 genes from *T. parva* isolates exhibited polymorphisms in both loci, including the epitope-containing regions. Results for sequence analysis showed that the overall nucleotide polymorphism ($\pi$) was 0.7% and 13.5% for Tp1 and Tp2, respectively. The Tajima's D and Fu's Fs test showed a negative value for both Tp1 and Tp2 genes, indicating deviations from neutrality due to a recent population expansion. The study further revealed a low to high level of genetic differentiations between populations and high genetic variability within populations. The study also revealed that most samples from the seven populations possessed several epitopes in antigens that were identical to those in the *T. parva* Muguga reference stock, which is the main component of the widely used live vaccine cocktail. Therefore, different strategic planning and cost-effective control measures should be implemented in order to reduce losses caused by ECF in the study areas.

## 1. Introduction

Tick-borne diseases (TBDs) are a serious constraint to cattle production in Tanzania and other sub-Saharan countries (Mukhebi et al., 1992; Makala et al., 2003; Kivaria, 2006). Among the TBDs, East Coast fever (ECF) is the most important disease known to cause significant economic losses in cattle in sub-Saharan Africa (Mukhebi et al., 1992; Kivaria, 2006). Economic losses associated with ECF include high mortality, lowered milk production, and the high cost of controlling the disease (Jongejan and Uilenberg, 2004). In Tanzania East Coast fever has been reported to kill approximately one million cattle annually (Kivaria, 2007). In unvaccinated Zebu calves raised under pastoral conditions, ECF is responsible for annual mortality of 40–80% (Homewood et al., 2006). East Coast fever is caused by the haemoprotozoan parasite, *Theileria parva* that is transmitted by a brown ear tick, *Rhipicephalus appendiculatus* (Neitz et al., 1955; reviewed in Nene et al., 2016). *Theileria parva* parasitize T and B cells of cattle and some wild animals including Cape buffalo (*Syncerus caffer*) causing classical ECF to cattle (Norval et al., 1991). *Syncerus caffer* is considered to be an important reservoir of *T. parva* parasite (Pienaar et al., 2011).

* Corresponding author at: Department of Animal, Aquaculture and Range Sciences, College of Agriculture, Sokoine University of Agriculture P.O. Box 3004, Morogoro, Tanzania.
*E-mail address:* isackkerario@sua.ac.tz (I.I. Kerario).

https://doi.org/10.1016/j.ttbdis.2019.05.007
Received 31 December 2018; Received in revised form 29 April 2019; Accepted 21 May 2019
Available online 22 May 2019
1877-959X/ 

*Theileria parva* alternates between cattle and the tick in its life cycle and has two main developmental stages (Mehlhorn and Schein, 1984). The sexual stage of development occurs in the gut of the tick after picking the piroplasms in the red blood cells of an infected cattle (Lawrence et al., 2005). The piroplasms differentiate into gametes and then the gametes undergo syngamy to form diploid zygotes which invade epithelial cells of the tick gut and develop into motile kinete and then migrate to the tick salivary glands. Subsequently, the kinetes transform into mature sporozoites, which are ready for transmission from the tick salivary glands to a host. The tick transmits the parasite to a new host when feeding as a nymph/adult and releasing sporozoites through saliva, which then enters the animal's body (Mehlhorn and Schein, 1984). Infected ticks inoculate infective sporozoites three days after attaching to the host and within a short period, sporozoites invade host lymphocytes. Lymphocytes of infected animal transform into lymphoblast which is a stimulation of the formation of schizonts that undergo division in synchrony with the host cell as mitosis occur (reviewed in Norval et al., 1992). Invasion and entry of the parasite into the cell is associated with the transformation of the infected cell to a state of uncontrolled proliferation (reviewed in Norval et al., 1992). Schizonts in infected cells undergo further differentiation to merozoites, and as the cell ruptures they invade erythrocytes and develop into piroplasms, the infective stage for ticks (reviewed in Norval et al., 1992).

Prevailing methods for the control of ECF include vector (tick) control (George et al., 2004; Nonga et al., 2012), use of tick-resistant breeds (Shyma et al., 2013; reviewed in Laisser et al., 2017), treating infected animals by therapeutic drugs (Musoke et al., 2004) and immunization (vaccination) through infection and treatment method (ITM) (Radley et al., 1975). Acaricide application is the most common method used to control ticks in the tropics (George et al., 2004; Nonga et al., 2012). However, this method has several drawbacks, such as development of resistance by ticks, food safety concerns and environmental contamination due to toxic residues obtained from these chemicals (George et al., 2004). Vaccination of cattle using live *T. parva* sporozoites and simultaneous treatment with long-acting tetracycline (ITM) leads to long-term immunity against the homologous parasite stocks (Radley et al., 1975). However, protection against challenge with heterologous parasite genotypes may not be sufficient (Di Giulio et al., 2009). Infection and treatment method using Muguga cocktail vaccine, which comprises parasites from three stocks namely Kiambu 5, Muguga and Serengeti-transformed, has been used to immunize cattle in Malawi, Tanzania, Uganda and Zambia (Musisi et al., 1992). In Tanzania, cattle immunization against ECF by ITM using the Muguga cocktail vaccine has been deployed successfully in pastoral farming systems in Eastern, Northern and Central zones (Di Giulio et al., 2009). The advantage of ITM in controlling ECF is that it increases the survival rate of calves and reduces the mortality rate down to 2% annually in the pastoral herds (Lynen et al., 2006). Despite the success of this method, there is a fear among veterinary authorities that the introduction of new parasite genetic material in a particular area through vaccination might generate novel and more virulent genotypes through sexual recombination (Morzaria and Williamson, 1999; Oura et al., 2007; Gibson et al., 2015). Previous reports have shown that the indiscriminate deployment of the live sporozoite vaccine has raised numerous concerns, including the possibility of introducing vaccine strains to unvaccinated cattle (Geysen et al., 1999; De Deken et al., 2007; Oura et al., 2007).

Recent identification of several *T. parva* antigens and epitopes recognized by CD8 + T-cells from *T. parva* immune cattle gives an opportunity to address the nature and selective pressure driving antigenic diversity relevant to immune protection (Graham et al., 2006, 2008). *Theileria parva* antigens labeled Tp1 to Tp10 have been identified (Akoolo et al., 2008; Graham et al., 2006, 2007, 2008; reviewed in Morrison et al., 2015). A comprehensive study by MacHugh et al. (2009) on immune response using two of these antigen genes, have demonstrated that Tp1 and Tp2 antigen genes are highly dominant targets of the CD8 + T-cell response in cattle expressing the A18 and A10 class I major histocompatibility complex (MHC) haplotypes, respectively. Previous studies have examined genetic diversity in cytotoxic T lymphocytes (CTL) antigens. Among the CTL determinants in *T. parva*, higher genetic diversity has been shown in Tp1 and Tp2 variants than in other antigens (reviewed in McKeever, 2009). A study conducted by Pelle et al. (2011) found that Tp1 and Tp2 gene sequences exhibit high ratio between the number of nonsynonymous substitutions per nonsynonymous site and number of synonymous substitutions per synonymous site. Furthermore, the study revealed a higher level of sequence diversity in *T. parva* isolates obtained from buffalo (*Syncerus caffer*) than in isolates derived from cattle (Pelle et al., 2011). Recently, Salih et al. (2017), reported that Tp1 and Tp2 genes exhibit sequence diversity in South Sudanese cattle populations, but most alleles were found to be similar to the Muguga component of live vaccine cocktail. In Tanzania, the use of Muguga cocktail has been promoted extensively (Di Giulio et al., 2009). Therefore, understanding the extent of genetic exchange that exists within natural populations and the levels of genetic variability as well as the structure of *T. parva* parasite populations may help in determining whether the Muguga vaccine would offer protection against the parasite stocks in the study areas. The current study reports the genetic diversity of genes encoding Tp1 and Tp2 antigens in *T. parva* isolates derived from cattle and buffalo in Tanzania. Findings from this work provide guidance regarding the deployment of ECF vaccine in the study areas.

## 2. Materials and methods

### 2.1. Study area

The study was carried out in five regions of Tanzania, namely, Mara in the Lake zone, Mbeya in the southern highlands (Kerario et al., 2017a, 2017b), Morogoro, Tanga, and Coast regions in the eastern zone.

**Table 1**
Populations, location of sampling areas, hosts and year of sample collection.

| Population | Location | Collected samples | Samples used | Host | Year collected | Reference | Live vaccine deployment |
|---|---|---|---|---|---|---|---|
| Mara | Mara | 216 | 27 | Cattle | 2015 | Kerario et al. (2017a) | No |
| Mbeya | Mbeya | 216 | 16 | Cattle | 2015 | Kerario et al. (2017b) | Yes |
| Morogoro | Morogoro | 39 | 23 | Cattle | 2012 | Mwega et al. (2014a) | No |
| Tanga | Tanga | 50 | 20 | Cattle | 2014 | Rukambile et al. (2016) | No |
| Pwani (Coast) | Pwani (Coast) | 50 | 12 | Cattle | 2014 | Rukambile et al. (2016) | Yes |
| Buffalo-associated | Ngorongoro Conservation Area | 50 | 13 | Cattle | 2012 | Mwega et al. (2014b) | No |
| | Mikumi National Park | 7 | 4 | Cattle | 2014 | Rukambile et al. (2016) | No |
| | Ruaha National park | 10 | 2 | Cattle | 2014 | Rukambile et al. (2016) | No |
| Buffalo-derived | Mikumi National Park | 25 | 11 | Buffalo | 2014 | Rukambile et al. (2016) | No |
| | Mkomazi National Park | 20 | 2 | Buffalo | 2014 | Rukambile et al. (2016) | No |

### 2.2. Blood sample collection

A total of 130 blood samples out of 683 collected in three different phases were used in this study. Table 1 shows populations, location of sampling areas, hosts, number of samples collected, number of samples used in the current study and year of sample collection. Briefly, in the first phase, the samples collected from Morogoro region originated from cattle which had no history of co-grazing with buffalo, while those collected from Ngorongoro conservation area were derived from cattle with a history of grazing together with buffalo (Mwega et al., 2014a, 2014b). In the second phase samples from Tanga and Coast regions were obtained from cattle with no history of co-grazing with buffalo. Seven samples from Mikumi National Park and ten from Ruaha National Park originated from cattle with a history of co-grazing with buffalo (Table 1). On the other hand, a total of 45 samples, 25 from Mikumi National Park and 20 from Mkomazi National Park were sampled from buffalo (Rukambile et al., 2016). In the third phase 432 samples, 216 samples from Mara and 216 from Mbeya were collected in 2015 and all of them originated from cattle which had no history of co-grazing with buffalo (Kerario et al., 2017b). The blood samples were collected from the jugular vein of each animal using labeled vacutainer tubes containing Ethylenediaminetetraacetic acid (EDTA) (Becton Dickson Vacutainer, Oxford, UK) and kept in a cooler box while in the field. Then they were transported from the field to Sokoine University of Agriculture and stored at -20 °C pending further analysis.

### 2.3. DNA extraction and detection of Theileria parva infection

Genomic DNA from each blood sample was extracted using Quick-gDNA™ Blood MiniPrep kit catalog No. D3073 (Zymo Research Cooperation, USA), according to the manufacturer's instructions and then eluted in 100 µl elution buffer. Afterward, the concentration of purified genomic DNA was determined using a spectrophotometer (NanoDrop™ One Spectrophotometer, Thermo Fisher Scientific). Detection of *T. parva* infection was performed following the procedures described by Kerario et al. (2017b).

### 2.4. Amplification and sequencing of Tp1 and Tp2 genes

Genomic DNA of 130 positive samples for *T. parva* were sent to Microsynth AG, Switzerland for amplification and sequencing using the protocol described below. Nested PCR was used to amplify Tp1 and Tp2 genes from the 130 samples that were found positive for *T. parva* (Table 1). The outer primers for Tp1 and Tp2 genes used in this study have previously been described by Pelle et al. (2011), while the inner primers were described by Salih et al. (2017). The PCR amplifications for both TP1 and Tp2 were carried out in a total volume of 25 µl, consisting of 7.5 µl KAPA2G Robust Hot start PCR kit, cat. no. KK5525, KAPA Biosystems Inc, Switzerland, 12.5 µl nuclease free water, 1.25 µl each forward and reverse primers (3 µM) and 2.5 µl of genomic DNA (20 ng) template added into the lyophilized pellet and briefly spinning down to dissolve the pellet. For the second round PCR, 1.0 µl of the primary PCR products was used as template and other reagents (primers and nuclease free water) were the same as in the primary PCR. The PCR conditions for both TP1 and Tp2 were performed as previously described by Salih et al. (2017), with few modifications. The initial denaturation was performed at 95 °C for 180 s, followed by 35 cycles of denaturation at 95 °C for 20 s, annealing at 50 °C for 20 s and extension at 72 °C for 100 s, and final extension at 72 °C for 45 s and subsequent cooling at 4 °C (Salih et al., 2017). The cycling conditions for the secondary PCR was as described in the primary amplification. The expected sizes of the primary PCR products were 432 bp for Tp1 and 525 bp for Tp2, while the expected sizes of the nested PCR (nPCR) products were 405 bp for Tp1 and 504 bp for Tp2. The nPCR products were checked for successful amplification using Qiaxcel platform. All amplified PCR products were purified by centrifugation using Wizard Genomic DNA purification kit, #TB308, USA (www.Promega.com) as per manufacturer's instruction and then Sanger-sequenced following the protocol described by Microsynth AG (https://www.microsynth.ch/standard-services.html).

### 2.5. Data analysis

Sequences were assembled and edited using Sequencher® version 5.4.6 DNA sequence analysis software, (Gene Codes Corporation, Ann Arbor, MI USA) (http://www.genecodes.com). The same program was also used to translate open reading frames present within the sequences generated from the amplified fragments into amino acid sequences and both DNA and amino acid sequences were converted into FASTA format. Multiple alignment of TP1 and TP2 sequences including those of Muguga cocktail isolate (XM 760490) and *Theileria annulata* (TA 19865) were performed by ClustalW (http://www.genome.jp/tools-bin/clustalw) (Thompson et al., 1994). The rooted phylogenetic tree for each locus was constructed by neighbor-joining method using MEGA 7 (http://megasoftware.net/) (Kumar et al., 2016). The evolutionary distance model was employed to construct the phylogenetic trees. The same program was used to determine the genetic distances (expressed as the number of nucleotide differences per 100 bases or per 100 amino acids together with the length polymorphism). In order to determine similarity among populations the genetic distances were used to perform Principal Component Analysis (PCA) using the Excel plug-in GenAIEx6.5 (http://biology.anu.Edu.au/GenAlEx) (Peakall and Smouse, 2012). Excel plug-in GenAIEx6.5 was also used for the analysis of molecular variance (AMOVA) in order to investigate the distribution of genetic variation among allelic sequences and to determine the level of population differentiation. DNA Sequence polymorphism (DnaSP6.10.04) (Rozas et al., 2017) was used to calculate DNA polymorphism and divergence as well as population size change. Genetic diversity and neutrality indices for *T. parva* population from nucleotide sequence of Tp1 and Tp2 genes were determined using the same program. Neutrality Indices were calculated by Tajima's D and Fu'sFs. Furthermore, the alleles were identified after generating haplotype data file using DnaSP6.10.04.

## 3. Results

### 3.1. Genetic diversity of T. Parva

To assess the level of genetic diversity of *T. parva* field isolates derived from cattle and buffalo, sequence polymorphism analysis of Tp1 and Tp2 genes (the CD8 + T cell target antigens) was conducted. The Tp1 region (405 bp) that was sequenced in this study is situated in the center of the gene and extends from nucleotides 537 to 941 of the reference Tp1 sequence (accession number XP_792973). In particular, the region encodes 134 amino acids containing 24.7% of the 543-residue Tp1 protein (Pelle et al., 2011). For the Tp1 gene, a total of 124 samples (accession numbers MH576702 to MH576825) were sequenced. Six samples out of 130 were not sequenced because of poor PCR product yield. We identified thirty alleles in this locus from the 124 samples that were sequenced (Fig. 1). Allele one, which is found in the *T. parva* Muguga reference sequence was observed in 59 (48%) out of the 124 samples. Allele four which was the second most frequent was observed in 24 samples, representing 19% of the total sequences analyzed. Allele two, six, 12, 13, and 18 were observed in seven, four, two, two and three samples, respectively. The remaining alleles were observed in one sample each. Of the 30 alleles observed in the Tp1 gene, two (alleles one and four) have been reported previously by Pelle et al. (2011) (accession number JF451936.1 and JF451973.1 for alleles one and four, respectively), three (alleles two, 11 and 15) by Salih et al. (2017), (accession numbers KJ566597.1, KJ566596.1 and KJ566599.1 for alleles two, 11 and 15, respectively), while two (alleles six and 17) were previously reported by Amzat et al. (2018-Unpublished) (accession


**Fig. 1.** Multiple sequence alignment showing thirty (30) alleles of the Tp1 gene identified in this study. The primer regions are boxed.

number MF449291.1 for allele six and MF449291.1 for allele 17). The remaining 23 alleles have not been identified before and are reported here for the first time (Fig. 1).

On the other hand, we sequenced the partial length of the Tp2 gene which encodes 168 amino acids. We were only able to obtain 82 sequences (accession numbers MH576826 to MH576907) out of 130 samples due to poor PCR product yield. A total of 35 alleles were obtained from the 82 sequences (Fig. 2). Allele five which is found in *T. parva* Muguga reference stock was the most frequent and was observed in 23 samples, representing 28% of the total sequences analyzed. Allele two, which was the second most frequent was observed in 18 samples (22%) out of 82 sequences analyzed. Alleles three, six, and 18 were observed in two samples each, while allele 13 was observed in six samples. The remaining alleles were observed in one sample each. Out of the 35 alleles identified in Tp2 gene, two alleles (allele five and allele 13) have been reported previously by Pelle et al. (2011), (accession numbers JF451856.1 (allele five), JF451880.1 (allele 13). The remaining 33 alleles identified in this study have not been previously


**Fig. 1.** (*continued*)


**Fig. 1.** (*continued*)

identified and are reported here for the first time.

### 3.2. Tp1 and Tp2 T. Parva epitope variants identified in this study

Comparison of the protein sequences encoded by the 30 alleles of the Tp1 gene showed five distinct antigen variants (Fig. 3A), resulting from amino acid changes at three residues. The comparison of the *T. parva* Muguga CD8 + T cell epitope sequence (VGYPKVEEML) located within the sequenced region of Tp1 revealed five epitope variants ending with ML, II, IL, MI and MV. This was a result of nucleotide substitutions at positions 133 and 134. Most samples analyzed (71 out of 124 (57%)), exhibited the epitope –ML, found in the *T. parva* Muguga reference isolate (Fig. 3A). Of the 71-ML variants, 16, 6, 17, 4, 5, 17 and 6 originated from Mara, Mbeya, Morogoro, Tanga, Coast, buffalo-associated and buffalo-derived isolates, respectively. The next most common variant –II, was observed in 41 samples obtained from Mara


**Fig. 2.** Multiple sequence alignment showing thirty-five (35) alleles of the Tp2 gene identified in this study. The primer regions are boxed.


**Fig. 2.** (*continued*)

(10), Mbeya (8), Tanga (13), Coast (7), buffalo-associated (1) and buffalo-derived (2), representing 33% of the total samples. Other protein variants identified in this study were IL, MI and MV and these were observed in four (3%), seven (6%) and one (1%) samples, respectively. The MI variant was observed in isolates obtained from Mbeya (2), Tanga (3), buffalo-associated (1) and buffalo-derived (10). All the four IL variants were observed in isolates collected from buffalo. The MV variant was observed in one isolate collected in Mara region. Of the five antigen variants discovered in this study, var-1(ML), var-2 (II)and var-3 (IL) have previously been reported by Pelle et al. (2011); antigen variant var-4 (MI) has been reported by Amzat et al. (2018 unpublished) (accession number AVM 41,545), while antigen variant Var-5 (MV) has not been reported before and is reported here for the first time. Alignment of sequences from the 124 samples of the Tp1 gene showed

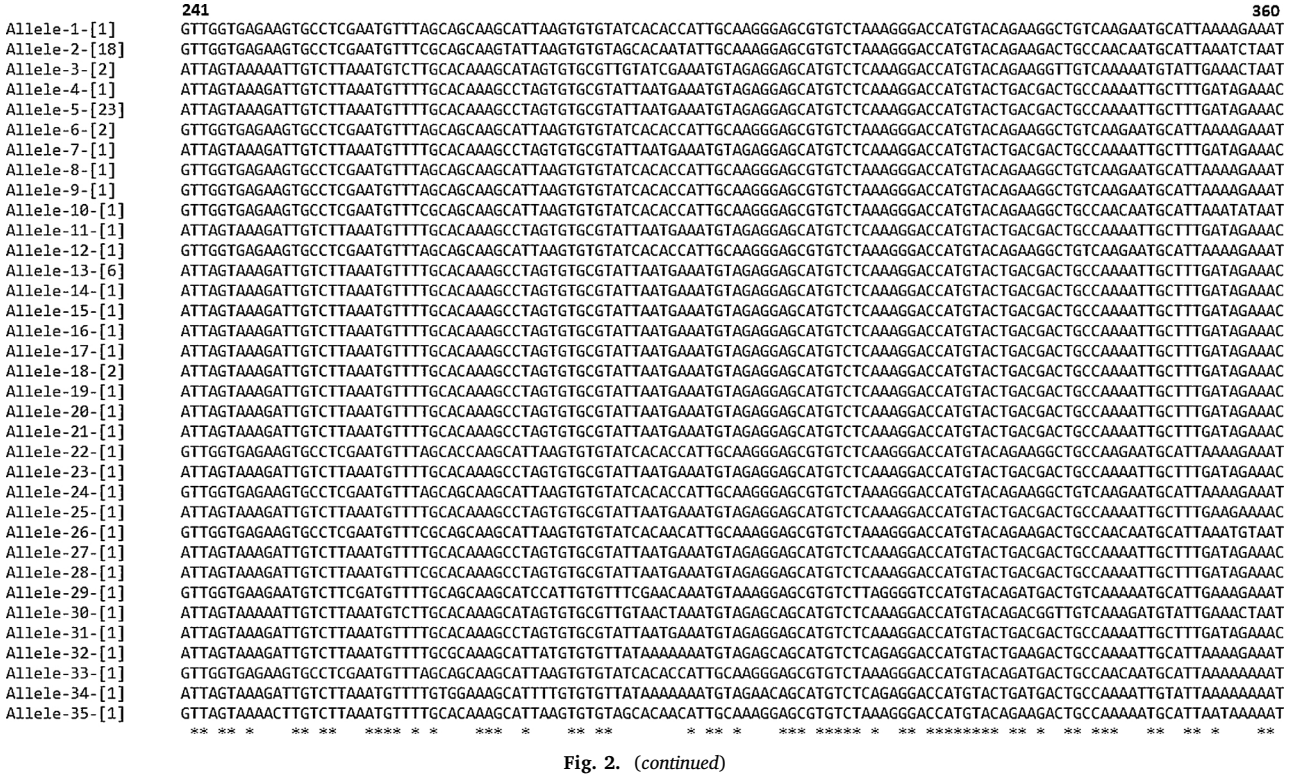


**Fig. 2.** (*continued*)

**Fig. 2.** (*continued*)

that 94.3% of the amino acid residues were conserved (Fig. 3A).

With regards to Tp2 locus, the 82 samples sequenced revealed 12 distinct antigen variants (Fig. 3B). The sequences of Tp2 gene revealed different number of variants in the six epitopes ranging from 6 to 12 per epitope. This gene encodes 168 amino acids and six distinct CD8 + T cell epitopes namely SHEELKKLGML-Tp2$_{227-237}$, DGFDRDALF-Tp2$_{240-248}$, KSSHGMGKVGK-Tp2$_{49-59}$, FAQSLVCVL-Tp2$_{96-104}$, QSLVCV-LMK-Tp2$_{98-106}$ and KTSIPNPCKW-Tp2$_{138-147}$, represented by MHC class I elements that restrict epitopes of Tp2 antigen (Graham et al., 2007, 2008; Akoolo et al., 2008). Epitope number one exhibited higher number of variants (12 variants) compared to the other epitopes. Epitope number five was the second highest encompassing nine variants, followed by epitope two and four with eight variants each. Epitope six comprised seven variants while epitope three had the least number of variants compared to the other epitopes. Of the 12 Tp2 antigen variants found in this study, var-1 and var-3 have previously been reported by Pelle et al. (2011) whereas the remaining 10 antigen variants have not been reported before and are reported here for the first time. Alignment of sequences from the 82 samples of the Tp2 gene showed that 55.2% of the amino acid residues were conserved (Fig. 3B).

### 3.3. Diversity and neutrality indices of T. Parva in Tp1 and Tp2 genes

Diversity and neutrality indices of *T. parva* in Tp1 and Tp2 genes are depicted in Tables 2A & 2B . Regarding the Tp1 locus, the number of polymorphic sites, mutations and number of haplotypes were relatively higher in *T. parva* isolates that originated from buffalo compared to the isolates derived from cattle (Table 2A). Among the regions, isolates from Mbeya exhibited relatively higher number of polymorphic sites, mutations and number of haplotypes, followed by those from the Coast region. There was only one haplotype observed in isolates originated from Morogoro. Mean haplotype diversity in all seven populations was

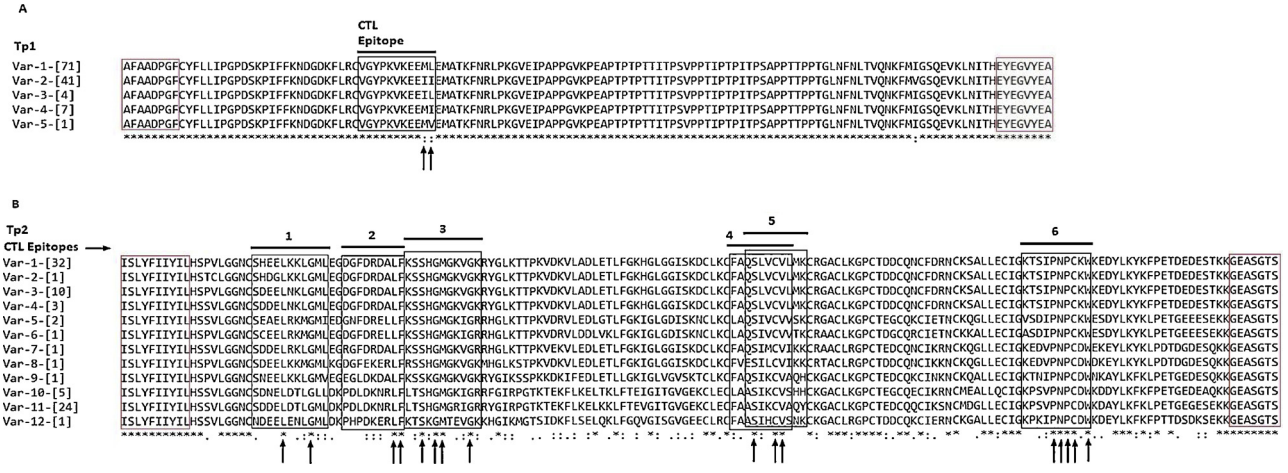

**Fig. 3.** Multiple amino acid sequence alignment of Tp1 and Tp2 antigen variants found in cattle and buffalo. (A) Multiple sequence alignment of the five Tp1 antigen variants. (B) Multiple sequence alignment of the 12 Tp2 antigen variants. The CD8 + T cell target epitopes are boxed in black while the flanked PCR primer regions not included in the estimations of % residues conserved are boxed in red. The polymorphic residues in the Tp1 epitope and the conserved amino acid residues in Tp2 epitopes are indicated by arrows.

**Table 2A**
Diversity and neutrality indices of *T. parva* populations obtained from cattle and buffalo.

| Locus | Population | No. Sequence | Variable sites | Mutation | No. Haplotype | K | HD ± SD | π ± SD | D | Fu'sFs |
|---|---|---|---|---|---|---|---|---|---|---|
| Tp1 | Cattle-derived | 92 | 10 | 12 | 15 | 2.167 | 0.701 ± **0.039** | 0.005 ± **0.000** | −0.215 | −4.295 |
| | Buffalo-associated | 19 | 9 | 9 | 5 | 1.743 | 0.386 ± 0.139 | 0.004 ± 0.002 | −1.128 | −0.236 |
| | Buffalo-derived | 13 | 21 | 21 | 12 | 5.359 | 0.987 ± 0.035 | 0.013 ± .003 | −0.959 | −6.873 |
| | **Total** | **124** | **25** | **28** | **30** | **2.970** | **0.735 ± 0.037** | **0.007 ± 0.001** | **−1.17** | **−17.33** |
| Tp2 | Cattle-derived | 71 | 177 | 208 | 27 | 66.540 | 0.855 ± 0.029 | 0.132 ± 0.007 | 1.896 | 19.250 |
| | Buffalo-associated | 7 | 161 | 175 | 5 | 53.143 | 0.857 ± 0.137 | 0.105 ± 0.047 | −1.802 | −1.895 |
| | Buffalo-derived | 4 | 144 | 157 | 4 | 84.167 | 1.000 ± 0.177 | 0.167 ± 0.039 | −0.334 | −0.249 |
| | **Total** | **82** | **226** | **298** | **35** | **68.286** | **0.872 ± 0.027** | **0.135 ± 0.008** | **−2.37*** | **−1.147** |

Statistical significance: * = significant at P < 0.05; Without an asterisk = Not significant at P < 0.05.
HD = Haplotype diversity; SD = Standard deviation K = Number of pairwise nucleotide differences; π = Nucleotide diversity; D = Tajima's D test statistics. Cattle derived = isolates derived from cattle without a history of co-grazing with buffalo; Buffalo-associated = isolates collected from cattle co-grazing with buffalo; Buffalo-derived = isolates obtained from buffalo.

0.735 ± 0.037 and was relatively higher in isolates obtained from buffalo compared to those originated from cattle (Table 2B). When comparison was made among the five regions, isolates from Coast (0.894 ± 0.054) showed relatively higher haplotype diversity compared to those from other regions (Table 2B). Average number of nucleotide differences, K was found to be higher in buffalo compared to other populations. When all populations were treated as one, K was found to be 2.970. The overall nucleotide polymorphism in the study area was π = 0.007 ± 0.001 (0.7%), with isolates from buffalo exhibiting slightly higher nucleotide polymorphism π = 0.013 ± 0.003 (1.3%) compared to the other populations (Table 2B).

Neutrality indices were calculated by Tajima's D and Fu'sFs. In all populations, with an exception of Morogoro region, the D and Fu'sFs values were negative and not significant (Mara, D = −0.702, Fu'sFs = -0.550; Mbeya, D = −1.808, Fu'sFs = −1.834; Coast, D = −1.643, Fu'sFs = −0.865; Tanga, D = −0.086, Fu'sFs = −0.034; buffalo-associated, D = −1.127, Fu'sFs = −0.236; buffalo-derived, D = −0.959, Fu'sFs = −6.875). When isolates from cattle that had no history of co-grazing with buffalo were treated as one, the D and Fu'sFs values were negative and not significant (Table 2A). When the data set from all populations were combined together, D and Fu'sFs values were negative and not significant, consistent with a population expansion.

Regarding the Tp2 gene, the number of polymorphic sites, mutations, and haplotypes were higher in samples collected from Mara than in other populations (Table 2B). The average number of nucleotide differences, K was higher in buffalo, followed by Coast and Mara regions. Isolate from Morogoro exhibited the lowest number of K compared to the other populations (Table 2B). When all samples were treated as a single population, K was found to be 68.286. The overall haplotype diversity in all populations was 0.872 ± 0.027, with samples from buffalo and Coast region exhibiting relatively higher haplotype diversity compared to other populations. Morogoro isolates demonstrated relatively lower haplotype diversity compared to other populations (Table 2B). The overall nucleotide diversity was π = 0.135 ± 0.008 (13.5%). The comparison among populations revealed that isolates from Mara exhibited slightly higher nucleotide polymorphism π = 0.151 ± 0.015 (15.1%) compared to other populations. Neutrality indices calculated by Tajima's D and Fu'sFs were positive and significant in Mara while in Mbeya, Morogoro, Coast and Tanga regions the values for D and Fu'sFs were positive but not significant. Moreover, the values for D and Fu'sFs for isolates from buffalo-associated and buffalo-derived were negative and not significant. When all populations were analyzed together, the value for D was negative and significant while the value for Fu'sFs was negative but not significant (Table 2A and 2B. Generally, higher nucleotide polymorphism was observed in Tp2 gene (13.5%) compared to Tp1 gene (0.7%).

### 3.4. Genetic differentiation and gene flow in Tp1 and Tp2 loci

The levels of genetic differentiation and gene flow among the seven *T. parva* populations are presented in Table 3 (A and B). Genetic differentiation statistics based on pairwise genetic distance ($F_{ST}$) and the number of migrants (*Nm*) revealed evidence of differentiation between most of the *T. parva* populations in both Tp1 and Tp2 genes. With regard to the Tp1 locus, the pairwise genetic distance ($F_{ST}$) in all seven populations ranged from negative and not significant value with

**Table 2B**
Diversity and neutrality indices of *T. parva* populations obtained from different regions in Tanzania.

| Locus | Population | No. Sequence | Variable sites | Mutation | No. Haplotype | K | HD ± SD | π ± SD | D | Fu'sFs |
|---|---|---|---|---|---|---|---|---|---|---|
| Tp1 | Mara | 27 | 4 | 6 | 5 | 1.917 | 0.641 ± 0.079 | 0.005 ± 0.001 | −0.702 | −0.55 |
| | Mbeya | 16 | 10 | 10 | 8 | 2.842 | 0.842 ± 0.075 | 0.007 ± 0.001 | −1.808 | −1.834 |
| | Morogoro | 17 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| | Coast | 12 | 5 | 5 | 6 | 2.394 | 0.894 ± 0.054 | 0.006 ± 0.001 | −1.644 | −0.865 |
| | Tanga | 20 | 4 | 4 | 5 | 1.526 | 0.568 ± 0.119 | 0.004 ± 0.001 | −0.086 | −0.034 |
| | Buffalo-associated | 19 | 9 | 9 | 5 | 1.743 | 0.386 ± 0.139 | 0.004 ± 0.002 | −1.128 | −0.236 |
| | Buffalo-derived | 13 | 21 | 21 | 12 | 5.359 | 0.987 ± 0.035 | 0.013 ± .003 | −0.959 | −6.873 |
| | **Total** | **124** | **25** | **28** | **30** | **2.970** | **0.735 ± 0.037** | **0.007 ± 0.001** | **−1.17** | **−17.33** |
| Tp2 | Mara | 17 | 165 | 189 | 13 | 76.068 | 0.963 ± 0.033 | 0.151 ± 0.015 | 1.437* | 1.778** |
| | Mbeya | 12 | 129 | 132 | 7 | 67.509 | 0.833 ± 0.100 | 0.134 ± 0.018 | 0.777 | 1.378 |
| | Morogoro | 23 | 3 | 3 | 3 | 0.798 | 0.423 ± 0.104 | 0.002 ± 0.000 | 0.844 | 0.925 |
| | Coast | 7 | 152 | 162 | 7 | 76.667 | 1.000 ± 0.076 | 0.152 ± 0.029 | 0.534 | 0.725 |
| | Tanga | 12 | 127 | 130 | 4 | 38.742 | 0.455 ± 0.170 | 0.077 ± 0.035 | 1.169 | 0.823 |
| | Buffalo-associated | 7 | 161 | 175 | 5 | 53.143 | 0.857 ± 0.137 | 0.105 ± 0.047 | −1.802 | −1.895 |
| | Buffalo-derived | 4 | 144 | 157 | 4 | 84.167 | 1.000 ± 0.177 | 0.167 ± 0.039 | −0.334 | −0.249 |
| | **Total** | **82** | **226** | **298** | **35** | **68.286** | **0.872 ± 0.027** | **0.135 ± 0.008** | **−2.37*** | **−1.147** |

Statistical significance: * = significant at P < 0.05; ** = Significant at P < 0.01; Without an asterisk = Not significant at P < 0.05.
HD = Haplotype diversity; SD = Standard deviation K = Number of pairwise nucleotide differences; π = Nucleotide diversity; D = Tajima's D test statistics. Buffalo-associated = isolates collected from cattle co-grazing with buffalo; Buffalo-derived = isolates obtained from buffalo.

**Table 3**
Pairwise genetic distance ($F_{ST}$ in upper diagonal) and gene flow (*Nm* in lower diagonal) between different populations of *T. parva* isolates calculated from nucleotide sequences of Tp1 (A) and Tp2 (B) genes.

A Tp1

| | Mara | Mbeya | Morogoro | Coast | Tanga | BA | BD |
|---|---|---|---|---|---|---|---|
| Mara | | 0.011 (0.178) | 0.336 (0.034) | 0.031(0.050) | 0.253 (0.006) | 0.160 (0.058) | 0.396 (0.000) |
| Mbeya | 23.260 | | 0.417 (0.033) | −0.005 (0.271) | 0.061(0.123) | 0.245 (0.088) | 0.373 (0.066) |
| Morogoro | 0.490 | 0.350 | | 0.448 (0.002) | 0.746 (0.000) | 0.080 (0.403) | 0.526 (0.003) |
| Coast | 7.740 | −47.990 | 0.310 | | 0.180 (0.020) | 0.242 (0.015) | 0.374 (0.095) |
| Tanga | 0.740 | 3.220 | 0.090 | 1.140 | | 0.533 (0.000) | 0.495 (0.007) |
| BA | 1.320 | 0.770 | 2.870 | 0.780 | 0.220 | | 0.336 (0.015) |
| BD | 0.380 | 0.420 | 0.230 | 0.420 | 0.260 | 0.490 | |

A Tp2

| | Mara | Mbeya | Morogoro | Coast | Tanga | BA | BD |
|---|---|---|---|---|---|---|---|
| Mara | | 0.055 (0.357) | 0.544 (0.002) | −0.011 (0.312) | 0.154 (0.183) | 0.295 (0.204) | 0.083 (0.179) |
| Mbeya | 4.300 | | 0.375 (0.000) | −0.078 (0.276) | 0.233 (0.266) | 0.129 (0.180) | 0.092 (0.099) |
| Morogoro | 0.21 | 0.420 | | 0.295 (0.000) | 0.811 (0.000) | 0.044 (0.015) | 0.455 (0.000) |
| Coast | −22.81 | −3.440 | 0.600 | | 0.258 (0.087) | 0.075 (0.233) | 0.044 (0.358) |
| Tanga | 1.370 | 0.820 | 0.060 | 0.720 | | 0.571(0.015) | 0.301 (0.025) |
| BA | 0.600 | 1.690 | 5.380 | 3.070 | 0.190 | | 0.196 (0.202) |
| BD | 2.760 | 2.470 | 0.300 | 5.400 | 0.580 | 1.030 | |

$F_{ST}$ = pairwise genetic distance, Nm = number of migrants between populations.
Statistical significance: *P < 0.05; **P < 0.01; *** P < 0.001, significance was determined using 1000 coalescent simulations. Number in brackets indicate p values.

negative Nm value between Mbeya and Coast isolates to the robust genetic differentiation with low number of migrants between Morogoro and Tanga isolates (Table 3A). When population of Mara isolates was compared with those of Tanga and buffalo-derived isolates, significant $F_{ST}$ values of 0.253, p = 0.006 and 0.396, p = 0.001, respectively, were observed, implying that these populations are sufficiently differentiated. Likewise, significant genetic differentiation was observed between Morogoro isolates and isolates from Coast and Tanga regions. Non-significant $F_{ST}$ values were observed between Mara and Mbeya, Mara and Coast, Mbeya and Tanga, and Morogoro and buffalo-associated isolates, implying moderate genetic differentiation with high gene flow (Table 3A).

For the Tp2 gene, pairwise genetic distance for all *T. parva* populations varied from negative and non-significant value with negative *Nm* value between Mara and Coast, to significantly higher level of genetic differentiation with small number of migrants between Morogoro and Tanga isolates (Table 3B). The $F_{ST}$ values between Mara and Coast and between Mbeya and Coast isolates were negative and not significant with negative values of migrants between Mara and Coast and between Mbeya and Coast signifying absence of genetic differentiation at this locus. Significant $F_{ST}$ values were observed between Mara and Morogoro, Mbeya and Mororgoro, Morogoro and Coast, Morogoro and buffalo-derived, Tanga and buffalo-associated, and Tanga and buffalo-derived, implying genetic differentiation with low number of migrants between populations (Table 3B). On the other hand, non-significant $F_{ST}$ values were observed between Mara and Mbeya, Mara and buffalo-derived, Mbeya and buffalo-derived, Coast and buffalo-associated and Coast and buffalo-derived. Significant $F_{ST}$ value was observed between Morogoro and buffalo-associated with number of migrants ranging from *Nm* = 2.47 (Mbeya and buffalo-derived) to *Nm* = 5.40 (Coast and buffalo-derived), suggesting moderated genetic differentiation with high gene flow (Table 3B).

### 3.5. Population structure of T. Parva isolates

To examine the relationships among the different allelic sequences identified in this study, neighbour-joining trees were constructed based on the sequences of Tp1 and Tp2 loci. This was done to assess whether the relationships among the different allelic sequences were linked to geographical origin or mammalian host species. In the Tp1 gene, phylogenetic analysis divided the 30 distinct alleles into two major clusters, Tp1 cluster 1 and Tp1 cluster 2 (Fig. 4). Tp1 cluster 1 consisted of three sub-clusters, namely 1A, 1B and 1C. Muguga F100 isolate was grouped with 59 isolates belonging to allele 1 and five samples belonging to allele 5, 11, 12, 17 and 26 in sub-cluster 1A. Sub-cluster 1A consisted of isolates from all the seven populations together with Muguga isolate. The second sub-cluster 1B consisted of 15 isolates, 12 from buffalo-derived and three from buffalo-associated (Fig. 4). The last sub-cluster 1C consisted of 16 isolates, five from Mara, five from Mbeya, four from Tanga, and two from Coast and most of them belonged to alleles two and six. The second cluster (Tp1 cluster 2) comprised 27 out of 124 isolates belonging to three distinct alleles, 24 isolates belonging to allele four, two isolates belonging to allele 13 and one isolate belonging to allele eight. In this group, 13 isolates originated from Tanga, six from Mara, four from Mbeya and another four from Coast. The remaining one sample belonging to allele nine was not grouped to any cluster (Fig. 4).

The phylogenetic analysis of the 35 distinct Tp2 alleles grouped the 82 isolates into two major clades, Tp2 cluster 1 and Tp2 cluster 2 (Fig. 5). Most samples (52 out of 82 samples) were found within the first clade (Tp2 cluster 1), which was further sub-divided into four sub clusters, 1A with 26 samples, 1B with 13 isolates, 1C with seven samples, and 1D with five samples. One sample belonging to allele 35 was not grouped to any of the four sub-clusters. This cluster (Tp2 cluster 1), comprised samples from all the seven populations together with the reference Tp2 Muguga isolate. The second clade (Tp2 cluster 2) was subdivided into two sub clusters, 2A with 11 isolates (seven from Mara, two from Coast and one from both Tanga and buffalo-associated) and 2B with 19 isolates (10 from Tanga, five from Mbeya, three from Mara, and one from Coast) (Fig. 6).

Analysis of molecular variance (AMOVA) was used to partition the genetic diversity of the seven populations based on Tp1 and Tp2 loci. In the first analysis, the full set of Tp1 sequences from the seven populations (Mara = 27, Mbeya = 16, Morogoro = 17, Coast = 12, Tanga = 20, buffalo-associated = 19, and buffalo-derived = 13) were tested (Table 4). The results indicate that 16% of the variation among Tp1 sequences was found among populations, whereas 84% of the variation was ascribed to differences within the populations. When only unique alleles from each population (Mara = 5, Mbeya = 8, Morogoro = 1, Tanga = 5, Coast = 6, buffalo-associated = 5, and buffalo-derived = 12) were used in the analysis, all the variation (100%) among Tp1 alleles was found within populations. For the Tp2 gene, the full set of sequences (Mara = 17, Mbeya = 12, Morogoro = 23, Tanga, = 7, Coast = 17, buffalo-associated = 7, and buffalo-derived = 4) was tested. It was observed that all the variation (100%)

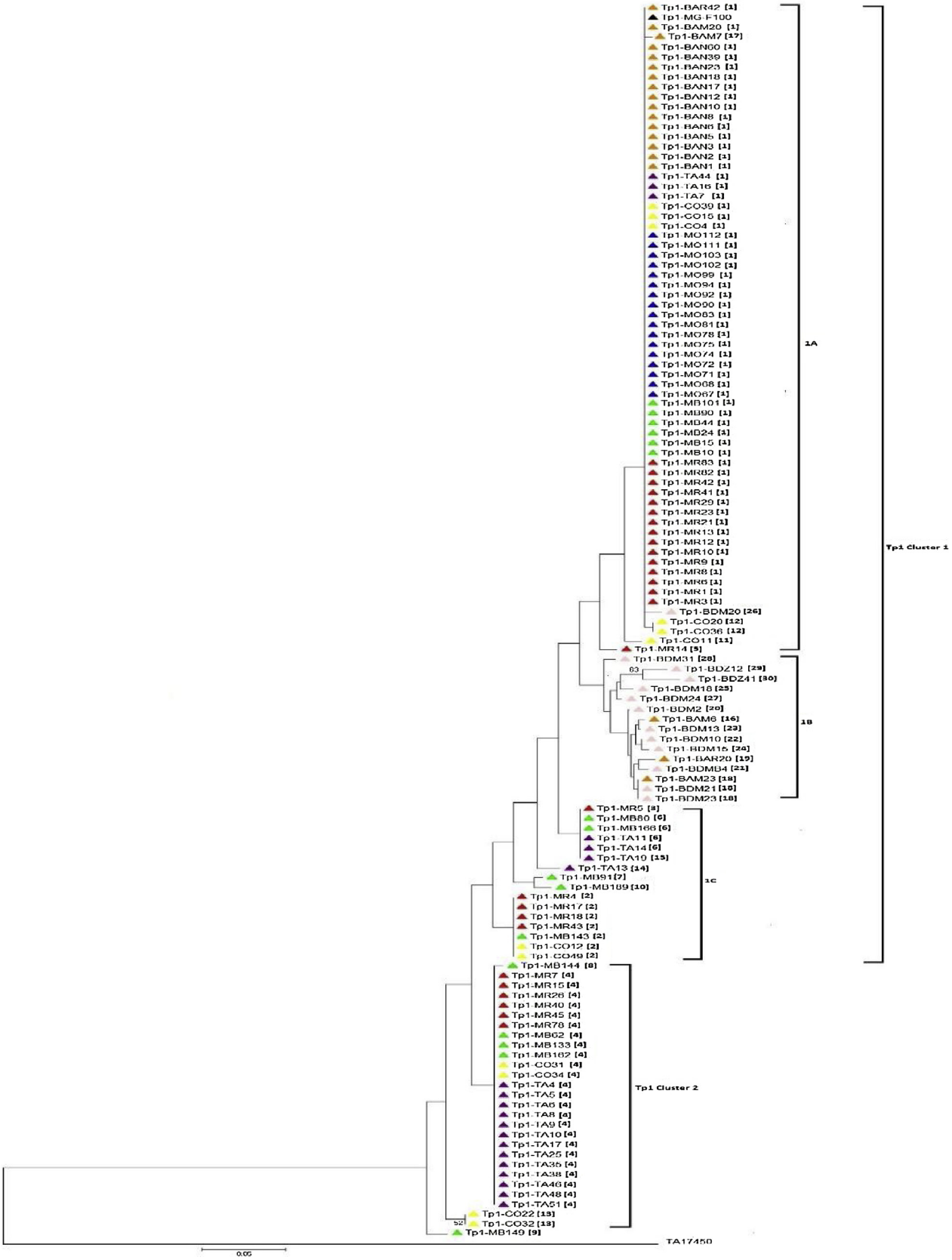


**Fig. 4.** Neighbour-joining tree of Tp1 gene sequences showing relationship among 124 cattle and buffalo-derived T. parva isolates. The isolates are labeled based on their geographical origin in Tanzania and alleles which are represented by these isolates are indicated in parentheses. Red colour represent samples from Mara, green from Mbeya, Blue from Morogoro, Yellow from Coast, Purple from Tanga, orange from buffalo-associated, and pink from buffalo-derived, and Bootstrap values > 50% are depicted above the nodes. The TP03_0849 (Tp1-F100-TpM) gene from the *T. parva* (Muguga) genome sequence shown in black color was also included in the analysis. The sequence of *T. annulata* Tp1 homologue (TA17450) was used to root the tree, and alleles which are represented by these samples are shown in brackets.

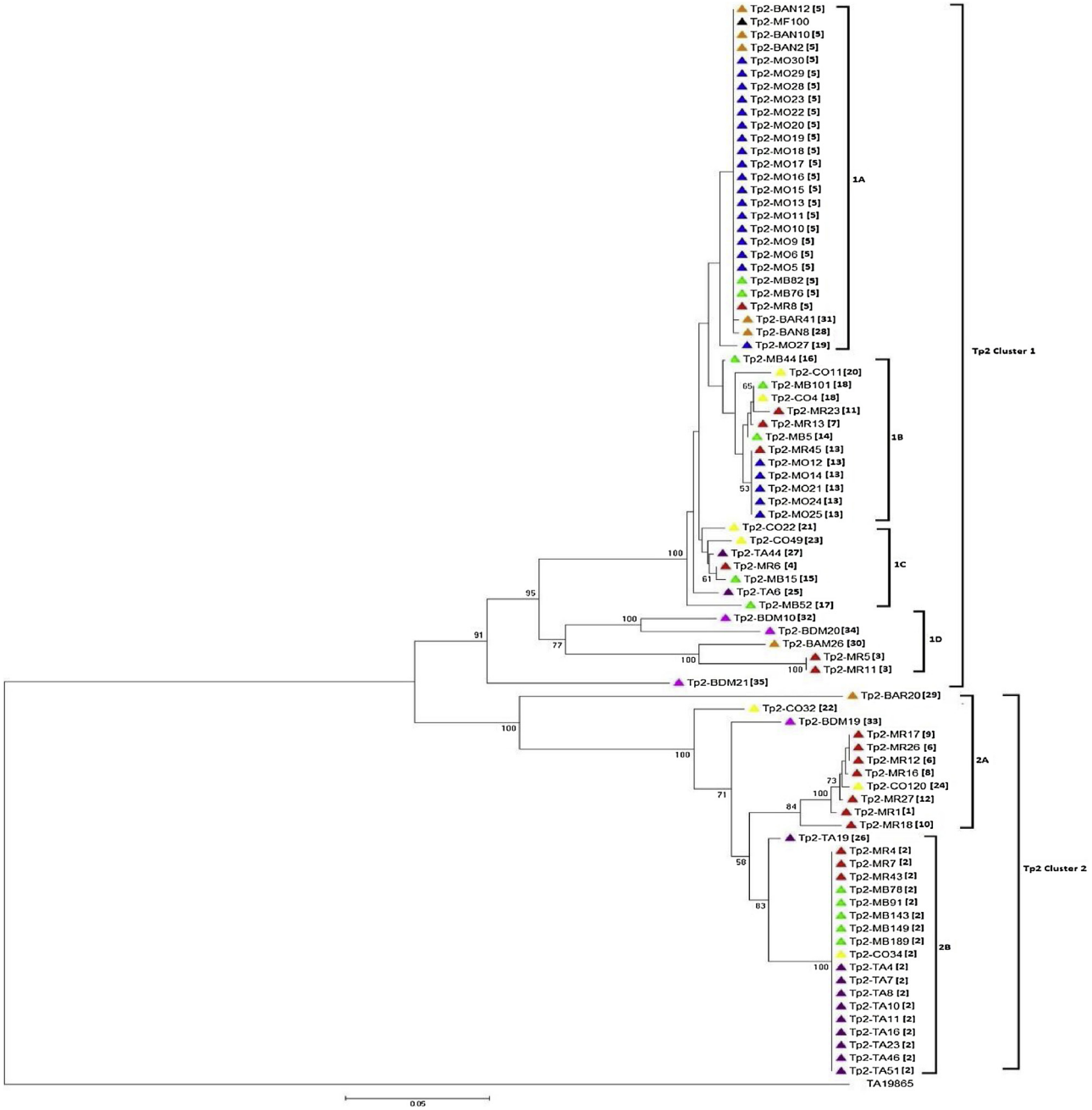


**Fig. 5.** Neighbour-joining tree of Tp2 gene sequences showing relationship among 82 cattle and buffalo-derived *T. parva* isolates. The isolates are labeled based on their geographical origin in Tanzania and alleles which are represented by these isolates are indicated in parentheses. Red color represents samples from Mara, green from Mbeya, Blue from Morogoro, Yellow from Coast, Purple from Tanga, orange from buffalo-associated, and pink from buffalo-derived. The TP01_0056 (Tp2-F100-TpM) gene from the *T. parva* (Muguga) genome sequence shown in blue color was also encompassed in the analysis. Bootstrap values > 50% are shown above the nodes. The Tp2 homologous from *T. annulata* (TA19865) was used as an outgroup to root the tree.

among Tp2 alleles occurred within population. Similarly, when only unique alleles from each population (Mara = 13, Mbeya = 7, Morogoro = 3, Tanga = 7, Coast = 4, buffalo-associated = 5, and buffalo-derived = 4) were used in the analysis, all variation (100%) among Tp2 alleles was found within populations. When Tp2 sequences from cattle derived isolates (Mara = 13, Mbeya = 7, Morogoro = 3, Tanga = 7, Coast = 4) were analyzed separately, 2% of the variation was found among populations, whereas 98% of the variation was attributed to differences within populations. On the other hand, when Tp2 sequences (5 from buffalo-associated and 4 from buffalo-derived) were tested the result showed that 100% of the variation was found within populations (Data not shown).

### 3.6. Principal component analysis (PCA) for Tp1 and Tp2 antigen genes

Principal component analysis (PCA) was conducted to determine the similarity among the isolate populations and the Muguga strain. Principal component analysis of the sequences for Tp1 gene, revealed two patterns of clustering. The first pattern involve isolates that clustered on the top right quadrant and consisted of the isolates from buffalo and few from cattle co-grazing with buffalo. The second pattern is the admix group containing samples from all populations, including Muguga isolate and is depicted on the left top and bottom quadrants. This observation is supported by phylogenetic analysis which showed that most of the samples from all populations in the Tp1 gene (Fig. 4)

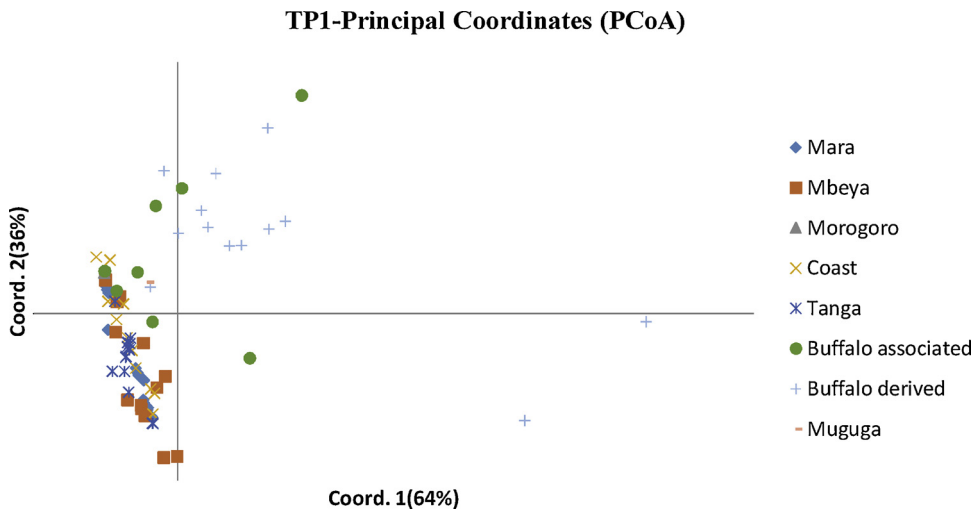


**Fig. 6.** Principal component analysis (PCA) of Tp1 allelic diversity. This diagram depicts relationship between cattle derived, buffalo-associated and buffalo-derived isolates together with the Muguga strain. The percentages of variation in the dataset explained by the 1st and 2nd principal components are indicated in parenthesis.

were clustered with Muguga isolate.

The results of the PCA for Tp2 gene indicate two main clusters, one on the left top and bottom quadrants and the other on the right top and bottom quadrants. There were sub-clusters within each main cluster (Fig. 7). The main cluster which is found on the left, comprised of isolates from all seven populations including the Muguga strain. The other cluster found on the right contained isolates from all populations, except samples from Morogoro region. Most of the isolates in the Tp2 gene occupied almost all the quadrants, suggesting weak geographical sub-structuring. In particular, the isolates from Mara and Mbeya occupied all the four quadrants while those from Morogoro occupied only the top left quadrant. These findings agree with the phylogenetic tree (Fig. 5). In addition, analysis of molecular variance (Table 4) supports this observation that all the variations among the Tp2 gene occurred within the population.

## 4. Discussion

To establish effective control measures of ECF and evaluate their effectiveness, information on genetic diversity of *T. parva* is very important. The application of molecular approaches, including sequencing of *T. parva* CD8 + T-cell target antigens, allows the analysis of parasite population diversity in areas where ECF is endemic. In this study analysis of sequence polymorphisms of Tp1 and Tp2 genes (the CD8 + T-cell target antigens) was conducted using 130 *T. parva* isolates derived from cattle which do not interact with buffalo, cattle co-grazing with buffalo and buffalo. In the current study, analysis of the sequences of Tp1 and Tp2 genes provided evidence of genetic diversity of *T. parva* isolates based on amino acid substitutions as well as among residues within CD8 + T-cell epitopes, mapped in the context of specific bovine and buffalo haplotypes. Analysis of nucleotide sequence of a section of the Tp1-encoding gene exhibited thirty (30) distinct alleles defined by point mutation. The predicted protein sequences identified five distinct variants of the antigen. The single CD8 + T-cell epitope identified in the Tp1 gene was relatively conserved at the amino acid level compared to the six known epitopes of Tp2 gene. The Tp1 gene had only five variants detected among 124 samples, linking coding changes at residues 10 and 11 of the epitope. On the other hand, sequence analysis revealed that the Tp2 antigen was highly polymorphic, with 35 distinct alleles and twelve predicted antigen variants among the 82 sequences analyzed. Analysis of the Tp2 CD8 + T-cell epitopes revealed a high number of amino acid substitutions, with less than half of the residues being conserved in each of the six epitopes across the 82 *T. parva* isolates examined. These findings agree with those reported by Pelle et al. (2011) in Kenya and Salih et al. (2017) in South Sudan who found high level of polymorphism in Tp2 antigen gene compared to Tp1 antigen gene. Moreover, a study conducted by Mwega et al. (2014a) in Eastern and Southern zones of Tanzania agrees with the current results that there is higher diversity in Tp2 antigen gene compared to Tp1 gene.

The current study revealed high haplotype diversity and low nucleotide diversity values at Tp1 and Tp2 antigen loci. The combination of high haplotype and low nucleotide diversity can be a signature of a rapid population expansion from a small effective population size as reported in a previous study (Avise, 2000). Nucleotide diversity in both Tp1 and Tp2 loci was higher in isolates derived from buffalo than in those obtained from cattle. These results are not surprising as they strongly support the assertion that buffaloes are believed to carry a heterogeneous population of parasites (Sibeko et al., 2011). Our observation agrees with the findings by Pelle et al. (2011) who reported high diversity in isolates derived from buffalo compared to the isolates from cattle. The observed higher diversity of cattle isolates from Mbeya and Coast regions compared to those from other regions (Mara, Morogoro, and Tanga) at the Tp1 locus is not surprising as the former regions (Mbeya and Coast) have been practicing immunization by ITM for many years. Therefore, isolates from Mbeya and Coast regions were expected to be more diverse than those from the other regions as

**Table 4**
Analysis of molecular variance (AMOVA) computed from seven *T. parva* populations.

| Locus | Source | Degree of freedom | Sum of square | Mean square | Estimated variation | Percentage variation |
|---|---|---|---|---|---|---|
| Tp1 | Among Populations | 6 | 1.939 | 0.323 | 0.014 | 16 |
| | Within Populations | 117 | 8.900 | 0.076 | 0.076 | 84 |
| | Total | 123 | 10.839 | | 0.09 | 100 |
| Tp2 | Among Populations | 6 | 0.071 | 0.012 | 0.000 | 0 |
| | Within Populations | 75 | 0.917 | 0.012 | 0.012 | 100 |
| | **Total** | 81 | 0.988 | | 0.0.012 | 100 |

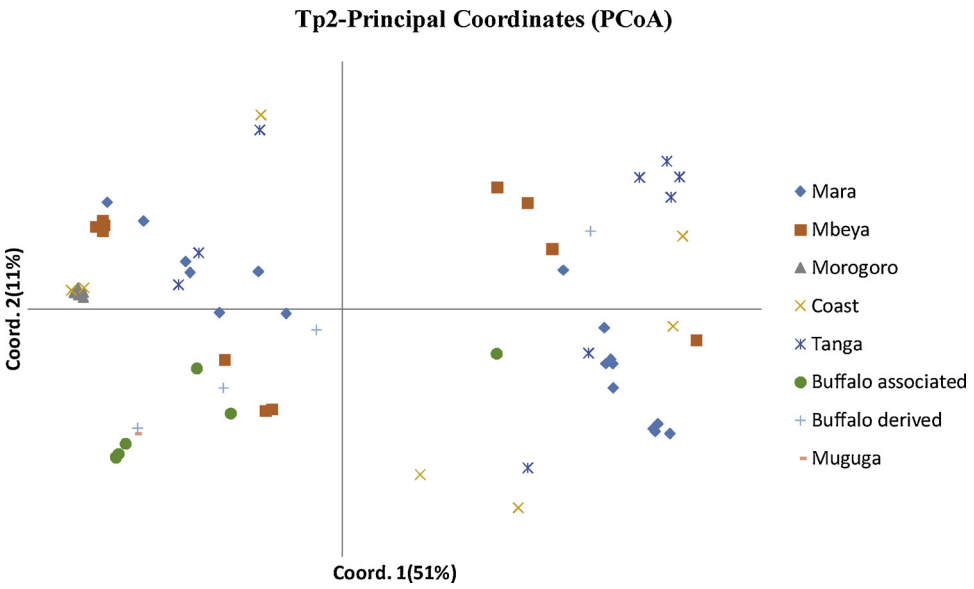


**Fig. 7.** Principal component analysis (PCA) of Tp2 allelic diversity. This diagram indicates the relationship among cattle derived, buffalo-associated and buffalo-derived populations together with Muguga strain. The percentages of variation in the dataset explained by the 1st and 2nd principal components are shown in parenthesis.

immunization increases parasite diversity (Weir et al., 2011).

In order to test selective neutrality of nucleotide polymorphism, several statistical tests have been developed and they were employed in this study to determine the population growth (Ramos-Onsins and Rozas, 2002). In the present study, Tajima's D test and Fu'sFs were used to find out if there is population expansion. Both tests are usually used to find the population expansion, but differ slightly in their approach (Ramos-Onsins and Rozas, 2002). The Tajima's D test focuses on comparison of the allelic frequency of segregating nucleotide sites (Tajima, 1989). A positive value of this test implies a bias towards intermediate frequency alleles while a negative value indicates a bias towards excess of the number of rare alleles (a signature of recent population expansion). On the other hand, the Fu'sFs is based on the alleles or haplotypes distribution. A negative value of this test indicates an excess number of alleles, meaning that there is a recent population expansion or genetic hitchhiking (Fu, 1997). In the present study, the overall Tajima's D test was negative and not significant for Tp1 gene and significantly negative for Tp2 gene. The Fu'sFs test was negative for both Tp1 and Tp2 genes but not significant. The overall negative values for both Tajima's D and Fu'sFs tests suggest an excess of the rare mutations in the populations, meaning that a recent population expansion of the *T. parva* parasite populations had happened. These results concur with those reported by Salih et al. (2017) in South Sudan which is consistent with population expansion.

Analysis of evolutionary relationship among different allelic sequences was done to test whether the relationships were linked to geographical origin or mammalian host species. In this study Muguga vaccine isolate clustered with genotypes from all populations as an admixed group. This was not expected since no vaccination against ECF (immunization) has been conducted in Mara, Morogoro, and Tanga up to the time the samples were collected. Sharing of parasite genotypes may be explained by the presence of unrestricted movements of cattle between regions which is a common phenomenon for agro-pastoralists and pastoralists in Tanzania, hence introducing new parasite genotypes. These results are congruent with those of Mwega et al. (2014a), which indicated sharing of genotypes among *T. parva* isolates from Eastern and Southern zones and Muguga isolates, even though no ECF vaccination has been conducted in the southern zone.

Principal Component Analysis of Tp1 antigen gene sequences revealed that buffalo and buffalo-associated *T. parva* isolates clustered together. This observation supports the assumption that parasites isolated from cattle co-grazing with buffalo originate from the buffalo reservoir (Pelle et al., 2011). The observation that isolates from buffalo were different from cattle isolates may affect the effectiveness of immunization through ITM, if parasite sharing occurs between these two hosts. The likelihood of sharing grazing areas between cattle and buffalo is higher since most pastoralists and agro-pastoralists graze their cattle in areas near game reserves and national parks, especially during the dry season when there is shortage of pastures in normal grazing areas. The current study showed low genetic differentiation with high number of migrants (gene flow) among *T. parva* parasites isolated from cattle and high genetic differentiation with low gene flow between isolates derived from buffalo and those obtained from cattle. The observed low genetic differentiation between cattle isolates could be attributed to unrestricted movement of cattle between regions which is a common phenomenon in the country. High genetic differentiation observed between *T. parva* isolated from buffalo and those derived from cattle suggests limited gene flow between cattle and buffalo populations.

Based on a study conducted by Hemmink et al. (2016), there is less diversity in CD8 antigens in the Muguga cocktail compared to the parasite strains circulating in the field. In their study they found a high similarity between Serengeti transformed and Muguga parasites which belong to the three components of Muguga cocktail. The main reason for this is that Muguga cocktail has been exposed to extended cattle-tick-cattle passages, following its isolation in the 1960s to 1970s when procedures for cryopreservation were developed (Cunningham et al., 1973). Since there is significant diversity in CD8 antigens in the field as observed in the current study, the deployment of Muguga cocktail may not be effective in controlling all *T. parva* strains circulating in cattle. Although immunizing parasites do not necessarily need to include all of the alleles of each polymorphic antigen to generate broad protection (reviewed in Morrison et al., 2015), we think that there is a need to produce an improved *T. parva* cocktail that contains a mixture of geographically diverse cloned parasites as suggested by Hemmink et al. (2016).

## 5. Conclusions

In this study, analysis of the sequences of Tp1 and Tp2 genes revealed evidence of genetic diversity of *T. parva* isolates from all seven populations. Nucleotide diversity in both Tp1 and Tp2 loci was higher in isolates derived from buffalo than in those obtained from cattle. The study also revealed that most samples possessed several epitopes in antigens that were identical to those in the *T. parva* Muguga reference stock, which is the main component of the widely used live vaccine

cocktail. The current study revealed low genetic differentiation with high number of migrants (gene flow) between *T. parva* parasites isolated from cattle and high genetic differentiation with low gene flow between isolates derived from buffalo and those obtained from cattle. Based on Tajima's D and Fu'sFs statistical tests the results of this study show that there is a recent population expansion of the *T. parva* parasite. In this study Muguga vaccine isolate clustered with genotypes from all populations as an admixed group. Since few samples and only two antigen encoding genes, were used in this study, and the fact that Tp1 and Tp2 sequences were obtained from the same parasite clone, these results may not be entirely representative of *T. parva* diversity in Tanzania. Extra molecular markers, including microsatellites and genome-wide single nucleotide polymorphisms as well as markers for buffalo-derived *T. parva*, could be used to improve the analysis of population genetic structure and diversity of *T. parva* in the study areas. Further study using larger sample size and inclusion of additional genetic markers, would provide better genetic profile of *T. parva* genotypes in the country. However, the presence of Tp1 and Tp2 epitope sequences similar to that of *T. parva* Muguga stock observed in this study suggests the applicability of ITM in the study areas.

## Funding

This study was funded by the INTRA-ACP Academic Mobility Scheme supported by European Union (Agreement no. 2012-3166/001) and the program for Enhancing Pro-Poor Innovation in Natural Resources and Agricultural Value Chains (EPINAV) at Sokoine University of Agriculture which was supported by the Norwegian Agency for International Development (NORAD).

## Ethics approval and consent to participate

The study was carried out with the full approval of households keeping cattle, district councils of the study areas, Sokoine University of Agriculture (SUA) and the University of Zambia (UNZA), School of Veterinary Medicine.

## Conflict of interest

The authors of this manuscript declare that they have no competing interests.

## Acknowledgement

We appreciate the support from local government authorities for their assistance during sample collection. We are very grateful to livestock farmers for their support and allowing us to use their animals during the entire period of sample collection. We also thank Microsynth AG, Switzerland for laboratory technical assistance.

## 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.

Avise, J.C., 2000. Phylogeography: the History and Formation of Species. Harvard University Press, Cambridge, pp. 447.

Cunningham, M.P., Brown, C.G.D., Burridge, M.J., Purnell, R.E., 1973. Cryopreservation of infective particles of *Theileria parva*. Int. J. Parasitol. 3, 583–587.

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 immunized 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.

Fu, Y.X., 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics 147, 915–925.

George, J.F., Pound, J.M., Davey, R.B., 2004. Chemical control of ticks on cattle and the resistance of these parasites to acaricides. Parasitology 129 (Suppl), S353–S366.

Geysen, D., Bishop, R., Skilton, R., Dolan, T.T., Morzaria, S., 1999. Molecular epidemiology of *Theileria parva* in the field. Trop. Med. Int. Health 4, A21–A27.

Gibson, W., Peacock, L., Ferris, V., Fischer, K., Livingstone, J., Thomas, J., Bailey, M., 2015. Genetic recombination between human and animal parasites creates novel strains of human pathogen. PLoS Negl. Trop. Dis. 9 (3).

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. Proc. Natl. Acad. Sci. U.S.A. 103, 3288–3291.

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., Taracha, E.L., 2007. A novel strategy for the identification of antigens that are recognized by bovine MHC class I restricted cytotoxic T cells in a protozoan infection using reverse vaccinology. Immunome Res. 3, 2.

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., Patel, E., 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.

Homewood, K., Pippa Trench, P., Randall, S., Lynen, G., Bishop, B., 2006. Livestock health and socio-economic impacts of a veterinary intervention in Maasai-land: infection and treatment vaccine against East Coast fever. Agric. Syst. 89, 248–271.

Jongejan, F., Uilenberg, G., 2004. The global importance of ticks. Parasitology 28 (Suppl. S3-14), 85–89.

Kerario, I.I., Simuunza, M.C., Chenyambuga, S.W., Koski, M., Hwang, S.G., Muleya, W., 2017b. Prevalence and Risk factors associated with *Theileria parva* infection in cattle in three regions of Tanzania. Trop. Anim. Health Prod. https://doi.org/10.1007/s11250-017-1367-8.

Kerario, I.I., Muleya, W., Chenyambuga, S., Koski, M., Hwang, S.G., 2017a. Abundance and distribution of Ixodid tick species infesting cattle reared under traditional farming systems in Tanzania. Afr. J. Agric. Res. 12, 286–299.

Kivaria, F.M., 2006. Estimated direct economic costs associated with tick borne diseases on cattle Tanzania. Trop. Anim. Health Prod. 38, 291–299.

Kivaria, F.M., 2007. The control of East Coast Fever in Africa: a constant battle for impoverished dairy farmers. Vet. J. 174, 221–222.

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

Laisser, E.L.K., Chenyambuga, S.W., Karimuribo, E.D., Msalya, G., Kipanyula, M.J., Mdegela, R., Mwilawa, A.J., Kusiluka, L.J.M., 2017. A Review on prevalence, control measure and tolerance of Tanzania shorthorn Zebu cattle to East Coast fever in Tanzania. Trop. Anim. Health Prod. 49, 813–822.

Lawrence, J.A., Perry, B.D., Williamson, S.M., 2005. East Coast fever. In: Coetzer, J.A.W., Tustin, R.C. (Eds.), Infectious Diseases of Livestock. Oxford University Press, Cape Town, Southern Africa, pp. 1–21.

Lynen, G., Di Giulio, G., Homewood, K., Reid, R., Mwilawa, A., 2006. Deployment of a live vaccine in pastoral areas: lessons learned from Tanzania. In: Rege, E., Nyamu, A., Sendalo, D. (Eds.), The Role of Biotechnology in Animal Agriculture to Address Poverty in Africa: Opportunities and Challenges. Arusha, Tanzania, pp. 193–201.

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-cellresponses to *Theileria parva* are preferentially directed to a single dominant antigen: implications for parasite strain-specific immunity. Eur. J. Immunol. 39, 2459–2469.

Makala, L.H., Mangani, P., Fujisaki, K., Nagasawa, H., 2003. The current status of major tick-borne diseases in Zambia. Vet. Res. 34, 27–45.

McKeever, D.J., 2009. Bovine immunity: a driver for diversity in Theileria parasites? Trends Parasitol. 25, 269–276.

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

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.

Morzaria, S., Williamson, S., 1999. Live vaccines for *theileria parva*: deployment in Eastern, Central and Southern Africa. In: Morzaria, S., Williamson, S. (Eds.), Proceedings of an FAO, OAU– IBAR and ILRI Workshop Held 12 March 1997 at ILRI, Nairobi, Kenya. ILRI (International Livestock Research Institute, Nairobi, Kenya.

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

Musisi, F.L., Quiroga, J.C., Mutugi, J.J., Jacobsen, J., De Castro, J.J., Di Giulio, G., 1992. Immunization against East Coast fever: recent experiences with the trivalent vaccine in East and Central Africa. In: Paling, R.W. (Ed.), Bovine Theileriosis, 3rd Symposium on Tropical and Animal Health and Production. Office for International Co-operation, Utrecht, pp. 26–32.

Musoke, R.A., Tweyongyere, R., Bizimenyera, E., Waiswa, C., Mugisha, A., Biryomumaisho, S., McHardy, N., 2004. Treatment of East Coast fever of cattle with a combination of parvaquone and frusemide. Trop. Anim. Health Prod. 36, 233–245.

Mwega, E., Salih, D.H., Njahira, M., Rukambile, E., Robert, S., Gwakisa, P., 2014a. Genetic and antigenic diversity of *Theileria parva* in cattle in Eastern and Southern

zones of Tanzania. A study to support control of East Coast fever. Parasitology. https://doi.org/10.1017/S0031182014001784. Page 1 of 8. © Cambridge University Press.
Mwega, E., Gwakisa, P., Sibeko, K., Oosthuizen, M., Geysen, D., 2014b. Molecular characterization of *Theileria parva* field strains derived from cattle and buffalo sympatric populations of northern Tanzania. Am. J. Res. Commun. 2, 10–22.
Neitz, W.O., Canham, A.S., Kluge, E.B., 1955. Corridor disease: a fatal of Bovine theileriosis encountered in Zululand. J. South. Afr. Vet. Med. Assoc. 26, 79–87.
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.
Nonga, H.E., Muwonge, A., Mdegela, R.H., 2012. Tick infestations in extensively grazed cattle and efficacy trial of high-cis cypermethrin pour-on preparation for control of ticks in Mvomero district in Tanzania. BMC Vet. Res. 8, 224.
Norval, R.A.I., Lawrence, J.A., Young, A.S., Perry, B.D., Dolan, T.T., Scott, J., 1991. *Theileria parva*: influence of vector, parasite and host relationships on the epidemiology of theileriosis in southern Africa. Parasitology 102, 347–357.
Norval, R.A., Perry, B.D., Young, A.S., 1992. The Epidemiology of Theileriosis in Africa. Academic press, London.
Oura, C.A., Bishop, R., Asiimwe, B.B., Spooner, P., Lubega, G.W., Tait, A., 2007. *Theileria parva* live vaccination: parasite transmission, persistence and heterologous challenge in the field. Parasitology 134, 1205–1213.
Peakall, R., 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., 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.
Pienaar, R., Potgieter, F.T., Latif, A.A., Thekisoe, O.M.M., Mans, B.J., 2011. The Hybrid II assay: a sensitive and specific real-time hybridization assay for the diagnosis of *Theileria parva* infection in Cape buffalo (*Syncerus caffer*) and cattle. Parasitology 2, 1–10.
Radley, D.E., Brown, C.G.D., Burridge, M.J., Cunninggham, M.P., Kirimi, I.M., Purnell, R.E., Young, A.S., 1975. East Coast fever. Chemotherapy immunization of cattle against *Theileria parva* (Muguga) and five Theileria strains. Vet. Parasitol. 1, 35–41.
Ramos-Onsins, S.E., Rozas, J., 2002. Statistical properties of new neutrality tests against population growth. Mol. Biol. Evol. 19, 2092–2100.
Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J.C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S.E., Sánchez-Gracia, A., 2017. DnaSP v6: DNA sequence polymorphism analysis of large datasets. Mol. Biol. Evol. 34, 3299–3302.
Rukambile, E., Machuka, E., Njahira, M., Kyalo, M., Skilton, R., Mwega, E., Chota, A., Mathias, M., Sallu, R., Salih, D., 2016. Population genetic analysis of *Theileria parva* isolated in cattle and buffaloes in Tanzania using minisatelite and microsatellite markers. Vet. Parasitol. 224, 20–26.
Salih, D.A., Pelle, R., Mwacharo, J.M., Njahira, M.N., Marcellino, W.L., Kiara, H., Malak, A.K., Rahim, A., Hussein, E.L., 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 (2), e0171426. https://doi.org/10.1371/journal.pone.0171426.
Shyma, K.P., Gupta, J.P., Singh, V., 2013. Breeding strategies for tick resistance in tropical cattle: sustainable approach for tick control. J. Parasit. Dis. 39, 1–6.
Sibeko, K.P., Collins, N.E., Oosthuizen, M.C., Troskie, M., Potgieter, F.T., Coetzer, J.A.W., 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. Vet. Parasitol. 181, 120–123.
Tajima, F., 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics. 123, 585–595.
Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22, 4673–4680.
Weir, W., Karagenç, T., Gharbi, M., Simuunza, M., Aypak, S., Aysul, N., Darghouth, M.A., Shiels, B., Tait, A., 2011. Population diversity and multiplicity of infection in Theileria annulata. Int. J. Parasitol. 41, 193–203.