Genotyping of *Theileria parva* populations in vaccinated and non-vaccinated cattle in Malawi

Elisha Chatanga[a,b], Yuma Ohari[a], Walter Muleya[c], Kyoko Hayashida[d], Chihiro Sugimoto[d], Ken Katakura[a], Nariaki Nonaka[a], Ryo Nakao[a,*]

[a.] Laboratory of Parasitology, Graduate School of Infectious Diseases, Faculty of Veterinary Medicine, Hokkaido University, Kita-18, Nishi-9, Sapporo, Hokkaido 060-0818, Japan

[b.] Department of Veterinary Pathobiology, Lilongwe University of Agriculture and Natural Resources, P.O. Box 219, Lilongwe, Malawi

[c.] Department of Biomedical Sciences, School of Veterinary Medicine, University of Zambia, P.O. Box 32379, Lusaka, Zambia

[d.] Division of Collaboration and Education, International Institute for Zoonosis Control, Hokkaido University, Kita-20, Nishi-10, Sapporo, Hokkaido 001-0020, Japan

*** Author for correspondence: Ryo Nakao,** ryo.nakao@vetmed.hokudai.ac.jp

Abstract

Theileria parva is an apicomplexan protozoan parasite that causes bovine theileriosis (East Coast Fever; ECF) in central, eastern, and southern Africa. In Malawi, ECF is endemic in the northern and central regions where it has negatively affected the development of dairy industry. Despite its endemic status the genetic population structure of *T. parva* in Malawi is currently unknown. To obtain an understanding of *T. parva* in Malawi, we performed population genetics analysis of *T. parva* populations in cattle vaccinated with the Muguga cocktail live vaccine and non-vaccinated cattle using mini- and microsatellite markers covering all the four *T. parva* chromosomes. The *T. parva* Muguga strain was included in this study as a reference strain. Linkage disequilibrium was observed when all samples were treated as a single population. There was sub-structuring among the samples as shown by the principal coordinate analysis (PCoA). Majority of the samples clustered with the *T. parva* Muguga reference strain suggesting that the isolates in Malawi are closely related to the vaccine component which support the current use of Muguga cocktail vaccine to control ECF. The clustering of samples from non-endemic southern region with those from endemic central region suggests expansion of the distribution of *T. parva* in Malawi.

Key words: Genotyping, Malawi, population structure, *Theileria parva*

Key Findings

- There was sub-structuring among *T. parva* in Malawi into two separate clusters.
- Samples from areas with no *T. parva* Muguga cocktail immunisation history clustered with the *T. parva* Muguga reference strain.
- The flow of genetic material was likely to occur within population than between populations.

Introduction

Theileria parva, which causes East Coast fever (ECF), one of bovine theilerioses, is the most important tick-borne pathogen of cattle on the African continent (Lubembe *et al.* 2020). It is estimated that almost half of the cattle population on the African continent are at risk of infection with *T. parva* (Lubembe *et al.* 2020). Although adult indigenous cattle in Africa rarely develop clinical diseases, calves under 6 months old are highly susceptible (Moll *et al.* 1984, 1986). However, in exotic and crossbred cattle, the fatality rate may be as high as 80-100% and this has negatively impacted the development of the dairy industry in the endemic countries (Oura *et al.* 2007). The control of ECF in Malawi is mainly based on tick control through acaricide application, livestock movement restriction where cattle from endemic northern and central regions are not allowed to go to the non-endemic southern region. Chemotherapy using buparvaquone is used to manage clinical cases in the endemic areas and immunisation of cattle using the Muguga cocktail live vaccine. The infection and treatment method (ITM) which involves inoculating the animal with a dose of the Muguga cocktail vaccine and simultaneous treatment with long acting tetracycline is currently in use in the northern and central regions where the disease is endemic (DAHLD, 2006). However, in the southern region which has been considered non-endemic, the vaccine has not been deployed although sporadic cases of ECF have been reported (Chinombo *et al.* 1988). The whole genome sequencing of *T. parva* by

Gardner et al. (2005) and the identification of *T. parva* specific mini- and microsatellite markers which were first described and applied by Oura *et al.* (2003, 2005) and Katzer *et al.* (2006, 2010) have provided good markers for the genotyping and characterisation of the population structure of *T. parva*.

However, the deployment of vaccine that will be effective against the *T. parva* strains circulating in cattle requires genotypic and population genetics data in the endemic countries. Despite *T. parva* being widely spread in central, eastern and southern Africa, there is limited information on the parasite genotypes and population structure available in southern Africa. The information on *T. parva* parasites from cattle available is only from Zambia (Muleya *et al.* 2012) and South Africa (Lubembe *et al.* 2020), although the disease is controlled in South Africa. Lubembe *et al.* (2020) investigated the genotypes of *T. parva* from buffaloes in Mozambique and South Africa. The knowledge about the genotypes and population structure of *T. parva* on the African continent is important to assess the diversity of *T. parva* in the continent as these strains may affect the current vaccines being used to control ECF.

This study aimed at investigating the genetic population structure of *T. parva* among cattle in Malawi using *T. parva* mini- and microsatellite markers. The specific objectives were 1) to determine if genotypes of *T. parva* from vaccinated cattle with *T. parva* Muguga cocktail vaccine will be different from those without vaccination history. 2) to determine if there is presence of sub-structuring among the *T. parva* isolates in Malawi and 3) to determine if the population of *T. parva* in Malawi is panmictic.

Materials and methods

Sample collection and DNA extraction

The cattle blood samples from Kasungu (n = 20), Nkhotakota (n = 20) and Katete (Lilongwe) (n = 19) were obtained from our previous study (Chatanga et al., 2020). The other samples were collected from Likasi farm (33º 17' 05" E; 14º 02' 43" S) (n = 92) and Lilongwe University

of Agriculture and Natural Resources (LUANAR) student farm (33º 77' 83" E; 14º 17' 96" S,) (n = 53) in the central region and Mikolongwe (35º 12' 30" E; 15º 51' 49" S) (n = 28) in the southern region (Figure 1). The cattle sampled from Kasungu and Nkhotakota were the indigenous Malawi zebu with no history of *T. parva* immunisation or dipping to control ticks. Furthermore, these animals belong to smallholder farmers who use communal grazing land which allows mixing with herds from other farms.

The cattle sampled from Katete farm were Holstein Friesians only, while those from Likasi, LUANAR and Mikolongwe farms were Holstein Friesians, and crossbreeds with the indigenous Malawi zebu, respectively. At Katete, Likasi, and LUANAR farms, cattle are kept under semi-intensive management system and some were immunised with the *T. parva* Muguga cocktail vaccine 2 years and 3 months before the sampling period as previously described (Chatanga et al., 2020) and dipping is done weekly to control ticks. At Mikolongwe farm, which is located in the southern region, cattle are dipped weekly but are not vaccinated against *T. parva* as the farm falls in a non-endemic area. The animals from other farms in Malawi are not allowed into the breeding population at Katete farm for disease control purposes. The farm practices natural breeding and regularly change the bulls used by importing their breeding stock from South Africa where *T. parva* is not endemic. Although the animals at Likasi, LUANAR, and Mikolongwe farms do not mix with cattle from other herds, the introduction of Malawi zebu heifers breeding stock from smallholder farmers in Malawi for crossbreeding with Holstein Friesians is allowed.

Approximately 5 ml of cattle whole blood was collected in Ethylene Diamine Tetra Acetic acid (EDTA) vacutainer tubes, aseptically by swabbing the external jugular venepuncture site with methylated spirit swab. The samples were kept on ice before transporting to the laboratory. DNA was extracted from 200 µl of whole blood using Quick Gene DNA whole blood kit S (DB-S) (Kurabo Industries Ltd., Osaka, Japan) according to

manufacturer's instructions. The extracted DNA was stored at -20°C until required for use.

Theileria parva *screening*

The samples from Likasi, LUANAR and Mikolongwe were screened in this study for *T. parva* using nested PCR assays targeting the *T. parva*-specific 104-kD antigen gene (*p104*) as described previously by Odongo et al. (2010). The amplification was done using Tks Gflex DNA Polymerase (TaKaRa Bio Inc., Shiga, Japan), the reaction mixture and PCR conditions were set as previously described (Chatanga *et al.* 2020). Approximately, 10 ng of the sample was used as DNA template and molecular grade water instead of genomic DNA was used as negative control in each run for quality control.

PCR amplification of mini- and microsatellites

The primers used for satellite amplification and their annealing temperatures are shown in Table 1. The forward primer of each pair was fluorescently labelled with either 6-FAM, / VIC, and/ or PET at the 5' end. The amplifications were conducted in a 10 µL reaction mixture, containing 5.0 µL of 2×Gflex PCR Buffer (Mg2+, dNTP plus), 0.2 µL of Tks Gflex DNA Polymerase, 0.5 µL of 200 nM of each primer, 10.0 ng of the DNA template. The volume was adjusted using distilled water. Negative control containing distilled water instead of DNA template was used for quality control. The cycling conditions were set with an initial denaturation at 94°C for 3 min, followed by 45 cycles of denaturation at 98°C for 10 sec, annealing for 30 sec, and extension at 68°C for 1 min and final extension at 68°C for 5 min. The amplicons were electrophoresed in a 1.5% agarose gel stained with Gel-Red (Biotium, Hayward, CA, USA) and visualized under UV light.

Capillary electrophoresis

The capillary electrophoresis was conducted in a 11.5 µL reaction volume comprising of 0.5 µL GS-600 LIZ size standard (Applied Biosystems, CA, USA), 10.0 µL ABI HiDi formamide (Applied Biosystems, CA, USA) and 1.0 µL of 10-fold diluted microsatellite PCR product. The mixture was denatured at 95°C for 5 min and immediately cooled on ice before capillary electrophoresis on the ABI 3500xl genetic analyser (Applied Biosystems, CA, USA).

Fragment analysis

The DNA fragment sizes were analysed relative to the ROX-labelled GS 600 LIZ size standard using Gene Mapper software version 6 (Applied Biosystems, CA, USA). This facilitated the resolution of multiple products with 1 base pair (bp) difference in a single reaction. Multiple products from a single PCR reaction indicated the presence of mixed genotypes. The output data from the genetic analyser were provided as the area under the peak of each allele (quantitative measurement), with the predominant allele possessing the greatest peak area. In this way, the predominant allele at each locus was identified for each sample, and this data was combined to generate a multi-locus genotype (MLG) representing the most abundant genotype in each sample. Only the alleles with the prescribed base pair range were used to generate the MLG and samples from the same area were electrophoresed and gene scanned on the same plate.

Data analysis

An allele sharing co-efficient (Bowcock *et al.* 1994) in Excel microsatellite toolkit (http://animalgenomics.ucd.i.e./sdepark/ms-toolkit/) was used for the similarity comparison of the MLGs (Peakall & Smouse, 2006, 2012). Similarity analysis was determined by constructing a similarity matrix. This was then utilised to conduct principal coordinate analysis (PCoA) and analysis of molecular variance (AMOVA) using the Excel plug-in software GenAIEx6

(http://www.anu.edu.au/BoZo/GenAIEx/) (Peakall & Smouse, 2006, 2012). The FSTAT computer package version 2.9.3.2 was used to calculate estimates of F statistics for population genetic analysis (http://www2.unil.ch/popgen/softwares/fstat.htm). LIAN (http://adenine.biz.fh-weihenstephan.de/lian/) was used to test the null hypothesis of linkage equilibrium by calculating a quantification of linkage equilibrium/linkage disequilibrium called the standardized index of association (I_A^S) (Haubold & Hudson, 2000). The statistical independence of alleles at all pairwise combinations of loci under study characterizes linkage equilibrium (LE) and this independent assortment was initially tested by LIAN by determining the number of loci at which each pair of MLGs differs. The mismatch values from this distribution were then used to calculate the variance (V_D) which was then compared to the variance expected (V_E) for LE. Monte Carlo (MC) computer simulation was used to test the null hypothesis that VD = VE. The computer software calculates a 95% confidence limit L. When V_D was greater than critical limit L, the null hypothesis of LE is rejected.

Mixed infection of different genotypes was indicated by the presence of several alleles at a locus in one sample. To determine the multiplicity of infection at the nine loci used, the mean number of alleles in each sample was calculated. Finally, the mean for each population as well as the combined population from the index value of each sample was also calculated to show the overall multiplicity of infection for each population and combined population.

Results

Confirmation of Theileria parva *infection by PCR*

Only samples from Likasi, LUANAR and Mikolongwe (n = 173) were screened in this study while samples from Kasungu, Nkhotakota and Katete were screened in a previous study (Chatanga *et al.* 2020). The positive rates for Likasi, LUANAR and Mikolongwe were 18% (17/95), 32% (17/53), and 25% (7/28), respectively. Likasi farm had the least positive rate at

18% despite having high number of sampled animals. The findings of this study confirm for the first time, the presence of *T. parva* infection in cattle from the southern region using molecular techniques.

Satellite marker diversity and allelic variation

The number of alleles per locus ranged from 3 in marker MS77 to 33 in marker MS14 (Table 2), and the average number of alleles observed per locus was 18.67 (Table 3). Markers MS14, MS9, and MS19 had the highest number of alleles (33, 30, and 29 alleles, respectively). Shared alleles were observed at all the 9 loci investigated with no locus with unique alleles for a particular population (Figure 2). The number of shared alleles ranged from 2 at locus MS77 to 10 at loci MS14 and MS19 (Figure 2). Similarly, genetic diversity of the 9 loci investigated showed that markers MS48 and MS77 were the least diverse in all the six populations. The populations from Kasungu and Nkhotakota showed low or no diversity at the loci MS48 and MS77 (Table 2). Similarly, the population from LUANAR showed no diversity at locus MS77, but the diversity of marker MS47 could not be determined in this population due to the low number of samples (n = 3) successfully amplified (Table 2).

Population diversity structure

The PCoA of *T. parva* for the six cattle populations in Malawi and the *T. parva* Muguga reference strain showed that there were two clusters designated as cluster A and B (Figure 3). Cluster A had the majority of samples (n = 83) from all the six sampling sites and the *T. parva* Muguga reference strain while cluster B, had 17 samples from Katete (n = 14) and Likasi (n = 3) farms from the central region. Interestingly, it was observed that even samples from areas without *T. parva* Muguga cocktail vaccination history, clustered together with *T. parva* Muguga strain (Figure 3). The analysis of molecular variance (AMOVA) showed that the

genetic variation observed was mainly within populations (99%) while that due to differences between populations was only 1% (Table 4).

The analysis of the allelic profile data used to determine the linkage of *T. parva* in Malawi showed that when all the six sub-populations were combined as a single population the standardised index of association (I^S_A) was greater than zero (0.0286). Furthermore, the pairwise variance (V_D) was greater than the 95% critical L value (M_L) which indicated linkage disequilibrium (Table 3). However, when each sub-population was treated separately, it was observed that most populations (4 out of 6) had an I^S_A which was either negative or close to zero and the V_D was less than the M_L indicating linkage equilibrium (Table 3). Katete and Likasi populations showed linkage disequilibrium when treated separately (Table 3). When populations A and B were treated as separate and individual populations as per PCoA clustering (Figure 3), both populations were found to be in LE (Table 3).

To determine the differences among the sampled populations, the **estimated** heterozygosity (H_e) and mean number of genotypes/locus were calculated for each of the six population. The **estimated** heterozygosity (H_e) ranged from 0.6127 for Mikolongwe and LUANAR to 0.7900 for Likasi (Table 3). The mean number of genotypes/locus ranged from 3.56 to 7.22 for Mikolongwe and Likasi, respectively (Table 3). The overall **estimated** heterozygosity (H_e) and mean number of genotypes/locus for the combined population was 0.666 and 18.67, respectively.

To determine the degree of genetic differentiation, the Wright's F index was calculated for each of the two clusters based on the PCoA. Cluster A which had majority of the isolates had an F_{ST} value of 0.008 while cluster B with isolates from Likasi and Katete had an F_{ST} value of 0.273 (Table 3). This finding shows significant differences between these two clusters. The combine population had an F_{ST} value of 0.105 (Table 3).

Multiplicity of infection

Multiple genotypes of *T. parva* were observed in the majority of the samples regardless of the sampling site and region. The multiplicity of infection in the six sub-populations ranged from 1.14 to 1.40 for the population from Mikolongwe and Nkhotakota, respectively as estimated by the calculated mean number of genotypes/locus (Table 5). The standard deviation ranged from 0.21 to 0.36 for Mikolongwe and Kasungu populations, respectively. The combined population had the mean value of 1.32 and a low standard deviation of 0.29 (Table 5).

Discussion

Knowledge about the genetic information and population structure of *T. parva* is a pre-requisite to the conception of effective control measures and monitoring of current measures (Lubembe *et al.* 2020). However, this information is not available in most southern African countries including Malawi although it is one of the endemic regions. *Theileria parva* is the most important tick-borne pathogen in Malawi and is endemic in the central and northern regions (DAHLD, 2006). Current control measures in use include immunisation using *T. parva* Muguga cocktail live sporozoite vaccine for exotic, crossbreed cattle and Malawi zebu calves below 6 months in the endemic areas (Perry, 2016; Lawrence *et al.* 1996). To understand the genetic composition and population structure of *T. parva* in Malawi, nine mini- and microsatellite markers were employed to examine 100 *T. parva* samples from six populations; five in the central region where ECF is endemic and one in the southern region which has traditionally been considered as non-endemic until this study. To control the spread of ECF in Malawi, movement of animals from the ECF endemic northern and central regions to the non-endemic southern region is not allowed and vaccination with *T. parva* Muguga cocktail is not permitted in the latter (DAHLD, 2006).

Almost all the populations investigated had high mean genetic diversity (**estimated** heterozygosity) that ranged from 0.612 for Mikolongwe and LUANAR to 0.790 for Likasi. The lower **estimated** heterozygosity at Mikolongwe may be due to limited number of samples as it had the least number of samples among the populations examined. Further, this observation may result from the pathogen being newly introduced in this region as reported previously in South Sudan (Salih *et al.* 2018; Marcellino *et al.* 2017). The lower He for LUANAR could be due to reduced exposure of the animals to the parasites as the animal are kept for a short period of time. The animals are disposed after some studies are concluded at the student farm unlike the other farms which keep animals for longer periods. The high **estimated** heterozygosity at Likasi farm is also supported by the PCoA results which showed that the isolates at the farm separated into two clusters and majority were in cluster A while the rest were in cluster B. The higher expected genetic diversity observed at Likasi may result from longer history of genetic recombination than the other populations as it was used as a trial site when the *T. parva* Muguga cocktail live vaccine was being developed for the eastern and southern Africa region (Lawrence *et al.* 1996; Dolan, 1988). The overall **estimated** heterozygosity for Malawi combined population of 0.67 is close to the values reported in Zambia (0.75) (Muleya *et al.* 2012), South Sudan (0.73) (Salih *et al.* 2018) and Tanzania (0.79) (Rukambile *et al.* 2016). However, it is lower than the 0.91 reported in Burundi (Atuhaire *et al.* 2021) and the 0.81 from the study combining eastern and southern Africa isolates from cattle and buffaloes (Lubembe *et al.* 2020). This observation may be due to the fact that the study either included samples from buffaloes or involved sampling from a wider geographical area resulting in a higher mean genetic diversity.

The observation of shared alleles in most samples from most of the sampled populations at all loci investigated shows that almost similar strains of *T. parva* are circulating among cattle in Malawi. This has also been strongly supported by the PCoA in which the majority of samples

289 from the six populations together with the *T. parva* Muguga reference strain clustered in
290 population A. This finding is in accordance with a previous study based on *Tp1* and *Tp2* genes,
291 which showed that most sequences were either identical or similar to *T. parva* Muguga, *T.*
292 *parva* Kiambu-5 and *T. parva* Serengeti transformed, the components of *T. parva* Muguga
293 cocktail vaccine (Chatanga *et al.* 2020). However, the presence of a separate minority
294 population (cluster B), different from the *T. parva* Muguga reference strain, was also indicated
295 as per PCoA results and the F_{ST} value of 0. 271 between population A and B. The majority of
296 samples from this minority population originated from Katete (n = 17) with Likasi only
297 contributing three (3). The separate population at Katete may result from it being a closed
298 population that does not allow introduction of animals from other farms in Malawi which may
299 introduce other *T. parva* strains circulating in other cattle populations in Malawi. Since some
300 of the sampled animals were only immunised 2 years and 3 months before the sampling, it may
301 help to explain why only a few samples from the farm clustered in majority population The
302 other isolates at Katete farm that clustered in population A may be due to immunisation with
303 the *T. parva* Muguga cocktail vaccine. Although, the other farms also have restricted access to
304 other animals, they allow breeding stock from smallholder farmers in Malawi into their
305 population which is likely to explain the relatedness with those populations from smallholder
306 farmers. The finding of mixed genotypes within a sample shows that mixed infections of *T.*
307 *parva* are common among cattle in Malawi. This may be due to prolonged exposure to ticks
308 and the complex life cycle to *T. parva* which include sexual and asexual stages in the vector
309 tick and host cattle, respectively (Katzer *et al.* 2010). The genetic recombination phenomenon
310 of *T. parva* in the vector tick which it has adopted as a strategy to survive and drive genetic
311 diversity ensures that host cattle can still be infected with new strains (Oura *et al.* 2003).
312 AMOVA showed that *T. parva* isolates in cattle in Malawi have high chance of
313 exchanging genetic material within each population than between populations. However, the

314 population at Katete was exceptional to this observation. This finding may be due to the point
315 that the immunisation with Muguga cocktail vaccine was just done 2 years and 3 months before
316 sampling in some animals which may not have provided enough time for the exchange of
317 genetic material to occur. This was supported by the high genetic variation observed within the
318 population (99%) than between populations (1%). This finding is also in agreement with
319 previous studies which reported similar findings in other countries (Lubembe *et al.* 2020; Salih
320 *et al.* 2018). As reported previously by Salih *et al.* (2018) in South Sudan, the findings of this
321 study do not support the presence of any correlation between genetic structure of a population
322 and the geographical location of the isolates.

323 Linkage analysis showed that the majority of the individual populations were in linkage
324 equilibrium and thus in panmixia, further supporting the results of PCoA and AMOVA that
325 gene flow is likely to occur within a particular population. Although this finding is not in
326 agreement with that reported in Zambia (Muleya *et al*. 2012), it is however supported by the
327 findings by Lubembe *et al.* (2020) that individual populations in southern Africa are in linkage
328 equilibrium. When the populations were combined, linkage disequilibrium was observed. This
329 may be due to Katete farm being a closed population that makes it difficult for random mating
330 to occur even with other populations within a 50 km radius. When the population from Katete
331 was excluded, the combined population was in linkage equilibrium even with those from
332 Mikolongwe with almost 500 km distance from the other population. Currently, there seems to
333 an increase in the detection of infections in the south of the country, this might be attributed to
334 human assisted movement of animals from endemic to no-endemic regions despite having the
335 livestock movement ban in place. Thus, there is a need to revise the current control measures
336 so that this trend can be reversed.

337 In conclusion, this study has for the first time provided molecular evidence of the
338 presence of *T. parva* infection in cattle in the southern region of Malawi. Further, the findings

have shown that the *T. parva* population in southern Malawi is closely related to the populations in the central region which is endemic to ECF. The study has also shown that *T. parva* genotypes circulating in both the vaccinated and unvaccinated cattle in Malawi are closely related. It has also been shown that there is presence of sub-structuring among the samples investigated. Generally, the population of *T. parva* in Malawi is not panmictic but individual populations are in panmixia. Furthermore, it has been shown that in Malawi there are at least two clusters of *T. parva*, one that is closely related to *T. parva* Muguga reference strain and another that is separate and different from the *T. parva* Muguga cocktail vaccine strain. It is therefore important to carry out a nationwide wide study to provide a comprehensive genetic population structure of *T. parva* in Malawi and to include samples from the neighbouring countries to determine the relatedness of the isolates in the region. This will help to assess the impact of the *T. parva* strains that are different from the *T. parva* Muguga cocktail vaccine strain on the control of ECF in Malawi in particular and southern Africa as a region in general.

Authors Contribution

EC, WM & RN conceived and designed the study. EC conducted the experiments. EC, YO & WM performed the population genetic and statistical analyses. EC wrote the original manuscript. WM, KH, CS & RN provided resources. RN & CS obtained funding. KH, CS, KK, NN & RN supervision, writing review and editing. All authors approved the submitted final version of the manuscript.

Acknowledgements

The authors wish to thank the management of the African Union Centre of Excellence for Ticks and Tick-borne Diseases (CTTBDs), Lilongwe, Malawi for providing laboratory space during

the sampling period. We also wish to extend our sincere appreciation to Professor Richard Bishop for providing the *Theileria parva* Muguga reference strain to Professor Chihiro Sugimoto which was used as a reference strain.

Ethical Standards

This study was approved by Ministry of Agriculture, Irrigation and Water Development (MoAIWD) in Malawi through the Department of Animal Health and Livestock Development (DAHLD) reference number 10/15/32/D and informed consent was obtained from the owners/custodians.

Conflicts of Interest

The authors declare there are no conflicts of interest.

Financial Support

This study was financially supported in part by JSPS KAKENHI (grant numbers 19H03118, 20K21358 and 20KK0151). International Collaborative Research Programme for Tackling Neglected Tropical Disease (NTD) Challenges in African countries (JP18jm0510001) of Japan Agency for Medical Research and Development (AMED). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Data

All the data generated in this study has been provided in this paper.

References

388 **Atuhaire, DK, Muleya, W, Mbao, V, Niyongabo, J, Nyabongo, L, Nsanganiyumwami, D,**
389 **Salt, J, Namangala, B and Musoke, AJ** (2021). Molecular characterization and
390 population genetics of *Theileria parva* in Burundi's unvaccinated cattle: Towards the
391 introduction of East Coast fever vaccine. *PLoS ONE* **16**, e0251500. DOI:
392 10.1371/journal.pone.0251500

393 **Bowcock, AM, Ruiz-Linares, A, Tomfohrde, J, Minch, E, Kidd, JR and Cavalli-Sforza,**
394 **LL** (1994). High resolution of human evolutionary trees with polymorphic
395 microsatellites. *Nature* 368, 455-457. DOI: 10.1038/368455a0

396 **Chatanga, E, Hayashida, K, Muleya, W, Kusakisako, K, Moustafa, MAM, Salim, B,**
397 **Katakura, K, Sugimoto, C, Nonaka, N, and Nakao, R** (2020). Genetic diversity and
398 sequence polymorphism of two genes encoding *Theileria parva* antigens recognized
399 by CD8+ T cells among vaccinated and unvaccinated cattle in Malawi. *Pathogens* **9**,
400 334. DOI: 10.3390%2Fpathogens9050334

401 **Chinombo, DO, Mzoma, FJ and Musisi, FL** (1988). East Coast Fever in Malawi (1983–
402 1987). Theileriosis in Eastern, Central and Southern Africa. In Proceedings of a
403 Workshop on East Coast Fever Immunization, Lilongwe, Malawi, 20–22 September
404 1988, Dolan, T.T, Ed, ILRAD: Nairobi, Kenya, pp. 12–16.

405 **DAHLD** (2006). Policy Document on Livestock in Malawi. Reducing poverty through
406 sustainable, private sector/farmer driven livestock production in Malawi. Department
407 of Animal Health and Livestock Development. Ministry of Agriculture, Lilongwe,
408 Malawi.

409 **Dolan, TT** (Ed.) (1988). Theileriosis in Eastern, Central and Southern Africa. In Proceedings
410 of a Workshop on East Coast Fever Immunization, Lilongwe, Malawi, 20–22
411 September 1988; ILRAD: Nairobi, Kenya.

Gardner, MJ, Bishop, R, Shah, T, De Villiers, EP, Carlton, JM, Hall, N, Ren, Q, Paulsen, IT, Pain, A, Berriman, M, Wilson, RJM, Sato, S, Ralph, SA, Mann, DJ, Xiong, Z, Shallom, SJ, Weidman, J, Jiang, L, Lynn, J, Weaver, B, Shoaibi, A, Domingo, AR, Wasawo, D, Crabtree, J, Wortman, JR, Haas, B, Angiuoli, SV, Creasy, TH, Lu, C, Suh, B, Silva, JC, Utterback, TR, Feldblyum, TV, Pertea, M, Allen, J, Nierman, WC, Taracha, ELN, Salzberg, SL, White, OR, Fitzhugh, HA, Morzaria, S, Venter, JC, Fraser, CM, Nene, V (2005) Genome sequence of *Theileria parva*, a bovine pathogen that transforms lymphocytes *Science* 309, 134-137 https://doi.org/10.1126/science.1110439

Haubold, B and Hudson, RR (2000). LIAN 3.0: detecting linkage disequilibrium in multilocus data. Linkage analysis. *Bioinformatics* **16**, 847-848. DOI: 10.1093/bioinformatics/16.9.847

Katzer, F, Daniel, N, Chris, O, Bishop, RP, Evans, LN, Taracha, ARW and Declan, JM (2006). Extensive genotypic diversity in a recombining population of the apicomplexan parasite *Theileria parva*. *Infection and Immunity* 74, 5456–5464. DOI: 10.1128/IAI.00472-06

Katzer, F, Ngugi, D, Walker, AR and McKeever, DJ (2010). Genotypic diversity, a survival strategy for the apicomplexan parasite *Theileria parva*. *Veterinary Parasitology* 167, 236–243. DOI: 10.1016/j.vetpar.2009.09.025

Lawrence, JA, Musisi, FL, Mfitilodze, MW, Tjornehoj, K, Whiteland, AP, Kafuwa, PT and Chamambala, KE (1996). Integrated tick and tick-borne disease control trials in crossbred dairy cattle in Malawi. *Tropical Animal Health and Production* **28**, 280–288. DOI: 10.1007/BF02240818

Lubembe, DM, Odongo DO, Salih, DA and Sibeko-Matjila, KP (2020). Microsatellite and minisatellite genotyping of *Theileria parva* population from southern Africa reveals

possible discriminatory allele profiles with parasites from eastern Africa. *Ticks and Tick-Borne Diseases* **11**, 101539. DOI: 10.1016/j.ttbdis.2020.101539

Marcellino, WL, Salih, DA, Njahira, MN, Ndiwa, N, Araba, A, El Hussein, AM, Seitzer, U, Ahmed, JS, Bishop, RP and Skilton, RA (2017). The emergence of *Theileria parva* in Jonglei state, South Sudan: confirmation using molecular and serological diagnostic tools. *Transboundary Emerging Diseases* **64**, 1229-1235. DOI: 10.1111/tbed.12495

Moll, G, Lohding, A and Young, A (1984) Epidemiology of theilerioses in the Trans-Mara division Kenya: Husbandry and disease background and preliminary investigations on theilerioses in calves. *Preventive Veterinary Medicine* **2,** 801–831, 1984

Moll, G, Lohding, A, Young, A and Leitch, B (1986) Epidemiology of theileriosis in calves in an endemic area of Kenya. *Veterinary Parasitology* **19**, 255–273, 1986

Muleya, W, Namangala, B, Simuunza, M, Nakao, R, Inoue, N, Kimura, T, Ito, K, Sugimoto, C and Sawa, H (2012). Population genetic analysis and sub-structuring of *Theileria parva* in the northern and eastern parts of Zambia. *Parasites & Vectors* **5**, 255. DOI: 10.1186/1756-3305-5-255

Mwega, E, Salih, DH, Njahira, M, Rukambile, E, Skilton, R and Gwakisa, P (2015). 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* **142**, 698–705. DOI: 10.1017/S0031182014001784

Odongo, D, Sunter, JD, Kiara, HK, Skilton, RA and Bishop, RP (2010). A nested PCR assay exhibits enhanced sensitivity for detection of *Theileria parva* infections in bovine blood samples from carrier animals. *Parasitology Research* **106**, 357–365. DOI: 10.1007/s00436-009-1670-z

461 **Oura, CA, Asiimwe, BB, Weir, W, Lubega, GW and Tait, A** (2005). Population genetic
462 analysis and sub-structuring of *Theileria parva* in Uganda. *Molecular and Biochemical*
463 *Parasitology* **140**, 229–239. DOI: 10.1016/j.molbiopara.2004.12.015
464 **Oura, CA, Odongo, DO, Lubega, GW, Spooner, PR, Tait, A and Bishop, RP** (2003). A
465 panel of microsatellite and minisatellite markers for the characterisation of field
466 isolates of *Theileria parva*. *International Journal for Parasitology* **33**, 1641–1653.
467 DOI: 10.1016/S0020-7519(03)00280-7
468 **Oura, CAL, Bishop, R, Asiimwe, BB, Spooner, P, Lubega, GW and Tait, A** (2007).
469 *Theileria parva* live vaccination: parasite transmission, persistence and heterologous
470 challenge in the field. *Parasitology* **134**, 1205–1213. DOI:
471 10.1017/S0031182007002557
472 **Peakall, R and Smouse, PE** (2006). GENALEX 6: Genetic analysis in excel. Population
473 genetic software for teaching and research. *Molecular and Ecology Notes* **6**, 288-295.
474 DOI: 10.1093/bioinformatics/bts460
475 **Peakall, R and Smouse, PE** (2012). GenAlEx 6.5: genetic analysis in Excel. Population
476 genetic software for teaching and research–an update. *Bioinformatics* **28**, 2537-2539.
477 DOI: 10.1093/bioinformatics/bts460
478 **Perry, B** (2016). The control of East Coast fever of cattle by live parasite vaccination: A
479 science-to-impact narrative. *One Health* 2, 103–114. DOI:
480 10.1016/j.onehlt.2016.07.002
481 **Rukambile, E, Machuka, E, Njahira, M, Kyalo, M, Skilton, R, Mwega, E, Chota, A,**
482 **Mathias, M, Sallu, R, and Salih. D** (2016). Population genetic analysis of *Theileria*
483 *parva* isolated in cattle and buffaloes in Tanzania using minisatellite and microsatellite
484 markers. *Veterinary Parasitology* **224**, 20-26. DOI: 10.1016/j.vetpar.2016.04.038

Salih, DA, Mwacharo, JM, Pelle, R, Njahira, MN, Odongo, DO, Mbole-Kariuki, MN, Marcellino, WL, Malak, AK, Kiara, H, El Hussein, ARM, Bishop, RP, and Skilton, RA (2018) Genetic diversity and population structure of *Theileria parva* in South Sudan. *Ticks and Tick-Borne Diseases* **9,** 806-813. DOI: 10.1016/j.ttbdis.2018.03.002

Appendix

Table 1. The list of satellite marker primers used in this study.

Marker ID	Primer name	Sequence 5'→ 3'	Chromosome	Annealing temperature (°C)	Amplicon size (bp)	Reference
ms1	ms1F	TGAGGCAGTGTAGAGCGCATAAC	1	60	235-368	Oura et al. 2003
	ms1R	AAATCCGCAACGCTATTGCCGAGG				
MS9	MS9F	CTGGTTCCTCATCTTCACACTA	3	60	230	Katzer et al., 2006
	MS9R	CTTTCCAGAACCTACAATCAC				
MS14	MS14F	ATGCCAATTCGGTAAAGGTCTCCG	2	60	360-600	Katzer et al. 2010
	MS14R	GCATATCTCAGTCAAGCCAACATC				
MS19	MS19F	CCAGACACCTCAAATCCCAAGTA	2	60	304	Oura et al. 2003
	MS19R	CCACACTGCCACCTAATACAAA				
MS39	MS39F	CCAATCAACATCAACTACTCC	4	60	263	Katzer et al. 2010
	MS39R	CGAACTCCAAACGATCTAAAC				
ms47	M47F	GTCACAAGGGAAATCATGTCACTC	1	60	398	Katzer et al. 2006
	M47R	GAGCCTTGAGTAGGTCTAAATTTG				
MS48	M48F	CTACTTCTGGATCAGGTGTGGTGG	1	60	223	Katzer et al. 2006
	M48R	GATTGAGACGATCCCGGTAGTCCT				
MS68	M68F	TCACATCGGGTAACAAGAA	1	60	469	Katzer et al. 2010
	M68R	TATTTATCGACCCCAAATCG				
MS77	MS77F	GGTAACCAACAACCACATTT	2	60	270	Katzer et al. 2010
	MS77R	TGCTTATGAACTCAATCATCTC				

ms = minisatellite, MS = microsatellite, F = forward, R = reverse, and bp = base pair.

495 Table 2. Allelic variation among *Theileria parva* from cattle in Malawi.

	Study site	N	Satellite markers								
			MS1	MS9	MS14	MS19	MS39	MS47	MS48	MS68	MS77
Alleles within population	Kasungu	20	3	11	13	12	6	9	1	5	1
	Nkhotakota	20	4	7	13	12	8	8	2	5	1
	Katete	19	4	8	5	8	6	2	2	7	2
	Likasi	17	6	11	11	11	8	6	4	6	2
	LUANAR	17	4	7	10	9	3	1	4	7	1
	Mikolongwe	7	3	5	6	4	3	4	2	3	2
	Overall	100	9	30	33	29	22	20	7	15	3
Genetic diversity	Kasungu	20	0.511	0.860	0.963	0.971	0.542	0.945	0.000	0.660	0.000
	Nkhotakota	20	0.616	0.858	0.949	0.961	0.641	0.886	0.118	0.625	0.000
	Katete	19	0.654	0.808	0.405	0.669	0.758	0.154	0.425	0.857	0.118
	Likasi	17	0.794	0.949	0.941	0.952	0.848	0.893	0.467	0.848	0.400
	LUANAR	17	0.714	0.802	0.895	0.936	0.700	NA	0.495	0.818	0.000
	Mikolongwe	7	0.667	0.905	1.000	0.900	0.700	1.000	0.571	0.600	0.571

496 N = number of samples, NA = not determined

497 Table 3. Population genetic analyses of *Theileria parva* in Malawi.

Population	N	H_e	Mean number of genotypes/ locus	I_A^S	V_D	L	P-value	Linkage	F_{ST}
Kasungu	20	0.666	6.78	-0.0177	1.2222	1.7936	8.20×10^{-01}	LE	
Nkhotakota	20	0.6947	6.67	0.0038	1.3644	1.6290	3.00×10^{-01}	LE	
Katete	19	0.6368	4.89	0.0988	3.1507	2.4566	$< 1.00 \times 10^{-02}$	LD	
Likasi	17	0.7900	7.22	0.0762	2.0248	1.6100	$< 1.00 \times 10^{-02}$	LD	
LUANAR	17	0.6127	5.11	0.0082	1.2739	1.5109	3.30×10^{-01}	LE	
Mikolongwe	7	0.6127	3.56	-0.0334	1.0000	2.2000	8.70×10^{-01}	LE	
Pop A	83	0.7019	16.33	0.0034	1.4721	1.5777	3.20×10^{-01}	LE	0.008
Pop B	17	0.7019	3.89	0.0034	1.4721	1.6322	3.90×10^{-01}	LE	0.273
Combined population	100	0.666	18.67	0.0286	1.6642	1.4900	0.067	LD	0.105

498 N= number of samples, H_e = estimated heterozygosity, I_A^S = standard index of association, V_D = mismatch variance (linkage analysis), L = upper

499 95% confidence critical limits of Monte Carlo simulation, LD = linkage disequilibrium, LE = linkage equilibrium, F_{ST} = Wright's fixation index

500

501 Table 4. The analysis of molecular variance (AMOVA) of *Theileria parva* in Malawi.

Source	*df*	Sum of squares	Mean Sum of squares	Variance components.	% of variation
Among Populations	5	456442.9	91288.58	835.277	1%
Among Individuals	94	6002033	63851.41	0	0%
Within Individuals	100	9308005	93080.05	93080.05	99%
Total	199	15766480		93915.32	100%
	Fixed indices				
	F_{SC}	0.011			
	F_{ST}	-0.186			
	F_{CT}	-0.174			

502 df = degress of freedom, F_{SC} = variation among populations within groups, F_{ST} = measure of population differentiation due to genetic structure,

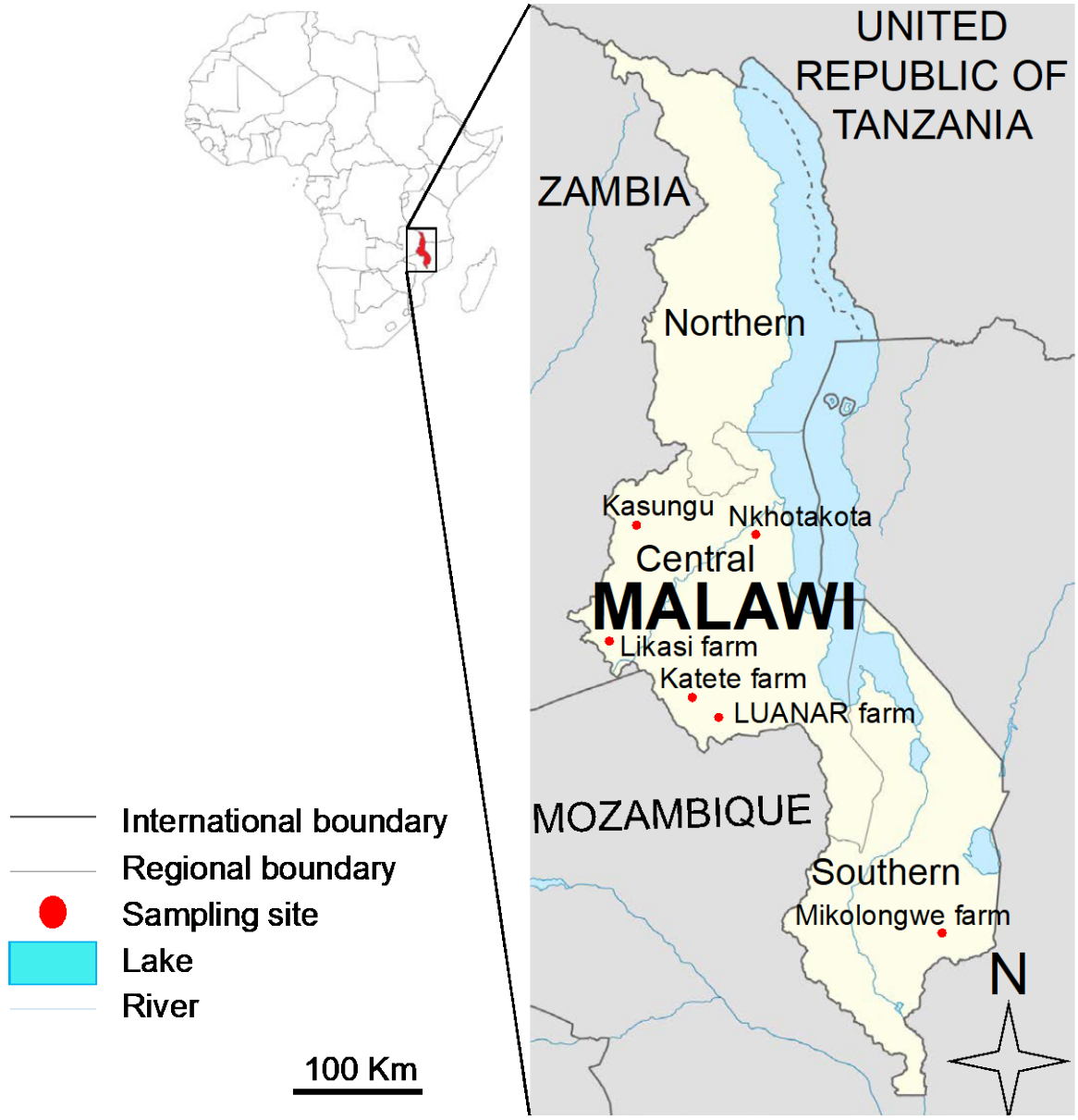
503 F_{CT} = Variation among groups of populations.

Table 5. Multiplicity of infection of *Theileria parva* in Malawi.

Population	Number of samples	Multiplicity of infection				Total number of alleles identified on all loci
		Mean	SD	Min.	Max.	
Kasungu	20	1.34	0.36	0.78	2.22	61.00
Nkhotakota	20	1.40	0.27	0.89	1.89	60.00
Katete	19	1.31	0.33	0.44	1.88	44.00
Likasi	17	1.39	0.30	0.67	1.89	65.00
LUANAR	17	1.14	0.28	0.67	1.56	46.00
Mikolongwe	7	1.31	0.21	1.00	1.56	32.00
Overall	100	1.32	0.29	0.74	1.83	51.33

SD: standard deviation, Min. and Max. refer to the minimum and maximum, respectively for the number of alleles identified per locus per sample.

The summary of the number of the alleles per locus for each population is provided in Table 2.



516

517

518 **Figure 1**. Map of Malawi showing the sample collection sites.

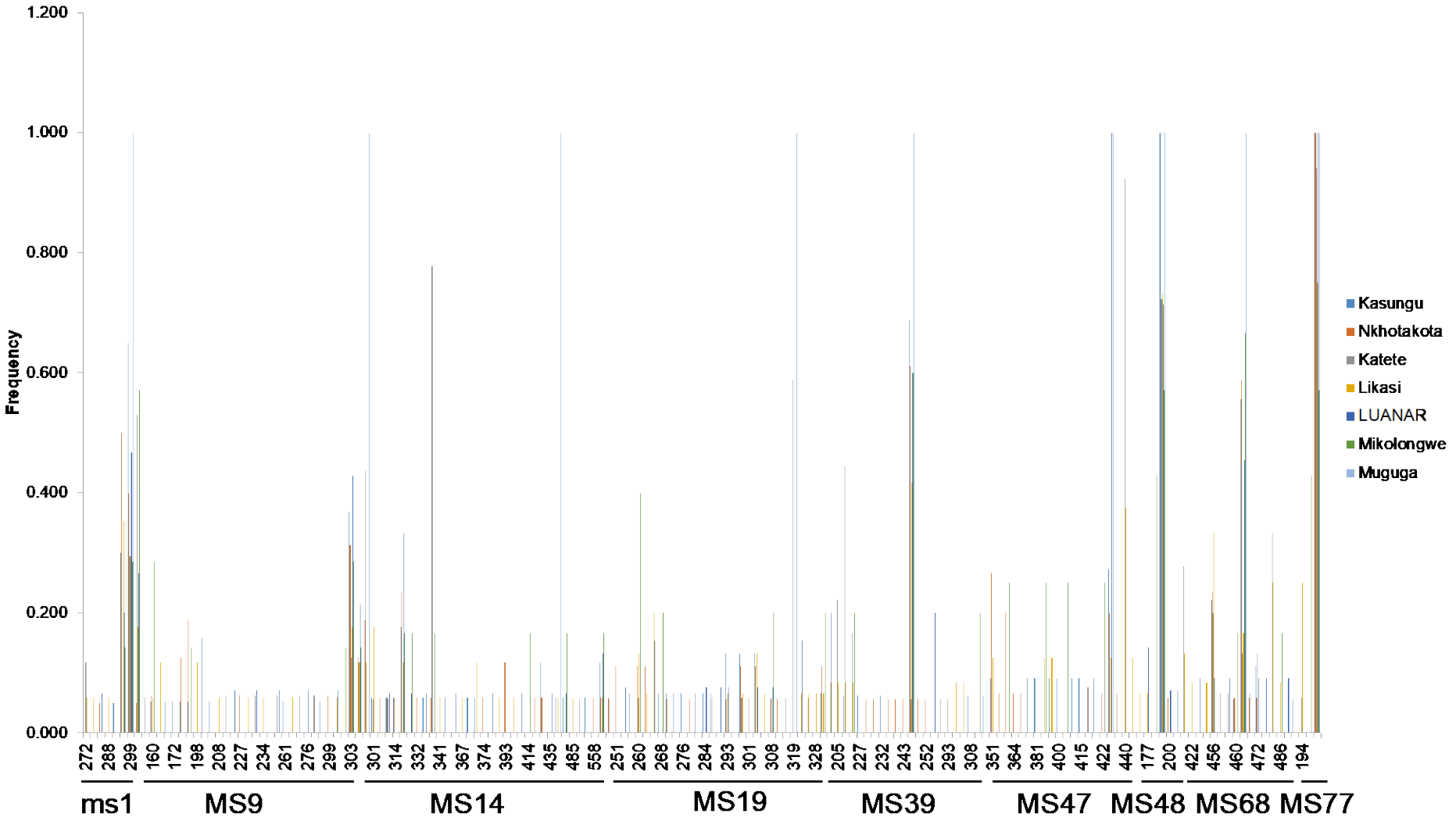
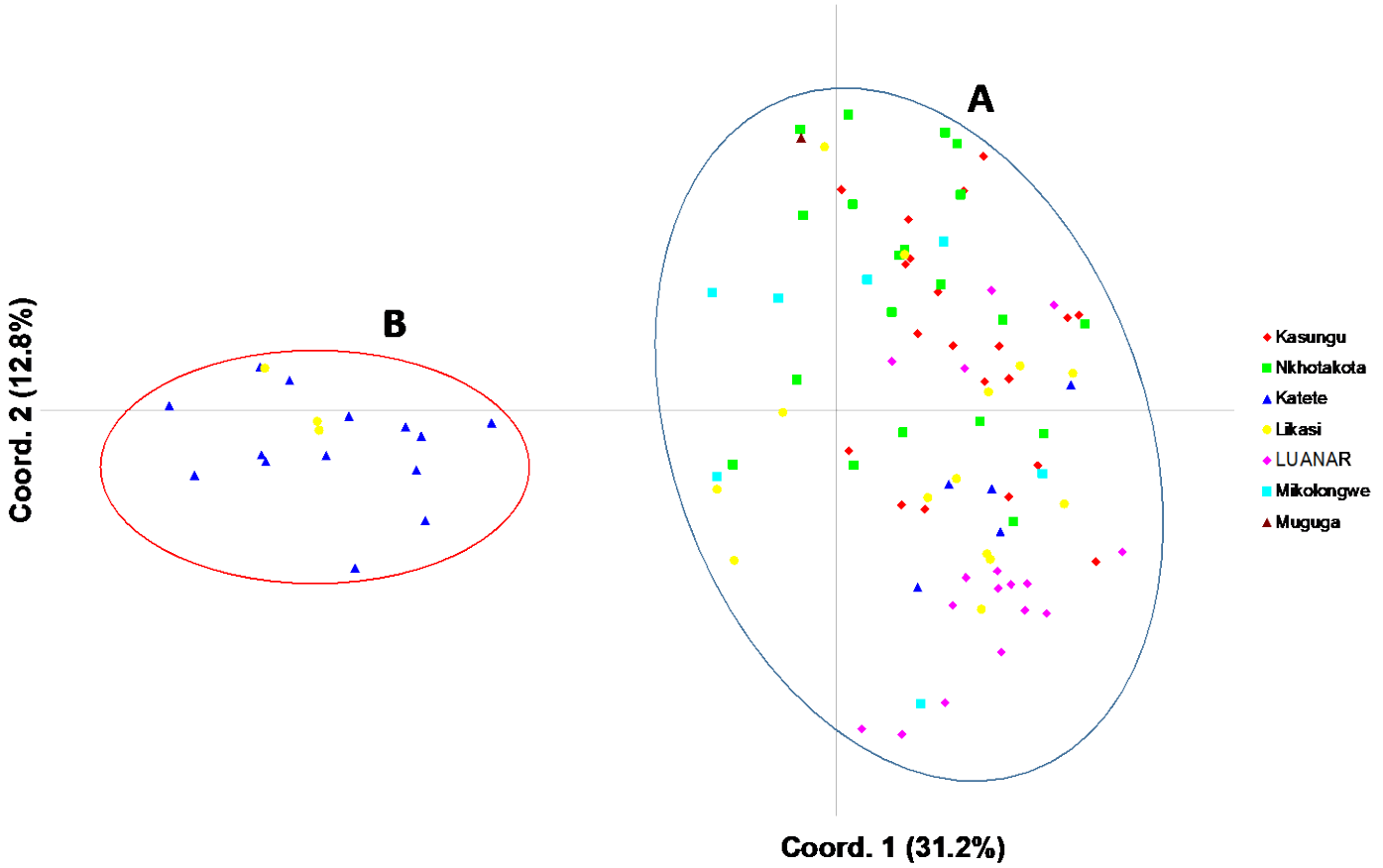


Figure 2. The allele frequencies of alleles in field populations of *Theileria parva* from Kasungu, Nkhotakota, Katete, Likasi, LUANAR and Mikolongwe and the Muguga reference laboratory strain. Shared alleles were observed at all the loci investigated. Multi-locus genotype (MLG) data was used to generate the histogram. The frequency of each predominant allele was calculated as a proportion of the total of each mark



523

524 **Figure 3**. Principal coordinate analysis (PCoA) of *Theileria parva* populations from Malawi and the Muguga reference laboratory strain. The
525 proportion of variation in the population dataset explained by each axis is shown in parentheses. The PCoA was performed using multi-locus
526 genotype data from Kasungu, Nkhotakota, Katete, Likasi, LUANAR and Mikolongwe and Muguga laboratory reference strain.