



Conservation and variation in the region of the *Theileria parva* p104 antigen coding gene used for PCR surveillance of the parasite

Isaiah Obara^{1,2} · Peris Makori³ · Kgomotso P. Sibeko⁴ · Richard P. Bishop⁵ · Ard M. Nijhof^{1,2} · Micky Mwamuye^{1,6}

Received: 17 January 2023 / Accepted: 4 April 2023 / Published online: 21 April 2023
© The Author(s) 2023

Abstract

The range of the protozoan parasite *Theileria parva*, which causes East Coast fever in cattle, has been expanding to countries where it has not previously been detected, as a result of cross-border domestic cattle movement. Countries where *T. parva* has not previously been observed until recently include Cameroon and South Sudan. This raises the issue of the conservation of the p104 antigen gene, on which the nested PCR assay that is widely used for *T. parva* surveillance in the blood of infected cattle is based. We sampled 40 isolates from six countries widely distributed across the geographical range of the parasite, including eastern, central and southern Africa, for p104 sequence polymorphism. These included parasites from both domestic cattle and the Cape buffalo (*Syncaerus caffer*) wildlife reservoir. The most frequent allelic variants were present in cattle transmissible isolates from multiple widely separated geographical regions in Zambia, Uganda, Kenya, Tanzania, Rwanda and South Africa. These frequent p104 variants were also present in the three component stocks of the Muguga cocktail used for the infection and treatment live immunisation procedure to control *T. parva* in the field. Other isolates exhibited unique alleles. This includes some of the p104 sequences from Cameroon, which is outside the known range of the *Rhipicephalus* tick vector and whose origin is therefore unclear. The nested primer oligonucleotides used to generate the amplicons were universally conserved in cattle-derived parasites and a majority of buffalo-derived isolates across the geographical range of the parasite. However, some rare South African buffalo-derived isolates exhibited one or two mismatches with the primer sequences. It therefore remains possible that some p104 alleles may be so divergent that they do not amplify with the current diagnostic primers and are not detectable in surveys, hence the need for increasing knowledge of genetic heterogeneity of diagnostic targets. There was no evidence for positive selection among those p104 mutations that resulted in residue changes. Importantly, the data indicate that the p104-based PCR detection assay should be effective across the majority of the range of *T. parva*, and if the one or two mismatches are shown in future to result in the primers annealing less efficiently, then the assay can be further improved by introduction of degenerate bases to enable amplification of the less frequent South African buffalo-derived variant p104 genes.

Keywords East Coast fever · *Theileria parva* · p104 antigen gene · Muguga cocktail

Section Editor: Abdul Jabbar

Isaiah Obara and Peris Makori have joint first authorship.

✉ Isaiah Obara
i.b.obara@gmail.com

¹ Department of Veterinary Medicine, Institute for Parasitology and Tropical Veterinary Medicine, Freie Universität Berlin, Berlin, Germany

² Department of Veterinary Medicine, Veterinary Centre for Resistance Research, Freie Universität Berlin, Berlin, Germany

³ University of Eldoret, Eldoret, Kenya

⁴ Vector and Vector-Borne Disease Research Programme, Department of Veterinary Tropical Diseases, Faculty of Veterinary Science, University of Pretoria, Pretoria, South Africa

⁵ Department of Veterinary Microbiology and Pathology, Washington State University, Pullman, WA, USA

⁶ Department of Environment and Natural Resource Management, Africa Nazarene University, Kajiado, Kenya

Introduction

The apicomplexan *Theileria parva* immortalizes bovine lymphocytes resulting in a severe disease known as East Coast fever (ECF) in susceptible cattle, particularly exotic breeds. The disease results in an estimated one million cattle deaths annually in eastern, central and southern Africa (Norval et al. 1992) and is endemic in 12 countries extending from South Sudan to southern Africa (Irvin and Morrison 1987). Furthermore, unregulated trans-boundary movements of cattle with subclinical long-term infections is thought to be responsible for the spread of *T. parva* from endemic areas north into Southern Sudan (Malak et al. 2012; Marcellino et al. 2017) and west into Cameroon (Silatsa et al. 2020). Cameroon, the central African country located along a major cattle trade route between eastern and western Africa countries, is unique in the sense that *T. parva* was recently discovered there for the first time in asymptomatic cattle in four out of the five major agro-ecological zones, yet *R. appendiculatus*, the main tick vector, has not been reported. The presence of *T. parva* infections among cattle in Cameroon and the risk of spread of the infection to additional countries, such as the Central African Republic, Southwest Ethiopia and Nigeria as a result of movements of cattle with long-term asymptomatic infections, indicate that *T. parva* can be considered an emerging transboundary pathogen.

The occurrence of *T. parva* infections in previously undocumented geographical locations has increased demand for robust testing procedures to support epidemiological studies on prevalence and distribution. The most widely used serological test utilises an indirect ELISA to detect antibodies against the polymorphic immunodominant molecule (PIM), a protein antigen expressed by *T. parva* schizonts and sporozoites (Katende et al. 1998). The interpretation of the PIM-based indirect ELISA involves expressing the test sera optical density score as a percentage of the positive control serum. However, the assay can sometimes result in false positive or inconclusive results. Consequently, validation using molecular methods often becomes necessary.

The reverse line blot (RLB) hybridisation assay is a sensitive technique for detection of tick-borne pathogens, such as *Babesia bovis* that multiply in erythrocytes, whereas for *T. parva*, RLB percentage positivity is likely to represent a minimum figure because most multiplication of *T. parva* occurs in T cells, which are several orders of magnitude less prevalent than erythrocytes in blood. Given this quantitative bias, the overall percentage of cattle infected with *T. parva*, as assessed using RLB, is typically likely to be an underestimate.

More sensitive surveillance based on detection of *T. parva* genomic DNA became possible with the development and evaluation of a nested set of primers targeting

the gene sequence of the *T. parva* 104 kDa (p104) rhoptry antigen (Odongo et al. 2010). However, currently there is a lack of information regarding p104 polymorphisms over the entire geospatial distribution of the parasite.

Theileria parva diversity has primarily been assessed using sequence polymorphism in the so-called Tp CD8 + T cell target antigen genes (Graham et al. 2006) and by analysis of a panel of variable number tandem repeat sequences (VNTRs) initially developed by Oura et al. (2003). In addition, analysis of single nucleotide polymorphisms (SNPs) through next-generation whole genome sequencing is emerging as a tool for the future (Hayashida et al. 2013; Henson et al. 2012) with the ultimate goal of performing population genomics studies. A recent study from Cameroon (Silatsa et al. 2020) surprisingly revealed seven p104 genotypes, contrasting with earlier data indicating widespread conservation in the section of the gene used for PCR-based diagnostics (Skilton et al. 2002). Some of the less frequent variants detected in the Cameroonian study had not been described previously. Only two of these variants are present in the complete genome sequences of the Muguga cocktail component stocks used for the infection and treatment immunisation (ITM), while the most frequent genotype was identical to that of the vaccine stock *T. parva* Marikebuni that was tested on several farms in the western Kenya highlands (Wanjohi et al. 2001). Given that this data indicates that additional p104 genotypes are identifiable in cattle, there is therefore a need to consider whether or not p104 polymorphism is extensive in the endemic region and can potentially impact the reliability of the PCR assay.

We report a multi-regional characterisation of the diversity present within the *T. parva* p104 gene and assess whether any genetic variation is likely to result in a discrepancy between the prevalence indicated by p104 PCR and the true presence of *T. parva* infections among cattle and buffalo in the field.

Materials and methods

T. parva isolates

The study utilized a range of *T. parva* isolates available at the Institute of Parasitology and Tropical Veterinary Medicine, Freie University Berlin and also isolates recently obtained from the field. We also analysed a further 26 near full-length p104 sequences from seven South African *T. parva* isolates obtained through a recent independent study. These isolates were selected from a collection of 111 isolates for which PIM PCR–RFLP profiles had been

determined. To maximize capture of isolate diversity, the selection focused on samples that were representative of the different PCR–RFLP profile clusters. These comprised six buffalo-derived *T. parva* isolates from Kruger National Park (KNP) in Mpumalanga province (KNPAB7, KNPB10, KNP W8, KNP Y4) and Hluhluwe-iMfolozi Park in KwaZulu-Natal province (HIP 05, HIP 19). A cattle-derived isolate from Bloemfontein, Free State Province (Bloeb) was also selected (Sibeko et al. 2011).

For all samples used in this study, their origin, the material used for DNA extraction and where available, the relevant reference for archived samples is provided in Table 1.

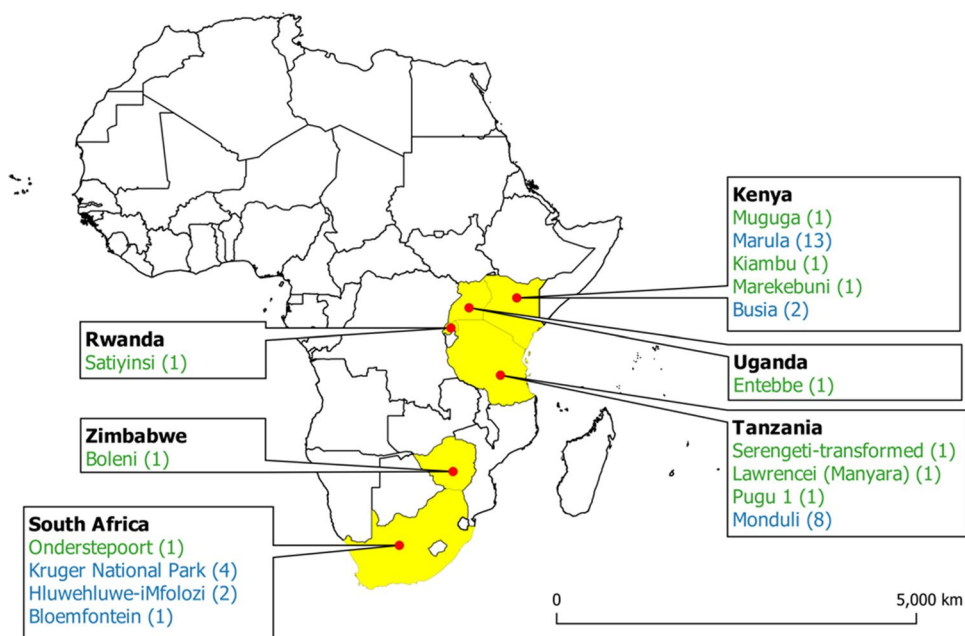
The Marula blood samples were collected as part of a field study undertaken to explore protection afforded to immunized animals exposed to *T. parva* challenge as described in Bishop et al. (2015).

Figure 1 shows the geospatial distribution of the parasite isolates analysed.

Table 1 Samples used in this study and details of their origin. Cattle blood was used for DNA extraction from all field samples

No	Sample	Origin	Material used	Reference
1	Ser — (Serengeti)	Tanzania	<i>T. parva</i> tick stabilate	Young and Purnell (1973)
2	Manyara buffalo	Tanzania	<i>T. parva</i> tick stabilate	Schreuder et al. (1977)
3	Pug — (Pugu)	Tanzania	Cell culture	Uilenberg et al. (1976)
4	M13 — Monduli	Tanzania	Blood	-
5	M14 — Monduli	Tanzania	Blood	-
6	M15 — Monduli	Tanzania	Blood	-
7	M36 — Monduli	Tanzania	Blood	-
8	M29 — Monduli	Tanzania	Blood	-
9	M16 — Monduli	Tanzania	Blood	-
10	M136 — Monduli	Tanzania	Blood	-
11	M141 — Monduli	Tanzania	Blood	-
12	Bol — (Boleni)	Zimbabwe	<i>T. parva</i> tick stabilate	Lawrence and Mackenzie (1980)
13	Entebbe	Uganda	Cell culture	Robson et al. (1978)
14	Sat — (Satinsyi)	Rwanda	<i>T. parva</i> tick stabilate	-
15	B11 — (Busia)	Kenya	Blood	-
16	B12 — (Busia)	Kenya	Blood	-
17	Mug — (Muguga)	Kenya	Tick salivary glands	Brocklesby et al. (1961)
18	Kia — (Kiambu)	Kenya	Tick salivary glands	Irvin et al. (1974)
19	Mar — (Marikibuni)	Kenya	<i>T. parva</i> tick stabilate	Irvin et al. (1983)
20	N4 — (Marula)	Kenya	Blood	-
21	M30 (Marula)	Kenya	Blood	-
22	N62 Marula	Kenya	Blood	-
23	N11 Marula	Kenya	Blood	-
24	N7 Marula	Kenya	Blood	-
25	N75 Marula	Kenya	Blood	-
26	N25 Marula	Kenya	Blood	-
27	N81 Marula	Kenya	Blood	-
28	N49 Marula	Kenya	Blood	-
29	N40 Marula	Kenya	Blood	-
30	N44 Marula	Kenya	Blood	-
31	N12 Marula	Kenya	Blood	-
32	N47 Marula	Kenya	Blood	-
33	Ond — (Onderstepoort)	South Africa	<i>T. parva</i> tick stabilate	Neitz (1948)
34	KNPAB7 - Kruger National park	South Africa	Blood	Sibeko et al. (2011)
35	KNPB10 - Kruger National park	South Africa	Blood	Sibeko et al. (2011)
36	KNP W8 - Kruger National park	South Africa	Blood	Sibeko et al. (2011)
37	KNP Y4 - Kruger National park	South Africa	Blood	Sibeko et al. (2011)
38	HIP 05 - Hluhluwe-iMfolozi Park	South Africa	Blood	Sibeko et al. (2011)
39	HIP19 - Hluhluwe-iMfolozi Park	South Africa	Blood	Sibeko et al. (2011)
40	Bloeb - Bloemfontein	South Africa	Blood	Sibeko et al. (2011)

Fig. 1 Map showing origin of *T. parva* samples used in this study. The number of isolates from each region analysed is shown in brackets. The blue font represents field samples



Genomic DNA extraction and amplification of the *T. parva* p104 gene

Genomic DNA was extracted from blood, frozen ground-up tick supernatants (GUTS), salivary glands and cell cultures using the Qiagen DNeasy Blood and Tissue Kit according to the manufacturer's instructions and used as template for amplification of the *T. parva* p104 gene as previously described (Skilton et al. 2002).

Briefly, the use of oligonucleotide primers IL3231 5'-ATTTAA GGAACCTGACGTGACTGC-3' (forward) and IL755 5'-TAAGAT GCCGACTATTAATGACACC-3' (reverse) permitted the amplification of a 496-bp fragment of the p104 gene. The amplifications were carried out using the Phusion High-Fidelity PCR reagents (Thermo Scientific) in 50 μ L reaction volumes composed of 20 ng of DNA, 1 \times Phusion HF buffer, 200 μ M of each dNTP, 0.5 μ M of forward and reverse primer and 1 unit of Phusion DNA polymerase. The PCR cycling conditions were 94 $^{\circ}$ C for 1 min, 40 cycles of 1 min at 94 $^{\circ}$ C, 1 min at 60 $^{\circ}$ C and 1 min at 72 $^{\circ}$ C, followed by a final elongation period of 2 min at 72 $^{\circ}$ C. Known positive and negative controls were included in each run. The amplification products were visualized by UV trans-illumination in 1.5% TAE-agarose gels stained with GRGreen (Excellgene, Monthey, Switzerland).

Analysis of p104 polymorphisms

The amplicons were purified using the Zymoclean Gel DNA Recovery Kit before bidirectional Sanger sequencing. Editing and assembly of chromatograms for downstream analyses (variant calling and assessing evidence of selection) were accomplished using Geneious Prime 2021.

Assessing evidence of selection in the *T. parva* p104 gene

We used CODEML from the PAML4 package (Yang 2007) to evaluate if there were positions in the *T. parva* p104 gene encoding residues that exhibited an excess of non-synonymous (dN) relative to synonymous substitutions (dS) in order to see whether there was evidence for positive selection for amino acid substitutions. Since models of nucleotide substitution can bias the accuracy of the phylogenetic inference, we utilized both the likelihood scores and estimated model parameters for Akaike information criterion (AIC) selection of the best-fit model from the candidate set of nucleotide substitution models. The following models were evaluated: M1a — two discrete categories, one for purifying selection where $\omega < 1$, and the other for neutral selection where $\omega = 1$ (Nielsen and Yang 1998; Yang et al. 2005); M2a — an extension of M1a model, with an additional category for positive selection where $\omega > 1$ (Nielsen and Yang 1998; Yang et al. 2005); M7 — a continuous beta distribution of ω restricted to the interval (0;1), no positive selection allowed (Yang et al. 2000); and M8 — extension of M7 model, with additional, discrete category for positive selection (Yang et al. 2000). Akaike weights were used to evaluate model fit, and whenever the best-fit model is M2a or M8, sites under positive selection are determined through the Bayes empirical Bayes (BEB) approach.

To construct the input tree files used to run CodeML, we selected the best-fitting nucleotide substitution model based on the Akaike information criterion corrected for small sample size (AICc). The parameters included the

following: nucleotide substitution rate parameters, equal or unequal base frequencies (+F), a proportion of invariable sites (+I) and rate variation among sites (+G). Likelihood calculations for the nucleotide substitution models were performed with PhyML_3.0_linux6 (Guindon et al. 2010). All the model evaluation steps described above were implemented in jModelTest 2.1.10 (Darriba et al. 2012). Maximum-likelihood tree-search algorithms were implemented in PAUP 4.0 beta version using the parameter estimates for the best-fit model identified as described above (Swofford 2003).

Results and discussion

p104 gene sequence variation from multiple geographically separated regions revealed limited polymorphism which does not correlate with geographic origin

The p104 antigen is the most frequently used target gene for *T. parva* detection and surveillance. The assay is based on a nested PCR procedure designed to provide enhanced sensitivity for detection of *T. parva* genomic DNA in cattle and buffalo blood, using an outer primer pair for the first amplification, and an additional set of inner primers for the second amplification. In the present study, the outer primers and DNA extracted from a range of *T. parva* isolates ($n=40$) obtained from multiple geographically separated areas in the endemic region were used for the amplification of the p104 gene. This study revealed 36 p104 genotypes with a mean pairwise identity at the nucleotide sequence level of 97.3%.

Although the frequently observed p104 variants were present in the three component stocks of the Muguga cocktail used for the ITM live immunisation procedure, the data is consistent with the possibility that parasites with similar p104 variants to Muguga cocktail have been circulating

naturally in the field, rather than disseminated following ITM deployment. This is because the most frequent p104 alleles were also present in isolates from multiple geographically widely separated regions in Zambia, Uganda, Kenya, Tanzania, Rwanda and South Africa (Table 2), whereas the Muguga cocktail version of ITM has only been deployed extensively in Tanzania and Kenya. Only small-scale vaccination trials using the Muguga cocktail have been undertaken in Uganda. Vaccination campaigns in Zambia, use locally isolated stocks, which provides a successful alternative to the use of the trivalent Muguga cocktail version of ITM and the frequent p104 genotypes also match those present in the Chitongo and Katete vaccine stocks (Table 2). The most frequent p104 variants also match those in the Marikebuni vaccine developed by the National Veterinary Research Institute in Kenya, which has not been sufficiently widely used for ITM in the field to provide support for the vaccination dissemination hypothesis. It is also relevant to note that analysis of variation in the *T. parva* sporozoite surface antigen p67 in South African field isolates has also shown that alleles similar to those of the *T. parva* Muguga stock were circulating naturally in the field in KwaZulu-Natal Province (Sibeko et al 2011).

Previous molecular epidemiology and population genetics studies using polymorphic variable number tandem repeat (VNTR) sequences and nucleotide sequencing of genes that are targets of bovine CD8+ T cell responses have also demonstrated that parasite genetic diversity does not correlate with geographic origin (reviewed in Bishop et al. 2020). However, it is also important to note that in the present study, the analysis of all sequences highlights that South African buffalo-derived *T. parva* have p104 alleles that are distinct from other buffalo- and cattle-derived *T. parva*. This finding is consistent with genome and other antigen gene sequence data (Maboko et al. 2021; Sibeko et al. 2011) and is discussed in the next section with respect to single-base mismatches within the p104 primer target sequences.

Table 2 p104 alleles, frequency and sharing among the isolates used in this study

Accession	Allele frequency (No. of isolates)	Isolates sharing allele
MZ798150	12	Chitongo (Zambia), Marikebuni (Kenya), Iganga01 (Uganda), Budako02 (Uganda), Gomba02 (Uganda), Katete (Zambia), Satinsyi (Rwanda), Kiambu (Kenya), Busia (Kenya), Marula — 25, 30, 47(Kenya)
MZ798149	10	Muguga (Kenya), Onderstepoort (South Africa), Serengeti (Tanzania), Monduli — 13,15,16,29,36,136,141 (Tanzania)
MZ798151	3	Koupa-Matapit (Cameroon), Pugu (Tanzania), Manyara (Tanzania buffalo)
MZ798152	2	BuffaloZ5E5 (Zambia), Marula 12 (Kenya)
MZ798153	1	Marula 75 (Kenya)
MZ798154	1	Marula 7 (Kenya)
MZ798155	1	Marula 44 (Kenya)
MZ798156	1	Marula 11 (Kenya)
MZ798157	1	Marula 40 (Kenya)
MZ798158	1	Boleni (Zimbabwe)

As regards sequence variation across the 496-bp amplicon, the mean number of bp differences between the genotypes was 6 (± 2.2 SD; range = 2–10; Median = 5). This limited variation is illustrated by the box plot in Fig. 2 that shows the distribution of base pair differences among the p104 alleles in the isolates studied herein.

Single-base mismatches within the primer target sequences in rare South African buffalo-derived *T. parva* isolates

The African buffalo is the primary host for *T. parva* and is distributed across sub-Saharan Africa throughout the region where *Rhipicephalus appendiculatus*, the main tick vector of *T. parva*, occurs, but in contrast to the deadly disease suffered by cattle, buffalo remain asymptomatic. However, buffalo represent a reservoir of *T. parva* for transmission to cattle and are therefore an important consideration in assessing the extent of conservation of the *T. parva* p104 antigen gene primers used for PCR surveillance of the parasite.

Studies of genotypic diversity of *T. parva* populations in cattle and buffalo have shown that buffalo harbour parasites of much greater diversity than those in cattle and are almost always infected with multiple strains. The known diversity of *T. parva* in buffalo (reviewed by Bishop et al. 2004) and the problems with breakthroughs in ITM-vaccinated animals in areas with a cattle-buffalo interface support the suggestions that *T. parva* strains transmissible between cattle represent a subset of the overall *T. parva* population found in buffalo.

Recent studies examining the genome and antigen gene sequences of South African buffalo-derived *T. parva* have drawn the important conclusion that they are distinct from other buffalo- and cattle-derived *T. parva* (Maboko et al. 2021; Sibeko et al. 2011). Since surveillance in areas where buffalo are present in South Africa will require assays that

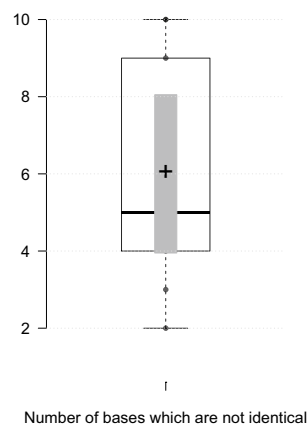
are effective for the range of genetic variants, we have used 26 near full-length p104 sequences recently generated from seven samples representative of South African buffalo-derived field isolates (Sibeko et al. 2011), to assess the extent of conservation of the primers used for PCR surveillance. The analysis shows that the p104 primers target sequences are relatively well conserved in South African buffalo-derived isolates as no mismatches were detected in the most frequently observed p104 variant. However, multiple alignments (Fig. 3) showed one or two mismatches within the primer target sequences in some of the rare South African buffalo-derived isolates.

This may still be a rather limited picture, and the impact of the one or two mismatches in the primer target sequences on the performance of the p104 assay in the detection of rare variant, so far identified only in Cape buffalo, remains to be determined. One possibility is that the mismatches will have little effect on assay performance, and similar detection efficiencies will be observed. It may also be the case that primers may anneal less efficiently. In the latter case, the assay can potentially be improved to accommodate increasing knowledge of genetic heterogeneity of diagnostic targets by introduction of degenerate bases to allow effective amplification.

The phylogenetic relatedness inferred from the p104 sequences does not correlate with the geographical origin of the isolates

Phylogenetic inference from the p104 sequence data was based on a maximum-likelihood algorithm. Our dataset for phylogenetic inference composed of the following: (i) p104 allelic variants derived from the isolates genotyped herein, (ii) eight p104 genotypes recently described in Cameroon. Two hypotheses have been advanced about the presence of *T. parva* infections among cattle in Cameroon in the absence of the main tick vector *Rhipicephalus appendiculatus*. The first is the importation of cattle that are naturally infected or have been ITM vaccinated since Cameroon is located on a major cattle trade route between eastern and western Africa countries and therefore connected to ECF endemic countries. This hypothesis is strengthened by similarity in the amino acid sequences of *T. parva* p104 and Tp1 between Cameroonian, East African and live vaccine stocks in the case of some, but not all, Cameroonian isolates. The second hypothesis, which is considered less likely, is that *T. parva* may be actively transmitted in the field by other tick vectors, since the principal field vector has not been reported in Cameroon. It is important to emphasize that these hypotheses are speculations, and further longitudinal studies are needed for a better understanding of *T. parva* infections in Cameroon; (iii) 26 p104 sequences were generated

Fig. 2 Base pair differences among ten p104 genotypes. Centre lines show the medians; box limits indicate the 25th and 75th percentiles as determined by R software; whiskers extend 1.5 times the interquartile range from the 25th and 75th percentiles; outliers are represented by dots; crosses represent sample means; bars indicate 95% confidence intervals of the means. $n = 10$ sample points



	Outer forward primer	Nested forward primer	Nested Reverse primer	Outer Reverse primer
KNPW8 Cl8(11)	ATTTAAGGAACCTGACGTGACTGC	GGCCAAGGTCTCCTTCAGAATACG	G CAGATGACGAGGAAACACACCCA	GGTGT CATT AATAGTCGGCATCTAA
KNPY4 Cl1(2)G.....
KNPY4 Cl2(1)C.....
KNPY4 Cl4(1)G.....
KNPY4 Cl6(1)G.....
KNPY4 Cl8(1)T.....	..T.....G.C.....
KNPAB7 Cl8(1)G.....
KNPAB7 Cl5(1)T.....G.....
KNPAB7 Cl1(1)	G.....T.....
KNPAB7 Cl3(1)A.....T.....G.....
KNPAB7 Cl2(1)
HIP19 Cl1(1)T.....
HIP19 Cl2(1)T.....C.....
KNPB10 Cl3(1)T.....
KNPB10 Cl4(1)T.....
BloeB Cl1(1)

Fig. 3 Primer matches to South African buffalo-derived *T. parva* p104 sequences

from a representative sample of South African buffalo and cattle-derived field isolates. These representative isolates were selected from a collection of 111 isolates taking into account the PIM and PCR–RFLP profiles; (iv) reference p104 sequences were distributed across the geographical range of the parasite available in the GenBank. The AIC ranking of the models resulted in the most support for the K80 + I model (Akaike weight = 0.239072). The likelihood scores, model selection criteria and numerical values for model parameters are summarized in Table 3, and the resulting tree is shown in Fig. 4.

It is relevant to note that our understanding of the molecular epidemiology and population genetics of *T. parva* has advanced significantly in recent years (reviewed in Bishop et al. 2020). In particular, evidence is accumulating, based on polymorphic variable number tandem repeat (VNTR) sequences and nucleotide sequencing of genes that

are targets of the CD8 + T cell responses from multiple geographically separated regions, that parasite diversity does not correlate with geographic origin.

Next, we used CODEML from the PAML4 package to evaluate if there are positions in the p104 sequences coding for residues that show an excess of non-synonymous (dN) over synonymous substitutions (dS). dN/dS (ω) > 1 is the hallmark of positive selection for amino acid substitutions. Models that were fitted to the data included the following: M1a — two discrete categories, one for purifying selection, where $\omega < 1$, and the other for neutral selection where $\omega = 1$; M2a — an extension of M1a model, with an additional category for positive selection, where $\omega > 1$; M7 — a continuous beta distribution of ω restricted to the interval (0; 1), no positive selection allowed; and M8 — extension of M7 model, with additional, discrete category for positive selection.

Table 3 AICc ranking of a candidate set of nucleotide substitution models and parameter estimates for the model with most support

Model	–lnL	K	AICc	Delta	Weight	cumWeight
K80 + I	595.59456	68	1371.874834	0.000000	0.239072	0.239072
HKY + I	590.89823	71	1373.187764	1.312930	0.124002	0.363075
JC + I	598.37941	67	1373.943654	2.068820	0.084975	0.448050
TPM1 + I	595.10147	69	1374.423036	2.548201	0.066864	0.514914
TrNef + I	595.20330	69	1374.626696	2.751861	0.060391	0.575305
TPM2 + I	595.29533	69	1374.810756	2.935921	0.055081	0.630386
F81 + I	593.59783	70	1374.984122	3.109287	0.050508	0.680893
TPM3 + I	595.52627	69	1375.272636	3.397801	0.043723	0.724616
TPM2uf + I	590.27243	72	1375.573986	3.699152	0.037607	0.762223
TrN + I	590.31083	72	1375.650786	3.775952	0.036190	0.798413

–lnL negative log likelihood, K number of estimated parameters, AICc corrected Akaike information criterion, delta AICc difference, Weight AICc weight, cumWeight cumulative AICc weight



Fig. 4 Maximum likelihood tree depicting the phylogenetic relationships of *Theileria parva* p104 sequences. The dataset includes (i) p104 allelic variants derived from the isolates investigated in the present study; (ii) p104 sequences from earlier field studies widely distributed across the geographical range of the parasite, includ-

ing eastern central and southern Africa. These included parasites of both cattle and Cape buffalo wildlife reservoir origin. Numbers on the branches refer to bootstrap percentages above 50% based on 1000 replicates with *Theileria equi* p104 gene used as an outgroup

Likelihood ratio tests were used to evaluate model support. This analysis did not reveal any sites with higher rates of non-synonymous to synonymous nucleotide substitutions than expected under neutral evolution. Both M2a and M8 models used similar numbers of parameters and the same sites under selection were identified, albeit, not supported by the BEB analysis as depicted by < 95% posterior probabilities of positive selection. A summary of the selection results is shown in Table 4.

Collectively, the data shows that polymorphisms in the p104 gene are unlikely to impact the PCR procedure for detection of *T. parva* in cattle and buffalo in eastern and southern Africa, since despite the existence of greater polymorphism than was previously realized within the amplicon, the primers are typically conserved. Any potential problems related to primers annealing less efficiently to the apparently rare South African buffalo-derived isolates can potentially be remedied by introduction of

Table 4 Summary of selection results showing model parameter estimates and positively selected sites under the M2a and M8 models using Bayes empirical Bayes (BEB) analysis

Model	K	LnL	Parameter estimates				Positive sites
			ω	p0	p1	p2	
M2a	68	-345.2150	2.25648	0.74700	0.00000	0.25300	46 A(0.633), 51 A(0.516), 52 S(0.581)
M8	68	-345.2150	2.25648	0.74700	0.00500	0.25300	46 A(0.752), 51 A(0.648), 52 S(0.707)

K indicates number of estimated parameters; LnL refers to maximized log likelihoods; p0 (purifying), p1 (neutral) and p2 (positive) denote the proportion of codons belonging to each site class, while ω represents the dN/dS for the positive selection site class only. The brackets () enclose the posterior probabilities of positive selection. The codon numbers are identified with reference to sequence AB739681

degenerate bases at one or two sites within each primer. It is also important to emphasize that increasing knowledge of genetic heterogeneity of diagnostic targets is an indisputably important element of molecular diagnostics and widening out into other populations remains a priority for future research.

Acknowledgements We acknowledge Prof. David Odongo and Prof. Paul Gwakisa for providing some of the samples used in this study.

Author contribution Conceptualisation: Micky Mwamuye, Isaiah Obara and Richard P. Bishop; methodology: Micky Mwamuye and Peris Makori; validation: Micky Mwamuye and Isaiah Obara; formal analysis: Isaiah Obara; investigation: Micky Mwamuye and Peris Makori; resources: Ard Nijhof; data curation: Micky Mwamuye, Kgomotso P. Sibeko-Matjila and Isaiah Obara; writing—original draft preparation, Peris Makori; writing—review and editing: Micky Mwamuye, Kgomotso P. Sibeko-Matjila, Isaiah Obara, Ard Nijhof, Richard P. Bishop; visualisation: Micky Mwamuye; funding acquisition: Ard Nijhof and Isaiah Obara.

Funding Open Access funding enabled and organized by Projekt DEAL. Part of the work was supported by the Deutsche Forschungsgemeinschaft (DFG) (Project No. CL166/4–2). Isaiah Obara was supported by the DFG grant OB 490/2–1. Micky Mwamuye was supported by a doctoral scholarship of the Kenyan government (National Research Fund—NRF) and the Germany Academic Exchange Service (DAAD).

Data availability The p104 sequences generated from this study are deposited in NCBI database under accession numbers: MZ798149–MZ798158.

Declarations

Ethics approval The archived biological material utilized in this study was previously obtained in compliance with the Animal Ethics research guidelines of the institutions involved and have been used in previously published studies. As for the field samples, the cattle from Kenya and Tanzania were sampled according to standard operating procedures approved by the Kenya National Commission for Science, Technology and Innovation (NACOSTI) and Sokoine University, respectively.

Consent to participate Not applicable.

Consent for publication All authors have read and approved the final version of this paper for publication.

Conflict of interest The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Bishop R, Musoke A, Morzaria S, Gardner M, Nene V (2004) *Theileria*: intracellular protozoan parasites of wild and domestic ruminants transmitted by ixodid ticks. *Parasitology* 129 Suppl:S271–83. <https://doi.org/10.1017/s0031182003004748>
- Bishop R, Hemmink JD, Morrison WI, Weir W, Toye PG, Sitt T, Spooner PR, Musoke AJ, Skilton RA, Odongo DO (2015) The African buffalo parasite *Theileria. sp.* (buffalo) can infect and immortalize cattle leukocytes and encodes divergent orthologues of *Theileria parva* antigen genes. *Int J Parasitol: Parasites Wildl* 4(3):333–342. <https://doi.org/10.1016/j.ijppaw.2015.08.006>
- Bishop RP, Odongo D, Ahmed J, Mwamuye M, Fry LM, Knowles DP, Nanteza A, Lubega G, Gwakisa P, Clausen PH, Obara I (2020) A review of recent research on *Theileria parva*: implications for the infection and treatment vaccination method for control of East Coast fever. *Transbound Emerg Dis* 67 Suppl 1:56–67. <https://doi.org/10.1111/tbed.13325>
- Brocklesby DW, Barnett SF, Scott GR (1961) Morbidity and mortality rates in east coast fever (*Theileria parva* infection) and their application to drug screening procedures. *Br Vet J* 117(12):529–531. [https://doi.org/10.1016/s0007-1935\(17\)43303-3](https://doi.org/10.1016/s0007-1935(17)43303-3)
- Darriba D, Taboada GL, Doallo RR, Posada D (2012) JModelTest 2: more models, new heuristics and parallel computing. *Nat Methods* 9(8):772. <https://doi.org/10.1038/nmeth.2109>
- Graham SP, Pellé R, Honda Y, Mwangi DM, Tonukari NJ, Yamage M, Glew EJ, De Villiers EP, Shah T, Bishop R, Abuya E, Awino E, Gachanja J, Luyai AE, Mbwika F, Muthiani AM, Ndegwa DM, Njahira M, Nyanjui JK, ... Taracha ELN (2006) *Theileria parva* candidate vaccine antigens recognized by immune bovine cytotoxic T lymphocytes. *Proc Natl Acad Sci USA* 103(9):3286–3291. <https://doi.org/10.1073/pnas.0511273103>
- Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O (2010) New algorithms and methods to estimate

- maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* 59(3):307–321. <https://doi.org/10.1093/sysbio/syq010>
- Hayashida K, Abe T, Weir W, Nakao R, Ito K, Kajino K, Suzuki Y, Jongejan F, Geysen D, Sugimoto C (2013) Whole-genome sequencing of *Theileria parva* strains provides insight into parasite migration and diversification in the African continent. *DNA Res* 20(3):209–220. <https://doi.org/10.1093/dnares/dst003>
- Henson S, Bishop RP, Morzaria S, Spooner PR, Pelle R, Poveda L, Ebeling M, Küng E, Certa U, Daubenberger CA, Qi W (2012) High-resolution genotyping and mapping of recombination and gene conversion in the protozoan *Theileria parva* using whole genome sequencing. *BMC Genomics* 13(1):503. <https://doi.org/10.1186/1471-2164-13-503>
- Irvin AD, Morrison WI (1987) Immunopathology, immunology and immunoprophylaxis of *Theileria* infections. In: Soulsby EJL (ed) Immune responses in parasitic infections: immunology, immunopathology, and immunoprophylaxis. Vol. III: Protozoa. CRC Press, Boca Raton, FL, pp 223–274
- Irvin AD, Purnell RE, Brown CGD, Cunningham MP, Ledger MA, Payne RC (1974) The application of an indirect method of infecting ticks with piroplasmids for use in the isolation of field infections. *Br Vet J* 130(3):280–287. [https://doi.org/10.1016/S0007-1935\(17\)35894-3](https://doi.org/10.1016/S0007-1935(17)35894-3)
- Irvin AD, Dobbelaere DA, Mwamachi DM, Minami T, Spooner PR, Ocamo JG (1983) Immunisation against East Coast fever: correlation between monoclonal antibody profiles of *Theileria parva* stocks and cross immunity in vivo. *Res Vet Sci* 35(3):341–346. [https://doi.org/10.1016/s0034-5288\(18\)32030-7](https://doi.org/10.1016/s0034-5288(18)32030-7)
- Katende J, Morzaria S, Toye P, Skilton R, Nene V, Nkongwe C, Musoke A (1998) An enzyme-linked immunosorbent assay for detection of *Theileria parva* antibodies in cattle using a recombinant polymorphic immunodominant molecule. *Parasitol Res* 84(5):408–416. <https://doi.org/10.1007/s004360050419>
- Lawrence J, Mackenzie PK (1980) Isolation of a non-pathogenic *Theileria* of cattle transmitted by *Rhipicephalus appendiculatus*. *Zimb Vet J* 11:27–35
- Maboko BB, Sibeko-Matjila KP, Pierneef R, Chan WY, Josemans A, Marumo RD, Mbizeni S, Latif AA, Mans BJ (2021) South African Buffalo-derived *Theileria parva* is distinct from other Buffalo and cattle-derived *T. parva*. *Front Genet* 12:666096. <https://doi.org/10.3389/fgene.2021.666096>
- Malak AK, Mpoke L, Banak J, Muriuki S, Skilton RA, Odongo D, Sunter J, Kiara H (2012) Prevalence of livestock diseases and their impact on livelihoods in Central Equatoria State, southern Sudan. *Prev Vet Med* 104(3–4):216–223. <https://doi.org/10.1016/j.prevetmed.2011.12.001>
- Marcellino WL, Salih DA, Njahira MN, Ndiwa N, Araba A, El Hussein AM, Seitzer U, Ahmed JS, Bishop RP, Skilton RA (2017) The emergence of *Theileria parva* in Jonglei State, South Sudan: confirmation using molecular and serological diagnostic tools. *Transbound Emerg Dis* 64(4):1229–1235. <https://doi.org/10.1111/tbed.12495>
- Neitz WO (1948) Studies on East Coast fever. *S.Afr. Sci* 1:133–135
- Nielsen R, Yang Z (1998) Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* 148(3):929–936. <https://doi.org/10.1093/genetics/148.3.929>
- Norval RAI, Perry BD, Young A (1992) The epidemiology of theileriosis in Africa. Academic Press
- Odongo DO, Sunter JJD, Kiara H, Skilton RRA, Bishop RPR (2010) A nested PCR assay exhibits enhanced sensitivity for detection of *Theileria parva* infections in bovine blood samples from carrier animals. *Parasitol Res* 106(2):357–365. <https://doi.org/10.1007/s00436-009-1670-z>
- Oura CAL, Odongo DO, Lubega GW, Spooner PR, Tait A, Bishop RP (2003) A panel of microsatellite and minisatellite markers for the characterisation of field isolates of *Theileria parva*. *Int J Parasitol* 33(14):1641–1653. [https://doi.org/10.1016/S0020-7519\(03\)00280-7](https://doi.org/10.1016/S0020-7519(03)00280-7)
- Robson J, Pedersen V, Odeke GM, Kamya EP, Brown CGD (1978) East coast fever immunisation trials in Uganda: field exposure of zebu cattle immunised with three isolates of *Theileria parva*. *Trop Anim Health Prod* 10(1):62. <https://doi.org/10.1007/BF02235307>
- Schreuder BE, Uilenberg G, Tondeur W (1977) Studies on Theileriidae (Sporozoa) in Tanzania. VIII. Experiments with African buffalo (*Syncerus caffer*). *Tropenmedizin Und Parasitologie* 28(3):367–371. <https://europepmc.org/article/med/410128>
- Sibeko KP, Collins NE, Oosthuizen MC, Troskie M, Potgieter FT, Coetzer JAW, 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* 181:120–130. <https://doi.org/10.1016/j.vetpar.2011.04.035>
- Silatsa BA, Simo G, Githaka N, Kamga R, Oumarou F, Keambou Tiambo C, Machuka E, Domelevo JB, Odongo D, Bishop R, Kuate JR, Njiokou F, Djikeng A, Pelle R (2020) First detection of *Theileria parva* in cattle from Cameroon in the absence of the main tick vector *Rhipicephalus appendiculatus*. *Transbound Emerg Dis* 67(S1):68–78. <https://doi.org/10.1111/tbed.13425>
- Skilton RA, Bishop RP, Katende JM, Mwaura S, Morzaria SP (2002) The persistence of *Theileria parva* infection in cattle immunized using two stocks which differ in their ability to induce a carrier state: analysis using a novel blood spot PCR assay. *Parasitology* 124(3):265–276. <https://doi.org/10.1017/S0031182001001196>
- Swofford D (2003) PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4. In: Sinauer Associates, Sunderland, Massachusetts. Sinauer Associates, Sunderland
- Uilenberg G, Schreuder BE, Silayo RS, Mpangala C (1976) Studies on Theileriidae (Sporozoa) in Tanzania. IV. A field trial on immunization against East Coast fever (*Theileria parva* infection of cattle). *Tropenmedizin Und Parasitologie* 27(3):329–336. <https://www.ncbi.nlm.nih.gov/pubmed/982549>
- Wanjohi JM, Ngeranwa JN, Rumberia RM, Muraguri GR, Mbogo S K (2001) Immunization of cattle against East Coast fever using *Theileria parva* (Marikibuni) and relaxation of tick control in North Rift, Kenya. In *Kenya. Onderstepoort Journal of Veterinary Research* (Vol. 68). Published jointly by the Agricultural Research Council, ARC-Onderstepoort Veterinary Institute and the Faculty of Veterinary Science, University of Pretoria. <https://repository.up.ac.za/handle/2263/18947>
- Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24(8):1586–1591. <https://doi.org/10.1093/molbev/msm088>
- Yang Z, Wong WSW, Nielsen R (2005) Bayes empirical Bayes inference of amino acid sites under positive selection. *Mol Biol Evol* 22(4):1107–1118. <https://doi.org/10.1093/molbev/msi097>
- Yang Z, Nielsen R, Goldman N, Pedersen AM (2000) Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* 155(1):431–449. <https://doi.org/10.1093/genetics/155.1.431>
- Young AS, Purnell RE (1973) Transmission of *Theileria lawrencei* (Serengeti) by the ixodid tick, *Rhipicephalus appendiculatus*. *Trop Anim Health Prod* 5(3):146–152. <https://doi.org/10.1007/BF02251383>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.