

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

^{1*} Kgomotso P Sibeko, ¹ Nicola E Collins, ¹Marinda C Oosthuizen, ¹Milana Troskie, ²Frederick T Potgieter, ¹Jacobus A W Coetzer, ³Dirk Geysen

¹ Department of Veterinary Tropical Diseases, Faculty of Veterinary Science, University of Pretoria, Private Bag X04, Onderstepoort, 0110, South Africa

² Agricultural Research Council-Onderstepoort Veterinary Institute, Private Bag X5, Onderstepoort, 0110, South Africa

³ Department of Animal Health, Institute of Tropical Medicine, 155 Nationalestraat, Antwerp B-2000, Belgium

*** Corresponding author:**

E-mail address: kgomotso.sibeko@up.ac.za

Postal address: Department of Veterinary Tropical Diseases, Faculty of Veterinary Science, University of Pretoria, Private Bag X04, Onderstepoort, 0110, South Africa

Telephone number: +27 (0) 12 529 8402

Fax number: +27 (0) 12 529 8312

Abstract

Restriction fragment length polymorphism analysis of PCR products (PCR-RFLP) and sequencing of the variable region of the p104 and PIM genes was performed on samples obtained from South African *T. parva* parasites originating from cattle on farms with suspected theileriosis and from buffalo. p104 and PIM PCR-RFLP profiles similar to those of the *T. parva* Muguga stock, an isolate that causes ECF in Kenya, were obtained from three of seven cattle samples collected on a farm near Ladysmith in KwaZulu-Natal Province. Amino acid sequences of the p104 and PIM genes from two of these samples were almost identical to the *T. parva* Muguga p104 and PIM sequences. This result supports findings from a recent p67 study in which p67 alleles similar to those of the *T. parva* Muguga stock were identified from the same samples. While these results suggest the presence of a cattle-derived *T. parva* parasite, reports of cattle-to-cattle transmission could not be substantiated and ECF was not diagnosed on this farm. Although extensive diversity of p104 and PIM gene sequences from South African *T. parva* isolates was demonstrated, no sequences identical to known cattle-type p104 and PIM alleles were identified from any of the buffalo *T. parva* samples analyzed. 'Mixed' PIM alleles containing both cattle- and buffalo-type amino acid motifs were identified for the first time, and there appeared to be selection of cattle-type and 'mixed'-type PIM sequences in the cattle samples examined.

Keywords: *Theileria parva*, p104, polymorphic immunodominant molecule (PIM), characterization, Corridor disease, East Coast fever

1. Introduction

Cattle theileriosis caused by *Theileria parva* is a disease of major economic importance in eastern, central and parts of southern Africa (Young *et al.*, 1988). Infections of cattle by *T. parva* parasites result in three recognized disease syndromes, East Coast fever (ECF), January disease and Corridor disease. The natural tick vectors of *T. parva* in South Africa are *Rhipicephalus appendiculatus* and *R. zambeziensis* (Lawrence *et al.*, 1983), and, in the presence of *T. parva*-infected buffalo, the vector ticks can transmit the parasite to naïve cattle, resulting in Corridor disease outbreaks. Corridor disease remains important in South Africa where it is a controlled disease. No cases of ECF have been reported in South Africa since its eradication in 1955 (Anonymous, 1981). It is not known whether ECF-causing *T. parva* parasites were transmitted to South African buffalo during the ECF epidemic, or if there exists in buffalo an ancestral subpopulation of *T. parva* parasites that could adapt to cattle.

In a recent study, size differentiation and sequence variation of the central region of the *T. parva* antigen gene, p67, were used to characterize South African *T. parva* field samples (Sibeko *et al.*, 2010). A p67 allele (allele 1) identical to that of Muguga, a *T. parva* stock from Kenya which causes ECF, was obtained from cattle samples collected on a farm near the town Ladysmith in KwaZulu-Natal Province, where cattle-to-cattle transmission of *T. parva* was suspected (Thompson *et al.*, 2008). This finding is of concern to the cattle industry in South Africa, as Potgieter *et al.* (1988) showed that buffalo-derived *T. parva* parasites, causing Corridor disease, can be maintained by passage between cattle and the tick vector. The persistence of *T. parva* infections in cattle in South Africa could eventually result in the selection of *T. parva* parasites adapted to cattle. It remains a concern that ECF could re-emerge and therefore a serious need exists to establish if there are cattle-type *T. parva* parasites in buffalo in South Africa.

Restriction fragment length polymorphism analyses of PCR products (PCR-RFLP) of *T. parva* antigen genes, PIM, p104, p150 and p67, have been used for characterization of *T. parva* stocks (Geysen *et al.*, 1999; Bishop *et al.*, 2001). The extensively characterized *T. parva* PIM (Baylis *et al.*, 1993; Toye *et al.*, 1996) is encoded by a single copy gene and its structure consists of a central variable region, flanked by conserved 5' and 3' termini (Toye *et al.*, 1995a; 1995b; Geysen *et al.*, 2004). Previous studies have identified a number of characteristics which could be used to distinguish PIM sequences from cattle-derived *T. parva* isolates from those from buffalo-derived isolates (Toye *et al.*, 1995a; Geysen *et al.*, 2004). More recently, mini- and micro-satellite markers

have been developed which enable detection of higher levels of polymorphism in *T. parva* stocks than PCR-RFLP methods (Oura *et al.*, 2003; 2005). However, most field isolates, particularly those from buffalo, contain complex mixtures comprising multiple *T. parva* genotypes, making it impossible to determine the genotypes of individual parasites directly using microsatellite markers.

The genes coding for the p104 and PIM antigens of South African *T. parva* field isolates were examined in this study since they have previously been exploited for discrimination between *T. parva* stocks in other countries (Geysen *et al.*, 1999; Bishop *et al.*, 2001). PCR-RFLP and sequence analyses of the variable regions of these genes were used to identify buffalo-type and cattle-type alleles.

2. Materials and methods

2.1. Sample collection and screening of *T. parva*-positive samples

Blood samples (n=166) were collected from buffalo from different game parks, and from cattle from farms with suspected theileriosis. The samples were collected in EDTA vacutainer tubes and placed at -20°C, for long term storage, or 4°C for short term storage, before extraction of DNA using the High Pure PCR Template Preparation kit (Roche Diagnostics, Mannheim, Germany). Total DNA was extracted from 200 µl of EDTA blood, according to the method described by the manufacturer, except that extracted DNA was eluted in 100 µl. The presence of *T. parva* DNA was demonstrated in field samples using the *T. parva*-specific real-time PCR assay as previously described (Sibeko *et al.*, 2008). Two DNA samples from cattle-derived *T. parva* stocks, Muguga (Brocklesby *et al.*, 1961) and Katete (Geysen, 2000), from Kenya and Zambia, respectively, were used as reference samples. *Theileria parva* (Schoonspruit), an isolate obtained from a bovine infected during the ECF epidemic in South Africa (Neitz, 1948), was also used for reference.

2.2. Amplification of p104 and PIM genes from *T. parva* samples

The variable regions of the *T. parva* PIM and p104 genes were amplified using the semi-nested PCR-RFLPs described by De Deken *et al.* (2007). SYBR[®] green (SIGMA-ALDRICH, USA) was used for DNA detection.

2.3. PCR-RFLP analysis of p104 and PIM amplicons

RFLP patterns were analysed by visual inspection and using BioNumerics version 5.1 (Applied Maths, Kortrijk, Belgium). RFLP profiles were normalised using the 100 bp DNA Ladder

(Fermentas Life Sciences, Germany). DNA fragments of less than 100 bp were excluded from the analysis as these could not be accurately estimated using the molecular weight marker and in some instances had run out of the gel. Samples with similar RFLP profiles obtained from different animals were defined as clusters.

2.4. Cloning and sequencing of p104 and PIM PCR products

The p104 and PIM PCR products from selected *T. parva* samples were cloned into pCR[®]2.1-TOPO[®] cloning vector (Invitrogen, Carlsbad, USA); at least 24 clones were screened for each sample. The presence of inserts in the recombinants was confirmed by colony PCR following the secondary PCR protocol described by De Deken *et al.* (2007). Amplicons produced from the PIM colony PCR were digested with *Bcl*I to produce RFLP profiles for individual clones. Only clones that produced amplicons which successfully digested were considered for further analysis. The ABI Big Dye Terminator Cycle Sequencing kit version 3.1 (Applied Biosystems, Foster City, CA) was used for sequencing reactions including 300 to 450 ng of plasmid DNA. Sequencing of the PIM gene was performed by INQABA Biotechnologies (South Africa) and p104 gene sequencing was done at the Genetic Service Facility of the University of Antwerp (Belgium).

2.5. Sequence analysis

The p104 and PIM gene sequences were assembled and edited using the GAP4 program of the Staden package (version 1.6.0 for Windows) (Bonfield *et al.*, 1995; Staden, 1996; Staden *et al.*, 2000). p104 sequences were aligned with previously published *T. parva* p104 sequences (Table 1). A multiple sequence alignment of p104 amino acid sequences was performed using MAFFT version 6 (Kato *et al.*, 2002) (<http://align.bmr.kyushu-u.ac.jp/mafft/software>), and maximum parsimony and Bayesian analyses were used to produce phylogenetic trees. The maximum parsimony analysis was performed in PAUP*4.0b10 (Swofford, 2003) with 1000 random addition sequence followed by bisection-reconnection (TBR) branch swapping and branch support was assessed with 100 bootstrap replicates. The Bayesian analysis was performed using MrBayes v3.1.1 (Ronquist and Huelsenbeck, 2003). A Markov chain Monte Carlo run of five million generations consisting of four parallel MCMC chains was performed.

PIM sequences were aligned with previously published *T. parva* PIM sequences (Table 1) using MacClade v4.0 (Maddison and Maddison, 1992). The alignment was adjusted manually because of the highly polymorphic structure of the PIM gene. It was impossible to perform phylogenetic analyses for the PIM sequences because of the polymorphic nature of this gene.

3. Results

Theileria parva DNA was detected in 111/166 (67%) samples using real-time PCR (Sibeko *et al.*, 2008). These comprised 103 field samples collected from buffalo and eight collected from cattle (Table 2).

3.1. PCR-RFLP profile analysis

p104 gene analysis: A p104 PCR product of approximately 800 bp in size was obtained from 101 samples analyzed in this study (results not shown). From visual analysis, cattle samples, Lad 02, Lad 06, Lad 10 (from Ladysmith) and *T. parva* Schoonspruit (Neitz, 1948) had an identical PCR-RFLP profile to that of *T. parva* Muguga, a stock from Kenya responsible for ECF (Fig. 1).

The p104 PCR-RFLP cluster analysis using BioNumerics revealed two major groups, A and B, each of which could be divided into three subgroups (Fig. A available online). The clustering correlated broadly with geographic origin of the samples: group A consisted mainly of samples obtained from Hluhluwe-iMfolozi Game Park (39/50, 78%) and group B consisted largely of samples from Kruger National Park (KNP) (42/51, 82.4%), although KNP samples were also present in group A. Samples from Hluhluwe-iMfolozi produced a relatively homogeneous fingerprint, with most (38/42, 90.5%) appearing in subgroup A1, while those from KNP buffalo samples were heterogeneous and appeared in all six fingerprint groups. The p104 RFLP profiles from four cattle samples, Lad 02, Lad 06, Lad 10 and *T. parva* Schoonspruit, which were identical to that of *T. parva* Muguga from visual inspection, grouped with *T. parva* Muguga in subgroup A2. These cattle samples clustered amongst p104 profiles mainly from buffalo samples from Hluhluwe-iMfolozi in group A. Other cattle samples, Lad 11, Lad 15, Lad 17, Bloe B and Lyd N254 grouped with samples in group B which contained mainly buffalo samples from KNP. The *T. parva* Muguga-like cattle-type p104 fingerprints were not obtained from any of the buffalo samples analyzed in this study.

PIM gene analysis: The PIM gene PCR products obtained from 109 *T. parva* positive samples analyzed in this study ranged in size from 0.7 to 1.2 kb (results not shown). Since PIM is a single copy gene (Toye *et al.*, 1995a), multiple PCR products obtained from a single sample indicated the presence of mixed infections. Mixed infections were evident from 56/109 (51%) samples.

PIM PCR-RFLP profiles from all 48 samples from KNP were heterogeneous by visual inspection (Fig. 2a), while the profiles obtained from 23/39 (59%) of the Hluhluwe-iMfolozi samples from buffalo were relatively homogeneous (Fig. 2b). Similarly 4/6 (67%) samples from Mabalingwe and all 10 from Ithala produced relatively homogeneous profiles (Fig. 2c and 2d). Further analysis of PIM PCR-RFLP profiles by cluster analysis using BioNumerics was not possible, since these profiles were too complex as a result of multiple infections with different *T. parva* strains. Furthermore, PCR products were present at different concentrations resulting in multiple bands of different intensities, and it was difficult to distinguish between bands from incompletely digested amplicons and authentic bands. Therefore, 27 samples representative of *T. parva* samples from buffalo that produced homogeneous and heterogeneous profiles as well as six cattle samples (Table 2) were selected for cloning in order to produce RFLP profiles from individual clones.

More than one profile was obtained from the cloned PIM amplicons from 30/33 (91%) of the field samples. Different PCR-RFLP profiles were also obtained from clones produced from some of the samples which had a single band when analyzed on the agarose gel. Three samples from Ladysmith (Lad 02, Lad 06 and Lad 10) were exceptions, as all clones from these samples produced only one profile and this profile was identical to that of *T. parva* Muguga and *T. parva* Schoonspruit (Fig. 3). When RFLP profiles obtained from clones produced from samples which had homogeneous overall PIM-RFLP profiles were visually analysed, it was observed that there were dominant profiles that were responsible for the apparently homogeneous overall profile. For example, among other profiles, three profiles were found to be dominant in clones produced from the 10 samples from Ithala; the three profiles were obtained in, respectively, 20/61 (33%), 14/61 (23%) and 11/61 (18%) clones produced from four different samples.

Cluster analysis of PIM PCR-RFLP profiles from 259 clones produced from both buffalo and cattle *T. parva* samples identified five cluster groups, A, B, C, D and E, (Fig. B available online). Cluster A was the largest group with 105/259 (40.5%) clones followed by cluster D with 73/259 (28%), then C (42/259, 16%), B (26/259, 10%) and E (15/259, 6%). No correlation with geographic distribution could be established from the major cluster groups. Profiles from clones produced from cattle samples Lad 02, Lad 06 and Lad 10 grouped with *T. parva* Muguga and *T. parva* Schoonspruit profiles in cluster A. Although most PIM profiles in cluster A were obtained from buffalo samples, 42/48 (88%) of the PIM profiles from clones obtained from cattle samples also occurred in this group. PIM profiles from clones produced from the other cattle *T. parva* samples

from Ladysmith and Bloemfontein grouped closely with profiles obtained from buffalo samples from Hluhluwe-iMfolozi and Ithala in different subgroups within cluster A.

Specific 'signatures' composed of several small fragments (less than 150 bp) were produced in the PIM PCR-RFLP profiles obtained from some *T. parva* field samples from KNP (results not shown) and Hluhluwe-iMfolozi (Fig. 2b). Two specific 'signatures' were associated with some RFLP profiles from clones from KNP samples and both 'signatures' co-occurred with other bands of larger sizes. The Hluhluwe-iMfolozi 'signature' was also apparent in profiles from field samples obtained from buffalo from Ithala (Fig. 2d), and from three bovines from Ladysmith (Lad 17, Lad 1438 and Lad M119) and a bovine from Bloemfontein (results not shown).

3.2. Sequence analysis

p104 gene analysis: To confirm results obtained by RFLP profile cluster analysis and to establish p104 alleles that occur in the different subgroups, PCR products from 18 field samples (Table 2) representative of each cluster group were cloned and sequenced. Fifty-three good quality p104 sequences were obtained. None of these was identical to the four p104 alleles previously reported by Skilton *et al.* (2002) from different *T. parva* stocks of cattle and buffalo in East Africa (Table 1), although they were very similar with sequence identities ranging from 92-99% in the region analyzed.

Both maximum parsimony and Bayesian analyses produced similar groupings of p104 sequences; however, the maximum parsimony analysis did not resolve the differences within groups containing alleles 2 and 3. Therefore only the phylogenetic tree based on Bayesian analysis is presented (Fig. 4). The p104 sequences obtained in this study could be grouped into two main clades, A and B, and each of the major clades could be divided into two sub-clades, A1 and A2, and B1 and B2. Alleles 2 and 3 grouped in sub-clade A2 and alleles 1 and 4 in sub-clade B1. The p104 amino acid sequence obtained from Lad10 was identical to the p104 sequence from *T. parva* Schoonspruit. The Lad10 and *T. parva* Schoonspruit p104 sequences grouped with the *T. parva* Muguga p104 sequence (allele 1) and both had sequence identity of 99% to the *T. parva* Muguga p104 amino acid sequence. In addition to these, a p104 sequence obtained from a buffalo *T. parva* sample, KNP B10_3.0.2, had 97% sequence identity to that of *T. parva* Muguga and grouped closely to p104 sequences from Lad10, *T. parva* Schoonspruit and *T. parva* Muguga in sub-clade B1. Apart from *T. parva* Katete, no sequences grouped closely to p104 allele 2; similarly no sequences grouped

closely with allele 3. Sequence identities of 96-97% were obtained from sequences that grouped closely with p104 allele 4 (KNP102 8 3, KNPW8 8.0.2, KNPAB47 8.0.1).

The groups obtained from PCR-RFLP cluster analysis did not correlate with the clades identified by phylogenetic analysis of p104 amino acid sequences. Unfortunately, DNA from samples of reference sequences was not available for PCR-RFLP analysis and it was therefore not possible to determine where the p104 profiles from these samples would fit in the cluster analysis.

PIM gene analysis: Clones from different cluster groups were selected for sequencing. A total of 97 PIM sequences were obtained from cloned amplicons produced from 26 selected *T. parva*-positive samples (Table 2). Analysis of the amino acid alignment of the PIM sequences revealed three groups of PIM sequences, cattle-type, buffalo-type and 'mixed'-type (Fig. C available online).

Cattle-type PIM sequences:

A tetrapeptide repeat, QPEP, in the variable region of the PIM gene was previously identified by Toye *et al.* (1995b) and found to be characteristic of cattle-derived *T. parva* PIM sequences (Geysen *et al.*, 2004). The amino acid sequences between positions 29 and 259 and positions 493 and 497 were also identified as exclusive to cattle-type PIM alleles in the present study. In this way, six PIM sequences obtained from two field samples investigated in this study (Lad 02 and Lad 10) were identified as cattle-type *T. parva* PIM sequences. These sequences were identical to the PIM sequence obtained from *T. parva* Schoonspruit. These sequences had 98% identity to the *T. parva* Muguga PIM sequence, with one to three amino acid differences; the major difference being a deletion of eight amino acids at position 420-427 in the South African sequences. Cattle-type PIM sequences were not identified from samples obtained from buffalo in this study.

Buffalo-type PIM sequences:

Toye *et al.* (1995a) identified a 20-amino-acid insert (VDQQQPVQQPSQDQPSGPDS) as characteristic to buffalo-type PIM amino acid sequences. In addition to this, two other buffalo-type amino acid motifs and other unique inserts were identified in our study. The 20-amino-acid insert identified by Toye *et al.* (1995a) from PIM sequences from buffalo-derived *T. parva* 7014 and Hluhluwe stocks PIM sequences, was missing from 9/53 (17%) of the buffalo-type PIM sequences obtained in this study. 54.6% (53/97) of the PIM sequences obtained in our study were identified as buffalo-type PIM sequences and were obtained from *T. parva* samples collected from both cattle and buffalo.

'Mixed'-type PIM sequences:

The use of the amino acid motifs identified in our study allowed identification of 37 'mixed' PIM sequences. Two subtypes of 'mixed' PIM sequences were identified, M-I (32/37) and M-II (5/37) (Fig. D available online). Subtype M-I consisted of sequences characteristic of cattle-type *T. parva* PIM sequences at the amino-terminus and buffalo-type sequences at the carboxy-terminus. Subtype M-II comprised sequences characteristic of buffalo-type *T. parva* PIM sequences at the amino-terminus and cattle-type sequences at the carboxy-terminus. The 37 'mixed' PIM sequences were obtained from samples collected from both cattle and buffalo.

Of the 22 *T. parva* PIM sequences obtained from cattle samples in this study (*T. parva* Schoonspruit and six field samples), 31.8% (7/22) were cattle-type, 59.1% (13/22) were 'mixed'-type, and only 9.1% (2/22) were buffalo-type. In comparison, no cattle-type PIM sequences were identified from 20 buffalo samples, 32% (24/75) of the PIM sequences obtained from buffalo samples were mixed-type, and 68% (51/75) were buffalo-type.

4. Discussion

Identification of cattle-type p104 and PIM alleles from cattle T. parva samples

Cattle-to-cattle transmission of *T. parva* was suspected on a farm near Ladysmith in the KwaZulu-Natal Province following cattle theileriosis outbreaks in several consecutive years in the apparent absence of infected buffalo (Thompson *et al.*, 2008). Three cattle samples from the Ladysmith farm and *T. parva* Schoonspruit had p104 and PIM PCR-RFLP profiles and inferred amino acid sequences almost identical to those of *T. parva* Muguga. These results could imply that there was selection of cattle-type alleles in some of the cattle on the farm, suggesting that cattle-to-cattle transmission may indeed have occurred. However, the original source of *T. parva* infection remains undiscovered and ECF was not diagnosed in animals on this farm. There have not been reports of any officially recognized ECF outbreaks in South Africa since its eradication in the early 1950s. While the presence, in South African cattle, of a parasite apparently similar to *T. parva* Muguga and the Schoonspruit isolate made during the ECF epidemic is a concern, there is no evidence to suggest that genotypes at individual loci such as p104, PIM and p67, can be linked to the pathogenicity of the isolate or to the disease syndrome it causes.

p104 and PIM alleles obtained from *T. parva* samples from several other animals from this farm were either buffalo-type or 'novel' (or mixed-type in the case of PIM) suggesting that the parasite population infecting cattle on this farm originated from buffalo, although contact between these cattle and *T. parva*-infected buffalo could not be confirmed (Thompson *et al.*, 2008). The Hluhluwe-iMfolozi 'signature' obtained in PIM PCR-RFLP profiles from cattle samples from the Ladysmith farm, similar to that obtained from buffalo samples from Hluhluwe-iMfolozi game park, provides further support for the hypothesis that there may have been contact between cattle on this farm and *T. parva*-infected buffalo.

p104 and PIM PCR-RFLP fingerprints obtained from a sample collected from a bovine which died of Corridor disease (Bloe B) on a vector-free game farm near Bloemfontein on which there was an infected buffalo breeding herd, grouped with profiles from buffalo samples, suggesting that these *T. parva* parasites may have originated from the buffalo on the farm. This hypothesis is supported by the presence of the 'signature' in the Bloe B PIM profile characteristic of most Hluhluwe-iMfolozi profiles and the sequence analyses, all of which suggest that the *T. parva* parasite characterized from the Bloemfontein bovine sample was similar to *T. parva* parasites from Hluhluwe-iMfolozi. In fact, the infected buffalo breeding herd on the Bloemfontein farm originated from Hluhluwe-iMfolozi. Therefore, it might be possible to use PIM PCR-RFLP profiles to track the origin of *T. parva* infections, especially when the profiles have a specific signature that characterizes a particular parasite population.

Identification of cattle-type alleles from buffalo T. parva samples

Although no p104 or PIM alleles identical to *T. parva* cattle-type alleles previously reported (Iams *et al.*, 1990; Toye *et al.*, 1995a; 1995b; Skilton *et al.*, 2002) were obtained from buffalo samples analyzed in this study, variants of p67 allele 1 (Sibeko *et al.*, 2010) as well as variants of p104 allele 1 (both cattle-type alleles) were obtained. These findings suggest that parasites possessing cattle-type p67 and p104 alleles may not be exclusively associated with cattle-derived *T. parva* parasites. While these results could suggest that there may have been transmission of cattle-derived *T. parva* parasites to buffalo, they could also be indicative of an ancestral buffalo-derived *T. parva* subpopulation with characteristics that might aid the parasite to adapt to cattle.

Since we identified cattle-type p104 and PIM alleles in cattle, we might have expected to see cattle-type p104 and/or PIM sequences in *T. parva* samples from buffalo. Buffalo are believed to carry a heterogeneous population of parasites and, as original hosts of the parasite, cattle-derived *T. parva*

parasites might have originated in buffalo. It is possible that the pool of buffalo samples investigated was not large enough to allow detection of all *T. parva* alleles that occur in buffalo. Therefore, a larger pool of samples will have to be characterized in order to determine if there are cattle-type alleles in buffalo.

It is still not clear if the parasite that caused ECF in South Africa in the early part of the 20th century was transmitted to buffalo during the ECF epidemic, or whether South African strains of *T. parva* existing in buffalo could eventually become adapted to cattle to cause ECF. To date, there are no reports where the source of infection was determined in cases where both cattle- and buffalo-derived *T. parva* parasites occur, and more reliable markers to distinguish between cattle-derived and buffalo-derived parasites would be needed in order to perform such experiments. There are reports indicating that cattle-derived parasites can be transmitted to buffalo (Walker, 1932; Lewis, 1943; Brocklesby, 1964; Barnett and Brocklesby, 1966a; 1966b). In this study, variants of cattle-type alleles were obtained from samples originating from buffalo, however there is no evidence that these were transmitted from cattle. Such transmission, if it did occur, could play an important role in the evolution of genetic polymorphism in *T. parva* parasites and in the selection for genetically variant *T. parva* strains.

Genetic diversity among South African T. parva parasite populations:

The sequence identities of p104 sequences obtained from buffalo in this study, when compared with published sequences (Skilton *et al.*, 2002), ranged from 94 to 97% suggesting that *T. parva* p104 alleles in South African buffalo are more diverse and that variations in the p104 gene may not be as limited as previously thought (Geysen *et al.*, 1999).

None of the PIM gene sequences obtained from *T. parva* field samples from buffalo characterized in this study were identical, providing further evidence that the PIM gene evolves at an extremely high rate (Toye *et al.*, 1995a). Extensive sequence heterogeneity among PIM sequences was demonstrated in the South African *T. parva* samples, confirming the extensive genetic diversity reported previously in *T. parva* parasites in buffalo (Collins and Allsopp, 1999; Sibeko *et al.*, 2010). Furthermore, sequence variants of buffalo-type PIM sequences were identified which have never been reported before. Novel alleles have been reported to arise from re-shuffling of important epitopes as a result of gene conversion and reciprocal intergenic exchanges (Dormoy *et al.*, 1997). While the discriminative characteristics of the PIM gene render it a good candidate for exploitation in discriminatory assays used for *T. parva* isolates (Bishop *et al.*, 2001; De Deken *et al.*, 2007), it is

advisable that assays based on this gene should be continuously evaluated because of the rapid evolution of the PIM gene.

In this study, for the first time, ‘mixed’ PIM sequences with characteristics of both cattle-type and buffalo-type PIM sequences were identified. It is possible that hybrid PCR products could arise during PCR amplification in samples containing mixed infections, as a result of template switching (Wang and Wang, 1997; Thompson *et al.*, 2002). These PCR artifacts are identifiable by sequence identities in hybrid sequences when compared with sequences of other amplicons produced in the same amplification reaction. To confirm whether the ‘mixed’ sequences obtained in this study were authentic, the PIM nucleic acid sequences in the more conserved regions, i.e. sequences flanking the central variable region, were compared to those of other sequences obtained from the same sample. None of the ‘mixed’ sequences were identical to any of the other PIM sequences obtained from the same sample in these conserved regions, suggesting that the ‘mixed’ sequences were therefore genuine and could not have resulted from PCR artifacts.

The combinations of different “blocks” of sequence observed in the PIM gene, in this study and elsewhere (Geysen *et al.*, 2004), are reminiscent of the structure that has previously been shown for the precursor to the major merozoite surface antigens (PMMSA) in *Plasmodium falciparum* (Peterson *et al.*, 1988). Recombination within the conserved blocks in the PMMSA is thought to result in reassembling of the variable blocks and accounts for much of the antigenic variation in this molecule. Studies of the population structure of *T. parva* parasites in Uganda showed that genetic exchange occurs frequently between isolates of *T. parva*, confirming the existence of a sexual cycle (Oura *et al.*, 2005). Evidence for recombination between *T. parva* parasites, in a form of mosaic segments, has previously been observed in the internal transcribed spacer (ITS) region (Collins and Allsopp, 1999) and sexual recombination between different *T. parva* stocks has been demonstrated in the laboratory (Morzaria *et al.*, 1993; Bishop *et al.*, 2002). It seems likely that the ‘mixed’ PIM sequences have arisen through recombination between cattle-type and buffalo-type PIM sequences.

While the ‘mixed’ PIM sequences identified in this study may well represent recombination events, we do not know the full extent of the recombination in the rest of the genome in these parasites. Recombination between buffalo-type *T. parva* parasites and cattle-type parasites can only occur where there has been contact between buffalo and cattle, and it has been reported that it is unlikely for recombinant parasites to become established in the cattle population (Geysen, 2000; Geysen *et al.*, 2004). Nonetheless, ‘mixed’ PIM sequences were obtained from samples originating from both

buffalo and cattle in this study, and the results indicate that ‘mixed’ PIM sequences are more prevalent in *T. parva* samples from cattle than in *T. parva* samples from buffalo (51% of sequences obtained from cattle samples were ‘mixed’ type PIM alleles compare to 32% from buffalo samples). This could suggest that *T. parva* parasites with the ‘mixed’ PIM allele might be more likely to establish in cattle. In fact, very few buffalo-type PIM sequences were identified in cattle, suggesting that there may have been selection for both ‘mixed’-type and cattle-type PIM sequences in cattle.

Our results demonstrate that p104 and PIM amino acid sequences very similar to those of *T. parva* Muguga were present in *T. parva* parasites on a farm in South Africa, where pathogenic cattle theileriosis was reported. The possibility of cattle-to-cattle transmission of *T. parva* in South Africa and the selection of cattle-type parasites in such animals should be further investigated. Although no cattle-type p104 or PIM alleles were identified from samples obtained from buffalo in this study, more samples need to be examined to confirm this finding. The challenge to identify more reliable and specific markers that can be directly associated with different disease syndromes caused by *T. parva* still remains. This is crucial for the design and implementation of control measures in South Africa to protect cattle against disease caused by *T. parva* infections.

6. Conflict of interest statement

None of the authors has a financial or personal relationship with other people or organisations that could inappropriately influence or bias the paper entitled “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”.

7. Acknowledgements

The authors are grateful to the BioPAD Bric (BP040), the Belgian Directorate General for Development Co-operation Framework agreement ITM/DGCD and Institute of Tropical Medicine (ITM) grants for financial support. Our sincere gratitude to J De Witte and L Vermeiren of ITM, for technical support and the Institute of Tropical Medicine, Belgium and the Agricultural Research Council-Onderstepoort Veterinary Institute, South Africa for making their facilities and material available for the project.

8. References

- Anonymous, 1981. The eradication of East Coast fever in South Africa. *J. S. Afr. Vet. Assoc.* 52, 71-73.
- Barnett, S.F. and Brocklesby, D.W., 1966a. The susceptibility of the African buffalo (*Syncerus caffer*) to infection with *Theileria parva* (Theiler 1904). *British Vet. J.* 122, 379-386.
- Barnett, S.F. and Brocklesby, D.W., 1966b. The passage of "*Theileria lawrencei* (Kenya)" through cattle. *British Vet. J.* 122, 361-37.
- Baylis, H.A., Allsopp, B.A., Hall, R. and Carrington, M., 1993. Characterisation of a glutamine-rich and proline-rich protein (QP protein) from *Theileria parva*. *Mol. Biochem. Parasitol.* 61, 171-178.
- Bishop, R.P., Geysen, D., Spooner, P., Skilton, R., Nene, V., Dolan, T. and Morzaria, S., 2001. Molecular and immunological characterization of *Theileria parva* stocks which are components of 'Muguga cocktail' used for vaccination against East Coast fever in cattle. *Vet. Parasitol.* 94, 227-237.
- Bishop, R., Geysen, D., Skilton, R., Odongo, D., Nene, V., Allsopp, B., Mbogo, S., Spooner, P. and Morzaria, S., 2002. Genomic polymorphism, sexual recombination and molecular epidemiology of *Theileria parva*. In: Mckeever, D. and Dobbelaere, D. (Eds.), *Theileria*. Kluwer Academic, Dordrecht, pp. 23-40.
- Bonfield, J. K., Smith, K.F. and Staden, R., 1995. A new DNA sequence assembly program. *Nucleic Acids Res.*, 23, 4992-4999.
- Brocklesby, D.W., Barnett, S.F. and Scott, G.R., 1961. Morbidity and mortality rates in East Coast fever (*Theileria parva* infection) and their application to drug screening procedures. *Br. Vet. J.* 117, 529-531.
- Brocklesby, D.W., 1964. Parasites of the family *Theileridae* of the African buffalo occurring in East Africa. D. V. M. Thesis, University of Zurich, pp114.
- Collins, N.E. and Allsopp, B.A., 1999. *Theileria parva* ribosomal internal transcribed spacer sequences exhibit extensive polymorphism and mosaic evolution: application to the characterisation of parasites from cattle and buffalo. *Parasitology*, 83, 541-551.
- De Deken, R., Martin, V., Saido, A., Madder, M., Bradt, J. and Geysen, D., 2007. An outbreak of East Coast Fever on the Comoros: A consequence of the import of immunised cattle from Tanzania? *Vet. Parasitol.* 143, 245-253.
- Dormoy, A., Reviron, D.V., Froelich, N., Weiller, P.J., Mercier, P.J. and Tongio, M.M., 1997. Birth of a new allele in a sibling: *cis* or *trans* gene conversion during meiosis? *Immunogenetics*, 46, 520-523.

- Geysen, D., 2000. The application of Molecular Biology techniques to analyse diversity in *Theileria parva* populations in Zambia. Ph.D. Thesis. Brunel University, UK.
- Geysen, D., Bishop, R., Skilton, R., Dolan, T.T. and Morzaria, S., 1999. Molecular epidemiology of *Theileria parva* in the field. *Trop. Med. Int. Health*, 4, A21-A27.
- Geysen, D., Bazarusanga, T., Brandt, J. and Dolan, T.T., 2004. An unusual mosaic structure of the PIM gene of *Theileria parva* and its relationship to allelic diversity. *Mol. Biochem. Parasitol.* 133, 163-174.
- Iams, K.P., Hall, R., Webster, P. and Musoke, A.J., 1990. Identification of λ gt11 clones encoding the major antigenic determinants expressed by *Theileria parva* sporozoites. *Infect. Immun.* 58, 1828-1834.
- Katoh, K., Misawa, K., Kuma, K. and Miyata, T., 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 30, 3059-3066.
- Lawrence, J.A., Norval, R.R. and Uilenberg, G., 1983. *Rhipicephalus zambeziensis* as a vector of bovine theileriae. *Trop. Anim. Health Prod.* 15, 39-42.
- Lewis, E.A., 1943. East Coast fever and the African buffalo, the eland and the bushbuck. *East African Agric. J.* 9, 90-92.
- Maddison, W.P. and Maddison, D.R., 1992. MacClade: Analysis of phylogeny and character evolution. Version 3. Sinauer Associates, Sunderland, Massachusetts.
- Morzaria, S.P., Young, J.R., Spooner, P.R., Dolan, T.T. and Bishop, R.P., 1993. *Theileria parva*: a restriction map and genetic recombination. In: Morzaria, S.P. (Ed.), *Genome Analysis of Protozoan Parasites*, ILRAD, Nairobi, pp. 67-73.
- Neitz, W.O., 1948. Studies on East Coast fever. *South African Journal of Science*, 1, 133-135.
- Oura, C.A., Odongo, D.O., Lubega, G.W., Spooner, P.R., Tait, A. and Bishop, R.P., 2003. A panel of microsatellite and minisatellite markers for the characterisation of field isolates of *Theileria parva*. *Int. J. Parasitol.* 33, 1641-1653.
- Oura, C.A., Asimwe, B.B., Weir, W., Lubega, G.W. and Tait, A., 2005. Population genetic analysis and sub-structuring of *Theileria parva* in Uganda. *Mol. Biochem. Parasitol.* 140, 229-239.
- Peterson, M.G., Coppel, R.L., Moloney, M.B. and Kemp, D.J. 1988. Third form of the precursor to the major merozoite surface antigens of *Plasmodium falciparum*. *Mol. Cell Biol.* 8, 2664-2667.
- Potgieter, F.T., Stoltz, W.H., Blouin, E.F. and Roos, J.A., 1988. Corridor Disease in South Africa: a review of the current status. *J. S. Afr. Vet. Assoc.* 59, 155-160.
- Ronquist, F. and Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, 19, 1572-1574.

- Sibeko, K.P., Oosthuizen, M.C., Collins, N.E., Geysen, D., Rambritch, N.E., Latif, A.A., Groeneveld, H.T., Potgieter, F.T. and Coetzer, J.A.W., 2008. Development and evaluation of a real-time polymerase chain reaction test for the detection of *Theileria parva* infections in Cape buffalo (*Syncerus caffer*) and cattle. *Vet. Parasitol.* 155, 37-48.
- Sibeko, K.P., Geysen, D., Oosthuizen, M.C., Mathee, C.A., Troskie, M., Potgieter, F.T., Coetzer, J.A.W. and Collins, N.E., 2010. Four p67 alleles identified in South African *Theileria parva* field samples. *Vet. Parasitol.* 167, 244-254.
- Skilton, R.A., Bishop, R.P., Katende, J.M., Mwaura, S and Morzaria, S.P., 2002. The persistence of *Theileria parva* infection in cattle immunized using two stocks which differ in their ability to induce a carrier state: analysis using a novel blood spot PCR assay. *Parasitology*, 124, 265-276.
- Staden, R. 1996. The Staden Sequence Analysis Package. *Mol. Biotechnol.* 5, 233-241.
- Staden, R., Beal, K.F. and Bonfield, J.K., 2000. The Staden package, 1998. *Meth. Mol. Biol.* 132, 115-130.
- Swofford, D.L., 2003. *Phylogenetic analysis using Parsimony (*and Other Methods)*. Version 4. Sinauer Associates, Sunderland, Massachusetts.
- Thompson, J.R., Marcelino, L.A., Polz, M.F., 2002. Heteroduplexes in mixed-template amplifications: formation, consequence and elimination by 'reconditioning PCR'. *Nucleic Acids Res.* 30, 2083-2088.
- Thompson, B.E., Latif, A.A., Oosthuizen, M.C., Troskie, M. and Penzhorn, B.L., 2008. Occurrence of *Theileria parva* infection in cattle on a farm in the Ladysmith district, Kwa-Zulu-Natal, South Africa. *J. S. Afr. Vet. Assoc.* 79, 31-35.
- Toye, P.G., Gobright, E., Nyanjui, J., Nene, V. and Bishop, R., 1995a. Structure and sequence variation of the genes encoding the polymorphic immunodominant molecule (PIM), an antigen of *Theileria parva* recognized by inhibitory monoclonal antibodies. *Mol. Biochem. Parasitol.* 73, 165-177.
- Toye, P.G., Metzelaar, M.J., Wijngaard, P.J.L., Nene, V., Iams, K., Roose, J., Nyanjui, J.K., Gobright, E., Musoke, A.J. and Clevers, H.C., 1995b. Characterization of the gene encoding the polymorphic immunodominant molecule, a neutralizing antigen of *Theileria parva*. *J. Immunol.* 155, 1370-1381.
- Toye, P.G., Nyanjui, J., Goddeeris, B. and Musoke, A.J., 1996. Identification of neutralization and diagnostic epitopes on PIM, the polymorphic immunodominant molecule of *Theileria parva*. *Infect. Immun.* 64, 1832-1838.
- Walker, J. 1932. Report of the Chief Veterinary Officer, Kenya, Nairobi, p 296.

- Wang, G.C-Y. and Wang, Y., 1997. Frequency of formation of chimeric molecules as a consequence of PCR coamplification of 16S rRNA genes from mixed bacterial genomes. *Appl. Environ. Microb.* 63, 4645-4650.
- Young, A.S., Grocock, C.M. and Kariuki, D.P., 1988. Integrated control of ticks and tick-borne diseases of cattle in Africa. *Parasitology*, 96, 403-432.

Table legends

Table 1: Previously published p104 and PIM sequences

Table 2: Geographical origin and source of blood samples (n=111) used for characterization of *T. parva* parasites

Figure legends

Fig. 1: p104 gene *AluI* RFLP profiles showing the *T. parva* Muguga RFLP profile and identical profiles obtained from cattle samples from a farm near Ladysmith.

Fig. 2: Representative PIM gene *BclI* PCR-RFLP profiles showing (a) heterogeneous profiles obtained from buffalo *T. parva* samples from KNP, (b), (c) and (d) homogeneous profiles obtained from buffalo samples from Hluhluwe-iMfolozi, Mabalingwe and Ithala, respectively.

Fig. 3: PIM gene *BclI* PCR-RFLP profiles obtained from (a) cattle *T. parva* samples from Ladysmith and (b) clones produced from cattle sample, Lad 10.

Fig. 4: Phylogenetic relationship of *T. parva* strains as revealed by p104 amino acid sequence analysis. The phylogenetic tree was inferred by Bayesian analysis; the first 10,000 trees were discarded as burnin and the majority-rule consensus tree was generated.

Legends for online supplementary figures

Fig. A: A similarity dendrogram generated by BioNumerics v5.1 cluster analysis of PCR-RFLP profiles from p104 amplicons using the Dice Coefficient analysis.

Fig. B: A similarity dendrogram generated by BioNumerics v5.1 cluster analysis of PCR-RFLP profiles from cloned PIM amplicons using the Dice Coefficient analysis.

Fig. C: Multiple sequence alignment of selected PIM amino acid sequences, representative of cattle-, buffalo- and 'mixed'-type sequences. The inferred PIM sequences obtained from buffalo and cattle *T. parva* samples collected from different geographical areas in South Africa were aligned with previously published *T. parva* PIM amino acid sequences (Table 1). The sequence alignment was constructed manually because of the extreme polymorphism in the central region of the PIM gene. Amino acid motifs characteristic of cattle-type PIM sequence are shown in solid-line blocks including the tetrapeptide repeat characteristic of the

central variable region of cattle-derived PIM sequences at positions 428-447. All amino acid motifs characteristic of buffalo-type PIM sequence are shown in broken-line blocks.

Fig. D: Multiple sequence alignment of representative amino acid sequences of 'mixed'-type *T. parva* PIM sequences, showing subtypes M-I and M-II. Regions typical of buffalo-type and cattle-type PIM sequences are indicated in broken-line and solid-line blocks, respectively. This alignment excludes the region between positions 101 and 300, which has buffalo-type PIM sequences in all subtypes.

Table 1

Isolate name	Accession number and reference	
	p104	PIM
<i>T. parva</i> Muguga	Allele 1: M29954 (Iams <i>et al.</i> , 1990)	L06323 (Baylis <i>et al.</i> , 1993)
<i>T. parva</i> Marikebuni	Allele 2: AY034069 (Skilton <i>et al.</i> , 2002)	L41148 (Toye <i>et al.</i> , 1995a)
<i>T. parva</i> Boleni	Allele 3: AY034070 (Skilton <i>et al.</i> , 2002)	-
<i>T. parva</i> 7014	Allele 4: AY034071 (Skilton <i>et al.</i> , 2002)	L41833 (Toye <i>et al.</i> , 1995b)

Table 2

Geographical location	Province	Sample Name	Host of blood sample	Date of collection/References
Hluhluwe-iMfolozi Park (n=39)	KwaZulu-Natal	HIP 01, HIP 03, HIP 04, HIP 05 , HIP 06, HIP 07, HIP 08, HIP 09, HIP 10, HIP 11, HIP 12, HIP 13, HIP 14, HIP 15, HIP 16, HIP 18, <i>HIP 19</i> , HIP 20, HIP 21, HIP 22 , HIP 23, HIP 24, HIP 25, HIP 26, HIP 27, HIP 28, HIP 30, HIP 31, HIP 32 , HIP 33, HIP 34, HIP 35, HIP 36, HIP 37, HIP 38, HIP 39 , HIP 41, HIP 42, HIP 49	Buffalo	2004
Kruger National Park (KNP) (n=48)	Mpumalanga	KNP 39, KNP 42, KNP 43 , KNP 47, KNP 48, KNP 49, KNP 50, KNP 61, KNP 62, <i>KNP 63</i> , KNP 66, KNP 67, KNP 68, KNP 102 , KNP AA5, <i>KNP AB47</i> , KNP AC10, KNP AD3, <i>KNP B10</i> , KNP B22, KNP D11, KNP D24, KNP E7, KNP E18, KNP F9, KNP G2, KNP G11, KNP H8, KNP J5, <i>KNP L6</i> , KNP L27, KNP M2, KNP M12, KNP M2706, KNP N1, KNP N8, KNP O1, KNP O11 , <i>KNP P7</i> , KNP S17, KNP U3, KNP U20, KNP V5 , <i>KNP W8</i> , KNP X4, <i>KNP Y4</i> , KNP Y19, KNP Z4	Buffalo	2003
Ladysmith (n=7)	KwaZulu-Natal	Lad 02 , Lad 06, <i>Lad 10</i> , <i>Lad 15</i> , <i>Lad 17</i> Lad M119, Lad I438	Bovines	2003, Thompson <i>et al.</i> (2008)
Mabalingwe Game Reserve (n=6)	Limpopo	<i>Mab A13</i> , <i>Mab A22</i> , Mab B21 , <i>Mab BB37</i> , <i>Mab BB38</i> , <i>Mab BB43</i>	Buffalo	2004
[#] Ithala Game Reserve (n=10)	KwaZulu-Natal	Itha 1, Itha 2, Itha 3, Itha 4, Itha 5, Itha 6, Itha 7, Itha 8, Itha 9, Itha 10	Buffalo	2005/6
Bloemfontein (n=1)	Free-State	<i>Bloe B</i>	Bovine	2004

Samples in bold were selected for cloning of the PIM gene to produce RFLP profiles from individual clones and only 26 of the 33 were used for sequencing.

[#] Samples from this Game Reserve (Ithala) were only analyzed for the p104 gene.

Samples in italic were used for sequencing of the p104 gene variable region.

Samples in italic and bold were used for sequencing of both the p104 and PIM genes.

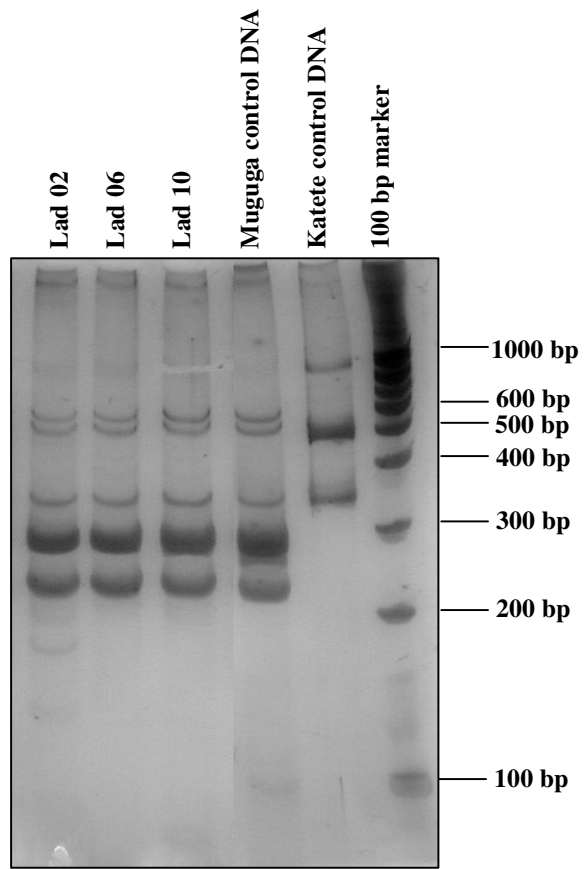


Figure 1

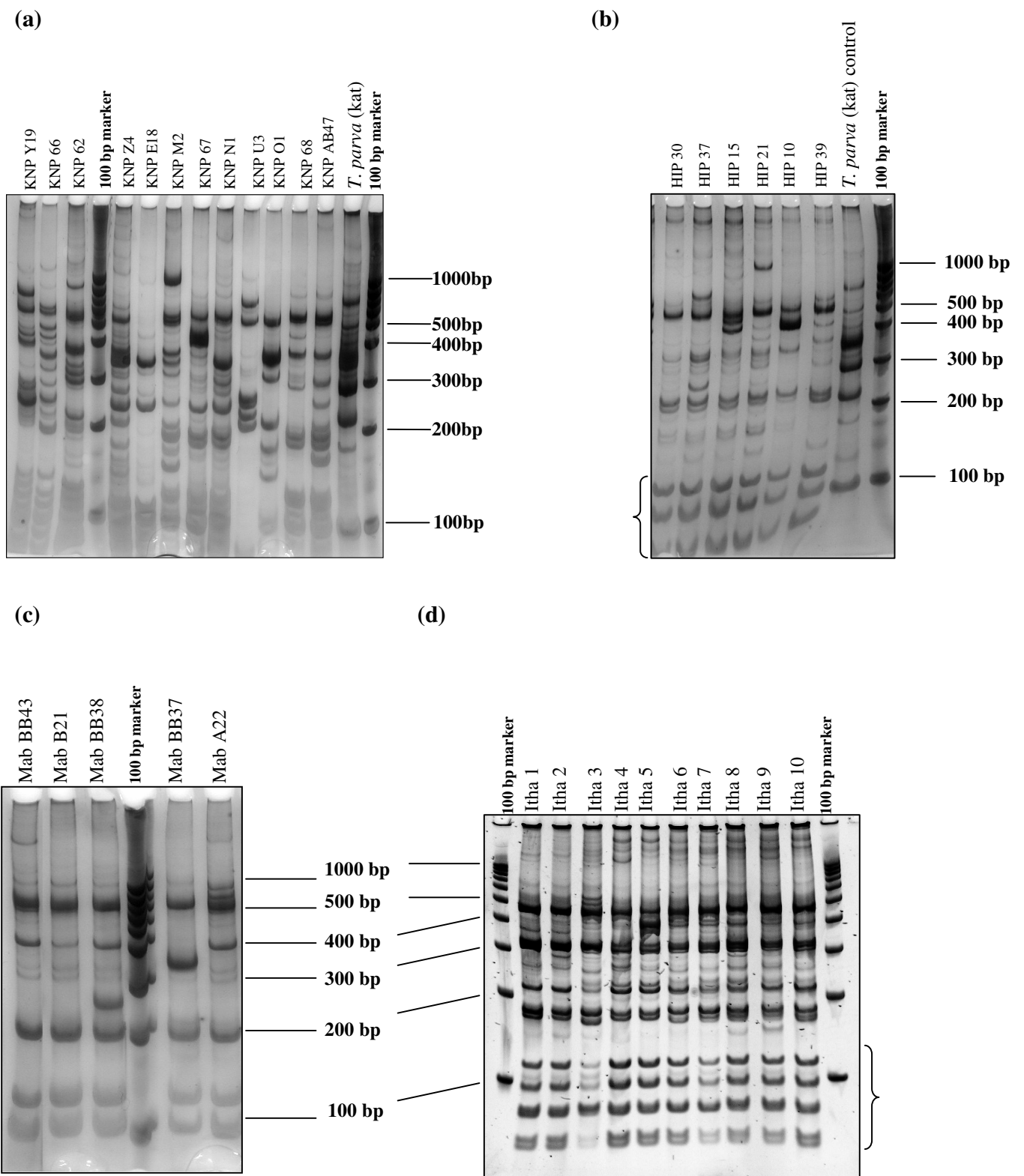


Figure 2

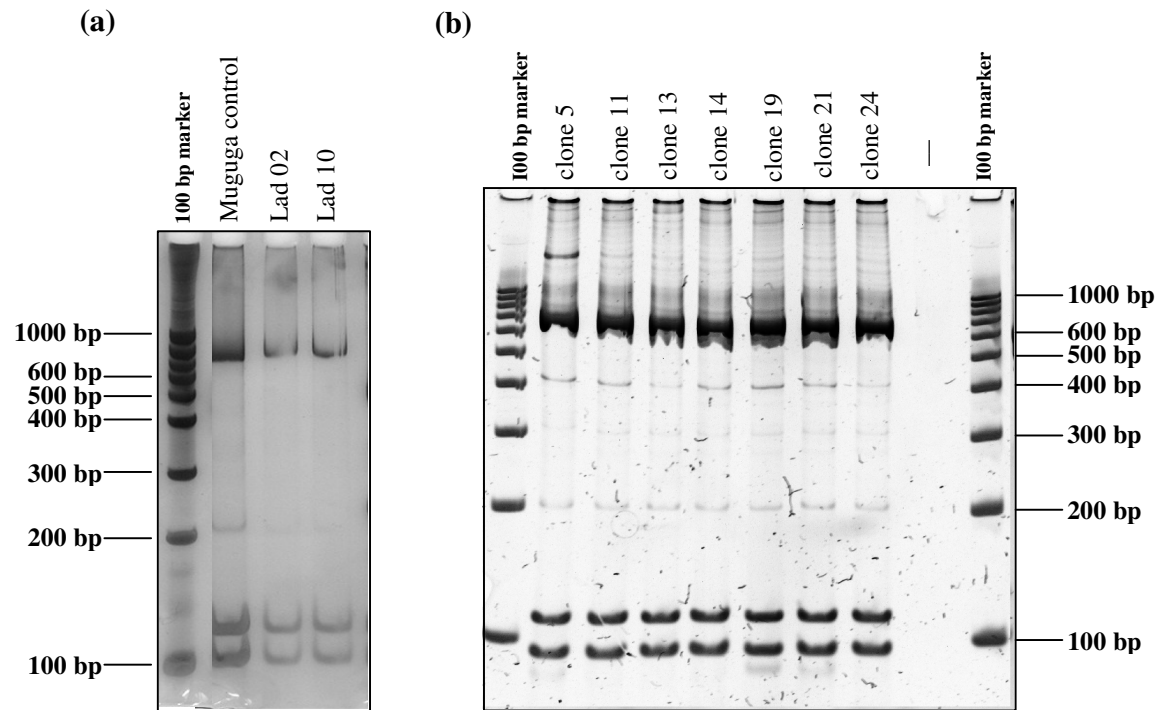


Figure 3

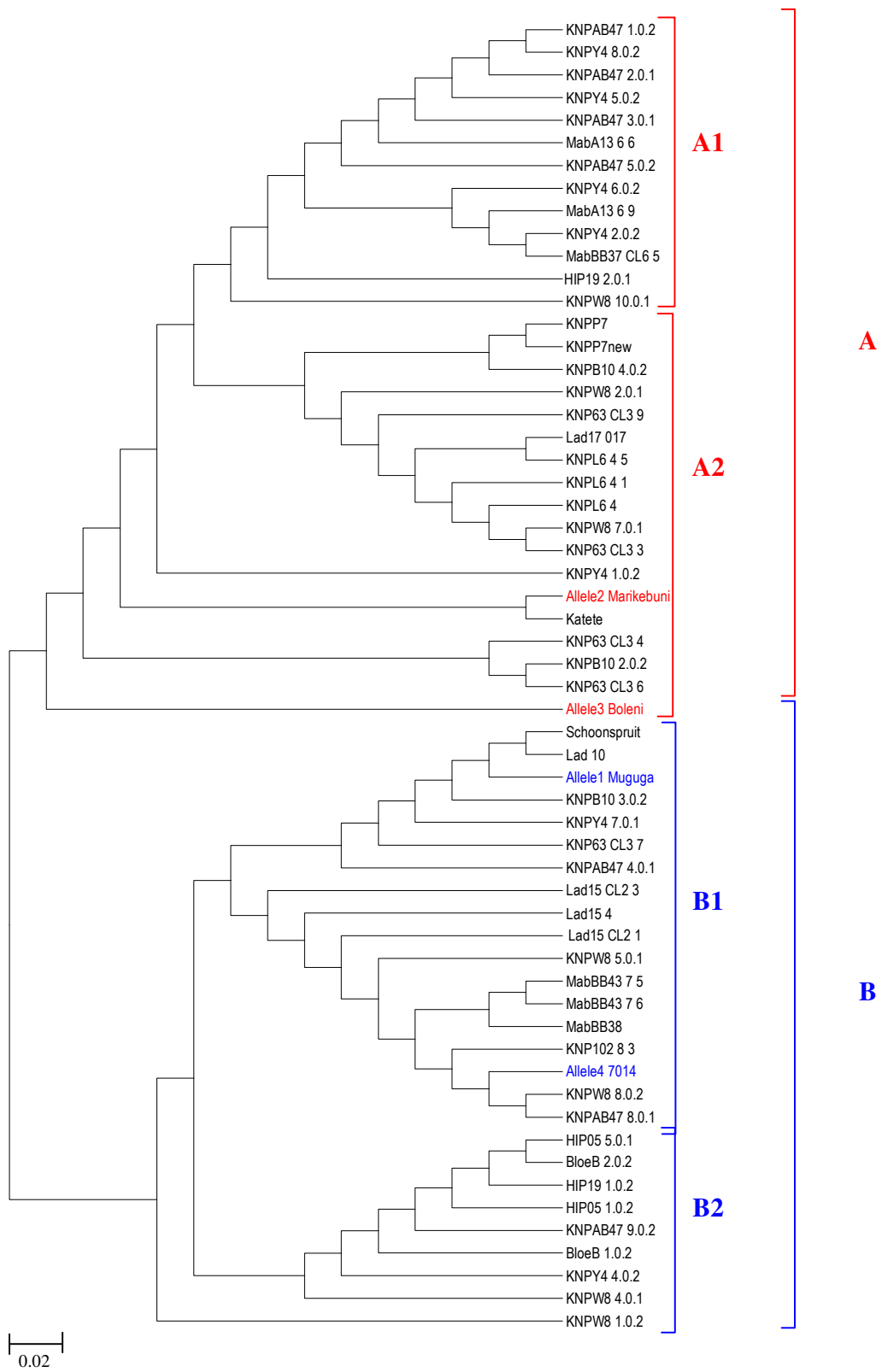


Figure 4

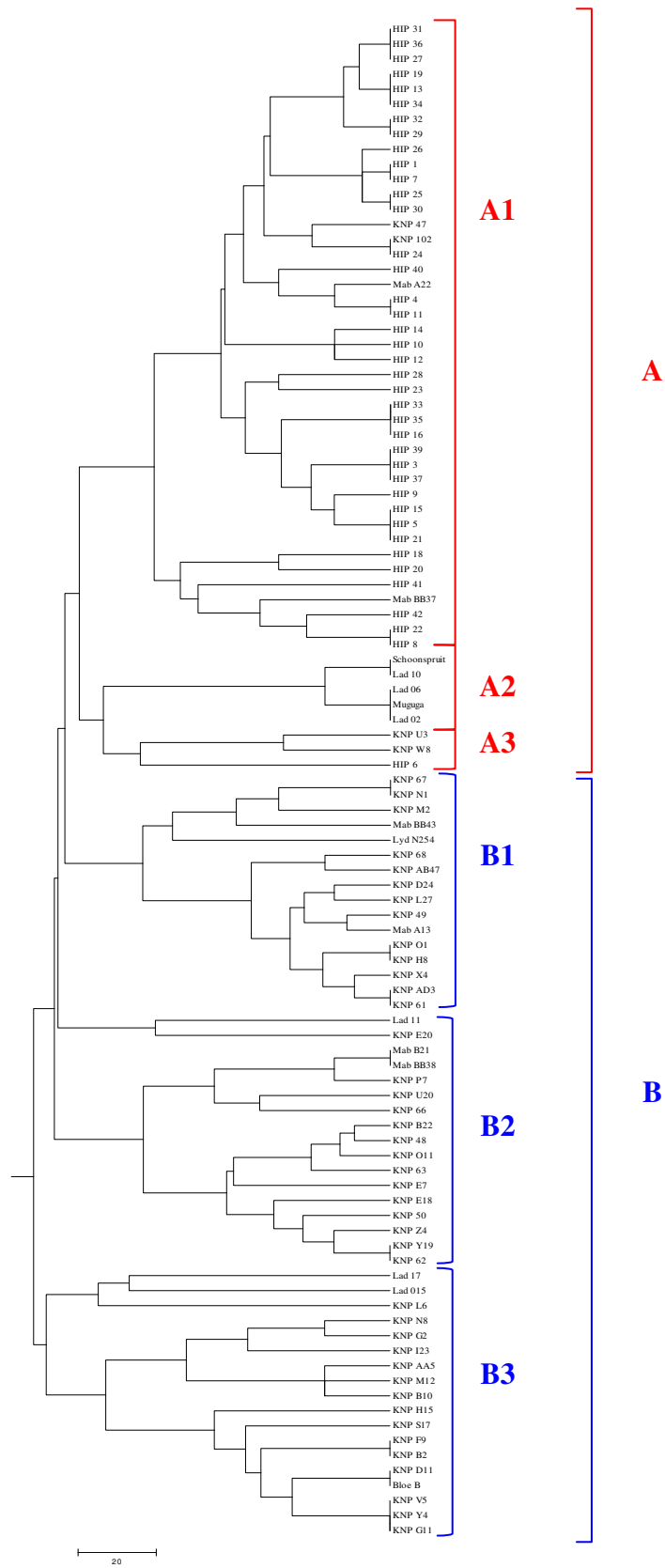


Figure A

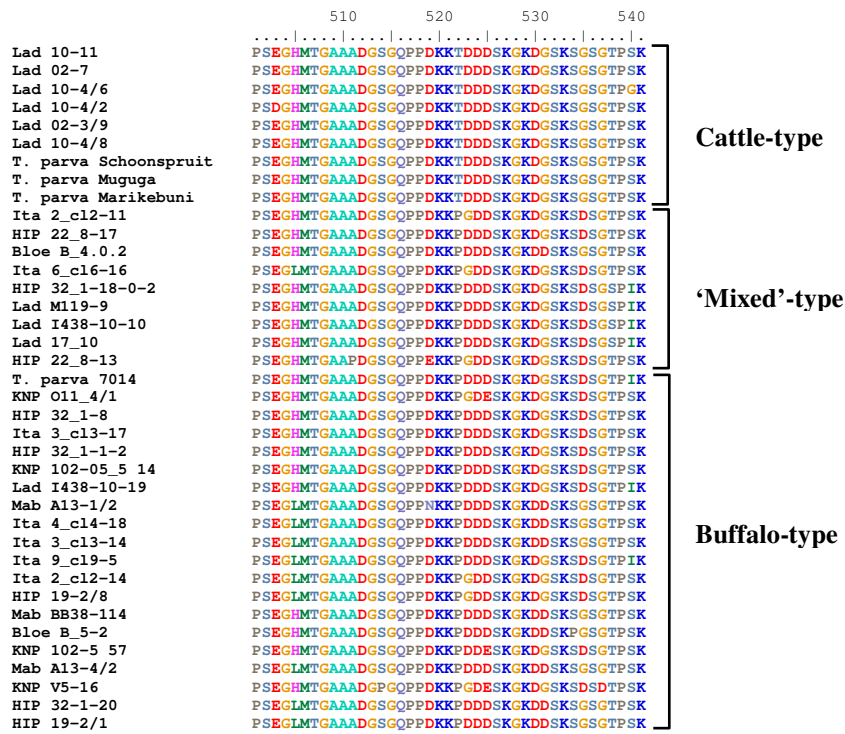


Figure C

