Investigating the diversity of the 18S SSU rRNA hyper-variable region of *Theileria* in cattle and Cape buffalo (*Syncerus caffer*) from southern Africa using a next generation sequencing approach

Ben J. Mans[1,a,b,c]*, Ronel Pienaar[1,a], John Ratabane[a], Boitumelo Pule[a], Abdalla A. Latif[a,d] ,

[a]Parasites, Vectors and Vector-Borne Diseases, Agricultural Research Council-Onderstepoort Veterinary Institute, Onderstepoort 0110, South Africa

[b]The Department of Veterinary Tropical Diseases, University of Pretoria, Pretoria, South Africa.

[c]Department of Life and Consumer Sciences, University of South Africa, South Africa

[d]School of Life Sciences, University of Kwa Zulu-Natal, South Africa

*Corresponding author: mansb@arc.agric.za

[1]These authors contributed equally

Abstract

Molecular classification and systematics of the *Theileria* is based on the analysis of the 18S rRNA gene. Reverse line blot or conventional sequencing approaches have disadvantages in the study of 18S rRNA diversity and a next-generation 454 sequencing approach was investigated. The 18S rRNA gene was amplified using RLB primers coupled to 96 unique sequence identifiers (MIDs). *Theileria* positive samples from African buffalo (672) and cattle (480) from southern Africa were combined in batches of 96 and sequenced using the GS Junior 454 sequencer to produce 825711 informative sequences. Sequences were extracted based on MIDs and analysed to identify *Theileria* genotypes. Genotypes observed in buffalo and cattle were confirmed in the current study, while no new genotypes were discovered. Genotypes showed specific geographic distributions, most probably linked with vector distributions. Host specificity of buffalo and cattle specific genotypes were confirmed and prevalence data as well as relative parasitemia trends indicate preference for different hosts. Mixed infections are common with African buffalo carrying more genotypes compared to cattle. Associative or incongruent co-infection profiles were observed between genotypes that may have implications for speciation and systematics: specifically that more *Theileria* species may exist in cattle and buffalo than currently recognized. Analysis of primers used for *Theileria parva* diagnostics indicate that no new genotypes will be amplified by the current primer sets confirming their specificity. *Theileria parva* SNP variants that occur in the 18S rRNA hypervariable region were confirmed. A next generation sequencing approach is useful in obtaining comprehensive knowledge regarding 18S rRNA diversity and prevalence for the *Theileria*, allowing for the assessment of systematics and diagnostic assays based on the 18S gene.

Keywords: 18S SSU, Diversity, Next-generation sequencing, species, *Theileria*, *Theileria parva*

1. Introduction

The *Babesia* and *Theileria* are part of the phylum Apicomplexa (Order Piroplasmorida) (Mans et al. 2015). The *Theileria* are distinguished by their life cycle, where tick transmitted sporozoites infect host leukocytes to form schizonts, which mature to merozoites that infect red blood cells to establish the piroplasm carrier stage (Uilenberg, 2006). Parasite morphology, host cell specificity, schizonts and piroplasm parasitemia levels, clinical disease presentation, serological and molecular methods, host and vector specificity have been used for classification and systematic purposes (Mans et al. 2015). Classification and diagnostics of the *Theileria* is important since they are the causative agents for a number of debilitating diseases in domestic and wild animals (Bishop et al. 2004), such as for example, *Theileria parva*, the causative agent of East Coast fever, Corridor disease and January disease in East and southern Africa.

A significant proportion of the systematic study of the Piroplasmorida comprise analysis of the 18S rRNA gene (Allsopp et al. 1994; Chae et al. 1999; Chansiri et al. 1999; Criado-Fornelio et al. 2004; Reichard et al. 2005; Criado et al. 2006; Altay et al. 2007; Bhoora et al. 2009). The reverse line blot (RLB) method (Gubbels et al. 1999), based on simultaneous detection of the 18S rRNA gene for various *Babesia* and *Theileria* species, has been extensively used for surveillance purposes (Georges et al. 2001; Almerìa et al. 2002; Oura et al. 2004; Nijhof et al. 2003, 2005; Altay et al. 2007; M'ghirbi et al. 2008; Altay et al. 2008; Matjila et al. 2008; Oosthuizen et al. 2008; Almerìa et al. 2009; Salih et al. 2010; Yusufmia et al. 2010; Tomassone et al. 2012; Ceci et al. 2014; Githaka et al. 2014; Eygelaar et al. 2015; Njiiri et al. 2015). The method relies on the amplification of the V4 hyper-variable region of the 18S rRNA gene, followed by hybridization to probes for various *Babesia* and *Theileria* species. Inclusion of a *Babesia*/*Theileria* catch-all allows detection of

novel genotypes when other probes are negative, but is hampered when mixed-infections are present (Mans et al. 2011a).

The ultimate gold standard for detection of genetic diversity based on the 18S rRNA gene remains direct sequencing (Mans et al. 2011a), since this is the only direct way to confirm novel genotypes. This approach has been used to study 18S rRNA genetic diversity of *Theileria* species in African buffalo and cattle from southern Africa (Mans et al. 2011a). Genotypes unique to buffalo included *T.* sp. (bougasvlei), *T.* sp. (buffalo), *T. mutans*-like 1, *T. mutans*-like 2, *T. mutans*-like 3, *T. buffeli* C, *T. sinensis*-like (*T. buffeli*-like D) and *T. velifera* B. Genotypes unique to cattle included *T. taurotragi, T. sp.* (sable)-like*, T. buffeli* Warwick and *T. buffeli-like B*. Genotypes shared between buffalo and cattle included *T. mutans*, *T. mutans* MSD, *T. parva* and *T. velifera*. Other studies supported these observations for *T.* sp. (buffalo) and *T.* sp. (bougasvlei) (Chaisi et al. 2011; Mans et al. 2011b; Pienaar et al. 2011a; Pienaar et al. 2014), *T. mutans*-like 1, *T. mutans*-like 2, *T. mutans*-like 3, *T velifera* and *T. velifera* B (Chaisi et al. 2013), *T. buffeli* C and *T. sinensis*-like (Chaisi et al. 2014). Sequence variants for *T. parva* in the V4 hyper-variable region of *T. parva* were also detected (Mans et al. 2011a). This study, however, only sampled 62 buffalo and 49 cattle and sequenced 10 clones from each, yielding ~1000 informative sequences. The current study expanded this by developing a next-generation sequencing approach for investigating 18S rRNA diversity in the *Theileria* for African buffalo and cattle and to determine whether any novel genotypes are circulating in southern Africa and what their geographic distribution may be.

The 18S rRNA gene is also important in diagnostic analysis of *Theileria* species (Mans et al. 2015). In this regard, diagnostic real-time PCR assays for *T. parva* based on the 18S rRNA gene was developed and include the hybridization and Hybrid II assays (Sibeko et al. 2008; Pienaar et al. 2011b). These assays are currently used for diagnoses in South Africa

to control and monitor Corridor disease and buffalo movement as mandated by the Department of Agriculture, Forestry and Fisheries (Animal Disease Act 1984, Act No. 35). The hybridization assay employed *T. parva* specific forward and a *Theileria* genus specific reverse primers that amplifies *T. parva*, *T.* sp. (buffalo) and *T.* sp. (bougasvlei) genotypes (Sibeko et al. 2008). The Hybrid II assay uses *Theileria* genus specific forward and *T. parva* specific reverse primers, optimized to specifically amplify *T. parva* (Pienaar et al. 2011b). The current study aimed to assess whether other genotypes may be amplified by the primers used in these assays, since it was shown that multiple templates may affect sensitivity of the diagnostic assays (Pienaar et al. 2011a).

2. Materials and Methods

2.1 Sample selection and DNA extraction

EDTA-blood samples for buffalo (n=1029) were sourced from National Parks, for cattle (n=828) were sourced from dip tanks or commercial farms, from various regions in southern Africa (Fig. 1; supplementary Table 1). Dip tanks refers to localities where communal farmers bring their animals for dipping purposes for State assisted tick control initiatives. EDTA-blood samples were extracted using the MagNAPure Large blood extraction kit using automated MagNAPure procedures and purified DNA eluted into 100 ul elution buffer (Mans et al. 2011).

2.2 Reverse line blot PCR for preparation of next-generation sequencing amplicons

A first round of PCR was performed using the RLB primers, RLBF (GAGGTAGTGACAAGAAATAACAATA) and RLBR (TCTTCGATCCCCTAACTTTC) to amplify the V4 hypervariable 18S rRNA region (Gubbels et al. 1999). GreenTaq DNA polymerase (25 ul) were used for amplification (10 pmol primer, 2.5 ul sample DNA) in a

final volume of 50 ul. The PCR consisted of initial denaturation (94°C, 2 min), 10 cycles of denaturation (94°C, 30s), touchdown annealing (67°C-57°C, 30s, one degree step down) and extension (72°C, 1 min), followed by 30 cycles of denaturation (94°C, 30s), annealing (57°C, 30s) and extension (72°C, 1 min). The PCR was rounded off at 72°C (7 min). Products were analysed using agarose gel electrophoresis and samples with bands at 400-500 bp were selected as positives. A second assembly PCR were performed using 2.5 ul of this PCR product. The forward primer consisted of a universal sequencing adapter A (CGTATCGCCTCCCTCGCGCCATCAG) linked to 96 different multiplex identifiers (MID), linked to the RLBF primer (supplementary Table S2). The reverse primer consisted of a universal sequencing adapter B (CTATGCGCCTTGCCAGCCCGCTCAG) linked to the RLBR primer. PCR consisted of initial denaturation (94°C, 2 min), 15 cycles of denaturation (94°C, 30s), annealing (57°C, 30s) and extension (72°C, 1 min), followed by extension at 72°C (7 min). Products were analysed and quantitated using agarose gel electrophoresis and visually confirmed to have incorporated the sequencing adapters by band shift analysis. Products (96 samples per batch, each with unique MID primer) were combined in equimolar ratios (10-20ul of PCR product from each sample), mixed and separated on preparative agarose gel electrophoresis by loading 10 lanes. Bands with the correct size (~600 bp) were cut from the gels, purified using a gel extraction clean-up kit, combined and quantified using nanodrop spectrophotometry and agarose gel electrophoresis. Batches were submitted to Inqaba Biotechnologies Ltd (South Africa), for sequencing on the Roche GS Junior as 96 combined samples using the universal forward primer.

2.3 Bioinformatic analysis of the sequenced amplicons

After 454 GS-Junior sequencing, samples were filtered for quality, the universal sequencing adapters trimmed and reads extracted based on their unique MID tags into 96 individual sample sets. Sequences smaller than 280 bp were discarded, since these did not cover the 18S rRNA V4 hyper-variable region. A non-redundant database of *Theileria* and *Babesia* 18S rRNA sequences were constructed by retrieving all 18S rRNA sequences using BLASTN analysis (Altschul et al. 1990). Neighbor-joining analysis using Mega 5 (Tamura et al. 2011) was performed to identify a representative from each monophyletic clade to remove redundant sequences. Sequences were trimmed to include the V4 hyper-variable region and served as non-redundant database. Sequences from all batches were queried against this database using BLASTN analysis. Sequences that aligned across the V4 hyper-variable region and showed 98-100% identity (1-2 SNP's) were considered as positive hits. Sequences that did not give positive hits were retrieved and assembled to identify novel genotypes. Assembly was performed using the BLAST tool and CAP3 assembler (Ribeiro et al. 2006). Sequences were clustered which showed 99% identity over 95% of the sequence. Consensus sequences derived from the clustered contigs were then manually analysed by BLASTN analysis against the non-redundant database, followed by neighbor-joining analysis to determine whether they are novel. Genotypes considered to be novel were then confirmed using conventional cloning and sequencing of the 18S rRNA gene as described (Mans et al. 2011a). Genotypes were then included in the non-redundant dataset and BLAST analysis of the various batched sequences were repeated until no novel genotypes were retrieved. Positive sequences for various genotypes were counted to determine the total number of sequences/genotype in each sample.

After genotypes were identified, sequence identifiers unique to each genotype (electronic signature; supplementary Table 3) were used to scan all sequences and confirm genotype identities and identify possible chimeric sequences. Chimeric sequences (those that

possess more than one genotype specific identifier) that indicate potential PCR-derived chimers were discarded from the analysis. In the case of *T. parva* variants, all *T. parva* positive sequences were extracted and SNP's were identified using specific sequence identifiers (supplementary Table 3), as well as by clustering and alignment of *T. parva* sequences and manual inspection of the hyper-variable region.

2.4 Determination of a cut-off value for each genotype

Parasitemia ranges for *T. parva*, *T.* sp. (buffalo) and *T.* sp. (bougasvlei) were shown to exhibit normal distribution curves (Pienaar et al. 2011a; Pienaar et al. 2014). Assuming that parasitemia ranges for all *Theileria* species will exhibit normal distribution curves (Mans et al. 2015), frequency distribution plots were constructed for all genotypes and cut-offs were determined to include 97% of all sequences for each genotype (supplementary Fig. S1). In some cases the lower 3% of all sequences made up an extensive proportion of the sequences below the cut-off and implies that the carrier-state parasitemia for these genotypes falls over the cut-off range. For these genotypes the prevalence will be under-estimated, but will minimize false positives. To validate this method, results were also compared to the real-time hybridization assay for *T. parva*, currently used as diagnostic assay in the PVVD laboratory (Pienaar et al. 2011). Real-time hybridization PCR assays for *T. parva* was performed as described (Sibeko et al. 2008; Pienaar et al. 2011a).

3. Results and Discussion

3.1 Sample screening using RLB primers

Samples were screened using the RLB primer set and this identified 1150 as *Theileria* or *Babesia* positive, of which 672 were buffalo and 478 cattle. This represent a ~10 fold increase in sample coverage from the conventional sequencing approach and a larger geographic sampling of southern Africa (Mans et al. 2011a). Of interest was that buffalo and cattle from the Northern Cape region, as well as buffalo from the Free State were all negative, suggesting that the main vector ticks found in this region, namely, *Hyalomma rufipes*, *H. truncatum*, *Rhipicephalus decoloratus* and *R. evertsi* (Howell et al. 1978; Horak et al. 2007; Spickett, 2013; Horak et al. 2015), are not major vectors for buffalo or cattle *Theileria* species, or that bovines from these areas were not previously exposed to infected ticks.

3.2 Sequence statistics

The 672 positive buffalo samples were allocated to 7 batches of 96 samples each (Batch 1-7) and the 478 positive cattle samples to 5 batches (Batch 8-12) (supplementary Table 1). Each batch was sequenced on a GS Junior 454 run, which is expected to yield ~100 000 reads with an average length of ~400 bp (Glenn, 2011). The total reads (87692 ± 13446) and read length (381 ± 31 bp) was consistent with this estimate (Table 1). However, for informative reads (> 280 bp) the read length was larger (424 ± 22 bp). Sequences that possessed more than one electronic signature comprised from 0.004-0.15%, indicating that chimeric sequences were minimal and were removed from the final analysis (Table 1). After removal of sequences that fall below the cut-off, the total sequences obtained were ~79% of the initial sequences and ~94% of the informative sequences. This comprised 825711 total sequences and ~778 fold increase in sequence coverage from the conventional sequencing approach that sampled ~1061 clones from 62 buffalo and 49 cattle (Mans et al. 2011a).

3.3 Comparison of cut-off values and the real-time PCR assay

Results using the determined cut-off values compared well to the gold-standard real-time PCR assay for *T. parva*, even though it is less sensitive (Fig. 2A). Of 668 buffalo samples tested with the *T. parva* real-time PCR, 97% correlation was found for those samples considered to be true negatives with 55% having zero sequences and 42% sequences below the cut-off. Three percent (6 samples) had sequences above the cut-off. For the true positive samples, 83% gave sequences above the cut-off, while 21.7% gave sequences below the cut-off and would be considered false-negatives. This correlated well with the cut-offs determined using the frequency distribution plots and suggest that the latter method of determining a cut-off is appropriate as well as conservative, i.e. the method would under-estimate the number of true positive samples to an extent of ~20%, while detection of false positives will be limited. No correlation between the real-time PCR parasitemia calculated from the CP values (Pienaar et al. 2011a), and the number of sequences obtained using 454 sequencing could be found (Fig. 2B). This indicates that the 454 approach would not be suitable for absolute quantification and may be ascribed to the presence of mixed infections, PCR competition and PCR suppression by more abundant templates. This was observed for RLB analysis, which only detected 37% positive samples, compared to ~80% with the real-time PCR (Pienaar et al. 2011).

3.4 Genotypes described

Genotypes for four main clades were found in buffalo and cattle in southern Africa, namely the *T. taurotragi* clade (*T. parva*, *T.* sp. (buffalo), *T.* sp. (bougasvlei) and *T. taurotragi*), the *T. mutans* clade (*T. mutans*, *T. mutans* MSD, *T. mutans*-like 1, *T. mutans*-like 2 and *T. mutans*-like 3), *T. velifera* clade (*T. velifera*, *T. velifera* A and *T. velifera* B) and the

T. buffeli clade (*T. buffeli* Warwick, *T. buffeli* B, *T. buffeli* C and *T. sinensis*-like) (Mans et al. 2011a; Mans et al. 2015). The current study identified eighteen genotypes: eight buffalo specific, five cattle specific and five found in buffalo and cattle, as well as *T. parva* variants 1, 3, 5 and 7 (Fig. 3). Genotypes unique to buffalo included *T.* sp. (buffalo), *T.* sp. (bougasvlei), *T. mutans*-like 1, *T. mutans*-like 2, *T. mutans*-like 3, *T. buffeli* C, *T. sinensis*-like and *T. velifera* A (Fig. 3). Genotypes unique to cattle included *T. buffeli* Warwick, *T. buffeli* B, *T. taurotragi*, *B. bigemina* and *B. bovis*. Genotypes found in buffalo and cattle included *T. parva*, *T. parva* variant 1, *T. parva* variant 3, *T. parva* variant 5, *T. parva* variant 7, *T. mutans*, *T. mutans* MSD, *T. velifera* and *T. velifera* B. No unique genotypes were obtained in the present study, indicating, perhaps surprisingly, that sampling of 49 cattle, 62 buffalo and sequencing 10 clones from each had sufficient sequencing depth to cover the sequence diversity observed in buffalo and cattle (Mans et al. 2011a). Conventional sequencing for genetic diversity characterization of the *Theileria* may be useful, where next-generation sequencing technology is not available.

3.5 Geographic distribution of genotypes

Babesia species

No *Babesia* genotypes were found in buffalo and corresponds to other RLB studies (Mans et al. 2011a; Chaisi et al. 2011; Pienaar et al. 2014). Conversely, *B. bigemina* and *B. bovis* were prevalent in cattle with geographic distributions similar to the main tick vectors for these species in South Africa, namely *R. decoloratus* and *R. microplus* (Combrink et al. 2014).

The T. taurotragi clade

The geographic distribution of *T.* sp. (buffalo) and *T.* sp. (bougasvlei) (Fig. 3), were similar as found using real-time PCR assays (Pienaar et al. 2014). The incongruent distribution (Pienaar et al. 2014), was observed with the 454 data in that only ~9% of all positives showed mixed infections for *T.* sp. (buffalo) and *T.* sp. (bougasvlei), even in sympatric regions, such as the Kruger National Park and Niassa National Reserve. Other localities were exclusive for *T.* sp. (buffalo) (Hluhluwe Nature Reserve, Marakele National Park), or *T.* sp. (bougasvlei) (Chobe National Park, Hwangwe National Park, Waterberg Platue Park). This was mainly attributed to genetic incompatibility in the same vector tick species, though a number of other possibilities were considered (Pienaar et al. 2014). No cattle were positive for *T.* sp. (buffalo) or *T.* sp. (bougasvlei), similar to other studies (Yusufmia et al. 2010; Pienaar et al. 2014), suggesting that cattle in southern Africa are not natural carriers. Conversely, *T.* sp. (buffalo) has been detected by RLB analysis in cattle co-grazing with buffalo in Kenya and are able to infect and transform cattle lymphocytes (Githaka et al. 2014; Bishop et al. 2015).

Theileria taurotragi was only found in cattle where *R. appendiculatus*, its vector tick is present (Pienaar et al. 2011a). In contrast, buffalo sampled in the Corridor disease endemic region was infected with *T. parva*, while the only cattle infected with *T. parva* correspond to Corridor disease outbreaks in Kwa-Zulu Natal (Mbizeni et al. 2013). This suggest that buffalo are not carriers of *T. taurotragi*, and detection in buffalo may be incidental (Chaisi et al. 2013).

A number of SNP's in the 18S hypervariable probe region of the hybridization and Hybrid II assays for *T. parva* are of particular interest with regard to diagnostics and epidemiology (Mans et al. 2011a). *T. parva* variants 1, 3 and 5 were found in cattle samples, while variant 1, 3 and 7 were found in buffalo. Variants 2, 4 and 6 were not found, suggesting that these were probably sequencing errors as suggested (Mans et al. 2011a). *T. parva* variant

1 (61 animals), 3 (17 animals), 5 (9 animals) and 7 (3 animals) occurred in 14.6%, 4.0%, 2.1% and 0.7%, respectively, of all *T. parva* infected animals. These results corresponds with results which indicated that variants 1, 3 and 5 could be confirmed by aberrant melting curves using real-time PCR (Mans et al. 2011a), and suggest that these may be true circulating SNP's in the *T. parva* population. Of interest is the high prevalence of *T. parva* variant 1 in the Chobe, Hwangwe and Marakele National Park. It was suggested that infection of buffalo in Marakele could have been a founder effect (Pienaar et al. 2011), which may explain the high prevalence of variant 1 in these parks. Deeper sequencing and larger sampling would, however, be necessary to determine the prevalence of specific variants for either buffalo or cattle.

The T. mutans clade

Theileria mutans-like 1, *T. mutans*-like 2 and *T. mutans*-like 3 were found in buffalo in all localities except for Niassa National Park, Karoo Nature Reserve, Waterberg Platue Park, parks from the Eastern Cape region (Addo Elephant Park, E'zulu Game Reserve, Great Fish River Reserve) and the North West Province (Borakalalo National Park, Molemane Nature Reserve). *Theileria mutans* and *T. mutans* MSD were found in buffalo, but at lower prevalence in buffalo than to cattle compared to the buffalo specific genotypes. All genotypes from the *T. mutans* clade have been shown to be vectored by *Amblyomma* spp. (Uilenberg 1974; Purnell et al. 1975; Uilenberg, 1976; Young et al. 1977; Young et al. 1978; Perie et al. 1979; de Vos and Roos, 1981). It is therefore of interest that *T. mutans* and *T. mutans* MSD were extensively found in the Eastern Cape in buffalo and cattle, whereas buffalo specific genotypes were absent from this region. The main vector in southern Africa for these genotypes is considered to be *A. hebraeum* (de Vos et al. 1981), which have a large geographic distribution (Horak et al. 2007; Spickett, 2013), with the implication that the buffalo specific genotypes would have been expected to occur in the Eastern Cape. The

absence of buffalo specific genotypes may be due to different tick vectors, or historical segregation of Corridor disease positive (Kruger National Park and Hluhluwe stocks) and negative (Addo Elephant Park stocks) buffalo in South Africa (Laubscher and Hoffman, 2012). The absence of the *T. mutans* clade genotypes in Niassa National Reserve (Mozambique) and Waterberg Platue Park (Namibia), would support the latter possibility, since both *A. hebraeum* and *A. variegatum* are absent from these regions (Walker and Olwage, 1987).

The T. buffeli clade

For the *T. buffeli* clade, a clear demarcation of genotypes exist. *Theileria buffeli* C and *T. sinensis*-like (*T. buffeli* D-like, Mans et al. 2011a), were found in buffalo, while *T. buffeli* Warwick and *T. buffeli* B were found in cattle. The distribution of *T. buffeli* C and *T. sinensis*-like was confined to Chobe National Park, Niassa National Reserve and the Eastern Cape region (Addo Elephant Park, E'zulu Game Reserve, Great Fish River Reserve). A few positive animals were found in Hluhluwe Nature Reserve, Marakele National Park, Borakalalo National Park and Molemane Nature Reserve. In contrast, *T. buffeli* and *T. buffeli* B was mainly found in the Vryheid district of KwaZulu Natal, Hazyview diptanks in Limpopo and the Port Elizabeth region (Eastern Cape). This suggest that the buffalo and cattle specific genotypes may be different species and that the *T. buffeli* clade consist of more species than recognized (Mans et al. 2011a; Mans et al. 2015). Various *Haemaphysalis* species have been implicated in transmission of *T. buffeli* clade genotypes and include *Haemaphysalis bancrofti*, *H. bispinosa*, *H. concinna*, *H. cornigera*, *H. humerosa, H. japonica*, *H. longicornis, H. mageshimaensis*, *H. punctate* and *H. qinghaiensis* (Shastri et al. 1985; Stewardt et al. 1987; Stewardt et al. 1989; Fujisaki et al. 1994; Yin et al. 2004; Hammer et al. 2015). None of these species occur in Africa and the only species from this genus that parasitize cattle or buffalo is *H. silacea* which occurs in the Eastern Cape and parts

of Kwa-Zulu Natal (Walker 1991; Horak et al. 2007). It could be a potential vector for *T. buffeli* Warwick/*T. buffeli* B in cattle and *T. buffeli* C/*T. sinensis*-like in buffalo (Chaisi et al. 2014). The extralimital presence of the different genotypes in Limpopo, Botswana and Mozambique suggest that other tick species may be vectors as well.

The T. velifera clade

Theileria velifera and *T. velifera* B were found in buffalo and cattle, while *T. velifera* A was found in buffalo. This differed from studies where *T. velifera* B were only found in buffalo (Mans et al. 2011a; Chaisi et al. 2013). *Amblyomma* spp. transmits *T. velifera* (Uilenberg and Schreuder, 1976), and the absence of *T. velifera* and *T. velifera* B in Niassa National Reserve (Mozambique) and Waterberg Platue Park (Namibia) fits with this, similar to the *T. mutans* clade. Of interest, was the presence of *T. velifera* A in Mozambique, suggesting that this genotype is vectored by a different tick species. This genotype was also found in buffalo from Niassa using a conventional sequencing approach (Mans et al. 2011a). The *T. velifera* A genotype was also found in 12 Nyala, 7 Eland and 10 Kudu samples from South Africa, using the 454 next-generation sequencing approach, while no *T. velifera* or *T. velifera* B genotypes were found in these samples (manuscript in preparation). Of interest, is that RLB screening of Nyala, Eland and Kudu also did not detect *T. velifera* in areas endemic for *A. hebraeum* and *A. variegatum* (Oura et al. 2011; Pfitzer et al. 2011; Berggoetz et al. 2014). This suggest that the preferential host of *T. velifera* A may be antelope and an antelope tick vector.

Genotypes not found

It should be noted that *T.* sp. (sable) genotypes were not found in the current study, even though several RLB based studies indicated the presence of this genotype in cattle and buffalo (Nijhof et al. 2005; Yusufmia et al. 2010; Eygelaar et al. 2015; Njiiri et al. 2015). It

was indicated that the *T.* sp. (sable) probe cross-hybridize with the *T. velifera* probe (Mans et al. 2011a). The current study confirmed this, indicating that RLB based analysis should be approached with caution. The only sequence confirmed case for a cattle sample infected with a *T.* sp. (sable)-like genotype, were an animal that showed apparent signs of theileriosis (Mans et al. 2011a), suggesting an immuno-compromised incidental infection.

3.6 Prevalence and relative parasitemia trends

Relative trends in parasitemia may be observed from the data. As such, buffalo specific genotypes are not only prevalent, but also show a higher numbers of sequences compared to genotypes found in both cattle and buffalo (Fig. 4). More sequences were found for *T. mutans*-like 1/2/3 compared to *T. mutans* and *T. mutans* MSD, suggesting that these latter genotypes generally occur at lower parasitemia in buffalo than buffalo specific genotypes. This is of interest since, RLB based studies indicated a high prevalence of *T. mutans* in buffalo (Oura et al. 2011; Chaisi et al. 2011; Pienaar et al. 2014), but this has probably been confounded by the non-specificity of the *T. mutans* probe that also detect *T. mutans*-like 1/2/3 (Mans et al. 2011a). In this regard, buffalo specific genotypes show sub-structuring with regard to co-infection compared to *T. mutans* and *T. mutans* MSD. Buffalo specific genotypes show co-infection rates of 80-87% (Fig. 5). In contrast, co-infection with *T. mutans* and *T. mutans* MSD range from 4-5% which is comparable to that observed for *T.* sp. (buffalo) and *T.* sp. (bougasvlei) (Pienaar et al. 2014), suggesting that discrepant co-infection patterns also occur in the *T. mutans* clade.

While *T. velifera* is prevalent in both buffalo and cattle, higher numbers of sequences are retrieved in cattle compared to buffalo samples, suggesting that this genotype is dominant in cattle samples, but not in buffalo. Better estimations of parasitemia ranges may in future be

possible using genotype specific real-time PCR assays or improved next-generation sequencing technologies. In this regard, specific assays for *T. parva*, *T.* sp. (buffalo) and *T.* sp. (bougasvlei) indicated that their parasitemia levels range from 0.001%-0.1% in carrier animals (Pienaar et al. 2011a; Pienaar et al. 2011b; Pienaar et al. 2014). The latter observations are supported by data from the current study and suggest that genotypes from the *T. mutans* and *T. velifera* clades are present at much higher parasitemia.

3.7 *Number of genotypes per animal and co-infection patterns*

Previously it was shown that mixed infections were prevalent in buffalo and cattle and conventional sequencing found more genotypes than RLB analysis, while the average number of genotypes were higher in buffalo compared to cattle (Mans et al. 2011a). The current study confirmed these general observations, but found up to 7 genotypes in cattle (91% from 1-5 genotypes, peak at 4 genotypes) and 10 genotypes in buffalo (75% from 3-8 genotypes, peak at 7 genotypes) (Fig. 5A). For both this was higher than the previous observed number of genotypes found per animal (Mans et al. 2011a), but confirmed the observation that buffalo on average carries a higher diversity of genotypes than cattle.

A number of studies have commented on the patterns of co-infection observed between different parasite genotypes or species (Dib et al. 2008; Vaumourin et al. 2014; Njiiri et al. 2015). Given that the *Theileria* may have life-long carrier status, co-infection or lack thereof can be ascribed to the same vector or vector distribution, vector infectivity and transmission efficiency, similar host preferences, host age, parasitemia ranges, ability to propagate asexually in the host, mechanisms of immune evasion or host immunity, competition for the same resources or genetic incompatibility between parasites (Pienaar et al. 2014). In the current study we define levels of co-infection as low (<10%), moderate (10-

40%) and high (>40%) (Fig. 5). Low levels of co-infection may represent poor host preference as seen for the buffalo and cattle specific genotypes. Differences in vector distribution could account for low levels of co-infection as seen for the buffalo specific *T.* sp. (buffalo), *T.* sp. (bougasvlei), *T. parva*, *T. mutans* 1/2/3 versus the *T. buffeli* C/*T. sinensis*-like genotypes or the cattle specific *T. taurotragi* versus *T. buffeli* Warwick/*T. buffeli* B genotypes. Genetic incompatibility may account for low levels of co-infection as suggested for *T.* sp. (buffalo) and *T.* sp. (bougasvlei) (Pienaar et al. 2014). The current study also suggest that this may be the case for *T. mutans* 1/2/3 versus *T. mutans* and *T. mutans* MSD or *T. velifera* A versus *T. velifera* and *T. velifera* B. In contrast, in areas considered to be endemic for certain genotypes, moderate to high levels of co-infection were observed. This was the case with most buffalo specific genotypes (*T. parva*, *T.* sp. (buffalo), *T.* sp. (bougasvlei), *T. mutans* 1/2/3, *T. velifera* and *T.* velifera B) and the same or common vector distributions probably account for this. In some instances, genotypes from the same clades show high levels of co-infection as observed for *T. mutans* 1/23, *T. velifera*/*T. velifera* B, *T. buffeli* C/*T. sinensis*-like, *T. buffeli* Warwick/*T. buffeli* B. This may be accounted for by the same vectors or possibly different genes in the same genome as discussed below.

3.8 Are unique genotypes unique species?

A central question in the study of 18S rRNA diversity are whether a unique genotype represents a unique species? It is generally recognized that the 18S rRNA gene is conserved within species and that as little as 3 nucleotide differences may differentiate *Theileria* species (Mans et al. 2011a; Mans et al. 2015). In the *T. taurotragi* clade, different species are well recognized, even though this clade shows the least intra-specific genetic diversity for the 18S rRNA gene. Conversely, while the *T. buffeli*, *T. mutans* and *T. velifera* clades show much

higher intra-specific diversity, they are generally considered as representing a single species, i.e. *T. buffeli*, *T. mutans* and *T. velifera* (Mans et al. 2015). If this is indeed the case, an explanation for the genetic diversity observed is needed.

The 18S rRNA gene occurs as two copies in all *Theileria* genomes, including *T. annulata*, *T. equi*, *T. parva* and *T. orientalis* (Gardner et al., 2005; Pain et al., 2005; Hayashida et al., 2012; Kappmeyer et al., 2012). There is therefore no reason to postulate that the extensive genetic diversity observed is due to more than two 18S rRNA copies, as found in *Babesia* and *Plasmodium* (Gunderson et al. 1987; Lau, 2009; Laughery et al. 2009). The two gene copies are also conserved due to concerted evolution (Eickbush and Eickbush, 2007). If the genotypes existed as multiple genes in the genome, they should be co-detected in samples and their quantitative ratio relative to each other would represent their gene copy number. Pairwise comparison of intra-specific co-localization of genotypes found in the different clades indicates a poor correlation that does not support multiple genes in the same genome (Fig. 5). As such, *T. mutans*-like 1/2/3 and *T. mutans*/*T. mutans* MSD do not co-occur at high incidence. In the case of *T. mutans*-like 1/2/3, a high level of co-occurrence was observed that ranged from 80-87%. However, *T. mutans*-like 3 was absent from some geographic localities. Similarly, although high levels of co-occurrence was found for *T. velifera* and *T. velifera* B (63%), it hardly approached levels expected for co-occurrence in the same genome. The same holds for the *T. buffeli*-like genotypes. Evidence for different genotypes in the same genome is therefore not convincing with the current dataset.

If genotypes were alleles, their prevalence will be determined by population genetic factors and prevalence data cannot distinguish species. However, if genotypes within clades represent the same species, host specificity would not be expected. Specificity of *T. mutans*-like1/2/3, *T. buffeli* C/*T. sinensis*-like and *T. velifera* A for buffalo argue against them being alleles of *T. mutans*, *T. mutans* MSD and *T. velifera*, respectively. This suggest that different

species do exist within the *T. buffeli*, *T. mutans* and *T. velifera* clades, with the presence of at least buffalo and cattle specific species. Whether the genotypes specific to cattle or buffalo represent the same species cannot as yet be resolved until more genes are analysed.

3.9 Assessment of the hybridization and Hybrid II assays for T. parva detection

BLASTN analysis and motif detection of the *T. parva* specific primer sequences against a database of the 454 sequences were performed. The *T. parva* specific forward primer (5'-CTGCATCGCTGTGTCCCTT-3') used in the hybridization assay was only detected in *T. parva*, *T.* sp. (buffalo) and *T.* sp. (bougasvlei) sequences. Similarly, the *T. parva* specific reverse primer (5'-AAAGTAAACATCCAGACAAAGCG-3') employed in the Hybrid II assay was only detected in *T. parva* sequences. This confirmed that these primers are specific for the current genotypes amplified and validate the specificity parameters of these assays.

4. *Conclusions*

The study of 18S rRNA diversity has practical application in the design of diagnostics assays based on this gene. Finite knowledge regarding the genotypes that circulate in buffalo and cattle allows for an objective assessment of the specificity of real-time PCR assays such as the hybridization and Hybrid II assays. The current study confirmed the prevalence and identity of *Theileria* genotypes that circulate in buffalo and cattle in southern Africa. It confirmed that the only templates that possess the primers used in the hybridization PCR assay remain *T. parva*, *T.* sp. (buffalo) and *T.* sp. (bougasvlei), and that *T. parva* is the only genotype that possess the Hybrid II assay primer pair. No other genotype circulating in buffalo or cattle in southern Africa will as such, affect the hybridization or Hybrid II assay

with regard to specificity or PCR suppression. No novel *T. parva* variants within the probe areas were detected in the current study. These assays therefore remain methods of choice in the diagnoses of buffalo-derived *T. parva* and determination of the disease free status of buffalo herds in South Africa.

Acknowledgements

We thank Dr. Roy Bengis, Kruger National Park for buffalo samples. This project was funded by the Theileria diagnostics project ARC-OVI (OV15/08/1P01) and the Department of Agriculture, Forestry and Fisheries (OV21/03/C148).

Figures

Figure 1: Sampling sites for cattle and buffalo samples used in study. The number of animals sampled per site are indicated in circles (buffalo) or rectangles (cattle). Provinces in South Africa are indicated in dotted circles and include Western Cape (WC), Eastern Cape (EC), Northern Cape (NC), Free State (FS), Kwa-Zulu Natal (KZN), North-West (NW), Gauteng (GP), Limpopo (LP) and Mpumalanga (MP). Refer to Table S1 to identify specific localities indicated in the map.

Figure 2: Comparison of the 454 data and the hybridization PCR assay for *T. parva*. A) Indicated are the expected true negative and true positive samples as determined with the hybridization PCR assay and the number of samples above or below the determined cut-off threshold (10 sequences) for the 454 data. The horizontal dashed line indicate the cut-off threshold and the vertical dashed line indicate the demarcation between true negative and positive samples. B) Comparison of parasitemia and number of sequences obtained using 454 sequencing. Indicated are the R^2 value for the linear regression fit.

Figure 3: A presence-absence heat map of different *Theileria* and *Babesia* genotypes. Vertical axes: Buffalo and cattle samples and their origins. Horizontal axes: Genotypes (1-22) divided into buffalo specific, buffalo and cattle genotypes and cattle specific genotypes. 1) *T.* sp. (buffalo), 2) *T.* sp. (bougasvlei), 3) *T.* mutans-like 1, 4) *T.* mutans-like 2, 5) *T.* mutans-like 3, 6) *T. buffeli* C, 7) *T. sinensis*-like, 8) *T. velifera* A, 9) *T. parva*, 10) *T. parva variant* 1, 11) *T. parva* variant 3, 12) *T. parva* variant 5, 13) *T. parva* variant 7, 14) *T. mutans*, 15) *T. mutans* MSD, 16) *T. velifera*, 17) *T. velifera* B, 18) *T. buffeli* Warwick, 19) *T. buffeli* B, 20) *T. taurotragi*, 21) *B. bovis*, 22) *B. bigemina*. Grey shading indicates presence and white absence.

Figure 4: Comparison of the total number of animals and sequences for each genotype. Indicated are the number of animals (top), number of sequences (middle) and average number of sequences per animal.

Figure 5: The number of genotypes found per animal. A) Indicated are a frequency distribution curve for the number of genotypes found per animal in cattle and buffalo. B) Co-infection observed between different genotype pairs. Indicated are the percentage co-infection observed for low (white), moderate (gray) and high (dark gray) for different genotype pairs. Genotypes are indicated for *T.* sp. (buffalo) (TsBuff), *T.* sp. (bougasvlei) (TsBgvl), *T. mutans*-like 1 (Tm1), *T. mutans*-like 2 (Tm2), *T. mutans*-like 3 (Tm3), *T. buffeli* C (TbC), *T. sinensis*-like (Tcfsin), *T. velifera* A (TvA), *T. parva* (Tpar), *T. mutans* (Tm), *T. mutans* MSD (TmMSD), *T. velivera* (Tv), *T. velifera* B (TvB), *T. buffeli* Warwick (TbW), *T. buffeli* B (TbB), *T. taurotragi* (Ttau), *B. bigemina* (Bbig), *B. bovis* (Bbov).

References

Allsopp, M.T., Cavalier-Smith, T., De Waal, D.T., Allsopp, B.A., 1994. Phylogeny and evolution of the piroplasms. Parasitology 108, 147-152.

Almeria, S., Castella, J., Ferrer, D., Gutierrez, J.F., Estrada-Pena, A., Sparagano, O., 2002. Reverse line blot hybridization used to identify hemoprotozoa in Minorcan cattle. Ann. N. Y. Acad. Sci. 969, 78-82.

Almeria, S., Delgado-Neira, Y., Adelantado, C., Huguet, M., Vinent, J., Nicolas, A., 2009. Mediterranean theileriosis and other tick transmitted piroplasmoses in cattle in Minorca (Balearic Islands, Spain): the effect of tick control on prevalence levels analyzed by reverse line blot (RLB) macroarrays. J. Parasitol. 95, 598-603.

Altay, K., Aydin, M.F., Dumanli, N., Aktas, M., 2008. Molecular detection of *Theileria* and *Babesia* infections in cattle. Vet. Parasitol. 158, 295-301.

Altay, K., Dumanli, N., Aktas, M., 2007. Molecular identification, genetic diversity and distribution of *Theileria* and *Babesia* species infecting small ruminants. Vet. Parasitol. 147, 161-165.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403-410.

Berggoetz, M., Schmid, M., Ston, D., Wyss, V., Chevillon, C., Pretorius, A.M., Gern, L., 2014. Tick-borne pathogens in the blood of wild and domestic ungulates in South Africa: interplay of game and livestock. Ticks Tick Borne Dis. 5, 166-175.

Bhoora, R., Franssen, L., Oosthuizen, M.C., Guthrie, A.J., Zweygarth, E., Penzhorn, B.L., Jongejan, F., Collins, N.E., 2009. Sequence heterogeneity in the 18S rRNA gene within *Theileria equi* and *Babesia caballi* from horses in South Africa. Vet. Parasitol. 159, 112-120.

Bishop, R.P., Hemmink, J.D., Morrison, W.I., Weir, W., Toye, P.G., Sitt, T., Spooner, P.R., Musoke, A.J., Skilton, R.A., Odongo, D.O., 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, 333-342.

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, S271-S283.

Ceci, L., Iarussi, F., Greco, B., Lacinio, R., Fornelli, S., Carelli, G., 2014. Retrospective study of hemoparasites in cattle in southern Italy by reverse line blot hybridization. J. Vet. Med. Sci. 76, 869-875.

Chae, J.S., Allsopp, B.A., Waghela, S.D., Park, J.H., Kakuda, T., Sugimoto, C., Allsopp, M.T., Wagner, G.G., Holman, P.J., 1999. A study of the systematics of *Theileria* spp. based upon small-subunit ribosomal RNA gene sequences. Parasitol. Res. 85, 877-883.

Chaisi, M.E., Collins, N.E., Oosthuizen, M.C., 2014. Phylogeny of *Theileria buffeli* genotypes identified in the South African buffalo (*Syncerus caffer*) population. Vet. Parasitol. 204, 87-95.

Chaisi, M.E., Collins, N.E., Potgieter, F.T., Oosthuizen, M.C., 2013. Sequence variation identified in the 18S rRNA gene of *Theileria mutans* and *Theileria velifera* from the African buffalo (*Syncerus caffer*). Vet. Parasitol. 191, 132-137.

Chaisi, M.E., Sibeko, K.P., Collins, N.E., Potgieter, F.T., Oosthuizen, M.C., 2011. Identification of *Theileria parva* and *Theileria* sp. (buffalo) 18S rRNA gene sequence

variants in the African Buffalo (*Syncerus caffer*) in southern Africa. Vet. Parasitol. 182, 150-162.

Chansiri, K., Kawazu, S., Kamio, T., Terada, Y., Fujisaki, K., Philippe, H., Sarataphan, N., 1999. Molecular phylogenetic studies on *Theileria* parasites based on small subunit ribosomal RNA gene sequences. Vet. Parasitol. 83, 99-105.

Combrink, M.P., Troskie, P.C., Pienaar, R., Latif, A.A., Mans, B.J., 2014. Genotypic diversity in *Babesia* bovis field isolates and vaccine strains from South Africa. Vet. Parasitol. 199, 144-152.

Criado, A., Martinez, J., Buling, A., Barba, J.C., Merino, S., Jefferies, R., Irwin, P.J., 2006. New data on epizootiology and genetics of piroplasms based on sequences of small ribosomal subunit and cytochrome b genes. Vet. Parasitol. 142, 238-247.

Criado-Fornelio, A., Gonzalez-del-Rio, M.A., Buling-Sarana, A., Barba-Carretero, J.C., 2004. The "expanding universe" of piroplasms. Vet. Parasitol. 119, 337-345.

De Vos, A.J., Roos, J.A., 1981. Observations on the transmission of *Theileria mutans* in South Africa. Onderstepoort J. Vet. Res. 48, 1-6.

Dib, L., Bitam, I., Tahri, M., Bensouilah, M., De Meeus, T., 2008. Competitive exclusion between piroplasmosis and anaplasmosis agents within cattle. PLoS Pathog. 4, e7.

Eickbush, T.H., Eickbush, D.G., 2007. Finely orchestrated movements: evolution of the ribosomal RNA genes. Genetics 175, 477-485.

Eygelaar, D., Jori, F., Mokopasetso, M., Sibeko, K.P., Collins, N.E., Vorster, I., Troskie, M., Oosthuizen, M.C., 2015. Tick-borne haemoparasites in African buffalo (*Syncerus caffer*) from two wildlife areas in Northern Botswana. Parasit. Vectors 8, 26-014-0627-y.

Fujisaki, K., Kawazu, S., Kamio, T., 1994. The taxonomy of the bovine *Theileria* spp. Parasitol. Today 10, 31-33.

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

Georges, K., Loria, G.R., Riili, S., Greco, A., Caracappa, S., Jongejan, F., Sparagano, O., 2001. Detection of haemoparasites in cattle by reverse line blot hybridisation with a note on the distribution of ticks in Sicily. Vet. Parasitol. 99, 273-286.

Githaka, N., Konnai, S., Bishop, R., Odongo, D., Lekolool, I., Kariuki, E., Gakuya, F., Kamau, L., Isezaki, M., Murata, S., Ohashi, K., 2014. Identification and sequence characterization of novel *Theileria* genotypes from the waterbuck (*Kobus defassa*) in a *Theileria parva*-endemic area in Kenya. Vet. Parasitol. 202, 180-193.

Glenn, T.C., 2011. Field guide to next-generation DNA sequencers. Mol. Ecol. Resour. 11, 759-769.

Gubbels, J.M., de Vos, A.P., van der Weide, M., Viseras, J., Schouls, L.M., de Vries, E., Jongejan, F., 1999. Simultaneous detection of bovine *Theileria* and *Babesia* species by reverse line blot hybridization. J. Clin. Microbiol. 37, 1782-1789.

Gunderson, J.H., Sogin, M.L., Wollett, G., Hollingdale, M., de la Cruz, V.F., Waters, A.P., McCutchan, T.F., 1987. Structurally distinct, stage-specific ribosomes occur in *Plasmodium*. Science 238, 933-937.

Hammer, J.F., Emery, D., Bogema, D.R., Jenkins, C., 2015. Detection of *Theileria orientalis* genotypes in *Haemaphysalis longicornis* ticks from southern Australia. Parasit. Vectors 8, 229-015-0839-9.

Hayashida, K., Hara, Y., Abe, T., Yamasaki, C., Toyoda, A., Kosuge, T., Suzuki, Y., Sato, Y., Kawashima, S., Katayama, T., Wakaguri, H., Inoue, N., Homma, K., Tada-Umezaki, M., Yagi, Y., Fujii, Y., Habara, T., Kanehisa, M., Watanabe, H., Ito, K., Gojobori, T., Sugawara, H., Imanishi, T., Weir, W., Gardner, M., Pain, A., Shiels, B., Hattori, M., Nene, V., Sugimoto, C., 2012. Comparative genome analysis of three eukaryotic parasites with differing abilities to transform leukocytes reveals key mediators of *Theileria*-induced leukocyte transformation. MBio 3, e00204-12.

Horak, I.G., Golezardy, H., Uys, A.C., 2007. Ticks associated with the three largest wild ruminant species in southern Africa. Onderstepoort J. Vet. Res. 74, 231-242.

Horak, I.G., Jordaan, A.J., Nel, P.J., van Heerden, J., Heyne, H., van Dalen, E.M., 2015. Distribution of endemic and introduced tick species in Free State Province, South Africa. J. S. Afr. Vet. Assoc. 86, 1255.

Howell, C.J., Walker, J.B., Nevill, E.M., 1978. Ticks, mites and insects infesting domestic animals in South Africa. Part 1. Descriptions and biology.

Kappmeyer, L.S., Thiagarajan, M., Herndon, D.R., Ramsay, J.D., Caler, E., Djikeng, A., Gillespie, J.J., Lau, A.O., Roalson, E.H., Silva, J.C., Silva, M.G., Suarez, C.E., Ueti, M.W., Nene, V.M., Mealey, R.H., Knowles, D.P., Brayton, K.A., 2012. Comparative genomic analysis and phylogenetic position of *Theileria* equi. BMC Genomics 13, 603-2164-13-603.

Lau, A.O., 2009. An overview of the *Babesia*, *Plasmodium* and *Theileria* genomes: a comparative perspective. Mol. Biochem. Parasitol. 164, 1-8.

Laubscher, L., Hoffman, L., 2012. An overview of disease-free buffalo breeding projects with reference to the different systems used in South Africa. Sustainability 4, 3124-3140.

Laughery, J.M., Lau, A.O., White, S.N., Howell, J.M., Suarez, C.E., 2009. *Babesia bovis*: transcriptional analysis of rRNA gene unit expression. Exp. Parasitol. 123, 45-50.

Mans, B.J., Pienaar, R., Latif, A.A., 2015. A review of *Theileria* diagnostics and epidemiology. Int. J. Parasitol. Parasites Wildl. 4, 104-118.

Mans, B.J., Pienaar, R., Latif, A.A., Potgieter, F.T., 2011. Diversity in the 18S SSU rRNA V4 hyper-variable region of *Theileria* spp. in Cape buffalo (*Syncerus caffer*) and cattle from southern Africa. Parasitology 138, 766-779.

Mans, B.J., Pienaar, R., Potgieter, F.T., Latif, A.A., 2011. *Theileria parva*, T. sp. (buffalo) and T. sp. (bougasvlei) 18S variants. Vet. Parasitol. 182, 382-383.

Matjila, P.T., Leisewitz, A.L., Jongejan, F., Penzhorn, B.L., 2008. Molecular detection of tick-borne protozoal and ehrlichial infections in domestic dogs in South Africa. Vet. Parasitol. 155, 152-157.

Mbizeni, S., Potgieter, F.T., Troskie, C., Mans, B.J., Penzhorn, B.L., Latif, A.A., 2013. Field and laboratory studies on Corridor disease (*Theileria parva* infection) in cattle population at the livestock/game interface of uPhongolo-Mkuze area, South Africa. Ticks Tick Borne Dis. 4, 227-234.

M'ghirbi, Y., Hurtado, A., Barandika, J.F., Khlif, K., Ketata, Z., Bouattour, A., 2008. A molecular survey of *Theileria* and *Babesia* parasites in cattle, with a note on the distribution of ticks in Tunisia. Parasitol. Res. 103, 435-442.

Nijhof, A.M., Penzhorn, B.L., Lynen, G., Mollel, J.O., Morkel, P., Bekker, C.P., Jongejan, F., 2003. *Babesia bicornis* sp. nov. and *Theileria bicornis* sp. nov.: tick-borne parasites associated with mortality in the black rhinoceros (*Diceros bicornis*). J. Clin. Microbiol. 41, 2249-2254.

Nijhof, A.M., Pillay, V., Steyl, J., Prozesky, L., Stoltsz, W.H., Lawrence, J.A., Penzhorn, B.L., Jongejan, F., 2005. Molecular characterization of *Theileria* species associated with mortality in four species of African antelopes. J. Clin. Microbiol. 43, 5907-5911.

Njiiri, N.E., Bronsvoort, B.M., Collins, N.E., Steyn, H.C., Troskie, M., Vorster, I., Thumbi, S.M., Sibeko, K.P., Jennings, A., van Wyk, I.C., Mbole-Kariuki, M., Kiara, H., Poole, E.J., Hanotte, O., Coetzer, K., Oosthuizen, M.C., Woolhouse, M., Toye, P., 2015. The epidemiology of tick-borne haemoparasites as determined by the reverse line blot hybridization assay in an intensively studied cohort of calves in western Kenya. Vet. Parasitol. 210, 69-76.

Oosthuizen, M.C., Zweygarth, E., Collins, N.E., Troskie, M., Penzhorn, B.L., 2008. Identification of a novel *Babesia* sp. from a sable antelope (*Hippotragus niger* Harris, 1838). J. Clin. Microbiol. 46, 2247-2251.

Oura, C.A., Bishop, R.P., Wampande, E.M., Lubega, G.W., Tait, A., 2004. Application of a reverse line blot assay to the study of haemoparasites in cattle in Uganda. Int. J. Parasitol. 34, 603-613.

Oura, C.A., Tait, A., Asiimwe, B., Lubega, G.W., Weir, W., 2011. *Theileria parva* genetic diversity and haemoparasite prevalence in cattle and wildlife in and around Lake Mburo National Park in Uganda. Parasitol. Res. 108, 1365-1374.

Pain, A., Renauld, H., Berriman, M., Murphy, L., Yeats, C.A., Weir, W., Kerhornou, A., Aslett, M., Bishop, R., Bouchier, C., Cochet, M., Coulson, R.M., Cronin, A., de Villiers, E.P., Fraser, A., Fosker, N., Gardner, M., Goble, A., Griffiths-Jones, S., Harris, D.E., Katzer, F., Larke, N., Lord, A., Maser, P., McKellar, S., Mooney, P., Morton, F., Nene, V., O'Neil, S., Price, C., Quail, M.A., Rabbinowitsch, E., Rawlings, N.D., Rutter, S., Saunders, D., Seeger, K., Shah, T., Squares, R., Squares, S., Tivey, A., Walker, A.R., Woodward, J., Dobbelaere, D.A., Langsley, G., Rajandream, M.A., McKeever, D., Shiels, B., Tait, A., Barrell, B., Hall, N., 2005. Genome of the host-cell transforming parasite *Theileria annulata* compared with *T. parva*. Science 309, 131-133.

Perie, N.M., Uilenberg, G., Schreuder, B.E., 1979. *Theileria mutans* in Nigeria. Res. Vet. Sci. 26, 359-362.

Pfitzer, S., Oosthuizen, M.C., Bosman, A.M., Vorster, I., Penzhorn, B.L., 2011. Tick-borne blood parasites in nyala (*Tragelaphus angasii*, Gray 1849) from KwaZulu-Natal, South Africa. Vet. Parasitol. 176, 126-131.

Pienaar, R., Latif, A.A., Thekisoe, O.M., Mans, B.J., 2014. Geographic distribution of *Theileria* sp. (buffalo) and *Theileria* sp. (bougasvlei) in Cape buffalo (*Syncerus caffer*) in southern Africa: implications for speciation. Parasitology 141, 411-424.

Pienaar, R., Potgieter, F.T., Latif, A.A., Thekisoe, O.M., Mans, B.J., 2011. The Hybrid II assay: a sensitive and specific real-time hybridization assay for the diagnosis of *Theileria parva* infection in Cape buffalo (*Syncerus caffer*) and cattle. Parasitology 138, 1935-1944.

Pienaar, R., Potgieter, F.T., Latif, A.A., Thekisoe, O.M., Mans, B.J., 2011. Mixed *Theileria* infections in free-ranging buffalo herds: implications for diagnosing *Theileria parva* infections in Cape buffalo (*Syncerus caffer*). Parasitology 138, 884-895.

Purnell, R.E., Young, A.S., Payne, R.C., Mwangi, J.M., 1975. Development of *Theileria mutans* (Aitong) in the tick *Amblyomma variegatum* compared to that of *T. parva* (Muguga) in *Rhipicephalus appendiculatus*. J. Parasitol. 61, 725-729.

Reichard, M.V., Van Den Bussche, R.A., Meinkoth, J.H., Hoover, J.P., Kocan, A.A., 2005. A new species of *Cytauxzoon* from Pallas' cats caught in Mongolia and comments on the systematics and taxonomy of piroplasmids. J. Parasitol. 91, 420-426.

Ribeiro, J.M., Alarcon-Chaidez, F., Francischetti, I.M., Mans, B.J., Mather, T.N., Valenzuela, J.G., Wikel, S.K., 2006. An annotated catalog of salivary gland transcripts from *Ixodes scapularis* ticks. Insect Biochem. Mol. Biol. 36, 111-129.

Salih, D.A., El Hussein, A.R., Ahmed, J., Seitzer, U., 2010. Comparison between reverse line blot and enzyme-linked immunosorbent assay in diagnosis of major tick-borne diseases of cattle in Southern Sudan. Transbound Emerg. Dis. 57, 61-62.

Shastri, U.V., Jadhav, K.V., Pathak, S.V., Shastri, S.R., Deshpande, P.D., 1985. Studies on *Theileria* sp. from buffalo, *Bubalis bubalis*, from Maharashtra. Indian Journal of Parasitology 9, 275-279.

Sibeko, K.P., Oosthuizen, M.C., Collins, N.E., Geysen, D., Rambritch, N.E., Latif, A.A., Groeneveld, H.T., Potgieter, F.T., Coetzer, J.A., 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.

Spickett, A.M., 2013. Ixodid ticks of major economic importance and their distribution in South Africa.

Stewart, N.P., De Vos, A.J., Shiels, I., McGregor, W., 1987. The experimental transmission of *Theileria buffeli* of cattle in Australia by *Haemaphysalis humerosa*. Aust. Vet. J. 64, 81-83.

Stewart, N.P., de Vos, A.J., Shiels, I.A., Jorgensen, W.K., 1989. Transmission of *Theileria buffeli* to cattle by *Haemaphysalis bancrofti* fed on artificially infected mice. Vet. Parasitol. 34, 123-127.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. 28, 2731-2739.

Tomassone, L., Grego, E., Calla, G., Rodighiero, P., Pressi, G., Gebre, S., Zeleke, B., De Meneghi, D., 2012. Ticks and tick-borne pathogens in livestock from nomadic herds in the Somali Region, Ethiopia. Exp. Appl. Acarol. 56, 391-401.

Uilenberg, G., Robson, J., Pedersen, V., 1974. Some experiments on the transmission of *Theileria mutans* (Theiler, 1906) and *Theileria parva* (Theiler, 1904) by the ticks *Amblyomma variegatum* (Fabricius, 1794) and *Rhipicephalus appendiculatus* Neumann, 1901, in Uganda. Tropenmedizin Und Parasitologie 25, 207-216.

Uilenberg, G., Schreuder, B.E., 1976. Studies on Theileriidae (Sporozoa) in Tanzania. I. Tick transmission of *Haematoxenus veliferus*. Tropenmedizin Und Parasitologie 27, 106-111.

Uilenberg, G., Schreuder, B.E., Mpangala, C., 1976. Studies on Theileriidae (Sporozoa) in Tanzania. III. Experiments on the transmission of *Theileria mutans* by *Rhipicephalus appendiculatus* and *Amblyomma variegatum* (Acarina, Ixodidae). Tropenmedizin Und Parasitologie 27, 323-328.

Uilenberg, G., 2006. *Babesia*--a historical overview. Vet. Parasitol. 138, 3-10.

Vaumourin, E., Vourc'h, G., Telfer, S., Lambin, X., Salih, D., Seitzer, U., Morand, S., Charbonnel, N., Vayssier-Taussat, M., Gasqui, P., 2014. To be or not to be associated: power study of four statistical modeling approaches to identify parasite associations in cross-sectional studies. Front. Cell. Infect. Microbiol. 4, 62.

Walker, J.B., 1991. A review of the ixodid ticks (Acari, Ixodidae) occurring in southern Africa. Onderstepoort J. Vet. Res. 58, 81-105.

Walker, J.B., Olwage, A., 1987. The tick vectors of *Cowdria ruminantium* (Ixodoidea, Ixodidae, genus *Amblyomma*) and their distribution. Onderstepoort J. Vet. Res. 54, 353-379.

Yin, H., Luo, J., Schnittger, L., Lu, B., Beyer, D., Ma, M., Guan, G., Bai, Q., Lu, C., Ahmed, J., 2004. Phylogenetic analysis of *Theileria* species transmitted by *Haemaphysalis qinghaiensis*. Parasitol. Res. 92, 36-42.

Young, A.S., Burridge, M.J., Payne, R.C., 1977. Transmission of a *Theileria* species to cattle by the ixodid tick, *Amblyomma cohaerens* Donitz 1909. Trop. Anim. Health Prod. 9, 37-45.

Young, A.S., Purnell, R.E., Payne, R.C., Brown, C.G., Kanhai, G.K., 1978. Studies on the transmission and course of infection of a Kenyan strain of *Theileria mutans*. Parasitology 76, 99-115.

Yusufmia, S.B., Collins, N.E., Nkuna, R., Troskie, M., Van den Bossche, P., Penzhorn, B.L., 2010. Occurrence of *Theileria parva* and other haemoprotozoa in cattle at the edge of Hluhluwe-iMfolozi Park, KwaZulu-Natal, South Africa. J. S. Afr. Vet. Assoc. 81, 45-49.

Figure 1

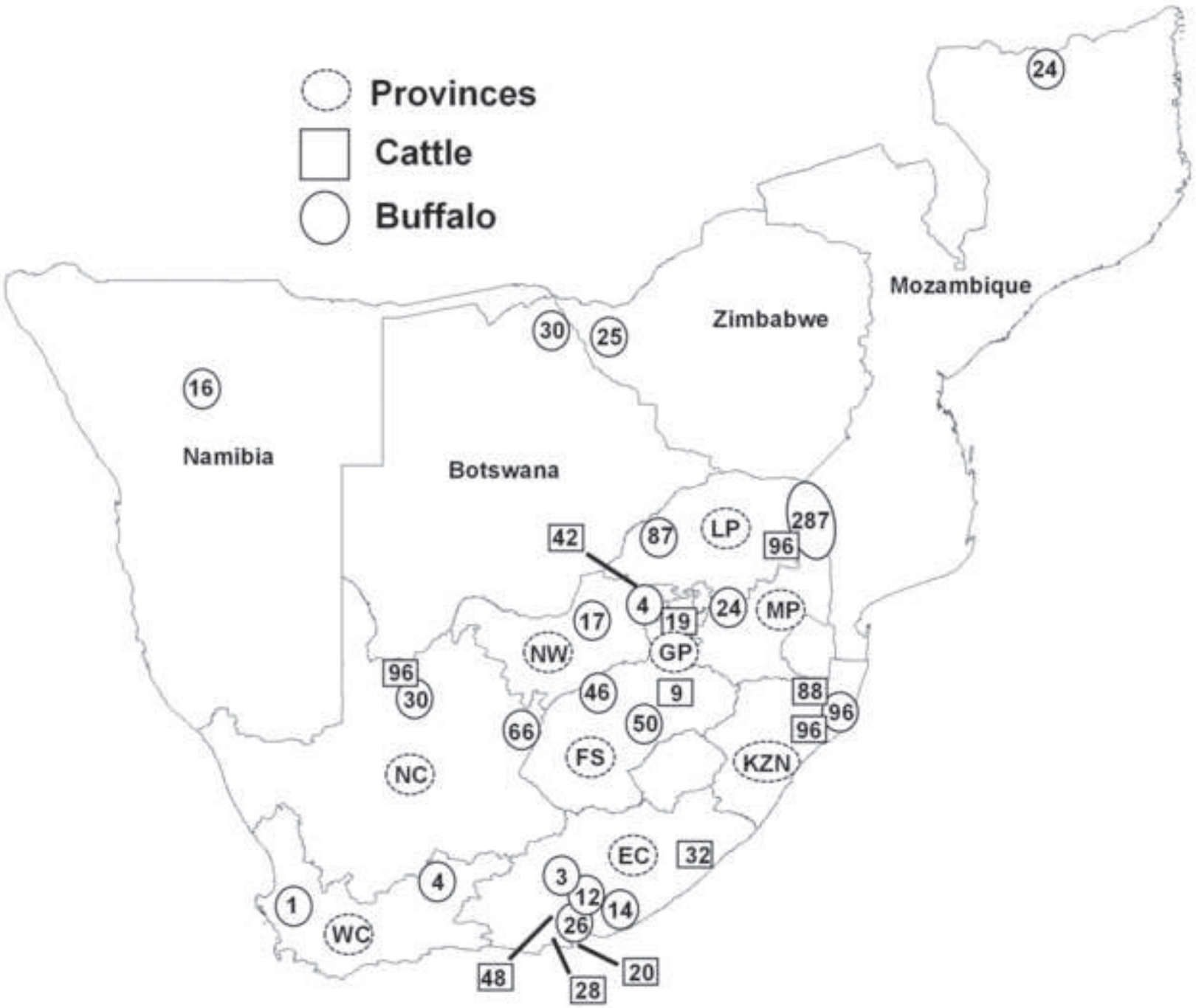

Provinces
Cattle
Buffalo
Mozambique
Zimbabwe
Namibia
Botswana
24
30
25
16
42
87
LP
287
96
4
24
MP
19
17
NW
GP
96
30
46
9
88
96
50
96
66
FS
KZN
NC
EC
32
4
3
12
14
1
26
WC
48
28
20

Figure 2

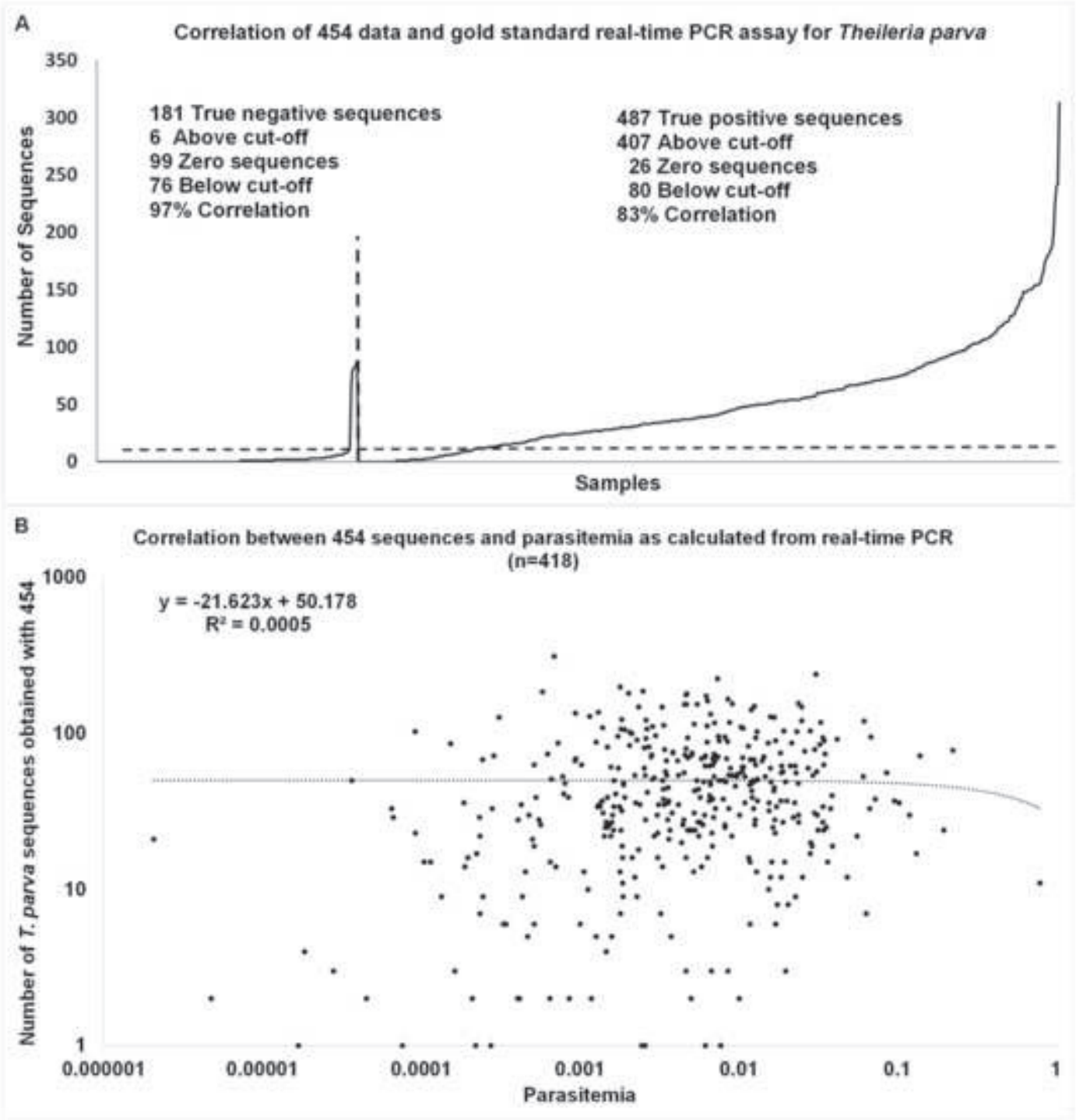

A
Correlation of 454 data and gold standard real-time PCR assay for *Theileria parva*
181 True negative sequences
6 Above cut-off
99 Zero sequences
76 Below cut-off
97% Correlation
487 True positive sequences
407 Above cut-off
26 Zero sequences
80 Below cut-off
83% Correlation
Number of Sequences
350
300
250
200
150
100
50
0
Samples
B
Correlation between 454 sequences and parasitemia as calculated from real-time PCR (n=418)
y = -21.623x + 50.178
R² = 0.0005
Number of *T. parva* sequences obtained with 454
1000
100
10
1
0.000001
0.00001
0.0001
0.001
0.01
0.1
1
Parasitemia

Figure 3

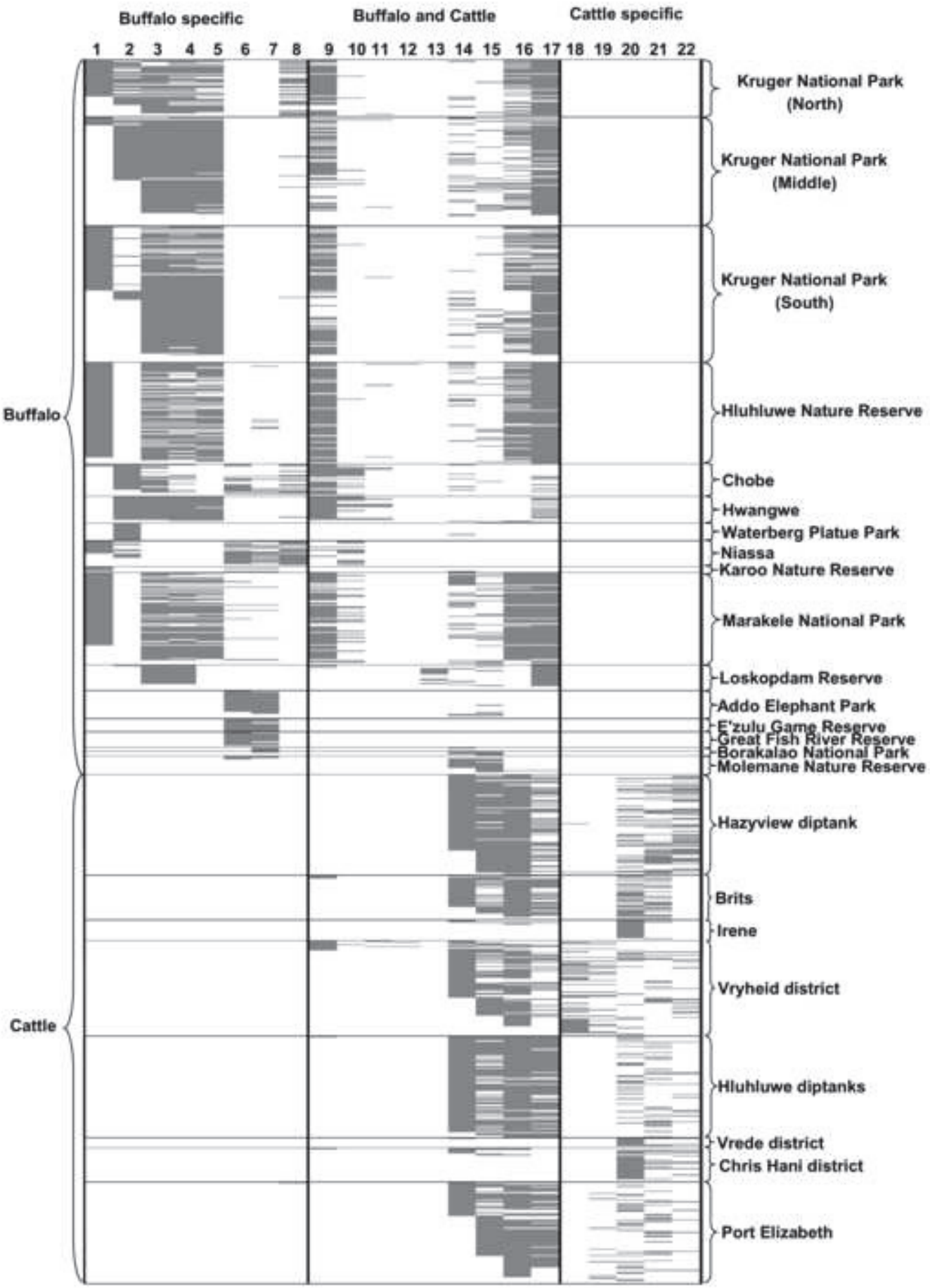

Buffalo specific
Buffalo and Cattle
Cattle specific
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22
Buffalo
Cattle
Kruger National Park (North)
Kruger National Park (Middle)
Kruger National Park (South)
Hluhluwe Nature Reserve
Chobe
Hwangwe
Waterberg Platue Park
Niassa
Karoo Nature Reserve
Marakele National Park
Loskopdam Reserve
Addo Elephant Park
E'zulu Game Reserve
Great Fish River Reserve
Borakalao National Park
Molemane Nature Reserve
Hazyview diptank
Brits
Irene
Vryheid district
Hluhluwe diptanks
Vrede district
Chris Hani district
Port Elizabeth

Figure 4

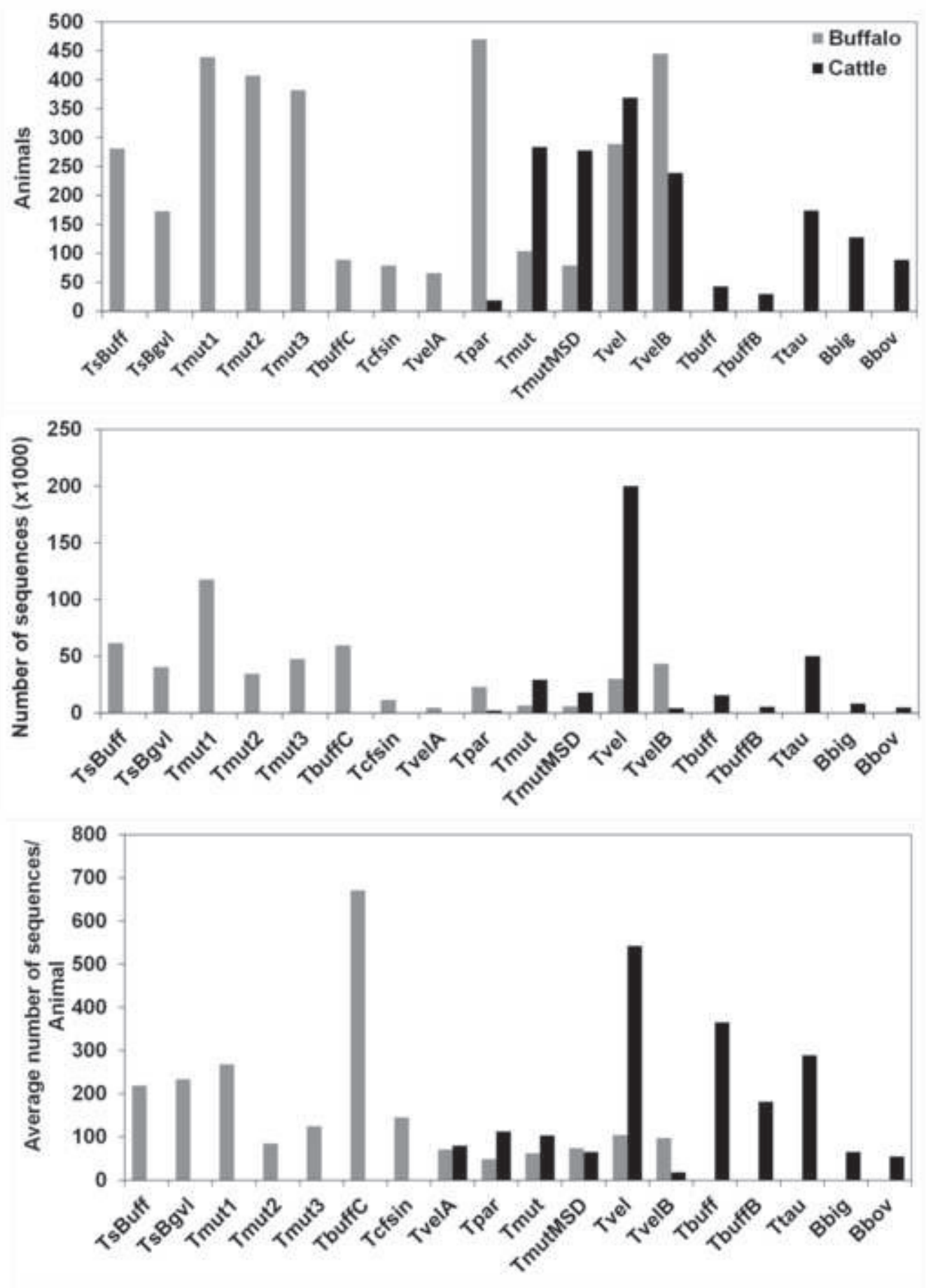
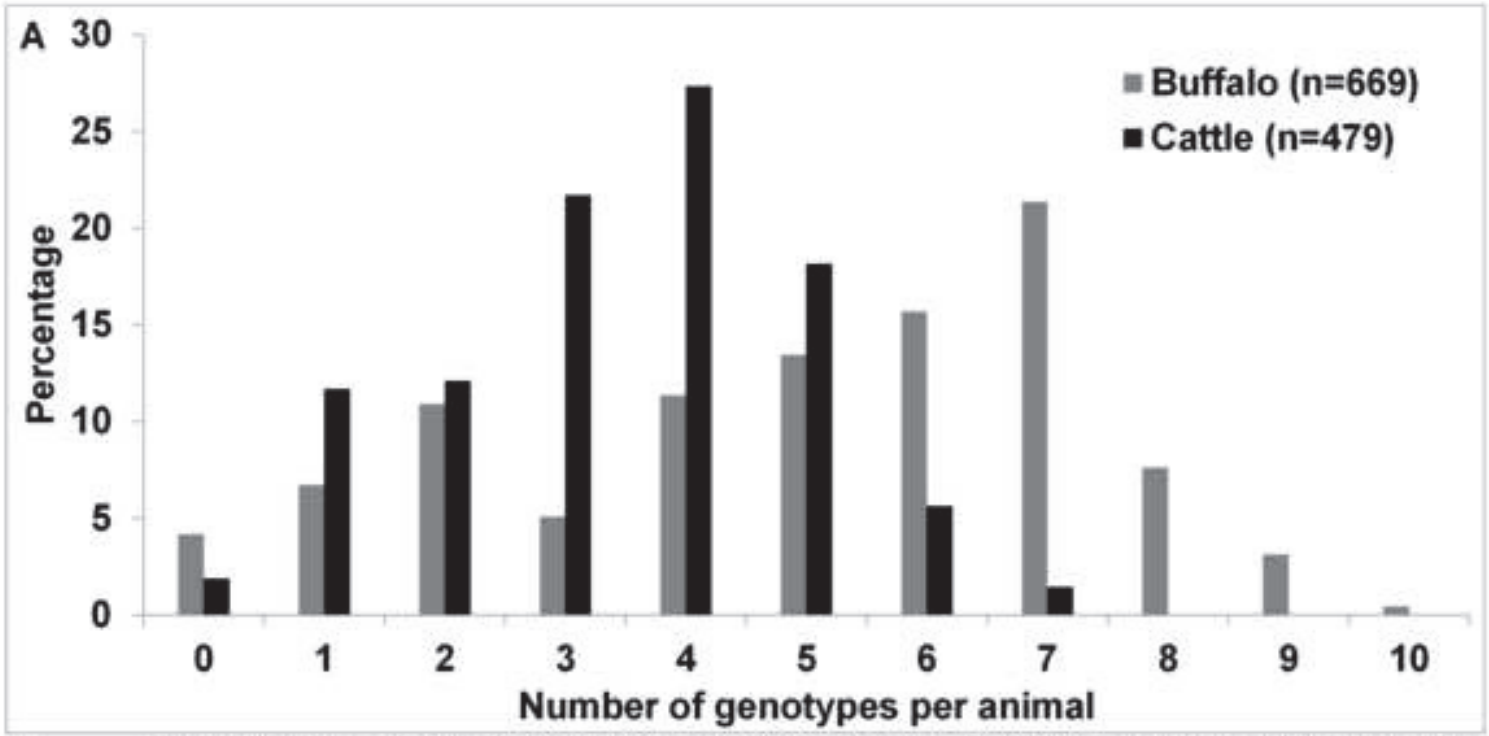

Buffalo
Cattle
Animals
500
450
400
350
300
250
200
150
100
50
0
TsBuff
TsBgvl
Tmut1
Tmut2
Tmut3
TbuffC
Tcfsin
TvelA
Tpar
Tmut
TmutMSD
Tvel
TvelB
Tbuff
TbuffB
Ttau
Bbig
Bbov
Number of sequences (x1000)
250
200
150
100
50
0
Average number of sequences/ Animal
800
700
600
500
400
300
200
100
0

Figure 5



B	TsBuff	TsBgvl	Tm1	Tm2	Tm3	TbC	Tcfsin	TvA	Tpar	Tm	TmMSD	Tv	TvB	TbW	TbB	Ttau	Bbig
TsBgvl	9.4																
Tm1	40.4	27.8															
Tm2	38.2	26.4	87.6														
Tm3	40.8	24.4	80.8	82.2													
TbC	4.8	6.9	3.1	1.4	0.6												
Tcfsin	4.7	2.9	2.6	1.7	1.5	64.7											
TvA	9.4	22.4	7.9	6.8	5.4	20.0	11.5										
Tpar	49.1	27.4	66.9	60.6	61.7	4.8	3.5	12.0									
Tm	7.2	3.7	8.2	8.6	8.1	1.7	1.1	1.3	9.1								
TmMSD	4.4	1.3	5.0	4.8	4.2	0.5	0.7	0.5	6.3	39.8							
Tv	26.9	5.9	30.3	28.0	31.3	0.7	1.2	3.0	31.6	39.5	41.2						
TvB	31.9	14.2	53.7	51.0	49.2	1.0	1.5	5.2	46.5	29.1	28.9	63.5					
TbW	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.7	5.6	3.6	3.7	0.3				
TbB	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.4	3.7	3.2	2.7	0.7	40.4			
Ttau	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.3	19.3	18.0	16.4	10.1	1.9	1.5		
Bbig	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2	17.0	23.4	15.2	10.0	1.2	0.6	17.5	
Bbov	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	14.1	19.3	10.8	5.0	4.8	2.6	14.3	30.7

Table 1: Statistics for sequences generated using GS-Junior 454 sequencing.

	Total reads	**Average length total reads**	**Total reads >280 bp**	**Average length >280 bp**	**Total after chimeric removal**	**Total after cut-off**	**Percentage total reads**	**Percentage informative reads**
Batch 1	92991	363 ± 106	77049	403 ± 55	76997	73528	79.1	95.4
Batch 2	95034	392 ± 126	80294	448 ± 37	80177	74912	78.8	93.3
Batch 3	75346	357 ± 120	59680	409 ± 59	59639	54072	71.8	90.6
Batch 4	86426	341 ± 121	63011	402 ± 49	62969	58880	68.1	93.4
Batch 5	77302	401 ± 103	75970	433 ± 51	75915	62604	80.9	82.4
Batch 6	114138	382 ± 124	92952	434 ± 54	92909	87424	76.6	94.0
Batch 7	89164	400 ± 97	79546	430 ± 47	79489	75225	84.4	94.6
Batch 8	101262	409 ± 103	90405	442 ± 47	90394	87273	86.2	96.5
Batch 9	70733	397 ± 112	60376	437 ± 60	60371	58297	82.4	96.6
Batch 10	68980	357 ± 108	57190	396 ± 58	57185	55204	80.0	96.5
Batch 11	97754	334 ± 132	68290	399 ± 66	68286	66694	68.2	97.7
Batch 12	83178	440 ± 92	73257	462 ± 51	73254	71598	86.1	97.7
Total	1052308	N/A	878020	N/A	877585	825711	N/A	N/A
Average ± SD	87692 ± 13446	381 ± 31	73168 ± 11796	424 ± 22	73132 ± 11785	68809 ± 11482	79 ± 6	94 ± 4