Coultous, R., McDonald, M., Raftery, A. G., Shiels, B. , Sutton, D. G.M. and Weir, W. (2020) Analysis of Theileria equi diversity in The Gambia using a novel genotyping method. *Transboundary and Emerging Diseases*, 67(3), pp. 1213-1221. (doi: 10.1111/tbed.13454).

There may be differences between this version and the published version. You are advised to consult the publisher's version if you wish to cite from it.

This is the peer reviewed version of the following article:
Coultous, R., McDonald, M., Raftery, A. G., Shiels, B. , Sutton, D. G.M. and Weir, W. (2020) Analysis of Theileria equi diversity in The Gambia using a novel genotyping method. *Transboundary and Emerging Diseases*, 67(3), pp. 1213-1221, which has been published in final form at 10.1111/tbed.13454. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Self-Archiving.

http://eprints.gla.ac.uk/205013/

Deposited on: 05 December 2019

Enlighten – Research publications by members of the University of Glasgow
http://eprints.gla.ac.uk

Title

Analysis of *Theileria equi* diversity in The Gambia using a novel genotyping method

Authors

Robert M Coultous[a], Michael McDonald[b], Alexandra G Raftery[b], Brian R Shiels[a], David G M Sutton[b] and William Weir[b]

Affiliations

[a] Institute of Biodiversity Animal Health and Comparative Medicine
College of Medical, Veterinary and Life Sciences
University of Glasgow
Bearsden Road
Glasgow
G61 1QH
United Kingdom

[b] School of Veterinary Medicine
College of Medical, Veterinary and Life Sciences
University of Glasgow
Bearsden Road
Glasgow
G61 1QH
United Kingdom

Corresponding author:
Robert Coultous
robert.coultous@glasgow.ac.uk
Telephone +44 (0)141 330 6922

Fax +44 (0)141 330 2271
Institute of Biodiversity Animal Health and Comparative Medicine
College of Medical, Veterinary and Life Sciences
Room 217, Henry Wellcome Building
University of Glasgow
Bearsden Road
Glasgow
G61 1QH
United Kingdom

Summary

Theileria equi, one of the primary pathogens causing equine piroplasmosis, has previously been sub-classified into a number of clades on the basis of 18S SSU rRNA gene sequence diversity. This partitioning of the parasite population has potential implications for host immunity, treatment and vaccine development. To detect and identify different clade genotypes among and within individual equine blood samples, a novel PCR based technique was designed and optimised. *Theileria equi* has only recently been described in The Gambia, and the developed genotyping technique was used to analyse blood samples taken from 42 piroplasmosis-positive horses and donkeys within the country. Three different *T. equi* genotypes were detected within the population, including the same genotype as the recently described *Theileria haneyi*, with 61.9 % of individuals found to be infected with more than one genotype. Overall, there was a trend that males were more likely to have a multiple genotype infection.

Thus, the novel genotyping technique has been shown to be effective in analysis of field populations, and offers researchers a rapid method of identifying multiple *T. equi* genotypes both within individuals and equine populations in epidemiological studies.

Keywords

Theileria equi, equine, piroplasmosis, Gambia, genotyping

Introduction

Theileria equi is one of the primary pathogens causing equine piroplasmosis, a disease of significant global welfare and economic concern, with a ubiquitous prevalence throughout much of the developing and developed world (Wise et al., 2013). The exact phylogenetic positioning of *T. equi* has been contested since its first description over a century ago (Laveran, 1901), and modern molecular techniques have led to several authors documenting a remarkably high degree of intraspecies genetic polymorphism at the typically highly conserved 18S SSU rRNA locus (Bhoora et al., 2009; Knowles et al., 2018; Nagore et al., 2004). Bhoora *et al.* partitioned *T. equi* into three groups, clades A-C, based on their findings in South African samples (Bhoora et al., 2009), and this was expanded by Salim *et al.* following their Sudanese survey, with an additional clade D (Salim et al., 2010). Qablan *et al.* later proposed a fifth clade E, after an investigation of Jordanian samples (Qablan et al., 2012; Qablan et al., 2013), however contemporary authors have maintained the four clade (A-D) framework (Alanazi et al., 2014; Hall et al., 2013), considering clade E to be a part of clade B (Alanazi et al., 2014).

Research interest in *T. equi* genetic diversity, and its implications, has recently been rekindled following the description of *Theileria haneyi* (Knowles et al., 2018), which has been proposed as a new species of *Theileria*.

Infection of equines with different *T. equi* clade genotypes has been noted, with studies of naturally infected populations showing differences in the severity of clinical signs between individuals depending on the infecting genotype (Manna et al. 2018). Additionally, recent experimental work has shown that super-infection with both *T. equi* and the newly described *T. haneyi* is possible, and can be maintained for many months (Sears et al. 2019). In comparison to *T. haneyi* infection, *T. equi* infection results in

82 greater haematological effect, with a more pronounced neutropenia which becomes
83 more marked following experimental super-infection with *T. haneyi* (Sears et al., 2019).
84 However, there has been no record to date of multi-clade genotype infection within a
85 single equine host in a naturally infected population (Knowles et al., 2018).
86 Consequently, the effect of mixed strain infection on the pathogenesis, host immunity
87 and treatment of infection has yet to be fully ascertained (Knowles et al., 2018; Sears et
88 al., 2019).
89 The reduced sensitivity of commercially available serological diagnostics has also been
90 attributed to genotypic diversity within the species (Bhoora et al., 2010), greatly
91 increasing the biosecurity challenge of detecting the carrier horses at the time of
92 importation.
93 Current methods for *T. equi* clade-level genotyping are laborious; PCR products must be
94 generated, purified and submitted for Sanger sequencing and the presence of multiple
95 allelic types in samples necessitates an additional molecular cloning step. Bioinformatic
96 comparison of the resulting sequences to existing DNA sequence databases must then be
97 undertaken manually to determine the isolate's clade type.
98 This paper describes the development and optimisation of a novel PCR based technique
99 for the rapid and cost-effective determination of the clade-level genotype of *T. equi*
100 isolates. The characteristics of such a test should include the capacity to (1) explicitly
101 detect *T. equi*; (2) qualitatively categorise samples to the described A-D clades and (3)
102 identify the presence of multiple subtypes in a single equine blood sample.
103 Following optimisation and parameter validation, the test was then used to analyse a set
104 of Gambian equine blood samples. The Gambia is the smallest country in mainland
105 Africa, and our recent research has identified piroplasmosis in over 50 % of equine
106 blood samples from the region (Raftery et al., 2019). The primary aims of this part of the
107 study were (1) identification of *T. equi* subtypes within the equine population and any

individuals infected with multiple parasite genotypes, and (2) identification of risk factors associated with multiple *T. equi* genotype infections within individual animals.

Materials and methods

Clade genotyping test

The developed genotyping method consists of two general steps: a PCR based amplification stage, and a probe-based clade identification stage, using real-time PCR technology.

i) PCR amplification

Initial amplification utilised a modified *Babesia*/*Theileria* catch-all primer set (BT1-F and BTH-1R) (Criado-Fornelio et al., 2003) (Table 1), which targets conserved areas flanking the hyper-variable V4 region of the 18S SSU rRNA gene, on which *T. equi* clade classification is based. Targeting this particular area, therefore, afforded the ability to design clade-specific probes. These primers have been reported to effectively detect a range of *Theileria/Babesia* species, and have been previously used for the detection of *T. equi* and *B. caballi* (Coultous et al., 2019; Raftery et al., 2019). Reaction conditions for this first round PCR were performed in a final volume of 20 µl with GeneAmp 10X PCR Buffer (Life Technologies Applied Biosystems) (final concentrations of 10 mM Tris-HCl, 50 mM KCl, and 1.5 mM $MgCl_2$), Deoxynucleotide (dNTP) Solution Mix (New England Biolabs) (final concentrations of 0.2 mM dATP, 0.2 mM dCTP, 0.2 mM dGTP and 0.2 mM dTTP), 0.5 µM of each primer, 0.025 units/µl DNA polymerase (AmpliTaq DNA Polymerase (Life Technologies Applied Biosystems)) and 2 µl of template genomic DNA solution. The cycling conditions for the primary reaction were an initial denaturation of 94 °C for 5 min, then 30 cycles of denaturation at 94 °C for 45 s, annealing at 67 °C for 60 s, and extension at 72 °C for 60 s, with a final extension at 72 °C for 5 min.

ii) Novel PCR primer design

A nested PCR (nPCR) approach has been shown to improve *T. equi* detection sensitivity (Nicolaiewsky et al., 2001) and so this method was selected for the clade identification step. To facilitate primer design, the NCBI database (GenBank) was searched for sequences containing the terms '18S' and either '*Babesia equi*' or '*Theileria equi*'. This resulted in the identification of a total of 533 sequences, which were downloaded for analysis. Following alignment and removal of incomplete sequences, the sequences were trimmed to the inner margins of the outer PCR primer sites (BT1-F and BTH-1R). This resulted in 162 remaining sequences. The hyper-variable region was identified following alignment of these sequences, and a pair of new primers (PCR-FWD and PCR-REV) (Table 1) was designed to flank this region, targeting sites that were conserved among all 162 sequences. Sequences were aligned using MUSCLE (Edgar, 2004) within the AliView alignment viewer and editor (Larsson, 2014). Phylogenetic trees were computed using the MEGA7 package (Kumar et al., 2016).

iii) Novel clade-specific probe design

The original 533 GenBank sequences were aligned and this time trimmed to the new PCR primers (PCR-FWD and PCR-REV), leaving 258 sequences. These sequences were used to identify a number of potential probe target sequences, designed to be specific for either clade A, B, C, or D.

Using an Applied Biosystems AB7500 real-time PCR cycler, a number of probe designs were trialled, and the reaction conditions optimised. Cross-reactivity between the probes was assessed using four DNA reference isolates, each representative of one of the described *T. equi* clades. The reference isolates were sourced from a variety of geographical origins: clade A (UK), clade B (Ireland), clade C (Morocco) and clade D (The Gambia). These reference samples were selected based on two main criteria: the repeatable production of a strong electrophoresis gel band of the correct size following

amplification by a previously described *Babesia/Theileria* catch-all nPCR (Coultous et al., 2019), and confirmation of single clade infection following cloning and sequencing of PCR amplicons. The PCR products for each reference isolate were subsequently cloned into the sequencing vector: pDrive Cloning Vector (Qiagen). For each isolate, four colonies were selected, and plasmid DNA was isolated from each cultured clone using a proprietary kit (QIAprep Spin Miniprep Kit, Qiagen). The recovered DNA was prepared for Sanger sequencing (Eurofins, Germany), and the T7 primer site in the vector flanking sequence used to generate sequence reads. Sanger sequencing confirmed the presence of a single clade in each isolate.

The final selected probes are described in Table 2. The finalised PCR reaction was performed in a total volume of 20 µl with Takyon MasterMix (Eurogentec), 300 nM forward primer (PCR-FWD), 900 nM reverse primer (PCR-REV), 50 nM clade-specific probe and 2.5 µl DNA template solution. The DNA template was a 1:10 dilution of the product from the *Babesia*/*Theileria* catch-all reaction. The cycling conditions were an initial master mix activation at 50°C for 2 min, then 95°C for 3 min, followed by 40 cycles of denaturation at 95°C for 10 s, and annealing/extension at 64°C for 1 min. Each probe was tested individually.

Analysis of field population samples

To assess its effectiveness on a field population, the optimised genotyping method was used to analyse DNA isolated from individual blood samples taken from a previously described Gambian equine population (Raftery et al., 2019). A sample set of 42 animals was identified to be piroplasmosis positive by a previously described *Babesia/Theileria* catch-all nPCR (Coultous et al., 2019; Raftery et al., 2019), and identified as free from trypanosomiasis by a previously described PCR (Raftery et al., 2019). Each sample had associated clinical data, primarily species, sex, age and the presence or absence of pyrexia. Genomic DNA was extracted from the EDTA blood samples using a QIAamp

DNA Blood Maxi Kit (Qiagen) with a standardised technique previously described (Raftery et al., 2019). The extracted DNA samples were then analysed using the described *T. equi* primers (Table 1) and probes (Table 2) with the stated reaction conditions.

The statistical data analysis for odds ratios and Fisher's exact test were performed using the epiDisplay package (https://CRAN.R-project.org/package=epiDisplay), and the binomial test using the 'binom.test' function, in RStudio (RStudio Team, 2016).

Results

Clade genotyping test

In developing the PCR based test for identification and discrimination of *T. equi* clades, available 18S SSU rRNA *T. equi* GenBank sequences were downloaded and aligned. This alignment was then used to design clade-specific probes that were well conserved between sequences, and to confirm the position of the four available clade reference samples. The position of the reference samples within their respective clades is shown in Figure 1. An alignment of the designed probes with the reference sample sequences is shown in Figure 2.

Using the available 258 GenBank sequences, alignments to assess intra-clade polymorphism illustrated that the probes perfectly match their respective clade reference samples, and were well conserved within, and specific to, their respective clades (Table 3).

Each probe was tested against the reference DNA sample for each clade to check for off-target amplification. The probes were used at the optimised concentration of 50 nM, using the same standardised reaction conditions, and a cut-off CT value of 0.2. The amplification curve for each probe and sample pairing is shown in Figure 3. Based on assay testing with the four reference isolates, all probes showed good specificity for their respective target clade, with good amplification curves and minimal background

noise. Additionally, the finalised probes showed no cross-reactivity outside their specified clades. Finally, the probes were tested against extracted DNA samples of other equine haemoparasites including *Babesia caballi*, *Anaplasma phagocytophylum* and *Trypanosoma brucei brucei*. No amplification was noted in species other than *T. equi*. Analysis of the finalised probes indicated that the test was sufficiently specific and robust to be deployed to determine clade types within field samples of *T. equi* from a naturally infected population of equines.

Analysis of field population samples

Forty-two blood samples were available from Gambian equines identified to be trypanosome PCR negative and EP PCR positive. The basic demography of this sample group was: 66.7 % horses (n = 28), 33.3 % donkeys (n = 14); 38.1 % males (n = 16), 61.9 % females (n = 26); 21.4 % aged 0-2 years (n = 9), 38.1 % aged >2 - <10 years (n = 16), 40.5 % aged >10 years (n = 17).

The newly developed genotyping assay was used to assess *T. equi* clade diversity in this sample collection and the results are shown in Figure 4. The genotyping workflow was successful in identifying up to three *T. equi* genotypes in a single sample, and a total of three genotypes were detected overall within the sample population (clades A, C and D). Overall, 61.9 % (n = 26) of individuals presented with infection of more than one clade. The most frequent presentation was a combined clade AC infection (52.4 %, n = 22), followed by clade C single infection (28.6 %, n = 12) and then clade A single infection (9.5 %, n = 4). Clade D was only detected in four samples, and never as a single clade infection.

Given the dominance of combined A and C infection in this population, an assessment was made of whether the observed probability of this co-infection in the initial week 1 samples (0.57 (n = 24/42)) differed from that expected, assuming infection events occur as independent events. To do so, a binomial test was performed, assuming a null

hypothesis of independent infections based on the product of the observed parasite clade prevalences in the study population; 67 % (n = 28) and 90 % (n = 38) for A and C respectively, giving a 60 % predicted prevalence of A and C co-infection. The results were in agreement with the null hypothesis, showing no indication that genotypes A and C were mixing non-randomly in this equine population (probability of A/C co-infection = 0.57, 95 % CI = 0.40 – 0.72, P = 0.754).

An analysis was also performed to identify whether associations existed between host parameters and the presence of multiple *T. equi* genotypes within an individual. Table 4 shows the odds ratios and Fisher's exact test results for the exposure variables of species, sex, age and the clinical sign of pyrexia. Although not statistically significant at P < 0.05, there was an increased odds ratio for a multi-clade infection in males of 4.33 (CI 0.99-18.89, p = 0.056), when compared to females.

Discussion

The preliminary trial of the optimised qualitative *T. equi* genotyping workflow was successful. It showed the technique to be robust in the detection of multi-genotype infections in individual animals in a naturally infected population. Therefore, the developed test was successful in achieving its stated goals, demonstrating promise for further refinement and exploitation as a rapid and cost-effective method for *T. equi* genotyping on a large scale. Future optimisation would involve the standardisation of CT values between the different clade-specific probes, to allow quantitative estimation of the proportions of sub-types in mixed infections as part of a qPCR methodology. There may also be scope for the inclusion of degenerate probes in order to enhance intra-clade sensitivity. The current probes possess different detector dyes and have all been optimised to function under the same reaction conditions, so following CT value standardisation, further optimisation could be undertaken to produce a multiplex 'one-pot' reaction, further reducing the technique's cost and labour time. A further addition

to the technique's detection range would be the inclusion of a *B. caballi* specific probe, the second apicomplexan species responsible for equine piroplasmosis. However, the number of fluorescence detection channels present on some currently available real-time PCR cyclers limits the number of individual probes, so this may not be feasible for many laboratories. Nonetheless, this novel technique has potential as a useful tool to assist future research efforts focused on *T. equi* genetic diversity, population biology and epidemiology.

Given the high *T. equi* prevalence within the Gambian population sample set, it was hypothesised that the parasite has a well-established endemic status in the country; this contrasts with the situation for *B. caballi,* which has yet to be identified in The Gambia (Raftery et al., 2019). It has been demonstrated in other *Theileria* species such as *Theileria parva* and *Theileria annulata*, that parasite populations historically endemic in a geographical area have a high level of genotypic diversity (Hemmink et al., 2018; Weir et al., 2011).

The new *T. equi* subtyping technique detected the presence of parasites from clades A, C and D. Mixed genotype infections were seen in 61.9 % (n = 26) of individuals, and in one sample representative genotypes from all three clades were detected. The high frequency of mixed infections indicates that exposure to different *T. equi* genotypes, particularly genotypes representing clades A and C, is common, indicating a mixed parasite population within the geographic area. The presence of multiple *T. equi* genotypes, along with their high prevalence, supports the concept of a long established *T. equi* parasite population in this area of The Gambia. However, it should be appreciated that the genotyping system employed in this study has relatively low resolution. Thus, an alternative, but less likely explanation is that the parasite population is derived from a recently introduced founder population composed of several clade types from a larger genetically diverse EP population. The use of high resolution markers, such as the comprehensive mini- and micro-satellite genotyping

297 systems employed in the population genetics of other *Theileria* species (Weir et al.,
298 2011), would be required to fully address this hypothesis.
299 It was extremely interesting to note the presence of both clade A and clade C genotypes
300 in a large proportion of individuals. The newly proposed *T. haneyi* (Knowles et al., 2018)
301 appears to be positioned in clade C, and the previously described *T. equi* reference
302 genome belongs to clade A (Hall et al., 2013). Knowles *et al.* note that they were yet to
303 find natural co-infection of animals with these two types (Knowles et al., 2018),
304 although recent experimental work by the group has shown super-infection is possible
305 (Sears et al., 2019). Since this screening of Gambian samples has only focused on
306 genotyping based on a single genetic locus, it is impossible to confirm if the clade C
307 samples detected here correspond to *T. haneyi per se*. Notably, the frequent finding of
308 mixed infection illustrates the possibility of co-ingestion by feeding ticks of multiple *T.*
309 *equi* genotypes, a necessary underlying factor required to facilitate mating of parasites
310 from different clades. However, the extent to which inter-clade genetic exchange occurs
311 in the field is unknown and refinement of polymorphic genetic markers would be
312 required to investigate this issue.
313 Although not statistically significant at $P < 0.05$, the field sample analysis showed there
314 is an indication for males to be more likely to harbor a multi-clade infection (see Table
315 4). Previous studies have noted a statistically significant higher level of EP seropositivity
316 in males (Munkhjargal et al., 2013; Sevinc et al., 2008), which might be extrapolated to
317 multiplicity of sub-type infection if males are more susceptible to EP infection, generally.
318 Authors have attributed this sex bias to male equines preferentially being used for
319 riding and physically intense activities, with this physical stress resulting in a degree of
320 immunosuppression and an increased susceptibility to EP infection (Munkhjargal et al.,
321 2013). Indeed, in other host-parasite models the role of testosterone levels in mature
322 males causing increased parasitism through relative immunosuppression has been
323 extensively discussed (Klein, 2004). However, this reported sex bias is not consistent

amongst EP risk factor studies (Aharonson-Raz et al., 2014; Guidi et al., 2015; Piantedosi et al., 2014).

Age has been identified as a risk factor for EP seropositivity in several studies (Garcia-Bocanegra et al., 2013; Guidi et al., 2015; Munkhjargal et al., 2013; Rapoport et al., 2014), and also noted in other *Theileria* species (Weir et al., 2011). The hypothesis is that older animals have had a longer period of exposure to infected ticks. It was unusual that age was not identified as a significant risk factor for multi-genotype infections, as it could be assumed that older animals would have had a longer exposure to multiple strains of the parasite. Although possibly due to a small sample size, and hence low power, the lack of statistical significance in this study may indicate a degree of acquired immunity or an ability to regulate the parasitaemia of certain strains. Further study with a larger sample size would be required to investigate this hypothesis.

Presence of pyrexia was considered as an indicator of clinical disease, however no statistically significant relationship was noted between animals with single or multi-clade infection. Recent experimental infections of *T. equi* and *T. haneyi* (clades A and C) indicated that super-infection produced relatively minor changes in clinical condition, which was limited to more marked neutropenia (Sears et al. 2019), however this parameter was not assessed in the current study. Further studies using a larger sample size and recording more detailed clinical parameters would be required to detect clade-associated or multiplicity of infection-associated subtle effects on disease presentation.

In conclusion, the novel genotyping workflow developed was successfully used to identify single and multiple *T. equi* subtype infections in a population of naturally infected individuals, underlining its potential as a tool for future EP epidemiology and population biology studies.

Acknowledgements

The Horserace Betting Levy Board (HBLB) funded the cost of the PCR screening, and RMC was supported by an HBLB research scholarship (VET/RS/254). The Donkey Sanctuary funded the sample collection, and the Gambia Horse and Donkey Trust helped to facilitate sample collection.

Ethics statement

The authors confirm that the ethical policies of the journal, as noted on the journal's author guidelines page, have been adhered to and the appropriate ethical review committee approval has been received. Ethical approval for the sample collection was granted by the University of Glasgow School of Veterinary Medicine ethics committee (reference 21a/13), The Donkey Sanctuary ethics board and the Gambian Ministry of Agriculture.

Conflict of interest

None declared

References

Aharonson-Raz, K., Rapoport, A., Hawari, I. M., Lensky, I. M., Berlin, D., Zivotofsky, D., Klement, E., & Steinmen, A. (2014). Novel description of force of infection and risk factors associated with *Theileria equi* in horses in Israel and in The Palestinian Authority. *Ticks and Tick-Borne Diseases*, *5*, 366–372. http://doi.org/10.1016/j.ttbdis.2014.01.002

Alanazi, A. D., Said, A. E., Morin-Adeline, V., Alyousif, M. S., & Slapeta, J. (2014). Quantitative PCR detection of *Theileria equi* using laboratory workflows to detect asymptomatic persistently infected horses. *Veterinary Parasitology*, *206*, 138–145. http://doi.org/10.1016/j.vetpar.2014.09.019

Bhoora, R., Franssen, L., Oosthuizen, M. C., Guthrie, A. J., Zweygarth, E., Penzhorn, B. L.,

377 Jongejan, F., & Collins, N. E. (2009). Sequence heterogeneity in the 18S rRNA gene
378 within *Theileria equi* and *Babesia caballi* from horses in South Africa. *Veterinary*
379 *Parasitology*, *159*, 112–120. http://doi.org/10.1016/j.vetpar.2008.10.004
380 Bhoora, R., Quan, M., Matjila, P. T., Zweygarth, E., Guthrie, A. J., & Collins, N. E. (2010).
381 Sequence heterogeneity in the equi merozoite antigen gene (ema-1) of *Theileria*
382 *equi* and development of an ema-1-specific TaqMan MGB assay for the detection of
383 *T. equi*. *Veterinary Parasitology*, *172*, 33–45.
384 http://doi.org/10.1016/j.vetpar.2010.04.025
385 Coultous, R. M., Phipps, P., Dalley, C., Lewis, J., Hammond, T. A., Shiels, B. R., Weir, W., &
386 Sutton, D. G. M. (2019). Equine piroplasmosis status in the UK: an assessment of
387 laboratory diagnostic submissions and techniques. *The Veterinary Record*, *184*, 95–
388 95. http://doi.org/10.1136/vr.104855
389 Criado-Fornelio, A., Martinez-Marcos, A., Buling-Sarana, A., & Barba-Carretero, J. C.
390 (2003). Molecular studies on *Babesia, Theileria* and *Hepatozoon* in southern Europe.
391 Part I. Epizootiological aspects. *Veterinary Parasitology*, *113*, 189–201.
392 http://doi.org/10.1016/S0304-4017(03)00078-5
393 Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high
394 throughput. *Nucleic Acids Research*, *32*, 1792–1797.
395 http://doi.org/10.1093/nar/gkh340
396 Garcia-Bocanegra, I., Arenas-Montes, A., Hernández, E., Adaszek, L., Carbonero, A.,
397 Almería, S., Jaén-Téllez, J. A., Gutiérrez-Palomino, P., & Arenas, A.. (2013).
398 Seroprevalence and risk factors associated with *Babesia caballi* and *Theileria equi*
399 infection in equids. *Veterinary Journal*, *195*, 172–178.
400 http://doi.org/10.1016/j.tvjl.2012.06.012
401 Guidi, E., Pradier, S., Lebert, I., & Leblond, A. (2015). Piroplasmosis in an endemic area:
402 analysis of the risk factors and their implications in the control of Theileriosis and
403 Babesiosis in horses. *Parasitology Research*, *114*, 71–83.

http://doi.org/10.1007/s00436-014-4161-9

Hall, C. M., Busch, J. D., Scoles, G. A., Palma-Cagle, K. A., Ueti, M. W., Kappmeyer, L. S., & Wagner, D. M. (2013). Genetic characterization of *Theileria equi* infecting horses in North America: evidence for a limited source of US introductions. *Parasites & Vectors*, *6*. http://doi.org/10.1186/1756-3305-6-35

Hemmink, J. D., Sitt, T., Pelle, R., de Klerk-Lorist, L. M., Shiels, B., Toye, P. G., Morrison, W. I., & Weir, W. (2018). Ancient diversity and geographical sub-structuring in African buffalo *Theileria parva* populations revealed through metagenetic analysis of antigen-encoding loci. *International Journal for Parasitology*, *48*, 287–296. http://doi.org/10.1016/j.ijpara.2017.10.006

Klein, S.L. (2004). Hormonal and immunological mechanisms mediating sex differences in parasite infection. *Parasite Immunology 26*, 247–264. http://doi.org/10.1111/j.0141-9838.2004.00710.x

Knowles, D. P., Kappmeyer, L. S., Haney, D., Herndon, D. R., Fry, L. M., Munro, J. B., Sears, K., Ueti, M. W., Wise, L. N., Silva, M., Schneider, D. A., Grause. J., White, S. N., Tretina, K., Bishop, R. P., Odongo, D. O., Pelzel-McCluskey, A. M., Scoles, G. A., Mealey, R. H., & Silva, J. C. (2018). Discovery of a novel species, *Theileria haneyi* n. sp., infective to equids, highlights exceptional genomic diversity within the genus *Theileria*: implications for apicomplexan parasite surveillance. *International Journal for Parasitology*, *48*, 679–690. http://doi.org/10.1016/j.ijpara.2018.03.010

Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular Biology and Evolution*, *33*, 1870–1874. http://doi.org/10.1093/molbev/msw054

Larsson, A. (2014). AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics*, *30*, 3276–3278. http://doi.org/10.1093/bioinformatics/btu531

Laveran. (1901). Contribution to the study of the *Piroplasma equi*. *Comptes Rendus Des*

Seances De La Societe De Biologie Et De Ses Filiales, 53, 385–388.

Manna, G., Cersini, A., Nardini, R., Bartolomé Del Pino, L. E., Antognetti, V., Zini, M., Contia, R., Lorenzettia, R., Venezianoc, V., Autorinoa, G. L., & Sciclunaa, M. T. (2018). Genetic diversity of *Theileria equi* and *Babesia caballi* infecting horses of Central-Southern Italy and preliminary results of its correlation with clinical and serological status. *Ticks and Tick-Borne Diseases*, *9*, 1212–1220. http://doi.org/10.1016/j.ttbdis.2018.05.005

Munkhjargal, T., Sivakumar, T., Battsetseg, B., Nyamjargal, T., Aboulaila, M., Purevtseren, B., Bayarsaikhan, D., Byambaa, B., Terkawi, M. A., Yokoyama, N., & Igarashi, I. (2013). Prevalence and genetic diversity of equine piroplasms in Tov province, Mongolia. *Infection, Genetics and Evolution : Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases*, *16*, 178–185. http://doi.org/10.1016/j.meegid.2013.02.005

Nagore, D., García-Sanmartín, J., García-Pérez, A. L., Juste, R. A., & Hurtado, A. (2004). Detection and identification of equine *Theileria* and *Babesia* species by reverse line blotting: epidemiological survey and phylogenetic analysis. *Veterinary Parasitology*, *123*, 41–54. http://doi.org/10.1016/j.vetpar.2004.04.010

Nicolaiewsky, T. B., Richter, M. F., Lunge, V. R., Cunha, C. W., Delagostin, O., Ikuta, N., Fonseca, A. S., Da Silva, S. S., & Ozaki, L. S. (2001). Detection of *Babesia equi* (Laveran, 1901) by nested polymerase chain reaction. *Veterinary Parasitology*, 101, 9–21. https://doi.org/10.1016/S0304-4017(01)00471-X

Piantedosi, D., D'Alessio, N., Di Loria, A., Di Prisco, F., Mariani, U., Neola, B., Santoro, M., Montagnaro, S., Capelli, G., & Veneziano, V. (2014). Seroprevalence and risk factors associated with *Babesia caballi* and *Theileria equi* infections in donkeys from Southern Italy. *Veterinary Journal*, *202*, 578–582. http://doi.org/10.1016/j.tvjl.2014.09.025

Qablan, M. A., Obornik, M., Petrželková, K. J., Sloboda, M., Shudiefat, M. F., Horin, P.,

458 Lukes, J., & Modry, D. (2013). Infections by *Babesia caballi* and *Theileria equi* in
459 Jordanian equids: epidemiology and genetic diversity. *Parasitology*, *140*, 1096–
460 1103. http://doi.org/10.1017/S0031182013000486
461 Qablan, M. A., Sloboda, M., Jirku, M., Obornik, M., Dwairi, S., Amr, Z. S., Horin, P., Lukes, J.,
462 & Modry, D. (2012). Quest for the piroplasms in camels: identification of *Theileria*
463 *equi* and *Babesia caballi* in Jordanian dromedaries by PCR. *Veterinary Parasitology*,
464 *186*, 456–460. http://doi.org/10.1016/j.vetpar.2011.11.070
465 Raftery, A. G., Jallow, S., Rodgers, J., & Sutton, D. G. M. (2019). Safety and efficacy of three
466 trypanocides in confirmed field cases of trypanosomiasis in working equines in The
467 Gambia: a prospective, randomised, non-inferiority trial. *PLoS Neglected Tropical*
468 *Diseases*, *13*, e0007175. http://doi.org/10.1371/journal.pntd.0007175
469 Rapoport, A., Aharonson-Raz, K., Berlin, D., Tal, S., Gottlieb, Y., Klement, E., & Steinman, A.
470 (2014). Molecular characterization of the *Babesia caballi* rap-1 gene and
471 epidemiological survey in horses in Israel. *Infection, Genetics and Evolution : Journal*
472 *of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases*, *23*, 115–
473 120. http://doi.org/10.1016/j.meegid.2014.01.033
474 RStudio Team. (2016). RStudio: Integrated Development for R. Boston, USA. Retrieved
475 from http://www.rstudio.com/
476 Salim, B., Bakheit, M. A., Kamau, J., Nakamura, I., & Sugimoto, C. (2010). Nucleotide
477 sequence heterogeneity in the small subunit ribosomal RNA gene within *Theileria*
478 *equi* from horses in Sudan. *Parasitology Research*, *106*, 493–498.
479 http://doi.org/10.1007/s00436-009-1691-7
480 Sears, K.P., Kappmeyer, L.S., Wise, L.N., Silva, M., Ueti, M.W., White, S., Reif, K.E. &
481 Knowles, D.P. (2019). Infection dynamics of *Theileria equi* and *Theileria haneyi*, a
482 newly discovered apicomplexan of the horse. *Veterinary Parasitology*, *271*, 68-75,
483 https://doi.org/10.1016/j.vetpar.2019.06.009
484 Sevinc, F., Maden, M., Kumas, C., Sevinc, M., & Ekici, O. D. (2008). A comparative study on

485 the prevalence of *Theileria equi* and *Babesia caballi* infections in horse sub-
486 populations in Turkey. *Veterinary Parasitology*, *156*, 173–177.
487 http://doi.org/10.1016/j.vetpar.2008.06.006
488 Weir, W., Karagenc, T., Gharbi, M., Simuunza, M., Aypak, S., Aysul, N., Darghouth, M. A.,
489 Shiels, B., & Tait, A. (2011). Population diversity and multiplicity of infection in
490 *Theileria annulata*. *International Journal for Parasitology*, *41*, 193–203.
491 http://doi.org/10.1016/j.ijpara.2010.08.004
492 Wise, L. N., Kappmeyer, L. S., Mealey, R. H., & Knowles, D. P. (2013). Review of equine
493 piroplasmosis. *Journal of Veterinary Internal Medicine / American College of*
494 *Veterinary Internal Medicine*, *27*, 1334–1346. http://doi.org/10.1111/jvim.12168
495

Primer	Direction	Name	Sequence	Melting temperature (°C)	Expected amplicon size (base pairs)
Babesia/Theileria catch-all	Forward	BT1-F	5' GGT TGA TCC TGC CAG TAG T	56.7	1077
	Reverse	BTH-1R	5' TTG CGA CCA TAC TCC CCC CA	61.4	
T. equi genotyping assay	Forward	PCR-FWD	5' CCA GCT CCA ATA GCG TAT ATT A	56.5	147 - 153 (dependant on clade)
	Reverse	PCR-REV	5' TTC CAT GCT RAA GTA TTC AAG	53.0	

Table 1 - Primers for the Theileria equi genotyping assay.

Probe	Name	Sequence	5' - Dye	Wavelength	3' - Quencher	Tm (°C)
Clade A	Probe-A	5'- CAT CGT TGC GGC TTG GTT GGG TTT CGA TTA	FAM	450 - 490 nM	BHQ1	68.1
Clade B	Probe-B	5'- TAG TTG GGG CAT GTT TTC ATG ACT CGA CGT	HEX	515 - 535 nM	BHQ1	66.8
Clade C	Probe-C	5'- GTT ATG GCT TAG TTG GGT CAC TTT GTG TCC	Cy5	620 - 650 nM	BBQ650	66.8
Clade D	Probe-D	5'- TTT CCT CTG CTT GAC AGT TGG ATT TCG TTA	Texas Red	560 - 590 nM	BHQ2	64.0

498

499 *Table 2 - Theileria equi clade-specific probes.*

Probe	Maximum number of intra-clade base mismatches	Maximum identity to off-target clade sequences (%)			
		A	**B**	**C**	**D**
A	3	-	66.7	63.3	70.0
B	0	59.4	-	66.7	61.3
C	2	65.7	69.7	-	76.5
D	2	56.7	54.8	83.3	-

Table 3 - Probe target clade specificity. The maximum number base pair mismatches within the target clade of each probe are shown in the second column. The maximum identity to sequences in off-target clades are shown as a percentage in subsequent columns.

	Single clade Infection	Multi-clade Infection	Odds Ratio	Lower 95% CI	Upper 95% CI	Fisher's Exact Test P value
Species						
Horse	32.1% (9/28)	67.9% (19/28)	REF			
Donkey	50.0% (7/14)	50.0% (7/14)	2.11	0.57	7.86	0.322
Sex						
Male	18.8% (3/16)	81.2% (13/16)	REF			
Female	50.0% (13/26)	50.0% (13/26)	4.33	0.99	18.89	0.056
Age						
>10 years	47.1% (8/17)	52.9% (9/17)	REF			
0-2 years	33.3% (3/9)	66.7% (6/9)	1.74	0.26	14.45	
>2 - <10 years	31.3% (5/16)	68.7% (11/16)	1.92	0.38	10.42	0.711
Pyrexia (Horses >38.5 °C) (Donkeys>37.8 °C)						
Yes	41.2% (7/17)	58.8% (10/17)	REF			
No	33.3% (8/24)	66.7% (16/24)	0.71	0.2	2.58	0.745

503

504 *Table 4 - Risk factors of species, sex, age and pyrexia on the multiplicity of T. equi infection in a working equid population.*

Figure legends

Figure 1 – Theileria equi clade phylogeny. This neighbour-joining tree shows the position of the clade reference sequences. The position of Theileria haneyi is highlighted, and a representative sequence of Theileria parva is included to root the tree. Bootstrap values are shown as percentages, as generated from 1,000 replications. The previously described T. equi clades A-D have been annotated.

Figure 2 - Alignment of Theileria equi reference sequences showing the location of the PCR primers and clade-specific probes.

Figure 3 - Probe cross-reactivity testing. The x-axis shows cycle number and the y-axis fluorescence (Delta Rn) on each graph. The horizontal line shows a cut-off count threshold (CT) value of 0.2. The highlighted boxes show the result of each probe with the target clade reference sequence.

Figure 4 - Summary of Theileria equi clade detection in a population of horses and donkeys (n = 42) in the Gambia.