Ticks and Tick-borne Diseases 10 (2019) 101276

Contents lists available at ScienceDirect

Ticks and Tick-borne Diseases

journal homepage: www.elsevier.com/locate/ttbdis

Original article

Development of a deep amplicon sequencing method to determine the species composition of piroplasm haemoprotozoa

Umer Chaudhry[a,*], Qasim Ali[b], Imran Rashid[b], Muhammad Zubair Shabbir[b], Muhammad Ijaz[b], Muhammad Abbas[c], Mike Evans[a], Kamran Ashraf[b], Ivan Morrison[a], Liam Morrison[a], Neil D. Sargison[a]

[a] Roslin Institute, Easter Bush Veterinary Centre, University of Edinburgh, Scotland, UK
[b] University of Veterinary and Animal Sciences, Lahore, Pakistan
[c] Disease Diagnostic Laboratory, Livestock and Dairy Department, Okara, Pakistan

ARTICLE INFO

Keywords:
Piroplasmosis
Theileria
Babesia
Deep amplicon sequencing

ABSTRACT

Piroplasmosis is caused by tick-borne haemoprotozoa of the genera *Theileria* and *Babesia*. These parasitic infections can seriously impact on the health of livestock and production. Piroplasms of multiple species can be present in a single host, but reliable molecular diagnostic tools are needed in order to understand the composition of these complex parasite communities. *Theileria* and *Babesia* vary in their epidemiology, drug sensitivity, pathogenicity and interaction with co-infecting species, but are similar in that infected animals become persistent carriers after recovery from primary infection, acting as reservoir hosts. Here, we describe for the first time the use of a deep amplicon sequencing platform to identify proportions of piroplasm species in co-infecting communities and develop the concept of a "haemoprotobiome". First, four phenotypically-verified species of *Theileria* and *Babesia* were used to prepare mock DNA pools with random numbers of the parasites amplified by four different numbers of PCR cycles to assess sequence representation for each species. Second, we evaluated the detection threshold of the deep amplicon sequencing assay for each of the four species and to assess the accuracy of proportional quantification of all four species. Finally, we applied the assay to the field samples to afford insight of the species composition of piroplasm communities in small and large ruminants in the Punjab province of Pakistan. The "haemoprotobiome" concept has several potential applications in veterinary and human research, including understanding of responses to drug treatment; parasite epidemiology and ecology; species interactions during mixed infections; and parasite control strategies.

1. Introduction

Haemoprotozoa are ubiquitous and amongst the most successful organisms in the biosphere, possessing an incredible ability to adapt to many different host environments and niches (Nene et al., 2016). Among these, *Theileria* and *Babesia* are arguably the most economically important livestock parasites, and some species are highly pathogenic, causing theileriosis and babesiosis in cattle. Different species of *Theileria* and *Babesia* occur globally, with their distribution in part determined by the presence of relevant tick vector species. A number of species of these parasites significantly impact on the health, welfare and production of livestock across this range of tick distribution (Jabbar et al., 2015). It is estimated that about approximately 80% of the world's cattle population is infected with *Theileria* and/or *Babesia*, causing economic loss due to high morbidity and mortality and threatening food security in many livestock dependent communities (de Castro, 1997). Theileriosis and babesiosis are particularly important in subtropical regions, where efficient ruminant livestock production is critical to the wellbeing and poverty alleviation of smallholder farmers (Gubbels et al., 1999).

Next generation genomic resources have potential applications in the diagnosis, surveillance, treatment and control of haemoprotozoan diseases, as well as in the evaluation of parasite population responses to drug treatments and other control strategies. Determination of sequence variations in the hyper-variable 18S rDNA cistron can discriminate between haemoprotozoan piroplasm parasites (Gubbels et al., 1999) and overcome limitations of traditional gross parasitological methods for the diagnosis of haemoprotozoa at species level (Agudelo et al., 2013; Haanshuus et al., 2013; Lee et al., 2015; Lefterova et al., 2015; Mens et al., 2006; Rougemont et al., 2004; Steenkeste et al.,

* Corresponding author at: University of Edinburgh, The Roslin Institute, Easter Bush Veterinary Centre, EH25 9RG, UK.
E-mail address: uchaudhr@exseed.ed.ac.uk (U. Chaudhry).

https://doi.org/10.1016/j.ttbdis.2019.101276
Received 3 May 2019; Received in revised form 13 August 2019; Accepted 23 August 2019
Available online 24 August 2019
1877-959X/

2009). Various PCR methods (reverse line blot (RLB)-PCR, quantitative (q)PCR, multiplex PCR) have been described to amplify the 18S region for sequence determination (Bilgic et al., 2013; Chaisi et al., 2013; Gubbels et al., 1999; Kundave et al., 2018), but these are low throughput, hence relatively expensive, and can be error-prone. These methods depend on the use of species-specific primers and probes, meaning that only the tested species is identified, resulting in limited available data on the true circulating species diversity. In contrast, high throughput deep amplicon sequencing using the Illumina Mi-Seq platform is relatively low-cost and potentially less error prone. A 454 pyrosequencing assay Mans et al. (2016) has similar advantages. The method has transformed the study of bacterial (microbiome) (Gloor et al., 2010; Rogers and Bruce, 2010) and nematode parasites (nemabiome) (Avramenko et al., 2015; Chaudhry and Sargison, 2018), and has the potential to open new areas of research in the study of protozoan parasite communities. The method has the potential to accurately identify and provide relative quantification of co-infecting species of haemoprotozoa. The use of primers binding to conserved sites and analysis of up to 600 bp sequence reads allows protozoan species to be detected. The use of barcoded primers allows a large number of samples to be pooled and sequenced in a single Mi-Seq run, making the technology suitable for high-throughput analysis (Avramenko et al., 2015). By multiplexing the barcoded primer combinations, it is possible to run 384 samples at once on a single Illumina Mi-Seq flow cell, helping to reduce the cost.

Here we report the development of a method of deep amplicon sequencing of the hyper-variable 18S rDNA region using Illumina Mi-Seq platform and the validation of its accuracy for species quantification of *Theileria* and *Babesia* communities present in field samples.

2. Materials and methods

2.1. Parasite material and genomic DNA preparation

Theileria and *Babesia* species of large and small ruminants were chosen in this study for a number of reasons; firstly, the availability of laboratory isolates allowed the creation of 'mock pools' of the parasites to assess the sequence representation and the proportions of each species and secondly, piroplasm-positive Asian buffalo, cattle and sheep blood samples allowed us to demonstrate the applicability of the method to field samples.

Five stocks of piroplasm isolates (*Theileria annulata, Theileria parva, Theileria lestoquardi, Babesia bigemina* and *Babesia bovis*) were available at the Roslin Institute, University of Edinburgh, UK. - The study used a cell line infected with a stock of *Theileria annulata* originally isolated from Ankara, Turkey, a cell line infected with the Muguga stock of *Theileria parva* originally isolated in Kenya and a cell line infected with the Shiraz 1 stock of *Theileria lestoquardi* originally isolated from infected sheep in Iran. A culture of an Australian isolate (Lismore) of *Babesia bovis* and blood containing a Kenyan isolate of *Babesia bigemina* were used. gDNA was prepared using 1000 µl Direct PCR lysis reagent (Viagen), 50 µl proteinase K solution (Qiagen), and 50 µl 1 M dithiothreitol (DTT). To extract gDNA, 50 µl of infected cell line from each of the species was transferred into a fresh tube and centrifuged for 5 min, before removing the supernatant and mixing with 25 µl lysis buffer (Viagen) and Proteinase K (New England BioLabs). Lysates were prepared as previously described by Chaudhry et al. (2015). Mock pools of the four isolates (*T. annulata, T. parva, B. bigemina* and *B. bovis*) were created. Three replicates of the four mock pools were amplified using four different numbers of PCR cycles (25X, 30X, 35X and 40X), to assess the sequence representation of each species. Three replicates of nine mock pools were used to test the detection threshold of the deep amplicon sequencing method and to show the proportions of each of the species present, which comprised the following: [Mix1 (*T. annulata*, *T. parva*, *B. bigemina* and *B. bovis*), Mix2 (*T. annulata*, *T. parva* and *B. bovis*), Mix3 (*T. annulata*, *T. parva* and *B. bigemina*), Mix4 (*T. parva*, *B. bigemina* and *B. bovis*), Mix5 (*T. annulata*, *B. bigemina* and *B. bovis*), Mix6 (*B. bovis*), Mix7 (*B. bigemina*), Mix8 (*T. annulata*), Mix9 (*T. parva*)].

A total of 79 piroplasm-positive Asian buffalo blood samples, 86 piroplasm-positive cattle blood samples and 18 piroplasm-positive sheep blood samples was collected from veterinary clinics throughout the Punjab province of Pakistan between 2017 and 2018. Moreover, 5 piroplasm-negative cattle blood samples were provided by Dr Tim Connelly, Roslin Institute. The procedures involved jugular venipuncture and withdrawal of 5 ml of intravenous blood into an EDTA tubes, followed by storage at -20°C. Discussions were held with key administrative and community leaders to raise awareness of the study, and encourage households to participate. Samples were taken by trained para-veterinary workers under the supervision of local veterinary staff. The study was approved by the Institutional Review Board of the University of Veterinary and Animal Sciences Punjab, Pakistan. Thick and thin blood smears were stained with 10% Giemsa solution and examined under oil immersion (100 x lens) to identify piroplasms. To extract gDNA, 50 µl of blood from each of the piroplasm positive samples was used as template, and the DNA was extracted according to the protocols described in the TIANamp blood DNA kit (Beijing Biopeony Co. Ltd).

2.2. Deep amplicon sequencing

The overall scheme of the deep amplicon sequencing approach using Illumina Mi-Seq platform is described below and shown in Fig. 1.

2.2.1. Adapter PCR amplification of 18S rDNA locus

463 to 504 bp fragments encompassing parts of the 18S rDNA spanning V4 hyper-variable region of haemoprotozoa were targeted for 1st round adapter PCR amplification. The primers were altered from the normal primers set (RLB_For, RLB_Rev) previously described by Gubbels et al. (1999). Adapters were added to each primer to allow the successive annealing and N is the random number of nucleotides included between the adapter and primer set (Supplementary Table S1). Equal proportions of four forward (RLB_For, RLB_For-1N, RLB_For-2N, RLB_For-3N) and four reverse primers (RLB_Rev, RLB_Rev-1N, RLB_Rev-2N, RLB_Rev-3N) were mixed and used for the adapter PCR with following conditions: 10 uM forward and reverse adapter primer, 10 mM dNTPs, 0.5 U DNA polymerase enzyme, 5X buffer (KAPA Biosystems, USA) and 1 µl of gDNA. The thermocycling conditions were 95 °C for 2 min, followed by 35 cycles of 98 °C for 20 s, 60 °C for 15 s, 72 °C for 15 s and a final extension of 72 °C for 5 min. The PCR product was purified with AMPure XP Magnetic Beads (1X) according to the manufacturer's instructions (Beckman Coulter).

2.2.2. Barcoded PCR amplification of 18S rDNA locus

The 2nd round barcoded PCR was performed by using sixteen forward and twenty-four reverse barcoded primers obtained from Illumina Mi-Seq protocols (Supplementary Table S2). Repetitions of the forward and reverse barcoded primers in different samples were avoided. The barcoded PCR was performed with the following conditions: 10 u M barcoded forward (N501 to N508, N510 to N517) and reverse (N701 to N712 and N713 to N724) primers, 10 mM dNTPs, 0.5 U DNA polymerase enzyme, 5X buffer (KAPA Biosystems,USA) and 2 µl of adaptor PCR product as DNA template. The thermocycling conditions were 98 °C for 45 s, followed by 7 cycles of 98 °C for 20 s, 63 °C for 20 s, and 72 °C for 2 min. The PCR product was purified with AMPure XP Magnetic Beads (1X) according to the manufacturer's instructions (Beckman Coulter).

2.2.3. Illumina Mi-Seq run, data handling and bioinformatics analysis

The pooled library was prepared with 10 µl of barcoded PCR bead purified product from each individual sample and measured with the KAPA qPCR library quantification kit (KAPA Biosystems, USA), before being run on an Illumina Mi-Seq Sequencer using a 600-cycle pair end

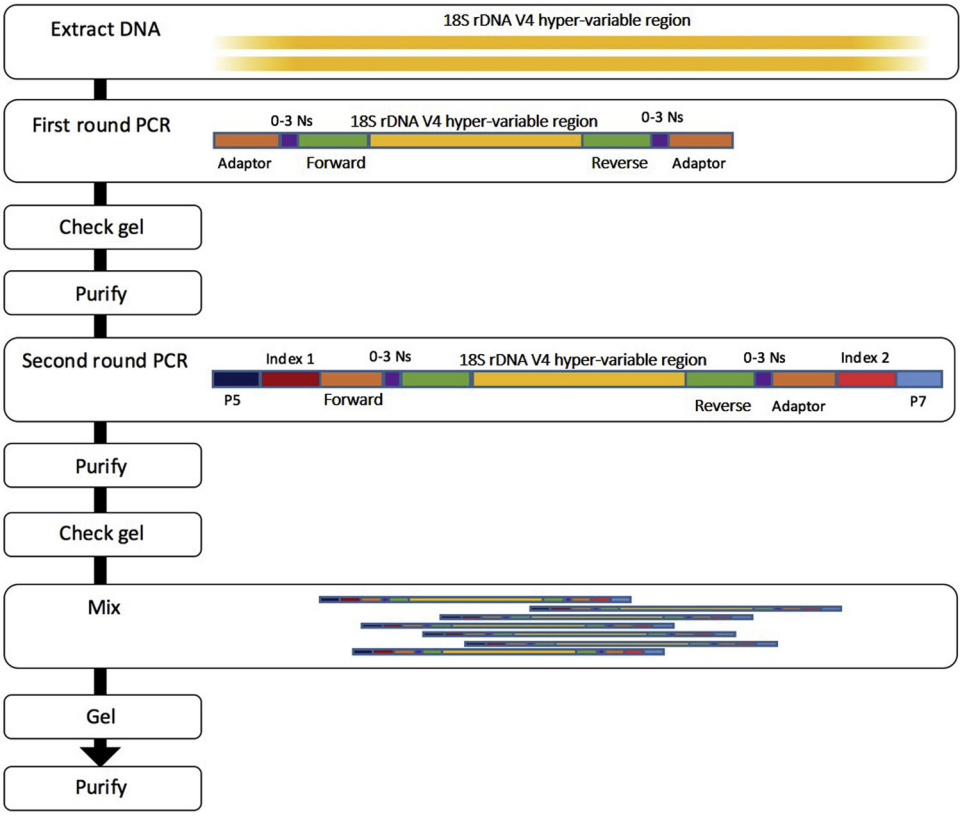

Fig. 1. Schematic representation of the preparation of the Illumina sequencing library. In the first-round adaptor PCR amplification, overhanging primers were used to amplify the 18S rDNA to generate a fragment that includes the V4 hyper-variable region as well as the specific primer regions. The adaptor regions provide the target sites for the primers used for sequencing the fragment. The 0–3 random nucleotides ('N's) are inserted between adaptor sequences and the binding site of the primers to offset the reading frame, thereby increasing diversity when the amplicons are sequenced to prevent oversaturation of the Mi-Seq sequencing channels. A second round barcoded limited-cycle PCR amplification was then performed using overhanging primers that bind to the adaptor tags of the amplicon to add indices as well as the P5 and P7 regions required to bind to the Illumina flow cell.

reagent kit (Mi-Seq Reagent Kits v2, MS-103-2003) at a concentration of 15 nM with addition of 25% Phix Control v3 (Illumina, FC-11-2003). The Mi-Seq separates all sequencing data by sample during post-run processing using the barcoded indices to generate FASTQ files. The data were analysed with our own bioinformatics Command Prompt pipeline. Illumina Mi-Seq data analysis was performed using Illumina Mi-seq SOPs (Kozich et al., 2013) and Mothur v1.39.5 software (Schloss et al., 2009). Briefly, raw paired reads were run into make.contigs command, and those that were too long, or ambiguous bases were removed. The sequence data were trimmed before the region of forward and reverse primers and aligned with the consensus sequence library (see below) using align.seqs command. The sequence data that did not align with the 18S rDNA consensus sequence taxonomy library was discarded as a non-specific amplification. To achieve this, summary.seqs command was used to identify the 463 to 504 bp fragments encompassing parts of the 18S rDNA spanning the V4 hyper-variable region of *Theileria and Babesia* species. Finally, all bulk 18S rDNA sequences overlaps in the same regions were run on the screen.seqs command to give the 18S rDNA matched sequences of *Theileria and Babesia* species. All the bulk sequences overlap regions were further run on the screen.seqs command to generate 18S rDNA matched sequences of *Theileria and Babesia* species. Overall, thousands of 18S rDNA reads were generated from the data set of five *Theileria* and *Babesia* isolates stock and piroplasm positive field samples [buffalo (n = 79), cattle (n = 86) and sheep (n = 18)].

2.3. Phylogenetic analysis of 18S rDNA sequences

Deep amplicon sequencing of the 18S rDNA of three *Theileria* species (*T. annulata, T. parva, T. lestoquardi*), and two *Babesia* species (*B. bigemina and B. bovis*) is described in the previous section. The generated reads were compared to the 18S sequences of other *Theileria* and *Babesia* species published in NCBI GenBank to account for any additional genetic diversity. The obtained 18S rDNA sequences were first aligned in Geneious v10.2.5 (Biomatters Ltd, New Zealand) and then imported into the CD-HIT online software to calculate the number of consensus sequences generated from each species (for more detail see Supplementary Table S3). A phylogenetic tree of the 18S rDNA consensus sequences was constructed by HKY + G model of substitution using the Maximum Likelihood method in the MEGA 5.05 software (Huson and Bryant, 2006). The program jModeltest 12.2.0 was used to select the appropriate model of nucleotide substitutions for Maximum Likelihood analysis (Posada, 2008). Branch supports were obtained by 1000 bootstraps of the data.

3. Results

3.1. Consensus sequence library preparation and the assessment of 18S rDNA genetic diversity in Theileria and Babesia species

A total of 43 consensus sequences of the 18S rDNA locus were identified among eight large ruminant *Theileria* species (*T. annulata = 12, T. parva = 6, T. orientalis = 5, T. sergenti = 2, T. buffeli = 5, T. velifera = 4, T. taurotragi = 2, T. mutans = 7*) and 43 consensus sequences were identified among six large ruminant *Babesia* species (*B. bigemina = 17, B. bovis = 15, B. orientalis = 2, B. occultans = 4, B. ovata = 3, B. major = 2*) (Fig. 2A, Supplementary Table S3). A phylogenetic tree of the 86 consensus sequences demonstrates a distinct clustering between species (Fig. 2A). Comparison of the genetic distance of 18S rDNA revealed 87% to 98% identity within the eight *Theileria* species and varied from 77% to 97% identity within the six *Babesia* species. The genetic distance also revealed differences in identity ranging from 72% to 86% between *Theileria* and *Babesia* genera (Supplementary Table S4). The most closely related species of *Theileria* and *Babesia* could still be reliably differentiated by virtue of 18S rDNA sequence variations (Supplementary Table S5).

A total of 21 consensus sequences of the 18S rDNA locus were identified among four small ruminant *Theileria* species (*T. lestoquardi = 3, T. ovis = 8, T. uilenbergi = 2, T. luwenshuni = 8*) and four consensus sequences in *B. ovis* (Fig. 2B, Supplementary Table S3). The

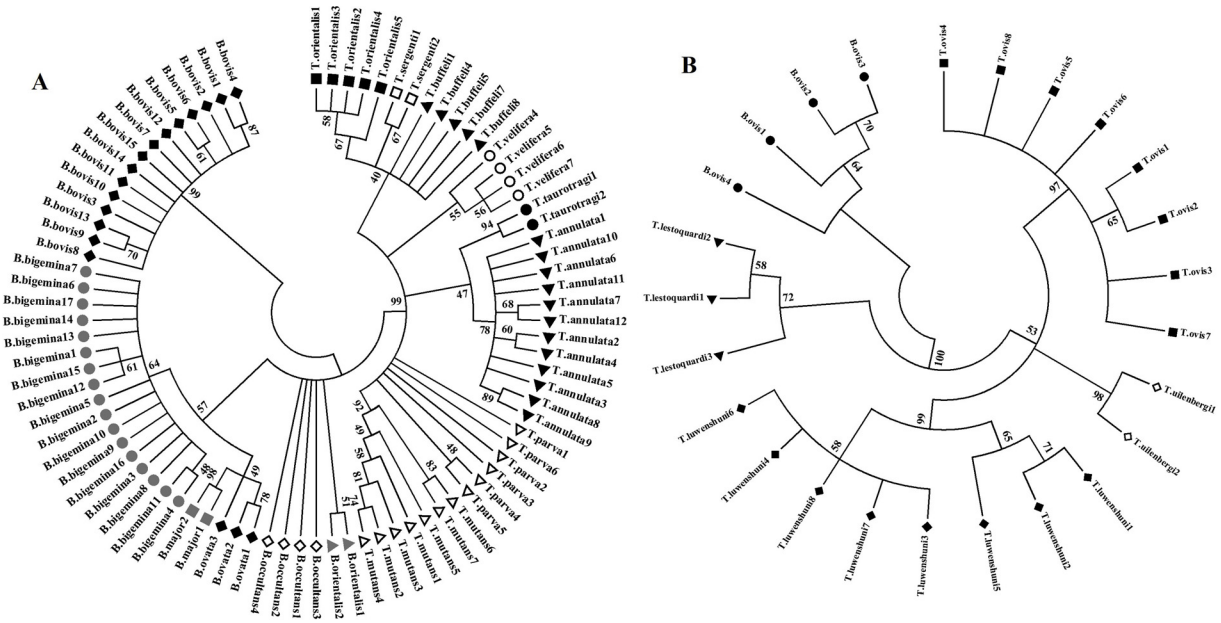

Fig. 2. Maximum Likelihood tree of 86 and 25 consensus sequences were obtained from large and small ruminants *Theileria* and *Babesia* species. The NCBI GeneBank sequences were first aligned and calculate the number of consensus sequences generated from each species (for more detail see Supplementary Table S3). (A) 86 consensus sequences of the 18S rDNA were identified among eight *Theileria* and six *Babesia* species of large ruminants. (B) 25 consensus sequences were identified among four *Theileria* and one *Babesia* species of small ruminants. The neighbor-joining tree (Kimura 2-parameter model) was computed with 1000 bootstrap replicates using MEGA5 software created by Biomatters. Each species was identified with different shades bars.

phylogenetic tree of the 25 consensus sequences showed distinct clustering between species (Fig. 2B). Genetic distance comparison of the 18S rDNA revealed 98% to 95% identity within the four *Theileria* species and differences in identity ranging from 81% to 83% between *Theileria* and *Babesia* (Supplementary Table S4).

3.2. Evaluation of the sequence representation of each of the four species (T. annulata, T. parva, B. bigemina and B. bovis) using different number of PCR cycles

The deep amplicon sequencing assay was performed three independent times from four isolates to assess the sequence representation of each species (Fig. 3, Supplementary Table S6). The number of cycles for the adapter PCR had no impact on the sequence representation of each species. Each of the species was significantly represented in each number of cycles, based on the number of sequences generated from each mock pool (Fig. 3A). The sequence results of *B. bovis* was abundant in the mock pool samples. This is probably reflecting the fact that they were in erythrocyte culture of an Australian isolate (Lismore) and hence the proportion of parasites and gDNA is much lower than in the other species samples. Nonetheless, there was no statistically significant differences in the proportion of each species between different numbers of PCR cycles (overall $H_{(3)} < 0.04$, $p > 0.9$; Kruskal-Wallis rank sum test). Calculating the mean values based on each cycle of amplification removed further deviation, indicating that level of amplification does not influence the proportional representation of each species (Fig. 3B).

3.3. Validation of the deep amplicon sequencing assay using mock pools mixes of T. annulata, T. parva, B. bigemina and B. bovis

Three replicates each of known admixtures were created from gDNA from four isolates, to validate the deep amplicon sequencing assay (Fig. 4, Supplementary Table S7). The mixing of different mock pools demonstrates the accurate detection ability of the deep amplicon sequencing method and to show the proportions of each of the species

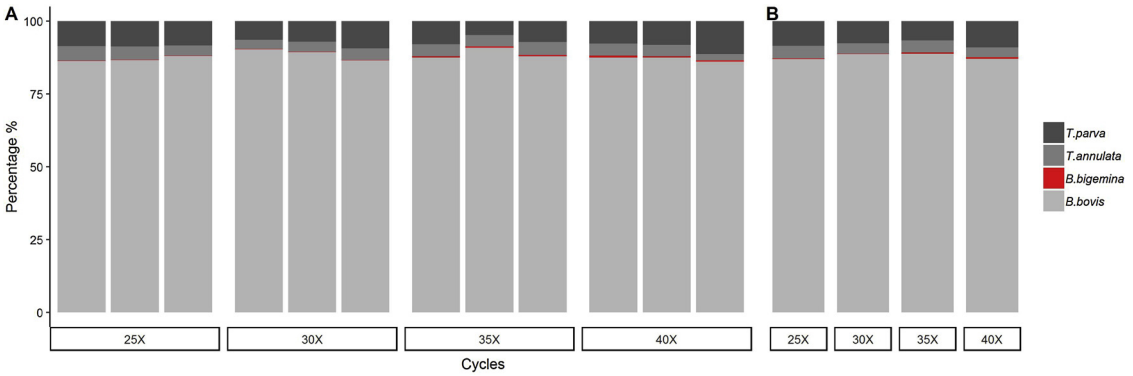

Fig. 3. Assessment of species representation bias for the deep amplicon sequencing assay. A gDNA was prepared from a mock pool of the random number of parasites from each of the following species; *T. annulata, T. parva, B. bigemina and B. bovis*. 2A. The sequencing assay was applied three times at 25, 30, 35 and 40 cycles of amplification respectively as denoted on the X-axis and Y-axis shows the percentage proportions of each species. 2B. Replicates were grouped and averaged based on cycles of amplification.

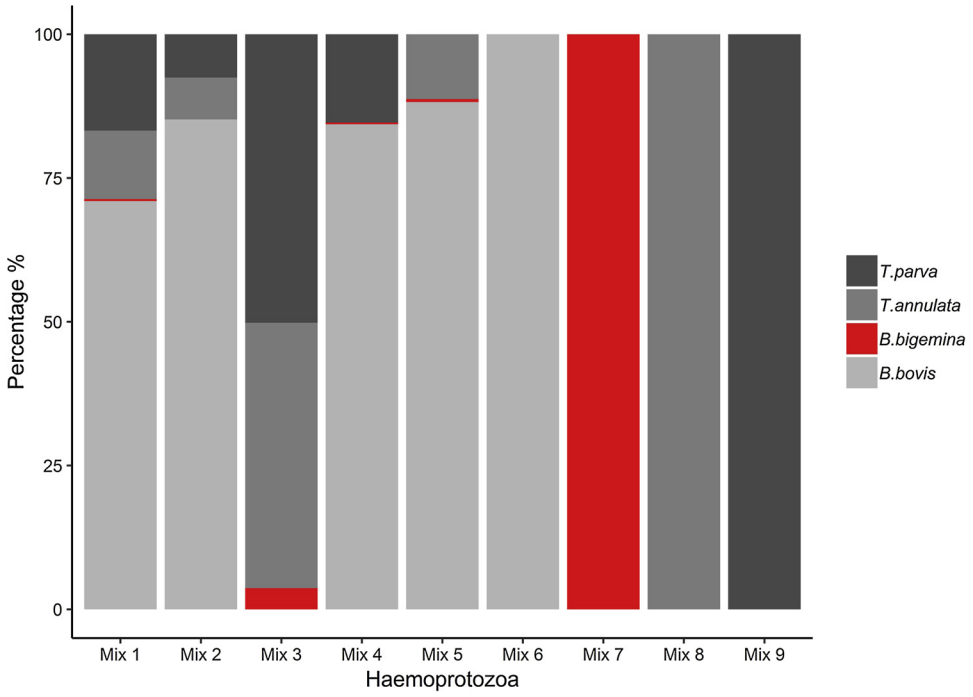

Fig. 4. Validation of the deep amplicon sequencing assay. Nine separate mock pools of the random number of parasites from each the four haemoprotozoa species were created. The panel shows the species proportions estimated from the results of the deep amplicon sequencing assay. In each case, the Y-axis shows the percentage proportions of each species and the X-axis indicates the mixing of the mock pools from each species; Mix1(*T. annulata*, *T. parva*, *B. bigemina* and *B. bovis*), Mix2(*T. annulata*, *T. parva* and *B. bovis*), Mix3(*T. annulata*, *T. parva* and *B. bigemina*), Mix4(*T. parva*, *B. bigemina* and *B. bovis*), Mix5(*T. annulata*, *B. bigemina* and *B. bovis*), Mix6(*B. bovis*), Mix7(*B. bigemina*), Mix8(*T. annulata*), Mix9(*T. parva*).

present. There was no statistically substantial difference in the proportions of each of the species in different admixtures (Kruskal-Wallis rank sum test; Mix1: $\chi2(1)$ 0.038462, p = 0.981; Mix2: $\chi2(1)$ 0.27305, p = 0.8724; Mix3: $\chi2(1)$ 0.039007, p = 0.9807; Mix4: $\chi2(1)$ 0.039007, p = 0.9807; Mix5: $\chi2(1)$ 0.04923, p = 0.7211; Mix6: $\chi2(1)$ 0, p = 1; Mix7: $\chi2(1)$ 0, p = 1; Mix8: $\chi2(1)$ 0, p = 1; Mix9: $\chi2(1)$ 0, p = 1). The Mix 1 pool contained all four isolates, the results were perfectly matched (Fig. 4). The Mix 2 contained *T. annulata, T. parva* and *B. bovis*, Mix 3 contained *T. annulata, T. parva* and *B. bigemina*, Mix 4 contained *T. parva, B. bigemina* and *B. bovis* and Mix 5 contained *T. annulata, B. bigemina* and *B. bovis,* the results were accurate, with no significant variations between replicates (Fig. 3). For the pools of 100% of each species (Mix 6, Mix 7, Mix 8, Mix 9), the results are also consistent (Fig. 4).

3.4. Assessment of the deep amplicon sequencing in the quantification of piroplasm positive blood samples

The deep amplicon sequencing was applied to the piroplasm-positive blood samples of Asian buffalo, cattle and sheep to detect the presence and relative prevalence of infections with these *Theileria* and *Babesia* species in the field (Fig. 5, Supplementary Table S8). Overall, 183-piroplasm positive field samples (buffalo (n = 79), cattle (n = 86) and sheep (n = 18) were examined to identify the species of *Theileria* and *Babesia* involved in the infections (Fig. 5). Those samples yielding more than 1000 reads (implying sufficient gDNA for accurate amplification) were included in the analysis (Supplementary Table S8). The results demonstrated that the prevalence of *T. annulata* infection was higher than that of *B. bigemina* and *B. bovis* infection in the buffalo and cattle samples. All 79 buffalo and 86 cattle samples were *T. annulata*-positive, 10/79 buffalo and 10/86 cattle samples were *B. bovis* positive, and 3/79 buffalo and 3/86 cattle samples were *B. bigemina* positive. In the sheep samples, the prevalence of *T. lestoquardi* and *T. ovis* infections were higher than that of *B. ovis*. All 18 samples were *T. lestoquardi* positive, 15/18 samples were *T. ovis* positive, and 4/18 were *B. ovis* positive.

4. Discussion

Several blood-borne protozoan diseases cause important health problems, resulting in economic losses and reduced productivity in domestic livestock worldwide (Nene et al., 2016). Among these, theileriosis is ranked to be the most significant disease among the tick-borne pathogens of ruminants (Mans et al., 2015). The most pathogenic species are *T. annulata, T. parva* and *T. lestoquardi*, while *T. mutans*, *T. orientalis*, *T. taurotragi*, and *T. ovis*, mostly cause asymptomatic infections in livestock (Nene et al., 2016). Depending on the species of *Theileria*, disease is transmitted by ticks of the genera *Hyalomma*, *Rhipicephalus, Haemaphysalis* and *Amblyomma*. *Theileria* is distributed worldwide, and is an important cause of disease in livestock in tropical and subtropical regions of Asia and Africa (Sivakumar et al., 2014). Babesiosis infects a wide range of domestic livestock resulting in severe economic losses (Bock et al., 2004). The most pathogenic species are *B. bovis*, *B. bigemina*, *B. divergens* and *B. ovis* (Sivakumar et al., 2016), while *B. occultans*, *B. ovata*, *B. orientalis* and *B. major* are considered to be less pathogenic (Decaro et al., 2013; He et al., 2017; Ros-Garcia et al., 2011; Sivakumar et al., 2016). Depending on the species of *Babesia*, ticks of the genera *Rhipicephalus* (*Boophilus*) and *Ixodes* are involved in disease transmission. *Babesia* are distributed worldwide, especially in tropical and subtropical regions (Jabbar et al., 2015). The severity of disease is influenced by tick infestation, climatic changes, geographical regions, co-grazing and time of the year. Haemoprotozoan species vary in terms of disease epidemiology, drug sensitivity, disease pathogenicity and control managements (Mans et al., 2015).

Giemsa-stained blood smears are the standard method for the detection of haemoprotozoa in the blood of infected animals. This method is generally useful in acutely infected animals, but insensitive in the detection of chronically infected carriers, where the level of parasitaemia is low. This method is laborious and error prone in inexperienced hands (Bose et al., 1995). Serological tests have been developed to detect circulating antibodies against the parasites, but these generally have poor sensitivity and specificity due to cross-reactivity or non-specific immune responses (Passos et al., 1998), and only detect previous exposure as opposed to current infection. Conventional PCR methods are useful in the detection of particular haemoprotozoan species for which the reagents and conditions have been developed, but

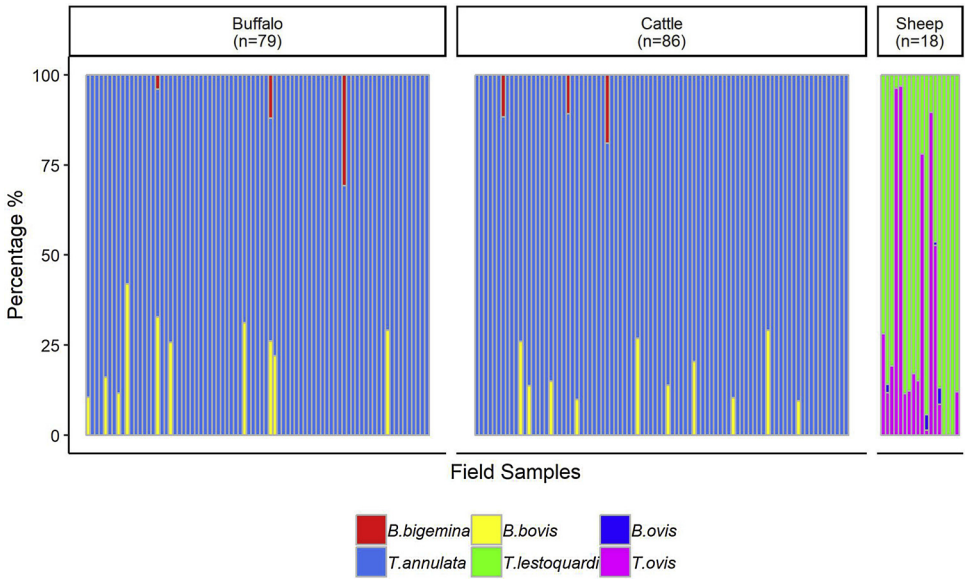

Fig. 5. Deep amplicon sequencing for the quantification of haemoprotozoan parasites from field samples. A total of 79 positive blood samples of buffalo, 86 from cattle and 18 from sheep were collected from veterinary clinics throughout the Punjab province of Pakistan. The samples were collected into an EDTA tube and stored at -20 °C. Thick and thin blood smears were examined to identify piroplasm before extracting gDNA. The amplicon sequencing assay was applied with unknown parasitemia level of each sample. The bar of each samples shows the proportion of each species as estimated by deep amplicon sequencing assay. The Y-axis shows the percentage proportions of each species.

have limitations in terms of lacking scalability (Bilgic et al., 2013; Chaisi et al., 2013; Gubbels et al., 1999; Kundave et al., 2018). In contrast, the deep amplicon sequencing method described in this study potentially provides a more reliable and accurate automated high-throughput analysis of all haemoprotozoan piroplasm species in a sample. Therefore, we have evaluated a deep amplicon sequencing method to identify the presence of *Theileria* and *Babesia* species in large and small ruminants by: (a) assessing sequence representation in different numbers of PCR cycles, (b) validating the assay using mock pools of each species and (c) applying the method to quantify the species present in piroplasm-positive blood samples from field.

The use of a single PCR, utilizing primers conserved between the parasite species, provides tools to measure the relative sequence representation of each parasite species in each sample. Our initial experiments with pools of laboratory-maintained protozoan parasites, showed that the relative representation of the different parasite sequences was unaffected by either the parasite species composition of the sample or the number of PCR cycles employed. In a previous study of nematode parasites (Avramenko et al. (2015) a bias was observed in sequence representation in PCR products used for sequencing, arising from the number of first round PCR cycles. We tested the ability of the assay to accurately determine the relative species proportions in various combinations of parasite species. To do this, we generated a mock pools containing different estimated proportions of *Theileria* and *Babesia* and showed no significant variations between replicates. Having validated the Illumina Mi-Seq platform using mock pools of piroplasm laboratory isolates, we applied the method to field samples collected from buffalo, cattle and sheep in the Punjab province of Pakistan, where theileriosis caused by *T. annulata*, *T. lestoquardi*, *T. ovis* and babesiosis caused by *B. bovis*, *B. bigemina*, *B. ovis* have been reported to infect buffalo, cattle and sheep (Jabbar et al., 2015). Our findings are consistent with these previous reports, confirming that theileriosis and babesiosis are the major infectious piroplasm diseases of large and small ruminants in Pakistan. Qayyum et al. (2010) and Durrani et al. (2006) described the biochemical and hematological profile of cattle and buffalo with clinical signs consistent with theileriosis and babesiosis and found significant changes in haemoglobin, total erythrocyte count, packed cell volume and total leukocyte count, globulins and serum protein level. In addition, theileriosis and babesiosis have been identified using stained blood smears as a diagnostic method (Jabbar et al., 2015). However, few studies have utilised conventional PCR based methods for the detection of *Theileria* and *Babesia* species in large and small ruminants (Durrani and Kamal, 2008; Khan et al., 2013; Shahnawaz et al., 2011).

To summarise, we describe for the first time the use of deep amplicon sequencing using an Illumina Mi-Seq platform to identify *Theileria* and *Babesia* species in mixed species populations and demonstrate its reliability on the field samples. This work was undertaken to explore the possibilities for the application of a high throughput practical method to determine the dynamics of piroplasm co-infections in particular *Theileria* and *Babesia* parasites within mixed species field populations. Our results are proof of concept for a a 'haemoprotobiome' method for use in disease surveillance in livestock and humans, similar to current trends in 'nemobiome' and 'microbiome' approaches that have revolutionised the study of gastrointestinal nematodes and bacteria (Avramenko et al., 2015; Chaudhry and Sargison, 2018; Gloor et al., 2010; Rogers and Bruce, 2010). This work was undertaken to explore the possibilities for the application of a high throughput practical method to determine the dynamics of disease epidemiology including *Theileria* and *Babesia* co-infections within mixed species field populations. Accurate description of the 'haemoprotobiome' has also applications in monitoring changes in parasite diversity after emergence and spread of drug resistance, and, or in understanding the impact of host immune selection, for example following vaccination against specific species (Altay et al., 2008; Bock et al., 2004; Nayel et al., 2012; Nene et al., 2016; Sivakumar et al., 2014).

Declaration of Competing Interest

None.

Acknowledgment

The study was financially supported by Roslin Institute uses facilities funded by the Carnegie Trust Scotland and Biotechnology and Biological Sciences Research Council (BBSRC). Work at the University of Veterinary and Animal Science Pakistan uses facilities funded by the Higher Education Commission of Pakistan. The authors of this study would like thanks Dr. Talat Naseer Pasha - Vice Chancellor of the University of Veterinary and Animal Science Lahore Pakistan for his great support in the arrangements of the sample collections.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.ttbdis.2019.101276.

References

Agudelo, O., Arango, E., Maestre, A., Carmona-Fonseca, J., 2013. Prevalence of gestational, placental and congenital malaria in north-west Colombia. Malaria J. 12, 1475–2875.

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

Avramenko, R.W., Redman, E.M., Lewis, R., Yazwinski, T.A., Wasmuth, J.D., Gilleard, J.S., 2015. Exploring the gastrointestinal "Nemabiome": deep amplicon sequencing to quantify the species composition of parasitic nematode communities. PLoS One 10, e0143559.

Bilgic, H.B., Karagenc, T., Simuunza, M., Shiels, B., Tait, A., Eren, H., Weir, W., 2013. Development of a multiplex PCR assay for simultaneous detection of *Theileria annulata*, *Babesia bovis* and *Anaplasma marginale* in cattle. Exp. Parasitol. 133, 222–229.

Bock, R., Jackson, L., de Vos, A., Jorgensen, W., 2004. Babesiosis of cattle. Parasitology 129, S247–269.

Bose, R., Jorgensen, W.K., Dalgliesh, R.J., Friedhoff, K.T., de Vos, A.J., 1995. Current state and future trends in the diagnosis of babesiosis. Vet. Parasitol. 57, 61–74.

Chaisi, M.E., Janssens, M.E., Vermeiren, L., Oosthuizen, M.C., Collins, N.E., Geysen, D., 2013. Evaluation of a real-time PCR test for the detection and discrimination of *Theileria* species in the African buffalo (Syncerus caffer). PLoS One 8.

Chaudhry, U., Redman, E., Abbas, M., Muthusamy, R., Ashraf, K., Gilleard, J., 2015. Genetic evidence for hybridisation between *Haemonchus contortus* and *Haemonchus placei* in natural field populations and its implications for interspecies transmission of anthelmintic resistance. Int. J. Parasitol. 45, 149–159.

Chaudhry, U., Sargison, N., 2018. The Nemabiome of Feral Native Ponies in the West of Scotland. In-preparation. .

de Castro, J.J., 1997. Sustainable tick and tickborne disease control in livestock improvement in developing countries. Vet. Parasitol. 71, 77–97.

Decaro, N., Larocca, V., Parisi, A., Losurdo, M., Lia, R.P., Greco, M.F., Miccolis, A., Ventrella, G., Otranto, D., Buonavoglia, C., 2013. Clinical bovine piroplasmosis caused by *Babesia occultans* in Italy. J. Clin. Microbiol. 51, 2432–2434.

Durrani, A.Z., Kamal, N., Khan, M.S., 2006. Incidence of Theileriosis and estimation of packed cell volume, total erythrocyte count and hemoglobin in buffaloes. J. Anim. Plant Sci. 16, 85–88.

Durrani, A.Z., Kamal, N., 2008. Identification of ticks and detection of blood protozoa in friesian cattle by polmerase chain reacton test and estimation of blood parameters in district Kasur, Pakistan. Trop. Anim. Health Prod. 40, 441–447.

Gloor, G.B., Hummelen, R., Macklaim, J.M., Dickson, R.J., Fernandes, A.D., MacPhee, R., Reid, G., 2010. Microbiome profiling by illumina sequencing of combinatorial sequence-tagged PCR products. PLoS One 5, 0015406.

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.

Haanshuus, C.G., Mohn, S.C., Morch, K., Langeland, N., Blomberg, B., Hanevik, K., 2013. A novel, single-amplification PCR targeting mitochondrial genome highly sensitive and specific in diagnosing malaria among returned travellers in Bergen, Norway. Malar. J. 12, 1475–2875.

He, L., Liu, Q., Yao, B., Zhou, Y., Hu, M., Fang, R., Zhao, J., 2017. A Historical Overview of Research on *Babesia orientalis*, a protozoan parasite infecting water Buffalo. Front. Microbiol. 8.

Huson, D.H., Bryant, D., 2006. Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. 23, 254–267.

Jabbar, A., Abbas, T., Sandhu, Z.U., Saddiqi, H.A., Qamar, M.F., Gasser, R.B., 2015. Tick-borne diseases of bovines in Pakistan: major scope for future research and improved control. Parasit. Vectors 8, 283.

Khan, M.K., He, L., Hussain, A., Azam, S., Zhang, W.J., Wang, L.X., Zhang, Q.L., Hu, M., Zhou, Y.Q., Zhao, J., 2013. Molecular epidemiology of *Theileria annulata* and identification of 18S rRNA gene and ITS regions sequences variants in apparently healthy buffaloes and cattle in Pakistan. Infect. Genet. Evol. 13, 124–132.

Kozich, J.J., Westcott, S.L., Baxter, N.T., Highlander, S.K., Schloss, P.D., 2013. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the miseq illumina sequencing platform. Appl. Environ. Microbiol. 79, 5112–5120.

Kundave, V.R., Ram, H., Banerjee, P.S., Garg, R., Mahendran, K., Ravikumar, G., Tiwari, A.K., 2018. Development of multiplex PCR assay for concurrent detection of tick borne haemoparasitic infections in bovines. Acta Parasitol. 63, 759–765.

Lee, P.C., Chong, E.T., Anderios, F., Al Lim, Y., Chew, C.H., Chua, K.H., 2015. Molecular detection of human *Plasmodium* species in Sabah using PlasmoNex multiplex PCR and hydrolysis probes real-time PCR. Malar. J. 14, 015–0542.

Lefterova, M.I., Budvytiene, I., Sandlund, J., Farnert, A., Banaei, N., 2015. Simple Real-Time PCR and amplicon sequencing method for identification of *Plasmodium* Species in human whole blood. J. Clin. Microbiol. 53, 2251–2257.

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., Ratabane, J., Pule, B., Latif, A.A., 2016. 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. Ticks Tick Borne Dis. 7, 869–879.

Mens, P.F., Schoone, G.J., Kager, P.A., Schallig, H.D., 2006. Detection and identification of human *Plasmodium* species with real-time quantitative nucleic acid sequence-based amplification. Malar. J. 5, 1475–2875.

Nayel, M., El-Dakhly, K.M., Aboulaila, M., Elsify, A., Hassan, H., Ibrahim, E., Salama, A., Yanai, T., 2012. The use of different diagnostic tools for *Babesia* and *Theileria* parasites in cattle in Menofia, Egypt. Parasitol. Res. 111, 1019–1024.

Nene, V., Kiara, H., Lacasta, A., Pelle, R., Svitek, N., Steinaa, L., 2016. The biology of *Theileria parva* and control of east coast fever – current status and future trends. Ticks Tick Borne Dis. 7, 549–564.

Passos, L.M., Bell-Sakyi, L., Brown, C.G., 1998. Immunochemical characterization of in vitro culture-derived antigens of *Babesia bovis* and *Babesia bigemina*. Vet. Parasitol. 76, 239–249.

Posada, D., 2008. JModelTest: phylogenetic model averaging. Mol. Biol. Evol. 25, 1253–1256.

Qayyum, Farooq, U., Samad, H.A., HR., C, 2010. Prevalence, clinicotherapeutic and prophylactic studies on Theileriosis in district Sahiwal (Pakistan). Anim. Plant. Sci. 20, 266–270.

Rogers, G.B., Bruce, K.D., 2010. Next-generation sequencing in the analysis of human microbiota: essential considerations for clinical application. Mol. Diagn. Ther. 14, 343–350.

Ros-Garcia, A., M'Ghirbi, Y., Bouattour, A., Hurtado, A., 2011. First detection of *Babesia occultans* in *Hyalomma* ticks from Tunisia. Parasitology 138, 578–582.

Rougemont, M., Van Saanen, M., Sahli, R., Hinrikson, H.P., Bille, J., Jaton, K., 2004. Detection of four *Plasmodium* species in blood from humans by 18S rRNA gene subunit-based and species-specific real-time PCR assays. J. Clin. Microbiol. 42, 5636–5643.

Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., Lesniewski, R.A., Oakley, B.B., Parks, D.H., Robinson, C.J., Sahl, J.W., Stres, B., Thallinger, G.G., Van Horn, D.J., Weber, C.F., 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl. Environ. Microbiol. 75, 7537–7541.

Shahnawaz, S., Ali, M., Aslam, M.A., Fatima, R., Chaudhry, Z.I., Hassan, M.U., Ali, M., Iqbal, F., 2011. A study on the prevalence of a tick-transmitted pathogen, *Theileria annulata,* and hematological profile of cattle from Southern Punjab (Pakistan). Parasitol. Res. 109, 1155–1160.

Sivakumar, T., Hayashida, K., Sugimoto, C., Yokoyama, N., 2014. Evolution and genetic diversity of *Theileria*. Infect. Genet. Evol. 27, 250–263.

Sivakumar, T., Igarashi, I., Yokoyama, N., 2016. Babesia ovata: taxonomy, phylogeny and epidemiology. Vet. Parasitol. 229, 99–106.

Steenkeste, N., Incardona, S., Chy, S., Duval, L., Ekala, M.T., Lim, P., Hewitt, S., Sochantha, T., Socheat, D., Rogier, C., Mercereau-Puijalon, O., Fandeur, T., Ariey, F., 2009. Towards high-throughput molecular detection of *Plasmodium*: new approaches and molecular markers. Malar. J. 8, 1475–2875.