THE UNIVERSITY of EDINBURGH

Edinburgh Research Explorer

A novel metabarcoded deep amplicon sequencing tool for disease surveillance and determining the species composition of Trypanosoma in cattle and other farm animals

Citation for published version:
Yasein, G, Zahid, O, Minter, E, Ashraf, K, Rashid, I, Shabbir, MZ, Betson, M, Sargison, N & Chaudhry, U 2022, 'A novel metabarcoded deep amplicon sequencing tool for disease surveillance and determining the species composition of Trypanosoma in cattle and other farm animals', *Acta Tropica*, vol. 230, 106416. https://doi.org/10.1016/j.actatropica.2022.106416

Digital Object Identifier (DOI):
10.1016/j.actatropica.2022.106416

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
Acta Tropica

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 26. Mar. 2023

A novel metabarcoded deep amplicon sequencing tool for disease surveillance and determining the species composition of *Trypanosoma* in cattle and other farm animals

Ghulam Yasein[1], Osama Zahid[2], Emily Minter[2], Kamran Ashraf[1], Imran Rashid[1], Muhammad Zubair Shabbir[1], Martha Betson[3], Neil D. Sargison[2], Umer Chaudhry[2,3]*

[1] University of Veterinary and Animal Sciences Lahore, Pakistan
[2] Royal (Dick) School of Veterinary Studies, University of Edinburgh, UK
[3] School of Veterinary Medicine, University of Surrey, UK

*Corresponding author:
Umer Chaudhry, Department of Veterinary Epidemiology and Public Health, School of Veterinary Medicine, University of Surrey, UK.
Email: u.chaudhry@surrey.ac.uk

49 **Abstract**

50 The World Health Organization (WHO) and the Food and Agriculture Organization (FAO) have
51 developed strategies to control trypanosomiasis in humans and livestock in endemic areas. These
52 require a better understanding of the distribution of different *Trypanosoma* species and improved
53 predictions of where they might appear in the future, based on accurate diagnosis and robust
54 surveillance systems. Here, we describe a metabarcoding deep amplicon sequencing method to identify
55 and determine the *Trypanosoma* species in co-infecting communities. First, four morphological
56 verified *Trypanosoma* species (*T. brucei, T. congolense, T. vivax* and *T. theileri*) were used to prepare
57 test DNA pools derived from different numbers of parasites to evaluate the method's detection
58 threshold for each of the four species and to assess the accuracy of their proportional quantification.
59 Having demonstrated the accurate determination of species composition in *Trypanosoma*
60 communities, the method was applied to determine its detection threshold using blood samples
61 collected from cattle with confirmed *Trypanosoma* infections based on a PCR assay. Each sample
62 showed a different *Trypanosoma* species composition based on the proportion of MiSeq reads. Finally,
63 we applied the assay to field samples to develop new insight into the species composition of
64 *Trypanosoma* communities in cattle, camels, buffalo, horses, sheep, and goat in endemically infected
65 regions of Pakistan. We confirmed that *Trypanosoma evansi* is the major species in Pakistan and for
66 the first time showed the presence of *Trypanosoma theileri*. The metabarcoding deep amplicon
67 sequencing method and bioinformatics pathway have several potential applications in animal and
68 human research, including evaluation of drug treatment responses, understanding of the emergence
69 and spread of drug resistance, and description of species interactions during co-infections and
70 determination of host and geographic distribution of trypanosomiasis in humans and livestock.

71

72 **Keywords:** Trypanosomiasis*,* *Trypanosoma*, metabarcoding, deep amplicon sequencing.

73

74

75

76

77

78

79

80

81

82

83 **1. Introduction**

84 The genus *Trypanosoma* is a diverse group of haemoprotozoan parasites that belongs to the family
85 Trypanosomatidae (Levine, 1973). Several *Trypanosoma* species are the causative agents of animal
86 and human trypanosomiasis (Cayla et al., 2019). The most important species belongs to the salivarian
87 group are *Trypanosoma brucei, Trypanosoma evensi*, *Trypanosoma congolense, Trypanosoma simiae*
88 and *Trypanosoma vivax*. The other most important species belongs to the stercoraria group are
89 *Trypanosoma rangeli*, *Trypanosoma cruzi, Trypanosoma musculi, Trypanosoma theleri* and
90 *Trypanosoma lewisi*. These species are found in a wide range of mammalian hosts and are transmitted
91 by tsetse flies (*Glossina* spp.), horse flies (*Tabanids*), stable flies (*Stomoxys* spp.) and kissing bugs
92 (*Triatoma* spp.) (Desquesnes et al., 2013; Ventura et al., 2002).

93 *Trypanosoma congolense, T. vivax, T. simiae* and *T. brucei* are highly pathogenic parasites of
94 livestock that cause the disease known as nagana, widely distributed in the African subcontinent
95 (Desquesnes et al., 2013; Kabiri et al., 1999). *Tyrpansoma evansi* is a highly pathogenic parasite of
96 ruminants, camels, and horses that causes the disease called Surra, widespread outside the tsetse belt
97 including in North Africa, the Middle East and South Asia. It has been exported through animal
98 movement to Latin America and even to Australia and Europe (Desquesnes et al., 2013). The genetic
99 characteristics of *T. evansi* are tightly clustered with *T. brucei*, providing strong support with the
100 notation that *T. evansi* may be a subspecies of *T. brucei* (Molinari and Moreno, 2018; Wen et al.,
101 2016). Similarly, *T. theileri* has been identified in Europe, Africa, Asia, and America. The infection
102 produces transient parasitaemia but is generally considered to be a non-pathogenic parasite of
103 ruminants. However, under stress conditions, it may cause abortion and even death (Calzolari et al.,
104 2018; Schlafer, 1979).

105 Despite distinct differences between factors that influence the rate and success of completion of the
106 life cycles of different *Trypanosoma* species, there is considerable overlap between the vectors,
107 biological niches and host ranges (Peacock et al., 2012; Weber et al., 2019). The similarities in these
108 characteristics between species imply that the emergence of one *Trypanosoma* species in an area could
109 be indicative of the risk of emergence of other species within the same area. Therefore, surveillance of
110 different *Trypanosoma* species is crucial for the successful control of trypanosomiasis (Hamill et al.,
111 2017; Lambin et al., 2010), by providing insights into the host distribution, co-infection, transmission
112 dynamics and the multiplicity of infection (Woolhouse et al., 2015).

113 Giemsa-stained blood smears are the standard method for the detection of *Trypanosoma* in the blood
114 samples of the infected animals. This method is generally useful in acutely infected animals, but
115 insensitive at the species level. The method is laborious and error-prone in inexperienced hands
116 (Kirchhoff et al., 1996). The antibody ELISA test is not reliable for differentiating between current

and past infections or assessing treatment outcomes (Hopkins et al., 1998). Similarly, the antigen ELISA test has an inadequate sensitivity due to cross-reactivity or non-specific immune responses (Eisler et al., 1998). The sensitivity and specificity of the molecular diagnostic tests for trypanosomiasis have been improved in recent years, but the reliability of infection reports vary considerably between affected areas because of the different PCR based methods used (Aregawi et al., 2019; Boulangé et al., 2017; Gaithuma et al., 2019; Isaac et al., 2017). Conventional PCR methods are useful for accurate detection of *Trypanosoma* but depend heavily on sensitive and specific primers and have limitations in terms of lacking scalability (Thumbi et al., 2008). There is a universal single test to detect all *Trypanosoma* species with equal reliability is needed to resolve these issues and improve surveillance systems.

A high throughput deep amplicon sequencing using the Illumina Mi-Seq platform is relatively low-cost and potentially less error-prone. The method has transformed the study of bacteria (microbiome) (Gloor et al., 2010; Rogers and Bruce, 2010), nematode (nemabiome) (Avramenko et al., 2015; Costa-Junior et al., 2021; Evans et al., 2021), trematode (tremobiome) (Rehman et al., 2021; Rehman et al., 2020; Sargison et al., 2019) parasites, and has the potential to open new areas of research to improve surveillance of separate *Trypanosoma* species as previously demonstrated by the concept of a 'haemoprotobiome' for the quantification of the bloodborne protozoan species (Chaudhry et al., 2019; Wahab et al., 2020). This approach targets genetic variations within defined regions of the haemoprotozoan genome to detect and quantify any species belonging to the parasite of interest. *Trypanosoma* 18S rDNA is a suitable genetic target due to the multiple genome-wide distributions and high copy number of species-specific variable sequences flanked by highly conserved sequences to enable universal primer binding and discrimination between species (Chaudhry et al., 2019; Dario et al., 2017; Wahab et al., 2020). The use of primers binding to conserved sites and analysis of up to 600 bp sequence reads allows *Trypanosoma* 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 tool suitable for high-throughput analysis. 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.

The aim of the present study was to describe the species composition of *Trypanosoma* communities present in blood samples collected from cattle and other farm animals. This involved designing metabarcoding Illumina Mi-Seq primers and developing an appropriate bioinformatic pathway to quantify the sequence reads generated for the different *Trypanosoma* species present. The validity of the approach was tested using "test pool" samples generated from known proportions of available positive control *T. brucei, T. congolense, T. vivax,* and *T.theileri* samples. The approach was then used

150 to analyse the composition of *Trypanosoma* species in field blood samples collected from seven
151 endemically infected regions of Pakistan.

152 **2. Methods and Materials**

153 *2.1. Positive control samples and genomic DNA isolation*

154 Four blood culture stocks of *Trypanosoma* isolates (*T. brucei, T. congolense, T. vivax* and *T.*
155 *theileri*) were provided by Professor Liam Morrison, Roslin Institute, University of Edinburgh. The
156 Lister427 stock of *T. brucei* was originally isolated from cattle in the Kiboko region of Kenya (Turner
157 et al., 1990). The IL3000 stock of *T. congolense* was originally isolated from cattle in the Transmara
158 region of Kenya (Wellde et al., 1974). The Y486 stock of *T. vivax* was originally isolated from
159 naturally infected cattle in the Zaria region of Nigeria (Leeflang et al., 1976). The origins of the *T.*
160 *theileri* stock are unknown. To extract the gDNA of *T. brucei,* and *T. congolense*, 50 µl of blood culture
161 stock from each of the species was transferred into a fresh 1.5 ml tube and centrifuged for 5 minutes,
162 before removing the supernatant and mixing with 25µl of lysis buffer (Viagen Biotech), Proteinase K
163 (New England BioLabs) and 50µl 1M dithiothreitol (DTT) (Chaudhry et al., 2019; Wahab et al., 2020).
164 *T. vivax* and *T. theileri* gDNA was extracted using the DNeasy Blood & Tissue Kits (Qiagen Co. Ltd
165 UK).

166 Genomic DNA pools of the four isolates (*T. brucei, T. vivax, T. congolense,* and *T. theileri*) were
167 created. Three replicates of the ten "test pools" were used to identify the detection threshold of the
168 deep amplicon sequencing method and to show the proportions of each of the species absent or present
169 in each "test pool". These were comprised of the following: [Mix1 (*T. brucei, T. vivax, T. congolense*),
170 Mix2 (*T. brucei, T. congolense*), Mix3 (*T. brucei, T. vivax*), Mix4 (*T. congolense, T. vivax*), Mix5 (*T*
171 *brucei*), Mix6 (*T. congolense*), Mix7 (*T. vivax*), Mix8 (*T. theileri, T. congolense, T. brucei*), Mix9 (*T.*
172 *theileri, T. brucei*), Mix10 (*T. theileri*)].

173 Additional 21 *Trypanosoma* samples provided by Professor Liam Morrison, Roslin Institute,
174 University of Edinburgh, were also used as a positive control initially collected from cattle in Tanzania.
175 Genomic DNA was extracted using protocols described in the DNeasy Blood & Tissue Kits using
176 100µl of blood from each sample (Qiagen Co. Ltd UK). The samples were selected *Trypanosoma*
177 positive or negative based on species-specific PCR (Cox et al., 2005) bands on the gel electrophoresis
178 (unpublished data).

179

180 *2.2. Study design, field sample collection and genomic DNA isolation*

181 A cross-sectional survey of *Trypanosoma* was conducted in seven known endemically infected
182 regions (Muzarffargarh, Lodhraan, Bahawalpur, D. G. Khan, Layyah, Rahim Yar Khan, Multan) of

the Punjab province of Pakistan during the peak trypanosomiasis transmission seasons in May to August 2018 to 2020. The study was approved by the Institutional Review Board of the University of Veterinary and Animal Sciences Punjab, Pakistan (UVAS-793-1). Discussions were held with key administrative and community leaders to raise awareness of the study, and encourage livestock farms to participate. Blood samples were collected from animals presenting at veterinary clinics from seven endemic regions. The procedures involved jugular venipuncture and withdrawal of 5 ml of intravenous blood into EDTA tubes, followed by storage at 20 C. A total of 288 blood samples from cattle, camels, buffalo, horses, sheep and goats were taken by trained para-veterinary workers under the supervision of local veterinary staff. Genomic DNA was extracted from field samples using protocols described in the TIANamp blood DNA kit (Beijing Biopeony Co. Ltd) using 50µl of blood from each sample (Chaudhry et al., 2019; Chaudhry et al., 2021).

2.3. PCR amplification and deep amplicon sequencing of Trypanosoma 18S rDNA

366 to 409 bp fragments encompassing the parts of 18S rDNA spanning the variable region of *Trypanosoma* species were targeted for PCR amplification. Adapters were added to each primer to allow subsequent annealing and N indicates the random number of nucleotides included between the adapter and primer set (Table 1). Equal proportions of *de novo* four forward (TRY- For1Ad, TRY-For1Ad1, TRY- For1Ad2, TRY- For1Ad3) and four reverse primers (TRY- Rev1Ad, TRY- Rev1Ad1, TRY- Rev1Ad2, TRY- Rev1Ad3) were mixed and used under the following conditions: 10 mM dNTPs, 0.5 U DNA polymerase, 5X buffer (KAPA Biosystems, USA) 10 µM forward and reverse adapter primer (Integrated DNA Technologies, IDT), and 12ng of positive control gDNA and 5 µl of field samples gDNA. The thermocycling conditions were 95ºC for 2 min, followed by 35 cycles of 98ºC for 20 s, 60ºC for 15 s, 72ºC for 15 s and a final extension of 72ºC for 5 min. The PCR products were purified with AMPure XP Magnetic Beads (1X) according to the manufacturer's instructions (Beckman Coulter) (Fig. 1A).

A barcoded primer set (Supplementary Table S1) was used in the second round of PCR amplification to add a fragment of unique sequence onto each purified product of the adapter PCR (Chaudhry et al., 2019; Chaudhry et al., 2021; Gibson, 2009; Hamilton et al., 2008; Hutchinson and Stevens, 2018; Wahab et al., 2020). Repetitions of the forward and reverse barcoded primers in different samples were avoided. The barcoded PCR was performed under the following conditions: 10 mM dNTPs, 0.5 U DNA polymerase, 5X buffer (KAPA Biosystems, USA), 10 µM barcoded forward (N501 to N517) and reverse (N701 to N724) primers (Integrated DNA Technologies, IDT) and 2 µl of adaptor PCR product as a DNA template. The thermocycling conditions were 98ºC for 45 s, followed by 7 cycles of 98ºC for 20 s, 63ºC for 20 s, and 72ºC for 2 min (Fig. 1A).

217 10 μl of each barcoded PCR product was combined to make a pooled library and separated by
218 agarose gel electrophoresis. The products were excised from the gel using commercial kits (QIAquick
219 Gel Extraction Kit, Qiagen, Germany) and 20 μl of eluted DNA was then purified using AMPure XP
220 Magnetic Beads (1X) (Beckman Coulter, Inc.), to form single purified DNA pooled library. The library
221 was measured with the KAPA qPCR library quantification kit (KAPA Biosystems, USA) and run on
222 an Illumina MiSeq Sequencer using a 600-cycle pair-end reagent kit (MiSeq Reagent Kits v2, MS-
223 103-2003) at a concentration of 15nM with the addition of 15% Phix Control v3 (Illumina, FC-11-
224 2003) (Chaudhry et al., 2019; Chaudhry et al., 2021; Wahab et al., 2020) (Fig. 1A).
225
226 *2.4. Bioinformatics pipeline for Trypanosoma 18S rDNA data analysis*
227 The Mi-Seq separates all sequencing data by the sample quality during post-run processing using
228 the barcoded indices to generate FASTQ files. The FASTQ files analysis was performed with a
229 bespoke pipeline using Mothur v1.39.5 software with modifications in the standard operating
230 procedures of Illumina Mi-Seq in the Command Prompt pipeline (Chaudhry et al., 2019; Costa-Junior
231 et al., 2021; Evans et al., 2021; Wahab et al., 2020). The complete Mothur script is included in Figure
232 1B and is freely available through the Mendeley database at DOI: 10.17632/d53sjjrwyn.3. Briefly, the
233 make.contigs command was used to read the forward and reverse FASTQ files and join to make a
234 contig (a pair of DNA sequences that overlap the region of *Trypanosoma* 18S rDNA locus). The reads
235 were screened using screen.seqs to ensure the sequence length ranged from 366 to 409 bp. Any reads
236 longer than the stated range or ambiguous were likely to exceed the expected length of the targeted
237 region and thus removed. The align.seqs command was used to align the reads against the closest
238 template sequence from the *Trypanosoma* 18S rDNA reference sequence taxonomy library (Mendeley
239 database at DOI: 10.17632/d53sjjrwyn.3). Sequence data that did not align with the 18S rDNA
240 reference sequence taxonomy library were discarded as non-specific amplification. Finally, the reads
241 were classified into *Trypanosoma* species using classify.seqs, which assigned the reads to their
242 corresponding taxonomy group and the 'summary.tax' command was used to generate the resultant
243 taxonomy file (freely available through the Mendeley database at DOI: 10.17632/d53sjjrwyn.3). In
244 order to reduce the impact of potential PCR or sequencing errors, reads fewer than 2000 across all
245 samples were removed. Individual samples with fewer than 500 reads across all taxonomic levels were
246 also removed. All FASTQ files have been made freely available through the Mendeley database at
247 DOI: 10.17632/tpcjkkgc2k.1. Overall, millions of 18S rDNA reads were generated from the data set
248 of four *Trypanosoma* positive control isolates and the *Trypanosoma*-positive field isolates.
249
250 *2.5. Statistical analysis of Trypanosoma 18S rDNA sequence reads*

251 The data of *Trypanosoma* positive control and field samples were used to analyse the accuracy of
252 *Trypanosoma* species detection using CompareTests: Correct for Verification Bias in Diagnostic
253 Accuracy and Agreement Statistics software (R package version 1.2.). The effect of the *Trypanosoma*
254 positive control “test pools” was analysed by running a Kruskal-Wallis rank-sum test for each
255 admixture. Any results with a p-value ≤0.05 were considered significant. The species composition of
256 *Trypanosoma* isolates in the field samples was calculated by dividing the number of sequence reads
257 for each species by the total number of reads in each sample.
258
259 *2.6. Phylogenetic analysis of Trypanosoma 18S rDNA haplotypes*
260 The consensus sequence library was developed using a filtered sequence reads generated from the
261 four *Trypanosoma* positive control isolates (*T. brucei, T. vivax, T. congolense,* and *T. theileri*) and 95
262 sequences of *T. brucei, T. evansi, T. congolense, T. vivax, T. simiae and T. theileri* extracted from the
263 NCBI GenBank to account for any additional genetic variations. The obtained 18S rDNA filtered
264 sequences were first aligned using the MUSCLE alignment tool of Geneious v10.2.5 (Biomatters Ltd,
265 New Zealand) and then imported into the FaBox v1.5 online software to calculate the number of
266 “consensus” sequences generated from each species (for more detail see Mendeley database at DOI:
267 10.17632/d53sjjrwyn.3). A phylogenetic tree of the 18S rDNA consensus sequences was constructed
268 by gamma-distributed Kimura 2-parameter (K2+G) model using the Maximum Likelihood method in
269 the MEGA X software, with a bootstrap value of 1000 (Kumar et al., 2018; Stecher et al., 2020).
270
271

272 **3. Results**
273 *3.1. Consensus sequence library preparation and the assessment of Trypanosoma 18S rDNA genetic*
274 *variations*
275 In total, 44 18S rDNA consensus sequences were identified among six *Trypanosoma* species (*T.*
276 *brucei=15, T. vivax=3, T. congolense=10, T. theileri=7, T. evansi=6, T. simiae=3*) (Fig. 2).
277 Comparison of genetic differentiation between 18S rDNA haplotypes revealed 83% to 98% identity
278 between the *Trypanosoma* species (Supplementary Table S2). The most closely related species of
279 *Trypanosoma* (*T. brucei* and *T. evansi*) could still be reliably differentiated by virtue of 18S rDNA
280 sequence variations (98%) (Supplementary Table S3, Supplementary Figure S2). For the remaining
281 species, there was a higher level of genetic variations between the 18S rDNA variants. A phylogenetic
282 tree of the 44 consensus sequences demonstrates a distinct clustering between species, apart from *T.*
283 *evansi* and *T. brucei* (Fig. 2). The clustering was supported by high bootstrap confidence intervals,

284 ranging from 72-95. A well-supported node is generally considered to have a bootstrap value greater
285 than 70, therefore the distinct clustering was well supported for each species.
286
287 *3.2. Validation of the deep amplicon sequencing assay using test pools of T. brucei, T. vivax, T.*
288 *congolense and T. theileri*
289 Three replicates of each known species were created from gDNA of four *Trypanosoma* isolates, to
290 validate the deep amplicon sequencing assay (Fig. 3). The mixing of different test positive controls
291 demonstrates the detection accuracy of the method and shows the different species compositions
292 present (Fig. 3). Overall, there was no statistically significant difference between the proportions of
293 the species detected in any of the replicates for the ten "test pools" (Kruskal-Wallis rank-sum test; Mix
294 1: $\chi^2(1)$ 0.021084, p=0.8845, Mix 2: $\chi^2(1)$ 0, p=1, Mix 3: $\chi^2(1)$ 0, p=1, Mix 4: $\chi^2(1)$ 0, p=1, Mix 5:
295 $\chi^2(1)$ 0, p=1, Mix 6: $\chi^2(1)$ 0, p=1, Mix 7: $\chi^2(1)$ 0, p=1, Mix 8: $\chi^2(1)$ 0.031818, p=0.8584, Mix 9: $\chi^2(1)$
296 0.21084, p=0.8845, Mix 10: $\chi^2(1)$ 0, p=1), indicating repeatability of the method. The Mix1 pool
297 contained *T. brucei, T. congolense*, *T. vivax* were perfectly matched (Fig. 3). The results of Mix 2, Mix
298 3, Mix 4 and Mix 9 were accurate with no significant difference between replicates (Mix 2 contained
299 *T. brucei, T. congolense*, Mix 3 contained *T. brucei, T. vivax*, Mix 4 contained *T. congolense, T. vivax*
300 and Mix 9 contained *T. theileri, T. brucei*) (Fig. 3). The results of Mix 5, Mix 6, Mix 7, and Mix 10
301 are also consistent for the pools of 100% of each species (Mix 5 contained *T. brucei*, Mix 6 contained
302 *T. congolense*, Mix 7 contained *T. vivax*, Mix 10 contained *T. theileri*) (Fig. 3).
303
304 *3.3. Validation of the deep amplicon sequencing assay for the detection of Trypanosoma infected*
305 *samples*
306 To further investigate the validity of the assay, additional 21 *Trypanosoma* positive controls
307 containing various species were analysed. Different proportions of reads aligning with *T. brucei, T.*
308 *congolense, T. theileri,* and *T. vivax* were present in each sample except *T. evansi* (Fig. 4). On the
309 analysis of each sample separately, Tryp6 was co-infected with *T. brucei, T. congolense, T. theileri,*
310 and *T. vivax*, Tryp1 and 5 were co-infected with *T. congolense, T. theileri,* and *T. vivax*, Tryp17 was
311 co-infected with *T. brucei, T. congolense,* and *T. theileri,* Tryp2, 18 and 21 were co-infected with *T.*
312 *brucei,* and *T. congolense*, Tryp1 was co-infected with *T. congolense* and *T. vivax,* Tryp3, 7, 10, 15,
313 19 and 20 were co-infected with *T. congolense, T. theileri* (Fig. 4). Single species infection of *T.*
314 *theileri* or *T. congolense* was identified in Tryp4, 8, 9, 11, 13, 14, 16 (Fig. 4). On the analysis of each
315 species separately, 17/21 samples were infected with *T. theileri,* followed by 15/21 samples were

316 infected with *T. congolense* making these the most common species. 5/21 and 4/21 samples were
317 infected with *T. brucei* and *T. vivax*, respectively.
318
319 *3.4. Species prevalence of Trypanosoma in field samples*
320 Blood samples from various age groups and animals of both sexes were collected from the
321 veterinary clinics were potentially considered positive for trypanosomiasis based on clinical signs.
322 Each sample was routinely analysed by microscopic examination under oil immersion (x100) of 10%
323 Giemsa-stained blood smears for the diagnosis of *Trypanosoma* infection. A total of 288 blood samples
324 were accepted positive for *Trypanosoma* genus based microscopic identification (Supplementary Fig.
325 S1). Deep amplicon sequencing was then applied to the 288 blood samples to detect the presence of
326 *Trypanosoma* species. 155 out of 288 samples (98 cattle, 21 camels, 19 buffalo, 10 horses, 5 sheep, 2
327 goats) were positive with *Trypanosoma* infections (Fig. 5, Supplementary Table S4).
328 The amplicon sequencing further demonstrated the presence of the *T. evansi* and *T. theileri* species,
329 but none of the samples contained *T. brucei*. The occurrence of *T. evansi* is higher than that of *T.*
330 *theileri* in cattle, camel and buffalo hosts in each of the seven endemic regions. From 98 cattle, 61
331 (66.24%) samples were harboured infections due to *T. evansi*, 26 (26.50%) due to *T. theileri* and 11
332 (11.22%) were co-infected by these two *Trypanosoma* species. From 21 camels, 19 (90.5%) samples
333 were *T. evansi* positive, 2 (9.5%) samples were *T. theileri* positive and no co-infections were identified.
334 From 19 buffalo, 16 (84.2%) samples were *T. evansi* positive, 2 (10.5%) samples were *T. theileri*
335 positive, and 1 (5.26%) sample was co-infected. *Trypanosoma evansi* was identified in 10 horses, 5
336 sheep and 2 goat samples, but neither *T. theileri* nor co-infections were confirmed in these host species
337 (Fig. 5).
338
339 **4. Discussion**
340 We have evaluated the high throughput amplicon sequencing method to identify the presence of
341 *Trypanosoma* species by (a) preparing a consensus sequence library of Trypanosoma 18S rDNA
342 genetic variations, (b) validating the assay using “test pools” of each *Trypanosoma* species, (c)
343 validating the assay for the detection of *Trypanosoma* infected samples, and (d) applying the method
344 to quantify the species present in *Trypanosoma*-positive blood samples from the field.
345 The 18S rDNA consensus sequences of six *Trypanosoma* species showed a high level of genetic
346 differentiation between species, except for between *T. brucei* and *T. evansi* in the present study. A
347 phylogenetic tree showed distinct clustering between species, apart from between *T. evansi* and *T.*
348 *brucei,* demonstrating the limitations for the diagnosis of closely related species, thus the need for new
349 genetic markers. Previous studies indicate that genetic and morphological characteristics of *T. evansi*

350 are tightly clustered with *T. brucei*, providing strong support that *T. evansi* may be a subspecies of *T.*
351 *brucei* (Molinari and Moreno, 2018; Wen et al., 2016). The morphology of *T. evansi* and *T. brucei* are
352 very similar, except *T. brucei* undergoes differentiation into stumpy bloodstream forms (Ou et al.,
353 1991). Moreover, it was demonstrated that maxicircle kDNA, microsatellite and random amplified
354 polymorphic markers are reliable tools use to differentiate *T. evansi* and *T. brucei* in natural field
355 conditions. There are few reports demonstrating the cocirculating of both species in the endemic
356 regions of Kenya and Tanzania (Hide et al., 1990; Kamidi et al., 2017; Kimenyi et al., 2022; Lun et
357 al., 1992; Lun et al., 2004).

358 We tested the capability of the amplicon sequencing assay to accurately determine the relative
359 species proportions in various combinations of four *Trypanosoma* species (*T. brucei, T. vivax, T.*
360 *congolense and T. theileri*). We generated the "test pools" containing different estimated proportions
361 of *Trypanosoma* and showed no significant difference between the proportions of the species detected
362 in each of the pairs of replicates for the ten "test pools". A previous study using "test pools" of
363 *Plasmodium* (Wahab et al., 2020), *Babesia* and *Theileria* (Chaudhry et al., 2019) haemoprotozoan
364 parasites showed that the relative sequence representation was unaffected by either the parasite species
365 composition of the sample or the number of PCR cycles employed (Chaudhry et al., 2019; Wahab et
366 al., 2020). No sequence representation bias was identified in PCR products used for high throughput
367 runs, arising from the number of first-round PCR cycles (Chaudhry et al., 2019). Having validated the
368 amplicon sequencing using "test pools" of *Trypanosoma* isolates, we applied the method to positive
369 control samples containing various *Trypanosoma* species collected from cattle and identified them as
370 *Trypanosoma* co-infections using PCR based assay (Cox et al., 2005). Trypanosomiasis caused by *T.*
371 *congolense* and *T. theileri* were the most common infections seen amongst these samples, followed by
372 *T. brucei* and *T. vivax* in the present study.

373 After validating the amplicon sequencing using "test pools" and positive control samples of
374 *Trypanosoma* isolates, we compared the assay with the microscopic method on the field samples
375 collected from seven endemic regions of the Punjab province of Pakistan. The microscopic analysis
376 showed that 288 blood samples were positive for *Trypanosoma* genus. In contrast, 155 out of 288
377 samples were shown positive with *Trypanosoma* infections in the amplicon sequencing analysis. The
378 microscopic method is generally insensitive at low parasitemia levels, more error-prone in
379 inexperienced hands and poor resource settings (Biéler et al., 2012), while high throughput sequencing
380 is more precise and relatively less error-prone (Chaudhry et al., 2019; Wahab et al., 2020). Amplicon
381 sequencing assays could, therefore, be helpful in epidemiological surveys to determine *Trypanosoma*
382 infection burdens in populations having low levels of parasitemia.

383 The amplicon sequencing assay demonstrated the presence of *T. evansi* and *T. theileri* in the field
384 samples, with the occurrence of *T. evansi* being higher than that of *T. theileri* in cattle, camels and
385 buffalo; and the presence of *T. evansi* (single species infection) being highest in horses, sheep and
386 goats. Our findings are consistent with previous reports, confirming that *T. evansi* is the major
387 infectious disease of animals in Pakistan (Rashid et al., 2017; Shahzad et al., 2012; Tehseen et al.,
388 2015; Ul Hasan et al., 2006). In this study, *T. theileri* was observed in cattle, camel and buffalo but
389 absent in horses, sheep and goats. To our knowledge, *T. theileri* has not been reported in Pakistan
390 before (Calzolari et al., 2018; Garcia et al., 2011; Hoyte, 1972; Schlafer, 1979). Whilst it is plausible
391 that *T. theileri* has recently emerged in Pakistan with the migration of livestock from south-east Asia
392 (Fernandez et al., 2010), or north Africa (Ngomtcho et al., 2017), and that the primary vectors are
393 Tabanid flies, is well established in Pakistan to maintain the infection in cattle and buffalo (Tehseen
394 et al., 2015). Previous studies have noted that the parasitaemia level of *T. theileri* is often low and
395 would, therefore, require a highly sensitive test to detect its presence (Kelly et al., 2017). The
396 diagnostic tests used in Pakistan may have low sensitivity, and therefore, would have been unlikely to
397 detect the presence of *T. theileri* at low parasitaemia in the previous studies (Chappuis et al., 2005).
398 There are obvious challenges affecting the ability to develop *Trypanosoma* surveillance
399 programmes and improve understanding of *Trypanosoma* epidemiology, especially in endemic areas.
400 The development of sensitive and specific molecular tools play an important role in the progression
401 towards overcoming these challenges. The epidemiology of trypanosomiasis is influenced by fly
402 infestation, geographical region, climatic conditions, co-grazing, host movement and time of the year.
403 Therefore, improvements in existing molecular tools would provide a better assessment of the
404 emergence or re-emergence of *Trypanosoma* infections. This would inform the correct administration
405 of trypanocidal drugs and might limit treatment failures (Hamill et al., 2017; Mossaad et al., 2017). In
406 addition, improvements in the molecular tools will enhance understanding of *Trypanosoma* co-
407 infection (Rodrigues et al., 2019).
408
409 **5. Conclusion**
410 We used the metabarcoding deep amplicon sequencing using an Illumina Mi-Seq platform to
411 reliably describe the *Trypanosoma* species present in blood samples collected from cattle, buffalo,
412 camels, horses, sheep and goats. Our results are the proof of concept for the use of this method in
413 disease surveillance programmes in cattle and farm animals their vectors in the resource-poor setting
414 of the endemic regions with the increasing capacity of the highly equipped laboratories. The plasticity
415 of this approach is not host specific and could have future application for the assessment of other

animals and human *Trypanosoma* species. This work will explore the possibilities for the application of a high throughput practical method to determine the disease epidemiology including *Trypanosoma* co-infections and sub-typing schemes. The accurate description of this technology has practical applications in monitoring changes in parasite diversity after the emergence and spread of drug resistance and the understanding of the multiplicity of infection described in other protozoan parasites (Chaudhry et al., 2021).

Acknowledgement

The Roslin Institute uses facilities funded by Biotechnology and Biological Sciences Research Council (BBSRC), UK. 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 to thank Professor Liam Morrison (Roslin Institute, University of Edinburgh, UK) for his great support and for providing positive control samples.

Conflict of interest

None

References

Aregawi, W.G., Agga, G.E., Abdi, R.D., Büscher, P., 2019. Systematic review and meta-analysis on the global distribution, host range, and prevalence of *Trypanosoma evansi*. Parasit Vectors 12, 67.

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.

Biéler, S., Matovu, E., Mitashi, P., Ssewannyana, E., Bi Shamamba, S.K., Bessell, P.R., Ndung'u, J.M., 2012. Improved detection of *Trypanosoma brucei* by lysis of red blood cells, concentration and LED fluorescence microscopy. Acta Trop 121, 135-140.

Boulangé, A., Pillay, D., Chevtzoff, C., Biteau, N., Comé de Graça, V., Rempeters, L., Theodoridis, D., Baltz, T., 2017. Development of a rapid antibody test for point-of-care diagnosis of animal African trypanosomosis. Vet Parasitol 233, 32-38.

Calzolari, M., Rugna, G., Clementi, E., Carra, E., Pinna, M., Bergamini, F., Fabbi, M., Dottori, M., Sacchi, L., Votýpka, J., 2018. Isolation of a Trypanosome Related to Trypanosoma theileri (Kinetoplastea: Trypanosomatidae) from Phlebotomus perfiliewi (Diptera: Psychodidae). Biomed Res Int 2018, 2597074.

Cayla, M., Rojas, F., Silvester, E., Venter, F., Matthews, K.R., 2019. African trypanosomes. Parasit Vectors 12, 190.

Chappuis, F., Loutan, L., Simarro, P., Lejon, V., Büscher, P., 2005. Options for field diagnosis of human african trypanosomiasis. Clin Microbiol Rev 18, 133-146.

Chaudhry, U., Ali, Q., Rashid, I., Shabbir, M.Z., Ijaz, M., Abbas, M., Evans, M., Ashraf, K., Morrison, I., Morrison, L., Sargison, N.D., 2019. Development of a deep amplicon sequencing method to determine the species composition of piroplasm haemoprotozoa. Ticks Tick Borne Dis 10, 101276.

Chaudhry, U., Ali, Q., Zheng, L., Rashid, I., Shabbir, M.Z., Numan, M., Ashraf, K., Evans, M., Rafiq, S., Oneeb, M., Morrison, L.J., Ivan Morrison, W., Sargison, N.D., 2021. Contrasting population genetics of co-endemic cattle- and buffalo- derived *Theileria annulata*. Ticks Tick Borne Dis 12, 101595.

460 Costa-Junior, L.M., Chaudhry, U.N., Silva, C.R., Sousa, D.M., Silva, N.C., Cutrim-Júnior, J.A.A., Brito,
461 D.R.B., Sargison, N.D., 2021. Nemabiome metabarcoding reveals differences between gastrointestinal
462 nematode species infecting co-grazed sheep and goats. Vet Parasitol 289, 109339.
463 Cox, A., Tilley, A., McOdimba, F., Fyfe, J., Eisler, M., Hide, G., Welburn, S., 2005. A PCR based assay for
464 detection and differentiation of African trypanosome species in blood. Exp Parasitol 111, 24-29.
465 Dario, M.A., Moratelli, R., Schwabl, P., Jansen, A.M., Llewellyn, M.S., 2017. Small subunit ribosomal
466 metabarcoding reveals extraordinary trypanosomatid diversity in Brazilian bats. PLoS neglected
467 tropical diseases 11, e0005790.
468 Desquesnes, M., Holzmuller, P., Lai, D.H., Dargantes, A., Lun, Z.R., Jittaplapong, S., 2013. *Trypanosoma*
469 *evansi* and surra: a review and perspectives on origin, history, distribution, taxonomy, morphology,
470 hosts, and pathogenic effects. Biomed Res Int 194176, 19.
471 Eisler, M.C., Lessard, P., Masake, R.A., Moloo, S.K., Peregrine, A.S., 1998. Sensitivity and specificity of
472 antigen-capture ELISAs for diagnosis of *Trypanosoma congolense* and *Trypanosoma vivax* infections
473 in cattle. Vet Parasitol 79, 187-201.
474 Evans, M.J., Chaudhry, U.N., Costa-Júnior, L.M., Hamer, K., Leeson, S.R., Sargison, N.D., 2021. A 4 year
475 observation of gastrointestinal nematode egg counts, nemabiomes and the benzimidazole resistance
476 genotypes of *Teladorsagia circumcincta* on a Scottish sheep farm. Int J Parasitol.
477 Fernandez, P., Baticados, A., Baticados, W., 2010. arasitological examination for *Trypanosoma theileri*
478 infection of cattle from Quirino Province, Philippines. . 2010 1:3-6.
479 Gaithuma, A.K., Yamagishi, J., Martinelli, A., Hayashida, K., Kawai, N., Marsela, M., Sugimoto, C., 2019. A
480 single test approach for accurate and sensitive detection and taxonomic characterization of
481 Trypanosomes by comprehensive analysis of internal transcribed spacer 1 amplicons. PLoS neglected
482 tropical diseases 13, e0006842.
483 Garcia, H.A., Kamyingkird, K., Rodrigues, A.C., Jittapalapong, S., Teixeira, M.M., Desquesnes, M., 2011. High
484 genetic diversity in field isolates of *Trypanosoma theileri* assessed by analysis of cathepsin L-like
485 sequences disclosed multiple and new genotypes infecting cattle in Thailand. Vet Parasitol 180, 363-
486 367.
487 Gibson, W., 2009. Species-specific probes for the identification of the African tsetse-transmitted trypanosomes.
488 Parasitology 136, 1501-1507.
489 Gloor, G.B., Hummelen, R., Macklaim, J.M., Dickson, R.J., Fernandes, A.D., MacPhee, R., Reid, G., 2010.
490 Microbiome profiling by illumina sequencing of combinatorial sequence-tagged PCR products. PLoS
491 One 5, 0015406.
492 Hamill, L., Picozzi, K., Fyfe, J., von Wissmann, B., Wastling, S., Wardrop, N., Selby, R., Acup, C.A., Bardosh,
493 K.L., Muhanguzi, D., Kabasa, J.D., Waiswa, C., Welburn, S.C., 2017. Evaluating the impact of
494 targeting livestock for the prevention of human and animal trypanosomiasis, at village level, in districts
495 newly affected with *T. b. rhodesiense* in Uganda. Infect Dis Poverty 6, 16.
496 Hamilton, P.B., Adams, E.R., Malele, II, Gibson, W.C., 2008. A novel, high-throughput technique for species
497 identification reveals a new species of tsetse-transmitted trypanosome related to the *Trypanosoma*
498 *brucei* subgenus, Trypanozoon. Infect Genet Evol 8, 26-33.
499 Hide, G., Cattand, P., LeRay, D., Barry, J.D., Tait, A., 1990. The identification of *Trypanosoma brucei*
500 subspecies using repetitive DNA sequences. Mol Biochem Parasitol 39, 213-225.
501 Hopkins, J.S., Chitambo, H., Machila, N., Luckins, A.G., Rae, P.F., van den Bossche, P., Eisler, M.C., 1998.
502 Adaptation and validation of antibody-ELISA using dried blood spots on filter paper for
503 epidemiological surveys of tsetse-transmitted trypanosomosis in cattle. Preventive veterinary medicine
504 37, 91-99.
505 Hoyte, H.M., 1972. The morphology of *Trypanosoma theileri* in the blood of cattle, and the rediscovery of
506 *Theileria mutans* in England. Z Parasitenkd 38, 183-199.
507 Hutchinson, R., Stevens, J.R., 2018. Barcoding in trypanosomes. Parasitology 145, 563-573.
508 Isaac, C., Ohiolei, J.A., Ebhodaghe, F., Igbinosa, I.B., Eze, A.A., 2017. Animal African Trypanosomiasis in
509 Nigeria: A long way from elimination/eradication. Acta Trop 176, 323-331.
510 Kabiri, M., Franco, J.R., Simarro, P.P., Ruiz, J.A., Sarsa, M., Steverding, D., 1999. Detection of *Trypanosoma*
511 *brucei gambiense* in sleeping sickness suspects by PCR amplification of expression-site-associated
512 genes 6 and 7. Trop Med Int Health 4, 658-661.

513 Kamidi, C.M., Saarman, N.P., Dion, K., Mireji, P.O., Ouma, C., Murilla, G., Aksoy, S., Schnaufer, A., Caccone,
514 A., 2017. Multiple evolutionary origins of *Trypanosoma evansi* in Kenya. PLoS neglected tropical
515 diseases 11, e0005895.
516 Kelly, S., Ivens, A., Mott, G.A., O'Neill, E., Emms, D., Macleod, O., Voorheis, P., Tyler, K., Clark, M.,
517 Matthews, J., Matthews, K., Carrington, M., 2017. An Alternative Strategy for Trypanosome Survival
518 in the Mammalian Bloodstream Revealed through Genome and Transcriptome Analysis of the
519 Ubiquitous Bovine Parasite Trypanosoma (Megatrypanum) theileri. Genome Biol Evol 9, 2093-2109.
520 Kimenyi, N.N., Kimenyi, K.M., Amugune, N.O., Getahun, M.N., 2022. Genetic connectivity of trypanosomes
521 between tsetse-infested and tsetse-free areas of Kenya. Parasitology 149, 285-297.
522 Kirchhoff, L.V., Votava, J.R., Ochs, D.E., Moser, D.R., 1996. Comparison of PCR and microscopic methods
523 for detecting *Trypanosoma cruzi*. J Clin Microbiol 34, 1171-1175.
524 Kozich, J.J., Westcott, S.L., Baxter, N.T., Highlander, S.K., Schloss, P.D., 2013. Development of a dual-index
525 sequencing strategy and curation pipeline for analyzing amplicon sequence data on the miseq illumina
526 sequencing platform. Applied and Environmental Microbiology.
527 Kumar, S., Stecher, G., Li, M., Knyaz, C., Tamura, K., 2018. MEGA X: Molecular Evolutionary Genetics
528 Analysis across Computing Platforms. Molecular biology and evolution 35, 1547-1549.
529 Lambin, E.F., Tran, A., Vanwambeke, S.O., Linard, C., Soti, V., 2010. Pathogenic landscapes: interactions
530 between land, people, disease vectors, and their animal hosts. Int J Health Geogr 9, 54.
531 Leeflang, P., Buys, J., Blotkamp, C., 1976. Studies on *Trypanosoma vivax*: infectivity and serial maintenance
532 of natural bovine isolates in mice. Int J Parasitol 6, 413-417.
533 Levine, N.D., 1973. Parasites: The Trypanosomes of Mammals. A Zoological Monograph. Cecil A. Hoare.
534 Blackwell, Oxford, England, 1972 (U.S. distributor, Davis, Philadelphia). xviii, 750 pp. + plates.
535 $34.50. Science 179, 60-60.
536 Lun, Z.R., Allingham, R., Brun, R., Lanham, S.M., 1992. The isoenzyme characteristics of *Trypanosoma evansi*
537 and *Trypanosoma equiperdum* isolated from domestic stocks in China. Ann Trop Med Parasitol 86,
538 333-340.
539 Lun, Z.R., Li, A.X., Chen, X.G., Lu, L.X., Zhu, X.Q., 2004. Molecular profiles of *Trypanosoma brucei*, *T.*
540 *evansi* and *T. equiperdum* stocks revealed by the random amplified polymorphic DNA method.
541 Parasitol Res 92, 335-340.
542 Molinari, J., Moreno, S.A., 2018. *Trypanosoma brucei* Plimmer & Bradford, 1899 is a synonym of T. evansi
543 (Steel, 1885) according to current knowledge and by application of nomenclature rules. Syst Parasitol
544 95, 249-256.
545 Mossaad, E., Salim, B., Suganuma, K., Musinguzi, P., Hassan, M.A., Elamin, E.A., Mohammed, G.E., Bakhiet,
546 A.O., Xuan, X., Satti, R.A., Inoue, N., 2017. *Trypanosoma vivax* is the second leading cause of camel
547 trypanosomosis in Sudan after *Trypanosoma evansi*. Parasit Vectors 10, 176.
548 Ngomtcho, S.C.H., Weber, J.S., Ngo Bum, E., Gbem, T.T., Kelm, S., Achukwi, M.D., 2017. Molecular
549 screening of tsetse flies and cattle reveal different *Trypanosoma* species including *T. grayi* and *T.*
550 *theileri* in northern Cameroon. Parasit Vectors 10, 631.
551 Ou, Y.C., Giroud, C., Baltz, T., 1991. Kinetoplast DNA analysis of four *Trypanosoma evansi* strains. Mol
552 Biochem Parasitol 46, 97-102.
553 Peacock, L., Cook, S., Ferris, V., Bailey, M., Gibson, W., 2012. The life cycle of *Trypanosoma (Nannomonas)*
554 *congolense* in the tsetse fly. Parasit Vectors 5, 109.
555 Rashid, I., Akbar, H., Gharbi, M., Riaz, F., Islam, S., Saleem, M.B., Shahzad, S., Shehzad, W., Rouatbi, M.,
556 Ashraf, K., 2017. First report of *Trypanosoma evansi* infection (surra) in a puma (Felis concolor) of
557 Lahore zoo, Pakistan. J Zoo Wildl Med 48, 918-921.
558 Rehman, Z.U., Martin, K., Zahid, O., Ali, Q., Rashid, I., Hafeez, M.A., Ahmad, N., Ashraf, K., Betson, M.,
559 Sargison, N.D., Chaudhry, U., 2021. High-throughput sequencing of *Fasciola* spp. shows co-infection
560 and intermediate forms in Balochistan, but only *Fasciola gigantica* in the Punjab province of Pakistan.
561 Infect Genet Evol 94, 105012.
562 Rehman, Z.U., Zahid, O., Rashid, I., Ali, Q., Akbar, M.H., Oneeb, M., Shehzad, W., Ashraf, K., Sargison, N.D.,
563 Chaudhry, U., 2020. Genetic diversity and multiplicity of infection in *Fasciola gigantica* isolates of
564 Pakistani livestock. Parasitol Int 76, 102071.
565 Rodrigues, C.M.F., Garcia, H.A., Sheferaw, D., Rodrigues, A.C., Pereira, C.L., Camargo, E.P., Teixeira,
566 M.M.G., 2019. Genetic diversity of trypanosomes pathogenic to livestock in tsetse flies from the Nech

Sar National Park in Ethiopia: A concern for tsetse suppressed area in Southern Rift Valley? Infect Genet Evol 69, 38-47.
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.
Sanches-Vaz, M., Temporão, A., Luis, R., Nunes-Cabaço, H., Mendes, A.M., Goellner, S., Carvalho, T., Figueiredo, L.M., Prudêncio, M., 2019. *Trypanosoma brucei* infection protects mice against malaria. PLoS Pathog 15, e1008145.
Sargison, N.D., Shahzad, K., Mazeri, S., Chaudhry, U., 2019. A high throughput deep amplicon sequencing method to show the emergence and spread of *Calicophoron daubneyi* rumen fluke infection in United Kingdom cattle herds. Vet Parasitol 268, 9-15.
Schlafer, D.H., 1979. *Trypanosoma theileri*: a literature review and report of incidence in New York cattle. Cornell Vet 69, 411-425.
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.
Shahzad, W., Mehmood, K., Munir, R., Aslam, W., Ijaz, M., Ahmad, R., Khan, M.S., Sabir, A.J., 2012. Prevalence and molecular diagnosis of *Fasciola hepatica* in sheep and goats in different districts of Punjab, Pakistan. Pak Vet J 32, 535-538.
Stecher, G., Tamura, K., Kumar, S., 2020. Molecular Evolutionary Genetics Analysis (MEGA) for macOS. Molecular biology and evolution 37, 1237-1239.
Tehseen, S., Jahan, N., Qamar, M.F., Desquesnes, M., Shahzad, M.I., Deborggraeve, S., Büscher, P., 2015. Parasitological, serological and molecular survey of *Trypanosoma evansi* infection in dromedary camels from Cholistan Desert, Pakistan. Parasit Vectors 8, 415.
Thumbi, S.M., McOdimba, F.A., Mosi, R.O., Jung'a, J.O., 2008. Comparative evaluation of three PCR base diagnostic assays for the detection of pathogenic trypanosomes in cattle blood. Parasit Vectors 1, 46.
Turner, C.M., Sternberg, J., Buchanan, N., Smith, E., Hide, G., Tait, A., 1990. Evidence that the mechanism of gene exchange in *Trypanosoma brucei* involves meiosis and syngamy. Parasitology 101 Pt 3, 377-386.
Ul Hasan, M., Muhammad, G., Gutierrez, C., Iqbal, Z., Shakoor, A., Jabbar, A., 2006. Prevalence of *Trypanosoma evansi* infection in equines and camels in the Punjab region, Pakistan. Ann N Y Acad Sci 1081, 322-324.
Ventura, R.M., Takeda, G.F., Silva, R.A., Nunes, V.L., Buck, G.A., Teixeira, M.M., 2002. Genetic relatedness among *Trypanosoma evansi* stocks by random amplification of polymorphic DNA and evaluation of a synapomorphic DNA fragment for species-specific diagnosis. Int J Parasitol 32, 53-63.
Wahab, A., Shaukat, A., Ali, Q., Hussain, M., Khan, T.A., Khan, M.A.U., Rashid, I., Saleem, M.A., Evans, M., Sargison, N.D., Chaudhry, U., 2020. A novel metabarcoded 18S ribosomal DNA sequencing tool for the detection of *Plasmodium* species in malaria positive patients. Infect Genet Evol 82, 104305.
Weber, J.S., Ngomtcho, S.C.H., Shaida, S.S., Chechet, G.D., Gbem, T.T., Nok, J.A., Mamman, M., Achukwi, D.M., Kelm, S., 2019. Genetic diversity of trypanosome species in tsetse flies (Glossina spp.) in Nigeria. Parasit Vectors 12, 481.
Wellde, B., Lötzsch, R., Deindl, G., Sadun, E., Williams, J., Warui, G., 1974. *Trypanosoma congolense*. I. Clinical observations of experimentally infected cattle. Exp Parasitol 36, 6-19.
Wen, Y.Z., Lun, Z.R., Zhu, X.Q., Hide, G., Lai, D.H., 2016. Further evidence from SSCP and ITS DNA sequencing support *Trypanosoma evansi* and *Trypanosoma equiperdum* as subspecies or even strains of Trypanosoma brucei. Infect Genet Evol 41, 56-62.
Woolhouse, M.E., Rambaut, A., Kellam, P., 2015. Lessons from Ebola: Improving infectious disease surveillance to inform outbreak management. Sci Transl Med 7, 307rv305.

Figure Legend

Figure 1: A flow diagram of the sample preparation (A) and the bioinformatics data handling (B) of the metabarcoded sequencing library. (A) The protocol for the amplification of the 18S rDNA region

621 of the *Trypanosoma* genome used similar protocols to those described by Chaudhry et al. (2019) and
622 Wahab et al. (2020). This method used both the adapter and barcoded primers. The 0-3Ns were added
623 between the adapter and barcoded primers. The 2nd round PCR used overhanging barcoded primers to
624 bind the index sequences and P5/P7 required for the Illumina flow cell. (B) Text files containing rDNA
625 18S sequence data (FASTQ files) were generated from the Illumina Mi-Seq binary raw data outputs,
626 and data analysis was performed using a bespoke modified pipeline in Mothur v1.39.5 software
627 (Schloss et al., 2009) and Illumina Mi-Seq standard procedures (Kozich et al., 2013) described by
628 Chaudhry et al. (2019) and Wahab et al. (2020).
629
630 **Figure 2**: Maximum likelihood phylogenetic tree of 44 consensus sequences generated from 95 18S
631 rDNA sequences of six *Trypanosoma* species shown in the adjacent table. Each species was identified
632 by their respective colour: *T. brucei* (grey), *T. evansi* (light green), *T. theileri* (blue), *T. vivax*
633 (turquoise), *T. simiae* (dark green) *T. congolense* (orange). The model was computed with 1000
634 bootstrap replicates and the values are shown on the branches. Bootstrap values >70 were considered
635 significant. The adjacent table shows the output of the bioinformatic analysis for the six *Trypanosoma*
636 species.
637
638 **Figure 3:** Species composition in "test pools" of *Trypanosoma* isolates. The result shows the average
639 percentage of *Trypanosoma* species detected in the 10 test pools of *Trypanosoma* samples from a set
640 of three replicates produced using the deep amplicon sequencing method. The number of reads for
641 each species in each sample mix was produced and this was subsequently translated into a percentage
642 of the total reads in each sample mix. The resultant percentages were used to create stacked bar charts
643 of the species composition in each sample using Rstudio software. The species are represented in
644 different colours; *T. vivax* (turquoise), *T. congolense* (orange), *T. brucei* (grey) and *T. theileri* (blue).
645 The adjacent table shows the output of the bioinformatic analysis for the 10 test pool samples.
646
647 **Figure 4:** Species composition in mixed infection of *Trypanosoma*-positive samples. The result shows
648 the percentage of reads for each species detected in analysis of 21 *Trypanosoma*-positive samples. The
649 species are represented in different colours; *T. vivax* (turquoise), *T. congolense* (orange), *T. brucei*
650 (grey) and *T. theileri* (blue). The number of reads for each species in each sample was produced and
651 this was subsequently translated into a percentage of the total reads in each sample. The resultant
652 percentages were used to create stacked bar charts of the species composition in each sample using
653 Rstudio software. The adjacent table shows the output of the bioinformatic analysis for the 21
654 *Trypanosoma*-positive samples.
655
656 **Figure 5:** Species composition of field samples originating from the seven endemic regions of the
657 Punjab province of Pakistan. A total of 98 blood samples from cattle, 21 from camels, 19 from buffalo,
658 10 from horses, 5 from sheep and 2 from goat were positive obtained in veterinary clinics. The samples
659 were collected into EDTA tubes and stored at -20℃. Thick and thin blood smears were examined to
660 identify *Trypanosoma* before extracting gDNA. The haemoprotobiome sequencing assay was applied
661 to each sample. The bar of each sample shows the proportion of each species as estimated by deep
662 amplicon sequencing assay. The Y-axis shows the percentage proportions of each species. The species
663 are represented in different colours: *T. theileri* (blue), *T. evansi* (light green). The number of reads for
664 each species in each sample was produced and this was subsequently translated into a percentage of
665 the total reads in each sample. The resultant percentages were used to create stacked bar charts of the
666 species composition in each sample using Rstudio software.
667
668
669

670 **Supplementary Figure Legend**

671

672 **Supplementary Figure S1:** Giemsa-stained blood smears were examined by 1000 x microscopy,
673 showing *Trypanosoma* positive samples in cattle, buffalo and camels.

674

675 **Supplementary Figure S2:** The alignment of sequences of the *T. evansi* and *T. brucei* ITS2 fragment.
676 Polymorphic sites in the ITS2 fragment were indicated with colour.

677

678