

Article

A Targeted “Next-Generation” Sequencing-Informatic Approach to Define Genetic Diversity in *Theileria orientalis* Populations within Individual Cattle: Proof-of-Principle

Anson V. Koehler *, Abdul Jabbar, Ross S. Hall and Robin B. Gasser

Department of Veterinary Biosciences, Melbourne Veterinary School, Faculty of Veterinary and Agricultural Sciences, The University of Melbourne, Parkville, VIC 3010, Australia; jabbara@unimelb.edu.au (A.J.); rossh@unimelb.edu.au (R.S.H.); robinbg@unimelb.edu.au (R.B.G.)

* Correspondence: anson.koehler@unimelb.edu.au; Tel.: +61-3973-12295

Received: 8 May 2020; Accepted: 3 June 2020; Published: 5 June 2020

Abstract: Oriental theileriosis is an economically important tickborne disease of bovines, caused by some members of the *Theileria orientalis* complex. Currently, 11 distinct operational taxonomic units (OTUs), or genotypes, are recognized based on their major piroplasm surface protein (MPSP) gene sequences. Two of these genotypes (i.e., *chitose* and *ikedada*) are recognized as pathogenic in cattle, causing significant disease in countries of the Asia-Pacific region. However, the true extent of genetic variation and associated virulence/pathogenicity within this complex is unknown. Here, we undertook a proof-of-principle study of a small panel of genomic DNAs ($n = 13$) from blood samples originating from individual cattle known to harbor *T. orientalis*, in order to assess the performance of a targeted “next-generation” sequencing-informatic approach to identify genotypes. Five genotypes (*chitose*, *ikedada*, *buffeli*, *type 4*, and *type 5*) were defined; multiple genotypes were found within individual samples, with dominant and minor sequence types representing most genotypes. This study indicates that this sequencing-informatic workflow could be useful to assess the nature and extent of genetic variation within and among populations of *T. orientalis* on a large scale, and to potentially employ panels of distinct gene markers for expanded molecular epidemiological investigations of socioeconomically important protistan pathogens more generally.

Keywords: *Theileria orientalis*; major piroplasm surface protein (MPSP) gene; multiplicity of infection (MOI); next-generation sequencing (NGS); SeekDeep

1. Introduction

Tickborne diseases (TBDs) substantially affect livestock in many parts of the world and have a major negative impact, particularly on resource-poor farming communities [1]. Key TBDs are caused by pathogens including *Anaplasma marginale* and *A. centrale* (anaplasmosis), *Babesia bovis*, *B. bigemina* and *B. divergens* (babesiosis), *Ehrlichia ruminantium* (cowdriosis/heartwater), as well as *Theileria annulata*, *T. parva*, and members of the *T. orientalis* complex (theileriosis), particularly in the tropics and subtropics [2].

Outbreaks of oriental theileriosis have had a significant, adverse economic impact on cattle farms in Australasia and parts of Asia [3–8]. The use of DNA-based diagnostic and analytical tools to establish the distribution and prevalence of distinct genetic variants (genotypes) within the *T. orientalis* complex has been central to the tracking and monitoring of oriental theileriosis outbreaks. Currently, 11 distinct *T. orientalis* genotypes, or operational taxonomic units (OTUs), including *chitose*

(= type 1), *ikedai* (= type 2), *buffeli* (= type 3), types 4 to 8, and N1-N3, have been reported worldwide, and only the genotypes *chitose* and *ikedai* are recognized to be pathogenic in cattle [8,9].

Molecular assays utilized to detect, characterize, differentiate, or quantitate members of the *T. orientalis* complex in blood samples include the conventional polymerase chain reaction (PCR), loop-mediated isothermal amplification (LAMP), multiplex tandem PCR (MT-PCR), nested-PCR, reverse line blot hybridization assay (RLB), and quantitative PCR (qPCR) assays [8]. These methods use one or more genetic markers, such as the major piroplasm surface protein gene (*MPSP*), piroplasm membrane protein genes (*p23*, *p32* to *p34*), the small subunit of the nuclear ribosomal RNA gene (*SSU*), internal transcribed spacers (ITS) of nuclear ribosomal DNA, β -*tubulin* genes, and the mitochondrial cytochrome *c* oxidase subunit 3 gene (*cox3*) (reviewed by [10]).

MPSP has been most widely used in PCR to identify and differentiate genotypes in the *T. orientalis* complex [8]. In some cases, Sanger-sequencing of PCR products has been employed to define genotypes (cf. [8]). This approach has a disadvantage that only the predominant consensus sequence (with or without polymorphism) is recorded, but distinct sequence types within individual amplicons cannot be discerned. Although such sequence types can be defined by PCR-based cloning and sequencing [11], this approach is indirect (via molecular cloning), time-consuming, costly, and can introduce errors [12]. By contrast, the direct sequencing of PCR products using "deep", short-read sequencing utilizing, for example, Illumina technology (www.illumina.com), could allow both sequence diversity within *T. orientalis* populations and infection intensity to be estimated.

Such a targeted "next-generation" (NGS) approach has been applied to *SSU* and other nuclear genetic markers to delineate a range of haemoprotozoan taxa and to explore genetic diversity within samples from individual host animals (e.g., [13–17]). However, for genetic markers that display marked sequence variation among genotypes, such as *MPSP*, there is an imperative to ensure that datasets produced are critically assessed for the presence of artefacts within amplicons (chimeras and jumping PCR errors) and in ensuing sequence data, preferably employing a bioinformatic workflow that is amenable to automation.

Recent studies of *Plasmodium* species (e.g., [18–20]) have demonstrated the exquisite sensitivity of targeted NGS, combined with the use of the sophisticated software, called SeekDeep v2.6.0 [21], to identify, record and quantitate, with confidence, distinct sequence types differing by as little as a point mutation within a PCR product. We expected that such a workflow could be applicable to and highly advantageous for the *T. orientalis* complex. Hence, we established, here, a proof-of-principle for a targeted NGS, i.e., a bioinformatic approach to reliably and directly discern genotypes within the *T. orientalis* complex using *MPSP* as the marker. This approach sets the scene for large-scale investigations of genetic variation in *Theileria* and a range of protistan taxa.

2. Results

2.1. Characteristics of Sequence Datasets, and Clustering of Reads to Define Genotypes

A total of 219,238 "short reads" were generated from all *MPSP* amplicons. After processing with the SeekDeep algorithm, 19,439 reads were excluded, as they represented either singletons, chimeras, or were of poor quality, leaving 199,799 reads, and a mean count of 15,329 reads per sample (Table 1). Of the 13 samples, all but two (A7 and A17) contained mixed genotypes, and one sample (A11) contained all five genotypes identified in this study. The numbers of reads that clustered into individual genotypes were as follows: 76,955 (*chitose*), 52,053 (*ikedai*), 52,236 (*buffeli*), 1905 (*type 4*), and 16,650 (*type 5*) (Tables 1 and S1); *chitose* was the commonest genotype, being identified in 11 of 13 samples; *ikedai* was recorded in nine samples; *type 5* in four samples; and *type 4* in two of the 13 samples (Table 1, Figure 1). Within genotype *chitose*, subgenotypes A and B (cf. [22]) were identified in six and eight samples, respectively.

Table 1. List of bovine blood DNA samples tested for *Theileria orientalis* using the next-generation sequencing-based bioinformatic approach, and the total number of MPSP sequence reads and multiplicity of infection (MOI, i.e., the number of distinct sequence types) in each sample.

Sample Code	Total No. of Reads (MOI)	Genotype					
		Chitose A	Chitose B	Ikeda	Buffeli	Type 4	Type 5
A2	18573 (4)	MT428436			MT428452		
A5	19215 (11)	MT428436		MT428433			
A6	20523 (9)		MT428435	MT428433	MT428453	MT428454	
A7	8167 (38)			MT428433			
A8	14038 (11)	MT428436	MT428435	MT428495	MT428434		MT428479
A9	8957 (39)			MT428433			MT428439
A10	26800 (5)		MT428435		MT428434		MT428439
A11	27608 (7)	MT428436	MT428435	MT428433	MT428434	MT428454	MT428439
A12	14171 (7)		MT428435	MT428433	MT428434		
A13	10032 (11)		MT428435		MT428434		
A14	8813 (11)		MT428440	MT428433	MT428452		
A15	13332 (6)	MT428436	MT428435	MT428433	MT428434		
A17	9570 (25)	MT428436					

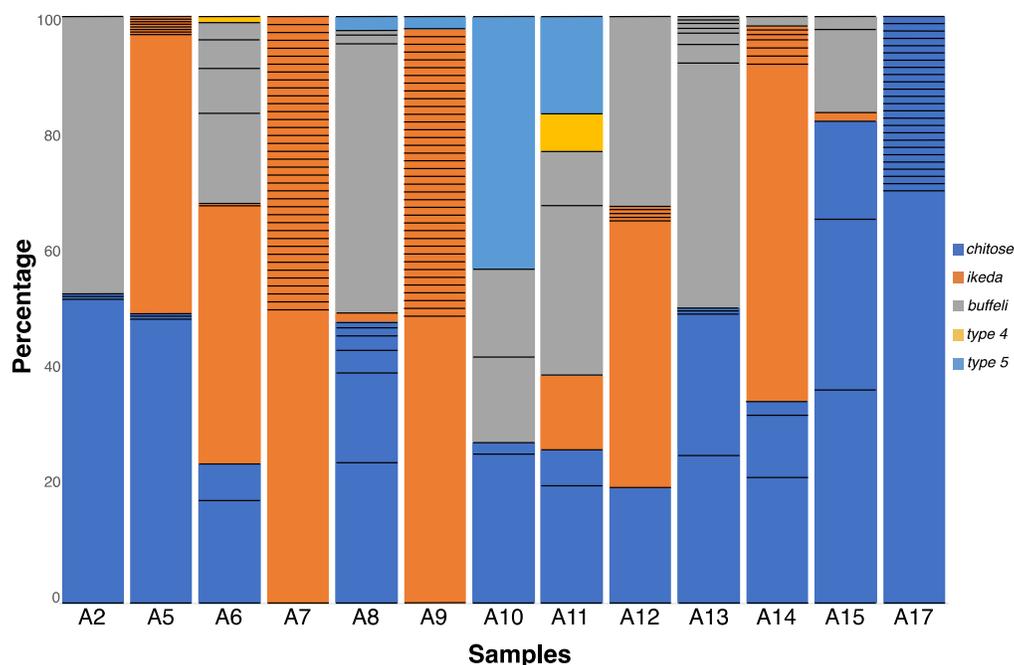


Figure 1. Genotypes of *Theileria orientalis* within individual samples, detected using the next-generation sequence-based bioinformatic approach, and presented as percentages of all MPSP sequence reads within individual samples. The numbers of thin, black, horizontal lines within individual bars (representing distinct nucleotide sequence types) define the multiplicity of infection (MOI) for individual samples (cf. Table 1).

2.2. Extensive MPSP Sequence Diversity within and among Samples

In total, 110 distinct MPSP sequences were identified in all 13 samples as follows: 37 for *chitose*, 58 for *ikeda*, 12 for *buffeli*, one for *type 4* and two for *type 5* (Tables 1 and S1; Figures 1 and S1). Of these 110 unique sequences, 72 (65.5%) were detected in a single sample (A17), whereas 38 (34.5%) were found in multiple samples. The number of different sequences, referred to as the “multiplicity of infection” (MOI) within a sample, ranged from four to 39 (Table 1 and Figure 1). Some samples

contained one genotype, yet many sequences representing the same genotype reflected a higher MOI (e.g., with A7 containing as many as 38 sequence variants representing *ikedda*). Similarly, samples A9 and A17 had high MOI values (39 and 25, respectively, Figure 1), with little or no genotypic diversity (A9 contained *ikedda* and *type5*, whereas A17 contained only *chitose*), contrasting sample A2 which had the lowest MOI (= 4) (Table 1 and Figure 1).

When conceptually-translated amino acid sequences were compared with the dominant (reference) sequence representing each of the five genotypes, 85 of the 110 amino acid sequences were unique. Therefore, most mutations were non-synonymous and were single point mutations (SNPs). Sequences representing each *buffeli* and *chitose* B differed at multiple nucleotide positions, whereas those representing each *ikedda* and *chitose* A varied only at single positions. There was only one sequence variant (representing *ikedda*) whose reading frame was disrupted by a premature stop codon (cf. Figure S1).

2.3. Genotypes within Samples Usually Represented by a Unique, Dominant Nucleotide Sequence

Individual genotypes within individual samples were mostly represented by a dominant “majority” sequence and, typically, sequences differed from this dominant sequence by only one or two nucleotides (Figure S1). For example, 38 sequence variants were identified in Sample A7, with a dominant sequence equating to 57% of 8167 reads, while the next commonest sequence was represented by 3.7% of all reads (Figure 1). For this sample, which contained solely the genotype *ikedda*, there was only a point mutation difference between the dominant sequence and all other sequence types. Likewise, sequences representing *chitose* A differed by single point mutations, whereas some representing *chitose* B and *buffeli* varied at multiple positions (Figure S1). For the 11 samples containing *chitose*, coinfections of subgenotypes A and B were detected in three samples, type A was found exclusively in three samples, and type B was found exclusively in five samples (Table 1). Of all 110 sequence variants, most ($n = 97$) were novel, meaning that they did not have an identical match to any published sequence available in the GenBank database.

3. Discussion

Here, we established proof-of-principle for a targeted NGS, i.e., a bioinformatic approach to reliably and directly discern genotypes within the *T. orientalis* complex using a portion of the *MPSP* gene as the marker. Using this approach, we simultaneously detected five genotypes of the *T. orientalis* complex and, surprisingly, a total of 110 distinct sequence variants in all samples analyzed. As SeekDeep has been shown to eliminate artefactual sequences and retain sequence variants with very high confidence [21], we were able to infer the multiplicity of infection (MOI) of particular genotypes. The findings revealed that mixed-genotypic infections were common, in spite of the small sample size tested in this study. The ability to reliably discern sequence types, and attribute them to particular populations should prove useful for comprehensive investigations of genetic variation within species of *Theileria*. The ability to associate sequence types and genotypes with clinical and asymptomatic cases would facilitate the identification and tracking of sources of infection, and the monitoring of outbreaks (cf. [9]).

NGS has been applied previously to *Theileria*, but in a manner distinct from this proof-of-principle study. Recently, a whole genome sequencing (WGS) method has been used to explore the *T. orientalis* complex in cattle [23], but such an approach is not presently scalable for routine monitoring or diagnostic purposes. Another WGS method was employed to investigate hemoparasites in ticks [24], but the amount of data produced was excessive and the procedure was costly as compared with targeted NGS which used one or a few loci to detect infections [25]. Targeted NGS approaches have been applied to selected *SSU* markers, resulting in the detection of 14 taxa of *Babesia* and *Theileria* in ruminants [14], *Theileria* species in African antelopes [26] and buffaloes [16,17], and *Hepatozoon*, *Babesia*, and *Trypanosoma* species in dogs [15]. However, the *SSU* region does not provide the resolution required for assessing genetic variation within taxa. An NGS approach with some similarity to ours has been applied previously to *Plasmodium falciparum* using regions of genes that encode the merozoite surface protein (*MSP*) [19,20], the *P. falciparum* apical membrane antigen 1

(*pfama1* [18,27]) and genome-wide SNPs (reviewed by [28]). In one study [21], the authors employed targeted NGS of multiple proposed vaccine candidate- and drug resistance-genes, along with mock microbiome datasets.

The gene encoding *MPSP*, originally described as a 32 kDa immunogenic protein with a promise to be a vaccine candidate [29], is orthologous to the *p32* gene of *T. parva* and the *Tams1* gene of *T. annulata*, which are both known for having a mosaic pattern of antigenic diversity [9,30]. Accordingly, the high levels of sequence (genetic) variation found within and among some samples in this study (Table 1 and Figure 1) had been assumed, but its extent was not appreciated based on published information for *T. orientalis*. Interestingly, Hemmink et al. [13] examined allelic variation for six antigen-encoding genes of *T. parva* using a targeted NGS method and found a substantial number of gene variants within blood samples from some cattle investigated, indicating multiple infection event(s) in these animals. Similarly, in our study, the detection of up to five distinct genotypes in some samples studied here (e.g., A11, Figure 1) suggests that multiple infection events had occurred via suitable tick vectors or a single transmission of multiple genotypes by a tick [13]. We hypothesize that this marked variation reflects antigenic drift or shift in the pathogen(s) to evade or subvert immunological attack by the host animal. This hypothesis warrants testing. Ectopic recombination, as reported for the "var" genes of *P. falciparum* (see [31–33]) could be an alternative possibility, although such recombination could result in marked chromosomal rearrangement and organismal destruction [34].

4. Materials and Methods

Blood samples ($n = 13$) from cattle (*Bos taurus* or *Bos indicus*) were available from previously published studies, approved by the Research and Ethics Committee of the College of Veterinary and Animal Sciences, University of Bahawalpur, Pakistan [35] and the Ethiopian National Research Ethics Review Committee [36]. These blood samples were known to contain *T. orientalis* DNA based on results of conventional PCR testing [35,36].

Genomic DNAs were extracted from 200 μ L of each of the whole-blood samples by sodium-dodecyl sulphate (SDS)-proteinase K digestion and mini-column purification using a commercial kit (DNeasy[®] Blood & Tissue, Qiagen, Hilden, Germany), following the manufacturer's instructions, and samples were stored at -20 °C until further analysis. Then, a 344 bp region of the *MPSP* gene was amplified from individual samples using primers MPSP-AJ-F (5-TTC ACT CCA ACA GTC GCC CAC A-3) and MPSP-AJ-R1 (5-ACG TAA ACT TTG ACT GCG GTG-3) [37] in 50 μ L containing 10 mM Tris-HCl (pH 8.4), 50 mM KCl (Promega, Madison, WI, USA), 3.5 mM MgCl₂, 6.25 M of each deoxynucleotide triphosphate (dNTP), 100 pmol of each primer, and 1 U of *GoTaq* polymerase (Promega, Madison, WI, USA). The PCR cycling conditions were as follows: an initial denaturation at 95 °C for 5 min, followed by 35 cycles at 95 °C for 30 s, 58 °C for 30 s, 72 °C for 1 min, followed by 72 °C for 5 min. No-DNA (negative) and known positive (*T. orientalis*) control samples were included in each PCR run. PCR products were examined following agarose (1.5%) gel electrophoresis, and then subjected to Illumina sequencing on a MiSeq platform at Novogene (Beijing, China; en.novogene.com) using the established protocol from Illumina. Paired-end sequence data (consensus length 301 bp) from which adaptors were trimmed were provided in Fastq files. Using SeekDeep v2.6.0 software [21], with settings designed specifically for Illumina data, the "clusters" algorithm was used to remove singletons, low quality reads, and chimeras. Then, sequences were clustered using the default cut-off (0.5%) for a fraction of all reads within a cluster to minimize errors. The "process clusters" algorithm was used to combine the reads from all amplicons. The unique sequences were converted into the FASTA file format. Values for the multiplicity of infection (MOI, the number of distinct sequences within an individual sample) and nucleotide diversity among all samples were calculated using DnaSP v5 [38]. All sequence data were deposited in the GenBank database (accession nos. MT428433 to MT428542).

5. Conclusions

In this study, first, we elected to critically assess the sequencing and informatic approach to convince ourselves of the reliability of the datasets and the ability of the SeekDeep algorithm to eliminate PCR-induced artefacts from the datasets. Our results indicate that this approach is useful and has the potential to be applied to a large panel of distinct genetic markers, preferably derived from well-assembled genomes. We believe that a focus on using multiple divergent single-copy genes would be advantageous, circumventing problems associated with in vitro recombination of paralogous sequences (representing multigene families) during PCR. We envisage that the use of a multiplex PCR-based NGS-informatic workflow, employing replicates to ensure repeatability and reproducibility and to gain further confidence in MOI data, should provide a unique platform for future molecular epidemiological surveys and population genetic analyses of protists of veterinary and medical importance.

Supplementary Materials: The following are available online at www.mdpi.com/2076-0817/9/6/448/s1, Figure S1: Nucleotide alignments of the sequence variants for the *Theileria orientalis* complex (including genotypes *buffeli*, *chitose* (A & B), and *ikedai*), produced by targeted NGS using a region of the major piroplasm surface protein gene (*MPSP*). Alignments are compared to the most frequent sequence variant for each genotype. Synonymous nucleotide changes are in bold, and nonsynonymous amino acid changes are highlighted in colour. Genotypes *type 4* and *type 5* were omitted here, as only three sequences were recorded, Table S1: Codes and GenBank accession numbers of all 110 (partial) *MPSP* sequences obtained from 13 bovine blood DNA samples tested for *Theileria orientalis* using next-generation sequencing-based bioinformatic approach. Total number of the sequence reads for individual genotypes in samples are listed.

Author Contributions: Conceptualization, A.J.; methodology, A.J., R.H., and A.V.K.; software, A.V.K. and R.S.H.; formal analysis, A.V.K. and R.S.H.; investigation, A.V.K. and R.S.H.; data curation, A.V.K.; writing—original draft preparation, A.V.K. and R.B.G.; writing—review and editing, R.B.G., A.V.K., and A.J.; visualization, A.V.K.; funding acquisition, R.B.G., A.V.K., and A.J. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported through a Collaborative Research Grant (no. 501101 to A.J.), an Early Career Research Grant (no. 601308 to A.J.) from The University of Melbourne, and grants from the Australian Research Council (LP180101085 to R.B.G. and LP160101299 to R.B.G. and A.V.K.).

Acknowledgments: The authors thank Michael F. Duffy for valuable comments on aspects of the draft manuscript.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Makala, L.H.; Mangani, P.; Fujisaki, K.; Nagasawa, H. The current status of major tick-borne diseases in Zambia. *Vet. Res.* **2003**, *34*, 27–45, doi:10.1051/vetres:2002056.
2. Jabbar, A.; Abbas, T.; Sandhu, Z.U.; Saddiqi, H.A.; Qamar, M.F.; Gasser, R.B. Tick-borne diseases of bovines in Pakistan: Major scope for future research and improved control. *Parasit Vectors* **2015**, *8*, 283, doi:10.1186/s13071-015-0894-2.
3. Kamau, J.; de Vos, A.J.; Playford, M.; Salim, B.; Kinyanjui, P.; Sugimoto, C. Emergence of new types of *Theileria orientalis* in Australian cattle and possible cause of theileriosis outbreaks. *Parasit Vectors* **2011**, *4*, 22, doi:10.1186/1756-3305-4-22.
4. Perera, P.K.; Gasser, R.B.; Anderson, G.A.; Jeffers, M.; Bell, C.M.; Jabbar, A. Epidemiological survey following oriental theileriosis outbreaks in Victoria, Australia, on selected cattle farms. *Vet. Parasitol.* **2013**, *197*, 509–521, doi:10.1016/j.vetpar.2013.06.023.
5. Perera, P.K.; Gasser, R.B.; Firestone, S.M.; Anderson, G.A.; Malmø, J.; Davis, G.; Beggs, D.S.; Jabbar, A. Oriental theileriosis in dairy cows causes a significant milk production loss. *Parasit Vectors* **2014**, *7*, 73, doi:10.1186/1756-3305-7-73.
6. Watts, J.G.; Playford, M.C.; Hickey, K.L. *Theileria orientalis*: A review. *N. Z. Vet. J.* **2016**, *64*, 3–9, doi:10.1080/00480169.2015.1064792.

7. Gebrekidan, H.; Nelson, L.; Smith, G.; Gasser, R.B.; Jabbar, A. An outbreak of oriental theileriosis in dairy cattle imported to Vietnam from Australia. *Parasitology* **2017**, *144*, 738–746, doi:10.1017/S0031182016002328.
8. Gebrekidan, H.; Perera, P.K.; Ghafar, A.; Abbas, T.; Gasser, R.B.; Jabbar, A. An appraisal of oriental theileriosis and the *Theileria orientalis* complex, with an emphasis on diagnosis and genetic characterisation. *Parasitol. Res.* **2020**, *119*, 11–22, doi:10.1007/s00436-019-06557-7.
9. Sivakumar, T.; Hayashida, K.; Sugimoto, C.; Yokoyama, N. Evolution and genetic diversity of *Theileria*. *Infect. Genet. Evol.* **2014**, *27*, 250–263, doi:10.1016/j.meegid.2014.07.013.
10. Mans, B.J.; Pienaar, R.; Latif, A.A. A review of *Theileria* diagnostics and epidemiology. *Int. J. Parasitol. Parasit. Wildl.* **2015**, *4*, 104–118, doi:10.1016/j.ijppaw.2014.12.006.
11. Perera, P.K.; Gasser, R.B.; Firestone, S.M.; Smith, L.; Roeber, F.; Jabbar, A. Semiquantitative multiplexed tandem PCR for detection and differentiation of four *Theileria orientalis* genotypes in cattle. *J. Clin. Microbiol.* **2015**, *53*, 79–87, doi:10.1128/JCM.02536-14.
12. Gasser, R.B. Molecular tools—Advances, opportunities and prospects. *Vet. Parasitol.* **2006**, *136*, 69–89, doi:10.1016/j.vetpar.2005.12.002.
13. Hemmink, J.D.; Sitt, T.; Pelle, R.; de Klerk-Lorist, L.M.; Shiels, B.; Toye, P.G.; Morrison, W.I.; Weir, W. Ancient diversity and geographical sub-structuring in African buffalo *Theileria parva* populations revealed through metagenetic analysis of antigen-encoding loci. *Int. J. Parasitol.* **2018**, *48*, 287–296, doi:10.1016/j.ijpara.2017.10.006.
14. Chaudhry, U.; Ali, Q.; Rashid, I.; Shabbir, M.Z.; Ijaz, M.; Abbas, M.; Evans, M.; Ashraf, K.; Morrison, I.; Morrison, L.; et al. Development of a deep amplicon sequencing method to determine the species composition of piroplasm haemoprotozoa. *Ticks Tick Borne Dis.* **2019**, *10*, 101276, doi:10.1016/j.ttbdis.2019.101276.
15. Huggins, L.G.; Koehler, A.V.; Ng-Nguyen, D.; Wilcox, S.; Schunack, B.; Inpankaew, T.; Traub, R.J. A novel metabarcoding diagnostic tool to explore protozoan haemoparasite diversity in mammals: A proof-of-concept study using canines from the tropics. *Sci. Rep.* **2019**, *9*, 1–10, doi:10.1038/s41598-019-49118-9.
16. Combrink, L.; Glidden, C.K.; Beechler, B.R.; Charleston, B.; Koehler, A.V.; Sisson, D.; Gasser, R.B.; Jabbar, A.; Jolles, A.E. Age of first infection across a range of parasite taxa in a wild mammalian population. *Biol. Lett.* **2020**, *16*, 20190811, doi:10.1098/rsbl.2019.0811.
17. Glidden, C.K.; Koehler, A.V.; Hall, R.S.; Saeed, M.A.; Coppo, M.; Beechler, B.R.; Charleston, B.; Gasser, R.B.; Jolles, A.E.; Jabbar, A. Elucidating cryptic dynamics of *Theileria* communities in African buffalo using a high-throughput sequencing informatics approach. *Ecol. Evol.* **2019**, *10*, 70–80, doi:10.1002/ece3.5758.
18. Miller, R.H.; Hathaway, N.J.; Kharabora, O.; Mwandagilirwa, K.; Tshetu, A.; Meshnick, S.R.; Taylor, S.M.; Juliano, J.J.; Stewart, V.A.; Bailey, J.A. A deep sequencing approach to estimate *Plasmodium falciparum* complexity of infection (COI) and explore apical membrane antigen 1 diversity. *Malar. J.* **2017**, *16*, 490, doi:10.1186/s12936-017-2137-9.
19. Zhong, D.; Lo, E.; Wang, X.; Yewhalaw, D.; Zhou, G.; Atieli, H.E.; Githeko, A.; Hemming-Schroeder, E.; Lee, M.C.; Afrane, Y.; et al. Multiplicity and molecular epidemiology of *Plasmodium vivax* and *Plasmodium falciparum* infections in east Africa. *Malar. J.* **2018**, *17*, 185, doi:10.1186/s12936-018-2337-y.
20. Metoh, T.N.; Chen, J.H.; Fon-Gah, P.; Zhou, X.; Moyou-Somo, R.; Zhou, X.N. Genetic diversity of *Plasmodium falciparum* and genetic profile in children affected by uncomplicated malaria in Cameroon. *Malar. J.* **2020**, *19*, 115, doi:10.1186/s12936-020-03161-4.
21. Hathaway, N.J.; Parobek, C.M.; Juliano, J.J.; Bailey, J.A. SeekDeep: Single-base resolution de novo clustering for amplicon deep sequencing. *Nucleic Acids Res.* **2017**, *46*, e21, doi:10.1093/nar/gkx1201.
22. Jenkins, C.; Micallef, M.; Alex, S.M.; Collins, D.; Djordjevic, S.P.; Bogema, D.R. Temporal dynamics and subpopulation analysis of *Theileria orientalis* genotypes in cattle. *Infect. Genet. Evol.* **2015**, *32*, 199–207, doi:10.1016/j.meegid.2015.03.017.
23. Bogema, D.R.; Micallef, M.L.; Liu, M.; Padula, M.P.; Djordjevic, S.P.; Darling, A.E.; Jenkins, C. Analysis of *Theileria orientalis* draft genome sequences reveals potential species-level divergence of the ikeda, chitose and buffeli genotypes. *BMC Genom.* **2018**, *19*, 298, doi:10.1016/j.meegid.2015.03.017.
24. Brinkmann, A.; Hekimoglu, O.; Dincer, E.; Hagedorn, P.; Nitsche, A.; Ergunay, K. A cross-sectional screening by next-generation sequencing reveals *Rickettsia*, *Coxiella*, *Francisella*, *Borrelia*, *Babesia*, *Theileria* and *Hemolivia* species in ticks from Anatolia. *Parasit Vectors* **2019**, *12*, 26, doi:10.1186/s13071-018-3277-7.
25. Goodwin, S.; McPherson, J.D.; McCombie, W.R. Coming of age: Ten years of next-generation sequencing technologies. *Nat. Rev. Genet.* **2016**, *17*, 333–351, doi:10.1038/nrg.2016.49.

26. Pienaar, R.; Josemans, A.; Latif, A.A.; Mans, B.J. The host-specificity of *Theileria* sp. (sable) and *Theileria* sp. (sable-like) in African bovidae and detection of novel *Theileria* in antelope and giraffe. *Parasitology* **2020**, *147*, 213–224, doi:10.1017/S003118201900132X.
27. Boyce, R.M.; Hathaway, N.; Fulton, T.; Reyes, R.; Matte, M.; Ntaro, M.; Mulogo, E.; Waltmann, A.; Bailey, J.A.; Siedner, M.J.; et al. Reuse of malaria rapid diagnostic tests for amplicon deep sequencing to estimate *Plasmodium falciparum* transmission intensity in western Uganda. *Sci. Rep.* **2018**, *8*, 10159, doi:10.1038/s41598-018-28534-3.
28. Koepfli, C.; Mueller, I. Malaria epidemiology at the clone level. *Trends Parasitol.* **2017**, *33*, 974–985, doi:10.1016/j.pt.2017.08.013.
29. Kobayashi, N.; Onuma, M.; Kirisawa, R.; Ohgitani, T.; Takahashi, K.; Sasaki, N.; Kawakami, Y. Monoclonal antibodies against intraerythrocytic merozoites (piroplasms) of *Theileria sergenti*. *Nihon Juigaku Zasshi* **1987**, *49*, 697–702, doi:10.1292/jvms1939.49.697.
30. Gubbels, M.J.; Katzer, F.; Hide, G.; Jongejan, F.; Shiels, B.R. Generation of a mosaic pattern of diversity in the major merozoite-piroplasm surface antigen of *Theileria annulata*. *Mol. Biochem. Parasitol.* **2000**, *110*, 23–32, doi:10.1016/s0166-6851(00)00253-x.
31. Freitas-Junior, L.H.; Bottius, E.; Pirrit, L.A.; Deitsch, K.W.; Scheidig, C.; Guinet, F.; Nehrbass, U.; Wellem, T.E.; Scherf, A. Frequent ectopic recombination of virulence factor genes in telomeric chromosome clusters of *P. falciparum*. *Nature* **2000**, *407*, 1018–1022, doi:10.1038/35039531.
32. Duffy, M.F.; Byrne, T.J.; Carret, C.; Ivens, A.; Brown, G.V. Ectopic recombination of a malaria *var* gene during mitosis associated with an altered *var* switch rate. *J. Mol. Biol.* **2009**, *389*, 453–469, doi:10.1016/j.jmb.2009.04.032.
33. Claessens, A.; Hamilton, W.L.; Kekre, M.; Otto, T.D.; Faizullahoy, A.; Rayner, J.C.; Kwiatkowski, D. Generation of antigenic diversity in *Plasmodium falciparum* by structured rearrangement of *var* genes during mitosis. *PLoS Genet.* **2014**, *10*, e1004812, doi:10.1371/journal.pgen.1004812.
34. Montgomery, E.; Charlesworth, B.; Langley, C.H. A test for the role of natural selection in the stabilization of transposable element copy number in a population of *Drosophila melanogaster*. *Genet. Res.* **1987**, *49*, 31–41, doi:10.1017/s0016672300026707.
35. Gebrekidan, H.; Abbas, T.; Wajid, M.; Ali, A.; Gasser, R.B.; Jabbar, A. Molecular characterisation of *Theileria orientalis* in imported and native bovines from Pakistan. *Infect. Genet. Evol.* **2017**, *47*, 19–25, doi:10.1016/j.meegid.2016.11.005.
36. Gebrekidan, H.; Gasser, R.B.; Baneth, G.; Yasur-Landau, D.; Nachum-Biala, Y.; Hailu, A.; Jabbar, A. Molecular characterization of *Theileria orientalis* from cattle in Ethiopia. *Ticks Tick Borne Dis.* **2016**, *7*, 742–747, doi:10.1016/j.ttbdis.2016.03.005.
37. Cufos, N.; Jabbar, A.; de Carvalho, L.M.; Gasser, R.B. Mutation scanning-based analysis of *Theileria orientalis* populations in cattle following an outbreak. *Electrophoresis* **2012**, *33*, 2036–2040, doi:10.1002/elps.201200082.
38. Librado, P.; Rozas, J. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **2009**, *25*, 1451–1452, doi:10.1093/bioinformatics/btp187.

