ABSTRACT

Babesia microti and *Babesia duncani* are the main causative agents of human babesiosis in the United States. While significant knowledge about *B. microti* has been gained over the past few years, nothing is known about *B. duncani* biology, pathogenesis, mode of transmission or sensitivity to currently recommended therapies. Studies in immunocompetent wild type mice and hamsters have shown that unlike *B. microti*, infection with *B. duncani* results in severe pathology and ultimately death. The parasite factors involved in *B. duncani* virulence remain unknown. Here we report the first known completed sequence and annotation of the apicoplast and mitochondrial genomes of *B. duncani*. We found that the apicoplast genome of this parasite consists of a 34 kb monocistronic circular molecule encoding functions that are important for apicoplast gene transcription as well as translation and maturation of the organelle's proteins. The mitochondrial genome of *B. duncani* consists of a 5.9 kb monocistronic linear molecule with two inverted repeats of 48 bp at both ends. Using the conserved cytochrome b (Lemieux) and cytochrome c oxidase subunit I (*coxI*) proteins encoded by the mitochondrial genome, phylogenetic analysis revealed that *B. duncani* defines a new lineage among apicomplexan parasites distinct from *B. microti*, *Babesia bovis*, *Theileria* spp. and *Plasmodium* spp. Annotation of the apicoplast and mitochondrial genomes of *B. duncani* identified targets for development of effective therapies. Our studies set the stage for evaluation of the efficacy of these drugs alone or in combination against *B. duncani* in culture as well as in animal models.

Keywords: Babesiosis, Genome, Mitochondria, Apicoplast, Sequencing, Annotation, Drug targets

1. Introduction

Human babesiosis is a global infectious disease transmitted by ticks, blood transfusion or congenitally (reviewed in Vannier and Krause (2012); Vannier et al. (2015)). The disease is caused by intraerythocytic parasites of the genus *Babesia*. Immunocompromised and asplenic individuals are at greatest risk of developing babesiosis symptoms, which include fever and headache and can advance to multi-system organ failure and death (Vannier et al., 2015). Most cases of babesiosis occur in the northeastern and midwestern regions of the United States and are caused by *Babesia microti* and transmitted by *Ixodes scapularis*, the same tick species that transmits the agents of Lyme disease, anaplasmosis and Powassan virus disease (Burgdorfer et al., 1982; Spielman et al., 1985). Beside these cases, several clinical reports from Washington State, USA and California, USA were linked to another *Babesia* sp, *Babesia duncani* (Kjemtrup and Conrad, 2000). The first case of human babesiosis reported to be caused by a new *Babesia* sp., named *WA1*, was of a 41-year-old man from Washington State (Quick et al., 1993). Since then 11 more cases of human babesiosis attributable to WA1 and WA1-like organisms have been reported in California and Washington State, and the etiological agent later identified as *B. duncani* (reviewed in Kjemtrup and Conrad (2000)). Two other cases preceding these 12 *B. duncani* cases were reported in California and were presumed to be caused by this pathogen (Scholtens et al., 1968; Bredt et al., 1981). Studies in hamsters and mice following infection with the parasite showed that *B. duncani* pathogenesis is different from that caused by *B. microti*. Infection with *B. microti* is characterized by an initial phase of high parasitemia, anemia and splenomegaly followed by rapid decline in parasitemia to undetectable levels, leaving animals surviving infection with little to no detectable symptoms (Cullen and Levine, 1987; Wozniak et al., 1996). In contrast, *B. duncani* infection of mice and hamsters results in both a rapid increase

in parasitemia and severe pathology with mortality rates of more than 95% in C3H, A/J, AKR/N and DBA/1J mice, between 40 to 50% in BALB/cJ, CBAJ and 129/J mice and less than10% in C57BL/6 and C57BL/10 mice (Dao and Eberhard, 1996; Moro et al., 1998).

While significant knowledge has been gained over the past several years about *B. microti*, nothing is known about the biology, genome composition and structure, or pathogenesis of *B. duncani*. Furthermore, recommended therapies have not been evaluated directly against this parasite in vitro or in animal models. Recent studies by Swei and colleagues (in press) suggest that the enzootic tick vector of *B. duncani* is *Dermacentor albipictus* and the reservoir host is likely the mule deer.

The data described in this study represent, to our knowledge, the first report of the completed sequence, assembly, and annotation of the apicoplast and mitochondrial genomes of *B. duncani*. Phylogenetic analysis using mitochondrial genes shows that *B. duncani* defines a new lineage in the Apicomplexa phylum distinct from *B. microti*. Our analysis further predicts potential therapeutic targets and new strategies to develop effective strategies to treat babesiosis resulting from infection by *B. duncani*.

2. Materials and methods

2.1. Animal ethics statement

All animal experimental protocols followed Yale University institutional guidelines for care and use of laboratory animals and were approved by the Institutional Animal Care and Use Committees (IACUC) at Yale University. Rules for ending experiments were to be enacted if animals showed any signs of distress or appeared moribund.

2.2. Sequencing assembly

The *B. duncani* WA1 isolate was obtained from BEI Resources (www.beiresources.org; Number: NR-12311). This parasite was isolated from the blood of the first reported case of babesiosis acquired in Washington State (Quick et al., 1993). The strain was injected into hamsters and infected red blood cells (RBCs) were purified and used to isolate *B. duncani* total DNA. This DNA was sequenced using both Illumina Hi-Seq 2500 paired-end 75 bp short-read sequencing and PacBio single molecule long-reads (>10kb) sequencing. Assembly was performed using the following steps: i) raw sequence reads were mapped using BWA-MEM (Li and Durbin, 2009) and BLASR (Chaisson and Tesler, 2012) against the host genome sequence (Golden Hamster MesAur1.0 genome, https://www.ncbi.nlm.nih.gov/assembly/GCF_000349665.1/) to remove host DNA contamination; ii) potential sequencing errors in PacBio reads were corrected using ectools (preprint at http://www.biorxiv.org/content/early/2014/06/18/006395) and then the corrected long reads were assembled using Celera Assembler v7 (Berlin et al., 2015); iii) samtools (Li et al., 2009) was used to extract only unmapped reads, and bedtools (Quinlan and Hall, 2010) bam2fastq to convert those back to fastq; iv) spades (Bankevich et al., 2012), with default parameters, was used to assemble the remaining reads into scaffolds. Two of the assembled contigs encompassed the entire apicoplast and mitochondrial genomes. The sequence of the mitochondrial genome was further validated using long-range PCRs and Sanger sequencing.

2.3. Annotation of the apicoplast and mitochondrial genomes

Annotation of the apicoplast and mitochondrial genomes was performed using Artemis (http://www.sanger.ac.uk/science/tools/artemis) (Rutherford et al., 2000) to identify all open

reading frames, and BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi) (Altschul et al., 1990) to identify homologous proteins in other organisms in the GenBank database. The prediction of tRNAs in both genomes was accomplished using tRNAscanSE version 1.21 and version 2.0 (http://lowelab.ucsc.edu/tRNAscan-SE/) with the search mode set to "default" and the source set to "Mito/Chloroplast" (Lowe and Eddy, 1997; Schattner et al., 2005; Lowe and Chan, 2016). Only tRNAs with a score above 30% were annotated on the genome. rRNA genes, large subunit (LSU) and small subunit (SSU), were determined by searching for their counterparts in *Babesia bovis, Babesia orientalis, B. microti* and *Theileria parva*. The circular genetic map for the apicoplast genome was designed using CGView (http://stothard.afns.ualberta.ca/cgview_server/) (Stothard and Wishart, 2005).

2.4. Phylogenetic analysis

A phylogenetic relationship was established using concatenated sequences of the *cob* and *cox1* mitochondrial genes. The sequences were first added consecutively in MEGA7 (Molecular Evolutionary Genetics Analysis version 7.0) (Kumar et al., 2016) before they were aligned using MUSCLE (Edgar, 2004). A Neighbor-Joining tree (Saitou and Nei, 1987) was then constructed using bootstrap analysis inferred from 1000 replicates (Felsenstein, 1985). The bootstrap consensus tree inferred from 1000 replicates was taken to represent the evolutionary history of the taxa analyzed (Felsenstein, 1985). The evolutionary distances were computed using the JTT matrix-based method and are in the units of the number of amino acid substitutions per site (Jones et al., 1992). Evolutionary analyses were conducted in MEGA7 (Kumar et al., 2016).

3. Results and Discussion

3.1. Identification of the B. duncani apicoplast and mitochondrial genomes

In order to obtain the genome sequence of the apicoplast and mitochondrial genomes of *B. duncani*, the WA1 clinical isolate was first propagated in hamsters and DNA isolated from infected RBCs. Search for coding sequences known to be encoded by the apicoplast and mitochondrial genomes of other apicomplexan parasites identified two contigs of ~34 kb (GenBank Accession no. **MH107388**) and ~6 kb (GenBank Accession no. **MH107387**) in length, respectively.

3.2. Annotation of the apicoplast genome of B. duncani

The apicoplast, a non-photosynthetic plastid organelle, is the result of a secondary endosymbiotic event and is a key characteristic of apicomplexan parasites (McFadden, 2011). Several metabolic processes have been shown to be active or are predicted to take place within this organelle (Lim and McFadden, 2010; Seeber and Soldati-Favre, 2010). The most well-characterized of these is the 1-deoxy-D-xylulose-5-phosphate (DOXP) pathway for the synthesis of isoprenoid precursors (Jomaa et al., 1999; Ralph et al., 2004). Parasite proteins involved in these metabolic machineries are all encoded by the nuclear genome and contain an N-terminal targeting motif required for their localization to the apicoplast. Some of these functions have been targeted by different classes of compounds including the antibiotic fosmidomycin (Jomaa et al., 1999). On the other hand, proteins encoded by the apicoplast genome are involved primarily in housekeeping functions (DNA replication, transcription and translation) and could be targeted by various antibiotics such as ciprofloxacin, rifampicin and thiostrepton, respectively (Gardner et al., 1991; Fichera and Roos, 1997; Chaubey et al., 2005). These antibiotics showed a remarkable inhibitory effect on the in vitro growth of *B. bovis, Babesia bigemina, Babesia caballi* and *Babesia equi* (AbouLaila et al., 2012). Additionally, thiostrepton inhibited the growth of *B.*

microti in vivo (AbouLaila et al., 2012). Our analysis of the *B. duncani* apicoplast genome revealed a circular molecule of 34,142 bp in size. Its size and structure are similar to those of *B. orientalis* (33.2 kb) and *B. bovis* (33 kb) (Brayton et al., 2007; Huang et al., 2015). The *B. duncani* apicoplast genome is 15.2% G+C rich and comprises 38 open reading frames (ORFs) encoding 17 ribosomal proteins, four subunits of RNA polymerase, one translation elongation factor Tu (EF-Tu), two copies of the ClpProtease and 14 hypothetical proteins (HypA-N) ranging in size between 103 and 305 amino acids (Table 1 and Fig. 1). All genes are oriented in the same direction and are monocistronic (Fig. 1). The genome also encodes one large (23S), and one small (16S) subunit RNA and a set of 21 tRNAs that facilitate translation of these genes. Our analysis revealed that all genes encoded by the apicoplast genome of *B. duncani* initiate with an AUG codon This contrasts with *B. microti* where 18 of the 31 CDSs encoded by the apicoplast geneome start with an AUG codon whereas the remaining genes start with an AUA codon (Cornillot et al., 2012). In both organisms, an A-rich region is found immediately upstream of the initiation codon and may play a role in the recruitment of the translation machinery (Garg et al., 2014). Translation in the *Plasmodium falciparum* apicoplast has been shown to initiate with the formation of the initiation complex involving two initiation factors, IF1 and IF3, and an unknown factor facilitating entry of initiator tRNA which could be the charged apicoplast-encoded $tRNA^{fMet}$ (Haider et al., 2015). Similar to *P. falciparum*, no initiation factors are encoded by the apicoplast genome and therefore are likely to be encoded by the nuclear genome and then recruited to the apicoplast (Haider et al., 2015). Translation termination is achieved by a single release factor, $PfRF2_{Api}$, which displays specific recognition of both UAA and UGA, the only two stop codons found in apicoplast ORFs (Vaishya et al., 2016). Interestingly, of the 38 ORFs encoded by the apicoplast genome of *B. duncani*, 35 carry a UAA stop codon, two (*rps3*

and the *rpl36*) carry a UGA stop codon, and one (*HypN*) ends with the UAG stop codon. This suggests that either the nuclear-encoded *B. duncani* ortholog of PfRF2Api recognizes all three stops codons or that translation termination in the apicoplast of this parasite requires more than one release factor (RF). While the majority of the apicoplast-encoded CDSs do not overlap, seven gene pairs were found to overlap. Among these, *rpl36* and *HypF* overlap by 44 bases and *CLpProtease2* and *HypJ* overlap by 65 bases, a feature unique to *B. duncani*.

Most hypothetical proteins identified in the apicoplast genome of *B. duncani* share no homologs with any other protein or contain no recognizable functional domains in available genome databases (Supplementary Table S1). Others such as HypF have orthologs in other parasites but their functions remain unknown. A noticeable hypothetical protein, HypJ, is one of the largest proteins encoded by the *B. duncani* apicoplast genome and is 305 amino acids in length. This protein, however, has no homologs in other organisms and its function remains to be determined.

Using tRNA-SE Scan to identify tRNAs in the apicoplast genome of *B. duncani*, 21 tRNAs were identified. Our analysis further identified 17 ribosomal proteins encoded by the *B. duncani* apicoplast genome. These include seven rpl (large) proteins and 10 rps (small) proteins and share high sequence similarity with apicoplast-encoded ribosomal proteins from other apicomplexan parasites (Huang et al., 2015). Therefore, any additional components needed for protein translation in the apicoplast are likely encoded by the nuclear genome and targeted to the apicoplast. For example, the *rff* gene encoding 5S rRNA in the chloroplast genome of *Chromera*, a distant ancestor of the apicoplast, was not detected in the apicoplast genome of *B. duncani*. Therefore, the 5S rRNA is either imported from the cytoplasm (shown to occur in mammalian mitochondria) or *B. duncani* encodes a highly divergent *rff* gene (Magalhães et al., 1998).

Interestingly, the apicoplast genome of *B. duncani* lacks the *rpl11* ribosomal gene, a gene also not found in *B. bovis, B. microti, T. parva* or *P. falciparum.* This suggests that *B. duncani* might have evolved a protein synthesis mechanism that is independent of the L11 protein. Alternatively, the *rpl11* protein might be encoded by the nuclear genome and then translocated to the apicoplast or another L11 protein is encoded by another *rpl11-like* gene which is radically divergent from the one found in other prokaryotes.

3.3. Comparison of the apicoplast genome of B. duncani to that of other apicomplexan species

The apicoplast genome of *B. duncani* was found to have four clusters in synteny with the original chloroplast genome of *Chomera* algae and other apicomplexan species (Fig. 1). Cluster 1 comprises the genes encoding ribosomal proteins, one hypothetical protein (HypF) and the elongation factor EF-Tu in an organization similar to that found in *Babesia* spp., *P. falciparum, Toxoplasma gondii* and *Chomerida* sp. (Fig. 2). Both *P. falciparum* and *Chromera* algae have retained the *rpl23* gene, whereas this gene is lacking in the apicoplast genome of *B. duncani* as well as all *Babesia* spp. sequenced to date. Whereas in Chromera sp. the *rps13* gene is found between *rps5* and *rpl36,* in the *B. duncani* apicoplast genome this locus contains the *HypF* gene. No tRNAs were found in Cluster 1 of *B. duncani* (Fig. 2).

Cluster 1 of the *B. duncani* genome is flanked by 11 tRNAs in Cluster 4 (Fig. 3). The tRNAs in *B. duncani* are clustered together with short nucleotide sequences separating them, whereas in *B. bovis, B. microti* and *T. parva* long non-coding nucleotide sequences are found between the tRNAs in this cluster (Supplementary Fig. S1). Four of the 11 tRNAs are conserved in *B. microti, T. parva* and *B. bovis*. On the other side of Cluster 1, closest to Cluster 2 and between *tufA* and CLp protease genes, only one conserved tRNA-Gln was found between *HypK*

and *HypJ* (Supplementary Fig. S1). Additionally, multiple hypothetical CDSs, more than those found in *B. microti*, *B. bovis* and *T. parva,* were found in the *B. duncani* apicoplast genome. Lastly, unlike most apicomplexan parasites, the *CLpProtease 2* of *B. duncani* is located downstream of *CLpProtease 1*. Overall, these data suggest that the regions on both ends of Cluster 1 may represent sites of frequent recombination events.

Cluster 2 in *B. duncani* consists of 13 hypothetical proteins, two ClpC genes and one tRNA-Gln (conserved in *B. bovis, T. parva* and *B. microti)* (Fig. 1). Both *ClpC* proteins of *B. duncani* (*CLpProtease 1* and *CLpProtease* 2) contain a AAA_2 ATPase domain (Fig. 1 and Supplementary Fig. S1). However, the tRNA-Ser (UGA) and tRNA-Trp (CCA), a key feature of this region of Cluster 2, conserved in *B. microti* and *B. bovis* (Brayton et al., 2007; Cornillot et al., 2012), is missing from *B. duncani* and was not found anywhere else in the apicoplast genome. No other known or so far sequenced apicoplast genomes show this arrangement of hypothetical proteins in this cluster, resulting in *B. duncani* having one of the largest Cluster 2 regions of the apicoplast genome.

Cluster 3 consists of four RNA polymerase genes (*rpoB, rpoC1, rpoC2a* and *rpoC2b*) and the *rps2* gene, which encodes an S2 ribosomal protein (Fig. 1). The content and arrangement of Cluster 3 mimics that in *B. orientalis, B. bovis, T. parva and B. microti* (Gardner et al., 2005; Brayton et al., 2007; Cornillot et al., 2012; Garg et al., 2014; Huang et al., 2015). The opposite orientation is found in *P. falciparum* and *T. gondii,* suggesting that perhaps an inversion took place during the evolution of piroplasmida, possibly even resulting in the loss of the *sufB* gene (Foth and McFadden, 2003).

Finally, Cluster 4 of the apicoplast genome of *B. duncani* contains a single set of *lsu* and *ssu* genes transcribed in the same direction (Fig. 3). Nine tRNAs are organized together at the

start of Cluster 4. Not only are there no large gaps between them, most apicomplexans have only eight tRNA genes, a difference that may have been created by simple recombination events. Similar to Cluster 3, the gene order and copy number vary between different species. In *Chromera,* a CDS separates the *ssu* and the *lsu* genes in both copies of the genes (Cornillot et al., 2012; Huang et al., 2015). Meanwhile, in *Toxoplasma* and *Plasmodium*, the *ssu* and the *lsu* genes are in opposite directions (Cornillot et al., 2012). In contrast, *B. duncani, B. bovis* and *T. parva* all have only one copy of the two ribosomal genes, both of which were transcribed in the same direction. Interestingly, *B. duncani* does not encode a tRNA upstream of the *ssu* and the *lsu* genes but it has preserved the tRNA-Thr (UGU) gene upstream of the *rps4* gene (Fig. 3).

3.4. Annotation of the mitochondrial genome of B. duncani

Mitochondria are vital organelles present in almost all eukaryotic organisms (Frederick and Shaw, 2007; Hikosaka et al., 2010; Kaczanowski et al., 2011; Taylor-Brown and Hurd, 2013). Their functions include energy generation, metabolism and cell growth (Kaczanowski et al., 2011; Taylor-Brown and Hurd, 2013). The mitochondrial genome varies in size, structure and organization between organisms and species (Feagin, 2000; Hikosaka et al., 2010). Structurally, two forms of the mitochondrial genome exist: a linear form and a circular form. The circular form is present in animals, at a size ranging between 15 and 20 kb, whereas the linear form has been reported in numerous apicomplexan parasites at a size of ~ 6 kb (Boore, 1999; Hikosaka et al., 2010; Hikosaka et al., 2012; Garg et al., 2014). Unlike animal mitochondrial genomes, the apicomplexan mitochondrial genomes encodes three proteins: *coxI*, cytochrome c oxidase subunit III (*coxIII*) and *cob* (Preiser et al., 1996). In the *P. falciparum* mitochondria, protein translation initiates with the formation of an initiation complex involving the ribosome,

initiation factor (IF)3, and IF2 which carries the initiator tRNA-Met. Termination of the peptide chain in mitochondria of apicomplexan parasites occurs due to the action of release factor RF1 that identifies the UAA stop codon in the mRNA from all three mitochondrial ORFs (Habib et al., 2016). tRNAs and other critical translation factors in *P. falciparum* are imported from the cytosol (Rusconi and Cech, 1996; Esseiva et al., 2004). The assembly of the mitochondrial genome of *Babesia duncani* identified a monocistronic linear genome of 5893 bp encoding three genes, *cob (Cytb), coxI and coxIII,* and six large rRNAs (Fig. 4 and Supplementary Table S2). All mitochondrial CDSs start with an ATG codon and end with a TAA codon. Similar to the mitochondrial genomes of other apicomplexan parasites, the *B. duncani* mitochondrial genome lacks tRNAs and translation factors, suggesting that they are encoded by the nuclear genome and imported into the mitochondria from the cytosol.

Most apicomplexan mitochondrial genomes contain a terminal inverted repeat sequence of approximately 440-450 bp (Hikosaka et al., 2010). However, in *B. duncani*, a TIR region of 48 bp was located at opposite ends of the genome (Fig. 4). Given that TIR can often lead to recombination of the genome, their position and sequence relative to the rest of the genome was validated using long-range PCRs and Sanger sequencing. A short CDS was found upstream of the 5' TIR sequence. A novel mitochondrial genome structure of *B. microti* and *B. rodhaini* was reported by (Hikosaka et al., 2012) with a dual flip-flop inversion system that generates four distinct linear genome structures. Since only one pair of inverted repeats was found at the terminal ends of the linear mitochondrial genome of *B. duncani* and no inverted repeats were found inside the molecule, we believe that the mitochondrial genome configuration in *B. microti* does not exist in *B. duncani*. Lastly, *cob* and *coxIII* are encoded on the forward strand while the rest of the features are on the reverse strand.

299

300 *3.5. Phylogenetic analysis reveals that B. duncani defines a distinct lineage among*
301 *Apicomplexan*

302 Mitochondrial proteins are routinely used to probe the evolutionary and phylogenetic
303 history of apicomplexan parasites due to their conserved nature (Hikosaka et al., 2011; Lin et al.,
304 2011; He et al., 2014; Alday et al., 2017). In this study, we used a concatenated sequence of *Cob*
305 and *Cox1* to determine the phylogenetic position of *B. duncani* (Table 2) (Hikosaka et al., 2011;
306 Lin et al., 2011). The *coxIII* gene was excluded from the concatenation and analysis because no
307 full-length sequence of the gene from *T. gondii, B. orientalis* or *P. falciparum* could be found in
308 the available genome databases. Additionally, *coxIII* has been identified in the nuclear genome
309 instead of the mitochondrial genome in some species such as *Tetrahymena thermophile* and *T.*
310 *gondii* (He et al., 2014). A neighbor-joining tree showed that *B. duncani* is a defining member of
311 a new clade compared with other apicomplexan parasites (Fig. 5). This finding confirms
312 previous analysis using 18S rRNA (Conrad et al., 2006). The 18S rRNA tree also showed that
313 *Babesia conradae*, a known canine pathogen that causes babesiosis in dogs, may fall in the same
314 clade as *B. duncani.* However, because the *cob* and *coxI* genes of *B. conradae* have not yet been
315 identified, this species was not included in our analysis. It is highly likely that both *B. duncani*
316 and *B. conradae* strains belong to the same distinct lineage recently referred to as "Western
317 *Babesia* group" (Schreeg et al., 2016).

318

319 *3.6. Therapeutic targets in the apicoplast and the mitochondrial genome*

320 A wide range of therapeutic drugs have been previously reported to show efficacy against
321 *Babesia* spp. by specifically targeting the apicoplast or mitochondria (Ralph et al., 2001;

322 AbouLaila et al., 2012). A thorough sequence analysis of the target proteins and rRNAs of these
323 drugs was performed to predict the sensitivity of *B. duncani* to those drugs (Table 3).

324 The apicoplast of *B. duncani* encodes an LSU sequence of 2726 bp. The GTPase
325 associated center of the 50S ribosome subunit can be targeted by thiostrepton, a thiazolyl peptide
326 antibiotic (Clough et al., 1997). Thiostrepton binds within a cleft between the 43rd and 44th
327 helices, and results in the perturbation of the binding of the elongation factor to ribosomes
328 (McConkey et al., 1997; Gupta et al., 2014). In the presence of the nucleotide adenine at position
329 1067 in the LSU gene of *P. falciparum* and *Escherichia coli*, thiostrepton is able to bind with a
330 high affinity (Clough et al., 1997). When this nucleotide is changed to a uracil or a guanine, 14%
331 and 35% of wild type *Plasmodium* showed reduced binding to thiostrepton, respectively (Table
332 3) (Clough et al., 1997). Sequence alignment revealed that nucleotide A-1067 in the LSU gene of
333 *P. falciparum* (Edgar, 2004) corresponds to nucleotide A-974 in the LSU gene of *B. duncani*,
334 suggesting that thiostrepton could inhibit the growth of *B. duncani*. It is predicted, however, that
335 the nucleotide sequence of 23S rRNA also has a crucial role in the binding of thiostrepton which
336 explains why a mutation at site 1067 of the apicoplast gene does not confer 100% resistance
337 (Clough et al., 1997). An alternative mode of action of thiostrepton was proposed by Aminake
338 and colleagues (Aminake et al., 2011). In their study, the compound was suggested to target the
339 proteasome, leading to rapid elimination of parasites prior to DNA replication (Aminake et al.,
340 2011). Whether thiostrepton has activity against *B. duncani* and whether it targets one or
341 multiple targets remain to be elucidated. Another LSU targeting drug is clindamycin, a
342 lincosamide that most protozoan parasites are uniquely sensitive to due to the presence of an
343 apicoplast. Clindamycin blocks the transpeptidation reaction of the apicoplast (Camps et al.,
344 2002). A point mutation in domain V of the apicoplast rRNA alters its predicted binding site

345 (Camps et al., 2002). As demonstrated in *T. gondii*, clones with strong and stable clindamycin
346 resistance displayed a uracil at position 1857 of the *lsu* gene instead of a guanine (Table 3)
347 (Camps et al., 2002). We can infer that *B. duncani* would also be sensitive to clindamycin since a
348 guanine residue is present at the homologous position (1877). The binding site of
349 chloramphenicol overlaps with the binding site of clindamycin (Gupta et al., 2014), suggesting
350 that this drug may also be considered as a potential inhibitor of *B. duncani*.

351 Azithromycin (AZ) is a broad-spectrum antibiotic that acts as an apicoplast-targeting
352 drug. Presently on the recommended list of drugs to treat human babesiosis, it inhibits protein
353 synthesis in the apicoplast by blocking the exit tunnel of the 50S ribosome polypeptide (Krause
354 et al., 2000; Sidhu et al., 2007). A mutation in the highly conserved ribosomal-protein rpl4 of *P.*
355 *falciparum* (*pfrpl4*) resulted in AZ-resistant lines (Sidhu et al., 2007). All strains were found to
356 harbor a U438C (G76V) mutation (Sidhu et al., 2007). Upon close inspection of the RPL4
357 protein of *B. duncani* and alignment with *Pfrpl4*, a conserved glycine residue at position 68 was
358 found, suggesting that *B. duncani* may also be sensitive to AZ (Table 3). Similarly, a conversion
359 of a glycine residue to aspartic acid at position 91 in the *pfrpl4* gene is also associated with a 57-
360 fold loss of sensitivity to AZ (Wilson et al., 2015). The corresponding region (position 83) in the
361 *B. duncani Bdrpl4* also has a glycine residue. AZ has also been shown to target other apicoplast-
362 encoding proteins such as the large ribosomal subunit (23S) and the L22 protein. AZ resistance
363 has also been associated with mutation in the LSU rRNA. In *P. falciparum*, C to T mutation at
364 base 2409 of the apicoplast 23S rRNA resulted in significant resistance (>10-fold) to prolonged
365 exposure to AZ (Goodman et al., 2013). Analysis of the *B. duncani* LSU rRNA showed the
366 presence of a cytosine residue at the equivalent base 2424. Other potential targets of

azithromycin described in *E. coli* include the L22. However no *rpl22* was found in the *B. duncani* apicoplast genome (Gupta et al., 2014).

Our analysis also suggests that rifampicin, a potent inhibitor of the β subunit of the RNA polymerases, encoded by the *rpoB* gene, may be considered as a possible inhibitor of *B. duncani*. In bacteria, multiple mutations in an 81 bp hotspot region, called the "rifampicin resistance determining region", alter sensitivity to the drug (Aubry-Damon et al., 1998). Moreover, trials with rifampicin against *Plasmodium vivax* infected patients showed that the drug, when given alone, lowered the fever and parasitemia but was not curative (Pukrittayakamee et al., 1994). The use of rifampicin in combination with other drugs also showed poor and slow therapeutic responses (Pukrittayakamee et al., 1994). Although rifampicin may act by inhibiting apicoplast transcription, a recent study in *P. falciparum* has shown that addition of isopyropentyl pyrophosphosphate (Berlin et al., 2015) does not rescue parasite inhibition (Uddin et al., 2018), suggesting that an alternative mode of inhibition may exist in these parasites and could take place outside the apicoplast.

Redox regulation in the mitochondria are central to energy metabolism of apicomplexan parasites (Kehr et al., 2010). Compounds that impede this function include atovaquone (ATV) and endochin-like quionlones (ELQ). ATV, a drug used clinically for treatment of human babesiosis, acts on the Q_o site of the cytochrome bc_1 complex in the parasite's mitochondrial membrane (Kessl et al., 2003; Birth et al., 2014). Similarly, ELQs inhibit the same complex as ATV but at the Q_i site (Stickles et al., 2015). Both compounds have been shown to be effective against apicomplexan parasites but when used alone, resistance rapidly emerges and results in treatment failure (Doggett et al., 2012; Stickles et al., 2015). In combination however, ATV and ELQ act synergistically, a direct result of their effect on the same enzyme complex (Lawres et

al., 2016). Together, these two drugs have been shown to eliminate *B. microti* parasites completely from immunosuppressed SCID mice, with a low dose of up to 5mg/kg, and prevent recrudescence of the parasite (Lawres et al., 2016). To predict the possible efficacy of these two drugs on *B. duncani*, we analyzed the sequence of the *cob* gene and identified residues associated with sensitivity to ATV and ELQs (Table 3). Although a phenylalanine exists at position 244 instead of an alanine in *B. duncani*, a similar residue is also found in *Saccharomyces cerevisiae*, which is also sensitive to specific classes of ELQ analogs (Stickles et al., 2015). Based on these analyses we predict that a combination of ATV and an endochin-like quinolone is likely to be effective against *B. duncani*.

In summary, our sequencing, assembly and annotation of the apicoplast and mitochondrial genomes of *B. duncani* revealed the unique nature of this parasite with respect to other *Babesia* spp. Furthermore, we show that *B. duncani* defines a new lineage in the apicomplexan phylum. Based on our analysis of the sequences of the apicoplast and mitochondrial genomes of *B. duncani*, we predict that this parasite may exhibit sensitivity to a wide range of antiparasitic drugs including thiostrepton, clindamycin, AZ, ATV and ELQs. Our analysis also showed that a large number of hypothetical proteins are encoded by the apicoplast and mitochondrial genes. Analysis of their functions could open new avenues for future studies aiming to characterize *B. duncani* biology, pathogenesis and therapy.

Ackowledgements

We thank Dr. Robert Molestina (BEI Resources, USA) for providing the *B. duncani* WA1 isolate. We thank members of the Yale Center for Genome Analysis, USA for help with Pac Bio and Illumina sequencing. Assistance with the analysis of the apicoplast and the mitochondrial genomes was provided by Dr. Emmanuel Cornillot. AV's research was supported by the Science, Technology and Research Scholars Program collectively funded by the Howard Hughes Medical Institute, USA and Yale College Dean's Office, USA. C.B.M.'s research is supported by grants from the National Institutes of Health, USA (AI097218, GM110506, AI123321 and R43AI136118) and the Bill and Melinda Gates Foundation, USA, grants (OPP1021571).

References

AbouLaila, M., Munkhjargal, T., Sivakumar, T., Ueno, A., Nakano, Y., Yokoyama, M., Yoshinari, T., Nagano, D., Katayama, K., El-Bahy, N., 2012. Apicoplast-targeting antibacterials inhibit the growth of *Babesia* parasites. Antimicrob Agents Chemother 56, 3196-3206.

Alday, P.H., Bruzual, I., Nilsen, A., Pou, S., Winter, R., Ben Mamoun, C., Riscoe, M.K., Doggett, J.S., 2017. Genetic Evidence for Cytochrome b Qi Site Inhibition by 4(1H)-Quinolone-3-Diarylethers and Antimycin in *Toxoplasma gondii*. Antimicrob Agents Chemother 61, pii: e01866-01816.

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

Aminake, M.N., Schoof, S., Sologub, L., Leubner, M., Kirschner, M., Arndt, H.D., Pradel, G., 2011. Thiostrepton and derivatives exhibit antimalarial and gametocytocidal activity by dually targeting parasite proteasome and apicoplast. Antimicrob Agents Chemother 55, 1338-1348.

Aubry-Damon, H., Soussy, C.-J., Courvalin, P., 1998. Characterization of mutations in therpob gene that confer rifampin resistance in *Staphylococcus aureus*. Antimicrob Agents Chemother 42, 2590-2594.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A.V., Sirotkin, A.V., Vyahhi, N., Tesler, G., Alekseyev, M.A., Pevzner, P.A., 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19, 455-477.

441 Berlin, K., Koren, S., Chin, C.S., Drake, J.P., Landolin, J.M., Phillippy, A.M., 2015. Assembling
442 large genomes with single-molecule sequencing and locality-sensitive hashing. Nat
443 Biotechnol 33, 623-630.
444 Birth, D., Kao, W.-c., Hunte, C., 2014. Structural analysis of mitochondrial cytochrome bc1
445 complex with atovaquone bound reveals the molecular basis of antimalarial drug action.
446 Malar J 13, P103.
447 Boore, J.L., 1999. Animal mitochondrial genomes. Nucleic Acids Res 27, 1767-1780.
448 Brayton, K.A., Lau, A.O., Herndon, D.R., Hannick, L., Kappmeyer, L.S., Berens, S.J., Bidwell,
449 S.L., Brown, W.C., Crabtree, J., Fadrosh, D., 2007. Genome sequence of *Babesia bovis*
450 and comparative analysis of apicomplexan hemoprotozoa. PLoS Pathog 3, e148.
451 Bredt, A.B., Weinstein, W.M., Cohen, S., 1981. Treatment of babesiosis in asplenic patients.
452 JAMA 245, 1938-1939.
453 Burgdorfer, W., Barbour, A.G., Hayes, S.F., Benach, J.L., Grunwaldt, E., Davis, J.P., 1982.
454 Lyme disease-a tick-borne spirochetosis? Science 216, 1317-1319.
455 Camps, M., Arrizabalaga, G., Boothroyd, J., 2002. An rRNA mutation identifies the apicoplast
456 as the target for clindamycin in *Toxoplasma gondii*. Mol Microbiol 43, 1309-1318.
457 Chaisson, M.J., Tesler, G., 2012. Mapping single molecule sequencing reads using basic local
458 alignment with successive refinement (BLASR): application and theory. BMC Bioinform
459 13, 238.
460 Chaubey, S., Kumar, A., Singh, D., Habib, S., 2005. The apicoplast of *Plasmodium falciparum* is
461 translationally active. Mol Microbiol 56, 81-89.
462 Clough, B., Strath, M., Preiser, P., Denny, P., Wilson, I.R., 1997. Thiostrepton binds to malarial
463 plastid rRNA. FEBS letters 406, 123-125.

464 Conrad, P.A., Kjemtrup, A.M., Carreno, R.A., Thomford, J., Wainwright, K., Eberhard, M.,
465 Quick, R., Telford Iii, S.R., Herwaldt, B.L., 2006. Description of *Babesia duncani* n.
466 sp.(Apicomplexa: Babesiidae) from humans and its differentiation from other piroplasms.
467 Int J Parasitol 36, 779-789.
468 Cornillot, E., Hadj-Kaddour, K., Dassouli, A., Noel, B., Ranwez, V., Vacherie, B., Augagneur,
469 Y., Bres, V., Duclos, A., Randazzo, S., 2012. Sequencing of the smallest Apicomplexan
470 genome from the human pathogen *Babesia microti*. Nucleic Acids Res 40, 9102-9114.
471 Cullen, J.M., Levine, J.F., 1987. Pathology of experimental *Babesia microti* infection in the
472 Syrian hamster. Lab Anim Sci 37, 640-643.
473 Dao, A.H., Eberhard, M.L., 1996. Pathology of acute fatal babesiosis in hamsters experimentally
474 infected with the WA-1 strain of *Babesia*. Lab Investigation 74, 853-859.
475 Doggett, J.S., Nilsen, A., Forquer, I., Wegmann, K.W., Jones-Brando, L., Yolken, R.H., Bordón,
476 C., Charman, S.A., Katneni, K., Schultz, T., 2012. Endochin-like quinolones are highly
477 efficacious against acute and latent experimental toxoplasmosis. Proc Natl Acad Sci U S
478 A 109, 15936-15941.
479 Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high
480 throughput. Nucleic Acids Res 32, 1792-1797.
481 Esseiva, A.C., Naguleswaran, A., Hemphill, A., Schneider, A., 2004. Mitochondrial tRNA
482 import in *Toxoplasma gondii*. J Biol Chem 279, 42363-42368.
483 Feagin, J.E., 2000. Mitochondrial genome diversity in parasites. Int J Parasitol 30, 371-390.
484 Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the bootstrap.
485 Evolution, 783-791.

Fichera, M.E., Roos, D.S., 1997. A plastid organelle as a drug target in apicomplexan parasites. Nature 390, 407-409.

Foth, B.J., McFadden, G.I., 2003. The apicoplast: a plastid in *Plasmodium falciparum* and other Apicomplexan parasites. Int Review Cytol 224, 57-110.

Frederick, R.L., Shaw, J.M., 2007. Moving mitochondria: establishing distribution of an essential organelle. Traffic 8, 1668-1675.

Gardner, M.J., Williamson, D.H., Wilson, R.J., 1991. A circular DNA in malaria parasites encodes an RNA polymerase like that of prokaryotes and chloroplasts. Mol Biochem Parasitol 44, 115-123.

Garg, A., Stein, A., Zhao, W., Dwivedi, A., Frutos, R., Cornillot, E., Ben Mamoun, C., 2014. Sequence and annotation of the apicoplast genome of the human pathogen *Babesia microti*. PLoS One 9, e107939.

Goodman, C.D., Useglio, M., Peiru, S., Labadie, G.R., McFadden, G.I., Rodriguez, E., Gramajo, H., 2013. Chemobiosynthesis of new antimalarial macrolides. Antimicrob Agents Chemother 57, 907-913.

Grant, J.R., Stothard, P., 2008. The CGView Server: a comparative genomics tool for circular genomes. Nucleic Acids Res 36, W181-W184.

Gupta, A., Shah, P., Haider, A., Gupta, K., Siddiqi, M.I., Ralph, S.A., Habib, S., 2014. Reduced ribosomes of the apicoplast and mitochondrion of *Plasmodium* spp. and predicted interactions with antibiotics. Open Biol 4, 140045.

Habib, S., Vaishya, S., Gupta, K., 2016. Translation in Organelles of Apicomplexan Parasites. Trends Parasitol 32, 939-952.

508 Haider, A., Allen, S.M., Jackson, K.E., Ralph, S.A., Habib, S., 2015. Targeting and function of
509 proteins mediating translation initiation in organelles of *Plasmodium falciparum*. Mol
510 Microbiol 96, 796-814.
511 He, L., Zhang, Y., Zhang, Q.-L., Zhang, W.-J., Feng, H.-H., Khan, M.K., Hu, M., Zhou, Y.-Q.,
512 Zhao, J.-L., 2014. Mitochondrial genome of *Babesia orientalis*, apicomplexan parasite of
513 water buffalo (*Bubalus babalis, Linnaeus*, 1758) endemic in China. Parasit Vectors 7, 82.
514 Hikosaka, K., Nakai, Y., Watanabe, Y.-i., Tachibana, S.-I., Arisue, N., Palacpac, N.M.Q.,
515 Toyama, T., Honma, H., Horii, T., Kita, K., 2011. Concatenated mitochondrial DNA of
516 the coccidian parasite *Eimeria tenella*. Mitochondrion 11, 273-278.
517 Hikosaka, K., Tsuji, N., Watanabe, Y.-i., Kishine, H., Horii, T., Igarashi, I., Kita, K., Tanabe, K.,
518 2012. Novel type of linear mitochondrial genomes with dual flip-flop inversion system in
519 apicomplexan parasites, *Babesia microti* and *Babesia rodhaini*. BMC Genomics 13, 622.
520 Hikosaka, K., Watanabe, Y.-i., Tsuji, N., Kita, K., Kishine, H., Arisue, N., Palacpac, N.M.Q.,
521 Kawazu, S.-i., Sawai, H., Horii, T., 2010. Divergence of the mitochondrial genome
522 structure in the apicomplexan parasites, *Babesia* and *Theileria*. Mol Biol Evol 27, 1107-
523 1116.
524 Huang, Y., He, L., Hu, J., He, P., He, J., Yu, L., Malobi, N., Zhou, Y., Shen, B., Zhao, J., 2015.
525 Characterization and annotation of *Babesia orientalis* apicoplast genome. Parasit Vectors
526 8, 543.
527 Janouskovec, J., Horak, A., Obornik, M., Lukes, J., Keeling, P.J., 2010. A common red algal
528 origin of the apicomplexan, dinoflagellate, and heterokont plastids. Proc Natl Acad Sci U
529 S A 107, 10949-10954.

Jomaa, H., Wiesner, J., Sanderbrand, S., Altincicek, B., Weidemeyer, C., Hintz, M., Turbachova, I., Eberl, M., Zeidler, J., Lichtenthaler, H.K., Soldati, D., Beck, E., 1999. Inhibitors of the nonmevalonate pathway of isoprenoid biosynthesis as antimalarial drugs. Science 285, 1573-1576.

Jones, D.T., Taylor, W.R., Thornton, J.M., 1992. The rapid generation of mutation data matrices from protein sequences. Computer applications in the biosciences: CABIOS 8, 275-282.

Kaczanowski, S., Sajid, M., Reece, S.E., 2011. Evolution of apoptosis-like programmed cell death in unicellular protozoan parasites. Parasit Vectors 4, 44.

Kehr, S., Sturm, N., Rahlfs, S., Przyborski, J.M., Becker, K., 2010. Compartmentation of redox metabolism in malaria parasites. PLoS Pathog 6, e1001242.

Kessl, J.J., Lange, B.B., Merbitz-Zahradnik, T., Zwicker, K., Hill, P., Meunier, B., Pálsdóttir, H., Hunte, C., Meshnick, S., Trumpower, B.L., 2003. Molecular basis for atovaquone binding to the cytochrome bc1 complex. J Biol Chem 278, 31312-31318.

Kjemtrup, A., Conrad, P.A., 2000. Human babesiosis: an emerging tick-borne disease. Int J Parasitol 30, 1323-1337.

Krause, P.J., Lepore, T., Sikand, V.K., Gadbaw Jr, J., Burke, G., Telford, S.R., Brassard, P., Pearl, D., Azlanzadeh, J., Christianson, D., 2000. Atovaquone and azithromycin for the treatment of babesiosis. N Engl J Med 343, 1454-1458.

Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Mol Biol Evol, msw054.

Lawres, L.A., Garg, A., Kumar, V., Bruzual, I., Forquer, I.P., Renard, I., Virji, A.Z., Boulard, P., Rodriguez, E.X., Allen, A.J., Pou, S., Wegmann, K.W., Winter, R.W., Nilsen, A., Mao, J., Preston, D.A., Belperron, A.A., Bockenstedt, L.K., Hinrichs, D.J., Riscoe, M.K.,

553 Doggett, J.S., Ben Mamoun, C., 2016. Radical cure of experimental babesiosis in
554 immunodeficient mice using a combination of an endochin-like quinolone and
555 atovaquone. J Exp Med 213, 1307-1318.
556 Lemieux, J., 2015. Genomic Analysis of Evolution in *Plasmodium falciparum* and *Babesia*
557 *microti*.
558 Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler
559 transform. Bioinformatics 25, 1754-1760.
560 Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G.,
561 Durbin, R., Genome Project Data Processing, S., 2009. The Sequence Alignment/Map
562 format and SAMtools. Bioinformatics 25, 2078-2079.
563 Lim, L., McFadden, G.I., 2010. The evolution, metabolism and functions of the apicoplast.
564 Philos Trans R Soc Lond B Biol Sci 365, 749-763.
565 Lin, R.-Q., Qiu, L.-L., Liu, G.-H., Wu, X.-Y., Weng, Y.-B., Xie, W.-Q., Hou, J., Pan, H., Yuan,
566 Z.-G., Zou, F.-C., 2011. Characterization of the complete mitochondrial genomes of five
567 *Eimeria* species from domestic chickens. Gene 480, 28-33.
568 Lowe, T.M., Chan, P.P., 2016. tRNAscan-SE On-line: integrating search and context for analysis
569 of transfer RNA genes. Nucleic Acids Res 44, W54-W57.
570 Lowe, T.M., Eddy, S.R., 1997. tRNAscan-SE: a program for improved detection of transfer
571 RNA genes in genomic sequence. Nucleic Acids Res 25, 955-964.
572 Magalhães, P.J., Andreu, A.L., Schon, E.A., 1998. Evidence for the presence of 5S rRNA in
573 mammalian mitochondria. Mol Biol Cell 9, 2375-2382.

McConkey, G.A., Rogers, M.J., McCutchan, T.F., 1997. Inhibition of *Plasmodium falciparum* protein synthesis Targeting the plastid-like organelle with thiostrepton. J Biol Chem 272, 2046-2049.

McFadden, G.I., 2011. The apicoplast. Protoplasma 248, 641-650.

Moro, M.H., David, C.S., Magera, J.M., Wettstein, P.J., Barthold, S.W., Persing, D.H., 1998. Differential Effects of Infection with a *Babesia*-Like Piroplasm, WA1, in Inbred Mice. Infect Immun 66, 492-498.

Preiser, P., Wilson, R., Moore, P., McCready, S., Hajibagheri, M., Blight, K., Strath, M., Williamson, D., 1996. Recombination associated with replication of malarial mitochondrial DNA. EMBO J 15, 684.

Pukrittayakamee, S., Viravan, C., Charoenlarp, P., Yeamput, C., Wilson, R.J., White, N.J., 1994. Antimalarial effects of rifampin in *Plasmodium vivax* malaria. Antimicrob Agents Chemother 38, 511-514.

Quick, R.E., Herwaldt, B.L., Thomford, J.W., Garnett, M.E., Eberhard, M.L., Wilson, M., Spach, D.H., Dickerson, J.W., Telford, S.R., Steingart, K.R., 1993. Babesiosis in Washington State: a new species of *Babesia*? Ann Intern Med 119, 284-290.

Quinlan, A.R., Hall, I.M., 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841-842.

Ralph, S.A., D'Ombrain, M.C., McFadden, G.I., 2001. The apicoplast as an antimalarial drug target. Drug Resist Updat 4, 145-151.

Ralph, S.A., van Dooren, G.G., Waller, R.F., Crawford, M.J., Fraunholz, M.J., Foth, B.J., Tonkin, C.J., Roos, D.S., McFadden, G.I., 2004. Tropical infectious diseases: metabolic

maps and functions of the *Plasmodium falciparum* apicoplast. Nat Rev Microbiol 2, 203-216.

Rusconi, C.P., Cech, T.R., 1996. Mitochondrial import of only one of three nuclear-encoded glutamine tRNAs in *Tetrahymena thermophila*. EMBO J 15, 3286-3295.

Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M.-A., Barrell, B., 2000. Artemis: sequence visualization and annotation. Bioinformatics 16, 944-945.

Saitou, N., Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4, 406-425.

Schattner, P., Brooks, A.N., Lowe, T.M., 2005. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. Nucleic Acids Res 33, W686-W689.

Scholtens, R.G., Braff, E.H., Healey, G.A., Gleason, N., 1968. A case of babesiosis in man in the United States. Am J Trop Med Hyg 17, 810-813.

Schreeg, M.E., Marr, H.S., Tarigo, J.L., Cohn, L.A., Bird, D.M., Scholl, E.H., Levy, M.G., Wiegmann, B.M., Birkenheuer, A.J., 2016. Mitochondrial Genome Sequences and Structures Aid in the Resolution of Piroplasmida phylogeny. PloS one 11, e0165702.

Seeber, F., Soldati-Favre, D., 2010. Metabolic pathways in the apicoplast of apicomplexa. Int Rev Cell Mol Biol 281, 161-228.

Sidhu, A.B.S., Sun, Q., Nkrumah, L.J., Dunne, M.W., Sacchettini, J.C., Fidock, D.A., 2007. In vitro efficacy, resistance selection, and structural modeling studies implicate the malarial parasite apicoplast as the target of azithromycin. J Biol Chem 282, 2494-2504.

Spielman, A., Wilson, M.L., Levine, J.F., Piesman, J., 1985. Ecology of Ixodes dammini-borne human babesiosis and Lyme disease. Annu Rev Entomol 30, 439-460.

618 Stickles, A.M., De Almeida, M.J., Morrisey, J.M., Sheridan, K.A., Forquer, I.P., Nilsen, A.,
619 Winter, R.W., Burrows, J.N., Fidock, D.A., Vaidya, A.B., 2015. Subtle changes in
620 endochin-like quinolone structure alter the site of inhibition within the cytochrome bc1
621 complex of *Plasmodium falciparum*. Antimicrob Agents Chemother 59, 1977-1982.
622 Stothard, P., Wishart, D.S., 2005. Circular genome visualization and exploration using CGView.
623 Bioinformatics 21, 537-539.
624 Taylor-Brown, E., Hurd, H., 2013. The first suicides: a legacy inherited by parasitic protozoans
625 from prokaryote ancestors. Parasit Vectors 6, 108.
626 Uddin, T., McFadden, G.I., Goodman, C.D., 2018. Validation of Putative Apicoplast-Targeting
627 Drugs Using a Chemical Supplementation Assay in Cultured Human Malaria Parasites.
628 Antimicrob Agents Chemother 62.
629 Vaishya, S., Kumar, V., Gupta, A., Siddiqi, M.I., Habib, S., 2016. Polypeptide release factors
630 and stop codon recognition in the apicoplast and mitochondrion of *Plasmodium*
631 *falciparum*. Mol Microbiol 100, 1080-1095.
632 Vannier, E., Krause, P.J., 2012. Human babesiosis. N Engl J Med 366, 2397-2407.
633 Vannier, E.G., Diuk-Wasser, M.A., Ben Mamoun, C., Krause, P.J., 2015. Babesiosis. Infect Dis
634 Clin North Am 29, 357-370.
635 Wilson, D.W., Goodman, C.D., Sleebs, B.E., Weiss, G.E., de Jong, N.W., Angrisano, F., Langer,
636 C., Baum, J., Crabb, B.S., Gilson, P.R., McFadden, G.I., Beeson, J.G., 2015. Macrolides
637 rapidly inhibit red blood cell invasion by the human malaria parasite, *Plasmodium*
638 *falciparum*. BMC Biol 13, 52.

639 Wozniak, E.J., Lowenstine, L.J., Hemmer, R., Robinson, T., Conrad, P.A., 1996. Comparative
640 pathogenesis of human WA1 and *Babesia microti* isolates in a Syrian hamster model. Lab
641 Anim Sci 46, 507-515.
642
643

Figure legends

Fig. 1. Graphical circular map of the apicoplast genome of *Babesia duncani*. The map was designed using the CGView Server - an online platform to generate circular genome maps (Grant and Stothard, 2008). From outside to center: 1) coding sequence (CDS), tRNAs, rRNAs and introns, 2) % G+C, 3) GC skew and 4) base coordinates. *HypA-N* refers to 14 hypothetical protein-encoding genes found in the apicoplast genome of *B. duncani*. The different clusters in *B duncani* were labelled using the identities of each cluster described in *Babesia microti* (Garg et al., 2014). Cluster 1 was identified based on the presence of ribosomal proteins and the EF-Tu elongation factor. Clusters 2 and 3 were labelled based on the presence of the ClpC chaperones and the "RNA Pol cluster", respectively. Cluster 4 consisted of the region with rDNA and a single set of *lsu* and *ssu* genes. Each cluster demonstrates synteny with other apicomplexan parasites as well as the *Chomera* genome (Janouskovec et al., 2010).

Fig. 2. Gene organization of Cluster 1 in the apicoplast genome of *Babesia duncani.* For comparison, Cluster 1 of *Babesia bovis* (T2Bo), *Babesia orientalis* (Wuhan), *Babesia microti* (R1), *Theileria parva* (Mugaga), *Plasmodium falciparum* (3D7), *Toxoplasma gondii* and the chloroplast genome of *Chromera* sp. (CCMP3155) is also provided. The black arrow at the top of the figure indicates the direction of transcription. Light grey boxes indicate highly divergent genes and white boxes correspond to genes restricted to one species.

Fig. 3. Organization of the ribosomal DNA region (Cluster 4 in the apicoplast genome) of *Babesia duncani* alongside the same region in other apicomplexan parasites. The tRNA genes that are present in all apicomplexan genomes are in bold.

668

669 **Fig. 4.** Linear map of the mitochondrial genome of *Babesia duncani*. White boxes represent
670 genes encoding proteins involved in the electron transport chain. Grey boxes indicate rRNA
671 subunits. TIR at the start and end of the genome depicts terminal inverted repeat regions.

672

673 **Fig. 5.** Evolutionary relationships of taxa. The evolutionary history was inferred using the
674 Neighbor-Joining method (Saitou and Nei, 1987). The analysis involved 13 different
675 apicomplexan species. All positions containing gaps and missing data were eliminated.
676 Following concatenation and alignment with MUSCLE of the *cob* and *cox1* genes, a total of 806
677 bases were used to generate the evolutionary tree. The evolutionary distances were computed
678 using the JTT matrix-based method (Jones et al., 1992) and are in the units of the number of
679 amino acid substitutions per site. The optimal tree with the sum of branch length = 2. 82680194
680 is shown. The percentage of replicate trees in which the associated taxa clustered together in the
681 bootstrap test (1000 replicates) are shown next to the branches (Felsenstein, 1985). The tree is
682 drawn to scale, with branch lengths in the same units as those of the evolutionary distances used
683 to infer the phylogenetic tree.

684

685

686

Supplementary figure legend

Supplementary Fig. S1. A schematic representation of the DNA regions surrounding Cluster 1 in *Babesia duncani*, *Babesia microti, Theileria parva* and *Babesia bovis.* A line connecting the two ends of Cluster 1 indicate possible recombination events. The tRNA genes that are present in all apicomplexan genomes are highlighted in bold. *rps4* is the gene coding for the ribosomal protein RPS4 while *clpC1* and *clpC2* are genes coding for the Clp chaperone proteins.

Table 1. A summary of all the features found in the apicoplast genome of *Babesia duncani*

Category	Genes
CDS	38 ORF Genes
Ribosomal Proteins (17)	rps2, 3, 4, 5, 7, 8, 11, 12, 17, 19 rpl2, 4, 5, 6, 14, 16, 36
Transfer RNAs[a] (21)	Lys^{AAA}, Ser^{AGC}, Cys^{UGC}, Tyr^{UAC}, Glu^{GAA}, Asp^{GAC}, Leu^{UUA}, His^{CAC}, Met^{AUG}, Gly^{GGA}, Thr^{ACA}, Arg^{CGU}, Ile^{AUC}, Ala^{GCA}, Val^{GUA}, Leu^{CUA}, Phe^{UUC}, Arg^{AGA}, Asn^{AAC}, Met^{AUG}, Gln^{CAA}
Ribosomal RNAs (2)	1 small subunit (SSU), 1 large subunit (LSU)
RNA Polymerase (4)	RNA Pol 1,2,3,4
Other Proteins (3)	CLpProtease 1, CLpProtease 2, TufA
Hypothetical proteins (14)	HypA-N

[a]tRNAs represented with three letter amino acid code and anticodon

Table 2. GenBank accession numbers of cytochrome b (*cob*) and cytochrome c oxidase subunit I (*coxI*) genes from species included in the evolutionary tree.

Organism	**Cytb/Cob**	**CoxI**
Babesia bigemina	BAI66164	BAI66162
Babesia bovis	BAI66173	BAI66171
Babesia caballi	BAI66167	BAI66165
Babesia canis canis	AGF95352	AGF95350
Babesia duncani	MH107387	MH107387
Babesia gibsoni	BAE94852	BAL72992
Babesia microti	BAM68222	BAM68223
Babesia orientalis	AHB82181	AHB82180
Babesia rodhaini	BAM68237	BAM68238
Plasmodium falciparum	AEK05566	AEK05565
Theileria orientalis	BAI66179	BAI66177
Theileria parva	YP001994287	BAI66174
Toxoplasma gondii	AFQ31674	AFQ31675

Table 3. Sequence-based target identification and predicted drug susceptibility.

Drug	**Target protein**	**Residue that confers resistance**	***Babesia duncani* residues**
Thiostrepton	LSU rRNA (Apicoplast)[a]	A_{1067}→U/G	A_{974}
Clindamycin	LSU rRNA (Apicoplast)[a]	G_{1857}→U	G_{1877}
Azithromycin	Ribosomal protein L4 (Apicoplast)	GLY_{76}→VAL GLY_{91}→ASP	GLY_{68} GLY_{83}
Atovaquone	Qo site of Cytochome bc1 complex (mitochondria)	MET_{138}→ILE	MET_{160}
ELQ 334	Qi site of Cytochome bc1 complex (mitochondria)	ASN_{32}→SER/TYR ALA_{222}→VAL	ASN_{54} PHE_{244}

[a] LSU refers to the large subunit rRNA

Figures

Figure 1.

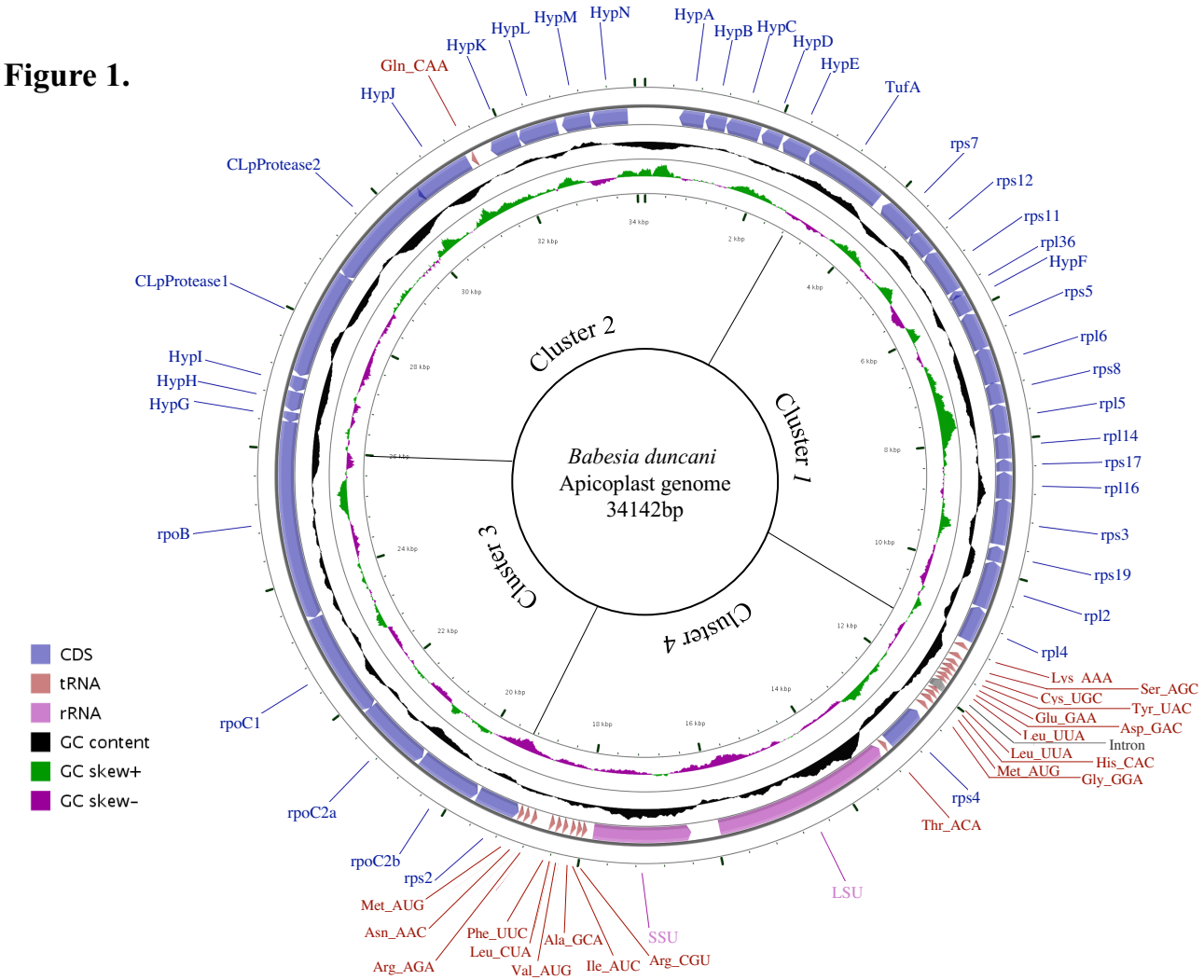


Figure 2.

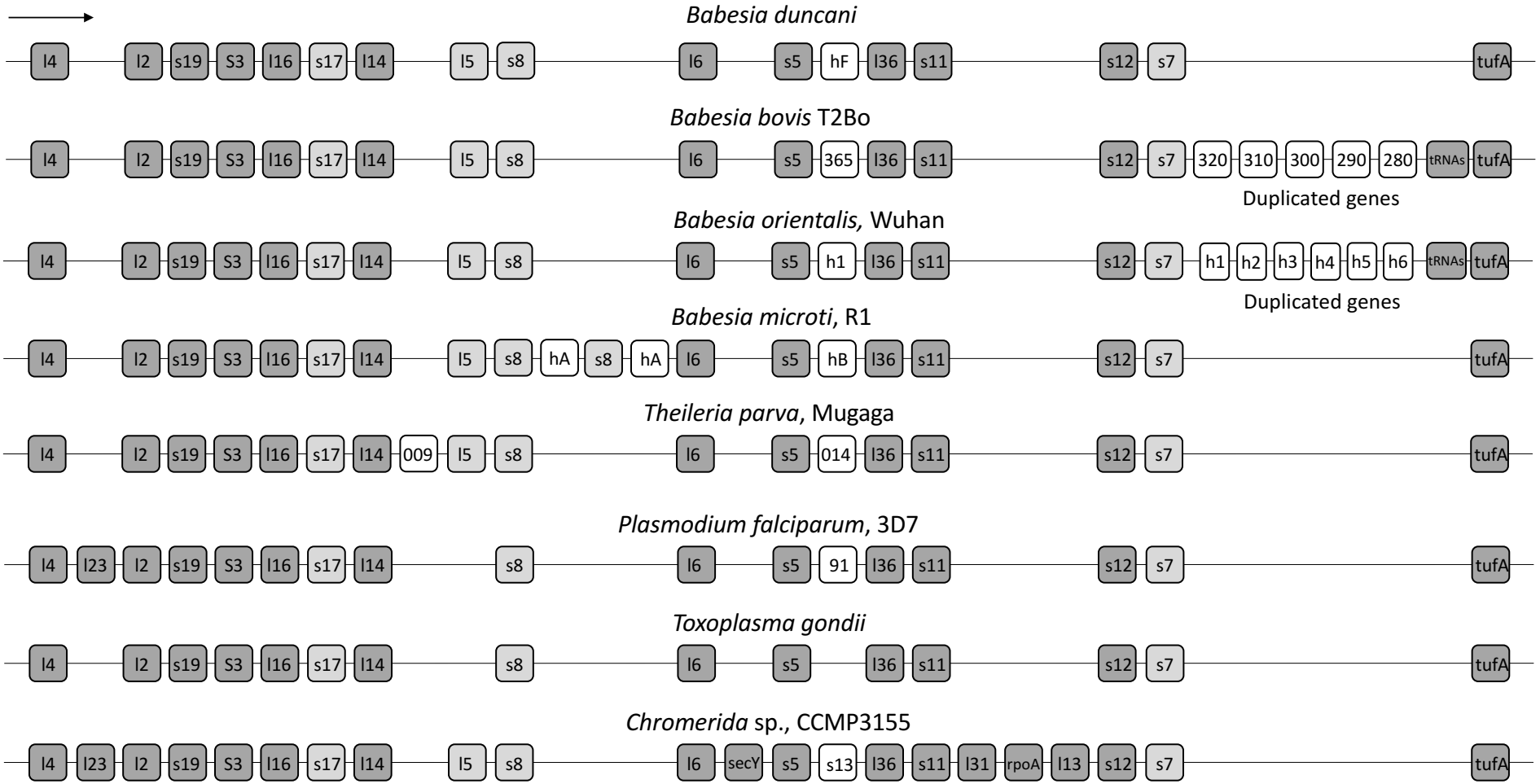

Babesia duncani
l4 l2 s19 S3 l16 s17 l14 l5 s8 l6 s5 hF l36 s11 s12 s7 tufA
Babesia bovis T2Bo
l4 l2 s19 S3 l16 s17 l14 l5 s8 l6 s5 365 l36 s11 s12 s7 320 310 300 290 280 tRNAs tufA
Duplicated genes
Babesia orientalis, Wuhan
l4 l2 s19 S3 l16 s17 l14 l5 s8 l6 s5 h1 l36 s11 s12 s7 h1 h2 h3 h4 h5 h6 tRNAs tufA
Duplicated genes
Babesia microti, R1
l4 l2 s19 S3 l16 s17 l14 l5 s8 hA s8 hA l6 s5 hB l36 s11 s12 s7 tufA
Theileria parva, Mugaga
l4 l2 s19 S3 l16 s17 l14 009 l5 s8 l6 s5 014 l36 s11 s12 s7 tufA
Plasmodium falciparum, 3D7
l4 l23 l2 s19 S3 l16 s17 l14 s8 l6 s5 91 l36 s11 s12 s7 tufA
Toxoplasma gondii
l4 l2 s19 S3 l16 s17 l14 s8 l6 s5 l36 s11 s12 s7 tufA
Chromerida sp., CCMP3155
l4 l23 l2 s19 S3 l16 s17 l14 l5 s8 l6 secY s5 s13 l36 s11 l31 rpoA l13 s12 s7 tufA

Figure 3.

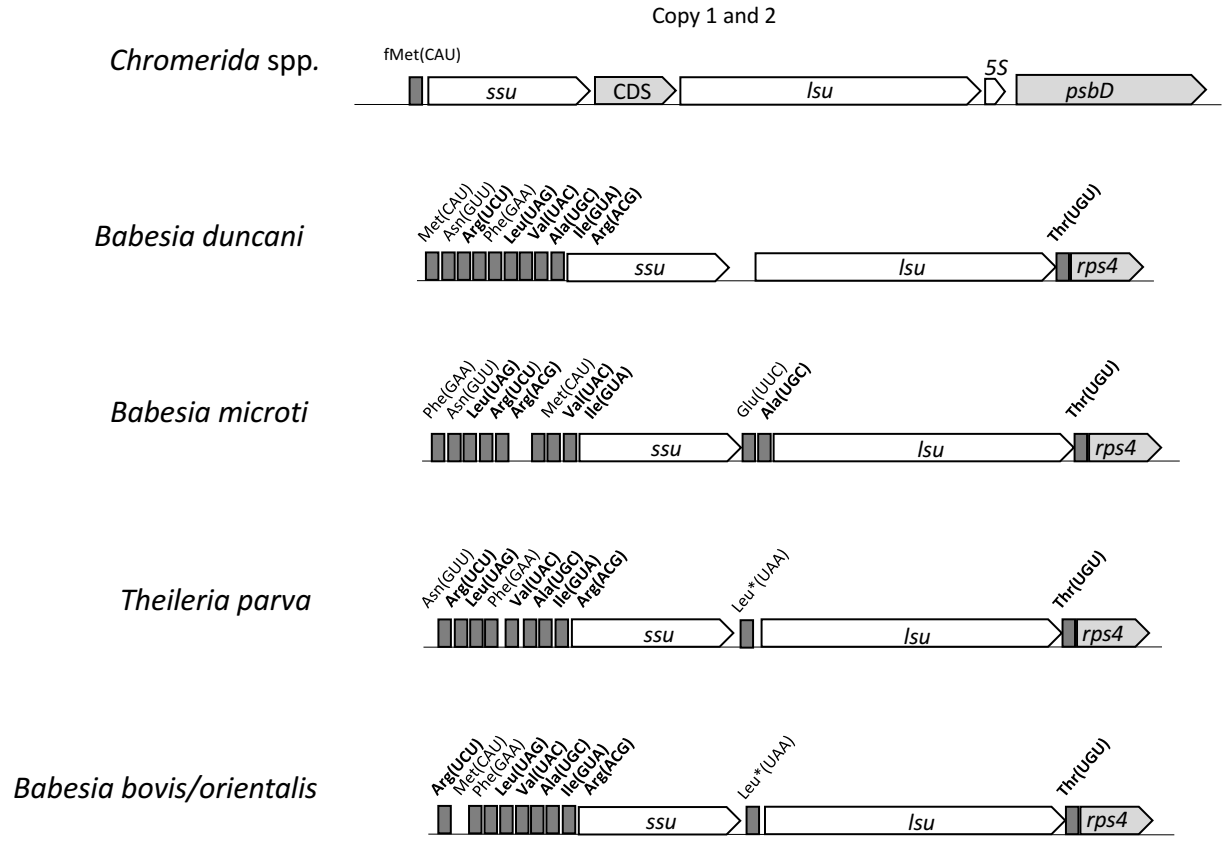
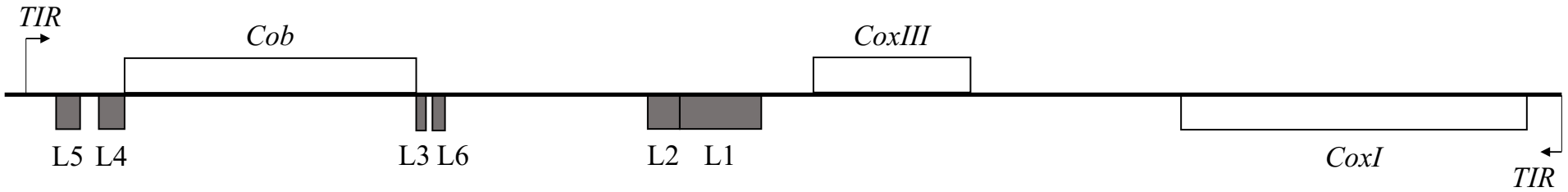
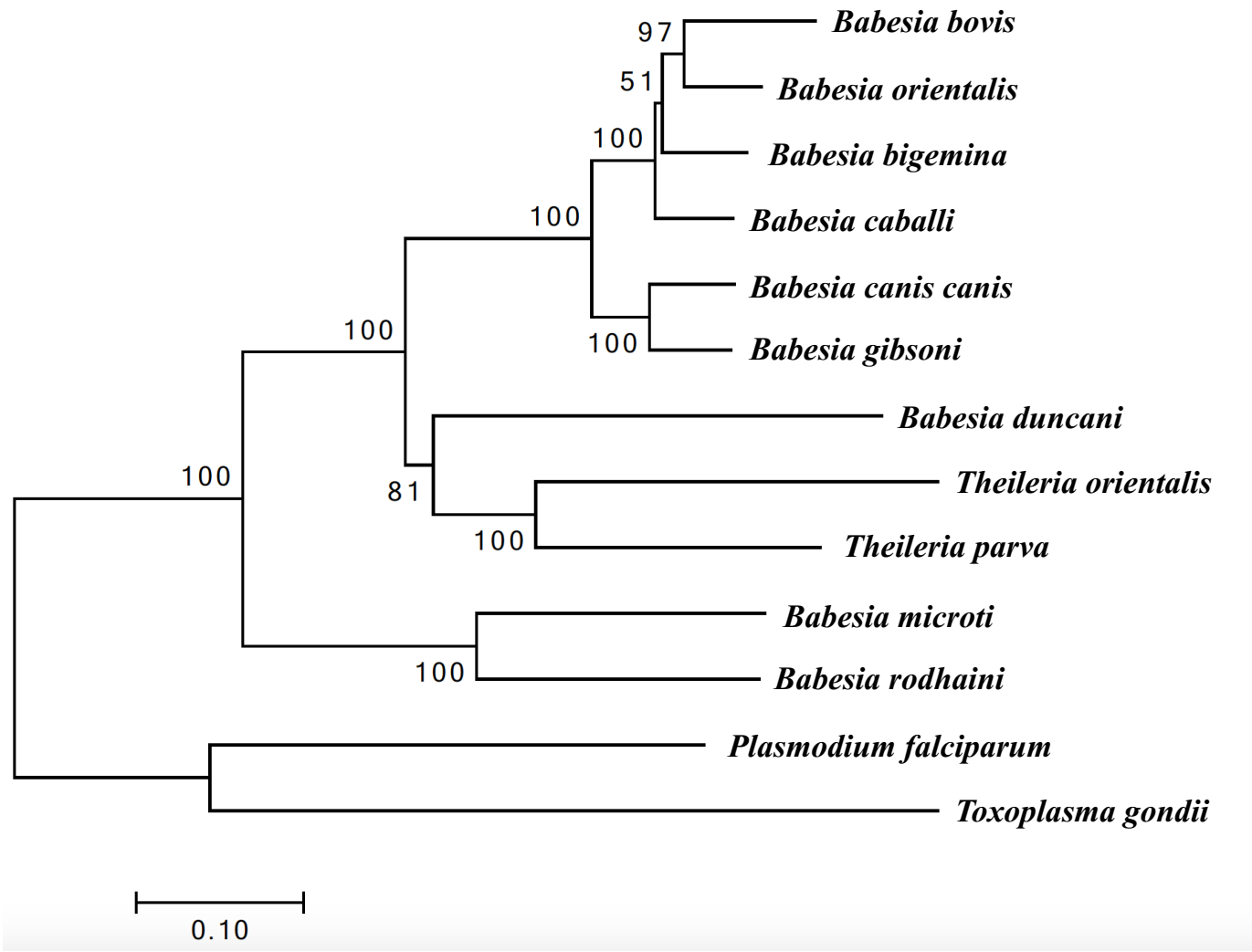

Copy 1 and 2
Chromerida spp.
fMet(CAU)
ssu
CDS
lsu
5S
psbD
Babesia duncani
Met(CAU)
Asn(GUU)
Arg(UCU)
Phe(GAA)
Leu(UAG)
Val(UAC)
Ala(UGC)
Ile(GUA)
Arg(ACG)
Thr(UGU)
ssu
lsu
rps4
Babesia microti
Phe(GAA)
Asn(GUU)
Leu(UAG)
Arg(UCU)
Arg(ACG)
Met(CAU)
Val(UAC)
Ile(GUA)
Glu(UUC)
Ala(UGC)
Thr(UGU)
ssu
lsu
rps4
Theileria parva
Asn(GUU)
Arg(UCU)
Leu(UAG)
Phe(GAA)
Val(UAC)
Ala(UGC)
Ile(GUA)
Arg(ACG)
Leu*(UAA)
Thr(UGU)
ssu
lsu
rps4
Babesia bovis/orientalis
Arg(UCU)
Met(CAU)
Phe(GAA)
Leu(UAG)
Val(UAC)
Ala(UGC)
Ile(GUA)
Arg(ACG)
Leu*(UAA)
Thr(UGU)
ssu
lsu
rps4

Figure 4.



Babesia duncani
Mitochondrial genome
5893 bp

Figure 5.


97
51
100
100
100
100
100
81
100
100
Babesia bovis
Babesia orientalis
Babesia bigemina
Babesia caballi
Babesia canis canis
Babesia gibsoni
Babesia duncani
Theileria orientalis
Theileria parva
Babesia microti
Babesia rodhaini
Plasmodium falciparum
Toxoplasma gondii
0.10