Contents lists available at ScienceDirect

International Journal for Parasitology

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

Insights into the evolution and drug susceptibility of *Babesia duncani* from the sequence of its mitochondrial and apicoplast genomes

Azan Z. Virji a, Jose Thekkiniath a, Wenxiu Ma b, Lauren Lawres a, James Knight c, Andrea Sweid d, Karine Le Roch a, Choukri Ben Mamoun b,c,e

a Department of Internal Medicine, Section of Infectious Diseases, Yale School of Medicine, New Haven, CT 06520, USA
b Department of Statistics, University of California, 900 University Avenue, Riverside, CA 92521, USA
c Department of Genetics, Yale School of Medicine, New Haven, CT 06520, USA
d Department of Biology, San Francisco State University, 1600 Holloway Ave, San Francisco, CA 94132, USA
e Department of Cell Biology & Neuroscience, 900 University Avenue, University of California, Riverside, CA 92521, USA

**A R T I C L E   I N F O**

Article history:
Received 22 January 2018
Received in revised form 4 April 2018
Accepted 3 May 2018
Available online 31 August 2018

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

**A B S T R A C T**

*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 (*Cytb*) 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.

© 2018 Australian Society for Parasitology. Published by Elsevier Ltd. All rights reserved.

1. Introduction

Human babesiosis is a global infectious disease transmitted by ticks, blood transfusion or congenitally (reviewed in Vannier and Krause (2012) and 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 characterised by an initial phase of high
parasitemia, anaemia 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 than 10% 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 (2019) 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 (>10 kb) 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 eccotools (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/Chloropl ast” (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 analysed (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-characterised 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 genome 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 tRNAfMet (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 in P. falciparum is achieved by a single release factor, PfRF2Api, 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-

### Table 1

<table>
<thead>
<tr>
<th>Category</th>
<th>Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDS</td>
<td>38 ORF Genes</td>
</tr>
<tr>
<td>Ribosomal Proteins (17)</td>
<td>rps2, 3, 4, 5, 7, 8, 11, 12, 17, 19</td>
</tr>
<tr>
<td></td>
<td>rpl2, 4, 5, 6, 14, 16, 36</td>
</tr>
<tr>
<td>Transfer RNAsa (21)</td>
<td>LysAAA, LeuAGC, CysAGC, TyrAGA, GluGGA, ArgGGA, IleAUC, AlaGCA, ValGUA, LeuUGA, PheUGC, ArgAUA, AsnAAC, MetAUG, GlyGGA, GlnCAA</td>
</tr>
<tr>
<td>Ribosomal RNAs (2)</td>
<td>1 small subunit (SSU), 1 large subunit (LSU)</td>
</tr>
<tr>
<td>RNA Polymerase (4)</td>
<td>RNA Pol 1,2,3,4</td>
</tr>
<tr>
<td>Other Proteins (3)</td>
<td>ClpProtease 1, ClpProtease 2, TufA</td>
</tr>
<tr>
<td>Hypothetical proteins (14)</td>
<td>HypN</td>
</tr>
</tbody>
</table>

* tRNAs represented with three letter amino acid code and anticodon.
encoded B. duncani ortholog of PRF2Api recognises 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 recognisable 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 then translocated to the apicoplast or by the rff gene encoding 5S rRNA in the chloroplast genome of Chromera spp. (Supplementary Table S1). Four of the 11 tRNAs are conserved in all apicomplexan genomes (Fig. 3). 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 genes encoding ribosomal proteins, one hypothetical protein (HypF) and the elongation factor EF-Tu in an organisation similar to that found in Babesia spp., P. falciparum, Toxoplasma gondii and Chromera 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 HypF (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 species

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 Chromera spp. 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 organisation similar to that found in Babesia spp., P. falciparum, Toxoplasma gondii and Chromera 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 HypF (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 an inversion took place during the evolution of piroplasmida, possibly even resulting in the loss of the supB gene (Foth and McFadden, 2002).

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 organised 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 organisation 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 linear form of the mitochondrial genome of Babesia duncani consists of four RNA polymerase genes (rpoB, rpoC1, rpoC2a and rpoC2b) and the rps2 gene, which encodes an S2 ribosomal protein (Fig. 1). The opposite orientation is found in P. falciparum and T. gondii, suggesting that an inversion took place during the evolution of piroplasmida, possibly even resulting in the loss of the supB gene (Foth and McFadden, 2002).

3.5 Phylogenetic analysis reveals that B. duncani defines a distinct lineage among apicomplexan parasite translation factors in P. falciparum are imported from the cytosol (Rusconi and Ceci, 1996; Esseiva et al., 2004). The assembly of the mitochondrial genome of B. duncani identified a monocistronic linear genome of 5893 bp encoding three genes, cob (Cytb), coxI and coxIII, and six large tRNAs (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 PCR 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.

<table>
<thead>
<tr>
<th>Organism</th>
<th>Cytb/Cob</th>
<th>CoxI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Babesia bigemina</td>
<td>BA66164</td>
<td>BA66162</td>
</tr>
<tr>
<td>Babesia bovis</td>
<td>BA66173</td>
<td>BA66171</td>
</tr>
<tr>
<td>Babesia caballi</td>
<td>BA66167</td>
<td>BA66165</td>
</tr>
<tr>
<td>Babesia canis canis</td>
<td>AGF95352</td>
<td>AGF95350</td>
</tr>
<tr>
<td>Babesia duncani</td>
<td>MB107387</td>
<td>MB107387</td>
</tr>
<tr>
<td>Babesia gibsoni</td>
<td>BA68452</td>
<td>BAL72992</td>
</tr>
<tr>
<td>Babesia microti</td>
<td>BAM68222</td>
<td>BAM68223</td>
</tr>
<tr>
<td>Babesia orientalis</td>
<td>AH882181</td>
<td>AH882180</td>
</tr>
<tr>
<td>Babesia rodhaini</td>
<td>BAM68237</td>
<td>BAM68238</td>
</tr>
<tr>
<td>Plasmodium falciparum</td>
<td>AE005266</td>
<td>AE005265</td>
</tr>
<tr>
<td>Theileria orientalis</td>
<td>BA66179</td>
<td>BA66177</td>
</tr>
<tr>
<td>Theileria parva</td>
<td>YP001949287</td>
<td>BA66174</td>
</tr>
<tr>
<td>Toxoplasma gondii</td>
<td>AFQ31674</td>
<td>AFQ31675</td>
</tr>
</tbody>
</table>

Fig. 4. Linear map of the mitochondrial genome of Babesia duncani. White boxes represent genes encoding proteins involved in the electron transport chain. Grey boxes indicate tRNA subunits. TIR at the start and end of the genome depicts terminal inverted repeat regions.
gene was excluded from the concatenation and analysis because no full-length sequence of the gene from *T. gondii*, *B. orientalis* or *P. falciparum* could be found in the available genome databases. Additionally, *coxIII* has been identified in the nuclear genome instead of the mitochondrial genome in some species such as *Tetrahymena thermophile* and *T. gondii* (He et al., 2014). A neighbor-joining tree showed that *B. duncani* is a defining member of a new clade compared with other apicomplexan parasites (Fig. 5). This finding confirms previous analysis using 18S rRNA (Conrad et al., 2006). The 18S rRNA tree also showed that *Babesia conradae*, a known canine pathogen that causes babesiosis in dogs, may fall in the same clade as *B. duncani*. However, because the *cob* and *coxI* genes of *B. conradae* have not yet been identified, this species was not included in our analysis. It is highly likely that both *B. duncani* and *B. conradae* strains belong to the same distinct lineage recently referred to as “Western Babesia group” (Schreeg et al., 2016).

### 3.6. Therapeutic targets in the apicoplast and the mitochondrial genome

A wide range of therapeutic drugs have been previously reported to show efficacy against *Babesia* spp. by specifically targeting the apicoplast or mitochondria (Ralph et al., 2001; AbouLaila et al., 2012). A thorough sequence analysis of the predicted target proteins and rRNA was performed to predict the sensitivity of *B. duncani* to those drugs (Table 3).

The apicoplast of *B. duncani* encodes an LSU sequence of 2726 bp. The GTPase associated center of the 50S ribosome subunit can be targeted by thiostrepton, a thiazolyl peptide antibiotic (Clough et al., 1997). Thiostrepton binds within a cleft between the 43rd and 44th helices, and results in the perturbation of the binding of the elongation factor to ribosomes (McConkey et al., 1997; Gupta et al., 2014). In the presence of the nucleotide adenine at position 1067 in the LSU gene of *P. falciparum* and *Escherichia coli*, thiostrepton is able to bind with a high affinity (Clough et al., 1997). When this nucleotide is changed to a uracil or a guanine, 14% and 35% of wild type *Plasmodium* showed reduced binding to thiostrepton, respectively (Clough et al., 1997). Sequence alignment revealed that nucleotide A-1067 in the LSU gene of *P. falciparum* (Edgar, 2004) corresponds to nucleotide A-974 in the LSU gene of *B. duncani*, suggesting that thiostrepton could inhibit the growth of *B. duncani*. It is predicted, however, that the nucleotide sequence of 23S rRNA also has a crucial role in the binding of thiostrepton which explains why a mutation at site

**Table 3**

<table>
<thead>
<tr>
<th>Drug</th>
<th>Target protein</th>
<th>Residue that confers resistance</th>
<th>Babesia duncani residues</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thiostrepton</td>
<td>LSU rRNA (Apicoplast)*</td>
<td>A&lt;sub&gt;1067&lt;/sub&gt; → U/G</td>
<td>A&lt;sub&gt;974&lt;/sub&gt;</td>
</tr>
<tr>
<td>Clindamycin</td>
<td>LSU rRNA (Apicoplast)*</td>
<td>G&lt;sub&gt;1857&lt;/sub&gt; → U</td>
<td>G&lt;sub&gt;1877&lt;/sub&gt;</td>
</tr>
<tr>
<td>Azithromycin</td>
<td>Ribosomal protein L4 (Apicoplast)</td>
<td>GLY&lt;sub&gt;76&lt;/sub&gt; → VAL</td>
<td>GLY&lt;sub&gt;58&lt;/sub&gt;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GLY&lt;sub&gt;91&lt;/sub&gt; → ASP</td>
<td>GLY&lt;sub&gt;61&lt;/sub&gt;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MET&lt;sub&gt;118&lt;/sub&gt; → ILE</td>
<td>MET&lt;sub&gt;60&lt;/sub&gt;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ASN&lt;sub&gt;32&lt;/sub&gt; → SER/TYR</td>
<td>ASN&lt;sub&gt;44&lt;/sub&gt;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ALA&lt;sub&gt;222&lt;/sub&gt; → VAL</td>
<td>PHE&lt;sub&gt;244&lt;/sub&gt;</td>
</tr>
</tbody>
</table>

* LSU refers to the large subunit rRNA.
1067 of the apicoplast gene does not confer 100% resistance (Clough et al., 1997). An alternative mode of action of thiostrepton was proposed by Aminake and colleagues (Aminake et al., 2011). In their study, the compound was suggested to target the proteasome, leading to rapid elimination of parasites prior to DNA replication (Aminake et al., 2011). Whether thiostrepton has activity against B. duncani and whether it targets one or multiple targets remain to be elucidated. Another LSU targeting drug is clindamycin, a lincosamide that most protozoan parasites are uniquely sensitive to due to the presence of an apicoplast. Clindamycin blocks the transpeptidation reaction of the apicoplast (Camps et al., 2002). A point mutation in domain V of the apicoplast rRNA alters its predicted binding site (Camps et al., 2002). As demonstrated in T. gondii, clones with strong and stable clindamycin resistance displayed a uracil at position 1857 of the LSU gene instead of a guanine (Table 3) (Camps et al., 2002). We can infer that B. duncani would also be sensitive to clindamycin since a guanine residue is present at the homologous position (1877). The binding site of chloramphenicol overlaps with the binding site of clindamycin (Gupta et al., 2014), suggesting that this drug may also be considered as a potential inhibitor of B. duncani.

Azithromycin (AZ) is a broad-spectrum antibiotic that acts as an apicoplast-targeting drug. Presently on the recommended list of drugs to treat human babesiosis, it inhibits protein synthesis in the apicoplast by blocking the exit tunnel of the 50S ribosome (Krause et al., 2000; Sidhu et al., 2007). A mutation in the apicoplast rRNA has shown that addition of isopyropentyl pyrophosphate (Berlin et al., 2015) does not rescue parasite growth. The R43AI136118) and the Bill and Melinda Gates Foundation, USA, support for Genome Analysis, USA for help with Pac Bio and Illumina sequencing. Assistance with the analysis of the apicoplast and mitochondrial genomes of B. duncani was provided by Dr. Emmanuel Cornilot. 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).

Acknowledgements

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 Cornilot. 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).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.ijpara.2018.05.008.

References


