International Journal for Parasitology 49 (2019) 375–387

Contents lists available at ScienceDirect

International Journal for Parasitology

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

Next generation sequencing from *Hepatozoon canis* (Apicomplexa: Coccidia: Adeleorina): Complete apicoplast genome and multiple mitochondrion-associated sequences ☆

Alexandre N. Léveillé [a], Gad Baneth [b], John R. Barta [a,*]

[a] Department of Pathobiology, Ontario Veterinary College, University of Guelph, 50 Stone Rd E, Guelph, ON N1G 2W1, Canada
[b] Koret School of Veterinary Medicine, Hebrew University of Jerusalem, P.O. Box 12, Rehovot 76100, Israel

ARTICLE INFO

Article history:
Received 13 July 2018
Received in revised form 29 November 2018
Accepted 1 December 2018
Available online 19 February 2019

Keywords:
Hepatozoon canis
Canine
Apicomplexa
Adeleorina
Apicoplast
Mitochondrion
Nuclear ribosomal RNA
Next generation sequencing

ABSTRACT

Extrachromosomal genomes of the adeleorinid parasite *Hepatozoon canis* infecting an Israeli dog were investigated using next-generation and standard sequencing technologies. A complete apicoplast genome and several mitochondrion-associated sequences were generated. The apicoplast genome (31,869 bp) possessed two copies of both large subunit (23S) and small subunit (16S) ribosomal RNA genes (rDNA) within an inverted repeat region, as well as 22 protein-coding sequences, 25 transfer RNA genes (tDNA) and seven open reading frames of unknown function. Although circular-mapping, the apicoplast genome was physically linear according to next-generation data. Unlike other apicoplast genomes, genes encoding ribosomal protein S19 and tDNAs for alanine, aspartic acid, histidine, threonine and valine were not identified. No complete mitochondrial genome was recovered using next-generation data or directed PCR amplifications. Eight mitochondrion-associated (215–3523 bp) contigs assembled from next-generation data encoded a complete cytochrome *c* oxidase subunit I coding sequence, a complete cytochrome *c* oxidase subunit III coding sequence, two complete cytochrome B coding sequences, a non-coding, pseudogene for cytochrome B and multiple fragmented mitochondrial rDNA genes (SSUA, SSUB, SSUD, LSUC, LSUG, RNA6, RNA10, RNA14, RNA18). The paucity of NGS reads generating each of the mitochondrion-like sequences suggested that a complete mitochondrial genome at typically high copy number was absent in *H. canis*. In contrast, the complete nuclear rDNA unit sequence of *H. canis* (18S rDNA to 28S rDNA, 6977 bp) had >1000-fold next-generation coverage. Multiple divergent (from 93.6% to 99.9% pairwise identities) nuclear 18S rDNA contigs were generated (three types with 10 subtypes total). To our knowledge this is the first apicoplast genome sequenced from any adeleorinid coccidium and the first mitochondrion-associated sequences from this serious pathogen of wild and domestic canids. These newly generated sequences may provide useful genetic loci for high-resolution species-level genotyping that is currently impossible using existing nuclear rDNA targets.

1. Introduction

The genus *Hepatozoon* is home to a group of heteroxenous blood parasites within the sub-order Adeleorina (reviewed in Smith, 1996; Barta, 2000). The Adeleorina remain a poorly understood group of biologically diverse apicomplexan parasites (Siddall, 1995; Barta, 1989, 2000; Smith et al., 2000) with clear association with other apicomplexan parasites but only poorly defined relationships among the constitutive adeleorinid taxa (Barta et al., 2012). *Hepatozoon canis* was first observed in the leukocytes of dogs in India and was initially named *Leucocytozoon canis* by James (1905) and later transferred to the genus *Hepatozoon* Miller 1908 by Wenyon (1926). The definitive host of *H. canis* is the ixodid tick, *Rhipicephalus sanguineus*, within which the parasite undergoes sexual maturation, fertilization in the digestive tract and sporogony in the haemocoel of infected ticks, resulting in the formation of infective oocysts (Baneth et al., 2007). *Hepatozoon canis* may also use *Rhipicephalus turanicus* or *Amblyomma ovale* as definitive hosts according to Giannelli et al. (2017) and Rubini et al. (2009), respectively. Wild and domestic canids act as intermediate hosts

☆ Note: Nucleotide sequence data reported in this paper are available in the GenBank database under the accession numbers: MH557086, MH557087, MH615002, MH615003, MH615004, MH615005, MH615006, MK091084, MK091085, MK091086, MK091087, MK091088, MK091089, MK091090, MK091091, MK091092, MK214282, MK214283, MK214284, MK214285.

* Corresponding author at: Department of Pathobiology, University of Guelph, 50 Stone Road East, Guelph, ON N1G 2W1, Canada. Fax: +1 519 824 5930.
E-mail address: jbarta@uoguelph.ca (J.R. Barta).

https://doi.org/10.1016/j.ijpara.2018.12.001
0020-7519/

and are infected by ingestion of a tick containing sporulated oocysts.

Merogony occurs within the bone marrow of the canid host. Merozoites then enter neutrophils and develop into macrogamonts and microgamonts. Gamonts remain dormant in neutrophils and circulate in the bloodstream awaiting ingestion by a blood feeding tick host (Baneth et al., 2007). *Hepatozoon canis* is prevalent in Africa (McCully et al., 1975; Ezeokoli et al., 1983), Asia (James, 1905; Rajamanickam et al., 1985; Murata et al., 1991; Baneth et al., 1996), Europe (Gabrielli et al., 2010; Farkas et al., 2014) and South America (O'Dwyer et al., 2001). *Hepatozoon canis* infections have been reported in North America (Allen et al., 2008; Li et al., 2008; Kistler et al., 2014) however another canid-infecting parasite, *Hepatozoon americanum*, is thought to be more prevalent in this geographical location (see Vincent-Johnson et al., 1997; Baneth et al., 2000). The molecular differentiation of *H. canis* and *H. americanum* was confirmed with partial sequences of the nuclear small subunit (SSU, 18S) ribosomal RNA gene (nu-rDNA) sequences (Baneth et al., 2000). Canine hepatozoonosis is a recognized as a clinically significant disease with fever, lethargy, myalgia and lameness occurring in severe cases (Elias and Homans, 1988). *Hepatozoon americanum* infections are highly pathogenic whereas *H. canis* infections can be relatively asymptomatic (with low level parasitemia) to severe (with high level parasitemia) (Elias and Homans, 1988; Baneth and Weigler, 1997; Little et al., 2009).

To date, 18S nu-rDNA sequences were the only molecular markers available from *Hepatozoon* spp. infecting canids. Several studies have analysed and compared 18S nu-rDNA sequences from various geographic locations (see Gabrielli et al., 2010; Cardoso et al., 2014; Farkas et al., 2014). The 18S nu-rDNA sequence is a popular molecular marker because it contains regions that are conserved across a wide range of organisms, making it easy to target in previously unsequenced taxa. Unfortunately, relatively conserved apicomplexan 18S nu-rDNA has been found to produce short branch lengths with little resolution among closely related species during molecular phylogenetic analyses (Zhao et al., 2001; Morrison et al., 2004). More seriously, the nuclear genomes of some apicomplexan parasites contain multiple, divergent ribosomal gene copies that result in high genetic divergence among paralogues within individual species (McCutchan et al., 1988; Goethert et al., 2006). For these reasons, the exclusive use of 18S nu-rDNA sequences can result in molecular phylogenetic trees that do not reflect the organismal evolutionary history among closely related taxa (see Ogedengbe et al., 2011, Ogedengbe et al., 2018). Intraspecific variation has been found to be problematic with 18S nu-rDNA molecular studies of canine hepatozoonosis (see Little et al., 2009; Gabrielli et al., 2010; Farkas et al., 2014; Modrý et al., 2017). Thus, molecular studies of canine hepatozoonosis could benefit from access to a wider selection of molecular targets.

Most members of the Apicomplexa possess two extrachromosomal genomes, apicoplast and mitochondrial, found frequently at comparatively high copy numbers (reviewed in Feagin, 1994; Wilson and Williamson, 1997). Coding regions of the mitochondrial genome have been used widely for generating sequences that are informative at a species level in a wide variety of eukaryotes (i.e. DNA barcoding, Hebert et al., 2003). The complete mitochondrial genome sequence (Léveillé et al., 2014) and short apicoplast-encoded small subunit (16S) rDNA sequence (Lang-Unnasch et al, 1998) have been generated from *Hepatozoon catesbianae*. Despite repeated attempts, the methods used for obtaining mitochondrial sequence data from other adeleorinid parasites (see Léveillé et al., 2014, 2019) were unsuccessful with *H. canis*.

The present work describes a next generation sequencing (NGS)-based approach to obtain nuclear, apicoplast and mitochondrial sequence data from *H. canis*-infected canine blood. A complete 18S-ITS1 (internal transcribed spacer 1)-5.8S-ITS2-28S nu-rDNA unit was assembled; variation among the 18S nu-rDNA copies included three distinct types with 10 unique subtypes identified. The first known complete apicoplast genome from an adeleorinid coccidium was generated. Finally, multiple mitochondrion-associated sequences were assembled from *H. canis*. These novel sequence data may provide phylogenetically informative loci from these genomes that can be exploited in future studies.

2. Materials and methods

2.1. Sample collection

A blood sample (code number 9992) was collected in July 2014 into an EDTA tube by venipuncture from a naturally infected mixed-breed dog with canine hepatozoonosis in central Israel and submitted to the Baneth laboratory at the Koret School of Veterinary Medicine, Hebrew University, Rehovot, Israel, for further evaluation and PCR analysis. The sample was found to be parasitemic with *Hepatozoon* sp. gamonts by light microscopy of a May-Grunwald Giemsa stained blood smear.

2.2. DNA extraction and sequence-based genotyping of isolate

DNA was extracted from the blood sample with a commercial purification kit (Illustra Blood GenomicPrep Mini Spin Kit; GE Healthcare, Buckinghamshire, UK), according to the manufacturer's instructions. Diagnostic PCR was performed with Piroplasmid-F (5′-CCAGCAGCCGCGGTAATTC-3′) and Piroplasmid-R (5′-CTTTCGC AGTAGTTYGTCTTTAACAAATCT-3′) primers to amplify and sequence an approximately 360 bp region of the 18S nu-rDNA of *Hepatozoon* spp. as previously described (Tabar et al., 2008; Baneth et al., 2013). The resultant partial 18S nu-rDNA sequence was compared with sequences deposited in GenBank using the Basic Local Alignment Search Tool (BLAST; Altschul et al., 1990).

2.3. Library preparation and NGS sequencing

Initially, multiple attempts were made to generate a mitochondrial genome sequence using various combinations of universal primers that target highly conserved regions of most apicomplexan mitochondrial genomes (as described in Léveillé et al., 2019). After all attempts failed, a NGS approach was pursued and 1.4 μg of the DNA sample was submitted for whole genome NGS using an Illumina HiSeq 2500 System (Illumina Inc., San Diego, California, USA) by the DNA Sequencing and Synthesis Facility of the Centre for Applied Genomics, Hospital for Sick Children (Toronto, Ontario, Canada). The DNA library was generated with a targeted fragment size of ~550 bp and the sequencing run used one lane to produce 126 bp paired-end sequence reads. Reads were assessed for quality and paired reads were linked prior to further analyses within Geneious ((Version 6.1 and later) BioMatters, Ltd., Auckland, New Zealand; Kearse et al., 2012).

2.4. Screening and assembly of NGS reads

The lack of reference sequences and the overwhelming excess of NGS data referable to the canine host rather than the *H. canis* parasite precluded the use of shotgun de novo assembly of all read pairs. Instead, paired NGS reads were first screened for similarities to a panel of reference organellar sequences from *H. catesbianae* (GenBank Accession number KF894962, mitochondrion; AF040972, apicoplast 16S rDNA), and *Klossiella equi* (MH203050, mitochondrion). A sequence for the complete nu-rDNA unit was

assembled by initial reference to existing 18S nu-rDNA sequences of *H. canis* (e.g. AY461378).

2.4.1. Apicoplast and nu-rDNA unit

For apicoplast and nu-rDNA reference sequences, read pairs generated in the Illumina HiSeq NGS run were screened against the reference sequences using the 'Map to Reference' function within Geneious. A high level of permitted mismatch per read (30%; complete assembly parameters are in Supplementary Table S1) was set to allow identification of putative *H. canis* homologs in the NGS reads to the assembly target sequences that shared only 70% identity; paired mates of any hits were also retained for further analysis. Selected NGS reads (and their paired mates) were then assembled de novo within Geneious to generate putative *H. canis*-specific assemblies and consensus sequences of the resulting contigs. Contigs were searched against publically available sequences using BLAST (Altschul et al., 1990; BLASTn from within Geneious); contigs with obvious similarity to vertebrate sequences were discarded.

Apicomplexa-associated consensus sequences were then used as the assembly seeds for repetitive rounds of assembly to reference (Geneious) to extend the contigs until no further elongation was possible. During each round of in silico extension of each contig, all reads and their paired mates were retained. Finally, the used reads and their mates associated with each seed were then de novo assembled using Geneious to produce robust assemblies for each target. After de novo assembly with Geneious, a contig alignment of the original reads was retained; this permitted the resulting alignment of NGS read pairs to be examined by eye to confirm that all reads were mapped correctly with respect to directions and expected distance between mapped paired reads.

In the case of the complete nu-rDNA unit assembly, evidence of multiple divergent copies was obvious when the resulting assembly was examined. To limit the final contig to a single complete nu-rDNA unit, all paired-end reads contributing to the penultimate assembly were de novo assembled one last time using only paired-end reads that mapped with 100% identity to generate a single representative sequence for the most abundant complete nuclear 18S-ITS1-5.8S-ITS2-28S rDNA unit.

2.4.2. Assembly and identification of 18S nu-rDNA variants

The most abundant 18S nu-rDNA sequence generated as part of the complete 18S-ITS1-5.8S-ITS2-28S rDNA unit was used as a reference for a highly permissive (30% mismatches, up to 20% gaps permitted) 'assemble to reference' (Geneious) with the complete NGS paired read dataset. Polymorphic sites within the resulting 18S nu-rDNA assembly were identified using the 'Find Variations/SNPs' function from within Geneious using a minimum variant frequency cutoff limit of 2.5%. Paired reads incorporating variant sites were then examined by eye to identify linked polymorphic sites; this process was repeated until complete 18S nu-rDNA variants were generated. To confirm each variant, NGS sequences were assembled to each putative variant using the 'Assemble to Reference' function within Geneious and checked for correct orientation (direction) and spacing of all NGS paired reads included in the assembly. Finally, all available NGS paired reads were assembled to all generated 18S nu-rDNA variants at low stringency (25% mismatch; 20% gaps permitted) to 'mine' all paired reads that have reasonable pairwise identities to *Hepatozoon* nu-rDNA. All of these reads were then used to estimate the relative abundance of each variant in the *H. canis* nuclear genome by conducting highly specific assembly to all variants (0% mismatch; no gaps permitted). After removing all reads that assembled to the confirmed variants, the remaining reads were de novo assembled (Geneious) to determine if any further variants were generated; none were found (data not shown).

2.4.3. Assembly of mitochondrion-associated sequences

Initial screening against mitochondrial reference sequences was conducted as described for the apicoplast and nu-rDNA unit above (see Section 2.4.1) using complete mitochondrial genomes or extracted coding regions (CDS; cytochrome *c* oxidase subunit I (COI), cytochrome *c* oxidase subunit III (COIII) and cytochrome B (CytB)) from the *H. catesbianae* and *K. equi* reference genomes. Additionally, short segments (20–40 bp) from highly conserved portions of the coding regions and some fragmented mitochondrial rDNA (LSUE, LSUF, LSUG, SSUB, RNA14, RNA18) served as references for screening using the same parameters. The majority of reads mapped to these targets originated from the host DNA. Individual reads demonstrating similarity to apicomplexan mitochondrial sequences (based on BLAST searches) were extracted together with their mates; both were used as new reference sequences to search the dataset for associated NGS reads using the 'map to reference' function within Geneious (5% mismatch; 20% gaps permitted), again retaining all read mates. After each round of screening, all paired reads were de novo assembled (Geneious; 'medium sensitivity/fast' – 30% mismatch, 15% gaps permitted) to produce putative *H. canis*-specific assemblies and consensus sequences of the resulting contigs. These contigs were then used as the assembly seeds for repetitive rounds of map to reference assembly (Geneious) to extend the contigs until no further elongation was possible as described above (see Section 2.4.1).

Contig sequences for all final assemblies of NGS reads were generated using the 'Highest Quality' criterion for generating a majority consensus sequence in Geneious; this setting totals the quality for each potential base call from the fastq data and, if the total contribution for a particular base exceeds 60% of the total, then the base is called. Otherwise, an ambiguous base code will be assigned.

2.5. PCR amplification and Sanger sequencing of the mitochondrial CDS regions

To confirm an NGS-derived mitochondrial protein CDS, PCR was used to amplify these CDS directly from mixed host/parasite DNA. Prior to use in PCR, the DNA sample was enriched for parasite DNA by selectively binding and removing the CpG-methylated canine host DNA using the NEBNext Microbiome DNA Enrichment Kit (New England Biolabs, Ipswich MA, USA). Briefly, 1 µg of mixed host/parasite DNA was enriched following the standard kit protocol. The resulting enriched DNA was suspended in 20 µl of Tris–EDTA (TE) buffer. PCR amplification primers (Table 1) were designed to span the CDS sequences (COI, COIII, CytB) identified by assembly of NGS reads.

High fidelity PCR was carried out in a MJ Mini thermal cycler (Bio-Rad Laboratories, Inc., Hercules, CA, USA) in a 25 µl reaction containing 1× Invitrogen Platinum SuperFi PCR Master Mix (Invitrogen, Carlsbad CA, USA), 0.5 mM of each amplification primer (Table 1) and 0.5 µl of *H. canis*-enriched DNA template. The PCR profile consisted of an initial melt at 98 °C for 30 s followed by 35 amplification cycles (denaturing at 98 °C for 10 s, annealing at 55 °C for 10 s and extension at 72 °C for 45 s (15–30 s/1 kb of expected product)) and then terminated with a final extension of 72 °C for 5 min to complete any partial products. Annealing temperatures were chosen using the Thermo Fisher Platinum SuperFi DNA polymerase Tm calculator (www.thermofisher.com). The PCR product was separated electrophoretically using a submarine 1.4% agarose gel with 1× Tris-acetate–EDTA (TAE) buffer (50 ml) and 2 µL of ethidium bromide dye (10 mg/ml, w/v). The GeneRuler 1 kb Plus DNA size ladder (Thermo Fisher Scientific, Waltham MA, USA) was used to determine the product length. The gel was then examined using an ultraviolet transilluminator and the DNA bands of the expected size (~1000–1500 bp) were excised using a new, sterile scalpel for each band. DNA was extracted from the gel slice

Table 1
PCR amplification and sequencing primers used to confirm the next-generation sequencing-derived *Hepatozoon canis* mitochondrion-associated coding sequences.

Primer name	Sequence	Anneal temperature[a]	Amplicon size (bp)
H_canis_COI_F	5′-GGTCTGGTATWCTAGTGGAT-3′	55 °C	1570
H_canis_COI_R	5′-GGTTATCTTGAATGGGGCTA-3′		
H_canis_COI_598F[b]	5′-CTGATCGGAACACTCCCCAT-3′	–	–
H_canis_COI_890R[b]	5′-ATCATGTGATGGCCCCARAC-3′	–	–
H_canis_COIII_F	5′-TGAATATTCTACTTCCATAGCTAGA-3′	55 °C	996
H_canis_COIII_R	5′-CCAACCAATACGTATACCGGT-3′		
H_canis_CytB1_F	5′-CAATAGCCCCGGATAGAAGA-3′	55 °C	1542
H_canis_CytB1_R	5′-GCTAGACAGAGAAGAGCCTG-3′		
H_canis_CytB2_F	5′-TCACTCTTCTTTAAGTAAGGCT-3′	55 °C	1418
H_canis_CytB2_R	5′-ACAGGATGAGACCCTCTAGT-3′		
Hcan_CytB_Uni_R[b]	5′-CCAGCTGGTTTSGATGGTAT-3′	–	–

COI, cytochrome *c* oxidase subunit I; COIII, cytochrome *c* oxidase subunit III; CytB, cytochrome B.
[a] Estimated for high fidelity PCR using Thermo Fisher Tm calculator for Platinum SuperFi DNA polymerase (www.thermofisher.com).
[b] Indicates the sequencing primers.

using the QIAquick Gel Extraction Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions.

The purified PCR amplicons were then sequenced in both directions with the forward and reverse amplification primers, supplemented with internal sequencing primers when additional coverage was needed (Table 1), using an ABI Prism 7000 Sequence Detection System (Applied Biosystems Inc., Foster City CA, USA) by the Molecular Biology Unit of the Laboratory Services Division, University of Guelph (Guelph ON, Canada). Chromatograms received from sequencing reactions were imported into Geneious for assembly and analyses.

2.6. Annotation

2.6.1. Protein CDS

Open reading frames (ORFs) in the assembled apicoplast genome and mitochondrion-associated sequences were identified using an ORF search utility within Geneious with either a bacterial translation table (trans_table 11, for apicoplast) or mold protozoan mitochondrial translation table (trans_table 4, for mitochondrion) to guide the search; these ORFs were translated using the same translation table, and the resulting amino acid residue (AA) sequences were searched against the public sequence databases using the protein BLAST (BLASTp) and the Domain enhanced lookup time accelerated BLAST (DELTA-BLAST) algorithms (Altschul et al., 1990; Gish and States, 1993; Boratyn et al., 2012) to identify each ORF with respect to similar AA sequences. Pairwise analyses of paralogous mitochondrial CDS and AA sequences were performed in Geneious using the 'Multiple Align' tool (CDS: cost matrix: 65%, gap open penalty: 12, gap extension penalty: 3; Protein: cost matrix: Blosum62, gap open penalty: 12, gap extension penalty: 3). In addition, transmembrane predictions for the translation products of the mitochondrial CDS were made using the Hidden Markov Model within Geneious and compared with the transmembrane predictions for the corresponding products from *H. catesbianae* (KF894962) and *Plasmodium falciparum* (M76611).

2.6.2. tRNA genes

Initially apicoplast-encoded tRNA gene (tDNA) sequences were identified using sequence comparisons with tDNA sequences previously annotated in *P. falciparum* (LN999985) and *Eimeria tenella* (AY217738) apicoplast genomes. All annotated tDNA were extracted from each of the publicly available annotated apicoplast genome sequences noted above, and these sequences were aligned pairwise, where possible, to the *H. canis* apicoplast genome sequence. Initial tDNA annotations were compared with entire genome scans with tRNAscan-SE software (Version 2.0; Lowe and Chan, 2016) using the general tRNA model (tRNAscan (strict) + EufindtRNA (relaxed) → Cove) and the TRNA2.cm covariance mode. The resulting tDNA sequences were then extracted individually and their secondary structure was verified using the DNA fold extension within Geneious. Sequence folds were computed using the Turner 2004 nearest neighbour parameters for RNA folding (Turner and Mathews, 2009) at a temperature of 37 °C. Sequences identified as transfer RNA that scored <50 on the 'Cove' parameter of tRNAscan-SE or did not fold properly were labelled as putative.

2.6.3. rRNA genes

The complete nu-rRNA gene (nu-rDNA) unit sequence (i.e. 18S-ITS1-5.8S-ITS2-28S region) was annotated by comparison with the same regions of several related apicomplexan parasites (i.e. *E. tenella* (AF026388), *Toxoplasma gondii* (L25635) and *Babesia microti* (AB190287)).

Apicoplast rDNA (pl-rDNA) sequences were initially identified based on pairwise sequence identity with previously annotated regions of *E. tenella* (AY217738), *Leucocytozoon caulleryi* (AP013071) and *T. gondii* (NC_001799) apicoplast genome sequences. Apicoplast rDNA sequences were extracted from each of the publicly available annotated apicoplast genome sequences noted above, and these sequences were aligned pairwise to the *H. canis* apicoplast sequence. Initial pl-rDNA annotations were refined by comparison with the results of the HMMER software (version 3.0; Finn et al., 2011) searched against the European rRNA database.

Hepatozoon canis fragmented mitochondrial rDNA (mt-rDNA) were annotated based on similarity to the fragmented mt-rDNA previously annotated in *P. falciparum* (M76611; Feagin et al., 2012) and *H. catesbianae* (KF894962, Léveillé et al., 2014) mitochondrial genomes. Any fragmented mt-rDNA that aligned with the *H. canis* mitochondrial sequence with a pairwise sequence identity below 70% were excluded from further analysis. Putative fragmented mt-rDNA annotations follow the nomenclature of Feagin et al. (2012) applied to the mitochondrial genome of *P. falciparum*.

2.7. Data accessibility

All sequence assembly files are available from Mendeley Data (doi: https://doi.org//10.17632/bs2z92449s.1).

3. Results

3.1. Confirmation of identity of H. canis in blood slides

Gamonts observed on the blood smear matched gamonts of *H. canis* described by Baneth et al. (2007, Fig. 1). The amplified partial

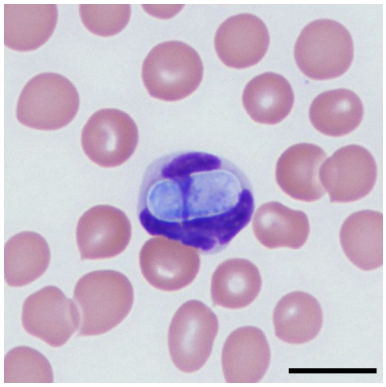

Fig. 1. Micrograph of *Hepatozoon canis* gamont within canine neutrophil. Gamonts observed on the blood smear matched gamonts of *H. canis* described by Baneth et al. (2007). Scale bar = 10 μm.

18S nu-rDNA sequence (346 bp) had 100% pairwise identity to a partial *H. canis* 18S nu-rDNA sequence from an Ethiopian dog (KF646812).

3.2. The H. canis apicoplast genome

The raw NGS data comprised 479,147,982 sequence reads (126 bp/read) resulting in 239,573,991 pairs after setting the sequence pairs using Geneious. After repeated assembly to reference iterations following initial seeding with a partial *H. catesbianae* apicoplast 16S rDNA sequence, a single contig of the complete *H. canis* apicoplast genome sequence was generated (31,869 bp; MH557086, Fig. 2). The final assembly utilized 8604 paired NGS reads with an average insert size of 508 bp (range 250–750 bp) and 34 ± 7-fold coverage. Although circular-mapping, the apicoplast genome of *H. canis* was found to be linear in structure. Initially a circular structure was considered since the 5′ and 3′ sequence termini overlapped by 79 bp with 100% pairwise identity. Further analysis revealed the absence of paired reads that spanned across the ends of the sequence; the 5′-end was composed exclusively of forward reads and the 3′-end was composed entirely of reverse reads (see Supplementary Fig. S1).

The apicoplast genome of *H. canis* possessed typical inverted repeat (IR) units at each end of the genome. The IR units were 4812 bp exact mirror copies that included a copy of the large (LSU, 23S) and small subunit (SSU, 16S) rRNA coding regions (Table 2) as well as six tDNA sequences (Table 3). The LSU of each IR contained a TA indel that varied from 14 to 22 bp in length. A copy length of 16 bp was in the highest abundance of reads mapped to both LSU copies (locations: 3217–3232 bp; 28,637–28,652 bp) and was chosen to represent the major copy (see Supplementary Data S1 (doi: https://doi.org//10.17632/bs2z92449s.1), Fig. 2). The region of both IR units interior to the ends of the LSU genes included the first 46 bp portion of the CDS for ribosomal protein S4 (rps 4; 4831 → 4876 bp) and cysteine desulfurase activator complex subunit (sufB; 27,038 → 26,993 bp). The region of both IR units external to the ends of the SSU genes, encoded at least the first 41 bp of the isoleucine tDNA. However, a complete isoleucine tDNA sequence was only found on the 3′-end of the genome with the remaining 31 bp of its sequence located immediately after the IR region; the 3′-end of the genome did not have a complete isoleucine tDNA.

Between the IR regions, 12 tDNA sequences were identified; however, alanine, aspartic acid, histidine, threonine and valine tDNA sequences were not found in the *H. canis* apicoplast genome. Five methionine tDNA sequences were identified. Two methionine tDNA sequences located within the IR region (4720 → 4794 bp and 27,149 → 27,075 bp) were in the same location as threonine tDNA sequences in other IR mapping apicoplast genomes. Twenty-two protein-coding genes and seven hypothetical ORFs were annotated (Table 4). A coding region for ribosomal protein S19 (rps 19) was not identified. The rps 19 coding sequence has been found in all other available apicoplast genomes. The *H. canis* coding sequence for ribosomal protein S2 (rps 2) and ribosomal protein S7 (rps 7) were divergent from corresponding sequences found in other apicoplast genomes. The rps 2 and rps 7 were not matched by BLASTp or DELTA-BLAST to any other apicomplexan sequences. The rps 2 was weakly (19% identity, 94% coverage) matched by BLASTp to the ribosomal protein S2 sequence of a *Chromerida* sp. (YP_003795399). The rps 7 was weakly BLASTp matched (23–28% identity, 82–94% coverage) to the 30S ribosomal protein S7 sequences of various bacteria (i.e. *Candidatus* sp., OHA07460). The *H. canis* rps 2 and rps 7 coding sequences were therefore labelled as putative. As observed in other apicoplast genomes, the *H. canis* coding sequence for RNA polymerase C subunit 2 (rpoC2) is interrupted halfway by a stop codon (TAA) (Wilson et al., 1996; Cai et al., 2003; Imura et al., 2014; Seeber et al., 2014); consequently, the *H. canis* rpoC2 coding sequence has been labelled as two coding sequences, rpoC2A and rpoC2B, following the convention proposed by Imura et al. (2014). The rpoC2B gene sequence also contained the only single nucleotide difference (SND) found in the assembly. At position 18,162 bp 19 reads called for a thymine and eight reads called for a cytosine. The SND was located in the third position of a codon and both options code for a leucine. The site was marked with the nucleotide ambiguity code "Y". In the *H. canis* apicoplast genome, there are seven hypothetical ORFs that have unknown functions. BLASTp searches of each of the unknown ORF sequences found no significant similarities to other protein-coding sequences. All other known apicoplast genomes also contain ORF sequences for which functions remain unknown. The unknown ORFs found in the *H. canis* apicoplast genome have been assigned alphabetical letters based on the order in which they appear in the genome sequence and these letters are not intended to correspond to any similarly labelled ORFs in other apicoplast genomes.

3.3. Hepatozoon canis mitochondrion-associated sequences

A complete mitochondrial genome for *H. canis* could not be obtained using methods successful for the apicoplast genome. Instead, eight sequences were generated that had sufficient similarity to other apicomplexan mitochondrial genome sequences to suggest that they were of mitochondrial origin (Fig. 3). All assemblies were generated from paired NGS reads when possible, however single reads that paired to reads that mapped outside of the assembly area were retained to improve coverage. Assembly extensions were limited by low coverage depth. In some cases, PCR amplifications were utilized to connect paired reads that spanned CDS regions with no NGS coverage. All similarities to known apicomplexan mitochondrial sequences were annotated accordingly. Mitochondrion-associated sequence 1 (mt-sequence 1) was 3207 bp in length and included a complete COI CDS that was 1479 bp in length and mapped from bases 885 to 2363 (Table 5). The complete COI coding sequence was confirmed by a 1570 bp PCR product that was sequenced in both directions using Sanger sequencing. The final assembly was composed of 88 paired and seven single (on assembly ends) NGS reads and four Sanger sequencing reads with an overall one to 12-fold sequence coverage (MH615002). The amplified complete COI CDS sequence, including mapped external and internal PCR primers (see Table 1), was also deposited to GenBank (MH557087). The COI CDS had a proposed ATG start codon, a proposed TAA stop codon and produced a predicted translation product of 492 AA. A DELTA-BLAST search of the COI AA sequence produced a 67% pairwise identity match

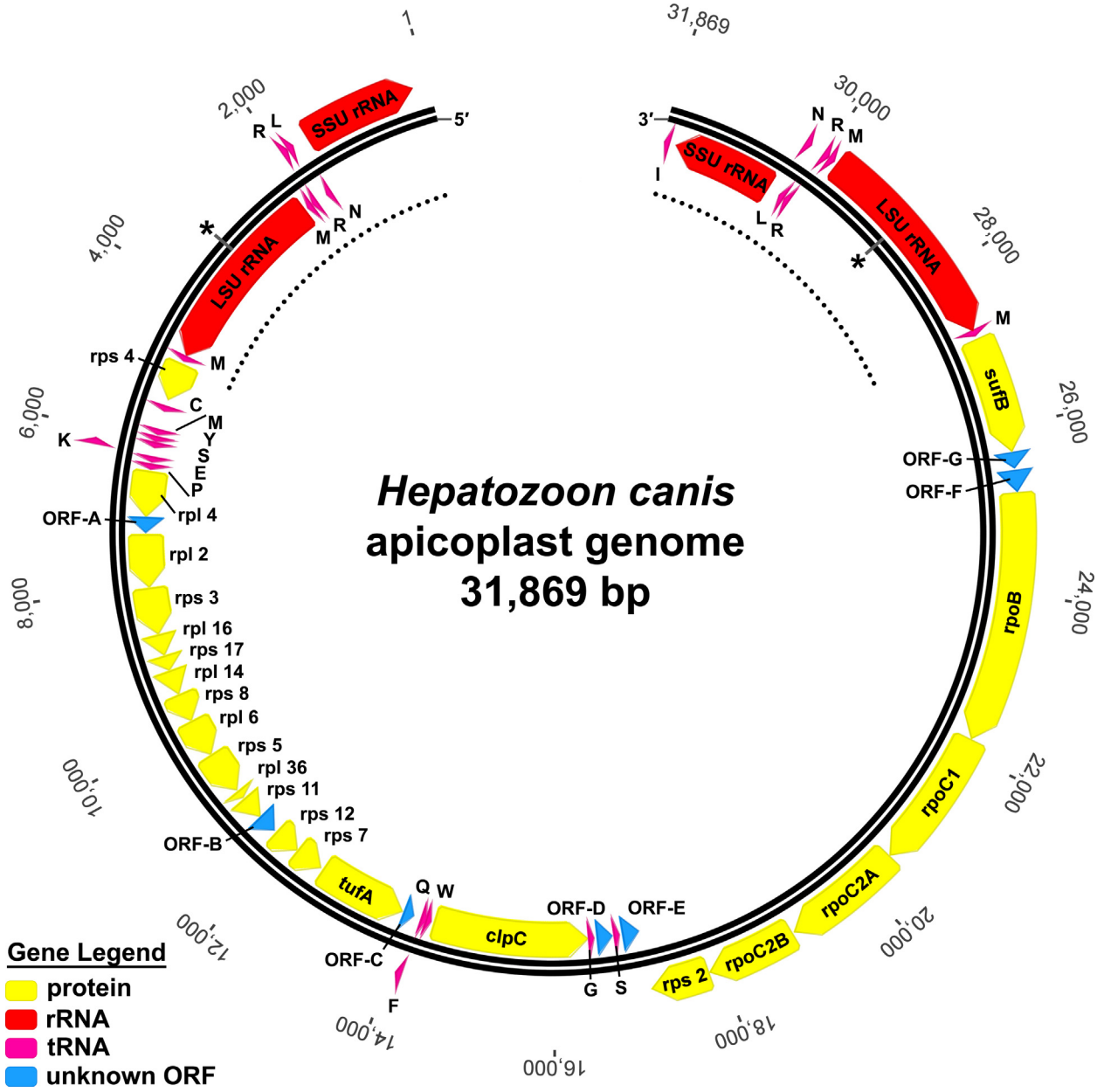

Fig. 2. Annotated complete apicoplast genome of *Hepatozoon canis* (GenBank Accession number MH557086). The final assembly utilized 8604 paired next-generation sequencing reads with an average insert size of 508 bp (range 250–750 bp) and 34 ± 7-fold coverage. Although circular-mapping, the apicoplast genome assembly data provided evidence for a linear structure (see Supplementary Fig. S1). The genome contained an inverted repeat region (dotted lines), gene content and genome organization most similar to the haemosporinid and coccidian apicoplast genomes. The large subunit of each inverted repeat contained a TA indel (marked by asterisks) that varied from 14 to 22 bp in length. A copy length of 16 bp was in the highest abundance of reads mapped to both LSU copies (locations: 3217–3232 bp; 28,637–28,652 bp) and was chosen to represent the major copy. Initial genome sequence image was generated in Geneious (www.geneious.com; Kearse et al., 2012). SSU, small subunit; clpC, clp protease ATP-binding subunit; sufB, cysteine desulfurase activator complex subunit; rpoB, RNA polymerase B; rpoC, RNA polymerase C; rpl, ribosomal protein; ORF, open reading frame, tufA, elongation factor Tu.

Table 2
rRNA gene sequences annotated on the *Hepatozoon canis* apicoplast genome sequence (GenBank Accession number MH557086) were determined based on pairwise sequence identity with previously annotated regions of *Eimeria tenella* (AY217738), *Leucocytozoon caulleryi* (AP013071) and *Toxoplasma gondii* (NC_001799) apicoplast genome sequences and followed up by HMMER software (version 3.0; Finn et al., 2011) searched against the 5S rRNA database and the European rRNA database. Bolded labels correspond to annotations in Fig. 2.

Apicoplast rRNA gene	Size (bp)	Location (bp)
Small subunit (**SSU rRNA**, 16S)	1399	1591 → 194
Small subunit (**SSU rRNA**, 16S)	1399	30,278 → 31,676
Large subunit (**LSU rRNA**, 23S)	2609	2131 → 4739
Large subunit (**LSU rRNA**, 23S)	2609	29,738 → 27,130

(96% query coverage) with the mitochondrial COI AA sequence of *K. equi* (AZL31646), a 65% pairwise identity match (94% query coverage) with the mitochondrial COI AA sequence of *H. catesbianae* (AIG55088) and a 58–59% pairwise identity match (94–96% query coverage) to various haemosporinid mitochondrial COI AA sequences. Transmembrane predictions showed that the *H. canis* COI translation had the same 12 transmembrane regions predicted for the COI AA sequences of *H. catesbianae* (KF894962) and *P. falciparum* (M76611). However, the *H. canis* COI AA sequence contained a thirteenth predicted transmembrane region (residues 78 to 98) not found in the other AA sequence transmembrane predictions. This transmembrane region was flanked by short internal and

Table 3
tRNA gene sequences annotated on the *Hepatozoon canis* apicoplast genome sequence (GenBank Accession number MH557086) were determined by comparison with tRNA gene sequences previously annotated in *Plasmodium falciparum* (LN999985) and *Eimeria tenella* (AY217738) and entire genome scans with tRNAscan-SE software (Version 2.0; Lowe and Chan, 2016) using the general tRNA model and the TRNA2.cm covariance mode. Bolded labels correspond to annotations in Fig. 2.

tRNA gene	Anti-codon	Length (bp)	Location
Alanine (tRNA-Ala; A)	–	–	Not found
Arginine (tRNA-Arg; **R**)	UCU	71	1918 → 1848[a]
		71	29,951 → 30,021[a]
	GCG	77	1927 → 2003[a,b]
		77	29,942 → 29,866[a,b]
Asparagine (tRNA-Asn; **N**)	GUU	72	1689 → 1760[a]
		72	30,180 → 30,109[a]
Aspartic Acid (tRNA-Asp; D)	–	–	Not found
Cysteine (tRNA-Cys; **C**)	GCA	71	5456 → 5526
Glutamine (tRNA-Gln; **Q**)	UUG	71	14,110 → 14,180
Glutamic Acid (tRNA-Glu; **E**)	UUC	73	6170 → 6242[b]
Glycine (tRNA-Gly; **G**)	UCC	71	16,427 → 16,497
Histidine (tRNA-His; H)	–	–	Not found
Isoleucine (tRNA-Ile, **I**)	GAU	72	31,764 → 31,835
Leucine (tRNA-Leu; **L**)	UAG	79	1841 → 1763[a,b]
		79	30,028 → 30,106[a,b]
Lysine (tRNA-Lys; **K**)	UUU	72	6139 → 6068
Methionine (tRNA-Met; **M**)	CAU	80	2010 → 2089[a,b]
		75	4720 → 4794[a,b,c]
		74	5781 → 5854
		75	27,149 → 27,075[a,b,c]
		80	29,859 → 29,780[a,b]
Phenylalanine (tRNA-Phe; **F**)	GAA	74	14,093 → 14,020
Proline (tRNA-Pro; **P**)	UGG	72	6265 → 6336
Serine (tRNA-Ser; **S**)	GCU	90	5975 → 6064[b]
	UGA	94	16,762 → 16,855[b]
Threonine (tRNA-Thr; T)	–	–	Not found
Tryptophan (tRNA-Trp; **W**)	CCA	71	14,182 → 14,252
Tyrosine (tRNA-Tyr; **Y**)	GUA	88	5867 → 5954[b]
Valine (tRNA-Val; V)	–	–	Not found

[a] Located within inverted repeat (IR) regions.
[b] Putative tDNA: scored <50 on the 'Cove' parameter of tRNAscan-SE or did not fold properly.
[c] Mapped as threonine in other apicoplast genomes.

Table 4
Protein coding gene sequences annotated on the *Hepatozoon canis* apicoplast genome sequence (GenBank Accession number MH557086) were translated using the bacterial translation table (trans_table 11) and the resulting amino acid sequences were searched against the public sequence databases using the BLASTp and DELTA-BLAST algorithms (Altschul et al., 1990; Gish and States, 1993; Boratyn et al., 2012) to identify each in relation to similar protein sequences. Bolded labels correspond to annotations in Fig. 2.

Protein coding sequence	Length (bp)	Location (bp)
clp protease ATP-binding subunit (**clpC**)	2154	14,288 → 16,441
Cysteine desulfurase activator complex subunit (**sufB**)	1404	27,038 → 25,635
Elongation factor Tu (**tufA**)	1233	12,616 → 13,848
Ribosomal protein L2 (**rpl 2**)	774	7235 → 8008
Ribosomal protein L4 (**rpl 4**)	630	6368 → 6997
Ribosomal protein L6 (**rpl 6**)	534	9876 → 10,409
Ribosomal protein L14 (**rpl 14**)	360	9132 → 9491
Ribosomal protein L16 (**rpl 16**)	297	8619 → 8915
Ribosomal protein L36 (**rpl 36**)	117	11,037 → 11,153
Ribosomal protein S2 – putative (**rps 2**)	723	17,836 → 17,114
Ribosomal protein S3 (**rps 3**)	615	8012 → 8626
Ribosomal protein S4 (**rps 4**)	513	4831 → 5343
Ribosomal protein S5 (**rps 5**)	591	10,420 → 11,010
Ribosomal protein S7 – putative (**rps 7**)	411	12,172 → 12,582
Ribosomal protein S8 (**rps 8**)	405	9488 → 9892
Ribosomal protein S11 (**rps 11**)	366	11,153 → 11,518
Ribosomal protein S12 (**rps 12**)	372	11,799 → 12,170
Ribosomal protein S17 (**rps 17**)	234	8912 → 9145
Ribosomal protein S19 (rps 19)[a]	–	Not found
RNA polymerase B (**rpoB**)	3024	25,146 → 22,123
RNA polymerase C – subunit C1 (**rpoC1**)	1701	22,111 → 20,411
RNA polymerase C – subunit C2A (**rpoC2A**)	1461	20,414 → 18,954
RNA polymerase C – subunit C2B (**rpoC2B**)	1134	18,944 → 17,811
Unknown ORF A (**ORF-A**)	219	7006 → 7224
Unknown ORF B (**ORF-B**)	267	11,502 → 11,768
Unknown ORF C (**ORF-C**)	183	13,856 → 14,038
Unknown ORF D (**ORF-D**)	261	16,539 → 16,799
Unknown ORF E (**ORF-E**)	228	16,894 → 17,121
Unknown ORF F (**ORF-F**)	300	25,448 → 25,149
Unknown ORF G (**ORF-G**)	237	25,623 → 25,387

ORF, open reading frame.
[a] rps19 was found in all other sequenced apicoplast genomes.

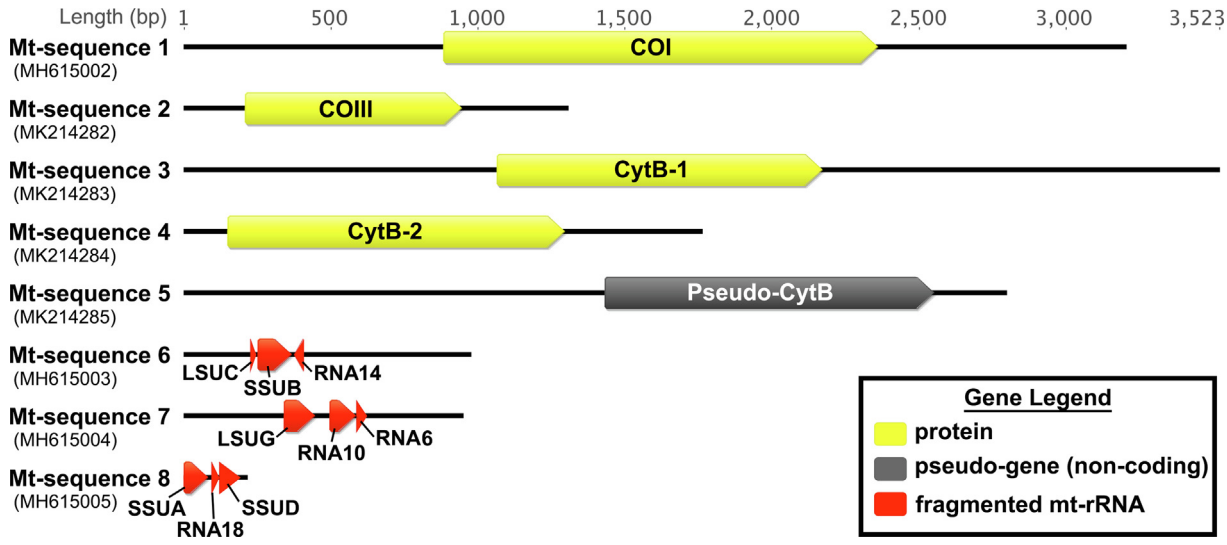

Fig. 3. Eight mitochondrion-associated sequences of *Hepatozoon canis*. Putative mitochondrion -associated sequences obtained from next-generation sequencing data supplemented with PCR and amplicon sequencing (GenBank Accession numbers in parentheses). All three protein-coding genes typical of apicomplexan mitochondrial genomes were detected (mt-sequences 1–5); three cytochrome B coding regions were detected but only two appeared to encode a potentially function CDS. Three small contigs (mt-sequences 6–8) possessed fragmented rDNAs that demonstrated sequence similarities to fragmented rDNA regions of apicomplexan mitochondrial genomes. Sequence images were generated in Geneious (www.geneious.com; Kearse et al., 2012). COI, cytochrome *c* oxidase subunit I; COIII, cytochrome *c* oxidase subunit III; LSU, large subunit; SSU, small subunit.

Table 5
Annotations of *Hepatozoon canis* mitochondrion-associated sequences. Cytochrome *c* oxidase subunit I, cytochrome *c* oxidase subunit III and cytochrome B coding sequences were translated using the metmt translation table (trans_table 4) and the resulting amino acid sequences were searched against the public sequence databases using the DELTA-BLAST algorithm (Altschul et al., 1990; Boratyn et al., 2012) to identify each in relation to similar protein sequences. Fragmented mitochondrial rRNA genes (mt-rDNA) were annotated based on similarity to the fragmented mt-rDNA annotations previously mapped in *Hepatozoon catesbianae* (GenBank Accession number KF894962) and *Plasmodium falciparum* (M76611) mitochondrial genomes. Bolded labels correspond to annotations in Fig. 3.

Mitochondrion-associated sequences	CDS	fragmented mt-rDNA identified[a]	Size (bp)	Location (bp)	% identity to *H. catesbianae*[b]	% identity to *P. falciparum*[b]
Mt-sequence 1 (3207 bp; MH615002)	**COI**	None	1479	885 → 2363	63.9	62.2
Mt-sequence 2 (1307 bp; MK214282)	**COIII**	None	738	205 → 945	56.6	59.9
Mt-sequence 3 (3523 bp; MK214283)	**CytB-1**	None	1110	1065 → 2174	62.6	62.6
Mt-sequence 4 (1762 bp; MK214284)	**CytB-2**	None	1149	150 → 1298	62.5	62.2
Mt-sequence 5 (2797 bp; MK214285)	**Pseudo-CytB** (non-coding)	None	1120	1435 → 2554	63.4	61.5
Mt-sequence 6 (976 bp; MH615003)	None	**LSUC**	22	225 → 246	77.3	77.3
		SSUB	116	252 → 367	75.0	78.0
		RNA14	32	406 → 375	96.8	93.8
Mt-sequence 7 (948 bp; MH615004)	None	**LSUG**	106	343 → 448	91.5	93.5
		RNA10	91	497 → 587	69.6	74.4
		RNA6	36	586 → 621	75.6	77.1
Mt-sequence 8 (215 bp; MH615005)	None	**SSUA**	85	1 → 84	72.7	81.8
		RNA18	25	94 → 118	92.0	90.0
		SSUD	72	120 → 191	81.6	77.8

LSU, large subunit; SSU, small subunit.
[a] Nomenclature follows Feagin et al. (2012).
[b] Analysed in Geneious using the 'Multiple Align' tool (cost matrix: 65%, gap open penalty: 12, gap extension penalty: 3; www.geneious.com, Kearse et al., 2012)

external loop structures and occurred in the same location as a large internal or external loop structure in the other AA sequence transmembrane predictions.

Mitochondrion-associated sequence 2 (mt-sequence 2) was 1307 bp in length and included a complete COIII CDS that was 738 bp in length and mapped from bases 208 to 945. The complete COIII CDS was confirmed by a 996 bp PCR product that was sequenced in both directions using Sanger sequencing. The final assembly was composed of 16 paired and five single (on one assembly end) NGS reads and two Sanger sequencing reads with an overall one to seven-fold sequence coverage (MK214282). The COIII CDS had a proposed ATG start codon, a proposed TAA stop codon and produced a predicted translation product of 245 AA. A DELTA-BLAST search of the COIII AA sequence produced a 35–38% pairwise identity match (96–97% query coverage) to various haemosporinid mitochondrial COIII AA sequences, a 37% pairwise identity match (96% query coverage) with the mitochondrial COIII AA sequence of *K. equi* (AZL31645), and a 36% pairwise identity match (96% query coverage) with the mitochondrial COIII AA sequence of *H. catesbianae* (AIG55090). Transmembrane predic-

tions showed that the *H. canis* COIII AA sequence had the same six transmembrane regions predicted for the COIII AA sequences of *H. catesbianae* and *P. falciparum*.

Mitochondrion-associated sequence 3 (mt-sequence 3) was 3523 bp in length and included a complete CytB CDS (CytB-1) that was 1110 bp in length and mapped from bases 1065 to 2174. The complete CytB-1 coding sequence was confirmed by a 1542 bp PCR product sequenced in both directions using Sanger sequencing. The final assembly was composed of 120 paired and seven single (on assembly ends) NGS reads and three Sanger sequencing reads with an overall one to 12-fold sequence coverage (MK214283). The CytB-1 CDS had a proposed TTG start codon, a proposed TAG stop codon and produced a predicted translation product of 369 AA. A DELTA-BLAST search of the translated CytB-1 AA sequence produced a 61% pairwise identity match (98% query coverage) with the mitochondrial CytB AA sequence of *K. equi* (AZL31647), a 61% pairwise identity match (97% query coverage) with the mitochondrial CytB AA sequence of *H. catesbianae* (AIG55089) and a 58–60% pairwise identity match (95–98% query coverage) to various haemosporinid mitochondrial CytB AA sequences. Transmembrane predictions showed that the *H. canis* CytB-1 AA sequence had the same nine transmembrane regions predicted for the CytB AA sequences of *H. catesbianae* and *P. falciparum*.

Mitochondrion-associated sequence 4 (mt-sequence 4) was 1762 bp in length and included a complete CytB CDS (CytB-2) that was 1149 bp in length and mapped from bases 150 to 1298. The complete CytB-2 coding sequence was confirmed by a 1418 bp PCR product and sequenced in both directions using Sanger sequencing. The final assembly was composed of 26 paired and seven single (on assembly ends) NGS reads and three Sanger sequencing reads with an overall one to eight-fold sequence coverage (MK214284). The CytB-2 CDS had a proposed TTG start codon, a proposed TAG stop codon and produced a predicted translation product of 382 AA. A DELTA-BLAST search of the translated CytB-2 AA sequence produced a 60% pairwise identity match (97% query coverage) with the mitochondrial CytB AA sequence of *K. equi* (AZL31647), a 59% pairwise identity match (93% query coverage) with the mitochondrial CytB AA sequence of *H. catesbianae* (AIG55089) and a 56–59% pairwise identity match (90–97% query coverage) to various haemosporinid mitochondrial CytB AA sequences. Transmembrane predictions showed that the *H. canis* CytB-2 translation had the same nine transmembrane regions predicted for the CytB AA sequences of *H. catesbianae* and *P. falciparum*. The CytB-2 CDS had a 78.8% pairwise identity to the CytB-1 CDS. The CytB-2 AA sequence had an 85.6% pairwise identity to the CytB-1 AA sequence.

Mitochondrion-associated sequence 5 (mt-sequence 5) was 2797 bp in in length and included a non-coding pseudo-CytB sequence that was 1120 bp in length and mapped from bases 1435 to 2554. The final assembly was composed of 94 paired and six single (on assembly ends) NGS reads with an overall one to 11-fold sequence coverage (MK214285). The pseudo-CytB CDS was determined to be non-coding due to one missense mutation and three frameshift mutations. Within the pseudo-CytB CDS, the 18th codon (bases 52–54) was a TAA stop codon as opposed to a CAA glutamine codon that was conserved in both CytB-1 and CytB-2 CDS. Following, there was a TT insert at bases 562 and 563, a G insert at base 660 and a T insert at base 715 that caused frameshift mutations observed when compared by pairwise alignments to CytB-1 and CytB-2 sequences. A hypothetical AA translation sequence was generated (372 AA) from a modified pseudo-CytB CDS edited to remove all frameshift mutations. A DELTA-BLAST search of the hypothetical pseudo-CytB AA sequence produced a 59% pairwise identity match (97% query coverage) with the mitochondrial CytB AA sequence of *K. equi* (AZL31647), a 56% pairwise identity match (95% query coverage) with the mitochondrial CytB AA sequence of *H. catesbianae* (AIG55089), a 55–57% pairwise identity match (93–94% query coverage) to various eimeriid mitochondrial CytB AA sequences, and a 53–56% pairwise identity match (93–97% query coverage) to various haemosporinid mitochondrial CytB AA sequences. The pseudo-CytB CDS had a 73.5% pairwise identity to the CytB-1 CDS and a 73.8% pairwise identity to the CytB-2 CDS. The hypothetical pseudo-CytB AA sequence had a 73.3% pairwise identity to the CytB-1 AA sequence and a 75.9% pairwise identity to the CytB-2 AA sequence.

Mitochondrion-associated sequence 6 (mt-sequence 6) was 976 bp in size and composed of 28 paired and 11 single NGS reads with a one to nine-fold sequence coverage. Three fragmented mt-rDNA (LSUC, SSUB, RNA14) were annotated on mt-sequence 6 (MH615003). Mitochondrion-associated sequence 7 (mt-sequence 7) was 948 bp in size and composed of 10 paired and 16 NGS reads with a one to eight-fold sequence coverage. Three fragmented mt-rDNA (LSUG, RNA10, RNA6) were annotated on mt-sequence 7 (MH615004). Mitochondrion-associated sequence 8 (mt-sequence 8) was 215 bp in size and composed of four single NGS reads with a one to-four fold sequence coverage. Three fragmented mt-rDNA (SSUA, RNA18, SSUD) were annotated on mt-sequence 8 (MH615005).

Forward and reverse primers were synthesized based on the mitochondrion-associated sequences and used in a wide variety of combinations in attempts to connect the mitochondrion-associated sequences using PCR amplifications. All such attempts to connect the mitochondrion-associated sequences by PCR were unsuccessful (data not shown).

3.4. Complete nu-rDNA unit sequence

Assembling the nu-rDNA unit sequence was also challenging. As observed with other members of the Apicomplexa (see Introduction), the nu-rDNA unit of *H. canis* included several divergent copies. Early analysis of an unrefined nu-rDNA unit assembly (6849 bp; 99% consensus; 102,158 paired-end reads; 1854 ± 670-fold coverage) revealed that the SSU 18S nu-rDNA sequence contained 20 positions with ambiguity (1.1% of 1816 bp; 1930 ± 505-fold coverage), the 5.8S nu-rDNA sequence contained two positions with ambiguity (1.3% of 157 bp; 770 ± 188-fold coverage) and the LSU 28S nu-rDNA sequence contained 45 positions with ambiguity (1.2% of 3731 bp; 2176 ± 320-fold coverage). The ITS sequences were more variable; the ITS1 region contained 40 positions with ambiguity (6.8% of 593 bp; 490 ± 174 (min: 136)-fold coverage and the ITS2 region contained 14 positions with ambiguity (4.8% of 291 bp; 937 ± 460 (min: 270)-fold coverage). The ITS regions were the most variable between nu-rDNA copies, causing them to have the lowest sequence coverage depth. Early assemblies revealed the presence of several indels being present among various ITS1 and ITS2 copies. The initial nu-rDNA unit assembly is available (see Supplementary Data S1 (doi: https://doi.org//10.17632/bs2z92449s.1)). No evidence of rDNA sequence belonging to any other apicomplexan protist was detected.

Further rounds of assembly refinements produced a complete nu-rDNA unit sequence (7500 bp) that was generated from 64,298 paired NGS reads (1077 ± 548-fold coverage; average insert size 510 bp, range 250–750 bp) containing no positions with ambiguity (MH615006). The nu-rDNA unit contained five rRNA coding sequences (18S-ITS1-5.8S-ITS2-28S; Table 6) and represented a reference assembly of the most prevalent copy of the nu-rDNA unit in the NGS dataset. As with prior assemblies, the ITS regions composed the areas of lowest sequence coverage in the final assembly. The ITS1 had 244 ± 107 (min: 61)-fold sequence coverage and the ITS2 had 559 ± 358 (min: 16)-fold sequence coverage.

Table 6
Nuclear rRNA gene sequences annotated on the reference sequence of the *Hepatozoon canis* complete rDNA unit sequence (7500 bp; GenBank Accession number MH615006) were assigned by comparison with the same regions of several related apicomplexan parasites (i.e. *Eimeria tenella* (AF026388), *Toxoplasma gondii* (L25635) and *Babesia microti* (AB190287)).

Nuclear rRNA gene	Size (bp)	Location (bp)
Small subunit (SSU rRNA, 18S)	1816	492 → 2307
Internal transcribed spacer 1 (ITS1)	593	2308 → 2900
5.8S ribosomal RNA (5.8S rRNA)	157	2901 → 3057
Internal transcribed spacer 2 (ITS2)	291	3058 → 3307
Large subunit (LSU rRNA, 28S)	3731	3349 → 7079

A BLAST search of the 18S nu-rDNA sequence produced a 99% pairwise identity match to many 18S nu-rDNA sequences identified as representing *H. canis* (e.g. AY461378, DQ439540, KU893118; 96% query coverage). There are currently no adeleorinid 28S nu-rDNA sequences available on GenBank. A BLAST search of the 28S nu-rDNA sequence produced an 88% pairwise identity match to *Sarcocystis singaporensis* (AF237617; 74% query coverage) and 85% pairwise identity match to *Cryptosporidium parvum* (AF040725; 82% query coverage) 28S nu-rDNA sequences. The *H. canis* 28S nu-rDNA sequence had two large insertions not present in *C. parvum* and three insertions not present in *S. singaporensis*.

3.5. Multiple variant copies of the small subunit 18S nu-rDNA sequence

The SSU 18S nu-rDNA sequence is important in the molecular diagnostics of *H. canis* infections; therefore further analysis was carried out on the ambiguous nucleotides found in the original assembly of this region. Ten distinct 18S nu-rDNA sequences were generated for *H. canis* (Table 7) belonging to three types that had up to 6.4% divergence between types but only up to 0.3% divergence among variants within one type (Table 8). The type-3 sequences each possessed the same 54 bp insert located within unpaired region 33 of the secondary structure model of *T. gondii* 18S nu-rDNA (see Clark, 1987; Gagnon et al., 1996).

Table 8
Pairwise sequence identity matrix of *Hepatozoon canis* 18S nuclear-rDNA sequence types and subtypes. Pairwise identity values were generated in Geneious using the 'Multiple Align' tool (cost matrix: 93%, gap open penalty: 12, gap extension penalty: 3; www.geneious.com, Kearse et al., 2012).

H. canis 18S nu-rDNA	Type-1 subtypes	Type-2 subtypes	Type-3 subtypes
Type-1 subtypes	99.7–99.9%	97.8–97.9%	95.4–95.5%
Type-2 subtypes		99.9%	93.6–93.7%
Type-3 subtypes			99.9%

4. Discussion

The apicoplast genome arrangement of *H. canis* was found to be conserved generally with known haemosporinid parasites (e.g. Wilson et al., 1996; Imura et al., 2014; but excluding the apicoplast of *Plasmodium chabaudi chabaudi* described by Sato et al., 2013), eimeriid coccidia (e.g. Cai et al., 2003) and sarcocystid coccidia (Seeber et al., 2014). With the exception of missing genes (tDNAs: alanine, aspartic acid, histidine, threonine and valine; CDS: rps 19) and the highly divergent rps 2 and rps 7 CDS, the complement of genes of the apicoplast of *H. canis* matched available apicoplast genomes from other apicomplexan protists. It is not known whether these missing genes would have an impact on the function of the apicoplast.

Unexpectedly, we could not confirm that the *H. canis* apicoplast genome had a circular or linear-concatenated structure despite the 'circular-mapping' nature of its apicoplast genome sequence. Substantial NGS paired-read data supported the conclusion that the apicoplast genome of *H. canis* was physically linear (see Supplementary Fig. S1). If the apicoplast genome of *H. canis* was either predominantly in circular form (e.g. *P. falciparum*, see Williamson et al., 2002) or linear concatemeric form (e.g. *T. gondii*, see Williamson et al., 2001), NGS paired-reads should have been detected spanning from the 3′-end to the 5′-end of the apicoplast genome sequence that we obtained.

Despite the conserved structure of many apicoplast genomes, the majority of the genes encoded on the apicoplast of *H. canis* and other apicomplexan parasites have high sequence divergence.

Table 7
Ten distinct 18S nuclear-rDNA sequences were identified for *Hepatozoon canis* belonging to three major types. Minimum base coverage reflects the prevalence of each subtype in the dataset. Subtype names identify single nucleotide differences from the associated major copy. The type-3 sequences each possessed the same 54 bp insert that was unique from other 18S nu-rDNA types.

H. canis 18S nu-rDNA subtype	Length (bp)	Minimum (mean) base coverage	Highest BLAST Match (max score; coverage; identity)	GenBank Accession No.
Type-1 (major copy)	1816	692 (1293 ± 391)	AY461378.2 – *H. canis* isolate Spain 2 18S rDNA (3208; 96%; 99%)	MH615006
Type-1-C541	1816	500 (1232 ± 389)	AY461378.2 – *H. canis* isolate Spain 2 18S rDNA (3203; 96%; 99%)	MK091084
Type-1-T509	1816	454 (1234 ± 396)	AY461378.2 – *H. canis* isolate Spain 2 18S rDNA (3203; 96%; 99%)	MK091086
Type-1-G1388	1816	264 (1198 ± 406)	AY461378.2 – *H. canis* isolate Spain 2 18S rDNA (3203; 96%; 99%)	MK091085
Type-1-T113-T797-T1382	1816	147 (988 ± 384)	AY461378.2 – *H. canis* isolate Spain 2 18S rDNA (3192; 96%; 99%)	MK091087
Type-1-T113-T178-A267-T797-T1382	1816	101 (902 ± 413)	AY461378.2 – *H. canis* isolate Spain 2 18S rDNA (3182; 96%; 99%)	MK091088
Type-2 (major copy)	1818	90 (189 ± 70)	KU893118.1 – *H. canis* isolate fox 1–2 18S rDNA (3009; 96%; 98%)	MK091089
Type-2- G282	1818	32 (174 ± 83)	KU893118.1 – *H. canis* isolate fox 1–2 18S rDNA (3014; 96%; 98%)	MK091090
Type-3 (major copy)	1874	176 (345 ± 160)	AY461378.2 – *H. canis* isolate Spain 2 18S rDNA (2119; 95%; 99%)	MK091091
Type-3-T146	1874	104 (338 ± 158)	AY461378.2 – *H. canis* isolate Spain 2 18S rDNA (2113; 95%; 99%)	MK091092

Most of the genes show too much sequence variation among major apicomplexan groups to be useful for universal primer design. As demonstrated by Lang-Unnasch et al. (1998), the 16S or 23S pl-rDNA sequences contained several conserved regions highly suited for phylum-wide primer design. Between these conserved regions are highly variable sequences. Therefore, these pl-rDNA targets could be ideal candidates for expanding the molecular tools available for studying the Apicomplexa.

Attempts to sequence the complete mitochondrial genome of *H. canis* were unsuccessful. It is uncertain why this was the case when there were no issues with generating complete sequences of the *H. canis* apicoplast genome and complete nu-rDNA sequence from the NGS paired-read dataset; both fold-coverage and overlap among paired reads allowed for robust assemblies of both the apicoplast and nuclear ribosomal sequences. Based on estimations made by flow cytometry and Southern blotting, the mitochondrial genomes of several *Plasmodium* spp. were present at 15 to 150 copies per cell whereas the apicoplast genome was present at only one to two copies per cell (reviewed in Wilson and Williamson, 1997). If a similar copy number ratio was conserved in *H. canis*, the assembly of the mitochondrial genome should have had a sequence coverage at least 10-fold that of the apicoplast genome. Since this was not observed, it is possible that *H. canis* does not encode a separate/discrete complete mitochondrial genome. All mitochondrion-associated sequences had similar sequence coverage 1–10 fold over their lengths indicating that if a complete mitochondrial genome was present it is at extremely low copy number, at least within the intraerythrocytic life cycle stages (gamonts) contributing to the NGS sequence data. Despite assembling eight mitochondrion-associated sequences totalling almost 15 kbp of sequence data confirmed with PCR amplifications covering various CDS regions, no evidence of an intact mitochondrial genome was found in *H. canis*.

Apicomplexan parasites are notable for their variation in mitochondrial genome structure. Despite this variation, most apicomplexan mitochondrial genomes have been found to be greatly reduced in size (frequently 6–7 kbp but some larger) and encode only three protein-coding genes: COI, COIII, and CytB. In addition, each mitochondrial genome also contains multiple fragments of LSU and SSU mt-rDNA scattered throughout each mitochondrial genome (for review see Feagin, 1994; Wilson and Williamson, 1997). Thus, most apicomplexan mitochondrial genomes could be expected to possess three protein-coding genes plus multiple fragmented mt-rDNA, usually in a circular-mapping (i.e. physically circular or linear concatemers) arrangement. Despite conservation of content, the structure and organization of apicomplexan mitochondrial genomes can be variable (see Léveillé et al., 2014). Differences in mitochondrial genome structure are most evident between major recognized groups within the Apicomplexa (e.g. between haemosporinids and eimeriid coccidia) with limited variation within a major apicomplexan group (i.e. among eimeriid coccidia). Mitochondrial genomes from other adeleorinid parasites (*H. catesbianae* and *K. equi*) both possess a complete circular-mapping genome that has COI, COIII and CytB coding sequences plus numerous fragmented mt-rDNA (see Léveillé et al., 2014, 2019). Based on the demonstrated relatedness of *H. canis* to other adeleorinid parasites (e.g. Barta et al., 2012; Karadjian et al., 2015), it would have been reasonable to expect that the mitochondrial genome of *H. canis* would be similar to those of other adeleorinid coccidia; clearly this was not observed in the present study.

The mitochondrion-associated CDS translation products of *H. canis* were most similar to the CDS translation products from the mitochondrial genomes of other adeleorinid coccidia, *H. catesbianae* and *K. equi*, as well as various haemosporinid parasites but with considerable and almost equal sequence divergence from both (see Section 3.3; Léveillé et al., 2014, 2019). The considerable genetic divergence observed among the few adeleorinid mitochondrial CDS reported to date strongly suggests that diversity among mitochondrial genomes from adeleorinid coccidia has only started to be explored.

Complete or partial loss of a 'typical' apicomplexan mitochondrial genome has been well documented in the Apicomplexa. *Cryptosporidium* spp. were found to lack a mitochondrial genome (Abrahamsen et al., 2004; Xu et al., 2004); similarly, mitochondrial genomes have not been identified from the few gregarines subjected to whole genome sequencing attempts (e.g. Templeton et al., 2010). Even among the closely related sarcocystid coccidia, attempts to sequence a complete mitochondrial genome for *T. gondii* and related sarcocystid coccidia have been complicated by the presence of many CytB and COI pseudogenes scattered throughout the nuclear genomes of these parasites (Ossorio et al., 1991; Gjerde, 2013; Ogedengbe et al., 2016). There are currently no complete mitochondrial genome sequences available for *T. gondii* and other sarcocystid relatives (Gjerde, 2013; Seeber et al., 2014).

The lack of a complete mitochondrial genome in some apicomplexan taxa does not necessarily result in the loss of the physical organelle. For example, *C. parvum* does retain a mitochondrion-like organelle. This double-membrane bound organelle was smaller than typical apicomplexan mitochondria but immunolocalization showed that the organelle contained some mitochondrion-associated proteins (see Putignani et al., 2004). Similarly, *T. gondii* retains a typical large mitochondrion containing tubular cristae (Figs. 3, 5 from Ferguson et al., 2005). *Hepatozoon canis* is believed to retain mitochondrion-like organelles that have been observed using transmission electron microscopy (e.g. Fig. 3 from Droleskey et al., 1993; Fig. 45 from Baneth et al., 2007; and, Figs. 11, 13 from *H. americanum*, Cummings et al., 2005). *Hepatozoon catesbianae*, with its typical compact mitochondrial genome (Léveillé et al., 2014), has mitochondria with ultrastructural features similar to other coccidia (see Figs. 12, 21, 22 from Desser et al., 1995).

If a complete mitochondrial genome is not present in *H. canis*, the mitochondrion-associated sequences generated in the present work could be genes of mitochondrial origin translocated to the nuclear genome. There were several lines of evidence to support this supposition. Although mitochondrion-associated coding regions were identified in the present study, most of these coding regions were flanked by extensive non-coding DNA sequences with no apparent relationship to any apicomplexan mitochondrial genome. Multiple PCR-based attempts to link up the mitochondrion-associated sequences that were obtained were conducted; no amplification products were produced, suggesting that these sequences were not found on a typically compact mitochondrial genome. Finally, the detection of two coding and one non-coding copies of a CytB CDS with considerable divergence among all copies (see Section 3.3, Fig. 3) supports the suggestion of gene transfer to the nuclear genome. Therefore, it is suggested that *H. canis* might not encode a complete organelle-localized mitochondrial genome and instead at least some mitochondrial genes have been translocated and dispersed throughout the nuclear genome. Future analyses will be necessary to confirm or dispute this interpretation.

The 18S nu-rDNA loci continue to be one of the easiest molecular targets to amplify from organisms in the Apicomplexa. The abundance of 18S nu-rDNA sequences available on public databases reflects the utility of this genetic locus as a diagnostic target as well as a starting point for phylogenetic studies. The 18S nu-rDNA remains an effective locus for initial molecular diagnostic protocols and identification of apicomplexan genera. However, considerable intraspecific variation was detected among nu-rDNA sequences in the present study. Indeed, three major 18S nu-rDNA types, perhaps belonging to independent nu-rDNA array copies,

were detected in the NGS-based assemblies. Such paralogous copies have been observed in *Plasmodium* spp. (see Gunderson et al., 1987) with stage-specific expression of the various paralogues giving rise to distinct 18S rRNA associated with particular life cycle stages. Divergent, paralogous 18S nu-rDNA copies within one apicomplexan species and limited sequence divergence of orthologs among species combine to make 18S nu-rDNA less appropriate for species delimitation (see Ross et al., 2008) for parasites in the phylum than other loci. All of the associated nu-rDNA targets (i.e. 18S, ITS1, 5.8S, ITS2, 28S) that form units with the 18S nu-rDNA would likely be subject to the same complications and should be used with care in species-level phylogenetic or taxonomic studies in the Apicomplexa. The diversity observed in the NGS-derived rDNA sequences in the present study (see Sections 3.4 and 3.5) suggest that rDNA sequences are likely to be ineffective for uncovering potential cryptic species among the canid-infecting *Hepatozoon* spp.

In conclusion, the diversity of nu 18S nu-rDNA sequences detected in the present study (10 subtypes detected within three divergent types) highlights the limited utility of this genetic locus for species level identifications. Novel sequences from extrachromosomal genomes may be better suited for uncovering the diversity of *Hepatozoon* spp. capable of infecting canids. Despite not sequencing a complete mitochondrial genome, the availability of mitochondrion-associated CDS from *H. canis* provides novel genetic targets that may prove useful for revealing undiscovered species and understanding the genetic diversity of canid-infecting *Hepatozoon* spp. Similarly, the sequenced apicoplast genome offers novel molecular targets that may prove useful for molecular phylogenetic applications. The divergence of the putative mitochondrial "genome" of *H. canis* from other adeleorinid coccidia sequenced to date is perplexing but may reflect the tremendous biological and genetic diversity displayed by parasites in the Adeleorina.

Acknowledgements

Dr. Galit Gichon-Levin is thanked for providing the blood sample used in this study. This research was supported by a Discovery Grant to J.R.B. from the Natural Sciences and Engineering Research Council of Canada (400566). A.N.L. was supported by a Queen Elizabeth II Graduate Scholarship in Science and Technology and an Ontario Graduate Scholarship from the Government of Ontario, and an OVC Scholarship from the Ontario Veterinary College, University of Guelph, Canada.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijpara.2018.12.001.

References

Abrahamsen, M.S., Templeton, T.J., Enomoto, S., Abrahante, J.E., Zhu, G., Lancto, C.A., Deng, M., Liu, C., Widmer, G., Tzipori, S., Buck, G.A., Xu, P., Bankier, A.T., Dear, P. H., Konfortov, B.A., Spriggs, H.F., Iyer, L., Anantharaman, V., Aravind, L., Kapur, V., 2004. Complete genome sequence of the apicomplexan, *Cryptosporidium parvum*. Science 304, 441–445.

Allen, K.E., Li, Y., Kaltenboeck, B., Johnson, E.M., Reichard, M.V., Panciera, R.J., Little, S.E., 2008. Diversity of *Hepatozoon* species in naturally infected dogs in the southern United States. Vet. Parasitol. 154, 220–225.

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.

Baneth, G., Shkap, V., Presentey, B.Z., Pipano, E., 1996. *Hepatozoon canis*: the prevalence of antibodies and gametocytes in dogs in Israel. Vet. Res. Commun. 20, 41–46.

Baneth, G., Weigler, B., 1997. Retrospective case-control study of hepatozoonosis in dogs in Israel. J. Vet. Intern. Med. 11, 365–370.

Baneth, G., Barta, J.R., Shkap, V., Martin, D.S., Macintire, D.K., Vincent-Johnson, N., 2000. Genetic and antigenic evidence supports the separation of *Hepatozoon canis* and *Hepatozoon americanum* at the species level. J. Clin. Microbiol. 38, 1298–1301.

Baneth, G., Samish, M., Shkap, V., 2007. Life cycle of *Hepatozoon canis* (Apicomplexa: Adeleorina: Hepatozoidae) in the tick *Rhipicephalus sanguineus* and domestic dog (*Canis familiaris*). J. Parasitol. 93, 283–299.

Baneth, G., Sheiner, A., Eyal, O., Hahn, S., Beaufils, J.P., Anug, Y., Talmi-Frank, D., 2013. Redescription of *Hepatozoon felis* (Apicomplexa: Hepatozoidae) based on phylogenetic analysis, tissue and blood form morphology, and possible transplacental transmission. Parasit Vectors 6, 1–10.

Boratyn, G.M., Schäffer, A.A., Agarwala, R., Altschul, S.F., Lipman, D.J., Madden, T.L., 2012. Domain enhanced lookup time accelerated BLAST. Biol. Direct. 2012, 1–14. https://doi.org/10.1186/1745-6150-7-12.

Barta, J.R., 1989. Phylogenetic analysis of the class Sporozoea (phylum Apicomplexa Levine, 1970): evidence for the independent evolution of heteroxenous life cycles. J. Parasitol. 75, 195–206.

Barta, J.R., 2000. Adeleorina. In: Lee, J.J., Leedale, G.F., Bradbury, P. (Eds.), An Illustrated Guide to the Protozoa. 2nd ed. Allen Press Inc., Lawrence, pp. 305–318.

Barta, J.R., Ogedengbe, J.D., Martin, D.S., Smith, T.G., 2012. Phylogenetic position of the adeleorinid coccidia (Myzozoa, Apicomplexa, Coccidia, Eucoccidiorida, Adeleorina) inferred using 18S rDNA sequences. J. Eukaryot. Microbiol. 59, 171–180.

Cai, X., Fuller, A.L., McDougald, L.R., Zhu, G., 2003. Apicoplast genome of the coccidian *Eimeria tenella*. Gene 321, 39–46.

Cardoso, L., Cortes, H.C., Eyal, O., Reis, A., Lopes, A., Vila-Viçosa, M., Rodrigues, P.A., Baneth, G., 2014. Molecular and histopathological detection of *Hepatozoon canis* in red foxes (*Vulpes vulpes*) from Portugal. Parasit. Vectors 7, 113.

Clark, C.G., 1987. On the evolution of ribosomal RNA. J. Mol. Evol. 25, 343–350.

Cummings, C.A., Panciera, R.J., Kocan, K.M., Mathew, J.S., Ewing, S.A., 2005. Characterization of stages of *Hepatozoon americanum* and of parasitized canine host cells. Vet. Pathol. 42, 788–796.

Desser, S.S., Hong, H., Martin, D.S., 1995. The life history, ultrastructure, and experimental transmission of *Hepatozoon catesbianae* n. comb., an apicomplexan parasite of the bullfrog, *Rana catesbeiana* and the mosquito, *Culex territans* in Algonquin Park. Ontario. J. Parasitol. 81, 212–222.

Droleskey, R.E., Mercer, S.H., DeLoach, J.R., Craig, T.M., 1993. Ultrastructure of *Hepatozoon canis* in the dog. Vet. Parasitol. 50, 83–99.

Elias, E., Homans, P.A., 1988. *Hepatozoon canis* infection in dogs: clinical and haematological findings; treatment. J. Small Anim. Pract. 29, 55–62.

Ezeokoli, C.D., Ogunkoya, A.B., Abdullahi, R., Tekdek, L.B., Sannusi, A., Ilemobade, A. A., 1983. Clinical and epidemiological studies on canine hepatozoonosis in Zaria, Nigeria. J. Small Anim. Pract. 24, 455–460.

Farkas, R., Solymosi, N., Takács, N., Hornyák, Á., Hornok, S., Nachum-Biala, Y., Baneth, G., 2014. First molecular evidence of *Hepatozoon canis* infection in red foxes and golden jackals from Hungary. Parasit. Vectors 7, 303.

Feagin, J., 1994. The extrachromosomal DNAs of apicomplexan parasites. Annu. Rev. Microbiol. 48, 81–104.

Feagin, J.E., Harrell, M.I., Lee, J.C., Coe, K.J., Sands, B.H., Cannone, J.J., Tami, G., Schnare, M.N., Gutell, R.R., 2012. The fragmented mitochondrial ribosomal RNAs of *Plasmodium falciparum*. PLoS One 7,. https://doi.org/10.1371/journal.pone.0038320 e38320.

Ferguson, D.J.P., Henriquez, F.L., Kirisits, M.J., Muench, S.P., Prigge, S.T., Rice, D.W., Roberts, C.W., McLeod, R.L., 2005. Maternal inheritance and stage-specific variation of the apicoplast in *Toxoplasma gondii* during development in the intermediate and definitive host. Eukaryot. Cell 4, 814–826.

Finn, R.D., Clements, J., Eddy, S.R., 2011. HMMER web server: Interactive sequence similarity searching. Nucleic Acids Res. 39, 29–37.

Gabrielli, S., Kumlien, S., Calderini, P., Brozzi, A., Iori, A., Cancrini, G., 2010. The first report of *Hepatozoon canis* identified in *Vulpes vulpes* and ticks from Italy. Vector Borne Zoonotic Dis. 10, 855–859.

Gagnon, S., Bourbeau, D., Levesque, R.C., 1996. Secondary structures and features of the 18S, 5.8S and 26S ribosomal RNAs from the apicomplexan parasite *Toxoplasma gondii*. Gene 173, 129–135.

Giannelli, A., Lia, R.P., Annoscia, G., Buonavoglia, C., Lorusso, E., Dantas-Torres, F., Baneth, G., Otranto, D., 2017. *Rhipicephalus turanicus*, a new vector of *Hepatozoon canis*. Parasitology 144, 730–737.

Gish, W., States, D.J., 1993. Identification of protein coding regions by database similarity search. Nat. Genet. 3, 266–272.

Gjerde, B., 2013. Characterisation of full-length mitochondrial copies and partial nuclear copies (numts) of the cytochrome b and cytochrome c oxidase subunit I genes of *Toxoplasma gondii*, *Neospora caninum*, *Hammondia heydorni* and *Hammondia triffittae* (Apicomplexa: Sarcocystidae). Parasitol. Res. 112, 1493–1511.

Goethert, H.K., Cook, J.A., Lance, E.W., Telford, S.R., 2006. Fay and Rausch 1969 revisited: *Babesia microti* in Alaskan small mammals. J. Parasitol. 92, 826–831.

Gunderson, J.H., Sogin, M.L., Wollett, G., Hollingdale, M., De La Cruz, V.F., Waters, A. P., McCutchan, T.F., 1987. Structurally distinct, stage-specific ribosomes occur in *Plasmodium*. Science 238, 933–937.

Hebert, P.D.N., Cywinska, A., Ball, S.L., DeWaard, J.R., 2003. Biological identifications through DNA barcodes. Proc. R. Soc. B Biol. Sci. 270, 313–321.

Imura, T., Sato, S., Sato, Y., Sakamoto, D., Isobe, T., Murata, K., Holder, A.A., Yukawa, M., 2014. The apicoplast genome of *Leucocytozoon caulleryi*, a pathogenic apicomplexan parasite of the chicken. Parasitol. Res. 113, 823–828.

James, S.P., 1905. On a parasite found in the white corpuscles of the blood of dogs. Sci. Mem. Off. Med. Sanit. Dep. Gov. India 14, 1–12.
Karadjian, G., Chavatte, J.-M., Landau, I., 2015. Systematic revision of the adeleid haemogregarines, with creation of *Bartazoon* n. g., reassignment of *Hepatozoon argantis* Garnham, 1954 to *Hemolivia*, and molecular data on *Hemolivia stellata*. Parasite 22, 31.
Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., et al., 2012. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28, 1647–1649.
Kistler, W.M., Brown, J.D., Allison, A.B., Nemeth, N.M., Yabsley, M.J., 2014. First report of *Angiostrongylus vasorum* and *Hepatozoon* from a red fox (*Vulpes vulpes*) from West Virginia, USA. Vet. Parasitol. 200, 216–220.
Lang-Unnasch, N., Reith, M.E., Munholland, J., Barta, J.R., 1998. Plastids are widespread and ancient in parasites of the phylum Apicomplexa. Int. J. Parasitol. 28, 1743–1754.
Léveillé, A.N., Bland, S.K., Carlton, K., Larouche, C.B., Kenney, D.G., Brouwer, E.R., Lillie, B.N., Barta, J.R., 2019. *Klossiella equi* infecting kidneys of Ontario horses: life cycle features and multi locus sequence based genotyping confirm the genus *Klossiella* belongs in the Adeleorina (Apicomplexa: Coccidia). J. Parasitol. 105, 29–40.
Léveillé, A.N., Ogedengbe, M.E., Hafeez, M.A., Tu, H.H.A., Barta, J.R., 2014. The complete mitochondrial genome sequence of *Hepatozoon catesbianae* (Apicomplexa: Coccidia: Adeleorina), a blood parasite of the green frog, *Lithobates* (formerly *Rana*) *clamitans*. J. Parasitol. 100, 651–656.
Li, Y., Wang, C., Allen, K.E., Little, S.E., Ahluwalia, S.K., Gao, D., Macintire, D.K., Blagburn, B.L., Kaltenboeck, B., 2008. Diagnosis of canine *Hepatozoon* spp. infection by quantitative PCR. Vet. Parasitol. 157, 50–58.
Little, S.E., Allen, K.E., Johnson, E.M., Panciera, R.J., Reichard, M.V., Ewing, S.A., 2009. New developments in canine hepatozoonosis in North America: a review. Parasit. Vectors 2, S5.
Lowe, T.M., Chan, P.P., 2016. tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. Nucleic Acids Res. 44, 54–57.
McCully, R.M., Basson, P.A., Bigalke, R.D., De Vos, V., Young, E., 1975. Observations on naturally acquired hepatozoonosis of wild carnivores and dogs in the Republic of South Africa. Onderstepoort J. Vet. Res. 42, 117–133.
McCutchan, T.F., de la Cruz, V.F., Lal, A.A., Gunderson, J.H., Elwood, H.J., Sogin, M.L., 1988. Primary sequences of two small subunit ribosomal RNA genes from *Plasmodium falciparum*. Mol. Biochem. Parasitol. 28, 63–68.
Modrý, D., Beck, R., Hrazdilová, K., Baneth, G., 2017. A review of methods for detection of *Hepatozoon* infection in carnivores and arthropod vectors. Vector-Borne Zoonotic Dis. 17, 66–72.
Morrison, D.A., Bornstein, S., Thebo, P., Wernery, U., Kinne, J., Mattsson, J.G., 2004. The current status of the small subunit rRNA phylogeny of the coccidia (Sporozoa). Int. J. Parasitol. 34, 501–514.
Murata, T., Shiramizu, K., Hara, Y., Inoue, M., Shimoda, K., Nakama, S., 1991. First case of *Hepatozoon canis* infection of a dog in Japan. J. Vet. Med. Sci. 53, 1097–1099.
O'Dwyer, L.H., Massard, C.L., Pereira De Souza, J.C., 2001. *Hepatozoon canis* infection associated with dog ticks of rural areas of Rio de Janeiro State, Brazil. Vet. Parasitol. 94, 143–150.
Ogedengbe, M.E., El-Sherry, S., Ogedengbe, J.D., Chapman, H.D., Barta, J.R., 2018. Phylogenies based on combined mitochondrial and nuclear sequences conflict with morphologically defined genera in the eimeriid coccidia (Apicomplexa). Int. J. Parasitol. 48, 59–69.
Ogedengbe, J.D., Hanner, R.H., Barta, J.R., 2011. DNA barcoding identifies *Eimeria* species and contributes to the phylogenetics of coccidian parasites (Eimeriorina, Apicomplexa, Alveolata). Int. J. Parasitol. 41, 843–850.
Ogedengbe, M.E., Ogedengbe, J.D., Whale, J.C., Elliot, K., Juárez-Estrada, M.A., Barta, J. R., 2016. Molecular phylogenetic analyses of tissue coccidia (sarcocystidae; apicomplexa) based on nuclear 18s rDNA and mitochondrial COI sequences confirms the paraphyly of the genus *Hammondia*. Parasitol. Open 2 (2e), 1–16. https://doi.org/10.1017/pao.2015.7.
Ossorio, P.N., Sibley, L.D., Boothroyd, J.C., 1991. Mitochondrial-like DNA sequences flanked by direct and inverted repeats in the nuclear genome of *Toxoplasma gondii*. J. Mol. Biol. 222, 525–536.
Putignani, L., Tait, A., Smith, H.V., Horner, D., Tovar, J., Tetley, L., Wastling, J.M., 2004. Characterization of a mitochondrion-like organelle in *Cryptosporidium parvum*. Parasitology 129, 1–18.
Rajamanickam, C., Wiesenhutter, E., Zin, F.M., Hamid, J., 1985. The incidence of canine haematozoa in peninsular Malaysia. Vet. Parasitol. 17, 151–157.
Ross, H.A., Murugan, S., Li, W.L.S., 2008. Testing the reliability of genetic methods of species identification via simulation. Syst. Biol. 57, 216–230.
Rubini, A.S., Paduan, K.S., Martins, T.F., Labruna, M.B., O'Dwyer, L.H., 2009. Acquisition and transmission of *Hepatozoon canis* (Apicomplexa: Hepatozoidae) by the tick *Amblyomma ovale* (Acari: Ixodidae). Vet. Parasitol. 164, 324–327.
Sato, S., Sesay, A.K., Holder, A.A., 2013. The unique structure of the apicoplast genome of the rodent malaria parasite *Plasmodium chabaudi chabaudi*. PLoS One 8, 1–10. https://doi.org/10.1371/journal.pone.0061778.
Seeber, F., Feagin, J.E., Parsons, M., 2014. The apicoplast and mitochondrion of *Toxoplasma gondii*. In: Weiss, L.M., Kim, K. (Eds.), *Toxoplasma gondii* (Second Edition): The Model Apicomplexan – Perspectives and Methods. Academic Press, London, Waltham and San Diego, pp. 297–350.
Siddall, M.E., 1995. Phylogeny of Adeleid blood parasites with a partial systematic revision of the haemogregarine complex. J. Eukaryot. Microbiol. 42, 116–125.
Smith, T.G., Kim, B., Hong, H., Desser, S.S., 2000. Intraerythrocytic development of species of *Hepatozoon* infecting ranid frogs: evidence for convergence of life cycle characteristics among apicomplexans. J. Parasitol. 86, 451–458.
Smith, T., 1996. The genus *Hepatozoon* (Apicomplexa: Adeleina). J. Parasitol. 82, 565–585.
Tabar, M.D., Altet, L., Francino, O., Sánchez, A., Ferrer, L., Roura, X., 2008. Vector-borne infections in cats: molecular study in Barcelona area (Spain). Vet. Parasitol. 151, 332–336.
Templeton, T.J., Enomoto, S., Chen, W.J., Huang, C.G., Lancto, C.A., Abrahamsen, M.S., Zhu, G., 2010. A genome-sequence survey for *Ascogregarina taiwanensis* supports evolutionary affiliation but metabolic diversity between a Gregarine and *Cryptosporidium*. Mol. Biol. Evol. 27, 235–248.
Turner, D.H., Mathews, D.H., 2009. NNDB: The nearest neighbor parameter database for predicting stability of nucleic acid secondary structure. Nucleic Acids Res. 38, 2009–2011.
Vincent-Johnson, N.A., Macintire, D.K., Lindsay, D.S., Lenz, S.D., Baneth, G., Shkap, V., Blagburn, B.L., 1997. A new *Hepatozoon* species from dogs: description of the causative agent of canine hepatozoonosis in North America. J. Parasitol. 83, 1165–1172.
Wenyon, C.M., 1926. Protozoology: A Manual for Medical Men, Veterinarians and Zoologists. William Wood and Company, New York.
Williamson, D.H., Denny, P.W., Moore, P.W., Sato, S., McCready, S., Wilson, R.J.M.I., 2001. The in vivo conformation of the plastid DNA of *Toxoplasma gondii*: implications for replication. J. Mol. Biol. 306, 159–168.
Williamson, D.H., Preiser, P.R., Moore, P.W., McCready, S., Strath, M., Wilson, R.J.M.I., 2002. The plastid DNA of the malaria parasite *Plasmodium falciparum* is replicated by two mechanisms. Mol. Microbiol. 45, 533–542.
Wilson, R.J., Denny, P.W., Preiser, P.R., Rangachari, K., Roberts, K., Roy, A., Whyte, A., Strath, M., Moore, D.J., Moore, P.W., Williamson, D.H., 1996. Complete gene map of the plastid-like DNA of the malaria parasite *Plasmodium falciparum*. J. Mol. Biol. 261, 155–172.
Wilson, R.J., Williamson, D.H., 1997. Extrachromosomal DNA in the Apicomplexa. Microbiol. Mol. Biol. Rev. 61, 1–16.
Xu, P., Widmer, G., Wang, Y., Ozaki, L., 2004. The genome of *Cryptosporidium hominis*. Nature 431, 1107–1112.
Zhao, X., Duszynski, D.W., Loker, E.S., 2001. Phylogenetic position of *Eimeria antrozoi*, a bat coccidium (Apicomplexa: Eimeriidae) and its relationship to morphologically similar *Eimeria* spp. from bats and rodents based on nuclear 18S and plastid 23S rDNA sequences. J. Parasitol. 87, 1120–1123.