A linear mitochondrial genome of *Cyclospora cayetanensis* (Eimeriidae, Eucoccidiorida, Coccidiasina, Apicomplexa) suggests the ancestral start position within mitochondrial genomes of eimeriid coccidia

Mosun E. Ogedengbe[a], **Yvonne Qvarnstrom**[b], **Alexandre J. da Silva**[b], **Michael J. Arrowood**[b], and **John R. Barta**[a,*]

[a]Department of Pathobiology University of Guelph, Guelph ON Canada

[b]United States Centers for Disease Control and Prevention, Atlanta, GA, USA

Abstract

The near complete mitochondrial (mt)genome for *Cyclospora cayetanensis* is 6184 bp in length with three protein-coding genes (Cox1, Cox3, CytB) and numerous lsrDNA and ssrDNA fragments. Gene arrangements were conserved with other coccidia in the Eimeriidae, but the *C. cayetanensis* mt genome is not circular-mapping. Terminal transferase tailing and nested PCR completed the 5'-terminus of the genome starting with a 21bp A/T-only region that forms a potential stem-loop. Regions homologous to the *C. cayetanensis* mt genome 5'-terminus are found in all eimeriid mt genomes available and suggest this may be the ancestral start of eimeriid mt genomes.

Cyclospora cayetanensis is a protistan disease agent of humans that has been responsible for waterborne and large scale foodborne outbreaks worldwide; this apicomplexan parasite is now recognized as an emerging intestinal pathogen of public health importance (Karanja et al., 2007). In developed countries, the protist has been incriminated in diarrheal illnesses linked to imported foods (fresh uncooked vegetables and soft skinned fruits) and has been associated with diarrhea acquired through increased travel to endemic tropical regions (Shields and Olson, 2003). Particularly in the U.S., the two most recent diarrheal outbreaks caused by *C. cayetanensis* were solely associated with the consumption of fresh produce and affected 631 and 304 persons in 2013 and 2014, respectively (http://www.cdc.gov/parasites/cyclosporiasis/outbreaks/index.html). *Cyclospora cayetanensis* was first reported in stools of

*Corresponding author. J. R. Barta, Department of Pathobiology,University of Guelph, 50 Stone Road East, Guelph ON Canada. Tel.: +1-519-824-4120 x54017; fax +1-519-824-5930. jbarta@uoguelph.ca.

individuals suffering from protracted intermittent watery non-bloody diarrhea (CDC Report, 1991). Ortega et al. (1993) first identified the agent as an apicomplexan protist on the basis of oocyst morphology following sporulation and later on intracellular developmental stages. The organism was concluded to belong to the coccidian genus *Cyclospora* (Schneider, 1881), family Eimeriidae (Minchin, 1903) based on morphology of the sporulated oocyst that contains two sporocysts, each possessing a Stieda body and containing two sporozoites.

A single apicomplexan mitochondrial (mt) genome copy is approximately 6 – 7 kb in length (Gray et al., 2004). A variety of mt genome forms have been described among these parasites including circular-mapping mitochondrial genomes (e.g. circular in haemosporinids (Wilson and Williamson, 1997; Feagin et al., 2012) or linear concatemers (multiple mt genome copies joined end to end) in coccidia (Hikosaka et al., 2011)) as well as linear genomes with terminal inverted telomeric repeats in piroplasms (Hikosaka et al., 2010, 2012). Regardless of their structural forms, apicomplexan mt genomes usually possess three protein coding genes encoding cytochrome *c* oxidase subunit I (Cox1 cytochrome *c* oxidase subunit III (Cox3) and cytochrome b (CytB) as well as fragmented ssrDNA and lsrDNA. The present study reports on the sequence and structure of the mt genome of *Cyclospora cayetanensis*.

Stool samples that were positive for the presence of *C. cayetanensis* by UV fluorescence microscopy were selected for the molecular studies. The samples were collected and used in accordance with the CDC Institutional Review Board (IRB) protocol entitled "Use of Human Specimens for Laboratory Methods Research". Three samples were used for the whole mt genome sequencing: two samples from different time points (2011 and 2013) from an endemic area of southeastern Asia, plus a sample collected during the 2013 outbreaks in the USA. Partial Cox1 and Cox3 genes were sequenced from five additional samples: two samples from the same southeastern Asian location (2012 and 2013); and, three samples from two different outbreaks in the USA during 2013. DNA was extracted using the Universal Nucleic Acid Extraction (UNEX) method as described by Shields et al. (2013) with some adjustments. Approximately 0.5 ml of stool was added to a matrix E bead beating tube (MP Biomedicals, Santa Ana, CA, USA) together with 60 μl of proteinase K (QIAGEN, Valencia, CA, USA) and 600 μl of UNEX buffer (Phthisis Diagnostics, Charlottesville, VA, USA). The tube was incubated at 56 °C for 15 min to allow for proteinase K activity. The mixture was homogenized in a FastPrep-24 tissue and cell disruptor instrument (MP Biomedicals) at a speed of 6.0 m/s for 1 min. The sample was then centrifuged at maximum speed (>13,000 *g*) for 1 min to pellet the debris. The supernatant was collected and passed through a DNeasy mini spin binding silica column (QIAGEN). Following two wash cycles using ethanol-containing wash buffers, the DNA was eluted from the column in 80 μl of AE buffer (QIAGEN). The eluted filtrate was further purified by passing through a Zymo-Spin IV-HRC column (Zymo Research Corp., Irvine, CA, USA). DNA samples were confirmed positive for *C. cayetanensis* by real-time PCR as described by Verweij et al. (2003).

Initial attempts to amplify near-complete mt genomes using methods that had worked reliably with various *Eimeria* spp. (see Ogedengbe et al., 2013, 2014) failed repeatedly with *C. cayetanensis*. Thereafter, shorter regions of the mt genome were amplified with primers

targeting conserved regions of other apicomplexan mitochondrial genomes. The resulting PCR products were purified, sequenced directly using internal sequencing primers as necessary, and readily assembled into a partial mt genome using the de novo assembler within Geneious (www.geneious.com). Repeated attempts were made to complete the mt genome by amplifying across the 'gap' (assuming a circular or linear concatenated genome) with three pairs of additional amplification primers (Table 1). These efforts failed repeatedly with *C. cayetanensis* but succeeded in the case of coccidia in the genera *Eimeria, Caryospora* and *Isospora* (e.g. Lin et al., 2011; Ogedengbe et al., 2013, 2014; 2015; Ogedengbe and Barta, 2015), all of which possess circular-mapping mt genomes; *Eimeria tenella* was used as positive control for all such PCRs due to its linear concatenated genome.

Under the assumption, based on PCR results, that the mt genome of *C. cayetanensis* was linear, bulk cellular DNA was tailed using a terminal deoxynucleotidyl transferase (TdT) tailing method previously described by Hikosaka et al. (2012) with minor modifications. Briefly, sample DNA (75 ng) in 5 μl of nuclease-free water was denatured at 94°C for 5 min and then immediately used in a 25 μl of 3'-tailing reaction consisting of 0.0125 μmol of dCTP (Clonetech Laboratories, Inc. Mountain View, CA USA), 7 U of TdT enzyme (Clonetech) and 0.02% BSA (w/v) in 1× TdT buffer (Clonetech) for 30 min at 37°C. At the conclusion of the tailing reaction, the TdT enzyme was heat inactivated at 65°C for 10 min (see Hikosaka et al., 2012). The poly-C-tailed genomic DNA was then used as template for a pair of nested PCRs. In the first PCR, 2 μl of the end-labeled genomic DNA (i.e. 6 ng) was used as template in a reaction mixture containing 1.25 units of Platinum Taq Polymerase (Invitrogen, Carlsbad CA, USA), 1× PCR buffer (Invitrogen), 2.5 mM $MgCl_2$, 200 μM dNTPs and 0.4 μM each of the required primers. For the amplification of the 5'-end of the genome, an mt genome-specific primer 'q_Eim_CytB_398R' (5'-CCCCAGWARCTCATYTGACCCCA-3') was used with a poly-G-containing anchor primer 'Telo_F_polyG' (5'-GGCCACGCGTCGACTAGTACGGGGGGGGGGGGGGGGGG-3'). For the amplification of the 3'-end of the genome, an mt genome-specific WG-MT_5416F (5'-GGTCCAGATAAGCGATCTCATG-3') was paired with the same poly-G-containing anchor primer ('Telo_F_polyG'). Cycling conditions were initial denaturation at 95°C for 2 min followed by 40 cycles of 94°C for 30 s, 60°C for 30 s, 72°C for 45 s and final extension step at 72°C for 7 min. Following the primary PCR, 1 μl from each of the first PCRs was used as template for amplification reactions of 5'- and 3'-ends. The 5'-end was amplified using amplification primer Telo_F (5'-GGCCACGCGTCGACTAGTAC-3') with mt genome-specific WG-MT_63R (5'-CTGGTATGGATGGATAACACT-3') under the following cycle conditions: initial denaturation at 95°C for 2 min, 40 cycles of 94°C for 30 s, 56°C for 30 s, 72°C for 20 s, and a final extension step at 72°C for 7 min. Similarly, the 3'-end was amplified with the primer Telo_F combined with mt genome-specific WG-MT_5813R (5'-AGGTGCTCAGGGTCTTACCG-3') under slightly different cycle conditions: initial denaturation at 95°C for 2 min, 40 cycles of 94°C for 30 s, 63°C for 30 s, 72°C for 40 s and final extension step at 72°C for 7 min. After electrophoresis, amplification products were excised, purified and submitted for sequencing using the appropriate *Cyclospora*-specific primer used in the final amplification. The sequence data resulting from the mt genome tailing and subsequent nested PCR products was mapped onto the assembled

partial *C. cayetanensis* mt genome obtained through direct PCR amplification. To allow alignment with related *Eimeria* sp. mt genome sequences, complete mt genome sequences from GenBank (conventionally notated as circular) were linearized such that position 1 of each sequence was at the region homologous to the 5'-terminus identified for *C. cayetanensis* (i.e. 5'-TATTTnnAA…). Start and stop codon positions for each of three long open reading frames were identified using the Geneious Version 6.1 (and later versions) software package. Annotation of the protein-coding regions within the near-complete mt genome sequence of *C. cayetanensis* followed that of the mt genome sequence from *Eimeria mitis* USDA 50 (Ogedengbe et al., 2013). Fragmented lsrDNA and ssrDNA were annotated following Feagin et al. (2012) as modified by Ogedengbe et al. (2015).

There were no observed sequence variations in the mt genomes of the *C. cayetanensis* isolates obtained from human feces examined in the present study regardless of the origin of the sample. Amplification of the most divergent regions of eimeriid mt genomes (the Cox1 and Cox3 loci) illustrated that the five tested human isolates of *C. cayetanensis* (i.e. two southeastern Asian samples from 2012 and 2013; three USA samples from 2012 and 2013) were identical over all of the 860 bp of sequence obtained from the Cox3 locus and all of the 761 bp of sequence obtained from the Cox1 locus. Other regions of the mt genome for which sequence data from multiple isolates were available showed no variation among any of the human isolates examined in this study (data not shown). The mt genome of *C. cayetanensis* is an apparently linear molecule approximately 6220 bp in length (see GenBank **KP658101** for a near-complete sequence). Five overlapping fragments obtained with five pairs of amplification primers (Table 1) were obtained (1749 bp, 809 bp, 1978 bp, 959 bp and 1209 bp). The full sequence length for the *C. cayetanensis* mt genome, including the completed 5'-end (but excluding 35 - 50 bp of the 3'-end estimated to be missing by comparison with the mt genome sequences of various *Eimeria* spp.) was 6184 bp (including the 3' primer); thus the estimated full length of the mt genome would be approximately 6220 bp. Base composition of the mt genome was A (30.4%), T (36%), C (16.7%), G (16.9%) and thus the GC content was 33.6%. There was an A/T-rich terminus to the 5'-end of the genome that appeared capable of forming a closed hairpin loop structure. The *C. cayetanensis* mt genome encodes three protein-coding genes (CytB 1080 bp (237 bp - 1316 bp), ATG start codon; Cox1 1449 bp (1351 bp - 2799 bp), TTA start codon and Cox3 753 bp (4362 bp - 5114 bp), ATG start codon). These coding regions in the mt genome of *C. cayetanensis* had high mean pairwise sequence identities (90.4% in CytB, 90.3% in Cox1 and 87.1% in Cox3) to the corresponding coding DNA sequence (CDS) found in 17 mt genomes of various *Eimeria* spp. Numerous regions were identified as fragmented rDNA that appeared to encode 19 lsrDNA and 14 ssrDNA fragments. The rDNA fragmentation was irregular, with fragments ranging in length from 17 to 192 bp with the majority (20) interspersed between Cox1 and Cox3 genes; the remaining 13 were located either before the CytB CDS (three fragments) or following the Cox3 CDS (10 fragments). The putative lsrDNA and ssrDNA fragments were annotated on the mt genome following the convention of Feagin et al. (2012) as extended by Ogedengbe et al. (2015). Overall, pairwise comparisons of the complete *C. cayetanensis* mt genome to all available complete coccidian mt genomes demonstrated that the mt genome of *C. cayetanensis* was most similar (92.5%

pairwise identity) to *Eimeria dispersa* which has been demonstrated to infect turkeys (*Meleagris gallopavo)* and Northern Bobwhite quail (*Colinus virginianus*).

The *C. cayetanensis* mt genome appears to exist as a linear monomeric genome. Terminal transferase poly-C tailing followed by nested PCR and sequencing demonstrated a non-repetitive, 21 bp long region consisting of only A and T bases at the 5'-end of the *C. cayetanensis* mt genome that has the potential to form a tight hairpin loop (i.e. 5'-TATTTTTAATATAAAAATTTA; complementary regions underlined). Alignment of this region with the homologous regions from the mt genomes of coccidia belonging to several genera (Fig. 1) illustrate that all of these parasites have A/T-rich regions with many having a similar ability to form hairpin structures with a single terminal-free base. The sequence and structure of the 3'-terminus of the mt genome of *C. cayetanensis* could not be determined using the same methods used to obtain the 5'-end of the genome despite repeated attempts. When the sequence data at the presumed 3' terminal end of other coccidia were examined, we observed a conserved, imperfect palindromic sequence motif 'CTGTTATTTTGTC' at the 3'-end of all of these linearized mt genomes (Fig. 1).

The mitochondrial genome of *C. cayetanensis* was highly conserved with respect to genome length, CDS content, transcriptional directions and lsrDNA and ssrDNA fragments compared with the mt genomes of *Eimeria* spp. infecting various hosts (e.g. Lin et al., 2011; Ogedengbe et al., 2013, 2014), an *Isospora* sp. infecting canaries (see Ogedengbe et al., 2015) or *Caryospora bigenetica* (see Ogedengbe and Barta, 2015) that infects viperid snakes and various mammals. The mt rDNA fragments of *C. cayetanensis* had pairwise sequence identities of 68.2% to 97.6% compared with homologous rDNA fragments of *P. falciparum* (M76611) annotated functionally by Feagin et al. (2012). Fragmented rRNA genes are a common feature of apicomplexan mt genomes (Hikosaka et al., 2010; Feagin et al., 2012); the reason for the rDNA fragmentation is unknown. Similar to all apicomplexan mt genomes sequenced thus far, transfer rRNA are absent in the *C. cayetanensis* mt genome. The start codon usage for Cox3 and CytB genes was conventional and corresponded to the start codons in *Eimeria* spp. (Ogedengbe et al., 2013, 2014). The start codon ATG for the Cox1 gene was identified at a position (1327 bp) in other *Eimeria* spp which does not exist in *C. cayetanensis*. Instead there is a TTA start codon two codons upstream of the putative start codon site found in all mt genomes of *Eimeria*, *Caryospora* and *Isospora* spp. examined to date (e.g. Hikosaka et al., 2011; Lin et al., 2011; Ogedengbe et al., 2013, 2014, 2015; Ogedengbe and Barta, 2015). There was a lack of sequence variation observed in a number of isolates of *C. cayetanensis* at the mt Cox1 and Cox3 loci even when obtained from geographically distant locations. Such lack of intraspecific variation at these loci has been reported for distinct geographic isolates of some *Eimeria* spp. (e.g. *E. tenella* and *Eimeria gallopavonis* have no intraspecific variation over the entire mt genome length) whereas *E. mitis* demonstrates limited genetic variation at these genetic loci. Based on alignments of complete, annotated mt genome sequences from many coccidia, any intraspecific variation among *C. cayetanensis* isolates is likely to be found in the intergenic regions rather than within the coding regions (for both proteins and rRNA fragments).

Initial attempts to amplify a near complete mt genome sequence of *C. cayetanensis* using outward-facing primers within the Cox1 CDS failed repeatedly; this approach had been used

successfully with numerous *Eimeria* spp. and other eimeriid coccidia. Consequently, starting with known partial Cox1 and Cox3 sequences, primers in highly conserved regions of the mt genomes of various eimeriid coccidia were designed until a near-complete genome was obtained. Ultimately, five overlapping PCR fragments were generated in this study to obtain a near-complete mt genome for *C. cayetanensis*. The physical structure of the *C. cayetanensis* mt genome apparently followed neither circular nor linear concatemer genome topology (i.e. circular-mapping); these latter forms usually permit the contiguous sequencing to obtain a single mt genome copy either by running around the circular genome or from one copy to the next in linear concatenated forms. Instead, the mt genome of *C. cayetanensis* appears to be an 'open circle' (i.e. linear) and this linear form made the amplification across the open region impossible. Different PCR attempts with reverse primers with confirmed binding sites near the 5'-end of the genome paired with forward primers with confirmed binding sites near the 3'-end of the genome all failed. To test whether a dimer consisting of a concatenated forward and reverse copy of the genome existed, PCRs using single forward primers near the 3-end of the known genome were attempted; none produced any amplification products (data not shown). Although we conclude that the mt genome of *C. cayetanensis* is a linear monomer, the number of copies of the mt genome in each parasite was not determined.

Unlike the mt genomes of other coccidia examined thus far, TdT tailing, nested PCR and sequencing were required to complete the 5'-end of the *C. cayetanensis* mt genome; the first 21 bp of the mt genome consists of only A and T bases that had the potential to form a hairpin loop with a single free terminal T. Alignment of the *C. cayetanensis* mt genome with a number of other coccidian mt genomes demonstrated homologous regions in all mt genomes examined, including an invariant "TATTT" region shared among all mt genomes that matched the 5'-terminus of *C. cayetanensis* (see Fig. 1). In each mt genome, a sequence permitting a similar hairpin structure was observed with sequence heterogeneity within the fold region of the loop. Our conclusion is that this hairpin region identified at the 5'-end of the *C. cayetanensis* mt genome is the likely start of all eimeriid mt genomes, even when arranged in concatemers. In future, all circular-mapping mt genomes from eimeriid coccidia should use this well-conserved "TATTT" region as the 5'-end of each genome to establish a standardized 'reading frame' for these eimeriid mt genomes. Interestingly, a 45 bp stretch of the mt genome of *Plasmodium falciparum* (bases 51 - 95 of GenBank **M76611**) possesses a region capable of forming a hairpin followed by a region of high sequence identity with the Cocci_WG-MT_F primer (RNA9/SSU8), suggesting that such a convention might be able to be applied more broadly within the Apicomplexa. Kosa et al. (2006) suggested that A/T-rich stretches at the 5'- and 3'-ends of the linear mt genomes of a number of *Candida* spp. may mediate homologous recombination events that can generate circular or concatenated mt genomes from ancestral linear mt genomes. Such an A/T rich sequence was found at the 5'-terminus of the *C. cayetanensis* mt genome but the 3'-end remains to be determined; repeated attempts to tail this end of the genome using TdT proved fruitless. Physical mapping of the genome may be useful to confirm the structure of this genome in future.

Molecular analyses of nuclear ssrDNA sequences suggest that *C. cayetanensis* is closely related to other coccidia, especially members of the genus *Eimeria* (e.g. Relman et al., 1996;

Shields and Olson, 2003) in the family Eimeriidae. Some authors have even suggested *C. cayetanensis* be reclassified as an *Eimeria* sp. (Pieniazek and Herwaldt, 1997), although this would necessitate a broadening of genus definition to include a range of oocyst morphologies. Although numerous *Cyclospora* spp. have been described, sequence data exist for only four (i.e. *C. cayetanensis*, *Cyclospora papionis, Cyclospora colobi* and *Cyclospora cercopitheci*) and these sequences are restricted to the nuclear ssrDNA and internal transcribed spacer (ITS) loci (Eberhard et al., 1999; Zhou et al., 2011). The potential presence of paralogous gene copies in the nuclear rDNA and ITS loci as well as high sequence divergence at the ITS locus (loci) makes these genetic targets less than ideal for delimiting closely related coccidian parasites (Adams et al., 2000; El-Sherry et al., 2013). In contrast, mt CDS sequences (e.g. partial or complete Cox1 sequences) have been used successfully for delimiting closely related coccidia (Ogedengbe et al., 2011, 2014; El-Sherry et al., 2013). Nuclear ssrDNA sequences place the protozoan parasite *C. cayetanensis* in the apicomplexan family Eimeriidae among various *Eimeria* and *Isospora* spp. The overall similarity of the mt genome content and sequence of *C. cayetanensis* supports the view that *Cyclospora* spp. are closely related to other eimeriid coccidia such as *Eimeria* spp.

Acknowledgements

References

Adam RD, Ortega YR, Gilman RH, Sterling CR. Intervening transcribed spacer region 1 variability in *Cyclospora cayetanensis*. J. Clin. Microbiol. 2000; 38:2339–2343. [PubMed: 10834999]

Centers for Disease Control Reports. Outbreaks of diarrheal illness associated with cyanobacteria (blue-green algae)-like bodies-Chicago and Nepal, 1989 and 1990. Morb. Mortal. Wkly Rep. 1991; 40:325–327.

Eberhard ML, da Silva AJ, Lilley BG, Pieniazek NJ. Morphologic and molecular characterization of new *Cyclospora* species from Ethiopian monkeys: *C. cercopitheci* sp.n., *C. colobi* sp.n., and *C. papionis* sp.n. Emerg. Infect. Dis. 1999; 5:651–658. [PubMed: 10511521]

El-Sherry S, Ogedengbe ME, Hafeez MA, Barta JR. Divergent nuclear 18S rDNA paralogs in a turkey coccidium, *Eimeria meleagrimitis*, complicate molecular systematics and identification. Int. J. Parasitol. 2013; 43:679–685. [PubMed: 23639264]

Feagin JE, Harrell MI, Lee JC, Coe KJ, Sands BH, Cannone JJ, Tami G, Schnare MN, Gutell RR. The fragmented mitochondrial ribosomal RNAs of *Plasmodium falciparum*. PLoS One. 2012; 7:e38320. doi:10.1371/journal. pone.0038320. [PubMed: 22761677]

Gray MW, Lang BF, Burger G. Mitochondria of protists. Annu. Rev. Genet. 2004; 38:477–524. [PubMed: 15568984]

Hikosaka K, Nakai Y, Watanabe Y, Tachibana S, Arisue N, Palacpac NM, Toyama T, Honma H, Horii T, Kita K, Tanabe K. Concatenated mitochondrial DNA of the coccidian parasite *Eimeria tenella*. Mitochondrion. 2011; 11:273–278. [PubMed: 21047565]

Hikosaka K, Tsuji N, Watanabe Y, Kishine H, Horii T, Igarashi I, Kita K, Tanabe K. Novel type of linear mitochondrial genomes with dual flip-flop inversion system in apicomplexan parasites, *Babesia microti* and *Babesia rodhaini*. BMC Genomics. 2012; 13:622. doi: 10.1186/1471-2164-13-622. [PubMed: 23151128]

Hikosaka K, Watanabe Y, Tsuji N, Kita K, Kishine H, Arisue N, Nirianne MQ, Palacpac SK, Sawai H, Horii T, Igarashi I, Tanabe K. Divergence of the mitochondrial genome structure in the apicomplexan parasites, *Babesia* and *Theileria*. Mol. Biol. Evol. 2010; 27:1107–1116. [PubMed: 20034997]

Karanja RM, Gatei W, Wamae N. Cyclosporiasis: An emerging public health concern around the world and in Africa. Afr. Health Sci. 2007; 7:62–67. [PubMed: 17594281]

Kosa P, Valach M, Tomaska L, Wolfe KH, Nosek J. Complete DNA sequences of the mitochondrial genomes of the pathogenic yeasts *Candida orthopsilosis* and *Candida metapsilosis*: Insight into the evolution of linear DNA genomes from mitochondrial telomere mutants. Nucleic Acids Res. 2006; 34:2472–2481. [PubMed: 16684995]

Lin RQ, Qiu LL, Liu GH, Wu XY, Weng YB, Xie WQ, Hou J, Pan H, Yuan ZG, Zou FC, Hu M, Zhu XQ. Characterization of the complete mitochondrial genomes of five *Eimeria* species from domestic chickens. Gene. 2011; 408:23–33.

Ogedengbe JD, Hanner RH, Barta JR. DNA barcoding identifies *Eimeria* species and contributes to the phylogenetics of coccidian parasites (Eimeriorina, Apicomplexa, Alveolata). Int. J. Parasitol. 2011; 41:843–850. [PubMed: 21515277]

Ogedengbe ME, Barta JR. The complete mitochondrial genome of *Caryospora bigenetica* (Eimeriidae, Eucoccidiorida, Coccidiasina, Apicomplexa). Mitochondr. DNA: Early Online. 2015:1–2. doi: 10.3109/19401736.2015.1015007.

Ogedengbe ME, Brash M, Barta JR. 2015. The complete mitochondrial genome sequence of an *Isospora* sp. (Eimeriidae, Eucoccidiorida, Coccidiasina, Apicomplexa) causing systemic coccidiosis in domestic Canaries (*Serinus canaria* Linn.). Mitochondr. DNA. doi: 10.3109/19401736.2015.1018201.

Ogedengbe ME, El-Sherry S, Whale J, Barta JR. Complete mitochondrial genome sequences from five *Eimeria* species (Apicomplexa; Coccidia; Eimeriidae) infecting domestic turkeys. Parasit. Vectors. 2014; 7:335. doi:10.1186/1756-3305-7-335. [PubMed: 25034633]

Ogedengbe ME, Hafeez AM, Barta JR. Sequencing the complete mitochondrial genome of *Eimeria mitis* strain USDA50 (Apicomplexa: Eimeriidae) suggests conserved start positions for mtCOI-and mtCOIII-coding regions. Parasitol. Res. 2013; 112:4129–4136. [PubMed: 24013344]

Ortega YR, Sterling CR, Gilman RH, Cama VA, Diaz F. *Cyclospora* species – a new protozoan pathogen of humans. N. Engl. J. Med. 1993; 328:1308–1312. [PubMed: 8469253]

Pieniazek NJ, Herwaldt BL. Re-evaluating the molecular taxonomy: Is the human associated *Cyclospora* a mammalian *Eimeria* species? Emerg. Infect. Dis. 1997; 3:381–383. [PubMed: 9284387]

Relman DA, Schmidt TM, Gajadhar A, Sogin M, Cross J, Yoder K, Sethabutr O, Echeverria P. Molecular phylogenetic analysis of *Cyclospora*, the human intestinal pathogen, suggests that it is closely related to *Eimeria* species. J. Infect. Dis. 1996; 173:440–445. [PubMed: 8568307]

Shields JM, Olson BH. *Cyclospora cayetanensis*: A review of an emerging parasitic coccidian. Int. J. Parasitol. 2003; 33:371–391. [PubMed: 12705931]

Shields JM, Joo J, Kim R, Murphy HR. Assessment of three commercial DNA extraction kits and a laboratory-developed method for detecting *Cryptosporidium* and *Cyclospora* in raspberry wash, basil wash and pesto. J. Microbiol. Methods. 2013; 92:51–58. [PubMed: 23147278]

Verweij JJ, Laeijendecker D, Brienen EA, van Lieshout L, Polderman AM. Detection of *Cyclospora cayetanensis* in travellers returning from the tropics and subtropics using microscopy and real-time PCR. Int. J. Med. Microbiol. 2003; 293:199–202. [PubMed: 12868656]

Wilson RJ, Williamson DH. Extrachromosomal DNA in the Apicomplexa. Microbiol. Mol. Biol. Rev. 1997; 61:1–16. [PubMed: 9106361]

Zhou Y, Lv B, Wang Q, Wang R, Jian F, Zhang L, Ning C, Fu K, Wang Y, Qi M, Yao H, Zhao J, Zhang X, Sun Y, Shi K, Arrowood MJ, Xiao L. Prevalence and molecular characterization of *Cyclospora cayetanensis*, Henan, China. Emerg. Infect. Dis. 2011; 17:1887–1890. [PubMed: 22000362]
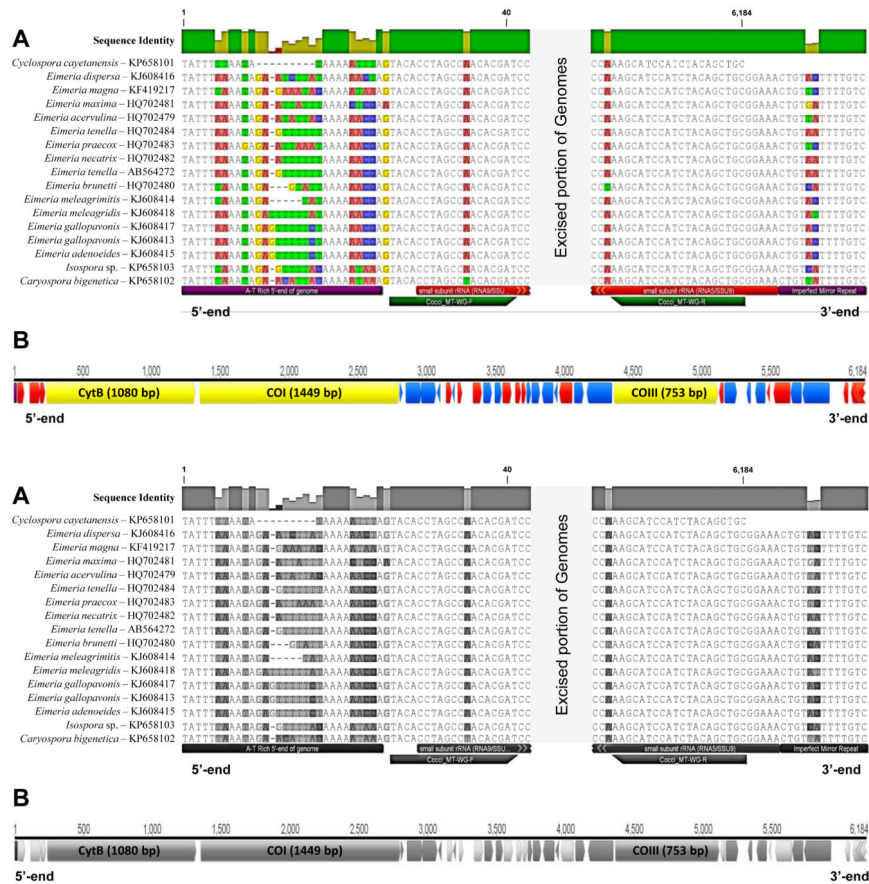
Fig. 1.

The mitochondrial (mt) genome of *Cyclospora cayetanensis* compared with the mt genomes of closely related eimeriid coccidia. (A) Aligned 5'- and 3'-termini of the *Cyclospora cayetanensis* mt genome aligned with the homologous regions in the mt genomes of other eimeriid coccidia; all comparative genomes have been linearized so that the first nucleotide of each mt genome matches the first base ('T') in the linear *C. cayetanensis* mt genome. A highly conserved region at the 5'-end with a consensus sequence of "5'-TATTTWWAADA…" was found in all genomes included in the alignment; all eimeriid genomes illustrated have the potential to form hairpins at the 5'-"end" of their circularly-mapped genomes. At the 3'-end of the eimeriid mt genome units in all coccidia other than *C. cayetanensis* (data not available) there was an imperfect mirror repeat sequence (i.e. … CTGTNHTTTTGTC -5', conserved mirrored bases underlined in the preceding consensus sequence). The role, if any, of this terminal mirror repeat motif in the replication of eimeriid mt genomes is not known. The annotations listed below the *Caryospora bigenetica* mt genome sequence (terminal features, Cocci_MT-WG-F and R primers, and RNA9 and RNA5 ssrDNA fragments) apply to all sequences in the alignment. (B) Map of the *C. cayetanensis* mt genome illustrating the locations, directions and lengths of the three protein-coding genes (cytochrome b (CytB), cytochrome *c* oxidase subunit I (Cox1) and cytochrome *c* oxidase subunit III (Cox3)). The A-T rich 5' end of the genome found in A is located at the start of the genome map (bp 1); the remaining smaller annotations illustrate locations and directions of the fragmented rRNA coding regions (lsrDNA fragments – dark

annotated regions; ssrDNA fragments – light annotated regions) interspersed amongst the three protein-coding genes.

Table 1

Amplification primers used to obtain the mitochondrial genome of C*yclospora cayetanensis* and confirm its linear nature.

FRAGMENT	PRIMER ID	NUCLEOTIDE SEQUENCE (5'-3')	SIZE in bp (position within mt genome[a])	REFERENCE
Fragment 1	Cocci_MT-WG-F Cyclo_COI_473R	TACACCTAGCCAACACGAT ATACCCGCAAGAGCTAAACC	1807 (23 - 1829)	Ogedengbe et al., 2014b Ogedengbe et al., 2013
Fragment 2	qPCR400-F COI_1202R	GDTCAGGTRTTGGTTGGAC CCAAKRAYHGCACCAAGAGATA	803 (1723 - 2525)	Ogedengbe et al., 2013 Ogedengbe et al., 2013
Fragment 3	Cyclo_COI-1085F Cyclo_CO3_113R	CTCCGCTCTAGATGTTGCTT TCACCATTCTTGCTCACTGT	2030 (2442 - 4471)	This Study This Study
Fragment 4	WG-MT_4140F Eim_CO3_799R	AGAAAACCTAAAATCATCATGT AAGTGAGTTCGCATGTTTAC	961 (4191-5151)	Ogedengbe et al., 2014b This Study
Fragment 5	Cyclo_CO3_219F Cocci_MT-WG-R	AGCTTCTTCTGGGGTGCATAC GCAGCTGTAGATGGATGCTT	1607 (4578 - 6184)	This Study Ogedengbe et al., 2014b
	WG_MT_344R Cocci_MT_WG_F (Positive Control)	GTAGGAATCTRAATTCCCAACC TACACCTAGCCAACACGAT	431 (23-453)	Ogedengbe et al., 2013 Ogedengbe et al., 2014b
PCR confirmation of linear genome	WG_MT_344R WG_MT_6219F	GTAGGAATCTRAATTCCCAACC GCATCCATCTACAGCTGCGG	No Product	Ogedengbe et al., 2013 Ogedengbe et al., 2013
	WG_MT_344R WG_MT_5416F	GTAGGAATCTRAATTCCCAACC GGTCCAGATAAGCGATCTCATG	No Product	Ogedengbe et al., 2013 Ogedengbe et al., 2013

[a]Positions reported are relative to the complete mitochondrial (mt) genome of *C. cayetanensis* (GenBank **KP658101**); complete 5'-end was generated using terminal deoxynucleotidyl transferase tailing followed by nested PCR and sequencing (see Section 2.3) but 3'-terminus has not been similarly completed.