Published in final edited form as:

Infect Genet Evol. 2015 June ; 32: 113–123. doi:10.1016/j.meegid.2015.03.003.

Highly divergent 18S rRNA gene paralogs in a *Cryptosporidium* genotype from eastern chipmunks (*Tamias striatus*)¹

Brianna L.S. Stenger^{a,b,c}, Mark E. Clark^{b,c}, Martin Kváček^{d,e}, Eakalak Khan^{c,f}, Catherine W. Giddings^a, Neil W. Dyer^g, Jessie L. Schultz^g, and John M. McEvoy^{a,c,*}

^aDepartment of Veterinary and Microbiological Sciences, North Dakota State University, Fargo, ND, USA

^bDepartment of Biological Sciences, North Dakota State University, Fargo, ND, USA

^cEnvironmental and Conservation Sciences Program, North Dakota State University, Fargo, ND, USA

^dInstitute of Parasitology, Biology Centre of the Academy of Sciences of the Czech Republic, eské Budějovice, Czech Republic

^eFaculty of Agriculture, University of South Bohemia in eské Budějovice, Czech Republic

^fDepartment of Civil and Environmental Engineering, North Dakota State University, Fargo, ND, USA

^gVeterinary Diagnostic Laboratory, North Dakota State University, Fargo, ND, USA

Abstract

Cryptosporidium is an apicomplexan parasite that causes the disease cryptosporidiosis in humans, livestock, and other vertebrates. Much of the knowledge on *Cryptosporidium* diversity is derived from 18S rRNA gene (18S rDNA) phylogenies. Eukaryote genomes generally have multiple 18S rDNA copies that evolve in concert, which is necessary for the accurate inference of phylogenetic relationships. However, 18S rDNA copies in some genomes evolve by a birth-and-death process that can result in sequence divergence among copies. Most notably, divergent 18S rDNA paralogs in the apicomplexan *Plasmodium* share only 89–95% sequence similarity, encode structurally distinct rRNA molecules, and are expressed at different life cycle stages. In the present study, *Cryptosporidium* 18S rDNA was amplified from 28/72 (38.9%) eastern chipmunks (*Tamias striatus*). Phylogenetic analyses showed the co-occurrence of two 18S rDNA types, Type A and Type B, in 26 chipmunks, and Type B clustered with a sequence previously identified as *Cryptosporidium* chipmunk genotype II. Types A and B had a sister group relationship but shared less than 93% sequence similarity. In contrast, actin and heat shock protein 70 gene sequences

¹Note: Supplementary data associated with this article

© 2015 Published by Elsevier B.V.

Corresponding author: John M. McEvoy, PO Box 6050, Dept. 7690, Fargo, ND, 58108-6050, USA, p: 701 231 8530; john.mcevoy@ndsu.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

were homogeneous in samples with both Types A and B present. It was therefore concluded that Types A and B are divergent 18S rDNA paralogs in *Cryptosporidium* chipmunk genotype II. Substitution patterns in Types A and B were consistent with functionally constrained evolution; however, Type B evolved more rapidly than Type A and had a higher G+C content (46.3% versus 41.0%). Oocysts of *Cryptosporidium* chipmunk genotype II measured $4.17\text{ }\mu\text{m}$ ($3.73\text{--}5.04\text{ }\mu\text{m}$) $\times$ $3.94\text{ }\mu\text{m}$ ($3.50\text{--}4.98\text{ }\mu\text{m}$) with a length-to-width ratio of 1.06 ± 0.06 , and infection occurred naturally in the jejunum, cecum, and colon of eastern chipmunks. The findings of this study have implications for the use of 18S rDNA sequences to infer phylogenetic relationships.

Keywords

Cryptosporidium; paralogy; 18S rRNA; 18S rDNA

1. Introduction

The apicomplexan *Cryptosporidium* parasitizes the major vertebrate groups (Kváč et al., 2014) and is a major cause of diarrheal disease in humans and livestock. Twenty-six species are currently recognized (Ryan et al., 2014), but these represent just a fraction of *Cryptosporidium* diversity. In addition the named species, more than 70 genotypes have been described from sequences of rRNA and protein coding genes. Genotypes are incomplete and temporary descriptors of *Cryptosporidium* that lack the morphological, biological, and genetic data to support full classification as species. However, despite their taxonomic uncertainty, genotype descriptions greatly enhance the understanding of *Cryptosporidium* diversity.

Phylogenetic relationships among eukaryotes are frequently inferred from sequences of the 18S rRNA gene (18S rDNA). Most genomes have multiple 18S rDNA copies arranged in tandem on a single chromosome, which evolve in concert because of frequent gene conversion and unequal crossover events. The sequence homogeneity among 18S rDNA copies that results from concerted evolution is necessary to accurately infer phylogenetic relationships. Some genomes have divergent 18S rDNA copies distributed across multiple chromosomes (Barthelemy et al., 2007; Gunderson et al., 1987; Krieger and Fuerst, 2002). Most notably, divergent 18S rDNA copies in the apicomplexan *Plasmodium* share only 89–95% sequence similarity (Qari et al., 1994). These paralogs encode structurally distinct rRNA molecules that are expressed at different life cycle stages (McCutchan et al., 1988; McCutchan et al., 1995). Rooney (2004) proposed that, in contrast to concerted evolution, these divergent 18S rDNA paralogs evolve by a birth-and-death process whereby new genes arise by duplication, evolve independently, and can acquire new functions, become non-functional, or be deleted. Although there have been reports of paralogous 18S rDNA copies in other apicomplexans, including *Cryptosporidium*, *Babesia*, *Theileria*, and *Eimeria*, the extent of their divergence has been considerably lower than that in *Plasmodium* (Bhoora et al., 2009; El-Sherry et al., 2013; Hill et al., 2008; Ikarashi et al., 2013; Le Blancq et al., 1997; Morgan et al., 2001; Santín and Fayer, 2007; Sevá Ada et al., 2011; Xiao et al., 1999).

We report the frequent co-occurrence of two highly divergent *Cryptosporidium* 18S rDNA types in eastern chipmunks (*Tamias striatus*) and present evidence that they are paralogs in a

single lineage that infects the intestinal epithelium. The identification of highly divergent 18S rDNA paralogs in *Cryptosporidium* has implications for the use of this gene to infer phylogenetic relationships.

2. Materials and Methods

2.1. Sample collection

Samples were collected from seventy-two eastern chipmunks (*Tamias striatus*) from five locations in Minnesota, USA, during a three-year period. Chipmunks were live-captured in Sherman box traps, ear-tagged, and released in compliance with North Dakota State University Institutional Animal Care and Use policies.

Traps were baited with sunflower seeds or rolled oats mixed with peanut butter. Voided feces were collected from the trap of each captured chipmunk. Recaptured chipmunks, identified by an ear-tag, were counted only once in prevalence analysis. In addition to voided feces, samples were collected from the rectum of two chipmunks (Table 1; Tast-1 and Tast-5), and from the stomach, small intestine, large intestine, and cecum of one chipmunk (Table 1; Tast-14). Samples were stored at 4°C without fixatives prior to DNA isolation.

2.2. DNA isolation and PCR

DNA was isolated from samples by alkaline digestion and phenol-chloroform extraction, and purified using the QIAamp DNA Stool Mini Kit (Qiagen, Valencia, CA) as described previously (Peng et al., 2003). DNA was stored at -20°C until use in PCR assays.

A genus-specific nested PCR assay was used to amplify ca. 830 bp of the *Cryptosporidium* 18S rRNA gene (Xiao et al., 2001). Based on a finding in the study that most 18S rDNA sequences belonged to one of two clades, which were named Type A and Type B, type-specific, semi-nested PCR assays were developed. Type A 18S rDNA was amplified using the primers TastIITAF1 (5'-GTG ACA TAC CAT TCA AGT TTC TGA CC) and TastIITAR (5'-GGG AAG GCT AAC AAC ATG TAA AGG C) in a primary reaction, and TastIITAF2 (5'-TCC TAA TAC AGG GAG GTA GTG AC) and TastIITAR in a secondary reaction. Type B 18S rDNA was amplified using primers TastIITBF1 (5'-ACG GAT CAC ACA CTT TCT CTG TG) and TastIITBR (5'-CTA AAG CCG TGC AGG AGC AAG AAA) in a primary reaction, and TastIITBF2 (5'-CGC GCA AAT TAC CCA ATC CTG ACA) and TastIITBR in a secondary reaction. Primary and secondary reactions for each target were prepared in a total volume of 50 µL, which included 1 µL of DNA template or 2.5 µL of primary product (secondary reactions), 200 µM dNTP, 1.5 mM MgCl₂, 1× PCR buffer, 2.5U of Taq (Promega GoTaq Flexi), and 0.4 µM of each primer. Reaction conditions for the Type A primary and secondary PCR were an initial denaturation at 94°C for 3 min; 35 cycles of 94°C for 45 s, 54°C for 45 s, and 72°C for 30 s; and a final elongation step of 72°C for 7 min. Type B primary and secondary PCR reaction conditions were as described for Type A, with the exception that the annealing temperatures were 55°C and 57°C in the primary and secondary reactions, respectively.

Nested PCR assays were used to amplify fragments of actin (ca. 1066 bp; Sulaiman et al., 2002) and heat shock protein 70 (HSP70; ca. 1800 bp; Sulaiman et al., 2000) genes.

DNA from *C. hominis* and *C. parvum* was included as positive controls in the genus-specific 18S rDNA, actin, and HSP70 PCR assays. Water was included instead of DNA template as a negative control in all PCR assays. Secondary PCR products were visualized using SYBR Green dye or ethidium bromide staining following agarose gel electrophoresis. Products of expected size were purified (Wizard SV, Promega, Madison, WI) and stored at 4°C until required for sequencing.

2.3. Sequencing and phylogenetic analyses

Purified amplicons of 18S rDNA, actin, and HSP70 were sequenced directly in both directions using a BigDye Terminator v3.1 cycle sequencing kit with secondary PCR primers in an ABI Prism 3130 genetic analyzer (Applied Biosystems, Carlsbad, CA). An additional internal primer (5' GCT GCT GCC ATT GCT TAT GGT CTT 3') was used to sequence HSP70 amplicons.

Sequences were assembled using SeqMan (DNASTar, Madison, WI), aligned with previously published sequences using the MAFFT version 7 online server with automatic selection of alignment mode (<http://mafft.cbrc.jp/alignment/server/>) (Kato and Standley, 2013). Sequences from this study have been deposited in GenBank under the accession numbers KC954211 to KC954269 and KJ816864 to KJ816871. The evolutionary history of aligned sequences was inferred using the Maximum Likelihood (ML) method (Saitou and Nei, 1987), with the substitution model that best fit the alignment selected using the Bayesian information criterion. The Tamura 3-parameter model (Tamura, 1992) was selected for 18S rDNA alignments, and the general time reversible model (Tavaré, 1986) was selected for actin and HSP70 alignments. Both models were used under an assumption that rate variation among sites was gamma distributed. A bootstrap consensus tree was inferred from 1000 pseudoreplicates. Phylogenetic analyses, including analysis of substitution model goodness of fit, were carried out using MEGA 6.0 (Tamura et al., 2013). Phylogenetic trees were edited for style using Adobe Illustrator CS5.1 (Adobe Systems, Inc., San Jose, CA).

MAFFT aligned 18S rDNA sequences were examined for recombination using the phylogeny-based Probabilistic Divergence Measures (PDM) method in TOPALi v2.5 (Husmeier et al., 2005; Milne et al., 2009). A Markov chain Monte Carlo simulation was used to examine differences in the marginal posterior distribution of tree topologies in a fixed size window of 200 moving in a step size of 10 along the length of aligned sequences. Bootstrap support was determined from 100 pseudoreplicates. Recombination breakpoints were predicted from the PDM analysis and ML trees were constructed from partitioned sequences using MEGA 6.0.

A structure-based alignment and secondary structure diagrams of 18S rDNA sequences were produced using SSU-ALIGN and SSU-DRAW, respectively (Nawrocki, 2009). SSU-ALIGN uses a eukaryotic 18S rRNA structure model inferred from available RNA sequence and crystal structure information (<http://www.rna.cccb.utexas.edu>). Secondary structure was

not predicted for the V4 region. Expansions of helix 11 in the V2 region model (helix numbering based on Wuyts et al., 2002) were handled poorly by the eukaryote secondary structure, so the stem-loop structure predicted by RNAfold (Hofacker et al., 1994) was used instead.

Nucleotide composition analysis and Tajima's relative rate test (Tajima, 1993) were carried out using MEGA 6.0.

2.4. Oocyst purification and morphometric analysis

Oocysts were isolated from the feces of two chipmunks, Tast-24 and Tast-25. Larger particles were removed by sieving through a US #35 mesh screen (opening 0.5 mm) followed by a US #325 mesh screen (opening 0.044 mm). Oocysts were subsequently purified using cesium chloride gradient centrifugation (Arrowood and Donaldson, 1996), and examined under differential interference contrast (DIC) and epifluorescence microscopy (Olympus, Center Valley, PA), following labeling with a *Cryptosporidium* genus-specific, FITC-conjugated antibody (Crypt-a-Glo, Waterborne Inc., LA).

Cell morphology was determined using digital analysis of images (MicroSuite™ Five; Olympus Soft Imaging Solution GmbH, Lakewood, CO) collected at 1,000x magnification using a 12-megapixel digital microscope camera (DP10; Olympus). Length and width were measured for nineteen oocysts and a shape index was calculated. Twenty oocysts of *C. parvum* (Waterborne Inc., New Orleans, LA) measured by the same person under the same conditions served as a control.

2.5. Tissue sample collection and histology

Tast-25 was euthanized immediately after capture, and gastrointestinal tissue sections were excised from the stomach, duodenum, proximal jejunum, middle jejunum, distal jejunum, ileum, cecum, colon, and rectum. Tissue samples were stored in 10% formalin at room temperature before embedding in paraffin. Five-micron cross-sections were prepared and deparaffinized from each tissue section, and stained with hematoxylin and eosin.

3. Results

3.1. *Cryptosporidium* prevalence in chipmunks

Out of seventy-two chipmunks sampled, 28 (38.9%) tested positive for *Cryptosporidium* on at least one occasion during the study. In total, 35 positive samples were obtained from the 28 positive chipmunks: a single positive sample was obtained from 25 chipmunks; three positive samples were obtained from one chipmunk, Tast-10, on different days; five positive samples were obtained from the stomach, small intestine, large intestine, cecum, and rectum of one chipmunk, Tast-14; and two positive samples were obtained from one chipmunk, Tast-17, on different days (Table 1).

3.2. Two *Cryptosporidium* 18S rDNA types occurred frequently in chipmunks

Sequences of 18S rDNA obtained from 28 out of 35 positive samples were used to construct a ML tree (Figure 1). Amplicons from the remaining seven samples were refractory to sequencing (Supplementary Table S1).

Sequences from 24 out of 28 samples clustered in two sister groups, which were named Type A and Type B (Figure 1). Types A and B shared $92.7 \pm 1.0\%$ sequence identity and were both identified in five samples using the *Cryptosporidium* genus-specific PCR assay. Sequence identity within the Type A group was $99.8 \pm 0.1\%$ (range: 99.1 to 100%). Within the Type B group, which included *Cryptosporidium* chipmunk genotype II from an eastern chipmunk in New York State (accession no. EU096238), sequence identity was $99.2 \pm 0.2\%$ (range: 97.6 to 100%). The W21 genotype (accession no. AY737589), which was previously isolated from storm water in New York (Jiang et al., 2005), was intermediate between the types, but was more similar to Type B ($96.2 \pm 0.7\%$ identity) than Type A ($94.6 \pm 0.9\%$ identity).

Four samples from two chipmunks had sequences other than Types A and B (Figure 1). Sequences from ID-3153 and ID-3154, which were obtained from the stomach and small intestine, respectively, of Tast-14, shared 100% identity with a *C. andersoni* isolate (accession no. EU825734) from the Potomac River Watershed (Yang et al., 2008). A novel genotype was identified in H-3117 and H-3130, which were obtained from Tast-10 on consecutive days in June 2011 (Figure 1). This genotype, which was named *Cryptosporidium* chipmunk genotype IV, clustered with sequences from an environmental isolate in the UK (accession no. HQ822143), *Cryptosporidium* muskrat genotype II (accession no. AY737570), *Cryptosporidium* deer mouse genotype IV (accession no. EF641019), *C. ubiquitum* (accession no. EF362479), and *C. suis* (accession no. AB449871) (Figure 1).

3.3. Types A and B are divergent 18S rDNA paralogs in a single lineage

Based on their sister group relationship and co-occurrence in five out of 24 samples, we hypothesized that Types A and B are divergent 18S rDNA paralogs in *Cryptosporidium* chipmunk genotype II. It was not feasible to test for paralogy by examining haploid genomes from single cells or clonally propagated isolates, so the approach taken was to determine a) how frequently Types A and B co-occurred in positive samples and b) to what extent other loci were divergent.

If Types A and B are paralogs in *Cryptosporidium* chipmunk genotype II, they would be expected to co-occur. The genus-specific 18S rDNA PCR assay is not useful for detecting individual members in mixed populations, so type-specific primers were designed. Combining data from the genus- and type-specific PCR assays, at least one of the types was detected in all 35 samples, and co-occurrence of Types A and B was detected in 30 out of 35 samples from 26 out of 28 *Cryptosporidium* positive chipmunks (Table 1 and Supplementary Tables S1 and S2). Type B was not detected in the only samples obtained from Tast-11 and Tast-12. Each of three chipmunks (Tast-10, Tast-14, and Tast-17) sampled on multiple occasions had a single sample from which only one type was detected. One of

the three fecal samples (H-3117) obtained from Tast-10 on separate occasions was negative for Type A; the remaining samples, including a sample collected three days later than H-3117, had both types present. Only Type A was detected in a fecal sample from Tast-14 (ID-3169); however, both types were detected in samples obtained from the stomach (ID-3153), small intestine (ID-3154), large intestine (ID-3155), and cecum (ID-3156) of this animal three days later. One sample from Tast-17 (H-3177) was negative for Type B, but both types were detected in sample ID-3175 collected from Tast-17 two days later. Based on their frequent co-occurrence, we concluded that Types A and B are either paralogs in a single *Cryptosporidium* lineage or are from divergent taxa that co-infect chipmunks.

Actin and HSP70 genes in *Cryptosporidium* tend to be more divergent than 18S rDNA (Table 2). Therefore, if the *Cryptosporidium* 18S rDNA types in eastern chipmunks diverged because of speciation (orthology) rather than gene duplication (paralogy), it would be expected that actin and HSP70 genes should exhibit similar or greater divergence.

Thirty-four actin amplicons were sequenced from 16 samples: single amplicons were sequenced from eight samples and two to five amplicons were sequenced from eight samples. Sequences from H-3117 and H-3130 clustered with *Cryptosporidium* deer mouse genotype (accession no. AY120925), *C. ubiquitum* (accession no. GQ337961), and *C. suis* (accession no. EF012373) (Figure 2), which is consistent with the clustering of 18S rDNA sequences from *Cryptosporidium* chipmunk genotype IV. All other sequences clustered as a single group that was most similar to *C. canis* (accession no. AY120926) and a bear genotype (accession no. AF382339). Sequences within this group shared $99.9 \pm 0.1\%$ identity (Table 2).

Nineteen HSP70 sequences from nine samples clustered in two groups (Figure 3). Consistent with the clustering of 18S rDNA and actin sequences, the H-3130 HSP70 sequence clustered with *C. ubiquitum* (accession no. DQ898163), *C. suis* (accession no. DQ898164), and a deer mouse genotype (accession no. AY120919). All other sequences formed a separate group that was most similar to sequences from *C. canis* (accession no. AY120920), a bear genotype (accession no. AF247536), and a giant panda genotype (accession no. JN588571). Sequences in this group shared $99.9 \pm 0.1\%$ identity (Table 2).

With the exception of *Cryptosporidium* chipmunk genotype IV sequences obtained from H-3117 and H-3130, actin and HSP70 sequences from chipmunks formed a single homogeneous group, which is inconsistent with Types A and B being from divergent taxa. A more parsimonious explanation is that Types A and B are divergent paralogs in *Cryptosporidium* chipmunk genotype II.

3.4. The pattern of nucleotide substitutions suggests functionally constrained evolution of Types A and B

The pattern of nucleotide substitutions in Types A and B was determined to assess the likelihood that one or both types are non-functional pseudogenes. The evolution of functional 18S rDNA is constrained by the conserved secondary and tertiary structure of the 18S rRNA molecule, which is maintained by the coordinated evolution of base pairs (co-variation). In contrast, non-functional pseudogenes are expected to accumulate mutations

randomly. The amplified *Cryptosporidium* 18S rDNA from HR-2039 Type A and HR-2039 Type B included three hypervariable stem-loop regions: part of V2 and the complete V3 and V4 regions (Figure 4). Sequence identity between Type A and Type B in Figure 4 is indicated by white font against a black background in the pairwise and structure alignments. Comparing secondary structure alignments, the majority of substitutions occurred in the V2, V3, and V4 regions. In the V2 region, a six-nucleotide insertion and two substitutions in Type B expanded the stem and loop of helix 11. A third substitution occurred in a single-stranded region immediately downstream of helix 11. Helix 16 had two substitutions: one in the single-stranded loop and one in the stem. The substitution in the stem resulted in a non-canonical C-A pairing in Type A, which is conserved in intestinal *Cryptosporidium* species (data not shown). The canonical C-G pairing in Type B is conserved in gastric *Cryptosporidium* species (data not shown). Helix 18, in the V3 region, had three substitutions: one in the loop; one at a covarying site in the stem, resulting in an A-U pairing in Type A and a G-U pairing in Type B; and one at a non-covarying site in the stem. Helix 25 had two substitutions, both occurring in a single stranded bulge. This pattern of nucleotide substitutions is consistent with the evolution of Types A and B under structural constraints, suggesting they are not pseudogenes.

3.5. Type B has evolved more rapidly than Type A

Tajima's relative rate test (Tajima, 1993) was carried out to test the hypothesis that Types A and B have undergone different rates of evolution (Table 3). The null hypothesis of equal evolution rates was rejected with a high level of significance in tests conducted using three different outgroups. Type B sequences exhibited significantly greater rates of evolution than Type A.

3.6. Types A and B show evidence of recombination

Divergent 18S rDNA sequences were examined for evidence of recombination using the PDM method in TOPALi v2.5. A recombination breakpoint was predicted in the conservative core region between the hypervariable V3 and V4 regions (immediately after nucleotide 297 in Supplementary Figure S1). Phylogenies inferred from nucleotides 1–297 (Figure 5A) and 298–629 (Figure 5B) are discordant because of the chimeric sequences: FL-2033 B (accession no. KC954239) clustered with Type A in Figure 5A and Type B in Figure 5B. Similarly, ID-3134 A (accession no. KC954222), and GM-2691 A (accession no. KC954231) clustered with Type B in Figure 5A, sharing 100% identity with W21, and Type A in Figure 5B. The likelihood that the chimeric sequences were artifacts was considered low because a) amplicons were not cloned prior to sequencing, b) the same chimeric sequence was obtained from separate amplicons of ID-3134 A and GM-2691 A, and c) GM-2691 A was detected using both genus- and type-specific primers (Supplementary Figure S2).

3.7. Types A and B have differing G+C contents

Type B sequences of *Cryptosporidium* chipmunk genotype II had a higher overall G+C content (46.3%) than Type A sequences (41.0%) (Table 4). The difference between the types was greatest in the V3 (3.6% difference) and V4 (13.1% difference) regions. The G+C

content of the conservative 18S rDNA core, i.e. outside of the hypervariable regions, differed by 0.7%. The G+C content of chimeric sequences (GM-2691 A, ID-3134 A, and FL-2033 B) matched that of the opposite type in the V3 region (boxed cells in Table 4).

3.8. *Cryptosporidium* chipmunk genotype II infects the intestinal epithelium of chipmunks

Cryptosporidium developmental stages were observed in the gastrointestinal epithelium of Tast-25, which was infected with *Cryptosporidium* chipmunk genotype II Types A and B. Endogenous stages were identified in the jejunum, cecum, and colon (Figure 6).

3.9. Oocysts of *Cryptosporidium* chipmunk genotype II are smaller than *C. parvum*

Oocysts of *Cryptosporidium* chipmunk genotype II from eastern chipmunks measured $4.17 \mu\text{m}$ ($3.73\text{--}5.04 \mu\text{m}$) $\times$ $3.94 \mu\text{m}$ ($3.50\text{--}4.98 \mu\text{m}$) with a length-to-width ratio of 1.06 ± 0.06 . Oocysts were smaller than *C. parvum*, which measured $5.06 \mu\text{m}$ ($4.49\text{--}5.41 \mu\text{m}$) $\times$ $4.75 \mu\text{m}$ ($4.36\text{--}5.35 \mu\text{m}$) with a length-to-width ratio was 1.07 ± 0.06 . Oocysts labeled with the FITC-conjugated antibody and examined by epifluorescence microscopy had a typical apple green, halo-like fluorescence (Figure 7).

4. Discussion

We found that highly divergent *Cryptosporidium* 18S rDNA types with a sister group relationship co-occurred in 26/28 chipmunks. This frequent co-occurrence together with the homogeneity of actin and HSP70 sequences support the conclusion that the divergent types are paralogs in a single *Cryptosporidium* lineage. Because Type B clusters with the previously described *Cryptosporidium* chipmunk genotype II, it is proposed that the types be named *Cryptosporidium* chipmunk genotype II Type A and Type B.

The less than 93% sequence similarity shared by Types A and B is comparable to that shared by orthologous 18S rDNA sequences from intestinal and gastric *Cryptosporidium* species. Intestinal and gastric species branched early during *Cryptosporidium* evolution and, in addition to being considerably different at the genomic level, they have different oocyst morphologies, biochemical characteristics, and gastrointestinal niches (Widmer and Sullivan, 2012).

The divergence of 18S rDNA paralogs in chipmunk genotype II is considerably greater than that reported in other *Cryptosporidium* spp. including *C. parvum* (99% similarity; Le Blancq et al., 1997), *C. andersoni* (99% similarity; Ikarashi et al., 2013), and *C. galli* (98.6–99.1% similarity; Morgan et al., 2001). Types A and B also are more divergent than the 18S rDNA paralogs in *Babesia bigemina* (99% similarity; Reddy et al., 1991), *Eimeria mitis* (98% similarity; Vrba et al., 2011), *Eimeria meleagridis* (97% similarity; El-Sherry et al., 2013), and *Theileria equi* (96% similarity; Bhoora et al., 2009). *Plasmodium* species have the most divergent 18S rDNA paralogs in the phylum Apicomplexa, and the paralogs are differentially expressed during parasite development: the A-type is expressed during asexual development in the vertebrate host and the S-type is expressed in sporozoites developing in the mosquito host (Gunderson et al., 1987; McCutchan et al., 1995). The divergence of Types A and B in the present study is higher than that of A- and S-types in *P. berghei* (95% similarity; Qari et al., 1994), similar to that of A- and S-types in *P. vivax* (92% similarity;

Qari et al., 1994), and lower than that of A- and S-types in *P. falciparum* (89% similarity; Qari et al., 1994).

Given the extent to which sequences of Types A and B have diverged in chipmunk genotype II, it is reasonable to ask if they have evolved different functions, similar to the A- and S-types in *Plasmodium* species. Addressing this question will require the development of a suitable infection model to study the development of *Cryptosporidium* chipmunk genotype II. Without these biological studies, we cannot exclude the possibility that one or both of the types is a non- functional pseudogene; however, the pattern of nucleotide substitutions observed in the present study suggests that both types have evolved under functional constraints. If we assume that the types are functional, one possibility is that they have adapted to different conditions encountered by the parasite in its hibernating host. During hibernation, chipmunks alternate between periods of torpor, which is characterized by no food consumption, low metabolism, and a body temperature below 10°C, and brief periods of activity, during which individuals eat from a food cache, experience an elevated metabolism, and reach normothermic body temperatures of 36–41°C (Careau et al., 2012; Landry-Cuerrier et al., 2008; Munro et al., 2005; Snyder, 1982; Wang and Hudson, 1971). Adaptation of the 18S rRNA molecule to different body temperatures could allow *Cryptosporidium* chipmunk genotype II to survive during chipmunk hibernation. The difference in G+C content of the types could support this hypothesis, as a correlation has been shown between the G+C content of 18S rRNA molecules and body temperature (Varriale et al., 2008).

The separation of 18S rDNA copies on different chromosomes limits the extent to which recombination can homogenize nucleotide sequences among copies. It has been proposed that apicomplexan rDNA paralogs evolve independently by a birth-and-death model, and similarity among copies is maintained by purifying selection rather than recombination (Rooney, 2004). However, our identification of chimeric sequences in chipmunk genotype II suggests that recombination could play a role in the evolution of *Cryptosporidium* 18S rDNA. El-Sherry et al. (2013) similarly detected chimeric 18S rDNA sequences in *Eimeria*, but concluded that these were PCR artifacts. While artifactual recombination cannot be ruled out in the present study, the consistent amplification of chimeras in different samples and using different amplification primers supports the conclusion that recombination occurred naturally.

The remarkable divergence of Types A and B in chipmunk genotype II suggests that one or both types have undergone accelerated evolution. Ohno (1970) proposed that redundancy frees duplicated genes from natural selection, allowing them to rapidly acquire mutations and, potentially, new functions. Using Tajima's relative rate test, Type B was shown to have evolved more rapidly than Type A, suggesting that Type B is a duplicated (paralogous) copy. In a *Cryptosporidium* phylogeny constructed using Type A but not Type B (data not shown), Type A clusters with intestinal *Cryptosporidium* species, which is consistent with the intestinal localization of isolates from this study. In contrast, Type B clusters outside the intestinal and gastric groups.

Studies on *Cryptosporidium* in wild populations can provide insight into the evolution and biology of *Cryptosporidium* parasites. This study has shown that highly divergent 18S rDNA paralogs can occur in *Cryptosporidium*. Identifying such occurrences will be important for the accurate inference of species relationships from 18S rDNA sequences. Future studies should aim to identify the factors contributing to 18S rDNA divergence in *Cryptosporidium* and determine the functions of divergent paralogs.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors gratefully acknowledge support from USDA National Institute of Food and Agriculture (project number: 2008-35102-19260), NIH (project numbers: 2P20 RR015566 and 1R15AI067284-01A1), and the Ministry of Education, Youth and Sports of the Czech Republic (project number: LH11061). The authors also gratefully acknowledge graduate research fellowship support from the North Dakota Water Resources Research Institute.

References

- Arrowood MJ, Donaldson K. Improved purification methods for calf-derived *Cryptosporidium parvum* oocysts using discontinuous sucrose and cesium chloride gradients. *J Eukaryot Microbiol.* 1996; 43:89S. [PubMed: 8822880]
- Barthelemy RM, Grino M, Pontarotti P, Casanova JP, Faure E. The differential expression of ribosomal 18S RNA paralog genes from the chaetognath *Spadella cephaloptera*. *Cell Mol Biol Lett.* 2007; 12:573–583. [PubMed: 17588220]
- Bhoora R, Franssen L, Oosthuizen MC, Guthrie AJ, Zweygarth E, Penzhorn BL, Jongejan F, Collins NE. Sequence heterogeneity in the 18S rRNA gene within *Theileria equi* and *Babesia caballi* from horses in South Africa. *Vet Parasitol.* 2009; 159:112–120. [PubMed: 19019541]
- Careau V, Réale D, Garant D, Speakman J, Humphries M. Stress-induced rise in body temperature is repeatable in free-ranging Eastern chipmunks (*Tamias striatus*). *J Comp Physiol [B].* 2012; 182:403–414.
- El-Sherry S, Ogedengbe ME, Hafeez MA, Barta JR. Divergent nuclear 18S rDNA paralogs in a turkey coccidium, *Eimeria meleagridis*, complicate molecular systematics and identification. *Int J Parasitol.* 2013; 43:679–685. [PubMed: 23639264]
- Feng Y, Alderisio KA, Yang W, Blancero LA, Kuhne WG, Nadareski CA, Reid M, Xiao L. *Cryptosporidium* genotypes in wildlife from a New York watershed. *Appl Environ Microbiol.* 2007; 73:6475–6483. [PubMed: 17720824]
- Gunderson JH, Sogin ML, Wollett G, Hollingdale M, de la Cruz VF, Waters AP, McCutchan TF. Structurally distinct, stage-specific ribosomes occur in *Plasmodium*. *Science.* 1987; 238:933–937. [PubMed: 3672135]
- Hill NJ, Deane EM, Power ML. Prevalence and genetic characterization of *Cryptosporidium* isolates from common brushtail possums (*Trichosurus vulpecula*) adapted to urban settings. *Appl Environ Microbiol.* 2008; 74:5549–5555. [PubMed: 18641156]
- Hofacker IL, Fontana W, Stadler PF, Bonhoeffer LS, Tacker M, Schuster P. Fast folding and comparison of RNA secondary structures. *Monatsh Chem.* 1994; 125:167–188.
- Husmeier D, Wright F, Milne I. Detecting interspecific recombination with a pruned probabilistic divergence measure. *Bioinformatics.* 2005; 21:1797–1806. [PubMed: 15572473]
- Ikarashi M, Fukuda Y, Honma H, Kasai K, Kaneta Y, Nakai Y. First description of heterogeneity in 18S rRNA genes in the haploid genome of *Cryptosporidium andersoni* Kawatabi type. *Vet Parasitol.* 2013; 196:220–224. [PubMed: 23369454]

- Jiang J, Alderisio KA, Xiao L. Distribution of *Cryptosporidium* genotypes in storm event water samples from three watersheds in New York. *Appl Environ Microbiol*. 2005; 71:4446–4454. [PubMed: 16085835]
- Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 2013; 30:772–780. [PubMed: 23329690]
- Krieger J, Fuerst PA. Evidence of multiple alleles in the nuclear 18S ribosomal RNA gene in sturgeon (Family: Acipenseridae). *J Appl Ichthyol*. 2002; 18:290–297.
- Kvá , M.; McEvoy, J.; Stenger, B.; Clark, M. Cryptosporidiosis in other vertebrates. In: Cacciò, SM.; Widmer, G., editors. *Cryptosporidium: parasite and disease*. Springer Vienna; Vienna, Austria: 2014. p. 237–323.
- Landry-Cuerrier M, Munro D, Thomas DW, Humphries MM. Climate and resource determinants of fundamental and realized metabolic niches of hibernating chipmunks. *Ecology*. 2008; 89:3306–3316. [PubMed: 19137938]
- Le Blancq SM, Khrantsov NV, Zamani F, Upton SJ, Wu TW. Ribosomal RNA gene organization in *Cryptosporidium parvum*. *Mol Biochem Parasitol*. 1997; 90:463–478. [PubMed: 9476794]
- McCutchan TF, de la Cruz VF, Lal AA, Gunderson JH, Elwood HJ, Sogin ML. Primary sequences of two small subunit ribosomal RNA genes from *Plasmodium falciparum*. *Mol Biochem Parasitol*. 1988; 28:63–68. [PubMed: 2836731]
- McCutchan TF, Li J, McConkey GA, Rogers MJ, Waters AP. The cytoplasmic ribosomal RNAs of *Plasmodium* spp. *Parasitol Today*. 1995; 11:134–138. [PubMed: 15275356]
- Milne I, Lindner D, Bayer M, Husmeier D, McGuire G, Marshall DF, Wright F. TOPALi v2: a rich graphical interface for evolutionary analyses of multiple alignments on HPC clusters and multi-core desktops. *Bioinformatics*. 2009; 25:126–127. [PubMed: 18984599]
- Morgan UM, Monis PT, Xiao L, Limor J, Sulaiman I, Raidal S, O'Donoghue P, Gasser R, Murray A, Fayer R, Blagburn BL, Lal AA, Thompson RC. Molecular and phylogenetic characterisation of *Cryptosporidium* from birds. *Int J Parasitol*. 2001; 31:289–296. [PubMed: 11226456]
- Munro D, Thomas DW, Humphries MM. Torpor patterns of hibernating eastern chipmunks *Tamias striatus* vary in response to the size and fatty acid composition of food hoards. *J Anim Ecol*. 2005; 74:692–700.
- Nawrocki, EP. PhD thesis. Washington University in Saint Louis, School of Medicine; 2009. Structural alignment and RNA homology search and alignment using covariance models.
- Ohno, S. Evolution by gene duplication. Springer; New York: 1970.
- Peng MM, Meshnick SR, Cunliffe NA, Thindwa BD, Hart CA, Broadhead RL, Xiao L. Molecular epidemiology of cryptosporidiosis in children in Malawi. *J Eukaryot Microbiol*. 2003; 50(Suppl): 557–559. [PubMed: 14736161]
- Qari SH, Goldman IF, Pieniazek NJ, Collins WE, Lal AA. Blood and sporozoite stage-specific small subunit ribosomal RNA-encoding genes of the human malaria parasite *Plasmodium vivax*. *Gene*. 1994; 150:43–49. [PubMed: 7959061]
- Reddy GR, Chakrabarti D, Yowell CA, Dame JB. Sequence microheterogeneity of the three small subunit ribosomal RNA genes of *Babesia bigemina*: expression in erythrocyte culture. *Nucleic Acids Res*. 1991; 19:3641–3645. [PubMed: 1840661]
- Rooney AP. Mechanisms underlying the evolution and maintenance of functionally heterogeneous 18S rRNA genes in apicomplexans. *Mol Biol Evol*. 2004; 21:1704–1711. [PubMed: 15175411]
- Ryan U, Fayer R, Xiao L. *Cryptosporidium* species in humans and animals: current understanding and research needs. *Parasitology*. 2014; 141:1667–1685. [PubMed: 25111501]
- Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol*. 1987; 4:406–425. [PubMed: 3447015]
- Santín M, Fayer R. Intra-genotypic variations in the *Cryptosporidium* sp cervine genotype from sheep with implications for public health. *J Parasitol*. 2007; 93:668–672. [PubMed: 17626362]
- da Sevá AP, Funada MR, Richtzenhain L, Guimarães MB, Souza SeO, Allegretti L, Sinhorini JA, Duarte VV, Soares RM. Genotyping of *Cryptosporidium* spp. from free-living wild birds from Brazil. *Vet Parasitol*. 2011; 175:27–32. [PubMed: 21035268]
- Snyder DP. *Tamias striatus*. *Mammalian Species*. 1982; 168:1–8.

- Sulaiman IM, Lal AA, Xiao L. Molecular phylogeny and evolutionary relationships of *Cryptosporidium* parasites at the actin locus. *J Parasitol.* 2002; 88:388–394. [PubMed: 12054017]
- Sulaiman IM, Morgan UM, Thompson RC, Lal AA, Xiao L. Phylogenetic relationships of *Cryptosporidium* parasites based on the 70-kilodalton heat shock protein (HSP70) gene. *Appl Environ Microbiol.* 2000; 66:2385–2391. [PubMed: 10831415]
- Tajima F. Simple methods for testing the molecular evolutionary clock hypothesis. *Genetics.* 1993; 135:599–607. [PubMed: 8244016]
- Tamura K. Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G+C-content biases. *Mol Biol Evol.* 1992; 9:678–687. [PubMed: 1630306]
- Tamura K, Stecher G, Peterson D, Filipinski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol.* 2013; 30:2725–2729. [PubMed: 24132122]
- Tavaré, S. Some probabilistic and statistical problems in the analysis of DNA sequences. In: Miura, RM., editor. *Some mathematical questions in biology: DNA sequence analysis (lectures on mathematics in the life sciences)*. American Mathematical Society; New York: 1986. p. 57-86.
- Varriale A, Torelli G, Bernardi G. Compositional properties and thermal adaptation of 18S rRNA in vertebrates. *RNA.* 2008; 14:1492–1500. [PubMed: 18567811]
- Vrba V, Poplstein M, Pakandl M. The discovery of the two types of small subunit ribosomal RNA gene in *Eimeria mitis* contests the existence of *E. mivati* as an independent species. *Vet Parasitol.* 2011; 183:47–53. [PubMed: 21767912]
- Wang LCH, Hudson JW. Temperature regulation in normothermic and hibernating eastern chipmunk, *Tamias striatus*. *Comp Biochem Physiol A.* 1971; 38:59–90. [PubMed: 4395846]
- Widmer G, Sullivan S. Genomics and population biology of *Cryptosporidium* species. *Parasite Immunol.* 2012; 34:61–71. [PubMed: 21595702]
- Wuyts J, Van de Peer Y, Winkelmans T, De Wachter R. The European database on small subunit ribosomal RNA. *Nucleic Acids Res.* 2002; 30:183–185. [PubMed: 11752288]
- Xiao L, Limor JR, Li L, Morgan U, Thompson RC, Lal AA. Presence of heterogeneous copies of the small subunit rRNA gene in *Cryptosporidium parvum* human and marsupial genotypes and *Cryptosporidium felis*. *J Eukaryot Microbiol.* 1999; 46:44S–45S. [PubMed: 10519242]
- Xiao L, Singh A, Limor J, Graczyk TK, Gradus S, Lal A. Molecular characterization of *Cryptosporidium* oocysts in samples of raw surface water and wastewater. *Appl Environ Microbiol.* 2001; 67:1097–1101. [PubMed: 11229897]
- Yang W, Chen P, Villegas EN, Landy RB, Kanetsky C, Cama V, Dearen T, Schultz CL, Orndorff KG, Prelewicz GJ, Brown MH, Young KR, Xiao L. *Cryptosporidium* source tracking in the Potomac River watershed. *Appl Environ Microbiol.* 2008; 74:6495–6504. [PubMed: 18776033]

Highlights

Two 18S rDNA types (A and B), sharing <93% similarity, were found in 26/28 chipmunks.

Type B included the previously described *Cryptosporidium* chipmunk genotype II

Actin and HSP70 gene sequences from samples with both types were homogeneous.

It is concluded that Types A and B are highly divergent 18S rDNA paralogs.

Types A and B exhibit different rates of evolution and have different G+C contents.

Figure 1. Maximum likelihood tree of 18S rDNA sequences

The Tamura-3-parameter substitution model was used. The tree with the highest log likelihood is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Only values >50% are shown. The tree on the left was rooted with 18S rDNA from *Monocystis agilis* (accession no. AF457127). Sequences from this study are bolded. Type A and Type B sequences are highlighted with black and gray backgrounds, respectively. Dashed lines with arrows link sequences from the same sample.

Figure 2. Maximum likelihood tree of actin gene sequences

The general time reversible substitution model was used. The tree with the highest log likelihood is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Only values >50% are shown. Sequences from this study are bolded. The tree was rooted with an actin gene sequence from *Plasmodium falciparum* (accession no. EF472536).

Figure 3. Maximum likelihood tree of heat shock protein 70 gene sequences

The general time reversible substitution model was used. The tree with the highest log likelihood is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Only values >50% are shown. Sequences from this study are bolded. The tree was rooted with a heat shock protein 70 gene sequence from *Plasmodium falciparum* (accession no. M19753).

Figure 4. Structure and pairwise sequence alignments of HR-2039 Type A and HR-2039 Type B 18S rDNA

Structure alignments of HR-2039 Type A (A) and HR-2039 Type B (B) were constructed using SSU-ALIGN and SSU-DRAW (Feng et al., 2007). Helix numbering (H11-H26) is based on a system described previously (Nawrocki, 2009). The inset (C) shows a MAFFT pairwise sequence alignment of HR-2039 Type A and HR-2039 Type B, with sequence identity indicated by white lettering against a black background. Identical sequences from the pairwise alignment are also indicated in the structure alignment by white lettering against a black background. Hypervariable V2, V3, and V4 regions are highlighted in gray in the structure and pairwise alignments.

Figure 5. Phylogenetic evidence of chimeric Type A and Type B sequences

Maximum likelihood trees were constructed from nucleotides 1-297 (A) and 298-629 (B) of Type A and Type B 18S rRNA gene sequences. Nucleotide 297 represents the recombination breakpoint determined using the Probabilistic Divergence Measures method in TOPALi v2.5. Based on goodness of fit, Jukes-Cantor with invariant rates and Tamura 3-parameter with gamma-distributed rates were selected as substitution models for alignments of nucleotides 1-297 (A) and 298-629 (B), respectively (Tamura et al., 2013). Chimeric sequences resulting in discordance in the tree topologies are highlighted with white text against a black background. The tree with the highest log likelihood is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Only values >50% are shown.

Figure 6. Endogenous *Cryptosporidium* chipmunk genotype II developmental stage (arrowhead) Hematoxylin and eosin stained *Cryptosporidium* developmental stage in the jejunum of an eastern chipmunk under A) 600× and B) 1000× total magnification.

Figure 7. *Cryptosporidium* chipmunk genotype II oocysts

Oocysts were stained with anti-*Cryptosporidium* FITC-conjugated antibody (FITC) and viewed under epifluorescence and differential interference contrast (DIC) microscopy.

Table 1
Chipmunk samples positive for *Cryptosporidium* spp. 18S rDNA

Samples positive for chipmunk genotype II Type A, chipmunk genotype II Type B, chipmunk genotype IV, and *C. andersoni* are indicated.

Chipmunk ID	Sample ID	chipmunk genotype II		chipmunk genotype IV	<i>C. andersoni</i>
		Type A	Type B		
Tast-1	ID-1985	+	+	-	-
Tast-2	FL-2033	+	+	-	-
Tast-3	HR-2038	+	+	-	-
Tast-4	HR-2039	+	+	-	-
Tast-5	ID-2165	+	+	-	-
Tast-6	GM-2688	+	+	-	-
Tast-7	GM-2689	+	+	-	-
Tast-8	GM-2690	+	+	-	-
Tast-9	GM-2691	+	+	-	-
Tast-10	H-3117 ^a	-	+	+	-
	H-3130 ^b	+	+	+	-
	H-3214 ^c	+	+	-	-
Tast-11	H-3181	+	-	-	-
Tast-12	ID-3134	+	-	-	-
Tast-13	ID-3135	+	+	-	-
Tast-14	ID-3169 ^d	+	-	-	-
	ID-3153 ^e	+	+	-	+
	ID-3154 ^f	+	+	-	+
	ID-3155 ^g	+	+	-	-
	ID-3156 ^h	+	+	-	-
Tast-15	ID-3166	+	+	-	-
Tast-16	ID-3168	+	+	-	-
Tast-17	ID-3177 ⁱ	+	-	-	-
	ID-3175 ^j	+	+	-	-

Chipmunk ID	Sample ID	chipmunk genotype II		chipmunk genotype IV	<i>C. andersoni</i>
		Type A	Type B		
Tast-18	ID-3344	+	+	-	-
Tast-19	ID-3347	+	+	-	-
Tast-20	ID-3348	+	+	-	-
Tast-21	ID-3349	+	+	-	-
Tast-22	ID-3350	+	+	-	-
Tast-23	ID-3351	+	+	-	-
Tast-24	ID-3353	+	+	-	-
Tast-25	ID-3356	+	+	-	-
Tast-26	ID-3358	+	+	-	-
Tast-27	ID-3359	+	+	-	-
Tast-28	ID-3362	+	+	-	-

Samples were of feces unless indicated otherwise. A sample was obtained from Tast-10 on ^a June 29, ^b June 30, and ^c September 1; A fecal sample was obtained from Tast-14 on ^d August 23 and samples of the ^e stomach, ^f small intestine, ^g large intestine, and ^h cecum were obtained following the animal's death on August 26; Samples were obtained from Tast-17 on ⁱ August 24 and ^j August 26.

+ Positive; – Negative.

Table 2
Percent sequence identity at 18S rDNA, actin, and heat shock protein-70 (HSP70) loci

Comparisons between *C. parvum* (accession nos. XM_625373, XM_001388245, and NC006986 for 18S rDNA, actin, and HSP70 loci, respectively), *C. hominis* (accession nos. XM_661662, XM_661095, and NW667442 for 18S rDNA, actin, and HSP70 loci, respectively), *C. muris* [accession nos. XM_002140816, XM_002140165, and EU245045 for 18S rDNA, actin, and HSP70 loci, respectively) and *Cryptosporidium* chipmunk genotype II Type A and Type B using the Kimura 2-parameter distance model.

Comparison	% sequence identity $\pm$ standard error		
	18S rDNA	Actin	HSP70
<i>C. parvum</i> and <i>C. hominis</i>	99.3 $\pm$ 0.3	98.5 $\pm$ 0.4	98.5 $\pm$ 0.3
<i>C. parvum</i> and <i>C. muris</i>	91.0 $\pm$ 1.1	81.4 $\pm$ 1.6	79.0 $\pm$ 1.2
Type A and Type B	92.7 $\pm$ 1.0	99.9 $\pm$ 0.1	99.9 $\pm$ 0.1
Type B and <i>C. parvum</i>	87.0 $\pm$ 1.5	88.7 $\pm$ 1.2	87.5 $\pm$ 0.9
Type A and <i>C. parvum</i>	93.6 $\pm$ 1.1	88.7 $\pm$ 1.2	87.5 $\pm$ 0.9
Type B and <i>C. muris</i>	85.6 $\pm$ 1.5	79.2 $\pm$ 1.7	73.5 $\pm$ 1.4
Type A and <i>C. muris</i>	89.4 $\pm$ 1.3	79.2 $\pm$ 1.7	73.5 $\pm$ 1.4

Table 3**Tajima's relative rate test**

Type A and Type B sequences from three samples were included in the analysis. Sequences from each sample were tested against three outgroups: *C. canis* (accession no. AB210854), *C. parvum* (accession no. AF161857), and *C. muris* (accession no. EU245045).

Outgroup	Sample	Number of unique sequence differences		Chi-squared statistic	P value
		Type A	Type B		
<i>C. canis</i>	ID-1985	2	36	30.4	<0.0001
	HR-2039	3	37	28.9	<0.0001
	GM-2688	3	37	28.9	<0.0001
<i>C. parvum</i>	ID-1985	1	37	34.1	<0.0001
	HR-2039	1	39	36.1	<0.0001
	GM-2688	1	39	36.1	<0.0001
<i>C. muris</i>	ID-1985	8	25	8.8	<0.01
	HR-2039	9	26	8.3	<0.01
	GM-2688	8	27	10.3	<0.01

Table 4**G+C content of 18S rRNA gene sequences**

Cryptosporidium 18S rRNA gene sequences are from this study and published sequences in GenBank. *Plasmodium vivax* A, S, and O-type 18S rRNA gene sequences from GenBank also are included.

Species/genotype	Type	Isolate [accession no.]	Percent G+C				
			All ^a	Core ^b	V3 ^c	V4 ^d	
chipmunk genotype II	A	FL-2033 A	[KC954238]	40.8	48.1	36.2	31.5
		GM-2688 A	[KC954236]	40.9	48.2	36.2	31.5
		GM-2691 A	[KC954231]	41.6	48.7	40.6	31.5
		H-3181 A	[KC954212]	40.8	48.1	36.2	31.5
		H-3214 A	[KC954211]	40.8	48.1	36.2	31.5
		HR-2038 A	[KC954229]	40.8	48.1	36.2	31.5
		HR-2039 A	[KC954226]	40.8	48.1	36.2	31.5
		ID-1985 A	[KC954224]	40.8	48.1	36.2	31.5
		ID-2165 A	[KC954223]	40.8	48.1	36.2	31.5
		ID-3134 A	[KC954222]	41.8	48.7	40.6	32.0
		ID-3135 A	[KC954220]	41.0	48.1	36.2	32.0
		ID-3155 A	[KC954218]	40.9	48.1	36.2	31.7
		ID-3156 A	[KC954219]	40.9	48.1	36.2	31.7
		ID-3168 A	[KC954217]	40.9	48.1	36.2	31.7
		ID-3169 A	[KC954216]	41.0	48.1	36.2	32.0
		ID-3344 A	[KJ769643]	40.8	48.1	36.2	31.5
		ID-3347 A	[KJ769644]	41.0	48.1	36.2	32.0
		ID-3348 A	[KJ769645]	40.7	48.1	36.2	31.0
		ID-3353 A	[KC954214]	41.0	48.1	36.2	32.0
		ID-3356 A	[KC954215]	41.0	48.1	36.2	32.0
		AVERAGE			41.0	48.1	36.6
chipmunk genotype II	B	Ec-14887	[EU096238]	45.6	48.7	40.6	43.1
		W21	[AY737589]	46.0	49.0	40.6	44.4

Species/genotype	Type	Isolate [accession no.]		Percent G+C		
				All ^a	Core ^b	V3 ^c V4 ^d
		FL-2033 B	[K0954239]	45.9	48.4	36.2 45.5
		GM-2688 B	[K0954237]	46.8	49.0	40.6 45.7
		GM-2689 B	[K0954235]	46.8	49.0	40.6 45.7
		GM-2690 B	[K0954233]	46.8	49.0	40.6 45.7
		GM-2690 B2	[K0954232]	45.1	48.4	40.6 42.6
		GM-2691 B	[K0954230]	46.8	49.0	40.6 45.7
		HR-2039 B	[K0954228]	46.6	49.0	40.6 45.2
		ID-1985 B	[K0954225]	46.5	48.7	40.6 45.2
		ID-3358 B	[KJ769647]	46.6	49.0	40.6 44.1
		AVERAGE		46.3	48.8	40.2 44.8
chipmunk genotype I (W17)			[EF641026]	35.9	48.1	29.0 20.7
chipmunk genotype III			[GQ121021]	36.0	48.1	28.2 21.6
chipmunk genotype IV		ID-3130	[K0954241]	36.0	47.8	27.5 21.1
<i>C. parvum</i> KSU-1	A		[AF015772]	35.8	48.1	29.0 20.7
	B		[AF308600]	36.1	47.8	29.0 21.0
<i>C. baileyi</i>			[AY954882]	39.1	48.7	36.2 26.2
<i>C. serpentis</i>			[AF093502]	40.0	49.0	36.2 26.9
<i>C. fragile</i>			[EU162751]	43.7	49.0	37.7 36.5
<i>Plasmodium vivax</i>	A		[U07367]	38.2	45.4	34.8 31.8
	S		[U03080]	37.7	44.4	30.0 32.7
	O		[U93095]	39.9	44.6	43.2 35.2

Using the nucleotide numbering from the alignment presented in Supplementary Figure S1, G + C content was determined from nucleotides 1-629 ^a, 32-215, 285-372, and 583-629 ^b, 216-284 ^c, and 373-582 ^d. Boxed cells highlight differences in G+C content of chimeric sequences