Molecular phylogeny of Toxoplasmatinae: comparison between inferences based on mitochondrial and apicoplast genetic sequences

Filogenia molecular de Toxoplasmatinae: comparação entre inferências baseadas em sequências genéticas mitocondriais e de apicoplasto

Michelle Klein Sercundes1; Samantha Yuri Oshiro Branco Valadas1; Lara Borges Keid2; Trícia Maria Ferreira Souza Oliveira2; Helena Lage Ferreira2; Ricardo Wágner de Almeida Vitor3; Fábio Gregori1; Rodrigo Martins Soares1*

1 Departamento de Medicina Veterinária Preventiva e Saúde Animal, Faculdade de Medicina Veterinária e Zootecnia, Universidade de São Paulo – USP, São Paulo, SP, Brasil
2 Departamento de Medicina Veterinária, Faculdade de Zootecnia e Engenharia de Alimentos, Universidade de São Paulo – USP, Pirassununga, SP, Brasil
3 Departamento de Parasitologia, Universidade Federal de Minas Gerais – UFMG, Belo Horizonte, MG, Brasil

Received November 24, 2015
Accepted February 16, 2016

Abstract

Phylogenies within Toxoplasmatinae have been widely investigated with different molecular markers. Here, we studied molecular phylogenies of the Toxoplasmatinae subfamily based on apicoplast and mitochondrial genes. Partial sequences of apicoplast genes coding for caseinolytic protease (clpC) and beta subunit of RNA polymerase (rpoB), and mitochondrial gene coding for cytochrome B (cytB) were analyzed. Laboratory-adapted strains of the closely related parasites Sarcocystis falcata and Sarcocystis neurona were investigated, along with Neospora caninum, Neospora hughesi, Toxoplasma gondii (strains RH, CTG and PTG), Besnoitia akodoni, Hammondia hammondi and two genetically divergent lineages of Hammondia heydorni. The molecular analysis based on organellar genes did not clearly differentiate between N. caninum and N. hughesi, but the two lineages of H. heydorni were confirmed. Slight differences between the strains of S. falcata and S. neurona were encountered in all markers. In conclusion, congruent phylogenies were inferred from the three different genes and they might be used for screening undescribed sarcocystid parasites in order to ascertain their phylogenetic relationships with organisms of the family Sarcocystidae. The evolutionary studies based on organellar genes confirm that the genus Hammondia is paraphyletic. The primers used for amplification of clpC and rpoB were able to amplify genetic sequences of organisms of the genus Sarcocystis and organisms of the subfamily Toxoplasmatinae as well.

Keywords: Sarcocystidae, molecular characterization, phylogeny, Toxoplasmatinae, apicoplast gene, mitochondrial gene.

Resumo

A filogenia da subfamília Toxoplasmatinae tem sido amplamente investigada com diversos marcadores moleculares. Neste estudo, a filogenia molecular da subfamília Toxoplasmatinae foi analisada através de genes de apicoplasto e mitocondriais. Foram analisadas sequências parciais de genes de apicoplasto codificadores da protease caseínilítica (clpC), e da subunidade beta da RNA polimerase (rpoB) e de gene mitocondrial codificador de citocromo B (cytB). Foram investigadas cepas adaptadas em laboratório de Sarcocystis neurona e Sarcocystis falcata, parasitos estreitamente relacionados, além de Neospora caninum, Neospora hughesi, Toxoplasma gondii (cepas RH, CTG e PTG), Besnoitia akodoni, Hammondia hammondi e duas linhagens geneticamente divergentes de Hammondia heydorni. A análise molecular, baseada em genes de organelas, não diferenciou claramente N. caninum de N. hughesi, porém foi possível confirmar as duas linhagens de H. heydorni. Foram encontradas pequenas diferenças entre as cepas adaptadas em laboratório de

*Corresponding author: Rodrigo Martins Soares. Departamento de Medicina Veterinária Preventiva e Saúde Animal – VPS, Faculdade de Medicina Veterinária e Zootecnia – FMVZ, Universidade de São Paulo – USP. Av. Prof. Dr. Orlando Marques de Paiva, 87, Cidade Universitária, CEP 05508-270, São Paulo, SP, Brasil. e-mail: rosoares@usp.br
S. falcataula e S. neurnona em todos os marcadores moleculares avaliados. Concluindo, filogenias congruentes foram reconstruídas com os três diferentes genes que podem ser úteis em triagem de parasitos sarcocistídeos não identificados, para identificar sua relação com organismos da família Sarcocystidae. Os estudos evolutivos com genes organelares confirmam que o gênero Hammondia é parafilético. Os primers utilizados para amplificação de clpC e rpoB foram capazes de amplificar sequências genéticas de organismos do gênero Sarcocystis e da subfamília Toxoplasmatinae.

Palavras-chave: Sarcocystidae, caracterização molecular, filogenia, Toxoplasmatinae, gene apicoplasto, gene mitocondrial.

Introduction

Mitochondrial genes are useful for phylogenetic analysis at different taxonomic levels of apicomplexan parasites (LIN et al., 2011; HE et al., 2014). Cytochrome oxidase II and cytochrome B (cytB) coding sequences have been used for this purpose (GJERDE, 2013a, b). Similar to mitochondria, apicoplasts are organelles of maternal inheritance (FERGUSON et al., 2005) that are commonly used to reconstruct evolutionary histories (FEAGIN, 1994; KOHLER et al., 1997).

Genes encoded in the apicoplast genome are useful markers for resolving the phylogeny of Plasmodium species, because these genes have greater phylogenetic signal than the mitochondrial genome for Plasmodium phylogeny (reviewed in ARISUE & HASHIMOTO, 2015). Among other genes within the plastid genome, two genes have greater phylogenetic signal than the mitochondrial genome for resolving the phylogeny of Toxoplasma (FEAGIN, 1994; KOHLER et al., 1997).

Materials and Methods

Samples

The DNA of tachyzoites and merozoites maintained in laboratory models of the following apicomplexan parasites were used: Toxoplasma gondii strains RH, CTG and PTG (archetypes I, II and III respectively) maintained in mice, Neospora caninum strain NC-1 maintained in VERO, Sarcocystis neurona strain 138 and Sarcocystis falcataula SF1 strain kept in CV-1 cells. Sarcocystis neurona strain 138 was obtained from the intestines of an opossum (Didelphis virginiana) that had been fed with skeletal muscles from a raccoon (Procyon lotor) that in turn had been fed with sporocysts that had been isolated in Ohio, United States (LINDSAY et al., 2004). Sarcocystis falcataula SF1 was obtained from tissues of budgerigars (Melopsittacus undulatus) that had been orally inoculated with sporocysts from the intestines of an opossum (D. virginiana) that had been isolated in California, United States (MARSH et al., 1997). Dr. David Lindsay kindly provided both the S. neurona and the S. falcataula strain.

DNA from Besnoitia akodoni was also used. This is a parasite isolated from a rodent (Akodon montensis) that is native to Brazil and was obtained as previously described (DUBEY et al., 2003b).

DNA from the Neospora hughesi Oregon isolate (DUBEY et al., 2001), Hammondia hammondi (isolate 300) and two genetically divergent lineages of Hammondia heydorni (isolates 376 and BR) were obtained as described elsewhere (MONTEIRO et al., 2007).

PCR and primers

The polymerase chain reaction was performed in accordance with the manufacturer’s recommendations for platinum Taq DNA polymerase (Invitrogen, Carlsbad, CA, USA), using the primers listed in Table 1. The annealing temperature of each primer pair is given in Table 1.

Primers for clpC were designed from a consensus sequence of clpC from EST database sequences of N. caninum and S. neurona and RNAm sequences of T. gondii. Primers for cytochrome B coding sequences (cytB) targeting Sarcocystis spp. were designed from genomic sequences of S. neurona and S. falcataula (VALADAS et al., 2016). A CytB primer targeting members of the Toxoplasmatinae subfamily was designed from EST database sequences of N. caninum and RNAm sequences of T. gondii. The rpoB primer targeting the Toxoplasmatinae subfamily was designed from RpoB sequences from the genomic database for S. neurona, S. falcataula, N. caninum, T. gondii, Besnoitia oryctofelisi and Besnoitia darling.

The PCR products were observed through electrophoresis on 2% agarose gel. The DNA of positive samples was extracted from the gel and purified by using GE Healthcare kits (Illustra, GFX PCR DNA and GEL Band purification kits) in accordance with the manufacturer’s instructions.

Sequencing and sequence analysis

The PCR products were sequenced using the ABI PRISM BigDye® Terminator v3.1 kit (Ready Reaction Cycle Sequencing, Applied Biosystems). The DNA was purified by means of alcohol...
precipitation and sequencing products were analyzed using an automated sequencer (Applied Biosystems 3500 Genetic Analyzer).

Sequences were assembled and the contig sequence was obtained by means of the Phred base-calling software and the Phrap assembly tool, which are available in the Codoncode aligner v.4.2.1 suite (Codoncode Corp. Dedham, MA, USA). The sequences were aligned by means of ClustalW, which is available in the BioEdit Sequence Alignment Editor suite (HALL, 1999), based on homologous sequences available in GenBank. Nucleotide sequences were submitted to GenBank under the accession numbers GenBank KP871703, KP871704, KP871707-KP871715, KP871716, KP871717, KP871723- KP871731, KP871732, KP871733, KP871734-KP871742.

Phylogenies and DNA polymorphism were evaluated by means of MEGA 6 (TAMURA et al., 2013) and DNAsp v.5 (LIBRADO & ROZAS, 2009), respectively. The phylogenetic reconstructions were inferred through maximum parsimony (MP) and maximum likelihood (ML) methods. Apicoplast gene segments were analyzed individually and concatenated. ML phylogenies were based on nucleotide substitution model calculated for each set of data. Initial tree(s) for the heuristic search were obtained by applying the Neighbor-Joining method to a matrix of pairwise distances estimated using the Maximum Composite Likelihood approach. The MP trees were obtained using the subtree-pruning-regrafting algorithm with search level 1 in which the initial trees were obtained by the random addition of sequences (10 replicates).

In all cases, the reliability of the inferred tree was tested by bootstrap resampling technique (FELSENSTEIN, 1985) after 1000 replicates. Briefly, if there are m sequences, each with n nucleotides, a phylogenetic tree can be reconstructed using some tree building method. From each sequence, n nucleotides are randomly chosen with replacements, giving rise to m rows of n columns each. These now constitute a new set of sequences. A tree is then reconstructed with these new sequences using the same tree building method as before. Next, the topology of this tree is compared to that of the original tree. Each interior branch of the original tree that is different from the bootstrap tree the sequence it partitions is given a score of 0; all other interior branches are given the value 1. This procedure of resampling the sites and the subsequent tree reconstruction is repeated several times (typically from 100 to 1000 times), and the percentage of times (the bootstrap value) each interior branch is given a value of 1 is noted.

Results and Discussion

The cytB phylogenetic reconstruction is shown in Figure 1. The two Hammondia heydorni isolate sequences were identical to each other, and the sequences of the genus Neospora were also identical to each other. Archetypes I, II and III of Toxoplasma gondii (strains RH, CTG and PTG, respectively) were identical.

Molecular analysis using cytB consistently separates all genera within the Toxoplasmatinae subfamily and shows high similarity (~100%) within the group. The genus Neospora is formed by two known species, Neospora caninum and Neospora hughesi, which differ markedly in relation to pathogenicity and antigenicity traits (MARSH et al., 1998, 1999). Although Neospora caninum has a known capacity to infect cattle and cause abortion problems, Neospora hughesi is related to neurological disease in horses. Reports on N. hughesi infections are scarce (VARDELEON et al., 2001; WOBESER et al., 2009; PUSTERLA et al., 2014), and little is known about this agent. The two Neospora species are identical at cytB, and the same is valid for the two Hammondia heydorni isolates, which were previously discriminated into two distinct genetic lineages (MONTEIRO et al., 2007). Hammondia heydorni can be separated into two molecularly divergent lineages when compared at the alpha-tubulin, HSP 70 and ITS-1 genes (ABEL et al., 2006; MONTEIRO et al., 2007; GJERDE & DAHLGREN, 2011; SOARES et al., 2011). A third lineage of Hammondia heydorni was recently reclassified as Hammondia triiflitti, due to its molecular divergence from the other lineages and to its characteristic of only using foxes, and not dogs, as its definitive host. Isolates classified as H. triflitti have cytB fragments identical to those of the H. heydorni lineages used in this study (GJERDE, 2013a). A single synonymous nucleotide substitution differentiates the sequence of SF1 from that of SN138.

ClpC phylogeny also discriminates the clades Toxoplasmatinae subfamily and Sarcocystis genus. It also resolves the phylogeny of Toxoplasmatinae but with less statistical support than the CytB
Figure 1. Phylogenies based on cytB, clpC, rpoB. [Apico]: clpC and rpoB concatenated. [ML]: Maximum Likelihood. [MP]: Maximum Parsimony. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. For ML analysis, Tamura 3-parameter model with Gamma distribution (T92+G) was selected by MEGA 6 model selection option for cytB-based phylogeny. Tamura 3-parameter with the rate variation model allowed for some sites to be evolutionarily invariable (T92+I) was selected for clpC, rpoB, and apico-based phylogenies. The MP trees were obtained using the subtree-pruning-regrafting algorithm with search level 1 in which the initial trees were obtained by the random addition of sequences (10 replicates). There were a total of 549, 423, 434, and 857 positions in the final dataset of cytB, clpC, rpoB and Apico, respectively.
The phylogenetic reconstruction is shown in Figure 1. *Toxoplasma gondii* strain RH differs from strains CTG and PTG in one synonymous substitution. A single synonymous substitution is found between *N. hughesi* and *N. caninum* sequences, whereas the two *H. heydorni* isolates differed in six nucleotides, among which one is a non-synonymous substitution and the rest are synonymous. Concerning the *clpC* locus, the topology of the phylogenetic reconstruction was similar to that obtained with *cytB*. *ClpC* was less conserved than *CytB*. *S. neurona* 138 and *S. falcatus* SF1 had differences of the same magnitude as seen in *T. gondii* variants and the two *Neospora* species. Meanwhile, the two *H. heydorni* lineages differed by magnitudes similar to those observed between *T. gondii* and *H. hammondii*, thus corroborating the divergence previously found in these two lineages of *H. heydorni* (MONTEIRO et al., 2007).

Similar to *clpC* analysis, *rpoB* phylogeny also allows reconstruction of Sarcocystidae phylogeny but with less statistical support than in the reconstruction with the *CytB* gene (Figure 1), in particular the node that segregates all other Toxoplasmatinae from Besnoitia akodoni. The distance between the two *Hammondia heydorni* isolates is the result of a single synonymous substitution. *N. hughesi* and *N. caninum* sequences are identical at this locus. Unlike the other organelle genome sequences, the *rpoB* sequences presented insertions and deletions among the organisms of the Toxoplasmatinae clade. After concatenate apicoplast derived genes, the resulting phylogeny have shown stronger statistic support and corroborated each individual genealogy.

The *rpoB* dataset consists of oocyst sequences from a California sea lion (*Zalophus californianus*) that were obtained from GenBank. These were previously identified as oocysts from the genus *Neospora*. The phylogenetic reconstruction in Figure 2 shows that these organisms (HM173319 and HM173318) are as closely related to *Neospora* spp. as to *Hammondia heydorni*, and probably represent a new genus within Toxoplasmatinae subfamily. One synonymous difference was found between the SF1 isolate and another two *S. falcatus* sequences from GenBank (AY164999, AY165001). This was the same difference as between SF1 and the *S. lindseyi* sequence (AY164997). SF1 and SN138 differ from each other by three synonymous nucleotide substitutions. On the other hand, SN138 differs from *S. neurona* AY165000 by three nucleotide differences, among which one is non-synonymous. This is a significantly greater difference than the one existing between the two *Neospora* species and is also greater than between the two lineages of *H. heydorni*.

Regarding the Toxoplasmatinae subfamily, *rpoB* molecular analysis reveals great similarity within groups, since no differences of more than one nucleotide exist between sequences of the same clade. By adding *Neospora*-like organisms (CARLSON-BREMER et al., 2012) that are shed by aquatic mammals, the phylogenetic reconstruction through parsimony displays these sequences in clades that are divergent from both *Neospora* spp. and *H. heydorni*. These two sequences are related to each other; they form a monophyletic group, but are considerably divergent from each other, and this can be observed from the length of the branches from its common ancestry. Therefore, these two sequences likely represent distinct species, which have not been yet described in the Toxoplasmatinae subfamily.

The DNA polymorphism between sequences of *Toxoplasma gondii* RH, *Hammondia heydorni* 376, *Hammondia hammondi* BR, *Hammondia hammondii*, *Neospora hughesi* and *Neospora caninum* NC1 showed that *rpoB* and *clpC* have similar diversity values and that both genes are less conserved than *cytB*. The *rpoB* sequences are the ones with the highest diversity in non-synonymous sites (Table 2).

As expected, the phylogenies reconstructed using organelar genes of Sarcocystidae are concordant. Thereby, as in other studies (GJERDE, 2013a, b), organelar genetic sequences can give a reliable estimate of the diversity among sarcocystid species. Gjerde (2013a), demonstrated unequivocally that mitochondrial markers have the ability to reconstruct evolutionary history within the Sarcocystidae family. The *cox2* sequences of *Toxoplasma gondii*, *Neospora caninum*, *Hammondia triffitti* and *Hammondia heydorni*
reported in the work above had nucleotide diversity of 0.06484, whereas the diversity in synonymous and non-synonymous sites was 0.21371 and 0.01569, respectively (data not shown). \(RpoB\) and \(clpC\) sequences had similar diversity, which was both inferior to \(cox2\) and superior to \(cytB\). However, of all the markers, \(rpoB\) sequences presented the highest diversity at non-synonymous sites, which suggests that this marker is more susceptible to positive selection pressure. On the other hand, \(cytB\) showed diversity at non-synonymous sites that was four to five times less than that of the others, and therefore was more subject to purifying selection.

In all of the evolutionary reconstructions of Toxoplasmatinae organisms, the clade holding the two \(Hammondia heydorni\) isolates has a basal position, compared with the other clades. Both \(H. heydorni\) and \(H. triffiti\) use canids as definitive hosts. The two \(H. heydorni\) isolates used in this study are molecularly different from each other (MONTEIRO et al., 2007) and from \(H. triffiti\) (GJERDE & DAHLGREN, 2011). According to the results, it would be possible to assume that the relationship between canids and felids might have favored adaptation of ancestral parasites, originally from canids, to felids, i.e. the group of animals that serves as the definitive host for both \(H. hammondii\) and \(T. gondii\).

However, in all of the phylogenetic reconstruction conducted in this study and other described elsewhere, moderate statistical support, less than 75, is shown for the groups gathering \(Neospora\) spp. and \(T. gondii\); \(H. heydorni\) and \(T. gondii\); or \(H. heydorni\) and \(Neospora\). The placement of \(Neospora\) spp. as sister group of \(T. gondii/H. hammondii\) was also difficult to demonstrate with Hsp70- and 18S rRNA phylogenies (ELLIS et al., 1999; MUGRIDGE et al., 2000; MONTEIRO et al., 2007).

Another reason for the occurrence of low statistical support for those branches is the possibility that genetic recombination might have occurred at remote times over the course of the evolution of these organisms. Genetic recombination is a plausible event in this species, due to the sexual reproduction of apicomplexan organisms.

The evolutionary history of Sarcocystisidae have been extensively study using 18S rRNA and 28S rRNA (ELLIS et al., 1999; MUGRIDGE et al., 1999, 2000; MORRISON et al., 2004). However, few nucleotide differences among species of the Toxoplasma–Neospora–Hammondia at these loci make difficult to resolve consistently their phylogenetic relationships. In fact, in order to be comparable to the rest of the coccidia Toxoplasma–Neospora–Hammondia would be separate species of a single genus based only on 18S rRNA (MORRISON et al., 2004). rRNA sequences cannot always be unambiguously aligned and details of the alignment are known to greatly affect the results of phylogenetic analyses involving these sequences (MUGRIDGE et al., 2000; MORRISON et al., 2004). Thus, other universal markers are required to resolve phylogenies especially those that codes for proteins, which allow for robust alignments (MONTEIRO et al., 2007). Moreover, analysis of more than one gene is necessary in order to obtain a robust species signal (MUGRIDGE et al., 2000).

Regarding the diversity within the \(H. heydorni\)-related lineages, for these lineages to be hierarchized at species level, there needs to be certainty that they are unable to exchange genes. If two lineages do exchange genes, this event would occur during the gametogony in the intestinal epithelium of the definitive host and the infection of the host by different lineages would have to occur within a short period, because gametogony in Toxoplasmatinae does not take very long. Nevertheless, gene exchange can only be confirmed through analysis on individual oocysts, or after their clonal expansion in biological systems.

In conclusion, congruent phylogenies were inferred from the three different genes and they might be used for screening undescribed sarcocystid parasites in order to ascertain their phylogenetic relationships with organisms of the family Sarcocystidae. The evolutionary studies based on organelar genes confirm that the genus \(Hammondia\) is paraphyletic, as already stated elsewhere (ELLIS et al., 1999; MONTEIRO et al., 2007). The primers used for amplification of \(clpC\) and \(rpoB\) were able to amplify genetic sequences of organisms of the genus \(Sarcocystis\) and organisms of the subfamily Toxoplasmatinae as well.

Acknowledgements

RMS is in receipt of grants from CNPq, SYOBV and MKS received scholarships from FAPESP (2011/10550-2 and 2007/56836-9, respectively).

Ethics

The study was approved by the Ethics Committee for Use of Animals of the School of Veterinary Medicine and Animal Science of the University of Sao Paulo, Sao Paulo, Brazil, under protocol 1241/2007.

Conflict of interest statement

We confirm that the manuscript has been read and approved by all the named authors and that there are no other persons who satisfied the criteria for authorship but are not listed. We further confirm that the order of authors listed in the manuscript has been approved by all of us.
References


Mugridge NB, Morrison DA, Heckereth AR, Johnson AM, Tenter AM. Phylogenetic analyses based on full-length large subunit ribosomal RNA gene sequence comparison reveals that Neopora caninum is more


