International Journal for Parasitology 42 (2012) 353–363

Contents lists available at SciVerse ScienceDirect

International Journal for Parasitology

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

Phylogeny and evolution of the Piroplasmida as inferred from 18S rRNA sequences

Justin B. Lack [a,*], Mason V. Reichard [b], Ronald A. Van Den Bussche [a]

[a] Department of Zoology, 501 Life Sciences West, Oklahoma State University, Stillwater, OK 74078, USA
[b] Department of Veterinary Pathobiology, 250 McElroy, Center for Veterinary Health Sciences, Oklahoma State University, Stillwater, OK 74078, USA

ARTICLE INFO

Article history:
Received 8 December 2011
Received in revised form 13 February 2012
Accepted 14 February 2012
Available online 13 March 2012

Keywords:
Babesia
Cytauxzoon
Theileria
Piroplasmida
18S rDNA
Phylogenetics
Duncani Group
Microti Group

ABSTRACT

The order Piroplasmida consists of several genera of tick-borne parasites that infect mammals, and to a lesser extent birds, and are therefore of medical and economic importance. Despite their importance, considerable confusion exists concerning the relationship among piroplasmid species, specifically concerning the number of genera and the intergeneric relationships. To examine evolutionary relationships among piroplasmids, we conducted phylogenetic analyses of 192 18S rDNA sequences from the genera *Theileria*, *Babesia* and *Cytauxzoon*. Our analyses revealed eight clades potentially representing distinct genera, and we distinguish the Duncani Group and Microti Group as genetically distinct groups of species requiring detailed analysis of morphology and life-history to allow formal generic description. The piroplasmid phylogeny revealed considerable host diversity and limited host specificity, suggesting piroplasmids have undergone frequent host switches during their evolution. Our analyses provide the first reported evolutionary timescale for piroplasmids independent of the assumption of parasite–host cospeciation, which is invalid for piroplasmids. Evolutionary rate analyses revealed considerable substitution rate heterogeneity, which we attribute to host switching and diversification. Finally, we call for a comprehensive phylogenetic, morphological and life-history analysis for these medically relevant taxa to resolve relationships and understand host specificity.

1. Introduction

The genus *Babesia* is composed of tick-borne apicomplexans that infect a diverse array of mammalian and a few avian hosts (Criado et al., 2006). Related to the malaria-causing *Plasmodium*, species of *Babesia* are responsible for babesiosis (Uilenberg, 2006; Chauvin et al., 2009) that ranges from being subclinical to causing fever and haemolytic anaemia in infected hosts. *Babesia* are common mammalian blood-borne parasites (Telford et al., 1993; Hunfeld et al., 2008), and are widespread parasites of domesticated mammals including cattle, horses and dogs (Uilenberg, 2006).

The origin of *Babesia* was with Babes (1888) when he reported microorganisms in erythrocytes of cattle following an outbreak of bovine hemoglobinuria in Romania. Babes subsequently found similar microorganisms in red blood cells of sheep. Smith and Kilbourne (1893) described *Pyrosoma bigemina*, the etiological agent of Texas fever in cattle. In the same year, Starcovici (1893) gave these three parasites the names of *Babesia bovis*, *Babesia ovis* and *Babesia bigemina*, respectively, as *Pyrosoma* was already occupied (see Uilenberg, 2006). The report from Smith and Kilbourne (1893) is also significant as it is the first experimental demonstration that an infectious agent of mammals could be transmitted by arthropods (e.g., the tick *Rhipicephalus* (*Boophilus*) *bovis*). There are currently over 100 described *Babesia* spp. (Gray and Weiss, 2008); however, this number is frequently increasing as more mammalian species are examined, indicating that this is a group with potentially high diversity (Criado-Fornelio et al., 2004).

Due to their ability to cause disease in domesticated mammalian species, *Babesia* have been a pathogen of concern with a significant economic impact, especially in the tropics (Collett, 2000; Kivaria et al., 2007; Hunfeld et al., 2008). *Babesia* have also been implicated in deadly human infections (Gorenflot et al., 1998; Homer et al., 2000; Herwaldt et al., 2003), with the first documented human death due to *Babesia* infection occurring in 1956 (Skrabalo and Deanovic, 1957). The majority of zoonotic *Babesia* infections have been attributed to *Babesia microti* and *Babesia divergens*, but our understanding of *Babesia* diversity and evolution is limited, especially relative to closely related *Plasmodium*. Moreover, evidence suggests that for many *Babesia* spp., there are few restrictions in the range of competent mammalian hosts (Hunfeld et al., 2008), and multiple *Babesia* spp. which are distinct from *B. microti* and *B. divergens* have been confirmed as infecting humans in both the US and Europe (Herwaldt et al., 2003; Conrad et al., 2006; Gray, 2006). Nonetheless, molecular studies investigating human and animal babesiosis have utilised limited datasets of only a few representatives of *Babesia* and/or have used *Theileria* as an outgroup taxon, which is inappropriate due to studies indicating

* Corresponding author. Tel.: +1 405 744 5555; fax: +1 405 744 7824.
E-mail address: justin.lack@okstate.edu (J.B. Lack).

0020-7519/$36.00
http://dx.doi.org/10.1016/j.ijpara.2012.02.005

that multiple species previously identified as *Babesia* are actually more closely related to *Theileria* in phylogenetic analyses (Criado-Fornelio et al., 2003; Schnittger et al., 2003; Allsopp and Allsopp, 2006). Without a comprehensive examination of phylogenetic relationships among all available *Babesia* isolates and other closely related piroplasmids, it is difficult to predict which species may be capable of human infection, understand the diversity of the genus, or provide a taxonomy reflective of the evolutionary history of piroplasmids.

Many *Babesia* identifications (as well as other piroplasmids) are based largely on morphology and serology, both of which are of limited utility in consistently identifying closely related apicomplexans (Gray, 2006). *Babesia* are typically differentiated from *Theileria* based on several life-cycle characteristics, including distinctions in their biology within tick vectors, the manner by which they are transmitted from vector to vertebrate host, and the location of replication in the vertebrate hosts (i.e., *Babesia* multiply only in red blood cells, while *Theileria* enter lymphocytes and develop into schizonts; see Uilenberg (2006) for a review of these characteristics). While these characters have been utilised in the past to distinguish *Babesia* from *Theileria* and other piroplasmids, multiple exceptions to these "rules" exist. Both *B. microti* and *Theileria*/*Babesia equi* have been suggested to produce schizonts and their status as members of the genus *Babesia* has been questioned (Schein et al., 1981; Mehlhorn and Schein, 1984), although these exceptions have yet to be substantiated by detailed life-history analyses.

Recent molecular studies have indicated that neither *Babesia* nor *Theileria* are monophyletic and systematic re-examination is required to determine the generic diversity of the piroplasmids (Penzhorn et al., 2001; Criado-Fornelio et al., 2003; Schnittger et al., 2003; Reichard et al., 2005; Allsopp and Allsopp, 2006; Morrison, 2009). In the most thorough phylogenetic examination to date, Criado-Fornelio et al. (2003) suggested that at least five distinct clades exist within Piroplasmida, with both *Babesia* and *Theileria* being polyphyletic and *Cytauxzoon* nested within a clade of *Theileria* spp. In addition, Criado-Fornelio et al. (2003) put forth a hypothesis for the evolution and diversification of the piroplasms. Reichard et al. (2005) also suggested five distinct clades within the piroplasmids, but these only partially agreed with those of Criado-Fornelio et al. (2003). Finally, Allsopp and Allsopp (2006)examined piroplasmid phylogenetic relationships, again with a taxonomically different dataset, and recovered five clades but with relatively low statistical support for significant portions of the phylogeny. For the above studies, incongruences were likely due to several factors: (i) the taxa utilised varied considerably among studies, limiting the utility of the resulting phylogeny in resolving conflict among genetic datasets and the life-history characters traditionally used to distinguish piroplasmid genera; (ii) the outgroups varied between studies (i.e., Criado-Fornelio et al. (2003) used the relatively distant—and likely inappropriate—*Plasmodium* lineage, while Reichard et al. (2005) and Allsopp and Allsopp (2006) used varying combinations of *Toxoplamsa*, *Neospora*, *Sarcocystis* and *Prorocentrum*), which has been shown to significantly affect ingroup topologies (Van Den Bussche and Hoofer, 2004); (iii) several previous analyses relied entirely on distance-based phylogenetic inference methods that can perform poorly on divergent datasets and/or utilised relatively few bootstrap replicates (i.e., 100) to assess support. Here, we have included comparisons among 18S rDNA sequences of all available isolates of *Babesia, Theileria* and *Cytauxzoon*, utilised the most closely related outgroup taxa available, and performed a thorough phylogenetic analysis to formally reorganise the piroplasmid taxonomy, test the hypothesis of Criado-Fornelio et al. (2003) for piroplasmid evolution, and examine the utility of various morphological and life-history characters in distinguishing genera.

2. Materials and methods

2.1. Taxonomic sampling

We obtained from GenBank all available 18S rDNA sequences from *Babesia*, *Theileria* and *Cytauxzoon* to examine the phylogenetic relationships among *Babesia* spp. and to evaluate the taxonomic status of the piroplasmid constituent taxa. The majority of available sequences were approximately 1.7 kb, but many deposited sequences were variable in length. To minimise the effects of missing data, we excluded sequences shorter than 900 bp. We also obtained complete 18S rDNA sequences for *Neospora caninum* and *Toxoplasma gondii* to serve as outgroup taxa. The total dataset resulted in 192 terminal taxa (including the two outgroup taxa), and all species identifications and accompanying GenBank accession numbers are provided in Figs. 1 and 2.

2.2. Sequence analysis

Sequences were aligned using the MAFFT aligner v6.85 (Katoh et al., 2002) application implemented in Geneious v5.3.6, and alignments were manually edited using MacClade v4.08 (Maddison and Maddison, 2000). Due to the difficulty in aligning several portions of the 18S rDNA fragment, we excluded any region where positional homology appeared questionable. All of these excluded regions included large indels that exhibited no consistency in alignment across multiple attempts. While frameworks exist to include indels in population genetics and phylogenetic analysis (i.e., Ohshima and Yoshizawa, 2011), the indels in our alignments varied considerably in length among the taxa included, leading to significant uncertainty in terms of length homology among taxa. While we took the most conservative approach to minimise erroneous resolution of relationships, it is possible that with the inclusion of these excluded regions and an improved ability to align them, additional resolution may have been obtained. Edited alignments and accompanying phylogenies have been deposited in TreeBASE (submission 12464). A Bayesian phylogenetic analysis was conducted on the total dataset using MrBayes v3.1.2 (Huelsenbeck and Ronquist, 2001). Akaike information criterion was used to identify the most appropriate model of nucleotide substitution for the Bayesian analysis in the program MrModeltest v2.2, and the GTR + I + Γ model of nucleotide substitution was indicated to be the most appropriate. The Bayesian analysis was run for 15,000,000 generations with phylogenies sampled every 1,000 generations, and values for the substitution model parameters were not defined a priori, but were treated as unknown variables with uniform priors. Resulting burn-in values (the point at which the model parameters and tree scores reached stationarity) were determined empirically by evaluating likelihood scores. The Bayesian analysis was checked for sufficient mixing, stable convergence on a unimodal posterior and effective sample sizes (Drummond et al., 2002) >200 for all parameters using TRACER v1.5.

We also conducted a maximum parsimony (MP) phylogenetic analysis using PAUP* v4.0b10. For the MP search settings, we used the heuristic search option with equal weighting of characters, the maximum number of trees retained set at 500, tree-bisection-reconnection branch-swapping, and 25 random additions of input taxa. For the MP phylogeny, estimates of nodal support were generated with 1,000 bootstrap replicates under the same heuristic search settings as described above. Pairwise maximum likelihood (ML) genetic distances were calculated to assess the extent of divergence among major clades using PAUP*. The likelihood ratio test (LRT) implemented in Modeltest v3.06 (Posada and Crandall, 1998) was used to identify the best-fit model of nucleotide substitution and estimate model parameters.