Veterinary Parasitology xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Veterinary Parasitology

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

Research paper

The apicoplast genomes of two taxonomic units of *Babesia* from sheep

Tao Wang[a,b,1], Guiquan Guan[a,1], Pasi K. Korhonen[b,1], Anson V. Koehler[b], Ross S. Hall[b], Neil D. Young[b], Hong Yin[a,*], Robin B. Gasser[b,*]

[a] State Key Laboratory of Veterinary Etiological Biology, Key Laboratory of Veterinary Parasitology of Gansu Province, Lanzhou Veterinary Research Institute, Chinese Academy of Agricultural Sciences, Gansu, PR China
[b] Faculty of Veterinary Science, The University of Melbourne, Parkville, Victoria, Australia

ARTICLE INFO

Article history:
Received 12 September 2016
Received in revised form 31 October 2016
Accepted 5 November 2016

Keywords:
Babesia/babesiosis
Illumina sequencing
Apicoplast genome
Bioinformatics
Relationships
Therapeutics

ABSTRACT

The apicoplast (ap) is a unique, non-photosynthetic organelle found in most apicomplexan parasites. Due to the essential roles that this organelle has, it has been widely considered as target for drugs against diseases caused by apicomplexans. Exploring the ap genomes of such parasites would provide a better understanding of their systematics and their basic molecular biology for therapeutics. However, there is limited information available on the ap genomes of apicomplexan parasites. In the present study, the ap genomes of two operational taxonomic units of *Babesia* (known as *Babesia* sp. Lintan [*Bl*] and *Babesia* sp. Xinjiang [*Bx*]) from sheep were sequenced, assembled and annotated using a massive parallel sequencing-based approach. Then, the gene content and gene order in these ap genomes (~30.7 kb in size) were defined and compared, and the genetic differences were assessed. In addition, a phylogenetic analysis of ap genomic data sets was carried out to assess the relationships of these taxonomic units with other apicomplexan parasites for which complete ap genomic data sets were publicly available. The results showed that the ap genomes of *Bl* and *Bx* encode 59 and 57 genes, respectively, including 2 ribosomal RNA genes, 25 transfer RNA genes and 30–32 protein-encoding genes, being similar in content to those of *Babesia bovis* and *B. orientalis*. Ap gene regions that might serve as markers for future epidemiological and population genetic studies of *Babesia* species were identified. Using sequence data for a subset of six protein-encoding genes, a close relationship of *Bl* and *Bx* with *Babesia bovis* from cattle and *B. orientalis* from water buffalo was inferred. Although the focus of the present study was on *Babesia*, we propose that the present sequencing-bioinformatic approach should be applicable to organellar genomes of a wide range of apicomplexans of veterinary importance.

1. Introduction

Babesiosis is a socioeconomically important disease of humans and other animals caused by tick-borne apicomplexans of the genus *Babesia*. This disease can have a major, adverse economic impact on the health and productivity of livestock animals, particularly ruminants, as a consequence death, reduced meat and milk production, increased sterility and abortion rates and/or costs associated with treatment or prevention (Bock et al., 2004; Uilenberg, 2006; Schnittger et al., 2012), and is an ongoing problem, particularly in tropical and subtropical regions of Australia, Asia, Africa and the Americas. Most economic impact worldwide appears to relate to babesiosis of cattle (Schnittger et al., 2012; Gohil et al., 2013), caused by *Babesia bovis*, *B. bigemina* and/or *B. divergens*, but the socioeconomic importance of babesiosis in small ruminants is also acknowledged to be considerable (Uilenberg, 2006). The main causative agents of sheep and goats are *B. ovis*, *B. motasi* and *B. crassa*, transmitted by ticks of the genera *Rhipicephalus* and *Haemaphysalis* (Uilenberg, 2006), each of which can cause relatively severe disease. However, in China, other distinct taxa of *Babesia* have been reported in small ruminants. For instance, *Babesia* sp. Lintan (*Bl*) (Guan et al., 2002) and *Babesia* sp. Xinjiang (*Bx*) (Guan et al., 2001) have been recorded; these taxa have marked differences in vector specificity, virulence and pathogenicity (Liu et al., 2007). Interestingly, while the former taxon (*Bl*) is transmitted by *Haemaphysalis* spp. and causes mild to severe disease, the latter (*Bx*) is transmitted by *Hyalomma anatolicum* and usually relates to subclinical infection (Liu et al., 2007).

Most apicomplexan protists, including *Babesia*, harbour a plastid-like organelle, termed the apicoplast (ap), which was

* Corresponding authors.
E-mail addresses: yinhong@caas.cn (H. Yin), robinbg@unimelb.edu.au (R.B. Gasser).
[1] equal first authors.

http://dx.doi.org/10.1016/j.vetpar.2016.11.004
0304-4017/

derived from a secondary endosymbiotic event with green algae (McFadden, 2011). This unique organelle is believed to play a critical role in essential metabolism of the parasite, including the synthesis of fatty acids, haem, iron-sulphur clusters and isoprenoids (Fichera and Roos, 1997; Vaishnava and Striepen, 2006; van Dooren and Striepen, 2013). Therefore, the ap genome represents a target for drugs against apicomplexans (Wiesner et al., 2008; Chakraborty, 2016). These ap genomes also provide data sets to explore the taxonomy and evolutionary relationships of apicomplexans.

However, there is surprisingly little information on ap genomes of apicomplexans and none for *Babesia* taxa of small ruminants. To date, complete ap genomes have been sequenced and/or characterised for *B. bovis*, *B. orientalis*, *B. microti*, *Cyclospora cayetanensis*, *Eimeria tenella*, *Leucocytozoon caulleryi*, *Plasmodium chabaudi chabaudi*, *Theileria parva*, *Th. equi* and *Toxoplasma gondii* (Table 1). While most of these studies used PCR and/or cloning-based approaches (Cai et al., 2003; Sato et al., 2013; Garg et al., 2014; Imura et al., 2014; Huang et al., 2015), some have utilized direct, deep sequencing of total genomic DNA (Gardner et al., 2005; Brayton et al., 2007; Kappmeyer et al., 2012; Tang et al., 2015).

In present study, Illumina technology was used to sequence the ap genomes of *Bl* and *Bx* directly from genomic DNA and a customised bioinformatics approach to annotated them. A phylogenetic analysis of the ap genomic data sets was conducted to assess the relationships of *Bl* and *Bx* with other apicomplexan parasites for which complete ap genomic data were publicly available. The results of the present study suggest that the sequencing-bioinformatic approach should be readily applicable to other protists of veterinary importance.

2. Materials and methods

2.1. Parasite materials and isolation of genomic DNA

Merozoites from clonal lines of *Babesia* sp. Lintan and *Babesia* sp. Xinjiang (designated *Bl* and *Bx*, respectively) were maintained separately in sheep erythrocytes in a continuous in vitro culture, and 'amplified' in parasite-free, splenectomised sheep (Guan et al., 2012). Animal experiments were approved (permit code: SYXK2010-0001) by the Chinese Academy of Agricultural Sciences, Gansu province, China. Merozoites were purified from blood (Guan et al., 2012), and high molecular genomic DNA was extracted using the Gentra Puregene kit (Qiagen). DNA amounts were measured using a fluorometer (Qubit, Invitrogen). DNA quality was assessed by agarose gel electrophoresis and using a BioAnalyzer (2100, Agilent).

2.2. Sequencing of ap genomes, assembly and annotation

For each *Bl* and *Bx*, one paired-end (500 bp insert size) and two mate-pair (2 kb and 5 kb) libraries (Illumina) were built from high molecular weight genomic DNA, assessed for quality and size distribution using a BioAnalyzer and then sequenced using Illumina technology (HiSeq; 2×100 reads for paired-end libraries, and 2×49 reads for mate-pair libraries). For each taxon, the genomic reads were filtered for quality using the program Trimmomatic v.0.36 (Bolger et al., 2014), assembled using the program SPAdes v.3.5.0 (Bankevich et al., 2012) and scaffolded using the program SSPACE v.3.0 (Boetzer et al., 2011). The ap genomes were extracted from the genomic assemblies and annotated. Each protein-encoding ap gene was identified by local alignments (six reading frames) using amino acid sequences conceptually translated from corresponding genes in the ap genomes of *B. bovis* (accession no. NC_011395) and *B. orientalis* (KT428643), and then manually verified. Any publicly available ap genome that had been misannotated was re-annotated. To predict ribosomal RNA genes, ap DNA sequences from *B. bovis* (NC_011395) and *B. orientalis* (KT428643) were used as queries, employing recommended algorithm parameters (Freyhult et al., 2007) in NCBI BLAST v.2.2 (Altschul et al., 1990). Transfer RNA genes were predicted using tRNAscan-SE v.1.21 (http://lowelab.ucsc.edu/cgi-bin/tRNAscan-SE/) employing the "default" setting and the "Mito/Chloroplast" source (Schattner et al., 2005). Annotated sequence data were first converted to the required format using the program tbl2asn (https://www.ncbi.nlm.nih.gov/genbank/tbl2asn2/), and the resultant annotation file was then sent to the GenBank database employing the submission toll BankIt (http://www.ncbi.nlm.nih.gov/WebSub/?tool=genbank). Nucleotide sequence data reported in this paper are available in the GenBank™ database under accession numbers: KX881914 and KX881915.

2.3. Sliding window analysis

First, nucleotide sequences of selected ap protein-encoding genes (i.e. *rpl*2, *rpl*14, *rpl*16, *rps*11, *rps*12 and *tuf*A) were concatenated and aligned (without ambiguity) among different apicomplexans (Table 1) using the program MUSCLE v.3.8 (Edgar, 2004). Then, a sliding window analysis (300 bp; 10 bp-steps) was used to estimate nucleotide diversity (π) between/among apicomplexans (Table 1) using the program DnaSP v.5 (Rozas et al., 2003). This diversity was plotted against midpoint positions, and gene boundaries were defined. Separating the analyses in this manner allowed a pairwise comparison of patterns of variation (Table 1), the identification of conserved regions for possible oligonucleotide primer design and of intervening areas for the definition of genetic markers with low, medium or high variability between/among taxa.

2.4. Phylogenetic analysis

Amino acid sequences conceptually translated from six selected protein-coding genes (i.e. *rpl*2, *rpl*14, *rpl*16, *rps*11, *rps*12 and *tuf*A) from each of the ap genomes (Table 1) were aligned using MUSCLE v.3.8, ensuring accurate alignment of homologous characters. Aligned blocks of sequences were concatenated, and the alignment was then manually adjusted. Subsequently, phylogenetic analysis of sequence data (partitioned in a gene-wise manner) was conducted separately using the methods Bayesian inference (BI) in MrBayes v.3.2.3 (Huelsenbeck and Ronquist, 2001) and maximum likelihood (ML) in RAxML v.8.0.24 (Stamatakis, 2014); *Chromera* sp. (accession no. HM222968) was used as the outgroup. For BI, following the model selection using the program ProTest v.3.4 (Darriba et al., 2011), the prior evolution model for amino acids was set to CpREV (for gene *tuf*A), JTT (for genes *rpl*14, *rps*11 and *rps*12) or RtREV (for genes *rpl*2 and *rpl*16), and the likelihood model was set to inv-gamma (Yang, 1993, 1994). From 200,000 Markov Chain Monte Carlo (MCMC) (Geyer, 1991) iterations, the first 50,000 were discarded as non-converged burn-in, and nodal support was given as a posterior probability (pp). For ML, the evolutionary models identified for BI analysis were used, and the partitioned alignment blocks were independently bootstrapped 100 times to infer nodal support values (bootstraps = bs). Unrooted trees were viewed in the program FigTree (http://tree.bio.ed.ac.uk/software/figtree/) and drawn in Inkscape (http://www.inkscape.org/en/).

3. Results and discussion

3.1. Genome features

The circular ap genomes of *Bl* and *Bx* were 30,738 bp and 30,729 bp in length, respectively (Fig. 1; Table 1). These

Please cite this article in press as: Wang, T., et al., The apicoplast genomes of two taxonomic units of *Babesia* from sheep. Vet. Parasitol. (2016), http://dx.doi.org/10.1016/j.vetpar.2016.11.004