Mitochondrial genomes of two Babesia taxa from sheep in China as a foundation for population genetic and epidemiological investigations *

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

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
c State Key Laboratory of Agricultural Microbiology, College of Veterinary Medicine, Huazhong Agricultural University, Wuhan, Hubei, 430070, China

* Note: Nucleotide sequence data reported in this article are publicly available in the GenBank database under accession nos. PRJNA338323 and PRJNA338210.
* Corresponding author.
E-mail addresses: robinbg@unimelb.edu.au (R.B. Gasser), yinhong@caas.cn (H. Yin).
1 equal first authors.
ABSTRACT

Here, we sequenced, assembled and annotated the mitochondrial (mt) genomes of two operational taxonomic units of Babesia from sheep from China using a deep sequencing-coupled approach. Then, we defined and compared the gene order of these mt genomes (~ 5.8 to 6.2 kb in size), assessed sequence differences in mt genes among Babesia taxa and evaluated genetic relationships among these taxa and related apicomplexans (Theileria) for which mt genomic data sets were available. We also identified mt genetic regions that might be useful as markers for future population genetic and molecular epidemiological studies of Babesia from small ruminants. We propose that the sequencing-bioinformatic approach used here should be applicable to a wide range of protists of veterinary importance.

Keywords:
Babesia/babesiosis
Illumina sequencing
Mitochondrial genome
Bioinformatics
Relationships
Systematics
1. Introduction

Babesiosis is an economically important disease caused by tick-borne apicomplexan protists of the genus *Babesia* (Schnittger et al., 2012; Uilenberg, 2006). This disease is characterised principally by haemolytic anaemia, haemoglobinuria, fever and icterus (Yabsley and Shock, 2013). Most economic impact worldwide appears to relate to babesiosis of cattle (Gohil et al., 2013; Schnittger et al., 2012), caused by *Babesia bovis*, *B. bigemina* and *B. divergens*, but the socioeconomic importance of babesiosis in small ruminants is also acknowledged to be considerable (Uilenberg, 2006).

In sheep and goats, the main causative agents 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 ovine *Babesia* have been recorded. For instance, *Babesia* sp. Lintan (*Bl*) (Guan et al., 2002) and *Babesia* sp. Xinjiang (*Bx*) (Guan et al., 2001) have been isolated, and are reported to display marked differences in vector specificity and virulence/pathogenicity (cf. Liu et al., 2007). Interestingly, while the former taxon is transmitted by *Haemaphysalis* spp. and causes mild to severe disease, the latter uses *Hyalomma anatolicum* and is associated with subclinical infection (Liu et al., 2007).

Although some basic research has been conducted to provide insights into the biology, epidemiology and immunology of *Bl* and *Bx* (Bai et al., 2002; Guan et al., 2001, 2002, 2010b, 2012b, 2012c; Niu et al., 2016), there have been few genetic studies (Guan et al., 2010a, 2015; Niu et al., 2013) and no genomic investigations of these two or any other *Babesia* taxa from small ruminants. Exploring their mitochondrial (mt) genomes would provide an avenue to better explore their systematic status and also defining a repertoire of genetic markers for population genetic studies (cf. Hikosaka et al., 2013). Previous studies have sequenced and characterised the mt genomes of *B. bigemina*, *B. bovis*, *B. orientalis* (bovids); *B. caballi* (equids); *B. gibsoni* (canids); *B. microti* and *B. rodhaini* (rodents) (Brayton et al., 2007; Cornillot et al., 2012; He et al., 2014; Hikosaka et al., 2010, 2012). While most of these studies used PCR and/or cloning-based approaches (He et al., 2014; Hikosaka et al., 2010, 2012), some have utilized direct, deep sequencing of total genomic DNA (Brayton et al., 2007; Cornillot et al., 2012).

In present study, we employed Illumina technology to directly sequence the mt genomes of *Bl* and *Bx* from genomic DNA, and a custom bioinformatics platform to annotated them. Using the mt genomic data sets, we then undertook a phylogenetic analysis to assess the relationships of *Bl* and *Bx* with all *Babesia* species and other related piroplasms (i.e., *Theileria*) for which mt genomic data were publicly available. Finally, we discussed the implications of these data sets of *Bl* and *Bx* for future epidemiological and population genetic applications, and the applicability of the present sequencing-bioinformatic approach for protists of veterinary importance.

2. Materials and methods

2.1. Parasite materials and isolation of genomic DNA

Merozoites representing clonal lines of each *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., 2012a); animal experiments were approved (permit SYXX2010-0001) by the Science and Technology Department of Gansu province, China. Merozoites were purified from blood as described previously (Guan et al., 2012a), and high molecular genomic DNA was isolated using the Gentra Puregene kit (Qiagen) according to the manufacturer’s protocol. DNA amounts were measured using a fluorometer (Qubit, Invitrogen), and genomic DNA quality was verified by agarose gel electrophoresis and using a BioAnalyzer (2100, Agilent).
2.2. Sequencing of mt 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; 2x100 reads for paired-end libraries, and 2x49 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 mt genomes were extracted from genomic assemblies and annotated using an established approach (Hikosaka et al., 2012). In brief, each protein-encoding mt gene was identified by local alignments (six reading frames) using amino acid sequences conceptually translated from corresponding genes from each mt genome of B. bovis (accession nos. EU075182 and AB499088; Brayton et al., 2007; Hikosaka et al., 2010) and B. bigemina (AB499085; Hikosaka et al., 2010). To predict ribosomal RNA genes, mt DNA sequences from B. bovis (EU075182 and AB499088) and B. bigemina (AB499085) were used as queries employing suggested algorithm parameters (Freyhult et al., 2007) in NCBI BLAST 2.2 (Altschul et al., 1990). Inverted repeat sequences were identified using a ‘self-against-self’ BLASTN search (Altschul et al., 1997) employing a threshold of >20 nucleotides.

Annotated sequence data were imported using the program SEQUIN (available via http://www.ncbi.nlm.nih.gov/Sequin/) for the final verification of the mt genome organization/annotation prior to submission to the GenBank database.

2.3. Sliding window analysis

Sliding window analysis was performed on aligned, concatenated nucleotide sequences of all genes of complete mt genomes representing nine recognised taxa of Babesia (cf. Table 1) using the program DnaSP v.5 (Rozas et al., 2003). The sequences were aligned using the program MUSCLE v.3.8 (Edgar, 2004); keeping the nucleotides in frame, there were no ambiguously aligned regions. A sliding window of 300 bp (10 bp-steps) was used to estimate nucleotide diversity (π) within and among members (pairwise) of each Babesia taxon using DnaSP v.5. Nucleotide diversity for the alignments was plotted against midpoint positions, and gene boundaries were defined. Separating the analyses in this manner allowed a pairwise comparison of general patterns among Babesia taxa as well as the identification of conserved regions and areas with potential for the definition of mt genetic markers with low, medium or high variability among taxa.

2.4. Phylogenetic analysis

Nucleotide or amino acid sequence conceptually translated from the protein-encoding genes from each of the mt genomes (cf. Table 1) were aligned using MUSCLE, ensuring accurate alignment of homologous characters. Aligned blocks of sequences were concatenated, and the alignment was manually adjusted. Subsequently, phylogenetic analysis of sequence data was conducted by Bayesian inference (BI) using Monte Carlo Markov Chain (MCMC) analysis in MrBayes v.3.2.3 (Huelsenbeck and Ronquist, 2001). Sequence data for the cox3 gene were excluded from phylogenetic analysis, due to high divergence among Babesia species (see Section 3). The likelihood parameters set for BI analysis were based on the Akaike Information Criteria (AIC) test in jModeltest v.2.1.7 (Darriba et al., 2012). The number of substitutions (Nst) was set at 6, with a gamma-distribution and a proportion of invariant sites. Posterior probability (pp) values were calculated by running 2,000,000 generations with four simultaneous tree-building
chains. Trees were saved every 100th generation. At the end of each run, the standard deviation of split frequencies was < 0.01, and the potential scale reduction factor approached one. A 50% majority rule consensus tree for each analysis was constructed based on the final 75% of trees generated by BI. To ensure convergence and insensitivity to priors, analyses were run three times. Plasmodium falciparum was used as the outgroup. 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. Characteristics of the mt genomes of the two Babesia taxa, Bl and Bx

The linear mt genomes of Bl and Bx were 5,790 bp and 6,020 bp in length, respectively (Fig. 1; Table 1), similar to those published previously for B. bovis, B. bigemina, B. orientalis, B. caballi and B. gibsoni (5,847-6,005 bp; Brayton et al., 2007; He et al., 2014; Hikosaka et al., 2010, 2012) and other piroplasms, including Theileria annulata, T. orientalis and T. parva (5,905-5,957 bp; Hikosaka et al., 2010; Pain et al., 2005), and smaller than those of B. microti, B. rodhaini and T. equi (6,929-11,109 bp; Cornillot et al., 2012; Hikosaka et al., 2012). The gene order of the two mt genomes was the same (Fig. 1). Similar to other Babesia mt genomes (Hikosaka et al., 2010), the overall A+T contents of the mt genomes of Bl and Bx were high, ranging from 70.0% to 71.3% (Supplementary Table 1). Both mt genomes contained three protein-encoding genes, namely cox1 and cox3 and cytochrome b (cob), five or six regions (designated as L1-L6) of the large subunit of the rRNA gene (LSU), but no tRNA genes. Interestingly, fragment L5 is missing from the mt genome of Bx compared with other Babesia species studied to date. In addition, two terminal inverted repeats (TIRs) of 56-307 bp were present (Fig. 1), in accord with the mt genomes of other Babesia species (Brayton et al., 2007; He et al., 2014; Hikosaka et al., 2010, 2012). The transcriptional direction and order of the protein-encoding genes are the same as those in B. bovis (accession nos. EU075182 and AB499088), B. bigemina (AB499085), B. orientalis (KF218819); B. caballi (AB499086); B. gibsoni (AB499087); and T. parva (AB499089), but distinctly different from those of B. microti (AB624353), B. rodhaini (AB624357), T. orientalis (AB499090), T. equi (AB499091) and P. falciparum (KT119882) (Fig. 1) (cf. Brayton et al., 2007; He et al., 2014; Hikosaka et al., 2010, 2012).

3.2. Predicted mt proteins and codon usages

Totals of 1,064 and 1,052 amino acids were encoded in the mt genomes of Bl and Bx, respectively. The initiation and termination codons predicted for the three protein-encoding genes (i.e., cox1, cox3 and cob) were compared among all nine Babesia taxa (cf. Supplementary Table 2). The ATG was the commonest start codon, followed by ATA, ATT, TTG and TTA. Most of the three mt protein-encoding genes had TAA, TGA or TAG as a translation termination codon (Supplementary Table 2). Subsequently, the codon usages of these protein-encoding genes were compared (Supplementary Table 3); most of the 64 possible codons were used in the mt genomes of Bl (n = 60) and Bx (n = 57). The preferred nucleotide usage at the third codon positions in the mt protein-encoding genes reflects the overall nucleotide composition of these two mt genomes. At this position, A is the most frequently, and C the least frequently used nucleotide. For both Bl and Bx, the frequency of codons ending in an A is higher than the codons ending in a G. These findings are consistent with those reported for mt genomes of Babesia spp. (cf. Brayton et al., 2007;
The AT richness of the Bl and Bx mt genomes was reflected in the predicted proteins. AT-rich codons represent the amino acids Leu, Phe, Ile, Met, Asn and Tyr, and GC-rich codons represent Pro, Gly, Ala and Ser (Supplementary Table 3). The TTA (Leu), TTT (Phe), ATT (Ile), GGA (Gly), GTT (Val) and ATG (Met) were the most frequently used codons. On the other hand, the GCC (Ala), GCG (Ala), GGC (Gly), GGG (Gly), CTC (Leu), CTG (Leu), CCG (Pro), CGC (Arg), AGC (Ser) and ACG (Thr) were least frequently used. The comparison of synonymous codons within the AT-rich group, such as Phe (TTT: 2.04% to 2.24%; TTC: 0.57% to 0.92%) and Ile (ATT: 1.52% to 1.57%; ATC: 0.16% to 0.21%), showed that a significantly lower frequency if C was at the third position.

3.3. Genetic differences and divergences

At the nucleotide level, the mt genomes of Bl and Bx differed by 86.8%, and pairwise comparisons of these taxa with the seven other Babesia species (B. bovis, B. bigemina, B. orientalis, B. caballi, B. gibsoni, B. microti and B. rodhaini), for which data are available, revealed differences ranging from 6.9% to 78.5% (Supplementary Table 4). The nucleotide sequence differences among all nine species ranged from 10.3% to 36.9% for cox1, from 8.9% to 57.6% for cox3, and from 8.9% to 50.9% for cob (Supplementary Table 5). By contrast, amino acid sequence differences and divergences ranged from 5.9% to 37.3%, and 17.9% to 75.5%, respectively, with COX1 being the most and COX3 being the least conserved amino acid sequences (Supplementary Table 6). These findings showed that cox1 and cob genes were more conserved than cox3. The different levels of nucleotide sequence variability in the genes suggested that some mt genome regions could be used as markers for systematic or population genetic studies.

Therefore, we used sliding window analysis to characterise the nucleotide diversity patterns across all mt genomes of Babesia (for which mt gnomic data are available; Table 1 and Fig. 2). For each taxon, due to the extensive sequence differences in TIRs, these regions could not be unambiguously aligned. Although TIRs are present in the mt genomes of all Babesia taxa studied to date (Brayton et al., 2007; Cornillot et al., 2012; He et al., 2014; Hikosaka et al., 2010, 2012), it is challenging to establish positional homology due to their non-coding nature and likely rapid evolutionary rates. In contrast, with the conservation of gene length and relative gene order of other regions of the mt DNAs, base positional homology could be readily established. The peaks and troughs of nucleotide diversity in protein-encoding regions (Fig. 2) revealed mainly transversions and transitions, with a few indels of 42-147 nucleotides. Using sliding window analyses, the nucleotide diversity patterns were found to be similar among species of Babesia, with greater amplitude of diversity between more distantly related Babesia taxa (Fig. 2). For Babesia, high variability was seen at the 5’-end of the cox1 gene, and at the 3’-ends of the cox3 and cob genes (Fig. 2).

3.4. Phylogenetic relationships

A phylogenetic analysis was conducted using nucleotide and amino acid sequence data representing Babesia and Theileria taxa for which data sets are publicly available. The BI trees constructed from concatenated cox1, cob and cox1 + cob sequence data were consistent, with no major differences in pp values. The BI trees constructed with cox1 + cob sequences are presented (Fig. 3). The analysis showed a clear estimate of interrelationships of Bl, Bx and all other taxa, with each node being strongly supported by pp values (Fig. 3). Consistent with previous studies (Criado-Fornelio et al., 2003; Hikosaka et al., 2012, 2013), three monophyletic groups were formed, i.e., babesids (Bl, Bx, B. bovis, B. bigemina, B. orientalis, B. caballi and B. gibsoni),
theilerids (T. annulata, T. parva, T. orevialis and T. equi) and archaeopiroplasms (B. microti and B. rodhaini) (Fig. 3). Within the babesids group, the taxa from ruminants grouped separately from those of other host species with complete support (pp = 1.00). Within the ‘ruminant’ Babesia clade, Bl and B. bigemina (pp = 1.00) grouped together, to the exclusion of all three other Babesia taxa that clustered. B. bovis and B. orientalis grouped together (pp = 1.00) to the exclusion of Bx. Within the ‘non-ruminant’ Babesia clade (pp = 1.00), B. caballi showed a closer branching position (pp = 1.00) to the ‘ruminant’ Babesia clade than did B. gibsoni.

3.5. Implications and concluding remarks

Using massively parallel sequencing and bioinformatics, we characterised the first mt genomes of Babesia taxa from small ruminants, Babesia sp. Lintan (Bl) (Guan et al., 2002) and Babesia sp. Xinjiang (Bx) (Guan et al., 2001). This approach was considerably more time-effective and practical to conduct than long PCR-based sequencing (cf. He et al., 2014; Hikosaka et al., 2010, 2012). As we sequenced from cloned parasite lines, nucleotide sequence variability in the mt genes within Bl and Bx was limited. Minor diversity might relate with varying substitution rates in mt genomes among mitochondria within a species. Nonetheless, it is necessary to determine levels of nucleotide variability within and among field isolates, which will usually contain many merozoites and/or other developmental stages, to exclude the presence of mixed-taxon infections in animals and make correct conclusions concerning nucleotide substitution rates.

The mt genomes of Bl and Bx were similar, in terms of the lengths of protein-encoding genes and the numbers of genes, to those reported previously for other Babesia taxa (e.g., Hikosaka et al., 2010). Nevertheless, differences in the lengths and sequences of TIRs were observed. The magnitude of genetic diversity detected in TIRs indicates potential for their use in population genetic studies of individual Babesia taxa. Comparative analysis of protein-encoding genes revealed relatively extensive sequence variability in the cox3 gene, moderate variability in that of cob and relative conservation in cox1; this information indicates that cox1 is likely to be useful for specific identification and/or differentiation, and the other ‘variable’ genes might be suited for recognizing lineages (strains) within Babesia species.

In a laboratory setting, by designing oligonucleotide primers to conserved regions (cf. Fig. 2), PCR-coupled assays (e.g., multiplexed tandem PCR; Gebrekidan et al., 2016) could be set up for the specific detection and differentiation of Babesia taxa in blood or other tissue samples. However, direct Illumina-based or single-molecule sequencing (cf. Gawad et al., 2016; Koboldt et al., 2013) followed by automated data processing and bioinformatic analyses is also now within reach, depending on cost. The development of both sequencing and bioinformatic techniques indicates that such an approach should be feasible, especially for authorities charged with the prevention and control of babesiosis and other socioeconomically important vector borne diseases. Establishing such an approach could allow rapid responses to disease outbreaks, and enable research to understand transmission patterns.

In this study, we also established the genetic relationships of Bl and Bx with selected piroplasms using mt data sets. Previously, Tian et al. (2013) suggested the phylogenetic relationships of Babesia species and genotypes using partial cob gene (550 bp) data (cf. Fig. 2), and undertook a direct comparison of the same isolates using data for nuclear small subunit (SSU) and internal transcribed spacer (ITS) rDNA. These authors concluded that SSU sequences could distinguish Babesia species and that the partial cob gene could better recognize lineages within Babesia species. Although this study was informative, the distance of some clades and/or taxa, such as Bl vs. B. bovis and Bx vs. B. bovis, were not well supported statistically. The limited resolution of some relationships was likely due to limited phylogenetic signal. Nonetheless, the present study resolved the positions of Bl, Bx, B. bovis, B. bigemina and B. orientalis in relation
to all other taxa, with consistent and absolute nodal support (pp = 1.00).

In conclusion, the mitochondrial (mt) genomes characterised for Bl and Bx provide useful bar-
codes for future comparative genomic research of Babesia taxa from small ruminants, and a rich
source of markers for future population genetic and molecular epidemiological studies. From
technical perspective, the sequencing-bioinformatic approach used to characterise these genomes
should be readily applicable to a wide range of protists of veterinary importance.

Acknowledgements

This project was partially funded by the PiroVac (KBBE-3-245145; H.Y.), NSFC (31072130;
G.G.), ASTIP, FRIP (2014ZL010), CAAS (H.Y.) and NBCIS (CARS-38; H.Y.) of China and the
Australian Research Council (ARC) (R.B.G.); it was also supported by a Victorian Life Sciences
Computation Initiative (VLSCI) grant (VR0007; R.B.G.) on its Peak Computing Facility at the
University of Melbourne, an initiative of the Victorian Government. Other support from the State
Key Laboratory of Veterinary Etiological Biology Project, China (SKLVEB2008ZZKT019;
H.Y), the Australian Academy of Science, Alexander von Humboldt Foundation and Melbourne
Water Corporation is gratefully acknowledged. N.D.Y. is an NHMRC Early Career Research
(ECRF) Fellow.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/XXXXXXXXX

References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J.,

large Babesia sp. from sheep and goats in the eastern part of Gansu Province, China. Parasitol

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M.,
Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A.V., Sirotkin, A.V., Vyahhi, N., Tesler,

Boetzer, M., Henkel, C.V., Jansen, H.J., Butler, D., Pirovano, W., 2011. Scaffolding pre-
assembled contigs using SSPACE. Bioinformatics 27, 578-579.


Brayton, K.A., Lau, A.O., Herndon, D.R., Hannick, L., Kappmeyer, L.S., Berens, S.J., Bidwell,
S.L., Brown, W.C., Crabtree, J., Fadrosh, D., Feldblum, T., Forberger, H.A., Haas, B.J.,
Howell, J.M., Khouri, H., Koo, H., Mann, D.J., Norimine, J., Paulsen, I.T., Radune, D., Ren,
Q., Smith, R.K., Jr., Suarez, C.E., White, O., Wortman, J.R., Knowles, D.P., Jr., McElwain,
T.F., Nene, V.M., 2007. Genome sequence of Babesia bovis and comparative analysis of


Cornillot, E., Hadj-Kaddour, K., Dassouli, A., Noel, B., Ranwez, V., Vacherie, B., Augagneur,


Guan, G., Moreau, E., Liu, J., Ma, M., Rogniaux, H., Liu, A., Niu, Q., Li, Y., Ren, Q., Luo, J., Chauvin, A., Yin, H., 2012c. BQP35 is a novel member of the intrinsically unstructured protein (IUP) family which is a potential antigen for the sero-diagnosis of Babesia sp. BQ1 (Lintan) infection. Vet Parasitol 187, 421-430.


FIGURE LEGENDS:

**Fig. 1.** Schematic representation of the linear mt genomes of (A) *Babesia* sp. Lintan (*Bl*) (5,790 bp) and (B) *Babesia* sp. Xinjiang (*Bx*) (6,020 bp) and related piroplasms, including (C) *B. bovis* (accession no. AB499088), (D) *B. bigemina* (AB499085) and (E) *Theileria parva* (AB499089). Genes above the horizontal line are transcribed from left to right, and those below the line are transcribed from right to left. Boxes indicate the protein-encoding genes, *cox1*, *cox3* and *cob*. Regions of the LSU gene (*L1* to *L6*) and terminal inverted repeats (TIRs) at the 5’- and 3’-ends of the mt genomes are indicated.

**Fig. 2.** Sliding window analysis of protein-encoding genes (*cox1*, *cob* and *cox3*) in the mt genomes among nine *Babesia* taxa (cf. Table 1). Nucleotide diversity, measured iteratively over 300 bp windows of aligned sequence data (every 10 bp), indicate troughs and peaks of nucleotide variability. Gene region employed by Tian et al. (2013) is shaded in grey.

**Fig. 3.** The phylogenetic tree constructed using Bayesian inference employing nucleotide sequence data for the two mt protein-encoding genes *cox1* and *cob* from *Babesia* sp. Lintan (*Bl*) and *Babesia* sp. Xinjiang (*Bx*) as well as 11 recognised species of *Babesia* and *Theileria* for which complete mitochondrial genome sequences are currently publicly available. *Plasmodium falciparum* was used as the outgroup. Host groups of *Babesia* spp. are indicated (right). All posterior probability (pp) values are 1.00, with the exception of that indicated (0.93).
**Table 1**

Taxa of piroplasms, their host and geographical origins, and mt genomes used in the present study.

<table>
<thead>
<tr>
<th>Taxon</th>
<th>Original host</th>
<th>Country of origin</th>
<th>Mt genome size (bp)</th>
<th>Accession</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Babesia sp. Lintan (Bl)</em></td>
<td>Bovidae (<em>Ovis</em>)</td>
<td>China</td>
<td>5,790</td>
<td>PRJNA338323</td>
<td>This study</td>
</tr>
<tr>
<td><em>Babesia sp. Xinjiang (Br)</em></td>
<td>Bovidae (<em>Ovis</em>)</td>
<td>China</td>
<td>6,020</td>
<td>PRJNA338210</td>
<td>This study</td>
</tr>
<tr>
<td><em>Babesia bovis</em></td>
<td>Bovidae (<em>Bos</em>)</td>
<td>United States</td>
<td>6,005</td>
<td>EU075182</td>
<td>Brayton et al. (2007)</td>
</tr>
<tr>
<td><em>B. bovis</em></td>
<td>Bovidae (<em>Bos</em>)</td>
<td>Japan</td>
<td>5,970</td>
<td>AB499088</td>
<td>Hikosaka et al. (2010)</td>
</tr>
<tr>
<td><em>B. bigemina</em></td>
<td>Bovidae (<em>Bos</em>)</td>
<td>Japan</td>
<td>5,924</td>
<td>AB499085</td>
<td>Hikosaka et al. (2010)</td>
</tr>
<tr>
<td><em>B. orientalis</em></td>
<td>Bovidae (<em>Bubalus</em>)</td>
<td>China</td>
<td>5,996</td>
<td>KF218819</td>
<td>He et al. (2014)</td>
</tr>
<tr>
<td><em>B. caballi</em></td>
<td>Equidae (<em>Equus</em>)</td>
<td>United States</td>
<td>5,847</td>
<td>AB499086</td>
<td>Hikosaka et al. (2010)</td>
</tr>
<tr>
<td><em>B. gibsoni</em></td>
<td>Canidae (<em>Canis</em>)</td>
<td>Japan</td>
<td>5,865</td>
<td>AB499087</td>
<td>Hikosaka et al. (2010)</td>
</tr>
<tr>
<td><em>B. microti</em></td>
<td>Rodentia (<em>Mus</em>)</td>
<td>Germany</td>
<td>11,109</td>
<td>AB624353</td>
<td>Hikosaka et al. (2012)</td>
</tr>
<tr>
<td><em>B. rodhania</em></td>
<td>Rodentia (<em>Mus</em>)</td>
<td>Australia</td>
<td>6,929</td>
<td>AB624357</td>
<td>Hikosaka et al. (2012)</td>
</tr>
<tr>
<td><em>Theileria annulata</em></td>
<td>Bovidae (<em>Bos</em>)</td>
<td>Turkey</td>
<td>5,905</td>
<td>NT167255</td>
<td>Pain et al. (2005)</td>
</tr>
<tr>
<td><em>T. parva</em></td>
<td>Bovidae (<em>Bos</em>)</td>
<td>Kenya</td>
<td>5,895</td>
<td>AB499089</td>
<td>Hikosaka et al. (2010)</td>
</tr>
<tr>
<td><em>T. orientalis</em></td>
<td>Bovidae (<em>Bubalus</em>)</td>
<td>Kenya</td>
<td>5,957</td>
<td>AB499090</td>
<td>Hikosaka et al. (2010)</td>
</tr>
<tr>
<td><em>T. equi</em></td>
<td>Equidae (<em>Equus</em>)</td>
<td>United States</td>
<td>8,246</td>
<td>AB499091</td>
<td>Hikosaka et al. (2010)</td>
</tr>
<tr>
<td><em>Plasmodium falciparum</em></td>
<td>Hominidae (<em>Homo</em>)</td>
<td>India</td>
<td>5,967</td>
<td>KT119882</td>
<td>Tyagi and Das (2015)</td>
</tr>
</tbody>
</table>