Divergence of the Mitochondrial Genome Structure in the Apicomplexan Parasites, Babesia and Theileria

Kenji Hikosaka, Yoh-ichi Watanabe, Naotoshi Tsuji, Kiyoshi Kita, Hiroe Kishine, Nobuko Arisue, Nirianne Marie Q. Palacpac, Shin-ichi Kawazu, Hiromi Sawai, Toshihiro Horii, Ikuo Igarashi, and Kazuyuki Tanabe

Abstract

Mitochondrial (mt) genomes from diverse phylogenetic groups vary considerably in size, structure, and organization. The genus Plasmodium, causative agent of malaria, of the phylum Apicomplexa, has the smallest mt genome in the form of a circular and/or tandemly repeated linear element of 6 kb, encoding only three protein genes (cox1, cox3, and cob). The closely related genera Babesia and Theileria also have small mt genomes (6.6 kb) that are monomeric linear with an organization distinct from Plasmodium. To elucidate the structural divergence and evolution of mt genomes between Babesia/Theileria and Plasmodium, we determined five new sequences from Babesia bigemina, B. caballi, B. gibsoni, Theileria orientalis, and T. equi. Together with previously reported sequences of B. bovis, T. annulata, and T. parva, all eight Babesia and Theileria mt genomes are linear molecules with terminal inverted repeats (TIRs) on both ends containing three protein-coding genes (cox1, cox3, and cob) and six large subunit (LSU) ribosomal RNA (rRNA) gene fragments. The organization and transcriptional direction of protein-coding genes and the rRNA gene fragments were completely conserved in the four Babesia species. In contrast, notable variation occurred in the four Theileria species. Although the genome structures of T. annulata and T. parva were nearly identical to those of Babesia, an inversion in the 3-kb central region was found in T. orientalis. Moreover, the T. equi mt genome is the largest (8.2 kb) and most divergent with unusually long TIR sequences, in which cox3 and two LSU rRNA gene fragments are located. The T. equi mt genome showed little synteny to the other species. These results suggest that the Theileria mt genome is highly diverse with lineage-specific evolution in two Theileria species: genome inversion in T. orientalis and gene-embedded long TIR in T. equi.

Key words: mitochondrial, mitochondrial genome, Babesia, Theileria, Plasmodium, Apicomplexa.

Introduction

Mitochondria, organelles essential for energy transduction and cellular functions, are present in almost all eukaryotes. Like nuclear genomes of eukaryotes, mitochondrial (mt) genomes from diverse phylogenetic groups vary considerably in size, structure, and organization as well as in the number of genes (Gray et al. 2004). The largest mt genome is found in land plants, in which the size ranges from 180 to 2,400 kb (Ward et al. 1981; Palmer et al. 1992), and the smallest is the 6-kb genome of the genus Plasmodium, causative agents of malaria. Plasmodium belongs to the phylum Apicomplexa, which includes >5,000 species, all of which are parasites of clinical or economic importance (Levine 1988). Veterinary and opportunistic pathogens include Babesia, which causes babesiosis in ruminants and humans; Theileria, causal agents for tropical theileriosis and East Coast fever in cattle; Cryptosporidium, responsible for cryptosporidiosis in humans and animals; and Toxoplasma, causing toxoplasmosis in immunocompromised patients and congenitally infected fetuses.

Relatively few apicomplexan mt genomes have been studied, and available data suggest that they are remarkably diverse in structure and genome organization. In Plasmodium, the mt genome is in the form of a circular and/or tandemly repeated, predominantly linear 6-kb element (Preiser et al. 1996; Wilson and Williamson 1997). The 6-kb element encodes only three mt protein-coding genes (cytochrome c oxidase subunits I and III: cox1 and cox3 and cytochrome b: cob) in addition to large subunit (LSU) and small subunit (SSU) ribosomal RNAs (rRNAs). The two rRNA genes are extensively fragmentated and rearranged with 20 identified rRNA pieces, and curiously, no transfer RNA genes have
yet been identified (Feagin et al. 1997). The mt genomes of closely related apicomplexan parasites Babesia and Theileria (Lau 2009) are 6.6 kb in size and monomeric linear molecules with terminal inverted repeats (TIRs), indicative of telomeres (Kairo et al. 1994; Brayton et al. 2007). Similar to Plasmodium, mt genomes of Babesia and Theileria contain the three protein-coding genes, but gene order and transcriptional direction are clearly different from Plasmodium (Kairo et al. 1994; Brayton et al. 2007). Thus, the mt genomes of Plasmodium and Babesia/Theileria are structurally highly divergent regardless of their close relatedness (Kuo et al. 2008). For Toxoplasma gondii, the mt genome remains to be isolated and analyzed, although multiple copies of partial mt genes (cox1 and cob) were found to be scattered throughout the nuclear genome (Ossorio et al. 1991). In Cryptosporidium parvum, the mitochondrion is degenerative and lacks any DNA (Abrahamsen et al. 2004). Clearly, the phylum Apicomplexa provides interesting materials for telomere study.

It remains unknown how the remarkable structural divergence between Plasmodium and Babesia/Theileria mentioned above was generated. Gathering enough data set will also help provide further insights on the extent at which the mt genomes have evolved in the different genera as well as in the phylum. In this study, we determined five new mt genome sequences from Babesia and Theileria species. Analyses of the genome structures show that although the mt genome structure is conserved in Babesia species, it varies notably in both size and genome organization in Theileria species, with lineage-specific evolution in two Theileria species: genome inversion in T. orientalis and gene-embedded long TIR in T. equi.

Materials and Methods

Parasite Species

Mitochondrial genome sequences were determined from the following seven parasite species: Babesia bigemina (Kochinda stock) (Fujinaga et al. 1980), B. caballi (USDA strain) (Avarzed et al. 1997), B. gibsoni (National Research Center for Protozoan Diseases strain) (Ishimine et al. 1978), B. bovis (Miyama stock) (Fujinaga et al. 1980), Theileria orientalis (Ikeda stock) (Kim et al. 2004), T. equi (USDA strain) (Avarzed et al. 1998), and T. parva (Muguga stock) (Kairo et al. 1994). Their host animals are cattle for B. bigemina, B. bovis, T. orientalis, and T. parva; horses for B. caballi and T. equi; and dogs for B. gibsoni (supplementary table S2a, Supplementary Material online). The primers were designed by aligning reported mt genome sequences of B. bovis (DDBL/EMBL/GenBank accession number EU075182), T. parva (Z23263), and T. annulata (NW_001091933). Amplification was carried out in a 20 μl reaction mixture containing 0.2 μM each of forward and reverse primers, 400 μM each of deoxynucleotide triphosphate (dNTP), 1 U of LA-Taq (Takara, Shiga, Japan), 2 μl of 10× PCR buffer, 2.5 mM of MgCl2, and 1 μl of genomic DNA. PCR conditions were as follows: initial denaturation at 94 °C for 1 min and amplification for 40 cycles at 94 °C for 30 s, 55–68 °C (depending on primers used) for 30 s, and 72 °C for 1–6 min (depending on amplicon size, 1 min/kb), followed by a final extension at 72 °C for 10 min.

Sequences of the T. equi mt telomeric regions were determined by using the terminal deoxynucleotidyl transferase (TdT) tailing method (Bah et al. 2004) with some modifications. Briefly, the 3’ end was tailed with cytosine by initial denaturation of genomic DNA (150 ng) for 5 min at 95 °C and then incubated for 30 min at 37 °C in a reaction mixture containing 200 μM deoxycytidine triphosphate, 1 U of TdT (Takara), 20 mM Tris–HCl (pH 8.4), 50 mM KCl, and 1.5 mM MgCl2, followed by heat inactivation of TdT at 65 °C for 10 min. The first PCR was done in a 50 μl reaction mixture containing 2 μl of the tailed DNA fragments, 1.25 U of AmpliTaq DNA Polymerase (Applied Biosystems, Foster City, CA), 2.5 mM MgCl2, 200 μM dNTPs, 0.4 μM of an mt genome–specific primer (supplementary table S2b, Supplementary Material online), and a selective anchor primer (5’-CTACTACTACTAGCCACCGTGCTGACTAGTACC GGGGGGGGGGGGG-3’). Initial denaturation was at 95 °C for 2 min, followed by 40 cycles at 94 °C for 30 s and 62 °C for 3 min, and with a final extension step at 72 °C for 10 min. The second PCR was performed using 1 μl of the first PCR product in a 50 μl reaction mixture mentioned above, containing a nested primer (supplementary table S2b, Supplementary Material online) and a universal amplification primer (5’-CTACTACTACTAGCCACCGTGCTGACTAGTAC-3’). The second PCR amplification was at 95 °C for 2 min, and 25 cycles of 94 °C for 30 s, 62 °C for 2 min, followed by an extension step at 72 °C for 10 min. This method would not work for the other Babesia and Theileria samples. It can be surmised that relatively high (A+T) content in TIRs of the other samples may have caused some problems. Multiple palindromes in TIR reported for T. parva (Shukla and Nene 1998) may also be contributing factors in the difficulty to determine telomeric sequences. In T. equi, unlike other Babesia and Theileria species, the TIR has a relatively low (A+T) content with no apparent multiple palindromes and, surprisingly, contains cox3 and two fragments of rRNA gene (see Results).

TIR sequences of other Babesia and Theileria species were determined using an “inverted PCR,” in which primers leading toward telomere ends (supplementary table S2b, Supplementary Material online) were used. We assumed that small inverted sequences, probably present in TIRs as reported for T. parva (Shukla and Nene 1998), could
self-anneal in opposite direction, enabling amplification of two telomeric regions encompassed by two outward primers when Taq polymerase with an exonuclease activity, that could excise unpaired bases (such as LA-Taq), was used. This inverted PCR successfully amplified specific DNA bands for all (Babesia and Theileria) but one species (i.e., T. equi) examined here.

PCR products were purified using QIAquick PCR Purification Kit (QIAGEN). DNA sequencing was performed directly from two independent PCR products using the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) and an ABI 3130 Genetic Analyzer (Applied Biosystems). Sequencing primers were designed to cover target regions in both directions. DDBL/EMBL/GenBank accession numbers of sequences obtained in this study are given in supplementary table S1 (Supplementary Material online). The T. parva sequence obtained here has a 24-bp inconsistency with the reported T. parva sequence (Z23263). The B. bovis sequence in this study has a 7-bp difference from the reported B. bovis sequence (EU075182). These differences may be due to polymorphism because uncloned stock (T. parva) and different parasite strains (B. bovis) were used. We used our sequences of T. parva and B. bovis for analysis.

**Gene Annotation**

Nucleotide sequences of obtained mt genomes from Babesia and Theileria species and their deduced amino acid sequences were aligned together with reported sequences from B. bovis (EU075182), T. parva (Z23263), and T. annulata (NW_001091933) by ClustalW (Thompson et al. 1994). Alignment was manually corrected. Protein-coding genes were predicted using previously annotated sequences from T. parva and B. bovis.

To identify putative rRNA genes, mitochondrial DNA (mtDNA) sequences or annotated rRNA gene fragments from B. bovis (EU075182) and T. parva (Z23263) were used as queries under the suggested algorithm parameters (Freyhult et al. 2007) in NCBI BLAST 2.2 (Altschul et al. 1990). In silico analysis was also done with ProAlign beta version 1.2 (Roshan et al. 2008) and SSEARCH 3.5 (Pearson 1991) using known rRNA gene fragments and suggested advanced search options (Freyhult et al. 2007; Roshan et al. 2008). Newly identified candidate rRNA genes were, likewise, used as input sequences. The information from sequence alignments using ClustalW (Thompson et al. 1994) and putative base-pairings between fragments proposed for T. parva mitochondrial ribosomal RNA fragments (Kairo et al. 1994) were considered in assigning the termini of the candidate genes. Similar searches using some of the rRNA fragment sequences from Plasmodium falciparum (Feagin et al. 1997) detected additional candidate gene regions.

**Southern Blot Hybridization**

Genomic DNA of B. gibsoni, T. orientalis, and T. equi, either undigested or digested with PvuII, HindIII, or XhoI, was electrophoresed on 0.8% agarose gels in Tris–acetate–ethylene-diaminetetraacetic acid (40 mM Tris–acetate and 1 mM ethylenediaminetetraacetic acid) and then transferred to a positively charged nylon membrane (Amersham Hybrid-N+; GE Healthcare, Little Chalfont, England). PCR products amplified specifically from B. gibsoni, T. orientalis, and T. equi genomic DNA (supplementary table S2c, Supplementary Material online) were labeled with digoxigenin-deoxyuridine triphosphate using the DIG High Prime DNA Labeling and Detection Starter Kit II (Roche Diagnostics, Rotkreuz, Switzerland). The digoxigenin-labeled DNA probes were used for overnight hybridization. Blots were washed twice with 2× saline–sodium citrate (SSC) and 0.1% sodium dodecyl sulfate (SDS) and twice with 0.5× SSC and 0.1% SDS at 65 °C for 15 min. Hybridization signals were detected using the Detection Starter Kit II.

**RNA Preparation and Analysis**

Transcription of Cox1, Cox3, and Cob in B. gibsoni, T. orientalis, and T. equi was analyzed by reverse transcriptase-PCR (RT-PCR). Total RNA was extracted with RNeasy Mini Kit (QIAGEN). DNase I treatment was done to remove any residual DNA before cDNA synthesis. Using specific primers (supplementary table S2d, Supplementary Material online), cDNA synthesis and DNA amplification were carried out using PrimeScript High Fidelity RT-PCR Kit (Takara). RNA extracts that were not treated with reverse transcriptase gave no PCR products.

For northern blot analysis, total RNA including short RNAs from B. gibsoni was prepared with mirVana miRNA Isolation Kit (Ambion, Austin, TX). Total RNA (10 μg) was subjected to 8.3 M urea–12% (w/v) polyacrylamide gel electrophoresis. After electrophoresis, the gel was stained with ethidium bromide and photographed. RNA was electrophoretically blotted onto Biodyne Plus (Pall, Glen Cove, NY) using a semi-dry blotter NA-1515B (Nihon Eido, Tokyo, Japan) according to the manufacturer’s protocol. The blotted membrane was ultraviolet treated for cross-linking (Brown et al. 2004) and incubated in hybridization solution (200 mM sodium phosphate [pH 7.2]–7% (w/v) SDS) for 30 min at 37 °C. The oligo probe was 5′ labeled with T4 polynucleotide kinase and [γ-32P] adenosine triphosphate according to the enzyme supplier’s instruction (Takara) and purified based on Brown et al. (2004). After overnight hybridization at 37 °C, the membrane was then washed twice with 2× SSC-0.5% (w/v) SDS at 37 °C, twice at 47 °C, and finally twice with 0.2× SSC-0.5% (w/v) SDS at 47 °C. The membrane was exposed to an Imaging Plate (Fujifilm, Tokyo, Japan), and the plate was scanned with a BAS2500 Bioimaging Analyzer (Fujifilm).

**Phylogenetic Analysis**

The concatenated amino acid sequences of Cox1 and Cob were used for phylogenetic analysis. (Sequences of Cox3 were not used due to very high divergence in Babesia/Theileria species [see Results].) The data set of 834 amino acid positions, comprising 474 COX1 and 360 COB amino acids, was analyzed using the PROML program in PHYLIP version
3.68 (Felsenstein and Churchill 1996) under the Jones, Taylor, and Thornton model (Jones et al. 1992) with the amino acid frequencies of the data set used to infer the maximum likelihood (ML) tree. Corresponding sequence of *P. falciparum* was used as an outgroup. To take the evolutionary rate heterogeneity, the R option was set to utilize discrete *I* distribution with eight categories for approximating the site rate distribution. CODEML program in PAML version 4.2 (Yang 2007) was used to estimate the *I* shape parameter value $a$. Bootstrap analysis was done by applying PROML to 100 resampled data sets produced by SEQBOOT program in PHYLIP. Bootstrap proportion (BP) values were calculated for internal branches of the inferred ML tree using CONSENSE in PHYLIP.

LSU sequences (592 sites in total: 265 bp for LSU1; 35 bp for LSU2; 111 bp for LSU3; 82 bp for LSU4; 64 bp for LSU5; and 35 bp for LSU6) were analyzed using the ML method performed with PAUP* 4.0 b10 (Swofford 2002). The appropriate nucleotide substitution model was first determined using the Modeltest (version 3.7) estimations, including both the proportion of invariable sites and the *I* shape parameter (Posada and Crandall 1998). For branch support of the ML tree, bootstrap probability was estimated from 1,000 heuristic replicates with single random addition replicates. All trees were reconstructed with TreeView 1.6.6 (Page 1996). For statistical comparisons among the ML best tree and its alternatives, $P$ values of Kishino–Hasegawa (KH) test (Kishino and Hasegawa 1989) and Shimodaira–Hasegawa (SH) test (Shimodaira and Hasegawa 1999) were obtained.

**Results**

**Mitochondrial Genome Organization**

We obtained 5.8- to 5.9-kb sequences from each of *B. bigemina, B. caballi, B. gibsoni,* and *B. bovis,* in which three protein-coding genes, *cox1, cob,* and *cox3,* and five fragments of the LSU rRNA gene were identified (fig. 1a). TIR sequences were also found on both ends of the predicted linear mt genomes. Although full-length sequences of TIRs were not successfully determined, the size of the TIR was inferred to be 440–450 bp from results of Southern blot hybridization analysis (see below). Additionally, we identified one LSU fragment of rRNA gene, LSU6, that showed a high sequence similarity to the 3′ part of RNA10 of the *P. falciparum* rRNA gene fragment (Feagin et al. 1997) (underlined sequence in the following *B. gibsoni* LSU sequence are identical nucleotides to *P. falciparum* Figure 1. Structure of the mitochondrial genomes of *Babesia gibsoni* (a), *Theileria orientalis* (b), and *T. equi* (c). Mitochondrial (mt) genome structure was completely conserved among *B. gibsoni, B. bigemina* (not shown), *B. caballi* (not shown), *B. bovis* (not shown), and *T. parva* (d). The *T. orientalis* mt genome has an inversion in the 3-kb central region. The *T. equi* mt genome has a relatively long TIR and contains a *cox3* gene and rRNA gene fragments. Shown for comparison is the mt genome of *Plasmodium falciparum* (e) (Feagin et al. 1997). Genes shown above the bold line in each genome are transcribed left to right and those below are transcribed from right to left. Two small gray lines in the *T. orientalis* genome indicate an inversion. Vertical broken lines indicate the boundaries of the 3-kb inversion. Dark and light gray boxes indicate fragments of LSU and SSU rRNA genes, respectively. *cox1,* cytochrome c oxidase subunit I; *cox3,* cytochrome c oxidase subunit III; *cob,* cytochrome *b*. **FIG. 1.** Structure of the mitochondrial genomes of *Babesia gibsoni* (a), *Theileria orientalis* (b), and *T. equi* (c). Mitochondrial (mt) genome structure was completely conserved among *B. gibsoni, B. bigemina* (not shown), *B. caballi* (not shown), *B. bovis* (not shown), and *T. parva* (d). The *T. orientalis* mt genome has an inversion in the 3-kb central region. The *T. equi* mt genome has a relatively long TIR and contains a *cox3* gene and rRNA gene fragments. Shown for comparison is the mt genome of *Plasmodium falciparum* (e) (Feagin et al. 1997). Genes shown above the bold line in each genome are transcribed left to right and those below are transcribed from right to left. Two small gray lines in the *T. orientalis* genome indicate an inversion. Vertical broken lines indicate the boundaries of the 3-kb inversion. Dark and light gray boxes indicate fragments of LSU and SSU rRNA genes, respectively. *cox1,* cytochrome c oxidase subunit I; *cox3,* cytochrome c oxidase subunit III; *cob,* cytochrome *b*. **FIG. 1.** Structure of the mitochondrial genomes of *Babesia gibsoni* (a), *Theileria orientalis* (b), and *T. equi* (c). Mitochondrial (mt) genome structure was completely conserved among *B. gibsoni, B. bigemina* (not shown), *B. caballi* (not shown), *B. bovis* (not shown), and *T. parva* (d). The *T. orientalis* mt genome has an inversion in the 3-kb central region. The *T. equi* mt genome has a relatively long TIR and contains a *cox3* gene and rRNA gene fragments. Shown for comparison is the mt genome of *Plasmodium falciparum* (e) (Feagin et al. 1997). Genes shown above the bold line in each genome are transcribed left to right and those below are transcribed from right to left. Two small gray lines in the *T. orientalis* genome indicate an inversion. Vertical broken lines indicate the boundaries of the 3-kb inversion. Dark and light gray boxes indicate fragments of LSU and SSU rRNA genes, respectively. *cox1,* cytochrome c oxidase subunit I; *cox3,* cytochrome c oxidase subunit III; *cob,* cytochrome *b*.
RNA10: 5′-ATGCCGGAGTACGTAAGGAATAGGAAA-GATTACCGCTATCA-3′). The organization and predicted transcriptional direction of the three protein-coding genes and the six LSU rRNA gene fragments were completely conserved among *B. bigemina*, *B. caballi*, *B. gibsoni*, *B. bovis*, *T. parva*, and *T. annulata* (fig. 1a). Southern blots probed with a 4.9-kb portion (P1) of the *B. gibsoni* mt genome produced a clear signal at 6.6 kb against *B. gibsoni* genomic DNA and two bands at 3.7 and 2.9 kb against DNA digested with *Pvu* II (fig. 2a and e). The sizes of the two bands were identical to those predicted from the sequence. These results confirm the monomeric linear 6.6 kb of *B. gibsoni* mt genome similar to that reported for *T. parva* and *T. annulata* (Hall et al. 1990; Kairo et al. 1994).

Interestingly, the *T. orientalis* mt genome, aside from having three protein-coding genes, six rRNA gene fragments, and TIRs similar to four Babesia species, *T. parva*, and *T. annulata* (fig. 1b) showed an inversion at the 3.0-kb central region containing *cox3*, LSU1, LSU3, LSU6, and LSU2. No sequences that potentially form secondary structures such as a hairpin structure were apparent near the boundaries of this inverted sequence, but instead, unique insertions of 21–22 and 84–102 bp were noted (fig. 1b), making the *T. orientalis* mt genome slightly longer (112–168 bp) than the other six Babesia and Theileria mt genomes (supplementary table S3, Supplementary Material online). As predicted from the sequence, hybridization using a *T. orientalis* probe (P2) that spans the central 5.0-kb region yielded a band at 6.7 kb against undigested DNA and two bands at 3.8 and 2.9 kb against *Hind*III-digested DNA (fig. 2b and f). From these results, the *T. orientalis* mt genome shows a 6.7-kb monomeric linear structure.

Strikingly distinctive from all other mt genomes described above is *T. equi*. First, the size of the *T. equi* genome is 8.2 kb, that is, 1.6–1.7 kb longer than that of other species. Second, TIR sequences are large (1,563 bp), compared with the 440–450 bp TIRs of the other seven species, and, interestingly, contained *cox3* and two LSU rRNA gene fragments, LSU4 and LSU5. TIR sequences on both ends showed complete identity. Third, the protein-coding genes and rRNA gene fragments showed little synteny to other species (fig. 1c). Hybridization with a probe corresponding to a 1.0-kb region in *cox1* (P3) produced a band at 8.2 kb against undigested DNA, a band at 5.6 kb against *Hind*III-digested DNA, and a band at 6.4 kb against *Xho*I-digested DNA (fig. 2c and g). Another *T. equi* probe using a 1-kb region in the TIR (P4) produced a band at 8.2 kb against undigested DNA, two bands at 5.6 and 2.7 kb against *Hind*III-digested DNA, and a band at 0.9 kb against

**Fig. 2.** Monomeric linear structure of the mitochondrial genomes of *Babesia gibsoni*, *Theileria orientalis*, and *T. equi*. Genomic DNA of *B. gibsoni* (a), *T. orientalis* (b), and *T. equi* (c and d) was hybridized with a *B. gibsoni* probe P1 (e), a *T. orientalis* probe P2 (f), and *T. equi* probes P3 and P4 (g), respectively. Undigested DNA (lanes 1, 3, 5, and 8) or DNA digested with *Pvu* II (lane 2), *Hind*III (lanes 4, 6, and 9), or *Xho*I (lanes 7 and 10) was fractionated on 0.8% agarose gels.
$\text{XhoI}$-digested DNA ($\text{fig. 2d}$ and g). The sizes of these bands match the predicted $T. \text{equi}$ sequence, indicating an 8.2-kb monomeric linear structure of the $T. \text{equi}$ mt genome.

**Transcription**

RT-PCR using three separate primer sets targeting about 500-bp sequences of $\text{cox1}$, $\text{cox3}$, and $\text{cob}$ of $B. \text{gibsoni}$ gave the expected transcript size using cDNA but not using RNA ($\text{fig. 3}$). Similarly, expected PCR sized fragments were obtained using primers specific to $T. \text{orientalis}$ or $T. \text{equi}$ for $\text{cox1}$, $\text{cox3}$, and $\text{cob}$. Results confirm the transcription of the three protein-coding genes, including $\text{cox3}$ in the 3-kb inverted region of $T. \text{orientalis}$ and $\text{cox3}$ in TIR of $T. \text{equi}$.

Transcription of LSU6 was confirmed by northern blot analysis. Probing with two oligonucleotides (18-mer and 23-mer at the 5′ and 3′ end, respectively) complementary to LSU6 against $B. \text{gibsoni}$ total RNA produced approximately 90-nt RNA signal ($\text{fig. 4}$), suggesting that the LSU6 was actually transcribed and the corresponding stable transcript existed as a short RNA fragment. Due to extreme difficulties in obtaining an adequate amount of parasites from infected hosts and/or in vitro culture limitations for $B. \text{bigemina}$, $B. \text{caballi}$, $T. \text{orientalis}$, and $T. \text{equi}$, we were unable to perform additional northern blot analyses. Together with the report of Kairo et al. (1994) on the transcription of the three protein-coding genes and five known LSU rRNA gene fragments (LSU1–LSU5) in $T. \text{parva}$, however, these results suggest that both protein genes and rRNA gene fragments are transcribed in Babesia and Theileria.

**Sequence Similarity and Phylogeny**

$\text{cox1}$ and $\text{cob}$ pairwise sequence similarity scores among $B. \text{bovis}$, $B. \text{bigemina}$, $B. \text{caballi}$, $B. \text{gibsoni}$, $T. \text{parva}$, $T. \text{orientalis}$, $T. \text{annulata}$, and $T. \text{equi}$ were comparable: 67–88% for $\text{cox1}$ and 60–85% for $\text{cob}$ at the amino acid sequence level (supplementary table S4, Supplementary Material online). Conserved sequence regions in COX1 and COB correspond to the 12 and 9 transmembrane domains as inferred from bovine COX1 (SWISS-PROT protein database accession number P00396) and COB (P00157), respectively, in which multiple heme-binding histidine residues that form catalytic sites are perfectly conserved (Widger et al. 1984; Yun et al. 1991; Castresana et al. 1994; Ferguson-Miller and Babcock 1996). For $\text{cox3}$, pairwise sequence similarity was relatively low, and $T. \text{equi}$ $\text{cox3}$ in particular has considerably low similarity to $\text{cox3}$ of other species (37–41% compared with the 57–80% similarity of other $\text{cox3}$ in Babesia and Theileria species). Nevertheless, the predicted $T. \text{equi}$ COX3 amino acid sequence shows seven transmembrane domains (from I to VII) and the C-terminal hydrophobic domain VII (P00415) that are highly conserved among a wide variety of organisms from prokaryotes to plants (Halitza et al. 1991). In contrast to these protein-coding genes, pairwise sequence similarity of six LSUs was very high, 75–96% among all the Babesia and Theileria species examined here (supplementary table S5, Supplementary Material online).

The ML tree was inferred from concatenated COX1 and COB amino acid sequences using $P. \text{falciparum}$ as an outgroup ($\text{fig. 5}$). Monophyletic relationships were observed with high BP values (98%) for 1) $B. \text{bovis}$, $B. \text{bigemina}$, $B. \text{caballi}$, and $B. \text{gibsoni}$ and 2) $T. \text{annulata}$, $T. \text{parva}$, and $T. \text{orientalis}$. Theileria $\text{equi}$ was located at the branch.
leading to the common ancestor of Babesia species with moderate BP value (65%). The other possible branching positions of *T. equi* at the ancestral branch of Theileria (arrow A in fig. 5) or at the ancestral branch of Theileria and Babesia (arrow B in fig. 5) were, however, not rejected by either the KH or the SH tests (supplementary table S6, Supplementary Material online), thus indicating that the evolutionary position of *T. equi* is yet unclear with the present data set.

The ML tree constructed using LSU sequences was basically the same as that of *cox1* and *cob* (supplementary fig. S1, Supplementary Material online), although, in this case, unrooted trees were compared because an appropriate outgroup species is not available for LSU. Monophyletic relationships of the four Babesia species and of *T. annulata*, *T. parva*, and *T. orientalis* were supported with high BP values. Within the clade of Babesia, however, the relationship between *B. bovis*, *B. bigemina*, and *B. caballi* was not supported with high BP values and not highly consistent to that in the *cox1/cob* tree. Notably, when the data set of LSU was applied to the two topologies, one for the *cox1/cob* tree and the other for the LSU tree, the two trees were not significantly different (data not shown), suggesting that the relationship of the three Babesia species was not clearly separable using the LSU sequences.

Partial TIR sequences (47–1,563 bp) (supplementary table S3, Supplementary Material online) were highly divergent among all eight Babesia and Theileria species examined here. In noncoding regions excluding TIR, we identified 14 sequence regions that were highly similar among all the Babesia and Theileria species (fig. 6 and supplementary tables S7 and S8, Supplementary Material online). Pairwise sequence similarity of these 14 regions ranged from 81% to 97%, values comparable with LSUs. The array and direction of the 14 regions were well conserved among all species except *T. equi*, in which case the 14 regions were extensively

![Fig. 5.](image-url) The ML phylogenetic tree of the two mitochondrial protein-coding genes, *cox1* and *cob*, from eight Babesia and Theileria species. *Plasmodium falciparum* was used as an outgroup. Concatenated amino acid sequences (834 sites) were used with 1,000 heuristic replicates under a Jones, Taylor, and Thornton model ($x = 0.72$). The numbers shown along nodes represent bootstrap values. Arrows, A and B, indicate alternative positions of the *Theileria equi* sequence, whose possibilities were statistically compared by the SH and KH tests. Five parasite sequences obtained in this study are underlined.
rearranged (fig. 6). Within species, these 14 regions did not show sequence similarity to each other.

Discussion

This study presents for the first time evidence of the highly divergent mt genomes in the genus Theileria and a high conservation in the genus Babesia. *Theileria annulata* and *T. parva* have an mt genome structure nearly identical to that of Babesia, whereas the mt genomes of *T. orientalis* and *T. equi* were clearly distinctive: a large genomic region is inverted in *T. orientalis*, and in *T. equi*, there occurs an unusually long TIR, containing *cox3* and two LSU rRNA gene fragments. The phylogenetic tree of *cox1/cob* showed two separate clades: one for three Theileria species (*T. parva, T. annulata*, and *T. orientalis*) and another for four Babesia species. The mt genome structures of these Theileria species are highly conserved. This suggests that an inversion event occurred specifically in the *T. orientalis* lineage after divergence from a common ancestor of *T. annulata, T. parva*, and *T. orientalis*. The phylogenetic position of *T. equi* was not clearly determined in the *cox1/cob* tree. This unclear positioning of *T. equi* is consistent with the tree based on 18S rRNA gene, a nuclear genome gene (Criado-Fornelio et al. 2003). Thus, it remains to be solved whether a common ancestor of *T. equi*, the Theileria clades, and Babesia clades possessed an mt genome with gene-embedded long TIR. Phylogenetic separation of *T. equi* from Theileria clade as well as the identical life cycle, namely, the presence of schizogony in lymphocytes and lack of transovarial transmission by vector (Uilenberg 2006), suggests that, most likely, the mt genome of the common ancestor contained a long TIR.

Compared with moderate sequence similarity in *cox1* and *cob* of Theileria and Babesia species, *cox3* is highly divergent. Similarly, *cox3* is more divergent than *cox1* and *cob* in Plasmodium species (Wilson and Williamson 1997; Perkins 2008). COB is a subunit of complex III, essential for electron transfer in the mitochondrial respiratory chain (Xia et al. 1997). COX1 (subunit I) (and also COX2 [subunit II]), which contains heme and copper and is essential for electron transfer, is 1 of 13 subunits of cytochrome c oxidase (complex IV), a terminal oxidase in the respiratory chain (Castresana et al. 1994). COX3, which is not directly involved in the electron transport, is considered to stabilize the complex of COX1 and COX2 (Haltia et al. 1991; Tsukihara et al. 1996). Such an accessory role of COX3 in the function of cytochrome c oxidase may relax constraints, causing acceleration of evolutionary rate and low sequence similarity.

In addition to the five previously known LSU rRNA gene fragments (LSU1–LSU5) (Kairo et al. 1994; Brayton et al. 2007), we newly identified an LSU fragment, LSU6, in all Babesia and Theileria species, whose transcription was confirmed by northern blot analysis. Kairo et al. (1994) have previously mapped LSU1–LSU5 (by comparative secondary structure modeling) to the 3′ half of *Escherichia coli* 23S (LSU) rRNA. LSU5 and 5′ region of LSU1 form the domain IV of LSU, whereas 3′ region of LSU1, LSU2, LSU3, and LSU4 forms the domain V of the LSU, which contains the peptidyl transferase center. We mapped the newly identified LSU6 to positions 2640–2670, which forms a part of the domain VI (alpha-sarcin/ricin stem loop) of the LSU. Contrary to the other LSU fragments, the borders of LSU6 were unclear because the flanking sequences have no recognizable complementarities to other fragments. In addition to LSU6, we found a small sequence region in all Babesia and Theileria mtDNA, which showed a high sequence similarity to a small subunit fragment (SSUF) of the *P. falciparum* rRNA gene. We were not, however, able to confirm transcription of the SSUF (data not shown), although the region is regarded as one of the highly similar noncoding regions (supplementary table S7, Supplementary Material online). The mtDNA of Babesia/Theileria still contains large unannotated sequence regions (about one-third sequence regions of the genome), and those sequences are highly conserved among parasite species of the two genera (supplementary table S8, Supplementary Material online). It is thus likely that unidentified rRNA gene fragments as well as other genes exist in these regions. Comprehensive and detailed mtDNA transcript analysis of Babesia/ Theileria would be required for further identification of genes and gene fragments as was done for *P. falciparum* (Feagin et al. 1997).
The observed high divergence of TIR sequences among Babesia and Theileria species was not surprising because no universal signature has been reported for TIR sequences of linear mt genomes from several unicellular flagellates: *Polytomella capuana, Po. parva,* and *Chlamydomonas reinhardtii* in the Reinhardtii clade of chlorophycean green algae (Smith and Lee 2008). Inverted complementary sequences at both ends of a linear mt genome are believed to play important roles in replication and stabilization (Nosek and Tomaska 2003). In TIR, small repeats with forward and reverse directions frequently occur, which are likely to form secondary structure. In general, sequence regions containing small repeats diverge rapidly compared with nonrepeat sequences. As a consequence, little sequence conservation is expected in TIR of a linear mt genome. An exception is *T. equi* TIR, in which sequence regions of *cox3* and two LSU rRNA gene fragments are relatively conserved, perhaps due to functional constraints.

In conclusion, our studies show the remarkable structural diversity in the mt genomes of Theileria species, in contrast to the highly conserved genome among Babesia species. The finding of a high structural divergence of the *T. equi* mt genome from other Babesia and Theileria species suggests the occurrence of lineage-specific evolution of mt genomes within the closely related Babesia and Theileria genera. Further investigations into mt genomes of other Theileria species and other related genera, such as those that belong to the Archaeopiroplasmsids group, should provide insights into the evolutionary origin of a major structural divergence between Plasmodium (circular/concatemer genome) and Babesia/Theileria (linear genome) in the same phylum Apicomplexa.

**Supplementary Material**

Supplementary material, tables S1–S8, and figure S1 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

**Acknowledgments**

We wish to thank T. Toyama for technical advice and H. Inokuma for kindly providing parasite samples. This work was supported by Grant-in-Aids for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology of Japan (18073013) and from Japan Society for Promotion of Sciences (18GS03140013 and 20390120).

**References**


Yun CH, Crofts AR, Gennis RB. 1991. Assignment of the histidine axial ligands to the cytochrome b6 and cytochrome b5 components of the bc1 complex from *Rhodobacter sphaeroides* by site-directed mutagenesis. *Biochemistry* 30:6747–6754.