between fragments proposed for *T. parva* mt rRNA 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 *Pvu*II, *Hind*III or *Xho*I, was electrophoresed on 0.8% agarose gels in TAE (40 mM Tris-acetate, 1 mM EDTA) and then transferred to a positively charged nylon membrane (Amersham Hybond-N+, GE Healthcare, Little Chalfont, England). PCR products amplified specifically from *B. gibsoni*, *T. orientalis* and *T. equi* genomic DNA (supplementary table 2c, Supplementary Material online) were labeled with digoxigenin-dUTP using the DIG High Prime DNA Labeling and Detection Starter Kit II (Roche Diagnostics, Rotkreuz, Switzerland). The DIG-labeled DNA probes were used for overnight hybridization. Blots were washed twice with 2 x SSC, 0.1% SDS and twice with 0.5 × SSC, 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 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 2d, 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 electroblotted on 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 UV-treated for cross-linking (Brown, Mackey, and Du 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-^{32}P] ATP 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 2X SSC-0.5% (w/v) SDS at 37 °C, twice at 47 °C, and finally twice with 0.2X SSC-0.5% (w/v) SDS at 47 °C. The membrane was exposed to an Imaging Plate (Fujifilm, Tokyo, Japan); and the Plate 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 *Theileria/Babesia* species (see Results).) The dataset 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 JTT model (Jones, Taylor, and Thornton 1992) with the amino acid frequencies of the dataset used to infer the maximum likelihood (ML) tree. Corresponding sequence of *Plasmodium falciparum* was used as an outgroup. To take the evolutionary rate heterogeneity, the R option was set to utilize discrete Γ distribution with eight categories for approximating the site-rate distribution. CODEML program in PAML version 4.2 (Yang 2007) was used to estimate the Γ shape parameter value α. Bootstrap analysis was done by applying PROML to 100 re-sampled datasets 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; 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 (ver 3.7) estimations, including both the proportion of invariable sites and the Γ 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 Simodaira-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. cabali*, *B. gibsoni* and *B. bovis*, in which three protein coding genes, *cox1*, *cox3* and *cob*, and five fragments of the large subunit (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, which 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* RNA10: 5′-ATAGCCGGAGTACGTAAGGAATAGGAAAGATTAACCGCTATCA-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 kb and 2.9 kb against DNA digested with *Pvu*II (fig. 2a and 2e). 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 bp 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 3, Supplementary Material online). As predicted from the sequence, southern 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 kb and 2.9 kb against *Hind*III-digested DNA (fig. 2b and 2f). 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, i.e., 1.6-kb to 1.7-kb longer than that of other species. Second, TIR sequences are large (1,563 bp), as compared to 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 2g). 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 kb and 2.7 kb against *Hind*III-digested DNA, and a band at 0.9 kb against *Xho*I-digested DNA (fig. 2d and 2g). The sizes of these bands match the predicted *T. equi* sequence, indicating an 8.2-kb monomeric linear structure of the *T. equi* mt genome.

Transcription

RT-PCR using three separate primer sets targeting about 500 bp sequences of *cox1*, *cox3* and *cob* of *B. gibsoni*, gave the expected transcript size using cDNA but not RNA (fig. 3). Similarly, expected PCR sized fragments were obtained using primers specific to *T. orientalis* or *T. equi* for *cox1*, *cox3* and *cob*. Results confirm the transcription of the three protein coding genes, including *cox3* in the 3-kb inverted region of *T. orientalis* and *cox3* in TIR of *T. 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. gibsoni* total RNA produced approximately 90-nt RNA signal (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. bigemina*, *B. caballi*, *T. orientalis* and *T. equi*, we were unable to perform additional Northern blot analysis. 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 to LSU5) in *T. parva*, however, these results suggest that both protein genes and rRNA gene fragments are transcribed in *Babesia* and *Theileria*.

Sequence similarity and phylogeny

cox1 and *cob* pair-wise sequence similarity scores among *B. bovis*, *B. bigemina*, *B. caballi*, *B. gibsoni*, *T. parva*, *T. orientalis*, *T. annulata*, and *T. equi* were comparable: 67% to 88% for *cox1* and 60% to 85% for *cob* at the amino acid sequence level (supplementary table 4, 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 data base 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, Crofts, and Gennis 1991; Castresana et al. 1994; Ferguson-Miller and Babcock 1996). For *cox3*, pair-wise sequence similarity was relatively low, and *T. equi cox3* in particular has considerably low similarity to *cox3* of other species (37% to 41% compared to the 57% to 80% similarity of other *cox3* in *Babesia* and *Theileria* species). Nevertheless, the predicted *T. 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 (Haltia, Saraste, and Wikstrom 1991). In contrast to these protein coding genes, pair-wise sequence similarity of six LSUs was very high, 75% to 96% among all the *Babesia* and *Theileria* species examined here (supplementary table 5, Supplementary Material online).

The ML-tree was inferred from concatenated COX1 and COB amino acid sequences using *P. falciparum* as an outgroup (fig. 5). Monophyletic relationships were observed with high BP values (98%) for (i) *B. bovis*, *B. bigemina*, *B. caballi* and *B. gibsoni*; and (ii) *T. annulata*, *T. parva* and *T. orientalis*. *T. 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 and SH tests (supplementary table 6, Supplementary Material online), thus indicating that

the evolutionary position of *T. equi* is yet unclear with the present dataset.

The ML tree constructed using LSU sequences was basically the same as that of *cox1* and *cob* (supplementary fig. 1, 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. cabali* was not supported with high BP values, and not highly consistent to that in the *cox1/cob* tree. Notably, when the dataset 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 bp to 1563 bp) (supplementary table 3, Supplementary Material online) were highly divergent among all eight *Babesia* and *Theileria* species examined here. In non-coding regions excluding TIR, we identified 14 sequence regions that were highly homologous among all the *Babesia* and *Theileria* species (fig. 6 and supplementary tables 7 and 8, Supplementary Material online). Pair-wise sequence similarity of these 14 regions ranged from 81% to 97%, values comparable to LSUs. The array and direction of the 14 regions were well conserved amongst (including the inverted region of *T. orientalis*), but one, species. *T. equi* was the exception, in which case the 14 regions were extensively 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*. *T. annulata* and *T. parva* have a

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 a 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 (Perkins 2008; Wilson and Williamson 1997). 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 one 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, Saraste, and Wikstrom 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 – 5) (Kairo et al. 1994; Brayton et al. 2007), we newly identified a 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 to LSU5 (by comparative secondary structure modelling) to the 3′ half of *Escherichia coli* 23S (LSU) rRNA. LSU5 and 5′ region of LSU1 form the domain IV of LSU, while 3′ region of LSU1, LSU2, LSU3 and LSU4 form the domain V of the LSU which contains the peptidyltransferase 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 SSU fragment (SSUF) of the *P. falciparum* rRNA gene. We were not, however, able to confirm transcription of the SSU fragment (data not shown); although the region is regarded as one of the highly homologous non-coding regions (supplementary table 7, Supplementary Material online). The mtDNA of *Babesia/Theileria* still contains large unannotated sequence regions (about 1/3 sequence regions of the genome) and those sequences are highly conserved among parasite species of the two genera (supplementary Table 8, 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 have been reported for TIR sequences of linear mt genomes from several unicellular flagellates: *Polytomella capuana, P. 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 non-repeat 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 Archaeopiroplasmids 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 tables 1 to 8 and supplementary FIG. 1 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, 20390120).

Literature Cited

Abrahamsen MS, Templeton TJ, Enomoto S, et al. (20 co-authors). 2004. Complete genome sequence of the apicomplexan, *Cryptosporidium parvum*. Science. 304:441-445. *(Please note that 20 includes all authors of the article.)*

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol. 215:403-410.

Avarzed A, Igarashi I, De Waal DT, et al. (12 co-authors). 1998. Monoclonal antibody against *Babesia equi*: characterization and potential application of antigen for serodiagnosis. J Clin Microbiol. 36:1835-1839. *(Please note that 12 includes all authors of the article.)*

Avarzed A, Igarashi I, Kanemaru T, Hirumi K, Omata Y, Saito A, Oyamada T, Nagasawa H, Toyoda Y, Suzuki N. 1997. Improved in vitro cultivation of *Babesia caballi*. J Vet Med Sci. 59:479-481.

Bah A, Bachand F, Clair E, Autexier C, Wellinger RJ. 2004. Humanized telomeres and an attempt to express a functional human telomerase in yeast. Nucleic Acids Res. 32:1917-1927.

Brayton KA, Lau AO, Herndon DR, et al. (28 co-authors). 2007. Genome sequence of *Babesia bovis* and comparative analysis of apicomplexan hemoprotozoa. PLoS Pathog. 3:1401-1413. *(Please note that 28 includes all authors of the article.)*

Brown T, Mackey K, Du T. 2004. Analysis of RNA by northern and slot blot hybridization. Curr Protoc Mol Biol Chapter 4:Unit 4.9.

Castresana J, Lubben M, Saraste M, Higgins DG. 1994. Evolution of cytochrome oxidase, an enzyme older than atmospheric oxygen. Embo J 13:2516-2525.

Criado-Fornelio A, Martinez-Marcos A, Buling-Sarana A, Barba-Carretero JC. 2003. Molecular studies on *Babesia*, *Theileria* and *Hepatozoon* in southern Europe. Part II. Phylogenetic analysis and evolutionary history. Vet Parasitol 114:173-194.

Feagin JE, Mericle BL, Werner E, Morris M. 1997. Identification of additional rRNA fragments encoded by the *Plasmodium falciparum* 6 kb element. Nucleic Acids Res. 25:438-446.

Felsenstein J, Churchill GA. 1996. A Hidden Markov Model approach to variation among sites in rate of evolution. Mol Biol Evol. 13:93-104.

Ferguson-Miller S, Babcock GT. 1996. Heme/Copper Terminal Oxidases. Chem Rev 96:2889-2908.

Freyhult EK, Bollback JP, Gardner PP. 2007. Exploring genomic dark matter: a critical assessment of the performance of homology search methods on noncoding RNA. Genome Res. 17:117-125.

Fujinaga T, Minami T, Ishihara T. 1980. Serological relationship between a large *Babesia* found in Japanese cattle and *Babesia major*, *B. bigemina* and *B. bovis*. Res Vet Sci. 29:230-234.

Gray MW, Lang BF, Burger G. 2004. Mitochondria of protists. Annu Rev Genet. 38:477-524.

Hall R, Coggins L, McKellar S, Shiels B, Tait A. 1990. Characterisation of an extrachromosomal DNA element from *Theileria annulata*. Mol Biochem Parasitol. 38:253-260.

Haltia T, Saraste M, Wikstrom M. 1991. Subunit III of cytochrome *c* oxidase is not involved in proton translocation: a site-directed mutagenesis study. Embo J 10:2015-2021.

Ishimine T, Makimura S, Kitazawa S, Tamura S, Suzuki N. 1978. Pathophysiological findings on blood of Beagles experimentally infected with *B. gibsoni*. Jpn J Trop Med Hyg. 6:15-26.

Jones DT, Taylor WR, Thornton JM. 1992. The rapid generation of mutation data matrices from protein sequences. Comput Appl Biosci. 8:275-282.

Kairo A, Fairlamb AH, Gobright E, Nene V. 1994. A 7.1 kb linear DNA molecule of *Theileria parva* has scrambled rDNA sequences and open reading frames for mitochondrially

encoded proteins. Embo J. 13:898-905.

Kim JY, Yokoyama N, Kumar S, Inoue N, Inaba M, Fujisaki K, Sugimoto C. 2004. Identification of a piroplasm protein of *Theileria orientalis* that binds to bovine erythrocyte band 3. Mol Biochem Parasitol. 137:193-200.

Kishino H, Hasegawa M. 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in hominoidea. J Mol Evol 29:170-179.

Kuo CH, Wares JP, Kissinger JC. 2008. The apicomplexan whole-genome phylogeny: an analysis of incongruence among gene trees. Mol Biol Evol. 25:2689-2698.

Lau AO. 2009. An overview of the *Babesia*, *Plasmodium* and *Theileria* genomes: a comparative perspective. Mol Biochem Parasitol. 164:1-8.

Levine ND. 1988. Progress in taxonomy of the Apicomplexan protozoa. J Protozool. 35:518-520.

Nosek J, Tomaska L. 2003. Mitochondrial genome diversity: evolution of the molecular architecture and replication strategy. Curr Genet. 44:73-84.

Ossorio PN, Sibley LD, Boothroyd JC. 1991. Mitochondrial-like DNA sequences flanked by direct and inverted repeats in the nuclear genome of *Toxoplasma gondii*. J Mol Biol. 222:525-536.

Page RD. 1996. TreeView: an application to display phylogenetic trees on personal computers. Comput Appl Biosci. 12:357-358.

Palmer JD, Soltis D, Soltis P. 1992. Large size and complex structure of mitochondrial DNA in two nonflowering land plants. Curr Genet. 21:125-129.

Pearson WR. 1991. Searching protein sequence libraries: comparison of the sensitivity and selectivity of the Smith-Waterman and FASTA algorithms. Genomics. 11:635-650.

Perkins SL. 2008. Molecular systematics of the three mitochondrial protein-coding genes of malaria parasites: corroborative and new evidence for the origins of human malaria.

Mitochondrial DNA 19:471-478.

Posada D, Crandall KA. 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics. 14:817-818.

Preiser PR, Wilson RJ, Moore PW, McCready S, Hajibagheri MA, Blight KJ, Strath M, Williamson DH. 1996. Recombination associated with replication of malarial mitochondrial DNA. Embo J. 15:684-693.

Roshan U, Chikkagoudar S, Livesay DR. 2008. Searching for evolutionary distant RNA homologs within genomic sequences using partition function posterior probabilities. BMC Bioinformatics. 9:61.

Shimodaira H, Hasegawa M. 1999. Multiple Comparisons of Log-Likelihoods with Applications to Phylogenetic Inference. Mol Biol Evol 16:1114-1116.

Shukla GC, Nene V. 1998. Telomeric features of *Theileria parva* mitochondrial DNA derived from cycle sequence data of total genomic DNA. Mol Biochem Parasitol. **95**:159-163.

Smith DR, Lee RW. 2008. Mitochondrial genome of the colorless green alga *Polytomella capuana*: a linear molecule with an unprecedented GC content. Mol Biol Evol. 25:487-496.

Swofford DL. 2002. PAUP*. Phylogenetic analysis using parsimony (and other methods). Sinauer Associates, Sunderland, MA.

Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22:4673-4680.

Tsukihara T, Aoyama H, Yamashita E, Tomizaki T, Yamaguchi H, Shinzawa-Itoh K, Nakashima R, Yaono R, Yoshikawa S. 1996. The whole structure of the 13-subunit oxidized cytochrome c oxidase at 2.8 A. Science 272:1136-1144.

Uilenberg, G. 2006. Babesia--a historical overview. Vet Parasitol 138:3-10.

Ward BL, Anderson RS, Bendich AJ. 1981. The mitochondrial genome is large and variable in

a family of plants (cucurbitaceae). Cell. 25:793-803.

Widger WR, Cramer WA, Herrmann RG, Trebst A. 1984. Sequence homology and structural similarity between cytochrome *b* of mitochondrial complex III and the chloroplast b_6-*f* complex: position of the cytochrome *b* hemes in the membrane. Proc Natl Acad Sci USA 81:674-678.

Wilson RJ, Williamson DH. 1997. Extrachromosomal DNA in the Apicomplexa. Microbiol Mol Biol Rev. 61:1-16.

Xia, D., C. A. Yu, H. Kim, J. Z. Xia, A. M. Kachurin, L. Zhang, L. Yu, and J. Deisenhofer. 1997. Crystal structure of the cytochrome bc_1 complex from bovine heart mitochondria. Science 277:60-66.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 24:1586-1591.

Yun, C. H., A. R. Crofts, and R. B. Gennis. 1991. Assignment of the histidine axial ligands to the cytochrome b_H and cytochrome b_L components of the bc_1 complex from *Rhodobacter sphaeroides* by site-directed mutagenesis. Biochemistry 30:6747-6754.

Figure legends

FIG. 1. Structure of the mitochondrial genomes of *B. gibsoni* (a), *T. 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 terminal inverted repeat (TIR) and contains a *cox3* gene and rRNA gene fragments. Shown for comparison is the mt genome of *P. falciparum* (e) (Feagin et al. 1997). Genes shown above the bold line in each genome are transcribed left to right and those below 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 large subunit (LSU) and small subunit (SSU) rRNA genes, respectively. Abbreviations are: *cox1*, cytochrome *c* oxidase subunit I; *cox3*, cytochrome *c* oxidase subunit III; *cob*, cytochrome *b*.

FIG. 2. Monomeric linear structure of the mitochondrial genomes of *B. gibsoni, T. 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.

FIG. 3. Transcription of the three protein coding genes, *cox1* (a), *cox3* (b) and *cob* (c), in the mitochondrial genome of *B. gibsoni, T. orientalis* and *T. equi*. See Method section for details.

FIG. 4. Transcription of the *B. gibsoni* mt genome LSU6. *B. gibsoni* total RNA was probed

with oligonucleotides complementary to the putative LSU6 fragment. Probe sequences: lane 1; TGATAGCGGTTAATCTTTCCTAT, lane 2; CTTACGTACTCCGGCTAT. Positions and sizes (in nucleotides (nt)) of marker RNAs (DynaMarker RNA Low II, BioDynamics Laboratory) are shown.

FIG. 5. The Maximum likelihood phylogenetic tree of the two mitochondrial protein coding genes, *cox1* and *cob* from eight *Babesia* and *Theileria* species. *P. falciparum* was used as an outgroup. Concatenated amino acid sequences (834 sites) were used with 1,000 heuristic replicates under a JTT model ($\alpha = 0.72$). The numbers shown along nodes represent bootstrap values. Arrows, A and B, indicate alternative positions of the *T. equi* sequence, whose possibilities were statistically compared by the SH and KH tests. Five parasite sequences obtained in this study are underlined.

FIG. 6. Arrangement of small homologous non-coding sequences in the mitochondrial genome of *T. equi* (a), *B. gibsoni* (b) and *T. orientalis* (c). Homologous sequences are indicated by matched colored bars and arrows. A 3.0-kb inversion in the *T. orientalis* mt genome is shown by green bars with a circular arrow in (b) and (c). For details of abbreviations, see FIG. 1.

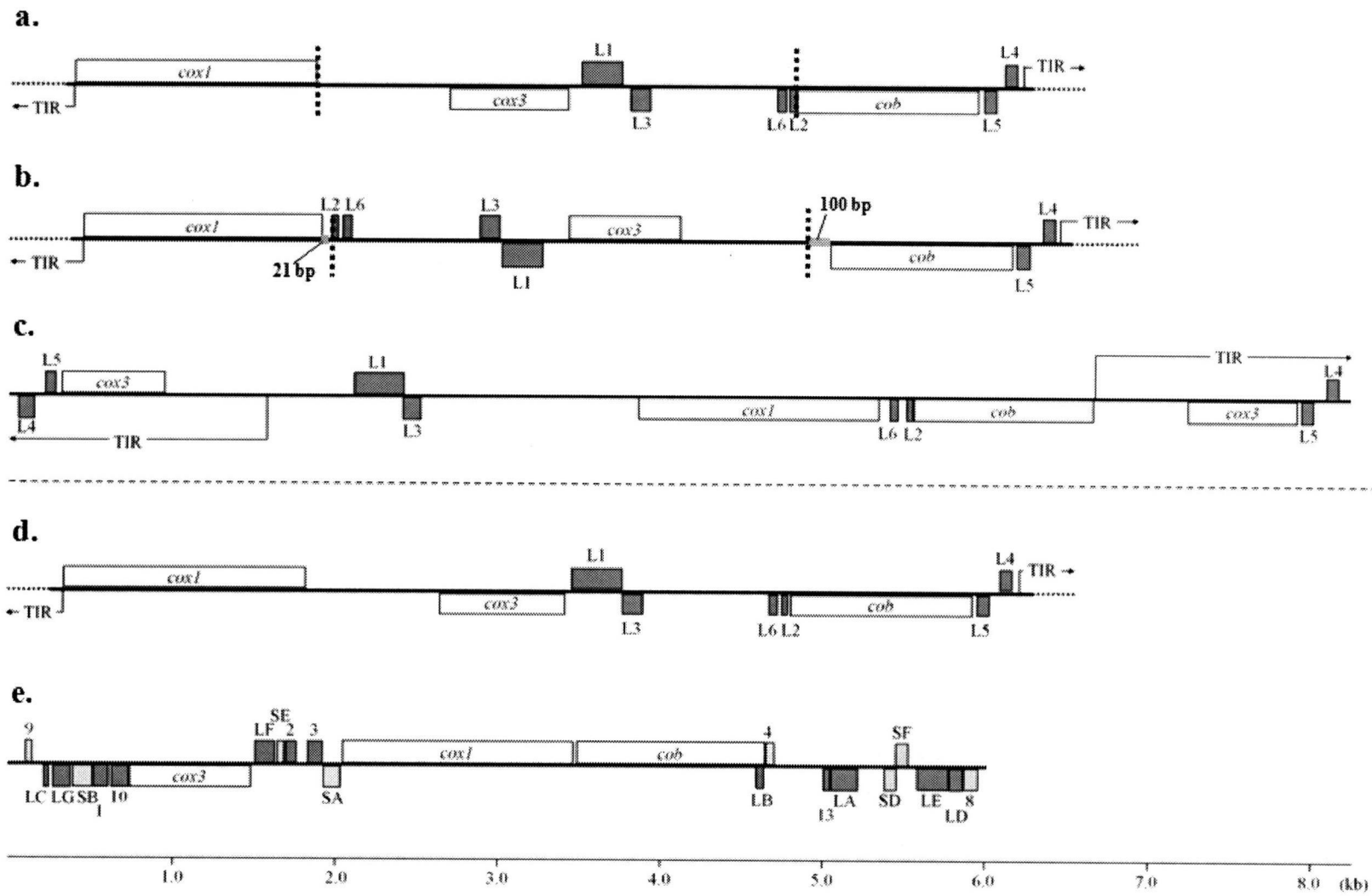
a.
cox1
cox3
cob
TIR
TIR
L1
L3
L6 L2
L4
L5
b.
cox1
cox3
cob
TIR
TIR
L2 L6
21 bp
L3
L1
100 bp
L4
L5
c.
L5
cox3
L4
TIR
L1
L3
cox1
cob
L6 L2
TIR
L4
cox3
L5
d.
cox1
cox3
cob
TIR
TIR
L1
L3
L6 L2
L4
L5
e.
9
SE
LF 2
3
4
SF
cox1
cox3
cob
LC LG SB 1 10
SA
LB
LA 13
SD
LE
8
LD
1.0
2.0
3.0
4.0
5.0
6.0
7.0
8.0
(kb)

FIG. 1.

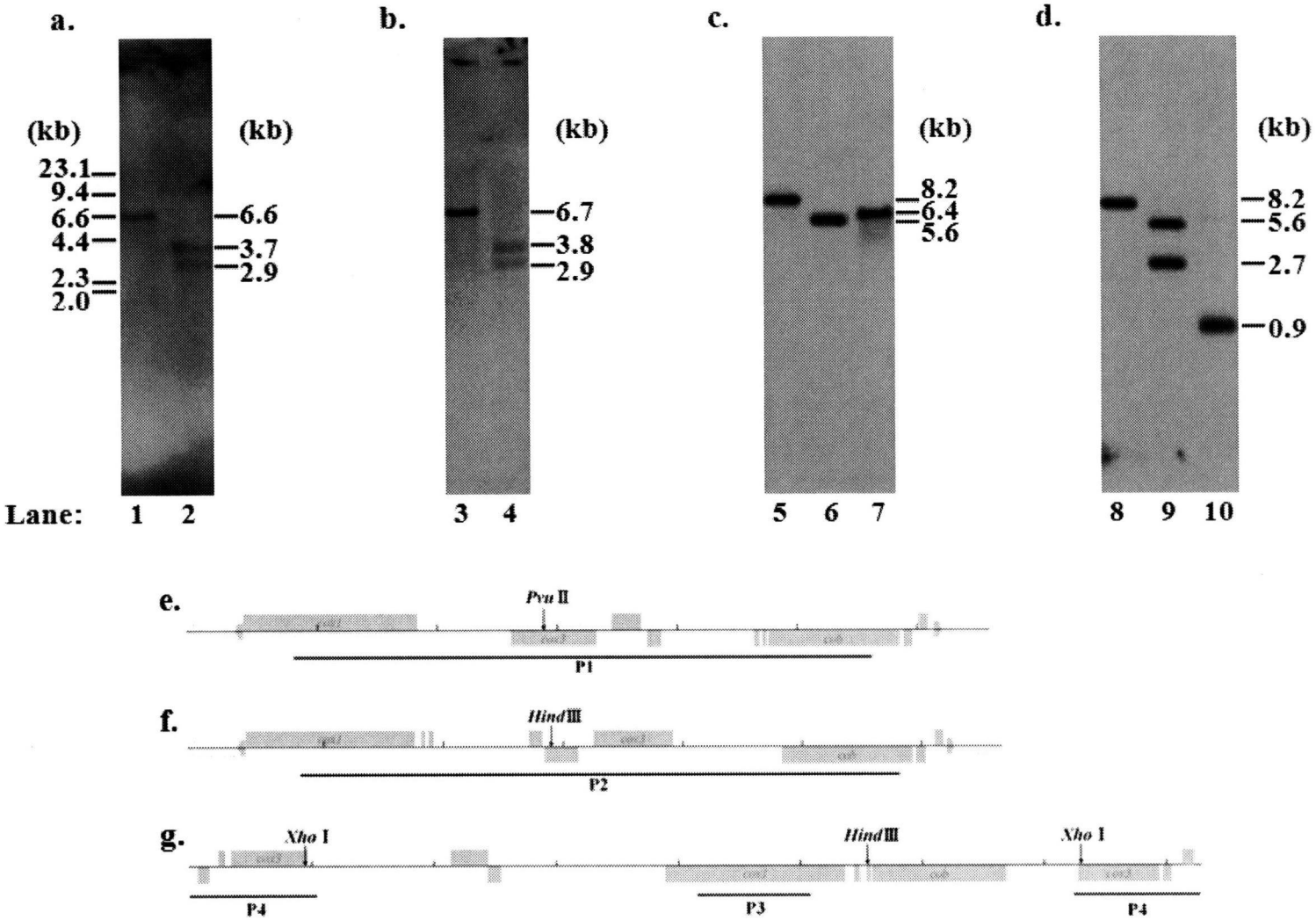
a.
(kb)
23.1
9.4
6.6
4.4
2.3
2.0
(kb)
6.6
3.7
2.9
Lane: 1 2
b.
(kb)
6.7
3.8
2.9
3 4
c.
(kb)
8.2
6.4
5.6
5 6 7
d.
(kb)
8.2
5.6
2.7
0.9
8 9 10
e.
Pvu II
P1
f.
HindIII
P2
g.
Xho I
HindIII
Xho I
P4
P3
P4

FIG. 2.