The *Plasmodium* Apicoplast Genome: Conserved Structure and Close Relationship of *P. ovale* to Rodent Malaria Parasites

Nobuko Arisue,*1 Tetsuo Hashimoto,2 Hideya Mitsui,2,3 Nirianne M.Q. Palacpac,1 Akira Kaneko,4,5,6 Satoru Kawai,7 Masami Hasegawa,†3 Kazuyuki Tanabe,1 and Toshihiro Horii1

1Department of Molecular Protozoology, Research Institute for Microbial Diseases, Osaka University, Suita, Osaka, Japan
2Graduate School of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Japan
3Department of Biosystems Science, Graduate University for Advanced Studies (Sokendai), Hayama, Kanagawa, Japan
4Department of Microbiology, Tumor and Cell Biology, Karolinska Institutet, Stockholm, Sweden
5Department of Parasitology, Osaka City University, Graduate School of Medicine, Osaka, Japan
6Institute of Tropical Medicine, Nagasaki University, Nagasaki, Japan
7Laboratory of Tropical Medicine and Parasitology, Dokkyo University School of Medicine, Tochigi, Japan

*Corresponding author: E-mail: arisue@biken.osaka-u.ac.jp.
Associate Editor: Richard Thomas

Abstract

Apicoplast, a nonphotosynthetic plastid derived from secondary symbiotic origin, is essential for the survival of malaria parasites of the genus *Plasmodium*. Elucidation of the evolution of the apicoplast genome in *Plasmodium* species is important to better understand the functions of the organelle. However, the complete apicoplast genome is available for only the most virulent human malaria parasite, *Plasmodium falciparum*. Here, we obtained the near-complete apicoplast genome sequences from eight *Plasmodium* species that infect a wide variety of vertebrate hosts and performed structural and phylogenetic analyses. We found that gene repertoire, gene arrangement, and other structural attributes were highly conserved. Phylogenetic reconstruction using 30 protein-coding genes of the apicoplast genome inferred, for the first time, a close relationship between *P. ovale* and rodent parasites. This close relatedness was robustly supported using multiple evolutionary assumptions and models. The finding suggests that an ancestral host switch occurred between rodent and human *Plasmodium* parasites.

Key words: malaria, *Plasmodium ovale*, apicoplast genome, molecular phylogeny.

Widely found in apicomplexan parasites, the apicoplast is a nonphotosynthetic plastid, a product of secondary endosymbiosis of algal origin (McFadden 2011). The plastid retains critical metabolic pathways such as type II fatty acid synthesis, isoprenoid synthesis, and heme synthesis. The 35-kb circular apicoplast genome of *Plasmodium falciparum*, the most virulent human malaria parasite, possesses translation- and transcription-related protein genes, rRNAs, tRNAs, and several other genes. Elucidation of the evolution of the apicoplast is important to better understand the functions of the organelle. However, the complete apicoplast genome sequence is available from only a few apicomplexan parasites (e.g., Wilson et al. 1996; Cai et al. 2003; Brayton et al. 2007). Among malaria parasites, *P. falciparum* is the only species whose apicoplast genome was completely sequenced (Wilson et al. 1996). Thus, little is known on the divergence of the apicoplast genome in the genus *Plasmodium*. In this study, we report structural and phylogenetic analyses of the near-complete apicoplast genome obtained from eight species, which, together with *P. falciparum*, represent the different well-known *Plasmodium* lineages. *Plasmodium falciparum* has recently been shown to be closely related to parasites of Great Apes (e.g., Duval et al. 2010; Krief et al. 2010; Liu et al. 2010), while *P. vivax*, the human parasite most prevalent outside Africa, positioned within the Asian Old World Monkey parasites (Mu et al. 2005; Escalante et al. 2005). The phylogenetic positions of other human malaria parasites *P. ovale* and *P. malariae*, however, remain unresolved. Also, relationship between *P. falciparum* and *P. gallinaceum*, an avian malaria parasite, remain unsettled. Therefore, we specifically address these issues using the apicoplast genome sequences.

In this work, the apicoplast genome sequences were obtained from the following eight *Plasmodium* species (strain name enclosed): a bird parasite *P. gallinaceum* (AB), three rodent parasites *P. yoelii* (17XL), *P. berghei* (ANKA), and *P. chabaudi* (AS), three human parasites *P. malariae* (Kisii 67), *P. ovale* (Nigeria II) and *P. vivax* (Salvador I), and a macaque parasite *P. coatneyi* (CDC) (AB649417–AB649424, supplementary fig. S1 and materials and methods, Supplementary Material online). Except for a small region around tRNA-Ile in the inverted repeats (IRs), the sequenced regions cover the near-complete apicoplast genome (fig. 1; supplementary fig. S2 and table S1, Supplementary Material online). Among the nine species, variations in the genome size based on comparable sequenced regions were from 33,896 to 33,630 bp, except for *P. chabaudi* (29,600 bp). The slight size difference (except for *P. chabaudi*) was mostly due to indels in genes and intergenic regions. All the *Plasmodium* apicoplast genomes contained genes for small subunit (SSU)
and large subunit (LSU) rRNAs, 25 species of tRNAs, 17 ribosomal proteins, 3 subunits of RNA polymerase, elongation factor Tu (tufA), caseinolytic protease C (clpC), sulfur mobilizing protein B (sufB), and 7 open reading frames of unknown function (supplementary table S1, Supplementary Material online), all tightly packed in the genome with very short intergenic regions. Notably, gene repertory, gene arrangement, and other structural attributes were almost identical among *Plasmodium* species, but the region, shown in pale gray using dashed lines, was almost deleted in *Plasmodium chabaudi*. AT contents of nine species are almost constant, ranging from 84.9% (*P. berghei*) to 86.9% (*P. falciparum*). For details, see supplementary figure S2 (Supplementary Material online).

We performed phylogenetic analyses for 30 protein-coding genes, rRNA genes, and tRNA genes. Due to high sequence similarity (91.9–100.0%) (supplementary table S2, Supplementary Material online), sequences of both rRNA and tRNA genes did not possess sufficient phylogenetic signals to resolve relationships among *Plasmodium* lineages (data not shown). We then focused on the 30 protein-coding genes, whose sequence similarity was relatively low (72.0–92.3% at the nucleotide level). For all protein-coding genes, except rpoD gene, both nucleotide and amino acid sequences were unambiguously aligned. The rpoD gene contained a highly variable region (supplementary fig. S3, Supplementary Material online), and thus this region was excluded from further analysis. The DNA data set of the 30 protein-coding genes (20,823 nucleotides/6,941 amino acids), although extremely AT-rich, in particular at the third codon position (>96%), showed no extreme difference in compositional bias among the nine species for both the DNA and protein data sets (P > 0.05, $\chi^2$ test) (supplementary fig. S4, Supplementary Material online).

Optimal maximum likelihood (ML) trees of RAxML analyses using the DNA and protein data sets were identical in their topologies. As representative, the tree of protein data set with JTT + F + G model is shown in figure 2A. The tree
clearly reconstructed the following previously established phylogenetic relationships: that is, the monophyly of rodent parasites (e.g., Martiensen et al. 2008), the close affinity of *P. vivax* with an Asian macaque parasite *P. coatneyi* (e.g., Escalante and Ayala 1994), each with 100% bootstrap proportion value (BV) in both DNA and protein analyses. A close relationship was also found between *P. falciparum* and *P. gallinaceum*, being consistent with previous studies (e.g., Hayakawa et al. 2008; Nishimoto et al. 2008). Importantly, the present analysis revealed a sister-group relationship between *P. ovale* and the rodent parasites with BV of 100% and 91% in the DNA and the protein analyses, respectively. For *P. malariae*, its close relatedness to the clade of rodent parasites and *P. ovale* was supported by DNA analysis (85% BV) but was not well supported by protein analysis (47% BV).

When the *Plasmodium* tree was rooted by using *Toxoplasma* and *Eimeria* as outgroups, both DNA and protein analyses (using a data set of 11,565 nucleotides/3,855 amino acids) inferred a tree in which *P. gallinaceum* was located at the root position of the *Plasmodium* tree; although it cannot also be ruled out if the root is on the *P. falciparum* branch or on the common ancestral branch of *P. falciparum* and *P. gallinaceum* (fig. 2B). As well as in the unrooted tree in figure 2A, a close relationship between *P. ovale* and rodent parasites was reconstructed also in the rooted tree (fig. 2B); while a close relationship between *P. falciparum* and *P. gallinaceum* was not supported.

To confirm that the highly supported clade combining *P. ovale* and rodent parasites was not an artifact related to model misspecification, we performed further analyses of the unrooted tree under a variety of substitution models and partitioning strategies. Fifteen alternative trees made from five lineages, that is, *P. vivax*/*P. coatneyi*, *P. malariae*, *P. ovale*, three rodent malaria species, and the rest (*P. falciparum* and *P. gallinaceum*) were exhaustively compared (table 1). Regardless of the different substitution models assumed, the DNA and protein data sets of 30 concatenated or separated genes consistently favored Tree 1 in table 1 as the ML tree among the 15 alternative trees. The clade combining *P. ovale* and the rodent parasites was supported by high BVs in any model analyzed (supplementary table S3, Supplementary Material online). Comparison of the Akaike information criterion values between different analyses revealed that the concatenate analysis of the 30-
gene data set with codon + Γ model was by far the most appropriate among the five alternatives (Model C in supplementary table S3, Supplementary Material online). With the protein data set, RAxML analyses using three substitution models (supplementary materials and method, Supplementary Material online) consistently revealed the same optimal tree shown in figure 2, where the close relatedness between P. ovale and the rodent parasites was supported with high BVs (data not shown). Also in the exhaustive 15-tree analyses, both the concatenate and the separate analyses using the JTT + F + Γ model supported the P. ovale–rodent parasite relationship. Bootstrap analysis and P values of the Approximately unbiased (AU) test for 15 trees were shown for the DNA analysis using Model C, which was selected as the most appropriate model (supplementary table S3: concatenate analysis with codon + Γ model, Supplementary Material online; table 1). BVs for Trees 1–3 were 99.2%. Approximately unbiased (AU) test rejected Trees 4–15, with the significance level P = 0.02, when compared with the ML tree (Tree 1). Remarkably, only Trees 2 and 3 have no significant log-likelihood difference and together with Tree 1, select for topologies where P. ovale is a sister group to rodent parasites.

The present study revealed the highly conserved structure of the apicoplast genome in the genus Plasmodium. Similarly, the mitochondrial genome of Plasmodium species was highly conserved (Hikosaka et al. 2011). In contrast, the Plasmodium nuclear genome shows substantial structural difference in gene order and arrangements in the 14 chromosomes among species (Carlton et al. 2008). In its life cycle, Plasmodium has the sexual stage where chromosome shuffling and meiotic recombination occur, thus, the difference in the mode of genome inheritance between the nuclear and organelle genomes is probably associated with conservation/variation of genome structures. Both the apicoplast and mitochondria genomes show uniparental inheritance (McFadden 2011).

The present phylogenetic analyses consistently supported the close relationship of P. ovale and rodent parasite species. This finding is different with the ML tree previously inferred from mitochondrial genome sequences (Hayakawa et al. 2008; Krief et al. 2010), which showed a close relationship between P. ovale and P. malariae. Pacheco et al. (2011) suggested an earlier divergence of P. ovale than P. malariae. However, these relationships were not robustly supported, and the results varied depending on the different gene- and taxon-sampling, as well as on the phylogenetic methods employed. Our analysis of the mitochondrial data set comparing the 15 trees in table 1 with the same taxon sampling as in the apicoplast data set revealed that the close relationship between P. ovale and rodent malaria species was not significantly rejected (data not shown).

To our knowledge, this is the first robust inference on the phylogenetic position of P. ovale. The close relatedness suggests a third example of a host switch associated with a human malaria parasite. Based on recent evidence, P. falciparum became a human parasite by host switch from Great Apes in Africa (e.g., Duval et al. 2010; Krief et al. 2010; Liu et al. 2010), while P. vivax was derived by introgression from Old World Monkey parasites in Southeast Asia (Mu et al. 2005; Escalante et al. 2005). The present study, however, cannot reject the hypothesis of a host switch from ancestral humans to rodents, thus, a host switch at both directions: from human to rodent or vice versa is possible. Plasmodium ovale has always been associated to be closely related with P. vivax because both parasites have the liver dormant form or hypnozoite in their life cycle and show Schüffner’s dots in infected erythrocytes. The present study puts forward an example where biological and morphological resemblances have no correlation with evolutionary relationship in Plasmodium.

**Supplementary Material**

Supplementary materials and methods, figures S1–S4, and tables S1–S3 are available at Molecular Biology and Evolution online (http://www.mbe.oxfordjournals.org/).

**Acknowledgments**

The authors thank Dr R. Culleton (Nagasaki University, Japan) for genomic DNA of P. ovale (Nigeria II). This work was supported by grants from the Ministry of Education, Culture, Sports, Science, and Technology of Japan (18073013, 18GS03140013, 20390120, 21570238, 23117005, and Frontier Biomedical Science Underlying Organelle Network Biology).

**References**


