Published in final edited form as:

Mol Phylogenet Evol. 2010 October ; 57(1): 135–143. doi:10.1016/j.ympev.2010.06.004.

Malaria Parasite Sequences from Chimpanzee Support the Co-Speciation Hypothesis for the Origin of Virulent Human Malaria (*Plasmodium falciparum*)

Austin L. Hughes^{1,*} and Federica Verra²

¹ Department of Biological Sciences, University of South Carolina, Columbia SC 29208 USA

² Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London WC1E 7HT, UK

Abstract

Phylogenetic analyses of the mitochondrial cytochrome b (*cytb*), apicoplast caseinolytic protease C (*clpC*), and 18S rRNA sequences of *Plasmodium* isolates from chimpanzees along with those of the virulent human malaria parasite *P. falciparum* showed that the common chimpanzee (*Pan troglodytes*) malaria parasites, assigned by Rich et al. (2009; Proc. Natl. Acad. Sci. USA 106, 14902–14907) to *P. reichenowi*, constitute a paraphyletic assemblage. The assumption that *P. falciparum* diverged from *P. reichenowi* as recently as 5,000–50,000 years ago would require a rate of synonymous substitution/site/year in *cytb* and *clpC* on the order of 10^{-5} – 10^{-6} , several orders of magnitude higher than any known from eukaryotic organelle genomes, and would imply an unrealistically recent timing of the most recent common ancestor of *P. falciparum* mitochondrial genomes. The available data are thus most consistent with the hypothesis that *P. reichenowi* (in the strict sense) and *P. falciparum* co-speciated with their hosts about 5–7 million years ago.

For many years, only a single species of malaria parasites, named *Plasmodium reichenowi* (Bray 1956), was known to infect common chimpanzees (*Pan troglodytes*); and this species was represented by a single isolate (CDC1). Escalante and Ayala (1994) were the first to apply phylogenetic analysis to sequence data from *P. reichenowi* CDC1. Their phylogenetic tree based on 18S SSU rRNA sequences placed *P. reichenowi* as a sister taxon of *P. falciparum*, the most virulent human malaria parasite (Escalante and Ayala 1994). It had previously been difficult to determine the phylogenetic position of *P. falciparum*, which appeared not to be closely related to other *Plasmodium* species infecting primates, including the other human malaria parasites (Waters et al. 1991, 1993). A natural consequence of the close relationship between *P. falciparum* and *P. reichenowi* was the co-speciation hypothesis, first proposed by Escalante and Ayala (1994). According to this hypothesis, the divergence of *P. reichenowi* and *P. falciparum* occurred around the time that the human and chimpanzee lineages diverged (5–7 Mya). As illustrated by comparisons of a number of orthologous genes from these two *Plasmodium* species, the assumption that *P. falciparum* and *P. reichenowi* diverged when their hosts diverged provided estimates of the synonymous substitution rate similar to those seen in other eukaryotes (Hughes and Verra 2001; Tanabe et al. 2004). Similarly, a mitochondrial

*Address correspondence to: Austin L. Hughes, Ph.D., Department of Biological Sciences, University of South Carolina, Coker Life Sciences Bldg., 715 Sumter St., Columbia SC 29208 USA, Tel: 1-803-777-9186, Fax: 1-803-777-4002, austin@biol.sc.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

molecular clock for *Plasmodium* calibrated on the divergence of African and Asian Old World monkey parasites yielded a time estimate for the *P. falciparum*-*P. reichenowi* divergence consistent with the cospeciation hypothesis (Hayakawa et al. 2008).

Rich and colleagues (2009) reported partial sequences of three genes from a number of new isolates of malaria parasites from chimpanzees: the mitochondrial cytochrome b (*cytb*) gene; the apicoplast caseinolytic protease C (*clpC*) gene; and the 18S ribosomal RNA gene (*18S rRNA*) gene. These authors claimed that their sequence data provided evidence against the co-speciation hypothesis, indicating instead that *P. falciparum* is nested within *P. reichenowi* and that the former may be recently derived from the latter, perhaps quite recently. Rich et al. (2009) did not attempt to date the alleged transfer event, but they cited as plausible hypotheses both a date within the past 50,000 years (Rich et al. 1998) or even a date within the past 5,000 years, after the origin of agriculture in West Africa (Volkman et al. 2001). Indeed, Rich et al. (2009) deemed the latter dating the “most probable” (p. 14905). The latter dates were originally proposed for the transfer to humans of a single haploid *P. falciparum* genotype (Rich et al. 1998; Volkman et al. 2001), a hypothesis that is inconsistent with the extensive sequence polymorphism observed in *P. falciparum* (Hughes and Verra 2001, 2002; Mu et al. 2002; Volkman et al. 2001; Neafsey et al. 2008).

Rich et al. (2009) assigned all of the sequences they obtained from chimpanzees to “*P. reichenowi*,” along with the previously sequenced *P. reichenowi* CDC1; but the *cytb* phylogeny they presented indicated that by this definition, “*P. reichenowi*” is paraphyletic. Additional partial sequences of *Plasmodium* mitochondrial genes have been obtained from the common chimpanzee and from the gorilla (*Gorilla gorilla*), certain of which have been tentatively assigned to a new species, *P. gaboni* (Ollomo et al. 2009; Prugnolle et al. 2010). If multiple *Plasmodium* species infect the common chimpanzee, the hypothesis of a recent origin of *P. falciparum* remains in doubt.

Here we reanalyze the data of Rich et al. (2009), along with those of Ollomo et al. (2009) and Prugnolle et al. (2010), in order to provide a further test of the co-speciation hypothesis. We conduct phylogenetic analyses to test whether the malaria parasites sequenced from chimpanzee can be considered a monophyletic group. Moreover, we estimate rates of synonymous nucleotide substitution in the *cytb* and *clpC* genes, assuming various divergence times between *P. falciparum* and the malaria parasites derived from chimpanzees. These estimates can be used to test the plausibility of different divergence times of *P. falciparum* assuming rates of synonymous substitution in the mitochondrial and apicoplast genomes consistent with those seen in organelle genomes of other eukaryotes.

Methods

Sequences were aligned using the Megalign procedure in the DNASTAR package (Burland 2000). In all analyses, any site at which the alignment postulated a gap were excluded from all pairwise comparisons (“complete deletion” Tamura et al. 2007) so that all evolutionary distance estimates were based on comparable numbers of sites. Sequences were analyzed from the mitochondrial *cytb* gene (477 aligned sites), the apicoplast *clpC* gene (315 aligned sites), and the nuclear A-type *18S rRNA* gene (336 aligned sites). The *cytb* sequence from Ollomo et al. (2009) has a four-codon deletion (codons 22–25) relative to other *Plasmodium cytb* genes; thus, following the policy of complete deletion, these four codons were excluded in analyses of *cytb*. We also conducted a further phylogenetic analysis of *cytb* sequences, including 12 sequences from Prugnolle et al. (2010); because the latter sequences only partially overlapped other partial *cytb* sequences, only 349 aligned sites were available for the latter analysis.

In *Plasmodium* there are distinct *18S rRNA* genes expressed in different developmental stages (Nishimoto et al. 2008). Of seven *18S rRNA* genes sequenced from chimpanzee-derived malaria parasites by Rich et al. (2009), six were of the A-type, which is expressed in the asexual stage. We analyzed only the six A-type sequences, along with previously sequenced A-type genes from *P. falciparum* and from *P. reichenowi* CDC1.

Accession numbers of sequences used in phylogenetic analyses are given in Figures 1–4, except those of *cytb* from *P. falciparum* (N = 121) and *P. vivax* (N = 285), which were too numerous to be shown in the figures. The accession numbers of the latter sequences are listed in Supplementary Table S1.

Phylogenetic trees were constructed by a variety of methods, including maximum likelihood, maximum parsimony, and minimum evolution, all of which yielded similar results. Therefore, we include here only minimum evolution (ME) trees (Rzhetsky and Nei 1992) based on the number of nucleotide substitutions per site (d), estimated by the Tamura-Nei (TN) method (Tamura and Nei 1993) and the maximum composite likelihood (MCL) method (Tamura et al. 2004, 2007). The TN model takes into account both transitional bias and nucleotide content bias, while the MCL method fits this model simultaneously to the set of sequences analyzed (Tamura et al. 2004, 2007). As with most *Plasmodium* genes, the present data showed strongly biased nucleotide content. In *cytb*, the mean percent A+T was 86.3% at third codon positions and 72.8% overall; and in *clpC*, the mean percent A+T was 94.6% at third codon positions and 74.7% overall. In the *18S rRNA* genes, mean percent A+T was 65.9%. Mean transition:transversion ratio was 1.2 in *cytb*, 0.2 in *clpC*, and 1.3 in *18S rRNA*. Reliability of branching patterns within trees was tested by the internal branch test, with the standard error of the branch length computed by the bootstrap method (Nei and Kumar 2000); 1000 bootstrap samples were used. In each case, the trees of *P. falciparum* and *P. reichenowi* sequences were rooted by sequences from other *Plasmodium* species; the latter were chosen so as to maximize the number of aligned sites within each gene.

Because of the bias in transition:transversion ratios, we used Li's (1993) method to estimate the number of synonymous substitutions per synonymous site (d_S) and the number of nonsynonymous substitutions per nonsynonymous site (d_N). In preliminary analyses, other methods (Nei and Gojobori 1986; Yang and Nielsen 2000) yielded similar estimates, as was found in previous studies of *Plasmodium* mitochondrial protein-coding genes (Jongwutiwes et al. 2005), probably because the biases in the present data were not very great. Within species, we computed the mean of all pairwise d_S values, designated the synonymous nucleotide diversity (π_S); and the mean of all pairwise d_N values, designated the nonsynonymous nucleotide diversity (π_N). In the case of *18S rRNA* genes, we computed the mean of all pairwise d values, estimated by the MCL method; this value is designated nucleotide diversity (π). Standard errors of mean d , d_S and d_N were estimated by the bootstrap method (Nei and Kumar 2000); 1000 bootstrap samples were used.

Rates of synonymous substitution per year (r) were estimated by the relationship $d_S = 2rt$, where t is the divergence time in years. The time of the most recent common ancestor (MRCA) of *P. falciparum cytb* sequences was estimated by the synonymous divergence at the deepest branch point within the subtree of *P. falciparum* sequences. Note that d_S at the deepest branch point is not the same as the synonymous nucleotide diversity (π_S), which is the average of all pairwise comparisons.

Results

Phylogenetic Analyses

In the ME tree of *cytb* sequences based on the MCL distance, the sequences obtained from chimpanzee (chimp-derived sequences) and assigned to *P. reichenowi* by Rich et al. (2009) formed two separate clades (designated A and B in Figure 1). *P. reichenowi* CDC1 fell within clade B, although this clade did not receive significant support from the internal-branch test (Figure 1). *P. gaboni* (Ollomo et al. 2009) fell within clade A, and this clade received highly significant support (99% confidence level) in the internal branch test (Figure 1).

A slightly different topology was seen in the ME tree based on the TN distance (Figure 2); the latter topology was similar to that obtained by Rich et al. (2009) in a *cytb* tree based on a smaller number of aligned sites. In this tree, the *P. reichenowi* CDC1 isolate clustered separately with *P. falciparum* (with 93% confidence level; Figure 2). The remaining chimp-derived sequences formed two separate clades (designated A and B in Figure 2). In the TN-based tree, Clade B did not include *P. reichenowi* CDC1, but otherwise its membership was the same as in the MCL-based tree (Figures 1 and 2). *P. gaboni* (Ollomo et al. 2009) fell within clade A in the TN-based tree, and the membership of clade A was the same in both MCL- and TN-based trees (Figures 1 and 2). As in the MCL-based tree, in the TN-based tree, clade A received highly significant support (99% confidence level) in the internal branch test (Figure 2). Thus, although differing in details, neither tree supported the hypothesis of monophyly of “*P. reichenowi*,” a result consistent with the phylogenetic analysis of *cytb* reported by Rich et al. (2009).

Because of the smaller number of aligned sites, the number of branches receiving statistically significant support was fewer when the sequences from Prugnolle et al. (2010) were included (Figure 3; 349 sites) in phylogenies than when the latter sequences were not included (Figures 1–2; 477 sites). In the tree based on the MCL distance for the data set including the sequences of Prugnolle et al. (2010), a cluster of chimp-derived sequences (Chimp-derived A; Figure 3) that formed an outgroup to both other chimp-derived and gorilla-derived sequences (Figure 3). The Chimp-derived A group, formed two distinct clusters, each supported by a significant internal branch; each of these included sequences assigned to *P. gaboni* (Figure 3). *P. reichenowi* CDC1 again fell in the cluster closest to *P. falciparum* (Figure 3). For the data set including the sequences of Prugnolle et al. (2010), the tree based on the TN distance (not shown). The sequences of Prugnolle et al. (2010) included three sequences of *P. falciparum* from gorilla (GU045311, GU045312, and GU045313); these fell within the cluster of *P. falciparum* sequences from human. Indeed two of the three sequences were identical to numerous *P. falciparum cytb* sequences previously isolated from humans.

The phylogenetic analysis of *clpC* based on both MCL (Figure 4) and TN (not shown) distances likewise did not show monophyletic grouping of chimp-derived sequences, although the support was weak. The phylogenetic analyses based on *18S rRNA* genes, using both the MCL (Figure 5) and TN (not shown) distances, showed strong support for paraphyly of chimp-derived sequences. The interior-branch test provided 98% support for a clade including *P. falciparum* and two chimp-derived sequences (one of which was *P. reichenowi* CDC1; Figure 5). Two additional clades of chimp-derived sequences fell outside the latter clade (Figure 5). Thus, the phylogenetic analysis of *18S rRNA* sequences provided strong support for paraphyly of the isolates assigned to “*P. reichenowi*” by Rich et al. (2009).

Nucleotide Diversity

Synonymous nucleotide diversity (π_S) in all 10 *cytb* sequences obtained from chimpanzees was about 65 times as great as that for 121 *P. falciparum cytb* sequences; and the difference between the two values was highly significant ($P < 0.001$; Table 1). Similarly, nonsynonymous

nucleotide diversity (π_N) for the 10 chimp-derived sequences was about 19 times as great as that for *P. falciparum*; and again there was a highly significant difference between the two values ($P < 0.05$; Table 1).

Considering just clade A of chimp-derived *cytb* (Figures 1 and 2) π_S and π_N were greater than that of *P. falciparum*, but the difference was not significant (Table 1). On the other hand, both π_S and π_N were significantly greater in clade B of chimp-derived *cytb* (Figures 1 and 2) than in *P. falciparum*, whether or not *P. reichenowi* CDC1 was included in the former (Table 1). Note, however, that the π_S and π_N in *P. falciparum* were not extremely low; in fact, both were an order of magnitude greater than the corresponding values based on 285 *P. vivax* sequences, although the difference was not statistically significant. In the case of *clpC*, π_S for the 6 chimp-derived sequences was 0.0994 ± 0.0383 ; and π_N was 0.0448 ± 0.0031 .

In the case of the 8 “*P. reichenowi*” 18S *rRNA* sequences, π was 0.1688 ± 0.0191 . By contrast, d between the two *P. falciparum* A-type 18S *rRNA* sequences used in the phylogenetic analysis (Figure 4) was only 0.0042 ± 0.0019 . Likewise, d between the two *P. vivax* A-type 18S *rRNA* sequences used in the phylogenetic analysis (Figure 4) was only 0.0062 ± 0.0026 .

Substitution Rates

Based on the assumption of different divergence times between *P. falciparum* and the various clades of chimp-derived *cytb* and *clpC* sequences, we estimated rates of synonymous substitution in these genes (Table 2). Divergence time estimates of 200,000 years or less implied extraordinarily high rates of synonymous substitution (Table 2). For example, in the case of *cytb*, if *P. falciparum* diverged from *P. reichenowi* CDC1 and/or from other clade B (Figures 1 and 2) sequences 50,000 years ago, we must assume a rate of synonymous substitution around 2×10^{-6} substitutions/site/year (Table 2). This would in turn imply that the most recent common ancestor (MRCA) of *P. falciparum* mitochondrial genomes is only about 2,000 years ago (Table 2). If *P. falciparum* diverged from *P. reichenowi* CDC1 and/or from other clade B (Figures 1 and 2) sequences 5,000 years ago, we must assume a rate of synonymous substitution around 2×10^{-5} substitutions/site/year (Table 2). This would imply that the MRCA of *P. falciparum* mitochondrial genomes is only about 200 years.

Similarly, in the case of *clpC*, if *P. falciparum* diverged from the most closely related group of chimp-derived sequences (clade D; Figure 3) 50,000 years ago, this would imply a synonymous substitution rate of at least about 1×10^{-6} substitutions per site per year (Table 2). If the divergence took place only 5,000 years ago, the rate would be about 1×10^{-5} substitutions per site per year (Table 2). On the other hand, if *P. falciparum* diverged from the most closely related clade of “*P. reichenowi*” 5–7 million years ago, as predicted under the co-speciation hypothesis, the rate of synonymous substitution per site would be around $1\text{--}2 \times 10^{-8}$ substitutions/site/year (Table 2). Such a rate is consistent with a MRCA for currently extant *P. falciparum* mitochondrial lineages about 200,000 years ago (Table 2).

Discussion

Phylogenetic analyses of *cytb*, *clpC*, and 18S *rRNA* sequences from *Plasmodium* isolates from chimpanzees along with those of *P. falciparum* showed that the chimpanzee malaria parasites, assigned by Rich et al. (2009) to *P. reichenowi*, constitute a paraphyletic assemblage, as indeed did the *cytb* phylogeny reported by those authors. In each case, all *P. falciparum* sequences clustered together within the *P. reichenowi* assemblage (Figure 1–5). Given this phylogeny, the assignment of all of the chimpanzee-derived sequences to *P. reichenowi* and the human-derived sequences to *P. falciparum* does not seem defensible. In our phylogenetic trees of *cytb*, certain of the sequences assigned by Rich et al. (2009) clustered with those recently

assigned to *P. gaboni* (chimp-derived A; Figures 1–3). However, certain other sequences (chimp-derived B; Figures 1–3) did not form a monophyletic group with *P. gaboni*.

One alternative might be to group all chimpanzee-derived sequences, including those assigned to *P. gaboni* and *P. reichenowi*, in a single species with *P. falciparum*. This alternative might be reasonable if the transfer of *P. falciparum* to its human host occurred very recently; for example, within the last 5,000–50,000 years, as proposed by Rich et al. (2009). However, the sequence data show a level of sequence divergence within the chimpanzee-derived *Plasmodium* species that argues against their inclusion in a single species. In all of the chimpanzee-derived *cytb* sequences, the synonymous nucleotide diversity was over 20% (Table 1). This value seems high for within-species mitochondrial synonymous nucleotide diversity. For example, only a small fraction of vertebrate or invertebrate animal species showed π_S as high as 20% in the extensive survey of mitochondrial genes by Bazin et al. (2006).

Moreover, the A-type *18S rRNA* sequences assigned to *P. reichenowi* by Rich et al. (2009) showed nucleotide diversity of nearly 17%. Typically, divergence at the *18S rRNA* gene is assumed to be only about 1–2 % or less within species of single-celled eukaryotes (Moon-van der Stay et al. 2001; Allen et al. 2008). A similar assumption regarding the extent of within-species divergence has been made in the case of the homologous *16S rRNA* gene of bacteria (Fox et al. 1992). Consistent with this assumption, A-type *18S rRNA* sequences of other *Plasmodium* species are generally less than 1% different within species.

Thus, it seems likely that the chimpanzee-derived *Plasmodium* sequences are derived from as many as two or three different species; see for example, the three clades of chimp-derived sequences in our tree of A-type *18S rRNA* genes (Figure 5). However, we believe that it is premature at present to attempt to define species boundaries within this assemblage. Indeed it may prove challenging to reconcile new taxa identified by sequence data with taxa infecting the chimpanzees that were previously defined on morphological grounds, such as *P. schwetzi* and *P. rodhaini* (Brumpt 1939; Coatney 1971). Nonetheless, our phylogenetic analyses suggest that the group of sequences including *P. reichenowi* CDC1 is the sister taxon to *P. falciparum*.

In the phylogenetic analysis including *cytb* sequences from Prugnolle et al. (2009), three sequences assigned to *P. gaboni* clustered with three of the sequences assigned to *P. reichenowi* by Rich et al. (2009). This clustering pattern received highly significant support (Figure 3). On the other hand, three sequences assigned to *P. gaboni* fell outside this cluster, suggesting the possibility that the latter group may correspond to yet another distinct species. In any event, it seems plausible that the clades of chimp-derived *clpC* (Figure 3) and *18S rRNA* (Figure 4) sequences that clustered farthest out from *P. falciparum* may also belong to *P. gaboni* or a closely related species.

A synonymous substitution rate on the order of $1\text{--}2 \times 10^{-8}$ substitutions/site/year in both mitochondrial and apicoplast genomes is consistent with mitochondrial substitution rates reported for other eukaryotes. For example, estimates of the rate of synonymous substitutions/site/year in primates (Horai et al. 1995; Kivisild et al. 2006; Seo et al. 2004) have fallen in the range $1\text{--}5 \times 10^{-8}$ substitutions/site/year. Such a synonymous substitution rate is consistent with divergence of *P. falciparum* from its closest chimpanzee-associated relatives around 5–7 million years ago and thus with the co-speciation hypothesis. An estimate of the mitochondrial synonymous substitution rate on the order of 10^{-8} substitutions/site/year is also consistent with a MRCA of extant *P. falciparum* mitochondrial lineages about 200,000 years ago. The latter figure is consistent with previous estimates of the MRCA of *P. falciparum* based on nuclear (Hughes and Verra 2001, 2002; Mu et al. 2002) and mitochondrial (Jongwutiwes

et al. 2005) genes, as well as consistent with the substantial level of nucleotide sequence polymorphism in the *P. falciparum* genome uncovered by recent SNP analyses (Neafsey et al. 2008; Volkman et al. 2007). By contrast, the assumption that *P. falciparum* diverged from the most closely related chimpanzee parasite 50,000 years ago or less would require a synonymous substitution rate on the order of at least 10^{-6} , which is much higher than rates reported for any nuclear or organelle genome of eukaryotes (Li 1997). Rich et al. (2009) deemed “most probable” the hypothesis that the transfer of *P. falciparum* malaria to humans coincided with the advent of agriculture to West Africa; in other words within the last 5,000 years or so. This would require a synonymous substitution rate on the order of at least 10^{-5} , which is similar to that of DNA viruses (Li 1997). This in turn would imply that the mitochondrial MRCA of *P. falciparum* occurred within the last 200 years, which seems inconsistent with the nuclear genomic diversity and wide geographic range of *P. falciparum*.

The hypothesis that the advent of *P. falciparum* as a human disease is linked to the beginnings of agriculture in West Africa has a long history (Livingstone 1958). Evidence adduced for this hypothesis includes the estimated ages of the sickle-cell and G6PD polymorphisms (Livingstone 1958; Rich et al. 2009). However, the age of an evolved response to selection cannot be used to estimate the age of the selective agent, because the former depends on the occurrence of an appropriate mutation, which is a random event independent of selection pressure. By providing a marked increase in the human population, agriculture may indeed have increased the probability that such mutations would occur and spread; but there is no reason to suppose that the selective agent itself, *P. falciparum*, was not present in the human population well prior to that time.

As reviewed by Ho and Larsen (2006), some studies have indicated differences in molecular clocks over different time scales, with much faster rates being observed over the short times involved in pedigree studies. These differences are probably explained by the following factors: (1) the occurrence of slightly deleterious variants within populations, which will eventually be purged by purifying selection over evolutionary time; and (2) the saturation of highly variable sites over long periods of evolutionary time (Ho and Larsen 2006). Since slightly deleterious variants in coding regions are more likely to be nonsynonymous than synonymous (Hughes 2008), time estimates based on synonymous sites (as used here) are unlikely to be strongly affected by the former factor. Saturation of synonymous sites may indeed be a problem in inferring molecular clocks from mitochondrial genes, but probably only over much longer periods than the 5–7 million year time frame since the human-chimpanzee divergence. Note also that use of the same synonymous substitution rate both within *P. falciparum* and between *P. falciparum* and *P. reichenowi* provided results consistent with other studies based on nuclear genes (Hughes and Verra 2001, 2002; Mu et al. 2002). Moreover, if we follow Rich et al. (2009) in assuming a very recent date for the *P. falciparum*-*P. reichenowi* split, we should expect that the same rate will apply to divergence between these taxa and within-species polymorphism; yet it is this very assumption that leads to implausibly recent estimates of the MRCA of *P. falciparum* mitochondrial genomes.

The occurrence of natural *Plasmodium falciparum* infections in gorilla (Prugnolle et al. 2010) is of great interest for the question of the origin of human *P. falciparum*. One possible interpretation of these findings is that the gorilla constitutes a reservoir for *P. falciparum*, from which human infections are derived. An alternative interpretation is that the human species constitutes a reservoir from which occasional gorilla infections are derived; indeed, infections of gorillas by *P. falciparum* may be a very recent phenomenon, due to the recent expansion of the human population in West Africa. A prediction of the latter hypothesis is that all gorilla-derived *P. falciparum* isolates should show close relationships with human-derived isolates; and this prediction is supported by the limited available data. In our phylogenetic analyses, the three available gorilla-derived *P. falciparum* *cytb* sequences clustered within human *P.*

falciparum, and two of the three sequences were identical to *P. falciparum* *cytb* haplotypes widespread in the human population. On the other hand, if gorilla constitutes a reservoir for human infections, *P. falciparum* diversity in humans should constitute a subsample of that seen in gorilla. Thus, more extensive sequencing of genes from gorilla-derived malaria parasites should make it possible to decide between these two hypotheses.

In conclusion, our analyses provide support for the co-speciation hypothesis and indicate that it would be premature to rule out this hypothesis on the basis of present knowledge. Rather, the results suggest that *P. falciparum* may indeed be a sister species of *P. reichenowi* in the strict sense (including the CDC1 isolate) and that the divergence time of the two species may have been around the time of the human-chimpanzee divergence. Other *Plasmodium* taxa infecting chimpanzees probably represent at least two other species, which may have diverged from the common ancestor of *P. falciparum* and *P. reichenowi* 10–12 million years ago or more (Table 2). Recent paleontological evidence regarding the complex history of African hominids (White et al. 2009) makes it unsurprising that multiple speciation events occurred among the malaria parasites infecting hominids. The sequences analyzed here support the hypothesis (Ollomo et al. 2009; Prugnolle et al. 2010) that a number of anciently diverged lineages of *Plasmodium* continue to infect the common chimpanzee today. However, the amount of sequence data available for analysis remains small, and more extensive sequencing will be necessary to resolve the evolutionary history of malaria parasites infecting African hominids.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was supported by grant GM43940 from the National Institutes of Health to A.L.H.

References

- Allen KE, Li Y, Kaltenbock B, Johnson EM, Reichard MV, Panciera RJ, Little SE. Diversity of Hepatozoon species in naturally infected dogs in the southern United States. *Vet Parasitol* 2008;154:220–225. [PubMed: 18485602]
- Bazin E, Glémin S, Galtier N. Population size does not influence mitochondrial genetic diversity in animals. *Science* 2006;312:570–572. [PubMed: 16645093]
- Bray RS. Studies on malaria in chimpanzees. I. The erythrocytic forms of *Plasmodium reichenowi*. *J Parasitol* 1956;42:588–592. [PubMed: 13398864]
- Brumpt E. Les parasites du paludisme des chimpanzés. *C R Soc Biol (Paris)* 1939;130:847–840.
- Burland TG. DNASTAR'S lasergene sequence analysis software. *Methods Mol Biol* 2000;132:71–91. [PubMed: 10547832]
- Coatney GR. The simian malarias: zoonoses, anthroponoses, or both? *Am J Trop Med Hyg* 1971;20:795–803. [PubMed: 5002245]
- Escalante AA, Ayala FJ. Phylogeny of the malarial genus *Plasmodium*, derived from rRNA gene sequences. *Proc Natl Acad Sci USA* 1994;91:11373–11377. [PubMed: 7972067]
- Fox GE, Wisotzkey JD, Jurtshuk P Jr. How close is close: 16S rRNA sequence identity may not be sufficient to guarantee species identity. In *J Syst Bacteriol* 1992;42:166–170.
- Hayakawa T, Culleton R, Otani H, Horii T, Tanabe K. Big bang in the evolution of extant malaria parasites. *Mol Biol Evol* 2008;25:2233–2239. [PubMed: 18687771]
- Ho SY, Larsen G. Molecular clocks: when times are a-changin'. *Trends Genet* 2006;22:79–83. [PubMed: 16356585]

- Horai S, Hayasaka K, Kondo R, Tsugane K, Takahata N. Recent African origin of modern humans revealed by complete sequences of hominoid mitochondrial DNAs. *Proc Natl Acad Sci USA* 1995;92:532–536. [PubMed: 7530363]
- Hughes AL. Near neutrality: leading edge of the neutral theory of molecular evolution. *Ann NY Acad Sci* 2008;1133:162–179. [PubMed: 18559820]
- Hughes AL, Verra F. Very large long-term effective population size in the virulent human malaria parasite *Plasmodium falciparum*. *Proc R Soc Lond B* 2001;268:1855–1860.
- Hughes AL, Verra F. Extensive polymorphism and ancient origin of *Plasmodium falciparum*. *Trends Parasitol* 2002;18:348–351. [PubMed: 12377283]
- Jongwutiwes S, Putaporntip C, Iwasaki T, Ferreira MU, Kanbara H, Hughes AL. Mitochondrial genome sequences support ancient population expansion in *Plasmodium vivax*. *Mol Biol Evol* 2005;22:1733–1739. [PubMed: 15901839]
- Kivisild T, Shen P, Wall DP, Do B, Sung R, Davis K, Passarino G, Underhill PA, Scharfe C, Torroni A, Scozzari R, Modiano D, Coppa A, de Knijff P, Feldman M, Cavalli-Sforza LL, Oefner PJ. The role of selection in the evolution of human mitochondrial genomes. *Genetics* 2006;172:373–387. [PubMed: 16172508]
- Li WH. Unbiased estimates of the rates of synonymous and nonsynonymous substitution. *J Mol Evol* 1993;36:96–99. [PubMed: 8433381]
- Li, W-H. Molecular evolution. Sinauer associates; Sunderland MA: 1997.
- Livingstone FB. Anthropological implications of sickle cell gene distribution in West Africa. *Amer Anthropol* 1958;60:531–561.
- Moon-van der Stay SY, De Wachter R, Vaulot D. Oceanic 18S rRNA sequences from picoplankton reveal unsuspected eukaryotic diversity. *Nature* 2001;409:607–610. [PubMed: 11214317]
- Mu J, Duan J, Markova K, Joy D, Huynh CQ, Branch OH, Li WH, Su X. Chromosome-wide SNPs reveal an ancient origin for *Plasmodium falciparum*. *Nature* 2002;418:323–326. [PubMed: 12124624]
- Neafsey DE, Schaffner SF, Volkman SK, Park D, Montgomery P, Milner DA Jr, Lukens A, Rosen D, Daniels R, Houde N, Cortese JF, Tyndall E, Gates C, Stange-Thomann N, Sarr O, Ndiaye D, Ndir O, Mboup S, Ferreira MU, do Lago Moraes S, Dash AP, Chitnis CE, Wigand RC, Hartl DL, Birren BW, Lander ES, Sabeti PC, Wirth DF. Genome-wide SNP genotyping highlights the role of natural selection in *Plasmodium falciparum* population divergence. *Genome Biol* 2008;9:R171.
- Nei M, Gojobori T. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol* 1986;3:418–426. [PubMed: 3444411]
- Nei, M.; Kumar, S. Molecular evolution and phylogenetics. Oxford University Press; New York: 2000.
- Nishimoto Y, Arisue N, Kawai S, Escalante AA, Horii T, Tanabe K, Hashimoto T. Evolution and phylogeny of the heterogeneous cytosolic SSU rRNA genes in the genus *Plasmodium*. *Mol Phyl Evol* 2008;47:45–53.
- Ollomo B, Durand P, Prugnolle F, Douzery E, Arnathau C, Nkoge D, Leroy E, Renaud F. A new malaria agent in African hominids. *PLoS Pathogens* 2009;5(5):e1000446. [PubMed: 19478877]
- Prugnolle F, Durand P, Neel C, Ollomo B, Ayala FJ, Arnathau C, Etienne L, Mpoudi-Ngole E, Nkoghe D, Leroy E, Delaporte E, Peeters M, Renaud F. African great apes are natural hosts of multiple related malaria species, including *Plasmodium falciparum*. *Proc Natl Acad Sci USA* 2010;107:1458–1463. [PubMed: 20133889]
- Rich SM, Licht MC, Hudson RR, Ayala FJ. Malaria's Eve: evidence of a recent bottleneck throughout the world populations of *Plasmodium falciparum*. *Proc Natl Acad Sci USA* 1998;95:4425–4430. [PubMed: 9539753]
- Rich SM, Leendertz FH, Xu G, LeBreton M, Djoko CF, Aminake MN, Takang EE, Diffio JL, Pike BL, Rosenthal BM, Formenty P, Boesch C, Ayala FJ, Wolfe ND. The origin of malignant malaria. *Proc Natl Acad Sci USA* 2009;106:14902–14907. [PubMed: 19666593]
- Rzhetsky A, Nei M. A simple method for estimating and testing minimum-evolution trees. *Mol Biol Evol* 1992;9:945–967.
- Tamura K, Nei M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol* 1993;10:512–526. [PubMed: 8336541]

- Tamura K, Nei M, Kumar S. Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proc Natl Acad Sci USA* 2004;101:11030–11035. [PubMed: 15258291]
- Tamura K, Dudley J, Nei M, Kumar S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* 2007;24:1596–1599. [PubMed: 17488738]
- Tanabe K, Sakihama N, Hattori T, Ranford-Cartwright L, Goldman I, Escalante AA, Lal AA. Genetic distance in housekeeping genes between *Plasmodium falciparum* and *Plasmodium reichenowi* and within *P. falciparum*. *J Mol Evol* 2004;59:687–694. [PubMed: 15693624]
- Volkman SK, Barry AE, Lyons EJ, Nielsen KM, Thomas SM, Choi M, Thakore SS, Day KP, Wirth DF, Hartl DL. Recent origin of *Plasmodium falciparum* from a single progenitor. *Science* 2001;293:482–484. [PubMed: 11463913]
- Volkman SK, Sabeti PC, DeCaprio D, Neafsey DE, Shaffner SF, Milner DA Jr, Daily JP, Sarr O, Ndiaye D, Ndir O, Mboup S, Duraisingh MT, Lukens A, Derr A, Stange-Thomann N, Waggoner S, Onofrio R, Ziaugra L, Mauceli E, Gnerre S, Jaffe DB, Zainoun J, Wiegand RC, Birren BW, Hartl DL, Galagan JE, Lander ES, Wirth DF. A genome-wide map of diversity in *Plasmodium falciparum*. *Nature Genet* 2007;39:113–119. [PubMed: 17159979]
- Waters AP, Higgins DG, McCutchan TF. *Plasmodium falciparum* appears to have arisen as a result of lateral transfer between avian and human hosts. *Proc Natl Acad Sci USA* 1991;88:3140–3144. [PubMed: 2014232]
- Waters AP, Higgins DG, McCutchan TF. Evolutionary relatedness of some primate models of *Plasmodium*. *Mol Biol Evol* 1993;10:914–923. [PubMed: 7689135]
- White TD, Asfaw B, Beyene Y, Haile-Selassie Y, Lovejoy CO, Suwa G, WoldeGabriel G. *Ardipithecus ramidus* and the paleobiology of early hominids. *Science* 2009;326:75–86. [PubMed: 19810190]
- Yang Z, Nielsen R. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Mol Biol Evol* 2000;17:32–43. [PubMed: 10666704]

Figure 1.

ME tree of *Plasmodium cytb* sequences (477 aligned sites) based on the MCL distance. Clusters of *P. falciparum* and *P. vivax* sequences are condensed for ease of presentation; for accession numbers see Supplementary Table S1. Numbers on the branches are confidence levels of the standard-error test; only values $\geq 90\%$ are shown.

Figure 2.

ME tree of *Plasmodium cytb* sequences (477 aligned sites) based on the TN distance. Clusters of *P. falciparum* and *P. vivax* sequences are condensed for ease of presentation; for accession numbers see Supplementary Table S1. Numbers on the branches are confidence levels of the standard-error test; only values $\geq 90\%$ are shown.

Figure 3.

ME tree of *Plasmodium cytb* sequences, including those of Prugnolle et al. (2010; 349 aligned sites) based on the MCL distance. Clusters of *P. falciparum* and *P. vivax* sequences are condensed for ease of presentation; for accession numbers see Supplementary Table S1. Numbers on the branches are confidence levels of the standard-error test; only values $\geq 90\%$ are shown.

Figure 4. ME tree of *Plasmodium clpC* sequences based on the MCL distance. Numbers on the branches are confidence levels of the standard-error test; only values $\geq 90\%$ are shown.

Figure 5. ME tree of *Plasmodium* 18S rRNA sequences based on the MCL distance. Numbers on the branches are confidence levels of the standard-error test; only values $\geq 90\%$ are shown.

Table 1

Synonymous (π_S) and nonsynonymous (π_N) nucleotide diversity in *Plasmodium cytb* genes.

Source of sequences	N	$\pi_S \pm \text{S.E.}$	$\pi_N \pm \text{S.E.}$
<i>P. falciparum</i>	121	0.00337 $\pm$ 0.00170	0.00057 $\pm$ 0.000044
<i>P. vivax</i>	285	0.00046 $\pm$ 0.00021	0.00005 $\pm$ 0.00004
Chimp-derived			
All	10	0.21973 $\pm$ 0.03843 ^{***}	0.01110 $\pm$ 0.00414 [*]
Clade A (Figures 1 and 2)	4	0.03765 $\pm$ 0.02036	0.00456 $\pm$ 0.00251
Clade B (including CDC1; Figure 1)	6	0.10695 $\pm$ 0.02709 ^{***}	0.00512 $\pm$ 0.00224 [*]
Clade B (excluding CDC1; Figure 2)	5	0.07598 $\pm$ 0.02380 ^{**}	0.00614 $\pm$ 0.00269 [*]

Z-tests of the hypothesis that π_S or π_N equals the corresponding value for *P. falciparum*:

* $P < 0.05$;

** $P < 0.01$;

*** $P < 0.001$.

Table 2

Mean number of synonymous substitutions per site (d_S) in *Plasmodium falciparum* and chimp-derived coding sequences and estimated rates of synonymous substitutions on the basis of different divergence times.

Gene	Comparison	$d_S \pm \text{S.E.}$	Divergence time (years)	Rate (substitutions/site/year)
cytb	<i>P. falciparum</i> vs. <i>P. reichenowi</i> CDC1	0.1899 ± 0.0576	5,000	1.89×10^{-5}
			50,000	1.89×10^{-6}
			200,000	4.75×10^{-7}
			5,000,000	1.89×10^{-8}
			7,000,000	1.36×10^{-8}
	<i>P. falciparum</i> vs. Chimp-derived B clade (excluding CDC1 (Figure 2)	0.2703 ± 0.0643	5,000	2.71×10^{-5}
			50,000	2.71×10^{-6}
			200,000	6.77×10^{-7}
			5,000,000	2.71×10^{-8}
			7,000,000	1.93×10^{-8}
	<i>P. falciparum</i> vs. " <i>P. reichenowi</i> " A clade (Figure 2)	0.4053 ± 0.0947	5,000	4.05×10^{-5}
			50,000	4.05×10^{-6}
			200,000	1.01×10^{-6}
			5,000,000	4.05×10^{-8}
			7,000,000	2.90×10^{-8}
			10,000,000	2.03×10^{-8}
			12,000,000	1.69×10^{-8}
	<i>P. falciparum</i> MRCA	0.0074 ± 0.0055	5,000	7.44×10^{-7}
			50,000	7.44×10^{-8}
			200,000	1.86×10^{-8}
clpC	<i>P. falciparum</i> vs. Chimp-derived D clade (Figure 3)	0.1279 ± 0.0704	5,000	1.28×10^{-5}
			50,000	1.28×10^{-6}
			200,000	3.20×10^{-7}
			5,000,000	1.28×10^{-8}
			7,000,000	9.14×10^{-9}
	<i>P. falciparum</i> vs. Chimp-derived C clade (Figure 3)	0.2620 ± 0.0984	5,000	2.62×10^{-5}
			50,000	2.62×10^{-6}
			200,000	6.55×10^{-6}
			5,000,000	2.62×10^{-8}
			7,000,000	1.87×10^{-8}
			10,000,000	1.31×10^{-8}
			12,000,000	1.09×10^{-8}