NIH Public Access
Author Manuscript
Parasitology. Author manuscript; available in PMC 2011 November 1.

Published in final edited form as:
Parasitology. 2011 November ; 138(13): 1737–1749. doi:10.1017/S0031182010001575.

Genome sequences reveal divergence times of malaria parasite lineages

JOANA C. SILVA[1,2], **AMY EGAN**[2], **ROBERT FRIEDMAN**[3], **JAMES B. MUNRO**[1,2], **JANE M. CARLTON**[4], and **AUSTIN L. HUGHES**[3,*]

[1] Department of Microbiology and Immunology, University of Maryland School of Medicine, Baltimore, MD 21201, USA

[2] Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, MD 21201, USA

[3] Department of Biological Sciences, University of South Carolina, Columbia, SC 29208, USA

[4] Department of Medical Parasitology, New York University School of Medicine, New York, NY 10011, USA

SUMMARY

Objective—The evolutionary history of human malaria parasites (genus *Plasmodium*) has long been a subject of speculation and controversy. The complete genome sequences of the two most widespread human malaria parasites, *P. falciparum* and *P. vivax*, and of the monkey parasite *P. knowlesi* are now available, together with the draft genomes of the chimpanzee parasite *P. reichenowi*, three rodent parasites, *P. yoelii yoelli, P. berghei* and *P. chabaudi chabaudi*, and one avian parasite, *P. gallinaceum*.

Methods—We present here an analysis of 45 orthologous gene sequences across the eight species that resolves the relationships of major *Plasmodium* lineages, and provides the first comprehensive dating of the age of those groups.

Results—Our analyses support the hypothesis that the last common ancestor of *P. falciparum* and the chimpanzee parasite *P. reichenowi* occurred around the time of the human-chimpanzee divergence. *P. falciparum* infections of African apes are most likely derived from humans and not the other way around. On the other hand, *P. vivax*, split from the monkey parasite *P. knowlesi* in the much more distant past, during the time that encompasses the separation of the Great Apes and Old World Monkeys.

Conclusion—The results support an ancient association between malaria parasites and their primate hosts, including humans.

INTRODUCTION

An accurate account of the evolutionary history of parasite species is essential to understand the acquisition of novel parasite life-history traits and the emergence of new human diseases (Lefevre *et al.* 2007; Wolfe *et al.* 2007). Despite the fact that malaria is one of the most devastating infectious diseases currently affecting the human species (Hotez *et al.* 2006), key aspects in the phylogeny of the *Plasmodium* genus remain elusive (Waters *et al.* 1991; Escalante and Ayala, 1994; McCutchan *et al.* 1996; Qari *et al.* 1996; Escalante *et al.* 1998, 2005; Perkins and Schall, 2002; Martinsen *et al.* 2008; Garamszegi, 2009), and a

*Corresponding author: Tel: +1-803-777-9186. austin@biol.sc.edu.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

comprehensive assessment of the age of the major *Plasmodium* lineages infecting mammals has never been attempted.

P. falciparum and *P. vivax* are the two most important malaria parasites of humans, together accounting for several hundred million cases per year (Kappe *et al.* 2010). Human infections with *P. knowlesi*, a simian parasite previously reported to infect humans occasionally, may be increasing as a result of forest clearing in Southeast Asia (Cox-Singh *et al*. 2008). Although *P. vivax* is known to be more closely related to *P. knowlesi* than to *P. falciparum*, the apparent absence of a correlation with host divergence has hampered attempts to estimate the *P. vivax*–*P. knowlesi* divergence time. Estimates based on the mitochondrial genome (which, in *Plasmodium*, contains only three protein-coding genes) and on small numbers of nuclear genes placed the *P. vivax*–*P. knowlesi* divergence time within the past 7 million years (My) (Escalante *et al*. 2005; Jongwutiwes *et al*. 2005), leading to the hypothesis that *P. vivax* has arisen from a macaque monkey parasite through a host switch event in Southeast Asia.

P. falciparum, by contrast, shows evidence of a close relationship to *P. reichenowi*, a parasite of the common chimpanzee, supporting an African origin for the former species (Escalante and Ayala, 1994; Hughes and Hughes, 1995; McCutchan *et al.* 1996; Qari *et al*. 1996; Escalante *et al*. 1998; Perkins and Schall, 2002; Martinsen *et al.* 2008). Thus, it has frequently been assumed that these two *Plasmodium* species co-speciated with their mammalian hosts and thus diverged 5–7 million years ago (Mya) (Escalante and Ayala, 1994), the divergence time of human and chimpanzee lineages (Glazko and Nei, 2003). Recent sequencing of mitochondrial genes from wild populations of African apes has complicated this picture by showing that African apes are infected by a number of *Plasmodium* taxa related to *P. falciparum* and *P. reichenowi* (Ollomo *et al.* 2009; Rich *et al.* 2009; Krief *et al.* 2010; Prugnolle *et al*. 2010). Rich *et al*. (2009) proposed that *P. falciparum* and *P. reichenowi* diverged very recently, possibly as recently as 10,000 years ago. However, it has been argued that the degree of sequence divergence at mitochondrial and apicoplast loci between *P. falciparum* and *P. reichenowi* is too great to be consistent with a very recent date for the divergence of these two species (Hughes and Verra, 2010; Ricklefs and Outlaw, 2010), but the latter conclusion was based on a small number of sequences.

Krief *et al.* (2010) reported mitochondrial genomic sequences from the bonobo (*Pan paniscus*) that appear to belong to *P. falciparum*. On this basis, they proposed that *P. falciparum* arose in bonobos and subsequently colonized humans by a host switch. On this scenario, it might be supposed that the divergence of *P. falciparum* from *P. reichenowi* occurred at the time of the divergence of the bonobo from the common chimpanzee (about 1·3 Mya; Caswell *et al.* 2008). An alternative hypothesis remains that *P. falciparum* infecting bonobos originated from humans, rather than the other way around (Hughes and Verra, 2010). Similarly, Liu *et al.* (2010) argued that *P. falciparum* was recently transferred from gorillas (*Gorilla gorilla*) to humans; but these authors did not attempt to rule out the alternative hypothesis that gorilla infections with *P. falciparum* originated from humans.

The three most studied rodent-infecting malaria parasite species form a monophyletic clade in phylogenetic analyses, but the phylogenetic position of this group in the *Plasmodium* tree is not consistent across analyses (Fig. 1; Waters *et al.* 1991;McCutchan *et al.* 1996;Perkins and Schall, 2002). Finally, the relationship of the bird malaria parasite *P. gallinaceum* to *Plasmodium* species infecting mammals, has been the source of repeated revision and controversy (Fig. 1;Waters *et al*. 1991;Siddall and Barta, 1992;McCutchan *et al.* 1996;Qari *et al.* 1996;Perkins and Schall, 2002;Roy and Irimia, 2008). The ability to address these phylogenetic questions has been hampered by the small number of loci sampled and the lack

NIH-PA Author Manuscript

of a suitable outgroup. Further increasing the difficulty of phylogenetic reconstruction are the large evolutionary distances involved and the strong bias in nucleotide composition of some *Plasmodium* nuclear genomes, which are conducive to homoplasy (Dávalos and Perkins, 2008). On the other hand, recent analyses of rare changes in mitochondrial genomes strongly support the monophyly of *Plasmodium* species infecting mammals (Roy and Irimia, 2008).

Based on the genomic sequence data now available for eight *Plasmodium* species, we generated a set of 45 highly conserved, single-copy orthologous nuclear genes containing sequences from all species, for a total of ~15,400 aligned amino acid residues. Here we use this data-set to estimate divergence times in the genus *Plasmodium* and a subset to infer a phylogeny. We also use data on polymorphism at these loci within *P. falciparum* to estimate the timing of the most recent common ancestor (MRCA) of *P. falciparum* relative to the divergence of *P. falciparum* and *P. reichenowi*. These analyses thus provide a further test of the hypothesis that *P. falciparum* and *P. reichenowi* co-speciated with their hosts, which in turn provides a calibration for estimates of the divergence times of other *Plasmodium* taxa. In addition, we analyze sequences of complete mitochondrial genomes in order to test the hypothesis that *P. falciparum* originated from a parasite of bonobos (Krief *et al.* 2010).

MATERIALS AND METHODS

Genomic sequence data

We obtained from PlasmoDB5.4 the genomic sequence data for six *Plasmodium* species: *P. gallinaceum, P. falciparum, P. reichenowi, P. vivax, P. knowlesi, P. berghei, P. chabaudi* and *P. yoelii*. We also downloaded the accompanying GFF files describing gene, protein coding and exon features. In addition we obtained the draft genome sequences for *P. gallinaceum* and *P. reichenowi*, with permission, from the WT Sanger Institute (http://www.sanger.cak.uk/resources/downloads/protozoa). A set of 1285 high-quality orthologous clusters was determined for *P. vivax, P. knowlesi, P. falciparum* and *P. yoelii*, which included exactly one copy in each of the species (i.e. no paralogs were detected at a Jacquard filter cut-off of 0·6). For each of these genes, the *P. falciparum* and *P. yoelii* sequence lengths were within 10% of each other; and the same was true for *P. vivax* and *P. yoelii*. Orthologs from *P. gallinaceum* and *P. reichenowi* were obtained by BLAST homology search of their respective partial genome shotgun sequence using the *P. falciparum* amino acid or nucleotide sequences as queries, respectively. A similar procedure was used to find orthologs from *P. berghei* and *P. chabaudi*, using *P. yoelii* sequences as queries. A total of 45 genes were obtained from all species that satisfied the following criteria: (1) each gene was a complete single-exon gene; (2) the BLAST alignment encompassed >97% of the *P. falciparum* or *P. yoelii* sequence with an E value $<1\times10^{-30}$); and (3) no paralogs of these genes were found in any of the species. By homology search we identified full-length orthologs of 29 of the 45 genes in the genomes of *Theileria parva* and *T. annulata*; the subset of 29 genes (Supplementary Table S1; Supplementary Fig. S1, available at http://journals.cambridge.org/PAR) was used in phylogenetic analyses, so that the *Theileria* species could be used to root the *Plasmodium* in-group.

Phylogenetic analyses

The sequences for each locus were aligned with ClustalW (Thompson *et al.* 1994), using default parameters, followed by manual inspection. Alignments were usually straightforward as a consequence of our method of choosing of orthologs (see above), which selected for genes that were highly conserved in both sequence and length. The draft nature of several of the assemblies led to the presence of certain problematic sequences, such as those in which indels interrupted the open reading frame (ORF). Loci with such problematic sequences

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

were generally discarded, except in cases where the source of error could be easily identified and corrected. For example, in some cases polynucleotide runs disrupted the ORF in draft genome assembles, but these were assumed to be sequencing or assembly errors and were trimmed to preserve the ORF.

The concatenated amino acid sequence alignments of the 29 genes (Supplementary Fig. S1) were used to reconstruct the species phylogeny using the following methods: (1) Neighbor-joining (NJ) implemented in MEGA V4.1 (Tamura *et al.* 2007) using the JTT distance matrix and gamma-distributed rates with the shape parameter α=0·6, as estimated by ProTest V10.2 (Abascal *et al.* 2005); (2) Bayesian, using both the MrBayes V3.12 (Ronquist and Huelsenbeck, 2003) and Beast V1.5.4 (Drummond and Rambaut, 2007) programs; and (3) maximum likelihood (ML), using RAxML V7.2.6 (Stamatakis, 2006). In the NJ and ML analyses, bootstrapping (Felsenstein, 1985) was used to test support for branching patterns; 5,000 and 1,000 bootstrap samples were used, respectively. The ML and Bayesian analyses allowed for partitioning of the 29 genes, while the NJ data-set was unpartitioned.

We conducted multiple preliminary runs of MrBayes via the CIPRES Portal V2.2 (Miller *et al.* 2010) and the Computational Biology Service Unit from Cornell University, to optimize the temperature (0·05) and mixing of chains (86 to 87%). Three analyses consisting of 4 runs, each with 6 chains, were executed for 10, 12·5, and 15 million generations on the Texas A&M University Brazos HPC cluster. All parameters for the 29 gene partitions were unlinked, and MrBayes was allowed to estimate all parameters. Burn-in was set to 25% of the trees sampled, and convergence was assessed by evaluating (1) average standard deviation of split frequencies (0·006583, 0·008690 and 0·00601, respectively); (2) the –Ln cold-chain score of the four runs; (3) the Potential Scale Reduction Factor for the branch lengths (~1·0); and (4) the slide and compare commands of AWTY (Wilgenbusch *et al.* 2004). In all three analyses, posterior probability of amino acid models was 1·00 (SD =0·000) for the JTT model and 0·00 (SD =0·000) for all other models.

Seven runs of BEAST were launched, one of which ran for 15 million generations, and another three each for 20 and 30 million generations. The runs were based on the concatenated amino acid alignment, with the following parameters: JTT substitution matrix, with gamma-distributed rates (modeled by 6 classes) with alpha and the proportion of invariant sites estimated from the data. We used an uncorrelated relaxed lognormal clock, with the mean amino acid substitution rate estimated from the data, uniformly sampled between 0–2%/MY. Tree priors used include a Yule process, the monophyly of *P. falciparum* and *P. reichenowi* with divergence time uniformly distributed between 5–7 MY, and the divergence time between *Plasmodium* and *Theileria* uniformly sampled from the interval 50–500 MY. Appropriate mixing values for operators were determined based on several preliminary runs. Convergence was monitored with Tracer v1.5 (http://tree.bio.ed.ac.uk/software/tracer/). When appropriate, log and tree files were combined using LogCombiner v1.5.4 (distributed with BEAST), after removal of 2000 samples (2 million steps) from each run. Convergence was measured by the value of the effective sample size (ESS).

An initial test analysis using the frequency-based criterion to allow RaxML to halt automatically the number of boostraps (BS) showed that 550 BS were adequate. An analysis of 1000 BS, followed by ML optimization to find the best-scoring tree was conducted for the partitioned data-set, which applied the LG model (the best-fit model for the majority of genes suggested by ProtTest (Abascal *et al.* 2005), estimated all free model parameters with the gamma model of rate heterogeneity (alpha=0·9065), and applied empirical base frequencies to all 29 partitions.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

P. falciparum–P. reichenowi calibration

In order to test the hypothesis that *P. falciparum* and *P. reichenowi* co-speciated with their hosts, we compared polymorphism within *P. falciparum* at the 45 loci with divergence between *P. falciparum* and *P. reichenowi*. Polymorphism data for the 45 loci in *P. falciparum* was obtained with permission from the *Plasmodium falciparum* Database at the Broad Institute (http://www.broadinstitute.org/annotation/genome/plasmodium_falciparum_spp/GenomesIndex.html). In order to obtain a high quality sequence data set, we excluded isolates from which the available orthologous sequence was substantially less than full-length or included one or more frameshifts. For the 45 genes, the mean number of isolates present in our data set was 6·0 (±0·2 S.E; median=6·0; range= 3–12).

We computed the number of synonymous substitutions per synonymous site (d_S) and the number of non-synonymous substitutions per non-synonymous site (d_N) by Nei and Gojobori's (1986) method. We used this simple method to analyze the data from *P. falciparum* and *P. reichenowi* because the numbers of substitutions per site were few. When the numbers of substitutions per site are few, this method and more complicated methods yield essentially identical results (Hughes and French, 2007); yet the variance of the simpler method is lower (Nei and Kumar, 2000). Within *P. falciparum*, we estimated the synonymous nucleotide diversity (mean of all pairwise d_S values, designated π_S) and the non-synonymous nucleotide diversity (mean of all pairwise d_N values, designated π_N). In order to estimate the age of the most recent common ancestor (MRCA) of *P. falciparum*, we used the maximum d_S value for each gene. In order to compute the rate of synonymous substitution under different assumptions for the divergence time of *P. falciparum* from *P. reichenowi*, we computed mean d_S between *P. reichenowi* and *P. falciparum* sequences for each gene.

As a further test of this calibration, we conducted a phylogenetic analysis of complete or nearly complete mitochondrial genomes (4695 aligned nucleotide sites) from *P. falciparum* isolated from bonobos (Krief *et al.* 2010) along with *P. falciparum* sequences isolated from humans. The phylogenetic tree was reconstructed by the NJ method based on the maximum composite likelihood (MCL) distance (Tamura *et al.* 2007). The data-set included 4 isolates of *P. falciparum* from bonobos, 104 isolates of *P. falcparum* from humans, two isolates of *P. reichenowi*, 3 isolates assigned to *P. billbrayi*, and 3 isolates assigned to *P. billcollinsi* (Supplementary Table S2, available at http://journals.cambridge.org/PAR). 99 of the human isolates were mostly from Joy *et al.* (2003) and others were from additional Genbank accessions. 30 isolates were of African origin, 31 from Southeast Asia, 11 from Papua New Guinea, 30 from Central and South America, and 2 of unknown origin.

Other divergence times

Other divergence times were estimated by a number of methods; in each case, a time of 5–7 Mya for the *P. falciparum–P. reichenowi* divergence was used as a calibration. The following methods were used: (1) The linearized tree method (Takezaki *et al.* 1995) as implemented in the MEGA 4.1 program (Tamura *et al.* 2007); (2) the non-parametric rate smoothing method of Sanderson (1997), as implemented in the r8s program, version 1.71; (3) a linear regression method (Hughes and Nei, 1990) applied to d_S and d_N values; (4) a Bayesian method, with the posterior distribution of node ages approximated by MCMC simulations, as implemented in BEAST V1.5.4 (Drummond and Rambaut, 2007). In the regression method, the divergence time between species *A* and *B* was estimated by regressing d_S or d_N between the two species against the corresponding value for the orthologous comparison between *P. falciparum* and *P. reichenowi*. The regression line was forced through the origin, and the slope of the line used to estimate the ratio of the species

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

A-species B divergence time to the *P. falciparum–P. reichenowi* divergence time. Estimations based on d_S were only done for close relatives because synonymous sites are saturated for all other branches. In these analyses, because more distantly related sequences were involved, d_S and d_N were estimated by the method of Yang and Nielsen (2000) implemented in PAML (Yang, 2007), which takes into account nucleotide composition and mutational biases. PAML analyses were launched, monitored and results visualized, using IDEA (Egan *et al.* 2008). These biases can have a substantial effect on estimates of d_S and d_N when the latter are large (Nei and Kumar, 2000).

RESULTS

Phylogenetic analyses

The NJ method with JTT distance recovered the topology illustrated in Fig. 1a; that is a tree in which the bird parasite *P. gallinaceum* clustered with the primate parasites *P. falciparum* and *P. reichenowi.* NJ analyses using other distance methods and joining algorithms yielded the same result (not shown). The same tree was also obtained as the majority rule consensus of a bootstrap analysis, using JTT distances and 5000 replicates. However, the node establishing this topology (node 1 in Fig. 2a) received only 74% bootstrap support. The node clustering *P. falciparum, P. reichenowi*, and *P. gallinaceum* with *P. vivax* and *P. knowlesi* (node 2 in Fig. 2a) received only 58% bootstrap support. On the other hand, bootstrap support for all other nodes in the tree was 100%.

Bayesian analyses using MrBayes and BEAST, as well as the RAxML analyses, recovered the topology in Fig. 2b. With MrBayes, the posterior probability for all tree nodes was 1·00 (SD =0·000; PSRF of branch lengths=1·0), except for nodes 7 and 8, each of which had posterior probability=0·97 (SD =0·023, 0·030, and 0·021, respectively; PSRF of branch length=1·00). Similarly, bootstrap node support was 100% for all nodes save 7 and 8, which were 53% and 48%, respectively. Likewise, in the BEAST analyses the tree topology represented in the posterior sample is that shown in Fig. 2b. Five of the seven runs converged to this tree (ESS value for each individual run>2000), with log likelihood score of −86,630. The posterior, prior and parameter values were all identical for these five runs, which were then combined (posterior ESS>500). The divergence between the *Plasmodium* clade and the *Theileria* outgroup was estimated at 294 My (ESS~295). However, two runs of BEAST (with 15 million and 20 million generations, respectively), stabilized in a different region of tree and parameter space, with a topology corresponding to that of Fig. 2a. Posterior, prior and parameter values in these two runs were also identical to each other, and the two runs were combined. This tree had a slightly lower posterior log likelihood score (−86,636; ESS>2000) than that found in the other five runs. This topology places the divergence between *Plasmodium* and *Theileria* at an earlier date of 314 Mya (ESS~111).

These results exemplify the mixed signal in *Plasmodium* data that has led to much of the discussion regarding the position of bird and reptile malaria parasites relative to that of mammalian parasites (Dávalos and Perkins, 2008). Amino acid sequence convergence has taken place among the *Plasmodium* species with the highest AT nucleotide content in a degree sufficient to confound distance methods such as NJ, and to create a sub-optimal peak in the likelihood surface that is captured by a small number of runs of BEAST. In addition, our method of identifying orthologs may have preferentially selected genes from *P. gallinaceum* with high sequence similarity to *P. falciparum*, thus magnifying the problem. Nonetheless, both ML and Bayesian methods, those that are probably most impervious to the effects of sequence convergence, mostly capture the topology supported by rare mutational events, such as indels (Roy and Irimia, 2008).

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

P. falciparum–P. reichenowi calibration

Mean π_S in *P. falciparum* for the 45 loci analyzed here was 0·0007±0·0002 (Table 1); this value is comparable to estimates of synonymous nucleotide diversity in humans (Aquadro *et al.* 2001;Li and Sadler, 1991; The International SNP Map Working Group, 2001). Though somewhat lower than some early estimates of nucleotide diversity in *P. falciparum* (Hughes and Verra, 2001,2002), this value is very similar to recent estimates based on genome-wide SNPs (Volkman *et al.* 2007).

The ratio of the maximum d_S value within *P. falciparum* (0·0029±0·0009; Table 1) to mean d_S between *P. falciparum* and *P. reichenowi* (0·0595± 0·0056; Table 1) was 0·0487. Using this ratio, we estimated the MRCA of *P. falciparum* based on different hypotheses for the time of the *P. falciparum–P. reichenowi* divergence (Table 2). If the latter divergence is placed at 4–7 Mya, the MRCA of *P. falciparum* is around 200,000–300,000 years. This is consistent with previous estimates of the MRCA of *P. falciparum* (Hughes and Verra, 2001,2002), and it would place the MRCA of *P. falciparum* around the time of the origin of modern humans (Tamura and Nei, 1993). A very recent divergence of *P. falciparum* and *P. reichenowi* (5,000–50,000 years ago) yielded estimates of the MRCA of *P. falciparum* that seem implausible (Table 2).

When the synonymous substitution rate was estimated on the basis of different divergence times for *P. falciparum* and *P. reichenowi*, recent times (2 Mya or less) yielded very high estimates of the synonymous substitution rate, which were more comparable to those of DNA viruses (10^{-6} to 10^{-8} substitutions/site/year; Table 2) than to those known from eukaryotes (Li, 1997). On the other hand, a divergence time between 5 and 7 Mya, which is consistent with the co-speciation hypothesis, yielded rate estimates (4 to 6×10^{-9} substitutions/site/year; Table 2), which are consistent with estimates from eukaryotic groups having good fossil records, including vertebrates (Li, 1997) and diatoms (Sorhannus and Fox, 1999), the latter being the only protists with a substantial fossil record. We therefore used the calibration of 5–7 Mya between *P. falciparum* and *P. reichenowi* for estimation of divergence times of other *Plasmodium* taxa.

Mitochondrial genomes

A neighbor-joining tree (Fig. 3) was constructed from mitochondrial genome sequences of *P. falciparum* from humans and bonobos, along with sequences of *P. reichenowi* and others attributed to *P. billcollinsi* and *P. billbrayi* by Krief *et al.* (2010). The four bonobo-derived sequences all fell within the cluster of human-derived *P. falciparum* sequences. This topology did not support the hypothesis that human infection with *P. falciparum* results from an ancient host transfer event from bonobos (Krief *et al.* 2010). Rather, the topology is most consistent with the hypothesis that bonobo infections by *P. falciparum* are derived from the *P. falciparum* population infecting humans.

This interpretation is further supported by analysis of the pattern of nucleotide sequence difference among these mitochondrial genomes (Supplementary Table S2, available at http://journals.cambridge.org/PAR). Although none of the sequences originating from bonobos was identical to any sequence yet recorded from humans, each was less than 1% different at the nucleotide level from every *P. falciparum* sequence so far found in humans; and none of the bonobo-derived sequences was more than 0·34% different from the closest human sequence. Moreover, each bonobo-derived *P. falciparum* sequence showed greater sequence identity to certain human-derived sequences than it did to any other bonobo-derived sequence. For example, GQ355475 from bonobo was only 0·19% different from 6 human-derived sequences, but the closest other bonobo-derived sequence (GQ355472) was 0·38% different. The latter bonobo-derived sequence was only 0·17% different from a total

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

of 23 different human-derived sequences, while it was no closer to any other bonobo-derived sequence than to GQ355475 (0·38% different). Likewise, the bonobo-derived sequence GQ355474 was 0·34% different from 23 human-derived sequences, but 0·51% different from the closest other bonobo-derived sequence; and the bonobo-derived sequence GQ355473 was 0·23% different from 27 human-derived sequences, but 0·40% different from the closest other bonobo-derived sequence. The nucleotide diversity of the four bonobo-derived sequences (0·00474±0·00075) was significantly greater than that of the 104 human derived sequences (0·00049±0·00015; Z-test; $P<0·001$). However, only 30 of the 104 available human sequences (28·8%) were from Africa.

Other divergence time estimates

We applied a variety of methods to estimate the age of strongly supported nodes on the phylogeny using the 5–7 My divergence between *P. falciparum* and *P. reichenowi* as a calibration point (Table 3; node 3 in Fig. 2b). Although divergence time estimates varied among methods (Table 3), there was broad agreement on several main points. All methods placed the time of the *P. vivax–P. knowlesi* divergence between 15 and 46 Mya (Table 3; node 4 in Fig. 2b), a period corresponding to the early radiation of the Old World primate lineage. Estimates of the divergence of New World primates (Platyrrhini) and Old World primates (Catarrhini) range between 37 and 54 Mya, while those for the divergence of the Old World monkeys (Cercopithecoidea) from the great ape/human lineage (Hominoidea) range between 26 and 30 Mya (Steiper and Young, 2009).

Because estimates based on mitochondrial genomes yielded a much earlier estimate of the divergence time between *P. vivax* and *P. knowlesi* (Jongwutiwes *et al.* 2005), we examined potential alternative explanations for our results. One hypothesis might be that *P. vivax* and *P. knowlesi* diverged from each other recently and that the differences between the two species at nuclear loci have increased due to an unusually high mutation rate. If this was the case, and since the effects of a high mutation rate should be seen more strongly at synonymous than at non-synonymous sites (because of purifying selection on the latter), one would expect the mean value of d_N/d_S between *P. falciparum* and *P. reichenowi* to be significantly higher than that between *P. vivax* and *P. knowlesi*. Nevertheless, the two ratios are quite similar, with the former ratio being 0·0612 and the latter 0·0616 (Supplementary Table S3, available at http://journals.cambridge.org/PAR).

All methods of divergence time estimation supported the conclusion that the diversification of the three rodent malaria parasites (node 6 in Fig. 1b) has taken place within the last 36 My at most, and the regression and MCMC methods placed the divergences of these species within the past 25 My (Table 3). The known natural hosts of these rodent malaria parasites are primarily members of the genus *Thamnomys* in the rodent family Muridae (Killick-Kendrick, 1968;Yoeli and Most, 1964), and the origin of the Muridae has been estimated at about 24 Mya (Honeycutt, 2009).

DISCUSSION

The origin and evolutionary relationships of malaria parasites have been controversial for decades. The present analyses suggest that, although many questions remain unresolved, the overall pattern within *Plasmodium* parasites of mammals has been one where parasites and hosts lineages have co-evolved, with no evidence of frequent transfer between hosts across wide taxonomic distances (such as between different mammalian orders). However, within groups of closely related host species, transfers across species boundaries may be fairly common, as is seen in field infections of Southeast Asia primates (Putaporntip *et al.* 2010) and as inferred to have occurred over evolutionary time in Southeast Asia (Garamszegi, 2009).

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Recently, some authors have revived the old idea that *P. falciparum* is a new human parasite on an evolutionary time scale (Rich *et al.* 2009). This hypothesis was originally based on the estimations that the sickle-cell gene arose in the human population of West Africa about 5,000 years ago, around the time that agriculture appeared in that region; and it was presumed that *P. falciparum* malaria must have first infected humans around that time (Livingstone, 1958). The inference was thus based on a misunderstanding of the evolutionary process, since it was assumed that an adaptive mutation (the sickle-cell gene) must have occurred in response to the selection imposed by the parasite. Mutation is a random event, and an adaptive mutation need not coincide in time with the selective pressure to which it represents an adaptive response. However, if agriculture enabled an increase in the human population in West Africa, it would have increased the probability that an adaptive mutation would occur on at least one chromosome in the population, as well as the probability that it would be able to increase in response to selection rather than being lost to drift (Hughes and Verra, 2010).

The present results from 45 nuclear genes showed that the assumption of a recent divergence of *P. falciparum* and *P. reichenowi* requires the assumption of both (1) an implausibly high synonymous substitution rate, and (2) an implausibly recent MRCA of worldwide *P. falciparum*. Likewise, the divergence of *P. falciparum* from *P. reichenowi* around the time of the divergence of the bonobo from the common chimpanzee (about 1·3 Mya) would require a nuclear synonymous substitution rate (Table 2) over 4 times that of mammals (4–5×10–9; Li, 1997), which seems unlikely. These results are thus consistent with earlier analyses of mitochondrial and apicoplast sequences (Hughes and Verra, 2010) and are consistent with the hypothesis of co-speciation of *P. falciparum* and *P. reichenowi* with their respective human and chimpanzee hosts. These results are also consistent with earlier results showing that *P. falciparum* has substantial genetic diversity and thus an ancient MRCA, probably around the time of the origin of modern *Homo sapiens* 100,000–300,000 years ago (Hughes and Verra, 2001,2002;Mu *et al.* 2002;Volkman *et al.* 2007). The estimation of the MRCA of *P. falciparum* in humans at 100,000 years ago or earlier is in turn consistent with evidence that this parasite accompanied modern humans in their first emergence from Africa to Asia (Tanabe *et al.* 2010).

Recent PCR-based studies of malaria parasites from natural populations of non-human primates have revealed surprising diversity, suggesting that there may be numerous undescribed species infecting primates in both Southeast Asia (Putaporntip *et al.* 2010) and Africa (Duval *et al.* 2010; Prugnolle *et al.* 2010). The relationships and divergence times of these complex assemblages are far from being resolved. Because *P. falciparum* sequences have been found in African apes, Duval *et al.* (2010) suggested that chimpanzee and gorillas constitute 'reservoirs' of malaria that might continue to maintain the parasite even if it were exterminated in humans. Similarly, Krief *et al.* (2010) suggested that *P. falciparum* in humans arose as a result of a host transfer from bonobo. Analyses reported here show that all available bonobo-derived complete mitochondrial genomes are more similar to human-derived *P. falciparum* sequences than they are to other bonobo-derived sequences. It is true that nucleotide diversity of bonobo-derived *P. falciparum* mitochondrial genomes was greater than that of available human-derived mitochondrial genomes. However, the vast majority (69·2%) of available human-derived complete mitochondrial genome sequences originated in Asia, New Guinea or the Americas. Since there is no doubt that *P. falciparum* originated in Africa, it is expected that the most diverse *P. falciparum* populations infecting humans would be seen in Africa and that *P. falciparum* underwent a severe population bottleneck as it accompanied its human host out of Africa. Tanabe *et al.* (2010) have reported evidence from nuclear DNA sequences in support of the hypothesis that *P. falciparum* accompanied humans out of Africa and underwent substantial reduction in genetic diversity. An even more severe bottleneck is likely to have occurred when *P.*

falciparum was introduced to the Americas as a result of the African slave trade (Anderson *et al*. 2000).

On the other hand, it is also possible that there has been a recent worldwide reduction in diversity of the mitochondrial genome of *P. falciparum* infecting humans as the result of a selective sweep, perhaps in response to treatment with anti-mitochondrial drugs such as atovaquone (Warhurst, 1999). Point mutations in the *cytochrome b* gene are known to be associated with resistance to the latter drug (Warhurst, 1999). In any event, a complete understanding of the population history of the mitochondrial genome of *P. falciparum* will require a much more extensive sequencing of human-derived isolates from Africa than has so far been undertaken. Moreover, understanding the origin of *P. falciparum* in different primate hosts in Africa will require extensive sequencing of nuclear as well as mitochondrial genes, since the pattern of sequence polymorphism at nuclear loci will provide independent information regarding the species' history.

Liu *et al.* (2010) argued that *P. falciparum* was transferred from gorilla to human, on the basis of an analysis of mitochondrial *cytochrome b* genes in which *P. falciparum*-like sequences from gorilla clustered with human-derived *P. falciparum*. Like the phylogeny of bonobo-derived mitochondrial genomes (Fig. 3), this phylogeny is also consistent with the hypothesis of a very recent transfer from human to gorilla. In Fig. 4, we represent schematically the portion of Liu *et al*.'s (2010, p. 422) *cytochrome b* phylogeny that concerns the subgenus *Laverania*, to which *P. falciparum* and *P. reichenowi* belong. In this rooted phylogenetic tree, *P. falciparum* and *P. reichenowi* form part of a clade within *Laverania*, most members of which are found in *Pan* (the genus including the two extant chimpanzee species; Fig. 4). This topology can be explained by the following hypotheses: (1) that the clade including *P. falciparum* and *P. reichenowi* was present in the common ancestor of *Pan* and *Homo*; (2) that *P. falciparum* and *P. reichenowi* diverged when their host species diverged; and (3) that *P. falcipartum* sequences found in present-day gorilla, bonobo, and chimpanzee represent recent host transfers from human.

Alternatively, if *P. falciparum* was originally a parasite of gorilla, one might hypothesize that *P. falciparum* and *P. reichenowi* diverged when *Pan* and *Gorilla* diverged (6–8 Mya; Glazko and Nei, 2003). Note that, because this divergence time is just slightly earlier than that of *Pan* and *Homo*, the earlier divergence time as a calibration would have little effect on our other divergence time estimates (Table 3). However, if *P. falciparum* was originally a parasite of gorillas, one would also have to assume that *Homo* had no *Laverania* parasites until its acquisition of *P. falciparum* from gorilla. This seems unlikely given that *Homo* inhabited Africa and diverged 5–7 Mya from *Pan*, which is parasitized by several *Laverania* species.

Thus, although further sequencing of *P. falciparum* from human hosts in Africa is needed, currently available results favour the hypothesis that humans are reservoirs for the apes, not vice-versa. Indeed ape infections with *P. falciparum* may be very recent phenomena, occurring only as a result of forest destruction that has brought human and ape populations into greater proximity (Harcourt, 1981; Struhsaker, 1981; Morgan *et al.* 2006). The claim that ape-derived malaria infections may be common in humans (Duval *et al.* 2010; Liu *et al.* 2010) is premature and may have unfortunate consequences for great ape survival. The African great apes are subject to numerous threats (Caldecott and Miles, 2005), and the Western gorilla (*Gorilla gorilla*) is listed as critically endangered (Hopkin, 2007). Unsubstantiated assertions that these primates are a source of human malaria infections can only worsen the situation by providing a convenient excuse to those in Africa who see no value in great ape conservation. On the other hand, if *P. falciparum* infections in great apes are derived from humans, that fact also may be of considerable importance for conservation,

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

further emphasizing the need to establish refuges for the great apes that are safe from human intrusion.

As regards the most widespread human malaria parasite, *Plasmodium vivax*, numerous studies have suggested that it originated in Southeast Asia, since it shows close affinities to malaria parasites of Southeast Asian primates (Escalante *et al.* 2005; Garamszegi, 2009). A previous comparison of complete mitochondrial genomes placed the divergence of *P. vivax* from *P. knowlesi* as recently as 1·5 Mya (Escalante *et al.* 2005), whereas the present analysis of nuclear genes placed it at 15–40 Mya. The most likely explanation for this discrepancy is that synonymous sites in the mitochondrial genome saturate rapidly with changes, but extremely biased codon usage (McIntosh *et al.* 1998) leads to a high degree of homoplasy that in turn masks the effects of saturation. Hughes and Verra (2010) estimated the synonymous substitution rate in the mitochondrial *cytb* gene between *P. falciparum* and *P. reichenowi* to be about $1–2\times10^{-8}$ substitutions/site/year. The present estimates of the synonymous substitution rate in nuclear genes of the same taxa are at least 20 lower (6×10^{-9} substitutions/site/year, assuming a divergence time of 5 My; Table 2). If the rates are similar in *P. vivax* and *P. knowlesi*, saturation of synonymous sites in mitochondrial genes after 20 million years would be expected.

The recent analyses of Ricklefs and Outlaw (2010) provide additional support for the hypothesis of rapid saturation in *Plasmodium* mitochondrial genes. Using a rate estimation based on closely related pairs of bird species, those authors used mitochondrial *cytochrome b* sequences to obtain estimates of the deep divergence times within the Haemosporidia that were far more recent than those we obtained using nuclear sequences (Table 3). For instance, Ricklefs and Outlaw (2010) provided an estimate of only 8·7 Mya for the basal split within mammalian *Plasmodium*, which by our estimates must have occurred substantially before 15–46 Mya, the estimated age of the *P. vivax*–*P. knowlesi* and rodent malaria clades (Table 3). The phenomenon of saturation was particularly likely to be a factor in this case because Ricklefs and Outlaw (2010) based their estimates on both synonymous and non-synonymous sites in the *cytochrome b* gene, although non-synonymous sites are known to be subject to strong purifying selection (Hughes and Verra, 2010).

A much earlier date for the divergence of *P. vivax* and *P. knowlesi* places that divergence long before the origin of the genus *Macaca*, around the time of the early radiation of the Old World primates. This does not necessarily imply that *P. vivax* has been a parasite of hominoids from such an early date. Rather, most phylogenetic analyses show that *P. vivax* forms part of a well-supported clade along with malaria parasites from Asian monkeys (Duval *et al.* 2010; Escalante *et al.* 1998, 2005; Hayakawa *et al.* 2008; Jongwutiwes *et al.* 2005; Mitsui *et al.* 2010; Mu *et al.* 2005; Roy and Irimia, 2008). So far, no member of that clade has been found in Africa, other than *P. vivax* in humans, although interestingly *cytb* and *cox1* sequences from a *Plasmodium* isolate apparently closely related to that clade has been isolated from a Prosimian in Madagascar (Duval *et al.* 2010). Thus, at the present time, it still seems most likely that *P. vivax* became a hominid parasite when hominids first reached Southeast Asia, although this host transfer may have occurred at the *Homo erectus* stage (Jongwutiwes *et al.* 2005).

The present analyses suggest that the radiation of *P. berghei, P. chabaudi* and *P. yoelii*, the three rodent malaria parasite species with completely sequenced genomes, coincided with the radiation of the rodent family Muridae. The observation that non-human primates are infected by complex communities of malaria parasites (Duval *et al.* 2010; Prugnolle *et al.* 2010; Putaporntip *et al.* 2010) suggests that the same is likely to be true of rodents as well. Further examination of *Thamnomys* and related African rodents can be expected to yield new information that will be critical for understanding the evolutionary and ecological

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

relationships of rodent malaria parasites; such an understanding may in turn enhance the usefulness of the rodent malaria parasites as laboratory models for understanding basic malarial biology.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank D. Harris and J. Gillespie for methodological discussions and comments on the manuscript.

FINANCIAL SUPPORT

This work was supported by grant GM43940 from the National Institutes of Health to A.L.H., and 5R01 GM070793-03 to J.M.C. We thank the Wellcome Trust Sanger Institute for access to the draft genome assemblies of *P. gallinaceum* and *P. reichenowi*, and the Texas A&M University Brazos HPC cluster and the Computational Biology Service Unit from Cornell University for providing computational resources.

References

Abascal F, Zardoya R, Posada D. ProtTest: Selection of best-fit models of protein evolution. Bioinformatics. 2005; 21:2104–2105. [PubMed: 15647292]

Anderson TJ, Haubold B, Williams JT, Estrada-Franco JG, Richardson L, Mollinedo R, Bockarie M, Mokili J, Mharakurwa S, French N, Whitworth J, Velez ID, Brockman AH, Nosten F, Ferreira MU, Day KP. Microsatellite markers reveal a spectrum of population structures in the malaria parasite *Plasmodium falciparum*. Molecular Biology and Evolution. 2000; 17:1467–1482. [PubMed: 11018154]

Aquadro CF, Dumont VB, Reed FA. Genome-wide variation in the human and fruitfly: a comparison. Current Opinion in Genetics and Development. 2001; 11:627–634. [PubMed: 11682305]

Caldecott, J.; Miles, L. World Atlas of Great Apes and their Conservation. University of California Press; Berkeley: 2005.

Caswell J, Mallick S, Richter DJ, Neubauer J, Schirmer C, Gnerre S, Reich D. Analysis of chimpanzee history based on genome sequence alignments. PLoS Genetics. 2008; 4:e1000057. [PubMed: 18421364]

Cox-Singh J, Davis TM, Lee KS, Shamsul SS, Matusop A, Ratnam S, Rahman HA, Conway DJ, Singh B. *Plasmodium knowlesi* malaria in humans is widely distributed and potentially life threatening. Clinical Infectious Diseases. 2008; 46:165–171. [PubMed: 18171245]

Dávalos LM, Perkins SL. Saturation and base composition bias explain phylogenomic conflict in *Plasmodium*. Genomics. 2008; 91:433–442. [PubMed: 18313259]

Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evolutionary Biology. 2007; 7:214. [PubMed: 17996036]

Duval L, Fourment M, Nerrienet E, Rousset D, Sadeuh SA, Goodman SM, Andriaholinirina NV, Randrianarivelojosia M, Paul RE, Robert V, Ayala FJ, Ariey F. African apes as reservoirs of *Plasmodium falciparum* and the origin and diversification of the *Laverania* subgenus. Proceedings of the National Academy of Sciences, USA. 2010; 107:10561–10566.

Egan A, Mahurkar A, Crabtree J, Badger JH, Carlton JM, Silva JC. IDEA: Interactive Display for Evolutionary Analyses. BMC Bioinformatics. 2008; 9:524. [PubMed: 19061522]

Escalante AA, Ayala FJ. Phylogeny of the malarial genus *Plasmodium*, derived from rRNA gene sequences. Proceedings of the National Academy of Sciences, USA. 1994; 91:11373–11377.

Escalante AA, Cornejo OE, Freeland DE, Poe AC, Durrego E, Collins WE, Lal AA. A monkey's tale: the origin of *Plasmodium vivax* as a human malaria parasite. Proceedings of the National Academy of Sciences, USA. 2005; 102:1980–1985.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Escalante AA, Freeland DE, Collins WE, Lal AA. The evolution of primate malaria parasites based on the gene encoding cytochrome b from the linear mitochondrial genome. Proceedings of the National Academy of Sciences, USA. 1998; 95:8124–8129.

Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap. Evolution. 1985; 39:783–791.

Garamszegi LZ. Patterns of co-speciation and host switching in primate malaria parasites. Malaria Journal. 2009; 8:110. [PubMed: 19463162]

Glazko GV, Nei M. Estimation of divergence times for major lineages of primate species. Molecular Biology and Evolution. 2003; 20:424–434. [PubMed: 12644563]

Harcourt AH. Can Uganda's gorillas survive? – A survey of the Bwindi Forest Reserve. Biological Conservation. 1981; 19:269–282.

Hayakawa T, Culleton R, Otani H, Horii T, Tanabe K. Big bang in the evolution of extant malaria parasites. Molecular Biology and Evolution. 2008; 25:2233–2239. [PubMed: 18687771]

Honeycutt, RL. Rodents (Rodentia). In: Hedges, SB.; Kumar, S., editors. The Timetree of Life. Oxford University Press; New York: 2009. p. 490-494.

Hopkin M. Gorillas on the list. Nature. 2007; 449:127. [PubMed: 17851485]

Hotez PJ, Molyneux DH, Fenwick A, Ottesen E, Ehrlich Sachs S, Sachs JD. Incorporating a rapid-impact package for neglected tropical diseases with programs for HIV/AIDS, tuberculosis, and malaria. PLoS Medicine. 2006; 3:e102. [PubMed: 16435908]

Hughes AL, French JO. Homologous recombination and the pattern of nucleotide substitution in *Ehrlichia ruminantium*. Gene. 2007; 387:31–37. [PubMed: 17005333]

Hughes AL, Nei M. Evolutionary relationships of class II major-histocompatibility-complex genes in mammals. Molecular Biology and Evolution. 1990; 7:491–514. [PubMed: 2126590]

Hughes AL, Verra F. Very large long-term effective population size in the virulent human malaria parasite *Plasmodium falciparum*. Proceedings of the Royal Society of London B. 2001; 268:1855–1860.

Hughes AL, Verra F. Extensive polymorphism and ancient origin of *Plasmodium falciparum*. Trends in Parasitology. 2002; 18:348–351. [PubMed: 12377283]

Hughes AL, Verra F. Malaria parasite sequences from chimpanzee support the co-speciation hypothesis for the origin of virulent human malaria (*Plasmodium falciparum*). Molecular Phylogenetics and Evolution. 2010; 57:135–143. [PubMed: 20541613]

Hughes MK, Hughes AL. Natural selection on *Plasmodium* surface proteins. Molecular and Biochemical Parasitology. 1995; 71:99–113. [PubMed: 7630387]

Jongwutiwes S, Putaporntip C, Iwasaki T, Ferreira MU, Kanbara H, Hughes AL. Mitochondrial genome sequences support ancient population expansion in *Plasmodium vivax*. Molecular Biology and Evolution. 2005; 22:1733–1739. [PubMed: 15901839]

Joy DA, Feng X, Mu J, Furuya F, Chotivanich K, Krettli AU, Ho M, Wang A, White NJ, Suh E, Beerli P, Su X. Early origin and recent expansion of *Plasmodium falciparum*. Science. 2003; 300:318–321. [PubMed: 12690197]

Kappe SH, Vaughn AM, Boddey JA, Cowman AF. That was then but this is now: malaria research in the time of an eradication agenda. Science. 2010; 328:862–866. [PubMed: 20466924]

Killick-Kendrick R. Malaria parasites of *Thamnomys rutilans* (Rodentia, Muridae) in Nigeria. Bulletin of the World Health Organization. 1968; 38:822–824. [PubMed: 5303337]

Krief S, Escalante AA, Pacheco MA, Mugisha L, Andre C, Halbwax M, Fischer A, Krief JM, Kasenene JM, Crandfield M, Cornejo OE, Chavatte JM, Lin C, Letourneur F, Gruner AC, McCutchan TF, Renia L, Snounou G. On the diversity of malaria parasites in African apes and the origin of *Plasmodium falciparum* from Bonobos. PLoS Pathogens. 2010; 6:e1000765. [PubMed: 20169187]

Lefevre T, Sanchez M, Ponton F, Hughes D, Thomas F. Virulence and resistance in malaria: who drives the outcome of the infection? Trends in Parasitology. 2007; 23:299–302. [PubMed: 17493874]

Li, WH. Molecular Evolution. Sinauer; Sunderland, MA: 1997.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Li WH, Sadler LA. Low nucleotide diversity in man. Genetics. 1991; 129:513–523. [PubMed: 1743489]

Liu W, Li Y, Learn GH, Rudicell RS, Robertson JD, Keele BF, Ndjango JBN, Sanz CM, Morgan DB, Locatelli S, Gonder MK, Kranzusch PJ, Walsh PD, Delaporte E, Mpoudi-Ngole E, Georgiev AV, Müller MN, Shaw GM, Peeters M, Sharp PM, Rayner JC, Hahn BH. Origin of the human malaria parasite *Plasmodium falciparum* in gorillas. Nature. 2010; 467:420–425. [PubMed: 20864995]

Livingstone FB. Anthropological implications of sickle cell gene distribution in West Africa. American Anthropologist. 1958; 60:531–561.

Martinsen ES, Perkins SL, Schall JJ. A three-genome phylogeny of malaria parasites (*Plasmodium* and closely related genera): Evolution of life-history traits and host switches. Molecular Phylogenetics and Evolution. 2008; 47:261–273. [PubMed: 18248741]

McCutchan TF, Kissinger JC, Touray MG, Rogers MJ, Li J, Sullivan M, Braga EM, Krettli AU, Miller LH. Comparison of circumsporozoite proteins from avian and mammalian malarias: biological and phylogenetic implications. Proceedings of the National Academy of Sciences, USA. 1996; 93:11889–11894.

McIntosh MT, Srivasta R, Vaidya AB. Divergent evolutionary constraints on mitochondrial and nuclear genomes of malaria parasites. Molecular and Biochemical Parasitology. 1998; 95:69–80. [PubMed: 9763290]

Miller, MA.; Holder, MT.; Vos, R.; Midford, PE.; Liebowitz, T.; Chan, L.; Hoover, P.; Warnow, T. The CIPRES Portals. CIPRES. 2010 [Accessed: 2009-08-04]. 2009-08-04. URL: http://www.phylo.org/sub_sections/portal(Archived by WebCite(r) at http://www.webcitation.org/5imQlJeQa)

Mitsui H, Arisue N, Sakihama N, Inagaki Y, Horii T, Hasegawa M, Tanabe K, Hashimoto T. Phylogeny of Asian primate malarias inferred from apicoplast genome-encoded genes with special emphasis on the positions of *Plasmodium vivax* and *P. fragile*. Gene. 2010; 450:32–38. [PubMed: 19818838]

Morgan D, Sanz C, Ononga JR, Strindberg S. Ape abundance and habitat use in the Gouaougo Triange, Republic of Congo. International Journal of Primatology. 2006; 27:147–179.

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]

Mu J, Joy DA, Duan J, Huang Y, Carlton J, Walker J, Barnwell J, Beerli P, Charleston MA, Pybus OG, Su X. Host switch leads to emergence of *Plasmodium vivax* malaria in humans. Molecular Biology and Evolution. 2005; 22:1686–1693. [PubMed: 15858201]

Nei M, Gojobori T. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Molecular Biology and Evolution. 1986; 3:418–426. [PubMed: 3444411]

Nei, M.; Kumar, S. Molecular Evolution and Phylogenetics. Oxford University Press; New York: 2000.

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:e1000446. [PubMed: 19478877]

Perkins SL, Schall JJ. A molecular phylogeny of malarial parasites recovered from cytochrome b gene sequences. Journal of Parasitology. 2002; 88:972–978. [PubMed: 12435139]

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*. Proceedings of the National Academy of Sciences, USA. 2010; 107:1458–1463.

Putaporntip C, Jongwutiwes S, Thongaree S, Seethamchai S, Grynberg P, Hughes AL. Ecology of malaria parasites infection Southeast Asian macaques: evidence from cytochrome b sequences. Molecular Ecology. 2010; 19:3466–3476. [PubMed: 20646216]

Qari SH, Shi YP, Pieniazek NJ, Collins WE, Lal AA. Phylogenetic relationship among the malaria parasites based on small subunit rRNA gene sequences: monophyletic nature of the human malaria parasite, *Plasmodium falciparum*. Molecular Phylogenetics and Evolution. 1996; 6:157–165. [PubMed: 8812316]

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Rich SM, Leendertz FH, Xu G, LeBreton M, Djoko CF, Aminake MN, Takang EE, Diffo JL, Pike BL, Rosenthal BM, Formenty P, Boesch C, Ayala FJ, Wolfe ND. The origin of malignant malaria. Proceedings of the National Academy of Sciences, USA. 2009; 196:14902–14907.

Ricklefs RE, Outlaw DC. A molecular clock for malaria parasites. Science. 2010; 329:226–229. [PubMed: 20616281]

Ronquist F, Huelsenbeck JP. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics. 2003; 19:1572–1574. [PubMed: 12912839]

Roy SW, Irimia M. Origins of human malaria: rare genomic changes and full mitochondrial genomes confirm the relationship of *Plasmodium falciparum* to other mammalian parasites but complicate the origins of *Plasmodium vivax*. Molecular Biology and Evolution. 2008; 25:1192–1198. [PubMed: 18359945]

Sanderson MJ. A nonparametric approach to estimating divergence timescales in the absence of rate constancy. Molecular Biology and Evolution. 1997; 15:1218–1231.

Siddall ME, Barta JR. Phylogeny of *Plasmodium* species: estimation and inference. Journal of Parasitology. 1992; 78:567–568. [PubMed: 1597812]

Sorhannus U, Fox M. Synonymous and nonsynonymous substitution rates in diatoms: a comparison between chloroplast and nuclear genes. Journal of Molecular Evolution. 1999; 48:209–212. [PubMed: 9929389]

Stamatakis A. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics. 2006; 22:2688–2690. [PubMed: 16928733]

Steiper, ME.; Young, NM. Primates (Primates). In: Hedges, SB.; Kumar, S., editors. The Timetree of Life. Oxford University Press; New York: 2009. p. 482-486.

Struhsaker TT. Forest and primate conservation in East Africa. African Journal of Ecology. 1981; 19:99–114.

Tamura K, Dudley J, Nei M, Kumar S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Molecular Biology and Evolution. 2007; 24:1596–1599. [PubMed: 17488738]

Tamura K, Nei M. Estimation of the numberof nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Molecular Biology and Evolution. 1993; 10:512–526. [PubMed: 8336541]

Tanabe K, Mita T, Jombart T, Eriksson A, Horibe S, Placpac N, Ranford-Cartwright L, Sawai H, Sakihama N, Ohmae S, Nakamura M, Ferreira MU, Escalante AA, Prugnolle F, Björkman A, Färnet A, Kaneko A, Horii T, Manica A, Kishino H, Balloux F. *Plasmodium falciparum* accompanied the human expansion out of Africa. Current Biology. 2010; 20:1–7. [PubMed: 20036540]

Takezaki N, Rzhetsky A, Nei M. Phylogenetic test of the molecular clock and linearized trees. Molecular Biology and Evolution. 1995; 12:823–833. [PubMed: 7476128]

The International SNP Map Working Group. A map of human genome sequence variation containing 1·42 million single nucleotide polymorphisms. Nature. 2002; 409:928–933.

Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improvement of the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Research. 1994; 22:4673–4680. [PubMed: 7984417]

Volkman SK, Sabeti PC, DeCaprio D, Neafsey DE, Schaffner SF, Milner AA 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 Genetics. 2007; 39:113–119. [PubMed: 17159979]

Warhurst DC. Drug resistance in *Plasmodium falciparum* malaria. Infection. 1999; 27:S55–S58. [PubMed: 10885831]

Waters AP, Higgins DG, McCutchan TF. *Plasmodium falciparum* appears to have arisen as a result of lateral transfer between avian and human hosts. Proceedings of the National Academy of Sciences, USA. 1991; 88:3140–3144.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

Wilgenbusch, JC.; Warren, DL.; Swofford, DL. AWTY: A system for graphical exploration of MCMC convergence in Bayesian phylogenetic inference. 2004. http://king2.scs.fsu.edu/CEBProjects/awty/awty_start.php

Wolfe ND, Dunavan CP, Diamond J. Origins of major human infectious diseases. Nature. 2007; 447:279–283. [PubMed: 17507975]

Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. Molecular Biology and Evolution. 2007; 24:1586–1591. [PubMed: 17483113]

Yang Z, Nielsen R. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. Molecular Biology and Evolution. 2000; 17:32–43. [PubMed: 10666704]

Yoeli M, Most H. A study of *Plasmodium berghei* in *Thamnomys surdaster*, and in other experimental hosts. American Journal of Tropical Medicine and Hygiene. 1964; 13:659–663. [PubMed: 14205883]

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

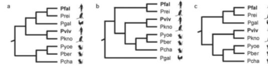

Fig. 1.
Alternative hypotheses for the phylogenetic relationships among *Plasmodium* species. Pfal=*P. falciparum*; Prei=*P. reichenowi*; Pgal=*P. gallinaceum*; Pviv=*P. vivax*; Pkno=*P. knowlesi*; Pyoe=*P. yoelii*; Pber=*P. berghei*; Pcha=*P. chabaudi* (See text for further discussion of each hypothesis).

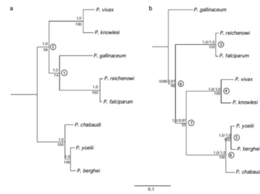

Fig. 2.

Phylogenetic relationships among *Plasmodium* species. a: Tree topology recovered in NJ, and in two out of seven BEAST runs (posterior log likelihood score=−86,636, ESS >2000). b: Tree topology recovered in RAxML, MrBayes and five out of seven BEAST analyses (posterior log likelihood score=−86,630, ESS>2000). Branch lengths as per BEAST analyses. Nodes referenced in the text are circled. Posterior probability for each node is shown above the branches (BEAST/MrBayes; the latter only available for b). Percent bootstrap support in NJ or ML analyses is shown below branches in trees a and b, respectively. Out-group (*T. parva* and *T. annulata*) not shown.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

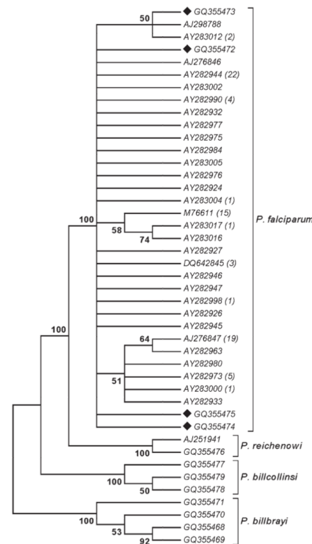

Fig. 3.

Condensed NJ tree of *Plasmodium* mitochondrial genomes based on MCL distance. Diamonds indicate *P. falciparum* sequences derived from bonobo; other *P. falciparum* sequences are from human. Only unique sequences are shown; when the data-set included more sequences from human identical to a given sequence, the number of those additional sequences is indicated in parentheses. Numbers on the branches are percentages of 1000 boostrap samples supporting the branch; only branches with 50% support or more are shown.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

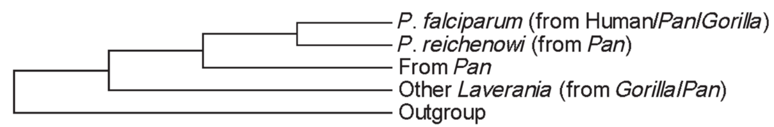

Fig. 4.

Schematic representation of the phylogeny of the subgenus *Laverania* from the phylogenetic analyses of Liu *et al*. (2010).

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 1

Mean synonymous and non-synonymous nucleotide diversity (π_S and π_N) within *P. falciparum* and synonymous divergence (d_S and d_N) between *P. falciparum* and *P. reichenowi* at 45 protein-coding loci

P. falciparum	π_S ± S.E.	0·0007±0·0002
	π_N ± S.E.	0·0007±0·0002
	Maximum d_S ± S.E.	0·0029±0·0009
P. falciparum vs. *P. reichenowi*	d_S ± S.E.	0·0595±0·0056
	d_N ± S.E.	0·0051±0·0008

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

Table 2

Synonymous substitution rates of *Plasmodium falciparum* and *P. reichenowi* nuclear genes and estimated most recent common ancestor (MRCA) of *P. falciparum* based on hypothetical *P. falciparum–P. reichenowi* divergence times

Divergence time (years)	Synonymous substitutions/site/year[1]	95% C.I. for synonymous substitution rate	MRCA of *P. falciparum*[2]
7,000,000	$4{\cdot}3\times10^{-9}$	$3{\cdot}5$–$5{\cdot}1\times10^{-9}$	341,000
5,000,000	$6{\cdot}0\times10^{-9}$	$4{\cdot}9$–$7{\cdot}1\times10^{-9}$	244,000
4,000,000	$7{\cdot}4\times10^{-9}$	$6{\cdot}0$–$8{\cdot}8\times10^{-9}$	195,000
2,000,000	$1{\cdot}5\times10^{-8}$	$1{\cdot}2$–$1{\cdot}8\times10^{-8}$	97,500
1,300,000	$2{\cdot}3\times10^{-8}$	$1{\cdot}9$–$2{\cdot}7\times10^{-8}$	63,300
500,000	$6{\cdot}0\times10^{-8}$	$4{\cdot}9$–$7{\cdot}1\times10^{-8}$	24,400
50,000	$6{\cdot}0\times10^{-7}$	$4{\cdot}9$–$7{\cdot}1\times10^{-7}$	2,440
5,000	$6{\cdot}0\times10^{-6}$	$4{\cdot}9$–$7{\cdot}1\times10^{-6}$	244

[1] Based on d_S values in Table 1.

[2] Based on maximum d_S values in Table 1.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 3

Summary of divergence time estimates, in millions of years

	Method				
Divergence	**Linearized tree (JTT)**	**NPRS**	**Regression (d_N)**	**Regression (d_S)**	**Bayesian MCMC (BEAST)**
P. falciparum–P. reichenowi[1]	5–7	5–7	5–7	5–7	5·8 (5–7)
P. vivax–P. knowlesi	26–36	46	22–31	28–39	15·3
P. berghei–P. yoelii	12–17	17	9–12	11–15	6·8
P. chabaudi–(P. berghei, P. yoelii)	24–33	36	20–27	18–25	15·5

[1] Calibration.