Host Switch Leads to Emergence of *Plasmodium vivax* Malaria in Humans

Jianbing Mu,[*][1] *Deirdre A. Joy,*[*][1] *Junhui Duan,*[†] *Yaming Huang,*[‡] *Jane Carlton,*[§] *John Walker,*[‖]
John Barnwell,[¶] *Peter Beerli,*[#] *Michael A. Charleston,*[**] *Oliver G. Pybus,*[††] *and*
Xin-zhuan Su[*]

[*]Laboratory of Malaria and Vector Research and [†]Malaria Vaccine Development Unit, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Rockville; [‡]Guangxi Zhuang Autonomous Region Center for Disease Prevention and Control, GuangXi, P. R. China; [§]Parasite Genomics Group, The Institute for Genomic Research, Rockville, Maryland; [‖]Parasitology Section, Centre for Infectious Diseases and Microbiology Laboratory Services ICPMR, Westmead Hospital, Westmead NSW, Australia; [¶]Division of Parasitic Diseases, Centers for Disease Control and Prevention, Atlanta; [#]School of Computational Science and Department of Biological Sciences, Florida State University; [**]School of Information Technologies and SUBIT, University of Sydney, Sydney NSW, Australia; and [††]Department of Zoology, Oxford University, Oxford, United Kingdom

The geographical origin of *Plasmodium vivax*, the most widespread human malaria parasite, is controversial. Although genetic closeness to Asian primate malarias has been confirmed by phylogenetic analyses, genetic similarities between *P. vivax* and *Plasmodium simium*, a New World primate malaria, suggest that humans may have acquired *P. vivax* from New World monkeys or vice versa. Additionally, the near fixation of the Duffy-negative blood type (FY $\times$ B^{null}/FY $\times$ B^{null}) in West and Central Africa, consistent with directional selection, and the association of Duffy negativity with complete resistance to vivax malaria suggest a prolonged period of host-parasite coevolution in Africa. Here we use Bayesian and likelihood methods in conjunction with cophylogeny mapping to reconstruct the genetic and coevolutionary history of *P. vivax* from the complete mitochondrial genome of 176 isolates as well as several closely related *Plasmodium* species. Taken together, a haplotype network, parasite migration patterns, demographic history, and cophylogeny mapping support an Asian origin via a host switch from macaque monkeys.

Introduction

The coevolutionary history of primate malarias and their hosts has long been of interest because of the potential for insight into the pathogenicity and life-history traits of human malarias, as well as the zoonotic potential of non-human malarias (Garnham 1966; Coatney et al. 2003). Of the four malaria species that infect humans, *Plasmodium vivax* is the most geographically widespread and is found throughout South and Central America, Asia, the Middle East, and parts of Africa. *Plasmodium vivax* infects an estimated 70–80 million people annually, and despite being generally nonfatal, the disease and economic burden it exerts on humans have been enormous (Carter and Mendis 2002). Compared with the more virulent *Plasmodium falciparum*, *P. vivax* tolerates a wide range of temperature environments (minimum: 16°C vs. 21°C for *P. falciparum*), which may explain its broader distribution. Alternatively, this distribution may reflect a longer historical association with humans. In fact, a previous study has shown a greater level of genetic diversity in *P. vivax* than in *P. falciparum*, suggesting that the former is older (Feng et al. 2003). On the other hand, low microsatellite and tandem repeat (TR) variability may indicate that *P. vivax* has only recently (<10,000 years ago) infected humans (Leclerc et al. 2004), although a different study based on polymorphisms in two nuclear and one plastid gene places the origin between 45,000 and 81,000 years ago (Escalante et al. 2005).

Relatively little is known about the evolutionary history and genetic structure of *P. vivax*, partly due to the difficulty of obtaining adequate genetic material in the absence of a sustainable in vitro culture system. Thus, questions about the age and geographic origin of *P. vivax* remain largely unresolved. Phylogenetic analyses have consistently placed *P. vivax* among the Asian primate malarias (Garnham 1966; Escalante et al. 1998; Perkins and Schall 2002; Escalante et al. 2005). However, near genetic identity between *P. vivax* and *Plasmodium simium*, a New World primate malaria, raises the possibility that humans acquired *P. vivax* from New World monkeys or vice versa (Ayala, Escalante, and Rich 1999; Coatney et al. 2003). Indirect evidence based on the high prevalence (up to 99%) of the Duffy-negative blood type (FY $\times$ B^{null}/FY $\times$ B^{null}) in Western and Central Africa where *P. vivax* is almost nonexistent, coupled with the association of Duffy negativity with complete resistance to vivax malaria, suggest that *P. vivax* has been coevolving with Africans longer than with other human populations (Miller et al. 1976; Carter and Mendis 2002). To date, only a few studies have tested the Duffy selection hypothesis, and though evidence for a selective sweep has been found, the question of the geographical origin of *P. vivax* has not been directly addressed (Hamblin and Di Rienzo 2000; Hamblin, Thompson, and Rienzo 2002; Seixas, Ferrand, and Rocha 2002). In this study we have combined population and phylogenetic analyses of the mitochondrial (mt) genome to reconstruct a comprehensive genetic history of *P. vivax* and its association with humans.

Methods

Parasite DNA Samples and DNA Sequencing

Previously collected *P. vivax*-infected human blood samples were obtained from South and Central America, Africa, Southeast (SE) Asia, and Melanesia (Supplementary table 1, Supplementary Material online). Parasite DNA

---

was extracted from filter paper using QIAGEN DNA kits. Oligonucleotide primers were synthesized based on *P. vivax* mitochondrial DNA (mtDNA) deposited in Genbank (Sharma, Pasha, and Sharma 1998).

The 6-kb mt genome was amplified and sequenced for 176 isolates and five simian malaria species (*Plasmodium cynomolgi*, *Plasmodium simiovale*, *Plasmodium gonderi*, *P. simium*, and the undescribed species *Plasmodium* sp. (mandrill)). DNA was amplified by polymerase chain reaction (PCR) using a reaction mixture containing the following: 4 μl DNA (~5 ng), 0.5 μl (~50 pM) primer, and 45 μl PCR mix containing 5 μl of 10 × PCR buffer, 1.0 μl deoxynucleoside triphosphate (10 mM), and 0.1 μl (5 units/μl) of *Taq* polymerase (Invitrogen, Carlsbad, Calif.). We employed the following cycling conditions: 94°C for 2 min, 35 cycles of 94°C for 20 s, 55°C for 10 s, 50°C for 10 s, and 65°C for 1.5 min, and a final extension step of 65°C for 5 min. For sequencing, PCR products were first treated with 1 μl of Exo-SAP-IT (U.S. Biochemical, Cleveland, Ohio) at 37°C for 15 min and at 80°C for an additional 15 min. Sequencing reactions containing approximately 2–5 μl PCR product and BigDye terminator chemistry were run on an ABI 3730xl automatic sequencer (Applied Biosystems, Foster City, Calif.). Sequence cycling conditions were as follows: 94°C for 2 min, 25 cycles of 94°C for 20 s, 50°C for 5 s, 48°C for 5 s, and 60°C for 3 min, and a final extension at 60°C for 5 min. DNA sequences were trimmed and aligned using the Phred/Phap package, available at http://www.phrap.org/, and Sequencher 4.0 (Gene Codes Corp., Ann Arbor, Mich.).

Tests of Neutrality and Recombination

To assess the suitability of the parasite mt genome for addressing evolutionary questions, we tested for selection, recombination, and rate heterogeneity. To detect selection acting on specific codons in the three protein-coding genes, we used the likelihood methods implemented in codeml program in the PAML package (Yang 1997). Models that describe the variation in the nonsynonymous-synonymous substitution rate (ω) among sites (Nielsen and Yang 1998; Yang, Wong, and Nielsen 2005) were compared using likelihood ratio tests (LRTs). Specifically, we compared null models with those that allow positive selection (ω > 1) (M1a vs. M2a and M7 vs. M8, see Supplementary table 2 [Supplementary Material online] for model details). Three methods were used to test for recombination within geographical populations. First, a GENECONV analysis was performed on the global alignment to detect potential recombinant fragments (Sawyer 1989). In addition, we estimated recombination using the population gene conversion parameter (γ) and the composite likelihood method of Hudson (2001) as implemented in LDhat (McVean, Awadalla, and Fearnhead 2002). A minor allele cutoff frequency of 0.05 was used for this analysis. Finally, we measured the correlation between r^2 and physical distance using LDhat as well. An LRT of the molecular clock assumption was employed to measure rate heterogeneity among *P. vivax* lineages.

Haplotype Network Reconstruction

A haplotype network was reconstructed from a subset of 145 isolates using the program TCS 1.18 (Clement, Posada, and Crandall 2000) to determine the maximum allowable number of connections within a 95% parsimony limit (Templeton, Crandall, and Sing 1992). Thirty-one Melanesian isolates were randomly removed to avoid biasing outgroup probabilities due to unequal sample sizes. Outgroup probabilities are estimates of the relative haplotype age based on the haplotype frequency and the number of connections to other haplotypes in the network (Castelloe and Templeton 1994).

Migration Rate Estimates

Migration rates among geographical populations were estimated using the maximum likelihood (ML) program MIGRATE 1.7.6 (Beerli 2002). Markov chain Monte Carlo (MCMC) search parameters were as follows: 10 short chains (visiting 21,000 genealogies and sampling 1,000 per chain) and 3 long chains (visiting 201,000 genealogies and sampling 10,000 per long chain). To guarantee better mixing of the MCMC searches, adaptive heating with four temperatures and combination over ten replicate runs were employed. Start parameters were obtained from F_{ST} estimates.

Phylogenetic Analysis

A *Plasmodium* tree was reconstructed for use in cophylogeny mapping from mt genome sequences (eight species) or cytochrome *b* sequences (Escalante et al. 1998) when the entire mt genome was not available. The tree was inferred with ML as implemented in PAUP v4.0b10 (Swofford 1998) and with Bayesian inference (BI) in MrBayes v3.0b4 (Huelsenbeck and Ronquist 2001). The general time-reversal substitution model with a proportion of invariable sites and a gamma distribution of rate variation among sites (GTR + I + Γ model) was chosen based on LRTs (Posada and Crandall 1998). For BI, model parameters were optimized separately for each partition (three genes, ribosomal RNA). We conducted two independent runs of the Markov chain at four temperatures (three ''heated'') for one million generations, sampling every 100 generations. The two runs converged on similar log-likelihood scores and reached stationarity by 100,000 generations. After discarding the first 1,000 trees from each analysis, the remaining samples were combined to generate a 50% majority consensus tree, with the percentage of samples recovering a clade representing that clade's posterior probability. The primate host tree was adopted from Page and Goodman (2001).

Cophylogeny Mapping of Malaria Parasites and Their Primate Hosts

We used cophylogeny mapping (Charleston 1998) as implemented in TreeMap v2.0 (Charleston and Page 2002) to compare alternative hypotheses about the origin of *P. vivax*. This approach accommodates noncodivergent events (NCEs) such as host switching, parasite lineage duplication within a host, and parasite lineage loss and is guaranteed to recover all potentially Pareto-optimal reconciliations of the host and parasite trees given a certain weighting scheme. (Charleston 2003). We used the default weighting scheme of codivergence = 0 and NCEs = 1.

Table 1
Bayesian Estimates of Population Size and TMRCA Under Four Different Demographic Models

Demographic Model and Its Equation	TMRCA (years)	Current Population Size, N_0	Mean Log-likelihood	AIC Score
Constant ($N_t = N_0$)	186,000 (121,000–265,000)	587,000 (435,000–761,000)	−11269.7	22551.4
Expansion				
$(N_t = N_0 \{[1 + \alpha]/[1 + \alpha \cdot \exp(-rt)]\})$	182,000 (116,000–263,000)	599,000 (428,000–788,000)	−11269.1	22554.2
Exponential ($N_t = N_0 \cdot \exp(-rt)$)	82,400 (53,000–117,000)	2,594,000 (1,146,000–5,227,000)	−11290.1	22594.2
Logistic				
$(N_t = N_0 \{[1 + c]/[1 + c \cdot \exp(-rt)]\})$	87,400 (54,000–134,000)	4,345,000 (731,000–426,600,000)	−11286.1	22588.2

NOTE.—Parameter estimates are median posterior density values. Confidence intervals in parentheses are highest posterior density intervals. AIC $= -2\ln(L) + 2q$, where q is the number of model parameters.

For each reconciliation, the parasite tree was randomized 100 times to test the null hypothesis that the actual parasite tree is no more congruent with the host tree than randomized parasite trees. Due to memory constraints, we were unable to conduct all randomizations on the total data set. We therefore removed the *Presbytis* host lineage. We ran the analysis three additional times adding in one of the *Presbytis*-parasite associations each time to test whether removing *Presbytis* had any effect on our main conclusions. Cophylogenetic mapping requires fully resolved phylogenies; we therefore ran the program several times with different arrangements of poorly supported branches to test whether this altered our main conclusions. Parasites associated with multiple hosts were treated as monophyletic clades with zero-length branches.

Estimating the Most Recent Common Ancestor

A Bayesian coalescent framework was used to estimate the time to the most recent common ancestor (TMRCA) and current effective population size for *P. vivax* (Drummond and Rambaut 2003). Preliminary analysis suggested the Hasegawa-Kishino-Yano substitution model with a gamma distribution of rate variation among sites (HKY-Γ model) provided the best fit to the within–*P. vivax* data. We estimated a mutation rate for inclusion in the analysis from closely related *Plasmodium* species for which we had the complete mtDNA sequence (*P. gonderi*, *P.* sp. (mandrill), *P. simiovale*, and *P. cynomolgi*), using host divergence time to calibrate rates. For the *Macaca*–other *Papionina* (including *Cercocebus*) branch point, we used the previously reported divergence time of 6.8–7.0 MYA (Page and Goodman 2001). Two assumptions are implicit in this approach: codivergence of parasite and host lineages and similar mutation rates among parasite species. The first assumption is supported by cophylogeny mapping (see below). We tested the second assumption using Tajima's relative rates test (Tajima 1993).

We also estimated the parameters of four models of population growth for inclusion in the Bayesian analysis: constant size, exponential growth, expansion growth, and logistic growth (see table 1 for equations). Two MCMC searches were performed in BEAST v1.1.2 (Drummond and Rambaut 2003) under each demographic model, and convergence was monitored using measures of effective sample size. The initial 10% of each run was discarded. Model fit was assessed by comparing Akaike information criterion (AIC) scores and coalescent likelihood distributions for each model.

Results

Worldwide mt Genome Diversity

We identify 71 single-nucleotide polymorphisms (SNPs) and one microsatellite size polymorphism (Supplementary table 3, Supplementary Material online). The microsatellite was not included in subsequent population analyses. Fifty of the 71 SNPs are in the three protein-coding genes, cytochrome oxidase III (cox 3) having 16 (7 synonymous), cytochrome oxidase I (cox 1) having 20 (11 synonymous), and cytochrome *b* (cytb) having 14 (7 synonymous). The parasite populations in Melanesia, SE Asia, and India appear to be more diverse than those in Africa and the New World when measured both in terms of nucleotide diversity (π) and theta (θ) (Supplementary table 4, Supplementary Material online).

Tests of positive selection in the protein-coding genes did not reject null models in favor of those allowing for positively selected sites, suggesting a lack of sites under diversifying selection (Supplementary table 5, Supplementary Material online). Although analysis of gene conversion in LDhat indicated low levels of recombination for the SE Asian population ($\gamma = 3$, $P = 0.005$), GENECONV did not detect any putative recombination events, nor did we find a correlation between linkage disequilibrium and physical distance between polymorphic sites for any of the geographical populations (Supplementary table 6, Supplementary Material online). The *P. vivax* mt genome appears to be evolving at a relatively constant rate among isolates (LRT: $\chi^2 = 77.4220$, df $= 68$; $P > 0.05$).

Ancestral Haplotype Estimates and Network Structure

The *P. vivax* haplotype network is considerably more complex than that of *P. falciparum*, another human malaria (Joy et al. 2003). For example, although haplotypes show geographic clustering, there are exceptions, most notably China (with two clusters) and SE Asia, found throughout the network (fig. 1). New World *P. vivax* isolates, are not genetically distinct from Old World isolates nor is *P. simium* most closely related to Old World *P. vivax* (fig. 1) as has previously been reported (Li et al. 2001). Reminiscent of the low diversity in New World *P. falciparum*, our New World *P. vivax* sample comprises only five
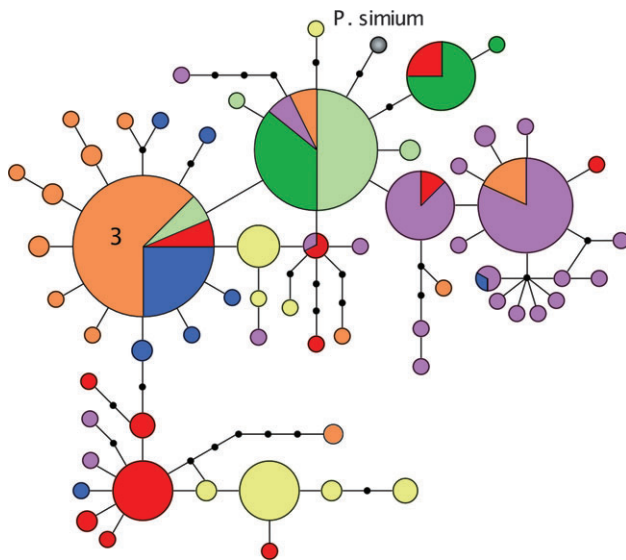
Fig. 1.—Minimum spanning haplotype network of the mt genome sequences showing genetic relationships among 145 *Plasmodium vivax* isolates. Each line represents one mutational step. Circle size is proportional to haplotype frequency. Black dots are hypothetical missing intermediates. Blue, Africa; light green, Central America; dark green, South America; red, SE Asia; orange, India; yellow, China; purple, Melanesia. Haplotypes cluster loosely by geographic region. *Plasmodium simium* clusters with South and Central American isolates. Number indicates haplotype with the highest root probability.
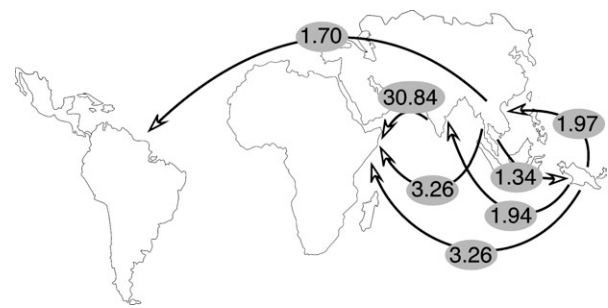

Fig. 2.—Migration among geographical regions (expressed as number of migrants per generation [$2Nm$] exchanged between geographic populations and estimated in MIGRATE 1.7.6). Only migration rates >1 are shown.

haplotypes, three of which were found at low frequencies. Haplotype 3 had the highest outgroup probability (0.121; Supplementary table 7, Supplementary Material online) and is the most widespread, having been collected from India (63%), Africa (25%), SE Asia (6%), and Central America (6%). This haplotype forms the center of a star-like structure (fig. 1) indicative of population expansion or possibly migration.

Elevated Migration Rates from Asia to Africa

The highest migration estimate was from India to Africa (fig. 2). In contrast, we detected no discernable migration out of Africa. Applying this information to the haplotype network suggests that the broadly distributed and potentially oldest haplotype 3 most likely originated in India, not Africa (fig. 1). Two other regions, SE Asia and Melanesia, also showed migration into Africa. In general, migration was not symmetric between regions with the exception of SE Asia and Melanesia (LRT: $\chi^2 = 0.8040$; df = 2; $P = 0.6689$). The New World showed the most isolation, with only very low rates of migration from SE Asia detected.

Plasmodium vivax is Closely Related to Asian Primate Malarias

ML and BI phylogenetic analyses of the mt genome (seven species) combined with previously collected cytochrome b sequences (six species) (Escalante et al. 1998) confirms the closeness of *P. vivax* and *P. simium* and their placement within an Asian parasite clade (fig. 3A). Most branch lengths within this clade are short and poorly supported, such that it is not possible to discern which species

is most closely related to the *P. vivax*–*P. simium* clade or which parasite is ancestral, consistent with previous phylogenetic studies based on the cytochrome b gene (Escalante et al. 1998; Perkins and Schall 2002). However, analysis of two nuclear and one plastid gene supports *Plasmodium fieldi* as the ancestral lineage within the Asian parasite clade (Escalante et al. 2005). The sister group to the Asian clade consists of African mangabey and mandrill parasites, *P. gonderi* and *P.* sp. (mandrill), and this relationship is strongly supported by both bootstrap replicates (90%) and posterior probabilities (>0.95) (fig. 3A).

Host Switch from Asian Monkeys to Humans

Several earlier attempts to reconstruct the coevolutionary history of primates and their malaria parasites have been made based on relationships among the hosts (Garnham 1963; Waters, Higgins, and McCutchan 1993; Coatney et al. 2003); however, as some of these authors have noted, the geographic distribution of parasites and the large evolutionary distances among some primate hosts do not suggest a simple codivergence scenario (fig. 3A). We detected a complex history of host-parasite associations including several host switches. The two solutions with the minimum number of NCEs were also the only solutions that were significantly more congruent than random trees ($P < 0.01$, standard error ≤ 0.00705) (fig. 3B). The inclusion of the *Presbytis* lineage did not alter these conclusions (Supplementary fig. 1, Supplementary Material online). The origin of *P. vivax* is inferred in both solutions as beginning with a host switch from *Macaca*, but the recipient host is either (1) humans (Asian origin) or (2) New World monkeys followed by a second switch to humans (New World origin) (fig. 3C). We reject the second scenario on the grounds that Old and New World monkeys split from each other around 35 MYA (Schrago and Russo 2003), much earlier than the estimated age of *P. vivax*. Additionally, low genetic diversity and lack of emigration argue against a New World *P. vivax* origin. Our data agree with previous reports, based on ecologic and historic grounds (Coatney et al. 2003), that the genetic similarities between *P. vivax* and *P. simium* are the result of a recent lateral transfer from humans to New World monkeys.

Hylobates, *Presbytis*, *Macaca*, and *Homo* all are believed to have originated in Africa and subsequently migrated to Asia (Stewart and Disotell 1998). Thus, any
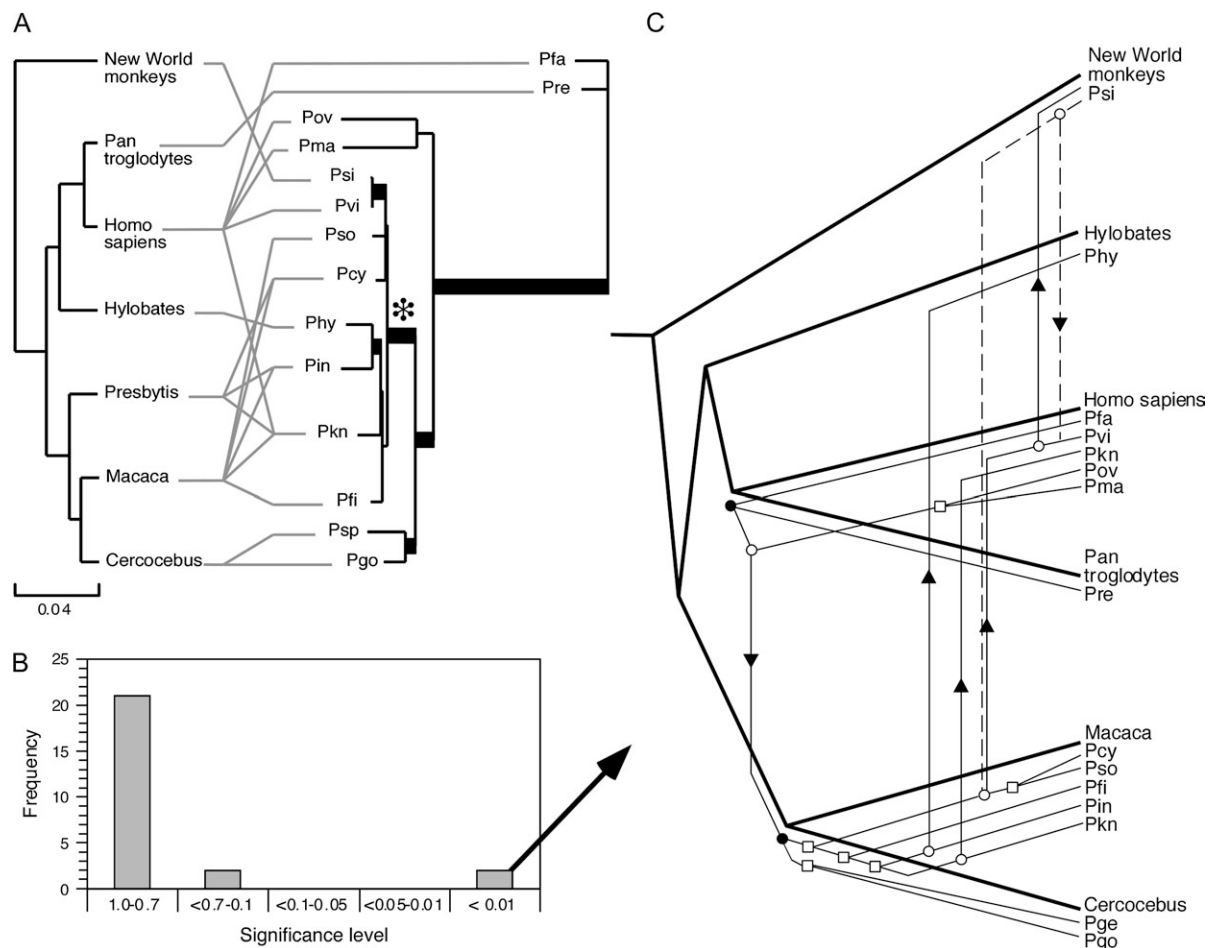
Fig. 3.—Cophylogeny mapping of primate hosts and their malaria parasites. (*A*) Tanglegram illustrating host-parasite associations. Associated taxa are joined by gray lines. Bold branches indicate posterior probabilities >0.95 and bootstrap values >80%. The parasite ML tree was inferred from the complete mt genome when available (*Plasmodium reichenowi*, *Plasmodium falciparum*, *Plasmodium cynomolgi*, *Plasmodium vivax*, *Plasmodium simiovale*, *Plasmodium simium*, *Plasmodium gonderi*, and *Plasmodium georgesi*) or from the mt cytochrome b gene when the complete mt genome was not available (Escalante et al. 1998). The host tree is adapted from Page and Goodman (2001). An "*" demarcates the Asian parasite clade. (*B*) Frequency distribution of significance values for all potentially optimal reconstructions of the primate-parasite trees. For each reconstruction (*n* = 25) the parasite tree was randomized 100 times to test the null hypothesis that the actual parasite tree is no more congruent with the host tree than random trees. (*C*) An overlay of the two reconstructions with $P < 0.01$. Dashed lines illustrate an alternative host switch pattern for *P. vivax*, representing the only difference between the two reconstructions (see text for details). Solid circles, codivergence; open circles followed by arrows, host switch; open squares, duplication within a host. Parasite names are as follows: Pfa, *P. falciparum*; Pre, *P. reichenowi*; Pov, *Plasmodium ovale*; Pma, *Plasmodium malariae*; Psi, *simium*; Pvi, *P. vivax*; Pso, *P. simiovale*; Pcy, *P. cynomolgi*; Phy, *Plasmodium hylobates*; Pin, *Plasmodium inui*; Pkn, *Plasmodium knowlesi*; Pfi, *Plasmodium fieldi*; Psp, *Plasmodium* sp. (mandrill) and Pgo, *P. gonderi*.

one of them could potentially have introduced primate malaria into Asia. Cophylogeny mapping favors a scenario in which primate malaria was introduced by macaque monkeys. However, because *Presbytis* was excluded from the mapping due to memory constraints, we cannot rule out this lineage either (but see Supplementary fig. 1, Supplementary Material online). Host switches from *Macaca* to *Hylobates* and humans were inferred on the solutions with minimum NCEs and significant congruence (fig. 3). Within the *Macaca*, four duplication events were also inferred. We interpreted these as likely codivergences because the *Macaca* lineage in our analysis consists of eight macaque species. We agree with previous authors that the macaque radiation provides a potential mechanism for the radiation of primate malaria parasites (Escalante et al. 1998; Coatney et al. 2003). Short branches within the Asian parasite clade

may indicate rapid speciation and are therefore consistent with cospeciation during the macaque radiation. Fossil and molecular genetic evidences place the start of the macaque radiation in Asia at <4.9 MYA (Delson 1980; Tosi, Morales, and Melnick 2003).

Estimated Timing of Host Switch to Humans

Tajima's relative rates test did not reject the molecular clock hypothesis for *P. vivax* and closely related species (out-group *P. gonderi*: $\chi^2 = 1.85$; $P = 0.174$; out-group *P.* sp. (mandrill): $\chi^2 = 0.17$; $P = 0.680$), allowing us to estimate a mutation rate for inclusion in the Bayesian analysis. The mutation rate estimate was 5.07×10^{-9} (standard deviation $\pm 5.73 \times 10^{-10}$) nucleotide substitutions/site/year. Among the demographic models tested for inclusion in the
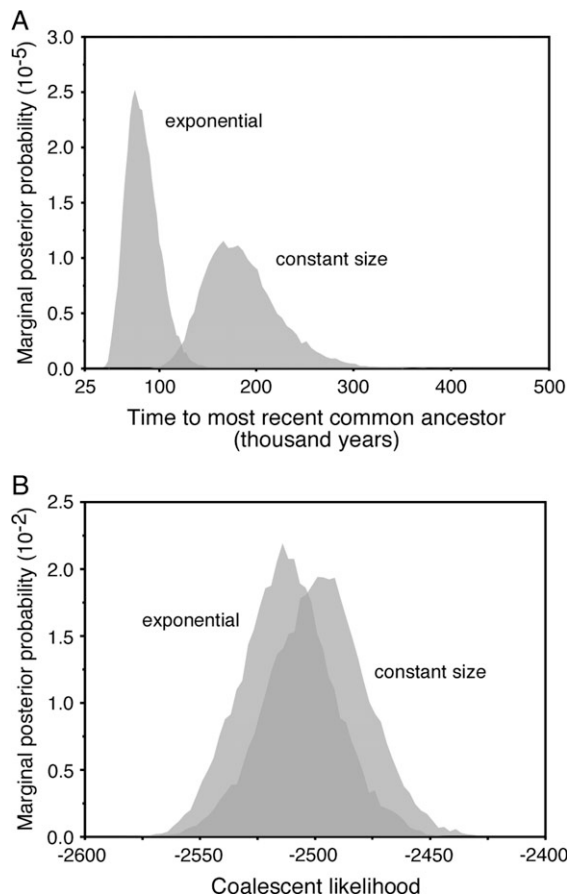
FIG. 4.—Bayesian estimates of TMRCA under constant size and exponential growth models. (*A*) Marginal posterior distribution of *Plasmodium vivax* TMRCA. (*B*) Coalescent likelihood distribution for each model.

Bayesian analysis, expansion and logistic growth showed no significant improvement over the simpler constant size and exponential growth models, suggesting a lack of power in our data to estimate the additional parameters in the more complex models. The marginal posterior distributions of the TMRCA under constant size and exponential growth as well as the coalescent likelihood distribution associated with these models are shown in figure 4. Although AIC scores suggest that constant size is a better fit than exponential growth (table 1) it is clear from figure 4*B* that there is extensive overlap between the likelihood distributions of the two models, suggesting that neither model presents a definitively better fit to our data. This conservatively gives a maximum TMRCA range of 53,000–265,000 years (table 1). Our estimate is older than the ~10,000 years recently proposed based on genomic microsatellites and TR (Leclerc et al. 2004), but overlaps with estimates from Escalante et al. (2005).

Discussion

Our study provides several lines of evidence in support of an Asian origin for *P. vivax*: (1) haplotype network outgroup probabilities suggest that haplotype 3, primarily from India and Africa, is the oldest, (2) historic migration patterns show high migration from India to Africa but not the reverse, and (3) cophylogeny mapping supports a host switch from Asian macaque monkeys to humans. Additionally, the possibility that *P. vivax* is a zoonosis from Asian monkeys is bolstered by the observation that *Plasmodium knowlesi*, an Asian macaque malaria, also infects humans under conditions of natural transmission (Chin et al. 1965; Singh et al. 2004).

Our TMRCA estimate is consistent with an Asian origin but does not exclude the possibility of an African one. Due to the low number of polymorphisms in the mt genome, we were not able to precisely estimate a TMRCA for *P. vivax*, although a minimum cutoff of 53,000 years is strongly supported. There is some evidence that modern humans were in Asia at this time (Shen et al. 2002) though the issue remains controversial (Jin and Su 2000). A selective sweep of the Duffy-negative allele in Africa is estimated to have occurred 33×10^3 years B.P. (95% confidence interval, 97×10^3 to 6×10^3 years B.P.) (Hamblin and Di Rienzo 2000), although more recent estimates have also been proposed (Seixas, Ferrand, and Rocha 2002). Thus, the timing of the selective sweep does not negate a possible Asian origin for *P. vivax*. However, Duffy negativity is a homozygous trait and therefore the period of selection leading up to the sweep would have been considerable, in turn suggesting that *P. vivax* must have been in Africa for an extended period as well (Carter 2003). But a prolonged association between *P. vivax* and Africans would only be necessary if *P. vivax* were the sole selective agent. Alternatively, preexisting $FY \times B^{null}$ alleles may have prevented the establishment of *P. vivax* in Africa to begin with (Livingstone 1984), or precipitated a selective sweep once *P. vivax* arrived. The Duffy blood group antigen functions as a receptor for chemokines, including interleukin 8, and has melanoma growth-stimulatory activity (Horuk et al. 1993), all of which suggests that factors other than *P. vivax* may have exerted selection pressure on this locus.

Not all primate malarias were available for this study. In particular, the phylogenetic placement of *Plasmodium schwetzi*, a parasite of African Great Apes, is unknown. It has been proposed as a close relative of either *P. vivax* (Garnham 1966) or *Plasmodium ovale* (Coatney et al. 2003) based on morphology and other biologic characteristics. Resolution of this question will no doubt provide insight regarding the relationships among Asian and African malaria parasites. Nonetheless, our study provides compelling evidence that *P. vivax* became a human parasite via a host switch from Asian macaques.

Literature Cited

Ayala, F. J., A. A. Escalante, and S. M. Rich. 1999. Evolution of *Plasmodium* and the recent origin of the world populations of *Plasmodium falciparum*. Parassitologia **41**:55–68.

Beerli, P. 2002. MIGRATE v1.7.6. (http://evolution.genetics.washington.edu/lamarc.html).

Carter, R. 2003. Speculations on the origins of *Plasmodium vivax* malaria. Trends Parasitol. **19**:214–219.

Carter, R., and K. N. Mendis. 2002. Evolutionary and historical aspects of the burden of malaria. Clin. Microbiol. Rev. **15**:564–594.

Castelloe, J., and A. R. Templeton. 1994. Root probabilities for intraspecific gene trees under neutral coalescent theory. Mol. Phylogenet. Evol. **3**:102–113.

Charleston, M. A. 1998. Jungles: a new solution to the host/parasite phylogeny reconciliation problem. Math. Biosci. **149**:191–223.

———. 2003. Recent results in cophylogeny mapping. Pp. 303–330 *in* D. T. J. Littewood, ed. Advances in parasitology: the evolution of parasitism—a phylogenetic perspective, 54. Elsevier Academic Press, Burlington, Mass.

Charleston, M. A., and R. D. M. Page. 2002. TreeMap v2.0.2. (http://taxonomy.zoology.gla.ac.uk/%7emac/treemap/index.html).

Chin, W., P. G. Contacos, G. R. Coatney, and H. R. Kimball. 1965. A naturally acquired quotidian-type malaria in man transferable to monkeys. Science **149**:865–866.

Clement, M., D. Posada, and K. A. Crandall. 2000. TCS: a computer program to estimate gene genealogies. Mol. Ecol. **9**:1657–1660.

Coatney, G. R., W. E. Collins, M. Warren, and P. G. Contacos. 2003. The primate malarias. CD-ROM. CDC, Atlanta, Ga [original book published 1971, U.S. Government Printing Office, Washington, D.C.].

Delson, E. 1980. Fossil macaques, phyletic relationships and a scenario of deployment. Pp. 10–29 *in* D. G. Lindburg, ed. The macaques—studies in ecology, behavior and evolution. Van Nostrand Reinhold Co., New York.

Drummond, A., and A. Rambaut. 2003. BEAST v1.1.2. (http://evolve.zoo.ox.ac.uk/beast/).

Escalante, A. A., O. E. Cornejo, D. E. Freeland, A. C. Poe, E. Durrego, W. E. Collins, and A. A. Lal. 2005. A monkey's tale: the origin of *Plasmodium vivax* as a human malaria parasite. Proc. Natl. Acad. Sci. USA **102**:1980–1985.

Escalante, A. A., D. E. Freeland, W. E. Collins, and A. A. Lal. 1998. The evolution of primate malaria parasites based on the gene encoding cytochrome *b* from the linear mitochondrial genome. Proc. Natl. Acad. Sci. USA **95**:8124–8129.

Feng, X., J. M. Carlton, D. A. Joy, J. Mu, T. Furuya, B. B. Suh, Y. Wang, J. W. Barnwell, and X.-Z. Su. 2003. Single-nucleotide polymorphisms and genome diversity in *Plasmodium vivax*. Proc. Natl. Acad. Sci. USA **100**:8502–8507.

Garnham, P. C. C. 1963. Distribution of simian malaria parasites in various hosts. J. Parasitol. **49**:905–911.

———. 1966. Malaria parasites and other Haemosporidia. Blackwell Scientific, Oxford.

Hamblin, M. T., and A. Di Rienzo. 2000. Detection of the signature of natural selection in humans: evidence from the Duffy blood group locus. Am. J. Hum. Genet. **66**:1669–1679.

Hamblin, M. T., E. E. Thompson, and A. D. Rienzo. 2002. Complex signatures of natural selection at the Duffy blood group locus. Am. J. Hum. Genet. **70**:369–383.

Horuk, R., C. E. Chitnis, W. C. Darbonne, T. J. Colby, A. Rybicki, T. J. Hadley, and L. H. Miller. 1993. A receptor for the malarial parasite *Plasmodium vivax*; the erythrocyte chemokine receptor. Science **261**:1182–1184.

Hudson, R. R. 2001. Two-locus sampling distributions and their application. Genetics **159**:1805–1817.

Huelsenbeck, J. P., and F. Ronquist. 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics **17**:754–755.

Jin, L., and B. Su. 2000. Natives or immigrants: modern human origin in East Asia. Nat. Rev. Genet. **1**:126–133.

Joy, D. A., X. Feng, J. Mu et al. (12 co-authors). 2003. Early origin and recent expansion of *Plasmodium falciparum*. Science **300**:318–321.

Leclerc, M. C., P. Durand, C. Gauthier, S. Patot, N. Billotte, M. Menegon, C. Severini, F. J. Ayala, and F. Renaud. 2004. Meager genetic variability of the human malaria agent *Plasmodium vivax*. Proc. Natl. Acad. Sci. USA: **101**: 14455–14460.

Li, J., W. E. Collins, R. A. Wirtz, D. Rathore, A. Lal, and T. F. McCutchan. 2001. Geographic subdivision of the range of the malaria parasite *Plasmodium vivax*. Emerg. Infect. Dis. **7**:35–42.

Livingstone, F. B. 1984. The Duffy blood groups, vivax malaria, and malaria selection in human populations: a review. Hum. Biol. **56**:413–425.

McVean, G. A. T., P. Awadalla, and P. Fearnhead. 2002. A coalescent-based method for detecting and estimating recombination from gene sequences. Genetics **160**:1231–1241.

Miller, L., S. Mason, D. Clyde, and M. McGuiness. 1976. The resistance factor to *Plasmodium vivax* in blacks. The Duffy-blood-group genotype, FyFy. N. Engl. J. Med. **295**: 302–304.

Nielsen, R., and Z. Yang. 1998. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. Genetics **148**:929–936.

Page, S. L., and M. Goodman. 2001. Catarrhine phylogeny: non-coding DNA evidence for a diphyletic origin of the mangabeys and for a human-chimpanzee clade. Mol. Phylogenet. Evol. **18**:14–25.

Perkins, S. L., and J. J. Schall. 2002. A molecular phylogeny of malarial parasites recovered from cytochrome *b* gene sequences. J. Parasitol. **88**:972–978.

Posada, D., and K. A. Crandall. 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics **14**:817–818.

Sawyer, S. 1989. Statistical test for detecting gene conversion. Mol. Biol. Evol. **6**:526–538.

Schrago, C. G., and C. A. M. Russo. 2003. Timing the origin of New World monkeys. Mol. Biol. Evol. **20**:1620–1625.

Seixas, S., N. Ferrand, and J. Rocha. 2002. Microsatellite variation and evolution of the human duffy blood group polymorphism. Mol. Biol. Evol. **19**:1802–1806.

Sharma, I., S. T. Pasha, and Y. D. Sharma. 1998. Complete nucleotide sequence of the *Plasmodium vivax* 6 kb element. Mol. Biochem. Parasitol. **97**:259–263.

Shen, G., W. Wang, Q. Wang, J. Zhao, K. Collerson, C. Zhou, and P. V. Tobias. 2002. U-Series dating of Liujiang hominid site in Guangxi, Southern China. J. Hum. Evol. **43**:817–829.

Singh, P. B., L. Kim Sung, A. Matusop, A. Radhakrishnan, S. S. G. Shamsul, J. Cox-Singh, A. Thomas, and D. J. Conway. 2004. A large focus of naturally acquired *Plasmodium knowlesi* infections in human beings. Lancet **363**:1017–1024.

Stewart, C.-B., and T. R. Disotell. 1998. Primate evolution—in and out of Africa. Curr. Biol. **8**:R582–R588.

Swofford, D. L. 1998. PAUP*: phylogenetic analysis using parsimony (*and other methods). Sinauer Associates, Sunderland, Mass.

Tajima, F. 1993. Simple methods for testing the molecular clock hypothesis. Genetics **135**:599–607.

Templeton, A. R., K. A. Crandall, and C. F. Sing. 1992. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. Genetics **132**:619–633.

Tosi, A. J., J. C. Morales, and D. J. Melnick. 2003. Paternal, maternal, and biparental molecular markers provide unique windows onto the evolutionary history of macaque monkeys. Evolution **57**:1419–1435.

Waters, A., D. Higgins, and T. McCutchan. 1993. Evolutionary relatedness of some primate models of *Plasmodium*. Mol. Biol. Evol. **10**:914–923.

Yang, Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput. Appl. Biosci. **13**:555–556.

Yang, Z., W. S. W. Wong, and R. Nielsen. 2005. Bayes empirical bayes inference of amino acid sites under positive selection. Mol. Biol. Evol. **22**(4):1107–1118.