

# *Plasmodium falciparum* Accompanied the Human Expansion out of Africa

Kazuyuki Tanabe,<sup>1,\*</sup> Toshihiro Mita,<sup>3</sup> Thibaut Jombart,<sup>4</sup> Anders Eriksson,<sup>5</sup> Shun Horibe,<sup>6</sup> Nirianne Palacpac,<sup>2</sup> Lisa Ranford-Cartwright,<sup>7</sup> Hiromi Sawai,<sup>1</sup> Naoko Sakihama,<sup>1</sup> Hiroshi Ohmae,<sup>8</sup> Masatoshi Nakamura,<sup>9</sup> Marcelo U. Ferreira,<sup>10</sup> Ananias A. Escalante,<sup>11</sup> Franck Prugnolle,<sup>12</sup> Anders Björkman,<sup>13</sup> Anna Färnert,<sup>13</sup> Akira Kaneko,<sup>13,14</sup> Toshihiro Horii,<sup>2</sup> Andrea Manica,<sup>5,\*</sup> Hirohisa Kishino,<sup>6</sup> and Francois Balloux<sup>4,\*</sup>

<sup>1</sup>Laboratory of Malariology

<sup>2</sup>Department of Molecular Protozoology  
Research Institute for Microbial Diseases, Osaka University,  
Osaka 565-0871, Japan

<sup>3</sup>Department of International Affairs and Tropical Medicine,  
Tokyo Women's Medical University, Tokyo 162-8666, Japan

<sup>4</sup>MRC Centre for Outbreak Analysis and Modelling,  
Department of Infectious Disease Epidemiology,  
Faculty of Medicine, Imperial College, London W2 1PG, UK

<sup>5</sup>Evolutionary Ecology Group, Department of Zoology,  
University of Cambridge, Downing Street,  
Cambridge CB2 3EJ, UK

<sup>6</sup>Graduate School of Agriculture and Life Sciences,  
University of Tokyo, Tokyo 113-8657, Japan

<sup>7</sup>Division of Infection and Immunity, Faculty of Biomedical and  
Life Sciences, University of Glasgow, Glasgow G12 8TA, UK

<sup>8</sup>Department of Parasitology, National Institute of Infectious  
Diseases, Tokyo 162-3640, Japan

<sup>9</sup>Department of Tropical Medicine and Parasitology, Dokkyo  
Medical University, Tochigi 321-0293, Japan

<sup>10</sup>Department of Parasitology, Institute of Biomedical  
Sciences, University of São Paulo, São Paulo 05508-900, Brazil

<sup>11</sup>School of Life Sciences, Arizona State University,  
Tempe, AZ 85287, USA

<sup>12</sup>Génétique et Evolution des Maladies Infectieuses,  
Institut de Recherche pour le Développement Montpellier,  
34394 Montpellier, Cedex 05, France

<sup>13</sup>Infectious Diseases Unit, Department of Medicine Solna,  
Karolinska Institutet, 17176 Stockholm, Sweden

<sup>14</sup>Global COE Program, Nagasaki University, Sakamoto,  
Nagasaki 852-8523, Japan

## Summary

*Plasmodium falciparum* is distributed throughout the tropics and is responsible for an estimated 230 million cases of malaria every year, with a further 1.4 billion people at risk of infection [1–3]. Little is known about the genetic makeup of *P. falciparum* populations, despite variation in genetic diversity being a key factor in morbidity, mortality, and the success of malaria control initiatives. Here we analyze a worldwide sample of 519 *P. falciparum* isolates sequenced for two housekeeping genes (63 single nucleotide polymorphisms from around 5000 nucleotides per isolate). We observe a strong negative correlation between

within-population genetic diversity and geographic distance from sub-Saharan Africa ( $R^2 = 0.95$ ) over Africa, Asia, and Oceania. In contrast, regional variation in transmission intensity seems to have had a negligible impact on the distribution of genetic diversity. The striking geographic patterns of isolation by distance observed in *P. falciparum* mirror the ones previously documented in humans [4–7] and point to a joint sub-Saharan African origin between the parasite and its host. Age estimates for the expansion of *P. falciparum* further support that anatomically modern humans were infected prior to their exit out of Africa and carried the parasite along during their colonization of the world.

## Results and Discussion

The genetic diversity of malaria parasites is central to their pathogenesis by facilitating immune evasion and drug resistance. As such, a better understanding of the worldwide distribution of the genetic diversity in *Plasmodium falciparum* is crucial for devising optimal drug- or vaccine-based malaria control strategies. Previous attempts at characterizing population structure in *P. falciparum* have pointed to considerable variation in genetic diversity and well-differentiated populations over the parasite's vast distribution range [8–14]. Moreover, no consensus has been reached over worldwide trends in the distribution of genetic diversity, with different studies identifying either Africa or South America as the region harboring the highest genetic diversity [8, 10, 13]. These inconsistencies between studies may be due to populations having been defined by pooling isolates sampled over heterogeneous geographic ranges. Alternatively, these discrepancies could be explained by the type of genetic markers that were deployed. Microsatellite markers may offer lower resolution for picking up patterns generated by relatively old demographic events as a result of their high mutation rate and high rate of back mutation (homoplasmy). Conversely, single nucleotide polymorphisms (SNPs) often suffer from biases induced by the selection of polymorphic markers on a small initial discovery panel. This phenomenon, referred to as ascertainment bias, leads to the genetic diversity of populations that are not well represented in the initial discovery panel being strongly underestimated [15, 16].

The age of *P. falciparum* is also highly disputed [17–22], with previous estimates for the time to the most recent common ancestor (TMRCA) spanning more than an order of magnitude with values ranging from about 10,000 years [19, 22] to over 300,000 years [20]. With *P. falciparum* being an exclusively human parasite with no known animal reservoir, we hypothesized that if *P. falciparum* had been associated with humans for over 50,000 to 60,000 years (the estimated date for the out-of-Africa migration of anatomically modern humans), its current population structure could still carry a signal of human settlement history. Within-population genetic diversity of native human populations decreases smoothly with geographic distance measured through landmasses from a sub-Saharan African origin [5], and genetic differentiation between populations also increases steadily with physical distance along landmasses [7, 23]. These smooth patterns in the

\*Correspondence: kztanabe@biken.osaka-u.ac.jp (K.T.), am315@cam.ac.uk (A.M.), fballoux@imperial.ac.uk (F.B.)

distribution of human genetic diversity have been ascribed to sequential bottlenecks of small amplitude during the colonization of the world by our ancestors from an African cradle. In contrast, the parasite population structure may primarily depend on variation in epidemiological settings between populations. In particular, *P. falciparum* populations are characterized by high variability in variation in transmission intensity [1, 10, 18, 24, 25], which could have affected local genetic diversity. Selective pressure imposed by antimalaria interventions that used drugs and insecticides might, likewise, have locally reduced genetic diversity.

To assess the relative importance of past human demography and recent epidemiological factors, we used a data set of 519 *P. falciparum* isolates from nine populations covering the entire distribution range of the parasite, with populations from sub-Saharan Africa, Southeast Asia, Oceania, and South America. Care was taken to obtain a representative picture of the genetic makeup of *P. falciparum* populations by trying to minimize the geographic and temporal range of the isolates pooled into populations. We also chose to sequence two housekeeping genes, P type  $\text{Ca}^{2+}$ -ATPase (*serca*) and adenylosuccinate lyase (*adsl*), in their entirety for all isolates to circumvent any possible ascertainment bias induced by marker selection.

We identified 49 and 14 SNPs in the *serca* (3630 bp) and *adsl* (1413 bp) genes, respectively (see Table S1 available online). The ratio of synonymous ( $d_S$ ) over nonsynonymous ( $d_N$ ) substitutions did not significantly deviate from neutral expectations overall or in any of the nine parasite populations considered individually (Table S1). Because there was no evidence for natural selection, we present all results on analyses using both the synonymous and nonsynonymous SNPs. However, using synonymous SNPs only did not affect any of the conclusions (data not shown). Comparison of average pairwise nucleotide differences ( $\theta_\pi$ ) and standardized number of segregating sites ( $\theta_S$ ) indicates an excess of rare alleles in line with the high frequency of variants observed in a single population (private alleles), except in the two South American populations (Table S1; Figure S1).

Of the populations studied, those from Africa are by far the most genetically diverse, with the exception of Brazil, which is as diverse as Tanzania but only for the  $\theta_\pi$  estimate for the *serca* gene (Table S1). However, we found no overall significant difference in within-population genetic diversity between the two genes. Thus, both genes were analyzed together in all subsequent analyses. The overall genetic differentiation is high, with a global  $F_{ST} = 0.21$ . Between-population genetic distances (pairwise  $F_{ST}$  estimates; Figure S2) are highest for the two South American populations, which are strongly differentiated from all other populations, with the largest of all pairwise differentiation observed between the two South American populations ( $F_{ST} = 0.47$ ; Figure S2).

To test for an association between humans and *P. falciparum* predating the out-of-Africa exit, we computed the shortest distance through landmasses between each sampled population and a grid of hypothetical origins covering the entire world. We did not include the two South American populations in this analysis. The Americas were first colonized by humans some 15,000–20,000 years ago through the Bering Strait connecting Siberia to Alaska. The climate found today in the arctic is far too cold for the development of *P. falciparum*, which is unable to fulfill its life cycle at a temperature below 16°C–18°C [26, 27]. Because the temperature at the Bering Strait was considerably colder for the previous

90,000 years than it has been over the last 10,000 years [28], a joint colonization of *P. falciparum* together with humans into the Americas is highly implausible. Thus, the Americas have probably been colonized by this parasite far more recently, possibly through the slave trade [10, 17]. Interestingly, the resulting admixture could explain the relatively high level of parasite genetic diversity observed within American populations. We assumed that the cost of moving along coastlines was half the friction of movement inland. We also allowed for a southern route of migration out of Africa [29] by creating land bridges on either side of the Arabic Peninsula at the Bab-el-Mandeb Strait and Hormuz Strait.

Using these geographic distances and within-population genetic diversities ( $\theta_\pi$ ), we searched for the hypothetical origin providing the best correlation between genetic diversity and the logarithm of geographic distance (Figure 1). We observed the highest correlation for a central sub-Saharan African origin ( $R^2 = 0.95$ ;  $p < 0.0001$ ; Figure 1). Although this origin coincides with the one previously inferred for anatomically modern humans based on both genetic and morphological data [30], this result should be taken with some caution. The exact position of the inferred origin depends heavily on the genetic diversity of two African *P. falciparum* populations in the data set. Although the pattern is robust to the removal of either of the African populations (i.e., one at a time), ultimate confirmation of a coinciding geographic origin for both the parasite and its human host will require the inclusion of additional populations.

As previously reported in humans, there is also a strong isolation by distance (IBD) pattern, with pairwise genetic distances (pairwise  $F_{ST}$ ) being tightly correlated with geographic distance, computed as above (Mantel  $r_M = 0.68$ ;  $p < 0.001$ ; Figure 2). The results are robust to the underlying assumptions in the analysis. We recover the same origin in sub-Saharan Africa with a correlation of  $R^2 = 0.95$  if we use  $\theta_S$  instead of  $\theta_\pi$  genetic diversity estimates. Similarly, assuming an equal cost between coastlines and inland or a 3× lower cost for coastlines again points to a sub-Saharan African origin, with respective variance explained of 0.98 and 0.93 and IBDs of 0.46 and 0.66. Finally, forcing a northern route out of Africa through the Sinai into the Levant by closing the land bridges in and out of the Arabic Peninsula has a negligible effect on the best supported origin, with a correlation between geography and genetic diversity of  $R^2 = 0.95$  and an IBD of  $r_M = 0.68$ .

The smooth patterns in the apportionment of *P. falciparum* genetic diversity mirroring the ones previously described in native human populations are suggestive of an extensive association between the parasite and its host. However, this is not sufficient to conclude that humans were infected prior to their migration out of Africa and carried the parasite along during their colonization of the Old World. *P. falciparum* genetic diversity may be driven by a variable that we did not consider but that is strongly correlated with distance from Africa. There are a variety of such candidate factors, ranging from the availability of competent insect vectors to resistance alleles in the human host. Because it is not feasible to consider sequentially a vast number of variables, we instead used the entomological inoculation rate (EIR), an inclusive metric capturing current transmission intensity, which has been previously shown to correlate with *P. falciparum* genetic diversity, as determined by microsatellite markers [10]. The intensity of malaria transmission varies greatly among geographic areas, with the highest transmission levels recorded in Africa and Oceania (particularly on Papua New Guinea and the

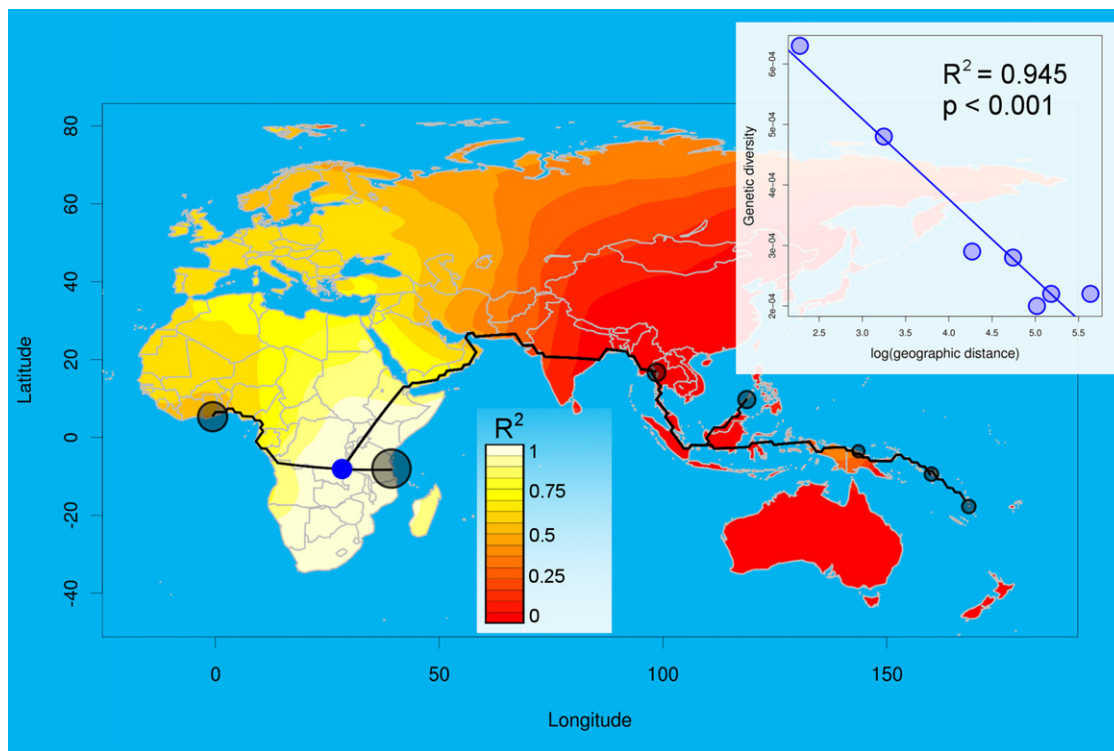


Figure 1. Map of Most Likely Origin for *Plasmodium falciparum*

Grey circles represent the geographic locations of the populations analyzed, with the surface of the circles proportional to within-population genetic diversity ( $\theta_{pi}$ ). The background color represents the strength of the correlation between geographic distance from various origins and genetic diversity, with light shades representing more likely origins. The best supported origin is indicated by a blue dot. We also represented the colonization routes from this origin through landmasses to all populations analyzed (see [Experimental Procedures](#)). The inset represents the correlation between geographic distance (measured as travel cost over friction routes) and genetic diversity for the best supported origin.

Solomon Islands) and much lower transmission intensity in Southeast Asia (Thailand and the Philippines) [25, 31]. EIR rates were obtained for all non-American populations from the literature (Table S2). The correlation between distance from Africa and genetic diversity remains highly significant after controlling for EIR ( $R^2 = 0.81$ ;  $p = 0.0015$ ), suggesting that sequential bottlenecks during the expansion out of Africa are the main determinant of the global patterns in within-population diversity of *P. falciparum*.

Alternatively, *P. falciparum* may have originated in sub-Saharan Africa much more recently than anatomically modern humans but may have followed similar colonization routes to its human host during its expansion. To test for this possibility, we developed a new approximate Bayesian computation (ABC) framework of forward simulations of serial population expansions. This allowed us to estimate, among other parameters, the timing of the initial expansion of *P. falciparum*, as well as colonization rates and subsequent migration between colonized demes.

To parameterize the simulations, we needed to infer a substitution rate. To this end, we sequenced the two housekeeping genes in the chimpanzee malaria parasite *P. reichenowi* and estimated the substitution rate for the *serca* and *adsl* genes. There is no consensus on the exact timing of the split between *P. falciparum* and *P. reichenowi*. However, a range between around 2.5 million years ago (Mya) and 6 Mya is generally considered to be credible [21, 32–34]. There have also been speculations that the split between the two parasites may have coincided with the advent of agriculture in West Asia

and could be as recent as 10,000 years ago [19, 22]. Although the latter calibration date leads to an implausibly high substitution rate for eukaryotic coding genes of about  $10^{-6}$ /site/year, we still decided to consider this hypothesis of a very recent host transfer.

We ran 4 million simulations for each of the three substitution rates (based on the assumption of a 10,000 year, 2.5 Mya and 6 Mya split between the two malaria species). The 10,000 year split turned out to be highly implausible given the data, because we obtained not a single simulation with a reasonable fit to the observed pattern. Conversely, the model provided excellent fits between predicted and observed within-population genetic diversity  $\ln(\theta_{pi})$ , with  $R^2 = \sim 0.98$  for the best combination of parameters for both scenarios with fast (split at 2.5 Mya) and slow (split at 6 Mya) mutation rates. The model with a faster mutation rate pointed to the start of the spread out of Africa at around 40 thousand years ago (Kya) (90% confidence interval: 19–77 Kya), whereas the slower mutation rate gave 80–90 Kya (33–96 Kya) (Figure 3; Table S3). These dates for the initial expansion of *P. falciparum* are compatible with the human out-of-Africa expansion some 50–60 Kya ago. Importantly, the 90% credibility intervals clearly exclude the advent of agriculture.

The demographic parameters were remarkably similar between the two scenarios (Figure S3). The only exception was the size of the ancestral population ( $K_0$ ), which was greater for the slower mutation rate. This is the result of the TMRCA being accounted for by a combination of the starting time of the expansion and the TMRCA in the founding

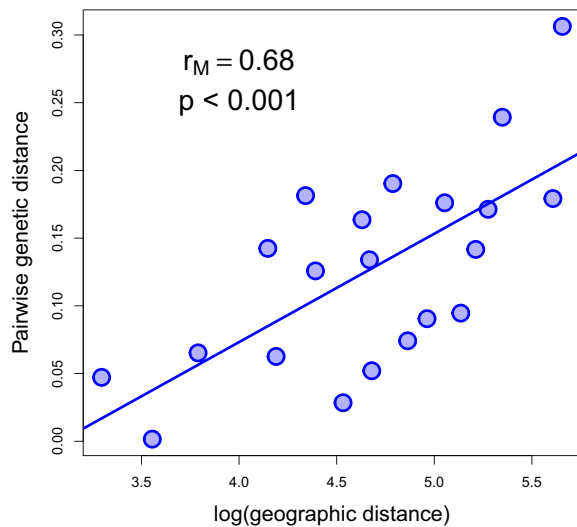


Figure 2. Plot of Isolation by Distance

Logarithm of pairwise geographic distances along landmasses (measured as travel costs over the friction routes) plotted against pairwise genetic distances ( $F_{ST}$ ) between populations.

population (determined by  $K_0$ ). In general, both scenarios were characterized by a large ancestral population and relatively large carrying capacities for the demes during the expansion. Intense bottlenecks (Figure S3) were needed to achieve the sharp decrease in within-population genetic variability observed in the data. Population growth rate (which defines the speed of expansion) was relatively fast. Conversely, migration among demes was low, indicating that the pattern in genetic diversity was shaped primarily during the colonization of the tropics by *P. falciparum* rather than by subsequent exchanges between populations. Interestingly, the scenario we recover for the expansion of *P. falciparum* mirrors previous results for the human expansion out of Africa, with the same pattern of intense bottlenecks and low subsequent migration [35].

Our results establish that the genetic diversity of *P. falciparum* has been primarily shaped by human demography and does not provide evidence for a significant effect of contemporary nationwide malaria interventions based on mass drug administration and/or widespread spraying of DDT. For example, intensive malaria control programs in the Solomon Islands, initiated in the 1970s and since discontinued, led to a massive temporary reduction in parasite incidence [36]. Despite this, the genetic diversity found in the Solomon Islands does not fall below the curve for the genetic diversity, as predicted by distance from Africa (Figure 1). The results also suggest that geographic variation in the distribution of insect vector species may have played only a minor role in shaping the population structure of *P. falciparum*, despite extensive variation in anthropophily and behavior between *Anopheles* species [25].

In this paper, we have shown that the population genetic structure of *P. falciparum* outside of the Americas is primarily explained by geography, with 95% of the variance in within-population genetic diversity explained by physical distance from a sub-Saharan African origin alone. We further recovered strong patterns of isolation by distance and age estimates for the spread of *P. falciparum* coinciding with the colonization by anatomically modern humans of Africa, Asia, and Oceania.

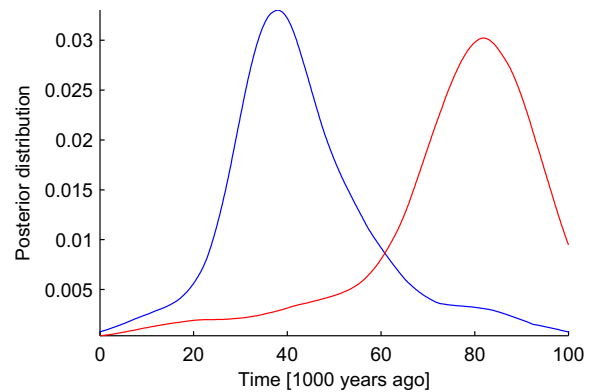


Figure 3. Posterior Distributions of the Date of Expansion of *P. falciparum* These curves represent the posterior distributions of the date of expansion of *P. falciparum* obtained by the approximate Bayesian computation analysis. The two curves have been generated under the assumption of a split between the two *Plasmodium* species (*P. falciparum* and *P. reichenowi*) at 2.5 (blue) and 6 (red) million years ago.

Taken together, our results confirm that *P. falciparum* had already infected humans before the out-of-Africa migration and followed our ancestors in their expansion throughout the tropics, with the exception of South America, which *P. falciparum* probably only reached much more recently through the slave trade. Although an association between humans and *P. falciparum* predating the exit out of Africa has been suggested before [18, 20, 37], our results additionally demonstrate the dramatic impact of past human migrations on the current apportionment of genetic diversity in the parasite.

#### Experimental Procedures

##### DNA Sample Collection

*P. falciparum* isolates were collected from nine countries: Tanzania, Ghana, Thailand, The Philippines, Papua New Guinea (PNG), the Republic of Vanuatu, the Solomon Islands, Brazil, and Venezuela. Details of the samples from Tanzania, Thailand, The Philippines, the Solomon Islands, Vanuatu, and Brazil used in this study have been described previously [31, 38–44]. Briefly, in Tanzania, blood samples were collected from infected individuals in the Rufiji River Delta in eastern coastal Tanzania in 1993, 1998, and 2003 [40]. In Thailand, blood samples were collected from malaria patients attending a malaria clinic in Mae Sot near the northwestern Thailand-Myanmar border in 1995 [42]. In the Philippines, samples were collected from malaria patients attending hospitals in Palawan Island in 1997 [31]. In the Solomon Islands, samples were collected from infected individuals in northeastern Guadalcanal Island in 1995–1996 [39]. In Vanuatu, samples were collected from four islands, Malakula, Gaua, Esprit Santo, and Pentecost, between 1996 and 1998 [38, 42]. In Brazil, isolates were collected from malaria patients in five states, Acre, Rondonia, Mato Grosso, Para, and Amapa, between 1985 and 1999 [43]. Additional samples were collected from Acre in 2004–2005 [44]. Venezuelan isolates were collected from Upper Orinoco, Amazonas State, Venezuela in 1997 (A.A.E., unpublished data). In all cases, ethical clearance for sampling was obtained from relevant ethical committees. Isolates from Ghana and PNG were collected specifically for this study. In Ghana, 182 *P. falciparum*-infected blood samples were collected during malaria surveys from 0- to 15-year-old children in three villages (Okyereko, Mpota, and Apam) near Winneba, a western coastal region, in November 2004. The study was approved by the Ministry of Health/Ghana Health Service. In PNG, 195 malaria-infected blood samples were collected during surveys in five villages in Wewak, East Sepik Province along the northeast coast: Kiniambu in August 2001, Jawia and Witupe in September 2001, and Boiken and Wingei in February 2002. The study was approved by the National Department of Health Medical Research Advisory Committee of PNG and the Tokyo Women's Medical University Ethical Committee. Informed consent was obtained from the patients or their

parents. In both Ghana and PNG, finger-prick blood was collected on Whatman 31ETCHR filter paper. Parasite genomic DNA was extracted from filter blots using the EZ1 DNA Investigator kit on the EZ1 BioRobot (QIAGEN).

#### DNA Sequencing

Isolates infected with mixed *msp1* haplotypes, as determined by polymerase chain reaction (PCR)-based haplotyping [31], were excluded from further analysis, and only those with a single *msp1* haplotype infection ( $n = 519$ ) were used for sequencing. Full-length sequences were obtained for two housekeeping genes,  $\text{Ca}^{2+}$ -transporting ATPase gene (*serca*) and adenylosuccinate lyase gene (*adsl*). The  $\text{Ca}^{2+}$ -transporting ATPase of *P. falciparum* has recently been suggested to be a potential target of artemisinins [45], antimalarial drugs currently widely used for treatment in many endemic countries. However, all isolates in this study were collected before the adoption of this drug. Genomic DNA was subjected to two independent PCR amplifications, the products of which were directly sequenced in both directions. Procedures and conditions used for PCR amplification and sequencing of *serca* have been described elsewhere [46]. For amplification of *adsl*, the first PCR was run using primers ASL-F3 (5'-TATACTCCCC AAAACAAAACCACTAAAATGT) and ASL-R4 (5'-AAAGCGGTACATGTTATA AGGTCCT), followed by a nested PCR using primers ASL-F2 (5'-ATTTATA TATATTCCTTATTATATAGTCA) and ASL-R3 (5'-TGGGAGTGCCCAACTG CAGTGTCT). Full-length *adsl* sequence was also obtained from *P. reichenowi*, a chimpanzee malaria parasite closely related to *P. falciparum*. Whole-genome-amplified *P. reichenowi* genomic DNA [44] was subjected to PCR using three primer sets targeting the 5', central, and 3' regions of the gene: ASL-F3 and ASL-R6 (5'-CAATTATATAAGCATAAACCATATGCT), ASL-F6 (5'-AAATTGGAAGTAGTACCATGCCACA) and ASL-R4, and ASL-F5-2 (5'-GACAAATCATGATGTTAAGCGGTTGA) and ASL-R5-2 (5'-CTTAT TGGTAATTTGGAATAAATAACTTGA). PCR conditions were identical for *serca*. Whenever there was an inconsistency between two sequences after independent amplifications, a third round of PCR and sequencing was performed. Only isolates with a single genotype infection, as judged by the lack of overlapping peaks on electropherograms, were used for analysis. The sequences reported in this study have been deposited in the DNA Data Bank of Japan/European Molecular Biology Laboratory/GenBank database (accession numbers AB501575-AB502442, AB AB519183, and AB520081-AB520237).

#### Statistical Genetics Analyses

Two DNA sequence regions of Asn-codon tandem repeats in *serca* were excluded from the analyses. Nucleotide diversity was estimated by  $\theta_{\pi}$ , the average pairwise nucleotide distance, and by  $\theta_S$ , the standardized number of polymorphic sites per site (Watson's estimator) using DnaSP version 4.10 [47]. A Z test was applied to test for a difference of  $\theta_S$  and  $\theta_{\pi}$  between populations. The mean numbers of synonymous substitutions (*dS*) and non-synonymous substitutions (*dN*) were estimated by the Nei and Gojobori method [48] with a Jukes and Cantor correction, as implemented in the MEGA software version 3.1 [49]. Standard error was determined by 1000 bootstrap replications, and *dN* and *dS* were compared with a Z test using MEGA. The interpopulation variance in allele frequencies,  $F_{ST}$ , was calculated using Weir and Cockerham's  $\theta$  estimator [50] with Arlequin version 3.1 [51].

#### Spatially Explicit Analyses

All geographic distances were computed via graph theory [5, 23] as shortest distances along landmasses within a spherical referential of 40,962 vertices. We assumed that the friction cost along coastlines was half that of moving inland. Land bridges were created between the Malay Peninsula and Australia, connecting the major Indonesian islands. We further assumed two land bridges on either side of the Arabic peninsula at the Bab-el-Mandeb Strait and Hormuz Strait to allow for a southern route of colonization out of Africa [29]. The best supported origin was inferred by searching for the shortest routes to all analyzed populations from 312 hypothetical origins on land arranged on a regular grid. We controlled for EIRs using partial correlation. We tested whether within-population diversity was still significantly correlated to geographic distance from Africa once EIRs were accounted for first. For the isolation by distance analysis, a matrix of pairwise physical distances was computed for all populations using the distance through landmasses described above.

#### Approximate Bayesian Computation

We modeled the expansion of malaria out of Africa by considering a one-dimensional stepping stone, an approach successfully used to investigate

the out-of-Africa expansion of anatomically modern humans, as well as *Helicobacter pylori* [6, 7, 35, 52]. Although the real expansion would have followed a two-dimensional spread, numerical studies suggest that a one-dimensional framework is a reasonable approximation as long as the distance between samples is large. We used 300 demes of equal sizes to represent the coastal route from the origin in Africa to Oceania (c.f. Figure 1). In order to avoid potential boundary effects at the origin, we added 50 demes to the African end (making the origin deme 51 in the chain).

Before the spread of *P. falciparum*, we imagine a well-mixed population of size  $K_0$  existed that represents the early presence of *P. falciparum* in Africa. At the onset of the spread, a seed population of size  $c_0 K_0$  was placed at the origin. This population increased linearly with rate  $r$  until it reached size  $K$ , the carrying capacity for all demes in the simulation (in other words,  $K$  is the effective population size of a deme at carrying capacity). At each time step, demes that had reached their maximum size were allowed to send out colonists to adjacent empty demes and migrants to adjacent demes that had already been colonized. The number of colonists was given by the colonization rate  $c$  multiplied by the deme size  $K$ , and the number of migrants by the migration rate  $m$  multiplied by  $K$  (migrants were shared equally between the two adjacent neighbors). Once a deme was colonized by some individuals, its population increased linearly with growth rate  $r$  until it reached its maximum size,  $K$ . *P. falciparum* undergoes about six generations a year [10]. After testing that generation time (scaled for mutations) up to 60 $\times$  slower (1/10 years) had no qualitative impact on the simulation results, we settled for a computationally reasonable compromise of one generation per year.

From the demography described above, we generated gene genealogies for the two unlinked genes (*serca* 3630 bp and *adsl* 1413 bp) according to the Wright-Fisher model: individuals were assumed to be randomly mating within each deme, and generations were nonoverlapping. Assuming a Poisson process, we then simulated mutations on the gene genealogies; three mutation rates,  $9.18 \times 10^{-7}$ /site/year,  $3.67 \times 10^{-9}$ /site/year, and  $1.53 \times 10^{-9}$ /site/year, were estimated from the data, assuming a divergence between *P. falciparum* and *P. reichenowi* at 10,000, 2.5 million, or 6 million years ago [19, 21, 22, 32-34]. For each simulated gene genealogy, we then computed for each deme  $\theta_{\pi}$ , the average number of pairwise differences per site between sequences within a deme. Rather than using the raw values of  $\theta_{\pi}$ , which tend to have a highly skewed distribution, we applied a logarithmic transformation on  $\theta_{\pi}$ , giving us a variable with an approximately symmetrical distribution.

We estimated the best parameter values that describe the spread of *P. falciparum* via ABC, including weighted local regression [53], using within-population  $\ln(\theta_{\pi})$  as our summary statistics. The stepping-stone parameter values for each simulation were sampled from uniform prior distributions of the log values of the following ranges:  $K$   $10^{-10}$ – $10^6$ ,  $K_0$   $10^{-10}$ – $10^6$ ,  $m$   $10^{-6}$ – $0.5$ ,  $c$   $10^{-6}$ – $0.5$ ,  $c_0$   $10^{-6}$ – $1$ , and  $r$   $10^{-3}$ – $1$ . The start of the simulation  $t$  had a uniform distribution from 1,000 to 100,000 years ago. We also enforced two constraints:  $c K \geq 1$  and  $c_0 K_0 \geq 1$ . We ran 4 million simulations for each of the three mutation rates, with an acceptance criterion of  $R^2 > 0.7$  between observed and predicted  $\ln(\theta_{\pi})$ . For the fastest mutation rate (referring to a split between the *Plasmodium* species 10,000 years ago), we were unable to obtain any simulation that fitted the data well ( $R^2 > 0.7$ ). For this reason, no results are shown, because this mutation rate is implausible given the data. For the other two mutation rates, we accepted 6352 and 8943 out of 4 million simulations for the split at 2.5 Mya and 6 Mya, respectively.

#### Accession Numbers

All sequences analyzed in this work have been deposited in the GenBank database under accession numbers AB501575-AB502442, AB519183, and AB520081-AB520237.

#### Supplemental Information

Supplemental Information includes three tables and three figures and can be found with this article online at doi:10.1016/j.cub.2010.05.053.

#### Acknowledgments

This study is dedicated to the late David Walliker. We thank all those who participated in the epidemiological studies for their kind cooperation, particularly I. Rooth, M. Dzodzomenyo, H. Eto, T. Tsukahara, F. Hombhanje, H. Osawa, I. Hwaihwanje, K. Hirayama, K. Na-Bangchang, A. Palanca, Jr.,

R. Espina, and B. Bakote'e. We also want to acknowledge three anonymous reviewers for insightful suggestions. This work was supported by the Ministry of Education, Culture, Sports, Science and Technology, Japan (18073013), Japan Society for the Promotion of Sciences (18GS03140013, 20390120), the Ministry of Health, Labour and Welfare (H21-Shinkou-ippan), and Takeda Science Foundation, Japan. A.E. and A.M. were funded by the Leverhulme Trust. F.B. and A.M. acknowledge financial support from the Biotechnology and Biological Sciences Research Council and the Medical Research Council.

Received: December 16, 2009

Revised: May 11, 2010

Accepted: May 12, 2010

Published online: June 17, 2010

## References

- Hay, S.I., Guerra, C.A., Gething, P.W., Patil, A.P., Tatem, A.J., Noor, A.M., Kabaria, C.W., Manh, B.H., Elyazar, I.R.F., Brooker, S., et al. (2009). A world malaria map: *Plasmodium falciparum* endemicity in 2007. *PLoS Med.* 6, e1000048.
- Snow, R.W., Guerra, C.A., Noor, A.M., Myint, H.Y., and Hay, S.I. (2005). The global distribution of clinical episodes of *Plasmodium falciparum* malaria. *Nature* 434, 214–217.
- World Health Organization. (2008). World Malaria Report, Volume WHO/HTM/GMP/2008.1 (Geneva: WHO Press).
- Handley, L.J.L., Manica, A., Goudet, J., and Balloux, F. (2007). Going the distance: Human population genetics in a clinal world. *Trends Genet.* 23, 432–439.
- Prugnolle, F., Manica, A., and Balloux, F. (2005). Geography predicts neutral genetic diversity of human populations. *Curr. Biol.* 15, R159–R160.
- Liu, H., Prugnolle, F., Manica, A., and Balloux, F. (2006). A geographically explicit genetic model of worldwide human-settlement history. *Am. J. Hum. Genet.* 79, 230–237.
- Ramachandran, S., Deshpande, O., Roseman, C.C., Rosenberg, N.A., Feldman, M.W., and Cavalli-Sforza, L.L. (2005). Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. *Proc. Natl. Acad. Sci. USA* 102, 15942–15947.
- Mu, J.B., Awadalla, P., Duan, J.H., McGee, K.M., Joy, D.A., McVean, G.A.T., and Su, X.Z. (2005). Recombination hotspots and population structure in *Plasmodium falciparum*. *PLoS Biol.* 3, e335.
- Mu, J.B., Awadalla, P., Duan, J.H., McGee, K.M., Keebler, J., Seydel, K., McVean, G.A.T., and Su, X.Z. (2007). Genome-wide variation and identification of vaccine targets in the *Plasmodium falciparum* genome. *Nat. Genet.* 39, 126–130.
- Anderson, T.J.C., Haubold, B., Williams, J.T., Estrada-Franco, J.G., Richardson, L., Mollinedo, R., Bockarie, M., Mokili, J., Mharakurwa, S., French, N., et al. (2000). Microsatellite markers reveal a spectrum of population structures in the malaria parasite *Plasmodium falciparum*. *Mol. Biol. Evol.* 17, 1467–1482.
- Jeffares, D.C., Pain, A., Berry, A., Cox, A.V., Stalker, J., Ingle, C.E., Thomas, A., Quail, M.A., Siebenthal, K., Uhlemann, A.C., et al. (2007). Genome variation and evolution of the malaria parasite *Plasmodium falciparum*. *Nat. Genet.* 39, 120–125.
- Volkman, S.K., Sabeti, P.C., DeCaprio, D., Neafsey, D.E., Schaffner, S.F., Milner, D.A., Jr., Daily, J.P., Sarr, O., Ndiaye, D., Ndir, O., et al. (2007). A genome-wide map of diversity in *Plasmodium falciparum*. *Nat. Genet.* 39, 113–119.
- Neafsey, D.E., Schaffner, S.F., Volkman, S.K., Park, D., Montgomery, P., Milner, D.A., Jr., Lukens, A., Rosen, D., Daniels, R., Houde, N., et al. (2008). Genome-wide SNP genotyping highlights the role of natural selection in *Plasmodium falciparum* population divergence. *Genome Biol.* 9, R171.
- Volkman, S.K., Barry, A.E., Lyons, E.J., Nielsen, K.M., Thomas, S.M., Choi, M., Thakore, S.S., Day, K.P., Wirth, D.F., and Hartl, D.L. (2001). Recent origin of *Plasmodium falciparum* from a single progenitor. *Science* 293, 482–484.
- Rogers, A.R., and Jorde, L.B. (1996). Ascertainment bias in estimates of average heterozygosity. *Am. J. Hum. Genet.* 58, 1033–1041.
- Romero, I.G., Manica, A., Goudet, J., Handley, L.L., and Balloux, F. (2009). How accurate is the current picture of human genetic variation? *Heredity* 102, 120–126.
- Hartl, D.L. (2004). The origin of malaria: Mixed messages from genetic diversity. *Nat. Rev. Microbiol.* 2, 15–22.
- Joy, D.A., Feng, X.R., Mu, J.B., Furuya, T., Chotivanich, K., Krettli, A.U., Ho, M., Wang, A., White, N.J., Suh, E., et al. (2003). Early origin and recent expansion of *Plasmodium falciparum*. *Science* 300, 318–321.
- Rich, S.M., Licht, M.C., Hudson, R.R., and Ayala, F.J. (1998). Malaria's Eve: Evidence of a recent population bottleneck throughout the world populations of *Plasmodium falciparum*. *Proc. Natl. Acad. Sci. USA* 95, 4425–4430.
- Hughes, A.L., and Verra, F. (2001). Very large long-term effective population size in the virulent human malaria parasite *Plasmodium falciparum*. *Proc. Biol. Sci.* 268, 1855–1860.
- Escalante, A.A., and Ayala, F.J. (1994). Phylogeny of the malarial genus *Plasmodium*, derived from rRNA gene sequences. *Proc. Natl. Acad. Sci. USA* 91, 11373–11377.
- Rich, S.M., Leendertz, F.H., Xu, G., LeBreton, M., Djoko, C.F., Aminake, M.N., Takang, E.E., Diffo, J.L.D., Pike, B.L., Rosenthal, B.M., et al. (2009). The origin of malignant malaria. *Proc. Natl. Acad. Sci. USA* 106, 14902–14907.
- Manica, A., Prugnolle, F., and Balloux, F. (2005). Geography is a better determinant of human genetic differentiation than ethnicity. *Hum. Genet.* 118, 366–371.
- Kelly-Hope, L.A., and McKenzie, F.E. (2009). The multiplicity of malaria transmission: A review of entomological inoculation rate measurements and methods across sub-Saharan Africa. *Malar. J.* 8, 19.
- Kiszewski, A., Mellinger, A., Spielman, A., Malaney, P., Sachs, S.E., and Sachs, J. (2004). A global index representing the stability of malaria transmission. *Am. J. Trop. Med. Hyg.* 70, 486–498.
- Molineaux, L. (1988). The epidemiology of human malaria as an explanation of its distribution, including some implications for its control. In *Malaria: Principles and Practice of Malariology*, W.H. Wernsdorfer and I. McGregor, eds. (Edinburgh: Churchill Livingstone), pp. 913–998.
- Onori, E., and Grab, B. (1980). Indicators for the forecasting of malaria epidemics. *Bull. World Health Organ.* 58, 91–98.
- De Boer, A.M., and Nof, D. (2004). The Bering Strait's grip on the northern hemisphere climate. *Deep Sea Res.* 51, 1347–1366.
- Macaulay, V., Hill, C., Achilli, A., Rengo, C., Clarke, D., Meehan, W., Blackburn, J., Semino, O., Scozzari, R., Cruciani, F., et al. (2005). Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. *Science* 308, 1034–1036.
- Manica, A., Amos, W., Balloux, F., and Hanihara, T. (2007). The effect of ancient population bottlenecks on human phenotypic variation. *Nature* 448, 346–348.
- Sakihama, N., Nakamura, M., Palanca, A.A., Jr., Argubano, R.A., Realon, E.P., Larracas, A.L., Espina, R.L., and Tanabe, K. (2007). Allelic diversity in the merozoite surface protein 1 gene of *Plasmodium falciparum* on Palawan Island, the Philippines. *Parasitol. Int.* 56, 185–194.
- Ollomo, B., Durand, P., Prugnolle, F., Douzery, E., Amathau, C., Nkoghe, D., Leroy, E., and Renaud, F. (2009). A new malaria agent in African hominids. *PLoS Pathog.* 5, e1000446.
- Hayakawa, T., Culleton, R., Otani, H., Horii, T., and Tanabe, K. (2008). Big bang in the evolution of extant malaria parasites. *Mol. Biol. Evol.* 25, 2233–2239.
- Krief, S., Escalante, A.A., Pacheco, M.A., Mugisha, L., André, C., Halbwax, M., Fischer, A., Krief, J.-M., Kasenene, J.M., Crandfield, M., et al. (2010). On the diversity of malaria parasites in African apes and the origin of *Plasmodium falciparum* from Bonobos. *PLoS Pathog.* 6, e1000765.
- Deshpande, O., Batzoglou, S., Feldman, M.W., and Cavalli-Sforza, L.L. (2009). A serial founder effect model for human settlement out of Africa. *Proc. Biol. Sci.* 276, 291–300.
- Over, M., Bakote'e, B., Velayudhan, R., Wilikai, P., and Graves, P.M. (2004). Impregnated nets or ddt residual spraying? Field effectiveness of malaria prevention techniques in Solomon Islands, 1993–1999. *Am. J. Trop. Med. Hyg.* 71(2, Suppl), 214–223.
- Escalante, A.A., Barrio, E., and Ayala, F.J. (1995). Evolutionary origin of human and primate malaria: Evidence from the circumsporozoite protein gene. *Mol. Biol. Evol.* 12, 616–626.
- Tanabe, K., Sakihama, N., and Kaneko, A. (2004). Stable SNPs in malaria antigen genes in isolated populations. *Science* 303, 493.
- Sakihama, N., Ohmae, H., Bakote'e, B., Kawabata, M., Hirayama, K., and Tanabe, K. (2006). Limited allelic diversity of *Plasmodium*

- falciparum* merozoite surface protein 1 gene from populations in the Solomon Islands. *Am. J. Trop. Med. Hyg.* 74, 31–40.
40. Tanabe, K., Sakihama, N., Rooth, I., Björkman, A., and Färnert, A. (2007). High frequency of recombination-driven allelic diversity and temporal variation of *Plasmodium falciparum* msp1 in Tanzania. *Am. J. Trop. Med. Hyg.* 76, 1037–1045.
  41. Sakihama, N., Kaneko, A., Hattori, T., and Tanabe, K. (2001). Limited recombination events in merozoite surface protein-1 alleles of *Plasmodium falciparum* on islands. *Gene* 279, 41–48.
  42. Sakihama, N., Kimura, M., Hirayama, K., Kanda, T., Na-Bangchang, K., Jongwutiwes, S., Conway, D., and Tanabe, K. (1999). Allelic recombination and linkage disequilibrium within Msp-1 of *Plasmodium falciparum*, the malignant human malaria parasite. *Gene* 230, 47–54.
  43. Silva, N.S., Silveira, L.A., Machado, R.L.D., Póvoa, M.M., and Ferreira, M.U. (2000). Temporal and spatial distribution of the variants of merozoite surface protein-1 (MSP-1) in *Plasmodium falciparum* populations in Brazil. *Ann. Trop. Med. Parasitol.* 94, 675–688.
  44. Ferreira, M.U., Karunaweera, N.D., da Silva-Nunes, M., da Silva, N.S., Wirth, D.F., and Hartl, D.L. (2007). Population structure and transmission dynamics of *Plasmodium vivax* in rural Amazonia. *J. Infect. Dis.* 195, 1218–1226.
  45. Eckstein-Ludwig, U., Webb, R.J., Van Goethem, I.D.A., East, J.M., Lee, A.G., Kimura, M., O'Neill, P.M., Bray, P.G., Ward, S.A., and Krishna, S. (2003). Artemisinin target the SERCA of *Plasmodium falciparum*. *Nature* 424, 957–961.
  46. Tanabe, K., Sakihama, N., Hattori, T., Ranford-Cartwright, L., Goldman, I., Escalante, A.A., and Lal, A.A. (2004). Genetic distance in house-keeping genes between *Plasmodium falciparum* and *Plasmodium reichenowi* and within *P. falciparum*. *J. Mol. Evol.* 59, 687–694.
  47. Rozas, J., Sánchez-DelBarrio, J.C., Messeguer, X., and Rozas, R. (2003). DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 19, 2496–2497.
  48. Nei, M., and Gojobori, T. (1986). Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.* 3, 418–426.
  49. Kumar, S., Tamura, K., and Nei, M. (2004). MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief. Bioinform.* 5, 150–163.
  50. Weir, B., and Cockerham, C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution* 38, 1358–1370.
  51. Excoffier, L., Laval, G., and Schneider, S. (2005). Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Evol. Bioinform. Online* 1, 47–50.
  52. Linz, B., Balloux, F., Moodley, Y., Manica, A., Liu, H., Roumagnac, P., Falush, D., Stamer, C., Prugnolle, F., van der Merwe, S.W., et al. (2007). An African origin for the intimate association between humans and *Helicobacter pylori*. *Nature* 445, 915–918.
  53. Beaumont, M.A., Zhang, W., and Balding, D.J. (2002). Approximate Bayesian computation in population genetics. *Genetics* 162, 2025–2035.