19. Niogi, K. K. Photoprotection revisited: Genetic and molecular approaches. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **50,** 333–359 (1999).
20. Greene, R. M., Geider, R. J., Kolber, Z. & Falkowski, P. G. Iron-induced changes in light-harvesting and photochemical energy conversion processes in eukaryotic marine algae. *Plant Physiol.* **100,** 565–575 (1992).
21. Kieber, D. J., Jiao, J., Kiene, R. P. & Bates, T. S. The impact of dimethyl sulfide photochemistry on methyl sulfur cycling in the Equatorial Pacific Ocean. *J. Geophys. Res.* **101,** 3715–3722 (1996).
22. Simó, R. & Pedrós-Allió, C. Role of vertical mixing in controlling the oceanic production of dimethyl sulphide. *Nature* **402,** 396–399 (1999).
23. Dacey, J. W., Howse, F. A., Michaels, A. F. & Wakeham, S. G. Temporal variability of dimethylsulfide and dimethylsulfoniopropionate in the Sargasso Sea. *Deep-Sea Res. I* **45,** 2085–2104 (1998).
24. Naninga, H. J. & Tyrrell, T. Importance of light for the formation of algal blooms by *Emiliania huxleyi*. *Mar. Ecol. Prog. Ser.* **136,** 195–206 (1996).
25. Simó, R., Grimalt, J. O., Pedrós-Alio, C. & Albaiges, J. Occurrence and transformation of dissolved dimethyl sulfur species in stratified seawater (western Mediterranean Sea). *Mar. Ecol. Prog. Ser.* **127,** 291–299 (1995).
26. Belviso, S. *et al.* Size distribution of dimethylsulfoniopropionate (DMSP) in areas of the tropical northeastern Atlantic Ocean and the Mediterranean Sea. *Mar. Chem.* **44,** 55–71 (1993).
27. Sunda, W. G. & Huntsman, S. A. Iron uptake and growth limitation in oceanic and coastal phytoplankton. *Mar. Chem.* **50,** 189–206 (1995).
28. Kiene, R. P. & Gerard, G. Determination of trace levels of dimethyl sulfoxide (DMSO) in seawater and rainwater. *Mar. Chem.* **47,** 1–12 (1994).
29. Ahner, B. A., Wei, L., Oleson, J. R. & Ogura, N. Glutathione and other low molecular weight thiols in marine phytoplankton under metal stress. *Mar. Ecol. Prog. Ser.* **232,** 93–103 (2002).
30. Dabrowski, K. *Ascorbate in Aquatic Organisms* (CRC, Boca Raton, 2001).

Supplementary Information accompanies the paper on *Nature*'s website (http://www.nature.com/nature).

Acknowledgements

We thank D. Blackwood, P. Griffin and L. Linn for technical assistance.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to W.S. (e-mail: bill.sunda@noaa.gov).

Genetic diversity and chloroquine selective sweeps in *Plasmodium falciparum*

John C. Wootton*, Xiaorong Feng†, Michael T. Ferdig‡, Roland A. Cooper†, Jianbing Mu†, Dror I. Baruch†, Alan J. Magill†§ & Xin-zhuan Su†

** Computational Biology Branch, National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland 20894-6075, USA*
† Laboratory of Malaria and Vector Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20892-0425, USA
‡ Department of Biological Sciences, University of Notre Dame, Notre Dame, Indiana 46556-016, USA
§ Division of Communicable Disease and Immunology, Walter Reed Army Institute of Research, Washington DC 20307-5100, USA

Widespread use of antimalarial agents can profoundly influence the evolution of the human malaria parasite *Plasmodium falciparum*. Recent selective sweeps for drug-resistant genotypes may have restricted the genetic diversity of this parasite, resembling effects attributed in current debates[1–4] to a historic population bottleneck. Chloroquine-resistant (CQR) parasites were initially reported about 45 years ago from two foci in southeast Asia and South America[5], but the number of CQR founder mutations and the impact of chlorquine on parasite genomes worldwide have been difficult to evaluate. Using 342 highly polymorphic microsatellite markers from a genetic map[6], here we show that the level of genetic diversity varies substantially among different regions of the parasite genome, revealing extensive linkage disequilibrium surrounding the key CQR gene *pfcrt*[7] and at least four CQR founder events. This disequilibrium and its decay rate in the *pfcrt*-flanking region are consistent with strong directional selective sweeps occurring over only ~20–80 sexual generations, especially a single resistant *pfcrt* haplotype spreading to very high frequencies throughout most of Asia and Africa. The presence of linkage disequilibrium provides a basis for mapping genes under drug selection in *P. falciparum*.

We examined microsatellite markers covering the 14 haploid chromosomes, at an average interval of ~75 kilobases (kb), from 87 *P. falciparum* worldwide isolates (Supplementary Information Fig. 1). Eighty-five per cent of the markers are highly polymorphic with 10 or more alleles. To evaluate whether genotypes reflect the geographical origins of the isolates, we tested the hypothesis of no allele sharing within a continental subset of isolates (Fig. 1a). The American isolates show very significant allele sharing for all chromosomes ($P = 1.5 \times 10^{-37}$ genome-wide), as do Asian isolates for most chromosomes except 1, 2, 8 and 14 ($P = 2.15 \times 10^{-7}$). In contrast, African isolates are very diverse and have a level of allele sharing similar to the bootstrap controls. This analysis adds genome-wide support to previous claims[8–10] that distinct parasite population structures exist on different continents, and that African parasites could be the common ancestors of *P. falciparum* parasites worldwide because of their higher diversity.

However, substantial differences in allelic diversity exist among chromosomes (Fig. 1a), possibly reflecting recent directional selection superimposed on the patterns of geographical diversity. The CQR isolates from Africa and Asia shared many more alleles on chromosome 7 (where *pfcrt* is located) than the chloroquine sensitive (CQS) isolates (Fig. 1b), suggesting common origins for CQR parasites. To further explore the extent of allele sharing on chromosome 7, we analysed the *pfcrt* alleles and flanking

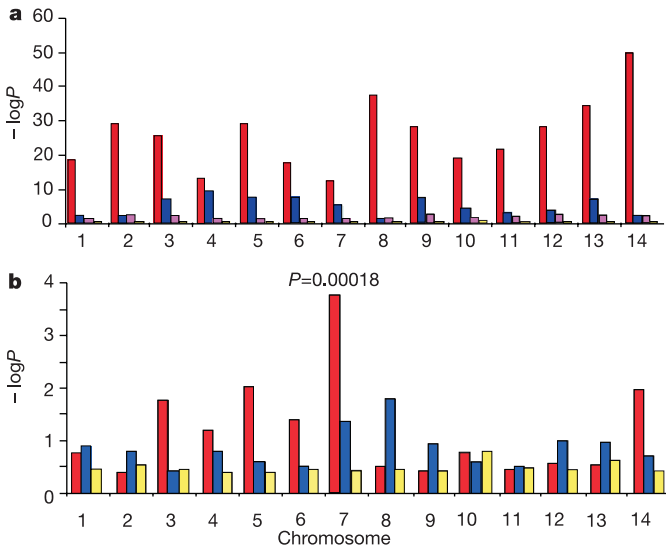

Figure 1 Genome-wide allele sharing analysis of *P. falciparum* isolates from various geographical regions or CQR/CQS (chloroquine-resistant/chloroquine sensitive) subsets. For the individual chromosomes, bars represent the negative $\log_{10}(P)$ values for obtaining by chance the amount of allele sharing observed, based on 20,000 bootstrap samples (see Methods; higher bars indicate more significant allele sharing). **a**, Allele sharing among parasites from Asia (blue), Africa (purple) and South America (red). Yellow, bootstrap controls. **b**, Higher allele sharing among CQR (red) than CQS (blue) parasites from Africa and Asia for chromosome 7 ($P = 1.76 \times 10^{-4}$) and to a lesser extent for chromosomes 3, 5 and 14. Yellow, bootstrap controls.

microsatellite haplotypes.

We found that nearly all CQR isolates from both Papua New Guinea and the Brazilian/Peruvian Amazon have the same *pfcrt* allele, coding for five amino-acid substitutions (S-MNT-H-S-Q-D-L-R, Fig. 2a). This allele probably evolved independently in South America and Papua New Guinea because the *pfcrt*-flanking microsatellite haplotypes are notably different (Fig. 2a); and the genome-wide allele sharing between Papua New Guinea and the Amazon is not significantly different from the background sharing between other Asian and American isolates ($P = 0.34$). In contrast, a distinct allele (C-MET-Q-S-Q-N-I-T) in Columbian parasite JAV evolved with an ECU-like flanking haplotype and genetic background (see Fig. 2a). African and Asian (excluding Papua New Guinea) CQR parasites have a near-identical *P. falciparum* chloroquine-resistance transporter (PfCRT) with 6–8 amino-acid substitutions and a common *pfcrt*-flanking haplotype (Fig. 2a). Asian isolates (except FCB and its close relatives) have a C-IET-H-S-E-S-T-I variant, whereas most African isolates and FCB have C-IET-H-S-E-S-I-I, suggesting that a back mutation in an Asian CQR parasite pre-dated the spread of CQR parasites in Africa. These *pfcrt* alleles, the flanking microsatellite markers, and the lack of admixture shown by genome-wide haplotypes demonstrate—in contrast to the two origins previously proposed[5]—at least four independent CQR founder events: one in Asia spreading to Africa, one in Papua New Guinea, and two in South America.

A hallmark of a selective sweep is a chromosomal region with reduced allelic diversity associated with a specific phenotype. One example is the *pfcrt* allele (C-IET-H-S-E-S-T-I) that has overwhelmed most of Asia and Africa with current allele frequencies of ~40–100% (refs 11, 12). Chromosomal segments spanning >200 kb surrounding *pfcrt* have been maintained in most of the CQR parasites, but decline gradually with distance, whereas those in the CQS parasites are highly diverse in these flanking regions (Fig. 2a–c). On the basis of the high recombination rates inferred

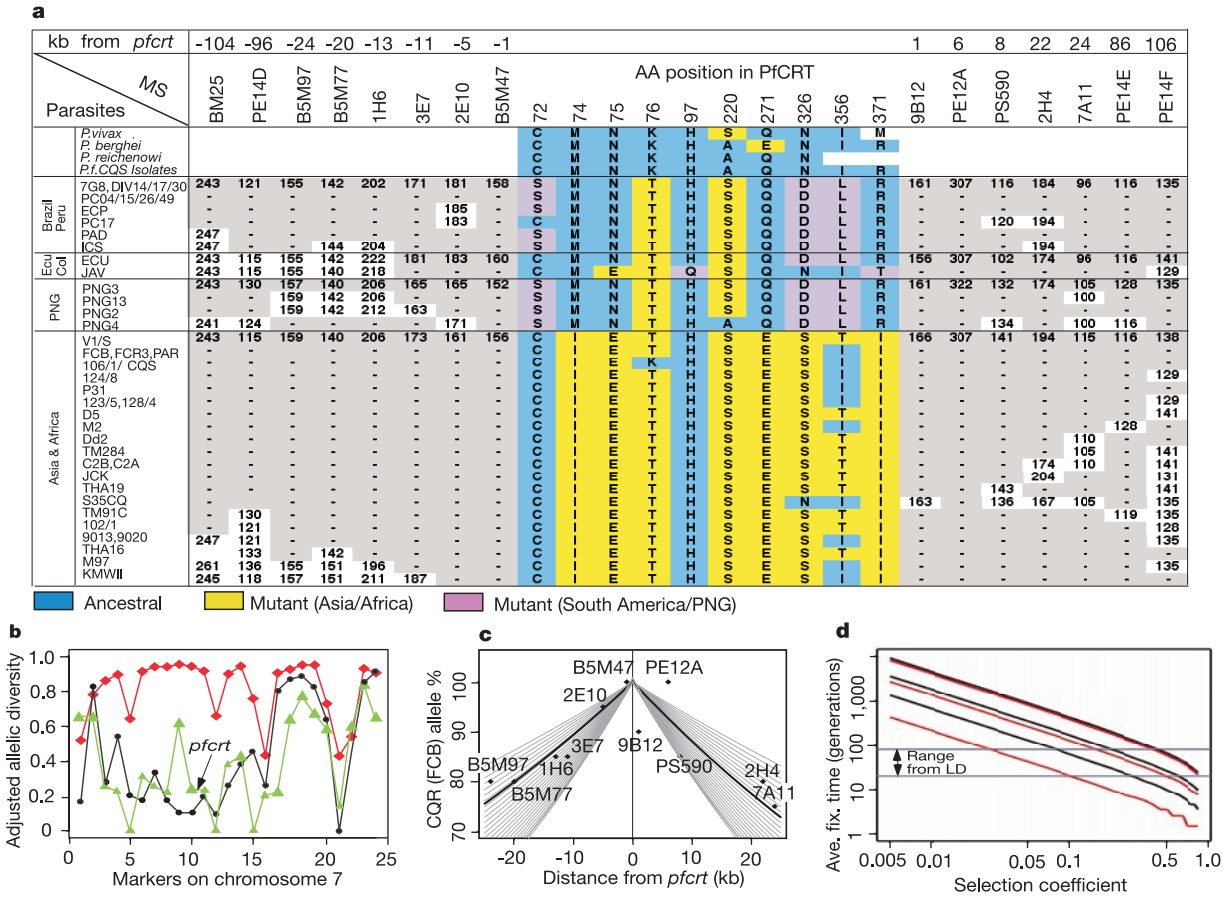

Figure 2 Extensive linkage disequilibrium (LD), microsatellite (MS) haplotypes flanking *pfcrt*, and the rate of LD decline in CQR isolates. **a**, Amino-acid substitutions in PfCRT and its flanking MS haplotypes. MS and their distances from *pfcrt* are indicated at the top. 'Pf CQS isolates' represents 38 CQS isolates having the ancestral C-MNK-H-A-Q-N-I-R allele (blue shading), but with highly variable flanking MS haplotypes (Supplementary Information Table 1). CQR parasites are grouped according to their genotypes and geographic origins (Ecu, Ecuador; Col, Columbia); and dashes represent the same MS allele as the one at the top of each section. The substitutions in some *P. falciparum* isolates and the orthologues from *Plasmodium vivax*, *Plasmodium berghei*, and *Plasmodium knowlesi* are from refs 7 and 20. **b**, Decreased diversity in CQR isolates surrounding *pfcrt* (arrow) suggests chloroquine selection. Black, CQR parasites from Asia and Africa; green, CQR parasites from South America; red, CQS parasites from Asia and Africa. **c**, Decline of LD from *pfcrt* (at position 0) in Asian/African CQR isolates. The points represent the percentage of shared alleles of each marker. The thin lines are the expected LD decay curves obtained by simulation from equation (2), using known recombination rate[6], an inbreeding coefficient of 0.5 and a range, in steps of 2, of 25 (top line) to 75 sexual generations. The thick lines are the best-fit regression curves corresponding to ~40 generations for inbreeding coefficient of 0.5. **d**, Estimation of the selection coefficients and numbers of generations required to fix a drug resistant allele under different initial allele proportions (p_0): 10^{-18} (upper lines); 10^{-6} (middle lines); 10^{-1} (lower lines). Final allele proportions (p_g): 0.5 (red lines); 0.99 (black lines). The results show no possibility of generating the observed high CQR *pfcrt* allele frequencies by genetic drift in the 20–80 generation period (horizontal lines) inferred from LD.

from a genetic cross[6] and field populations[8,13], and inbreeding coefficients previously estimated for Africa and Asia as 0.3–0.9 (refs 14–16), we estimate that this linkage disequilibrium decay has occurred over 20–80 parasite sexual generations. The timescale corresponding to 20–80 generations is ~6–30 years in endemic areas, emphasizing the profound influence of chloroquine on recent *P. falciparum* evolution.

Fixation of this CQR haplotype by genetic drift alone would require hundreds or thousands of generations (Fig. 2d). Although genetic drift, especially during epidemic expansions, could fix neutral alleles in isolated populations, neutral mechanisms cannot fix the same allele as observed for CQR on the continental scale across many parasite populations with different endemicity and transmission levels. Simulations using a wide range of initial allele frequencies to model the effects of different rates of migration and mutation showed that very high selection coefficients of about 0.1–0.7 are necessary to produce the level of linkage disequilibrium observed (Fig. 2c, d, and Supplementary Information Fig. 2). These are substantially larger than selection coefficients (<0.0002) estimated from effective population sizes[3,8] that would typically fix effectively-neutral alleles by stochastic dynamics. From computer simulation (Supplementary Information Fig. 2), linkage disequilibrium segment lengths encompassing the *pfcrt* locus after 40 generations of directional selection are consistent with the 10–200 kb observed.

Only a small repertoire of amino-acid substitutions in PfCRT appears to be compatible with the constraints imposed by drug selection and the biological function of the protein. A minimum of three changes, at positions that are highly conserved in different *Plasmodium* species and CQS parasites, are present in addition to the critical 76T substitution (Fig. 2a). These substitutions may have evolved to compensate deleterious effects of 76T and/or as quantitative modulators of the chloroquine response. To investigate if compensatory mutations elsewhere in the genome have evolved under chloroquine selection together with *pfcrt* mutations, we mapped potential regions of linkage disequilibrium using genome-wide scans (Fig. 3). We found evidence of reduced allelic diversity, additional to *pfcrt*, especially on chromosomes 1, 5, 6, 7, 10 and 12 (Fig. 3a, b). However, none of these loci (except the *pfcrt* locus) can be attributed directly to the chloroquine selective sweep because most of the peaks are present in both CQS and CQR isolates. The flat baselines (Fig. 3c) in African isolates suggest that no other genes with effects similar to *pfcrt* spread throughout the continent, although genes with minor effects could escape our detection at present marker resolution.

Our genome-wide analysis shows highly diverse *P. falciparum* populations, with variable levels of polymorphisms in different chromosomal regions. Such a genomic structure is consistent with a history of multiple selective sweeps, and raises questions concerning the debate on a recent major bottleneck of *P. falciparum* evolution[1–4]. Successive selective sweeps or a bottleneck could present a picture of limited genetic diversity if relatively few loci are tested in local populations. Concerted genome-wide analyses of both single nucleotide polymorphism and microsatellite variation may give further insight into these and other questions related to diversifying, directional, stabilizing and balancing selection in *P. falciparum* evolution.

The reduced diversity and shared chromosomal segments at the *pfcrt* locus show that linkage-disequilibrium based methods could be used to map *P. falciparum* genes under drug selection. Association studies with other antimalarial drugs, singly or in combination, could reveal new drug-response genes and contribute further information on the diversity and evolutionary potential of the *P. falciparum* genome. □

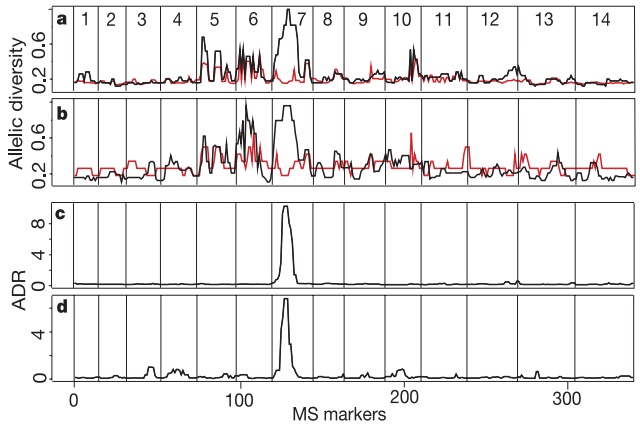

Figure 3 Genome-wide scans for loci of reduced diversity and association of reduced diversity with the CQR phenotype. The numbered panels represent the 14 chromosomes covered by 342 markers. **a**, **b**, Allelic diversity, plotted as (1 − median adjusted allelic diversity of five contiguous markers), for CQS (red) and CQR (black) isolates from Africa (**a**) and Asia (**b**, excluding CQR PNG isolates). Peaks represent regions with reduced diversity. **c**, **d**, allelic diversity ratio (ADR, plotted as the median log likelihood statistic over three contiguous markers) comparing CQR and CQS from Africa (**c**) and Asia (**d**), respectively. In both cases, a highly significant peak of log likelihood ratio at the *pfcrt* locus on chromosome 7 was correctly identified, demonstrating the power of this approach for detecting drug-resistant genes in malaria parasites. Peaks with ADR < 3 are not statistically significant.

Methods

Genotyping, DNA sequencing and drug assay

Parasite culture and DNA extraction were as described[17,18]. For microsatellite typing, polymerase chain reaction (PCR) products were labelled with fluorescent dyes and detected on an ABI377 automatic sequencer as described[6], using ROX 500XL size standards (PE Biosystem) and Genotyper software. PCR products amplified from genomic DNA were sequenced directly without cloning. Half maximal inhibitory concentrations (IC_{50}s) of the isolates in responses to chloroquine were determined as described[18] after culturing the parasites for 48 h followed by the addition of ^{3}H hypoxanthine for another 24 h.

Statistical and computational analysis

To analyse allele sharing in a geographical or phenotypic subset of n isolates (Fig. 1), an average haplotype distance was measured:

$$D_{av} = \left(\frac{1}{n^2}\right)\sum_{i=1}^{n}\sum_{j=1}^{n}\left(M_{T,ij}/M_{S,ij}\right) \quad (1)$$

M_T is the total number or markers compared between isolate i and j, and M_S is the number of shared alleles. Negative log probabilities of D_{av} were estimated using empirical distributions obtained from 20,000 bootstrap samples of randomly picked isolates.

Decline of linkage disequilibrium was calculated using[19]

$$L_i = 100[1 - d_i c(1 - F)]^g \quad (2)$$

where d_i is the distance (kb) between the gene under selection (*pfcrt*) and the ith marker, L_i is the percentage of isolates in which the entire progenitor haplotype was maintained over d_i, c is the crossover frequency per meiosis (one crossover per 1,700 kb per meiosis[6]), F is the inbreeding coefficient, and g is the number of sexual generations.

Changes in frequencies of two alleles (R and S) conferring drug resistance versus sensitivity were modelled for haploid inheritance using:

$$\ln\left(\frac{p_g}{1-p_g}\right) = \ln\left(\frac{p_0}{1-p_0}\right) + g\ln\left(\frac{1}{1-s}\right) \quad (3)$$

where p_0 and p_g are the proportions of the R allele, respectively, initially and at generation g, and s is the selection coefficient against the S allele (relative fitness of R and S: 1 and $(1 - s)$ respectively). To encompass the effects of a wide range of population structures and initial conditions, p_0 was varied from 0.1 (high introduction frequency of the R allele by human migration) to 10^{-24} (the estimated frequency of a double or triple mutation arising *de novo*). Values of p_g were computed over 10,000 generations for s values ranging from 0.0005 (near-neutrality) to 1, to determine the average times required to give R allele p_g values of 0.5 to 0.99.

The 'adjusted allelic diversity' (Fig. 3a, b) is heterozygosity normalized for the marker-specific stepwise mutational component. The estimated number (A) of 2-bp or 3-bp steps available to a marker was computed from the range and intervals observed in its allele size

distribution. Then:

$$Y = \frac{A}{A+1}\left[1 - \sum_i p_i^2\right] \quad (4)$$

where p_i is the proportion of the ith allele.

The allelic diversity ratio (ADR, Fig. 3c, d) was used to detect significant diversity differences between two phenotypically distinct sets of isolates (CQS and CQR):

$$\mathrm{ADR} = \frac{\left(\frac{n_1}{n_1+1}\right)\sum_{i=1}^{k_1} p_i^2}{\left(\frac{n_2}{n_2+1}\right)\sum_{j=1}^{k_2} p_j^2} \quad (5)$$

where the isolates of the two phenotypes have n_1 and n_2 individuals and k_1 and k_2 alleles at proportions p_i and p_j. Log likelihoods for ADR were obtained empirically using 1,000 permutations of the indices of all isolates over their fixed genome-wide microsatellite haplotypes together with an empirical correction for the degree of relatedness among isolates. ADR, which cancels out biases owing to different intrinsic microsatellite mutation modes and homoplasy rates, was determined for all individual markers, then plotted (Fig. 3c, d) as the median filtered log likelihood over three-marker windows.

Received 26 September 2001; accepted 28 March 2002; doi:10.1038/nature00813.

1. Rich, S. M. & Ayala, F. J. The recent origin of allelic variation in antigenic determinants of *Plasmodium falciparum*. *Genetics* **150,** 515–517 (1998).
2. Volkman, S. K. *et al.* Recent origin of *Plasmodium falciparum* from a single progenitor. *Science* **293,** 482–484 (2001).
3. Hughes, A. L. & Verra, F. Very large long-term effective population size in the virulent human malaria parasite *Plasmodium falciparum*. *Proc. R. Soc. Lond. B* **268,** 1855–1860 (2001).
4. Hey, J. Parasite populations: the puzzle of *Plasmodium*. *Curr. Biol.* **9,** R565–R567 (1999).
5. Payne, D. Spread of chloroquine resistance in *Plasmodium falciparum*. *Parasitol. Today* **3,** 241–246 (1987).
6. Su, X. *et al.* A genetic map and recombination parameters of the human malaria parasite *Plasmodium falciparum*. *Science* **286,** 1351–1353 (1999).
7. Fidock, D. A. *et al.* Mutations in the *P. falciparum* digestive vacuole transmembrane protein PfCRT and evidence for their role in chloroquine resistance. *Mol. Cell* **6,** 861–871 (2000).
8. Anderson, T. J. *et al.* Microsatellite markers reveal a spectrum of population structures in the malaria parasite *Plasmodium falciparum*. *Mol. Biol. Evol.* **17,** 1467–1482 (2000).
9. Escalante, A. A., Barrio, E. & Ayala, F. J. Evolutionary origin of human and primate malarias: evidence from the circumsporozoite protein gene. *Mol. Biol. Evol.* **12,** 616–626 (1995).
10. Conway, D. J. *et al.* Origin of *Plasmodium falciparum* malaria is traced by mitochondrial DNA. *Mol. Biochem. Parasitol.* **111,** 163–171 (2000).
11. Djimde, A. *et al.* A molecular marker for chloroquine-resistant *falciparum* malaria. *N. Engl. J. Med.* **344,** 257–263 (2001).
12. Dorsey, G., Kamya, M. R., Singh, A. & Rosenthal, P. J. Polymorphisms in the *Plasmodium falciparum* pfcrt and pfmdr-1 genes and clinical response to chloroquine in Kampala, Uganda. *J. Infect. Dis.* **183,** 1417–1420 (2001).
13. Conway, D. J. *et al.* High recombination rate in natural populations of *Plasmodium falciparum*. *Proc. Natl Acad. Sci. USA* **96,** 4506–4511 (1999).
14. Hill, W. G., Babiker, H. A., Ranford-Cartwright, L. C. & Walliker, D. Estimation of inbreeding coefficients from genotypic data on multiple alleles, and application to estimation of clonality in malaria parasites. *Genet. Res.* **65,** 53–61 (1995).
15. Walliker, D., Babiker, H. & Ranford Cartwright, L. in *Malaria: Parasite Biology, Pathogenesis, and Protection* (ed. Sherman, I. W.) 235–252 (American Society for Microbiology, Washington DC, 1998).
16. Paul, R. E. *et al.* Mating patterns in malaria parasite populations of Papua New Guinea. *Science* **269,** 1709–1711 (1995).
17. Trager, W. & Jensen, J. B. Human malaria parasites in continuous culture. *Science* **193,** 673–675 (1976).
18. Su, X., Kirkman, L. A., Fujioka, H. & Wellems, T. E. Complex polymorphisms in an approximately 330 kDa protein are linked to chloroquine-resistant *P. falciparum* in Southeast Asia and Africa. *Cell* **91,** 593–603 (1997).
19. Dye, C. & Williams, B. G. Multigenic drug resistance among inbred malaria parasites. *Proc. R. Soc. Lond. B* **264,** 61–67 (1997).
20. Nomura, T. *et al.* Evidence for different mechanisms of chloroquine resistance in 2 *Plasmodium* species that cause human malaria. *J. Infect. Dis.* **183,** 1653–1561 (2001).

Supplementary Information accompanies the paper on *Nature*'s website (http://www.nature.com/nature).

Acknowledgements

We thank various investigators who provided the isolates over the years, S. Davis-Hayman, D. Joy, K. Hayton and B. Marshall for critical reading of the manuscript and editorial assistance, and T. Wellems, L. Miller and D. Lipman for support and encouragement. The opinions of the authors do not necessarily reflect those of the US army or the Department of Defense.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to X.-z.S. (e-mail: xsu@niaid.nih.gov).

Chromosome-wide SNPs reveal an ancient origin for *Plasmodium falciparum*

Jianbing Mu*, Junhui Duan*, Kateryna D. Makova†, Deirdre A. Joy*, Chuong Q. Huynh‡, Oralee H. Branch§, Wen-Hsiung Li† & Xin-zhuan Su*

** Laboratory of Malaria and Vector Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20892-0425, USA*

† Department of Ecology and Evolution, University of Chicago, Chicago, Illinois 60637, USA

‡ Information Engineering Branch, § Computational Biology Branch, National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland 20894-6075, USA

The Malaria's Eve hypothesis, proposing a severe recent population bottleneck (about 3,000–5,000 years ago) of the human malaria parasite *Plasmodium falciparum*, has prompted a debate about the origin and evolution of the parasite[1–6]. The hypothesis implies that the parasite population is relatively homogeneous, favouring malaria control measures. Other studies, however, suggested an ancient origin and large effective population size[5,7–10]. To test the hypothesis, we analysed single nucleotide polymorphisms (SNPs) from 204 genes on chromosome 3 of *P. falciparum*. We have identified 403 polymorphic sites, including 238 SNPs and 165 microsatellites, from five parasite clones, establishing chromosome-wide haplotypes and a dense map with one polymorphic marker per ~2.3 kilobases. On the basis of synonymous SNPs and non-coding SNPs, we estimate the time to the most recent common ancestor to be ~100,000–180,000 years, significantly older than the proposed bottleneck. Our estimated divergence time coincides approximately with the start of human population expansion[11], and is consistent with a genetically complex organism able to evade host immunity and other antimalarial efforts.

The launching of the international project to sequence the genome of *P. falciparum*, and the completion of chromosomes 2 and 3 (refs 12, 13), have enabled us to search directly for SNPs from the parasite genes. Here we analyse coding sequences (202 kilobases, kb) and non-coding sequences (16 kb) on chromosome 3 from five independent parasite clones originating from Southeast Asia (Dd2), Africa (3D7, sequences obtained from PlasmoDB; see Methods), South America (7G8), Central America (Hb3) and Papua New Guinea (D10). Their identities and geographical origins were verified by genotyping the parasite DNA with 348 microsatellite (MS) markers, which showed highly diverse genetic backgrounds (mean genetic diversity $H = 0.8776 \pm 0.0079$). Oligonucleotide primers were synthesized to amplify 1.2–1.5 kb DNA fragments from each predicted gene. Polymerase chain reaction (PCR) products were sequenced directly without cloning into bacteria, and SNPs were identified after alignment of the DNA sequences. We also independently amplified and sequenced DNA covering 56 SNPs, including 32 randomly chosen and all 24 SNPs with single strand coverage, to validate the accuracy of the SNPs. Fifty-four of the SNPs were verified (~96%). Both of the two false SNPs were base call errors from regions of single strand coverage.

Despite not being distributed evenly, SNPs occur across the chromosome, with most of the genes (80.5%) containing only one or two SNPs (Fig. 1a, b). Of the 238 SNPs identified from 118 of 204 predicted genes (57.8%), including 153 complete or partial introns (Table 1), 207 were in coding regions (cSNP) and 31 in non-coding regions (ncSNP), although the frequency of SNPs was greater in non-coding regions (Table 1). There were more than