Multiple populations of artemisinin-resistant *Plasmodium falciparum* in Cambodia

Olivo Miotto¹–³, Jacob Almagro-Garcia¹,³,⁴, Magnus Manske³, Bronwyn MacInnis³, Susana Campino³, Kirk A Rockett¹,³,⁴, Chanaki Amaratunga⁵, Pharah Lim³,⁶, Seila Suon⁶, Sokunthea Sreng⁶, Jennifer M Anderson⁵, Socheat Duong⁶, Chea Nguon⁶, Char Meng Chuo⁶, David Saunders⁷, Youry Se⁸,⁹, Chantap Lon⁸, Mark M Fukuda⁷,⁹, Lucas Amenga-Etego¹⁰, Abraham V O Hodgson¹⁰, Victor Asola¹⁰, Mallika Imwong²,¹¹, Shannon Takala-Harrison¹², François Nosten²,¹³,¹⁴, Xin-zhuan Su⁵, Pascal Ringwald¹⁵, Frédéric Ariey¹⁶, Christiane Dolecek¹⁴,¹⁷, Tran Tinh Hien¹⁴,¹⁷, Maciej F Boni¹⁴,¹⁷, Cao Quang Thai¹⁷, Alfred Amambua-Ngwa¹⁸, David J Conway¹⁸,¹⁹, Abdoulaye A Djiméto²⁰, Ogobara K Doumbo²⁰, Issaka Zongo²¹, Jean-Bosco Ouedraogo²¹, Daniel Alcock³, Eleanor Drury³, Sarah Auburn²², Oliver Koch¹, Mandy Sanders³, Christina Hubbart⁴, Gareth Maslen³, Valentin Ruano-Rubio³,⁴, Dushyanth Jothy³, Alistair Miles¹,⁴, John O’Brien⁴, Chris Gamble²³, Samuel O Oyola³, JULIAN C Rayner³, Chris I Newbold¹,³,²⁴, Matthew Berriman³, Chris C A Spencer¹,⁴, Samuel O Oyola³, JULIAN C Rayner³, Chris I Newbold¹,³,²⁴, Matthew Berriman³, Chris C A Spencer¹,⁴, Gideon McVean⁴, Nicholas P Day²,¹⁴, Nicholas J White²,¹⁴, Delia Bethell⁷, Arjen M Dondorp²,¹⁴, Christopher V Plowe¹², Rick M Fairhurst⁵ & Dominic P Kwiatkowski¹,³,⁴

We describe an analysis of genome variation in 825 *P. falciparum* samples from Asia and Africa that identifies an unusual pattern of parasite population structure at the epicenter of artemisinin resistance in western Cambodia. Within this relatively small geographic area, we have discovered several distinct but apparently sympatric parasite subpopulations with extremely high levels of genetic differentiation. Of particular interest are three subpopulations, all associated with clinical resistance to artemisinin, which have skewed allele frequency spectra and high levels of haplotype homozygosity, indicative of founder effects and recent population expansion. We provide a catalog of SNPs that show high levels of differentiation in the artemisinin-resistant subpopulations, including codon variants in transporter proteins and DNA mismatch repair proteins. These data provide a population-level genetic framework for investigating the biological origins of artemisinin resistance and for defining molecular markers to assist in its elimination.

The malaria parasite *P. falciparum* has shown a considerable capacity to develop resistance to antimalarial drugs by evolutionary adaptation. For reasons that remain poorly understood, successive global waves of antimalarial drug resistance have originated in western Cambodia¹. The most common form of chloroquine resistance was observed there in the late 1950s before it spread around the world², and the most common forms of clinically significant pyrimethamine resistance³ and sulfadoxine resistance⁴ are thought to have originated in the same region. Clinical resistance to artemisinin and its derivatives is now well established in the *P. falciparum* population of western Cambodia⁵–⁸ and appears to be emerging in neighboring regions⁹,¹⁰. These recent developments have grave implications for public health, as artemisinin derivatives are the mainstay of malaria treatment worldwide. There is an urgent need to discover the parasite genetic factors that cause artemisinin resistance¹¹ and to identify effective markers to monitor its spread¹².

1¹Medical Research Council (MRC) Centre for Genomics and Global Health, University of Oxford, Oxford, UK. ²Mahidol-Oxford Tropical Medicine Research Unit, Mahidol University, Bangkok, Thailand. ³Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK. ⁴Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK. ⁵National Institute of Allergy and Infectious Diseases, US National Institutes of Health, Bethesda, Maryland, USA. ⁶National Center for Parasitology, Entomology and Malaria Control, Phnom Penh, Cambodia. ⁷Department of Immunology and Medicine, US Army Medical Component, Armed Forces Research Institute of Medical Sciences (USAMC-AFRIMS), Bangkok, Thailand. ⁸USAMC-AFRIMS, Phnom Penh, Cambodia. ⁹Armed Forces Health Surveillance Center, Silver Spring, Maryland, USA. ¹⁰Navrongo Health Research Centre, Navrongo, Ghana. ¹¹Department of Molecular Tropical Medicine and Genetics, Faculty of Tropical Medicine, Mahidol University, Bangkok, Thailand. ¹²Howard Hughes Medical Institute, University of Maryland School of Medicine, Baltimore, Maryland, USA. ¹³Shoklo Malaria Research Unit, Mae Sot, Tak, Thailand. ¹⁴Centre for Tropical Medicine, University of Oxford, Oxford, UK. ¹⁵Global Malaria Programme, World Health Organization, Geneva, Switzerland. ¹⁶Unité d’Immunologie Moléculaire des Parasites, Institut Pasteur, Paris, France. ¹⁷Oxford University Clinical Research Unit, Wellcome Trust Major Overseas Programme, Ho Chi Minh City, Vietnam. ¹⁸MRC Laboratories, Fajara, The Gambia. ¹⁹London School of Hygiene and Tropical Medicine, London, UK. ²⁰Mahidol-Oxford Tropical Medicine Research Unit, Navrongo Health Research Centre, Navrongo, Ghana. ²¹Institut de Recherche en Sciences de la Santé, Direction Régionale de l’Ouest, Bobo-Dioulasso, Burkina Faso. ²²Menzies School of Health Research, Charles Darwin University, Darwin, Northern Territory, Australia. ²³Department of Statistics, University of Oxford, Oxford, UK. ²⁴Weatherall Institute of Molecular Medicine, University of Oxford, Oxford, UK. ²⁵Deceased. Correspondence should be addressed to D.P.K. (dominic@sanger.ac.uk).

Received 30 November 2012; accepted 4 April 2013; published online 28 April 2013; doi:10.1038/ng.2624
A central question is whether the genetic epidemiology of the *P. falciparum* population in Cambodia offers insights into the emergence of drug resistance and of artemisinin resistance in particular. To address this question, we analyzed patterns of genome variation in 825 *P. falciparum* samples collected at 10 locations in West Africa and Southeast Asia. These locations include Ghana, Mali, Burkina Faso, The Gambia, Thailand, Vietnam and four locations in Cambodia (Table 1). An estimated median of 2 Gb of sequence read data was obtained for each sample using the Illumina Genome Analyzer platform, and SNP genotype calls were made using a well-validated set of algorithms and quality control procedures whose development and evaluation are described in detail elsewhere.13. Here, we focus on a set of 86,158 coding SNPs that could be genotyped with confidence in other populations in this sample set. To put this in geographic context, low levels of malaria transmission14, the extent of population structure in Southeast Asia, consistent with relatively recent migration (Fig. 1c). Within Cambodia, there was further separation between a group of samples from northeastern Cambodia and Vietnam, close to samples from Thailand, whereas three distinct groups of western Cambodian samples were identified as outliers from this core group by the first, second and third principal components. An inspection of principal components beyond PC3 did not identify additional outlier groups. Other samples were intermediate between the core and outlier groups, suggesting admixture between the groups. To characterize this population structure in more detail, we used a process of ‘chromosome painting’, a probabilistic method of reconstructing the chromosomal pattern of each individual sample from homologous segments of DNA in samples representative of the core group and the three outlier groups (Fig. 2 and Online Methods). This process yielded a matrix of ancestral similarity in which the three outlier clusters could be clearly distinguished from other samples, and complex admixture patterns were apparent (Supplementary Fig. 3). Using the ancestral proportions determined by chromosome painting, each Cambodian sample was classified as belonging to one of the following subpopulations: KH1, representing the core group that clustered with samples from neighboring countries; KH2, KH3 and KH4, representing the three outlier groups; and KHA, representing samples with apparently mixed ancestry, defined here as having <80% ancestral content from any one of the main groups (Fig. 3a). Using these definitions, 63 samples were classified as KH1, 55 as KH2, 23 as KH3, 17 as KH4 and 135 as KHA.

Other methods of population structure analysis corroborated these findings. A neighbor-joining tree of Cambodian and Vietnamese samples showed clusters corresponding to KH1, KH2, KH3 and KH4, interspersed with KHA samples (Supplementary Fig. 4). When the ADMIXTURE program was used to model the Cambodian samples as admixture of four putative subpopulations, there was 82% concordance between the classification of samples by ADMIXTURE and by the chromosome painting method (Online Methods and Supplementary Fig. 5). Both analyses confirmed that most samples from northeastern Cambodia and Vietnam are very similar to KH1 samples in western Cambodia. It was evident from all methods that our classification of samples into KH1, KH2, KH3, KH4 and KHA populations should be regarded as a first approximation, as the population structure is clearly complex and will require further epidemiological sampling to understand in detail.

The frequencies of the subpopulations at different locations in Cambodia are shown in Table 2. All of the subpopulations were present in western Cambodia, whereas KH1 was predominant in northeastern Cambodia. Samples came from four independent field studies conducted over the past 5 years: there was clear evidence of population structure within each study, which was not limited to a particular site or sampling year (Supplementary Tables 1–3). The discovery of several distinct subpopulations of *P. falciparum* coexisting at the same location is unexpected and requires a biological explanation.

### Artemisinin resistance

An obvious question arising from these findings is whether different parasite subpopulations in Cambodia have different levels of resistance to artemisinin and its derivatives. To address this question, we analyzed parasite clearance rates in patients with acute falciparum malaria...
after artesunate treatment. In vivo estimation of parasite clearance rate from frequent peripheral blood parasite counts is becoming the standard method used for surveillance of artemisinin resistance, as no laboratory method has yet been developed to reliably correlate in vitro results with clinical resistance. In vivo clearance data were available for 212 of the samples in this study: 30 from Pailin, 47 from Tasanh, 89 from Pursat and 46 from Ratanakiri (Table 2). For each sample, we calculated the parasite clearance half-life on the basis of the slope of the linear portion of the parasite clearance curve. This analysis identified a significant prolongation of parasite clearance half-life in the KH2, KH3 and KH4 subpopulations compared to the KH1 subpopulation (Fig. 3b,c). The difference was evident from analysis of all 212 Cambodian samples and also when restricted to the 166 samples from western Cambodia. The median parasite clearance half-life in 35 KH2

![Population structure in the sample set analyzed. (a) Neighbor-joining tree of all 825 samples, based on a pairwise distance matrix. (b) Three-dimensional plot of PCA using all 825 samples, showing the first 3 components (PC1, PC2 and PC3). (c,d) PCA plots using only Southeast Asian samples, showing PC1 versus PC2 (c) and PC2 versus PC3 (d). The initial identification of three outlier clusters in Cambodia (labeled KH2, KH3 and KH4) was conducted by dividing the plot into quadrants, as shown. A core group (labeled KH1) of northeastern and western Cambodian samples was identified by its proximity to samples from other Southeast Asian countries, and the remaining samples with intermediate principal-component values were labeled KHA.](image)

![Chromosome painting. (a) A high-level view of chromosome painting across all 14 chromosomes is shown. Each line represents a sample; samples are organized by subpopulation in the order KH1, KHA, KH2, KH3 and KH4, starting from the bottom. The colors of segments indicate the cluster found to be the most probable donor of each segment. (b-d) Callouts show details for some features of the painting plots. (b) KH1 samples contain many short segments of DNA sequence labeled as KH2, KH3 and KH4, consistent with KH1 representing the ancestral population from which the KH2, KH3 and KH4 subpopulations are derived. (c) KH2, KH3 and KH4 samples comprise very long segments labeled as belonging to other populations, suggesting recent admixture. (d) KHA samples contain both short and long segments of DNA sequence.](image)
samples (all from western Cambodia) was 6.8 h compared to 4.1 h in 6 KH1 samples from western Cambodia (P = 2 × 10⁻⁴ by Mann-Whitney test) and 2.7 h in all 50 KH1 samples from Cambodia (P = 2 × 10⁻¹⁴). More detailed results for all subpopulations are given in Table 3. Each box represents the interquartile range of values, split at the median; whiskers extend to the furthest points that are within 1.5 times the length of the box. (c) An analogous plot derived from an analysis of western Cambodian parasites only (excluding the Ratanakiri site; n = 166). Parasite clearance half-lives with P values are given in Supplementary Table 1.

Genetic and phenotypic differentiation between Cambodian subpopulations. (a) Ancestry analysis of the 293 Cambodian samples, based on chromosome painting analysis of all Cambodian samples (Online Methods). Each vertical bar represents a single sample, and the bar is colored according to the proportion of the genome that was determined to have originated in each of the four KH1–KH4 clusters, as defined by PCA (magenta, KH1; blue, KH2; red, KH3; orange, KH4). Samples were ordered according to these proportions, and any sample with ≥80% of the genome originating from a single cluster was assigned the label of that cluster. Any sample that was not labeled as belonging to one of the resulting four subpopulations was assigned to the KHA group. (b) Box plot showing the distribution of artemisinin response phenotypes (parasite clearance half-life) in the Cambodian parasite subpopulations thus characterized (n = 212). To put this in an epidemiological context, recent clinical studies have shown that parasite clearance rates after artesunate treatment are significantly faster in northeastern Cambodia (P. L., D. Dek, V. Try, R. Eastman, S. Chy et al., unpublished data) and in Vietnam than in western Cambodia. It seems likely that these geographic differences in parasite clearance rate can largely be attributed to the predominance of KH2, KH3, KH4 and KHA subpopulations in western Cambodia, with KH1 predominant in northeastern Cambodia.

Founder effects
A notable feature of the KH2, KH3 and KH4 subpopulations is their strong genetic differentiation relative to KH1 and also relative to each other. This was evident from comparisons of genome-wide estimates of the fixation index (FST) between subpopulations: for example, the pairwise FST for KH2 versus KH4 (0.38) was considerably higher than that for Thailand versus Ghana samples (0.16) (Supplementary Table 4). It was also evident from counting the number of SNPs showing high levels of FST between subpopulations: for example, the number of SNPs with FST > 0.5 was 11 for Thailand samples versus the KH1 subpopulation, 562 for Thailand versus Ghana samples and 1,200 for KH2 versus KH4 (Supplementary Table 5). In summary, we found higher levels of genetic differentiation between subpopulations within western Cambodia than between typical parasite populations on different continents.

Several features of the KH2, KH3 and KH4 subpopulations indicate that their strong genetic differentiation is due to founder effects. Consistent with this explanation, we observed a marked reduction in the number of low-frequency alleles in KH2, KH3 and KH4 samples compared to KH1 samples and to parasites from other parts of Southeast Asia and West Africa (Fig. 4a). When SNPs with high levels of differentiation between subpopulations were compared to SNPs with low levels of differentiation, both groups had a similar ratio of nonsynonymous to synonymous polymorphisms (overall p/S was 2.4 for SNPs with FST < 0.2 and 2.1 for SNPs with FST > 0.8; Supplementary Table 6), suggesting that differentiation is likely to have arisen from founder effects.

As further evidence for founder effects, the KH2, KH3 and KH4 subpopulations were observed to have markedly reduced haplotype diversity. This was evident from higher levels of linkage disequilibrium (LD) in the KH2, KH3 and KH4 populations compared to KH1 and other populations (Fig. 4b). The loss of haplotype diversity was also quantified more directly: for example, levels of haplotype homozygosity, measured in a sliding window across the genome (Supplementary Note), were considerably higher for KH2, KH3 and KH4 populations than for the KH1 population (Fig. 4c and Supplementary Fig. 7). The level of haplotype diversity varied considerably across the genome, and, in some cases, it was found that a single haplotype predominated across a large chromosomal region: for example, KH4 had essentially a single haplotype across most of...
Table 3  Artemisinin response characteristics of Cambodian populations

|                  | All Cambodian samples | Western Cambodian samples |
|------------------|-----------------------|--------------------------|
|                  | N  | Half-life | P      | N  | Half-life | P      |
| KH1              | 50 | 2.7 h    | –      | 6  | 4.1 h    | –      |
| KH2              | 35 | 6.8 h    | 2 x 10^{-14} | 35 | 6.8 h    | 2 x 10^{-4} |
| KH3              | 18 | 6.4 h    | 7 x 10^{-9}  | 18 | 6.4 h    | 6 x 10^{-3}  |
| KH4              | 12 | 6.4 h    | 1 x 10^{-7}  | 12 | 6.4 h    | 2 x 10^{-3}  |
| KHA              | 97 | 6.6 h    | 7 x 10^{-16} | 95 | 6.7 h    | 4 x 10^{-3}  |

Indicated are the number of samples (N), the median parasite clearance half-life and the P value of a Mann-Whitney test comparing the half-life distribution for the subpopulation against that of KH1. Analysis results are shown using all Cambodian samples and also using samples from western Cambodia alone (Pailin, Tasanh and Pursat sites). Even compared against the smaller KH1 sample in western Cambodia, clearance half-lives are significantly longer in the KH2, KH3 and KH4 subpopulations and in the KHA recombinant population.

chromosome 9, and KH2 had essentially a single haplotype across approximately half of chromosome 13 (1.4–3.4 Mb), in marked contrast with KH1 (Supplementary Fig. 8).

An analysis of within-host diversity, quantified in terms of the F_{WS} metric\(^{13}\), showed that the vast majority of samples in the KH2, KH3 and KH4 populations were essentially clonal, as were those in the recombinant KHA population (Supplementary Fig. 9). Although within-sample diversity was found to be generally lower in Southeast Asian populations than in West African samples, the artemisinin-resistant samples were clearly more likely to be clonal than those in KH1 or in Thailand. This suggests a high level of inbreeding in the KH2, KH3 and KH4 populations.

We conclude that western Cambodia is home to at least three distinct founder populations of \textit{P. falciparum}, each of which is artemisinin resistant, coexisting with a more diverse and widely distributed subpopulation to which they are ancestrally related (Fig. 1a). Although KH2, KH3 and KH4 populations maintain extreme genetic differentiation, analyses of population structure and chromosome painting indicate that there is some gene flow between these populations (Fig. 2). This is most evident in KHA, which appears to be a mosaic of the other populations. This appearance might have arisen from recent admixture, past evolutionary events or some combination of the two. Future studies may be able to elucidate the evolutionary and demographic processes that underlie these observations.

Genetic markers and candidate loci

These data provide a set of genetic markers to map the geographic distribution of the artemisinin-resistant founder populations and to monitor their spread outside western Cambodia. The SNPs that differentiate the founder populations from other populations analyzed here are listed in Supplementary Tables 7–9.

The potential usefulness of this data set is illustrated by its application to a recent report that low levels of artemisinin resistance observed in northwestern Thailand are associated with a genomic region of low haplotype diversity on chromosome 13, located around 1.7–1.8 Mb\(^{7}\). We found that the KH2 population had essentially a
single haplotype extending across half of chromosome 13, from 1.4 Mb to 3.4 Mb (Supplementary Fig. 8). At chromosome 13 position 2,075,035, we observed a codon variant (p.Tyr175Ser) in an ABC transporter gene (PF13_0271) for which all samples from West Africa carried the ancestral allele, whereas almost all KH2 samples had the derived (non-ancestral) allele. Because this gene has been experimentally associated with chloroquine and quinine resistance\(^\text{16}\) and has previously been reported to lie within a genomic region of positive selection\(^\text{17}\), it would seem to be an important candidate for further investigation.

Transporter genes that have highly differentiated sequences in the founder populations are listed in Supplementary Table 10, as these are prime candidates for the causation of antimalarial drug resistance\(^\text{16,18}\). Apart from PF13_0271, they include PFE1150w (pfmdr1, an ABC transporter known to be associated with drug resistance), PFC0725c (a putative formate-nitrate transporter), PFE0805w (pfAT-Pase1, a cation-transporting ATPase) and PF13_0252 (pfnt1, a nucleoside transporter).

The founder populations also serve as a reservoir for a variety of known antimalarial resistance alleles (Supplementary Table 11). In pfcr\(^\text{T}\), almost all KH2, KH3 and KH4 samples had the resistance-associated haplotype CVIET, whereas KH1 had a wider variety of haplotypes\(^\text{20,21}\). Comparison of pfmdr1 (refs. 22–24) sequence coverage suggested that most samples in KH3 and KH4 had higher copy numbers than those in KH1 and KH2 (Supplementary Table 12).

In pfldhs,\(^\text{25}\) KH2 had a high frequency of the triple-mutant haplotype encoding p.Lys540Asn, p.Ala581Gly and p.Ala437Gly alterations\(^\text{26,27}\), whereas KH3 and KH4 had a high frequency of the triple-mutant haplotype encoding p.Ser436Ala, p.Lys540Glu and p.Ala437Gly alterations\(^\text{28}\). In pfldhf,\(^\text{25,29}\) the codon variant encoding p.Ile146Leu, which characterizes a highly resistant triple- or quadruple-mutant haplotype\(^\text{30}\), was absent in KH1 and KH4 but near fixation in KH2 and at an intermediate frequency in KH3.

It has been proposed that antimalarial resistance emerges rapidly in Southeast Asia because of the existence of hypermutable parasites, an idea that is based on experimental evidence that a \(P. falciparum\) strain originating from Indochina shows accelerated resistance to multiple structurally unrelated drugs\(^\text{31}\). It will therefore be of interest to researchers in this area that a number of the most differentiated nonsynonymous SNPs in the KH2 population were located in genes involved in DNA mismatch repair pathways (Supplementary Table 13). These included the genes encoding PMS1 and MLH, the two proteins that make up MutL, a DNA mismatch repair heterodimer that is highly conserved throughout evolution\(^\text{32}\), and UvrD, a helicase that is critical for methyl-directed mismatch repair and interacts physically with MutL, which loads it onto DNA\(^\text{33–35}\). This observation could be highly relevant, as MutL variants in bacteria have been shown to cause hypermutation and the rapid evolution of antibiotic resistance. However, uncertainty over the evolutionary timeline of the KH2 population means that further studies will be needed to establish whether these parasites are indeed hypermutable.

**DISCUSSION**

\(P. falciparum\) population structure is known to be related to the intensity of malaria transmission. Parasites are transmitted to humans by \(Anopheles\) mosquitoes, and a crucial step in this process is the sexual recombination of parasites within the mosquito vector. Previous studies have shown that higher rates of inbreeding and greater population structure occur in regions of low transmission intensity, such as Southeast Asia, than in regions with high transmission intensity, such as West Africa\(^\text{36,37}\).

The population structure observed in western Cambodia is unusual in several respects. Within this relatively small geographic area, we find multiple distinct but apparently sympatric parasite subpopulations with exceptionally high levels of genetic differentiation. Three of the subpopulations have skewed allele frequency spectra and high levels of haplotype homozygosity, indicative of founder effects and recent population expansion. These subpopulations are associated with clinical resistance to artesinin. Two key questions arise from these findings: what caused the founder effects and what is the biological relevance of their association with artesinin resistance?

Here, we offer a hypothesis to explain these unusual findings. Our starting assumption is that a parasite population under drug pressure will eventually acquire a number of genetic variants with the potential to cause drug resistance. However, the majority of such variants will have relatively small effects when acting individually. Moreover, many potential resistance-conferring alleles will be associated with a biological fitness cost in the absence of the drug. We postulate that the process of sexual recombination occasionally produces a parasite possessing a ‘winning combination’ of alleles that jointly confer high levels of drug resistance together with high levels of biological fitness. The inbred progeny of such a parasite, which carry the full set of alleles required for optimum biological fitness, are at a selective advantage compared to the rest of the parasite population. When such a parasite outcrosses with other parasite lines, the progeny are at risk of losing the winning combination of alleles and thereby losing biological fitness. In brief, our hypothesis is that the founder effects observed in western Cambodia represent the recent expansion of artesinin-resistant parasite lines, whose biological fitness requires a particular combination of alleles to be maintained at multiple loci across the genome.

If this hypothesis is broadly correct, it follows that multiple environmental and ecological factors will affect the emergence and spread of artesinin resistance. A multigenic resistance phenotype is more likely to emerge if the individual component alleles are already prevalent in the local parasite population, as determined by considering the history of previous antimalarial drug usage. Low transmission intensity and other factors that favor inbreeding could assist the propagation of a multigenic resistance phenotype once it has emerged. High rates of inbreeding might also arise because of physical isolation, for example, if a group of parasites becomes isolated in a remote area of the jungle, or because of some form of reproductive isolation, for example, it is conceivable that different parasite subpopulations are preferentially transmitted by different \(Anopheles\) vector species\(^\text{38}\).

These observations provide a population-level genetic framework for revisiting the longstanding debate about why western Cambodia is a global hotspot for antimalarial drug resistance. Drug pressure was particularly high in Pailin in the late 1950s and early 1960s as a result of mass administration of chloroquine and pyrhythmamine\(^\text{39,40}\). This may explain why subpopulations from this region possess a variety of haplotypes that are highly resistant to antimalarials other than artesinin—in other words, it is possible that we are witnessing the most recent episode in a longer and more complex chain of events. Human demographic factors relevant to western Cambodia should also be considered, such as the physical isolation of human settlements due to poor road infrastructure in forested mountain areas or to restricted human movement in the period of Khmer Rouge resistance (1979–1998), which might have provided a favorable environment for parasite inbreeding.

A major objective of the World Health Organization (WHO) Global Plan for Artemisinin Resistance Containment is to stop the spread of resistant parasites\(^\text{41}\). The discovery of multiple subpopulations of
resistant parasites, with widely different genetic characteristics, suggests that there could be multiple forms of resistance, each of which will need to be controlled. The definition of genetic markers for these subpopulations will assist efforts to eliminate major foci of transmission through the characterization of their geographic distribution and the ecological niches that they occupy. This will require the acquisition of genomic epidemiological data with a high level of spatial and temporal resolution, and the data presented here should not be regarded as a definitive account of the current situation, which is clearly in flux.

URLs
Information page for this study, http://www.malarialogen.net/resource/12; information page for the SNP list used, http://www.malarialogen.net/resource/10; European Nucleotide Archive (ENA) sequencing read repository, http://www.ebi.ac.uk/ena/data/search?query=plasmodium; WWARN Parasite Clearance Estimator, http://www.wwarn.org/toolkit/data-management/parasite-clearance-estimator; Burrows-Wheeler Aligner (BWA), http://bio-bwa.sourceforge.net/; PlasmoDB, http://www.plasmodb.org/; R language, http://www.r-project.org/; R analysis of phylogenetics and evolution (ape) package, http://ape.mpl.irf.fr/; ChromoPainter, http://www.mathsbris.ac.uk/~madjl/finestructure/chromopainter.html; ADMIXTURE, http://www.genetics.ucla.edu/software/admixture/; PLINK, http://pngu.mgh.harvard.edu/~purcell/plink/.

METHODS
Methods and any associated references are available in the online version of the paper.

Accession codes. All sequence data are available online at the European Nucleotide Archive (ENA), accessible at http://www.ebi.ac.uk/ena/data/search?query=plasmodium. ENA accession numbers for all samples used in this study are listed in Supplementary Table 14.

Note: Supplementary information is available in the online version of the paper.

ACKNOWLEDGMENTS
We would like to thank V. Cornelius and R. Giacomantonio for their support in producing and reviewing this manuscript; S. Uk and E.S. Phelps for their contributions in the Cambodian studies; and T. Anderson for helpful review comments. The sequencing, genotyping and analysis components of this study were supported by the Wellcome Trust through core funding of the Wellcome Trust Sanger Institute (090851), core funding of the Wellcome Trust Centre for Human Genetics (090532/2/09/Z) and a Strategic Award (090770/2/09/Z) and the MRC through the MRC Centre for Genomics and Global Health (G0600718) and an MRC Professorship to D.P.K. (G19/9). Other parts of this study were partly supported by the Wellcome Trust; the MRC; the Division of Intramural Research, National Institute of Allergy and Infectious Diseases, US National Institutes of Health; and a Howard Hughes Medical Institute International Scholarship (55005502) to A.A.D. P.R. is a staff member of the World Health Organization; he alone is responsible for the views expressed in this publication, and they do not necessarily represent the decisions, policy or views of the World Health Organization.

AUTHOR CONTRIBUTIONS
S.C., C.A., P.L., S. Suon, S. Sreng, J.M.A., S.D., C.N., C.M.C., D.S., Y.S., C.L., M.M.F., L.A.-E., A.V.O.H., V.A., M.I., F.N., X.S., P.R., F.A., C.D., T.T.H., M.F.B., C.Q.T., A.A.-N., D.J.G.C., A.A.D., O.K.D., I.Z., J.-B.O., S.A., N.P.D., N.J.W., D.B., A.M.D., C.V.P. and R.M.F. carried out field and laboratory studies and obtained Plasmodium falciparum samples for sequencing. C.A., P.L., S. Suon, S. Sreng, J.M.A., S.D., C.N., C.M.C., D.S., Y.S., C.L., M.M.F., F.N., X.S., P.R., F.A., N.J.W., D.B., A.M.D., C.V.P. and R.M.F. carried out clinical studies to obtain ART phenotype data. S.C., D.A., E.D., M.S., S.A., O.K., S.O.O., B.M., C.L.N. and M.B. developed and implemented methods for sample processing and sequencing library preparation. O.M., J.A.-G., M.M., G. Maslen, V.R.-R., D.J. and A.M. developed software for data management and visualization. K.A.R., C.H., D.A. and M.M. carried out validation experiments. C.V.P., S.T.-H., G. McVean and R.M.F. contributed to development of the project. B.M., M.B., C.L.N. and J.C.R. provided project management and oversight. O.M., J.A.-G., M.M., J.O., C.G. and C.C.A.S. carried out data analyses. D.P.K., O.M. and J.A.-G. wrote the manuscript. All authors reviewed the manuscript.

COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. White, N.J. Artemisinin resistance—the clock is ticking. Lancet 376, 2051–2052 (2010).
2. Payne, D. Spread of chloroquine resistance in Plasmodium falciparum. Parasitol. Today 3, 241–246 (1987).
3. Roper, C. et al. Intercontinental spread of pyrimethamine-resistant malaria. Science 287, 1124–1128 (2000).
4. Mita, T. et al. Limited geographical origin and global spread of sulfadoxine-resistant dhps alleles in Plasmodium falciparum populations. J. Infect. Dis. 204, 1980–1988 (2011).
5. Noedl, H. et al. Evidence of artemisinin-resistant malaria in western Cambodia. N. Engl. J. Med. 359, 2619–2620 (2008).
6. Dondorp, A.M. et al. Artemisinin resistance in Plasmodium falciparum malaria. N. Engl. J. Med. 361, 455–467 (2009).
7. Cheeseman, I.H. et al. A major genome region underlying artemisinin resistance in malaria. Science 336, 79–82 (2012).
8. Amaratunga, C. et al. Artemisinin-resistant Plasmodium falciparum in Pursat province, western Cambodia: a parasite clearance rate study. Lancet Infect. Dis. 12, 851–858 (2012).
9. Phyo, A.P. et al. Emergence of artemisinin-resistant malaria on the western border of Thailand: a longitudinal study. Lancet 379, 1960–1966 (2012).
10. Tran, T.H. et al. In vivo susceptibility of Plasmodium falciparum to artesunate in Binh Phuoc Province, Vietnam. Malar. J. 11, 355 (2012).
11. Anderson, T.J. et al. High heritability of malaria parasite clearance rate indicates a genetic basis for artemisinin resistance in western Cambodia. J. Infect. Dis. 201, 1326–1330 (2010).
12. Dondorp, A.M. et al. The threat of artemisinin-resistant malaria. N. Engl. J. Med. 365, 1073–1075 (2011).
13. Manske, M. et al. Analysis of Plasmodium falciparum diversity in natural infections by deep sequencing. Nature 487, 375–379 (2012).
14. Hay, S.I. et al. A world malaria map. Plasmodium falciparum endemicity in 2007. PLoS Med. 6, e1000408 (2009).
15. Flegg, J.A., Guerin, P.J., White, N.J. & Stepniewska, K. Standardizing the measurement of parasite clearance in falciparum malaria: the parasite clearance estimator. Malar. J. 10, 339 (2011).
16. Mu, J. et al. Multiple transporters associated with malaria parasite responses to chloroquine and quinine. Mol. Microbiol. 49, 977–989 (2003).
17. Mu, J. et al. Plasmodium falciparum genome-wide scans for positive selection, recombination hot spots and resistance to antimalarial drugs. Nat. Genet. 42, 268–271 (2010).
18. Anderson, T.J. et al. Are transporter genes other than the chloroquine resistance locus (pfcrt) and multidrug resistance gene (pfmdr) associated with antimalarial drug resistance? Antimicrob. Agents Chemother. 49, 2180–2188 (2005).
19. Fidock, D.A. et al. Mutations in the P. falciparum digestive vacuole transmembrane protein PIRC and evidence for their role in chloroquine resistance. Mol. Cell 6, 861–871 (2000).
20. Lim, P. et al. PfCRT polymorphism and chloroquine resistance in Plasmodium falciparum strains isolated in Cambodia. Antimicrob. Agents Chemother. 47, 97–94 (2003).
21. Isozumi, R. et al. Longitudinal survey of Plasmodium falciparum infection in Vietnam: characteristics of antimalarial resistance and their associated factors. J. Clin. Microbiol. 48, 70–77 (2010).
22. Flegg, J.A., Baric, R.S. & Klenerman, D. A strong association between mefloquine and halofantrine resistance and amplification, overexpression, and mutation in the P-glycoprotein gene homolog (pfmdr1) of Plasmodium falciparum in vitro. Am. J. Trop. Med. Hyg. 51, 648–658 (1994).
23. Pechappanikui, P. et al. Plasmodium falciparum pfmdr1 amplification, mefloquine resistance, and parasite fitness. Antimicrob. Agents Chemother. 53, 1509–1515 (2009).
24. Powe, C.V. et al. Mutations in Plasmodium falciparum dihydrofolate reductase and dihydropyrimidine synthase and epidemiological patterns of pyrimethamine-sulfadoxine use and resistance. J. Infect. Dis. 176, 1590–1596 (1997).
25. Vinayak, S. et al. Origin and evolution of sulfadoxine resistant Plasmodium falciparum. PLoS Pathog. 6, e1000830 (2010).
26. Kulin, J.G. et al. Molecular assays for surveillance of antifolate-resistant malaria. Lancet 351, 1629–1630 (1998).
27. Te Linh, T., Wang, P., Sims, P.F., Hyde, J.E. & Cowman, A.F. Allelic exchange at the ars gene in Plasmodium falciparum. Nature 398, 3807–3815 (1998).
28. Peterson, D.S., Walliker, D. & Wellems, T.E. Evidence that a point mutation in dihydropteroate synthase confers resistance to pyrimethamine in falciparum malaria. Proc. Natl. Acad. Sci. USA 85, 9114–9118 (1988).
30. Plowe, C.V. The evolution of drug-resistant malaria. *Trans. R. Soc. Trop. Med. Hyg.* 103 (suppl. 1), S11–S14 (2009).

31. Rathod, P.K., McErlean, T. & Lee, P.C. Variations in frequencies of drug resistance in *Plasmodium falciparum*. *Proc. Natl. Acad. Sci. USA* 94, 9389–9393 (1997).

32. Jiricny, J. Replication errors: challenging the genome. *EMBO J.* 17, 6427–6436 (1998).

33. Shankar, J. & Tuteja, R. UvrD helicase of *Plasmodium falciparum*. *Proc. Natl. Acad. Sci. USA* 94, 9389–9393 (1997).

34. Jiricny, J. Replication errors: challenging the genome. *EMBO J.* 17, 6427–6436 (1998).

35. Shankar, J. & Tuteja, R. UvrD helicase of *Plasmodium falciparum*. *Proc. Natl. Acad. Sci. USA* 94, 9389–9393 (1997).

36. Shankar, J. & Tuteja, R. UvrD helicase of *Plasmodium falciparum*. *Gene* 410, 223–233 (2008).

37. Hall, M.C., Jordan, J.R. & Matson, S.W. Evidence for a physical interaction between the *Escherichia coli* methyl-directed mismatch repair proteins MutL and UvrD. *EMBO J.* 17, 1535–1541 (1998).

38. Mechanic, L.E., Frankel, B.A. & Matson, S.W. *Escherichia coli* MutL loads DNA helicase II onto DNA. *J. Biol. Chem.* 275, 38337–38346 (2000).

39. 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).

40. Dye, C. & Williams, B.G. Multigenic drug resistance among inbred malaria parasites. *Proc. Biol. Sci.* 264, 61–67 (1997).

41. Sinko, M.E. et al. The dominant *Anopheles* vectors of human malaria in the Asia-Pacific region: occurrence data, distribution maps and bionomic precis. *Parasit. Vectors* 4, 89 (2011).

42. Verdrager, J. Epidemiology of the emergence and spread of drug-resistant falciparum malaria in South-East Asia and Australasia. *J. Trop. Med. Hyg.* 89, 277–289 (1986).

43. Payne, D. Did medicated salt hasten the spread of chloroquine resistance in *Plasmodium falciparum? Parasitol. Today* 4, 112–115 (1988).

44. World Health Organization. Recommendations. in *Global Plan for Artemisinin Resistance Containment* 23–27 (World Health Organization, Geneva, 2011).
We determined gene reading frames and exon boundaries by using PlasmoDB 5.5 annotation (see URLs) of the 3D7 genome. Each SNP in the V1.0 SNP Catalogue was classified as synonymous or nonsynonymous according to whether an amino acid change occurred when substituting the reference allele with the non-reference allele at that SNP in the 3D7 reference genome sequence without any other changes.

Chromosome painting. Chromosome painting is a probabilistic method of reconstructing the ancestry of individuals in a recombinant population. Whereas commonly used methods for discovering population structure (such as ADMIXTURE) tend to consider SNPs independently, chromosome painting uses linkage information in the data set in its probability computation, on the basis of the assumption that groups of linked alleles tend to be exchanged together during recombination. Hence, given a set of candidate recombinant ‘donor’ populations, chromosome painting assigns the most probably ancestral population to a segment of DNA from a given individual on the basis of its similarity to corresponding DNA segments in individuals from the donor populations (rather than performing this analysis on each SNP independently). This strategy results in visualizations of the recombination patterns, assigning a color to each donor population and using these colors to paint each DNA segment according to the most probable assignment.

We performed chromosome painting using the ChromoPainter software package (see URLs). The parameters \( N_e \) (effective population size) and \( \mu \) (emission probability) were estimated in Cambodian samples by running the expectation-maximization algorithm with 40 iterations for each sample and chromosome. For this analysis, a small number of samples were excluded to ensure that no two samples in the subset were more than 99% similar. The final values for the parameters (\( N_e = 8,824, \mu = 0.000492 \)) were computed as means weighted by chromosome length. These parameters were used in conjunction with a uniform recombination map to infer linkage information. We painted the chromosomes of each sample by running ChromoPainter with 40 expectation-maximization iterations, maximizing over the proportion of genetic material copied from each donor population with the restriction that only individuals from KH1, KH2, KH3 and KH4 were considered as donors (KHA was assumed to be an admixture population). Plots were generated by assigning a color to each founder population, with each segment being painted with the color of the population to which the donor sample belonged.

Ancestry analysis. We used ADMIXTURE to estimate the ancestry of the Cambodian samples. ADMIXTURE is a high-performance program for estimating ancestry in a model-based manner from large autosomal SNP genotype data sets. The ADMIXTURE model assumes low LD between markers (SNPs). Accordingly, we excluded SNP pairs that appeared to be linked using two filtering steps. First, we removed all SNPs with extremely low MAF (MAF ≤ 0.01), as these SNPs are less informative in the inference process. In a second step, we discarded SNPs according to the observed sample correlation coefficients. Using the PLINK tool set (see URLs), we scanned the genome with a sliding window of 100 SNPs in size, advanced in steps of 10 SNPs, and removed any SNP with correlation coefficient ≥ 0.02 with any other SNP within the window. The 5,484 SNPs that remained after this filtering (list available upon request) were used to run ADMIXTURE with 5-fold cross-validation, 1,000 replicates for bootstrapping and \( K \) (number of putative ancestral populations) = 4.

**F\( ST \) estimation.** To estimate \( F\( ST \)\) between two populations at a given SNP, we used the equation

\[
F_{ST} = 1 - \frac{\hat{\pi}_S}{\hat{\pi}_T}
\]

where \( \hat{\pi}_S \) is the average probability that two samples chosen at random from the same population carry different alleles at the SNP and \( \hat{\pi}_T \) is the average probability that two samples chosen at random from the joint population carry different alleles. If one of the alleles is observed with frequencies \( p_1 \) and \( p_2 \) in the two populations, we estimate these average probabilities as

\[
\hat{\pi}_S = \frac{1}{2} \left( 2p_1(1-p_2) + 2p_2(1-p_1) \right)
\]

\[
\hat{\pi}_T = 2 \times \frac{(p_1 + p_2)}{2} \times \left( 1 - \frac{(p_1 + p_2)}{2} \right)
\]
To estimate genome-wide $F_{ST}$ between two populations, we used the equation $F_{ST} = 1 - (\Pi_S / \Pi_T)$, where $\Pi_S$ is the average correspondence within the two populations and $\Pi_T$ is the average correspondence in the joint population. Given a genome-wide set of $M$ SNPs, the above can be expressed as

$$\text{Genome-wide } F_{ST} = 1 - \frac{\sum_{v=1}^{M} \pi_S}{\sum_{v=1}^{M} \pi_T}$$

From applying the estimators for $\pi_S$ and $\pi_T$ at each of the 86,128 SNPs, we derive genome-wide $F_{ST}$ estimates.

42. Bethell, D. et al. Artesunate dose escalation for the treatment of uncomplicated malaria in a region of reported artemisinin resistance: a randomized clinical trial. *PLoS ONE* 6, e19283 (2011).
43. Aurrecoechea, C. et al. PlasmoDB: a functional genomic database for malaria parasites. *Nucleic Acids Res.*, D539-D543 (2009).
44. Lawson, D.J., Hellenthal, G., Myers, S. & Falush, D. Inference of population structure using dense haplotype data. *PLoS Genet.* 8, e1002453 (2012).
45. Alexander, D.H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19, 1655–1664 (2009).