

Published in final edited form as:

Nat Genet. 2013 June ; 45(6): . doi:10.1038/ng.2624.

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

Olivo Miotto<sup>1,2,3</sup>, Jacob Almagro-Garcia<sup>1,3,4</sup>, Magnus Manske<sup>1,3</sup>, Bronwyn MacInnis<sup>1,3</sup>, Susana Campino<sup>1,3</sup>, Kirk A Rockett<sup>1,3,4</sup>, Chanaki Amarunga<sup>5</sup>, Pharath Lim<sup>5,6</sup>, Seila Suon<sup>6</sup>, Sokunthea Sreng<sup>6</sup>, Jennifer M Anderson<sup>5</sup>, Socheat Duong<sup>6</sup>, Chea Nguon<sup>6</sup>, Char Meng Chuor<sup>6</sup>, David Saunders<sup>7</sup>, Youry Se<sup>8</sup>, Chantap Lon<sup>8</sup>, Mark M Fukuda<sup>7,9</sup>, Lucas Amenga-Etego<sup>10</sup>, Abraham VO Hodgson<sup>10</sup>, Victor Asoala<sup>10</sup>, Mallika Imwong<sup>2,11</sup>, Shannon Takala-Harrison<sup>12</sup>, Francois Nosten<sup>2,13,14</sup>, Xin-zhuan Su<sup>5</sup>, Pascal Ringwald<sup>15</sup>, Frédéric Ariey<sup>16</sup>, Christiane Dolecek<sup>14,17</sup>, Tran Tinh Hien<sup>14,17</sup>, Maciej F Boni<sup>14,17</sup>, Cao Quang Thai<sup>17</sup>, Alfred Amambua-Ngwa<sup>18</sup>, David J Conway<sup>18,19</sup>, Abdoulaye A Djimdé<sup>20</sup>, Ogobara K Doumbo<sup>20</sup>, Issaka Zongo<sup>21</sup>, Jean-Bosco Ouedraogo<sup>21</sup>, Daniel Alcock<sup>3</sup>, Eleanor Drury<sup>3</sup>, Sarah Auburn<sup>22</sup>, Oliver Koch<sup>1</sup>, Mandy Sanders<sup>3</sup>, Christina Hubbard<sup>4</sup>, Gareth Maslen<sup>1,3</sup>, Valentin Ruano-Rubio<sup>1,3,4</sup>, Dushyanth Jyothi<sup>1,3</sup>, Alistair Miles<sup>1,4</sup>, John O'Brien<sup>1,4</sup>, Chris Gamble<sup>23</sup>, Samuel O Oyola<sup>3</sup>, Julian C Rayner<sup>3</sup>, Chris I Newbold<sup>1,3,24</sup>, Matthew Berriman<sup>3</sup>, Chris CA Spencer<sup>4</sup>, Gilean McVean<sup>4</sup>, Nicholas P Day<sup>2,14</sup>, Nicholas J White<sup>2,14</sup>, Delia Bethel<sup>7</sup>, Arjen M Dondorp<sup>2,14</sup>, Christopher V Plowe<sup>12</sup>, Rick M Fairhurst<sup>5</sup>, and Dominic P Kwiatkowski<sup>1,3,4,\*</sup>

<sup>1</sup>MRC Centre for Genomics and Global Health, University of Oxford, Oxford OX3 7BN, UK

<sup>2</sup>Mahidol-Oxford Tropical Medicine Research Unit, Mahidol University, Bangkok 10400, Thailand

<sup>3</sup>Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1SA, UK <sup>4</sup>Wellcome Trust Centre

for Human Genetics, University of Oxford, Oxford OX3 7BN, UK <sup>5</sup>National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA <sup>6</sup>National

Center for Parasitology, Entomology and Malaria Control, Phnom Penh, Cambodia <sup>7</sup>Department

of Immunology and Medicine, US Army Medical Component, Armed Forces Research Institute of

Medical Sciences (USAMC-AFRIMS), Bangkok, Thailand <sup>8</sup>US Army Medical Component, Armed

Forces Research Institute of Medical Sciences (USAMC-AFRIMS), Phnom Penh, Cambodia

<sup>9</sup>Armed Forces Health Surveillance Center, Silver Spring MD 20904, USA <sup>10</sup>Navrongo Health

Research Centre, Navrongo, Ghana <sup>11</sup>Department of Molecular Tropical Medicine and Genetics,

Faculty of Tropical Medicine, Mahidol University, Bangkok 10400, Thailand <sup>12</sup>Howard Hughes

Medical Institute, University of Maryland School of Medicine, Baltimore, Maryland 21201, USA

<sup>13</sup>Shoklo Malaria Research Unit, Mae Sot, Tak 631 10, Thailand <sup>14</sup>Centre for Tropical Medicine,

University of Oxford, Oxford OX3 7LJ, UK <sup>15</sup>Global Malaria Programme, World Health

Organization, Geneva, Switzerland <sup>16</sup>Unité d'Immunologie Moléculaire des Parasites, Institut

Pasteur, Paris 75015, France <sup>17</sup>Oxford University Clinical Research Unit, Wellcome Trust Major

Overseas Programme, Ho Chi Minh City, Vietnam <sup>18</sup>MRC Laboratories, Fajara, The Gambia

<sup>19</sup>London School of Hygiene and Tropical Medicine, London WC1E 7HT, UK <sup>20</sup>Malaria Research

and Training Center, Faculty of Pharmacy, University of Science, Techniques and Technologies

of Bamako, Mali <sup>21</sup>Institut de Recherche en Sciences de la Santé, Direction Régionale de l'Ouest,

Bobo-Dioulasso, Burkina Faso <sup>22</sup>Menzies School of Health Research, Charles Darwin University,

Darwin, Northern Territories 0811, Australia <sup>23</sup>Department of Statistics, University of Oxford,

Oxford, OX1 3TG, UK <sup>24</sup>Weatherall Institute of Molecular Medicine, University of Oxford, Oxford

OX3 9DS, UK

### Abstract

We describe an analysis of genome variation in 825 *Plasmodium falciparum* samples from Asia and Africa that reveals an unusual pattern of parasite population structure at the epicentre of

artemisinin resistance in western Cambodia. Within this relatively small geographical 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 remarkably high levels of haplotype homozygosity, indicative of founder effects and recent population expansion. We provide a catalogue of SNPs that show high levels of differentiation in the artemisinin-resistant subpopulations, including codon variants in various transporter proteins and DNA mismatch repair proteins. These data provide a population genetic framework for investigating the biological origins of artemisinin resistance and for defining molecular markers to assist its elimination.

## Introduction

The malaria parasite *Plasmodium falciparum* has a remarkable 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 Cambodia.<sup>1</sup> The most common form of chloroquine resistance was observed there in the late 1950's before it spread around the world<sup>2</sup> and the most common forms of clinically significant pyrimethamine resistance<sup>3</sup> and sulfadoxine resistance<sup>4</sup> 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 Cambodia<sup>5-8</sup> and appears to be emerging in neighbouring regions.<sup>9,10</sup> These recent developments have grave implications for public health, since artemisinin derivatives are the mainstay of malaria treatment worldwide. There is an urgent need to discover the parasite genetic factors that cause artemisinin resistance<sup>11</sup> and to identify effective markers to monitor its spread.<sup>12</sup>

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 analysed patterns of genome variation in 825 *P. falciparum* samples collected at 10 locations in West Africa and South-East Asia. These include Ghana, Mali, Burkina Faso, The Gambia, Thailand, Vietnam and 4 locations in Cambodia (Table 1). An estimated median of 2 Gbp of sequence read data were obtained for each sample using the Illumina Genome Analyser platform, and SNP genotype calls were made using a well-validated set of algorithms and quality control procedures whose development and evaluation is described in detail elsewhere.<sup>13</sup> Here we focus on a set of 86,158 coding SNPs that could be genotyped with confidence in the majority of samples, as detailed at <http://www.malariagen.net/resource/10>. For purposes of quality assurance, 733/825 (89%) of the samples analysed were independently assayed for 148 SNPs using the Sequenom platform: the genotype calls produced by Sequenom showed 99.4% concordance with the genotypes calls obtained from Illumina sequencing (see Supplementary Note).

## Population structure

As a starting point for the analysis of population structure, we constructed a neighbour-joining tree (see Supplementary Note) of all 825 samples based on a pairwise distance matrix (Figure 1a). Cambodian and Vietnamese samples formed a complex branching cluster of ancestrally related individuals, which could be clearly distinguished from Thai samples and, to a much greater extent, from West African samples. Principal components analysis (PCA, see Supplementary Note) revealed a surprising pattern of population structure within the total sample set: the first principal component (PC1) separated samples from different continents, but the second and third components (PC2 and PC3 respectively) represented variation in western Cambodia alone (Figure 1b). Although we expected to find

high levels of population structure in South-East Asia, consistent with relatively low levels of malaria transmission<sup>14</sup>, the strength of population structure observed in western Cambodia was remarkable when compared to the other populations in this sample set. To put this in geographical context, the western Cambodian sites sampled are located up to ~200km apart, while the other South-East Asian locations were separated by up to ~1000km, and the West African locations by up to ~1,500 km (Supplementary Figure 1). Thus the population structure observed in western Cambodia cannot be explained simply by geographical distance. We considered the possibility that it might be an artefact of increased variance introduced by the relatively large number of samples from Cambodia, but the signal was found to be robust when correcting for sample size by sub-sampling (Supplementary Figure 2).

The existence of multiple parasite subpopulations in western Cambodia became evident when PCA was restricted to South-East Asia (Figure 1c, d). A core group of western Cambodian samples clustered with samples from north-eastern Cambodia and Vietnam, close to samples from Thailand, while three distinct groups of western Cambodian samples were identified as outliers from this core group by the first, second and third principal components, respectively. An inspection of principal components beyond PC3 did not reveal further outlier groups. Other samples were intermediate between the core and outlier groups, suggesting admixture between the groups. To characterise this in more detail we used a process of *chromosome painting*, a probabilistic method of reconstructing the chromosomes of each individual sample from homologous segments of DNA in samples representative of the core group and three outlier groups (see Figure 2, Methods). This 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 Figure 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 neighbouring 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 (Figure 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 neighbour-joining tree of Cambodian and Vietnamese samples showed clusters corresponding to KH1, KH2, KH3 and KH4, interspersed with KHA samples (Supplementary Figure 4). When the ADMIXTURE program was used to model the Cambodian samples as admixture of 4 putative subpopulations, there was 82% concordance between the classification of samples by ADMIXTURE and by the chromosome painting method described above (see Methods, Supplementary Figure 5). Both analyses confirmed that most samples from north-eastern Cambodia and Vietnam are very similar to KH1 samples in western Cambodia. It was evident from all methods that our classification of samples into KH1/2/3/4/A should be regarded as a first approximation, since the population structure is clearly complex and will require further epidemiological sampling to understand in detail.

Table 2 shows the frequency of subpopulations at different locations in Cambodia. All of the subpopulations were present in western Cambodia, while KH1 was predominant in north-eastern Cambodia. The samples came from four independent field studies conducted over the past 5 years: there was clear evidence of population structure within each study, and it was not limited to a particular site or sampling year (Supplementary Tables 1-3). The discovery of several distinct subpopulations of *P. falciparum* co-existing 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 analysed parasite clearance rates in patients with acute falciparum malaria following artesunate treatment.<sup>6,8</sup> *In vivo* estimation of parasite clearance rate from frequent peripheral blood parasite counts is becoming the standard method used for surveillance of artemisinin resistance, since no laboratory method has yet been developed to correlate reliably 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 based on the slope of the linear portion of the parasite clearance curve<sup>15</sup>. This revealed a significant prolongation of parasite clearance half-life in the KH2, KH3 and KH4 subpopulations compared to the KH1 subpopulation (Figure 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 samples (all from western Cambodia) was 6.8 hours compared to 4.1 hours in 6 KH1 samples from western Cambodia ( $P=2\times 10^{-4}$  by Mann-Whitney test) and 2.7 hours in all 50 KH1 samples from Cambodia ( $P=2\times 10^{-14}$ ). More detailed results for all subpopulations are given in Table 3. KHA samples showed wide variance in parasite clearance half-life, consistent with their genetic makeup as an admixture of artemisinin-sensitive and artemisinin-resistant types. The long median clearance half-life in this group is consistent with selection of recombinants with an ART-R genetic background. However, we observed no direct correlation between admixture proportions and parasite clearance half-life (Supplementary Figure 6), which suggests that the acquisition of ART-R phenotypes depends on receiving a limited set of specific DNA segments from resistant ancestors.

To put this in an epidemiological context, recent clinical studies show that parasite clearance rates following artesunate treatment are significantly faster in north-eastern Cambodia (Fairhurst *et al.*, manuscript in preparation) and in Vietnam<sup>10</sup> than in western Cambodia. It appears likely that these geographical differences in parasite clearance rate can be largely attributed to the predominance of KH2, KH3, KH4 and KHA in western Cambodia, whereas KH1 predominates in north-eastern Cambodia and Vietnam.

## Founder effects

A striking 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 genome-wide estimates of the fixation index ( $F_{ST}$ ) between subpopulations: e.g., the pairwise  $F_{ST}$  for KH2 vs KH4 (0.38) was considerably higher than that for Thailand vs Ghana (0.16) (Supplementary Table 4). It was also evident from counting the number of SNPs showing high levels of  $F_{ST}$  between subpopulations: e.g. the number of SNPs with  $F_{ST}>0.5$  was 11 for Thailand vs KH1, 562 for Thailand vs Ghana and 1200 for KH2 vs 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 of low frequency alleles in KH2, KH3 and KH4, compared to KH1 and to parasites from other parts of South-East Asia and West Africa (Figure 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 non-synonymous to synonymous

polymorphisms (overall pN/pS was 2.4 for SNPs with  $F_{ST} < 0.2$ , and 2.1 for SNPs with  $F_{ST} > 0.8$ ; 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 in KH2, KH3 and KH4 compared to KH1 and other populations (Figure 4b). The loss of haplotypic diversity was also quantified more directly, e.g. levels of haplotype homozygosity, measured in a sliding window across the genome (see Supplementary Note), were considerably higher for KH2, KH3 and KH4 than for KH1 (Figure 4c, Supplementary Figure 7). The level of haplotype diversity varied considerably across the genome, and in some cases it was found that a single haplotype predominates across a large chromosomal region, e.g. KH4 has essentially a single haplotype across most of chromosome 9, and that KH2 has essentially a single haplotype across approximately half of chromosome 13 (1.4-3.4Mb), in marked contrast with KH1 (Supplementary Figure 8).

An analysis of within-host diversity, quantified in terms of the  $F_{WS}$  metric<sup>13</sup>, shows that the vast majority of samples in KH2, KH3 and KH4 are essentially clonal, as are those in the recombinant KHA population (Supplementary Figure 9). Although within-sample diversity was found to be generally lower in Southeast Asian populations than in West Africa, the artemisinin-resistant samples are clearly more likely to be clonal than those in KH1 or in Thailand. This suggests a high level of inbreeding in the KH2/3/4 populations.

We conclude that western Cambodia is home to at least three distinct founder populations of *P. falciparum*, each of which is artemisinin resistant, co-existing with a more diverse and widely-distributed subpopulation to which they are ancestrally related (Figure 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 (Figure 2). This is most evident in KHA, which appears as a mosaic of the other populations. This appearance might have arisen from recent admixture or from 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. Supplementary Tables 7-9 lists SNPs that differentiate the founder populations from other populations analysed here.

To illustrate the potential utility of this dataset, it has recently been reported that low levels of artemisinin-resistance observed in North-Western Thailand are associated with a genomic region of low haplotype diversity on chromosome 13, around the 1.7-1.8 Mb position<sup>7</sup>. We find that KH2 has essentially a single haplotype extending across half of chromosome 13, from the 1.4Mb to the 3.4Mb position (Supplementary Figure 8). At position 13:2075035 we observe a codon variant (Y175S) in an ABC transporter gene (PF13\_0271) for which all samples from West Africa carry the ancestral allele, whereas almost all KH2 samples have the derived (non-ancestral) allele. Since this gene has been experimentally associated with chloroquine and quinine resistance<sup>16</sup> and has previously been reported to lie within a genomic region of positive selection<sup>17</sup>, it would appear to be an important candidate for further investigation.

Supplementary Table 10 lists transporter genes that are highly differentiated in the founder populations, as these are prime candidates for the causation of antimalarial drug

resistance.<sup>16,18</sup> 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 (a cation-transporting ATPase 1) and PF13\_0252 (a nucleoside transporter).

The founder populations also serve as a reservoir for a variety of known antimalarial resistance alleles (Supplementary Table 11). In *pfcr1*<sup>19</sup>, almost all samples in KH2, KH3 and KH4 have the resistant haplotype CVIET, whereas KH1 has a wider variety of haplotypes.<sup>20,21</sup> A comparison of *pfmdr1*<sup>22-24</sup> sequence coverage suggests that most samples in KH4 and KH3 have higher copy numbers than those in KH1 and KH2 (Supplementary Table 12). In *pfdhps*<sup>25</sup>, KH2 has a high frequency of the triple mutant haplotype comprising K540N, A581G and A437G<sup>26,27</sup>, whereas KH3 and KH4 have a high frequency of the triple mutant haplotype comprising S436A, K540E and A437G<sup>28</sup>. In *pfdhfr*<sup>25,29</sup> the codon variant I164L, which characterizes a highly resistant triple or quadruple mutant haplotype<sup>30</sup>, is 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 South-East Asia because of the existence of hypermutable parasites, based on experimental evidence that a *P. falciparum* strain originating from Indochina shows accelerated resistance to multiple structurally unrelated drugs<sup>31</sup>. It will therefore be of interest to researchers in this area that a number of the most differentiated nonsynonymous SNPs in the KH2 subpopulations are located in genes involved in DNA mismatch repair pathways (Supplementary Table 13). These include PMS1 and MLH, the two proteins that make up MutL, a DNA mismatch repair heterodimer that is highly conserved throughout evolution<sup>32</sup>; and UvrD, a helicase that is critical for methyl-directed mismatch repair and interacts physically with MutL, which loads it onto DNA.<sup>33-35</sup> This could be highly relevant, since MutL variants in bacteria have been shown to cause hypermutation and 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. The parasites are transmitted to humans by *Anopheles* mosquitoes, and a crucial step in this process is 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 high transmission regions such as West Africa.<sup>36,37</sup>

The population structure observed in Western Cambodia is unusual in several respects. Within this relatively small geographical 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 remarkably high levels of haplotype homozygosity, indicative of founder effects and recent population expansion. These subpopulations are associated with clinical resistance to artemisinin. Two key questions arise from these findings: what caused the founder effects and what is the biological significance of their association with artemisinin 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 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 recent expansion of artemisinin-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 artemisinin 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 the history of previous antimalarial drug usage. Low transmission intensity and other factors that favour inbreeding could assist the propagation of a multigenic resistance phenotype once it has emerged. High rates of inbreeding might also arise due to physical isolation, e.g. if a group of parasites became isolated in a remote area of the jungle, or due to some form of reproductive isolation, e.g. it is conceivable that different parasite subpopulations are preferentially transmitted by different *Anopheles* vector species.<sup>38</sup>

These observations provide a population 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 pyrimethamine.<sup>39,40</sup> This may explain why these subpopulations possess a variety of haplotypes that are highly resistant to antimalarials other than artemisinin - 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 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 favourable environment for parasite inbreeding.

A major objective of the WHO Global Plan for Artemisinin Resistance Containment is to stop the spread of resistant parasites.<sup>41</sup> 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 characterisation of their geographical distribution and the ecological niches that they occupy. This will require the acquisition of genomic epidemiological data at 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.<sup>9</sup>

## Methods

### Ethical Approval

All samples in this study were derived from blood samples obtained from patients with *P. falciparum* malaria, collected with informed consent from the patient or a parent or guardian. At each location, sample collection was approved by the appropriate local ethics committee; a full list is given in the Supplementary Note.

## Clinical Phenotype Determination in Cambodian Samples

In Pursat, patients aged 10-65 years had uncomplicated malaria with a *P. falciparum* density 10,000/ $\mu$ l whole blood.<sup>8</sup> Patients were treated with 4 mg/kg artesunate given orally at 0, 24 and 48 h, followed by 15 mg/kg mefloquine given orally at 72 h, and 10 mg/kg mefloquine given orally at 96 h. During treatment, thick blood films were made from finger-prick blood samples every 6 h until the asexual parasite density was undetectable. In Ratanakiri, patients were aged 2-65 years and their parasite density was counted every 6 h until undetectable or 48 h elapsed, whichever occurred first. Further details can be obtained from the trial registrations (NCT00341003 for Pursat; NCT01240603 for Ratanakiri).

In Tasanh and Pailin, two separate but coordinated studies were conducted within the Artemisinin Resistance Confirmation, Characterisation and Containment (ARC3) multi-site study, with harmonized methodologies. Depending on the study arm, patients recruited were aged 18-65, presenting acute *P. falciparum* malaria symptoms, with a parasite density of 1000-200,000/ $\mu$ L. In Tasanh patients were treated with artesunate, given as a single oral dose of 2, 4 or 6 mg/kg/day for 7 days. Parasite counts based on blood smears were taken up to 8 times on Day 0, and then 4 times a day until 2 consecutive thick blood films were negative for asexual parasites. Further details of the Tasanh study are given in<sup>42</sup>. In Pailin patients were treated with artesunate in a daily dose of 6 or 8 mg/kg administered as a single or split dose. Three day regimens were followed by a split dose of mefloquine at 72 and 96 hours. Peripheral blood parasitemia was assessed at 0, 2, 4, 6, 8, 12, and 18 hours after enrolment and then 6 hourly until two consecutive slides were negative. More details can be obtained from the trial registration (ISRCTN15351875).

For all sites, parasite clearance half-life estimates were computed from the frequent parasite counts obtained during treatment. The statistical models used to estimate the parasite clearance rate constant and lag phase duration were fitted using the Parasite Clearance Estimator<sup>15</sup> developed by the Worldwide Antimalarial Resistance Network (WWARN), accessible at <https://www.wwarn.org/research/parasite-clearance-estimator>.

## Sequencing and Genotyping

The sequencing and genotyping methods applied to the clinical samples were previously published<sup>13</sup> and are briefly summarized here; an extended description is provided in the Supplementary Note. For most samples, parasitized erythrocytes were obtained directly from the blood samples, after leukocyte depletion to remove the majority of human DNA, without culturing. A minority of samples from Thailand and Cambodia were adapted to *in vitro* culture prior to DNA extraction. Samples with  $>1 \mu$ g DNA and  $<60\%$  human DNA contamination were sequenced by the Illumina Genome Analyzer II, with read lengths ranging from 37 to 105 bp. From short sequence read alignments against the *P. falciparum* 3D7 reference, we obtained a set of 86,158 quality-controlled SNPs which passed a series of stringent quality filters<sup>13</sup> (the V1.0 SNP Catalogue). Samples were genotyped at each SNP by a single allele, based on sequencing read counts.

## Allele Frequency Analysis

Allele frequencies at any given SNP in any given population were determined by analysing all genotyped samples (i.e. samples with undetermined genotype at the SNP were excluded from the computation). The *non-reference allele frequency* (NRAF) was computed as the proportion of genotyped samples in the population whose genotype was not the reference allele. The *minor allele frequency* (MAF) was computed as the proportion of genotyped samples carrying the least common genotype in the population (i.e.  $MAF=NRAF$  if  $NRAF<0.5$ ;  $MAF=(1-NRAF)$  otherwise).

In analyses where *derived allele frequency* (DAF) was used, we determined the putative ancestral state alleles at each position by comparison of the allele in the *P. falciparum* 3D7 reference sequence with alleles in homologous sequences observed in *P. reichenowi* (*Pr*) as previously described<sup>13</sup>. The putative ancestral/derived alleles at each SNP are included in the online 86k SNP dataset. The *derived allele frequency* (DAF) was computed as the proportion of genotyped samples carrying the derived allele in a population; the DAF was undetermined at positions where the putative ancestral/derived state could not be established.

We determined gene reading frames and exon boundaries from the PlasmoDB 5.5 annotation ([www.plasmodb.org](http://www.plasmodb.org)) of the 3D7 genome<sup>43</sup>. Each SNP in our 86k set was classified as synonymous or non-synonymous according to whether an amino acid change occurs 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<sup>44</sup> is a probabilistic method of reconstructing the ancestry of individuals in a recombinant population. While commonly used methods for discovering population structure (such as ADMIXTURE<sup>45</sup>) tend to consider SNPs independently, chromosome painting utilizes linkage information in the dataset in its probability computation, based on 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 chunk of DNA from a given individual, based on its similarity to corresponding chunks in individuals from the donor populations (rather than perform this analysis on each SNP independently). This results in visualizations of the recombination patterns, which assign a colour to each donor population, and use these colours to paint each DNA chunk according to the most probable assignment.

We performed chromosome painting using the Chromopainter software package (<http://www.maths.bris.ac.uk/~madjl/finestructure/chromopainter.html>). The parameters  $N_e$  (effective population size) and  $\mu$  (global mutation rate) were estimated in Cambodian samples by running the expectation-maximization (E-M) algorithm for 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 ( $N_e=8824$ ,  $\mu=0.000492$ ) were computed as a weighted mean, according to chromosome length. These parameter values were used in conjunction with a uniform recombination map, to infer linkage information. We painted the chromosomes of each sample by running Chromopainter for 40 E-M iterations, maximizing over copying proportions, with the restriction that only individuals from KH1, KH2, KH3 and KH4 were considered as donors (i.e. KHA was assumed to be an admixture population). The plots were generated by assigning a colour to each founder populations (KH1=purple, KH2=blue, KH3=red and KH4=green), with each segment being painted with the colour of the population the donor sample belonged to.

### Ancestry Analysis

We used ADMIXTURE<sup>45</sup> (<http://www.genetics.ucla.edu/software/admixture/>) to estimate ancestry of the Cambodian samples. ADMIXTURE is a high-performance program for estimating ancestry in a model-based manner from large autosomal SNP genotype datasets. The ADMIXTURE model assumes low linkage disequilibrium between markers (SNPs). Accordingly, we excluded SNPs pairs that appeared to be linked, by two filtering steps. First, we removed all SNPs with extremely low MAF (MAF  $\leq$  0.01), as these SNPs are be

less informative towards the inference process. In a second step, we discarded SNPs according to the observed sample correlation coefficients. Using the plink tool-set (<http://pnu.mgh.harvard.edu/~purcell/plink/>), we scanned the genome with a sliding window of size 100 SNPs, advanced in steps of 10 SNPs, and removed any SNP with correlation coefficient  $> 0.02$  with any other SNP within the window. The 5484 SNPs that remained after this filtering (available upon request) was used to run ADMIXTURE with 5-fold cross validation, 1000 replicates for bootstrapping and  $K=4$ .

### $F_{ST}$ Estimation

To estimate the  $F_{ST}$  between two populations at a given SNP, we used

$$F_{ST} = 1 - \frac{\hat{\pi}_s}{\hat{\pi}_t}$$

where  $\hat{\pi}_s$  is the average probability that two samples chosen at random from the same population will carry different allele at the SNP, and  $\hat{\pi}_t$  is the average probability that two samples chosen at random from the joint population will carry different allele. If one of the alleles is observed with frequencies  $p_1$  and  $p_2$  in the two populations, we estimate

$$\hat{\pi}_s = \frac{1}{2} (2p_1(1-p_1) + 2p_2(1-p_2))$$

$$\hat{\pi}_t = 2 \cdot \frac{(p_1+p_2)}{2} \cdot \left(1 - \frac{(p_1+p_2)}{2}\right)$$

To estimate *genome-wide*  $F_{ST}$  between two populations, we used  $GW F_{ST} = 1 - \left(\frac{\hat{\pi}_s}{\hat{\pi}_t}\right)$  where  $\hat{\pi}_s$  is the average correspondence *within* the two populations, and  $\hat{\pi}_t$  is the average correspondence *in the joint* population. Given a genome-wide set of  $M$  SNPs, the above can be expressed as

$$GW F_{ST} = 1 - \frac{\sum_{\nu=1}^M \hat{\pi}_s}{\sum_{\nu=1}^M \hat{\pi}_t}$$

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

### Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

### Footnotes

\*Correspondence and requests for materials should be addressed to dominic@sanger.ac.uk

**Author Contributions** SC, CA, PL, SuS, SrS, JMA, SD, CN, CMC, DS, YS, CL, MMF, LA-E, AVOH, VA, MI, FN, X-ZS, PR, FA, CD, TTH, MFB, CQT, AA-N, DJC, AD, OKD, IZ, J-BO, SA, NPD, NJW, DB, AMD, CVP and RMF carried out field and laboratory studies to obtain *P. falciparum* samples for sequencing. CA, PL, SS, JMA, SD, CN, CMC, DS, YS, CL, MMF, FN, X-ZS, PR, FA, NJW, DB, AMD, CVP and RMF carried out clinical studies to obtain ART phenotype data. SC, DA, ED, MS, SA, OK, SO, BM, CIN and MB developed and implemented methods for sample processing and sequencing library preparation. OM, JA-G, MM, GM, VRR, DJ and AM developed software for data management and visualisation. KAR, CH, DA and MM carried out validation experiments. CVP, ST-H, GMV and RMF contributed to development of the project. BM, MB, CIN and JCR provided project management and oversight. OM, JA-G, MM, JO'B, CG and CS carried out data analyses. DPK, OM and JA-G wrote the manuscript. All Authors reviewed the manuscript.

**URLs** Information page for this study: <http://www.malariagen.net/resource/12>  
 Information page for the SNP list used: <http://www.malariagen.net/resource/10>  
 ENA archive records: <http://www.ebi.ac.uk/ena/data/search/?query=plasmodium>  
 WWARN Parasite Clearance Estimator: <https://www.wwarn.org/research/parasite-clearance-estimator>  
 bwa: <http://bio-bwa.sourceforge.net/>  
 PlasmoDB: <http://www.plasmodb.org/>  
 R language: <http://www.r-project.org/>  
 R ape package: <http://ape.mpl.ird.fr/>  
 Chromopainter: <http://www.maths.bris.ac.uk/~madjl/finestructure/chromopainter.html>  
 ADMIXTURE: <http://www.genetics.ucla.edu/software/admixture/>  
 plink: <http://pngu.mgh.harvard.edu/~purcell/plink/>

**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.

**Supplementary Information** The following Supplementary Information files are provided: a document containing the Supplementary Note, Supplementary Tables and Supplementary Figures referenced by the paper; and two Excel spreadsheets containing the electronic forms of Supplementary Table 7 and Supplementary Table 8.

The authors declare no competing financial interests.

## Acknowledgments

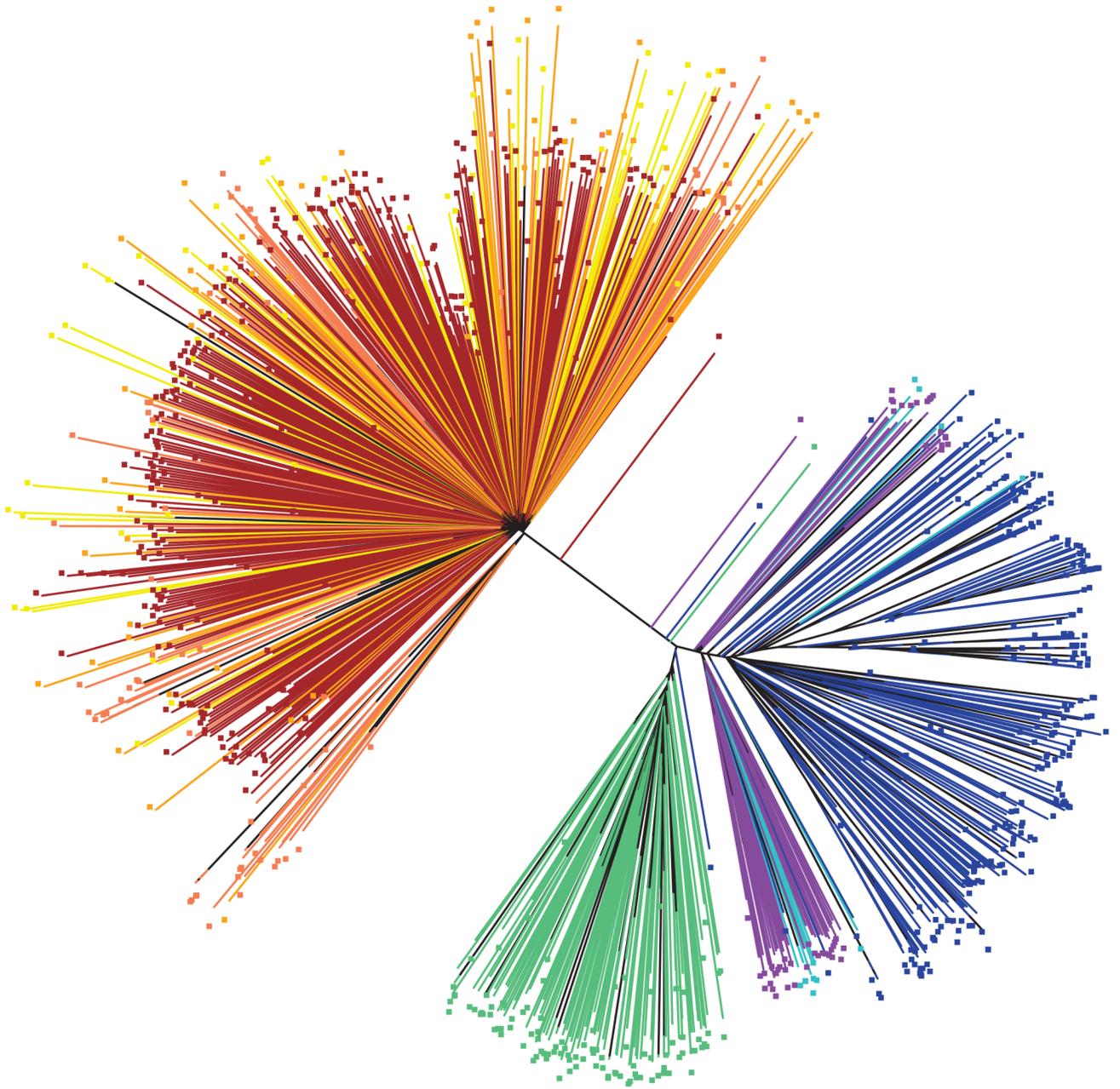
We would like to thank Sambunny Uk (National Center for Parasitology, Entomology and Malaria Control, Cambodia) and Erika S. Phelps (NIAID, NIH, USA) for their contributions in the Cambodian studies. We are grateful to Tim Anderson for helpful review comments. The sequencing, genotyping and analysis components of this study were supported by The Wellcome Trust through core funding of the Wellcome Trust Sanger Institute (077012/Z/05/Z; 098051), core funding of the Wellcome Trust Centre for Human Genetics (075491/Z/04; 090532/Z/09/Z) and a Strategic Award (090770/Z/09/Z); and the Medical Research Council through the MRC Centre for Genomics and Global Health (G0600230) and an MRC Professorship to Dominic Kwiatkowski (G19/9). Other parts of this study were partly supported by the Wellcome Trust; the Medical Research Council; the Division of Intramural Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health; and a Howard Hughes Medical Institute International Scholarship (55005502) to Abdoulaye Djimde. PR 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.

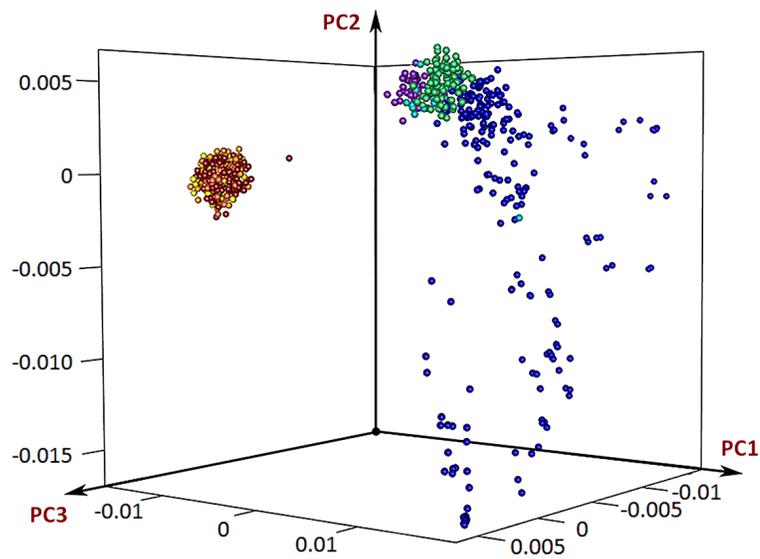
## References

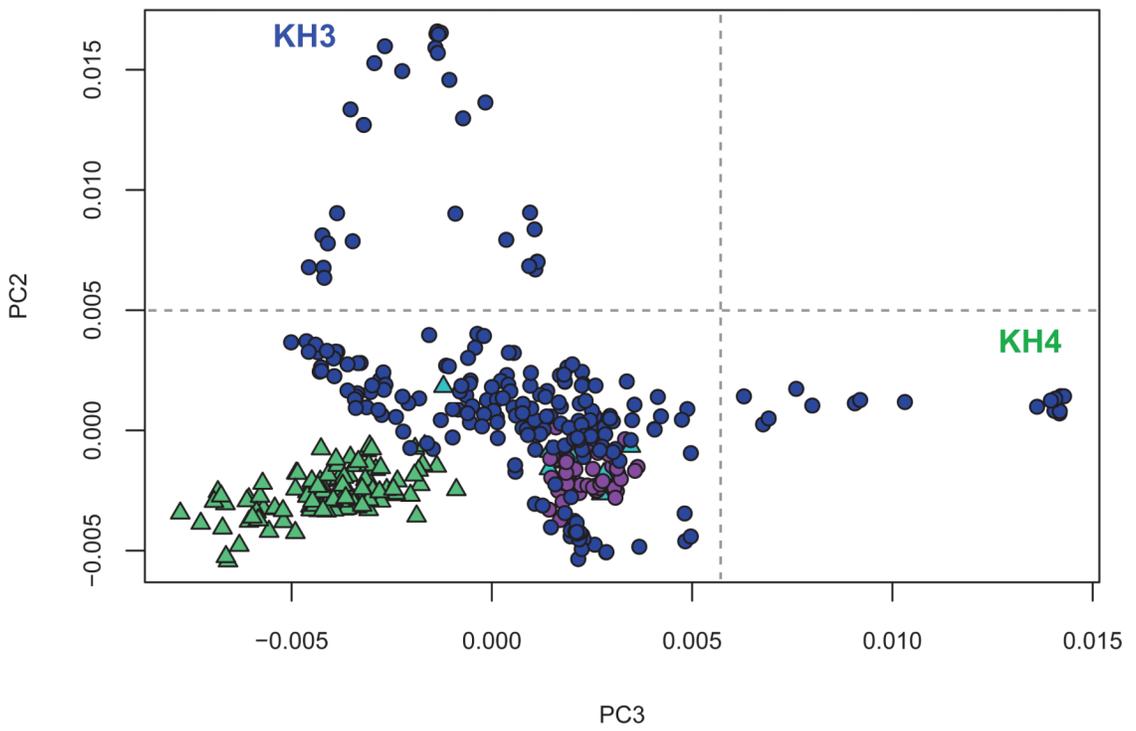
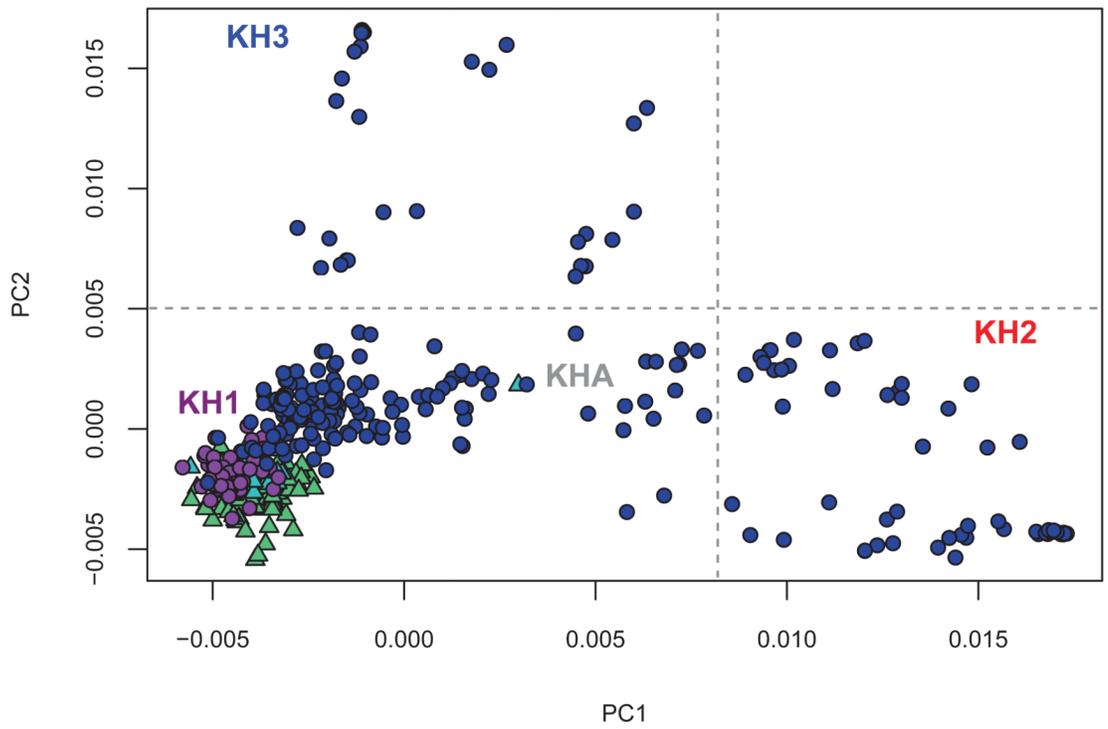
1. White NJ. Artemisinin resistance--the clock is ticking. *Lancet*. 2010; 376:2051–2. [PubMed: 21168039]
2. Payne D. Spread of chloroquine resistance in *Plasmodium falciparum*. *Parasitology today*. 1987; 3:241–6. [PubMed: 15462966]
3. Roper C, et al. Intercontinental spread of pyrimethamine-resistant malaria. *Science*. 2004; 305:1124. [PubMed: 15326348]
4. Mita T, et al. Limited geographical origin and global spread of sulfadoxine-resistant dhps alleles in *Plasmodium falciparum* populations. *The Journal of Infectious Diseases*. 2011; 204:1980–8. [PubMed: 22021623]
5. Noedl H, et al. Evidence of artemisinin-resistant malaria in western Cambodia. *The New England Journal of Medicine*. 2008; 359:2619–20. [PubMed: 19064625]
6. Dondorp AM, et al. Artemisinin resistance in *Plasmodium falciparum* malaria. *N Engl J Med*. 2009; 361:455–67. [PubMed: 19641202]
7. Cheeseman IH, et al. A major genome region underlying artemisinin resistance in malaria. *Science*. 2012; 336:79–82. [PubMed: 22491853]
8. Amaratunga C, et al. Artemisinin-resistant *Plasmodium falciparum* in Pursat province, western Cambodia: a parasite clearance rate study. *The Lancet infectious diseases*. 2012; 12:851–8. [PubMed: 22940027]

9. Phyo AP, et al. Emergence of artemisinin-resistant malaria on the western border of Thailand: a longitudinal study. *Lancet*. 2012;6736.
10. Hien TT, et al. In vivo susceptibility of *Plasmodium falciparum* to artesunate in Binh Phuoc Province, Vietnam. *Malaria journal*. 2012; 11:355. [PubMed: 23101492]
11. Anderson TJ, et al. High heritability of malaria parasite clearance rate indicates a genetic basis for artemisinin resistance in western Cambodia. *The Journal of infectious diseases*. 2010; 201:1326–30. [PubMed: 20350192]
12. Dondorp AM, et al. The threat of artemisinin-resistant malaria. *The New England Journal of Medicine*. 2011; 365:1073–5. [PubMed: 21992120]
13. Manske M, et al. Analysis of *Plasmodium falciparum* diversity in natural infections by deep sequencing. *Nature*. 2012; 487:375–9. [PubMed: 22722859]
14. Hay SI, et al. A world malaria map: *Plasmodium falciparum* endemicity in 2007. *PLoS medicine*. 2009; 6:e1000048. [PubMed: 19323591]
15. Flegg JA, Guerin PJ, White NJ, Stepniewska K. Standardizing the measurement of parasite clearance in falciparum malaria: the parasite clearance estimator. *Malaria journal*. 2011; 10:339. [PubMed: 22074219]
16. Mu J, et al. Multiple transporters associated with malaria parasite responses to chloroquine and quinine. *Mol Microbiol*. 2003; 49:977–89. [PubMed: 12890022]
17. Mu J, et al. *Plasmodium falciparum* genome-wide scans for positive selection, recombination hot spots and resistance to antimalarial drugs. *Nat Genet*. 2010; 42:268–71. [PubMed: 20101240]
18. Anderson TJ, et al. Are transporter genes other than the chloroquine resistance locus (*pfcr*) and multidrug resistance gene (*pfmdr*) associated with antimalarial drug resistance? *Antimicrob Agents Chemother*. 2005; 49:2180–8. [PubMed: 15917511]
19. Fidock DA, et al. Mutations in the *P. falciparum* digestive vacuole transmembrane protein PfCRT and evidence for their role in chloroquine resistance. *Mol Cell*. 2000; 6:861–71. [PubMed: 11090624]
20. Lim P, et al. *pfcr* polymorphism and chloroquine resistance in *Plasmodium falciparum* strains isolated in Cambodia. *Antimicrobial agents and chemotherapy*. 2003; 47:87–94. [PubMed: 12499174]
21. Isozumi R, et al. Longitudinal survey of *Plasmodium falciparum* infection in Vietnam: characteristics of antimalarial resistance and their associated factors. *Journal of clinical microbiology*. 2010; 48:70–7. [PubMed: 19889891]
22. Foote SJ, et al. Several alleles of the multidrug-resistance gene are closely linked to chloroquine resistance in *Plasmodium falciparum*. *Nature*. 1990; 345:255–258. [PubMed: 2185424]
23. Peel SA, Bright P, Yount B, Handy J, Baric RS. A strong association between mefloquine and halofantrine resistance and amplification, overexpression, and mutation in the P-glycoprotein gene homolog (*pfmdr*) of *Plasmodium falciparum* in vitro. *The American journal of tropical medicine and hygiene*. 1994; 51:648–58. [PubMed: 7985758]
24. Preechapornkul P, et al. *Plasmodium falciparum* *pfmdr1* amplification, mefloquine resistance, and parasite fitness. *Antimicrobial agents and chemotherapy*. 2009; 53:1509–15. [PubMed: 19164150]
25. Plowe CV, et al. Mutations in *Plasmodium falciparum* dihydrofolate reductase and dihydropteroate synthase and epidemiologic patterns of pyrimethamine-sulfadoxine use and resistance. *The Journal of Infectious Diseases*. 1997; 176:1590–1596. [PubMed: 9395372]
26. Vinayak S, et al. Origin and evolution of sulfadoxine resistant *Plasmodium falciparum*. *PLoS pathogens*. 2010; 6:e1000830. [PubMed: 20360965]
27. Kublin JG, et al. Molecular assays for surveillance of antifolate-resistant malaria. *Lancet*. 1998; 351:1629–30. [PubMed: 9620719]
28. Triglia T, Wang P, Sims PF, Hyde JE, Cowman AF. Allelic exchange at the endogenous genomic locus in *Plasmodium falciparum* proves the role of dihydropteroate synthase in sulfadoxine-resistant malaria. *The EMBO Journal*. 1998; 17:3807–3815. [PubMed: 9669998]
29. Peterson DS, Walliker D, Wellem TE. Evidence that a point mutation in dihydrofolate reductase-thymidylate synthase confers resistance to pyrimethamine in falciparum malaria. *Proceedings of the National Academy of Sciences of the United States of America*. 1988; 85:9114–8. [PubMed: 2904149]

30. Plowe CV. The evolution of drug-resistant malaria. *Transactions of the Royal Society of Tropical Medicine and Hygiene*. 2009; 103(Suppl 1):S11–4. [PubMed: 19084883]
31. Rathod PK, McErlean T, Lee PC. Variations in frequencies of drug resistance in *Plasmodium falciparum*. *Proceedings of the National Academy of Sciences of the United States of America*. 1997; 94:9389–93. [PubMed: 9256492]
32. Jiricny J. Replication errors: cha(lle)nging the genome. *The EMBO Journal*. 1998; 17:6427–36. [PubMed: 9822589]
33. Shankar J, Tuteja R. UvrD helicase of *Plasmodium falciparum*. *Gene*. 2008; 410:223–33. [PubMed: 18242886]
34. Hall MC, Jordan JR, Matson SW. Evidence for a physical interaction between the *Escherichia coli* methyl-directed mismatch repair proteins MutL and UvrD. *The EMBO Journal*. 1998; 17:1535–41. [PubMed: 9482750]
35. Mechanic LE, Frankel BA, Matson SW. *Escherichia coli* MutL loads DNA helicase II onto DNA. *The Journal of biological chemistry*. 2000; 275:38337–46. [PubMed: 10984488]
36. Anderson TJ, et al. Microsatellite markers reveal a spectrum of population structures in the malaria parasite *Plasmodium falciparum*. *Mol Biol Evol*. 2000; 17:1467–82. [PubMed: 11018154]
37. Dye C, Williams BG. Multigenic drug resistance among inbred malaria parasites. *Proc Biol Sci*. 1997; 264:61–7. [PubMed: 9061961]
38. Sinka ME, et al. The dominant *Anopheles* vectors of human malaria in the Asia-Pacific region: occurrence data, distribution maps and bionomic precis. *Parasites & vectors*. 2011; 4:89. [PubMed: 21612587]
39. Verdrager J. Epidemiology of the emergence and spread of drug-resistant *falciparum* malaria in South-East Asia and Australasia. *The Journal of tropical medicine and hygiene*. 1986; 89:277–89. [PubMed: 3543384]
40. Payne D. Did medicated salt hasten the spread of chloroquine resistance in *Plasmodium falciparum*? *Parasitology today*. 1988; 4:112–5. [PubMed: 15463062]
41. World Health Organisation. Global plan for artemisinin resistance containment. Vol. 87. World Health Organisation; Geneva: 2011.
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*. 2011; 6:e19283. [PubMed: 21603629]
43. Aurrecochea C, et al. PlasmoDB: a functional genomic database for malaria parasites. *Nucleic Acids Res*. 2009; 37:D539–43. [PubMed: 18957442]
44. Lawson DJ, Hellenthal G, Myers S, Falush D. Inference of population structure using dense haplotype data. *PLoS genetics*. 2012; 8:e1002453. [PubMed: 22291602]
45. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*. 2009; 19:1655–64. [PubMed: 19648217]



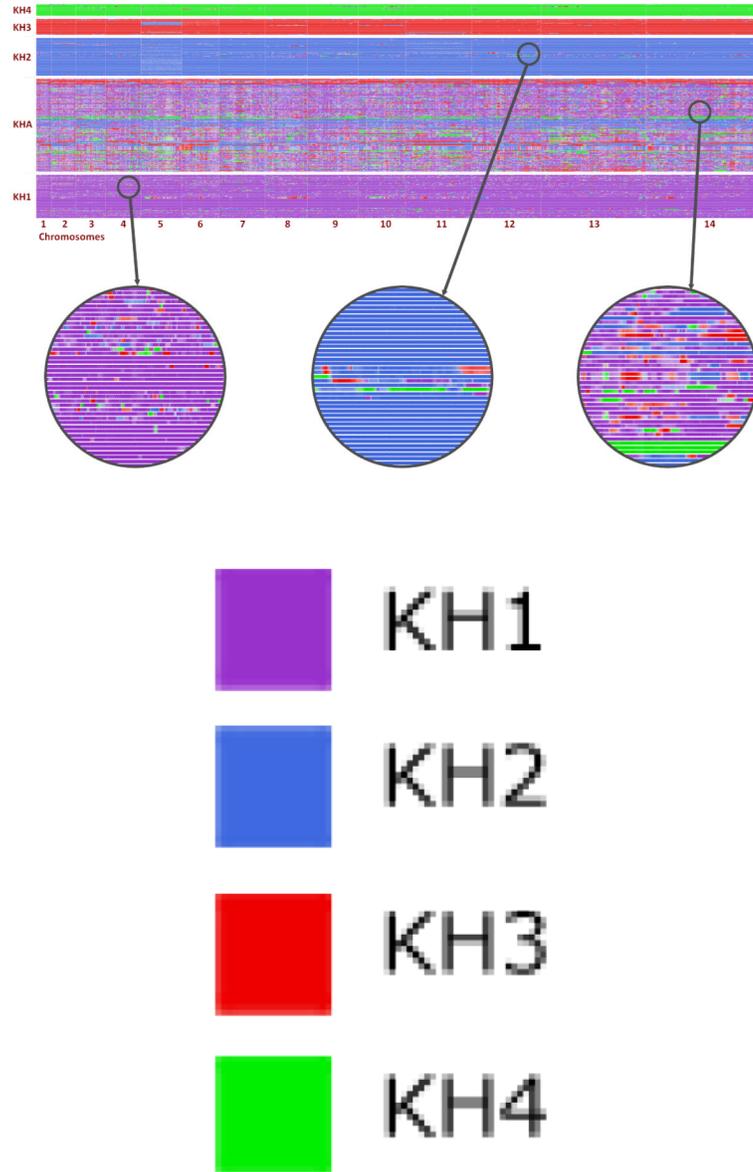




- ▲ Thailand
- ▲ Vietnam
- NE Cambodia
- W Cambodia

**Figure 1. Population Structure in the samples set analysed**

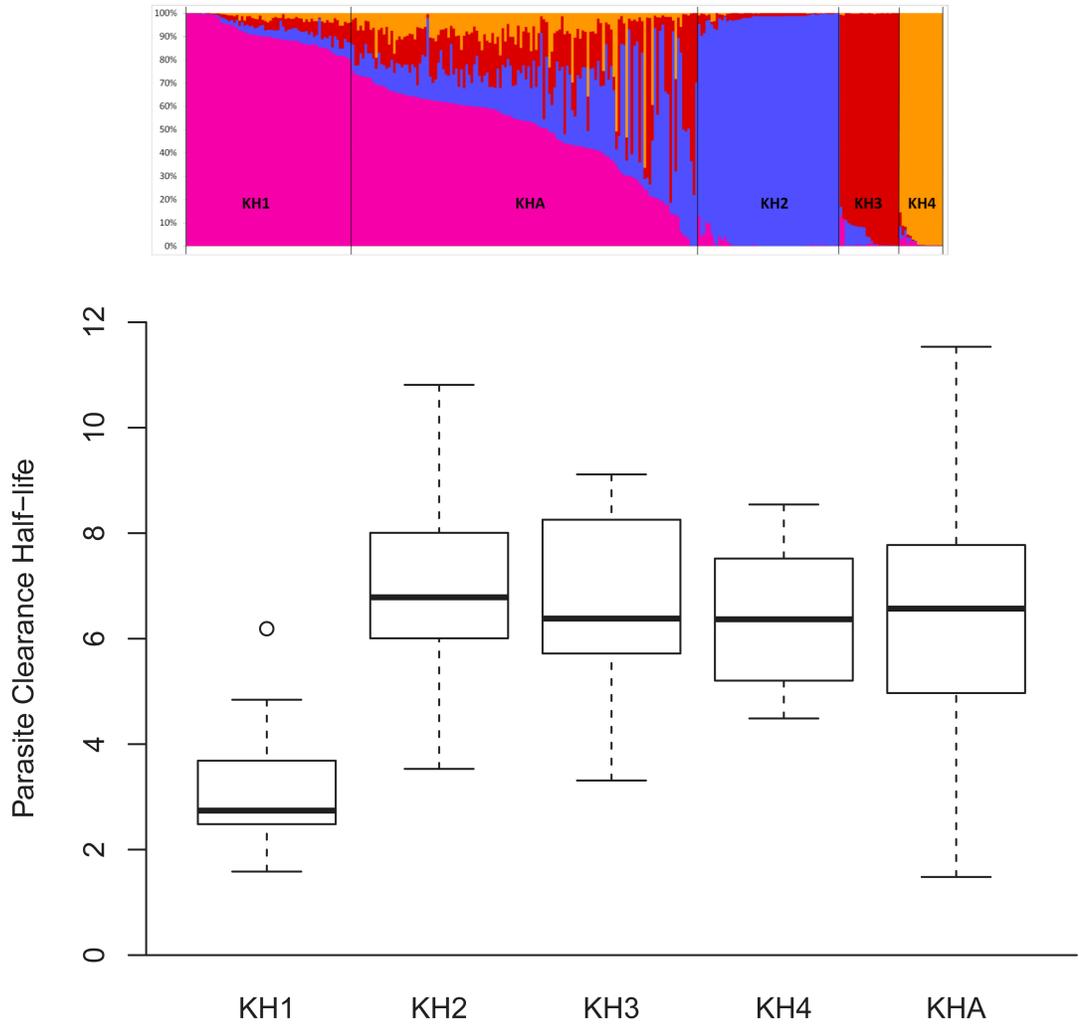
(a) Neighbour-joining (NJ) tree of all 825 samples, based on a pairwise distance matrix. The West African (WAF) samples are very clearly separated from South-East Asian (SEA) one. Within SEA, we observe that Thailand samples form a distinct cluster from Cambodian samples. Even within Cambodia, we see clear separation between a NE Cambodia/Vietnam group and a western Cambodian cluster. Samples in western Cambodia form a common clade, in spite of discernible population structure. (b) Three-dimensional plot of Principal components analysis (PCA) using all 825 samples, showing the first three components (PC1, PC2 and PC3). PC1 separates between continental clusters in Africa and Asia, while PC2 and PC3 largely account for the variability in western Cambodian samples (blue circles). (c, d) PCA plots using only SEA samples, showing (c) PC1 vs. PC2, and (d) PC2 vs. PC3. Once African samples are removed, all first three components are driven by variability in western Cambodia. Accordingly, we identified three “outlier” clusters in Cambodia (labelled KH2, KH3 and KH4), whose initial composition was determined by dividing the plot into quadrants. A “core” group (labelled KH1) is composed of north-eastern and western Cambodian samples, which cluster close to other SEA countries. Between the core and outlier group, several samples (labelled KHA) have intermediate PC values, suggesting admixture.

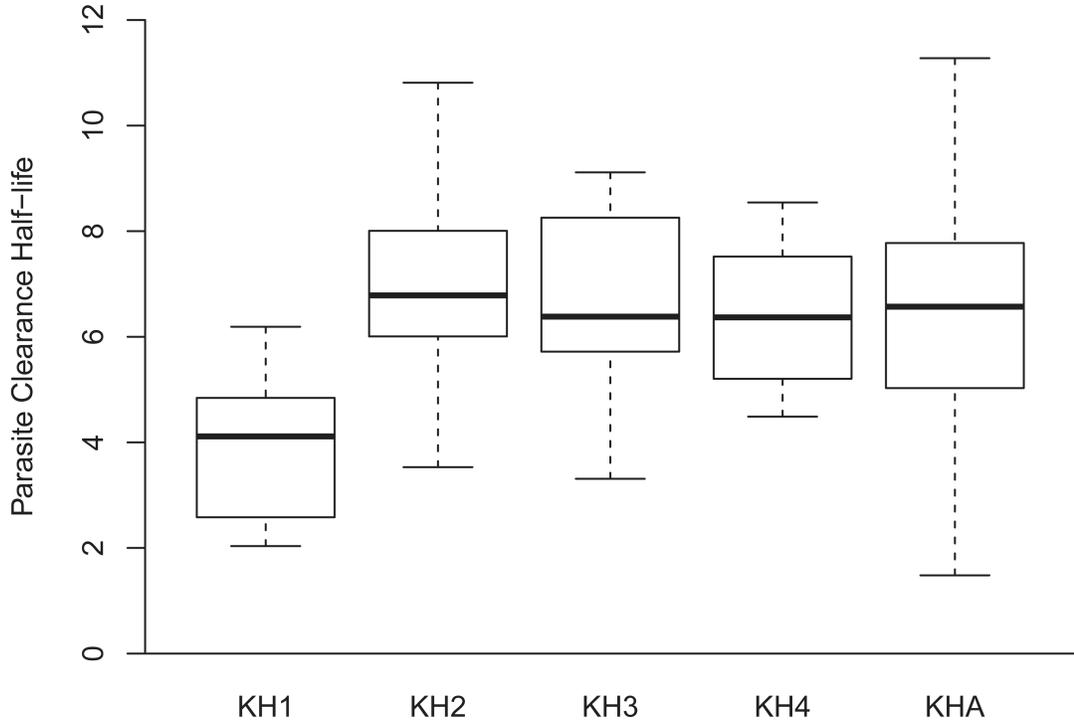


**Figure 2. Chromosome Painting**

A high-level view of chromosome painting across all 14 chromosomes is shown here. Each line represents a sample; samples are organized by subpopulation, in the order KH1, KHA, KH2, KH3 and KH4, starting from the bottom. The colours of segments indicate the cluster found to be the most probable donor of each segment (purple=KH1, blue =KH2, red=KH3, green =KH4). Callouts (a-c) show details of some important features of the painting plots. (a) KH1 samples contain many short segments of DNA sequence labelled as KH2/3/4, consistent with KH1 representing the ancestral population from which KH2/3/4 are derived. (b) KH2/3/4 samples comprise very long segments labelled as belonging to other populations, suggesting recent admixture. (c) KHA samples contain both short and long

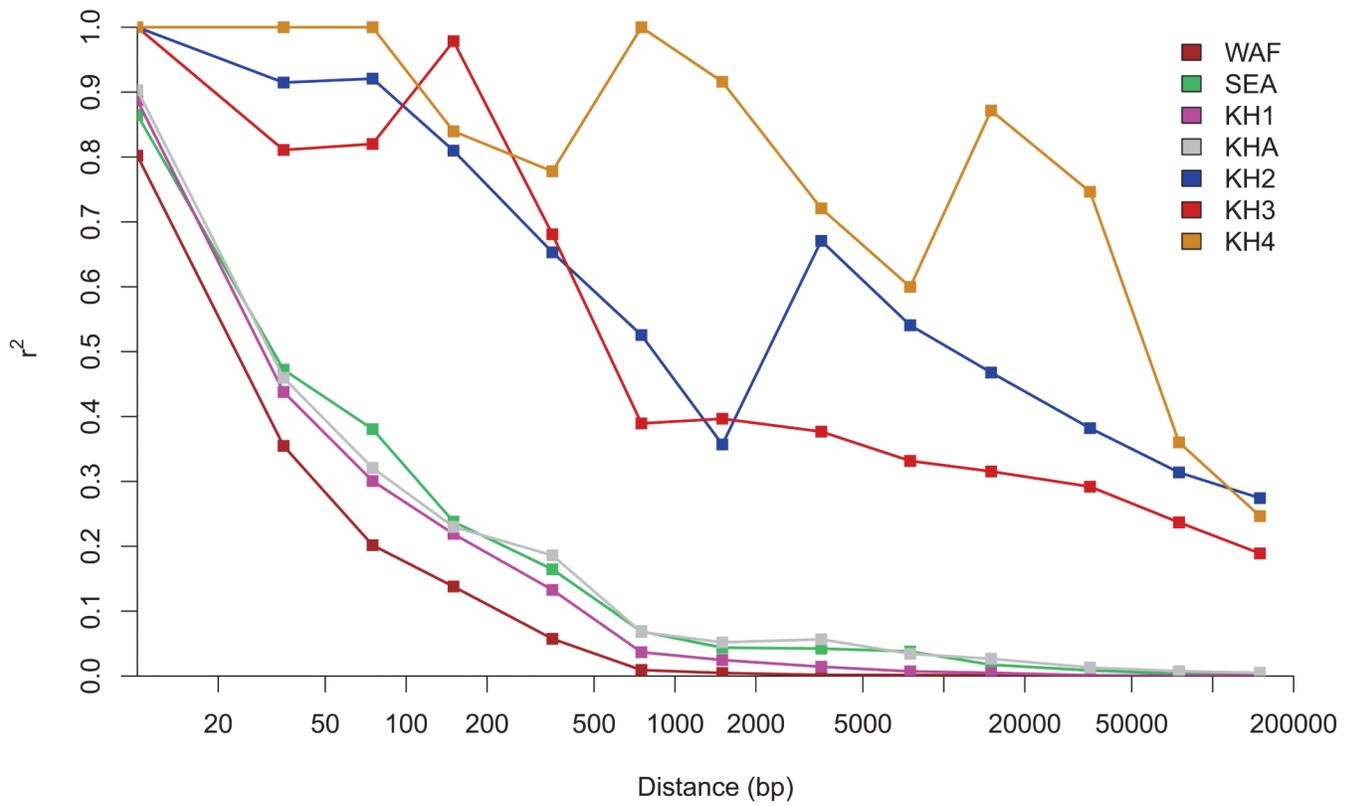
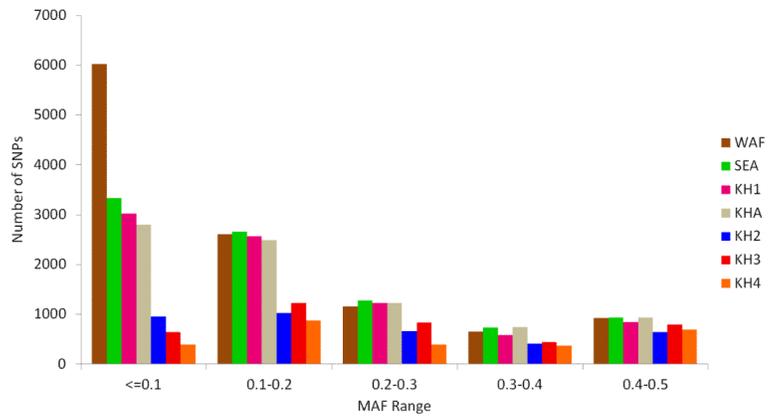
chunks of DNA labelled as KH2/3/4, consistent with the hypothesis that it is an admixed population, with contribution from these ART-R subpopulations.

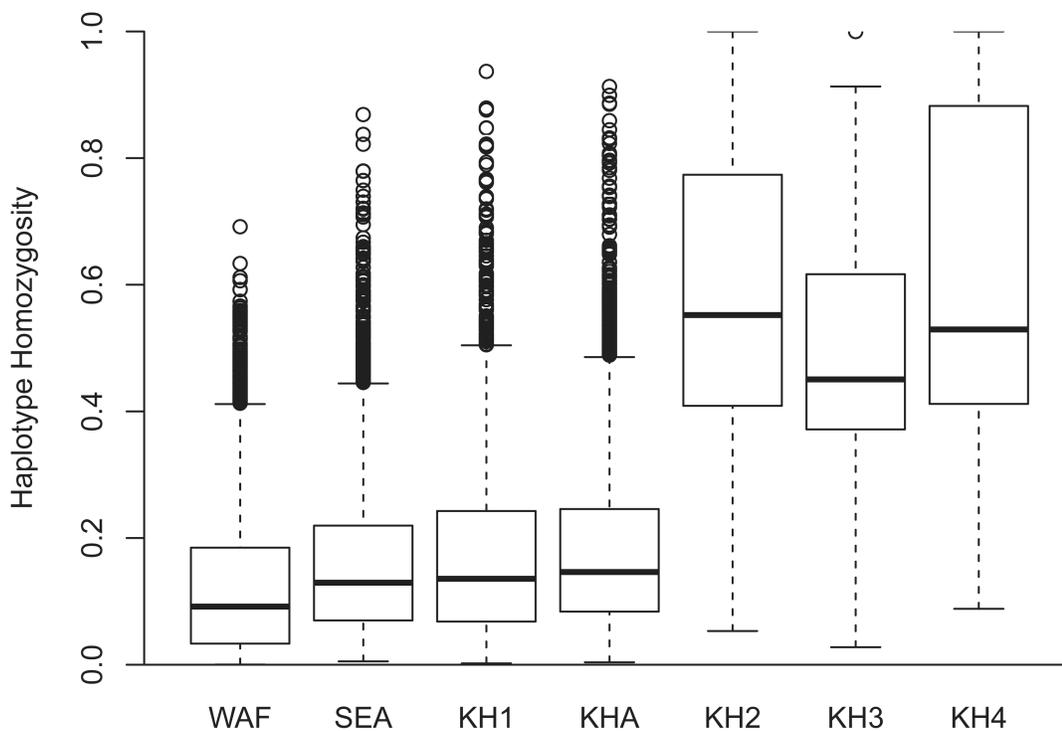




**Figure 3. Genetic and Phenotypic differentiation between Cambodian subpopulations**

(a) Ancestry analysis of the 293 Cambodian samples, based on chromosome painting analysis of all Cambodian samples (see Methods). In this figure, each vertical bar represents a single samples, and the bar is coloured 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 analysis (KH1=purple, KH2=blue, KH3=red, KH4=green). 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 labelled as belonging to one of the resulting four subpopulations was assigned to the KHA group. The final partitioning of samples into subpopulations is shown in the figure. Most of the samples classified as KHA appear to be composed of a mixture of KH1 and one or more of the remaining clusters, consistent with the hypothesis of a recombinant population. (b) Boxplot showing the distribution of artemisinin response phenotypes (parasite clearance half-life) in the Cambodian parasite subpopulations thus identified (n=212). Each box represents the interquartile range of values, split at the median; whiskers extend to the furthest point that is within 1.5 times the length of the box. (b) An analogous plot derived from an analysis of Western Cambodian parasites only (i.e. excluding the Ratanakiri site; n=166). A table of parasite clearance half-life with p-values is given in Supplementary Table 1.





**Figure 4. Evidence for multiple founder effects in Cambodian subpopulations**

(a) Minor allele frequencies (MAF) spectra for 7 populations. For each population, SNPs with  $MAF > 0$  were assigned to one of 5 equally-sized MAF bins, and we show counts in these bins. A sub-sampling procedure, averaged over 1000 iterations, was used to minimize confounding by differences in sample size. MAF frequency spectra of KH1 and KHA are consistent with the rest of SEA, with a high proportion of low-MAF SNPs. In contrast, the ART-R subpopulations (KH2, KH3 and KH4) have considerably fewer SNPs at low frequency, and more homogeneous frequency spectra. (b) Genome-wide decay of Linkage Disequilibrium (LD), expressed using a classic LD measure ( $r^2$ ). SNPs with  $MAF > 0.15$  were organized into equal-size allele frequency bins. Each bin was analyzed independently and corrected for effects of sample size and population structure, before bin results are aggregated (see Supplementary Note). African populations exhibit rapid LD decay, with no detectable LD within 1 kbp. In Cambodian ART-R subpopulations, LD decays over much greater distances than in other groups (including KH1 and KHA), suggesting highly reduced haplotype diversity. (c) Genome-wide distribution of haplotype homozygosity in different populations. Each box represents the interquartile range of values, split at the median; whiskers extend to the furthest point that is within 1.5 times the length of the box. A sliding window of width 201 SNP was used to construct local haplotypes for each sample. While “core” populations of South-East Asia (SEA and KH1) have a level of haplotype diversity consistent with that seen in Africa (WAF), the ART-R subpopulations (KH2, KH3 and KH4) exhibit much higher homozygosity, indicating reduced overall haplotype diversity.

**Table 1**

Geographical origin of the 825 samples in the dataset.

Region	Country Code	Country	Location	Sample size
West Africa	BF	Burkina Faso	Bobo-Dioulasso	48
	GM	Gambia	Banjul	67
	GH	Ghana	Navrongo	243
	ML	Mali	Bamako	56
South-East Asia	TH	Thailand	Mae Sot	106
	VN	Vietnam	Binh Phuoc Province	12
	KH	Cambodia (east)	Ratanakiri	50
			Pailin	49
		Cambodia (west)	Tasanh	47
Pursat			147	

**Table 2**  
**Distribution of the Cambodian clusters samples (as described in the text) by geographical locations**

Counts of samples for which a clinical artemisinin phenotype was available are shown in brackets after the counts of all samples.

Location	KH1	KH2	KH3	KH4	KHA	Total
<b>Pailin</b>	4 (1)	14 (5)	5 (3)		26 (21)	<b>49 (30)</b>
<b>Tasanh</b>	1 (1)	9 (9)	3 (3)		34 (34)	<b>47 (47)</b>
<b>Pursat</b>	10 (4)	32 (21)	15 (12)	17 (12)	73 (40)	<b>147 (89)</b>
<b>Ratanakiri</b>	48 (44)				2 (2)	<b>50 (46)</b>
<b>Total</b>	<b>63 (50)</b>	<b>55 (35)</b>	<b>23 (18)</b>	<b>17 (12)</b>	<b>135 (97)</b>	<b>293 (212)</b>

**Table 3**  
**Artemisinin response characteristics of Cambodian populations**

The columns indicate the number of samples (N), the median parasite clearance half-life (“Half-life”), and the p-value (P) of a Mann-Whitney test comparing the half-life distribution for the subpopulation against that of KH1. The 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 size in Western Cambodia, clearance half-lives are significantly longer in the KH2-KH4 subpopulation and in the KHA recombinant population.

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