Genetic markers associated with dihydroartemisinin–piperaquine failure in *Plasmodium falciparum* malaria in Cambodia: a genotype–phenotype association study

Roberto Amato, Pharath Lim, Olivo Miotto, Chanaki Amarasinghe, Dalin Dek, Richard D Pearson, Jacob Almagro-Garcia, Aaron T Neal, Sokunthea Sreng, Seila Suon, Eleanor Drury, Dushyanth Jyothi, Jim Stalker, Dominic P Kwiatkowski, Rick M Fairhurst

**Summary**

**Background** As the prevalence of artemisinin-resistant *Plasmodium falciparum* malaria increases in the Greater Mekong subregion, emerging resistance to partner drugs in artemisinin combination therapies seriously threatens global efforts to treat and eliminate this disease. Molecular markers that predict failure of artemisinin combination therapy are urgently needed to monitor the spread of partner drug resistance, and to recommend alternative treatments in southeast Asia and beyond.

**Methods** We did a genome-wide association study of 297 *P falciparum* isolates from Cambodia to investigate the relationship of 11630 exonic single-nucleotide polymorphisms (SNPs) and 43 copy number variations (CNVs) with in-vitro piperaquine 50% inhibitory concentrations (IC₅₀) s, and tested whether these genetic variants are markers of treatment failure with dihydroartemisinin–piperaquine. We then did a survival analysis of 133 patients to determine whether candidate molecular markers predicted parasite recrudescence following dihydroartemisinin–piperaquine treatment.

**Findings** Piperaquine IC₅₀ s increased significantly from 2011 to 2013 in three Cambodian provinces (2011 vs 2013 median IC₅₀ s: 20·0 nmol/L [IQR 13·7–29·0] vs 39·2 nmol/L [32·8–48·1] for Ratanakiri, 19·3 nmol/L [15·1–26·2] vs 66·2 nmol/L [49·9–83·0] for Preah Vihear, and 19·6 nmol/L [11·9–33·9] vs 81·1 nmol/L [61·3–113·1] for Pursat; all p≤10⁻³; Kruskal-Wallis test). Genome-wide analysis of SNPs identified a chromosome 13 region that associates with raised piperaquine IC₅₀ s. A non-synonymous SNP (encoding a Glu415Gly substitution) in this region, within a gene encoding an exonuclease, associates with parasite recrudescence following dihydroartemisinin–piperaquine treatment. Genome-wide analysis of CNVs revealed that a single copy of the mdr1 gene on chromosome 5 and a novel amplification of the plasmspsin 2 and plasmspsin 3 genes on chromosome 14 also associate with raised piperaquine IC₅₀ s. After adjusting for covariates, both exo-E415G and plasmspsin 2–3 markers significantly associate (p=3·0×10⁻⁴ and p=1·7×10⁻⁷, respectively) with decreased treatment efficacy (survival rates 0·38 [95% CI 0·25–0·51] and 0·41 [0·28–0·53], respectively).

**Interpretation** The exo-E415G SNP and plasmspsin 2–3 amplification are markers of piperaquine resistance and dihydroartemisinin–piperaquine failures in Cambodia, and can help monitor the spread of these phenotypes into other countries of the Greater Mekong subregion, and elucidate the mechanism of piperaquine resistance. Since plasmspsins are involved in the parasite’s haemoglobin-to-haemozoin conversion pathway, targeted by related antimalarials, plasmspsin 2–3 amplification probably mediates piperaquine resistance.

**Funding** Intramural Research Program of the US National Institute of Allergy and Infectious Diseases, National Institutes of Health, Wellcome Trust, Bill & Melinda Gates Foundation, Medical Research Council, and UK Department for International Development.

**Copyright © The Author(s).** Published by Elsevier Ltd. This is an Open Access article under the CC BY license.

**Introduction**

Artemisinin combination therapy, the use of a short-acting artemisinin derivative and a long-acting partner drug, is recommended worldwide for the treatment of *Plasmodium falciparum* malaria. Treatment with dihydroartemisinin–piperaquine, a current front-line artemisinin combination therapy in Cambodia, Vietnam, Thailand, Myanmar, China, and Indonesia is now failing in Cambodian provinces where artemisinin resistance has emerged and spread. This situation probably arose because the survival of parasites after artemisinin exposure increases the chance that they develop spontaneous genetic resistance to piperaquine, survive declining piperaquine plasma concentrations, and propagate via mosquitoes to other human beings. Through such processes, artemisinin resistance, which has evolved across the Greater Mekong subregion, threatens to compromise the efficacy of dihydroartemisinin–piperaquine and other artemisinin combination therapies in the global treatment and elimination of malaria. Treatment with dihydroartemisinin–piperaquine has been associated with higher treatment failure rates compared with other artemisinin combination treatments in southeast Asia and beyond. Genetic markers associated with dihydroartemisinin–piperaquine failures in *Plasmodium falciparum* in Cambodia and other regions are urgently needed to monitor the spread of resistance, and to recommend alternative treatments.
of malaria. The natural selection of low-frequency, pre-existing resistance to piperquine might also be occurring and further contributing to this problem.

Recent increases in treatment failures with dihydroartemisinin–piperquine and piperquine 50% inhibitory concentrations (IC\textsubscript{50}s) suggest that piperquine resistance has emerged in Cambodia. These findings, and the discovery that piperquine-resistant parasites are sensitive to the former artemisinin combination therapy partner drug mefloquine, have recently led Cambodia’s national malaria control programme and WHO to recommend artesunate–mefloquine as the first-line artemisinin combination therapy in ten Cambodian provinces, including Pursat and Preah Vihear. Molecular markers are urgently needed for large-scale surveillance programmes to predict treatment failures with dihydroartemisinin–piperquine in Cambodia and other countries in the Greater Mekong subregion, and investigate the molecular mechanism of piperquine resistance.

Several genetic variations have been associated with decreased piperquine susceptibility: a single copy of the \textit{mdr1} gene has been associated with dihydroartemisinin–piperquine treatment failures in Cambodian patients, whereas amplification of a region downstream of \textit{mdr1} and the \textit{crt} single-nucleotide polymorphism (SNP) C101F has been associated with raised piperquine IC\textsubscript{50}s in vitro. These genetic variants are not useful as molecular markers, however, because the first is wild type, the second is very uncommon, and the third has not yet been observed in MalariaGEN \textit{P. falciparum} Community Project’s global catalogue of variation in clinical samples.
dihydroartemisinin–piperaquine has not been used. Therefore, we sequenced the genomes and measured the piperaquine IC₅₀s of parasites collected in 2010–13 from patients with *P. falciparum* malaria in Pursat, Preah Vihear, and Ratanakiri, where the prevalences of *kelch13* mutations, genetic markers for artemisinin resistance, were 77%, 34%, and 11%, and where the prevalences of dihydroartemisinin–piperaquine treatment failures were 46%, 16%, and 2%, respectively, in 2012–13.

**Methods**

**Study design and participants**

To obtain samples for this GWAS study, we enrolled patients with uncomplicated *P. falciparum* malaria into parasite clearance rate and drug efficacy studies in 2010–13 in three provinces where piperaquine resistance is common (Pursat), emerging (Preah Vihear), or uncommon (Ratanakiri). Written informed consent was given by adult patients, or a parent or guardian of child patients. Protocols were approved by the Cambodian National Ethics Committee for Health Research and the National Institute of Allergy and Infectious Diseases Institutional Review Board and are registered with ClinicalTrials.gov, numbers NCT00341003, NCT01240603, and NCT01736319.

Using parasitised blood samples from these studies, we measured piperaquine IC₅₀s ex vivo or after short-term culture in vitro, and obtained whole-genome parasite sequence data, whenever possible. The GWAS was designed to identify genetic markers of raised piperaquine IC₅₀₅₀s, whereas a dihydroartemisinin–piperaquine efficacy study was used to test for association between GWAS candidate markers and parasite recrudescence. PCR genotyping using *msp1*, *msp2*, and *glurp* microsatellites as genetic markers distinguished recrudescences from newly acquired infections. A summary of samples, according to province of origin and year of collection, is shown in the appendix.

**GWAS and survival analyses**

The preparation, sequencing, genotyping, and phenotyping of samples are described in the appendix. The GWAS and correction for population structure were done using a linear mixed model, implemented in FaST-LMM version 2.06. We tested 11630 SNPs with minor allele frequency greater than 0·033, using genotypes encoded as the number of non-reference alleles (0 or 1), and excluding heterozygous calls to minimise confounding effects of mixed infections. Piperaquine IC₅₀ was used as the continuous dependent variable. When whole-genome sequence or phenotype data were available for both initial and recrudescent samples, we only analysed data from the recrudescent samples. Different approaches (using initial sample data only, using the phenotype and genotype of the same sample, averaging initial and recrudescent IC₅₀s) did not alter results. A relationship matrix was calculated using a subset of 6678 SNPs with minor allele frequency greater than 3% (options: –maf 0·03), missing data rate below 25% (options: –geno 0·25), and unlinked (in windows of 100 SNPs, shifted forward by ten SNPs each time, removing one from each pair of SNPs with linkage disequilibrium >0·3; options: –indep-pairwise 100 10 0·3) and extracted using PLINK version 1.02. In estimating the relationship matrix, we found that excluding proximal SNPs (within 10 kb or 100 kb from the tested variant) did not substantially affect results. Given the number of independent SNPs used, we applied Bonferroni correction to define a significance threshold of p of 8·6×10⁻⁷ or less for GWAS analyses. We also defined a suggestive threshold of p of 10⁻⁴ or less to identify relatively high-ranking loci.

To adjust for potential confounder effects, we treated the geographical origin of the sample, the presence of *mdr1* amplification, and the presence of *kelch13* resistance alleles collectively as covariates in the linear mixed model. These covariates reduced the genomic inflation factor λₚ to 1·013, suggesting that residual inflation is due to extended homozygosity haplotypes in the samples and thus has little effect on the GWAS.

We did a survival analysis using the R package survival. We fitted a Cox proportional hazard regression model, in which treatment success (recrudescent vs non-recrudescent infection) represented the survival status, and we added the age of the patient (in years), parasitaemia on day 0 (log-scaled), and dose of piperaquine given (mg/kg, in five-unit increments starting at 35 mg/kg) as covariates. We then included the two markers (*exo-E415G* and *plasmodin* 2–3) as covariates, and estimated adjusted hazard ratios (aHRs) and adjusted survival curves.

**Copy number variations**

We subsequently tested the association of copy number variations (CNVs) with piperaquine IC₅₀₅₀s using the same method and parameters just described. We tested 43 CNVs present in at least five samples; genotypes represented the presence or absence (encoded as 0 or 1) of the CNV. In this analysis, *mdr1* amplification was not included as a covariate in the linear mixed model. To call CNVs across the genome, we modified a procedure used previously. Briefly, we first divided the genome into 300 bp non-overlapping bins and calculated for each sample the number of reads whose alignment started within each bin. We then normalised these binned read counts by dividing by the median read count of the core regions of chromosome 9. We excluded bins where guanine–cytosine (GC) content was less than 20% due to coverage bias in most samples. Copy number state for each bin was predicted in each sample by fitting a
Gaussian hidden Markov model to the normalised coverage data.

After determining the most likely state for each window (Viterbi algorithm), in each sample independently we merged adjacent windows having the same state prediction and allowed a maximum gap of 5000 bp between them. We then further removed windows less than 700 bp or not supported by face-away reads, and data for 43 samples with excessive variation in read coverage. This led to the identification of a median six CNVs per sample. Finally, we merged all the CNV calls across the samples, by considering any partial overlap. In total, we identified 134 regions containing CNVs across the core genome; 43 of these CNVs were present in at least five samples and were tested for association with the phenotype, as described above. For the 133 samples in the five samples and were tested for association with the phenotype, as described above. For the 133 samples in the study and final responsibility for the decision to submit for publication.

Role of the funding source
The funders had no role in study design, data collection, analysis, interpretation, or report writing. The corresponding authors had full access to all data in the study and final responsibility for the decision to submit for publication.

Results
We analysed 486 *P falciparum* clinical isolates collected between July 9, 2010, and Dec 31, 2013, in three

Cambodian provinces where artemisinin resistance is entrenched (Pursat), emerging (Preah Vihear), and uncommon (Ratanakiri; appendix). To monitor the evolution of piperaquine resistance, we measured piperaquine IC₅₀s for 297 *P. falciparum* clinical isolates

<table>
<thead>
<tr>
<th>Locus</th>
<th>Chromosome number</th>
<th>Position</th>
<th>Gene ID</th>
<th>Gene description</th>
<th>N or S</th>
<th>Alteration</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Locus L13</td>
<td>13</td>
<td>2504550</td>
<td>PF3D7_1362500</td>
<td>Exonuclease, putative</td>
<td>N</td>
<td>p.Glu145Gly</td>
<td>2.69 × 10⁻⁸</td>
</tr>
<tr>
<td>Locus L12-01</td>
<td>12</td>
<td>418246</td>
<td>PF3D7_1208900</td>
<td>Conserved plasmid protein, unknown function</td>
<td>S</td>
<td>p.Glu137Pro</td>
<td>4.80 × 10⁻⁷</td>
</tr>
<tr>
<td>Locus L10-02</td>
<td>10</td>
<td>2728402</td>
<td>PF3D7_1068700</td>
<td>Mitochondrial carrier protein, putative</td>
<td>N</td>
<td>p.Asn252Asp</td>
<td>2.33 × 10⁻⁶</td>
</tr>
<tr>
<td>Locus L3</td>
<td>3</td>
<td>2512415</td>
<td>PF3D7_0362700</td>
<td>Conserved plasmid protein, unknown function</td>
<td>S</td>
<td>p.Gly707Asn</td>
<td>6.25 × 10⁻⁴</td>
</tr>
<tr>
<td>Locus L13</td>
<td>13</td>
<td>2519091</td>
<td>PF3D7_1362800</td>
<td>Conserved plasmid protein, unknown function</td>
<td>N</td>
<td>p.Gly202Asp</td>
<td>2.12 × 10⁻⁴</td>
</tr>
<tr>
<td>Locus L13</td>
<td>13</td>
<td>2447146</td>
<td>PF3D7_1361000</td>
<td>Arginine methyltransferase S, putative (PRMT5)</td>
<td>N</td>
<td>p.Arg40Ser</td>
<td>3.96 × 10⁻⁶</td>
</tr>
<tr>
<td>Locus L04</td>
<td>4</td>
<td>904088</td>
<td>PF3D7_0420000</td>
<td>Zinc finger protein, putative</td>
<td>S</td>
<td>p.Ile213Leu</td>
<td>7.67 × 10⁻⁷</td>
</tr>
<tr>
<td>Locus L14</td>
<td>14</td>
<td>2411942</td>
<td>PF3D7_1468700</td>
<td>Conserved plasmid protein, unknown function</td>
<td>N</td>
<td>p.Arg251Lys</td>
<td>1.27 × 10⁻⁴</td>
</tr>
<tr>
<td>Locus L13</td>
<td>13</td>
<td>2395762</td>
<td>PF3D7_1368300</td>
<td>Conserved plasmid protein, unknown function</td>
<td>N</td>
<td>p.Ile201Phe</td>
<td>2.03 × 10⁻⁴</td>
</tr>
<tr>
<td>Locus L13</td>
<td>13</td>
<td>2542366</td>
<td>PF3D7_1363300</td>
<td>Mitochondrial ribosomal protein L9 precursor, putative</td>
<td>N</td>
<td>p.Thr281Ile</td>
<td>2.48 × 10⁻⁴</td>
</tr>
<tr>
<td>Locus L12-02</td>
<td>12</td>
<td>1787279</td>
<td>PF3D7_1242000</td>
<td>Conserved plasmid protein, unknown function</td>
<td>N</td>
<td>p.Ser516Asn</td>
<td>2.90 × 10⁻⁴</td>
</tr>
<tr>
<td>Locus L04</td>
<td>4</td>
<td>866807</td>
<td>PF3D7_0419400</td>
<td>Conserved plasmid protein, unknown function</td>
<td>S</td>
<td>p.Val23Ile</td>
<td>3.05 × 10⁻²</td>
</tr>
<tr>
<td>Locus L06</td>
<td>6</td>
<td>866478</td>
<td>PF3D7_0621100</td>
<td>Conserved plasmid protein, unknown function</td>
<td>N</td>
<td>p.Asp1144Glu</td>
<td>4.04 × 10⁻⁴</td>
</tr>
<tr>
<td>Locus L12-01</td>
<td>12</td>
<td>407813</td>
<td>PF3D7_1208800</td>
<td>Zinc finger protein, putative</td>
<td>N</td>
<td>p.Ser439Asn</td>
<td>4.56 × 10⁻⁴</td>
</tr>
<tr>
<td>Locus L13</td>
<td>13</td>
<td>2656538</td>
<td>PF3D7_1366400</td>
<td>Rho protein (Rhop148)</td>
<td>N</td>
<td>p.Lys363Asn</td>
<td>5.02 × 10⁻³</td>
</tr>
<tr>
<td>Locus L14</td>
<td>14</td>
<td>2385590</td>
<td>PF3D7_1467900</td>
<td>Conserved plasmid protein, unknown function</td>
<td>S</td>
<td>p.Ser23Thr</td>
<td>8.00 × 10⁻⁴</td>
</tr>
<tr>
<td>Locus L04</td>
<td>4</td>
<td>865709</td>
<td>PF3D7_0419400</td>
<td>Conserved plasmid protein, unknown function</td>
<td>N</td>
<td>p.Leu323Ile</td>
<td>8.96 × 10⁻³</td>
</tr>
<tr>
<td>Locus L13</td>
<td>13</td>
<td>2746936</td>
<td>PF3D7_1369100</td>
<td>Conserved plasmid protein, unknown function</td>
<td>N</td>
<td>p.Leu430Arg</td>
<td>8.98 × 10⁻³</td>
</tr>
</tbody>
</table>

Significant SNPs (Bonferroni-corrected p = 6 × 10⁻⁷) and suggestive SNPs (p = 10⁻⁴) associated with piperaquine IC₅₀, according to increasing p value, are shown. For each SNP, we list the locus name, chromosome number, nucleotide position, gene ID, gene description, whether the SNP is non-synonymous (N) or synonymous (S), encoded amino acid alteration, and association p value.

Table 1: Genome-wide single-nucleotide polymorphisms (SNPs) most strongly associated with piperaquine 50% inhibitory concentrations (IC₅₀s)
obtained directly from patients (ex vivo, 275 [93%] of 297) or following short-term culture (in vitro, 22 [7%] of 297) in 2011, 2012, and 2013. In all three provinces, IC₅₀s increased significantly over time, especially when comparing province-stratified data from 2011 and 2013 (Kruskal-Wallis test, all p≤10⁻³; appendix). Despite parasites having comparable IC₅₀s at all sites in 2011 (medians 20·0 nmol/L [IQR 13·7–29·0], 19·3 nmol/L [15·1–26·2], and 19·6 nmol/L [11·9–33·9] for Ratanakiri, Preah Vihear, and Pursat, respectively), they were remarkably differentiated at all sites by 2013, showing about two, three, and four times increases in IC₅₀s (medians 39·2 nmol/L [IQR 32·8–48·1], 66·2 nmol/L [49·9–83·0], and 81·1 nmol/L [61·3–113·1], respectively). These regional differences are consistent with the relative prevalences of artemisinin-resistant parasites and
Articles

To investigate the genetic basis of piperquine resistance in vitro, we did a GWAS analysis of the 297 samples (appendix) for which we had both piperquine IC₅₀s and whole-genome sequences (see Methods). We tested 11,630 SNPs that were well covered in this set of samples, and where one of the two alleles was present in at least ten samples. IC₅₀s were treated as a continuous dependent variable in a linear mixed-model algorithm, and we treated the province of sample origin, presence of mdr1 amplification, and presence of kelch13 mutations as covariates; to correct for the confounding effect of population structure, we treated genetic similarity across samples as a random effect (appendix).

GWAS analysis identified one major locus on chromosome 13, containing two non-synonymous SNPs, that were significantly associated with raised IC₅₀s (p≤2.3×10⁻⁴; figure 2A, appendix). The strongest signal (2.55×10⁻⁶) is from a non-synonymous SNP (referred to as exo-E415G) producing a Asn252Asp substitution in a putative exonuclease. The second strongest signal (2.27×10⁻⁶) with low IC₅₀s. However, the most significant association (p=7.6×10⁻⁷) was found for a novel amplification on chromosome 14, encompassing two of the ten plasmspin genes in the P falciparum genome—plasmspin 2 and plasmspin 3—that encode aspartic proteases involved in the haemoglobin-to-haemozoin conversion pathway targeted by quinolines.

A detailed analysis of sequencing reads aligned within the plasmspin 2–3 locus revealed that the amplification boundaries were identical in all samples. The putative breakpoints lie near the 3’ end of both plasmspin 1 and plasmspin 3, so that each amplification creates an intact extra copy of plasmspin 2 together with a new chimeric

<table>
<thead>
<tr>
<th>plasmspin 2–3 CN1</th>
<th>plasmspin 2–3 CN2+</th>
</tr>
</thead>
<tbody>
<tr>
<td>exo-E415G WT Mut</td>
<td>exo-E415G WT Mut</td>
</tr>
</tbody>
</table>

The number of samples that are wild-type (WT) or mutant (Mut) for the exonuclease E415G mutation (exo-E415G) and have one (CN1) or multiple (CN2+) copies of plasmspin 2–3, according to province and year of collection, are shown. Samples where data for one of the two markers were not available or the marker was present in a mixed infection were excluded.

Table 2: Joint distribution of the two piperquine resistance markers in Cambodia

www.thelancet.com/infection Published online November 3, 2016 http://dx.doi.org/10.1016/S1473-3099(16)30409-1

recrudescent parasitaemias following dihydroartemisinin–piperquine therapy.

To gain a better picture of its frequency distribution in Cambodia, we used Sanger sequencing of PCR-amplified DNA fragments to genotype this SNP (ie, exo-E415G) in 168 additional samples for which whole-genome sequence data were not available, from the same initial cohort of 241 patients. As expected, given the low p value of this SNP in the GWAS, the phenotype distribution differs significantly between the wild-type and the mutant alleles. When stratified by province and year, the association seems to be particularly strong in Pursat and Preah Vihear (figure 1B, appendix), and from 2012 onwards (appendix). Consistent with these results, the frequency of exo-E415G appears to increase in Pursat and Preah Vihear over time (figure 1C, table 2). In summary, exo-E415G is the best predictor of raised IC₅₀s in the current dataset, although this finding alone is insufficient to infer a causal role for it in mediating piperquine resistance.

Since gene CNVs have also been associated with drug resistance in P falciparum, we investigated the association of piperquine IC₅₀s with 43 genomic regions exhibiting CNVs in at least five samples. We did a GWAS using the same method we used in the SNP GWAS, except that genotypes were the presence or absence of the CNV in each sample. We found that two CNVs strongly associated with IC₅₀s (p≤2.3×10⁻⁴; figure 2A, appendix). As expected, mdr1 amplification was associated (p=1.4×10⁻⁹) with low IC₅₀s. However, the most significant association (p=7.6×10⁻⁷) was found for a novel amplification on chromosome 14, encompassing two of the ten plasmspin genes in the P falciparum genome—plasmspin 2 and plasmspin 3—that encode aspartic proteases involved in the haemoglobin-to-haemozoin conversion pathway targeted by quinolines.
version of plasmepsin 3, with its 3’ end replaced by the 3’ end of plasmepsin 1 (appendix). Due to the degree of homology between plasmepsin 1 and plasmepsin 3, the aminoacid sequence of the chimeric plasmepsin 3 protein is identical to that of plasmepsin 3, except that an asparagine residue is replaced by two consecutive lysines. The prevalence of plasmepsin 2–3 amplification in our cohort shows a steady increase over time in Pursat and

Figure 2: Manhattan plot showing the statistical significance of copy number variation associations in the genome-wide association study (A) and piperazine 50% inhibitory concentrations according to mdr1 and plasmepsin 2–3 copy number (B, C). (A) Each point represents one of the 43 copy number variations (CNVs) present in at least five samples, alternately coloured red and blue according to chromosome. Genomic location is shown on the x-axis. The p value for each CNV’s association, calculated using a linear mixed model, is shown on the y-axis; point size is proportional to significance level. The province of sample origin, status of kelch13 (mutant vs wild type), and a genetic relatedness matrix were added as fixed effects to the analysis. Two CNVs reached the Bonferroni-corrected, genome-wide significance level (p≤2×10⁻⁴; above the horizontal blue line), one including plasmepsin 2 and plasmepsin 3, and one including mdr1. All 43 CNVs are marked by black lines at the top and are listed in the appendix. Dashed grey vertical lines are plotted every 500 kb from the beginning of each chromosome. Solid grey vertical lines mark telomeric, sub-telomeric, and internal hypervariable regions. (B, C) Each point represents the piperazine 50% inhibitory concentrations (IC₅₀) for a Plasmodium falciparum clinical isolate carrying wild-type (WT) or amplified (Amplification) mdr1 (B) or plasmepsin 2–3 (C) genes. Bold and thin horizontal lines indicate the median and IQR of each distribution, respectively. Filled circles identify samples also carrying exo-E415G.
Preah Vihear compared with Ratanakiri (appendix), and reflects our observations of increasing piperazine IC₅₀s and rising prevalence of exe-E415G in these two provinces (table 2, appendix). Despite being on different chromosomes, exe-E415G and plasmodin 2–3 amplification are in significant linkage disequilibrium (r²=0.56, empirical p=1.6×10⁻⁵; appendix), and have a very similar allele frequency distribution in these Cambodian data. In particular, of the 462 samples where both markers were reliably typed, 72% (n=334) have neither marker, 19% (n=86) have both markers, and only 9% (n=42) have one of the two markers. Although the occurrence of the two markers in the population is interesting and surprising, the few samples where the two markers are found separately makes any conclusion regarding their relative effect on the phenotype difficult to support statistically (appendix).

To further investigate whether the exe-E415G and plasmodin 2–3 amplification markers segregate with a newly described piperazine resistance phenotype, we subjected a subset of 12 clinical isolates in triplicate to the in-vitro piperazine survival assay.25 In this assay, parasites did not recrudesce after dihydroartemisinin–piperazine treatment (A), and recrudesced after dihydroartemisinin–piperazine treatment (B). The relationship between raised piperazine survival rates (median 61.6 nmol/L [IQR 56.8–67.0]; n=6) than the first (2.4 nmol/L [1.6–2.9]; n=6; p=0.002). The relationship between raised piperazine survival rate and the presence of either exe-E415G or plasmodin 2–3 amplification were fully concordant.

Given that exe-E415G and plasmodin 2–3 amplification are associated with raised piperazine IC₅₀s and survival rates, which, in turn, have been associated with parasite recrudescence,26 we directly tested their association with dihydroartemisinin–piperazine failures. Of the 241 samples with clinical outcome data, we analysed a subset of 133 samples for which we had complete genetic and clinical information (appendix). Both exe-E415G and plasmodin 2–3 amplification mutations showed a highly significant enrichment in recrudescent samples (p=1.6×10⁻⁸ and 1.8×10⁻¹¹, respectively, Fisher’s exact test; appendix), with aHRs of recrudescence of 13.4 (95% CI 5.3–33.5; p=3.0×10⁻⁸) and 16.7 (5.8–48.1; p=1.7×10⁻⁷), respectively. Furthermore, the aHR for plasmodin 2–3 amplification is still significant (5.2 [95% CI 1.5–17.7]; p=8.6×10⁻³) when only kelch13 mutants were considered, suggesting that whereas artemisinin resistance is certainly a risk factor, plasmodin 2–3 amplification potentially has an additional, independent effect on piperazine resistance and, ultimately, dihydroartemisinin–piperazine treatment failure. To explicitly clarify the effect of the two markers on treatment success, we analysed the survival rate in samples with or without the two markers. We found that the adjusted survival rates at 63 days in the presence or absence of exe-E415G were 0.38 (95% CI 0.25–0.51) and 0.93 (0.85–0.97; figure 3A), respectively. Samples with or without plasmodin 2–3 amplification had survival rates of 0.41 (95% CI 0.28–0.53) and 0.95 (0.87–0.98; figure 3B), respectively. Considering that non-recrudescent infections are not necessarily indicators of drug sensitivity, these values are likely underestimated. Together, these data identify exe-E415G and plasmodin 2–3 amplification as strongly predictive markers of dihydroartemisinin–piperazine treatment failures in Cambodia.

Discussion

In GWAS analyses of piperazine IC₅₀ phenotypes, we identified two genetic markers of piperazine resistance in vitro and of dihydroartemisinin–piperazine treatment failures in patients: a non-synonymous SNP encoding a Glu415Gly mutation in a putative exonuclease (exe-E415G), and amplification of the plasmodin 2 and plasmodin 3 genes (plasmodin 2–3 amplification). The prevalence of these two markers has increased substantially in recent years in Pursat and Preah Vihear, where artemisinin resistance is prevalent and where dihydroartemisinin–piperazine has been the front-line artemisinin combination therapy since 2008 and 2010, respectively. In a global dataset of SNP allele frequencies, the exe-E415G allele is observed only
in eastern mainland southeast Asia, where dihydroartemisinin–piperaquine is used, and completely absent where this artemisinin combination therapy has not been used.

Of the associated variations found in this study, \textit{plasmspsin} 2–3 amplification is the strongest candidate for mediating piperaquine resistance because of the role of plasmspsins in haemoozon synthesis pathways in the parasite food vacuole. Since piperaquine, like chloroquine, is believed to inhibit the conversion of toxic haem moieties to non-toxic haemoozon crystals during haemoglobin digestion, it is possible that piperaquine targets \textit{plasmspsin} 2, \textit{plasmspsin} 3, or both, and that overproduction of these plasmspsins counteracts the drug’s action. The SNP markers, \textit{exo-E415G} and \textit{mcp-N252D}, do not lend themselves to an equally simple explanation and, until functional studies are done, it will not be possible to determine whether either is directly involved in mediating piperaquine resistance, whether their role is compensatory for lost fitness in piperaquine-resistant mutants, or whether they are associated with piperaquine resistance simply because of their strong linkage to \textit{plasmspsin} 2–3 amplification or some other functional mutation in the Cambodian parasite populations.

Assigning a role to the single copy variant of \textit{mdr1}, which is associated with lower sensitivity to piperaquine, is equally problematic. Although it is possible that \textit{mdr1} plays a functional part in piperaquine resistance (eg, by restricting the amount of drug that enters the parasite’s food vacuole), recent changes in Cambodia’s antimalarial drug policy have caused a decline in mefloquine pressure, which might have promoted the loss of \textit{mdr1} amplifications in the parasite populations. In view of the pronounced population structure following the emergence of artemisinin resistance in Cambodia,\textsuperscript{27} it is possible that the association with a single copy of \textit{mdr1} is the result of piperaquine resistance having emerged in specific artemisinin-resistant, mefloquine-sensitive populations.

Taken together, these findings identify \textit{plasmspsin} 2–3 amplification as the most likely candidate to be a causal variant of piperaquine resistance. In the samples studied here, it is strongly linked to \textit{exo-E415G}, which is on a different chromosome and might represent some other functional component of the resistance phenotype. In practice, this means that the \textit{exo-E415G} SNP can be presently used as a predictive marker of piperaquine resistance in Cambodia; however, this might not be the case elsewhere. Therefore, \textit{plasmspsin} 2–3 amplification should be monitored in areas where dihydroartemisinin–piperaquine is used, although it is somewhat more laborious to genotype. Despite unresolved questions about causal mechanism, we now have tools to identify areas where treatment failures with dihydroartemisinin–piperaquine are likely to occur, and thereby to empower national malaria control programmes to make informed decisions about whether to switch to alternative artemisinin combination therapies for first-line treatment of \textit{P falciparum} malaria.

\textbf{Contributors}
RA, PL, OM, DPK, and RMF designed the study. RA, PL, CA, DD, ATN, SSR, SSu, and ED collected data. RA, PL, OM, RDP, JAG, DJ, JS, DPK, and RMF analysed and interpreted data. RA, PL, DPK, and RMF wrote the manuscript.

\textbf{Declaration of interests}
We declare no competing interests.

\textbf{Acknowledgments}
This work was funded by the Intramural Research Program of the National Institute of Allergy and Infectious Diseases, US National Institutes of Health; Bill & Melinda Gates Foundation (OPP1040643, OPP1118816); Medical Research Council (G0600718); and UK Department for International Development (M006212). This publication uses data from the MalariaGEN \textit{Plasmodium falciparum} Community Project. Genome sequencing was done by the Wellcome Trust Sanger Institute and the Community Project was coordinated by the MalariaGEN Resource Centre with funding from the Wellcome Trust (098051, 090770). We thank all patients for participating in this study; the staff of the WTSI Sample Logistics, Sequencing, and Informatics facilities for their contribution; Marcus Lee for helpful discussions; and Robert Gwadz and Thomas Wellem for their efforts in support of this work.

\textbf{References}


