Genetic surveillance in the Greater Mekong Subregion to support malaria control and elimination

The SpotMalaria platform - Technical Notes and Methods

V1.0, May 2020

Change History

Version	Date	Notes
1.0	18 May 2020	Details SpotMalaria V2.1 and previous versions. Released alongside Jacob CG *et al.* (2020)

Table of Contents

Introduction 3
SpotMalaria Outputs 3
Antimalarial Drug Resistance 3
Artemisinin 5
Chloroquine 5
Mefloquine 6
Piperaquine 6
Antifolates (Sulfadoxine, Pyrimethamine, and Cycloguanil) 6
Genetic Barcodes and Additional Loci 7
Mixed Species Infection 8
Platform Validation 8
Selection of Loci 10
Drug resistance loci 10
Genetic barcodes 10
Speciation 10
P. vivax loci 11
V1 Platform Implementation Methods 12
Agena Biosciences MassArray® assays 12
Assay design 12
Sample selection 12
Sample processing 12
Data Analysis 13
Kelch13 capillary sequencing 13
V2 Platform Implementation Methods 14
Amplicon Sequencing 14
Assay design 14
Sample selection 14
Sample processing 17
Primer preparation 17
Pre-sequencing PCRs 19
PCR1 20
PCR2 20
Product purification and size selection 21
Quantification and Sequencing 23
Primer re-balancing 24
Data Analysis 25
Abbreviations 27
URLs 27
List of Supplementary Figures and Tables 28
References 29

Introduction

This document describes the **SpotMalaria** genetic surveillance platform, which powers large-scale regional projects such as GenRe-Mekong. It is intended to provide a technical reference for this platform, which will be updated and versioned over time, and detail the platform's methods for reproducibility. The SpotMalaria platform is defined in terms of the data products it delivers, i.e. the set of genotypes and other experimental results produced by the platform. This means that the platform will be updated and versioned when new data products are added or modified (e.g. when a new marker of resistance to an antimalarial is discovered and added to the platform). It is important to note that the set of *data products* is distinct from the *methods* that are used to implement the platform. In other words, SpotMalaria is method-agnostic, and over time, new versions of the platform will use new or modified methods, refined to provide the most reliable outputs at the most reasonable costs in the way that best guarantees participation by the partners involved.

So far, two major versions of the SpotMalaria platform have been released:

- Version 1 operates with the massArray® system from Agena Bioscience Inc. (San Diego, USA) which relies on centralized large-scale processing of samples using mass-spectrometry facilities[1], and
- Version 2 uses an amplicon-sequencing approach, implemented on the widely available MiSeq mid-size sequencing machine from Illumina Inc. (San Diego, USA).

In this document we provide technical details of both platforms.

SpotMalaria Outputs

The following sections detail the different components of SpotMalaria outputs, and describe how they are implemented in different versions of the platform

Antimalarial Drug Resistance

One of the main issues affecting malaria control is wide spread drug resistance to most current first and second line treatments. According the WHO, 17 different antimalarial drugs are part of national treatment policies [2]. For *P. falciparum,* artemisinin derivatives are by far the most prescribed for uncomplicated cases and used primarily with a longer lasting partner drug which includes amodiaquine, lumefantrine, mefloquine, and piperaquine. For severe cases, both quinine and artemisinin are widely used, and for prevention during pregnancy– sulfadoxine and pyrimethamine. Much less is known about drug resistance in *P. vivax* and studies are ongoing to try and elucidate

markers and determinants of resistance. Potential candidate markers have been proposed and these were included in the multiplex assays. *P. vivax* treatment is predominantly by use of chloroquine and primaquine, but many other drugs are also used. All genotyped markers associated with drug resistance are listed in Tables SM1 and SM2 for *P. falciparum* and *P. vivax* respectively.

Table SM1 - Markers associated with drug resistance for *P. falciparum*, which are included in the SpotMalaria platform.

Target	Antimalarial	Gene	Core Amino Acid Positions	Wild Type* (3D7)	Additional Amino Acid Positions
K13	artemisinin	*pfkelch13*	any mutation found in BTB/POZ and propeller domains	WT	-
DHFR	Pyrimethamine, cycloguanil	*pfdhfr*	51, 59, 108, 164	NCSI	A16, S306
DHPS	sulfadoxine	*pfdhps*	436, 437, 540, 581, 613	SAKAA	-
EXO	piperaquine	*exonuclease*	415	E	-
MDR-1	chloroquine, amodiaquine, lumefantrine, mefloquine	*pfmdr1*	86, 184, 1246	NYD	S1034, N1042, F1226
PM2/3 Breakpoint (P23_BP)	piperaquine	*Plasmepsin 2/3*	-	WT	-
CRT	Chloroquine, piperaquine	*pfcrt*	72, 73, 74, 75, 76	CVMNK	T93, H97, I218, A220, Q271, T333, G353, R371
Artemisinin Genetic Background	artemisinin	*pfarps10*	127, 128	VDDNIT	-
		ferredoxin	193		
		pfcrt	326, 356		
		pfmdr2	484		

*wild type allele does not necessarily indicate "sensitive" allele

Table SM2 - Markers associated with drug resistance for *P. vivax*, which are included in the SpotMalaria platform.

Target	Antimalarial	Gene	Amino Acid Positions	Wild Type* (P01)
DHFR	Pyrimethamine	*pvdhfr*	57, 58, 61, 117	F, R, T, N
DHPS	Sulfadoxine	*pvdhps*	380, 382, 383,385, 553	E, S, G, Y, A
MDR-1	Amodiaquine, Chloroquine	*pvmdr1*	976	Y

*wild type allele does not necessarily indicate "sensitive" allele

Both space and methodological restraints limited the number of drug resistance loci we were able to genotype in version 1 of the assays. For Agena, the use of single base-extension primers forced each mutation to have its own primer, and in areas of sequential or very dense mutational sites like *pfcrt* 72-76 we had to rely on genotyping only critical loci, while using imputation where possible to fill in other information. Additionally, in version 1 mutations that had less evidence for association with drug resistance or had not yet been studied were left out due to limits in the size of multiplexes in the Agena platform. The transition to amplicon sequencing removed the need for the base-extension primers effectively removing issues with typing closely occurring mutations. Also, while the number of loci in amplicon sequencing is limited, it is increased from that of Agena and gave us the ability to incorporate more loci that have less validated association with drug resistance as well as including markers for future use in genotyping CNVs. Details about the selection of each marker for study can be found in the below.

Artemisinin

Artemisinin resistance is typified by a reduction in the rate of parasite clearing or the complete failure to clear parasites resulting in a recrudescence. Mutations in the *pfkelch13* (PF3D7_1343700) gene [3, 4] have been linked to this phenotype as well as a specific "genetic background" found in parasites from Southeast Asia [5], comprised of SNP in the *pfarps10* (PF3D7_1460900), *pffd* (PF3D7_1318100), *pfcrt* (PF3D7_0709000), and *pfmdr2* (PF3D7_1447900) genes. The large number of mutations in the *pfkelch13* gene make targeted approaches difficult to implement, and for the Agena multiplexes we opted to not type *pfkelch13*, but instead performed capillary sequencing on all samples. In the switch to version 2, we were able to sequence across the resistance sections of the *pfkelch13* gene in the amplicon sequencing multiplexes.

Chloroquine

Chloroquine is the canonical case study for antimalarial resistance and is no longer widely used for *P. falciparum* control. However, use in treating *P. vivax* can keep pressure on *P. falciparum* parasites in areas of dual-infections, and reversion of resistance in some areas has made it possible for re-introduction as a treatment [6, 7]. Resistance to chloroquine is primarily mediated through mutations in the *pfcrt* gene [8-10] (codons 72-76), with additional resistance being conferred by a mutations (codons N86 & Y184) in the *pfmdr1* (PF3D7_0523000) gene [11, 12]. Genotyping the main resistance locus in *pfcrt* by Agena single-base extension PCR proved difficult, as multiple mutations in proximity meant designing primers to overlap mutations which lowers binding efficiency. Our initial focus was on genotyping the K76 codon and we were able to successfully genotype codon N75 as well. For samples run on the Agena platform we performed imputation to infer the rest of the

five-codon haplotype based on geographic haplotype distribution and the N75 & K76 codons. The switch to amplicon sequencing allowed us to genotype all loci at the multivariant site as well as additional sites in *pfcrt* associated with resistance to chloroquine, including codons A220, Q271, T333, G353, & R371 [13, 14]. Even as one of the primary treatments for *P. vivax* malaria, chloroquine resistance in that species is largely unsolved. Currently, we only genotyped a single locus (Y976) in the *pvmdr1* gene (PVP01_1010900) which has been somewhat associated with resistance to chloroquine as well as amodiaquine in *P. vivax* [15, 16].

Mefloquine

Mefloquine resistance is modulated by both point mutations and CNVs, centered around the *pfmdr1* gene. This gene confers differing levels of resistance to multiple drugs including mefloquine [12, 17], chloroquine, quinine [18], lumefantrine [19, 20], and amodiaquine [19, 21]. In version 1 of the protocol we focused on codons N86 and D1246 [12, 22], but expanded the targets in version 2 to include codons S1034, N1042, and F1226 [22, 23]. For version 1, we experienced lower than average genotyping efficiency for codon 86 but amplicon sequencing assays in version 2 provided normal genotype calling efficiency.

Piperaquine

Two molecular markers of piperaquine resistance were initially targeted in the assays, the first being a SNP (codon E415) within the *pfexo* (PF3D7_1362500) gene [24], and the second was a copy-number polymorphism in the *pfplasmepsin2/3* (PF3D7_1408000 & PF3D7_1408100) genes [24, 25] which was able to be detected by a conserved break-point found in samples across Southeast Asia. In version 2 of the protocol using amplicon sequencing, we included additional loci in *pfcrt* (T93, H97, I218, & G353) that have been identified as further drivers of piperaquine resistance [26, 27]. Version 3 under development, will incorporate further mutations in the *pfcrt* gene, specifically the F145 and M343 codons [27, 28].

Antifolates (Sulfadoxine, Pyrimethamine, and Cycloguanil)

The antifolates sulfadoxine, pyrimethamine, and cycloguanil target genes in the folate synthesis pathway, specifically *pfdhfr* (PF3D7_0417200) and *pfdhps* (PF3D7_0810800). Version 1 of the genotyping assays targeted four mutations (codons, N51, C59, S108, & I164) in *pfdhfr* [29] which are most highly associated with resistance to pyrimethamine and five mutations (codons S436, G437, K540, A581, A613) in *pfdhps* [30, 31] which are linked with resistance to sulfadoxine. Resistance to the drug combination Sulfadoxine-Pyrimethamine (SP), most commonly used now as intermittent preventive treatment of malaria in pregnancy (IPTp), is typified by a “quintuple” mutation haplotype

of *pfdhfr* 51/59/108 with *pfdhps* 437/540. Version 1 had lower than average genotyping coverage for position S436 in *pfdhfr*. Version 2 corrected this inefficiency as well as adding two additional loci in *pfdhfr* at codons A16 linked with cycloguanil resistance [32], and position S306 which is seen to be highly differentiated in sample from Papua New Guinea and could play a role in resistance. Antifolate drugs are also used for treatment in cases of *P. vivax* infection. While there is limited evidence to support specific markers, we genotype four loci (F57, R58, T61, N117) in *pvdhfr* (PVP01_0526600) and five loci (E380, S382, G383, Y385, & A553) in *pvdhps* (PVP01_1429500).

Genetic Barcodes and Additional Loci

For *P. falciparum* we sought to create a genetic barcode for use in calculating basic parasite and population statistics. Using the MalariaGEN parasite genome variation database we started with a list of all SNPs variant across all major malaria endemic regions. The initial set of SNPs were scored and ranked based on their ability to differentiate populations and recapitulate sample heterozygosity (COI). The highest scored SNPs had a high average MAF (0.35-0.5) across the global populations and a lower population F_{ST} (<0.1). When testing, we determined barcodes of 20 or 50 SNPs were able to identify parasite lineages and provide comparable estimates of between sample genetic distances. A larger barcode of 100 was able to significantly decrease the error of the estimates when comparing to the same statistic calculated from whole-genome data. Given that there was little cost increase from performing two multiplexes on Agena compared to four, we opted for the larger barcode size. We were able to successfully design multiplexes containing 101 top rated barcode SNPs across 4 PCRs, in conjunction with the drug resistance mutations. Most SNPs performed well across all samples, with an average individual pass rate of 83% in samples that passed QC, with 3 SNPs having a poor pass rate below 20%. For designing the Amplicon sequencing assays, we were able to design amplicons for 100 of the 101 markers, with an average individual pass rate of 95% in samples that passed QC. Overall, the selected loci have performed well for identifying parasite lineages [33], COI, and also looking at migration [34] in surveillance samples from multiple studies.

A different approach for *P. vivax* was taken when developing a molecular barcode. Previously, a group at the Broad Institute developed a SNP barcode [35] for uses similar to those we've described above in *P. falciparum*. To harmonize our work with that done in previous studies, we chose to adapt the loci in their barcode to work for our system. Thirty-eight of the forty-two SNPs in the barcode were successfully designed and genotyped in version 1 of the *P. vivax* assays, and the four missing SNPs were able to be rescued in version 2. Each had good coverage across all SNPs with an average success rate of 94% in the Agena assays, and 97% in amplicon sequencing assays. The smaller *P.*

vivax barcode, and less numbers of putative drug resistance markers meant available spaces in multiplexes for the addition of more research oriented genotyping assays. From WGS analysis of several hundred *P. vivax* field isolates, multiple regions of the *P. vivax* genome exhibited signatures of selection [36]. From these regions we selected markers for genotyping to track their frequency over time in field populations. In addition to these regions, we included potential population markers for tracking parasite populations.

Mixed Species Infection

While *P. falciparum* is responsible for most malaria related deaths, other species frequently infect humans, and in much of the world *P. vivax* causes the majority of disease. Understanding the full extent of infections by alternate species is essential as regions push towards elimination. In our version 1 assays we strove to detect *P.vivax* and *P. knowlesi*, focused on infections from Southeast Asia. Detection was expanded in later iterations of version 1 to detect *P. ovale* and *P. malariae* to cover the major human infecting species. Detection of multiple species within a sample by molecular methods serves several purposes. It can help distinguish visually similar looking parasites observed through microscopy, it can also detect parasites at low levels in mixed infections. The ability to detect these parasites can help control organizations map the extent and frequency of understudied parasites like *P. ovale* and *P. malariae*, or the extent of zoonotic infections of *P. knowlesi*. For both version 1 and version 2 assay sets, targets were in the parasite's mitochondria relying on the shared conservation of sequence across the genus. Species specific polymorphisms were used to detect the infecting species.

Platform Validation

To validate both the Agena and amplicon sequencing assays we used a set of test samples consisting of lab cultured isolates and field collected samples. Preliminary validation was done with mostly lab isolates either individually or in mixtures and spiked into dilutions of human DNA to simulate typical sample qualities. Two rounds of preliminary validation were done after initial tests showed a small proportion of assays failed to amplify and generate genotypes. Following redesign, in *P. falciparum* all but two barcode SNPs were successfully genotyped, and two drug resistance SNPs (*pfdhps*:436 & *pfmdr1*:86) had lower than average yield. For *P. vivax*, four of the barcode SNPs were unable to be designed, but these were rescued in the amplicon sequencing multiplex. As an initial threshold for both *P. falciparum* and *P. vivax* we used a minimum concordance of 95% identity between WGS and genotypes, and across all assays the average concordance was >97%. Upon further examination it was noted that most incongruencies were within heterozygous positions, where in one method

there was single allele and the comparison method was a heterozygote. In these instances, it was primarily Agena where a single allele was being called and the WGS was showing heterozygous alleles. This was observed in ongoing use of these assays for both *P. falciparum* and *P. vivax* and can lead to underestimates of complexity of infection when using Agena genotypes. The reason behind the lower heterozygous rate in Agena calls is due primarily to background noise associated with using low volumes of parasite DNA in the mass spectrometry-based method not allowing distinction between low proportion minor alleles.

For validation of amplicon assays, we compared genotypes from amplicon sequencing to those from both WGS as well as Agena. For *P. vivax*, a >99% concordance was seen between amplicon sequencing and Agena and for *P. falciparum*, a 98% concordance between Agena and amplicon sequencing which raised to >99.9% when disregarding heterozygous positions. For comparisons of amplicon genotypes with those from WGS we looked at field samples processed from dried blood spots (DBS) as well as from whole-blood, with both having an overall concordance of >99%.

In every test plate we included mixtures of the reference strain *P. falciparum* 3D7 DNA with human DNA in various dilutions to test the lower limits of detection and genotype reliability. For both Agena and amplicon sequencing we were able to detect parasite DNA at a minimum concentration of 0.0006 ng/ul, and we were able to reliable genotype (with some missing loci) at ~0.00125 ng/ul, which approximates to 100 parasite genomes in the beginning reaction of 2µl. Genotypes from very low concentration samples failed to meet the 95% concordance threshold and parasites with <50% callable loci are considered low quality and not used for analysis. Originally, for Agena validation of speciation assays, only *P. vivax*, *P. falciparum*, and *P. knowlesi* were able to be differentiated by the assays. The different assays added to the later versions, including the amplicon sequencing panel were able to discern all five species. Test samples used in both the Agena and amplicon assay validation experiments were able to detect all mixtures of parasites. Further analysis of field samples has identified numerous mixed species infections in previously classified mono-infections [33], and was confirmed by searching WGS sequence files for unique reads aligning to the detected minor species.

Selection of Loci

Drug resistance loci

Extensive literature review identified key targets for genotyping polymorphisms associated with drug resistance to key antimalarial compounds. For *P. falciparum* (*Pf*), mutations outlined in Table SM1 are associated with resistance or have been validation as a molecular marker of resistance. Core mutations have had more extensive validation and associated evidence, while additional positions have less supporting evidence, or lessened effects. In addition, we genotyped a genetic "background" set of mutations that were associated with artemisinin resistant parasite populations in Southeast Asia. In *P. vivax* (*Pv*), less evidence exists for molecular markers of drug resistance, and we genotyped putatively associated markers found in literature searches as well as some *Pf* homologues (Table SM2).

Genetic barcodes

To better represent the global parasite population, we designed a novel genetic barcode for use in surveillance. A set of >3,000 *Pf* whole genomes [37] was used to compile a list of potential SNP targets variable in 8 global populations with a minor allele frequency (MAF) of >= 0.01 in every population. These SNPs were grouped into bins based on MAF and global F_{ST}. Random SNPs (1000 iterations) were selected to comprise barcodes of various sizes (1, 2, 5, 10, 20, 50, 100) and analyzed for their ability to recapitulate statistics calculated from whole-genome data. We tested pairwise genetic distance, population differentiation, and sample heterozygosity. Each SNP was scored by bin and iteration and ranked by score. All ranked SNPs (in order) were presented to the Agena design software (described below) and a total of 101 top-ranked SNPs (SpotMalaria Supplementary File 1, sheet "*P. falciparum* barcode SNP list") were able to be designed in multiplexes with the drug resistance and speciation SNPs. For COI estimation, the programs COIL [38] (default parameters) and The Real McCOIL [39] (maxCOI=20, threshold_ind=20, threshold_site=20, totalrun=1000, burnin=100, M0=5, e1=0.05, e2=0.05, err_method=1) were used. While it was necessary to design a barcode for Pf, a recently published *Pv* barcode [35] proved a good resource and was adapted for use on the Agena and amplicon sequencing platforms (SpotMalaria Supplementary File 1, sheet "*P. vivax* barcode SNP list").

Speciation

An important aspect of malaria control is the identification and treatment of mixed species infections. In most areas of the malaria endemic world multiple species of human infection plasmodium co-exist and are capable of causing illness, and in these cases a single drug or

combination may not be effective at killing both species. A set of assays were designed to target cross-species conserved areas of the mitochondrial genome to enable identification and speciation in mixed infections. Mitochondrial genome sequences for the five human infecting species of Plasmodium (*Pf*; Pf_M76611, *Pv*; PvP01_MIT_v1, *P. malariae*; PmUG01_MIT_v1, *P. ovale*; PocGH01_MIT_v2, *P. knowlesi*; PKNH_MIT_v2) were aligned using the online multi-sequence alignment tool MAFFT (version 7, strategy G-INS-i) [40]. Downloaded sequences were found to begin at different positions due to the tandem repeat nature of mitochondrial genomes. Prior to alignment all sequences were manually edited to begin at the same start of sequence as *Pf*, and *P. ovale* was reverse complemented from the published sequence (SpotMalaria Supplementary File 2). For the initial Agena set of assays two loci were selected to differentiate species based on fixed differences between species. For adaptation to amplicon sequencing, different loci were selected to increase the differentiation power between the 5 species (Figure SM1).

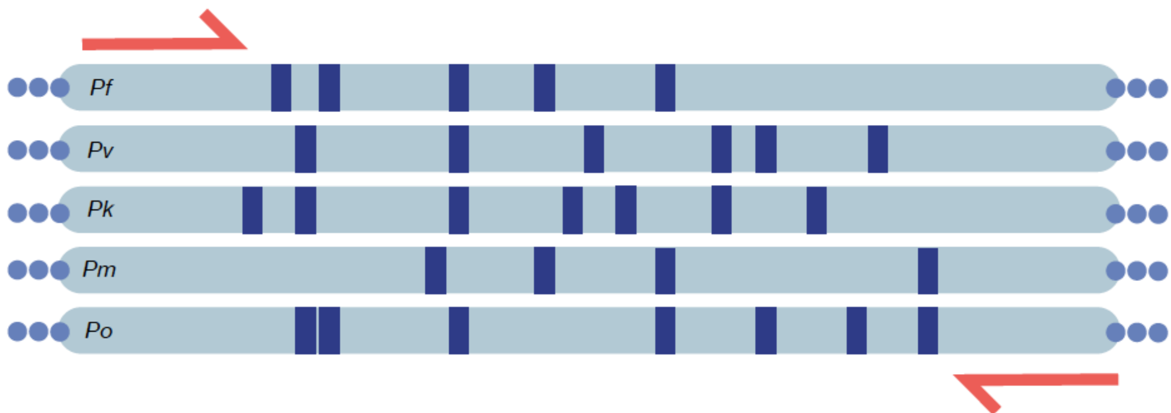

Figure SM1. Schematic of mitochondria-based speciation. Common primers (red arrows) bind to conserved sequence in all 5 species allowing for amplification and identification of species-specific variation (blue vertical bars).

P. vivax loci

The volume of genomic information that exists for *Pv* is much less than *Pf*, in large part due to lower overall parasitemia and lack of a protocol for long-term laboratory culture. To supplement the barcode and drug resistance loci, additional markers derived from tests of population differentiation (F_{ST} and cross-population extended haplotype homozygosity (XPEHH)) were included to help monitor regions of the genome possibly under-going selection as a result of human interventions [36]. Additional loci were included which had potential in being able to differentiate major populations of *Pv* parasites (S Auburn, personal communication).

V1 Platform Implementation Methods

Agena Biosciences MassArray® assays

Assay design

For designing the Agena Bioscience MassArray® platform (see URLs) multiplexes, all the loci of interest were formatted and processed using the Agena primer design software (MassARRAY® v3.1 software) which groups assays in multiplexes of up to 40 loci. *Pf* and *Pv* loci each form 4 multiplexes per species. The majority of loci were bi-allelic, with 12 and 9 tri-allelic (*Pf* and *Pv* respectively, and 1 tetra-allelic for *Pf* (SpotMalaria Supplementary File 1, sheets "*P. falciparum* Agena primers" and "*P. vivax* Agena primers").

Sample selection

For testing the *Pf* Agena assays success rate and accuracy we created multiple test plates which consisted of 230 *Pf* field isolates, 60 dilutions and replicates of *Pf* 3D7 with human DNA, 11 alternate individual *Pf* lab clones with human DNA, 55 mixtures and replicates of *Pf* clones in human DNA, and 10 negative human controls. All lab strains and field isolates had been whole-genome sequenced (WGS) for testing assay accuracy. For *Pv* we relied mostly on whole-genome sequenced field isolates, with 264 *Pv* samples tested. We also tested 6 human negative controls and 33 *Pf* samples to observe cross-reactivity and accuracy of the speciation assays.

Sample processing

All samples were extracted using commercially available kits (Qiagen QIAamp DNA Investigator Kit (see URLs) either by hand or robotics (QiaSymphony, Qiagen). Both *Pf* and *Pv* DNA underwent primer extension pre-amplification (PEP) [41] prior to genotyping on the Agena platform. PEP uses N15 random primers (Sigma-Aldrich; custom order, see URLs) to amplify whole-genome products. N15 primers for PEP were sourced from Sigma-Aldrich in 20mL of TE at 3,000 O.D. and diluted to 150 OD/ml before aliquoting (220µl) and storing at -20°C. Samples were whole-genome amplified on 96-well plates as previously described [42]. Briefly, a PCR reagent mixture of 45µL/well comprising; 2.2µL of 1:10 diluted N15 primers, 1.25µL 8mM pooled dNTP's (Bioline; Cat#BIO-39025, see URLs), 0.5µL 5U/µL BioTaq DNA polymerase (Bioline; Cat#BIO-21040), 2.5µL 50mM $MgCl^2$ (with BioTaq), 5µL of 10X BioTaq Buffer (with BioTaq), 33.55µL nuclease-free water (ThermoFisher; Cat#AM9930, see URLs) were added to each well of a 96-well skirted PCR plate (ThermoScientific; Cat#AB2800, see URLs). Five microliters of gDNA (1-2ng/µL) were added to the PEP PCR mixture and the plates were

sealed with Microseal "A" adhesive films (Bio-RAD; Cat#MSA5001, see URLs) before thermocycling using a MJ Tetrad (Bio-Rad) PCR conditions used are highlighted in Table SM3. Neat PEP material was stored at -20°C, and was able to withstand many freeze-thaw cycles. Once diluted for use, PEP material can be stored at 4 °C for a brief period (a few days at most), although it is recommended to store it frozen at -20 °C until used (and then only for a single freeze-thaw cycle).

<table>
<tr><th colspan="4">Table SM3 - PEP PCR procedure steps</th></tr>
<tr><td>Step 1</td><td>94°C</td><td>3min</td><td></td></tr>
<tr><td>Step 2</td><td>94°C</td><td>1min</td><td rowspan="4">50 cycles</td></tr>
<tr><td>Step 3</td><td>37°C</td><td>2min</td></tr>
<tr><td>Step 4</td><td colspan="2">ramp to 55°C at 0.1/sec</td></tr>
<tr><td>Step 5</td><td>55°C</td><td>4min</td></tr>
<tr><td>Step 6</td><td>72°C</td><td>5min</td><td></td></tr>
<tr><td>Step 7</td><td>4°C</td><td>∞</td><td></td></tr>
</table>

Data Analysis

MassArray® genotypes are called based on the intensity of peaks corresponding the molecular mass of the extension primer with a single base addition. The use of this system with low quantities of DNA typically seen in *Pf* samples (specifically dried blood spots (DBS)) makes signal to noise ratios lower than normally seen with other species for which this system is usually used. A custom analysis pipeline was therefore necessary to ensure accurate calling. Samples were processed in the lab in batches of up to 384 samples and analysis was done by batch to account for potential processing bias. Quantile normalization of allele intensity data was performed on raw intensities to account for differences in average sample intensity, and background correction of each assay was done using average allele intensities from blank wells in each run. Each assay was assessed independently in a 2-step training-based algorithm. Stringent calling identified the distribution of assay intensities for samples containing non-mixed genotypes (single infections) and the range of alternate allele intensities was used to calculate the per assay background noise to determine true genotypes during the second round of calling. Assays with three or four possible bases were assessed manually per batch after normalization and background calculations.

Kelch13 capillary sequencing

Due to the extensive list of mutations in *Pfkelch13* putatively associated with artemisinin resistance it made the Agena platform infeasible for genotyping. Instead, capillary sequencing of the core part of the *Pfkelch13* gene was performed on all samples using PCR products generated from whole-

genome amplified material with a modified published protocol [3]. Sequencing primers (FWD: 5'-ACG TTG GAT GAC TTC GCC ATT TTC TCC TCC -3', REV: 5'- ACG GGG GGG TTG ATG CAA ATA TTG CTA CTG -3' produced overlapping reads of ~800-900 bp. PCR and sequencing was carried out by Eurofins UK, Genomics Services (see URLs) as per their instructions. Sequence files were analyzed using the SeqMan Pro^{TM} software from DNASTAR (V14.0.0 (see URLs)). SNPs were called using PF3D7_1343700 (SpotMalaria Supplementary File 3) as the reference gene.

V2 Platform Implementation Methods

Amplicon Sequencing

Assay design

For the design of amplicon sequencing multiplex primers, we included; key loci used on the Agena platform; several non-“core” drug resistance loci; plus provisional *Pf* loci for use in detecting copy number variations. These loci were used to design multiplexes of non-conflicting primers. The program MPprimer [43] was used to create multiplexes using reference genomes as well as masking files of known polymorphisms in both *Pf* and *Pv*. The reference genomes used were *Pf* 3D7 v3 and *Pv* P01 v1 (plasmodb.org). Known polymorphisms were extracted from the MalariaGEN Pf3K dataset for *Pf* [44] and an update to the MalariaGEN Pv1 dataset [36], and all variation >5% frequency was included in the mask files. During validation the decision was made to transfer speciation assays from the second *Pf* multiplex to a separate multiplex due to high levels of assay failures. Final multiplex designs therefore consisted of 3 multiplex PCRs for *Pf* with 66, 68, and 2 assay per multiplex (PFA_GRC1, PFA_GRC2, and PFA_SPEC respectively), and a single multiplex of 114 assays (PVI_GRC1) for *Pv*.

Sample selection

For proof of concept testing, a plate of 96 *Pf* samples was created from a mixture of whole-genome sequenced lab strains and high-quality sequenced field isolates. This included laboratory isolates 3D7, Dd2, GB4, and 7G8 (used at concentrations of 4 ng/µL, 1 ng/µL, and 0.1 ng/µL) both individually and as mixtures, to test the ability to detect multiple strains; 36 field samples from leukocyte-depleted whole blood draws (for which high-quality full genome sequence data was available); and a dilution series of *Pf* 3D7 gDNA diluted in human DNA to simulate a natural sample of known “parasitemia” (4% *Pf* DNA w/w to 0.03125%) (Table SM4).

This sample set was used to test initial genotyping accuracy, by comparing the amplicon sequencing calls with both WGS data and the Agena platform's genotype calls. Additionally, this data was used to further refine the amplicon sequencing by adjusting the primer concentrations to ensure equal read coverage across all amplicons (Primer balancing). After primer balancing and initial quality assurance checks we performed dual genotyping of DBS parasite field samples using both amplicon sequencing and Agena genotyping to verify accuracy of the amplicon genotypes. After successful design and processing of *Pf* assays, *Pv* samples were tested and validated using the same methodology. Independent runs totaling over 1000 *Pv* samples were analyzed for concordance.

Table SM4 – Parasite/human test DNA dilution series

	3D7 gDNA (ng)	Human gDNA (ng)	final volu me	final (ng/ul)	3D7 ng/ul	genomes/ uL	genomes/ ng Hu DNA	genome equivalents per uL blood	parasitized RBC per uL blood	% parasitaemia	% parasitaemia
primary mixture	4	96	100	1	0.04	1559.6	16.2	649.8	649.8	0.01300	0.03249
1:2 dilution series	2	98	100	1	0.02	779.8	8.0	318.3	318.3	0.00637	0.01591
	1	99	100	1	0.01	389.9	3.9	157.5	157.5	0.00315	0.00788
	0.5	99.5	100	1	0.005	194.9	2.0	78.4	78.4	0.00157	0.00392
	0.25	99.75	100	1	0.0025	97.5	1.0	39.1	39.1	0.00078	0.00195
	0.125	99.875	100	1	0.00125	48.7	0.5	19.5	19.5	0.00039	0.00098
	0.0625	99.9375	100	1	0.000625	24.4	0.2	9.8	9.8	0.00020	0.00049
	0.03125	99.96875	100	1	0.0003125	12.2	0.1	4.9	4.9	0.00010	0.00024
	0.015625	99.984375	100	1	0.00015625	6.1	0.1	2.4	2.4	0.00005	0.00012
										assuming no anaemia	at anaemia threshold (~40% normal)

Sample processing

All samples were extracted as described above and then underwent a whole genome amplification step. For *Pf* we used a recently developed selective whole genome amplification (sWGA) [45] protocol, while for *Pv*, since no sWGA method was yet available, we used primer extension pre-amplification (PEP) (Described above).

Primer preparation

Three sets of primers were required for the amplicon sequencing protocol. The first PCR (PCR1) amplifies the specific target regions of the DNA and consists of primers designed for the amplicons in the various multiplexes (SpotMalaria Supplementary File 1, sheets "*P. falciparum* assay amplicon sequencing primers" and "*P. vivax* assay amplicon sequencing primers") with each containing a 33 base 5' tail sequence for priming in PCR2. Primers for PCR1 were ordered from IDT (Integrated DNA Technologies, Leuven Belgium; see URLs) at a 100nM synthesis scale with full yield and standard de-salt in plate format, with forward and reverse primers in corresponding wells of separate plates. Primers were reconstituted in 500µM Tris-EDTA buffer (T0.1E; 2.42g Tris base [Fisher Scientific, Cat#BP152-1] with 400 µl of 0.5M EDTA [pH 8.0] in 2L of MilliQ [Merck Millipore; see URLs] water, adjusted to a final pH of 8.0). Primers contained a penultimate 2'-O-Me RNA modification at the 3' end, denoted by a "m", which has been shown to enhance PCR specificity and reduce degradation by nucleases. A working plate of PCR1 primers was made by transferring 7.5µL of the stock forward primers to a new 96-well plate and then adding 7.5µL of the corresponding reverse primer (i.e. one primer pair per well with individual primers at 250µM in a total volume of 15µL). This 'stock' plate format was then used to create the initial pool for test runs, primer balancing, and final re-balanced pools. For primer balancing, a 400µL equivolume "working" pool of primers was created for each multiplex. The concentration of individual primer pairs within these pools was 40nM. To create these pools, 2µL of each primer pair (250µM) from the stock plate was collected into a single lo-bind Eppendorf tube (Eppendorf; see URLs) for the starting pool, and the concentration was calculated (250µM / number of primer pairs). For the PFA_SPEC multiplex, a 1:25 pre-dilution (1µL primer pool : 24µL T0.1E) was done to avoid small volume pipetting due to low primer numbers. To dilute primers to 40nM, we used differing amounts of T0.1E for each multiplex depending on the new pool concentration. The dilution factor was calculated by converting the pool concentration to nM (concentration x 1000) and dividing by 40nM. The amount of pooled primers to add to the working pool was 400 / the dilution factor, and the amount of diluent (T0.1E) was 400 minus the volume of primer pool used (Table SM5). Each working pool was aliquoted to 20µl volumes and stored at -20°C.

Table SM5 – Stock primer dilution

Multiplex	Stock Primer Concentration (µM)	Primer Pairs	Starting Pool Concentration (µM)	Predilution	New Pool Concentration (µM)	Dilution Factor	Volume of Pool to add (µL)	Voulme of Diluent (T0.1E) to add (µL)	Working Primer Concentration (nM)
PFA_GRC1	250	69	3.62	-	3.62	90.5	4.42	395.58	40
PFA_GRC2	250	68	3.68	-	3.68	91.9	4.35	395.65	40
PFA_SPEC	250	2	125	1:25	5	125	4	396	40
PVI_GRC1	250	114	2.19	-	2.19	54.8	7.3	392.7	40

The second PCR (PCR2) in the protocol incorporated multiplex barcodes (tags) and flowcell adapter sequences to the amplified products from PCR1. The flowcell adapter sequence allows for amplified products to be directly sequenced after PCR2 without further library preparation apart from clean-up and normalization. The multiplex and sample tags included in the primer sequences allows for multiple samples to be sequenced simultaneously on a single flow cell. Our protocol leveraged on the MiSeq's ability to support dual-indexing (a different tag at both ends of the amplified product) and can therefore easily process 3 multiplexes of 384 samples used in the *Pf* runs. The two multiplexing primers were the i7 [46], of which there are 96 in plate format, and the i5 of which there are 16 – giving a maximum multiplexing factor of 1536 samples (SpotMalaria Supplementary File 1, sheets "i7 multiplexing primers" and "i5 multiplexing primers"). Multiplexing primers are combined and lyophilized prior to use, and products from PCR1 are added to the dried primers for PCR2. To make the working stock plates of i5 + i7 primers, each was diluted and combined. i7 primers were ordered lyophilized in plates from IDT with full yield and standard de-salt, reconstituted to 300µM with nuclease-free water. For creating 'stamps' of the i7 plate, we transferred 7.5µL of the stock (300µM) primers to a new 96-well plate then diluted to a concentration of 50µM with TE (x1). We transferred 2.5 µL of the 50 µM dilution to individual plates for storage (4°C for immediate use, or -20°C for long term use). i5 primers were also ordered lyophilized from IDT with full yield and standard de-salt in individual tubes. These primers were reconstituted with nuclease free water to a concentration of 100µM. To make the working dilution of i5 primers we added 12,375µL of EB+TritonX-100 to a 15mL Falcon centrifuge tube, then added 125 µL of a single i5 primer. Each i5 primer is diluted into a separate Falcon tube. To make EB+TritonX-100 we added 40mL of Qiagen Buffer EB (Qiagen; Cat#:19086) to a 50mL Falcon tube. To the EB we added 4 µL of neat TritonX-100 (VWR PanReac Applichem; Cat#A4975.1000, see URLs) to the inside of the tube (i.e. not directly into the EB) and vortexed until fully in solution.

To create a combined i7+i5 stock plate, we added 122.5 µL of a single i5 primer to each well of a i7 stamp plate. This was repeated for each different i5 primer into a new i7 stamp plate to make 16 plates of unique i7+i5 (1µM) combinations. From each combined plate, we created a set of PCR2 stock plates by transferring 40 µL of the i7+i5 combination plate into sixteen new 96-well plates. To each of these stock plates we diluted the combined primers with 40 µL of EB+TritonX-100.

For creating the lyophilized 'stamps', we transferred 2 µL of stock primers to a new 96-well plate. Each of the sixteen stock plates for an individual i5 primer can create up to 40 dried stamps for PCR2. To dry, we took the plates with 2 µL of stock and dried in a fan oven (50°C) for at least one hour (until visually dry). Dried plates were sealed with Microseal "A" seal sticky lids (BIO-Rad; Cat#MSA5001) and stored together in sealed plastic bags in the dark in temperature-controlled rooms. Expected longevity of dried plates is at least six months. Combined (i7+i5) stock plates were aliquoted and dried as needed, otherwise we kept stocks frozen at -20°C.

Pre-sequencing PCRs

All PCRs are undertaken in either 96 or 384-well plates. To reduce the occurrence of artefacts such as primer dimers, it was important to reduce the number of no-sample (empty) wells to below 20% (76 wells of a 384-well plate), although at least 1 negative control well (T0.1E) was included per 96-well sample plate. When processing in 384-well plates, we transferred sample plates and primer plates using liquid-handling robots and utilized an interleaved plate layout system (Figure SM2) while maintaining detailed tracking of samples and barcoding primers.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
A	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2
B	3	4	3	4	3	4	3	4	3	4	3	4	3	4	3	4	3	4	3	4	3	4	3	4
C	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2
D	3	4	3	4	3	4	3	4	3	4	3	4	3	4	3	4	3	4	3	4	3	4	3	4
E	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2
F	3	4	3	4	3	4	3	4	3	4	3	4	3	4	3	4	3	4	3	4	3	4	3	4
G	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2
H	3	4	3	4	3	4	3	4	3	4	3	4	3	4	3	4	3	4	3	4	3	4	3	4
I	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2
J	3	4	3	4	3	4	3	4	3	4	3	4	3	4	3	4	3	4	3	4	3	4	3	4
K	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2
L	3	4	3	4	3	4	3	4	3	4	3	4	3	4	3	4	3	4	3	4	3	4	3	4
M	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2
N	3	4	3	4	3	4	3	4	3	4	3	4	3	4	3	4	3	4	3	4	3	4	3	4
O	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2
P	3	4	3	4	3	4	3	4	3	4	3	4	3	4	3	4	3	4	3	4	3	4	3	4

Figure SM2. 384 well plate interleaved format. Four 96-well plates (4 colors) are arranged in quadrants on plate with the A1 well from plates 1-4 in new wells A1, A2, B1, and B2 respectively.

PCR1

We prepared plates of 2µl of neat amplified (sWGA or PEP) sample DNA, one per amplicon pool (e.g. *Pf,* 3 multiplexes = 3 sample plates and *Pv*, 1 multiplex = 1 sample plate), which included at least one well of 2 µl of T0.1E as a negative control. We prepared the PCR1 mastermix in 15mL Falcon tubes appropriate for the number of samples and controls to be processed. For a full sample plate of 96-wells we prepared enough mastermix for 110 aliquots to allow for pipetting overages. The mastermix comprised of 5.5 µl Qiagen multiplex PCR polymerase mix (x2) (Qiagen, Cat#206145), 3.432 µl of nuclease-free water, and 0.068 µl of PCR1 multiplex primer pool (40nM). To each sample well we added 9 µl of mastermix mixing each thoroughly. Plates were sealed with adhesive PCR plate film (Thermo Scientific; Cat#AB-0558) and centrifuged at 1000g for 20 seconds at 4°C. Plates were loaded onto a MJ TetRAD thermocycler and run with PCR1 program (Table SM6). The plate was removed from the thermocycler immediately and placed on an ice/cold block and PCR2 was performed as soon as possible following PCR1. All procedures between PCR1 and PCR2 carried out chilled (on ice/cold blocks) and plates were kept on ice/cold blocks no longer than 3 hours before beginning PCR2.

Table SM6 - PCR1 procedure steps, all multiplexes			
Step 1	95°C	15min	
Step 2	95°C	20sec	5 cycles
Step 3	51°C	40min	
Step 4	60°C	3min	
Step 5	4°C	∞	

PCR2

To run PCR2, we added 5µl of PCR1 amplified product directly from the PCR1 plate to a unique lyophilized i5+i7 tag primer plate. Note that for each multiplex and each sample plate a different tag primer plate must be used. For *Pf* this comprised 3 multiplexes for 4 sample plates totaling 1152 unique sequence tags (96 samples x 12 tag plates), and for *Pv* this was 1 multiplex for 4 samples plates totaling 384 sequence tags. After adding the amplified product to the lyophilized primer plate, we thoroughly mixed to ensure resuspension of primers. Plates were sealed with adhesive PCR films and centrifuged at 1000g for 20 seconds at 4°C. For PCR2 we used a MJ TetRAD PCR machine which is capable of sub-cycling and we used PCR2 Program A (Table SM7), if a thermocycler with sub-cycling is not available, PCR2 program B (Table SM8) should be used. In either program (A or B), in step 1, 95°C must be reached BEFORE adding plate to the machine. Once the plate is placed on the

machine and the lid is secured continue the program. The program can be run overnight and plates can be stored at -20°C for up to one week.

Table SM7 - PCR2 Program A procedure steps (subcycling)				
Step 1	95°C	∞		
Step 2	95°C	20sec		31 cycles
Step 3	68°C	15sec	4 cycles	
Step 4	60°C	15sec		
Step 5	68°C	3min		
Step 6	4°C	∞		

Table SM8 - PCR2 Program B procedure steps (NO subcycling)			
Step 1	95°C	∞	
Step 2	95°C	20sec	31 cycles
Step 3	68°C	15sec	
Step 4	60°C	15sec	
Step 5	68°C	15sec	
Step 6	60°C	15sec	
Step 7	68°C	15sec	
Step 8	60°C	15sec	
Step 9	68°C	15sec	
Step 10	60°C	15sec	
Step 11	68°C	3min	
Step 12	4°C	∞	

Product purification and size selection

For each PCR2 plate, the contents were transferred to a 1.5 ml low-bind tube. If using 96-well plates, all four plates for a single multiplex were combined. Transfer 100µL of the pooled PCR2 product to a new 1.5mL LoBind tube, followed by 75µL of Ampure XP beads (SPRI) (Beckman Coulter; Cat#NC9959336, see URLs), brought to room temperature and mixed immediately prior to pipetting. Any remaining PCR2 product was stored at -20°C.

The PCR2 product and Ampure bead mixture was vortexed for 2-3 seconds before being placed on a NON-magnetic rack for 5 minutes. The PCR2 product-bead mixture was then transferred to a magnetic rack (ThermoScientific; Cat#12321D) and allowed to stand until the supernatant was

completely clear (~4-5 minutes). The supernatant was then carefully removed and discarded, and while still on the magnet, 700 µL of 75% ethanol was added to the tube without disturbing the pellet. After 30 seconds the ethanol was removed with a pipette and discarded. An additional 700 µL of ethanol was added to the tube and removed after a further 30 second incubation. After this second ethanol wash, the tube was pulse centrifuged and any remaining ethanol was removed off of the pellet. The tube was placed back on the magnetic and was left to air dry on the magnetic rack for 2-3 minutes. The tube was taken off the magnetic rack and 105 µL of EB buffer was added to resuspend the pellet and then incubated off-magnet after a quick vortex and spin. The tube was transferred back to the magnetic rack and the beads were captured for 4-5 minutes until the supernatant became clear. While the tube was still on the magnetic rack, the supernatant was extracted and transferred to a new 1.5 mL lo-bind tube. From the extracted supernatant, 100 µL was transferred to a new 1.5 mL tube for a second round of bead size-selection. To this 100 µL, we added 75 µl of fresh Ampure XP beads. After a 5-minute incubation off-magnet, the beads were captured and two identical ethanol washes of 700 µl was done as above. After the second ethanol wash and removal the dried beads were treated with 22 µl of EB buffer and mixed using the pipette. Beads were incubated for 2 minutes off-magnet, then moved to the magnetic rack for 3-4 minutes until the supernatant was clear. The supernatant contained the size selected DNA and was transferred to a new 1.5 µL lo-bind tube. Each multiplex underwent size selection at the same time in separate tubes. To assess the quality of the product it was necessary to observe the amplicon size profile using an Agilent Tapestation (Agilent; Model 2200, see URLs) (High Sensitivity DNA Chip and Reagents) (Cat#5067-5365, 5366, 5584, & 5585) using the manufacturer's recommended procedure. A typical profile had peaks near 25bp, 300-400bp, and a "smear" at ~1000bp. It was essential that there was no discernable peak near 160pb, which was indicative of large primer dimers that affect sequencing (Figure SM3). If a marked peak was visible at 160bp, Ampure bead cleaning was repeated an additional time and sizes re-checked by Agilent Tapestation.

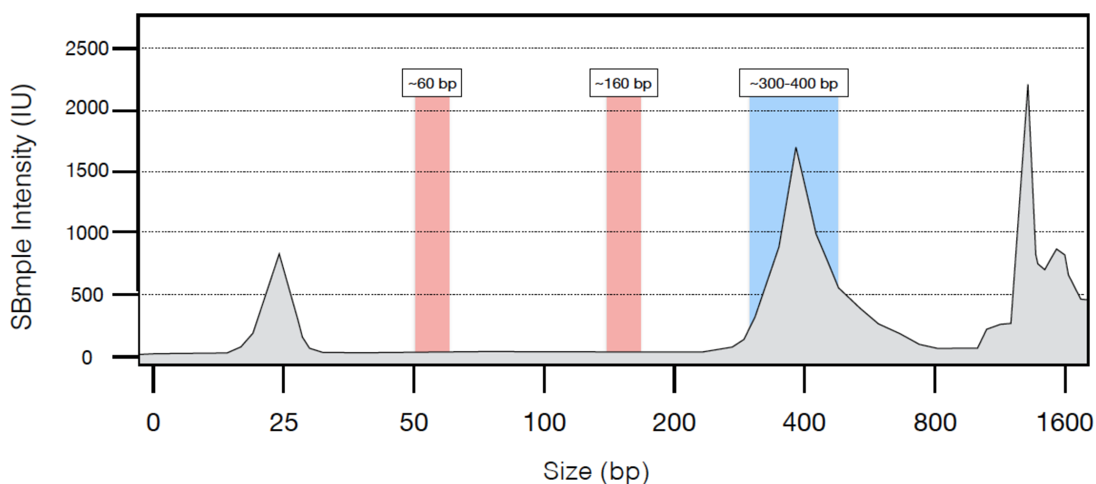

Figure SM3. Representative view of an ideal amplicon size profile from the Agilent Tapestation. Target amplicons are visible as a peak from 300-400 bp, while no peaks should be visible post size exclusion at 60bp and 160bp.

Quantification and Sequencing

Before sequencing, the size-selected samples (library) were quantified using the KAPA SYBR FAST ABI Prism qPCR kit (Illumina ABI, Cat#BCKA0004, see URLs). A universal master mix was prepared by mixing SYBR master mix (x2) with Library Quantification Primer Premix (x10) per manufacturer's instructions. Then a working master mix was made by mixing 12 µL of the universal master mix with 4 µl of nuclease-free water per well of the quantification plate. A 1:10 dilution was made by mixing 1 µL of stock PCR library with 9 µL of buffer EB. From this, 1 µL was taken and mixed with 99 µL of buffer EB to make the 1:1,000 dilution. Finally, 10 µL of the 1:1,000 dilution was mixed with 90 µL of buffer EB to give a 1:10,000 dilution. On the quantification plate, 16 µL of the working master mix was plated per well along with 4 µL of material to be quantified. Samples quantified were the 1:1,000 and 1:10,000 diluents of each multiplex, standards 0-6 from the KAPA qPCR kit, and a negative control of buffer alone (Figure SM4). All quantifications were done in triplicate wells. The quantification plate was run on a qPCR machine with absolute quantification and a standard curve setting. We performed thermocycling using the PCR program in Table SM9 and machine software was used to calculate relative quantities of each multiplex using the standard curve.

	1	2	3	4	5	6	7	8	9	10	11	12
B	StBndBrd 6	StBndBrd 6	StBndBrd 6	StBndBrd 5	StBndBrd 5	StBndBrd 5	StBndBrd 4	StBndBrd 4	StBndBrd 4	StBndBrd 3	StBndBrd 3	StBndBrd 3
B	StBndBrd 2	StBndBrd 2	StBndBrd 2	StBndBrd 1	StBndBrd 1	StBndBrd 1	StBndBrd 0	StBndBrd 0	StBndBrd 0	NegBtive Control	NegBtive Control	NegBtive Control
C	PFB_GRC1 1:1,000	PFB_GRC1 1:1,000	PFB_GRC1 1:1,000	PFB_GRC1 1:10,000	PFB_GRC1 1:10,000	PFB_GRC1 1:10,000						
D	PFB_GRC2 1:1,000	PFB_GRC2 1:1,000	PFB_GRC2 1:1,000	PFB_GRC2 1:10,000	PFB_GRC2 1:10,000	PFB_GRC2 1:10,000						
E	PFB_SPEC 1:1,000	PFB_SPEC 1:1,000	PFB_SPEC 1:1,000	PFB_SPEC 1:10,000	PFB_SPEC 1:10,000	PFB_SPEC 1:10,000						
F												
G												
H												

Figure SM4. Plate format for multiplex PCR quantification by qPCR for *Pf*.

Before loading prepared DNA onto the sequencing machine all multiplex PCR products were combined to a single pool of 4nM. To calculate the volume of each of the libraries/multiplexes to be added to the final pool we created a calculator in Microsoft Excel (SpotMalaria supplementary file 1, sheet “multiplex primer calculator”). In brief, using the concentration for the 1:1,000 dilution from the KAPA qPCR we calculated the average tag+target concentration per multiplex and used this to add relative quantities of each multiplex to the final pool. This 4nM library pool was used for sequencing on an Illumina MiSeq at the Wellcome Sanger Institute (see URLs) using the MiSeq v2 300 kit (Illumina, San Diego, USA, see URLs) using manufacturer’s instructions.

Table SM9 - KAPA qPCR Quantification procedure steps			
Step 1	95°C	15min	
Step 2	95°C	30sec	35 cycles
Step 3*	60°C	40sec	
Step 4	4°C	∞	

* Data acquisition step

Primer re-balancing

As the multiplex PCRs contain primers of various binding and amplification efficiencies it was necessary to adjust the input concentrations of each primer for use in subsequent runs to equalize the reads per amplicon. To calculate the new volumes of primers we analyzed a successful run performed with the equivolume pool. From this run we created a table of read counts (omitting

unaligned reads) of which the axes were tags (X) and targets (Y). This table was sorted first by tags then targets ensuring no target had zero reads across all tags which would have indicated a non-binding primer(s). The table was converted from raw read counts to target-read fractions, which was (number of reads for a tag+target) / (total reads for all targets for a single tag). Using the target-read fractions, we calculated the median read-fraction for each target. We scaled all median read-fractions by multiplying each median read-fraction by 1/sum of all median read-fractions. We calculated the target pool-weighting value as the scale value of the median read fraction raised to the power of -0.561 (pre-determined optimal value, unpublished). The minimum primer volume was placed at 1 µL to avoid errors associated with small volume pipetting, therefore all weighted values were divided by the minimum weight to scale the minimum volume to 1 µL. To avoid potential over-representation of primers in the new pool, we clipped all maximum values at 10 µL. For using the newly created variable volume pool in the multiplex PCR1, we estimated the central primer concentration of all primers in the new pool. To calculate the central value, we took the interquartile mean (IQM) and divided by the sum of all weights. We then used this value as the concentration to calculate the new volumes for the non-equivolume pool, shown below.

New_multiplex = (central value x 250µM) / 40nM = A
Pooled primers = 400 / A = B µl
Diluent (T0.1E) = 400 – B = C µl

Data Analysis

For amplicon sequencing, we used a series of open-source and custom analysis programs and scripts. To ensure future reproducibility across sites/institutions all analyses began with BCL files and the sequencing manifest from the Illumina MiSeq. The program bambi (see URLs) was used to convert the BCL directory into a BAM file, and was then used to separate reads into read groups based on the taglist. We used biobambam2 (see URLs) (bamadapterfind) to identify potential contamination by sequencing adaptors and then used samtools to split the BAM file into separate CRAM files by read group. Subsequent steps were carried out per CRAM files. Beginning with biobambam2, we collated and reset CRAM files into a pre-aligned state using bamcollate2 and bamreset. Using the output from bamadapterfind, we detected and removed adapters using bamadapterclip. We converted CRAM files to FASTQ files with bamtofastq in biobambam2 and aligned with BWA-MEM (see URLs) to a custom referenced which consists of all target amplicon regions in FASTA format. We then converted the resulting aligned SAM file into a BAM files and did a header replacement with samtools (see URLs) reheader and the FASTQ dictionary. Using biobambam2 again, we used bam12split to create single ranks and then merged using bam12uxmerger to create a single BAM again. On the single BAMs we performed a bamsort using

biobambam2. For genotyping we used BCFTools (see URLs) using the sorted BAM. We performed an mpileup to create a VCF using a predetermined set of loci which contained the list of typeable loci in amplicons used for targeted genotyping or all loci in regions used for *pfkelch13* sequencing and speciation. We performed two filters on the VCF files, the first with quality exclude rules of "%QUAL<15" and "MQ<20" and the second on depth with an exclude of "FORMAT/DP<8".

After genotyping, VCFs from the same sample were combined using the sequencing manifest tag list and genotypes failing VCF filters (low quality/low depth) were masked. Next, individual samples are assessed for sample quality using pass rates of genotype barcodes. In early tests, samples with low pass rates (<50%) in their barcodes SNPs were significantly more likely to produce spurious genotypes due most likely to PCR artefacts and low input DNA amounts. An exception was seen that speciation PCRs still performed when other loci failed and allowed us to identify samples that contained parasites (any species) but otherwise were unable to be genotyped. Any samples with less than 50% of their genotyping barcode able to be called were excluded from further analysis, and only speciation results were reported. Complexity of infection estimates using the genotype barcode were identical to those using Agena genotypes described above. For speciation, we aligned mitochondrial amplicons to both their *Pf* and *Pv* reference sequences and called genotypes. This was to account for sequence divergence across clades which in some cases caused read not to align. All loci variable between species were used to determine the probability of infection with a certain species. Any species whose reference sequence matched the informative genotyped loci at >95% of positions was deemed to be present in the infection.

Abbreviations

- WGS – whole-genome sequencing
- PEP – primer extension pre-amplification
- DBS – dried blood spot(s)
- sWGA – selective whole genome amplification

URLs

- QIAamp DNA Investigator Handbook: https://www.qiagen.com/gb/products/human-id-and-forensics/investigator-solutions/qiaamp-dna-investigator-kit/#resources;
- Qiagen: http://www.qiagen.com;
- Invitrogen: http://www.probes.com;
- Bioline: http://www.bioline.com;
- Sigma-Aldrich: https://www.sigmaaldrich.com;
- Agena Bioscience MassArray®: http://agenabio.com/products/massarray-system/
- ThermoFisher Scientific: https://www.thermofisher.com/uk/en/home.html
- Eurofins, UK: https://www.eurofinsgenomics.eu/en/custom-dna-sequencing.aspx
- DNASTAR INC: https://www.dnastar.com/software/lasergene/
- IDT: https://eu.idtdna.com/pages/products/custom-dna-rna/dna-oligos
- MAFFT, Multiple Sequence Alignment: https://mafft.cbrc.jp/alignment/server/
- Merck Millipore: https://www.merckmillipore.com/GB/en
- BIO RAD: https://www.bio-rad.com/
- Eppendorf: https://www.eppendorf.com/UK-en/
- VWR: https://uk.vwr.com/store/
- Beckman Coulter: https://mybeckman.uk
- Illumina: https://illumina.com
- Wellcome Sanger Institute: https://www.sanger.ac.uk/science/facilities
- Bambi: https://github.com/wtsi-npg/bambi
- Biobambam2: https://github.com/gt1/biobambam2
- Bcftools: https://samtools.github.io/bcftools/

List of Supplementary Figures and Tables

SpotMalaria Supplementary File 1: Targeted genotyping information

- *P. falciparum* barcode SNP list
- *P. vivax* barcode SNP list
- *P. falciparum* Agena primers
- *P. vivax* Agena primers
- *P. falciparum* assay amplicon sequencing primers
- *P. vivax* assay amplicon sequencing primers
- i7 multiplexing primers
- i5 multiplexing primers
- multiplex primer calculator

SpotMalaria Supplementary File 2: *Plasmodium* mitochondrial alignment for speciation assay design

SpotMalaria Supplementary File 3: *pfkelch13* reference gene sequence

References

1. Gabriel S, Ziaugra L, Tabbaa D: **SNP genotyping using the Sequenom MassARRAY iPLEX platform.** *Curr Protoc Hum Genet* 2009, **Chapter 2:**Unit 2 12.
2. Organization WGWH: **World Malaria Report 2018.** 2018.
3. Ariey F, Witkowski B, Amaratunga C, Beghain J, Langlois AC, Khim N, Kim S, Duru V, Bouchier C, Ma L, et al: **A molecular marker of artemisinin-resistant Plasmodium falciparum malaria.** *Nature* 2014, **505:**50-55.
4. Ashley EA, Dhorda M, Fairhurst RM, Amaratunga C, Lim P, Suon S, Sreng S, Anderson JM, Mao S, Sam B, et al: **Spread of artemisinin resistance in Plasmodium falciparum malaria.** *N Engl J Med* 2014, **371:**411-423.
5. Miotto O, Amato R, Ashley EA, MacInnis B, Almagro-Garcia J, Amaratunga C, Lim P, Mead D, Oyola SO, Dhorda M, et al: **Genetic architecture of artemisinin-resistant Plasmodium falciparum.** *Nat Genet* 2015, **47:**226-234.
6. Frosch AE, Laufer MK, Mathanga DP, Takala-Harrison S, Skarbinski J, Claassen CW, Dzinjalamala FK, Plowe CV: **Return of widespread chloroquine-sensitive Plasmodium falciparum to Malawi.** *J Infect Dis* 2014, **210:**1110-1114.
7. Laufer MK, Thesing PC, Eddington ND, Masonga R, Dzinjalamala FK, Takala SL, Taylor TE, Plowe CV: **Return of chloroquine antimalarial efficacy in Malawi.** *N Engl J Med* 2006, **355:**1959-1966.
8. Wellems TE, Plowe CV: **Chloroquine-resistant malaria.** *J Infect Dis* 2001, **184:**770-776.
9. Djimde A, Doumbo OK, Cortese JF, Kayentao K, Doumbo S, Diourte Y, Coulibaly D, Dicko A, Su XZ, Nomura T, et al: **A molecular marker for chloroquine-resistant falciparum malaria.** *N Engl J Med* 2001, **344:**257-263.
10. Fidock DA, Nomura T, Talley AK, Cooper RA, Dzekunov SM, Ferdig MT, Ursos LM, Sidhu AB, Naude B, Deitsch KW, et al: **Mutations in the P. falciparum digestive vacuole transmembrane protein PfCRT and evidence for their role in chloroquine resistance.** *Mol Cell* 2000, **6:**861-871.
11. Foote SJ, Kyle DE, Martin RK, Oduola AM, Forsyth K, Kemp DJ, Cowman AF: **Several alleles of the multidrug-resistance gene are closely linked to chloroquine resistance in Plasmodium falciparum.** *Nature* 1990, **345:**255-258.
12. Veiga MI, Dhingra SK, Henrich PP, Straimer J, Gnadig N, Uhlemann AC, Martin RE, Lehane AM, Fidock DA: **Globally prevalent PfMDR1 mutations modulate Plasmodium falciparum susceptibility to artemisinin-based combination therapies.** *Nat Commun* 2016, **7:**11553.
13. Sidhu AB, Verdier-Pinard D, Fidock DA: **Chloroquine resistance in Plasmodium falciparum malaria parasites conferred by pfcrt mutations.** *Science* 2002, **298:**210-213.
14. Dhingra SK, Gabryszewski SJ, Small-Saunders JL, Yeo T, Henrich PP, Mok S, Fidock DA: **Global Spread of Mutant PfCRT and Its Pleiotropic Impact on Plasmodium falciparum Multidrug Resistance and Fitness.** *MBio* 2019, **10**.
15. Marfurt J, de Monbrison F, Brega S, Barbollat L, Muller I, Sie A, Goroti M, Reeder JC, Beck HP, Picot S, Genton B: **Molecular markers of in vivo Plasmodium vivax resistance to amodiaquine plus sulfadoxine-pyrimethamine: mutations in pvdhfr and pvmdr1.** *J Infect Dis* 2008, **198:**409-417.
16. Suwanarusk R, Russell B, Chavchich M, Chalfein F, Kenangalem E, Kosaisavee V, Prasetyorini B, Piera KA, Barends M, Brockman A, et al: **Chloroquine resistant**

Plasmodium vivax: in vitro characterisation and association with molecular polymorphisms. *PLoS One* 2007, **2:**e1089.
17. Price RN, Uhlemann AC, Brockman A, McGready R, Ashley E, Phaipun L, Patel R, Laing K, Looareesuwan S, White NJ, et al: **Mefloquine resistance in Plasmodium falciparum and increased pfmdr1 gene copy number.** *Lancet* 2004, **364:**438-447.
18. Sidhu AB, Valderramos SG, Fidock DA: **pfmdr1 mutations contribute to quinine resistance and enhance mefloquine and artemisinin sensitivity in Plasmodium falciparum.** *Mol Microbiol* 2005, **57:**913-926.
19. Venkatesan M, Gadalla NB, Stepniewska K, Dahal P, Nsanzabana C, Moriera C, Price RN, Martensson A, Rosenthal PJ, Dorsey G, et al: **Polymorphisms in Plasmodium falciparum chloroquine resistance transporter and multidrug resistance 1 genes: parasite risk factors that affect treatment outcomes for P. falciparum malaria after artemether-lumefantrine and artesunate-amodiaquine.** *Am J Trop Med Hyg* 2014, **91:**833-843.
20. Malmberg M, Ferreira PE, Tarning J, Ursing J, Ngasala B, Bjorkman A, Martensson A, Gil JP: **Plasmodium falciparum drug resistance phenotype as assessed by patient antimalarial drug levels and its association with pfmdr1 polymorphisms.** *J Infect Dis* 2013, **207:**842-847.
21. Nsobya SL, Kiggundu M, Nanyunja S, Joloba M, Greenhouse B, Rosenthal PJ: **In vitro sensitivities of Plasmodium falciparum to different antimalarial drugs in Uganda.** *Antimicrob Agents Chemother* 2010, **54:**1200-1206.
22. Reed MB, Saliba KJ, Caruana SR, Kirk K, Cowman AF: **Pgh1 modulates sensitivity and resistance to multiple antimalarials in Plasmodium falciparum.** *Nature* 2000, **403:**906-909.
23. Veiga MI, Ferreira PE, Jornhagen L, Malmberg M, Kone A, Schmidt BA, Petzold M, Bjorkman A, Nosten F, Gil JP: **Novel polymorphisms in Plasmodium falciparum ABC transporter genes are associated with major ACT antimalarial drug resistance.** *PLoS ONE* 2011, **6:**e20212.
24. Amato R, Lim P, Miotto O, Amaratunga C, Dek D, Pearson RD, Almagro-Garcia J, Neal AT, Sreng S, Suon S, et al: **Genetic markers associated with dihydroartemisinin-piperaquine failure in Plasmodium falciparum malaria in Cambodia: a genotype-phenotype association study.** *Lancet Infect Dis* 2016.
25. Witkowski B, Duru V, Khim N, Ross LS, Saintpierre B, Beghain J, Chy S, Kim S, Ke S, Kloeung N, et al: **A surrogate marker of piperaquine-resistant Plasmodium falciparum malaria: a phenotype-genotype association study.** *Lancet Infect Dis* 2016.
26. Hamilton WL, Amato R, van der Pluijm RW, Jacob CG, Quang HH, Thuy-Nhien NT, Hien TT, Hongvanthong B, Chindavongsa K, Mayxay M, et al: **Evolution and expansion of multidrug-resistant malaria in southeast Asia: a genomic epidemiology study.** *Lancet Infect Dis* 2019, **19:**943-951.
27. Ross LS, Dhingra SK, Mok S, Yeo T, Wicht KJ, Kumpornsin K, Takala-Harrison S, Witkowski B, Fairhurst RM, Ariey F, et al: **Emerging Southeast Asian PfCRT mutations confer Plasmodium falciparum resistance to the first-line antimalarial piperaquine.** *Nat Commun* 2018, **9:**3314.
28. Agrawal S, Moser KA, Morton L, Cummings MP, Parihar A, Dwivedi A, Shetty AC, Drabek EF, Jacob CG, Henrich PP, et al: **Association of a Novel Mutation in the Plasmodium falciparum Chloroquine Resistance Transporter With Decreased Piperaquine Sensitivity.** *J Infect Dis* 2017, **216:**468-476.
29. Peterson DS, Walliker D, Wellems TE: **Evidence that a point mutation in dihydrofolate reductase-thymidylate synthase confers resistance to**

pyrimethamine in falciparum malaria. *Proc Natl Acad Sci U S A* 1988, **85:**9114-9118.
30. Gregson A, Plowe CV: **Mechanisms of resistance of malaria parasites to antifolates.** *Pharmacol Rev* 2005, **57:**117-145.
31. Picot S, Olliaro P, de Monbrison F, Bienvenu AL, Price RN, Ringwald P: **A systematic review and meta-analysis of evidence for correlation between molecular markers of parasite resistance and treatment outcome in falciparum malaria.** *Malar J* 2009, **8:**89.
32. Sirawaraporn W, Sathitkul T, Sirawaraporn R, Yuthavong Y, Santi DV: **Antifolate-resistant mutants of Plasmodium falciparum dihydrofolate reductase.** *Proc Natl Acad Sci U S A* 1997, **94:**1124-1129.
33. Jacob CG KD, Miotto O: **Genetic reconnaissance in the Greater Mekong Subregion to support malaria control and elimination.** *TBD* 2019.
34. Chang HH, Wesolowski A, Sinha I, Jacob CG, Mahmud A, Uddin D, Zaman SI, Hossain MA, Faiz MA, Ghose A, et al: **Mapping imported malaria in Bangladesh using parasite genetic and human mobility data.** *Elife* 2019, **8**.
35. Baniecki ML, Faust AL, Schaffner SF, Park DJ, Galinsky K, Daniels RF, Hamilton E, Ferreira MU, Karunaweera ND, Serre D, et al: **Development of a single nucleotide polymorphism barcode to genotype Plasmodium vivax infections.** *PLoS Negl Trop Dis* 2015, **9:**e0003539.
36. Pearson RD, Amato R, Auburn S, Miotto O, Almagro-Garcia J, Amaratunga C, Suon S, Mao S, Noviyanti R, Trimarsanto H, et al: **Genomic analysis of local variation and recent evolution in Plasmodium vivax.** *Nat Genet* 2016, **48:**959-964.
37. Malaria GENPfCP: **Genomic epidemiology of artemisinin resistant malaria.** *Elife* 2016, **5**.
38. Galinsky K, Valim C, Salmier A, de Thoisy B, Musset L, Legrand E, Faust A, Baniecki ML, Ndiaye D, Daniels RF, et al: **COIL: a methodology for evaluating malarial complexity of infection using likelihood from single nucleotide polymorphism data.** *Malar J* 2015, **14:**4.
39. Chang HH, Worby CJ, Yeka A, Nankabirwa J, Kamya MR, Staedke SG, Dorsey G, Murphy M, Neafsey DE, Jeffreys AE, et al: **THE REAL McCOIL: A method for the concurrent estimation of the complexity of infection and SNP allele frequency for malaria parasites.** *PLoS Comput Biol* 2017, **13:**e1005348.
40. Katoh K, Standley DM: **MAFFT multiple sequence alignment software version 7: improvements in performance and usability.** *Mol Biol Evol* 2013, **30:**772-780.
41. Zhang L, Cui X, Schmitt K, Hubert R, Navidi W, Arnheim N: **Whole genome amplification from a single cell: implications for genetic analysis.** *Proc Natl Acad Sci U S A* 1992, **89:**5847-5851.
42. Malaria Genomic Epidemiology N, Malaria Genomic Epidemiology N: **Reappraisal of known malaria resistance loci in a large multicenter study.** *Nat Genet* 2014, **46:**1197-1204.
43. Shen Z, Qu W, Wang W, Lu Y, Wu Y, Li Z, Hang X, Wang X, Zhao D, Zhang C: **MPprimer: a program for reliable multiplex PCR primer design.** *BMC Bioinformatics* 2010, **11:**143.
44. Djimde A ANA, Miles A, Craig A, Dondorp A, Newbold C, Neafsey D, Conway D, McVean G, Awandare G, O'Brien J, Zhu J, Amenga-Etego L, Berriman M, White NJ, Day N, Miotto O, Pearson R, Fairhurst R, Amato R, Maciuca S, Otto T, Iqbal Z, Kwiatkowski DP: **Pf3K.** (MalariaGEN ed., v5 edition. Hinxton, U.K.; 2016.
45. Oyola SO, Ariani CV, Hamilton WL, Kekre M, Amenga-Etego LN, Ghansah A, Rutledge GG, Redmond S, Manske M, Jyothi D, et al: **Whole genome sequencing of**

Plasmodium falciparum from dried blood spots using selective whole genome amplification. *Malar J* 2016, **15:**597.
46. Bronner IF, Quail MA, Turner DJ, Swerdlow H: **Improved Protocols for Illumina Sequencing.** *Curr Protoc Hum Genet* 2014, **80:**18 12 11-42.