SUPPLEMENTARY METHODS

MIP capture, amplification, and sequencing

Oligonucleotides (described in Supplementary Table 1) were synthesized as 200 nmole ultramers (Integrated DNA Technologies) with equimolar hand-mix option for random bases. The oligonucleotides were pooled at equal concentrations to create the MIP panel. Pools were 5' phosphorylated using 1 µl (10 units) T4 polynucleotide kinase (NEB) for every nmole of probe, in 1X T4 DNA ligase buffer (NEB). Phosphorylation reactions were split into 50 µl aliquots, incubated in a thermocycler at 37°C for 45 min, followed by heat inactivation at 65°C for 20 min. Split reactions were pooled together for homogeneity, re-aliquoted, and kept at -20°C. Probes were diluted 1:8 in TE buffer (10 mM Tris, 1 mM EDTA, pH 8) to bring them to 1 µM working solution.

Capture reactions were carried out as follows. 10 µl capture reactions for each sample containing Ampligase buffer (1X), Phusion DNA polymerase (0.0008 units/µl), Ampligase (0.04 units/µl), pooled MIPs (40 nM), dNTP (4 µM), and template DNA (5µl) were incubated in a preheated thermocycler at 95°C for10 min, 60°C for 1 h, and then 4°C. Next, 2 µl of exonuclease mix containing 1X Ampligase buffer, 10 units exonuclease I, and 50 units exonuclease III were added to reactions with incubations at 37°C for 1 h, 95°C for 2 min, and then 4°C.

The entire capture reaction (12 µl) was amplified in a 25 µl PCR reaction containing: 1X Phusion polymerase buffer, 1X Macromolecular Crowding (MMC) solution, 200 nM each dNTP, 0.02 units/µl Phusion DNA polymerase, and 500 nM each forward and reverse primer [1]. PCR was performed using a preheated thermocycler at 98°C for 30 s, 22 cycles (98°C 10 s, 63°C 30 s, 68°C 30 s), 68°C for 2 min, and then 4°C. 50 ml 5X MMC was prepared by mixing the following components in water and filter-sterilizing using a 0.2 µ nylon syringe filter: 3.75 g Ficoll 70 (GE Healthcare), 1.25 g Ficoll 400 (Sigma), and 0.125 g polyvinylpyrrolidone (Sigma).

Next, library pools were created by combining 5 µl of each PCR reaction in a single tube followed by clean-up and concentration using Ampure XP beads (Beckman Coulter) at 0.8x bead:DNA ratios using the manufacturer's protocol. This generally removed unwanted adapter/primer dimers < 200 bp. If dimers remained after bead

clean-up, the eluted DNA was loaded on a 1.5% agarose gel and the relevant band (~500 bp) was extracted from the gel using the Monarch DNA extraction kit (NEB). Libraries were sequenced on an Illumina Nextseq 550 instrument using 150 bp paired end sequencing with dual indexing using Nextseq 500/550 Mid-output Kit v2.

MIP data processing and variant calling

Sequencing data were processed using MIPWrangler software (v1.1.1-dev, github.com/bailey-lab/MIPWrangler) in combination with other software. Briefly, sequences were demultiplexed by their dual sample barcodes using bcl2fastq software (v2.20.0.422, Illumina). Paired end reads were then stitched together using MIPWrangler and filtered on expected length and per base quality scores, discarding a sequence if the fraction of quality scores > 30 was < 70%. Quality filtered stitched sequences were then further demultiplexed by target using the extension and ligation arm sequences to produce a file for each target for each sample. Target sequences for each sample were then corrected using their unique molecular identifiers (UMIs) by clustering sequences on their UMIs and creating a consensus sequence for each. This procedure removes PCR errors that occur in late cycles, including polymerase stutter and subsequent sequencing errors. UMI corrected sequences were then further clustered within MIPWrangler using the qluster algorithm derived from SeekDeep [2], allowing accurate detection of single base differences and indels at minor strain frequency of 1% or less. We set a minimum relative abundance threshold of 0.5% for a cluster to be included in final analysis. Variant calls were carried out using freebayes (v1.3.1) [3] following alignment to the reference genome with bwa (v0.7.17) [4]; outputs were annotated using snpEff software (v4.3t) [5]. Variants were filtered using MIPTools (v0.19.12.13) such that the variant site for a given sample had at least 5 UMI coverage and the non-reference allele had at least 2 UMIs supporting the call.

Copy numbers of *pfmdr1* and *plasmepsin-2/3* were estimated based on average depth of coverage for these genes using the cnv_caller module of MIPTools software (v0.20.07.28). A basic algorithm was used to normalize the UMI count data: 1) filter samples with sufficient UMI coverage for a given gene (minimum 25 UMI per probe on average), 2) sample normalize UMI counts based on the mean of all MIPs in the

sample to make samples with different total coverage comparable, 3) remove specific MIPs with high variability, 4) normalize the depth of each MIP across samples to the mean. For each sample, the average value of all MIPs for a given gene was considered the copy number for that gene. Average values were rounded to the nearest integer for discrete copy number calls.

References

[1] Aydemir O, Janko M, Hathaway NJ, et al. Drug-Resistance and population structure of plasmodium falciparum across the democratic Republic of Congo using high-Throughput molecular inversion probes. J Infect Dis. 2018; 218(6):946–955.

[2] Hathaway, N. J., Parobek, C. M., Juliano, J. J. & Bailey, J. A. SeekDeep: single-base resolution de novo clustering for amplicon deep sequencing. Nucleic Acids Res. (2017) doi:10.1093/nar/gkx1201.

[3] Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing. arXiv preprint arXiv:1207.3907 [q-bio.GN] 2012

[4] Li H. (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997v1 [q-bio.GN]

[5] Cingolani P, Platts A, Wang le L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin). 2012;6(2):80-92. doi:10.4161/fly.19695