

Sanger sequencing and deconvolution of polyclonal infections: a quantitative approach to monitor drug-resistant *Plasmodium falciparum*



Hamma Maiga,^{a,b,c,*} Robert D. Morrison,^{a,c} and Patrick E. Duffy^a

^aLaboratory of Malaria Immunology and Vaccinology (LMIV), National Institute of Allergy and Infectious Diseases (NIAID), National Institutes of Health (NIH), 29 Lincoln Drive, Bethesda, MD, 20892, USA

^bInstitut National de Santé Publique (INSP), Ministère de la Santé et du Développement Social (MSDS), Bamako, BP: 1771, Mali



Summary

Background Anti-malarial drug resistance in *Plasmodium falciparum* is a major public health problem in malaria-endemic regions. Although various technical improvements in sequencing methods have been introduced to identify SNPs, the conventional approach with current tools does not discriminate mixed infections, and thus can be improved for more sensitive surveillance of anti-malarial resistance to better inform control strategies.

Methods We developed a computational approach for deconvolution of chromatograms generated by standard Sanger sequencing of PCR amplicons in order to quantify molecular marker variants of anti-malarial drug resistance genes [*Plasmodium falciparum* dihydropteroate synthase (*Pf**dhps*) and *P. falciparum* dihydrofolate reductase (*Pf**dhfr*)]. We validated this computational approach using mixtures of V1/S and FCR3 at varying proportions between 0 and 100%, then applied it to field samples collected in Doneguebougou, Mali in 2018. We determined the mean fraction of resistance alleles in individual samples, as well as the prevalence of infections carrying resistant parasites.

Findings We observed a highly significant correlation between the predicted and measured proportions of V1/S and FCR3 alleles in mixed laboratory samples (all $p < 0.001$). Among field samples, the mean fraction of resistant *Pf**dhps* alleles was 4.7% 431V, 95.9% 436F/A, 49.9% 437G, 0.0% 540E, 1.2% 581G and 1.5% 613S/T; corresponding prevalences were 50.0%, 100%, 72.5%, 0.0%, 25.0%, and 12.5%, respectively. The mean fraction of resistant *Pf**dhfr* alleles was 0.6% 16V, 11.1% 50R, 89.0% 51I, 98.3% 59R, 74.7% 108T/N, 8.6% 140L and 8.7% 164L; corresponding prevalences were 12.5%, 75.0%, 100%, 100%, 100%, 50.0%, and 28.6%, respectively. We identified two new point mutations on the *Pf**dhps* gene at codons D484T and D545N.

Interpretation Computational deconvolution of sequencing chromatograms can discriminate varying proportions of antimalarial drug-sensitive versus -resistant alleles. This cost-effective and quantitative variant-sequencing approach will be useful for population-based surveys that characterize mixed infections at the individual level to survey known and unknown mutations in *P. falciparum* drug-resistance genes.

Funding This work was supported by the Division of Intramural Research of the National Institute of Allergy and Infectious Diseases, National Institutes of Health (NIH). HM was supported by the African Postdoctoral Training Initiative (APTI) Fellowship program jointly managed by the US NIH, The African Academy of Sciences (AAS) and Bill & Melinda Gates Foundation (BMGF); Grant Reference Number: APTI-18-01.

Copyright © 2024 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Keywords: Malaria; Drug resistance; Molecular marker quantification; Sequence deconvolution

Introduction

In 2020, an estimated 241 million cases of malaria resulting in 627,000 deaths were reported worldwide, with over 90% of malaria cases and deaths occurring in

Africa.¹ Artemisinin-based combination therapies, the recommended first-line treatment in all countries in which malaria is endemic, have contributed considerably toward reducing malaria burden.² However, the

*Corresponding author. Institut National de Santé Publique, Ministère de la Santé et du Développement Social, Bamako, BP: 1771, Mali.

E-mail address: hmaiga@icermali.org (H. Maiga).

^cThese authors contributed equally.

eBioMedicine

2024;103: 105115

Published Online 17 April 2024

<https://doi.org/10.1016/j.ebiom.2024.105115>

Research in context**Evidence before this study**

Molecular markers are used in epidemiological surveillance to monitor the emergence and spread of drug-resistant pathogens. These markers are comprised of either single nucleotide polymorphisms (SNPs) or concatenated SNPs in microbial genes that interact with drugs. Polyclonal malaria infections often present a mixed genotype at the loci of interest and are highly prevalent in malaria-endemic areas, particularly in Africa and Southeast Asia. In high transmission settings where polyclonal infections have been widely reported, mixed genotypes have often been binarily categorized as resistant, since any level of drug resistance might impact treatment response. To improve surveillance of antimalarial resistance control strategies, we sought an affordable solution that can quantify the proportions of all observed genotypes independently at each mutation site. Deconvolution of conventional Sanger sequencing chromatograms could provide the necessary quantification. We did multiple PubMed and Google literature searches, starting in 2019, using search terms including: deconvolution, Sanger, chromatogram. Search results revealed that some attempts at interpreting mixtures in Sanger chromatograms have been developed previously, but none properly address the needs for surveillance of drug-resistant genotypes. Some solutions, such as TRACY, deconvolute complete chromatograms into exactly 2 alleles. Other software, such as BCV, deconvolute complete chromatograms into user pre-specified indel variants. We found no reports of software solutions that focus on independently quantifying mixed genotype calls at multiple single amino acid codon sites within Sanger sequences in a simple standalone manner.

Added value of this study

Our study describes a robust deconvolution method for Sanger sequencing data that is a reliable, relatively less

expensive, and higher throughput for genotyping mutations associated with drug resistance. We targeted malaria genes *Plasmodium falciparum* dihydropteroate synthase (*Pfdhps*) and *P. falciparum* dihydrofolate reductase (*Pfdhfr*) genes in this proof-of-concept study, but the deconvolution computational approach is suitable for any protein from any species, being both gene and organism agnostic. This deconvolution method is unique in that the atomic units being modeled are the 3-base codons that code for amino acids, rather than either the entire chromatogram or just single nucleotides. The result of this deconvolution method is the relative percentages of observed protein amino acids, exactly the information needed for detailed epidemiological surveillance. Our results demonstrate quantitative discrimination of varying proportions of sensitive versus resistant alleles within samples. Compared to the binary call approach, the mean fraction or proportion determined by the deconvolution method may yield a more accurate estimate of drug-resistant parasites as a proportion of the total parasite reservoir in the community, as well as the heterogeneity of these proportions between individuals and across key groups such as children and pregnant women.

Implications of all the available evidence

Our computational approach to deconvolution of Sanger chromatograms at single amino acid resolution expands the set of tools for quantification of mixed mutations in polyclonal infections. This capability will improve population-based surveys to better characterize mixed infections as mean fraction or proportion at the individual level and may provide an earlier signal to public health officials of emerging resistance to better prepare for treatment policy changes.

emergence of artemisinin resistance poses a serious threat to malaria control worldwide.

Plasmodium falciparum, the most deadly species of malaria, displays a wide variety of resistance mechanisms in the field. Current malaria control methods rely on prevention, by minimizing exposure to mosquitoes or presumptive administration of antimalarial drugs to vulnerable groups. In endemic areas where large numbers of malaria cases are presently inevitable, anti-malarial drugs are relied upon to treat the disease. *P. falciparum* resistance has been reported against every anti-malarial drug currently in use in many parts of the world.³

Molecular markers are used in epidemiological surveillance to monitor the emergence and spread of drug-resistant pathogens. These markers are comprised of either single nucleotide polymorphisms (SNPs) or concatenated SNPs in microbial genes that interact with

drugs. Since its introduction in the 1970s, nucleic acid sequencing has allowed delineation of DNA sequences with extraordinary accuracy, and has undergone many important modifications and improvements.⁴ Improvements include longer reads (up to 1000 bp per analysis), better sequence accuracy, improved sensitivity with thermocycling protocols, full automation, higher speed, and substitution of radioactivity with fluorescent and other probes. Although various technical improvements in current sequencing methods have been introduced, including ultra-thin gel technology⁵ and capillary electrophoresis,⁶ throughput and cost of these technologies limit their use for large-scale surveillance.

P. falciparum resistance to different drugs has been associated with several alleles of *Pf* genes. *P. falciparum* multi drug resistance 1 (*Pfmdr1*) 86Y-184Y and *P. falciparum* chloroquine resistance transporter gene (*Pfcr1_72C-73V-74I-75E-76T* haplotype) are known as the

strongest amodiaquine and chloroquine resistance markers.^{7,8} Mutations in *Pfmdr1* have been associated with reduced susceptibility to mefloquine and lumefantrine⁹ particularly *Pfmdr1* Y184F mutation associated with lumefantrine resistance.¹⁰ In a genome-wide association study, the E415G mutation in the *P. falciparum* exonuclease gene (*Pfexo_E415G*) was associated with *ex-vivo* piperazine IC50 of parasite isolates from Cambodia. Molecular markers such as mutations in *P. falciparum* dihydrofolate reductase (*Pfdhfr*) and *P. falciparum* dihydropteroate synthetase (*Pfdhps*) are used as surveillance tools for resistance to sulfadoxine-pyrimethamine (SP).^{11,12} Resistance to pyrimethamine has been associated with numerous mutation combinations in the *Pfdhfr* gene, such as the *Pfdhfr* 511-59R-108N/T-164L genotypes¹³ while substitutions at codons 431, 436, 437, 540, 581 and 613 of the *Pfdhps* gene may confer resistance to sulfadoxine.¹⁴ Sulfadoxine-pyrimethamine (SP) remains in use for intermittent preventive treatment in pregnant women (IPTp)¹⁵ or infants (IPTi),¹⁶ seasonal malaria chemoprevention (SMC),¹⁷ and acute malaria treatment.¹⁸ Monitoring of *Pfdhfr* and *Pfdhps* mutations is less complicated in geographic areas where genetic complexity of *P. falciparum* is relatively low. In most of sub-Saharan Africa (SSA), however, not only do DHFR and DHPS mutations occur in many different combinations, but most infections are “polyclonal”,¹⁹ meaning multiple clonal populations of parasites are present in a single blood isolate. Patients are often infected with more than one distinct parasite strain (referred to as mixed infection, multiple infection, or complexity of infection), due to bites from multiple mosquitoes, mosquitoes carrying multiple genetic types, or a combination of both.²⁰ Mixed infections can lead to competition among co-existing strains and may influence disease development,²¹ transmission rates²² and the spread of drug resistance.²³

Polyclonal infections are often mixed at the loci of interest (e.g., individual infections carry some parasites with mutations at specific DHFR or DHPS codons, and some parasites without mutations at those codons). Mixed infections of different genotypes are highly prevalent in malaria-endemic areas, particularly in Africa and Southeast Asia.^{24–26} In high transmission settings where polyclonal infections have been reported in greater than 50% of sampled isolates,²⁷ isolates with mixed genotypes have often been classified as mutant-resistant, since any level of mutant resistance might impact treatment response.^{28–30} In some studies, polyclonal clinical samples were not included for calculation of allele prevalence and reconstruction of the various haplotypes.³¹

Many attempts at interpreting mixtures in Sanger chromatograms have been developed previously, but none properly address the needs for surveillance of drug-resistant genotypes. TRACY³² is a C++ command line application with a user-friendly suite of web-based

front end apps that facilitate basecalling, alignment, trace decomposition, variant calling, and trace assembly. It aligns a single chromatogram of interest to a genome or FASTA sequence to identify SNPs and indels, and decomposes the given chromatogram into exactly 2 allelic sequences with an estimated allelic fraction for each overall sequence. TRACY was not designed to independently deconvolute polyclonal infections at single amino acid codon resolution where each parasite may have distinct blends of drug resistance mutations. SnpEff³³ is a program for annotating and predicting the effect of SNPs on a genome. It requires a reference genome, and operates on variant call data more typical of Next Generation Sequencing, to predict amino acid mutations. It has no support for Sanger chromatogram data, and makes no attempt at deconvolution of mixed variants. Base-Calling with Vocabulary (BCV)³⁴ is a software method that computationally deciphers Sanger chromatograms obtained from mixed DNA samples. Its stated main goal is to resolve mixed traces due to indels, and requires as input a pre-defined dictionary of possible sequences containing the indels expected to be observed. Deconvolution in BCV is to identify entire sequences, and makes no effort to quantify the relative proportions observed at individual amino acid codons. Other analytical tools, such as DECODR,³⁵ TIDE,³⁶ ICE³⁷ and CRISP-ID³⁸ are aimed at identifying CRISP-induced indels from Sanger sequence data, but again not designed for deconvoluting single amino acid codons. MultiEditR³⁹ quantifies RNA-editing events in Sanger sequences, again focused on detecting indels instead of quantifying amino acid codon mutations. TraceTrack⁴⁰ provides for batch processing of multiple Sanger sequences to generate mutation calls against a reference sequence, but does not support deconvolution to quantify multiple codon calls within a single sequence. To date, no software solution exists that focuses on independently quantifying mixed genotype calls at multiple single amino acid codon sites within a Sanger sequence in a standalone manner.

To improve surveillance of antimalarial resistance control strategies, we implemented an approach that quantifies known and previously unknown markers of resistance in individual samples, based on the conventional Sanger sequencing chromatogram and advanced informatics tools. This approach allows the proportion of different genotypes to be estimated independently at each mutation site. We targeted *Pfdhfr* and *Pfdhps* genes in this proof-of-concept study, but the deconvolution computational approach is suitable for any protein from any species, being both gene and organism agnostic.

Methods

Parasite samples

P. falciparum strain V1/S is an in vitro, culture-adapted clone of the V1 strain originating in Vietnam, which

shows resistance to pyrimethamine and sulfadoxine. *P. falciparum* strain FCR-3/FMG (Gambia), also referred to as FCR-3/Gambia, was originally isolated in 1976 from the blood of a human patient collected in The Gambia, West Africa. FCR-3/FMG (Gambia) was identified as resistant to chloroquine *in vitro* and *in vivo* after nearly four years of continuous culture without chloroquine pressure, and is also reported as cycloguanil-resistant and pyrimethamine-susceptible.^{41–46} *Pfdhfr* and *Pfdhps* mutations for these parasite strains are summarized in Table 1. These two different parasites V1/S and FCR3/FMG were mixed in different ratios to obtain samples ranging from 0%/100%–100%/0%, respectively, in 11 mixtures with 10% successive increases/decreases in proportions (Table 2). One hundred percent corresponds to the individual strain that was not mixed. These mixture samples were generated and used for validation of the deconvolution methods.

Field samples for molecular analysis were collected from adults at Donegoubougou site, Mali in 2018 during the observational study “Community Dynamics of Malaria Transmission and Mosquito Feeding in Bancoumana and Donegoubougou, Mali” ([clinicaltrials.gov](https://clinicaltrials.gov/ct2/show/study/NCT03304704) ID NCT03304704). One mL of venous blood was collected in EDTA anticoagulant tubes among participants, and parasitized blood samples were stored at –80 °C until analysis.

DNA extraction

QIAamp DNA blood mini kits (Qiagen, Valencia, CA USA) were used according to the manufacturer’s instructions to extract genomic DNA from 200 µL EDTA whole blood which was then eluted in 60 µL of buffer AE (10 mM Tris-Cl and 0.5 mM EDTA). The concentration and purity of DNA were quantified using a NanoDrop2000 UV (Thermo Fischer, Wilmington, DE, USA).

PCR and sequencing

Pfdhfr and *Pfdhps* fragments were amplified by PCR. Briefly, Dhfr_Fwd (5’TTTATATTTTCCTTTTAA3’) and Dhfr_Rev (5’CATTTTATTATTCGTTTCT3’) oligonucleotides were used for single amplification of *Pfdhfr* sequences, producing a 681 bp fragment containing key mutations N51I, C59R and I164L. For *Pfdhps* sequences, dhps_Fwd (5’GGAATATTAATGTAAATTATGATTCT3’) and dhps_Rev (5’CGTCATGAACCTTAT-TAGATCTACCT3’) were used for single amplification,

producing a 986 bp fragment containing mutations S436 F/A, A437G and A613 S/T, previously associated with parasite responses to sulfonamides (e.g., sulfadoxine, sulfamethoxazole, and sulfathiazole).

The reaction comprised 5 µl of DNA, 0.5 µl of each primer (dhfr_Forward or dhps_Fwd and dhfr_Reverse or dhps_Rev with final concentration 0.5 µmol/L each), 25 µl of Q5 High/Hot-Fidelity 2X Master Mix and 29 µl of Nuclease-free Water. The reaction was cycled as follows: 5 min at 95 °C and then 44 repeated cycles of 30 s at 92 °C, 30 s at 45 °C and 45 s at 65 °C followed by 15 min at 72 °C. PCR products were purified by QIAamp DNA blood mini kits and concentration was measured using a NanoDrop2000 UV. The PCR products were sequenced by Eurofins Genomics Company.

Drug resistance calls

Mutation results can be reported by either of two different methods that impact the interpretation of drug resistance studies. The quantitative method (mean fraction of resistant allele in individual samples) gives the numeric proportion for the mutation of interest, while the conventional method (prevalence of resistant parasites in a population of samples) is instead based on binary calls, where any amount of detection of resistant allele labels the entire sample resistant, while a sample is only called susceptible if it has a total absence of any amount of resistant allele. Because the conventional method omits the fractional specificity of mixed infections, these two reporting methods may produce divergent findings.

Chromatogram deconvolution

Quantification by deconvolution of resistance mutation sites within a chromatogram used a multistep workflow, implemented as part of the DuffyTools R package (<https://github.com/robertdouglassmorrison/DuffyTools>), and visualized as a flowchart in Fig. 1. First, the raw binary ABI chromatogram file was imported into R using the sangerseqR package from Bioconductor,^{47,48} to extract the numeric trace matrix of signal intensities for all 4 base channels along the entire extent of the chromatogram, as well as the list of nucleotide base calls and their peak center locations within the trace matrix. The base call sequence was saved in both its original and reverse complement forms and translated into all 6 reading frames of possible protein sequence. The collection of DNA and amino acid calls, spectral peak center locations, and the full trace matrix of intensities in all 4 channels constituted the starting dataset for all subsequent mutation quantification steps. To facilitate automated assessment of the chromatogram sequence, each base call was augmented with a confidence score metric, similar to the Phred quality Q scores⁴⁹ that uses peak height, shape and uniqueness across the 4 channels to differentiate clean base calls from chromatogram noise. All the above steps are encapsulated into the DuffyTools function *loadChromatogram()* (Figure S1).

	<i>Pfdhfr</i>			<i>Pfdhps</i>		
	51	59	164	436	437	613
V1/S (resistant)	I	R	L	F	G	T
FCR3 (sensitive)	N	C	I	S	A	A

I, Isoleucine, R, Arginine, L, Leucine, N, Asparagine, C, Cysteine, S, Serine, A, Alanine, G, glycine, T, Threonine, F, Phenylalanine.

Table 1: Summary of the amino acid polymorphisms of dhfr and dhps *P. falciparum* V1/S and FCR3 strains.

	V1/S	M1	M2	M3	M4	M5	M6	M7	M8	M9	FCR3
Parasite mixtures (fractions)	1	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1	0
	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1

M, mixture of V1/S and FCR3 at different level test.

Table 2: Fractions of *P. falciparum* V1/S and FCR3 in strain mixtures (M1-M9).

The next step was to locate and isolate the small region of the chromatogram that contains the mutation site of interest, in a manner that tolerates the presence of unknown mutations that might be present in field parasites. Evaluations using mixtures of lab line parasites suggested that an 11 amino acid fragment motif, centered on the mutation site of interest, from the 3D7 reference protein, gave the best performance for finding the correct region of interest across a wide variety of chromatograms, proteins, and mutation sites. For example, to find the A581G mutation site in a chromatogram from *Pfdhps*, the motif 'IGLGF~~A~~KKHDQ' provided 5aa of hopefully conserved flanking context on both sides of the mutation of interest. Given this protein fragment motif, the raw chromatogram was searched in all 6 reading frames using the *pairwiseAlignment()* function from the Biostrings⁵⁰ package, to find the best scoring sequence match location while tolerating amino acid mismatches. Using that location and reading frame information, a small chromatogram subset was extracted from the original raw data containing only the 11

amino acid region that was found, in the correct strand and reading frame, discarding all extraneous portions of the trace matrix, peak locations, and base calls; positioning the 3-base codon of interest for the wanted mutation site at the exact center of this new smaller chromatogram. All the above methodology is encapsulated into DuffyTools function *motifSubsetChromatogram()* (Figure S2).

The final deconvolution quantification step was to model and measure the mutation site. First, the raw trace matrix of the observed 3bp DNA codon at the mutation site was extracted from the center of the motif chromatogram, using function *motifCodonChromatogram()*. Next, the collection of candidates 3bp codons for inclusion in the model was generated by gathering all possible codons that code for up to 3 different amino acids: the reference amino acid, the expected alternative mutation amino acid, and the observed chromatogram-called amino acid at the codon site. Each of these candidate codons was modeled as an idealized trace matrix containing 3 gaussian shaped synthetic

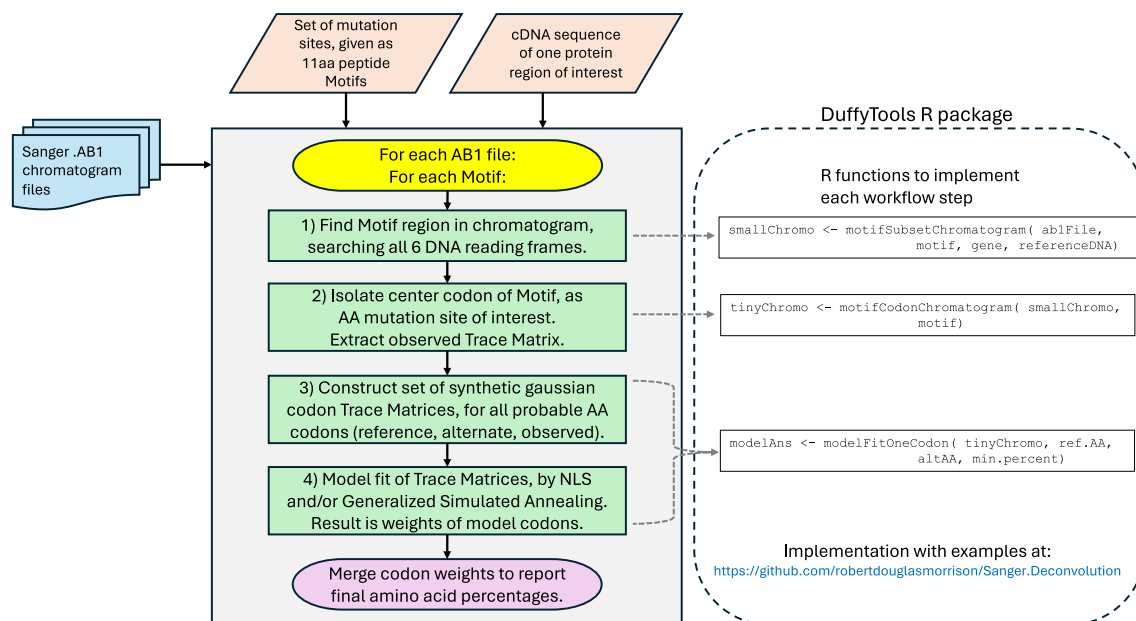


Fig. 1: Flowchart of deconvolution approach. The workflow of the deconvolution method is shown both descriptively, showing data inputs and processing steps to produce the mutation proportion calls, and as the actual R function calls to implement the workflow on an example chromatogram.

nucleotide peaks, using peak width and separation parameters measured from the observed chromatogram. This effectively makes a collection of minute synthetic trace matrices representing the signal intensity profile each candidate codon would generate (Figure S3). For example, the candidate ATG codon for methionine is constructed as a synthetic trace matrix with one fixed width gaussian peak in the 1st third of the A channel, a fixed width gaussian peak in the 2nd third of the T channel, a fixed width gaussian peak in the 3rd third of the G channel, and no peaks anywhere in the C channel. Lastly, these candidate codons and the observed chromatogram codon were passed to a model fitting step, using both a Nonlinear Least Squares (NLS)⁵¹ optimization method and a simulated annealing optimization method using the GenSA package.⁵² NLS was the preferred model fit algorithm, but because the observed data are sparse (only 3 DNA bases of spectral peaks), may contain noisy malformed peak artifacts, and the number of candidate codons might be large, the NLS method might fail to successfully converge. If NLS failed, the GenSA method was used instead, being more robust but at the expense of longer computational time and potentially returning a less optimal fit. The goal of the model fit is to explain the observed 3-nucleotide trace matrix as a linear additive blend of all the synthetic candidate codon trace matrices, adjusting the non-negative amplitude of each candidate codon as needed to minimize the residual error between the one observed and the sum of all synthetic trace matrices. The result of the model fit step is the subset of candidate codons having non-zero signal amplitudes (weights), as the best explanation of the observed codon chromatogram (Figure S4). All candidate codons with zero weights are discarded, and only those codons having non-zero weights were translated back to amino acids, scaling their weights proportionally to sum to 100%, giving the final proportion of amino acid calls that best model the observed raw data (Fig. 2). The methodology intentionally includes all possible codons in the model, and lets the model fit determine which small subset of codons best explain the observed data. Any attempt to down-select potential codons for inclusion prior to the model fit step would be based on either user choice or inspecting the observed data, both of which would introduce potential bias into the model. All the above deconvolution methodology and model fitting is encapsulated into the DuffyTools function `modelFitOneCodon()`.

Algorithm availability

The entire multistep algorithmic approach has been packaged into publicly available R markdown scripts suitable for RStudio⁵³ that can process a single chromatogram or batch process large sets of chromatograms. See <https://github.com/robertdouglassmorrison/Sanger.Deconvolution>. The GitHub page contains all

scripts, instructions, a clear description of the algorithm, example input chromatograms, and example result files in both HTML and CSV formats. Clone or download using GitHub's Code menu button.

Ethics

The study protocol was approved by the ethical committees of the Faculty of Medicine and Odontostomatology (FMOS)/Faculty of Pharmacy (FAPH)/University of Sciences, Techniques and Technologies of Bamako (USTTB) of Mali and US-NIH (clinicaltrials.gov ID NCT03304704). Community permission was obtained before to the study. Individual, written, informed consent was obtained of each volunteer prior to screening.

Statistical analysis

Field samples were selected randomly with blinding from parasite-positive visits. Data were collected on sequencing printout and entered into a Microsoft Excel™ database 2010 (Microsoft Inc., Washington, WA). Statistical analysis was performed using RStudio (R version 3.3.0). The mix at different proportion from V1/S to FCR3 was calculated with 95% Confidence interval [CI]. Logistic regression model was used for comparisons as appropriate with statistical significance set at p-value <0.05.

Role of funders

The funders had no role in study design, data collection, data analyses, interpretation, writing of the manuscript, or the decision to submit for publication.

Results

Laboratory samples

We validated the approach using mixtures of V1/S and FCR3 at varying proportions between 0 and 100% in increments/decrements of 10%. We interpreted the results as the mean fraction of resistance alleles in all samples or as the prevalence of samples carrying any detectable *Pfdhps* and *Pfdhfr* resistance genes.

Mean fraction and prevalence of *Pfdhps* alleles

As expected, the proportion of individual resistance alleles decreased from V1/S and M1 to M9 and FCR3, for S436F (Figure S5a), A437G (Figure S5b) and A613 S/T (Figure S5c) that followed the inverse order. The mean fraction of V1/S for resistant *Pfdhps* alleles was 60.3% 436F/A, 60.3% 437G and 47.6% 613S/T, respectively (Figure S6a). When applying the binary call approach that is common in field research, the resistance mutant prevalence was estimated at 91.0% for each codon (Figure S6b). We observed a significant correlation between the measured fraction of resistant *Pfdhps* alleles 436F/A, 437G and 613S/T versus the predicted fraction based on the mixed volumes ($p < 0.001$, Fig. 3). The

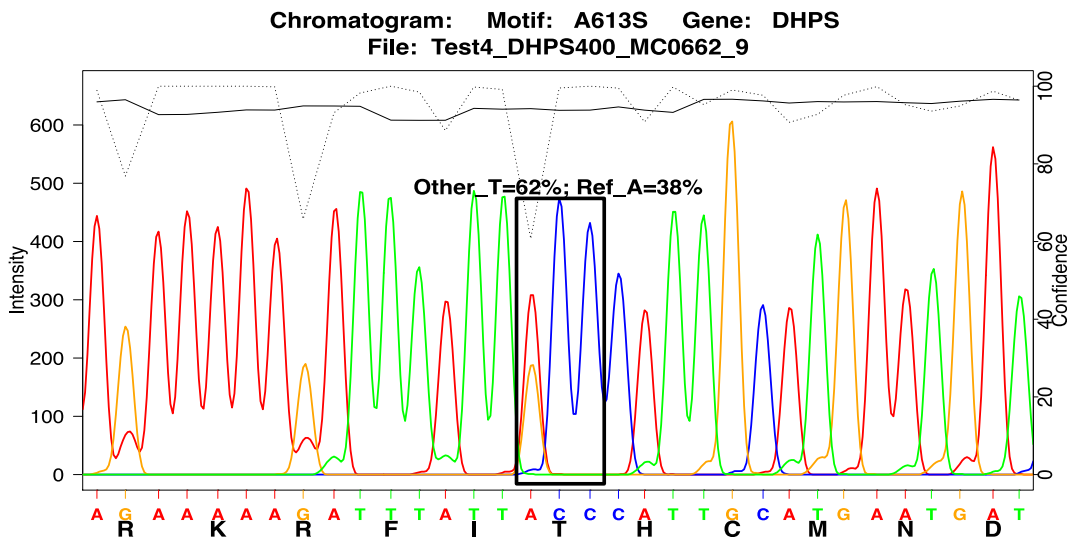


Fig. 2: Deconvolution model fitting result of *Pfdhps* A613S T. The final deconvolution model fit amino acid call for *Pfdhps* site A613S, visualized on the motif subset chromatogram. In this example from M4 mixture of V1S and FCR3 (Table 2), the method estimates 62% of DNA encodes Thr (ACC) and 38% encodes Ala (GCC), consistent with expected 60/40%.

results showed 100% of the *Pfdhps* K540 sensitive and *Pfdhfr* 108T/N resistant (data not shown).

and the measured proportion of resistant *Pfdhfr* alleles ($p < 0.001$, Fig. 4).

Mean fraction and prevalence of Pfdhfr alleles

The proportion of individual sensitive alleles of *Pfdhfr* in a) N511 b) C59R; and c) I164L was decreased from FCR3 and M9 to M1 and V1/S that followed the inverse order (Figure S7). The mean fraction of V1/S for resistant *Pfdhfr* alleles was 66.6% 51I, 62.0% 59R and 61.7% 164L, respectively (Figure S8a). The total prevalence of V1/S presented 91.0% for each resistant *Pfdhfr* alleles (51I, 59R and 164L) by binary call approach (Figure S8b). We observed a strong linear correlation between the proportion predicted by the mixed volumes

Comparison against published deconvolution tool TRACY

To evaluate how the method differs from existing deconvolution tools, we submitted the chromatogram of one of the DHPS lab mixtures (60% V1/S, 40% FCR3, as shown in Fig. 2, and Fig. S1 through Fig. S4) to the Indigo deconvolution tool of the TRACY³² suite. TRACY requires a reference DNA sequence as a comparator, so DHPS from *P.falciparum* reference isolate 3D7 was given. The result returned by TRACY declared that the chromatogram was best explained as being 64% reference 3D7 and 35% an alternative sequence that contained a

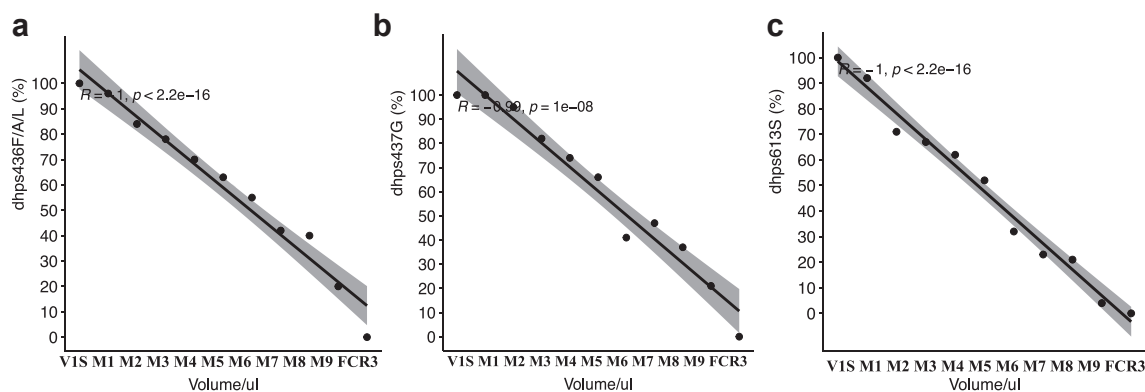


Fig. 3: Correlation of the quantified (%) versus predicted (Volume/ul) mutant allele of *Pfdhps* in mixture of V1S and FCR3 parasites. Correlation of the phenotype resistance proportion with mixture volume. V1/S and FCR3 parasites represent 100% mutant and sensitive, respectively. (3a) Mutant allele of *Pfdhps* 436F/A proportion, (3b) mutant allele of *Pfdhps* 437G proportion and (3c) mutant allele of *Pfdhps* 613S proportion.

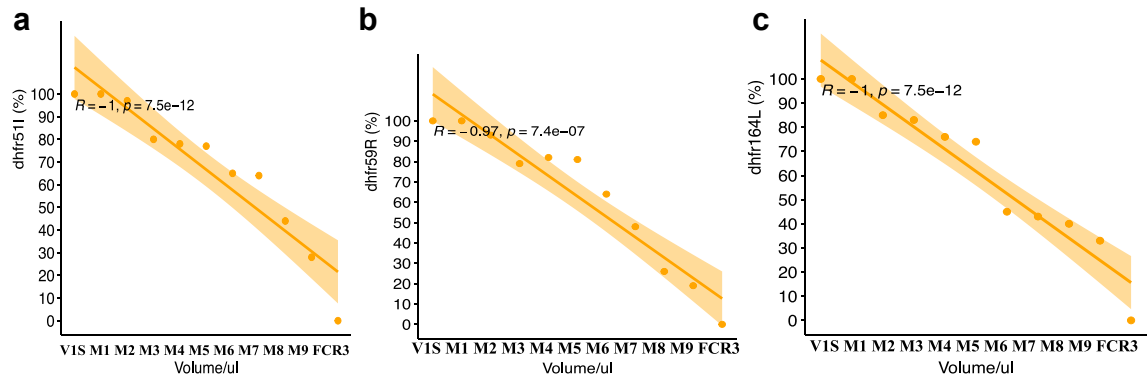


Fig. 4: Correlation of the quantified (%) versus predicted (Volume/ul) mutant allele of *Pfdhfr* in mixture of V1S and FCR3 parasites. Correlation between mixture volume and phenotype resistance proportion. (4a) Mutant allele of *Pfdhfr* 511 proportion, (4b) mutant allele of *Pfdhfr* 59R proportion and (4c) mutant allele of *Pfdhfr* 164L proportion.

mashup of various mutations from both V1/S and FCR3. At A613 S/T, where the known mutation call is 60%T/40%A, TRACY called 64%A/35%T. At mutation site S436F, where the known mutation call is 60%F/40%S, TRACY called 64%S/35%F. Overall, TRACY's proportion magnitudes were close to correct, but the major/minor variant was inverted in both cases. Furthermore, the TRACY output reveals it was incorporating different signal proportions at 18 separate nucleotide locations along the high quality portion of the chromatogram, to average all those differences into exactly 2 global proportions that describe the entire chromatogram. TRACY was unable to separately evaluate each mutation site of interest, and was confounded by requiring inclusion of every other mutation and chromatogram noise detected along the region into its result, regardless of their role (or lack thereof) in drug resistance.

Field samples

We applied the quantitative method on eight field samples to determine the mean fraction of resistance alleles among human isolates, as well as their prevalence defined as any detectable *Pfdhps* and *Pfdhfr* resistance genes.

Mean fraction and prevalence of *Pfdhps* alleles

The mean fraction of resistant *Pfdhps* alleles was 4.7% 431V, 95.9% 436F/A, 49.9% 437G, 0.0% 540E, 1.2% 581G and 1.5% 613S/T, respectively (Fig. 5a). By the binary call approach, the prevalence of mutant was 50.0%, 100%, 72.5%, 0.0%, 25.0% and 12.5% for the same resistant *Pfdhps* alleles, respectively (Fig. 5b). Mixed alleles (wild-type and mutant) were identified in 50.0%, 25.0%, 37.5%, 0.0%, 25.0% and 12.5% of samples at different *Pfdhps* loci (I431V, S436 F/A, A437G, K540E, A581G, and A613 S/T), respectively (Figure S9).

Mean fraction and prevalence of *Pfdhfr* alleles

The mean fraction of resistant *Pfdhfr* alleles was 0.6% 16V, 11.1% 50R, 89.0% 51I, 98.3% 59R, 74.7% 108T/N,

8.6% 140L and 8.7% 164L, respectively (Fig. 6a). The prevalence of mutant was 12.5%, 75.0%, 100%, 100%, 100%, 50.0% and 28.6% for the same resistant *Pfdhfr* alleles, respectively (Fig. 6b). Mixed alleles (wild-type and mutant) were identified in 12.5%, 75.0%, 50.0%, 28.6%, 100%, 50.0% and 28.6% of samples at different *Pfdhfr* loci (A16V, C50R, N51I, C59R, S108 T/N, V140L and I164L), respectively (Figure S10).

Discussion

DNA-sequencing is the gold standard for assessing molecular resistance to antimalarial drugs⁵⁴ and proportions of *P. falciparum* genotypes are of major importance for epidemiological surveys. However, the conventional method using binary calls for antimalarial resistance alleles within samples fails to characterize the proportion of a particular genotype in a mixed sample. Pyrosequencing is a quantitative method with a superior limit of detection (proportion of resistant alleles) compared to Sanger sequencing.⁵⁵ However, the pyrosequencing method requires the presence of at least 30% mutant DNA in mixed samples to detect *Pfdhfr* at codon 164L and uses many primers with high cost per sample.⁵⁶ For example, the cost estimate for pyrosequencing is \$2.28 per SNP⁵⁷ and for restriction fragment length polymorphism (RFLP) is \$6.90 per SNP.⁵⁸ Our approach enables estimation of the relative burden of individual molecular markers of SP drug resistance, and will be more cost-effective and sensitive to detect the emergence and spread of known and unknown resistance alleles which threaten public health.

In this paper, we report a bioinformatic approach to quantify the molecular markers of SP using sequence chromatograms, as a cost-effective alternative to guide decision-makers in the face of spreading resistance to anti-malarial drugs. Commonly used tools yield a binary or categorical call to define the molecular markers of antimalarial resistance, and do not allow quantitative assessment of variants within mixed infections. To

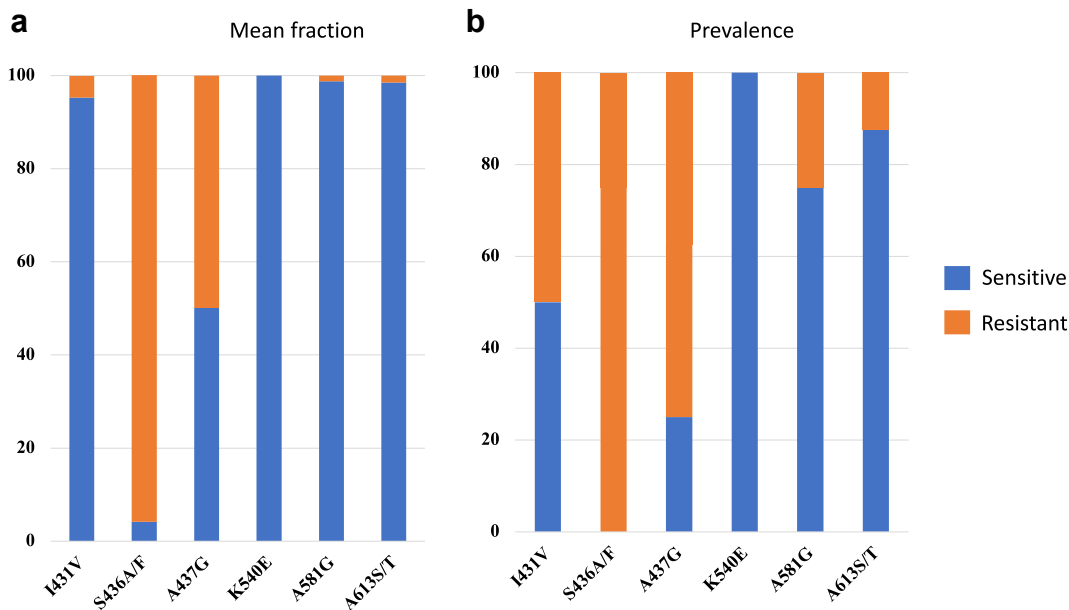


Fig. 5: Prevalence and Mean fraction of *Pf dhps* alleles at codons 431, 436, 437, 540, 581, and 613 of field samples. Presents a binary call result (S = sensitive; R = resistance); mean fraction presented in panel (5a), and prevalence presented in panel (5b).

improve surveillance of antimalarial resistance control strategies, our approach allows the proportion of different genotypes to be estimated from Sanger sequencing chromatograms. We evaluated the performance of the method to correctly estimate the true proportion of different alleles, first using laboratory

parasite mixtures with known fractions of resistant and wild-type alleles.

Using fixed ratios of laboratory parasites, our study demonstrated the ability of our deconvolution method to detect a minor allele at a proportion as low as 10%. As expected, the standard approach that assigns a binary

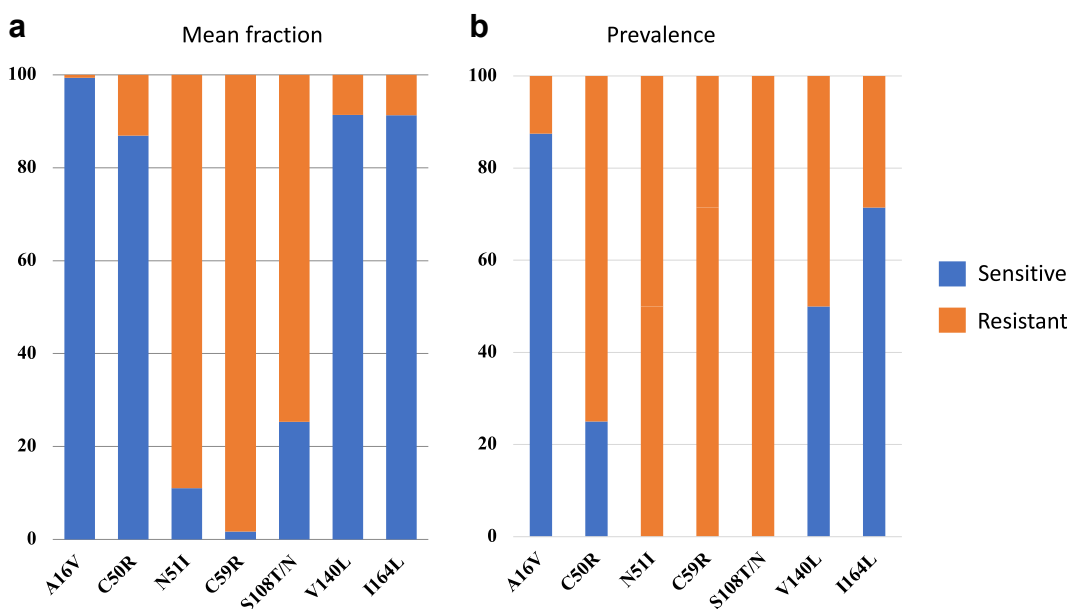


Fig. 6: Prevalence and mean fraction alleles of *Pf dhfr* at codons 16, 50, 51, 59, 108, 140 and 164 of field samples. Presents a binary call result (S = sensitive; R = resistance); mean fraction presented in panel (6a), and prevalence presented in panel (6b).

call would have designated all of our laboratory parasite mixtures as resistant at all *Pfdrhps* and *Pfdrhfr* loci,^{46,47} while the mean fraction of resistance markers using the quantitative approach was nearly 50% for all loci. We observed a highly significant correlation between the known proportion of each allele and the proportion measured by sequencing and deconvolution at all loci.

After qualifying our methods on laboratory samples, we applied them to field samples collected in Mali in 2018. In the field samples, the mean fraction of resistant *Pfdrhps* alleles (431V, 436F/A, 437G, 540E, 581G, and 613S/T) and of resistant *Pfdrhfr* alleles (16V, 50R, 51I, 59R, 108T/N, 140L and 164L) demonstrate quantitative discrimination of varying proportions of sensitive versus resistant alleles within samples. Polyclonal infection is a common phenomenon in malaria-endemic areas; Diakité et al. showed a very high genetic diversity in *P. falciparum* isolates in Mali consistent with previous findings throughout Africa.^{59,60} Complexity of infection has been associated with intensity of malaria transmission.^{61,62}

Our deconvolution method is based on standard sequencing while being more rapid, cost-effective by sample for all mutations (\$10 for six or seven SNPs we studied in the genes), and sensitive (down to at least a 10% proportion of the resistant allele in our sample mixture studies) compared to pyrosequencing and restriction fragment length polymorphism (RFLP). A pyrosequencing protocol was developed as a rapid and reliable method to identify the mutations of *Pfdrhfr* (at codons 16, 51, 59, 108, 140, and 164) and *Pfdrhps* (at codons 431, 436, 437, 540, 581 and 613) genes of *P. falciparum*,⁵⁵ but used ten and twelve primers to amplify *Pfdrhfr* and *Pfdrhps* genes, respectively, compared to four primers used in our deconvolution method for the same genes and all mutations. RFLP is a common method used for each SNP, using 32 primers for *Pfdrhfr* (at codons 16, 51, 59, 108, 140, and 164) and 24 primers for *Pfdrhps* (at codons 431, 436, 437, 540, 581 and 613) gene amplification. Pyrosequencing and RFLP methods are more time-consuming compared to the deconvolution method.

Zhou et al. noted that pyrosequencing and RFLP methods showed similar levels of sensitivity in detecting the mutant forms of *Pfdrhfr* SNPs, except for codon 164 in mixed samples.⁵⁵ Also, pyrosequencing required the presence of at least 30% mutant DNA in mixed samples to detect the 164L mutation, while RFLP was sensitive enough to detect this mutation with as little as 15% mutant DNA. Repeated RFLP experiments further escalated the cost of reagents for this procedure. The deconvolution method was sensitive enough to detect all mutations at proportions as low as 10%. The reagents used in the deconvolution method were relatively inexpensive compared to those for RFLP and pyrosequencing. RFLP was at least threefold more expensive than pyrosequencing in reagent costs alone, and reagent costs for our deconvolution method are inexpensive compared to those for pyrosequencing.

Our findings are limited to analysis of *Pfdrhps* and *Pfdrhfr* genes, above all the K540E and S108 T/N mutations that are most important in sub-Saharan Africa. While we did not analyze *Pfcrtr*, *Pfmdr1*, or *PfK13* genes in this study, we have developed the same method for other Pf genes (e.g., *MDR*, *CRT*, *CYTB*, *HRP2*, among others), and our method is not limited to “pairs of genes” but could be used for any number. Our pilot study was limited to a relatively small number of field samples, but also includes data from mixtures of laboratory isolates to demonstrate the linear range of quantitation. Further, resistance can be related to changes in the copy number of genes encoding the drug target or transport proteins involved in drug influx/efflux.⁶³ Our approach has not been designed to assess copy number.

The deconvolution method has the benefit that it uses a single sequencing reaction and can detect any potential new mutations that have not been previously described within the PCR-amplified region. Existing published methods such as TRACY fail to deconvolute results at single mutation resolution. For example, during our study we have discovered two new point mutations on the *Pfdrhps* gene at codons D484T and D545N. We are interested in future to assess any role for these identified SNPs in antimalarial resistance. This advantage of the deconvolution method equally pertains to resistance markers for other antimalarial drugs. An important future application could be to monitor artemisinin resistant parasites as they spread across Africa including emerging variants of the kelch 13 gene.⁶⁴

Other benefits of this deconvolution method derive from its flexibility. It is organism agnostic, taking only the reference DNA sequence of the expected PCR-amplified region and a list of amino acid motifs as the only required inputs other than the chromatograms. The list of motifs is fully user specified, so any coding region of any protein can be evaluated by this method. Empirical simulations using several *P. falciparum* cDNA sequences suggested that 11aa was an ideal motif length, providing enough local context around the mutation site of interest even when multiple nearby residues were also mutated, such that the search in all 6 reading frames always finds at most a single high scoring match to the motif. Shorter or longer motifs can be used, with the caveat that too short a motif may identify multiple non-unique matches across the 6 reading frames, while a too long motif may extend past the limits of the chromatogram sequence or span possible indel regions of the protein and thus fail to identify a single high scoring match. The only cause for this deconvolution method to fail is when the motif cannot be found in any of the 6 reading frames. The two most likely reasons a motif fails to be found are: noisy chromatograms without clear singular peak calls in the region of interest, or that the isolate being sequenced is highly mutated in that region such that the expected

motif is not present. The user is fully free to adjust the motif length or sequence for each mutation site of interest. This use of user-specified amino acid motifs allows the deconvolution method to locate the exact mutation sites of interest anywhere in the chromatogram in a fully automated method.

While sequencing capabilities in Africa have been hindered by logistical and infrastructure limitations in the past, many labs in Africa now have sequencing platforms, making this technology accessible for local scientists. Our technique offers practical advantages and is relatively easy to implement for surveillance of drug-resistant *P. falciparum* markers in the malaria elimination context. We know that molecular markers associated with antimalarial drug resistance, when validated, are invaluable to monitor geospatial distribution of resistant parasites in real time. The prevalence of molecular markers of drug resistance is often a good indicator of the level of clinical resistance.⁶³

Next Generation Sequencing (NGS) is a leading choice for many types of sequence variant discovery, but less ideal for this goal of easy low-cost drug-resistance surveillance. NGS is more expensive than Sanger sequencing, and requires more highly specialized technicians, equipment, and computational infrastructure. Individual NGS reads are discrete, having only ACGT fixed calls. Any attempt to quantify proportions of alleles requires aligning large numbers of reads to a reference genome to build consensus pileups and then using variant calling software to detect the relative proportions of each base call. One distinct advantage of NGS methods is in detecting the phasing of nearby mutations. For mutations that are closer together than the length of the NGS reads, by inspecting individual reads one could count how often particular mutation pairs were seen in the same DNA fragment. In our Sanger chromatogram approach, phasing can only be roughly estimated from the amino acid proportions.

In conclusion, the present study describes a deconvolution method that is a reliable, relatively less expensive, and higher throughput alternative for genotyping mutations of *Pf*dhfr and *Pf*dhps associated with drug resistance. Our results demonstrate quantitative discrimination of varying proportions of sensitive versus resistant alleles within samples. Compared to the binary call approach, the mean fraction determined by the deconvolution method may yield a more accurate estimate of drug-resistant parasites as a proportion of the total parasite reservoir in the community, as well as the heterogeneity of these proportions between individuals and across key groups such as children and pregnant women. This relatively simple approach for population-based surveys that characterizes mixed infections at the individual level may provide an earlier signal of emerging resistance to better prepare public health officials for treatment policy changes.

Contributors

H.M. and P.E.D. conceived the study. H.M. conducted the laboratory study and wrote the first draft of manuscript, which was revised by P.E.D. for initial submission. R.D.M. designed and implemented the chromatogram deconvolution algorithms, performed statistical data analysis, and edited the manuscript. All authors accessed and verified the underlying data. All authors reviewed, edited, and approved the final manuscript version after review, and agreed to the published version of the manuscript.

Data sharing statement

All data are included in the manuscript, figures, and supplemental data, and are available at the time of publication, from the authors upon reasonable request.

Declaration of interests

We declare no competing interests.

Acknowledgements

We thank Mr. J. Patrick Gorres of the National Institute of Health (NIH) and NIH Fellows Editorial Board for assistance in reviewing and editing the manuscript. We thank Ms. Holly Torano and Ms. Jillian Neal for technical support.

Appendix A. Supplementary data

Supplementary data related to this article can be found at <https://doi.org/10.1016/j.ebiom.2024.105115>.

References

- 1 World Health Organization. World malaria report. In: *World health organization*. Geneva: WHO; 2022.
- 2 Bhatt S, Weiss DJ, Cameron E, et al. The effect of malaria control on *Plasmodium falciparum* in Africa between 2000 and 2015. *Nature*. 2015;526:207–211.
- 3 Nsanzabana C, Djalle D, Guérin PJ, Ménard D, González IJ. Tools for surveillance of anti-malarial drug resistance: an assessment of the current landscape. *Malar J*. 2018;17:75.
- 4 Mulyasmita W, Christopher MM, Sarah H. Protein-Engineered Biomaterials: Synthesis and characterization. In: *Comprehensive Biomaterials II*. 2017:18–40.
- 5 Bharaj BS, Angelopoulou K, Diamandis EP. Rapid sequencing of the p53 gene with a new automated DNA sequencer. *Clin Chem*. 1998;44:1397–1403.
- 6 Carrilho E. DNA sequencing by capillary array electrophoresis and microfabricated array systems. *Electrophoresis*. 2000;21:55–65.
- 7 Ecker A, Lehane AM, Clain J, Fidock DA. PfCRT and its role in antimalarial drug resistance. *Trends Parasitol*. 2012;28:504–514.
- 8 Baro NK, Callaghan PS, Roepe PD. Function of resistance conferring *Plasmodium falciparum* chloroquine resistance transporter isoforms. *Biochemistry*. 2013;52:4242–4249.
- 9 Sidhu ABS, Valderramos SG, Fidock DA. pfmndr1 mutations contribute to quinine resistance and enhance mefloquine and artemisinin sensitivity in *Plasmodium falciparum*. *Mol Microbiol*. 2005;57:913–926.
- 10 Taylor AR, Flegg JA, Holmes CC, et al. Artemether-lumefantrine and dihydroartemisinin-piperaquine exert inverse selective pressure on *Plasmodium falciparum* drug sensitivity-associated haplotypes in Uganda. *Open Forum Infect Dis*. 2017;4:ofw229.
- 11 Ahmed A, Bararia D, Vinayak S, et al. *Plasmodium falciparum* isolates in India exhibit a progressive increase in mutations associated with sulfadoxine-pyrimethamine resistance. *Antimicrob Agents Chemother*. 2004;48:879–889.
- 12 Amato R, Lim P, Miotto O, et al. Genetic markers associated with dihydroartemisinin-piperaquine failure in *Plasmodium falciparum* malaria in Cambodia: a genotype-phenotype association study. *Lancet Infect Dis*. 2017;17:164–173.
- 13 Gregson A, Plowe CV. Mechanisms of resistance of malaria parasites to antifolates. *Pharmacol Rev*. 2005;57:117–145.
- 14 Wang P, Lee CS, Bayoumi R, et al. Resistance to antifolates in *Plasmodium falciparum* monitored by sequence analysis of dihydropteroate synthetase and dihydrofolate reductase alleles in a large number of field samples of diverse origins. *Mol Biochem Parasitol*. 1997;89:161–177.

- 15 Briand V, Cottrell G, Massougbojji A, Cot M. Intermittent preventive treatment for the prevention of malaria during pregnancy in high transmission areas. *Malar J*. 2007;6:160.
- 16 World Health Organization. *World malaria report*. Geneva: World Health Organization; 2009.
- 17 World Health Organization. *WHO Policy Recommendation: seasonal Malaria Chemoprevention (SMC) for Plasmodium falciparum malaria control in highly seasonal transmission areas of the Sahel sub-region in Africa*. World Health Organization; 2012.
- 18 World Health Organization. *Global technical strategy for malaria control 2016–2030*. Geneva: World Health Organization; 2015.
- 19 Färnert A, Rooth I, Svensson, Snounou G, Björkman A. Complexity of Plasmodium falciparum infections is consistent over time and protects against clinical disease in Tanzanian children. *J Infect Dis*. 1999;179:989–995.
- 20 Zhu SJ, Almagro-Garcia J, McVean G. Deconvolution of multiple infections in Plasmodium falciparum from high throughput sequencing data. *Bioinforma Oxf Engl*. 2018;34:9–15.
- 21 de Roode JC, Pansini R, Cheesman SJ, et al. Virulence and competitive ability in genetically diverse malaria infections. *Proc Natl Acad Sci U S A*. 2005;102:7624–7628.
- 22 Arnot D. Unstable malaria in Sudan: the influence of the dry season. Clone multiplicity of Plasmodium falciparum infections in individuals exposed to variable levels of disease transmission. *Trans R Soc Trop Med Hyg*. 1998;92:580–585.
- 23 de Roode JC, Culleton R, Bell AS, Read AF. Competitive release of drug resistance following drug treatment of mixed Plasmodium chabaudi infections. *Malar J*. 2004;3:33.
- 24 Lee SA, Yeka A, Nsobia SL, et al. Complexity of Plasmodium falciparum infections and antimalarial drug efficacy at 7 sites in Uganda. *J Infect Dis*. 2006;193:1160–1163.
- 25 Bendixen M, Msangeni HA, Pedersen BV, Shayo D, Bødker R. Diversity of Plasmodium falciparum populations and complexity of infections in relation to transmission intensity and host age: a study from the Usambara Mountains, Tanzania. *Trans R Soc Trop Med Hyg*. 2001;95:143–148.
- 26 Kobbe R, Neuhoﬀ R, Marks F, et al. Seasonal variation and high multiplicity of first Plasmodium falciparum infections in children from a holoendemic area in Ghana, West Africa. *Trop Med Int Health*. 2006;11:613–619.
- 27 Certain LK, Briceño M, Kiara SM, Nzila AM, Watkins WM, Sibley CH. Characteristics of Plasmodium falciparum dhfr haplotypes that confer pyrimethamine resistance, Kilifi, Kenya, 1987–2006. *J Infect Dis*. 2008;197:1743–1751.
- 28 Maiga H, Lasry E, Diarra M, et al. Seasonal malaria chemoprevention with sulphadoxine-pyrimethamine and amodiaquine selects Pfdhfr-dhps quintuple mutant genotype in Mali. *PLoS One*. 2016;11:e0162718.
- 29 Dokomajilar C, Lankoande ZM, Dorsey G, Zongo I, Ouedraogo JB, Rosenthal PJ. Roles of specific Plasmodium falciparum mutations in resistance to amodiaquine and sulfadoxine-pyrimethamine in Burkina Faso. *Am J Trop Med Hyg*. 2006;75:162–165.
- 30 Dokomajilar C, Nsobia SL, Greenhouse B, Rosenthal PJ, Dorsey G. Selection of Plasmodium falciparum pfmdr1 alleles following therapy with artemether-lumefantrine in an area of Uganda where malaria is highly endemic. *Antimicrob Agents Chemother*. 2006;50:1893–1895.
- 31 Grais RF, Laminou IM, Woi-Messe L, et al. Molecular markers of resistance to amodiaquine plus sulfadoxine-pyrimethamine in an area with seasonal malaria chemoprevention in south central Niger. *Malar J*. 2018;17:98.
- 32 Rausch T, Fritz MH-Y, Untergasser A, Benes V. Tracy: basecalling, alignment, assembly and deconvolution of sanger chromatogram trace files. *BMC Genom*. 2020;21:230.
- 33 Cingolani P, Platts A, Wang leL, 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*. 2012;6:80–92.
- 34 Fantin YS, Neverov AD, Favorov AV, et al. Base-calling algorithm with vocabulary (BCV) method for analyzing population sequencing chromatograms. *PLoS One*. 2013;8:e54835.
- 35 Bloh K, Kanchana R, Bialk P, et al. Deconvolution of complex DNA repair (DECODR): establishing a novel deconvolution algorithm for comprehensive analysis of CRISPR-edited sanger sequencing data. *CRISPR J*. 2021;4:120–131.
- 36 Brinkman EK, Chen T, Amendola M, van Steensel B. Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res*. 2014;42:e168.
- 37 Conant D, Hsiau T, Rossi N, et al. Inference of CRISPR edits from sanger trace data. *CRISPR J*. 2022;5:123–130.
- 38 Dehairs J, Talebi A, Cherifi Y, Swinnen JV. CRISP-ID: decoding CRISPR mediated indels by Sanger sequencing. *Sci Rep*. 2016;6:28973.
- 39 Kluesner MG, Tasakis RN, Lerner T, et al. MultiEditR: the first tool for the detection and quantification of RNA editing from Sanger sequencing demonstrates comparable fidelity to RNA-seq. *Mol Ther Nucleic Acids*. 2021;25:515–523.
- 40 Brazdilova K, Prihoda D, Ton Q, Klock H, Bitton DA. TraceTrack, an open-source software for batch processing, alignment and visualization of sanger sequencing chromatograms. *Bioinforma Adv*. 2023;3.
- 41 Yeo S-J, Liu D-X, Kim HS, Park H. Anti-malarial effect of novel chloroquine derivatives as agents for the treatment of malaria. *Malar J*. 2017;16:80.
- 42 Jensen JB, Capps TC, Carlin JM. Clinical drug-resistant falciparum malaria acquired from cultured parasites. *Am J Trop Med Hyg*. 1981;30:523–525.
- 43 Peterson DS, Milhous WK, Welles TE. Molecular basis of differential resistance to cycloguanil and pyrimethamine in Plasmodium falciparum malaria. *Proc Natl Acad Sci USA*. 1990;87:3018–3022.
- 44 Wootton JC, Feng X, Ferdig MT, et al. Genetic diversity and chloroquine selective sweeps in Plasmodium falciparum. *Nature*. 2002;418:320–323.
- 45 Mu J, Awadalla P, Duan J, et al. Recombination hotspots and population structure in Plasmodium falciparum. *PLoS Biol*. 2005;3:e335.
- 46 Biswas S. Associations of antifolate resistance in vitro and point mutations in dihydrofolate reductase and dihydropteroate synthetase genes of Plasmodium falciparum. *J Postgrad Med*. 2004;50.
- 47 Huber W, Carey VJ, Gentleman R, et al. Orchestrating high-throughput genomic analysis with Bioconductor. *Nat Methods*. 2015;12:115–121.
- 48 Hill JT, Demarest BL, Bisgrove BW, Su YC, Smith M, Yost HJ. Poly peak parser: method and software for identification of unknown indels using sanger sequencing of polymerase chain reaction products. *Dev Dyn*. 2014;243:1632–1636.
- 49 Ewing B, Hillier L, Wendl MC, Green P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res*. 1998;8:175–185.
- 50 Pagès H, Aboyoun P, Gentleman R, DebRoy S. Biostrings: efficient manipulation of biological strings. In: *Biostrings*. R package version 2.48.0; 2020.
- 51 Bates DMAW, D. G. N. R. A. and I. A., Wiley. 1988. 42 Yang Xiang SG, Brian S, Hoeng J. Generalized simulated annealing for global optimization: the GenSA package. *R J*. 2013;5(1):13–28.
- 52 Xiang Y, Gubian S, Suomela B, Hoeng J. Generalized simulated annealing for global optimization: the GenSA package. 2013;5:13–28.
- 53 Team Posit. *RStudio: Integrated development Environment for R*. (Boston, 2023).
- 54 Chahar M, Anvikar A, Valecha N. Development and evaluation of a novel HNB based isothermal amplification assay for fast detection of pyrimethamine resistance (S108N) in Plasmodium falciparum. *Int J Environ Res Publ Health*. 2019;16.
- 55 Harrington CT, Lin EI, Olson MT, Eshleman JR. Fundamentals of pyrosequencing. *Arch Pathol Lab Med*. 2013;137:1296–1303.
- 56 Zhou Z, Poe AC, Limor J, et al. Pyrosequencing, a high-throughput method for detecting single nucleotide polymorphisms in the dihydrofolate reductase and dihydropteroate synthetase genes of Plasmodium falciparum. *J Clin Microbiol*. 2006;44:3900–3910.
- 57 Martin L, Damaso N, Mills D. Detection of single nucleotide polymorphisms (SNP) in equine coat color genes using SNaPshotTM multiplex kit or pluronic F-108 tri-block copolymer and capillary electrophoresis. *Electrophoresis*. 2016;37:2862–2866.
- 58 LeClair NP, Conrad MD, Baliraine FN, Nsanabana C, Nsobia SL, Rosenthal PJ. Optimization of a ligase detection reaction-fluorescent microsphere assay for characterization of resistance-mediating polymorphisms in African samples of Plasmodium falciparum. *J Clin Microbiol*. 2013;51:2564–2570.
- 59 Babiker HA, Lines J, Hill WG, Walliker D. Population structure of Plasmodium falciparum in villages with different malaria endemicity in east Africa. *Am J Trop Med Hyg*. 1997;56:141–147.
- 60 Diakitè SAS, Traoré K, Sanogo I, et al. A comprehensive analysis of drug resistance molecular markers and Plasmodium falciparum

- genetic diversity in two malaria endemic sites in Mali. *Malar J.* 2019;18:361.
- 61 Happi CT, Gbotosho GO, Sowunmi A, et al. Molecular analysis of *Plasmodium falciparum* recrudescence malaria infections in children treated with chloroquine in Nigeria. *Am J Trop Med Hyg.* 2004;70:20–26.
- 62 Basco LK, Ringwald P. Molecular epidemiology of malaria in Yaounde, Cameroon. VII. Analysis of recrudescence and reinfection in patients with uncomplicated falciparum malaria. *Am J Trop Med Hyg.* 2000;63:215–221.
- 63 Menard D, Dondorp A. Antimalarial drug resistance: a threat to malaria elimination. *Cold Spring Harb Perspect Med.* 2017;7.
- 64 Matreivi SA, Tandoh KZ, Bruku S, et al. Novel pfk13 polymorphisms in *Plasmodium falciparum* population in Ghana. *Sci Rep.* 2022;12:7797.