Received: 15 September 2021 | Revised: 5 March 2022 | Accepted: 5 April 2022

DOI: 10.1111/1755-0998.13622

MOLECULAR ECOLOGY RESOURCES WILEY

RESOURCE ARTICLE

Design and implementation of multiplexed amplicon sequencing panels to serve genomic epidemiology of infectious disease: A malaria case study

Emily LaVerriere[1,2] | Philipp Schwabl[1,2] | Manuela Carrasquilla[1,2,3] | Aimee R. Taylor[2,4] | Zachary M. Johnson[1,2] | Meg Shieh[1,2] | Ruchit Panchal[1,2] | Timothy J. Straub[1,2] | Rebecca Kuzma[1,2] | Sean Watson[1] | Caroline O. Buckee[4] | Carolina M. Andrade[5] | Silvia Portugal[3,5] | Peter D. Crompton[6] | Boubacar Traore[7] | Julian C. Rayner[8] | Vladimir Corredor[9] | Kashana James[10] | Horace Cox[11] | Angela M. Early[1,2] | Bronwyn L. MacInnis[2] | Daniel E. Neafsey[1,2]

[1]Department of Immunology and Infectious Diseases, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA

[2]Infectious Disease and Microbiome Program, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA

[3]Max Planck Institute for Infection Biology, Berlin, Germany

[4]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA

[5]Centre of Infectious Diseases, Parasitology, Heidelberg University Hospital, Heidelberg, Germany

[6]Malaria Infection Biology and Immunity Section, Laboratory of Immunogenetics, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Rockville, Maryland, USA

[7]Mali International Center of Excellence in Research, University of Sciences, Technique and Technology of Bamako, Bamako, Mali

[8]Cambridge Institute for Medical Research, University of Cambridge, Cambridge, UK

[9]Departamento de Salud Pública, Facultad de Medicina, Universidad Nacional de Colombia, Bogotá, Colombia

[10]Guyana National Malaria Control Program, Ministry of Health, Georgetown, Guyana

[11]Guyana Vector Control Services, Ministry of Health, Georgetown, Guyana

Correspondence
Daniel E. Neafsey, Department of Immunology and Infectious Diseases, Harvard T.H. Chan School of Public Health, Boston, MA, USA.
Email: neafsey@hsph.harvard.edu

Funding information
National Institutes of Health, Grant/Award Number: R01AI141544; British Council; Broad Institute; National Institute of Allergy and Infectious Diseases; Bill and Melinda Gates Foundation; Division of Intramural Research, National Institute of Allergy and Infectious Diseases

Handling Editor: Sebastien Calvignac-Spencer

Abstract

Multiplexed PCR amplicon sequencing (AmpSeq) is an increasingly popular application for cost-effective monitoring of threatened species and managed wildlife populations, and shows strong potential for the genomic epidemiology of infectious disease. AmpSeq data from infectious microbes can inform disease control in multiple ways, such as by measuring drug resistance marker prevalence, distinguishing imported from local cases, and determining the effectiveness of therapeutics. We describe the design and comparative evaluation of two new AmpSeq assays for *Plasmodium falciparum* malaria parasites: a four-locus panel ("4CAST") composed of highly diverse antigens, and a 129-locus panel ("AMPLseq") composed of drug resistance markers, highly diverse loci for inferring relatedness, and a locus to detect *Plasmodium vivax*

Emily LaVerriere, Philipp Schwabl, Manuela Carrasquilla contributed equally to this work.

co-infection. We explore the performance of each panel in various public health use cases with *in silico* simulations as well as empirical experiments. The 4CAST panel appears highly suitable for evaluating the number of distinct parasite strains within samples (complexity of infection), showing strong performance across a wide range of parasitaemia levels without a DNA pre-amplification step. For relatedness inference, the larger AMPLseq panel performs similarly to two existing panels of comparable size, despite differences in the data and approach used for designing each panel. Finally, we describe an R package (paneljudge) that facilitates the design and comparative evaluation of genetic panels for relatedness estimation, and we provide general guidance on the design and implementation of AmpSeq panels for the genomic epidemiology of infectious disease.

KEYWORDS
amplicon sequencing, epidemiology, genome, genotyping, malaria, relatedness

1 | INTRODUCTION

Genetic data are a valuable resource for understanding microbial ecology and the epidemiology of infectious disease. The value of this data type has been highlighted by the COVID-19 pandemic, for which viral sequence analysis has greatly informed patterns of disease spread and evolution, influencing public health policy decisions around the world (Oude Munnink et al., 2021). Applications of genetic data in epidemiology extend from viral and bacterial outbreak management (Gardy & Loman, 2018) to the study of eukaryotic parasites underlying important diseases such as malaria, leishmaniasis and cryptosporidiosis (Cantacessi et al., 2015; Nader et al., 2019; Neafsey et al., 2021).

Many use cases (applications) of genetic data have been identified for malaria (Dalmat et al., 2019), the leading parasitic killer worldwide (WHO, 2019). These include tracking the spread of drug/insecticide resistance markers and diagnostic mutations (Chenet et al., 2016; Jacob et al., 2021; Kayiba et al., 2021; Lautu-Gumal et al., 2021; Miotto et al., 2020), assessing disease transmission levels (Daniels et al., 2015; Galinsky et al., 2015), identifying sources of infections and imported cases (Liu et al., 2020; Tessema et al., 2019), and estimating genetic connectivity among different populations (Taylor et al., 2017). Malaria parasite genetic data have also demonstrated utility in therapeutic efficacy studies, such as for distinguishing reinfections from recrudescent infections potentially indicative of drug inefficacy (Gruenberg et al., 2019; Jones et al., 2021). These applications in the malaria field are served by different types of genetic data produced at varying resolution, technical complexity and cost, ranging from genetic panels that comprise as few as 8–12 polymorphic microsatellites or 24 single nucleotide polymorphisms (SNPs) (Daniels et al., 2008; Yalcindag et al., 2012) to whole genome sequencing (WGS) data (Miotto et al., 2015; Takala-Harrison et al., 2015).

To be scalable and sustainable, genetic data should be produced at the minimum resolution that provides robust support for the intended analysis application. WGS data often provide the most complete population genomic perspective on an organism of interest. However, the cost and technical challenges of generating, storing and interpreting WGS data are impediments to scalability and widespread implementation for organisms with large genomes, or microbes with small genomes in samples dominated by host DNA. Targeted sequencing approaches that focus deep coverage on select genomic regions of interest using multiplexed PCR amplification (AmpSeq) are finding increased application in areas of conservation biology, fisheries science and evolutionary research (Baetscher et al., 2018; Bybee et al., 2011; Dupuis et al., 2018; Hargrove et al., 2021; Natesh et al., 2019; O'Neill et al., 2013; Schmidt et al., 2020). These approaches can also serve genomic epidemiology of infectious diseases by focusing sequencing coverage on the most informative regions of pathogen genomes.

While only a few AmpSeq protocols for eukaryotic parasites have been published to date, pioneer examples for malaria and trypanosomatid parasites have confirmed the viability of this approach with low-parasitaemia host and vector samples, where parasite DNA comprises a very small fraction of the total sample (Jacob et al., 2021; Ruybal-Pesántez et al., 2021; Schwabl et al., 2020; Tessema et al., 2020). Furthermore, one recent study has confirmed the value of designing amplicons to capture multi-SNP *Plasmodium falciparum* "microhaplotypes," which exhibit polyallelic rather than biallelic diversity to facilitate relatedness inference (Tessema et al., 2020). New relatedness-based analytical approaches for genomic epidemiology are currently developing for malaria parasites and other sexually recombining pathogens (Henden et al., 2018; Schaffner et al., 2018). The use of genomic data for estimation of recent common ancestry shared by pairs or clusters of parasites or mosquitoes has shown strong potential to provide epidemiologically useful insights over small geographical distances (10s to 100s of kilometres) and short timescales (weeks to months) relative to traditional population genetic parameters of population diversity and divergence (Cerqueira et al., 2017; Taylor et al., 2017). While many analyses of recent

common ancestry in malaria parasites to date have used WGS data, targeted genotyping of as few as 200 biallelic SNPs or 100 polyallelic loci (e.g., microsatellites or microhaplotypes) may also be used to infer relatedness with necessary precision (Taylor et al., 2019), making AmpSeq an excellent candidate for relatedness estimation.

However, there remains uncertainty in the molecular epidemiology field as to the suitability of existing panels for profiling pathogen populations in specific geographical locations that did not inform the original panel designs, and it is unclear which protocol features are most conducive to implementation in both high- and low-resource settings. Should each disease field adopt a common multiplexed amplicon protocol and panel, or should bespoke panels be implemented regionally to address genetically distinct pathogen populations and specific use cases?

To address these questions, here we describe the design and comparative evaluation of two new multiplexed amplicon assays for *P. falciparum* malaria parasites: a four-locus panel composed of highly diverse loci, useful for estimating the number of genetically distinct strains within an infection (complexity of infection; COI) as well as distinguishing between continuing and newly acquired infections in any geographical setting; and a 129-locus panel composed of drug resistance markers and many diverse loci for relatedness inference initially designed for application in South America (a region that did not inform previously published panel designs). Both assays use non-proprietary reagents (including standard PCR oligos) to maximize accessibility and affordability in malaria-endemic settings. The panels are supported by new open-source bioinformatic analysis pipelines to facilitate widespread use. We also show that the core sets of multiplexed PCR oligos can flexibly accommodate various new targets not included in the original designs, allowing for panel customization towards detecting locally relevant resistance markers, polymorphic loci and co-infecting parasite species. We use WGS data to explore the degree to which our newly described and previously published genotyping panels can serve studies in diverse geographies vs. the alternative of customizing panels with targets that are locally informative but not globally useful. We suggest there is value in genotyping panels that can be flexibly adapted to incorporate informative targets from pathogen populations of interest. The analyses and resources described here clarify the rapidly diversifying options for targeted microbial sequencing (Figure 1) by providing tools and guidance for the comparative evaluation and refinement of AmpSeq approaches.

2 | MATERIALS AND METHODS

2.1 | Panel designs

We developed a small multiplex of four highly polymorphic antigenic loci, dubbed "4CAST": *CSP*, *AMA1*, *SERA2* and *TRAP* (Figure 2). All four amplicons use previously published primer sequences (Miller et al., 2017; Neafsey et al., 2015), as no modification was required for successful multiplexing.

In designing the larger multiplexed amplicon panel we call "AMPLseq" (short for "Assorted Mix of *Plasmodium* Loci"), we first built a large pool of candidate loci, anticipating significant attrition of candidates due to primer incompatibility. We prioritized four classes of loci: loci within antigens of interest (Helb et al., 2015), loci with high population diversity for relatedness inference (Taylor et al., 2019), loci included in the SpotMalaria v1 panel (Chang et al., 2019; Jacob et al., 2021) and known antimalarial drug resistance markers. We contracted the services of GTseek LLC (https://gtseek.com) to design multiplexed oligo panels according to the criteria previously described for the Genotyping-in-Thousands by

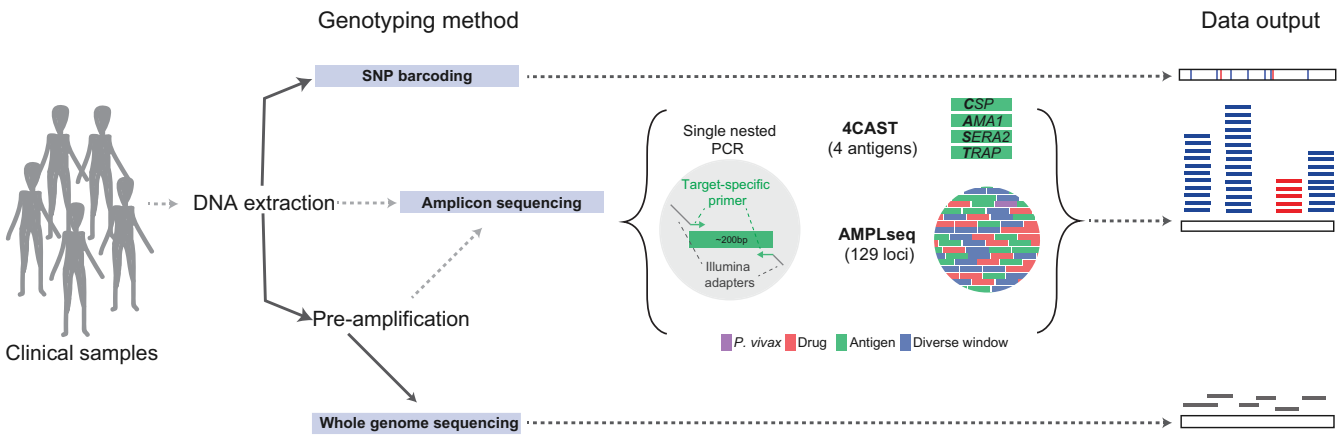


FIGURE 1 Amplicon sequencing and other genotyping approaches for genomic epidemiology of infectious diseases. Schematic of three common approaches for molecular surveillance data generation. Genomic DNA can be extracted from clinical samples and then processed using any of the three methods shown: SNP barcoding, amplicon sequencing or whole genome sequencing (WGS). Our two amplicon panels, AMPLseq and 4CAST, are shown with representations of their loci and amplification. Pre-amplification (selective whole genome amplification), which increases the ratio of parasite to human DNA in samples, is generally recommended for WGS and some amplicon sequencing panels (AMPLseq, but not 4CAST). SNP barcoding provides data in the form of variant calls at each SNP site; amplicon sequencing provides extremely deep coverage at select, small regions of the genome; and WGS generally provides shallower coverage of the entire genome

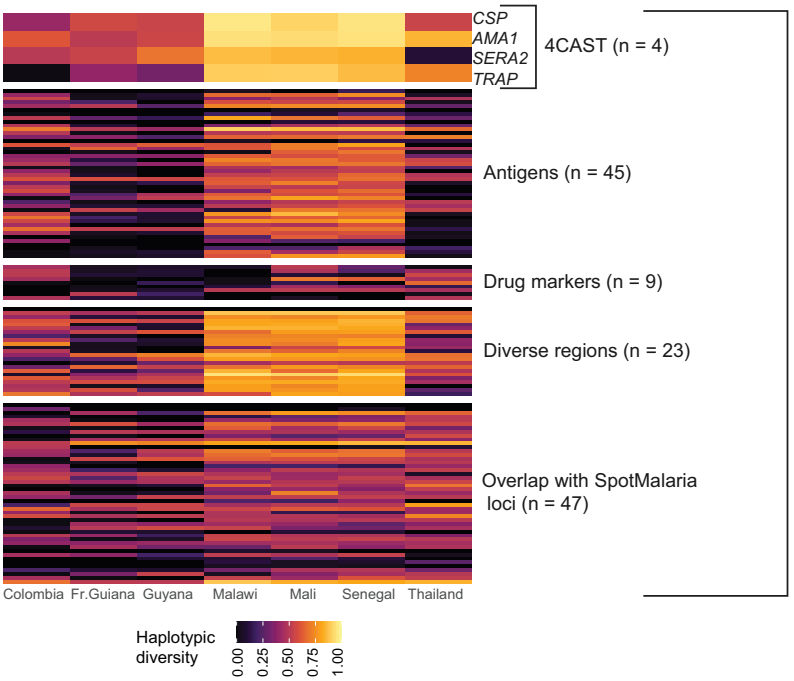


FIGURE 2 Global characterization of loci in the 4CAST and AMPLseq panels. Estimates of diversity of each locus in the 4CAST and AMPLseq panels, with one locus per row. We estimated haplotypic diversity from monoclonal *Plasmodium falciparum* WGS data from each country. The top four rows represent the 4CAST loci, which are also included in the AMPLseq panel. All 128 *P. falciparum* loci in the AMPLseq panel are shown; the single *P. vivax* locus is not shown

sequencing (GT-seq) protocol (Campbell et al., 2015; Supporting Information S1). We optimized the final primer set and reaction conditions through several sequencing runs and determined that the primers for the four 4CAST loci (*CSP*, *AMA1*, *SERA2*, *TRAP*) could be added to the panel without compromising amplification of the other loci. We also successfully added primers amplifying known markers of drug resistance within the genes *dhfr*, *dhps*, *mdr1* and *kelch13* (Table S1). Furthermore, we added previously described primers targeting a region within *Pvdhfr* (Lefterova et al., 2015) to identify *Plasmodium falciparum*/*P. vivax* co-infections undetected in preliminary screening by microscopy or rapid diagnostic test (RDT). The final panel contains this single *P. vivax* locus and 128 *P. falciparum* loci (Figure 2), with a median length across all amplicons of 276 bp (Figure S1).

2.2 | Panel protocols

To create the primer pool used in 4CAST PCR1, we combined 100 μM of each 4CAST primer (Table S2) and diluted the combined primer mix to 6.25 μM per primer in nuclease-free water (NF dH_2O). Each 10.5 μl PCR1 reaction incorporated 1.5 μl combined primer mix, 5 μl KAPA HiFi HotStart ReadyMix (2×) and 4 μl sample template. Each 12.2 μl PCR2 reaction incorporated 2.2 μl unique dual index (10 μM Illumina Nextera DNA UD Indexes), 5 μl KAPA HiFi HotStart ReadyMix (2×), 2 μl NF dH_2O and 3 μl PCR1 product. PCR cycling conditions are provided in Protocol S1. We combined PCR2 products in equal volumes and subjected the resultant library to double-sided size selection using Agencourt AMPure XP beads (Beckman Coulter). We verified size selection via Agilent Bioanalyzer 2100 and sequenced the selected library at 6 pM with >10% PhiX in paired-end, 500-cycle format using MiSeq Reagent Kit v2 (Protocol S1).

We followed a similar nested PCR and pooled clean-up procedure for AMPLseq library construction. Primer sequences, input volumes and concentrations are listed in Table S2 and PCR conditions and size selection steps are described in Protocol S2. As detailed therein, AMPLseq library construction differs to 4CAST library construction in a few minor aspects. For example, primer input quantities vary slightly (800 pmol ± 33%) to account for amplification rate differences among loci. PCR1 products are diluted prior to PCR2 and only single-sided (left-tailed) bead-based size selection is used to enhance yield. Sequencing also occurs via paired-end, 500-cycle MiSeq but with a higher final library loading concentration (12 pM) and a lower fraction of PhiX (8%).

2.3 | Mock samples

We generated mock samples from parasite lines 3D7 and Dd2, cultured at 3% haematocrit in commercially obtained red blood cells as previously described (Trager & Jensen, 1976). We extracted genomic DNA (gDNA) using the Qiagen Blood and Tissue Kit on cells previously lysed with 0.15% saponin. We generated positive control template representing DNA extractions from whole human blood infected with 10,000 monoclonal 3D7 parasites/μl by combining 13.76 ng/μl human genomic DNA (Promega) with 0.92 ng/μl 3D7 genomic DNA at a ratio (v/v) of 2.66 to 1, resulting in a mixture containing 10 ng/μl human genomic DNA and 0.25 ng/μl 3D7 gDNA. We used 0.25 ng to represent the mass of 10,000 *P. falciparum* genomes based on a 23-Mbp genome size and an average mass of 660 g per mol bp; this assumes one haploid parasite genome per infected cell, as expected for peripheral blood (the target profile). We generated further control templates representing 1000, 100 and 10 3D7 parasites/μl by serial

1:10 dilution of the 10,000 3D7 parasites/μl control with 10 ng/μl human gDNA. We also generated a 10,000 parasites/μl positive control as described above but using Dd2 instead of 3D7 strain gDNA. We generated mixed-strain control templates by combining the 10,000 3D7 parasites/μl control with this 10,000 Dd2 parasites/μl control at 1:1, 3:1 and 10:1 ratios (respectively). We serially diluted the 1:1 ratio to concentrations of 1000, 100 and 10 parasites/μl and diluted the 3:1 and 10:1 ratios to concentrations of 1000 and 100 parasites/μl using 10 ng/μl human gDNA diluent as before. We also applied selective whole genome amplification (sWGA) to all above control templates representing ≤1000 parasites/μl. The 50 μl sWGA reaction followed Oyola et al. (2016) with the exception of fixing template input volume to 10 μl. We purified sWGA products with Agencourt AMPure XP beads (Beckman Coulter) on the KingFisher Flex (Protocol S3) and verified amplification success via a NanoDrop (ThermoFisher Scientific).

2.4 | Clinical samples

We tested the panels on clinical dried blood spot (DBS) samples from Mali and Guyana. Tran and colleagues collected samples in Kalifabougou, Mali, between 2011 and 2013 as previously described (Tran et al., 2013). The Kalifabougou cohort study was approved by the Ethics Committee of the Faculty of Medicine, Pharmacy and Dentistry at the University of Sciences, Technique and Technology of Bamako, and the Institutional Review Board of the National Institute of Allergy and Infectious Diseases, National Institutes of Health (NIH IRB protocol no.: 11IN126; https://clinicaltrials.gov/; trial number NCT01322581). Written informed consent was obtained from parents or guardians of participating children before inclusion in the study. The Guyana Ministry of Health collected samples from Port Kaituma and Georgetown, Guyana, between May and August 2020 by spotting participants' whole blood onto Whatman FTA cards and storing the samples with individual desiccant packets at room temperature. Informed consent (or parental assent for minors) was obtained for all subjects according to protocols approved by ethics committees.

We punched DBS samples three to five times into a 96-well deep well plate using a DBS pneumatic card puncher (Analytical Sales and Services) equipped with a 3-mm cutter. We then extracted gDNA following the DNA purification from buccal swab section of the KingFisher Ready DNA Ultra 2.0 Prefilled Plates user guide (ThermoFisher Scientific) with minor modifications (Protocol S4). We used the same sWGA procedure as above on the extracted gDNA.

2.5 | Whole genome sequencing and variant calling

We performed whole genome sequencing on clinical samples collected in Guyana to validate 4CAST and AMPLseq outcomes. We performed sWGA on DNA samples as described above to enrich parasite DNA. We used the enriched DNA to construct Illumina sequencing libraries from the amplified material using the NEBNext Ultra II FS DNA Library Prep Kit (NEB #E6177) prior to sequencing on an Illumina HiSeqX instrument at the Broad Institute, using 150-bp paired-end reads and targeting a sequencing depth of at least 50×. We aligned reads to the *P. falciparum* v3 reference genome assembly using BWA-MEM (Li, 2013) and called SNPs and INDELs using the GATK HaplotypeCaller (DePristo et al., 2011; McKenna et al., 2010; Van der Auwera et al., 2013) according to the best practices for *P. falciparum* as determined by the Pf3k consortium (https://www.malariagen.net/resource/34). Analyses were limited to the callable segments of the genome (Miles et al., 2016) and excluded sites at which over 20% of monoclonal samples were multi-allelic. Data from these samples were submitted to the NCBI Sequence Read Archive (http://www.ncbi.nlm.nih.gov/sra) under accession PRJNA758191.

2.6 | Amplicon data analysis

We developed the application AmpSeQC (Supporting Information S2) to assess sequencing quality and yield (Figure S2). We also used AmpSeQC for *P. vivax* detection by concatenating the *P. falciparum* 3D7 and *P. vivax* PvP01 reference genomes during the BWA-MEM alignment step. For in-depth assessment of *P. falciparum* sequence variation, we processed paired-end Illumina sequencing data in the form of FASTQ files using a custom analysis pipeline (Supporting Information S2) that utilizes the Divisive Amplicon Denoising Algorithm (DADA2) tool designed by Callahan et al. (2016) to obtain microhaplotypes (Figure S2). We mapped microhaplotypes obtained from DADA2 against a custom-built database of 3D7 and Dd2 reference sequences for each amplicon locus and filtered microhaplotypes based on edit distance, length and chimeric identification using a custom R script (Supporting Information S2). We summarized observed sequence polymorphism into a concise format by converting individual microhaplotypes into pseudo-CIGAR strings using a custom python script. Microhaplotypes were discarded if supported by fewer than 10 read-pairs or by less than 1% total read-pairs within a locus, or if they exhibited other error features (Supporting Information S3).

We analysed native and pre-amplified mock samples to determine precision and sensitivity of the DADA2 pipeline and filters. We defined a true positive (TP) as a microhaplotype with a pseudo-CIGAR string identical to the reference strain (either 3D7 or Dd2). We defined a false positive (FP) as a microhaplotype with a pseudo-CIGAR string not matching 3D7 (in the case of samples containing only 3D7) or not matching 3D7 or Dd2 (in the case of the mixtures), and we defined a false negative (FN) as a locus without any correct microhaplotype representation. We defined precision as TP/(TP + FP), and sensitivity (recall) as TP/(TP + FN). Forty-five of 128 *P. falciparum* AMPLseq loci exhibited identical 3D7 and Dd2 reference sequences; we only included these in precision and sensitivity calculations for pure 3D7 controls (i.e., TP + FN = 128); precision and sensitivity calculations for strain mixtures considered only the 83 loci that differ between 3D7 and Dd2 reference sequences (i.e., TP + FN = 83).

All amplicon sequencing data were submitted to http://www.ncbi.nlm.nih.gov/sra under accession PRJNA758191.

2.7 | Comparator panels

We compared 4CAST and AMPLseq to two previously published AmpSeq panels for malaria molecular surveillance, Paragon HeOME v1 (Tessema et al., 2020) and SpotMalaria v2 (Jacob et al., 2021).

Paragon HeOME v1, designed using the CleanPlex algorithm (Paragon Genomics), contains 100 primer-pairs in a single pool. Primer design focused on 150 genetic windows that show high diversity and differentiation (Jost $D \geq 0.21$) among clinical isolates from throughout sub-Saharan Africa, resulting in 93 amplicon targets with high median heterozygosity in all malaria-endemic regions of the world. The panel also targets seven drug resistance-associated loci. A distinctive feature of HeOME library construction involves its requirement for bead-based clean-up and CleanPlex digestion of each sample between PCR1 and PCR2. The protocol therefore does not require sWGA prior to PCR1.

SpotMalaria v2, designed via Agena BioScience and MPprimer design software, contains 136 primer-pairs divided into three different pools. A majority of target loci are intended for genetic barcode generation and contain biallelic sites with ≥0.01 minor allele frequency (MAF) in each of eight global parasite populations studied via WGS (MalariaGEN Plasmodium falciparum Community Project, 2016). These barcoding loci were chosen based on their ability to recapitulate WGS-based inferences on pairwise genetic distance, population differentiation and sample heterozygosity at different spatial scales. Primers also target various drug resistance-associated loci and mitochondrial sequences with conserved primer binding sites among *Plasmodium* spp. Library construction requires sWGA prior to PCR1 but no special processing between PCR1 and PCR2.

We also compared our amplicon panels to a molecular barcode assay containing 24 SNPs (Daniels et al., 2008). The SNPs targeted by this Taqman qPCR-based assay were chosen principally for their high average MAF (>0.35) across parasite sample collections from Thailand and Senegal, with further filtering to remove tightly linked SNPs and to minimize the generation of identical barcodes for closely related strains (Daniels et al., 2008).

2.8 | paneljudge and *in silico* data simulations

We used WGS data to simulate genotypic panel data for simulations. This publication uses data from the MalariaGEN *Plasmodium falciparum* Community Project as described online pending publication and public release of data set Pf7 (https://www.malariagen.net/resource/34). Specifically, we used genomic data from monoclonal samples collected in Mali, Malawi, Senegal and Thailand (Zhu et al., 2019), and from Colombia and Venezuela (accession numbers in Table S3). We also used previously published monoclonal genomic data from Guyana (SRA BioProject PRJNA543530; Mathieu et al., 2020) and French Guiana (SRA BioProject PRJNA242182; Pelleau et al., 2015). We used the *scikit-allel* library (Miles et al., 2020) "read_vcf," "is_het," "haploidify_samples" and "distinct_frequencies" functions to process the data and estimate microhaplotype frequency and diversity metrics (Supporting Information S1).

We assessed the performance of different panels for relatedness inference using simulated data. We simulated data on pairs of haploid genotypes (equivalent to pairs of monoclonal malaria samples) using paneljudge, an R package that we built to simulate data under a hidden Markov model (HMM) (Taylor et al., 2019), which is the same HMM used in the single-population implementation of hmmIBD (Schaffner et al., 2018) (Supporting Information S2). For each panel, we calculated interlocus distances from the median nucleotide position of each locus and set distances as infinite between chromosomes. For each panel and population of interest, we calculated haplotype frequency estimates using *scikit-allel* as described above. Given these distances and frequency estimates, we simulated data using relatedness parameter values of 0.01 (unrelated), 0.50 (related) and 0.99 (clonal), and switch rate parameter values of 1, 5, 10 and 50. For each combination of panel, population, relatedness parameter and switch rate parameter, we simulated data on 100 haploid genotype pairs using paneljudge. For each haploid genotype pair, we then generated, also using paneljudge, estimates of the relatedness parameter, estimates of the switch rate parameter and 95% confidence intervals (CIs). The estimates were generated under the same parameterization of the HMM used to simulate the data; that is, frequencies and distances were unchanged. We next performed classification from the estimates of the relatedness parameters and their CIs. For pairs simulated using a relatedness parameter of 0.01, we generously classified estimates as correct if the lower limit of the 95% confidence interval around the relatedness estimate (LCI) was ≤0.01 and the upper limit of the 95% confidence interval (UCI) was <0.99. For pairs simulated using a parameter value of 0.50, we classified estimates of relatedness as correct if the LCI was >0.01 and the UCI was <0.99. For pairs simulated using a parameter value of 0.99, we classified estimates of relatedness as correct if the LCI was >0.01 and the UCI was ≥0.99. If the 95% confidence interval spanned both 0.01 and 0.99 (i.e., LCI < 0.01 and UCI > 0.99), then we denoted the estimate as unclassified. To evaluate panel performance in COI estimation, we combined monoclonal WGS data to engineer *in silico* polyclonal samples using VCFTOOLS (Danecek et al., 2011). We counted the number of distinct microhaplotypes observed at each locus per sample and estimated COI as the maximum number of distinct microhaplotypes observed at any locus within a sample.

To evaluate panel performance for geographical attribution, we identified microhaplotypes at loci as described above. We used the microhaplotype sequences themselves and visualized these data using the Rtsne package (Krijthe, 2015), with 5000 iterations, Θ of 0.0 and perplexity parameter of 10. We performed principal component analyses (PCAs) using the "prcomp" function in base R version 4.1.2 (R Core Team, 2021).

3 | RESULTS

3.1 | 4CAST and AMPLseq validation

We validated assay precision (defined as TP/(TP + FP)), sensitivity (defined as TP/(TP + FN)), and depth of coverage using 3D7 mock samples representing parasitaemia levels of between 10 and 10,000 parasites/μl in 10 ng/μl human DNA. Following automated and manual filtration steps (Supporting Information S3), both 4CAST and AMPLseq generated 3D7 microhaplotype calls with 100% precision for all parasitaemia levels assessed, both with and without pre-amplification by sWGA. 4CAST achieved high sensitivity and depth without preliminary sWGA, generating a median of 43 read-pairs per locus from native templates representing 10 parasites/μl (Figure 3a). Median depth increased to 443 and 1312.5 read-pairs per locus for native templates representing 100 and 1000 parasites/μl, respectively. Read-pair counts were also evenly distributed among 4CAST loci using native DNA (Figure 3a).

Unlike 4CAST, AMPLseq required sWGA for 3D7 mock samples representing 10 and 100 parasites/μl (Figure 3b). Following sWGA on mock samples representing 10 parasites/μl, the assay generated ≥10 read-pairs at a median of 126 loci, with a median of 465 read-pairs after excluding loci with fewer than 10 reads. Values were statistically similar for pre-amplified samples representing 100 parasites/μl and increased to 692 read-pairs for pre-amplified samples representing 1000 parasites/μl (Figure S3).

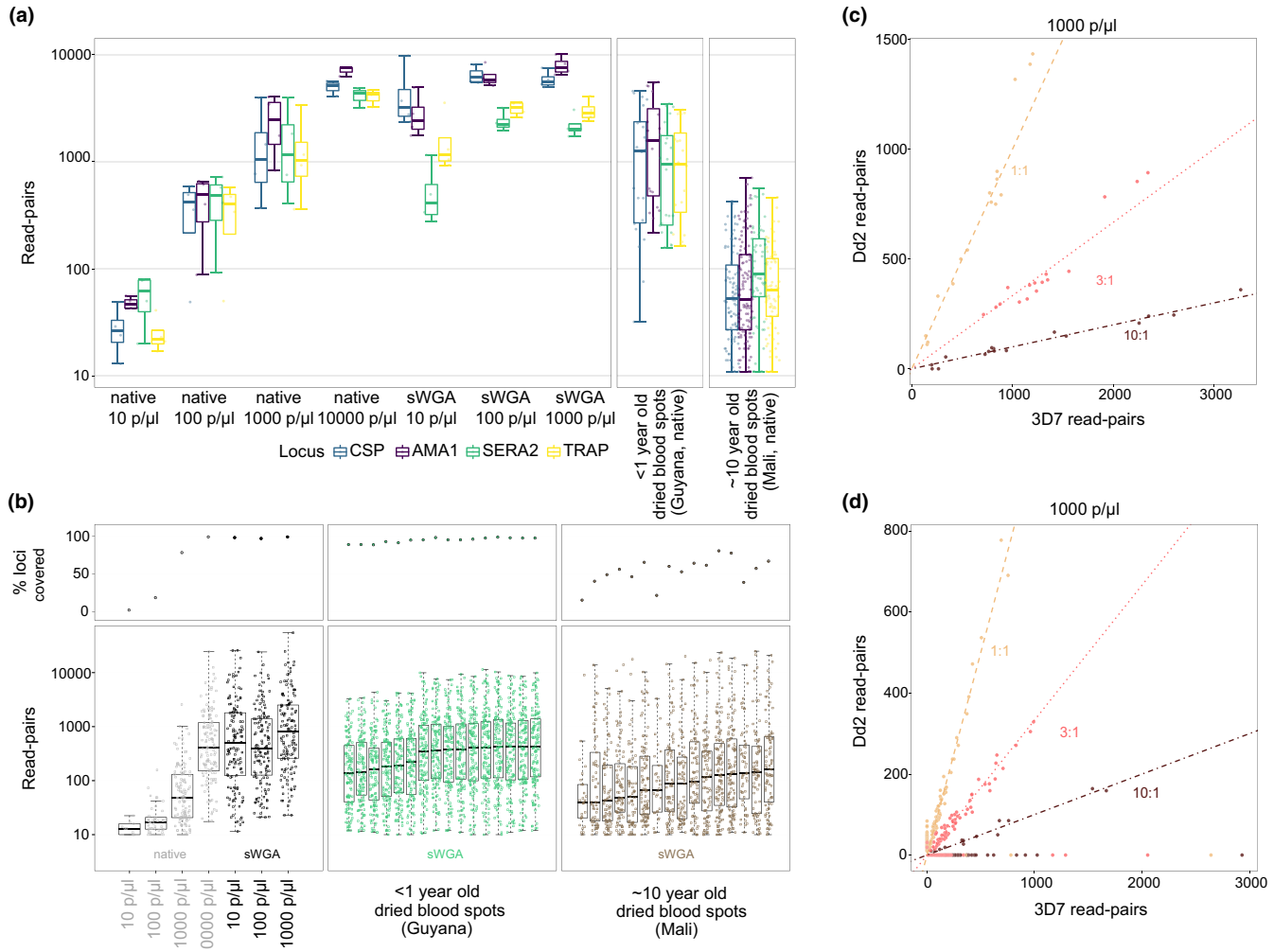


FIGURE 3 4CAST and AMPLseq panel validation with mock and clinical samples. (a) Boxplots of read-pairs per locus in the 4CAST panel. The first facet shows read-depth per locus across mock samples ranging from 10 to 10,000 parasites/μl, using both native and sWGA DNA ($n = 4$ per condition). The second and third facets show read-depth per locus across two sets of clinical samples, <1-year-old and ~10-year-old dried blood spots, respectively ($n = 32$ per sample set). (b) The same 3D7 mock sample sets were used to assess AMPLseq sensitivity (top left panel) and read-depth (bottom left panel). Each point in the bottom left panel represents read-pair support for one AMPLseq locus. Positions on the *y*-axis indicate median read-pair support across replicate samples. Low sensitivity observed using native templates (grey) representing 10 and 100 3D7 parasites/μl suggests that clinical samples should generally be pre-amplified with sWGA (results at right). (c) Ratio of 4CAST read-pairs from microhaplotypes assigned to 3D7 (*x*-axis) or Dd2 (*y*-axis) from 3D7 + Dd2 mock mixtures with strain ratios of 1:1 (tan), 3:1 (pink) and 10:1 (dark red), respectively. All samples contained 1000 parasites per μl in total, that is across both strains. Dashed lines represent the expected ratio, and each point represents a 4CAST locus per sample ($n = 4$ per condition). (d) AMPLseq read-pair ratios observed in native mock mixtures (1000 parasites/μl) plotted as above for 4CAST

We also validated the sensitivity of 4CAST and AMPLseq for genotyping polyclonal infections by using mock samples containing both 3D7 and Dd2 templates (likewise in 10 ng/μl human DNA). These mixtures featured Dd2 at 50% (i.e., 1:1 3D7:Dd2 ratio), 25% (3:1) and 9% (10:1) relative abundance. For all parasitaemia levels assessed in native and pre-amplified mixtures (between 10 and 10,000 parasites/μl), both 4CAST and AMPLseq generated microhaplotype calls with 100% precision at the 83 loci that are dimorphic between the 3D7 and Dd2 references (including all four 4CAST loci and an additional 79 loci in AMPLseq). This metric excludes two microhaplotypes classified as false positives in post-pipeline screening prior to precision analysis (Supporting Information S3).

4CAST showed high sensitivity for Dd2 without the need for sWGA. At 1000 parasites/μl, the assay detected Dd2-specific microhaplotypes at each of its four loci in all 1:1, 3:1 and 10:1 mixture replicates (Figure 3c). At 100 parasites/μl, median Dd2 sensitivity remained 100% at 1:1 and 3:1 ratios but was slightly lower (94%) at 10:1. At 10 parasites/μl, 1:1 ratios yielded a median of three target loci for 3D7 and a median of two target loci for Dd2; median sensitivity in these samples rose to 3.5 and 3 loci (respectively) following pre-amplification with sWGA, but this led to unbalanced read-pair support between the two strains (Figure S4A), possibly due to differential sWGA success on low-quality Dd2 vs. higher-quality 3D7 templates. 4CAST read-pair ratios generated from native templates, by contrast, showed a strong correlation with input ratios at 100 parasites/μl (Figure S4A) and 1000 parasites/μl (Figure 3c). Ratios were less informative at 10 parasites/μl (Figure S4A).

AMPLseq was also successful in detecting Dd2-specific microhaplotypes, but only at a maximum of 77 of 83 dimorphic loci (in the 1:1 ratio at 10,000 parasites/μl). With sWGA, Dd2-specific sequences were detected at a minimum of two dimorphic loci for all three input ratios (1:1, 3:1, 10:1) and parasitaemia levels (≥10 parasites/μl) assessed. Like with 4CAST, however, the use of sWGA decorrelated read-pair ratios from input ratios (Figure S4B). This decorrelation was not observed with native templates (Figure 3d) at 1000 and 10,000 parasites/μl for which AMPLseq achieved high read-pair support without the use of sWGA.

We also tested both panels on gDNA extracted from dried blood spots collected by the Guyana Ministry of Health in 2020 from individuals diagnosed as *P. falciparum*-positive via microscopy or RDT. Ten Guyanese samples were tested with both panels, and an additional six were tested with AMPLseq. Using 4CAST, we observed coverage across all loci in all samples, with a median depth per locus of 1162 read-pairs without sWGA (Figure 3a). Using AMPLseq (with sWGA), we observed a median of 122 loci with ≥10 read-pairs and a median depth of 298 read-pairs per covered locus (Figure 3b).

Additionally, we tested both panels on gDNA extracted from 16 dried blood spot samples collected in Mali in 2011 (Tran et al., 2013) and subsequently stored at room temperature for 10 years. Using 4CAST (without sWGA), we observed a median depth of 407 read-pairs per locus (Figure 3a). Using AMPLseq (with sWGA), we observed a median of 75 loci with ≥10 read-pairs and a median depth of 112 read-pairs per covered locus (Figure 3b).

3.2 | Evaluation of panel performance for relatedness

We used the R package paneljudge to assess *in silico* the impact of choosing a specific genotyping panel for relatedness inference. Considering the choice of panel, we evaluated relatedness estimation from data simulated on our 4CAST and AMPLseq panels, the SpotMalaria v2 (Jacob et al., 2021) and Paragon HeOME v1 (Tessema et al., 2020) amplicon panels, and a barcode of 24 SNPs (Daniels et al., 2008). When data were simulated using microhaplotype frequency estimates of Senegalese parasites, we found that almost all estimates of unrelated or clonal pairs were correctly classified, regardless of the panel (Figure 4a). All three large panels also performed similarly well in accurately identifying related (but not clonal) parasite pairs, despite being the product of three distinct design processes. Neither 4CAST nor the 24 SNP barcode estimated relatedness for partially related samples as well as with the larger panels. We also evaluated panel performance in less diverse parasite populations (Colombia and Thailand), including a population not used in the panel designs (Colombia). We repeated the simulations using microhaplotype frequencies estimated with these data. Again, we found that all panels performed well for estimating relatedness of clonal pairs, and that 4CAST and the 24 SNP barcode were less likely to have correctly classified estimates of nonclonal pairs. With the data simulated using Colombian microhaplotype frequencies from the Pacific Coast region, all three large panels performed well for all three relatedness values, despite the Colombian data not having informed the design of any of the panels.

Pairwise relatedness estimates (Schaffner et al., 2018) from AMPLseq correlated highly with those from WGS data available for the Guyanese sample set (Pearson's $r = .86$, slope = 1.01, $p < .001$; Figure 4b). Despite patient travel history metadata suggesting infections to have occurred in various geographical regions of Guyana (Table S4), AMPLseq relatedness estimates for the Guyanese sample set were significantly higher than those for the Malian sample set (Mann–Whitney U, $p < .001$), consistent with lower parasite population diversity anticipated in the Guiana Shield (Carrasquilla et al., 2022; Yalcindag et al., 2012). Nevertheless, the wide range of highly WGS-correlated AMPLseq relatedness estimates (0.007–1) observed among Guyanese sample comparisons suggests AMPLseq capacity to indicate epidemiologically relevant microstructure even in relatively unstructured parasite populations, as is achievable via WGS (Mathieu et al., 2020). For example, even within the very small Guyanese sample set analysed in this study, we could detect an enrichment of the first (lowest) quartile of pairwise relatedness estimates for comparisons involving A2-GUY and C5-GUY, two highly related samples that in WGS analysis show 50% relatedness with a sample from Venezuela (SPT26229, see Table S3). Larger AMPLseq sample sets from bordering countries such as Guyana and Venezuela may thus indicate cases of parasite importation or introgression between spatially proximate regions in which disease ecology and management is distinct.

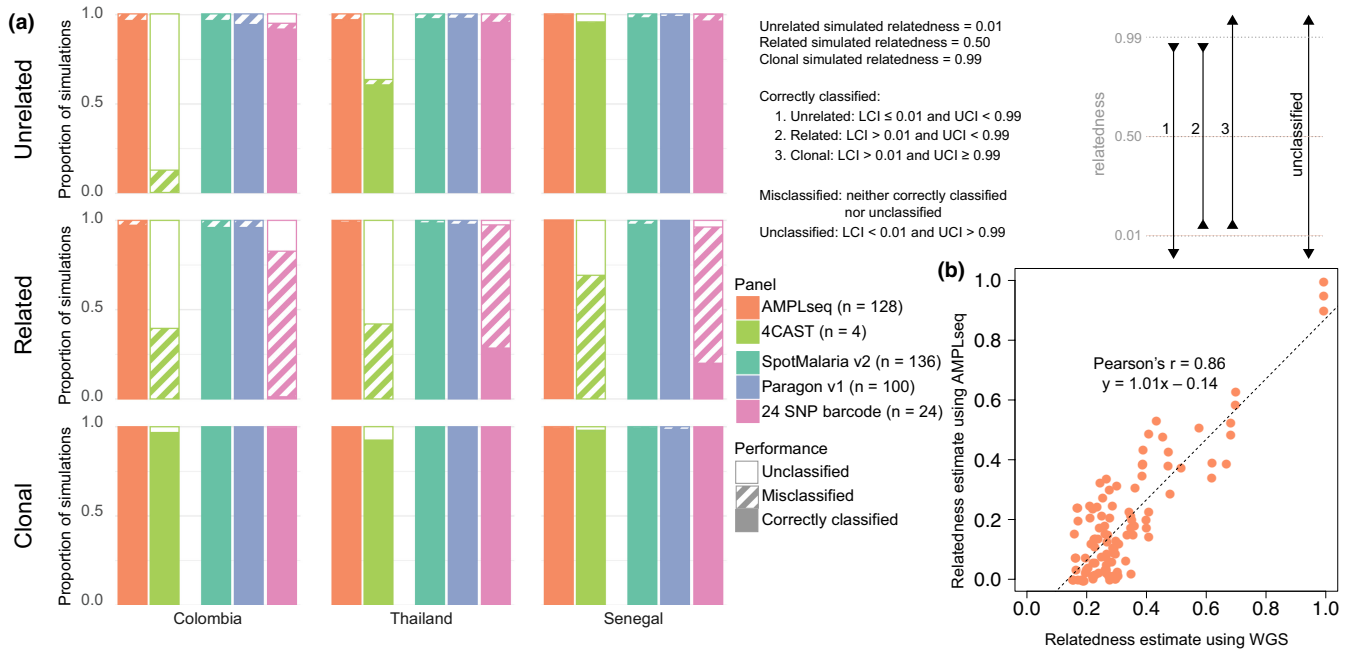


FIGURE 4 *In* silico relatedness estimation comparisons among panels and empirical AMPLseq validation against WGS. (a) Evaluation of relatedness estimation from data simulated on genotyping panels using the paneljudge R package. Pairs of haploid genotypes were simulated at each locus of a panel, using microhaplotype frequencies estimated from a given parasite population (Colombia, Thailand or Senegal, as shown in columns from left to right). Genotype pairs were simulated at three levels of relatedness: unrelated (relatedness = 0.01), related (relatedness = 0.50) and clonal (relatedness = 0.99), as shown in the rows from top to bottom. Relatedness estimates of these pairs were classified using their 95% confidence intervals (LCI = lower limit of the 95% confidence interval, UCI = upper limit of the 95% confidence interval). Estimates could be correctly classified, misclassified or unclassified, as described in the grey box. Each bar represents the proportion of simulations per condition ($n = 400$) classified in each category. Bars that are filled with a colour represent correctly classified simulations, bars that are hashed represent misclassified simulations and bars that are filled with white represent simulations that were unable to be classified. The colours of the bars represent the panel used in that set of simulations. (b) Empirical AMPLseq results recapitulate WGS-based relatedness inference. Points represent relatedness estimates (hmmIBD; Schaffner et al., 2018, "fract_sites_IBD" computed under default settings) for pairs of Guyanese samples using WGS ($n = 9408$ variants) vs. AMPLseq ($n = 220$ variants, from within 128 AMPLseq *Plasmodium falciparum* loci)

3.3 | Geographical attribution

We again engineered amplicon data *in silico* to evaluate the relative signal in genotyping panels for geographical attribution of samples. We subsampled WGS variant calls, called microhaplotypes, and evaluated these data using PCA (Figure S5) and t-SNE visualizations (Figure 5). We found that all three larger panels (AMPLseq, SpotMalaria, and Paragon) distinguished non-African samples by country, and these panels separated East African (Malawi) from West African samples (Mali/Senegal) to varying degrees; no panel was able to distinguish between Malian and Senegalese samples in these visualizations. Results from both the 24 SNP barcode (Daniels et al., 2008) and 4CAST distinguished samples by continent of origin, though not by country. We also added empirical AMPLseq data from five Guyanese samples (C3-GUY, C4-GUY, C5-GUY, C7-GUY and C8-GUY) and WGS data subsampled to AMPLseq coordinates for Venezuelan sample SPT26229 (Figure S6). The AMPLseq samples formed a small cluster beside the WGS-based Guyanese and French Guianese samples. The Venezuelan sample SPT26229 also placed on the perimeter of the Guyana/French Guiana sample cluster, sharing the same axis-2 position as the empirical AMPLseq points. The results suggest that empirical AMPLseq data can distinguish autochthonous samples from the Guiana Shield, and we expect geographical attribution in the region to improve as more data are collected from infections originating in Venezuela and other undersampled localities.

3.4 | COI estimation

We evaluated COI estimation based on 4CAST as opposed to the single locus *AMA1*, which is commonly used for this purpose, alone or together with a single additional locus (Lerch et al., 2017; Miller et al., 2017; Nelson et al., 2019). We engineered *in silico* samples with COI ranging from 2 to 10 (100 engineered samples per COI level) and used the maximum number of unique microhaplotypes present at any locus as a simple data summary method to estimate COI. 4CAST provided more accurate estimates of COI than *AMA1* alone in these simulated data, especially at simulated COI levels between 5 and 7. (Figure 6a). Estimation improved at engineered COI = 8 using AMPLseq (Figure S7), but to reap the full benefit of the larger panel in practice will require an inferential approach that accounts for both

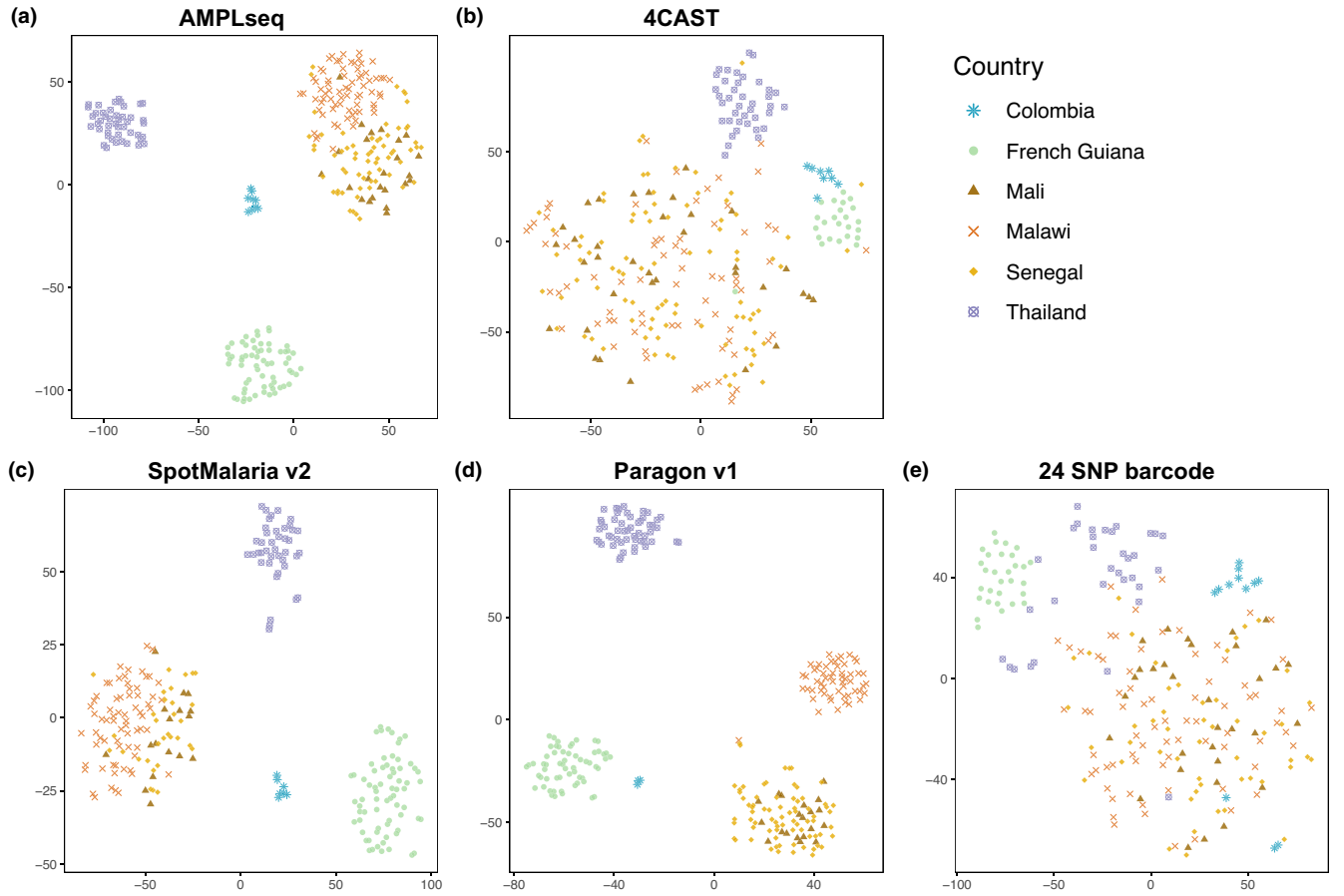


FIGURE 5 *In silico* geographical attribution comparison among panels. Visualization of WGS data subsetted to coordinates of genotyping panels. Microhaplotypes called at each locus were visualized using tSNE representation, with parameter Θ of 0.0, 5000 iterations, and a perplexity parameter of 30. Each point represents a single sample, with colour and shape representing its country of origin (which was not included in the tSNE algorithm). One genotyping panel is visualized in each plot: (a) AMPLseq, (b) 4CAST, (c) SpotMalaria v2, (d) Paragon v1, (e) 24 SNP barcode

chance sharing of alleles using population allele frequencies and for variable coverage/sensitivity among loci.

In the absence of such an algorithm, we proceeded with the simple data summary method above to classify COI in the Malian and Guyanese clinical samples. Only a single polyclonal infection (C6-GUY) occurred among Guyanese samples assayed by 4CAST. The repeated detection of two *CSP* and *SERA2* alternate alleles at depths ranging from 32 to 168 read-pairs enabled unambiguous COI = 2 classification for the sample. WGS sequencing coverage, by contrast, detected only moderately elevated SNP heterozygosity (1.9%) in C6-GUY and this elevation was not sufficient to classify COI > 1 via The Real McCoil (Chang et al., 2017) (Figure S8). AMPLseq also identified COI = 2 for C6-GUY but without consistent support (two loci presenting two alleles in replicate 1 and six loci presenting two alleles in replicate 2). Six additional Guyanese samples were assayed by AMPLseq and one was classified as COI = 2. This sample (A5-GUY) gave a stronger minor variant signal in both AMPLseq (15 loci presenting two alleles in both replicates) and WGS data (10.9% SNP heterozygosity; Figure S8).

For the Malian sample set, 4CAST and AMPLseq both classified samples E5-PST030 and C6-PST063 as monoclonal and all other samples as polyclonal based on the presence/absence of sample loci with multiple alleles. Sensitivity for minor strains was depth-dependent in both assays, as reducing 4CAST depth via subsampling reduced estimated COI (Figure 6b; Figure S9A) and increasing AMPLseq depth via sequencing batch size reduction increased estimated COI (Figure S9B,C). These results emphasize the importance of incorporating read-depth and expected diversity metrics per locus into COI inference algorithms, especially when depth is not concentrated or evenly spread across a panel's most polymorphic loci.

3.5 | Longitudinal sampling: distinguishing continuing vs. newly acquired infections

We used 4CAST to examine longitudinal samples that were likely to be diverse and polyclonal. We sequenced samples from two asymptomatic individuals in the longitudinal Mali cohort over multiple bi-weekly visits during the transmission season (July to December; Figure 7; Tran et al., 2013). In the first individual (Figure 7a), we detected a single microhaplotype at each locus that was present in the first two time points, suggesting a continued infection during

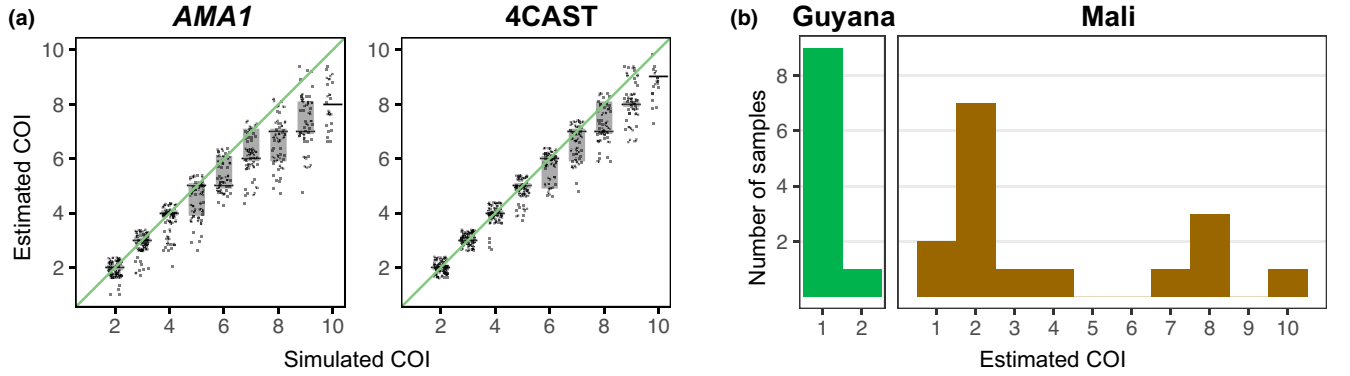
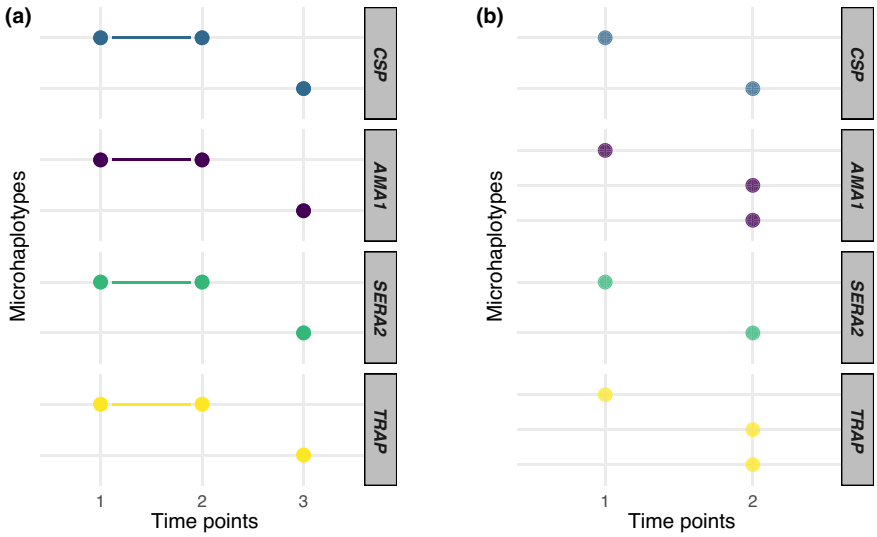


FIGURE 6 *In silico* and empirical complexity of infection (COI) inference. (a) Scatter plots of estimated COI for samples engineered *in silico* from combinations of monoclonal WGS data, subsetted to the loci of interest (*AMA1* locus or 4CAST loci). The *x*-axis represents the number of monoclonal genomes combined into each simulation, and the *y*-axis represents the COI estimated using the simulated data. COI was naively estimated as the maximum number of unique microhaplotypes present at any locus per sample ($n = 100$ samples per condition). Each point represents a sample, jittered for visibility. The black bars represent the median and light grey boxes represent the 25th–75th quantiles. (b) Estimated COI for clinical samples sequenced using 4CAST. COI was again estimated as the maximum number of unique microhaplotypes present at any locus in the sample

FIGURE 7 Longitudinal tracking of infections using 4CAST. Identification of distinct microhaplotypes present in samples from two individuals. The *x*-axis represents consecutive time points, and the *y*-axis represents the individual microhaplotypes identified, grouped by locus. Coloured points represent the presence of that microhaplotype, connected when present in multiple visits



the 2 weeks between sample collection. At the third time point, we detected a single, distinct microhaplotype at each locus, suggesting that a new infection had occurred and the original infection had disappeared or decreased below our limit of detection (<10 parasites/μl). In the second individual (Figure 7b), we detected a similar pattern of strain turnover: a monoclonal infection at the first time point, followed by a distinct, polyclonal infection at the second time point. We also examined two individuals over a longer series of visits (Figure S10), in which we detected a series of polyclonal infections, with some strains sustained over many time points (Figure S10A), and a series of distinct monoclonal infections (Figure S10B). In all cases, the individuals were asymptomatic and did not receive antimalarial treatment between visits; however, these simple examples demonstrate the clarity that 4CAST can bring to tracking infection turnover in longitudinal studies and suggests its potential in distinguishing recrudescence vs. reinfection in therapeutic efficacy studies.

3.6 | Drug resistance profiling

AMPLseq loci in *dhfr*, *mdr1*, *dhps*, *kelch13*, and *mdr2* contain 10 sequence regions that code for various amino acid (AA) polymorphisms that have previously been associated with resistance to antimalarial drugs (Ariey et al., 2014; Miotto et al., 2015; Mita et al., 2007; Veiga et al., 2016). Fourteen of these 18 positions of interest contained nonsynonymous mutations in Malian and Guyanese clinical samples of this study (Figure 8). Positions of interest that lacked mutations across both sample sets were *dhfr* AA 164, *dhps* AA 613, *kelch*13 AA 580 and *mdr*2 AA 484. All Guyanese sequences shared the same mutant alleles at many loci, suggesting fixed mutant alleles. Malian samples, by contrast, did not show fixed mutant alleles at any amino acid position of interest. A mix of mutant and wild-type alleles occurred among Malian samples for *dhfr* AA 51, 59 and 108; *mdr*1 AA 86, 184 and 1246; *dhps* AA 436 and 437; and *mdr*2 AA 492. A previously reported synonymous

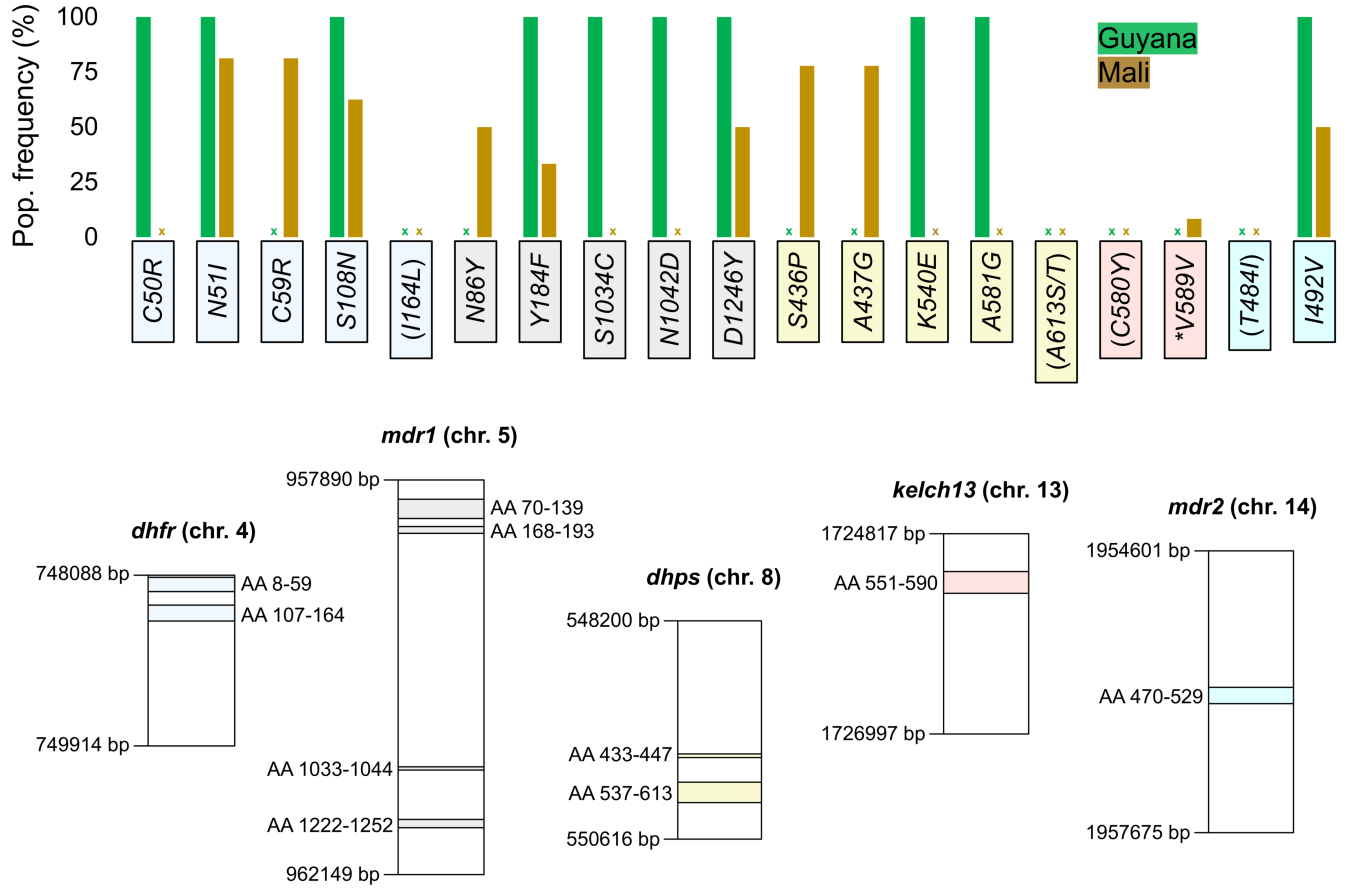


FIGURE 8 Drug resistance-associated sequence profiling in Guyanese and Malian clinical samples. Bars in the top plot indicate the occurrence of various drug resistance-associated amino acid changes within AMPLseq loci. Positions of interest assayed by AMPLseq but without mutant alleles (see x-marks) in the clinical samples profiled here are labelled in parentheses. Positions 484 and 492 in *mdr2* have been suggested to be involved in artemisinin resistance despite lack of experimental data showing an association with a clinical phenotype (Chenet et al., 2017; Miotto et al., 2015). Bottom plot indicates chromosomal and amino acid (AA) positions of each drug resistance-associated AMPLseq locus (excluding primer binding sites). Asterisk indicates synonymous mutation within *kelch13*

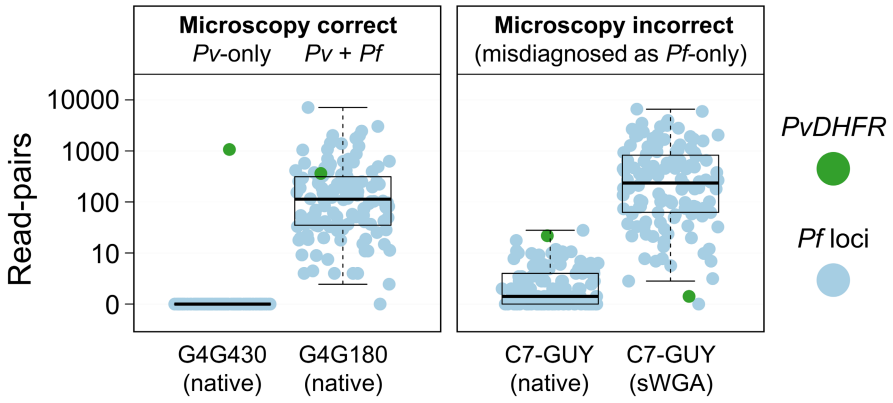


FIGURE 9 *Plasmodium vivax* detection by AMPLseq. The left panel demonstrates strong read-pair support for *PvDHFR* (green circle) in native control samples previously suggested to contain *P. vivax* (*Pv*; G4G430) and *P. vivax* + *P. falciparum* (*Pv* + *Pf*; G4G180) via microscopy. The right panel shows native and sWGA results for C7-GUY, a clinical sample that appears to have been misdiagnosed as *Pf*-only prior to AMPLseq. Blue circles represent read-pair support for *P. falciparum* loci. Positions on the *y*-axis indicate median read-pair support across two sample replicates. Box and whiskers indicate quartiles

polymorphism was observed in one Malian sample at *kelch*13 AA 589 (Taylor et al., 2015).

3.7 | *P. falciparum* and *P. vivax* co-infection detection

To test the ability of AMPLseq to detect *P. vivax* co-infections via co-amplification of *PvDHFR*, two additional Guyanese blood spot samples that had been diagnosed as *P. vivax*-only (G4G430) and *P. vivax* + *P. falciparum* co-infection (G4G180) via microscopy were included in the sample set. These samples did not undergo sWGA.

PvDHFR was detected at high depth in both samples (1068–1822 read-pairs for G4G430 and 234–560 read-pairs for G4G180) (Figure 9). Only G4G180 also showed read-pair support at *P. falciparum* loci (>10 read-pairs at 100–115 loci). *PvDHFR* was not detected in any native or pre-amplified 3D7 or mixed-strain (3D7 + Dd2) templates. This demonstrates high specificity of both *PvDHFR* and *P. falciparum* AMPLseq primers to their intended target species without any apparent amplification inhibition by the presence of congeneric DNA.

PvDHFR was also detected at low levels (16–30 read-pairs) in both native template replicates of C7-GUY, one of the 16 Guyanese samples previously diagnosed as *P. falciparum*-only via microscopy. Surprisingly, two *PvDHFR* read-pairs were also detected in one of the two sWGA replicates from the sample, despite the expectation that sWGA would primarily amplify *P. falciparum* sequences. The sensitivity of *PvDHFR* detection in pre-amplified samples could be enhanced by adding *PvDHFR* primers to the *P. falciparum* sWGA primer pool. *PvDHFR* detection did not occur in any Malian sample, consistent with low prevalence of *P. vivax in* West Africa relative to the Guiana Shield.

4 | DISCUSSION

The utility of AmpSeq for molecular surveillance of infectious diseases is evidenced by the growing number of protocols recently published or under development for *Plasmodium* and other pathogen taxa (Aydemir et al., 2018; Fola et al., 2020; Jacob et al., 2021; Mitchell et al., 2021; Moser et al., 2021; Ruybal-Pesántez et al., 2021; Schwabl et al., 2020; Tessema et al., 2020). Here, we demonstrate the performance of two new panels for *P. falciparum*, designed to serve different use cases and exhibiting different per-sample costs and levels of complexity. Our comparative analyses of these two new panels, 4CAST and AMPLseq, relative to previously published genotyping panels demonstrate that they perform comparably to existing panels of similar composition across use cases, in a diversity of geographical settings, despite different geographical representation in the population genomic data used to inform their designs. This suggests that *de novo* custom panel design may not be required for accurate COI and relatedness estimation in parasite populations from previously unstudied geographical regions. We therefore suggest that future implementation of these panels and future designs for other organisms should be guided by three criteria: (i) the intended use cases for the data; (ii) protocol complexity and compatibility with available instruments and expertise; and (iii) protocol customizability for locally relevant genetic loci.

Considering the first of these criteria, intended use case, our investigations above suggest a straightforward mapping of panels by size and feature to use case. The small 4CAST panel is well suited to COI estimation (Figure 6), and profiles four highly diverse antigens for the same effort and cost traditionally used to profile a single locus. Because of the very high diversity of the loci in the 4CAST panel in most parasite populations, this panel is also well suited to any application requiring genetic delineation of distinct parasite lineages (Figure 7). In therapeutic efficacy studies, for example, it is essential to determine whether subjects who become parasitaemic following drug treatment are exhibiting a recrudescence of an incompletely cleared strain from the initial infection (which could indicate treatment failure), or if they have become reinfected with a distinct parasite strain subsequent to treatment. We suggest that the 4CAST panel would be significantly more informative than traditional genotyping approaches used in therapeutic efficacy studies, such as profiling length polymorphisms or allele-specific amplification in the *msp1/msp2/glurp* genes (Reeder & Marshall, 1994; Snounou, 2002), especially if coupled with an inferential approach that accounts for some chance sharing of alleles dependent on their population frequencies. 4CAST is also more cost-effective than independent monoplex amplification and Illumina sequencing of individual loci (Early et al., 2019; Gruenberg et al., 2019; Lerch et al., 2017).

Our work demonstrates that the AMPLseq panel performs comparably to two existing multiplexed amplicon sequencing panels of similar size (Jacob et al., 2021; Tessema et al., 2020) for any use case reliant on estimation of parasite relatedness (Figure 4), despite different design criteria and data sets that informed the panels. Potential public health use cases that employ relatedness information include measuring the connectivity of parasites between locations to define units of control, and monitoring changes in the level of transmission (Cerqueira et al., 2017; Daniels et al., 2015; Knudson et al., 2020). The AMPLseq panel and its peers are well suited to detecting imported vs. local infections given their capacity to distinguish parasites from distinct countries, as long as population genetic differentiation is sufficiently high (Figure 5). Finally, the larger panels offer the capacity to monitor genetic markers associated with drug resistance (Figure 8) or, in some panels, detect co-infection with other *Plasmodium* species (Figure 9).

The second panel selection criterion, protocol complexity and compatibility with available instruments, should be prefaced with a reminder that all of these protocols employ nested PCRs as the fundamental mechanism to produce sequencing libraries targeting small genomic regions of interest. Equipped with a few key instruments, most laboratories with access to pre- and post-PCR hood space are probably capable of the nested PCR library construction approach. Key instruments needed are a centrifuge, thermocycler, vortexer, magnetic rack, fragment sizer (e.g., Bioanalyzer or TapeStation) and

DNA quantitation device (e.g., Qubit fluorometer). The latter items are necessary for careful clean-up of inappropriately large or small DNA molecules and precise quantification of libraries for optimal loading on the sequencing flow cell. Sequencing can be performed using a small instrument such as the Illumina iSeq 100 or on larger platforms shared by multiple groups. Batch sizes (i.e., number of samples pooled in a library) will depend on the available platform and on the read-depth needed for the study objective and transmission system at hand. COI inference in high-transmission regions, for example, typically requires greater depth (potentially smaller batch sizes) than does relatedness analysis focusing on monoclonal parasites. Another important consideration for most Illumina-based amplicon sequencing protocol is the high level of precaution required to avoid sample contamination. PCRs should be conducted in dedicated hoods using dedicated pipettes, ideally also with downstream sample processing confined to rooms or locations physically removed from those where native templates are processed. A centrifuge (or plate spinner) is listed as a requirement above primarily for its role in moving sample liquids away from the top seals of plates, which should always be handled with special care.

Though the AmpSeq protocols highlighted herein share many common features, they differ in other aspects that may impact implementation. Whereas the 4CAST and Paragon panels perform well on native DNA (both tested with samples as low as 10 parasites/μl), the AMPLseq and SpotMalaria panels require sWGA preamplification prior to the first PCR to ensure adequate performance for low-parasitaemia samples, which may represent a significant proportion of samples in some settings. The sWGA step occurs using an isothermal amplification protocol that is relatively simple to perform but requires expensive phi29 DNA polymerase and a magnetic bead-based post-reaction cleanup of individual samples, for an approximate additional cost of $8 USD per sample at the time of writing. Though not large in absolute terms, this cost is comparable to the cost of the AMPLseq or 4CAST protocols themselves, which range from $5 to $10 USD per sample without sWGA, depending on details of implementation such as sequencing instrument and sample indexing per run. The larger panels additionally employ differing numbers of first-round PCRs and require a varying number of magnetic bead-based cleanups to tailor the length profile of intermediate products (summarized in Table S5), which means that the local capacity for automating the bead-based cleanups is a relevant implementation consideration.

An additional limitation demonstrated here for AMPLseq regards incomplete panel amplification from relatively low-parasitaemia and/or older sample material. This limits potential application to older and/or degraded sample collections unless lower batch sizes (Figure S9C) or larger (more expensive) sequencing instruments are utilized. The impact of DNA integrity on panel performance should be further assessed in future work. Additional performance assessment using parasitaemias formulated prior to DNA extraction (e.g., diluting ring-stage parasites) would also enhance future sensitivity tests.

The third criterion for panel selection, customizability, may be most relevant for the drug resistance surveillance use case, given differences in the geographical distribution of important drug resistance markers, and varying coverage of known markers by the existing panels. All of the protocols are amenable to customization through the addition of independent target amplifications in the first round of PCR, which could be combined with other first-round PCR multiplex products prior to the second PCR. A more elegant customization approach would be to add (or subtract) targets from the first-round PCR. While complicated bioinformatic pipelines are typically necessary in the design of large multiplexes, in our experience, small multiplexes like 4CAST, which was made from preexisting primer-pairs designed independently, may simply function without optimization, and could presumably be augmented with a small number of additional loci. Though the AMPLseq multiplex of 129 PCR loci benefited from careful design of the original panel via GTseek, we successfully added 4CAST, drug resistance and PvDHFR loci to the GTseek target set without any primer modifications. The AMPLseq panel is thus probably receptive to further augmentation. As the AMPLseq and 4CAST protocols utilize unmodified, commercially available oligos as primers, further customization should be feasible in any setting. However, note that not all targets are amenable to incorporation into the multiplex, as we failed despite multiple attempts to include amplicons targeting the *crt* gene associated with chloroquine resistance (Fidock et al., 2000), or the *hrp2/3* genes, which can contain deletions that lead to false-negative diagnosis via rapid diagnostic test (Gamboa et al., 2010).

The proliferation of new AmpSeq protocols for molecular surveillance of infectious and noninfectious organisms raises the important question of whether it is valuable for each field to converge on a single approach or common panel per organism. Factors precluding a completely homogeneous approach include varying instrumentation, expertise and use cases for the data across settings, in addition to elucidation of new drug resistance markers of interest and an anticipated onward diversification of targeted genotyping technologies. For example, in the *Plasmodium* field, other productive multiplexed genotyping approaches include molecular inversion probes (Aydemir et al., 2018; Moser et al., 2021) or multicopy (*var*) gene metabarcoding analysis (Day et al., 2017). Factors favouring convergence within the AmpSeq domain include opportunities for coordinated procurement of instruments and reagents in low- and middle-income countries, and opportunities to directly compare observations between studies and surveillance efforts led by different groups. This latter factor, which we term portability of analyses, has the potential to provide regional or global insight through syntheses across studies. However, the portability of certain analyses is hampered by ascertainment bias, an inherent limitation of any targeted sequencing approach for analyses based on the genotypic state of select loci in different countries. That is, a panel designed based on observations of genetic diversity through WGS in countries A and B may not provide a fair means of comparing diversity in countries C vs. D, if diversity there is distributed differently in the genome than in countries A and B. WGS is the ultimate tool for avoiding this bias. However, the problems of comparing populations profiled with different panels may be mitigated by comparing inferred relatedness levels within

populations rather than actual genotypic diversity measures (Taylor et al., 2019). Overlap of loci among panels would further facilitate direct estimation of relatedness between samples included in different studies, as in Carrasquilla et al. (2022), where the importance of confidence intervals around relatedness estimates is highlighted. The AMPLseq panel we describe here contains a significant number ($n = 47$) of targets from the SpotMalaria panel, and we expect that future *P. falciparum* panel designs will also tend to exhibit some degree of overlap with other panels, both by deliberate design and through blind convergence based on key genomic features, such as high diversity and sequence amenability to PCR primer design.

As molecular surveillance efforts for malaria parasites and other organisms are more widely adopted and become increasingly diverse, it will be essential for the community to develop standardized approaches for the design, validation, interpretation and sharing of targeted amplicon sequencing data. The paneljudge R package described here provides an excellent means to comparatively evaluate existing and hypothetical panel performance via data collected from previous population genomic surveys, and the bioinformatic analysis pipelines we have developed are suitable for interpreting Illumina data from diverse targets and panels in different organisms. We anticipate the growth of this field and the development of new analytical tools to extract even more epidemiological and ecological insight from increasingly large AmpSeq data sets for diverse taxa.

CONFLICT OF INTERESTS

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

This project was funded in whole or in part with federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services, under Grant No. U19AI110818 to the Broad Institute. This project was also supported by an NIH R01 award to D.E.N. (R01AI141544), an award from the Bill and Melinda Gates Foundation to D.E.N. and C.O.B. (OPP1213366), and a Broad Institute NextGen Award to B.M. The Mali cohort study was funded by the Division of Intramural Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health. The Colombian cohort study was supported by British Council Newton-Caldas Fund Institutional Links Award G1854. We thank MalariaGen for use of the Colombian WGS data. We thank Annie Laws for project management. We thank Dr Nathan Campbell for assistance in AMPLseq panel design and evaluation.

AUTHOR CONTRIBUTIONS

Conceptualization: EL, PS, MC, ART, AME, BLM, DEN. Data curation: ZMJ, RP, TJS. Formal analysis: EL, PS, MC, ART, RP, TJS. Funding acquisition: COB, BLM, DEN. Investigation: PS, MC, ZMJ, MS, RK, SW. Methodology: EL, MC, PS, ART, ZMJ, MS, RP, TJS, RK. Project administration: BLM, DEN. Resources: CMA, SP, PDC, BT, JCR, VC, KJ, HC. Software: ART, RP, TJS, AME. Supervision: COB, BLM, AME, DEN. Validation: EL, PS, MC, ZMJ, MS, RP, TJS, RK. Visualization: EL, PS, MC, ART, RP. Writing—original draft preparation: EL, PS, MC, ART, DEN. Writing—review & editing (All authors).

OPEN RESEARCH BADGES

This article has earned an Open Data Badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. The data is available at http://www.ncbi.nlm.nih.gov/sra

DATA AVAILABILITY STATEMENT

All amplicon sequencing data, as well as WGS data from 16 Guyana samples, were submitted to the NCBI Sequence Read Archive (http://www.ncbi.nlm.nih.gov/sra) under accession PRJNA758191. This publication uses data from the MalariaGEN *Plasmodium falciparum* Community Project as described online pending publication and public release of data set Pf7 (https://www.malariagen.net/resource/34); additional accession numbers are available in Table S3. Previously published data from Guyana and French Guiana are available from SRA BioProjects PRJNA543530 and PRJNA242182, respectively. Software and documentation are available at https://github.com/broadinstitute/AmpSeQC (AmpSeQC pipeline), https://github.com/broadinstitute/malaria-amplicon-pipeline.git (Malaria amplicon pipeline) and https://github.com/artaylor85/paneljudge (paneljudge).

Benefits Generated: A research collaboration was developed with scientists from the countries providing genetic samples, and all principal collaborators are included as co-authors. The results of this research have been shared with the provider communities and the broader scientific community (see above), and the research addresses a priority concern, in this case the public health concern of malaria. More broadly, our group is committed to international scientific partnerships, as well as institutional capacity building.

ORCID

Emily LaVerriere https://orcid.org/0000-0002-4750-4199
Philipp Schwabl https://orcid.org/0000-0002-1244-945X
Manuela Carrasquilla https://orcid.org/0000-0002-7953-4376
Meg Shieh https://orcid.org/0000-0002-0483-5000
Ruchit Panchal https://orcid.org/0000-0002-4664-4501

REFERENCES

Ariey, F., Witkowski, B., Amaratunga, C., Beghain, J., Langlois, A.-C., Khim, N., Kim, S., Duru, V., Bouchier, C., Ma, L., Lim, P., Leang, R., Duong, S., Sreng, S., Suon, S., Chuor, C. M., Bout, D. M., Ménard, S., Rogers, W. O., … Ménard, D. (2014). A molecular marker of artemisinin-resistant Plasmodium falciparum malaria. *Nature*, *505*(7481), 50–55. https://doi.org/10.1038/nature12876

Aydemir, O., Janko, M., Hathaway, N. J., Verity, R., Mwandagalirwa, M. K., Tshefu, A. K., Tessema, S. K., Marsh, P. W., Tran, A., Reimonn, T., Ghani, A. C., Ghansah, A., Juliano, J. J., Greenhouse, B. R., Emch, M., Meshnick, S. R., & Bailey, J. A. (2018). Drug-resistance and population structure of plasmodium falciparum across the democratic republic of Congo using high-throughput molecular inversion

probes. *The Journal of Infectious Diseases*, *218*(6), 946–955. https://doi.org/10.1093/infdis/jiy223

Baetscher, D. S., Clemento, A. J., Ng, T. C., Anderson, E. C., & Garza, J. C. (2018). Microhaplotypes provide increased power from short-read DNA sequences for relationship inference. *Molecular Ecology Resources*, *18*(2), 296–305. https://doi.org/10.1111/1755-0998.12737

Bybee, S. M., Bracken-Grissom, H., Haynes, B. D., Hermansen, R. A., Byers, R. L., Clement, M. J., Udall, J. A., Wilcox, E. R., & Crandall, K. A. (2011). Targeted amplicon sequencing (TAS): A scalable next-gen approach to multilocus, multitaxa phylogenetics. *Genome Biology and Evolution*, *3*, 1312–1323. https://doi.org/10.1093/gbe/evr106

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, *13*(7), 581–583. https://doi.org/10.1038/nmeth.3869

Campbell, N. R., Harmon, S. A., & Narum, S. R. (2015). Genotyping-in-Thousands by sequencing (GT-seq): A cost effective SNP genotyping method based on custom amplicon sequencing. *Molecular Ecology Resources*, *15*(4), 855–867. https://doi.org/10.1111/1755-0998.12357

Cantacessi, C., Dantas-Torres, F., Nolan, M. J., & Otranto, D. (2015). The past, present, and future of Leishmania genomics and transcriptomics. *Trends in Parasitology*, *31*(3), 100–108. https://doi.org/10.1016/j.pt.2014.12.012

Carrasquilla, M., Early, A. M., Taylor, A. R., Knudson, A., Echeverry, D. F., Anderson, T. J., Mancilla, E., Aponte, S., Cárdenas, P., Buckee, C. O., Rayner, J. C., Sáenz, F. E., Neafsey, D. E., & Corredor, V. (2022). Resolving drug selection and migration in an inbred South American Plasmodium falciparum population with identity-by-descent analysis (p. 2022.02.18.480973). *bioRxiv*. https://doi.org/10.1101/2022.02.18.480973

Cerqueira, G. C., Cheeseman, I. H., Schaffner, S. F., Nair, S., McDew-White, M., Phyo, A. P., Ashley, E. A., Melnikov, A., Rogov, P., Birren, B. W., Nosten, F., Anderson, T. J. C., & Neafsey, D. E. (2017). Longitudinal genomic surveillance of Plasmodium falciparum malaria parasites reveals complex genomic architecture of emerging artemisinin resistance. *Genome Biology*, *18*(1), 78. https://doi.org/10.1186/s13059-017-1204-4

Chang, H.-H., Wesolowski, A., Sinha, I., Jacob, C. G., Mahmud, A., Uddin, D., Zaman, S. I., Hossain, M. A., Faiz, M. A., Ghose, A., Sayeed, A. A., Rahman, M. R., Islam, A., Karim, M. J., Rezwan, M. K., Shamsuzzaman, A. K. M., Jhora, S. T., Aktaruzzaman, M. M., Drury, E., … Buckee, C. (2019). Mapping imported malaria in Bangladesh using parasite genetic and human mobility data. *Elife*, *8*, e43481. https://doi.org/10.7554/eLife.43481

Chang, H.-H., Worby, C. J., Yeka, A., Nankabirwa, J., Kamya, M. R., Staedke, S. G., Dorsey, G., Murphy, M., Neafsey, D. E., Jeffreys, A. E., Hubbart, C., Rockett, K. A., Amato, R., Kwiatkowski, D. P., Buckee, C. O., & Greenhouse, B. (2017). THE REAL McCOIL: A method for the concurrent estimation of the complexity of infection and SNP allele frequency for malaria parasites. *PLoS Computational Biology*, *13*(1), e1005348. https://doi.org/10.1371/journal.pcbi.1005348

Chenet, S. M., Akinyi Okoth, S., Huber, C. S., Chandrabose, J., Lucchi, N. W., Talundzic, E., Krishnalall, K., Ceron, N., Musset, L., Macedo de Oliveira, A., Venkatesan, M., Rahman, R., Barnwell, J. W., & Udhayakumar, V. (2016). Independent emergence of the plasmodium falciparum Kelch propeller domain mutant allele C580Y in Guyana. *The Journal of Infectious Diseases*, *213*(9), 1472–1475. https://doi.org/10.1093/infdis/jiv752

Chenet, S. M., Okoth, S. A., Kelley, J., Lucchi, N., Huber, C. S., Vreden, S., Macedo de Oliveira, A., Barnwell, J. W., Udhayakumar, V., & Adhin, M. R. (2017). Molecular profile of malaria drug resistance markers of plasmodium falciparum in Suriname. *Antimicrobial Agents and Chemotherapy*, *61*(7), e02655–e2716. https://doi.org/10.1128/AAC.02655-16

Dalmat, R., Naughton, B., Kwan-Gett, T. S., Slyker, J., & Stuckey, E. M. (2019). Use cases for genetic epidemiology in malaria elimination. *Malaria Journal*, *18*(1), 1–11. https://doi.org/10.1186/s12936-019-2784-0

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., & Durbin, R. & 1000 Genomes Project Analysis Group (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Daniels, R. F., Schaffner, S. F., Wenger, E. A., Proctor, J. L., Chang, H.-H., Wong, W., Baro, N., Ndiaye, D., Fall, F. B., Ndiop, M., Ba, M., Milner, D. A., Taylor, T. E., Neafsey, D. E., Volkman, S. K., Eckhoff, P. A., Hartl, D. L., & Wirth, D. F. (2015). Modeling malaria genomics reveals transmission decline and rebound in Senegal. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(22), 7067–7072. https://doi.org/10.1073/pnas.1505691112

Daniels, R. F., Volkman, S. K., Milner, D. A., Mahesh, N., Neafsey, D. E., Park, D. J., Rosen, D., Angelino, E., Sabeti, P. C., Wirth, D. F., & Wiegand, R. C. (2008). A general SNP-based molecular barcode for Plasmodium falciparum identification and tracking. *Malaria Journal*, *7*(1), 223. https://doi.org/10.1186/1475-2875-7-223

DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., Philippakis, A. A., del Angel, G., Rivas, M. A., Hanna, M., McKenna, A., Fennell, T. J., Kernytsky, A. M., Sivachenko, A. Y., Cibulskis, K., Gabriel, S. B., Altshuler, D., & Daly, M. J. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, *43*(5), 491–498. https://doi.org/10.1038/ng.806

Dupuis, J. R., Bremer, F. T., Kauwe, A., San Jose, M., Leblanc, L., Rubinoff, D., & Geib, S. M. (2018). HiMAP: Robust phylogenomics from highly multiplexed amplicon sequencing. *Molecular Ecology Resources*, *18*(5), 1000–1019. https://doi.org/10.1111/1755-0998.12783

Early, A. M., Daniels, R. F., Farrell, T. M., Grimsby, J., Volkman, S. K., Wirth, D. F., MacInnis, B. L., & Neafsey, D. E. (2019). Detection of low-density Plasmodium falciparum infections using amplicon deep sequencing. *Malaria Journal*, *18*(1), 219. https://doi.org/10.1186/s12936-019-2856-1

Fidock, D. A., Nomura, T., Talley, A. K., Cooper, R. A., Dzekunov, S. M., Ferdig, M. T., Ursos, L. M., Sidhu, A. B., Naudé, B., Deitsch, K. W., Su, X. Z., Wootton, J. C., Roepe, P. D., & Wellems, T. E. (2000). Mutations in the P. falciparum digestive vacuole transmembrane protein PfCRT and evidence for their role in chloroquine resistance. *Molecular Cell*, *6*(4), 861–871. https://doi.org/10.1016/s1097-2765(05)00077-8

Fola, A. A., Kattenberg, E., Razook, Z., Lautu-Gumal, D., Lee, S., Mehra, S., Bahlo, M., Kazura, J., Robinson, L. J., Laman, M., Mueller, I., & Barry, A. E. (2020). SNP barcodes provide higher resolution than microsatellite markers to measure Plasmodium vivax population genetics. *Malaria Journal*, *19*(1), 375. https://doi.org/10.1186/s12936-020-03440-0

Galinsky, K., Valim, C., Salmier, A., de Thoisy, B., Musset, L., Legrand, E., Faust, A., Baniecki, M., Ndiaye, D., Daniels, R. F., Hartl, D. L., Sabeti, P. C., Wirth, D. F., Volkman, S. K., & Neafsey, D. E. (2015). COIL: a methodology for evaluating malarial complexity of infection using likelihood from single nucleotide polymorphism data. *Malaria Journal*, *14*(1), 4. https://doi.org/10.1186/1475-2875-14-4

Gamboa, D., Ho, M.-F., Bendezu, J., Torres, K., Chiodini, P. L., Barnwell, J. W., Incardona, S., Perkins, M., Bell, D., McCarthy, J., & Cheng, Q. (2010). A large proportion of *P. falciparum* Isolates in the Amazon Region of Peru Lack pfhrp2 and pfhrp3: Implications for malaria rapid diagnostic tests. *PLoS One*, *5*(1), https://doi.org/10.1371/journal.pone.0008091

Gardy, J. L., & Loman, N. J. (2018). Towards a genomics-informed, real-time, global pathogen surveillance system. *Nature Reviews Genetics*, *19*, 9–20. https://doi.org/10.1038/nrg.2017.88

Gruenberg, M., Lerch, A., Beck, H.-P., & Felger, I. (2019). Amplicon deep sequencing improves Plasmodium falciparum genotyping in clinical trials of antimalarial drugs. *Scientific Reports*, *9*(1), 1–12. https://doi.org/10.1038/s41598-019-54203-0

Hargrove, J. S., McCane, J., Roth, C. J., High, B., & Campbell, M. R. (2021). Mating systems and predictors of relative reproductive success in a Cutthroat Trout subspecies of conservation concern. *Ecology and Evolution*, *11*(16), 11295–11309. https://doi.org/10.1002/ece3.7914

Helb, D. A., Tetteh, K. K. A., Felgner, P. L., Skinner, J., Hubbard, A., Arinaitwe, E., Mayanja-Kizza, H., Ssewanyana, I., Kamya, M. R., Beeson, J. G., Tappero, J., Smith, D. L., Crompton, P. D., Rosenthal, P. J., Dorsey, G., Drakeley, C. J., & Greenhouse, B. (2015). Novel serologic biomarkers provide accurate estimates of recent Plasmodium falciparum exposure for individuals and communities. *Proceedings of the National Academy of Sciences*, *112*(32), E4438–E4447. https://doi.org/10.1073/pnas.1501705112

Henden, L., Lee, S., Mueller, I., Barry, A., & Bahlo, M. (2018). Identity-by-descent analyses for measuring population dynamics and selection in recombining pathogens. *PLoS Genetics*, *14*(5), e1007279. https://doi.org/10.1371/journal.pgen.1007279

Jacob, C. G., Thuy-Nhien, N., Mayxay, M., Maude, R. J., Quang, H. H., Hongvanthong, B., Vanisaveth, V., Ngo Duc, T., Rekol, H., van der Pluijm, R., von Seidlein, L., Fairhurst, R., Nosten, F., Hossain, M. A., Park, N., Goodwin, S., Ringwald, P., Chindavongsa, K., Newton, P., … Miotto, O. (2021). Genetic surveillance in the Greater Mekong subregion and South Asia to support malaria control and elimination. *Elife*, *10*, e62997. https://doi.org/10.7554/eLife.62997

Jones, S., Kay, K., Hodel, E. M., Gruenberg, M., Lerch, A., Felger, I., & Hastings, I. (2021). Should deep-sequenced amplicons become the new gold-standard for analysing malaria drug clinical trials? *Antimicrobial Agents and Chemotherapy*, *65*, AAC0043721. https://doi.org/10.1128/AAC.00437-21

Kayiba, N. K., Yobi, D. M., Tshibangu-Kabamba, E., Tuan, V. P., Yamaoka, Y., Devleesschauwer, B., Mvumbi, D. M., Okitolonda Wemakoy, E., De Mol, P., Mvumbi, G. L., Hayette, M.-P., Rosas-Aguirre, A., & Speybroeck, N. (2021). Spatial and molecular mapping of Pfkelch13 gene polymorphism in Africa in the era of emerging Plasmodium falciparum resistance to artemisinin: A systematic review. *The Lancet Infectious Diseases*, *21*(4), e82–e92. https://doi.org/10.1016/S1473-3099(20)30493-X

Knudson, A., González-Casabianca, F., Feged-Rivadeneira, A., Pedreros, M. F., Aponte, S., Olaya, A., Castillo, C. F., Mancilla, E., Piamba-Dorado, A., Sanchez-Pedraza, R., Salazar-Terreros, M. J., Lucchi, N., Udhayakumar, V., Jacob, C., Pance, A., Carrasquilla, M., Apráez, G., Angel, J. A., Rayner, J. C., & Corredor, V. (2020). Spatio-temporal dynamics of Plasmodium falciparum transmission within a spatial unit on the Colombian Pacific Coast. *Scientific Reports*, *10*(1), 3756. https://doi.org/10.1038/s41598-020-60676-1

Krijthe, J. H. (2015). *Rtsne: T-Distributed stochastic neighbor embedding using a barnes-hut implementation*. https://github.com/jkrijthe/Rtsne.

Lautu-Gumal, D., Razook, Z., Koleala, T., Nate, E., McEwen, S., Timbi, D., Hetzel, M. W., Lavu, E., Tefuarani, N., Makita, L., Kazura, J., Mueller, I., Pomat, W., Laman, M., Robinson, L. J., & Barry, A. E. (2021). Surveillance of molecular markers of Plasmodium falciparum artemisinin resistance (kelch13 mutations) in Papua New Guinea between 2016 and 2018. *International Journal for Parasitology: Drugs and Drug Resistance*, *16*, 188–193. https://doi.org/10.1016/j.ijpddr.2021.06.004

Lefterova, M. I., Budvytiene, I., Sandlund, J., Färnert, A., & Banaei, N. (2015). Simple real-time PCR and amplicon sequencing method for identification of plasmodium species in human whole blood. *Journal of Clinical Microbiology*, *53*(7), 2251–2257. https://doi.org/10.1128/JCM.00542-15

Lerch, A., Koepfli, C., Hofmann, N. E., Messerli, C., Wilcox, S., Kattenberg, J. H., Betuela, I., O'Connor, L., Mueller, I., & Felger, I. (2017). Development of amplicon deep sequencing markers and data analysis pipeline for genotyping multi-clonal malaria infections. *BMC Genomics*, *18*(1), 864. https://doi.org/10.1186/s12864-017-4260-y

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *ArXiv:1303.3997* [Preprint]. http://arxiv.org/abs/1303.3997

Liu, Y., Tessema, S. K., Murphy, M., Xu, S., Schwartz, A., Wang, W., Cao, Y., Lu, F., Tang, J., Gu, Y., Zhu, G., Zhou, H., Gao, Q., Huang, R., Cao, J., & Greenhouse, B. (2020). Confirmation of the absence of local transmission and geographic assignment of imported falciparum malaria cases to China using microsatellite panel. *Malaria Journal*, *19*(1), 244. https://doi.org/10.1186/s12936-020-03316-3

MalariaGEN Plasmodium falciparum Community Project (2016). Genomic epidemiology of artemisinin resistant malaria. *Elife*, *5*, e08714. https://doi.org/10.7554/eLife.08714

Mathieu, L. C., Cox, H., Early, A. M., Mok, S., Lazrek, Y., Paquet, J.-C., Ade, M.-P., Lucchi, N. W., Grant, Q., Udhayakumar, V., Alexandre, J. S., Demar, M., Ringwald, P., Neafsey, D. E., Fidock, D. A., & Musset, L. (2020). Local emergence in Amazonia of Plasmodium falciparum k13 C580Y mutants associated with in vitro artemisinin resistance. *Elife*, *9*, e51015. https://doi.org/10.7554/eLife.51015

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*(9), 1297–1303. https://doi.org/10.1101/gr.107524.110

Miles, A., Iqbal, Z., Vauterin, P., Pearson, R., Campino, S., Theron, M., Gould, K., Mead, D., Drury, E., O'Brien, J., Rubio, V. R., MacInnis, B., Mwangi, J., Samarakoon, U., Ranford-Cartwright, L., Ferdig, M., Hayton, K., Su, X., Wellems, T., … Kwiatkowski, D. (2016). Indels, structural variation, and recombination drive genomic diversity in *Plasmodium falciparum*. *Genome Research*, *26*(9), 1288–1299. https://doi.org/10.1101/gr.203711.115

Miles, A. pyup.io bot, Murillo, R., Ralph, P., Harding, N. J., Pisupati, R., Rae, S., & Millar, T. (2020). *cggh/scikit-allel: v1.3.2*. Zenodo. https://doi.org/https://doi.org/10.5281/zenodo.3976233

Miller, R. H., Hathaway, N. J., Kharabora, O., Mwandagalirwa, K., Tshefu, A., Meshnick, S. R., Taylor, S. M., Juliano, J. J., Stewart, V. A., & Bailey, J. A. (2017). A deep sequencing approach to estimate Plasmodium falciparum complexity of infection (COI) and explore apical membrane antigen 1 diversity. *Malaria Journal*, *16*(1), 490. https://doi.org/10.1186/s12936-017-2137-9

Miotto, O., Amato, R., Ashley, E. A., MacInnis, B., Almagro-Garcia, J., Amaratunga, C., Lim, P., Mead, D., Oyola, S. O., Dhorda, M., Imwong, M., Woodrow, C., Manske, M., Stalker, J., Drury, E., Campino, S., Amenga-Etego, L., Thanh, T.-N.- N., Tran, H. T., … Kwiatkowski, D. P. (2015). Genetic architecture of artemisinin-resistant *Plasmodium falciparum*. *Nature Genetics*, *47*(3), 226–234. https://doi.org/10.1038/ng.3189

Miotto, O., Sekihara, M., Tachibana, S.-I., Yamauchi, M., Pearson, R. D., Amato, R., Gonçalves, S., Mehra, S., Noviyanti, R., Marfurt, J., Auburn, S., Price, R. N., Mueller, I., Ikeda, M., Mori, T., Hirai, M., Tavul, L., Hetzel, M. W., Laman, M., … Mita, T. (2020). Emergence of artemisinin-resistant *Plasmodium falciparum* with kelch13 C580Y mutations on the island of New Guinea. *PLoS Path*, *16*(12), e1009133. https://doi.org/10.1371/journal.ppat.1009133

Mita, T., Tanabe, K., Takahashi, N., Tsukahara, T., Eto, H., Dysoley, L., Ohmae, H., Kita, K., Krudsood, S., Looareesuwan, S., Kaneko, A., Björkman, A., & Kobayakawa, T. (2007). Independent evolution of pyrimethamine resistance in *Plasmodium falciparum* isolates in melanesia. *Antimicrobial Agents and Chemotherapy*, *51*(3), 1071–1077. https://doi.org/10.1128/AAC.01186-06

Mitchell, R. M., Zhou, Z., Sheth, M., Sergent, S., Frace, M., Nayak, V., Hu, B., Gimnig, J., ter Kuile, F., Lindblade, K., Slutsker, L., Hamel, M. J., Desai, M., Otieno, K., Kariuki, S., Vigfusson, Y., & Shi, Y. P. (2021). Development of a new barcode-based, multiplex-PCR, next-generation-sequencing assay and data processing and analytical pipeline for multiplicity of infection detection of Plasmodium falciparum. *Malaria Journal*, *20*(1), 92. https://doi.org/10.1186/s12936-021-03624-2

Moser, K. A., Madebe, R. A., Aydemir, O., Chiduo, M. G., Mandara, C. I., Rumisha, S. F., Chaky, F., Denton, M., Marsh, P. W., Verity, R., Watson, O. J., Ngasala, B., Mkude, S., Molteni, F., Njau, R., Warsame, M., Mandike, R., Kabanywanyi, A. M., Mahende, M. K., … Bailey, J. A. (2021). Describing the current status of Plasmodium falciparum population structure and drug resistance within mainland Tanzania using molecular inversion probes. *Molecular Ecology*, *30*(1), 100–113. https://doi.org/10.1111/mec.15706

Nader, J. L., Mathers, T. C., Swain, M. T., Robinson, G., Chalmers, R. M., Hunter, P. R., van Oosterhout, C., & Tyler, K. M. (2019). Evolutionary genomics of anthroponosis in Cryptosporidium. *Nature Microbiology*, *4*, 14. https://doi.org/10.1038/s41564-019-0377-x

Natesh, M., Taylor, R. W., Truelove, N. K., Hadly, E. A., Palumbi, S. R., Petrov, D. A., & Ramakrishnan, U. (2019). Empowering conservation practice with efficient and economical genotyping from poor quality samples. *Methods in Ecology and Evolution*, *10*(6), 853–859. https://doi.org/10.1111/2041-210X.13173

Neafsey, D. E., Juraska, M., Bedford, T., Benkeser, D., Valim, C., Griggs, A., Lievens, M., Abdulla, S., Adjei, S., Agbenyega, T., Agnandji, S. T., Aide, P., Anderson, S., Ansong, D., Aponte, J. J., Asante, K. P., Bejon, P., Birkett, A. J., Bruls, M., … Wirth, D. F. (2015). Genetic diversity and protective efficacy of the RTS, S/AS01 malaria vaccine. *New England Journal of Medicine*, *373*(21), 2025–2037. https://doi.org/10.1056/NEJMoa1505819

Neafsey, D. E., Taylor, A. R., & MacInnis, B. L. (2021). Advances and opportunities in malaria population genomics. *Nature Reviews Genetics*, *22*(8), 502–517. https://doi.org/10.1038/s41576-021-00349-5

Nelson, C. S., Sumner, K. M., Freedman, E., Saelens, J. W., Obala, A. A., Mangeni, J. N., Taylor, S. M., & O'Meara, W. P. (2019). High-resolution micro-epidemiology of parasite spatial and temporal dynamics in a high malaria transmission setting in Kenya. *Nature Communications*, *10*, 5615. https://doi.org/10.1038/s41467-019-13578-4

O'Neill, E. M., Schwartz, R., Bullock, C. T., Williams, J. S., Shaffer, H. B., Aguilar-Miguel, X., Parra-Olea, G., & Weisrock, D. W. (2013). Parallel tagged amplicon sequencing reveals major lineages and phylogenetic structure in the North American tiger salamander (*Ambystoma tigrinum*) species complex. *Molecular Ecology*, *22*(1), 111–129. https://doi.org/10.1111/mec.12049

Oude Munnink, B. B., Worp, N., Nieuwenhuijse, D. F., Sikkema, R. S., Haagmans, B., Fouchier, R. A. M., & Koopmans, M. (2021). The next phase of SARS-CoV-2 surveillance: real-time molecular epidemiology. *Nature Medicine*, *27*(9), 1518–1524. https://doi.org/10.1038/s41591-021-01472-w

Oyola, S. O., Ariani, C. V., Hamilton, W. L., Kekre, M., Amenga-Etego, L. N., Ghansah, A., Rutledge, G. G., Redmond, S., Manske, M., Jyothi, D., Jacob, C. G., Otto, T. D., Rockett, K., Newbold, C. I., Berriman, M., & Kwiatkowski, D. P. (2016). Whole genome sequencing of Plasmodium falciparum from dried blood spots using selective whole genome amplification. *Malaria Journal*, *15*(1), 597. https://doi.org/10.1186/s12936-016-1641-7.

Pelleau, S., Moss, E. L., Dhingra, S. K., Volney, B., Casteras, J., Gabryszewski, S. J., Volkman, S. K., Wirth, D. F., Legrand, E., Fidock, D. A., Neafsey, D. E., & Musset, L. (2015). Adaptive evolution of malaria parasites in French Guiana: Reversal of chloroquine resistance by acquisition of a mutation in pfcrt. *Proceedings of the National Academy of Sciences*, *112*(37), 11672–11677. https://doi.org/10.1073/pnas.1507142112

R Core Team (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Reeder, J. C., & Marshall, V. M. (1994). A simple method for typing Trypanosoma falciparum merozoite surface antigens 1 and 2 (MSA-1 and MSA-2) using a dimorphic-form specific polymerase chain reaction. *Molecular and Biochemical Parasitology*, *68*(2), 329–332. https://doi.org/10.1016/0166-6851(94)90179-1

Ruybal-Pesántez, S., Sáenz, F. E., Deed, S., Johnson, E. K., Larremore, D. B., Vera-Arias, C. A., Tiedje, K. E., & Day, K. P. (2021). Clinical malaria incidence following an outbreak in Ecuador was predominantly associated with Plasmodium falciparum with recombinant variant antigen gene repertoires. *MedRxiv [preprint]*, 2021.04.12.21255093. https://doi.org/https://doi.org/10.1101/2021.04.12.21255093

Schaffner, S. F., Taylor, A. R., Wong, W., Wirth, D. F., & Neafsey, D. E. (2018). hmmIBD: Software to infer pairwise identity by descent between haploid genotypes. *Malaria Journal*, *17*(1), 196. https://doi.org/10.1186/s12936-018-2349-7

Schmidt, D. A., Campbell, N. R., Govindarajulu, P., Larsen, K. W., & Russello, M. A. (2020). Genotyping-in-Thousands by sequencing (GT-seq) panel development and application to minimally invasive DNA samples to support studies in molecular ecology. *Molecular Ecology Resources*, *20*(1), 114–124. https://doi.org/10.1111/1755-0998.13090

Schwabl, P., Maiguashca Sánchez, J., Costales, J. A., Ocaña-Mayorga, S., Segovia, M., Carrasco, H. J., Hernández, C., Ramírez, J. D., Lewis, M. D., Grijalva, M. J., & Llewellyn, M. S. (2020). Culture-free genome-wide locus sequence typing (GLST) provides new perspectives on Trypanosoma cruzi dispersal and infection complexity. *PLoS Genetics*, *16*(12), e1009170. https://doi.org/10.1371/journal.pgen.1009170

Snounou, G. (2002). Genotyping of Plasmodium spp. Nested PCR. *Methods in Molecular Medicine*, *72*, 103–116. https://doi.org/10.1385/1-59259-271-6:103

Takala-Harrison, S., Jacob, C. G., Arze, C., Cummings, M. P., Silva, J. C., Dondorp, A. M., Fukuda, M. M., Hien, T. T., Mayxay, M., Noedl, H., Nosten, F., Kyaw, M. P., Nhien, N. T. T., Imwong, M., Bethell, D., Se, Y., Lon, C., Tyner, S. D., Saunders, D. L., … Plowe, C. V. (2015). Independent emergence of artemisinin resistance mutations among Plasmodium falciparum in Southeast Asia. *The Journal of Infectious Diseases*, *211*(5), 670–679. https://doi.org/10.1093/infdis/jiu491

Taylor, A. R., Jacob, P. E., Neafsey, D. E., & Buckee, C. O. (2019). Estimating relatedness between malaria parasites. *Genetics*, *212*(4), 1337–1351. https://doi.org/10.1534/genetics.119.302120

Taylor, A. R., Schaffner, S. F., Cerqueira, G. C., Nkhoma, S. C., Anderson, T. J. C., Sriprawat, K., Pyae Phyo, A., Nosten, F., Neafsey, D. E., & Buckee, C. O. (2017). Quantifying connectivity between local *Plasmodium falciparum* malaria parasite populations using identity by descent. *PLoS Genetics*, *13*(10), e1007065. https://doi.org/10.1371/journal.pgen.1007065

Taylor, S. M., Parobek, C. M., DeConti, D. K., Kayentao, K., Coulibaly, S. O., Greenwood, B. M., Tagbor, H., Williams, J., Bojang, K., Njie, F., Desai, M., Kariuki, S., Gutman, J., Mathanga, D. P., Mårtensson, A., Ngasala, B., Conrad, M. D., Rosenthal, P. J., Tshefu, A. K., … Juliano, J. J. (2015). Absence of putative artemisinin resistance mutations among *Plasmodium falciparum* in Sub-Saharan Africa: A molecular epidemiologic study. *The Journal of Infectious Diseases*, *211*(5), 680–688. https://doi.org/10.1093/infdis/jiu467

Tessema, S. K., Hathaway, N. J., Teyssier, N. B., Murphy, M., Chen, A., Aydemir, O., Duarte, E. M., Simone, W., Colborn, J., Saute, F., Crawford, E., Aide, P., Bailey, J. A., & Greenhouse, B. (2020). Sensitive, highly multiplexed sequencing of microhaplotypes from

the *Plasmodium falciparum* heterozygome. *The Journal of Infectious Diseases*, *225*(7), 1227–1237. https://doi.org/10.1093/infdis/jiaa527

Tessema, S., Wesolowski, A., Chen, A., Murphy, M., Wilheim, J., Mupiri, A.-R., Ruktanonchai, N. W., Alegana, V. A., Tatem, A. J., Tambo, M., Didier, B., Cohen, J. M., Bennett, A., Sturrock, H. J., Gosling, R., Hsiang, M. S., Smith, D. L., Mumbengegwi, D. R., Smith, J. L., & Greenhouse, B. (2019). Using parasite genetic and human mobility data to infer local and cross-border malaria connectivity in Southern Africa. *Elife*, *8*, e43510. https://doi.org/10.7554/eLife.43510

Trager, W., & Jensen, J. B. (1976). Human malaria parasites in continuous culture. *Science (New York, N.Y.)*, *193*(4254), 673–675. https://doi.org/10.1126/science.781840

Tran, T. M., Li, S., Doumbo, S., Doumtabe, D., Huang, C.-Y., Dia, S., Bathily, A., Sangala, J., Kone, Y., Traore, A., Niangaly, M., Dara, C., Kayentao, K., Ongoiba, A., Doumbo, O. K., Traore, B., & Crompton, P. D. (2013). An intensive longitudinal cohort study of malian children and adults reveals no evidence of acquired immunity to *Plasmodium falciparum* infection. *Clinical Infectious Diseases: an Official Publication of the Infectious Diseases Society of America*, *57*(1), 40–47. https://doi.org/10.1093/cid/cit174

Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., Banks, E., Garimella, K. V., Altshuler, D., Gabriel, S., & DePristo, M. A. (2013). From FastQ data to high confidence variant calls: The Genome Analysis Toolkit best practices pipeline. *Current Protocols in Bioinformatics*, *43*, 11.10.1-11.10.33. https://doi.org/10.1002/0471250953.bi1110s43

Veiga, M. I., Dhingra, S. K., Henrich, P. P., Straimer, J., Gnädig, N., Uhlemann, A.-C., Martin, R. E., Lehane, A. M., & Fidock, D. A. (2016). Globally prevalent PfMDR1 mutations modulate Plasmodium falciparum susceptibility to artemisinin-based combination therapies. *Nature Communications*, *7*(1), 11553. https://doi.org/10.1038/ncomms11553

WHO (2019). *World Malaria Report*. https://www.who.int/publications-detail/world-malaria-report-2019

Yalcindag, E., Elguero, E., Arnathau, C., Durand, P., Akiana, J., Anderson, T. J., Aubouy, A., Balloux, F., Besnard, P., Bogreau, H., Carnevale, P., D'Alessandro, U., Fontenille, D., Gamboa, D., Jombart, T., Le Mire, J., Leroy, E., Maestre, A., Mayxay, M., … Prugnolle, F. (2012). Multiple independent introductions of *Plasmodium falciparum* in South America. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(2), 511–516. https://doi.org/10.1073/pnas.1119058109

Zhu, S. J., Hendry, J. A., Almagro-Garcia, J., Pearson, R. D., Amato, R., Miles, A., Weiss, D. J., Lucas, T. C., Nguyen, M., Gething, P. W., Kwiatkowski, D., & McVean, G. & for the Pf3k Project. (2019). The origins and relatedness structure of mixed infections vary with local prevalence of *P. falciparum* malaria. *eLife*, *8*, e40845. https://doi.org/10.7554/eLife.40845

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: LaVerriere, E., Schwabl, P., Carrasquilla, M., Taylor, A. R., Johnson, Z. M., Shieh, M., Panchal, R., Straub, T. J., Kuzma, R., Watson, S., Buckee, C. O., Andrade, C. M., Portugal, S., Crompton, P. D., Traore, B., Rayner, J. C., Corredor, V., James, K., Cox, H., … Neafsey, D. E. (2022). Design and implementation of multiplexed amplicon sequencing panels to serve genomic epidemiology of infectious disease: A malaria case study. *Molecular Ecology Resources*, *00*, 1–19. https://doi.org/10.1111/1755-0998.13622