Sensitive and modular amplicon sequencing of *Plasmodium falciparum* diversity and resistance for research and public health

Authors

Andrés Aranda-Díaz[1,2,§], Eric Neubauer Vickers[1*], Kathryn Murie[1*], Brian Palmer[1*], Nicholas Hathaway[1], Inna Gerlovina[1], Simone Boene[3], Manuel Garcia-Ulloa[4], Pau Cisteró[4], Thomas Katairo[5], Francis Ddumba Semakuba[5], Bienvenu Nsengimaana[5], Hazel Gwarinda[6], Carla García-Fernández[4], William Louie[1], Endashaw Esayas[7], Clemente Da Silva[3], Debayan Datta[4], Shahiid Kiyaga[5,8], Innocent Wiringilimaana[5], Sindew Mekasha Fekele[9,10], Adam Bennett[11], Jennifer L. Smith[12,13], Endalamaw Gadisa[7], Jonathan B. Parr[14], Melissa Conrad[15], Jaishree Raman[6,16,17], Stephen Tukwasibwe[5], Isaac Ssewanyana[5], Eduard Rovira-Vallbona[3], Cristina M. Tato[2], Jessica Briggs[1], Alfredo Mayor[3,4,18,19], Bryan Greenhouse[1]

Affiliations

[1]: EPPIcenter Research Program, Division of HIV, Infectious Diseases, and Global Medicine, Department of Medicine, University of California, San Francisco, California, USA

[2]: Chan Zuckerberg Biohub, San Francisco, California, USA

[3]: Centro de Investigação em Saúde de Manhiça, Maputo, Mozambique

[4]: ISGlobal, Barcelona, Spain

[5]: Infectious Diseases Research Collaboration, Kampala, Uganda

[6]: Laboratory for Antimalarial Resistance Monitoring and Malaria Operational Research (ARMMOR), Centre of Emerging Zoonotic and Parasitic Diseases, National Institute for Communicable Diseases, Johannesburg, South Africa

[7]: Armauer Hansen Research Institute, Addis Ababa, Ethiopia

[8]: Department of Immunology and Molecular Biology, College of Health Sciences, Makerere University, Kampala, Uganda

[9]: Ethiopian Public Health Institute, Addis Ababa, Ethiopia

[10]: Department of Environment and Genetics, La Trobe University, Melbourne, Australia

[11]: PATH, Seattle, Washington, USA

[12]: Malaria Elimination Initiative, Global Health Group, University of California, San Francisco, United States of America

[13]: Department of Epidemiology & Biostatistics, University of California, San Francisco, United States of America

[14]: Division of Infectious Diseases, University of North Carolina at Chapel Hill, North Carolina, USA[15]: Division of HIV, Infectious Diseases, and Global Medicine, Department of Medicine, University of California, San Francisco, California, USA

[16]: Wits Research Institute for Malaria, University of Witwatersrand, Johannesburg, South Africa

[17]: University of Pretoria Institute for Sustainable Malaria Control (UPISMC), University of Pretoria, Pretoria, South Africa.

[18]: Facultat de Medicina i Ciències de la Salut, Universitat de Barcelona (UB), Barcelona, Spain

[19]: Department of Physiologic Sciences, Faculty of Medicine, Universidade Eduardo Mondlane, Maputo, Mozambique

* These authors contributed equally

[§] Corresponding author: andres.arandadiaz@ucsf.edu

Abstract

Background

Targeted amplicon sequencing is a powerful and efficient tool for interrogating the *Plasmodium falciparum* genome, generating actionable data from infections to complement traditional malaria epidemiology. For maximum impact, genomic tools should be multi-purpose, robust, sensitive, and reproducible.

Methods

We developed, characterized, and implemented MAD^{4}HatTeR, an amplicon sequencing panel based on Multiplex Amplicons for Drug, Diagnostic, Diversity, and Differentiation Haplotypes using Targeted Resequencing, along with a bioinformatic pipeline for data analysis. Additionally, we introduce an analytical approach to detect gene duplications and deletions from amplicon sequencing data. Laboratory control and field samples were used to demonstrate the panel's high sensitivity and robustness.

Results

MAD^{4}HatTeR targets 165 highly diverse loci, focusing on multiallelic microhaplotypes, key markers for drug and diagnostic resistance (including duplications and deletions), and *csp* and potential vaccine targets. The panel can also detect non-*falciparum Plasmodium* species. MAD^{4}HatTeR successfully generated data from low-parasite-density dried blood spot and mosquito midgut samples, and detected minor alleles at within-sample allele frequencies as low as 1% with high specificity in high-parasite-density dried blood spot samples. Gene deletions and duplications were reliably detected in mono- and polyclonal controls. Data generated by MAD^{4}HatTeR were highly reproducible across multiple laboratories.

Conclusions

The successful implementation of MAD^{4}HatTeR in five laboratories, including three in malaria-endemic African countries, showcases its feasibility and reproducibility in diverse settings. MAD^{4}HatTeR is thus a powerful tool for research and a robust resource for malaria public health surveillance and control.

Keywords

Malaria; *Plasmodium falciparum*; Targeted amplicon sequencing; malaria; microhaplotype; antimalarial resistance; *hrp2* deletion; *csp*; identity-by-descent; complexity of infection

Background

Effective control and eventual elimination of *Plasmodium falciparum* malaria hinge on the availability and integration of data to inform research and public health strategies. Genomics can augment traditional epidemiological surveillance by providing detailed genetic information about infections[1]. Molecular markers of drug and diagnostic resistance can guide the selection of antimalarials and diagnostics, respectively[2–5]. Vaccine target sequences may shed light on vaccine efficacy and identify evidence of selective pressure[6]. Measures of genetic variation can provide insights into transmission intensity, rate and origin(s) of importation, and granular details of local transmission[7–14]. Differentiation of infections as either recrudescent or reinfections is critical for measuring outcomes of therapeutic efficacy studies that are used to guide antimalarial use worldwide[15–18]. Furthermore, the contribution of non-*falciparum* species to malaria burden is poorly characterized, and could complicate control and elimination efforts[19].

To maximize public health and research utility, genomic methods should be robust and provide rich information from field samples, which may be low-density and are often polyclonal in malaria-endemic areas of sub-Saharan Africa[13,20–22]. While traditional genotyping methods of length polymorphisms and microsatellites can characterize malarial infections, they suffer from low sensitivity and specificity, and difficulties in protocol standardization[23–25]. Single nucleotide polymorphism (SNP) barcoding approaches have improved throughput, sensitivity and standardization[26,27]. However, the biallelic nature of most targeted SNPs limits their discriminatory power to compare polyclonal infections. Sequencing of short, highly variable regions within the genome containing multiple SNPs (microhaplotypes) provides multiallelic information that overcomes many of those limitations[28]. Microhaplotypes can be reconstructed from whole-genome sequencing (WGS) data or amplified by PCR and sequenced. Low abundance variants, especially in low-density samples, may be missed by WGS due to low

130 depth of coverage. Amplicon sequencing offers much higher sensitivity and can target the most
131 informative regions of the genome, increasing throughput and decreasing cost. Several Illumina-
132 based multiplexed amplicon sequencing panels have been developed to genotype *P. falciparum*
133 infections. SpotMalaria is a panel that genotypes 100 SNPs, most of which are biallelic, for drug
134 resistance and diversity[26]. Pf AmpliSeq genotypes SNPs, currently focused on Peruvian genetic
135 diversity, and also targets drug and diagnostic resistance markers[27]. Panels that target
136 multiallelic microhaplotypes, including AMPLseq, provide greater resolution for evaluating
137 polyclonal infections and also include drug resistance markers[29,30]. Nanopore-based amplicon
138 panels enable the utilization of mobile sequencing platforms[31–33]. Thus, targeted amplicon
139 sequencing is a flexible approach that has the potential to address multiple use cases. To fully
140 realize this potential, a panel for research and public health would ideally include all necessary
141 targets to answer a wide range of questions, while remaining modular to allow flexible allocation
142 of sequencing resources.
143
144 Here, we developed MAD^{4}HatTeR, an Illumina-compatible, multipurpose, modular tool based on
145 Multiplex Amplicons for Drug, Diagnostic, Diversity, and Differentiation Haplotypes using
146 Targeted Resequencing. MAD^{4}HatTeR has 276 targets divided into two modules: A diversity
147 module with 165 targets to assess genetic diversity and relatedness; and a resistance module
148 consisting of 118 targets that cover 15 drug resistance-associated genes and assesses *hrp2*/*3*
149 deletions, along with current and potential vaccine targets. The modules also include targets for
150 non-*falciparum Plasmodium* species identification. We developed a bioinformatic pipeline to
151 report allelic data, and implemented laboratory and bioinformatic methods in several sites,
152 including countries in malaria-endemic sub-Saharan Africa. We then evaluated the panel's
153 performance on various sample types, including mosquito midguts, and showed that high quality
154 data can be consistently reproduced across laboratories, including from polyclonal samples with
155 low parasite density.

156 **Methods**

157

158 *Participating laboratories*

159

160 We generated data in five sites: the EPPIcenter at the University of California San Francisco
161 (UCSF), in collaboration with the Chan Zuckerberg Biohub San Francisco, California; Infectious
162 Diseases Research Collaboration (IDRC) at Central Public Health Laboratories (CPHL),
163 Kampala, Uganda; Centro de Investigação em Saúde de Manhiça (CISM), Manhiça,
164 Mozambique; National Institutes for Communicable Diseases (NICD), Johannesburg, South
165 Africa; and Barcelona Institute for Global Health (ISGlobal), Barcelona, Spain. The procedures
166 are described according to the workflows in San Francisco. Minor variations, depending on
167 equipment availability, were implemented at other institutions.

168

169 *Amplicon panel design*

170

171 We used available WGS data as of June 2021[3,30,34–42] to identify regions with multiple SNPs
172 within windows of 150-300 bp that lay between tandem repeats, using a local haplotype
173 reconstruction tool (Pathweaver[43]). We compiled a list of drug resistance-associated and
174 immunity-related SNPs (Tables 1 and 2) and identified regions of 150-300 bp between tandem
175 repeats in and around *hrp2* and *hrp3* to assess diagnostic resistance-related deletions, as well
176 as a region in chromosome 11 that is often duplicated in *hrp3*-deleted samples[44]. Paragon
177 Genomics, Inc. designed amplification primers in multiplexed PCR using the Pf3D7 genome as
178 a reference and related *Plasmodium* species and human genomes to design primers specific for
179 *P. falciparum*. Genome verions and their GeneBank or RefSeq accessions for each species are:
180 *P. falciparum* Pf3D7 (version=2020-09-01, GCA_000002765.3), *P. vivax* PvP01 (version=2018-
181 02-28, GCA_900093555.2), *P. malariae* PmUG01 (version=2016-09-19, GCA_900090045.1), *P.*

182 *ovale* PocGH01 (version=2017-03-06, GCA_900090035.2), *P. knowlesi* PKNH (version=2015-
183 06-18, GCA_000006355.2) and *Homo sapiens* GRCh37 (GCA_000001405.14). In addition to
184 the *P. falciparum* targets, we selected a target in the *ldh* gene (PF3D7_1325200) and its
185 homologs in the other 4 *Plasmodium* species listed above for identification of concurrent
186 infections with these species. To minimize PCR bias against longer amplicons, we restricted the
187 design to amplicons of 225-275 bp, which can be covered with a significant overlap in paired-
188 end sequencing in Illumina platforms with 300-cycle kits, except for targets around *hrp3* that
189 needed to be 295-300 bp long to design primers successfully. We excluded or redesigned
190 primers that contained more than 1 SNP (including non-biallelic SNPs) or indels in available
191 WGS data or aligned to tandem repeats. To increase coverage of SNPs close to each other, we
192 allowed for overlap in amplicons that targeted drug resistance and immunity-related markers.
193 Primers were grouped in modules, as outlined in the results section (Figure 1 and
194 Supplementary Table 1).
195
196 In silico *panel performance calculations*
197
198 Alleles were extracted from available WGS data as of July 2024[3,30,34–41,45]. SNPs, and
199 microhaplotypes were reconstructed using Pathweaver[43] for targets in MAD^{4}HatTeR,
200 SpotMalaria[26], AMPLseq[30], and AmpliSeq[27]. *In silico* heterozygosity was calculated using all
201 allele calls in available WGS data. Principal coordinate analysis was performed on the binary
202 distance matrix from presence/absence of alleles using alleles within loci present in both
203 samples for each pair.
204
205 To assess statistical power of testing if two (potentially polyclonal) infections are related, we
206 obtained within sample allele frequencies (WSAF) for the most variable SNP in each diversity
207 target (165, 111 and 100 total SNPs for MAD^{4}HatTeR, AMPLseq and SpotMalaria, respectively)

208 or microhaplotypes (161, 128 and 135, respectively) from WGS data for each of the three
209 panels, and simulated genotypes for mono- and polyclonal samples. In the simulations,
210 complexity of infection (COI) were fixed and ranged from 1 to 5, and we included genotyping
211 errors with a miss-and-split model[46]; missing and splitting parameters were 0.05 and 0.01,
212 respectively. Between two samples, only a single pair of parasite strains was related with
213 expected identity-by-descent (IBD) proportion varying from 1/16 to 1/2 (sibling level) to 1
214 (clones). We then analyzed these simulated datasets to obtain performance measures for
215 combinations of a panel, COI, and a relatedness level: first, we estimated COI and allele
216 frequencies using MOIRE[47]; we then used these to estimate pairwise interhost relatedness and
217 test the hypothesis that two infections are unrelated at significance level of 0.05 with Dcifer[46]
218 and calculated power as the proportion of 1000 simulated pairs where the null hypothesis was
219 correctly rejected.
220
221 *Samples*
222
223 We prepared control dried blood spots (DBS) using *P. falciparum* laboratory strains. We
224 synchronized monocultures in the ring stage. We made polyclonal controls by mixing cultured
225 strains (3D7, Dd2 MRA-156 and MRA-1255, D6, W2, D10, U659, FCR3, V1/S, and HB3), all
226 synchronized and ring-staged at various proportions. We mixed all monocultures and mixtures
227 with uninfected human blood and serially diluted them in blood to obtain a range of parasite
228 densities (0.1-100,000 parasites/μL). We spotted 20 μL of the mixture on filter papers and
229 stored them at -20 ºC until processing.
230
231 Finger-prick DBS samples were collected in Northern Ethiopia between 2022 and 2023 as part
232 of a mixed-methods study, which included a case-control study in two highland districts and
233 cross-sectional surveys in one lowland district[48]. In the case-control study, samples were

234 obtained from symptomatic patients presenting at health facilities. These included malaria cases
235 - individuals who tested positive for *P. falciparum* and/or *P. vivax* using a rapid diagnostic test
236 (Bioline Malaria Ag P.f/Pan by Abott, STANDARD Q Malaria P.f/P.v Ag by SD Biosensor, or
237 First Response Malaria Ag. P.f./P.v. Card Test) – as well as test-negative controls who were
238 later confirmed positive for malaria via qPCR. Cross-sectional surveys in lowland areas were
239 conducted at agricultural worksites and in households in nearby villages. The DBS were air-
240 dried, stored individually with desiccant, and kept at -20°C until laboratory processing. No
241 patient data collected during sampling was used in this analysis. We analyzed DNA extracts
242 from samples from previous studies, including 26 field samples from Ethiopia known to carry
243 deletions in the *hrp2* and/or *hrp3* genes[3], as well as 11 *P. falciparum* co-infections from Uganda
244 containing *P. malariae* and *P. ovale*[49]. Finally, we analyzed publicly available data from 436 field
245 samples from Mozambique[22]. The original works detail the sampling schemes and additional
246 sample processing procedures. We used genomic DNA from *P. knowlesi* Strain H, obtained
247 through BEI Resources, NIAID, NIH, contributed by Alan W. Thomas.
248
249 To assess performance of the assay for oocysts, we infected 9 *Anopheles gambiae* s.s.
250 mosquitos via direct membrane feeding with blood taken from participants who were diagnosed
251 with symptomatic malaria in transmission studies in Uganda. Briefly, mosquitos were fed via
252 direct membrane feeding with blood taken from participants in a cohort study based in
253 Nanongera and Busia districts (3 patients, 5 infected midguts) and from patients diagnosed with
254 P. falciparum malaria at Masafu General Hospital in Busia district (4 patients, 4 infected
255 midguts). [50,51] *P. falciparum* presence in blood samples was confirmed by varATS qPCR.
256 Oocysts were detected and quantified using mercurochrome staining and microscopy as
257 previously described[51].
258
259

260 *Library preparation*

261

262 We extracted DNA from control DBS and *P. falciparum*/*P. vivax* co-infections using the Chelex-
263 Tween 20 method[52]. Mosquito midgut DNA was extracted from dissected midguts using the
264 QIAGEN DNeasy blood and tissue DNA extraction kit as previously described[53]. *P. falciparum*
265 parasite density was quantified in all samples, including midgut extracts, by varATS[54] or 18S[55]
266 qPCR using standards made from DBS spotted with serial dilutions of cultured *P. falciparum* in
267 uninfected blood (Supplementary Text). *P. vivax* was quantified by 18S qPCR as previously
268 described[56].
269 Libraries were made with a minor adaptation of Paragon Genomics' CleanPlex Custom NGS
270 Panel Protocol[57] (Supplementary Text). A version of the protocol containing any updates can be
271 found at https://eppicenter.ucsf.edu/resources. Library pools were sequenced in Illumina MiSeq,
272 MiniSeq, NextSeq 550, or NextSeq 2000 instruments with 150 paired-end reads. We tested
273 different amplification cycles and primer pool configurations. Based on sensitivity and
274 reproducibility, the following are the experimental conditions we use as a default: primer pools
275 D1+R1.2+R2; 15 multiplexed PCR cycles for moderate to high parasite density samples
276 (equivalent to ≥ 100 parasites/μL in DBS) and 20 cycles for samples with lower parasite density;
277 0.25X and 0.12X primer pool concentration, respectively.

278

279 *Bioinformatic pipeline development and benchmarking*

280

281 We developed a Nextflow-based[58] bioinformatic pipeline to filter, demultiplex, and infer alleles
282 from fastq files (Supplementary Text). Briefly, the pipeline uses cutadapt[59] and DADA2[60] to
283 demultiplex reads on a per-amplicon basis and infer alleles, respectively. The pipeline further
284 processes DADA2 outputs to mask low-complexity regions, generate allele read count tables,
285 and extract alleles in SNPs of interest. We developed custom code in Python and R to filter out

286 low-abundance alleles and calculate summary statistics from the data. The current pipeline
287 version, with more information on implementation and usage, can be found at
288 www.github.com/EPPIcenter/mad4hatter.
289
290 We processed the data presented in this paper with release 0.1.8 of the pipeline.
291
292 We evaluated pipeline performance by estimating sensitivity (ability to identify expected alleles)
293 and precision (ability to identify only expected alleles) from monoclonal and mixed laboratory
294 controls with different proportions of strains (Supplementary Text). We tested the impact of
295 multiple parameters and features on allele calling accuracy, including DADA2's stringency
296 threshold OMEGA_A and sample pooling treatment for allele recovery, masking homopolymers
297 and tandem repeats, and post-processing filtering of low abundance alleles. Masking removed
298 false positives with the trade-off of masking real biological variation. We obtained the highest
299 precision and sensitivity using sample pseudo-pooling, highly stringent OMEGA_A (10^{-120}), and
300 a moderate postprocessing filtering threshold (minor alleles of > 0.75%). These results indicate
301 that bioinformatic processing of MAD^{4}HatTeR data can be optimized to retrieve accurate sample
302 composition with a detection limit of approximately 0.75% WSAF.
303
304 For analyses of allelic data from mixed controls, only samples with ≥ 90% of targets with > 50
305 reads (183 for diversity, and 165 for drug resistance markers) were included in the analysis. For
306 drug resistance markers, only SNPs with variation between controls were included (20/91
307 codons from 12/22 targets). Within a sample, targets with less than 100 reads were excluded as
308 alleles with a minor WSAF of 1% are very likely to be missed. The large majority of controls
309 (122/183 and 162/165 for diversity and drug resistance markers, respectively) had very good
310 coverage (at most 2 missing loci).
311

312 Species-specific *ldh* targets in the panel were used to identify non-*falciparum* species. Targets
313 with less than 5 reads were filtered out. *P. ovale ldh* target sequences were extracted from the
314 *P. ovale curtisi* (PocGH01, GCA_900090035.2) or *P. ovale wallikeri* (PowCR01,
315 GCA_900090025.2) genomes using target primer sequences. Observed sequences were then
316 aligned to these reference sequences using BLAST. Heterozygosity was estimated using
317 MOIRE[47] version 3.2.0.
318
319 *Deletions and duplications*
320
321 We used the following laboratory strains to benchmark deletion and duplication detection using
322 MAD[4]HatTeR data: *pfhrp2* deletions in Dd2 and D10, *pfmdr1* duplications in Dd2 and FCR3,
323 *pfhrp3* deletion in HB3, and *pfhrp3* duplication in FCR3[44]. We also used a set of field samples
324 from Ethiopia previously shown to have deletions in and around *pfhrp2* and *pfhrp3* at multiple
325 genomic breakpoints[3]. For sensitivity analysis using field samples, we estimated COI using
326 MOIRE[47] and excluded polyclonal samples due to the uncertainty in their true genotypes. Two
327 field samples were excluded from the analysis due to discordance in breakpoint classification,
328 possibly due to sample mislabeling and sequencing depth, respectively.
329 We applied a generalized additive model (Supplementary Text) to account for target length
330 amplification bias and differences in coverage across primer pools, likely due to pipetting error.
331 We fit the model on controls known not to have deletions or duplications to obtain correction
332 factors for targets of interest within sample batches. We then estimated read depth fold changes
333 from data for each gene of interest (*pfhrp2*, *pfhrp3* and *pfmdr1*). We did not have sufficient data
334 to validate duplications in plasmepsin 2 and 3.
335 For a subset of laboratory controls copy numbers were determined by qPCR using previously
336 described methods for *pfmdr1*[61], *pfhrp2*, and *pfhrp3*[62].
337

338 **Results**

339

340 *MAD^{4}HatTeR is a multi-purpose tool that exploits* P. falciparum *genetic diversity*

341

342 We designed primers to amplify 276 targets (Figure 1, Supplementary Tables 1-4) and
343 separated them into two modules: (1) Diversity module, a primer pool (D1) targeting 165 high
344 diversity targets and the *ldh* gene in *P. falciparum* and in 4 non-*falciparum Plasmodium* species
345 (*P. vivax*, *P. malariae*, *P. ovale*, and *P. knowlesi*); and (2) Resistance module, comprised of two
346 complementary and incompatible primer pools (R1 and R2) targeting 118 loci that genotype 15
347 drug resistance-associated genes (Table 1) along with *csp* and potential vaccine targets (Table
348 2), assess for *hrp2*/*3* deletion, and identify non-*falciparum* species. The protocol involves two
349 initial multiplex PCR reactions, one with D1 and R1 primers, and another with R2 primers
350 (Figure 1C, Supplementary Figure 1). After multiplexed PCR, subsequent reactions continue in
351 a single tube.

352

353 Based on publicly available WGS data, *P. falciparum* targets in the diversity module, excluding
354 *ldh*, had a median of 3 SNPs or indels (interquartile range [IQR] 2-5, N=165, Supplementary
355 Table 5). Most (140/165) targets were microhaplotypes (containing > 1 SNP or indel). Global
356 heterozygosity was high, with 35 targets with heterozygosity > 0.75 and 135 with heterozygosity
357 > 0.5. Within African samples, heterozygosity was > 0.75 in 40 targets, > 0.5 in 132 targets, and
358 we observed 2 to 20 unique alleles (median of 5, across a minimum of 3617 samples) in each
359 target. MAD^{4}HatTeR included more high-heterozygosity targets than other published panels
360 (Figure 2A, Supplementary Figures 2 and 3). Additionally, MAD^{4}HatTeR targets better resolved
361 geographical structure globally, within Africa, and even within a country[63] (Figure 2B).

362

363 We next evaluated the power of the diversity module to detect interhost relatedness between

364 parasites in pairs of simulated infections with COI ranging from 1 to 5. We selected one country
365 from each of three continents with the most publicly available WGS data and used
366 reconstructed genotypes for the analysis (Figure 3). MAD^{4}HatTeR identified partially related
367 parasites between polyclonal infections across a range of COI and geographic regions, and
368 generally performed as well or better than the other panels evaluated. For example, in simulated
369 Ghanaian infections sibling parasites (IBD proportion, r=½) were reliably detected with COI of 5
370 (82% power), half siblings (r=¼) in infections with COI of 3 (73% power), and less related
371 parasites (r=⅛) were still identifiable with COI of 2 (53% power). When using independent SNPs
372 instead of microhaplotypes, the power to identify related parasites between infections was much
373 lower, irrespective of the panel. Constraining the panel to the 50 targets with the highest
374 heterozygosity (mean heterozygosity of 0.8 ± 0.05) reduced the power to infer relatedness by as
375 much as 50%, highlighting the value of highly multiplexed microhaplotype panels for statistical
376 power.
377
378 *MAD^{4}HatTeR allows for genotyping of a variety of sample types and parasite densities*
379
380 We evaluated MAD^{4}HatTeR's performance using dried blood spots (DBS) containing up to 7
381 different cultured laboratory strains each. Sequencing depth was lower for samples amplified
382 with the original resistance R1 primer pool R1.1 than D1 (Supplementary Figure 4A), and primer
383 dimers comprised 58-98% of the reads for R1.1 compared to only 0.1-4% for D1. We thus
384 designed pool R1.2, a subset of targets from R1.1, by selecting the targets with priority public
385 health applications and discarding the primers that accounted for a significant portion of primer
386 dimers in generated data (Figure 1, Supplementary Table 2). Libraries prepared with pools
387 containing R1.2 instead of R1.1 showed higher depth across the range of parasitemia evaluated
388 (Supplementary Figure 4B). With the recommended set of primer pools (D1, R1.2, and R2),
389 sequencing provided > 100 reads for most amplicons from DBS with > 10 parasites/μL, with

390 depth of coverage increasing with higher parasite densities (Figure 4A). Samples with < 10
391 parasites/μL still yielded data albeit less reliably. Approximately 100,000 total unfiltered reads
392 (the output of sample demultiplexing from a sequencing run) were sufficient to get good
393 coverage across targets; on average, 95% of targets had at least 100 reads, and 98% had at
394 least 10 reads (Supplementary Figure 4 C,D). While results indicate that the protocol provides
395 consistently robust results, different experimental parameters may be optimal for different
396 combinations of primer pools and sample concentration.
397
398 Depth of coverage per amplicon was highly correlated within technical replicates
399 (Supplementary Figure 5A) with most deviations observed between primer pools. Importantly,
400 coverage was also reproducible when the same samples were tested across five laboratories on
401 3 continents, with minor quantitative but negligent qualitative differences in coverage
402 (Supplementary Figure 5B). Amplicon coverage was well balanced within a given sample, with
403 differences in depth negatively associated with amplicon length (Supplementary Figure 6). Nine
404 of the 15 worst-performing amplicons were particularly long (>297 bp, Supplementary Table 6).
405 The other worst-performing amplicons covered drug resistance markers in *mdr1* and *crt* (neither
406 covering *mdr1* N86Y or *crt* K76T), 2 high heterozygosity targets, and a target within *hrp2*. These
407 results indicate that robust coverage of the vast majority of targets can be consistently obtained
408 from different laboratories.
409
410 Given the high sensitivity of the method, we evaluated the ability of MAD^{4}HatTeR to generate
411 data from sample types where it is traditionally challenging to obtain high quality parasite
412 sequence data. We amplified DNA extracted from nine infected mosquito midguts with a median
413 *P. falciparum* DNA concentration equivalent to 0.9 parasites/μL from a DBS. On average, 58%
414 of amplicons had ≥100 reads, 84% had ≥10 reads, and only one sample did not amplify (Figure
415 4B). These results are comparable to libraries from DBS controls with 1-10 parasites/μL from

416 the same sequencing run, where 45-77% of amplicons with ≥100 reads. WSAF indicated that
417 some of the mosquito midguts contained several genetically distinct *P. falciparum* clones. These
418 data show the potential for applying MAD^{4}HatTeR to study a variety of sample types containing
419 *P. falciparum*.
420
421 *MAD^{4}HatTeR reproducibly detects genetic diversity, including for minority alleles in low density,*
422 *polyclonal samples*
423
424 We used DBS controls containing 2 to 7 laboratory *P. falciparum* strains with minor WSAF
425 ranging from 1 to 50% to evaluate sensitivity of detection and accuracy of WSAF estimation in
426 the diversity pool D1. We optimized and benchmarked the bioinformatic pipeline to maximize
427 sensitivity and precision, which included masking regions of low complexity (tandem repeats
428 and homopolymers) to avoid capturing PCR and sequencing errors in allele calls. Sensitivity to
429 detect minority alleles given that the locus amplified was very high, with alleles present at ≥ 2%
430 reliably detected in samples with > 1,000 parasites/μL and at ≥ 5% in samples with > 10
431 parasites/μL (Figure 4C). For very low parasitemia samples (< 10 parasites/μL), sensitivity was
432 still 82% for alleles expected at 10% or higher. Similar results were obtained for drug resistance
433 markers targeted by pools R1.2 and R2 (Figure 4D). Overall precision (reflecting the absence of
434 spurious alleles) was also high and could be increased by using a filtering threshold for
435 minimum WSAF. Each sample had a median of 3 false positive alleles (mean = 4.4, N = 161
436 targets) above 0.75% WSAF, a median of 1 (mean = 2.5) false positives over 2%, and a median
437 of 0 (mean = 0.7) over 5% (Supplementary Figure 7). A strong correlation between expected
438 and observed WSAF was observed in the diversity module targets at all parasite densities and
439 was stronger at higher parasite densities (R^2=0.99 for > 1,000 parasites/μL Figure 4E).
440

441 Reproducibility is an important feature in generating useful data, particularly given differences in
442 equipment and technique that often exists between laboratories. To evaluate this potential
443 source of variation, we generated data for the same mixed-strain controls in five different
444 laboratories on three continents. Reassuringly, the alleles obtained, along with their WSAF,
445 were highly correlated (Figure 4F). Missed alleles in one or more laboratories were mostly
446 present at < 2% within a sample. Finally, we tested MAD^{4}HatTeR's ability to recover expected
447 diversity in field samples. Observed genetic heterozygosity in samples from Mozambique[22] was
448 correlated with expected heterozygosity based on available WGS data (Figure 4G,
449 Supplementary Figure 8). These results highlight the reliability of MAD^{4}HatTeR as a method to
450 generate high quality genetic diversity data across laboratories.
451
452 *MAD^{4}HatTeR provides data on copy number variations and detection of non-*P. falciparum
453 *species*
454
455 In addition to detecting sequence variation in *P. falciparum*, amplicon sequencing data can be
456 used to detect gene deletions and duplications, as well as the presence of other *Plasmodium*
457 species. We tested the ability of MAD^{4}HatTeR to detect *hrp2* and *hrp3* deletions, and *mdr1* and
458 *hrp3* duplications (laboratory strain FCR3 has a duplication in *hrp3*[44]) in DBS controls consisting
459 of one or two laboratory strains, and field samples with previously known genotypes. We applied
460 a generalized additive model to normalize read depth and estimate fold change across several
461 targets per gene, accounting for amplicon length bias and pool imbalances, after using
462 laboratory controls to account for batch effects, e.g. running the assay in different laboratories
463 (Figure 5A, Supplementary Figure 9). The resulting depth fold changes for all loci assayed
464 correlated with the expected sample composition (Figure 5B). At 95% specificity, sensitivity was
465 100% for all controls composed of > 95% strains with duplications or deletions (Figure 5C).
466 Sensitivity was lower for samples with lower relative abundance of strains carrying duplications

467 or deletions, although this could be increased with a tradeoff in specificity (e.g. if used as a
468 screening test). Fold change data correlated well with quantification by qPCR, indicating that the
469 data obtained from MAD^{4}HatTeR are at a minimum semi-quantitative (Figure 5D). We could
470 also correctly detect deletions in field samples from Ethiopia previously shown to be *hrp2*- or
471 *hrp3*-deleted[3], and correctly classify the genomic breakpoint profiles within the resolution offered
472 by the targets included (Supplementary Figure 10). Finally, we detected *P. malariae* and *P.*
473 *ovale* in 11 samples from Uganda known to contain the corresponding species, as previously
474 determined by microscopy or nested PCR. We could distinguish *P. ovale wallikeri* from *P. ovale*
475 *curtisi* based on the alleles in the target sequence. The assay's sensitivity for detecting non-
476 *falciparum* species was evaluated using a set of field samples from Ethiopia containing *P.*
477 *falciparum* and *P. vivax*, with known parasite density for both species. Sensitivity depended on
478 the *P. falciparum* to *P. vivax* ratio within the sample and was estimated at 96% for samples with
479 more than 100 *P. vivax* 18S copies/µL (N = 148) and 90% for those with more than 10 18S
480 copies/µL (N = 170) for samples with a *P. falciparum* to *P. vivax* ratio below 100 (Supplementary
481 Figure 11). Furthermore, the *P. falciparum* to *P. vivax* ratio estimates obtained by qPCR and
482 MAD^{4}HatTeR were highly correlated. Specificity was 100% for all non-*falciparum* species,
483 based on *P. falciparum* controls (N = 368). These data highlight the potential of MAD^{4}HatTeR to
484 capture non-SNP genetic variation and to characterize mixed species infections.
485
486
487
488
489
490
491
492

493 **Discussion**

494

495 In this study, we developed, characterized and deployed a robust and versatile method to
496 generate sequence data for *P. falciparum* malaria genomic epidemiology, prioritizing information
497 for public health decision-making. The modular MAD^{4}HatTeR amplicon sequencing panel
498 produces high-resolution data on genetic diversity, key markers for drug and diagnostic
499 resistance, the C-terminal domain of the *csp* vaccine target, and presence of other *Plasmodium*
500 species. MAD^{4}HatTeR is highly sensitive, providing data for low parasite density DBS samples
501 and detecting minor alleles at WSAF as low as 1% with good specificity in high parasite density
502 samples; challenging sample types such as infected mosquitos were also successfully
503 amplified. MAD^{4}HatTeR has successfully generated data from field samples from Mozambique
504 and Ethiopia, with particularly good recovery rates for samples with > 10 parasites/µL
505 (~90%)[22,64]. Deletions and duplications were reliably detected in mono- and polyclonal controls.
506 The data generated by MAD^{4}HatTeR are highly reproducible and have been reliably produced
507 in multiple laboratories, including several in malaria-endemic countries. Thus, MAD^{4}HatTeR is a
508 valuable tool for malaria surveillance and research, offering policymakers and researchers an
509 efficient means of generating useful data.

510

511 The 165 diversity and differentiation targets in MAD^{4}HatTeR, of which the majority are
512 microhaplotypes, can be used to accurately estimate within-host and population genetic
513 diversity, and relatedness between infections. These data have promising applications:
514 evaluating transmission patterns, e.g. to investigate outbreaks[3]; characterizing transmission
515 intensity, e.g. to evaluate interventions[10,13,65] or surveillance strategies[22]; classifying infections in
516 low transmission areas as imported or local[11,66]; or classifying recurrent infections in antimalarial
517 therapeutic efficacy studies as recrudescence or reinfections[18]. The high diversity captured by
518 the current microhaplotypes could be further improved with updated WGS data to replace

519 targets with relatively low diversity and amplification efficiency. Fully leveraging the information
520 content of these diverse loci, which are particularly useful for evaluating polyclonal infections,
521 requires bioinformatic pipelines able to accurately call microhaplotype alleles and downstream
522 analysis methods able to incorporate these multi-allelic data. While some targeted sequencing
523 methods and pipelines similarly produce microhaplotype data[30,32,67–69], others only report
524 individual SNPs, resulting in the loss of potentially informative data[26,27] encoded in phased
525 amplicon sequences. Many downstream analysis tools are similarly limited to evaluating data
526 from binary SNPs[70–72]. Fortunately, methods to utilize these data are beginning to be developed,
527 providing statistically grounded estimates of fundamental quantities such as population allele
528 frequencies, COI[47], and IBD[46], and highlighting gains in accuracy and power provided by
529 analysis of numerous highly diverse loci.
530
531 Multiple targeted sequencing tools designed with different use cases and geographies in mind
532 are being used, raising questions about data compatibility. Comparing diversity metrics from
533 data generated using different target sets is feasible, provided that the panels have equivalent
534 performance characteristics and that the analysis methods appropriately account for differences
535 such as allelic diversity[47]. Comparing genetic relatedness between infections evaluated with
536 different panels, however, is limited to common loci. Over 25% of SNPs targeted by AMPLseq
537 or SpotMalaria diversity targets were intentionally included in MAD^{4}HatTeR. Other panels have
538 less or no overlap[27,67,69] (Supplementary Tables 9-10). Efforts to increase overlap between
539 future versions of amplicon panels would facilitate more direct comparison of relatedness
540 between infections genotyped by different panels.
541
542 MAD^{4}HatTeR genotypes several key drug resistance markers as well as vaccine targets. The
543 primer pool configuration recommended for optimized sensitivity covers markers of resistance to
544 artemisinin, artemisinin-based combination therapy partner drugs, and other drugs used in

545 treatment, chemoprevention, and other interventions. Additionally, it targets the C-terminal
546 domain of *csp*, present in the RTS,S and R21 malaria vaccines currently recommended for use
547 in children living in areas with moderate to high malaria transmission[73–75]. Other drug resistance
548 markers and vaccine targets can be genotyped in high parasite density samples using the full
549 primer pool configuration. Nevertheless, primer design and target prioritization have
550 necessitated some exclusions. For example, the central repeat region of *csp*, also targeted by
551 the RTS,S and R21 vaccines, is not covered. Future iterations of MAD^{4}HatTeR should aim to
552 include additional targets, such as evolving drug resistance markers and candidate vaccine
553 targets.
554
555 Depth of coverage and amplification biases were reproducible across samples, with most
556 deviations likely due pipetting volume differences and systematic differences in laboratory
557 equipment and reagent batches. Detection of *hrp2*/*3* deletions and *mdr1* duplications was
558 achieved by applying a model that accounts for these factors. MAD^{4}HatTeR detected deletions
559 and duplications in mono- and polyclonal samples, even at low parasitemia. Additional data and
560 analytical developments could improve MAD^{4}HatTeR's performance in deletion and duplication
561 analysis. The current approach does not make use of COI estimates for inference and relies on
562 controls known not to have duplications or deletions in the target genes within each library
563 preparation batch. While target retrieval was generally uniform, some samples showed target
564 drop-off, indicating the need for multiple targets to avoid falsely calling a deletion. Nonetheless,
565 in its current form, MAD^{4}HatTeR serves as an efficient screening tool for identifying putative
566 duplications and deletions, which can then be validated with gold-standard methodologies.
567
568 Continuous improvement of the allele-calling bioinformatic pipeline is planned to increase
569 accuracy and usability. Masking of error-prone regions (e.g. homopolymers and tandem
570 repeats) is useful in reducing common PCR and sequencing errors, but it also removes

571 biological variation. This can be optimized by tailored masking of error hotspots, rather than
572 uniformly masking all low-diversity sequences. To improve the detection of low-abundance
573 alleles, we currently conduct a second inference round using alleles observed within a run as
574 priors, but this approach may also increase the risk of incorporating low-level contaminant
575 reads. Improvements in experimental strategies to detect and prevent cross-contamination[76],
576 along with post-processing filtering, could mitigate this. Additionally, curating an evolving allele
577 database from ongoing empiric data generation could replace the run-dependent priors, thereby
578 improving the accuracy and consistency of allele inference.
579
580 Integrating genomics into routine surveillance and developing genomic capacity in research and
581 public health institutions in malaria-endemic countries is facilitated by efficient, cost-effective,
582 reliable and accessible tools. MAD^{4}HatTeR is based on a commercially available method for
583 multiplexed amplicon sequencing[77]. As such, while primer sequences are publicly available
584 (Supplementary Table 2), reagents are proprietary. However, procuring bundled, quality
585 controlled reagents to generate libraries is straightforward, including for laboratories in malaria
586 endemic settings. Procurement costs for laboratory supplies often vary significantly, making
587 direct comparisons with other methods challenging, but we have found the method to be cost-
588 effective compared with other methods. At the time of writing, the list price for all library
589 preparation reagents, excluding plastics, consumables used for other steps (e.g. DNA
590 extraction), sequencing costs, taxes, or handling, was $12-25 per reaction, depending on order
591 volume. Sequencing costs can vary considerably based on the scale of sequencer used. For
592 optimal throughput, we recommend multiplexing up to 96 samples using a MiSeq v2 kit to
593 achieve results comparable to those shown here; much greater efficiency can be obtained with
594 higher throughput sequencers.
595

596 This study includes data from five laboratories, three of which are located in sub-Saharan Africa.
597 Beyond this study, MAD^{4}HatTeR is also being used by four other African laboratories for
598 applications ranging from estimating the prevalence of resistance-mediating mutations to
599 characterizing transmission networks. Expertise and computational infrastructure for advanced
600 bioinformatics and data analysis remains a challenge, with fewer users demonstrating autonomy
601 in these areas compared to wet lab procedures. The robustness of the method, along with
602 detailed training activities and materials (available online[78]), has facilitated easier
603 implementation. Future developments could also expand accessibility, including adaptations for
604 other sequencing platforms and panels targeting a smaller set of key loci for public health
605 decision-making.
606
607 In summary, MAD^{4}HatTeR is a powerful and fit-for-purpose addition to the malaria genomic
608 epidemiology toolbox, well-suited for a wide range of surveillance and research applications.
609
610 **List of abbreviations**
611 MAD^{4}HatTeR: Multiplex Amplicons for Drug, Diagnostic, Diversity, and Differentiation
612 Haplotypes using Targeted Resequencing
613 SNP: Single nucleotide polymorphism
614 WGS: Whole-genome sequencing
615 COI: Complexity of Infection
616 IBD: Identity-by-descent
617 DBS: Dried blood spot
618 WSAF: Within sample allele frequency
619
620
621

622 **Declarations**

623 **Ethics approvals and consent to participate**

624 Ethical approval for the study that generated the 26 field samples from Ethiopia[3] was granted by
625 the Ethiopia Public Health Institute (EPHI) Institutional Review Board (IRB; protocol EPHI-IRB-
626 033-2017) and WHO Research Ethics Review Committee (protocol ERC.0003174 001).
627 Processing of de-identified samples and data at the University of North Carolina at Chapel Hill
628 (UNC) was determined to constitute non-human subjects research by the UNC IRB (study 17-
629 0155). The study was determined to be non-research by the Centers for Disease Control (CDC)
630 and Prevention Human Subjects office (0900f3eb81bb60b9).
631 Study protocols for the study that generated the data for the 436 field samples from
632 Mozambique[22] were approved by the ethical committees of CISM and Hospital Clínic of
633 Barcelona, and the Mozambican Ministry of Health National Bioethics Committee.
634 Ethical approval for the studies that included the collection of blood samples used in mosquito
635 feeding assays was received from the Uganda Council of Science and Technology, Makerere
636 University School of Medicine, the University of California, and the London School of Hygiene &
637 Tropical Medicine.
638 Ethical approval for the study that collected *P. falciparum* and *P. vivax* samples in Northern
639 Ethiopia in 2022-2023 was obtained from the National Research Ethical Review Committee,
640 Addis Ababa, Ethiopia (reference number: 02/256/630/14), AHRI/ALERT Ethics Review
641 Committee (protocol number: P0-08-22), Aklilu Lemma Institute of Pathobiology Institutional
642 Research Ethics Review Committee (reference number: ALIPB IRERC/111/2015/23) and the
643 WCG IRB approval (protocol number: 1769134-1; IRB tracking number: 20214694).
644 Participants, or guardians/parents in the case of minors, in all these studies provided written
645 informed consent. All research was performed in accordance with relevant guidelines and
646 regulations.
647

Consent for publication

Not applicable.

Availability of data and material

All data are available in the Sequencing Read Archive, accession code PRJNA1180199. Code is available in GitHub (https://github.com/EPPIcenter/mad4hatter and https://github.com/andres-ad/madh_utilities)

Competing interests

J.B.P. reports research support from Gilead Sciences, non-financial support from Abbott Laboratories, and consulting for Zymeron Corporation, all outside the scope of the current work. All other authors report no potential conflicts of interest.

Funding

This work was supported by several grants from the Bill & Melinda Gates Foundation (INV-019032, OPP1132226, INV-037316, INV-024346, INV-031512, INV-003212, INV-024346). This research is also part of the ISGlobal's Program on the Molecular Mechanisms of Malaria which is partially supported by the Fundación Ramón Areces. We acknowledge support from the grant CEX2023-0001290-S funded by MCIN/AEI/ 10.13039/501100011033, from the Generalitat de Catalunya through the CERCA Program, from the Departament d'Universitats i Recerca de la Generalitat de Catalunya (AGAUR; grant 2017 SGR 664) and from the Ministerio de Ciencia e Innovación (PID2020-118328RB-I00/AEI/10.13039/501100011033). CISM is supported by the Government of Mozambique and the Spanish Agency for International Development (AECID). The parent study from which the 2017-2018 Ethiopia samples were derived was funded by the Global Fund to Fight AIDS, Tuberculosis, and Malaria through the Ministry of Health - Ethiopia (EPHI5405) and by the Bill & Melinda Gates Foundation through the World Health Organization

674 (OPP1209843). A.A.-D. was supported by the Chan Zuckerberg Biohub Collaborative
675 Postdoctoral fellowship. B.G. was supported by NIH-NIAID K24AI144048. J.B. was supported
676 by NIH-NIAID K23AI166009. J.L.S. was supported by NIH-NIAID 5K01AI153555. J.B.P. was
677 supported by NIH-NIAD R01 AI77791. The funders had no role in study design, data collection
678 and analysis, decision to publish, or preparation of the manuscript.
679

680 **Authors' contributions**

681 Designed the study: A.A.-D., N.H., A.B., J.L.S., E.G., B.G
682 Developed and benchmarked bioinformatic pipeline: A.A.-D., K.M., B.P., M.G.U., D.D.
683 Managed samples and data: A.A.-D, E.N.V, B.P., N.H, S.B, M.G.U., H.G., S.K, I.W., S.M.F,
684 J.B.P., W.L., E.E.
685 Generated data: A.A.-D., E.N.V., S.B., P.C, T.K., F.D.S., B.N., H.G., C.G.F., C.D.S., A.E.,
686 S.M.F., W.L., E.E.
687 Analyzed data: A.A.-D., E.N.V., K.M., B.P., N.H., I.G., W.L.
688 Interpreted data: A.A.-D., E.N.V., K.M., B.P., N.H., I.G., M.G.U., M.C., J.R., S.T., I.S., E.R.-V.,
689 C.T., J.B., A.M., B.G., W.L
690 Drafted the manuscript: A.A.D., E.N.V., K.M., B.P., B.G
691 All authors reviewed the manuscript.
692

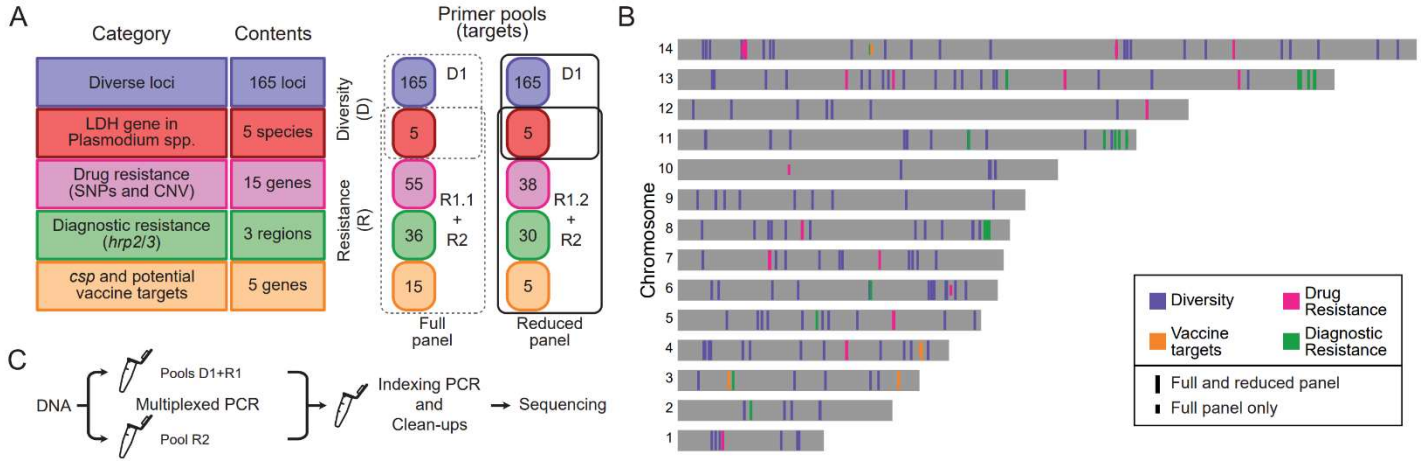
693 **Acknowledgments**

694 We thank Phil Rosenthal and Amy Bei for their input in panel design. We also thank members of
695 the EPPIcenter at UCSF, as well as the Rapid Response Team and the Genomics Platform at
696 the Chan Zuckerberg Biohub for valuable discussions.
697
698
699

700 **Figures**

701

702



703

704 **Figure 1**. **MAD^{4}HatTer is a multi-purpose malaria amplicon sequencing panel**.

705 **A**. Primer pools to amplify targets in 5 categories are grouped into two modules (Diversity and

706 Resistance). R1 refers to two primer pools: R1.1, the original pool, and R1.2, a reduced version

707 of primer pool R1.1 designed to increase sensitivity. The recommended configuration to

708 maximize information retrieval and sensitivity for low parasitemia samples are two mPCR

709 reactions, one with D1 and R1.2 primers, and one with R2 primers (solid lines). Supplementary

710 Tables 1-5 contain complete details on primer pools and targets.

711 **B**. Chromosomal locations of all targets in the *P. falciparum* genome (not including non-

712 *falciparum* targets). Note that the Diagnostic Resistance category includes targets in and around

713 *hrp2* and *hrp3* as well as targets in chromosome 11 that are often duplicated when *hrp3* is

714 deleted[44] and length controls in other chromosomes.

715 **C**. Simplified workflow for library preparation and sequencing, highlighting the need for two

716 multiplexed PCR reactions when using primer pools R1 and R2 which are incompatible due to

717 tiling over some genes of interest. A more detailed scheme can be found in Supplementary

718 Figure 1, and a full protocol, including didactic materials, can be found online[78].

719

720 **Table 1: SNPs associated with antimalarial resistance.**

721 SNPs of interest used to optimize target primer design that are covered by primer pools R1.1,
722 R1.2 or R2. The list excludes copy number variation markers, such as *plasmepsins* 2 and 3
723 (piperaquine) or *mdr1* (mefloquine). A full list of targets with the amino acid ranges they cover in
724 each gene can be found in the Supplementary Table 4.

Antimalarial†	Gene	Amino acids covered
Chloroquine and piperaquine	*crt*	72-76, 93, 97, 145, 218, 220, 271*, 326* 343, 350, 353, 356
Chloroquine	*aat1**	135*, 162*, 185*, 230*, 238*, 380*
Piperaquine	*exo*	415
Chloroquine and lumefantrine	*mdr1*	86, 184, 186, 371, 1034, 1042, 1246
Sulfadoxine	*dhps*	431, 436, 437, 540, 581, 613
Pyrimethamine and proguanil	*dhfr*	16, 51, 59, 108, 164
Artemisinin	*kelch13*	441, 446, 449, 458, 469, 476,481, 493, 515, 527, 537, 538, 539, 543, 553, 561, 568, 574, 580, 622, 675
	coronin	50, 100, 107
	fd	193
	arps10	127, 128
	mdr2	463, 484, 515
	PF3D7_1322700	236
	pib7	1484
	ubp-1	1525
	*pph**	1157*

725 † Antimalarial with validated or candidate markers[79], or with SNPs identified in GWAS studies[80].

726 * Included only in R1.1 and not in R1.2.

727

728

729

730

731

732

733

734 **Table 2: SNPs in *csp* and potential vaccine targets.**

735 SNPs of interest used to optimize target primer design that are covered by primer pools R1.1,
736 R1.2 or R2.1. A full list of targets with the amino acid ranges they cover in each gene can be
737 found in the Supplementary Table 4.

Gene	Amino acids covered
csp	280-398
*Ripr**	511, 673, 755, 985, 1039
*CyRPA**	339
*Rh5**	147, 170, 197, 203, 204, 221, 269, 350, 354, 357, 362
*P113**	106, 107, 234

738 * Included only in R1.1 and not in R1.2.

739

740

741

742

743

744

745

746

747

748

749

750

751

752

753

754

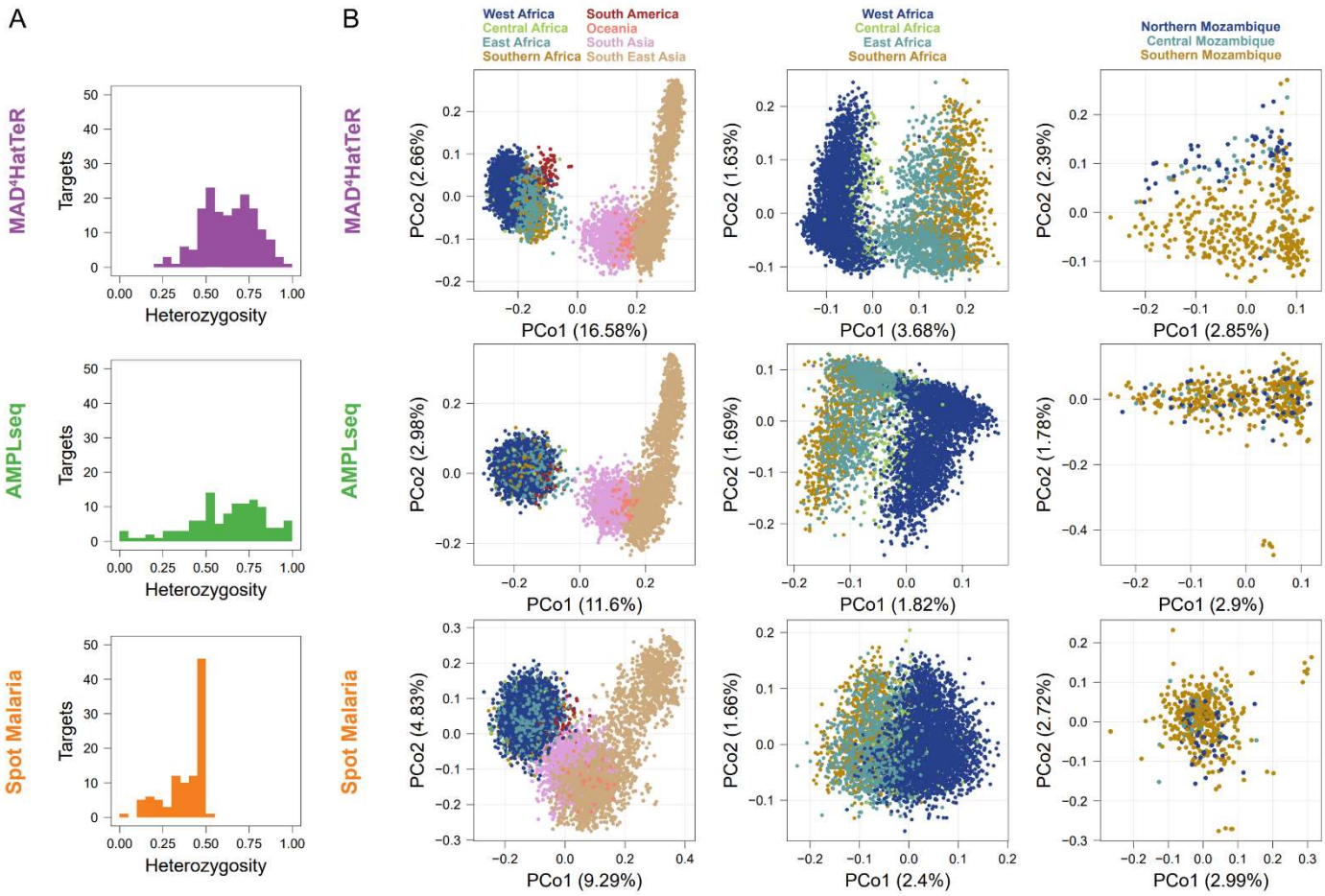


Figure 2. ***In silico* analysis demonstrates that MAD⁴HatTer's microhaplotypes capture high genetic diversity within African samples**. We reconstructed alleles (microhaplotypes) from publicly available WGS data to estimate genetic diversity. For SpotMalaria, SNP barcodes are used instead of microhaplotypes based on intended design and current usage. We note that additional information may be present within the amplified targets if microhaplotype sequences are accurately identifiable with appropriate bioinformatic processing. As such, alternate results for microhaplotypes reconstructed for the targets that contain the SNPs in each of those two panels are shown in Supplementary Figure 2.

A. Diversity module pool D1 includes more highly heterozygous targets than other published highly multiplexed panels. Only targets for diversity in each panel are included and heterozygosity is calculated for samples across Africa.

B. We performed principal coordinate analysis on alleles on global, African or Mozambican

data. The percentage of variance explained by each principal component is indicated in parentheses.

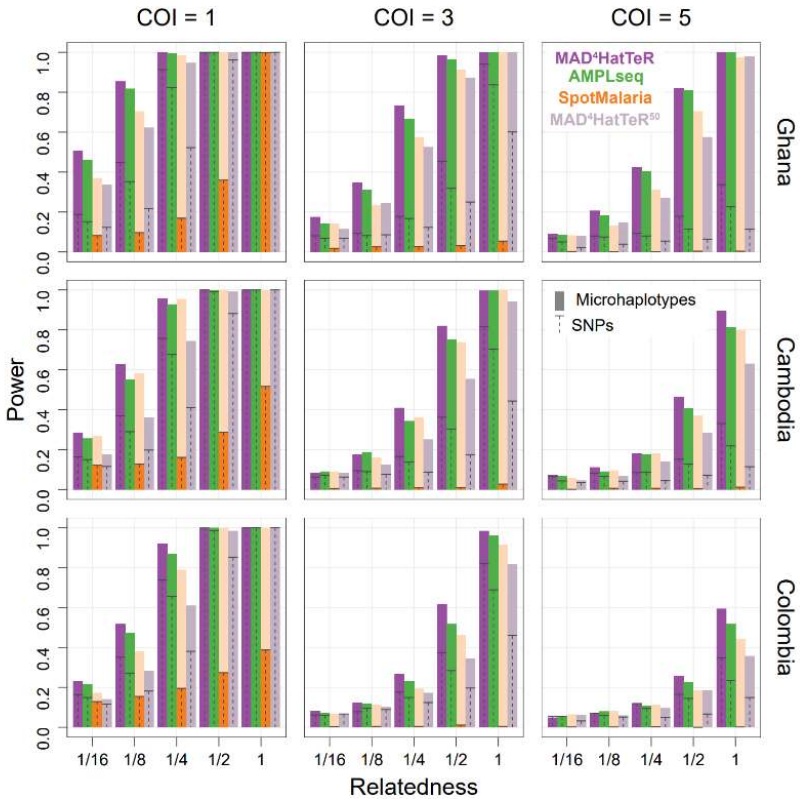
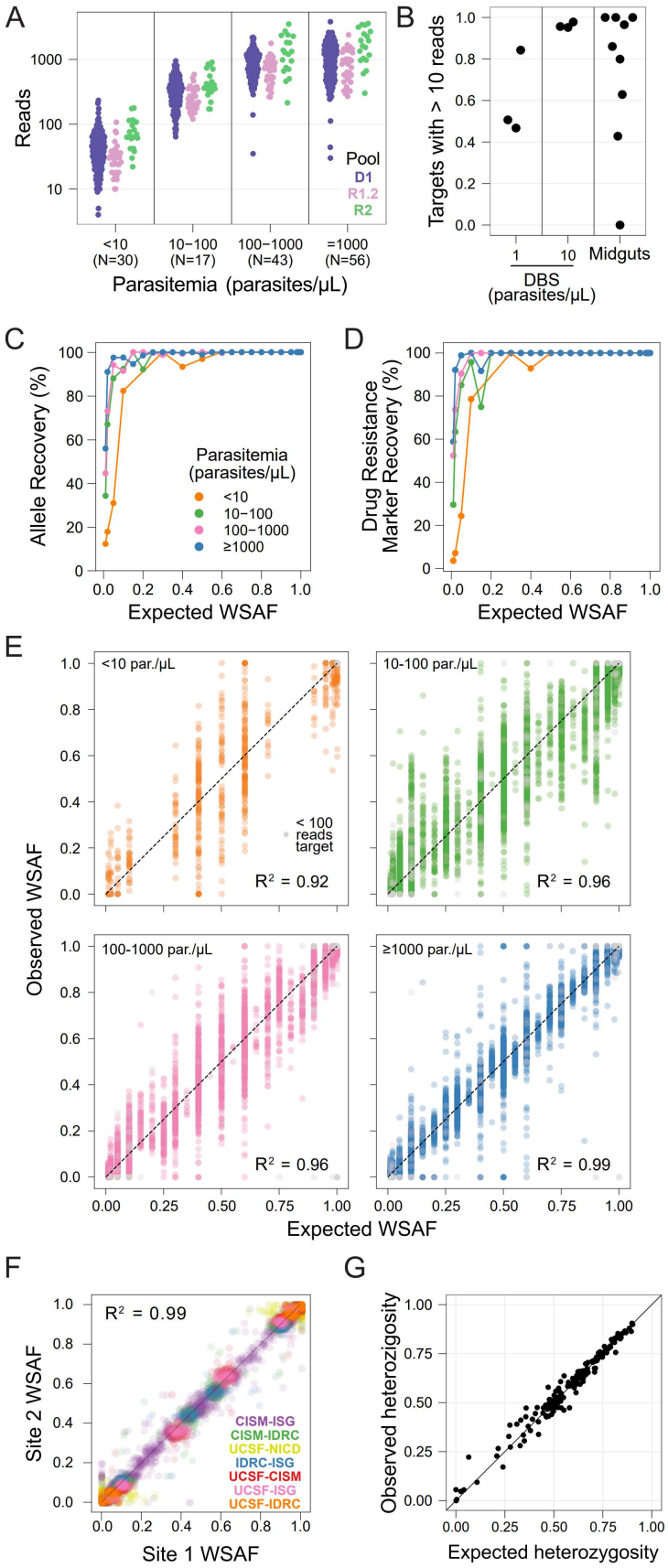
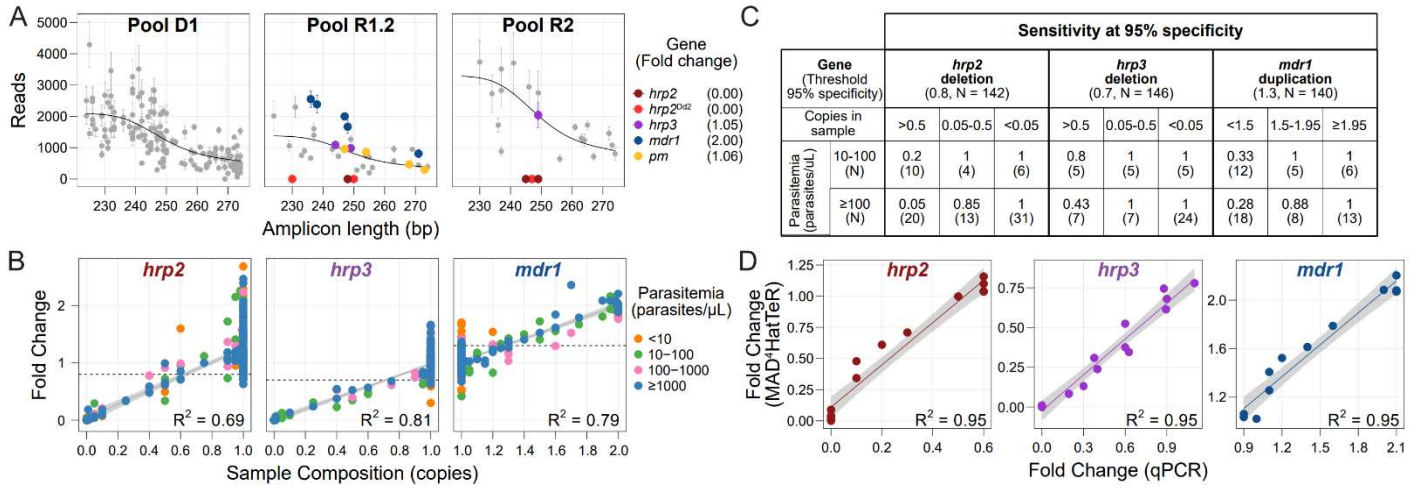


Figure 3. Power to identify relatedness of strains between infections is enhanced by highly multiplexed microhaplotypes. Simulated infections using population allele frequencies from available WGS data were used to estimate the power of testing if a pair of strains between infections is related. Countries in each of three continents with the most available WGS data were selected. Infections were simulated for a range of COI. Only one pair of strains between the infections was related with a given expected IBD proportion (r). The results were compared for reconstructed microhaplotypes and their most highly variable SNP for 3 panels (MAD^{4}HatTeR, SpotMalaria and AMPLseq). Note that SpotMalaria bioinformatics pipeline outputs a 100 SNP barcode, and thus its actual power (dark orange) is not reflective of the potential power afforded by microhaplotypes (light orange). Additionally, the 50 most diverse microhaplotypes and their corresponding SNPs were used to evaluate the effect of down-sizing MAD^{4}HatTeR (MAD^{4}HatTeR^{50}).


A
Reads
1000
100
10
Pool
D1
R1.2
R2
<10 (N=30)
10−100 (N=17)
100−1000 (N=43)
=1000 (N=56)
Parasitemia (parasites/µL)
B
Targets with > 10 reads
1.0
0.8
0.6
0.4
0.2
0.0
1
10
Midguts
DBS (parasites/µL)
C
Allele Recovery (%)
Parasitemia (parasites/µL)
<10
10−100
100−1000
≥1000
Expected WSAF
D
Drug Resistance Marker Recovery (%)
Expected WSAF
E
Observed WSAF
<10 par./µL
< 100 reads target
R² = 0.92
10-100 par./µL
R² = 0.96
100-1000 par./µL
R² = 0.96
≥1000 par./µL
R² = 0.99
Expected WSAF
F
Site 2 WSAF
R² = 0.99
CISM-ISG
CISM-IDRC
UCSF-NICD
IDRC-ISG
UCSF-CISM
UCSF-ISG
UCSF-IDRC
Site 1 WSAF
G
Observed heterozigosity
Expected heterozygosity

807

808 **Figure 4. MAD^{4}HatTeR produces reproducible and sensitive genetic data from a variety of**
809 **samples**
810 **A.** Mean read counts for each target in DBS controls (N in parenthesis in x-axis labels for each
811 parasitemia).
812 **B.** Proportion of targets with >10 reads in DBS controls with 1 and 10 parasites/μL and 9 midgut
813 samples (median parasite density equivalent to 0.9 parasites/μL in a DBS). 10 targets that
814 generally do not amplify well (>275 bp) were excluded.
815 **C-D.** Recovery within-sample allele frequency (WSAF) in the diversity module for 161 loci
816 across 183 samples (C), and biallelic SNPs in drug resistance markers across 20 codons in 165
817 samples (D).
818 **E**. Observed WSAF in laboratory mixed controls of known expected WSAF.
819 **F**. WSAF observed in libraries prepared and sequenced in different laboratories from the same
820 DBS mixed control. Participating laboratories are the EPPIcenter at the University of California
821 San Francisco (UCSF); Infectious Diseases Research Collaboration (IDRC), Uganda; Centro de
822 Investigação em Saúde de Manhiça (CISM), Mozambique; National Institutes for Communicable
823 Diseases (NICD), South Africa; and Barcelona Institute for Global Health (ISG), Spain.
824 **G.** Observed heterozygosity in field samples from Mozambique[22] and the respective expected
825 heterozygosity for each target obtained from available WGS data (which does not include the
826 MAD^{4}HatTeR-sequenced field samples).
827 False positives are excluded from C-G, as are targets with < 100 reads, except in E.
828



C

Gene (Threshold 95% specificity)		Sensitivity at 95% specificity								
		hrp2 deletion (0.8, N = 142)			*hrp3* deletion (0.7, N = 146)			*mdr1* duplication (1.3, N = 140)		
Copies in sample		>0.5	0.05-0.5	<0.05	>0.5	0.05-0.5	<0.05	<1.5	1.5-1.95	≥1.95
Parasitemia (parasites/uL)	10-100 (N)	0.2 (10)	1 (4)	1 (6)	0.8 (5)	1 (5)	1 (5)	0.33 (12)	1 (5)	1 (6)
	≥100 (N)	0.05 (20)	0.85 (13)	1 (31)	0.43 (7)	1 (7)	1 (24)	0.28 (18)	0.88 (8)	1 (13)

Figure 5. **MAD^{4}HatTeR can be used to screen for deletions and duplications**.

A. Technical replicates of Dd2 (a strain with *hrp2* deletion and *mdr1* duplication) with similar total reads were used to estimate fold changes in targets in and around *hrp2*, *hrp3*, *mdr1* and *plasmepsin2/3* (*pm*). A generalized additive model (black line) was applied to raw reads (Supplementary Figure 9) after correction by a control known not to have deletions or duplications in the genes of interest (3D7) to estimate fold changes in each of the genes. Note that there are two groups of *hrp2* targets, those that are deleted in field samples (*hrp2*) and those also deleted in Dd2 (*hrp2*^{Dd2}). Mean reads and fold changes are shown (N = 3); error bars denote standard deviation.

B. Estimated fold change for *hrp2*, *hrp3,* and *mdr1* loci in laboratory controls containing 1 or more strains at known proportions, or in field samples from Ethiopia[3] with known *hrp2* and *hrp3* deletions. Sample composition is estimated as the effective number of copies present in the sample based on the relative proportion of the strain carrying a deletion or duplication. Fold changes are obtained using the targets highlighted in A. Fold changes for Dd2-specific targets are shown in Supplementary Figure 10. Linear regression and R^2 values were calculated with data with parasitemia > 10 parasites/µL. The thresholds used to flag a sample as containing a duplication or deletion are shown in dashed black lines.

C. Sensitivity in detecting *hrp2* and *hrp3* deletions and *mdr1* duplications in controls, and field

847 samples from Ethiopia with known *hrp2* and *hrp3* deletions. Effective sample composition
848 (copies in sample) is estimated as in B. Sensitivity was calculated using a threshold to classify
849 samples with 95% specificity. Note that the small number of samples in the 0.05-0.5 copies
850 range may be responsible for the paradoxical lower sensitivity for higher parasitemia samples.
851 **D**. Estimated fold change for each gene correlates with qPCR quantification for the same
852 samples.
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872

873 **References**

874

875 1. Dalmat, R., Naughton, B., Kwan-Gett, T. S., Slyker, J. & Stuckey, E. M. Use cases for
876 genetic epidemiology in malaria elimination. *Malar. J.* **18**, 163 (2019).
877 2. Hamilton, W. L. *et al.* Evolution and expansion of multidrug-resistant malaria in southeast
878 Asia: a genomic epidemiology study. *Lancet Infect. Dis.* **19**, 943–951 (2019).
879 3. Feleke, S. M. *et al. Plasmodium falciparum* is evolving to escape malaria rapid diagnostic
880 tests in Ethiopia. *Nat. Microbiol.* **6**, 1289–1299 (2021).
881 4. Ndwiga, L. *et al.* A review of the frequencies of *Plasmodium falciparum* Kelch 13 artemisinin
882 resistance mutations in Africa. *Int. J. Parasitol. Drugs Drug Resist.* **16**, 155–161 (2021).
883 5. Rosenthal, P. J. *et al.* Cooperation in Countering Artemisinin Resistance in Africa: Learning
884 from COVID-19. *Am. J. Trop. Med. Hyg.* **106**, 1568–1570 (2022).
885 6. Neafsey Daniel E. *et al.* Genetic Diversity and Protective Efficacy of the RTS,S/AS01
886 Malaria Vaccine. *N. Engl. J. Med.* **373**, 2025–2037 (2015).
887 7. Wesolowski, A. *et al.* Mapping malaria by combining parasite genomic and epidemiologic
888 data. *BMC Med.* **16**, 190 (2018).
889 8. Tessema, S. *et al.* Using parasite genetic and human mobility data to infer local and cross-
890 border malaria connectivity in Southern Africa. *eLife* **8**, e43510 (2019).
891 9. Tessema, S. K. *et al.* Applying next-generation sequencing to track falciparum malaria in
892 sub-Saharan Africa. *Malar. J.* **18**, 268 (2019).
893 10. Watson, O. J. *et al.* Evaluating the Performance of Malaria Genetics for Inferring Changes in
894 Transmission Intensity Using Transmission Modeling. *Mol. Biol. Evol.* **38**, 274–289 (2021).
895 11. Daniels, R. F. *et al.* Genetic evidence for imported malaria and local transmission in Richard
896 Toll, Senegal. *Malar. J.* **19**, 276 (2020).

897 12. Mensah, B. A., Akyea-Bobi, N. E. & Ghansah, A. Genomic approaches for monitoring
898 transmission dynamics of malaria: A case for malaria molecular surveillance in Sub–
899 Saharan Africa. *Front. Epidemiol.* **2**, (2022).
900 13. Schaffner, S. F. *et al.* Malaria surveillance reveals parasite relatedness, signatures of
901 selection, and correlates of transmission across Senegal. *Nat. Commun.* **14**, 7268 (2023).
902 14. Fola, A. A. *et al.* Temporal and spatial analysis of *Plasmodium falciparum* genomics reveals
903 patterns of parasite connectivity in a low-transmission district in Southern Province, Zambia.
904 *Malar. J.* **22**, 208 (2023).
905 15. Yeka, A. *et al.* Comparative Efficacy of Artemether-Lumefantrine and Dihydroartemisinin-
906 Piperaquine for the Treatment of Uncomplicated Malaria in Ugandan Children. *J. Infect. Dis.*
907 **219**, 1112–1120 (2019).
908 16. Snounou, G. & Beck, H.-P. The Use of PCR Genotyping in the Assessment of
909 Recrudescence or Reinfection after Antimalarial Drug Treatment. *Parasitol. Today* **14**, 462–
910 467 (1998).
911 17. Uwimana, A. *et al.* Association of *Plasmodium falciparum kelch13* R561H genotypes with
912 delayed parasite clearance in Rwanda: an open-label, single-arm, multicentre, therapeutic
913 efficacy study. *Lancet Infect. Dis.* **21**, 1120–1128 (2021).
914 18. Schnoz, A. *et al.* Genotyping methods to distinguish *Plasmodium falciparum* recrudescence
915 from new infection for the assessment of antimalarial drug efficacy: an observational, single-
916 centre, comparison study. *Lancet Microbe* **5**, (2024).
917 19. Lover, A. A., Baird, J. K., Gosling, R. & Price, R. N. Malaria Elimination: Time to Target All
918 Species. *Am. J. Trop. Med. Hyg.* **99**, 17–23 (2018).
919 20. Mwesigwa, A. *et al. Plasmodium falciparum* genetic diversity and multiplicity of infection
920 based on *msp-1*, *msp-2*, *glurp* and microsatellite genetic markers in sub-Saharan Africa: a
921 systematic review and meta-analysis. *Malar. J.* **23**, 97 (2024).

922 21. Briggs, J. *et al.* Within-household clustering of genetically related *Plasmodium falciparum*
923 infections in a moderate transmission area of Uganda. *Malar. J.* **20**, 68 (2021).
924 22. Brokhattingen, N. *et al.* Genomic malaria surveillance of antenatal care users detects
925 reduced transmission following elimination interventions in Mozambique. *Nat. Commun.* **15**,
926 2402 (2024).
927 23. Viriyakosol, S. *et al.* Genotyping of *Plasmodium falciparum* isolates by the polymerase chain
928 reaction and potential uses in epidemiological studies. *Bull. World Health Organ.* **73**, 85–95
929 (1995).
930 24. Anderson, T. J. C., Su, X.-Z., Bockarie, M., Lagog, M. & Day, K. P. Twelve microsatellite
931 markers for characterization of *Plasmodium falciparum* from finger-prick blood samples.
932 *Parasitology* **119**, 113–125 (1999).
933 25. Anderson, T. J. C. *et al.* Microsatellite Markers Reveal a Spectrum of Population Structures
934 in the Malaria Parasite *Plasmodium falciparum*. *Mol. Biol. Evol.* **17**, 1467–1482 (2000).
935 26. Jacob, C. G. *et al.* Genetic surveillance in the Greater Mekong subregion and South Asia to
936 support malaria control and elimination. *eLife* **10**, e62997 (2021).
937 27. Kattenberg, J. H. *et al.* Molecular Surveillance of Malaria Using the PF AmpliSeq Custom
938 Assay for *Plasmodium falciparum* Parasites from Dried Blood Spot DNA Isolates from Peru.
939 *Bio-Protoc.* **13**, e4621 (2023).
940 28. Taylor, A. R., Jacob, P. E., Neafsey, D. E. & Buckee, C. O. Estimating Relatedness
941 Between Malaria Parasites. *Genetics* **212**, 1337–1351 (2019).
942 29. Tessema, S. K. *et al.* Sensitive, Highly Multiplexed Sequencing of Microhaplotypes From the
943 *Plasmodium falciparum* Heterozygome. *J. Infect. Dis.* **225**, 1227–1237 (2022).
944 30. LaVerriere, E. *et al.* Design and implementation of multiplexed amplicon sequencing panels
945 to serve genomic epidemiology of infectious disease: A malaria case study. *Mol. Ecol.*
946 *Resour.* **22**, 2285–2303 (2022).

947 31. de Cesare, M. *et al.* Flexible and cost-effective genomic surveillance of *P. falciparum*
948 malaria with targeted nanopore sequencing. *Nat. Commun.* **15**, 1413 (2024).
949 32. Holzschuh, A. *et al.* Using a mobile nanopore sequencing lab for end-to-end genomic
950 surveillance of *Plasmodium falciparum*: A feasibility study. *PLOS Glob. Public Health* **4**,
951 e0002743 (2024).
952 33. Girgis, S. T. *et al.* Drug resistance and vaccine target surveillance of *Plasmodium falciparum*
953 using nanopore sequencing in Ghana. *Nat. Microbiol.* **8**, 2365–2377 (2023).
954 34. Melnikov, A. *et al.* Hybrid selection for sequencing pathogen genomes from clinical samples.
955 *Genome Biol.* **12**, R73 (2011).
956 35. Villena, F. E., Lizewski, S. E., Joya, C. A. & Valdivia, H. O. Population genomics and
957 evidence of clonal replacement of *Plasmodium falciparum* in the Peruvian Amazon. *Sci.*
958 *Rep.* **11**, 21212 (2021).
959 36. Mathieu, L. C. *et al.* Local emergence in Amazonia of *Plasmodium falciparum k13* C580Y
960 mutants associated with in vitro artemisinin resistance. *eLife* **9**, e51015 (2020).
961 37. Cerqueira, G. C. *et al.* Longitudinal genomic surveillance of *Plasmodium falciparum* malaria
962 parasites reveals complex genomic architecture of emerging artemisinin resistance.
963 *Genome Biol.* **18**, 78 (2017).
964 38. Parobek, C. M. *et al.* Partner-Drug Resistance and Population Substructuring of Artemisinin-
965 Resistant *Plasmodium falciparum* in Cambodia. *Genome Biol. Evol.* **9**, 1673–1686 (2017).
966 39. Pelleau, S. *et al.* Adaptive evolution of malaria parasites in French Guiana: Reversal of
967 chloroquine resistance by acquisition of a mutation in *pfcrt*. *Proc. Natl. Acad. Sci.* **112**,
968 11672–11677 (2015).
969 40. Dara, A. *et al.* New var reconstruction algorithm exposes high var sequence diversity in a
970 single geographic location in Mali. *Genome Med.* **9**, 30 (2017).
971 41. Tvedte, E. S. *et al.* Evaluation of a high-throughput, cost-effective Illumina library
972 preparation kit. *Sci. Rep.* **11**, 15925 (2021).

973 42. Ahouidi, A., Ali, M. & et al. An open dataset of *Plasmodium falciparum* genome variation in
974 7,000 worldwide samples. *Wellcome Open Res.* **6**, 42 (2021).
975 43. Hathaway, N. A suite of computational tools to interrogate sequence data with local
976 haplotype analysis within complex *Plasmodium* infections and other microbial mixtures.
977 (2018) doi:10.13028/M2039K.
978 44. Hathaway, N. J. *et al.* Interchromosomal segmental duplication drives translocation and loss
979 of *P*. *falciparum* histidine-rich protein 3. *eLife* **13**, (2024).
980 45. MalariaGEN *et al.* Pf7: an open dataset of *Plasmodium falciparum* genome variation in
981 20,000 worldwide samples. *Wellcome Open Res.* **8**, 22 (2023).
982 46. Gerlovina, I., Gerlovin, B., Rodríguez-Barraquer, I. & Greenhouse, B. Dcifer: an IBD-based
983 method to calculate genetic distance between polyclonal infections. *Genetics* **222**, iyac126
984 (2022).
985 47. Murphy, M. & Greenhouse, B. MOIRE: a software package for the estimation of allele
986 frequencies and effective multiplicity of infection from polyallelic data. *Bioinformatics* **40**,
987 btae619 (2024).
988 48. Esayas, E. *et al.* Impact of nighttime human behavior on exposure to malaria vectors and
989 effectiveness of using long-lasting insecticidal nets in the Ethiopian lowlands and highlands.
990 *Parasit. Vectors* **17**, 520 (2024).
991 49. Asua, V. *et al. Plasmodium* Species Infecting Children Presenting with Malaria in Uganda.
992 (2017) doi:10.4269/ajtmh.17-0345.
993 50. Rek, J. *et al.* Asymptomatic School-Aged Children Are Important Drivers of Malaria
994 Transmission in a High Endemicity Setting in Uganda. *J. Infect. Dis.* **226**, 708–713 (2022).
995 51. Andolina, C. *et al.* Sources of persistent malaria transmission in a setting with effective
996 malaria control in eastern Uganda: a longitudinal, observational cohort study. *Lancet Infect.*
997 *Dis.* **21**, 1568–1578 (2021).

998 52. Teyssier, N. B. *et al.* Optimization of whole-genome sequencing of *Plasmodium falciparum*
999 from low-density dried blood spot samples. *Malar. J.* **20**, 116 (2021).
1000 53. Hugo, L. E. *et al.* Rapid low-resource detection of *Plasmodium falciparum* in infected
1001 *Anopheles* mosquitoes. *Front. Trop. Dis.* **5**, (2024).
1002 54. Hofmann, N. *et al.* Ultra-Sensitive Detection of Plasmodium falciparum by Amplification of
1003 Multi-Copy Subtelomeric Targets. *PLOS Med.* **12**, e1001788 (2015).
1004 55. Mayor, A. *et al.* Sub-microscopic infections and long-term recrudescence of *Plasmodium*
1005 *falciparum* in Mozambican pregnant women. *Malar. J.* **8**, 9 (2009).
1006 56. Molla, E. *et al.* Seasonal Dynamics of Symptomatic and Asymptomatic *Plasmodium*
1007 *falciparum* and *Plasmodium vivax* Infections in Coendemic Low-Transmission Settings,
1008 South Ethiopia. (2024) doi:10.4269/ajtmh.24-0021.
1009 57. Paragon Genomics Product Documents. *Paragon Genomics*
1010 https://www.paragongenomics.com/customer-support/product_documents/.
1011 58. Di Tommaso, P. *et al.* Nextflow enables reproducible computational workflows. *Nat.*
1012 *Biotechnol.* **35**, 316–319 (2017).
1013 59. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads.
1014 *EMBnet.journal* **17**, 10–12 (2011).
1015 60. Callahan, B. J. *et al.* DADA2: High-resolution sample inference from Illumina amplicon data.
1016 *Nat. Methods* **13**, 581–583 (2016).
1017 61. Gupta, H. *et al.* Drug-Resistant Polymorphisms and Copy Numbers in *Plasmodium*
1018 *falciparum*, Mozambique, 2015. *Emerg. Infect. Dis.* **24**, 40–48 (2018).
1019 62. Grignard, L. *et al.* A novel multiplex qPCR assay for detection of *Plasmodium falciparum*
1020 with histidine-rich protein 2 and 3 (pfhrp2 and pfhrp3) deletions in polyclonal infections.
1021 *EBioMedicine* **55**, 102757 (2020).

1022 63. da Silva, C. *et al.* Targeted and whole-genome sequencing reveal a north-south divide in *P.*
1023 *falciparum* drug resistance markers and genetic structure in Mozambique. *Commun. Biol.* **6**,
1024 1–11 (2023).
1025 64. Emiru, T. *et al.* Evidence for a role of *Anopheles stephensi* in the spread of drug- and
1026 diagnosis-resistant malaria in Africa. *Nat. Med.* **29**, 3203–3211 (2023).
1027 65. Daniels, R. F. *et al.* Modeling malaria genomics reveals transmission decline and rebound in
1028 Senegal. *Proc. Natl. Acad. Sci.* **112**, 7067–7072 (2015).
1029 66. Chang, H.-H. *et al.* Mapping imported malaria in Bangladesh using parasite genetic and
1030 human mobility data. *eLife* **8**, e43481 (2019).
1031 67. Holzschuh, A. *et al.* Multiplexed ddPCR-amplicon sequencing reveals isolated *Plasmodium*
1032 *falciparum* populations amenable to local elimination in Zanzibar, Tanzania. *Nat. Commun.*
1033 **14**, 3699 (2023).
1034 68. Hathaway, N. J., Parobek, C. M., Juliano, J. J. & Bailey, J. A. SeekDeep: single-base
1035 resolution de novo clustering for amplicon deep sequencing. *Nucleic Acids Res.* **46**, e21
1036 (2018).
1037 69. Lerch, A. *et al.* Development of amplicon deep sequencing markers and data analysis
1038 pipeline for genotyping multi-clonal malaria infections. *BMC Genomics* **18**, 864 (2017).
1039 70. Schaffner, S. F., Taylor, A. R., Wong, W., Wirth, D. F. & Neafsey, D. E. hmmIBD: software
1040 to infer pairwise identity by descent between haploid genotypes. *Malar. J.* **17**, 196 (2018).
1041 71. Henden, L., Lee, S., Mueller, I., Barry, A. & Bahlo, M. Identity-by-descent analyses for
1042 measuring population dynamics and selection in recombining pathogens. *PLOS Genet.* **14**,
1043 e1007279 (2018).
1044 72. Chang, H.-H. *et al.* THE REAL McCOIL: A method for the concurrent estimation of the
1045 complexity of infection and SNP allele frequency for malaria parasites. *PLOS Comput. Biol.*
1046 **13**, e1005348 (2017).

1047 73. Collins, K. A., Snaith, R., Cottingham, M. G., Gilbert, S. C. & Hill, A. V. S. Enhancing
1048 protective immunity to malaria with a highly immunogenic virus-like particle vaccine. *Sci.*
1049 *Rep.* **7**, 46621 (2017).
1050 74. Laurens, M. B. RTS,S/AS01 vaccine (Mosquirix™): an overview. *Hum. Vaccines*
1051 *Immunother.* **16**, 480–489 (2020).
1052 75. World Health Organization. Malaria vaccine: WHO position paper – May 2024. *Wkly.*
1053 *Epidemiol. Rec.* **19**, 225–248 (2024).
1054 76. Lagerborg, K. A. *et al.* Synthetic DNA spike-ins (SDSIs) enable sample tracking and
1055 detection of inter-sample contamination in SARS-CoV-2 sequencing workflows. *Nat.*
1056 *Microbiol.* **7**, 108–119 (2022).
1057 77. CleanPlex amplicon sequencing for targeted DNA and RNA Seq. *Paragon Genomics*
1058 https://www.paragongenomics.com/targeted-sequencing/amplicon-sequencing/cleanplex-
1059 ngs-amplicon-sequencing/.
1060 78. Resources | EPPIcenter. https://eppicenter.ucsf.edu/resources.
1061 79. Report on antimalarial drug efficacy, resistance and response: 10 years of surveillance
1062 (2010-2019). https://www.who.int/publications/i/item/9789240012813.
1063 80. Miotto, O. *et al.* Genetic architecture of artemisinin-resistant *Plasmodium falciparum*. *Nat.*
1064 *Genet.* **47**, 226–234 (2015).
1065