

1 **Microhaplotype deep sequencing assays to capture *Plasmodium vivax* infection lineages**

2

3 Mariana Kleinecke¹, Angela Rumaseb¹, Edwin Sutanto², Hidayat Trimarsanto^{1,20}, Kian Soon Hoon¹, Ashley
4 Osborne¹, Paulo Manrique^{3,4}, Trent Peters⁵, David Hawkes⁵, Ernest Diez Benavente⁶, Georgia Whitton⁷,
5 Sasha V Siegel⁷, Richard D Pearson⁷, Roberto Amato⁷, Anjana Rai¹, Nguyen Thanh Thuy Nhien⁸, Nguyen
6 Hoang Chau⁸, Ashenafi Assefa⁹, Tamiru S Degaga¹⁰, Dagimawie Tadesse Abate¹⁰, Awab Ghulam Rahim¹¹,
7 Ayodhia Pitaloka Pasaribu¹², Inge Sutanto¹³, Mohammad Shafiu Alam¹⁴, Zuleima Pava¹, Tatiana Lopera-
8 Mesa¹⁵, Diego Echeverry¹⁶, Tim William^{1,17}, Nicholas M Anstey¹, Matthew J Grigg¹, Nicholas Day^{18,19},
9 Nicholas J. White^{18,19}, Dominic P Kwiatkowski^{7*}, Rintis Noviyanti²⁰, Daniel Neafsey^{3,4}, Ric N Price^{1,18,19},
10 Sarah Auburn^{1,18 **}

11

- 12 1. Menzies School of Health Research and Charles Darwin University, Darwin, Australia
- 13 2. Exeins Health Initiative, South Jakarta, Indonesia
- 14 3. Harvard T.H. Chan School of Public Health, USA
- 15 4. Broad Institute, USA
- 16 5. Australian Genome Research Facility, Australia
- 17 6. Laboratory of Experimental Cardiology, Department of Cardiology, University Medical Center
18 Utrecht, Utrecht, the Netherlands
- 19 7. Wellcome Sanger Institute, Hinxton, UK
- 20 8. Oxford University Clinical Research Unit, Hospital for Tropical Diseases, Ho Chi Minh City,
21 Vietnam
- 22 9. Ethiopian Public Health Institute, Addis Ababa, Ethiopia
- 23 10. College of Medicine & Health Sciences, Arba Minch University, Arba Minch, Ethiopia
- 24 11. Nangarhar Medical Faculty, Nangarhar University, Ministry of Higher Education, Jalalabad,
25 Afghanistan
- 26 12. Universitas Sumatera Utara, Medan, Indonesia
- 27 13. Faculty of Medicine, University of Indonesia, Jakarta, Indonesia
- 28 14. Infectious Diseases Division, International Centre for Diarrhoeal Disease Research, Bangladesh
29 (icddr,b), Dhaka, Bangladesh
- 30 15. Universidad de Antioquia, Colombia
- 31 16. Departamento de Microbiología, Facultad de Salud, Universidad del Valle, Cali, Colombia
- 32 17. Queen Elizabeth Hospital, Malaysia

33 18. Centre for Tropical Medicine and Global Health, Nuffield Department of Medicine, University of
34 Oxford, Oxford, United Kingdom

35 19. Mahidol-Oxford Tropical Medicine Research Unit, Mahidol University, Bangkok, Thailand

36 20. Eijkman Molecular Biology Research Center, National Research and Innovation Agency,
37 Cibinong, Indonesia

38

39 *Deceased

40

41 **Corresponding author: A/Prof Sarah Auburn, Sarah.Auburn@Menzies.edu.au, Menzies School of
42 Health Research, PO Box 41096, Casuarina, Darwin, NT 0811, Australia; Tel: (+61) 8 8946 8503

43

44 **Key words**

45 Plasmodium vivax; malaria; microhaplotype; relatedness; identity-by-descent; recurrence; relapse;
46 amplicon sequencing

47

48 **Abstract**

49

50 The elimination of *Plasmodium vivax* is challenged by dormant liver stages (hypnozoites) that can
51 reactivate months after initial infection resulting in relapses that enhance transmission. Relapsing
52 infections confound antimalarial clinical efficacy trials due to the inability to distinguish between
53 recurrences arising from blood-stage treatment failure (recrudescence), reinfection or relapse. Genetic
54 relatedness of paired parasite isolates, measured by identity-by-descent (IBD), can provide important
55 information on whether individuals have had single or multiple mosquito inoculations, thus informing on
56 recurrence origin. We developed a high-throughput amplicon sequencing assay comprising 93 multi-SNP
57 (microhaplotype) *P. vivax* markers to determine IBD between *P. vivax* clinical isolates. The assay was
58 evaluated in 659 independent infections from the Asia-Pacific and Horn of Africa, including 108 pairs of
59 infections from a randomized controlled trial (RCT). A bioinformatics pipeline (vivaxGEN-MicroHaps) was
60 established to support data processing. Simulations using *paneljudge* demonstrated low error in
61 pairwise IBD estimation in all countries assessed (all RMSE <0.12) and IBD-based networks illustrated
62 strong clustering by geography. IBD analysis in the RCT demonstrated a higher frequency of suspected
63 relapses or recrudescence in patients treated with primaquine regimens (0.84) compared to those

64 without primaquine (0.60). Our results demonstrate the potential to derive information on *P. vivax* IBD
65 using amplicon sequencing, that informs policy-relevant treatment and transmission dynamics.

66

67 **Introduction**

68

69 Outside of sub-Saharan Africa, *Plasmodium vivax* is becoming the predominant cause of malaria¹. The
70 control and elimination of *P. vivax* is confounded by the parasite's ability to form dormant liver stages
71 (hypnozoites) that can reactivate weeks to months after initial infection, causing recurrent episodes of
72 malaria known as relapses². The radical cure of vivax malaria requires treatment with both a
73 schizontocidal antimalarial (chloroquine (CQ) or artemisinin combination therapy) combined with a
74 hypnozoitocidal agent (primaquine or tafenoquine)¹. *P. vivax* resistant to chloroquine (CQR) has been
75 documented in several endemic countries^{3,4}, but surveillance of the emergence and spread of CQR is
76 undermined by a lack of reliable molecular markers^{2,5}. Although chloroquine efficacy can be defined
77 using clinical efficacy trials to determine the risk of recurrent parasitaemia within 28 to 42 days of
78 treatment, estimates are confounded by an inability to distinguish whether recurrent infections are due
79 to schizontocidal treatment failure (recrudescence), reactivation of hypnozoites (relapses), or a new
80 mosquito inoculation (reinfection)^{6,7}. In the case a resistant infection is detected, traditional
81 surveillance tools, such as patient travel history, may also be limited in capacity to assess the origin or
82 spread of the given strains within a community². Molecular approaches provide great potential to
83 address these knowledge gaps^{6,8}.

84

85 Parasite genotyping of infection pairs (pre- and post-treatment), is well established for interpreting
86 antimalarial clinical efficacy for *P. falciparum*, to distinguish between recrudescence (same/homologous
87 pairs) and reinfection (different/heterologous pairs)⁹. However, a similar approach to defining clinical
88 efficacy for *P. vivax* infection is undermined by relapses that can be homologous or heterologous to the
89 pre-treatment isolate¹⁰⁻¹². Genomic studies of *P. vivax* have revealed that a proportion of paired clinical
90 isolates that are classified as heterologous using traditional genotyping methods, can share homology in
91 large segments of the genome, inferring recent common ancestry^{13,14}. These familial relationships can
92 be defined using genetic information on identity-by-descent (IBD). Closely related paired infections (e.g.,
93 siblings, with $\geq 50\%$ IBD) are more likely to come from a single mosquito inoculation and thus to be
94 relapses rather than reinfections. Genome-wide data are ideal for quantifying IBD. However, this is not
95 feasible in most malaria-endemic settings. Targeted next generation sequencing (NGS)-based methods,

96 such as amplicon sequencing, offer an affordable and versatile alternative¹⁵. Previously we have
97 established a bioinformatic framework for retrieving genome-wide panels of *P. vivax* microhaplotype
98 markers; short (200 bp) genomic segments comprising multiple high diversity single nucleotide
99 polymorphisms (SNPs)⁸. Using *in-silico* analyses, we demonstrated that ~100 *P. vivax* microhaplotypes
100 can reliably capture IBD between paired isolates and could be used to define spatio-temporal patterns
101 of *P. vivax* infection⁸.

102
103 In this study, we advanced our previous *in-silico* work by designing an NGS-based amplicon sequencing
104 assay for 98 *P. vivax* amplicons, including 93 microhaplotypes that capture IBD accurately in diverse
105 endemic areas. Our rhAmpSeq library preparation protocol incorporates all markers in a single plex
106 reaction and was able to generate high throughput, cost-effective, sensitive and specific data from
107 global isolates. Standard population-level genetic metrics could be applied to the data to highlight
108 potential use cases for National Malaria Control Programs (NMCPs) that will inform *P. vivax* treatment
109 options and transmission dynamics.

111 **Results**

113 ***A rhAmpSeq multiplex of 93 microhaplotypes with high sensitivity and specificity***

114 A two-step multiplex PCR-based library preparation assay was established using rhAmpSeq chemistry
115 (Integrated DNA Technologies, IDT). The multiplex amplifies one species confirmation marker, four
116 putative markers of drug resistance, and 93 microhaplotypes distributed relatively uniformly across the
117 *P. vivax* genome (Supplementary Data 1; Supplementary Figure 1). To evaluate the assays technical and
118 analytical performance, a total of 750 *P. vivax* infections were assessed, including 108 paired *P. vivax*
119 infections from a randomized controlled trial, two *P. vivax* serial dilution panels, and 10 non-*P. vivax*
120 control samples (Supplementary Data 2, Supplementary Figure 2).

121
122 The specificity of the assay was evaluated using a selection of 7 non-*vivax Plasmodium* spp. and 3
123 uninfected human DNA controls. None (0/10) of the control samples were permissible to successful read
124 pair calling using *DADA2*. Examination of amplicon coverage revealed that 58 markers (60% [58/97] after
125 excluding the mitochondrial species marker) exhibited amplicon coverage <0.9 in all negative controls
126 which would not yield successful read-pairs (Supplementary Figure 3a). Amongst the 39 markers with
127 coverage >0.9 in one or more negative controls, 37 (95%) had read counts <25 in all controls tested, with

128 only 2 markers (markers 307891 and 313080) exceeding 25 reads and only in the *P. knowlesi* controls
129 (Supplementary Figure 3b).

130
131 The assay sensitivity was assessed using serial dilutions of two *P. vivax*-infected patient blood samples
132 (KV3 and KV5) under high sample multiplexing (library pooled across 384 samples), using both the
133 standard PCR step 1 DNA input and reaction volumes (referred to as full chemistry with a 20 ul reaction
134 volume) and halving the PCR step 1 DNA input and reaction volumes (referred to as half chemistry with
135 a 10 ul reaction volume). Under full chemistry conditions, depths exceeding 100 reads per marker were
136 observed in the 70 parasite/ul (mean 95-160 across triplicates) and 96 parasite/ul (mean 74-138 across
137 triplicates) sample preparations, as well as in higher densities (Supplementary Figures 4a and c). Under
138 half chemistry conditions, similar depths were observed in the 70 parasite/ul (mean 235-280 across
139 triplicates) and 96 parasite/ul (mean 167-194 across triplicates) sample preparations and higher parasite
140 densities (Supplementary Figures 4b and d).

141
142 Parasite density defined by microscopic blood film examination may not correlate directly with parasite
143 DNA yield owing to the presence of multiple life cycle stages in *P. vivax* infections (schizonts, for
144 example, having greater DNA quantity than ring stages). To address this, we evaluated the threshold
145 cycle (Ct) values of the KV5 serial dilution using real-time PCR, targeting the *pvmtcox1* gene. Samples
146 with a Ct <30 (KV5 96 parasite/ul) yielded a mean of ≥ 100 reads or more, while samples with a Ct <33
147 (KV5 9.6 parasite/ul) resulted in a mean yield of ≥ 9 reads (Supplementary Figure 5, Supplementary Table
148 1).

149
150 ***Geographically diverse application potential***

151 A total of 659 *P. vivax* isolates were used to evaluate amplification efficacy from 8 countries across the
152 globe. Successful genotyping was defined as ≥ 25 reads for >80% of the 97 nuclear genome amplicons
153 and could be achieved in 627 (95.1%) isolates. 500 (79.7%) independent (non-recurrent) samples were
154 taken forward for evaluation of individual marker efficacy and country-specific amplification patterns. A
155 minimum of 5 isolates were available from 6 countries: Afghanistan (n=156), Bangladesh (n=5),
156 Colombia (n=5), Ethiopia (n=213), Indonesia (n=38) and Vietnam (n=83). Aside from 146 samples from
157 East Shewa, Ethiopia, and 38 samples from Sumatra, Indonesia, data from all isolates was generated
158 from a single 384-well run (run 3). One marker (419038) demonstrated poor amplification in all run 3
159 populations with both read (SNP-based) and read-pair (microhaplotype-based) analyses (Supplementary

160 Data 3, Figure 1). Marker 419038 was therefore excluded from several downstream analyses (see
161 Supplementary Figure 1 for usage). However, increasing the primer concentration for marker 419038
162 enhanced amplification in later runs, as demonstrated in the East Shewa and Sumatran samples from
163 runs 5 and 6. Three markers (64721, 354590 and 466426) exhibited a large drop in read-pair counts in a
164 moderate proportion of samples after denoising with *DADA2* (Figure 1b). Marker 354590 was most
165 impacted by the *DADA2* denoising and was excluded from all downstream read-pair analyses. The
166 remaining 93 markers exhibited >50% successful genotype calls in majority of populations assessed.

167

168 ***Accuracy in P. vivax variant calling***

169 Using a set of 83 isolates with both whole genome sequencing (WGS) and amplicon sequencing data, the
170 accuracy of variant calling (applying the default 10% minor allele threshold) was confirmed at the 425
171 SNPs within the 93 microhaplotypes. Concordance was observed at 98.3% (32,352/32,923) of the
172 genotype calls (Supplementary Table 2, Supplementary Data 4). Homozygous reference versus
173 homozygous alternate allele discordances contributed 0.06% (19/32,923) of all calls. Most discordant
174 calls reflected heterozygous versus homozygous calls (96.7% (552/571)), likely reflecting differences in
175 the limit of detection of minor clones between the datasets. On further exploring the concordance after
176 applying a 1% minor allele threshold to both datasets, a notable difference was more than doubling of
177 heterozygous amplicon sequencing calls that were homozygous in the WGS dataset (from 225 calls with
178 10% threshold to 550 with 1% threshold; Supplementary Table 3); this trend likely reflects a greater limit
179 of detection in the deep sequencing (amplicon) data.

180

181 ***High potential to capture within-host infection complexity***

182 Using a set of 83 isolates with both WGS and microhaplotype data, within-host infection complexity was
183 explored using the *Fws* and effective MOI (eMOI). As illustrated in Figures 2a and b, there was a strong
184 correlation between the genome-wide *Fws* and the microhaplotype-based eMOI in each geographic
185 region assessed (Spearman's rank correlation, $\rho = -0.700$ $p = 2.389 \times 10^{-13}$). However, amongst 59
186 infections with $Fws \geq 0.95$ (traditionally used as a cut-off to define monoclonal infections), 5 isolates had
187 >95% probability of being polyclonal according to MOIRE analysis of the microhaplotypes (eMOI range
188 1.08-1.93) (Supplementary Data 5). In contrast, none of the infections defined as polyclonal with
189 genomic data ($Fws < 0.95$) were monoclonal with the microhaplotypes (all 100% probability of
190 polyclonality).

191

192 The eMOI distributions in Figures 2c and d illustrate trends in within-host diversity at the country and
193 provincial level in a larger panel of 498 isolates from 4 populations which had over 30 isolates available:
194 Afghanistan (n=158), Indonesia (n=38), Ethiopia (n=213) and Vietnam (n=89). The data reveals
195 differences within-country in Afghanistan, where Laghman exhibited a trend (not significant) of lower
196 eMOI than Nangahar (Wilcoxon rank sum test, $p=0.324$). The lowest eMOI distribution was observed in
197 Sumatra, Indonesia, suggestive of low endemicity relative to the other sites.

198

199 ***Effective IBD capture***

200 To assess the ability of the microhaplotype panel to capture IBD accurately, we used *paneljudge* (an R
201 package to judge the performance of a panel of genetic markers using simulated data) to simulate the
202 relative mean square error (RMSE) in estimation of five pairwise IBD states: IBD=0 (unrelated infections),
203 IBD=0.25 (half-siblings), IBD=0.5 (siblings), IBD=0.75 (highly related), and IBD=1 (identical clones). The
204 simulations were run in the 4 major populations (Afghanistan, Ethiopia, Indonesia and Vietnam) using
205 data from the 92 assayable microhaplotypes, revealing moderately high diversity (mean diversity range
206 0.44-0.65) and low RMSE (<0.12 for all pairwise IBD states) in each population (Supplementary Figure 6,
207 Figure 3).

208

209 ***Potential of microhaplotype-based IBD to inform on recurrence***

210 To demonstrate the potential of the assay to inform on the origin of recurrences, the microhaplotype-
211 based IBD was determined using *DCifer* on data from 108 pairs of initial and recurrent infections
212 (Supplementary Data 6). These isolates came from individuals enrolled into a randomized controlled trial
213 at two sites in Ethiopia and treated with either chloroquine (CQ), CQ plus primaquine, artemether-
214 lumefantrine (AL) or AL plus primaquine¹⁶. As illustrated in Figure 4a, the patients treated without
215 primaquine had a higher frequency of highly related pairs (arbitrary IBD threshold ≥ 0.25), compared to
216 those treated with primaquine, consistent with a greater risk of relapsing and recrudescing infections. In
217 total, 84% (74/88) of paired isolates in patients not treated with primaquine had IBD $\geq 25\%$ compared to
218 60% (12/20) of those treated with primaquine ($\chi^2 = 4.44$, $p=0.0351$). Overall, 20% (18/70) of paired
219 isolates recurring within 120 days of treatment had low relatedness (IBD $<25\%$) compared to 55%
220 (11/20) recurring after 120 days; ($\chi^2 = 8.22$, $p=0.004$), consistent with a rising proportion of reinfections
221 at later time points (Figure 4b).

222

223 ***Spatial transmission potential***

224 The utility of the marker panel to define spatial transmission dynamics was explored using the isolates
225 from Afghanistan, Ethiopia, Indonesia and Vietnam. As illustrated in the PCoA and neighbour-joining
226 trees in Figures 5a and 5b, SNP-based identity-by-state (IBS) analyses on 280 infections (with <50%
227 probability of polyclonality) demonstrated differentiation between borders. In accordance with the low
228 eMOI distribution, isolates from Sumatra, Indonesia, revealed clusters of genetically identical infections
229 indicative of clonal expansion. IBD analyses on the microhaplotype panel conducted in both clonal and
230 polyclonal infections (n=498) from the same sites revealed evidence of local and inter-site transmission
231 networks and confirmed the clonal clustering in Sumatra (Figures 6a-d).

232

233 ***Effective Plasmodium species confirmation***

234 In addition to the microhaplotypes, a previously described mitochondrial amplicon was included in the
235 assay to confirm *Plasmodium* spp.¹⁷. The assay amplifies coordinates PvP01_MIT_V2:2904-3149, which
236 include species-specific SNPs and indels. Using *P. vivax* samples from a range of countries, and *P.*
237 *falciparum*, *P. malariae*, *P. ovale* spp. and *P. knowlesi* negative controls, we confirmed amplification of
238 the mitochondrial marker at moderate to high depth and coverage in all *Plasmodium* species (range 28-
239 5212) (Supplementary Table 4). Concordance between PCR-based and mitochondrial species
240 classification was confirmed for each of the *Plasmodium* samples.

241

242 ***Drug resistance candidates***

243 Our assay included a non-exhaustive selection of amplicons encompassing markers of antimalarial drug
244 resistance including the multidrug resistance 1 (*pvmdr1*) 976 and 1076 loci, and dihydropteroate
245 synthase (*pvdhps*) 383 and 553 loci that have been associated with *ex vivo* or clinical phenotypes (see
246 ¹⁸). The dihydrofolate reductase (*pvdhfr*) 57, 58, 61 and 117 loci did not multiplex effectively with the
247 other markers based on bioinformatic predictions and were thus not included. The prevalence of the
248 variants by population is summarized in Figure 7a and Supplementary Table 5. The prevalence of the
249 *pvmdr1* Y976F variant, the most widely characterized candidate of CQR, ranged from 0% in Afghanistan
250 to 100% in Sumatra, Indonesia¹⁹. The F1076L variant, which has also been implicated in CQR, exceeded
251 90% frequency in all 4 countries. The prevalence of A383G mutation in *pvdhps*, a marker of antifolate
252 resistance, varied highly, ranging from 2% in Afghanistan to 80% in Vietnam. No *pvdhps* A553G
253 mutations were observed in any country.

254

255 We also explored the prevalence of the *pvmdr1* Y976F variant in the Ethiopian RCT cases
256 (Supplementary Data 6, Figure 7b). Amongst 141 day 0 samples with successful microhaplotype and
257 *pvmdr1* 976 genotyping, Y976F prevalence was 52% (73/141). The prevalence of Y976F mutants
258 amongst recurrences occurring before day 63 in the CQ only arm with IBD ≥ 0.95 relative to the prior
259 infection (likely CQ recrudescence) was not significantly different from the day 0 samples, but sample
260 size was small (7/12, 58%, chi-square, $p=0.892$).

261

262 **Discussion**

263

264 We have established a highly multiplexed rhAmpSeq assay (up to 98 amplicons) to address critical
265 knowledge gaps in our understanding of antimalarial efficacy and transmission dynamics of *P. vivax*. Our
266 assay can be applied to low-density infections and to NGS platforms that are generally available in
267 reference laboratories in malaria-endemic settings.

268

269 A critical requirement of the assay was its ability to be implemented in surveillance frameworks in vivax-
270 endemic countries. This necessitated high sensitivity to genotype low-density *P. vivax* infections at
271 affordable cost using locally accessible sequencing platforms. With full chemistry, our analyses of serial
272 dilutions demonstrated yields of >100 mean read depth when parasite density was >70 per microliter
273 whole blood. Defining a meaningful assay sensitivity with amplicon sequencing approaches is
274 challenging as several factors other than parasite density can impact on read depth, including the yield
275 of the sequencing platform used, the level of sample multiplexing, the relative target biomass of other
276 samples on the same run, and whether a pre-amplification step is applied. To stress-test our assay, we
277 evaluated the sensitivity of target detection under conditions at the lower end of potential sequence
278 yield at the benefit of lower processing cost. Sensitivity evaluations were undertaken on a MiSeq
279 platform, which is the most widely available platform in malaria endemic settings. The paired end read
280 yield per run is approximately 20-30 million for the MiSeq compared to 260-800 million and up to 40
281 billion on the NextSeq and NovaSeq platforms, opening the possibility for even greater yield if the
282 setting permits. Furthermore, samples could be pooled and did not require pre-amplification with
283 selective Whole Genome Amplification (sWGA) or Primer Extension Preamplification (PEP)^{20,21}. Our
284 rationale was that the rhAmpSeq RNase H2 enzyme-dependent amplicon sequencing chemistry should
285 reduce the need for pre-amplification steps relative to standard AmpSeq as it minimizes spurious
286 primer-primer interactions and off-target amplification²².

287
288 Another essential data analysis requirement of our assay was the ability to generate genetic data that
289 effectively capture pairwise IBD in *P. vivax* infections with the aim of defining the origin of recurrent
290 infections. Previously described *P. vivax* amplicon sequencing assays, such as a widely used 42-SNP
291 barcode, have not comprised the density of SNPs needed to capture IBD accurately^{8,23,24}. Previously
292 using *in-silico* methods, we demonstrated that panels of ~100 *P. vivax* microhaplotypes can capture
293 pairwise IBD states ranging from 0 (unrelated) to 1 (clonally identical) with low error (RMSE under 0.12)
294⁸. In our current study, we applied the same *in-silico* models to demonstrate that, despite less optimal
295 spacing and slightly fewer markers than in the *in-silico* panels, our assayable set of microhaplotypes has
296 comparably low error in IBD estimation at simulated states of 0, 0.25, 0.5, 0.75 and 1 (RMSE under 0.12
297 with 92 microhaplotypes). We also note the adaptability of the rhAmpSeq protocol, enabling addition of
298 new markers, which could be selected to cover genomic regions with low marker density. The software
299 developed in our previous study supports microhaplotype selection that can be customized to specific
300 populations or for specific marker traits⁸.

301
302 In addition to the *in-silico* analysis, we explored microhaplotype-based IBD patterns between pre and
303 post treatment peripheral blood isolates collected from a clinical trial conducted in Ethiopia¹⁶. Our
304 results demonstrated a higher proportion of highly related but non identical (IBD 0.25-0.95) paired
305 isolates in patients not treated with primaquine, which we postulate reflects a greater risk of relapses in
306 the absence of radical cure. We also observed a higher frequency of clonally identical infection pairs
307 (IBD ≥ 0.95) in patients not treated with primaquine, likely reflecting a combination of relapses and
308 recrudescence events, that can be further dissected using time-to-event analyses²⁵. Lastly, we observed
309 a greater frequency of paired infections with low relatedness (IBD < 0.25) after 4 months consistent with
310 an increasing risk of reinfections during prolonged follow-up after antimalarial treatment. These findings
311 concur with expectations and thus contribute proof of concept on the utility of genetics to inform on
312 recurrence. Data from known relapse events, i.e. from individuals who were relocated to non-endemic
313 areas, will be needed to confirm these findings. Further work is also needed to dissect the potential
314 contribution of the splenic reservoir to recurrent *P. vivax* infections. Lastly, the incorporation of genetic
315 data with time-to-event information using modeling approaches that can handle polyclonal infections,
316 such as the PV3Rs software will greatly inform recurrence classification
317 (<https://github.com/aimeertaylor/Pv3Rs>). The data have been made open access to support further
318 developments in this space.

319

320 The microhaplotype data produced in this study also provides evidence of the potential for our assay to
321 inform spatial patterns of diversity and connectivity. At a macro-epidemiological level, there was a
322 distinct separation of populations by national boundaries, that could facilitate detection of cross-border
323 importation of infection. Within countries, trends in within-host diversity and connectivity highlighted
324 different transmission dynamics in Sumatra, Indonesia, relative to other sites. Relatively low prevalence
325 of polyclonal infections and large clonal clusters supported low transmission and subsequent inbreeding
326 in Sumatra relative to the other sites. The observed patterns infer that Sumatra is approaching the pre-
327 elimination phase and may be receptive to targeted rather than broad scale intervention approaches.
328 Within countries, we also observed evidence of moderate to large local transmission networks, largely
329 confined to specific sites but with some connections across sites. As the data repositories grow, it should
330 be possible to better define the major transmission networks within local regions and their
331 epidemiological context to inform on major reservoirs of infection and routes of infection spread
332 between communities.

333

334 Our panel included a non-exhaustive selection of *P. vivax* drug resistance candidates. The markers and
335 mechanisms of resistance to the more widely used drugs such as chloroquine, artemisinin and partner
336 drugs are not well understood in *P. vivax*⁵. The inference on treatment failure that can be made on the
337 available markers, including those incorporated in our panel, is therefore limited from an NMCP
338 perspective. However, the adaptability of the rhAmpSeq protocol ensures that new drug resistance
339 markers can be readily added to the panel as they arise. Insights from the current selection of
340 candidates included notable variation between countries in the prevalence of the *pvm-dr1* Y976F variant,
341 which has been implicated as a minor modulator of chloroquine (CQ) resistance¹⁹. The Y976F variant
342 was present at 100% frequency in Sumatra, Indonesia. To our knowledge, this is the first report on
343 Y976F in Sumatra, and aligns with the high prevalence (100%) of this variant in Papua province,
344 Indonesia, a region that has historically harbored high-grade CQ resistant *P. vivax* infections^{19,26,27}.
345 Although not at fixation, high prevalence of the Y976F variant was also observed in Vietnam (73%) and
346 Ethiopia (41%), but not in Afghanistan (0%). Therapeutic efficacy surveys report low level (<10%) CQ
347 resistance in Ethiopia, Vietnam and Afghanistan, while 16-65% failures by day 28 have been reported in
348 Sumatra, Indonesia²⁸⁻³⁶. The connection between the Y976F variant and clinical CQ efficacy is also
349 unclear. Our exploration of the Y976F variant in recurrent cases from the CQ only arm of an RCT
350 conducted in Ethiopia in 2015 revealed a non-significant difference in prevalence (58%) relative to the

351 baseline study set (51%). However, the sample size was small, with only 12 recurrence pairs in the CQ
352 arm meeting the IBD and pre-day 63 requirement. Further exploration of phenotypic associations
353 utilizing IBD and time-to-event data are needed in endemic settings with evidence of higher grade CQ
354 resistance.

355
356 The panel also includes a previously described mitochondrial amplicon, comprising SNPs and indels that
357 support *Plasmodium* spp. determination¹⁷. We used the amplicon to successfully confirm *Plasmodium*
358 spp. in a selection of non-*P. vivax* controls. A caveat of the assay is the inability to distinguish *P. ovale*
359 *curtisi* from *P. ovale wallikeri*. However, this could be addressed with the addition of other
360 mitochondrial amplicons³⁷. Although most assays displayed high specificity to *P. vivax*, several primers
361 yielded amplicons of *P. knowlesi* DNA with high coverage (>0.9) and moderate to high read depth.
362 Although *P. knowlesi* is less common in Southeast Asia outside of Malaysian and Indonesian Borneo and
363 considering that *P. vivax* has been eliminated from Malaysia, the *Plasmodium* spp marker may still be
364 helpful to detect potential mixed *P. vivax* and *P. knowlesi* infections which could yield inaccurate
365 genotypes at the specified markers.

366
367 Library preparation costs were approximately \$AUD 46 per sample for our study but can be
368 approximately halved (to \$AUD 23) by using half-chemistry reaction volumes for PCR 1. We observed
369 similar sensitivity between the full- and half-chemistry reactions in our serial dilution experiments.

370
371 In summary, the tools generated in our study provide a major in-road to establishing high-throughput
372 genetic data on *P. vivax* at reasonable cost that can provide policy relevant information on transmission
373 dynamics, parasite reservoirs and the spread of infection across national and provincial borders.

374

375 **Methods**

376

377 ***Marker selection and assay design***

378 193 microhaplotype markers, 5 drug resistance candidates, and two mitochondrial *Plasmodium* species-
379 confirmation markers, were selected for assay design (Supplementary Data 1). The set of 193
380 microhaplotypes were derived from two 100-microhaplotype panels (referred to as the high-diversity
381 and the random microhaplotype panels respectively based on the SNP filtering methods) that we
382 previously demonstrated exhibit high accuracy in IBD determination⁸. The rationale for selecting two

383 panels was based on expectation that some markers in the preferred panel (the high-diversity panel)
384 might not be assayable. The drug resistance candidates were not chosen to reflect an exhaustive list of
385 resistance candidates but rather a selection of SNPs that have previously been associated with *ex vivo* or
386 clinical phenotypes¹⁸. The species-confirmation markers have been described in a previous malaria
387 amplicon panel¹⁷. Primers and multiplex pools were designed by Integrated DNA Technologies (IDT).
388 Primer specifications included non-mapping to the human (GRCh38.p14), *P. falciparum* (Pf3D7), *P.*
389 *malariae* (PmUG01), *P. ovale wallikeri/curtisi* (GH01), and *P. knowlesi* (PKNH) reference genomes. A
390 variant calling format (VCF) file for the open-access MalariaGEN Pv4.0 dataset, comprising 911,901 high-
391 quality variants at 1,895 global *P. vivax* samples, was also provided to confirm non-mapping to highly
392 variable regions of the *P. vivax* genome³⁸. A set of 148 markers that met the described primer
393 specifications, and compatible within a single-plex reaction, were taken forward for a pilot experimental
394 run, comprising 176 positive and 16 negative controls (Supplementary Data 1). From the pilot data (not
395 presented here), a set of 98 markers were selected for the final assay (Supplementary Data 1).
396 Requirements for the final panel included high specificity to *P. vivax* and approximate uniform
397 microhaplotype distribution across the genome (Supplementary Figure 1).

398

399 **Patient samples**

400 The sensitivity and specificity of the 98-plex assay was evaluated using *P. vivax*-infected serial blood
401 dilutions and unmodified *P. vivax*-infected blood samples, stored as dried blood spots or refrigerated
402 whole blood in EDTA-coated microtainer or vacutainer tubes. The dilution series were prepared by
403 mixing each of the three *P. vivax*-infected whole blood samples (microscopy-determined parasite
404 densities of 740 to 960 parasites/ul blood) with uninfected human whole blood at 10-fold serial dilutions
405 (0.7 to 0.96 parasites/ul blood). Whole blood samples from uninfected human (n=3), *P. falciparum* (n=2),
406 *P. malariae* (n=2), *P. ovale* (n=1), and *P. knowlesi* (n=3) were included as negative controls. The samples
407 were collected within the framework of previously described clinical trials and cross-sectional surveys,
408 as well as returning travelers presenting with malaria at the Royal Darwin Hospital, Australia, and
409 represented 8 *vivax*-endemic countries (Supplementary Data 2)³⁹⁻⁴². DNA extraction was undertaken
410 using Qiagen's QIAamp kits for respective dried blood spots or whole blood.

411

412 **Real-time PCR**

413 The cycle threshold (Ct) values of *P. vivax* DNA samples were assessed using an Applied Biosystems®
414 QuantStudio™ 6 Flex Real-Time PCR System (ThermoFisher Scientific) with a TaqMan real-time PCR-

415 based assay, as adapted from ⁴³. This assay detects *P. vivax* by targeting a conserved region of the
416 mitochondrial cytochrome oxidase 1 gene (*pvmtcox1*) ⁴³. Each qPCR reaction mixture consisted of 10 ul
417 TaqMan Universal Mix II master mix (ThermoFisher Scientific), 1.6 ul of both the forward and reverse *Pv-*
418 *mtcox1* primers (10 uM), 0.8 μL of the *pvmtcox1* probe (10 uM), 2 ul H₂O, and 4 ul gDNA. The
419 thermocycling conditions included an initial denaturation at 95°C for 10 min, followed by 45 cycles of
420 denaturation at 95°C for 15 sec, and annealing and elongation at 60°C for 1 min. Data analysis was
421 performed using QuantStudio RT-PCR software to derive the threshold cycle (Ct) values.

422

423 **Whole genome sequencing (WGS) data**

424 A combination of newly generated and pre-existing WGS data was used in the study (Supplementary
425 Data 7). The pre-existing WGS data were derived from the MalariaGEN Pv4.0 repository ³⁸. Briefly, the
426 data were generated using 100-150 bp paired-end sequencing on a HiSeq or MiSeq instrument
427 (Illumina). The resultant reads were analyzed with the biallelic SNP pipeline used in this study, which
428 incorporated read mapping against the *P. vivax* P01 reference using BWA-MEM2 and variant calling
429 using the Genome Analysis Toolkit v4.5 ^{44,45}. A total of 911,901 high-quality biallelic SNP loci were
430 derived. Positions with less than 5 reads were considered genotype failures, with a minimum of two
431 reference and two alternate alleles, and a minimum of 10% of minor allele read frequency required to
432 define a genotype as heterozygous. The new WGS data were generated using microscopy-positive *P.*
433 *vivax* infections collected within a therapeutic efficacy survey of chloroquine, conducted in Arbaminch,
434 Ethiopia in 2019. Briefly, DNA was extracted from 1-2 ml white blood cell-depleted (Plasmodipur-
435 filtered) whole blood samples using Qiagen's QIAamp blood midi kits. Library preparation and
436 sequencing was conducted on the gDNA at the Wellcome Sanger Institute using the Illumina HiSeq
437 platform, generating 150bp paired reads. The resultant reads were mapped against the PvP01 reference
438 genome and genotype calling was conducted using the 911,901 MalariaGEN Pv4.0 SNPs, following the
439 same methods. Only samples with more than 50% genotype calls were included in the analysis.

440

441 **rhAmpSeq library preparation and sequencing**

442 Library preparation of the 98-plex marker panel was conducted using IDT's rhAmpSeq methodology.
443 Details on the protocol are provided in Supplementary Note 1 and on the protocols.io website
444 (<https://www.protocols.io/file/rfffce597.pdf>). In brief, genomic DNA samples were subject to a single
445 multiplexed (all 98 targets amplified within one pool) primer amplification using pre-balanced,
446 customized, rhAmpSeq primers. The product was diluted and subject to a second PCR, incorporating

447 indexes and Illumina sequencing adaptors. The products from the second PCR step were pooled
448 (incorporating up to 384 samples), resulting in the final library. The library was bead purified, assessed
449 for quality (fragment size and quantity), and sequenced with 150 bp paired-end clusters on a MiSeq
450 instrument (Illumina).

451

452 ***Amplicon data read mapping and variant calling***

453 Options for variant calling from fastq files at both biallelic SNPs and multiallelic microhaplotypes were
454 incorporated into the data processing pipeline, which is available at
455 <https://github.com/vivaxgen/MicroHaps> and described in Supplementary Note 2. In short, for the
456 biallelic SNP pipeline, vcfs are generated using a pipeline that conducts read mapping via *bwa-mem2*,
457 and variant calling with GATK v4.5. Positions with less than 25 reads were considered genotype failures,
458 with a minimum of two reference and two alternate alleles, and a minimum of 10% of minor allele read
459 frequency required to define a genotype as heterozygous. For the microhaplotype-calling pipeline, we
460 adapted an existing *P. falciparum* microhaplotype pipeline which utilizes the *DADA2* software for
461 denoising⁴⁶. The microhaplotype pipeline includes adapter and primer removal with *cutadapt*, and
462 custom scripts to postprocess *DADA2* output.

463

464 ***Population genetic measures***

465 For the resulting microhaplotype data, within-host infection diversity was characterized by the effective
466 multiplicity of infection (eMOI) metric using *MOIRE* software⁴⁷. For the WGS data, within-host diversity
467 was characterized using the F_{WS} score, calculated from biallelic SNP loci⁴⁸. The R-based *paneljudge*
468 package was used to assess the relative mean square error (RMSE) of the microhaplotype markers at
469 IBD levels of 0 (unrelated), 0.25 (half-siblings), 0.5 (siblings), 0.75 and 1.0 (identical)
470 (<https://github.com/aimeertaylor/paneljudge>). Between-sample IBD was measured using *DCifer*⁴⁹.
471 Network plots were generated using the R-based *igraph* package (<https://github.com/igraph>). Identity-
472 by-state (IBS)-based measures of genetic distance were calculated using the R-based *ape* package,
473 treating the microhaplotypes as multiallelic variants, and presented as neighbour-joining trees and
474 principal coordinate (PCoA) plots⁵⁰.

475

476 ***Plasmodium speciation based on genetic distance matrix***

477 To confirm the *Plasmodium* spp., amplicons targeting a universal mitochondrial sequence region were
478 mapped to a synthetic genome. The synthetic genome was constructed using reference genomes of five

479 *Plasmodium* species: PvP01_MIT_v2:2904-3149 (*P. vivax*), PmUG01_MIT_v1:1115-1361 (*P. malariae*),
480 PKNH_MIT_v2:1632-1875 (*P. knowlesi*), Pf3D7_MIT_v3:1634-1877 (*P. falciparum*), and
481 PocGH01_MIT_v2:2894-3137 (*P. ovale*). The synthetic genome was generated through an alignment
482 containing the specified regions within the reference genomes; aligned using muscle⁵¹.
483 For each sample with more than 10 reads and >0.9 coverage in the mitochondrial sequence region, the
484 reads were mapped to the synthetic genome. The mapped reads were used to call a consensus
485 sequence using iVar⁵². The consensus sequence was then compared pairwise to each sequence within
486 the multi-sequence alignment, containing all the reference genomes. The species of the sample was
487 determined based on the highest pairwise alignment score.

488

489 ***Inclusion and ethics***

490 The authors in this study combine researchers from the malaria-endemic countries represented in the
491 study (Afghanistan, Bangladesh, Colombia, Ethiopia, Indonesia and Vietnam) and nonmalaria endemic
492 areas. The researchers from malaria-endemic countries were involved throughout the research process
493 to ensure that the assays are implementable and impactful in local communities affected by malaria. All
494 samples in this study were derived from blood samples obtained from patients positive for malaria and
495 collected with informed consent from the patient, or patient parent / legal guardian where individuals
496 were less than 18 years of age. At each location, sample collection was approved by the appropriate
497 local institutional ethics committees. The following committees gave ethical approval for the partner
498 studies: Human Research Ethics Committee of NT Department of Health and Families and Menzies
499 School of Health Research, Darwin, Australia; Islamic Republic of Afghanistan Ministry of Public Health
500 Institutional Review Board, Afghanistan; ICDDR,B Ethical Review Committee, Bhutan; Research Ethics
501 Board of Health, at the Ministry of Health in Bhutan; Comit e Instiucional de Etica de Investigaciones en
502 Humanos, Colombia; Comit e de Bioetica Instituto de Investigaciones Medicas Facultad de Medicina
503 Universidad de Antioquia, Colombia; Armauer Hansen Research Institute Institutional Review Board,
504 Ethiopia; Addis Ababa University College of Natural Sciences, Ethiopia; Addis Ababa University, Aklilu
505 Lemma Institute of Pathobiology Institutional Review Board, Ethiopia; National Research Ethics Review
506 Committee of Ethiopia; Eijkman Institute Research Ethics Committee, Jakarta, Indonesia; Research
507 Review Committee of the Institute for Medical Research and the Medical Research Ethics Committee
508 (MREC), Ministry of Health, Malaysia; Scientific and Ethical Committee of the Hospital for Tropical
509 Diseases in Ho Chi Minh City, Vietnam; The Ministry of Health Evaluation Committee on Ethics in
510 Biomedical Research, Vietnam.

511

512 **Data availability**

513 The MalariaGEN Pv4.0 WGS data are available through the European Nucleotide Archive³⁸. The WGS
514 reads from the newly generated, high-quality, Ethiopian *P. vivax* infections are available through the
515 European Nucleotide Archive. The ENA accession numbers for all WGS samples used here are listed in
516 Supplementary Data 7. All high-quality (pass) microhaplotype-based parasite reads will be made
517 available in the European Nucleotide Archive at publication (Supplementary Data 2).

518

519 **Code availability**

520 The pipelines for retrieving SNP data in variant call format (vcf), as well as microhaplotype data in
521 Compact Idiosyncratic Gapped Alignment Report (cigar) string format, are available on GitHub at
522 <https://github.com/vivaxgen/MicroHaps>.

523

524 **References**

525

- 526 1 WHO. World Malaria Report 2022. World Health Organization; Geneva 2022. (2022).
- 527 2 Auburn, S., Cheng, Q., Marfurt, J. & Price, R. N. The changing epidemiology of *Plasmodium vivax*:
528 Insights from conventional and novel surveillance tools. *PLoS medicine* **18**, e1003560,
529 doi:10.1371/journal.pmed.1003560 (2021).
- 530 3 Price, R. N. *et al.* Global extent of chloroquine-resistant *Plasmodium vivax*: a systematic review
531 and meta-analysis. *The Lancet. Infectious diseases* **14**, 982-991, doi:10.1016/S1473-
532 3099(14)70855-2 (2014).
- 533 4 Commons, R. J. *et al.* The Vivax Surveyor: Online mapping database for *Plasmodium vivax* clinical
534 trials. *International journal for parasitology. Drugs and drug resistance* **7**, 181-190,
535 doi:10.1016/j.ijpddr.2017.03.003 (2017).
- 536 5 Buyon, L. E., Elsworth, B. & Duraisingh, M. T. The molecular basis of antimalarial drug resistance
537 in *Plasmodium vivax*. *International journal for parasitology. Drugs and drug resistance* **16**, 23-37,
538 doi:10.1016/j.ijpddr.2021.04.002 (2021).
- 539 6 Popovici, J. *et al.* Recrudescence, Reinfection, or Relapse? A More Rigorous Framework to Assess
540 Chloroquine Efficacy for *Plasmodium vivax* Malaria. *The Journal of infectious diseases* **219**, 315-
541 322, doi:10.1093/infdis/jiy484 (2019).
- 542 7 White, N. J. & Imwong, M. Relapse. *Advances in parasitology* **80**, 113-150, doi:10.1016/B978-0-
543 12-397900-1.00002-5 (2012).
- 544 8 Siegel, S. V. *et al.* Lineage-informative microhaplotypes for recurrence classification and spatio-
545 temporal surveillance of *Plasmodium vivax* malaria parasites. *Nature communications* **15**, 6757,
546 doi:10.1038/s41467-024-51015-3 (2024).
- 547 9 Snounou, G. & Beck, H. P. The use of PCR genotyping in the assessment of recrudescence or
548 reinfection after antimalarial drug treatment. *Parasitol Today* **14**, 462-467, doi:10.1016/s0169-
549 4758(98)01340-4 (1998).

- 550 10 Imwong, M. *et al.* Relapses of Plasmodium vivax infection usually result from activation of
551 heterologous hypnozoites. *The Journal of infectious diseases* **195**, 927-933, doi:10.1086/512241
552 (2007).
- 553 11 Chen, N., Auliff, A., Rieckmann, K., Gatton, M. & Cheng, Q. Relapses of Plasmodium vivax
554 infection result from clonal hypnozoites activated at predetermined intervals. *The Journal of*
555 *infectious diseases* **195**, 934-941, doi:10.1086/512242 (2007).
- 556 12 Abdullah, N. R. *et al.* Plasmodium vivax population structure and transmission dynamics in
557 Sabah Malaysia. *PLoS one* **8**, e82553, doi:10.1371/journal.pone.0082553 (2013).
- 558 13 Bright, A. T. *et al.* A high resolution case study of a patient with recurrent Plasmodium vivax
559 infections shows that relapses were caused by meiotic siblings. *PLoS neglected tropical diseases*
560 **8**, e2882, doi:10.1371/journal.pntd.0002882 (2014).
- 561 14 Popovici, J. *et al.* Genomic Analyses Reveal the Common Occurrence and Complexity of
562 Plasmodium vivax Relapses in Cambodia. *mBio* **9**, doi:10.1128/mBio.01888-17 (2018).
- 563 15 Ruybal-Pesantez, S. *et al.* Molecular markers for malaria genetic epidemiology: progress and
564 pitfalls. *Trends in parasitology* **40**, 147-163, doi:10.1016/j.pt.2023.11.006 (2024).
- 565 16 Abreha, T. *et al.* Comparison of artemether-lumefantrine and chloroquine with and without
566 primaquine for the treatment of Plasmodium vivax infection in Ethiopia: A randomized
567 controlled trial. *PLoS medicine* **14**, e1002299, doi:10.1371/journal.pmed.1002299 (2017).
- 568 17 Jacob, C. G. *et al.* Genetic surveillance in the Greater Mekong subregion and South Asia to
569 support malaria control and elimination. *eLife* **10**, doi:10.7554/eLife.62997 (2021).
- 570 18 Price, R. N., Auburn, S., Marfurt, J. & Cheng, Q. Phenotypic and genotypic characterisation of
571 drug-resistant Plasmodium vivax. *Trends in parasitology* **28**, 522-529,
572 doi:10.1016/j.pt.2012.08.005 (2012).
- 573 19 Suwanarusk, R. *et al.* Chloroquine resistant Plasmodium vivax: in vitro characterisation and
574 association with molecular polymorphisms. *PLoS one* **2**, e1089,
575 doi:10.1371/journal.pone.0001089 (2007).
- 576 20 Cowell, A. N. *et al.* Selective Whole-Genome Amplification Is a Robust Method That Enables
577 Scalable Whole-Genome Sequencing of Plasmodium vivax from Unprocessed Clinical Samples.
578 *mBio* **8**, doi:10.1128/mBio.02257-16 (2017).
- 579 21 Arneson, N., Hughes, S., Houlston, R. & Done, S. Whole-Genome Amplification by Improved
580 Primer Extension Pre-amplification PCR (I-PEP-PCR). *CSH Protoc* **2008**, pdb prot4921,
581 doi:10.1101/pdb.prot4921 (2008).
- 582 22 Dobosy, J. R. *et al.* RNase H-dependent PCR (rhPCR): improved specificity and single nucleotide
583 polymorphism detection using blocked cleavable primers. *BMC Biotechnol* **11**, 80,
584 doi:10.1186/1472-6750-11-80 (2011).
- 585 23 Baniecki, M. L. *et al.* Development of a single nucleotide polymorphism barcode to genotype
586 Plasmodium vivax infections. *PLoS neglected tropical diseases* **9**, e0003539,
587 doi:10.1371/journal.pntd.0003539 (2015).
- 588 24 Taylor, A. R., Jacob, P. E., Neafsey, D. E. & Buckee, C. O. Estimating Relatedness Between Malaria
589 Parasites. *Genetics* **212**, 1337-1351, doi:10.1534/genetics.119.302120 (2019).
- 590 25 Taylor, A. R. *et al.* Resolving the cause of recurrent Plasmodium vivax malaria probabilistically.
591 *Nature communications* **10**, 5595, doi:10.1038/s41467-019-13412-x (2019).
- 592 26 Ratcliff, A. *et al.* Therapeutic response of multidrug-resistant Plasmodium falciparum and P.
593 vivax to chloroquine and sulfadoxine-pyrimethamine in southern Papua, Indonesia. *Transactions*
594 *of the Royal Society of Tropical Medicine and Hygiene* **101**, 351-359,
595 doi:10.1016/j.trstmh.2006.06.008 (2007).
- 596 27 Pearson, R. D. *et al.* Genomic analysis of local variation and recent evolution in Plasmodium
597 vivax. *Nature genetics* **48**, 959-964, doi:10.1038/ng.3599 (2016).

- 598 28 Assefa, M., Eshetu, T. & Biruksew, A. Therapeutic efficacy of chloroquine for the treatment of
599 Plasmodium vivax malaria among outpatients at Hossana Health Care Centre, southern Ethiopia.
600 *Malaria journal* **14**, 458, doi:10.1186/s12936-015-0983-x (2015).
- 601 29 Getachew, S. *et al.* Chloroquine efficacy for Plasmodium vivax malaria treatment in southern
602 Ethiopia. *Malaria journal* **14**, 525, doi:10.1186/s12936-015-1041-4 (2015).
- 603 30 Thuan, P. D. *et al.* A Randomized Comparison of Chloroquine Versus Dihydroartemisinin-
604 Piperaquine for the Treatment of Plasmodium vivax Infection in Vietnam. *The American journal*
605 *of tropical medicine and hygiene* **94**, 879-885, doi:10.4269/ajtmh.15-0740 (2016).
- 606 31 Phong, N. C. *et al.* Susceptibility of Plasmodium falciparum to artemisinins and Plasmodium vivax
607 to chloroquine in Phuoc Chien Commune, Ninh Thuan Province, south-central Vietnam. *Malaria*
608 *journal* **18**, 10, doi:10.1186/s12936-019-2640-2 (2019).
- 609 32 Kolaczinski, K., Durrani, N., Rahim, S. & Rowland, M. Sulfadoxine-pyrimethamine plus artesunate
610 compared with chloroquine for the treatment of vivax malaria in areas co-endemic for
611 Plasmodium falciparum and P. vivax: a randomised non-inferiority trial in eastern Afghanistan.
612 *Transactions of the Royal Society of Tropical Medicine and Hygiene* **101**, 1081-1087,
613 doi:10.1016/j.trstmh.2007.06.015 (2007).
- 614 33 Leslie, T. *et al.* Sulfadoxine-pyrimethamine, chlorproguanil-dapsone, or chloroquine for the
615 treatment of Plasmodium vivax malaria in Afghanistan and Pakistan: a randomized controlled
616 trial. *JAMA* **297**, 2201-2209, doi:10.1001/jama.297.20.2201 (2007).
- 617 34 Awab, G. R. *et al.* Dihydroartemisinin-piperaquine versus chloroquine to treat vivax malaria in
618 Afghanistan: an open randomized, non-inferiority, trial. *Malaria journal* **9**, 105,
619 doi:10.1186/1475-2875-9-105 (2010).
- 620 35 Sutanto, I. *et al.* Evaluation of chloroquine therapy for vivax and falciparum malaria in southern
621 Sumatra, western Indonesia. *Malaria journal* **9**, 52, doi:10.1186/1475-2875-9-52 (2010).
- 622 36 Fryauff, D. J. *et al.* The drug sensitivity and transmission dynamics of human malaria on Nias
623 Island, North Sumatra, Indonesia. *Ann Trop Med Parasitol* **96**, 447-462,
624 doi:10.1179/000349802125001249 (2002).
- 625 37 Higgins, M. *et al.* New reference genomes to distinguish the sympatric malaria parasites,
626 Plasmodium ovale curtisi and Plasmodium ovale wallikeri. *Scientific reports* **14**, 3843,
627 doi:10.1038/s41598-024-54382-5 (2024).
- 628 38 MalariaGen *et al.* An open dataset of Plasmodium vivax genome variation in 1,895 worldwide
629 samples. *Wellcome open research* **7**, 136, doi:10.12688/wellcomeopenres.17795.1 (2022).
- 630 39 Abreha, T. *et al.* Correction: Comparison of artemether-lumefantrine and chloroquine with and
631 without primaquine for the treatment of Plasmodium vivax infection in Ethiopia: A randomized
632 controlled trial. *PLoS medicine* **15**, e1002677, doi:10.1371/journal.pmed.1002677 (2018).
- 633 40 Taylor, W. R. J. *et al.* Short-course primaquine for the radical cure of Plasmodium vivax malaria:
634 a multicentre, randomised, placebo-controlled non-inferiority trial. *Lancet* **394**, 929-938,
635 doi:10.1016/S0140-6736(19)31285-1 (2019).
- 636 41 Grigg, M. J. *et al.* Artesunate-mefloquine versus chloroquine for treatment of uncomplicated
637 Plasmodium knowlesi malaria in Malaysia (ACT KNOW): an open-label, randomised controlled
638 trial. *The Lancet. Infectious diseases* **16**, 180-188, doi:10.1016/S1473-3099(15)00415-6 (2016).
- 639 42 Ley, B. *et al.* G6PD Deficiency and Antimalarial Efficacy for Uncomplicated Malaria in
640 Bangladesh: A Prospective Observational Study. *PLoS one* **11**, e0154015,
641 doi:10.1371/journal.pone.0154015 (2016).
- 642 43 Gruenberg, M. *et al.* Plasmodium vivax molecular diagnostics in community surveys: pitfalls and
643 solutions. *Malaria journal* **17**, 55, doi:10.1186/s12936-018-2201-0 (2018).

- 644 44 Auburn, S. *et al.* A new *Plasmodium vivax* reference sequence with improved assembly of the
645 subtelomeres reveals an abundance of *pir* genes. *Wellcome open research* **1**, 4,
646 doi:10.12688/wellcomeopenres.9876.1 (2016).
- 647 45 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform.
648 *Bioinformatics* **25**, 1754-1760, doi:10.1093/bioinformatics/btp324 (2009).
- 649 46 LaVerriere, E. *et al.* Design and implementation of multiplexed amplicon sequencing panels to
650 serve genomic epidemiology of infectious disease: A malaria case study. *Mol Ecol Resour* **22**,
651 2285-2303, doi:10.1111/1755-0998.13622 (2022).
- 652 47 Murphy, M. & Greenhouse, B. MOIRE: A software package for the estimation of allele
653 frequencies and effective multiplicity of infection from polyallelic data. *bioRxiv*,
654 doi:10.1101/2023.10.03.560769 (2024).
- 655 48 Auburn, S. *et al.* Characterization of within-host *Plasmodium falciparum* diversity using next-
656 generation sequence data. *PloS one* **7**, e32891, doi:10.1371/journal.pone.0032891 (2012).
- 657 49 Gerlovina, I., Gerlovin, B., Rodriguez-Barraquer, I. & Greenhouse, B. Dcifer: an IBD-based
658 method to calculate genetic distance between polyclonal infections. *Genetics* **222**,
659 doi:10.1093/genetics/iyac126 (2022).
- 660 50 Paradis, E., Claude, J. & Strimmer, K. APE: Analyses of Phylogenetics and Evolution in R language.
661 *Bioinformatics* **20**, 289-290 (2004).
- 662 51 Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput.
663 *Nucleic acids research* **32**, 1792-1797, doi:10.1093/nar/gkh340 (2004).
- 664 52 Grubaugh, N. D. *et al.* An amplicon-based sequencing framework for accurately measuring
665 intrahost virus diversity using PrimalSeq and iVar. *Genome biology* **20**, 8, doi:10.1186/s13059-
666 018-1618-7 (2019).
- 667

668 **Acknowledgements**

669 This work was supported, in whole or in part, by the Bill & Melinda Gates Foundation INV-043618
670 (supporting S.A., D.N and R.N.P). Under the grant conditions of the Foundation, a Creative Commons
671 Attribution 4.0 Generic License has already been assigned to the Author Accepted Manuscript version
672 that might arise from this submission. The study was also supported by the National Health and Medical
673 Research Council of Australia (APP2001083 supporting S.A. and S.V.S.). The whole genome sequencing
674 component of the study was supported by the Medical Research Council and UK Department for
675 International Development (award number M006212 to D.K.) and the Wellcome Trust (award numbers
676 206194 and 204911 to D.K.). The OPRA clinical trial was supported by the Bill & Melinda
677 Gates Foundation (OPP1054404 awarded to R.N.P.). We thank the patients who contributed their
678 samples to the study, and the health workers and field teams who assisted with the sample collections.
679 Whole genome sequencing was undertaken by the Wellcome Sanger Institute, and amplicon sequencing
680 was undertaken at the Australian Genome Research Facility (AGRF) and Menzies School of Health
681 Research. We thank the staff of the Wellcome Sanger Institute and AGRF for contributions to samples
682 logistics, sequencing and informatics. We thank Ruchit Panjal for his assistance with the pipeline set up.

683

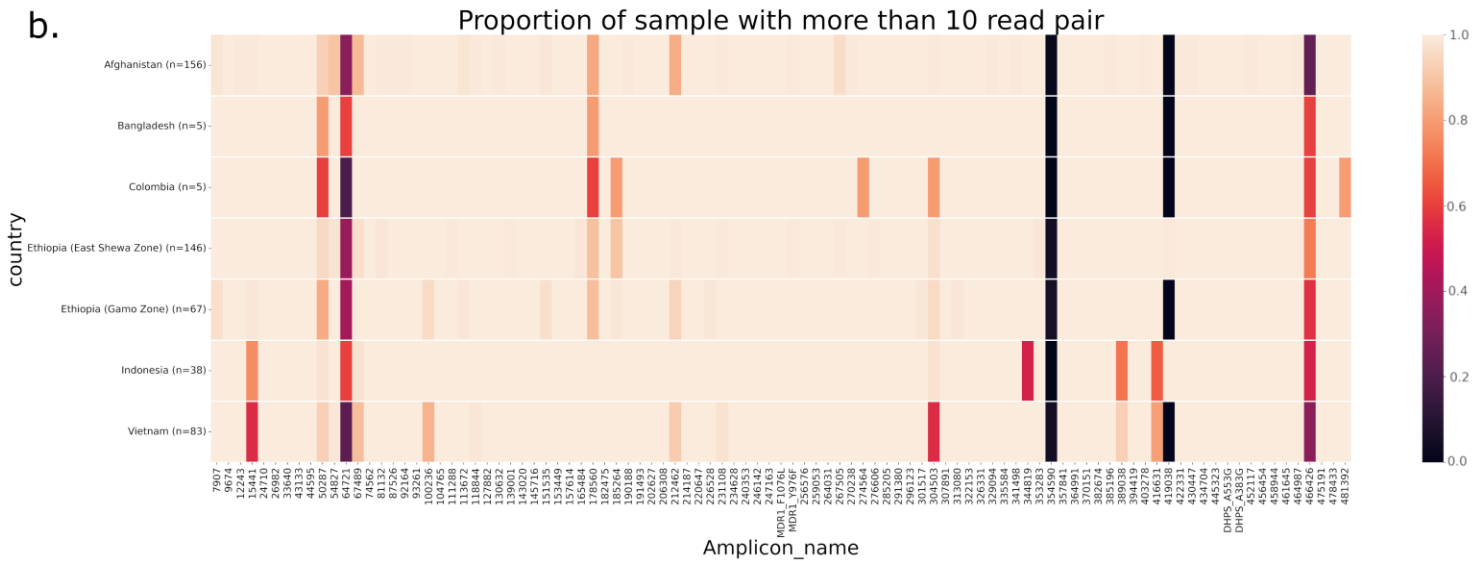
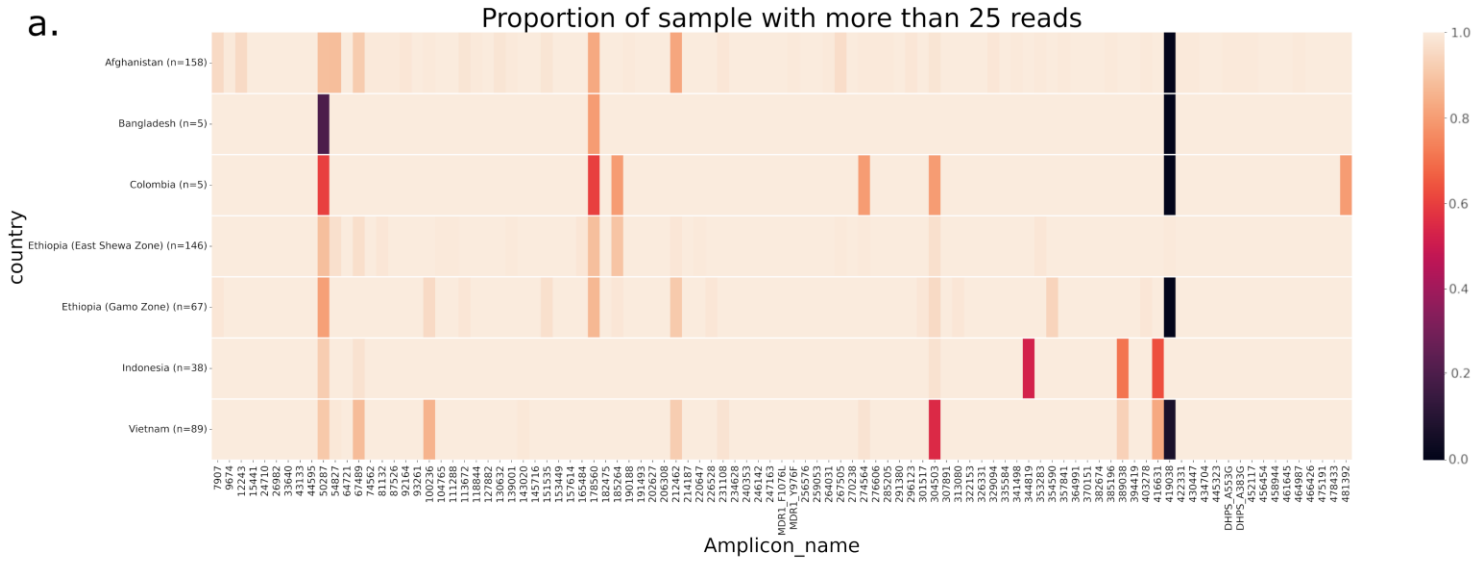
684 ***Author contributions***

685 S.A., D. N. and R.N.P. conceived the study. S.A., M.K., A Rumaseb, E.S., and H.T. designed the study. A.
686 Rumaseb, T.P., D.H., G.W., S.V.S, N.T.T.N. and S.A. contributed to the laboratory assay design. A
687 Rumaseb, T.P., D.H., A.O. and A Rai contributed to genotyping data generation. S.V.S, R.P., R.A. and
688 D.P.K. contributed whole genome sequencing data production and informatic support. M.K., H.T., A.O.,
689 P.M. and D.N. contributed to bioinformatic pipeline development. M.K., E.S., H.T., K.S.H. and S.A.
690 conducted data analysis. E.D.B., N.T.T.N, N.H.C., A.A., T.S.D., D.T.A., A.G.R., A.P.P., I.S., M.S.A., Z.P., T.L.,
691 D.E., T.W., N.A., M.G., N.D., N.J.W., D.P.K. and R.N. contributed essential field-based malaria collections
692 and metadata, and guidance on the study design and interpretation.

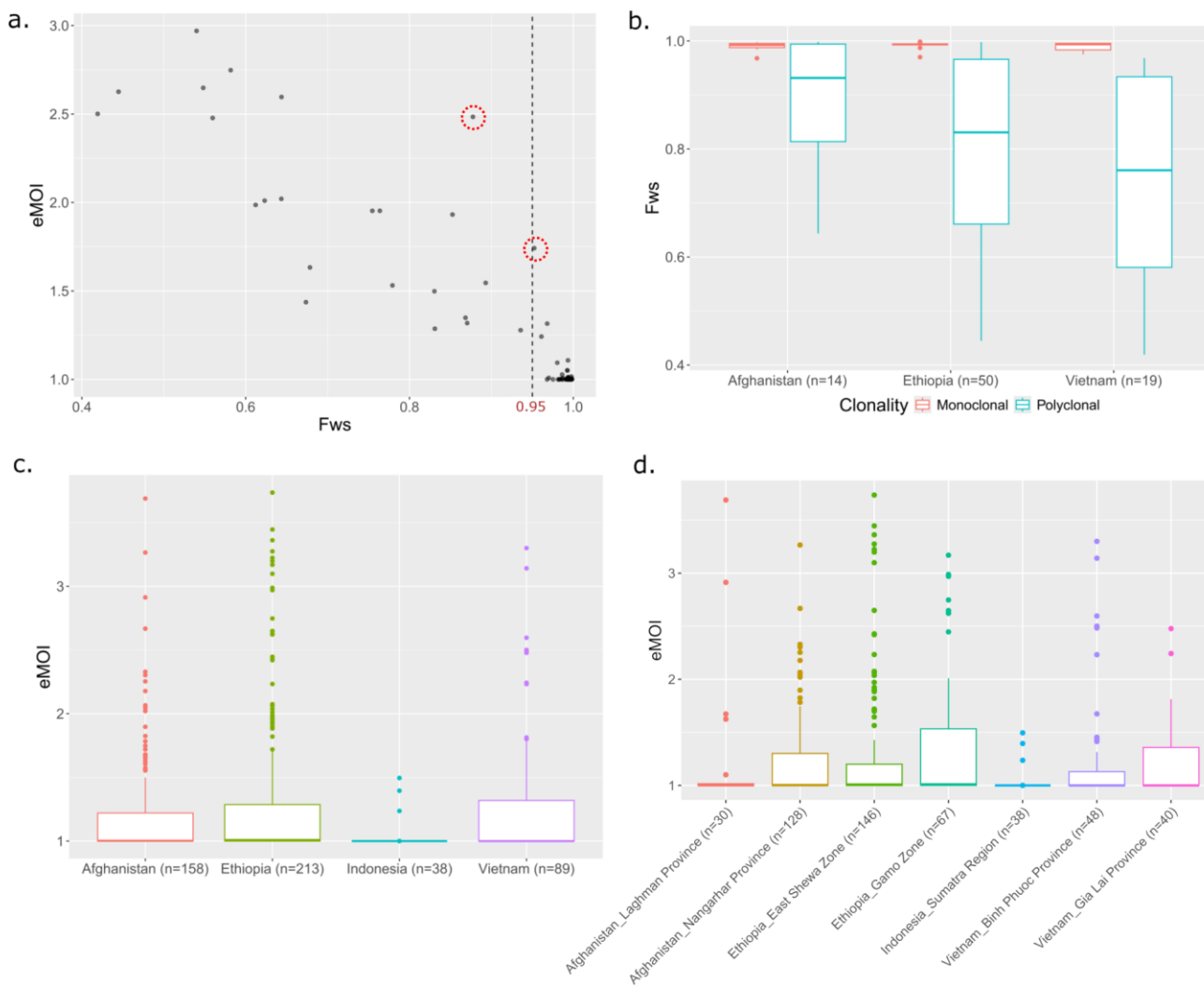
693

694 ***Competing Interests***

695 The authors declare no competing interests



699 **Figure 1. *P. vivax* proportion of samples with more than 25 reads or more than 10 read pairs by country.** Heat maps illustrate the proportion of
700 samples with more than 25 reads (panel a) and proportion of samples with more than 10 read pairs derived from *DADA2* (panel b) for each
701 marker by country. Markers (x-axis) were ordered by chromosome and coordinate. All samples with genotype failures (<25 reads or <10 read
702 pairs) are presented in black. Microhaplotype markers 64721, 354590, 419038 and 466426 displayed consistently low read-pair counts.
703 Microhaplotype marker 354590 exhibits high individual read counts (panel a) but low read-pair counts (panel b) in all countries and was
704 excluded from all downstream read pair analyses. Data are presented on a total of 500 independent (not including replicates) infections that
705 passed genotyping. The data from Ethiopia are split into dried blood spot (DBS) from a clinical trial conducted in East Shewa Zone and whole
706 blood (WB) extracts from a therapeutic efficacy survey conducted in Gamo Zone for comparative assessment. The DBS versus WB comparisons
707 reveal similar read and read pair counts at all loci.



708

709

710 **Figure 2. Microhaplotype-based within-host diversity trends.** Panels a) and b) illustrate the level of

711 concordance between genomic (as measured by the Fws) and microhaplotype (as measured by eMOI)

712 data in estimation of within-host *P. vivax* diversity. The boxplots present the median, interquartile range

713 and min and max value. Data are derived from 83 independent *P. vivax* cases. Overall, high concordance

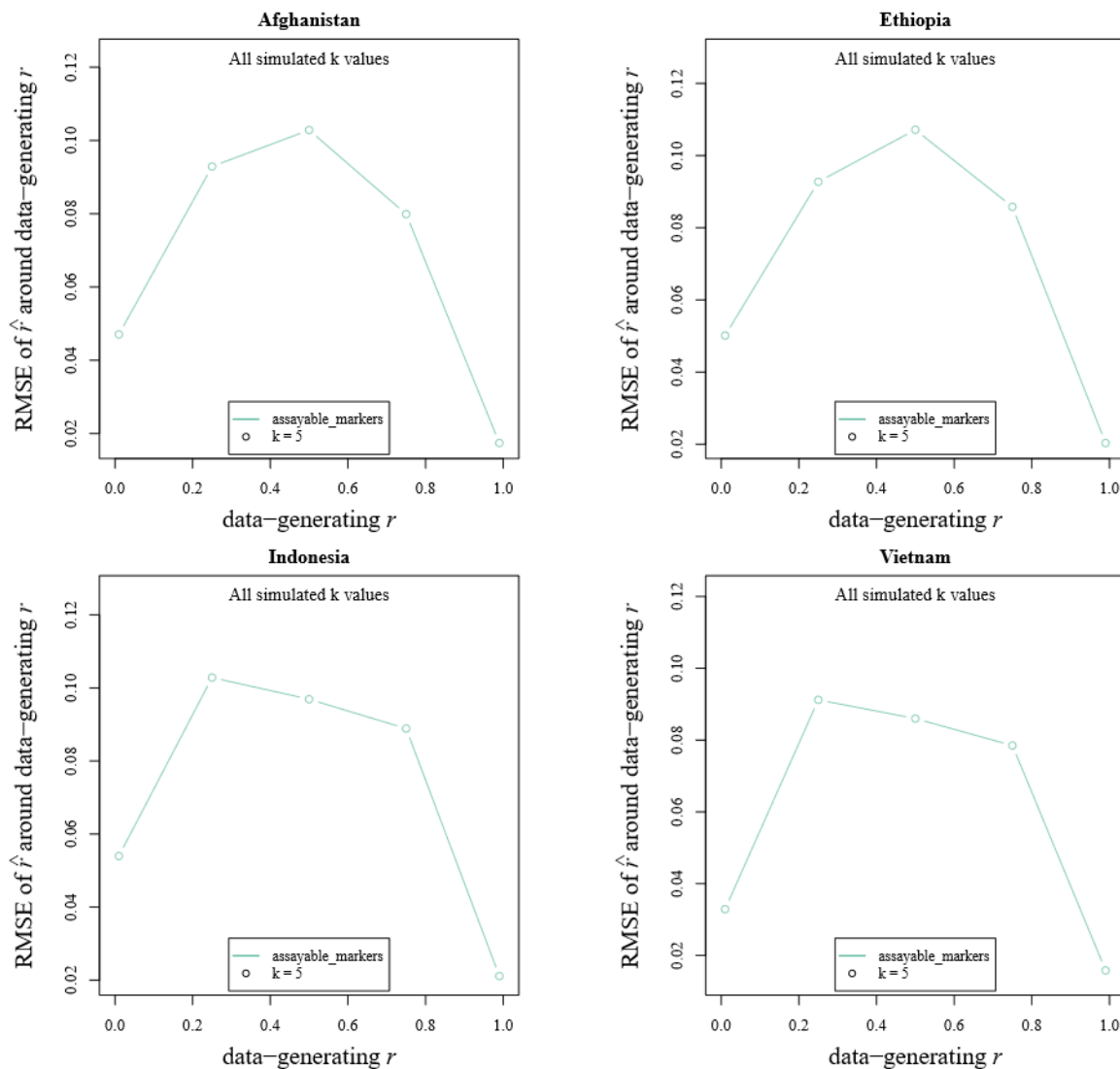
714 is observed between the two datasets. However, a few infections, including VN0001-321 and

715 VN001_030 circled in panel 4a) display notably higher within-host diversity with microhaplotype versus

716 whole genome sequence data. Panels c) and d) present the eMOI distributions at the country level (c)

717 and provincial level (d), illustrating variation between sites, likely reflecting local endemicity levels. The

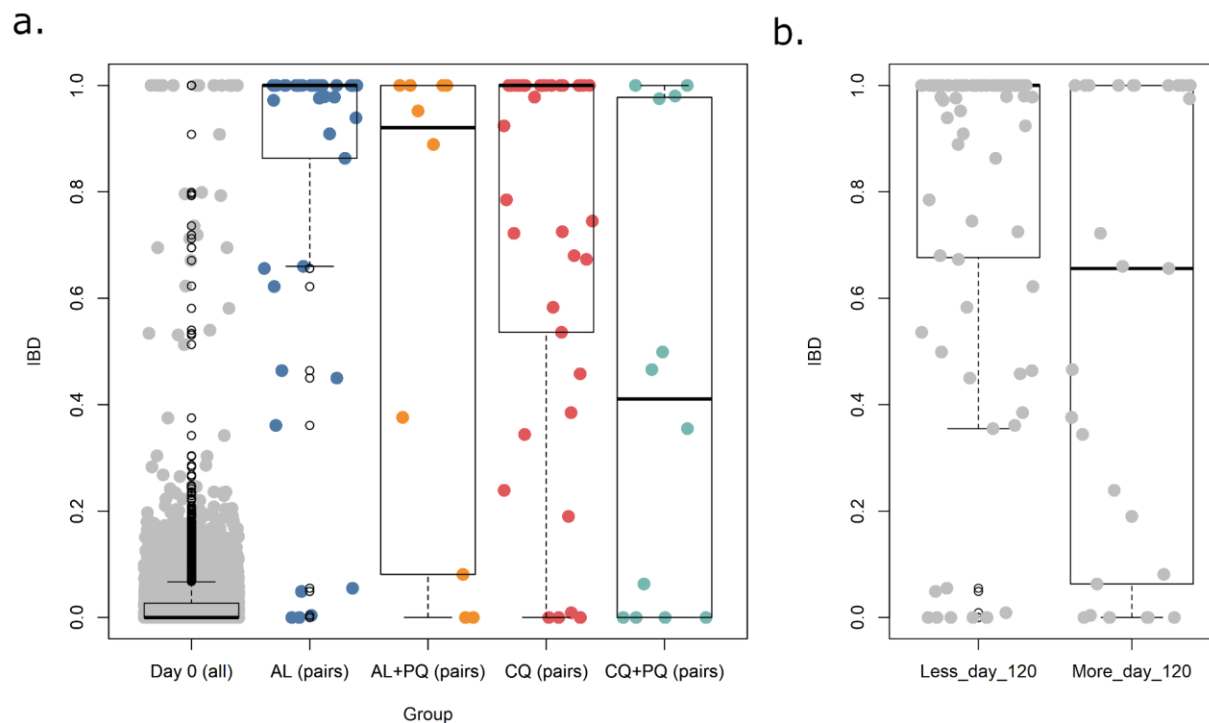
718 data in panels c) and d) are based on 498 independent cases.



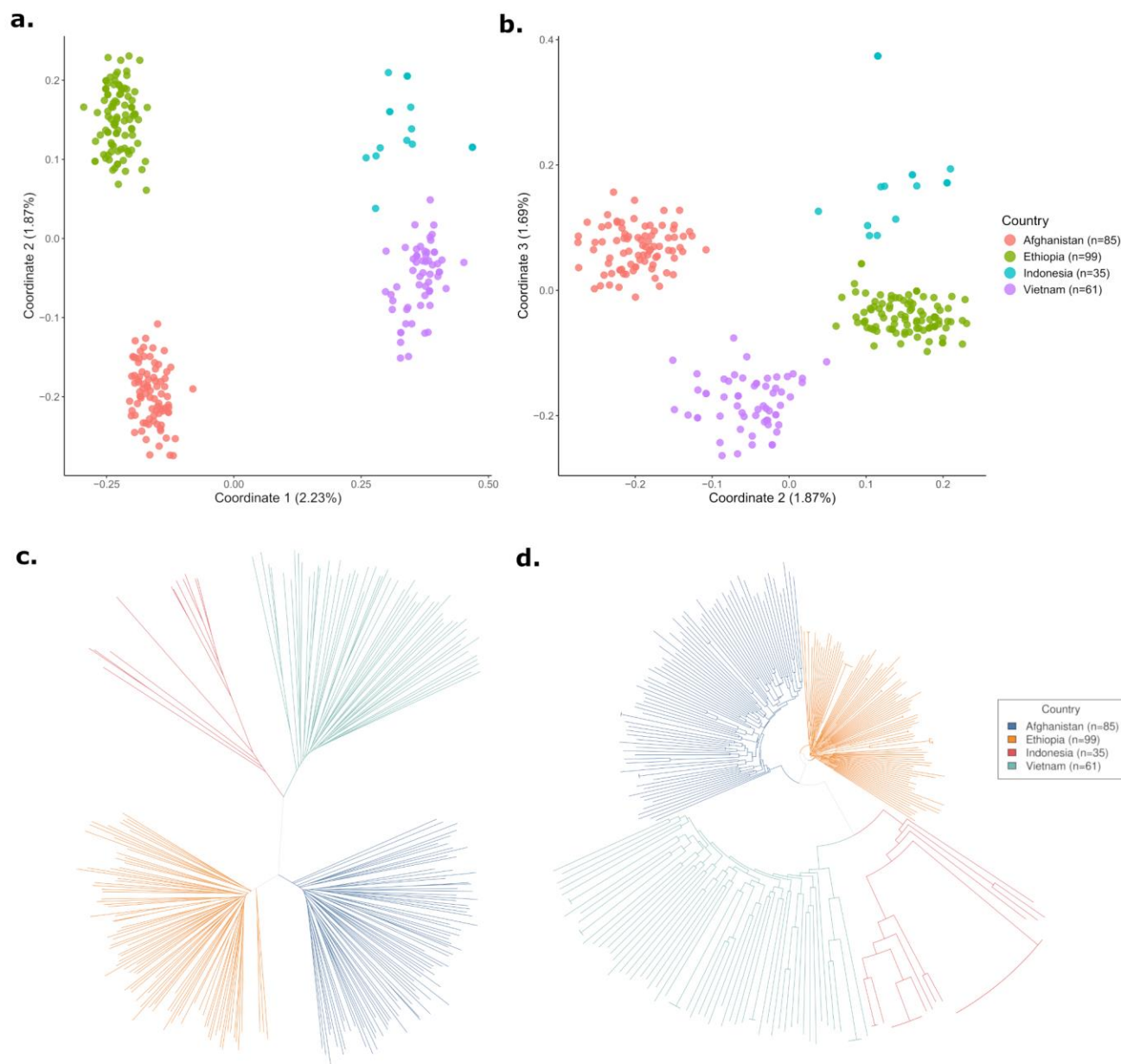
719
720
721
722
723
724
725
726
727
728
729

Figure 3. Simulations of IBD estimation in different geographic areas using the microhaplotype panel.

Root mean square error (RMSE) of relatedness estimates based on data simulated using five different data-generating relatedness estimates, r (specifically IBD of 0 [unrelated], 0.25 [half-sibling], 0.5 [sibling], 0.75 [highly related], and 1 [clonally identical]) with switch rate parameter k set to 5. Data were generated using *paneljudge* software on 91 high performance microhaplotypes and 4 drug resistance markers in independent infections from Afghanistan ($n=158$), Ethiopia ($n=213$), Indonesia ($n=38$) and Vietnam ($n=89$). In all populations, half-siblings and siblings had the highest RMSE, but this remained below 0.12 in all cases.



730
731 **Figure 4. IBD distributions by treatment and time to recurrence in a randomized controlled trial**
732 **conducted in Ethiopia.** Panel a) presents the IBD distributions in initial and recurrent infection pairs
733 across all pairs and grouped by treatment arm; AL (Artemether-Lumefantrine), CQ (Chloroquine), AL+PQ
734 (AL + Primaquine) and CQ+PQ. Panel b) presents the same IBD data grouped by recurrences occurring
735 less versus more than 120 days after the initial infection. Each boxplot presents the median,
736 interquartile range and min and max value. Data are presented on recurrence pairs from 108
737 independent patients, with only one infection pair presented per patient to avoid potential bias (n=108
738 patients, 216 samples). Majority (86%, 93/108) of pairs reflect day 0 and recurrence 1 time points.
739 Where the day 0 or recurrence 1 infections failed genotyping or had inconclusive clinical metadata,
740 consecutive pairs of recurrence 2 to 4 pairs were used instead as the patients received the same
741 treatment up to recurrence 4.
742



743

744

745 **Figure 5. Identity-by-state-based spatial patterns.** Panels a) and b) present Principal Coordinate (PCoA)

746 plots derived from distance matrices on the microhaplotype calls. High separation of countries is

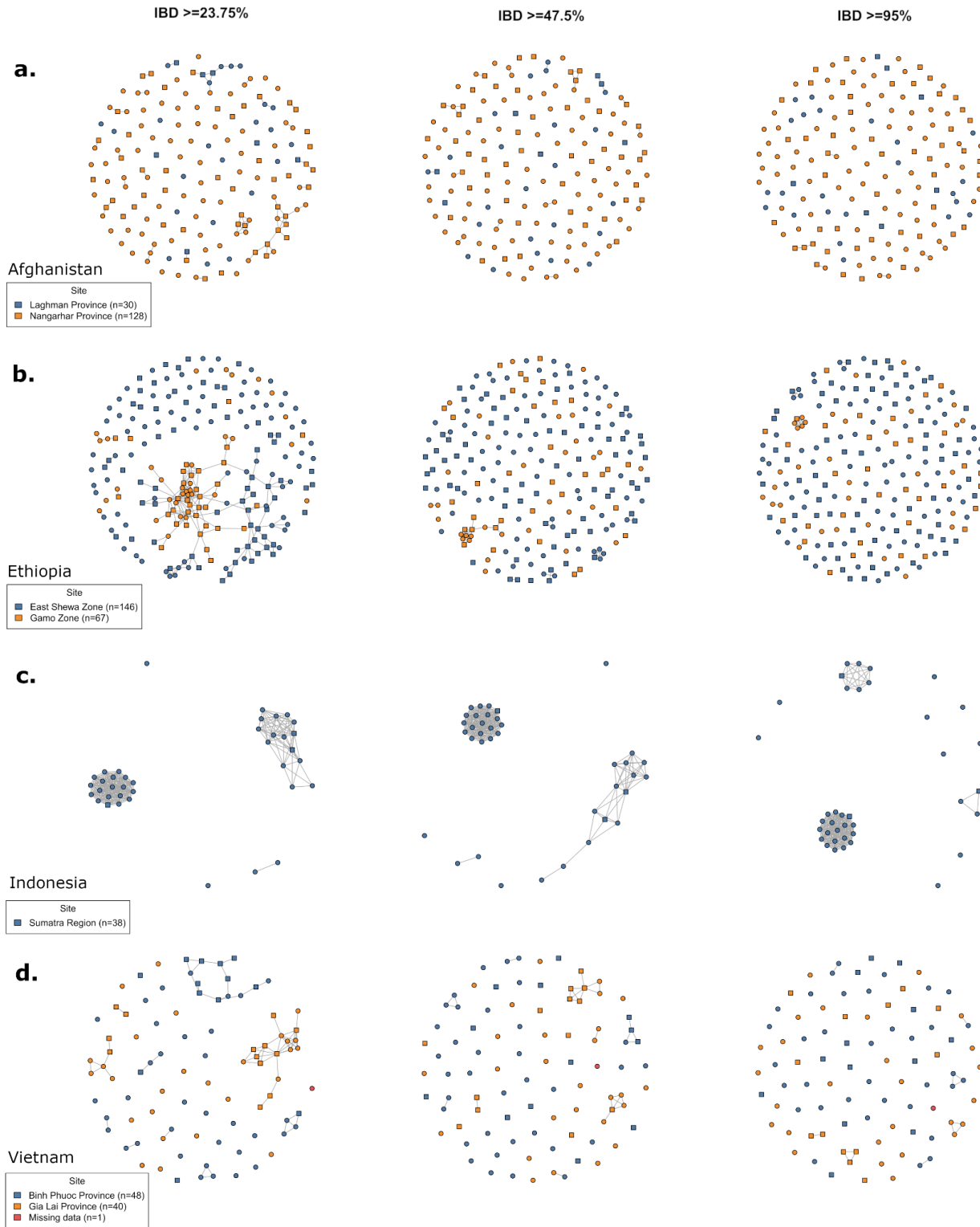
747 observed with principal coordinates 1 and 2, with even further separation of Indonesia from Vietnam on

748 principal coordinate 3. Panels c) and d) present unrooted and rooted, respectively, neighbour-joining

749 trees based on the same distance matrix as per the PCoA plots. The neighbour-joining trees further

750 illustrate distinct clustering by country and, additionally, highlight a clonally identical cluster of

751 infections in Sumatra, Indonesia. All plots were generated using data on 280 independent, monoclonal
752 infections.
753
754



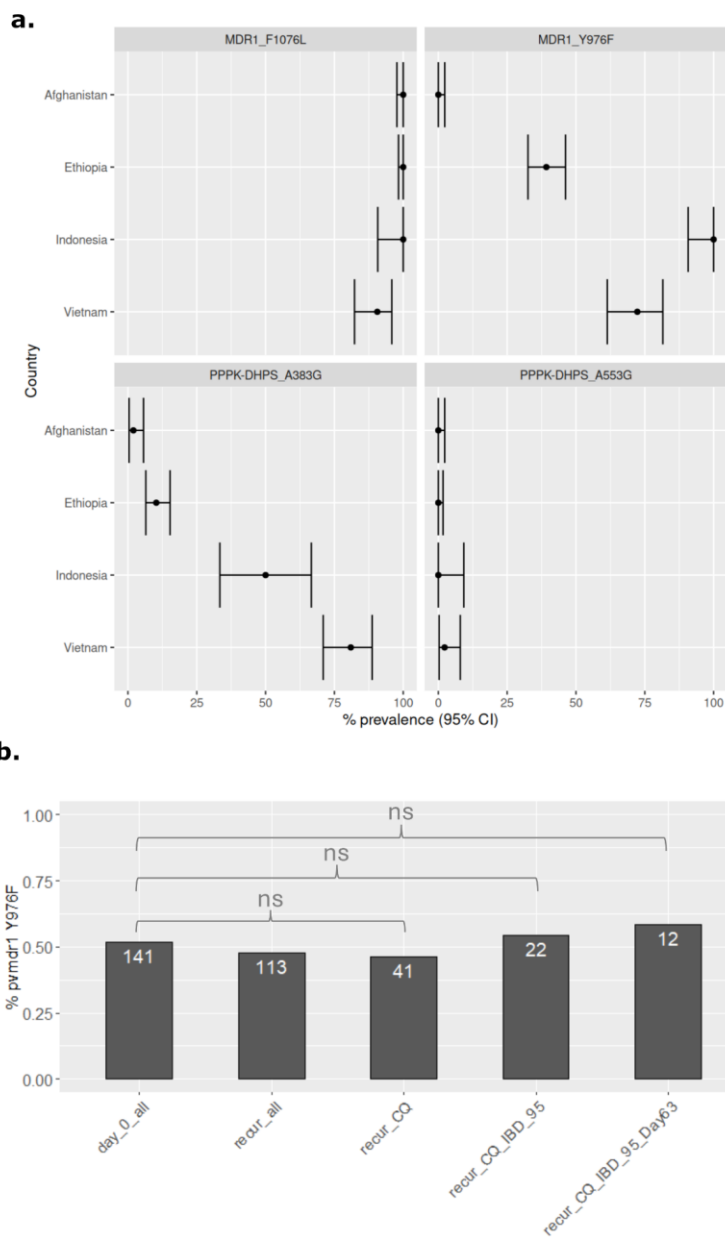
755

756

757 **Figure 6. IBD-based spatial patterns using microhaplotype data.** Panels a) to d) present networks

758 illustrating IBD-based connectivity between infections in Afghanistan (a), Ethiopia (b), Vietnam (c), and

759 Sumatra, Indonesia (d). Each shape reflects an infection, colour-coded by site, and with shapes reflecting
760 monoclonal (circle) versus polyclonal (square) infections. For each country, connectivity (illustrated by
761 connecting lines between shapes) is presented at IBD thresholds of minimum ~25%, 50% and 95% (left
762 to right). IBD measures were calculated on the microhaplotype calls using *DCifer* software. At the ~25%
763 IBD threshold, large networks (10 or more connected infections) are observed amongst infections from
764 the same site, with some connectivity between networks from different sites; these networks shrink
765 with increasing IBD in all sites except for Sumatra, Indonesia, where the networks appear to reflect
766 clonal clusters (retained at IBD $\geq 95\%$). All plots were generated using data on independent infections:
767 Afghanistan (n=158), Ethiopia (n=213), Vietnam (n=89) and Indonesia (n=38).



768
 769 **Figure 7. Amino acid frequencies at *P. vivax* drug resistance candidates.** Panel a) presents proportions
 770 and corresponding 95% confidence intervals (CIs) for the given amino acid changes in baseline
 771 population samples from each country. All frequencies reflect the suspected drug-resistance-conferring
 772 amino acid. All plots were generated using independent samples (n=498). Panel b) presents the
 773 proportions of the *pvmdr1* Y976F variant, which has been associated with CQ resistance, in the
 774 respective patient groups within the Ethiopian randomized control trial; all genotyped baseline (day 0)
 775 infections irrespective of treatment arm (day_0_all), all recurrent infections irrespective of treatment
 776 arm (recur_all), all recurrent infections in the CQ only arm (recur_CQ), all recurrent infections in the CQ

777 only arm with IBD>0.95 relative to the prior infection (enhancing probability of recrudescence based on
778 genetics; recur_CQ_IBD_095), and all recurrent infections in the CQ only arm occurring prior to day 63
779 and with IBD>0.95 relative to the prior infection (enhancing probability of recrudescence based on
780 genetic and time-to-event metadata, recur_CQ_IBD_095_Day63). Sample sizes are annotated in white
781 font over the respective bars. ns = non-significant difference in proportion between the indicated
782 groups.

