

1 **Genetic surveillance reveals low, sustained malaria transmission with clonal**  
2 **replacement in Sao Tome and Principe**

3

4 Ying - An Chen<sup>1,2</sup>, Peng-Yin Ng<sup>2</sup>, Daniel Garcia<sup>2,3</sup>, Aaron Elliot<sup>1</sup>, Brian Palmer<sup>1</sup>, Ronald  
5 Mendes Costa d' Assunção Carvalho<sup>4</sup>, Lien-Fen Tseng<sup>5</sup>, Cheng-Sheng Lee<sup>6</sup>, Kun-Hsien  
6 Tsai<sup>5,7</sup>, Bryan Greenhouse<sup>1</sup>, and Hsiao-Han Chang<sup>2\*</sup>

7

8 <sup>1</sup> EPPIcenter Research Program, Division of HIV, Infectious Diseases and Global Medicine,  
9 Department of Medicine, University of California, San Francisco, United States

10 <sup>2</sup> Institute of Bioinformatics and Structural Biology, College of Life Sciences and Medicine,  
11 National Tsing Hua University, Hsinchu, Taiwan

12 <sup>3</sup> Bioinformatics Program, Institute of Statistical Science, Taiwan International Graduate  
13 Program, Academia Sinica, Taipei, Taiwan

14 <sup>4</sup> Taiwanese Medical Mission, São Tomé, Democratic Republic of São Tomé and Príncipe

15 <sup>5</sup> Taiwan Anti-Malarial Advisory Mission, São Tomé, Democratic Republic of São Tomé and  
16 Príncipe

17 <sup>6</sup> Institute of Molecular and Cellular Biology, College of Life Sciences and Medicine, National  
18 Tsing Hua University, Hsinchu, Taiwan

19 <sup>7</sup> Institute of Environmental and Occupational Health Sciences, College of Public Health,  
20 National Taiwan University, Taipei, Taiwan

21

22 \* Corresponding author: Hsiao-Han Chang (hhchang@life.nthu.edu.tw)

23

24

25 **Abstract**

26

27 Despite efforts to eliminate malaria in Sao Tome and Principe (STP), cases have recently  
28 increased. Understanding residual transmission structure is crucial for developing effective  
29 elimination strategies. This study collected surveillance data and generated amplicon  
30 sequencing data from 980 samples between 2010 and 2016 to examine the genetic structure  
31 of the parasite population. The mean multiplicity of infection (MOI) was 1.3, with 11%  
32 polyclonal infections, indicating low transmission intensity. Temporal trends of these genetic  
33 metrics did not align with incidence rates, suggesting that changes in genetic metrics may  
34 not straightforwardly reflect changes in transmission intensity, particularly in low transmission  
35 settings where genetic drift and importation have a substantial impact. While 88% of samples  
36 were genetically linked, continuous turnover in genetic clusters and changes in drug-  
37 resistance haplotypes were observed. Principal component analysis revealed some STP  
38 samples were genetically similar to those from Central and West Africa, indicating possible  
39 importation. These findings highlight the need to prioritize several interventions such as  
40 targeted interventions against transmission hotspots, reactive case detection, and strategies  
41 to reduce the introduction of new parasites into this island nation as it approaches elimination.  
42 This study also serves as a case study for implementing genetic surveillance in a low  
43 transmission setting.

44

45

46

47

48

49

## 50 **Introduction**

51

52 Effective interventions have significantly reduced the malaria burden in the Democratic  
53 Republic of Sao Tome and Principe (STP), an island nation in the Gulf of Guinea, Central  
54 Africa <sup>1,2,3</sup>. In recognition of the substantial progress achieved, the World Health Organization  
55 (WHO) has included the country in the E-2025 initiative, aiming to eliminate malaria by the  
56 year 2025 <sup>4</sup>. Despite being in a pre-elimination phase, the annual incidence rate has  
57 consistently remained around 10 cases per 1,000 people over 7 years. Moreover, there was  
58 a 46% increase in case numbers reported in 2022 <sup>5</sup>. These setbacks pose challenges to  
59 achieving the E-2025 goal and underscore the necessity of understanding the genetic  
60 structure of residual malaria transmission in STP in order to tailor effective elimination  
61 strategies.

62

63 The main island, Sao Tome, consists of six administrative districts. The majority of cases  
64 occur in the capital district, Agua Grande, where approximately 40% of the population resides  
65 <sup>1</sup>. Malaria cases are reported year-round, with a slight decline often observed during the dry  
66 season from June to early September <sup>1,6</sup>. Located about 300 km off the coast of Gabon,  
67 STP's increase in international tourism and active trade facilitate potential parasite  
68 connectivity between STP and other countries in West and Central Africa <sup>7</sup>. However, there  
69 has been limited exploration of the role of importation on malaria transmission in STP <sup>6</sup>.  
70 Malaria interventions in STP have focused on vector control <sup>1,8</sup>, passive case surveillance <sup>6</sup>,  
71 and/or mass drug administration <sup>9</sup>, but have not included routine collection of travel history or  
72 performing reactive case detection, which may be useful strategies for preventing imported  
73 malaria and additional onward transmission in low-transmission settings <sup>10</sup>.

74

75 The current understanding of malaria transmission in STP includes epidemiological findings,  
76 the evaluation of intervention effectiveness, and the analysis of specific genes in the local  
77 vector (*Anopheles coluzzii*) and the malaria parasite (*Plasmodium falciparum*)<sup>1, 6, 11, 12, 13</sup>.  
78 However, a gap remains due to the lack of integration between epidemiological metadata  
79 and genomic information, which is crucial for understanding the structure of residual  
80 transmission in STP. With advancements and cost reductions in sequencing technology, an  
81 increasing number of countries are adopting genomic surveillance approaches to track  
82 transmission dynamics, potential importations, and the spread of drug-resistant parasites<sup>14,</sup>  
83 <sup>15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27</sup>. Several genetic metrics have been proposed to estimate  
84 malaria transmission, including the multiplicity of infection (MOI—which denotes the number  
85 of genetically distinct parasite strains within an individual), within-host relatedness,  $F_{ws}$ , and  
86 pairwise relatedness between infections<sup>22, 28, 29, 30, 31, 32, 33</sup>. Moreover, in settings with low  
87 transmission, genetic data can capture changes in circulating parasites that may not be easily  
88 detected through traditional epidemiological or clinical measures, thereby providing useful  
89 information for elimination efforts<sup>14, 34</sup>.

90

91 To understand the genetic structure of residual malaria transmission in STP, we performed  
92 amplicon sequencing to obtain the first genome-wide genetic dataset of malaria parasites in  
93 the country. By integrating parasite genomic data with epidemiological metadata, this study  
94 offers spatiotemporal inferences on transmission clusters that can help guide effective  
95 malaria control strategies. The fine-scale analysis of genetic metrics and transmission  
96 inferences can serve as a valuable case study for applying genetic epidemiology to malaria  
97 in a low-transmission setting, particularly as more countries experience transmission declines  
98 and approach malaria elimination.

99

100

## 101 **Materials and Methods**

102

### 103 *Ethics statement*

104

105 The transfer, shipment, and use of malaria dried blood spots (DBS) and encrypted case  
106 surveillance data for research analysis in Taiwan was approved by the Centro Nacional de  
107 Endemias (CNE) in STP (reference no. OF°N°20/P°CNE/2016). This study was reviewed and  
108 approved by the Research Ethics Committee of National Taiwan University Hospital  
109 (reference no. 201110023RD) and the Research Ethics Review Committee of National Tsing  
110 Hua University (reference no. 11012HM135).

111

### 112 *Study materials*

113

114 A total of 7,482 malaria dried blood spots (DBS) from the Central Hospital Ayres de Menezes  
115 (HAM) were obtained from a surveillance project conducted by the Taiwan Anti-Malarial  
116 Advisory Team in partnership with the Centro Nacional de Endemias (CNE) of STP (Fig. S1a).  
117 We randomly selected 1,629 DBS, evenly distributed across our study period nationwide,  
118 ensuring around 20 samples per month from 2010 to 2016. The average sample sizes by  
119 year in areas with relatively high and low malaria endemicity were 70 and 5 per district,  
120 respectively. The DBS, along with the basic characteristics of patients (including age, gender,  
121 onset date, residential district and village, treatment regime, and diagnostic results), were  
122 collected from patients who had either visited or reported to HAM and were confirmed positive  
123 for *Plasmodium falciparum* infections through microscopy and/or rapid diagnostic tests. The  
124 total numbers of samples from six administrative districts, Agua Grande (AG), Me-Zochi (MZ),  
125 Lobata (LO), Cantagalo (CT), Lemba (LE), and Principe (PR) were 936 (57%), 375 (23%),

126 236 (15%), 74 (5%), 6 (0.4%), and 2 (0.1%), respectively. Agua Grande is the capital district.  
127 The incidence data used in this study were provided by the Taiwan Anti-Malarial Advisory  
128 Team in STP <sup>1</sup>. The data supporting the findings of this study have been deposited in the  
129 GitHub repository, accessible at [https://github.com/hhc-lab/malaria\\_genetics\\_STP](https://github.com/hhc-lab/malaria_genetics_STP).

130

### 131 *DNA extraction and qPCR*

132

133 DNA was extracted from DBS using the Tween-Chelex method described in Teyssier et al.,  
134 2021 <sup>35</sup>. The 6-mm disc from the DBS punch underwent two-step washes with 0.5% Tween  
135 20 in 1X PBS and 1X PBS at 4 °C. After removing the wash solution, the DBS disc was  
136 covered with 10% Chelex 100 resin in water at 95 °C for 10 minutes and centrifuged at 15,000  
137 rpm for 10 minutes. The resulting supernatant, which contained the extracted DNA, was then  
138 transferred into a PCR tube and stored at -20 °C.

139

140 Parasite density was measured using the *varATS* qPCR protocol as previously described <sup>36</sup>.  
141 DNA extracted from mock DBS with known parasite densities of 1, 10, 100, 1,000, and 10,000  
142 parasites/ $\mu$ L served as qPCR standards. Duplicate sets of these standards were used to  
143 generate a standard curve for each run, with a linearity ( $R^2$ ) value exceeding 0.98 being  
144 considered a validated run. The parasite densities of the samples were then benchmarked  
145 using the established standard curve. A total of 1,478 samples were confirmed as positive  
146 with quantified parasite densities and were processed for amplicon sequencing.

147

### 148 *Amplicon sequencing*

149

150 The amplification of two pools, comprising 165 diversity amplicons, 38 drug-resistance

151 amplicons (containing 68 important drug-resistance sites), and 38 immune, diagnostic, and  
152 species-related amplicons, was obtained using the MAD4HatTeR V.3 protocol, building on a  
153 previously published method targeting diverse microhaplotypes<sup>37, 38</sup> and the CleanPlex  
154 library preparation kit from Paragon Genomics, CA, USA<sup>39</sup>. The two pools were combined  
155 and processed through clean-up steps with magnetic beads for each sample. Samples were  
156 then pooled together and gel-extracted based on parasite density and the presence of primer  
157 dimers. All sample pools underwent quality checks on a TapeStation (Agilent, CA, USA) and  
158 were sequenced with 150 bp paired-end reads on Illumina sequencers.

159

160 For quality assurance, each sample plate included at least one positive control (3D7) and  
161 several negative controls randomly placed in each plate to monitor amplification efficiency  
162 and detect potential contamination. If the positive controls did not perform well (less than 100  
163 reads per amplicon), or if the negative controls showed more than 10 reads per targeted  
164 amplicon on average, the entire plate was discarded and re-prepared until good-quality and  
165 contamination-free results were obtained.

166

### 167 *Bioinformatic pipeline and quality filtering*

168

169 The amplicon data were processed through the allele calling pipeline V0.1.8 (available at  
170 <https://github.com/EPPICenter/mad4hatter>). The pipeline consists of core modules designed to  
171 filter and correct demultiplexed reads, remove adapters, and enable accurate and precise  
172 identification of variant alleles. In this study, we focused on genetic diversity analysis using 165  
173 diversity amplicons to estimate parasite genetic metrics, and seven drug-resistance amplicons  
174 to infer drug-resistance haplotypes.

175

176 To ensure data quality, we applied the following criteria to select high-quality samples for  
177 analyses using diversity amplicons: (a) total reads per sample exceeding the number of  
178 amplified amplicons multiplied by 100, (b) amplicon coverage exceeding 75%, and (c) more  
179 than half of the amplicons having 100 or more reads (Fig. S2). These standards ensured  
180 uniform and adequate amplification across most amplicons per sample. For drug-resistance  
181 amplicons, the criterion was that more than half of them should have 10 or more reads. A  
182 total of 980 samples met these requirements and were included in the analysis (Fig. S1b).

183

#### 184 *Analysis of parasite genetic diversity*

185

186 We utilized the *MOIRE V3.1.0* R package to implement a Bayesian approach for estimating  
187 MOI from polyallelic data, accounting for experimental error<sup>28</sup>. We also estimated a new  
188 metric of diversity, the effective MOI (eMOI), by adjusting the true MOI for underlying within-  
189 host relatedness ( $r_w$ ), the average proportion of the genome that is identical by descent  
190 across all strains within the same host<sup>28</sup>. The eMOI was defined as  $(MOI - 1) * (1 - r_w) + 1$ .  
191 We used the means from the posterior distributions of MOI, eMOI, and  $r_w$  as their point  
192 estimates. A polyclonal infection was defined as having eMOI greater than 1.1.

193

194 We employed the *Dcifer V1.2.0* R package, an identity-by-descent (IBD) method, to estimate  
195 between-host relatedness from polyallelic data ( $r_b$ )<sup>31</sup>. Using the *MOIRE*-estimated MOI and  
196 allele frequencies as inputs, we obtained an estimation of between-host relatedness for each  
197 pair, ranging from 0 to 1, accompanied by a corresponding p-value for statistical significance.  
198 Genomic clusters were identified by applying various levels of  $\hat{r}_b$ , with 0.9 being the default  
199 and most stringent threshold and 0.3 being the least. The resulting 62 clusters using the 0.9  
200 threshold were then ordered and named by their size (from small to large; C1 to C62). The



201 clustering network was visualized, displaying only the links with a significance level that had  
202 a Bonferroni-corrected p-value of less than 0.05.

203

204 Given the seasonality of malaria transmission in STP, we classified annual seasons based  
205 on the cycle of transmission intensity (Table S1 and Fig. S3) instead of using the calendar  
206 year. The year was categorized from the beginning of the rainy season to the conclusion of  
207 the dry season, that is, from September of the prior year to August of the current year. Malaria  
208 incidence rates were also calculated based on the adjusted year and geographical groups  
209 (Capital or Others). Regression analysis was employed to identify the relationship between  
210 MOI, eMOI, and polyclonal infections and other characteristics, including the sampling year,  
211 season (rainy or dry), residential area (Capital or Others), location (Q1 and Q2, where Q1  
212 represents locations at the village administrative level with malaria case numbers ranking in  
213 the top 25% within each district and Q2 represents others), age, gender, parasite density  
214 from qPCR ( $\log_{10}$  transformed), and treatment regime (artemisinin-based combination  
215 therapy or quinine).

216

### 217 *Estimation of effective population size*

218

219 We estimated the effective population sizes based on temporal changes in allele frequencies  
220 between consecutive years using the temporal method in *NeEstimator V2.1*<sup>40, 41, 42</sup>. We used  
221 SNPs from diversity amplicons and excluded those with more than 50% missing data. Since  
222 the generation time for malaria parasites in STP was not previously estimated, we explored  
223 a range of values from 2 to 9 generations per year.

224

### 225 *Analysis of drug-resistance mutations*

226

227 A total of 9 amino acid sites across 3 resistance genes (*mdr1*, *dhfr*, and *dhps*) were analyzed.  
228 For each sample, we first identified the predominant amino acid at each site using read counts,  
229 then combined the major amino acids from six sites in *dhfr* and *dhps* and three sites in *mdr1*  
230 (Table S2) to determine the drug-resistance haplotype for SP (sulfadoxine-pyrimethamine)  
231 and ASAQ/AL (artesunate-amodiaquine/artemether-lumefantrine), respectively.

232

233 *Principal component analysis (PCA)*

234

235 We performed PCA for sequences from STP and six African countries, including Gabon,  
236 Nigeria, Ghana, Cameroon, Tanzania, and Kenya. Sequences from six African countries  
237 were obtained from the *Plasmodium falciparum* genomic variation dataset version 7 (Pf7) <sup>43</sup>.  
238 For each sample, sequences in the regions corresponding to the genomic locations of the  
239 diversity amplicons from STP samples were extracted using *BCFtools* <sup>44</sup>. These sequences  
240 were then merged with STP amplicon sequences to create a unified dataset. For STP  
241 samples with more than one sequence per amplicon per sample, the sequence with the  
242 highest read count was used. The concatenated sequences were aligned using the long  
243 sequence aligner *FAME* <sup>45</sup>. Sites with homology greater than 90% and sites containing more  
244 than 50% gaps were removed. To perform PCA <sup>46</sup>, aligned sequences were converted into a  
245 Pandas DataFrame, and bases were one-hot encoded using a base map.

246

## 247 **Results**

248

249 *Genetic metrics were consistent with low malaria transmission but did not temporally track*  
250 *with incidence rate*

251

252 With 980 sequenced samples collected nationwide over 7 consecutive years (Fig. S1b and  
253 Table S1), we examined four genetic metrics related to transmission intensity: MOI  
254 (multiplicity of infection), eMOI (effective MOI), within-host relatedness ( $\widehat{r}_w$ ), and the  
255 proportion of polyclonal infections. Overall, the MOI and eMOI among the samples were both  
256 low (mean = 1.3 [MOI] and 1.08 [eMOI]; Fig. 1a). Monoclonal infections were predominant,  
257 with polyclonal infections accounting for only 11% of samples (Fig. 1b). Among polyclonal  
258 infections, the average within host relatedness ( $\widehat{r}_w$ ) was high at 0.7 (Q1-Q3 = 0.6-0.8; Fig.  
259 1c). Moreover, the estimated effective population sizes based on allele frequency changes  
260 are below 100 across all years (Fig. S4). Collectively, these data indicating low within-host  
261 diversity are consistent with low transmission intensity in STP during the study period. While  
262 these genetic metrics varied over time, their temporal trend differed from that of the incidence  
263 rate (Spearman's rho < 0.6 and  $p > 0.2$  for all; Fig. 1a, 1b). During the study period, the  
264 incidence was highest in 2012, while MOI was higher before and after that (2011 and 2014).

265

266 In addition to temporal differences, we identified two other significant factors associated with  
267 the proportion of polyclonal infections, MOI, and eMOI (Table S3). A higher proportion of  
268 polyclonal infections (14% vs. 9%), along with higher MOI (1.34 vs. 1.24) and eMOI (1.10 vs.  
269 1.05), were observed in locations outside or bordering the capital district compared to within  
270 the capital (Fig. 2 and Table S3). This suggests that effective recombination rates were higher  
271 outside or bordering the capital district, despite the fact that the capital district harbored the  
272 highest number of malaria cases. Another significant factor was age — individuals aged  
273 above 5 years exhibited a higher proportion of polyclonal infections compared to children  
274 under 5 (Table S3). This trend could be attributed to increased mobility among older  
275 individuals, leading to greater exposure to infectious sources. Additionally, a higher

276 proportion of mild or asymptomatic infections among older individuals<sup>6, 47</sup> may result in  
277 delayed treatment, facilitating longer duration infections and the potential for superinfection.

278

279 *Genetic relatedness revealed sustained local transmission with continuous clonal*  
280 *replacement*

281

282 A large majority of samples (88%) were infected by parasites closely related to at least one  
283 other sample (between-host relatedness  $[\hat{r}_b] \geq 0.9$ ), forming 62 highly related clusters (Fig.  
284 3a). The high proportion of clustered samples and relatively small number of clusters  
285 identified each year (ranging from 8 to 31) suggest that our study captured the majority of a  
286 limited extent of population genetic diversity, despite only genotyping a representative sample  
287 of cases. In addition, these data suggest that most detected cases resulted from local  
288 transmission and, corroborating the incidence and within-host diversity data, are consistent  
289 with low transmission intensity. Despite low and seasonal transmission, the majority of  
290 clusters (63%) were found across multiple years, indicating sustained local transmission.

291

292 While clusters frequently persisted over multiple years, few spanned the entire study period.  
293 Generally, we observed turnover in the parasite population over time (Fig. 3a). We found  
294 transitional changes in the parasite population over the years across a range of relatedness  
295 cutoffs from relaxed ( $\hat{r}_b \geq 0.3$ ) to stringent ( $\hat{r}_b \geq 0.9$ ) (Fig. 4a). Parasites collected in the same  
296 year exhibited higher relatedness compared to those collected further apart in time (Fig. 4ab).  
297 The mean within-year relatedness peaked in 2016 (average  $\hat{r}_b = 0.43$ ), while the mean  
298 between-year relatedness between 2016 and earlier years, such as 2010 or 2011, was very  
299 low (average  $\hat{r}_b \leq 0.02$ ) (Fig. 4b). Drug resistance haplotypes also exhibited a temporal trend,  
300 with mutations suggesting a decreasing resistance level to SP (Fig. 5a) and an increasing

301 resistance level to AL (Fig. 5b).

302

303 Two predominant clusters, C61 and C62, accounted for 17 and 20% of the total samples,  
304 respectively, and persisted for 4-5 years, with C61 dominating before 2014 and C62  
305 dominating after 2014 (Fig. 3a and 4c). Upon closer examination, C61 comprises two  
306 subgroups and C62 consists of four subgroups. The samples linking these subgroups were  
307 all polyclonal, with MOI values ranging from 2.8 to 4.1 and  $\widehat{r}_w$  values ranging from 0.26 to  
308 0.5 (Fig. S5a). The intra-subgroup  $\widehat{r}_b$  was very high (mean  $\widehat{r}_b = 0.99$ ), contrasting with the  
309 very low inter-subgroup  $\widehat{r}_b$  (mean  $\widehat{r}_b = 0.03$ ; Fig. S5b). This indicates that the subgroups were  
310 largely unrelated, suggesting minimal effective recombination among them. The connector  
311 samples among subgroups were likely superinfections, which occurred infrequently and did  
312 not produce sufficient recombinants that circulated within the population to be frequently  
313 observed. When we lowered the  $\widehat{r}_b$  cutoff to 0.6, clear substructures remained within C61 and  
314 C62, and C61 and C62 became interconnected through few samples that were polyclonal  
315 (Fig. 4c). This, along with the low average relatedness between most clusters (Fig. 4d) and  
316 the bimodal distribution of within-year  $\widehat{r}_b$  (Fig. 4e), once again suggests that the parasite  
317 clusters in STP were loosely related, with limited yet present opportunities for superinfections  
318 and effective recombination, due to low transmission intensity. Moreover, consistent with  
319 evidence of limited effective recombination, the temporal changes in drug resistance  
320 haplotypes were mainly driven by the frequency changes of different subclusters to which  
321 they belong (Fig. S5c).

322

323 In addition to temporal changes in the parasite population, we also observed spatial  
324 substructure (Fig. 4f). Parasites in or near the capital (AG and LO), where most cases  
325 occurred, exhibited similar levels of relatedness within and between districts (average  $\widehat{r}_b =$

326 0.06-0.09). Conversely, two districts with lower incidence rates than the capital, Me-Zochi  
327 and Cantagalo, displayed the highest relatedness among themselves (average  $\hat{r}_b = 0.13$ -  
328 0.15), but lower relatedness with other districts in Sao Tome Island (average  $\hat{r}_b \leq 0.09$ ). While  
329 some general spatial substructure was evident, many clusters comprised samples from  
330 different districts, including Me-Zochi and Cantagalo, suggesting parasite flow across the  
331 entire region (Fig. S6). Interestingly, the two predominant clusters, C61 and C62, also  
332 exhibited different spatial distributions, with the former mainly existing in AG and MZ, while  
333 the latter primarily in AG and LO (Fig. 4f). With the predominant clusters changing from C61  
334 to C62 in 2014, the number of cases in MZ also greatly decreased (Fig. S1a), suggesting  
335 limited genetic diversity in MZ.

336

337 *The role of importation and the decreased parasite diversity in the end of the study*

338

339 Fig. 3 illustrates annual turnover in the parasite population, with old clusters disappearing  
340 and new clusters emerging. From 2011 to 2015, an average of 45% of clusters (12% of  
341 infections) were newly detected each year, representing a substantial amount. The newly  
342 emerged clusters may result from newly imported parasites, recombination among pre-  
343 existing parasites, and/or an increase in the frequency of infections from previously  
344 unsampled clusters. If effective recombination had occurred, we would expect to observe a  
345 group of recombinants connecting clusters. However, even after decreasing the  $\hat{r}_b$  cutoff to  
346 0.3, the subgroups were connected by few connector samples consistent with superinfections  
347 (Fig. S7). This suggests a limited role for recombination and a higher likelihood of importation.  
348 Additionally, PCA analysis revealed that samples from STP mainly clustered together with  
349 samples from Central and West Africa, situated between the clearly separated patterns of  
350 East and West African samples (Fig. 6), suggesting parasite flow between STP islands and

351 other countries in continental Africa.

352

353 We observed a decline in parasite diversity in the end of the study, reflected by both the  
354 number of clusters and parasite effective population size (Fig. 3a and S4). The number of  
355 clusters decreased during this period, suggesting that many clusters detected in previous  
356 years were either eliminated or reduced in size to undetectable levels (Fig. 3a). The  
357 proportion of clusters that were no longer detected in the following years increased after 2014  
358 (Fig. 3b). While newly identified clusters were substantial from 2011 to 2015, a shift was seen  
359 in 2016, with no new clusters identified (Fig. 3c). The proportion of non-clustered or uniquely-  
360 clustered samples in 2016 was the lowest (7%) across the years, with 74% of the samples  
361 belonging to the predominant cluster C62. These data suggest a decrease in the rate of  
362 importation and/or the rates at which imported cases were successfully propagated locally in  
363 2016.

364

## 365 **Discussion**

366

367 Our study presents the first genomic description of the malaria parasite population in STP  
368 using amplicon sequencing technology. We offer a comprehensive depiction of the parasite  
369 population dynamics across time and space, capturing the majority of parasite genetic  
370 diversity in the country. Our findings reveal sustained, low level local transmission over a  
371 span of 7 years with moderate turnover of the parasite population likely due to sporadic  
372 propagation of imported cases. Given the consistently limited population diversity observed  
373 in STP, extending malaria genomic surveillance could serve as a valuable resource for  
374 identifying potential cases imported into STP in the future.

375

376 Genetic analysis offers valuable insights into parasite population dynamics and the  
377 effectiveness of intervention policies, complementing traditional surveillance data such as  
378 incidence of malaria <sup>15, 17, 22, 48, 49, 50, 51</sup>. Incidence data may be influenced by health-seeking  
379 behaviors, and asymptomatic infections often go undetected. Furthermore, while incidence  
380 data provide changes in case numbers over time, they do not reveal the underlying structure  
381 of transmission, which is relevant when planning interventions. Our study shows that the  
382 majority of infections in STP are closely related to others, suggesting a limited number of  
383 parasite lineages circulating in STP. Moreover, while the incidence rate remained similar from  
384 2014 to 2016, genetic analysis revealed that parasite genetic diversity was the lowest in 2016.  
385 This decline may be due to the enhanced malaria control interventions in 2015 and 2016,  
386 such as improvements in case management <sup>6</sup> and the use of new vector control methods,  
387 including outdoor larval control and the use of alternative insecticide in Indoor Residual  
388 Spraying (IRS) <sup>1</sup>.

389  
390 After years of malaria control, transmission intensity in several countries or regions, including  
391 STP, has decreased to a low level, with the next goal being malaria elimination <sup>52, 53</sup>. In low  
392 transmission settings, genetic surveillance plays a crucial role in identifying imported cases  
393 <sup>54</sup>, yet questions remain regarding the accuracy of genetic metrics in reflecting changes in  
394 underlying transmission levels <sup>20</sup>. In our study, genetic metrics support the notion of low  
395 transmission intensity, but their changes over time, including the proportion of polyclonal  
396 infections, MOI, effective MOI, and effective population size, were inconsistent with changes  
397 in the incidence rate. While MOI is influenced by transmission intensity, which dictates rates  
398 of superinfection and co-transmission, it is also affected by other factors, including host age  
399 and immunity, population diversity, importation, and the genetic diversity of the locations from  
400 which imported cases originate <sup>55, 56</sup>. Effective population size estimated from allele frequency



401 changes is also influenced by importations, which increase fluctuations in allele frequencies  
402 and lead to underestimates in effective population size. In low transmission settings, genetic  
403 metrics are highly influenced by stochastic fluctuations in the parasite population. Therefore,  
404 caution is warranted when interpreting genetic metrics, and they are better considered in the  
405 context of other surveillance data and known or suspected drivers of transmission. Due to  
406 the low transmission intensity in STP, the parasite population is particularly susceptible to  
407 chance effects, and the continuous emergence of new clusters suggests the impact of  
408 importation on genetic metrics. Similarly, recent studies in Senegal and Zambia have also  
409 found a limited association between genetic metrics and incidence rates in areas with low  
410 transmission intensity <sup>20, 22</sup>.

411

412 To establish the genetic basis of malaria parasites in STP, efforts were made to evenly  
413 sample across time and space, aiming to reduce bias and capture as many different genetic  
414 lineages as possible (Fig. S1). We also redefined the boundaries between years based on  
415 the transmission season to better reflect changes in malaria transmission from season to  
416 season (Fig. S3). Utilizing 165 diversity amplicons, we successfully captured both temporal  
417 changes and spatial substructure, including the continuous turnover of parasite clusters as  
418 well as genetic differentiation between some districts. Additionally, we observed a higher  
419 proportion of polyclonal infections outside the capital compared to inside, despite the higher  
420 case numbers in the capital. These polyclonal infections from non-capital districts suggest  
421 higher rates of transmission leading to superinfection, which provide opportunities for  
422 recombination and potentially increase selection efficiency <sup>57</sup>. Given the connectivity between  
423 the capital and other districts (Fig. S6), these infections have the potential to spread into the  
424 capital, thereby complicating malaria control efforts.

425

426 Furthermore, evidence of a moderate degree of importation was inferred from the clustering  
427 patterns and PCA results. Despite the Sao Tome and Principe islands being approximately  
428 300 kilometers apart from the mainland of Africa, the increased international tourism and  
429 frequent trading with high malaria transmission countries like Cameroon, Nigeria, and Gabon  
430 in recent years <sup>7</sup> make it reasonable to expect importation. However, the absence of  
431 comprehensive travel surveys has made it difficult to identify imported or introduced cases <sup>58,</sup>  
432 <sup>59</sup>. Although this study utilized genetic relatedness to understand importation, it could not fully  
433 distinguish previously unsampled parasites present in the population from imported parasites.  
434 Therefore, incorporation of a systematic travel survey, routine reactive case detection, and  
435 more comprehensive genomic surveillance into the elimination program could improve  
436 monitoring and help design and monitor strategies to reduce the impact of importation <sup>60, 61,</sup>  
437 <sup>62, 63</sup>.

438  
439 Our study also characterized changes in drug resistance haplotypes of three genes, *mdr1*,  
440 *dhfr*, and *dhps* in STP, revealing a decrease of SP-resistant haplotypes and an increase in  
441 AL-resistant haplotypes. Sulfadoxine-pyrimethamine (SP) is used solely for intermittent  
442 preventive treatment (IPT) in pregnancy in STP [2] and its infrequent use may explain the  
443 decrease in SP resistance. According to WHO guidance <sup>64</sup>, AL, one of the artemisinin-based  
444 combination therapies (ACT), is used as a second-line drug in STP <sup>6</sup>. Among patients in our  
445 data, another ACT, AQ, is more frequently used and AL is rarely used. Thus, the increase in  
446 resistance to AL is unlikely driven by its use in STP, but is more likely led by the co-occurrence  
447 of AL-resistant and SP-lightly-resistant haplotypes in the same subgroups (Fig. S5c) and/or  
448 other non-selective factors, such as frequency changes of subgroups due to genetic drift or  
449 the emergence of new subgroups due to importation. While AL is rarely used in STP, it is  
450 recommended for treating uncomplicated infections in nearby countries such as Angola,

451 Gabon, Nigeria, and Cameroon <sup>5</sup>. Treatment failure of AL for uncomplicated *P. falciparum*  
452 malaria has been increasingly reported in sub-Saharan African countries <sup>65, 66, 67, 68</sup>. Since the  
453 frequency of drug-resistant parasites can change rapidly once they evolve in a low-  
454 transmission area, routine surveillance of drug-resistance mutations or haplotypes is crucial  
455 in STP.

456

457 The last mile for STP to eliminate malaria could be challenging, given the increasing number  
458 of cases recorded in the past three years <sup>5, 69</sup>. Our study revealed that the majority of malaria  
459 population was infected by highly clonal parasites with sustained transmission, while a  
460 minority were attributed to genetically distinct parasites likely originating from external  
461 sources. Therefore, we recommend disrupting the sustained local transmission by  
462 strengthening targeted interventions in common foci, and additionally implementing routine  
463 reactive surveillance and collecting travel histories, particularly for highly-mobile individuals.  
464 Lastly, this study also demonstrates how genomic surveillance can enhance our  
465 understanding of transmission dynamics. This is particularly important for low-transmission  
466 countries like STP, as it enables the establishment of strong early warning systems and aids  
467 in achieving the elimination goal.

468

## 469 **Conclusion**

470

471 Leveraging both genomic and epidemiological data allows us to capture fine-scale dynamics  
472 of the parasite population in a pre-elimination setting. In a low-transmission setting, parasites  
473 could have experienced significant changes in population structure and drug resistance, as  
474 our study demonstrated. Targeted interventions should not only be strengthened in common  
475 foci to eliminate prevalent strains but also involve reactive case detection and tracking of

476 travel history to prevent imported transmission in STP.

477

## 478 **Author contributions**

479

480 HHC, YAC, BG, and KHT conceptualized the study. RC, LFT, and KHT led the data and  
481 sample collection. YAC and CSL designed and supervised the experiments. YAC, PYN, and  
482 AE performed experiments. YAC, DG, and PYN conducted the analysis. YAC and BP  
483 assisted with the bioinformatic analysis. YAC, HHC, and BG contributed to the interpretation  
484 of the results. YAC and HHC wrote the manuscript with help from PYN, DG, CSL, and BG.

485

## 486 **Acknowledgements**

487

488 We would like to express our gratitude to the study participants and study teams for their  
489 cooperation, especially the Taiwan Anti-Malarial Advisory Mission and the Centro Nacional  
490 de Endemias (CNE) in STP. We also want to express our deepest gratitude to the leader of  
491 the Taiwan Anti-Malarial Advisory Mission team, Dr. Jih-Ching Lien, and to Dr. Arlindo  
492 Vicente de Assunção Carvalho from the local health ministry. Dr. Jih-Ching Lien (1927-2022)  
493 was a leading entomologist who made significant contributions to the identification of  
494 mosquito specimens. Dr. Arlindo Vicente de Assunção Carvalho (1961-2024) not only  
495 provided professional suggestions for the study but also assisted in coordinating with local  
496 departments. We thank Maxwell Murphy and Inna Gerlovina from the UCSF EPPIcenter for  
497 their comments on data analysis, Andres Aranda-Diaz for his guidance on the amplicon  
498 sequencing experiment and bioinformatic pipeline, and Ju-Hsuan Wang and Yu-Wen Huang  
499 for their help with data collection. This research was funded by the Taiwan Ministry of Foreign  
500 Affairs and the National Science and Technology Council (NSTC 113-2636-B-007-006). YAC

501 was supported by the Fulbright Program and the UCSF EPPIcenter. HHC was supported by  
502 Yushan Scholar Program. BG was supported by NIH/NIAID mentoring award K24 AI144048.

503

#### 504 **Competing interests**

505 The authors declare no competing interests.

506

#### 507 **References**

- 508 1. Chen YA, *et al.* Effects of indoor residual spraying and outdoor larval control on  
509 Anopheles coluzzii from São Tomé and Príncipe, two islands with pre-eliminated  
510 malaria. *Malar J* **18**, 405 (2019).  
511
- 512 2. Teklehaimanot HD, Teklehaimanot A, Kiszewski A, Rampao HS, Sachs JD. Malaria in  
513 Sao Tome and principe: on the brink of elimination after three years of effective  
514 antimalarial measures. *The American journal of tropical medicine and hygiene* **80**,  
515 133-140 (2009).  
516
- 517 3. Wang Y, Li M, Guo W, Deng C, Zou G, Song J. Burden of Malaria in Sao Tome and  
518 Principe, 1990-2019: Findings from the Global Burden of Disease Study 2019.  
519 *International journal of environmental research and public health* **19**, (2022).  
520
- 521 4. WHO. World malaria day: WHO launches effort to stamp out malaria in 25 more  
522 countries by 2025. Available from: [https://www.who.int/news/item/21-04-2021-world-](https://www.who.int/news/item/21-04-2021-world-malaria-day-who-launches-effort-to-stamp-out-malaria-in-25-more-countries-by-2025)  
523 [malaria-day-who-launches-effort-to-stamp-out-malaria-in-25-more-countries-by-2025.](https://www.who.int/news/item/21-04-2021-world-malaria-day-who-launches-effort-to-stamp-out-malaria-in-25-more-countries-by-2025)  
524 (2021).  
525
- 526 5. WHO. World malaria report 2023. (ed<sup>^</sup>(eds). World Health Organization (2023).  
527
- 528 6. Chen YA, *et al.* Dynamic changes in genetic diversity, drug resistance mutations, and  
529 treatment outcomes of falciparum malaria from the low-transmission to the pre-  
530 elimination phase on the islands of São Tomé and Príncipe. *Malaria Journal* **20**, 467  
531 (2021).  
532
- 533 7. Tralac. São Tomé and Principe: 2019 intra-Africa trade and tariff profile., (2020).  
534
- 535 8. Tseng LF, Chang WC, Ferreira MC, Wu CH, Rampão HS, Lien JC. Rapid control of  
536 malaria by means of indoor residual spraying of alphacypermethrin in the Democratic  
537 Republic of Sao Tome and Principe. *The American journal of tropical medicine and*  
538 *hygiene* **78**, 248-250 (2008).  
539
- 540 9. Li M, *et al.* Mass Drug Administration With Artemisinin-Piperaquine for the Elimination  
541 of Residual Foci of Malaria in São Tomé Island. *Front Med (Lausanne)* **8**, 617195  
542 (2021).  
543

- 544 10. Moonen B, *et al.* Operational strategies to achieve and maintain malaria elimination.  
545 *Lancet (London, England)* **376**, 1592-1603 (2010).  
546
- 547 11. Pinto J, *et al.* Genetic structure of *Anopheles gambiae* (Diptera: Culicidae) in Sao  
548 Tome and Principe (West Africa): implications for malaria control. *Mol Ecol* **11**, 2183-  
549 2187 (2002).  
550
- 551 12. Pinto J, *et al.* Malaria in Sao Tome and Principe: parasite prevalences and vector  
552 densities. *Acta tropica* **76**, 185-193 (2000).  
553
- 554 13. Lee PW, Liu CT, do Rosario VE, de Sousa B, Rampao HS, Shaio MF. Potential threat  
555 of malaria epidemics in a low transmission area, as exemplified by Sao Tome and  
556 Principe. *Malar J* **9**, 264 (2010).  
557
- 558 14. Nsanzabana C. Strengthening Surveillance Systems for Malaria Elimination by  
559 Integrating Molecular and Genomic Data. *Tropical medicine and infectious disease* **4**,  
560 (2019).  
561
- 562 15. Neafsey DE, Volkman SK. Malaria Genomics in the Era of Eradication. *Cold Spring*  
563 *Harb Perspect Med* **7**, (2017).  
564
- 565 16. Tessema SK, Raman J, Duffy CW, Ishengoma DS, Amambua-Ngwa A, Greenhouse  
566 B. Applying next-generation sequencing to track falciparum malaria in sub-Saharan  
567 Africa. *Malaria Journal* **18**, 268 (2019).  
568
- 569 17. Daniels R, *et al.* Genetic Surveillance Detects Both Clonal and Epidemic Transmission  
570 of Malaria following Enhanced Intervention in Senegal. *PLOS ONE* **8**, e60780 (2013).  
571
- 572 18. Nkhoma SC, *et al.* Population genetic correlates of declining transmission in a human  
573 pathogen. *Molecular Ecology* **22**, 273-285 (2013).  
574
- 575 19. Hendry JA, Kwiatkowski D, McVean G. Elucidating relationships between *P.falciparum*  
576 prevalence and measures of genetic diversity with a combined genetic-  
577 epidemiological model of malaria. *PLOS Computational Biology* **17**, e1009287 (2021).  
578
- 579 20. Schaffner SF, *et al.* Malaria surveillance reveals parasite relatedness, signatures of  
580 selection, and correlates of transmission across Senegal. *Nature Communications* **14**,  
581 7268 (2023).  
582
- 583 21. Carrasquilla M, *et al.* Resolving drug selection and migration in an inbred South  
584 American *Plasmodium falciparum* population with identity-by-descent analysis. *PLOS*  
585 *Pathogens* **18**, e1010993 (2022).  
586
- 587 22. Fola AA, *et al.* Temporal and spatial analysis of *Plasmodium falciparum* genomics  
588 reveals patterns of parasite connectivity in a low-transmission district in Southern  
589 Province, Zambia. *Malaria Journal* **22**, 208 (2023).  
590
- 591 23. Neafsey DE, Taylor AR, MaInnis BL. Advances and opportunities in malaria  
592 population genomics. *Nat Rev Genet* **22**, 502-517 (2021).  
593



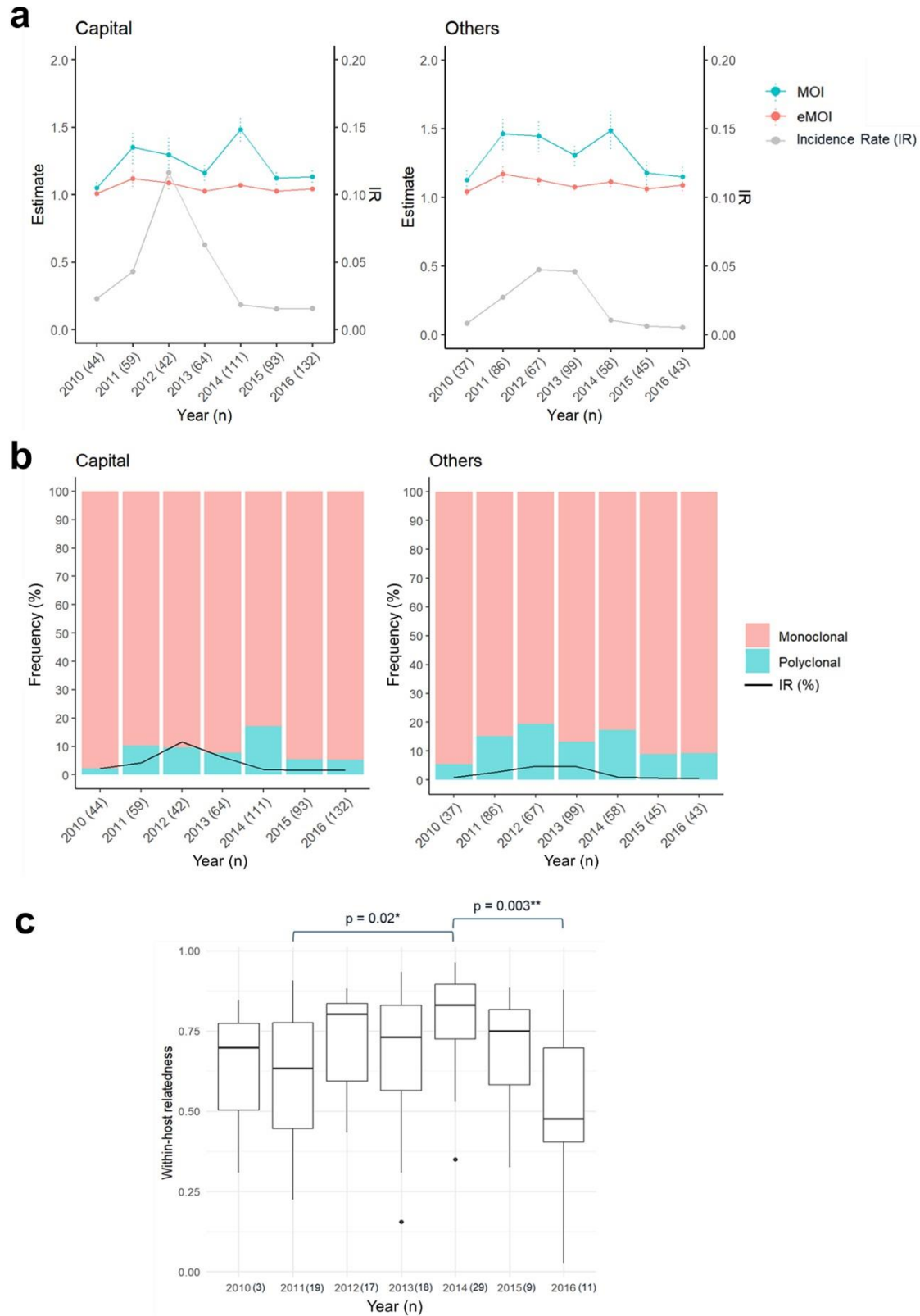
- 594 24. Ruybal-Pesántez S, McCann K, Vibin J, Siegel S, Auburn S, Barry AE. Molecular  
595 markers for malaria genetic epidemiology: progress and pitfalls. *Trends in parasitology*  
596 **40**, 147-163 (2024).  
597
- 598 25. Henden L, Lee S, Mueller I, Barry A, Bahlo M. Identity-by-descent analyses for  
599 measuring population dynamics and selection in recombining pathogens. *PLOS*  
600 *Genetics* **14**, e1007279 (2018).  
601
- 602 26. Campino S, *et al.* Population Genetic Analysis of Plasmodium falciparum Parasites  
603 Using a Customized Illumina GoldenGate Genotyping Assay. *PLOS ONE* **6**, e20251  
604 (2011).  
605
- 606 27. Connelly SV, *et al.* Strong isolation by distance and evidence of population  
607 microstructure reflect ongoing Plasmodium falciparum transmission in Zanzibar.  
608 *medRxiv : the preprint server for health sciences*, (2024).  
609
- 610 28. Murphy M, Greenhouse B. MOIRE: A software package for the estimation of allele  
611 frequencies and effective multiplicity of infection from polyallelic data. *bioRxiv : the*  
612 *preprint server for biology*, (2023).  
613
- 614 29. Nkhoma SC, *et al.* Co-transmission of Related Malaria Parasite Lineages Shapes  
615 Within-Host Parasite Diversity. *Cell host & microbe* **27**, 93-103.e104 (2020).  
616
- 617 30. Chang H-H, *et al.* THE REAL McCOIL: A method for the concurrent estimation of the  
618 complexity of infection and SNP allele frequency for malaria parasites. *PLOS Comput*  
619 *Biol* **13**, e1005348 (2017).  
620
- 621 31. Gerlovina I, Gerlovin B, Rodríguez-Barraquer I, Greenhouse B. Dcifer: an IBD-based  
622 method to calculate genetic distance between polyclonal infections.  
623 <https://eppicenter.github.io/dcifer/>. *Genetics* **222**, (2022).  
624
- 625 32. Schaffner SF, Taylor AR, Wong W, Wirth DF, Neafsey DE. hmmlBD: software to infer  
626 pairwise identity by descent between haploid genotypes. *Malaria Journal* **17**, 196  
627 (2018).  
628
- 629 33. Auburn S, *et al.* Characterization of Within-Host Plasmodium falciparum Diversity  
630 Using Next-Generation Sequence Data. *PLOS ONE* **7**, e32891 (2012).  
631
- 632 34. Dalmat R, Naughton B, Kwan-Gett TS, Slyker J, Stuckey EM. Use cases for genetic  
633 epidemiology in malaria elimination. *Malar J* **18**, 163 (2019).  
634
- 635 35. Teyssier NB, Chen A, Duarte EM, Sit R, Greenhouse B, Tessema SK. Optimization of  
636 whole-genome sequencing of Plasmodium falciparum from low-density dried blood  
637 spot samples. *Malaria Journal* **20**, 116 (2021).  
638
- 639 36. Gruenberg M, *et al.* Utility of ultra-sensitive qPCR to detect Plasmodium falciparum  
640 and Plasmodium vivax infections under different transmission intensities. *Malaria*  
641 *Journal* **19**, 319 (2020).  
642
- 643 37. Andres Aranda-Diaz ENV, Bryan Greenhouse. MAD4HatTeR V.3. Available at

- 644 <https://eppicenter.ucsf.edu/resources>. (ed<sup>^</sup>(eds) (2022).
- 645
- 646 38. Tessema SK, *et al.* Sensitive, Highly Multiplexed Sequencing of Microhaplotypes From  
647 the Plasmodium falciparum Heterozygome. *J Infect Dis*, (2020).
- 648
- 649 39. Paragon Genomics I. CleanPlex® Targeted Library Preparation User Guide For  
650 Amplicon Sequencing on Illumina® Sequencers.  
651 [https://www.paragongenomics.com/wp-content/uploads/2016/06/CleanPlex-Library-](https://www.paragongenomics.com/wp-content/uploads/2016/06/CleanPlex-Library-Kit_User_Guide_02111620A1-EN.pdf)  
652 [Kit User Guide 02111620A1-EN.pdf](https://www.paragongenomics.com/wp-content/uploads/2016/06/CleanPlex-Library-Kit_User_Guide_02111620A1-EN.pdf). (2018).
- 653
- 654 40. Do C, Waples RS, Peel D, Macbeth GM, Tillett BJ, Ovenden JR. NeEstimator v2: re-  
655 implementation of software for the estimation of contemporary effective population  
656 size ( $N_e$ ) from genetic data. *Molecular ecology resources* **14**, 209-214 (2014).
- 657
- 658 41. Jorde PE, Ryman N. Temporal allele frequency change and estimation of effective  
659 size in populations with overlapping generations. *Genetics* **139**, 1077-1090 (1995).
- 660
- 661 42. Waples RS. A generalized approach for estimating effective population size from  
662 temporal changes in allele frequency. *Genetics* **121**, 379-391 (1989).
- 663
- 664 43. Abdel Hamid MM, *et al.* Pf7: an open dataset of Plasmodium falciparum genome  
665 variation in 20,000 worldwide samples. *Wellcome open research* **8**, 22 (2023).
- 666
- 667 44. Li H, *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**,  
668 2078-2079 (2009).
- 669
- 670 45. Naznooshsadat E, Elham P, Ali S-Z. FAME: fast and memory efficient multiple  
671 sequences alignment tool through compatible chain of roots. *Bioinformatics* **36**, 3662-  
672 3668 (2020).
- 673
- 674 46. Konishi T, Matsukuma S, Fuji H, Nakamura D, Satou N, Okano K. Principal  
675 Component Analysis applied directly to Sequence Matrix. *Scientific Reports* **9**, 19297  
676 (2019).
- 677
- 678 47. Salgado C, *et al.* The prevalence and density of asymptomatic Plasmodium falciparum  
679 infections among children and adults in three communities of western Kenya. *Malaria*  
680 *Journal* **20**, 371 (2021).
- 681
- 682 48. Wesolowski A, *et al.* Mapping malaria by combining parasite genomic and  
683 epidemiologic data. *BMC Medicine* **16**, 190 (2018).
- 684
- 685 49. Chang H-H, *et al.* Mapping imported malaria in Bangladesh using parasite genetic and  
686 human mobility data. *eLife* **8**, e43481 (2019).
- 687
- 688 50. Daniels RF, *et al.* Genetic analysis reveals unique characteristics of Plasmodium  
689 falciparum parasite populations in Haiti. *Malar J* **19**, 379 (2020).
- 690
- 691 51. Verity R, *et al.* The impact of antimalarial resistance on the genetic structure of  
692 Plasmodium falciparum in the DRC. *Nat Commun* **11**, 2107 (2020).
- 693

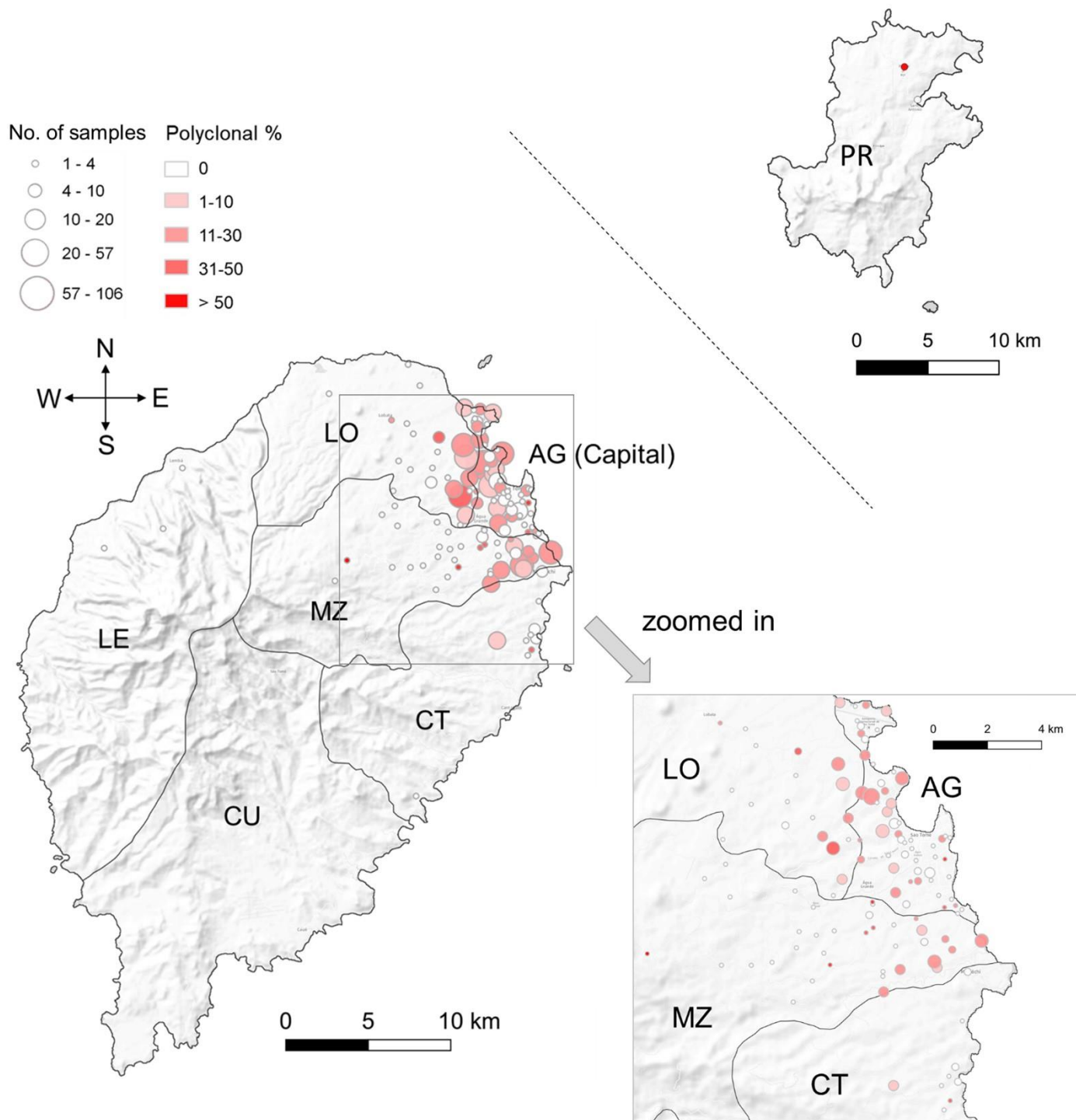


- 694 52. Lindblade KA, *et al.* Supporting countries to achieve their malaria elimination goals:  
695 the WHO E-2020 initiative. *Malaria Journal* **20**, 481 (2021).  
696
- 697 53. Wangmo LD, *et al.* Sustaining progress towards malaria elimination by 2025: Lessons  
698 from Bhutan & Timor-Leste. *The Lancet Regional Health – Western Pacific* **22**,  
699 (2022).  
700
- 701 54. WHO. Technical consultation on the role of parasite and anopheline genetics in  
702 malaria surveillance (2019).  
703
- 704 55. Roh ME, *et al.* High Genetic Diversity of Plasmodium falciparum in the Low-  
705 Transmission Setting of the Kingdom of Eswatini. *J Infect Dis* **220**, 1346-1354 (2019).  
706
- 707 56. Labbé F, *et al.* Neutral vs. non-neutral genetic footprints of Plasmodium falciparum  
708 multiclonal infections. *PLOS Computational Biology* **19**, e1010816 (2023).  
709
- 710 57. Rice WR, Chippindale AK. Sexual Recombination and the Power of Natural Selection.  
711 *Science* **294**, 555-559 (2001).  
712
- 713 58. Wesolowski A, *et al.* Quantifying the impact of human mobility on malaria. *Science*  
714 **338**, 267-270 (2012).  
715
- 716 59. Ruktanonchai NW, *et al.* Identifying Malaria Transmission Foci for Elimination Using  
717 Human Mobility Data. *PLoS Comput Biol* **12**, e1004846 (2016).  
718
- 719 60. Hagmann R, Charlwood JD, Gil V, Ferreira C, do Rosário V, Smith TA. Malaria and its  
720 possible control on the island of Príncipe. *Malar J* **2**, 15 (2003).  
721
- 722 61. Fasih BS, *et al.* Risk of imported malaria infections in Zanzibar: a cross-sectional study.  
723 *Infectious Diseases of Poverty* **12**, 80 (2023).  
724
- 725 62. Das AM, *et al.* Modelling the impact of interventions on imported, introduced and  
726 indigenous malaria infections in Zanzibar, Tanzania. *Nature Communications* **14**,  
727 2750 (2023).  
728
- 729 63. Guerra CA, *et al.* Human mobility patterns and malaria importation on Bioko Island.  
730 *Nature Communications* **10**, 2332 (2019).  
731
- 732 64. WHO. Artemisinin resistance and artemisinin-based combination therapy efficacy.  
733 (2018).  
734
- 735 65. Halsey ES, Plucinski MM. Out of Africa: Increasing reports of artemether-lumefantrine  
736 treatment failures of uncomplicated Plasmodium falciparum infection. *Journal of*  
737 *Travel Medicine* **30**, (2023).  
738
- 739 66. Silva-Pinto A, *et al.* Artemether-lumefantrine treatment failure of uncomplicated  
740 Plasmodium falciparum malaria in travellers coming from Angola and Mozambique.  
741 *International journal of infectious diseases : IJID : official publication of the*  
742 *International Society for Infectious Diseases* **110**, 151-154 (2021).  
743

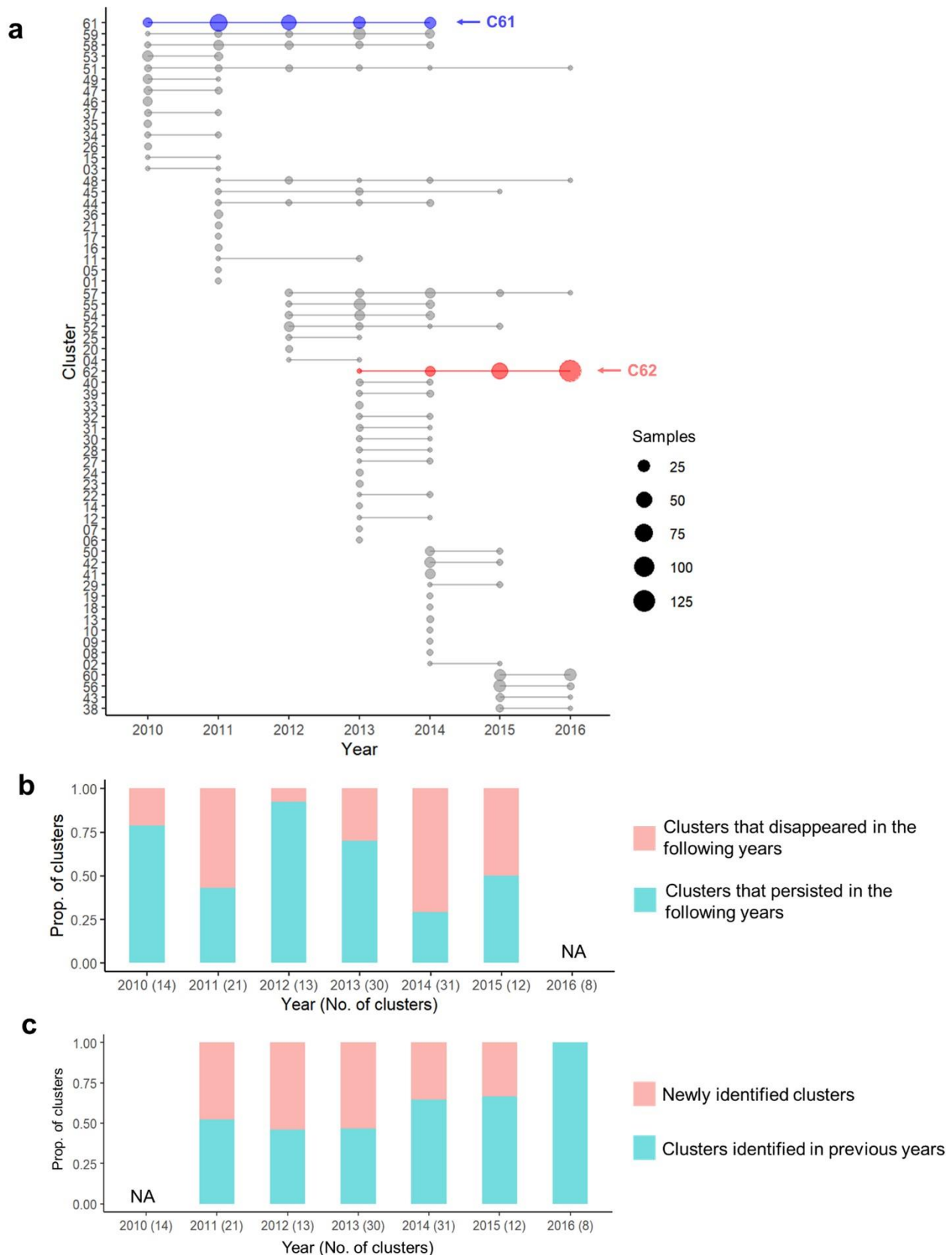
- 744 67. van Schalkwyk DA, *et al.* Treatment Failure in a UK Malaria Patient Harboring  
745 Genetically Variant *Plasmodium falciparum* From Uganda With Reduced In Vitro  
746 Susceptibility to Artemisinin and Lumefantrine. *Clinical Infectious Diseases*, (2023).  
747
- 748 68. Ebong C, *et al.* Efficacy and safety of artemether-lumefantrine and dihydroartemisinin-  
749 piperaquine for the treatment of uncomplicated *Plasmodium falciparum* malaria and  
750 prevalence of molecular markers associated with artemisinin and partner drug  
751 resistance in Uganda. *Malaria Journal* **20**, 484 (2021).  
752
- 753 69. WHO. Towards malaria elimination: Strengthening the quality of malaria diagnostic  
754 services in Sao Tome and Principe. (2023).  
755  
756



**Fig. 1. Temporal variation of parasite genetic metrics and malaria incidence rates.** (a) MOI, eMOI, and incidence rates by reclassified year in the capital district (left) and other districts (right). The dots represent the mean for each year, and the bars for MOI and eMOI indicate the standard errors. The sample size for each year is shown in brackets on the x-axis. (b) Proportion of polyclonal (cyan) and mono-clonal (pink) infections by year. Polyclonal infection is defined as eMOI greater than 1.1. (c) Boxplot of within-host relatedness ( $\widehat{r}_w$ ) for polyclonal infections each year. The difference in within-host relatedness between years was tested using Dunn's test.

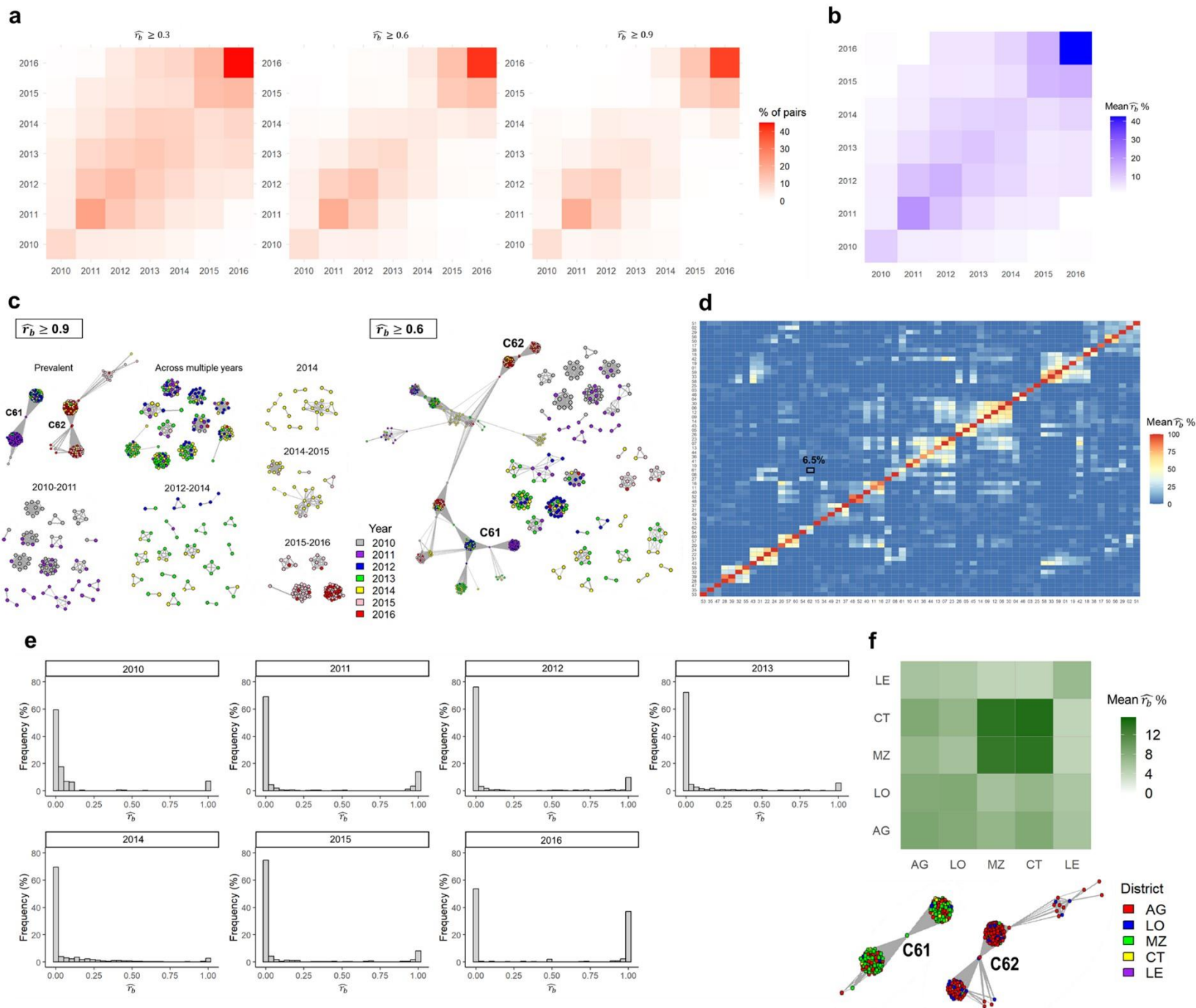


**Fig. 2. Distribution of sequenced samples and proportion of polyclonal infections.** Sao Tome Island comprises six administrative districts: Agua Grande (AG, capital), Lobata (LO), Me-Zochi (MZ), Cantagalo (CT), Caue (CU), and Lemba (LE). Principe (PR) island is located 173 km away from Sao Tome Island. The dots on the map were plotted at the centroids of each location, which may represent residential hamlets, specific places, or roads reported by infected individuals. The size of each dot is proportional to the number of samples sequenced from that location, while the color gradients indicate the level of polyclonal infections.

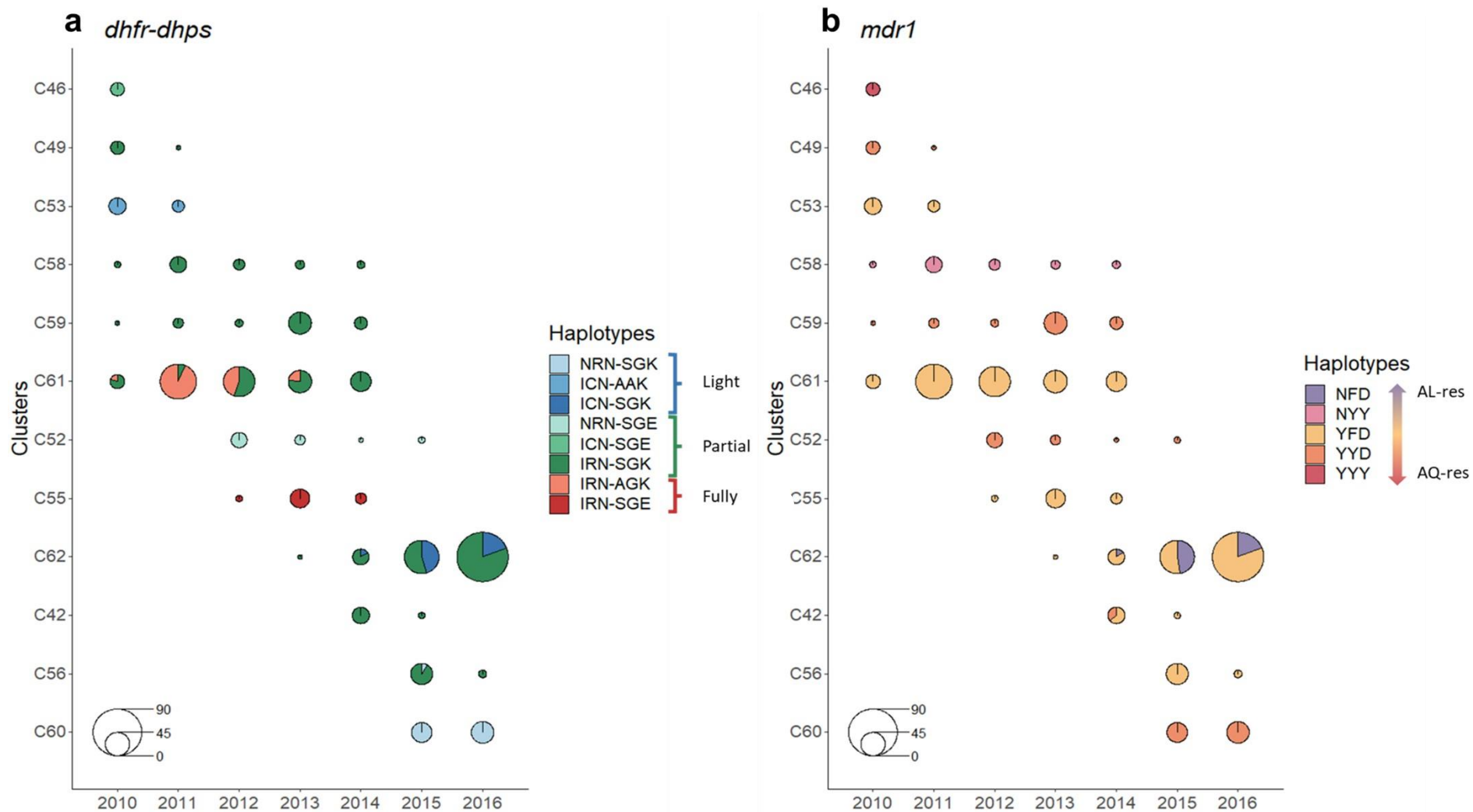


**Fig. 3. Characterization of clusters across years.** (a) Using a cutoff of  $\hat{r}_b \geq 0.9$ , 62 clusters were identified. The two most prevalent clusters are marked in blue (C61) and red (C62), respectively. (b) Proportion of clusters that disappeared in the following years (pink) and those that persisted (cyan). (c) Proportion of clusters that were newly identified in each year (pink) and those that were identified in previous years (cyan). NA = Not Applicable (since 2010 is the first year and 2016 is the last year, there is no information for previous years and following years, respectively).



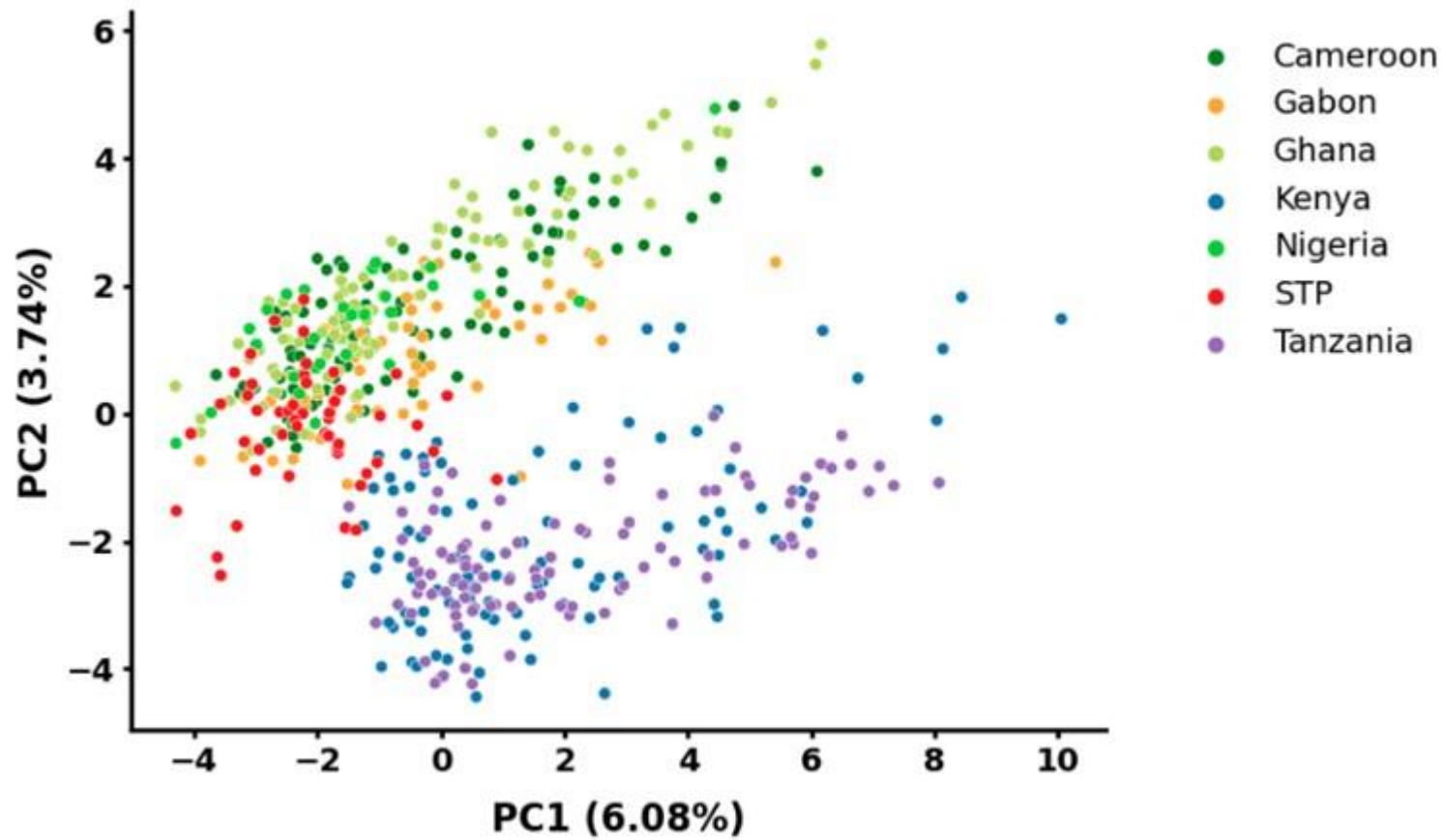


**Fig. 4. Estimated pairwise relatedness ( $\widehat{r}_b$ ) across time, space, and clusters.** (a) Proportion of pairs with  $\widehat{r}_b$  over 0.3, 0.6, or 0.9 between all pairs of years or within the same years. (b) Mean  $\widehat{r}_b$  across and within years. (c) Samples with significant relatedness above 0.9 (left) and 0.6 (right) with a Bonferroni-corrected p-value of less than 0.05 are connected. Two predominant clusters, C61 and C62, are connected by  $\widehat{r}_b \geq 0.6$  through some transitional infections in 2013 and 2014. (d) Mean  $\widehat{r}_b$  among 62 clusters identified using  $\widehat{r}_b \geq 0.9$  cutoff. The average relatedness between C61 and C62 is 6.5% (marked in square). (e) Distribution of  $\widehat{r}_b$  among samples within the same year. (f) Mean  $\widehat{r}_b$  between districts in Sao Tome main island and the composition of districts in two predominant clusters, C61 and C62.



**Fig. 5. Haplotype changes in (a) *dhfr-dhps* and (b) *mdr1* over the years.** The *dhfr-dhps* haplotypes (N511/C59R/S108N-S436A/A437G/K540E) are color-coded according to their resistance levels: blue for light resistance (3 mutants), green for partial resistance (4 mutants), and red for full resistance (5 mutants). The *mdr1* haplotypes (N86Y/Y184F/D1246Y) are colored purple and orange to indicate resistance to AL and ASAQ, respectively. Following the shift in predominant clusters after 2014, resistance against SP decreased from partial-full (green-red) to light-partial (blue-green). Additionally, the proportion of the AL-resistant haplotype (purple) in *mdr1* haplotypes emerged after 2014.





**Fig. 6. Principal component analysis (PCA) of samples from STP and six other African countries. Samples from STP are closely clustered with those from Central and West Africa.**