Duffy-Negative and Duffy-Positive *Plasmodium vivax* Shares Similar Microsatellite Gene Pool Indicative of Frequent Transmission in East Africa

Daniel Kepple[1], Alfred Hubbard[2], Musab M A. Albsheer[4], Beka Raya[3], Karen Lopez[2], Kareen Pestana[1], Daniel A. Janies[2], Guiyun Yan[5], Muzamil Mahdi Abdel Hamid[4], Delenasaw Yewhalaw[3], Eugenia Lo[1]

[1] Biological Sciences, University of North Carolina at Charlotte, USA

[2] Bioinformatics and Genomics, University of North Carolina at Charlotte, USA

[3] Tropical Infectious Disease Research Center, Jimma University, Ethiopia

[4] Department of Parasitology and Medical Entomology, Institute of Endemic Diseases, University of Khartoum, Khartoum, Sudan

Correspondence: Eugenia Lo <eugenia.lo@uncc.edu>, Biological Sciences, University of North Carolina at Charlotte

Running Title: Transmission patterns of Duffy-negative and Duffy-positive *Plasmodium vivax* in East Africa

Abstract – 150 words; Text – 3500 words

Abstract

Plasmodium vivax malaria has historically been rare in Africa due to the parasites requiring the Duffy antigen/chemokine receptor for erythrocyte invasion, which most Africans lack. However, recent reports have shown *P. vivax* cases in East Africa increasing at alarming rates and a portion of these cases were detected in Duffy-negative individuals. It is unclear if *P. vivax* in Duffy-negatives is genetically similar to that in Duffy-positives from the same region, and if there is frequent gene flow among populations in East Africa. This study examined genetic variation and transmission patterns of *P. vivax* from Duffy-negative and Duffy-positive individuals in Ethiopia and Sudan. Duffy-negative and Duffy-positive *P. vivax* revealed similar multiplicity of infections. Duffy-negative *P. vivax* showed lower genetic variation but higher genotypic evenness than that from Duffy-positives. No clear differentiated was found between Duffy-negative and Duffy-positive *P. vivax* coexisted in the same area, indicative of between-host transmission. A weak isolation-by-distance pattern was detected, and resistance surface analyses suggested that high road density and elevation enable *P. vivax* frequent gene flow. This study is the first to show that *P. vivax* can transmit to and from Duffy-negative individuals and provides critical insights into the widespread of *P. vivax* through sub-Saharan Africa.

Keywords: *Plasmodium vivax*, Duffy-Negative, Transmission patterns, East Africa, Genetic diversity

Introduction

43 *Plasmodium vivax* malaria is a neglected tropical disease, despite being more
44 geographically widespread than any other form of malaria [1] and causes 132–391 million
45 clinical infections each year [2]. *P. vivax* was previously thought to be rare or absent in
46 Africa because people of African descent often lack the Duffy blood group antigen, known
47 as the Duffy antigen-chemokine receptor (*DARC*), on the surface of red blood cells
48 (RBCs) that allows *P. vivax* to bind and invade human erythrocytes [3]. However, recent
49 studies have reported several cases of *P. vivax* infection in Duffy-negative people in
50 different parts of Africa where Duffy-negatives are predominant [4,5]. It is unclear whether
51 these cases are merely infections, meaning the parasite is unable to complete its life cycle
52 in Duffy-negative individuals, or whether *P. vivax* in Duffy-negative individuals can
53 develop into gametocytes and transmit to other individuals via mosquitoes and spread
54 through a population. The latter scenario constitutes a significant public health threat,
55 especially considering reports of antimalarial resistance [6-9] and the pathogen's unique
56 ability to form dormant-stage hypnozoites in the host liver, giving rise to relapse infections
57 from months to years later [10,11]. Compared to *P. falciparum*, *P. vivax* has a broader
58 temperature tolerance, an earlier onset of gametocyte development, and can re-emerge
59 as relapse [12], enabling *P. vivax* to spread through the diverse African climate [13] and
60 more difficult to control and eliminate [14].

61 Ethiopia and Sudan are two of the few African countries where *P. vivax* malaria is
62 prevalent [15], and Duffy-negative and Duffy-positive individuals coexist [3]. In Central
63 and West Africa, more than 97% of the population is Duffy-negative [3]. By contrast, in
64 Ethiopia and Sudan, about 35% and 50% of the general population are …., as well as
65 20% and 18% of the hospitalized malaria patients, respectively, are Duffy-negatives

66 [3,12,16,17]. Other East African countries such as Eritrea and Madagascar have also
67 reported significant *P. vivax* infections in Duffy-positive and Duffy-negative individuals
68 who live side-by-side [3,18], pointing to a potential of between-host transmission. In
69 addition, recent economic development has made Ethiopia a major hub for travelers,
70 further promoting the spread of *P. vivax* malaria across Africa [13,19,20].

71 The genetic diversity and population structure of *P. vivax* have been shown to be
72 substantially different from *P. falciparum* [10,21-23]. For example, *P. vivax* has a higher
73 nucleotide diversity compared to *P. falciparum* [21], which could be attributed to frequent
74 gene flow via human movement, higher transmission intensity and recombination, as well
75 as variation in host susceptibility [10,22-25]. In Papua New Guinea, *P. vivax* was shown
76 to have a 3.5-fold higher rate of polyclonality and nearly double the multiplicity of infection
77 (MOI) than *P. falciparum* infections [26]. Similarly, in Cambodia [27], the Indo-West Pacific
78 [28-30], and the Brazilian Amazon [31], *P. vivax* had a higher microsatellite diversity than
79 its sympatric *P. falciparum*. These findings highlight the ability of *P. vivax* to adapt to a
80 wide range of landscapes and climates. However, the questions of whether *P. vivax*
81 infecting Duffy-negative individuals is genetically similar to Duffy-positive ones from the
82 same region, and what landscape or environmental factors influence *P. vivax* gene flow
83 in Africa, remain unclear. This study examined genetic variation and source-sink
84 dynamics of *P. vivax* between Duffy-negative and Duffy-positive populations in Ethiopia
85 and Sudan. We further estimated how landscape and environmental factors influence
86 parasite gene flow. These data provide critical insights into whether and how *P. vivax*
87 evolves and spreads in Duffy-negative individuals, highlighting the need for *P. vivax*
88 diagnosis, tracking, and control in Africa.

89

90 ## Materials and Methods

91 ### Ethics Statement

92 Scientific and ethical clearance was obtained from the institutional scientific and ethical
93 review boards of Jimma University, Ethiopia, the ethical committee of Institute of Endemic
94 Diseases, Sudan (IEND; approval number 1/2014; 10 June 2014), the ethical committee
95 of state ministry of Khartoum and Kassala State Ministry of Health, Sudan, and University
96 of North Carolina, Charlotte, USA. Written informed consent/assent for study participation
97 was obtained from all consenting heads of households, parents/guardians (for minors
98 under 18 years old), and each individual who was willing to participate in the study.

99

100 ### Study areas and sample collection

101 Clinical samples (N???) were obtained from six different study sites, three from Ethiopia
102 and three from Sudan. Study sites in Ethiopia included Jimma (JM), Gojeb (GJ), and Arjo
103 (AJ); and study sites in Sudan included Khartoum (KH), River Nile (RN), and New Halfa
104 (NH). These sites are located around the Ethiopia-Sudan border area (Figure 1) and
105 experience low to moderate transmission [32]. Finger-prick blood samples were collected
106 from symptomatic malaria patients (who had a fever with axillary body temperature
107 >37.5°C and microscopy confirmed asexual stages for *P. vivax*). Thick and thin blood
108 smears were prepared for microscopic examination and three to four spots of blood,
109 equivalent to ~50μL, were blotted on Whatman 3MM filter paper.

110

111 I suggest to include a separate paragraph of

112 Plasmodium species identification and Duffy genotyping

113

114 Parasite DNA was extracted from dried blood spots by the Saponin/Chelex method [33].
115 Nested and quantitative PCR were performed to identify and confirm parasite species of
116 the infected samples. Duffy genotyping was performed using TaqMan real-time
117 quantitative PCR [16]. A total of 305 Duffy-positive (which ones are TT which ones are
118 CT genotype??) and 107 Duffy-negative samples were included in microsatellite
119 analyses. These included 150 Duffy-positive and 83 Duffy-negative *P. vivax* samples from
120 Ethiopia, and 155 Duffy-positive and 24 Duffy-negative samples from Sudan.

121

122

123

124 **Microsatellite genotyping**

125 Eight single-copy microsatellites with tri- or tetranucleotide repeats, which mapped to 6
126 chromosomes, were typed for *P. vivax* (S1 table). Alleles were PCR-amplified with
127 oligonucleotide primers [34-36]. For each PCR reaction, 4μL of genomic DNA were used
128 with 7μL of 2x dreamtaq master mix (Thermo Scientific, Waltham, MA), 2μL of PCR-grade
129 water, and 2μM of each primer (all forward primers were labeled with fluorescent dyes;
130 Applied Biosystems, Foster City, Ca) in a final volume of 20μL. PCR conditions were as

131 follows: 2 min, 94˚C; (30 sec, 94˚C; 40 sec, 58˚C; 50 sec, 72˚C) for 40 cycles; 5 min,
132 72˚C. After PCR amplification, products were separated on an Applied Biosystems 3130
133 Genetic Analyzer and all allele sizes were determined and visualized in GeneMapper. To
134 avoid background noise and potential artifacts, a threshold of 500 relative fluorescent
135 units was set for GeneMapper. For each sample, the dominant allele and any alleles at
136 least 50% of the height of the dominant allele were scored [34].

137 **Linkage disequilibrium and multiplicity of infections**

138 To examine whether the microsatellite loci are in linkage disequilibrium (LD), multilocus
139 LD was assessed among the parasite samples for each study site using the web-based
140 LIAN, version 3.7 [37]. The standardized index of association (I_A^S), which measures the
141 strength of linkage disequilibrium and views as a function of the rate of recombination
142 among samples, was calculated. Due to small sample size (N=2) of Duffy-negative
143 samples from New Halfa, Sudan, this population was excluded from LD analyses.

144 The percentage of polyclonal infections (samples with more than one allele at any
145 given locus), as well as average multiplicity of infection (MOI: the number of genetically
146 distinct clones present within a host), were estimated for each of the study sites and Duffy
147 populations. For each sample, MOI was scored as the maximum number of alleles
148 observed when all loci were considered.

149

150 **Genetic diversity and population structure**

151 Genotypic variation was calculated in GenoDive, version 2.0b27 [38]. We first calculated
152 genetic distances using the method of Smouse and Peakall, a squared Euclidean

153 distance based on the number of times a certain allele was found in the two individuals
154 [39]. The minimal distance class was set as the threshold to identify the following: (i) the
155 number of multilocus genotypes (*G*); (ii) Simpson's diversity index (*D*), also known as
156 Nei's genetic diversity corrected for sample size, which ranges from zero (where two
157 randomly-chosen individuals in a population share a single genotype) to one (where
158 individuals have different genotypes); and (iii) genotype evenness (*E*), which ranges from
159 zero (where one or a few genotypes dominate in a population) to one (where all genotypes
160 are of equal frequency in a population). In addition, the number of
161 effective alleles and expected heterozygosity were estimated for each study site.

162 A model-based Bayesian method implemented in STRUCTURE v2.3.4 [40] was
163 performed on all study sites to examine the partitioning of individuals to genetic clusters.
164 The number of clusters (*K*) was determined by simulating a range of *K* values from 1 (no
165 genetic differentiation among all sites) to 12 (all sites and Duffy populations were
166 genetically differentiated). The posterior probability of each value was then used to detect
167 the model value of *ΔK*, a quantity related to the second order of change with respect to *K*
168 of the likelihood function [41]. Posterior probability values were estimated using a Markov
169 Chain Monte Carlo (MCMC) method. A burn-in period of 100,000 iterations followed by
170 10^6 iterations of each chain was performed to ensure convergence of the MCMC. Each
171 MCMC chain for each value of *K* was run 15 times with the 'independent allele frequency'
172 option that allows individuals with ancestries in more than one group to be assigned into
173 one cluster. Individuals were assigned into *K* clusters according to membership coefficient
174 values (*Q*) ranged from 0 (lowest affinity to a cluster) to 1 (highest affinity to a cluster).
175 The partitioning of clusters was visualized with DISTRUCT [42].

176 In addition, neighboring-joining trees were constructed using the web-based
177 program MABL (Methods Algorithms Bioinformatics LIRMM) [43] and visualized using
178 FigTree v1.4.4 [44] to show the genetic relatedness between *P. vivax* samples from Duffy-
179 positive and Duffy-negative individuals in Ethiopia and Sudan. Finally, F_{ST} (fixation index)
180 was estimated for each pair of populations as θ, using GenoDive, version 2.0b27 [38].
181 The fixation index is a measure of genetic differentiation between two populations. The
182 resulting pairwise matrix of θ values was used to perform multidimensional scaling (MDS)
183 in two and three dimensions, using the R programming language, version 3.6.2. This
184 method was selected to complement the neighbor-joining trees created above, as MDS
185 uses a pairwise dissimilarity matrix to estimate the set of coordinates for each population
186 that best preserves the true "distances" between the populations. In this application, MDS
187 permits visualization of the genetic differences between populations with simple
188 scatterplots.

189

190 **Identification of source-sink dynamics and gene flow pathways**

191 To infer source-sink dynamics and gene flow pathways of *P. vivax*, the StrainHub
192 software [45] was used to generate a transmission network from the phylogeny of our
193 samples. StrainHub used the genetic relationships from the neighbor-joining tree and the
194 associated metadata states to reconstruct an estimate of ancestral states. In this case,
195 these states correspond to the geographic location and/or host Duffy status of the
196 common ancestors from the phylogenetic tree. These ancestral states were used to
197 determine state transitions, which were in turn used to visualize the transmission network.
198 To clearly illustrate transmission patterns, two networks were generated: one in which

199 samples were grouped according to country and host Duffy status, and one in which
200 samples were grouped according to site, with no consideration of host Duffy status.
201 StrainHub offers a handful of metrics for interpreting transmission networks, and the
202 Source Hub Ratio (SHR) was selected for this study. This is the ratio of transitions from
203 a given node over all transitions to or from that node. While this does not measure the
204 importance of a node in the network, it characterizes whether it is behaving as a source,
205 a sink, or a hub. Values near 1 indicate a source, values near 0 indicate a sink, and values
206 in the middle suggest the node is a hub.

207

208 **Influence of geographic and landscape factors**

209 To test for isolation-by-distance, the Mantel test was performed on the matrix of pairwise
210 F_{ST}-values. This test was conducted with the ade4 R package, version 1.7_15, with
211 10,000 permutations performed to estimate significance [46]. For this analysis, the
212 geographical coordinates of the six study sites were used for estimating physical
213 distances between sites, and Duffy-positives and Duffy-negatives were combined for
214 each study site.

215 To determine whether the observed patterns of genetic differentiation are
216 explained by landscape factors, resistance surfaces were fit to the genetic data for a
217 variety of gridded, environmental datasets (Suppl. Figure 1). Specifically, these were:
218 road density, as estimated by Meijer *et al.* [47]; elevation, as measured by the Shuttle
219 Radar Topography Mission [48]; and land cover, obtained from the MCD12Q1 dataset
220 [49]. Road density was selected as a proxy for human movement, whereas elevation and

land cover were chosen to be representative of mosquito habitat. The malaria samples were gathered between 2017 and 2019, so land cover data for 2018 was chosen as representative. The other two datasets are static. All three were projected to a consistent coordinate system (UTM 36N with WGS84 datum) and resampled to 8,000 km spatial resolution, selected to approximate the resolution of the coarsest dataset (road density). This preprocessing was performed using the Geospatial Data Abstraction Library, version 3.0.4 [50].

Resistance surfaces based on these three environmental datasets that best explain the observed genetic distances (pairwise F_{ST}-values) were modeled using the ResistanceGA R package, version 4.0-14 [51]. A resistance surface is a gridded, spatial dataset in which the value of each cell is proportional to the degree of obstruction (e.g. resistance) to gene flow posed by that space. The ResistanceGA software uses a genetic algorithm to find parameter values for each resistance surface that optimize the fit of a linear mixed effects model between resistances and genetic distances [51]. Resistance distances are obtained by estimating the path an organism (or the organism's genetic information, to be precise) would take between two points on the surface, based on the resistance values. Numerous other methods for creating resistance surfaces exist, including those that incorporate expert knowledge of the relationship between certain environmental features and the organism in question, but these methods are also more vulnerable to the bias of the researcher. This optimization procedure only incorporates bias in the choice of which environmental variables to include, as all parameter values are fit during the modeling process. The resulting resistance surfaces allow spatial

243 mapping of the degree to which the environmental variables in question are predicted to
244 obstruct gene flow.

245 Resistance surfaces were fit for each individual environmental variable, and
246 composite resistance surfaces were fit for each possible combination of multiple
247 environmental variables. Resistance distances between each sample location were
248 computed using the commute distance algorithm. Resistance distances computed using
249 circuit theory are generally considered to be best, but commute distance produces results
250 that are almost perfectly correlated with those generated by circuit theory and requires
251 roughly half the computations [51]. The corrected Akaike information criteria (AICc) was
252 used as the fit criteria for parameter optimization. Optimization was performed twice with
253 different random seeds to ensure convergence of the parameter values. Bootstrapping
254 was not performed, as it was not deemed meaningful with only six sample locations.

255

256
Results

257
Linkage disequilibrium and multiplicity of infections

258 Significant LD was detected for all pairwise combinations of microsatellite loci (Bonferroni-
259 corrected $P<0.05$). When all loci were pooled together, *P. vivax* from Duffy-negatives in
260 both countries showed a slightly higher level of linkage and/or rate of recombination than
261 samples from Duffy-positives (Table 1). I_A^S values of *P. vivax* from Duffy-negatives ranged
262 from 0.03 (Jimma, Ethiopia) to 0.77 (Khartoum, Sudan), whereas I_A^S values ranged from
263 0.04 (Gojeb, Ethiopia) to 0.25 (Khartoum, Sudan) in Duffy-positives.

264 *P. vivax* from Duffy-negatives and Duffy-positives showed a comparable rate of
265 polyclonal infections (15% and 14%, respectively; Table 1). Among the samples from
266 Duffy-negatives, the highest rate of polyclonal infections was observed in Jimma, Ethiopia
267 (JM: 17.2%), and no polyclonal infections in Khartoum, Sudan (KH) and Gojeb, Ethiopia
268 (GJ), although there were few samples from Duffy-negatives in these locations. Among
269 the samples from Duffy-positives, the highest rate of polyclonal infections was observed
270 in Gojeb, Ethiopia (30.7%), followed by River Nile, Sudan (21%); whereas the lowest was
271 observed in Khartoum (5.8%; Table 1). MOI was also similar between the Duffy-negative
272 (mean = 1.17) and Duffy-positive (mean = 1.16; Table 1) infections, with a maximum of
273 three clones detected within a sample. Because we were unable to confidently
274 differentiate the genotypes of the different clones in samples with more than 1 alleles in
275 two or more loci, 16 of the polyclonal samples were discarded in the genetic analyses
276 (Suppl. Table 1).

277

278 **Genetic diversity and population structure**

279 In both Ethiopia and Sudan, Duffy-positive infections showed high levels of genotypic and
280 allelic diversity than Duffy-negative infections (Table 2). The level of genetic diversity
281 observed in *P. vivax* from Duffy-positives was similar between Ethiopia and Sudan.
282 However, the diversity observed in the *P. vivax* from Duffy-negatives in Ethiopia was
283 much higher than that in Sudan, likely due to the differences in sample size. Interestingly,
284 our findings revealed higher genotypic evenness for Duffy-negative infections in both
285 countries (Table 2), suggesting a more even distribution of genotypes in populations
286 infecting Duffy-negatives.

287 At the population level, neighbor-joining trees did not indicate a clear distinction
288 among samples from Duffy-positives or Duffy-negatives (Figure 2a). The tree had
289 relatively short internodes with long terminal branches, suggesting the parasite lineages
290 rapidly diverged from one another. Duffy-positive and Duffy-negative infections from the
291 same area did not cluster in the same clade. Due to the concern of low sample size of
292 Duffy-negative infections in sites GJ, NH, and KH that skew branch length and the
293 observed clustering pattern, a second tree was constructed without these three
294 populations. The same genetic relationships were observed (Suppl. Figure 2a). The MDS
295 results mirror these clustering patterns, showing no clear distinction among samples by
296 host Duffy status or country (Figure 2b). MDS was also performed without the three
297 populations containing the smallest sample size (i.e., *P. vivax* from Duffy-negatives at
298 sites GJ, NH, and KH) and this did not change the clustering pattern (Suppl. Figure 2b).

299 At the individual level, STRUCTURE analyses indicated two most probable genetic
300 clusters among all samples (purple and yellow clusters; Figure 3). In Ethiopia, most of the
301 samples from Arjo, Gojeb, and Jimma predominately belonged to the purple cluster. No
302 distinguishable differences were detected between samples from Duffy-positive and
303 Duffy-negative infections. In Sudan, however, a mixed clustering pattern was observed in
304 the three study sites. For instance, ~80% (69 out of 86) of the samples from Duffy-
305 positives in Khartoum belonged to the yellow cluster (Figure 3), whereas samples from
306 Duffy-positives in New Halfa and River Nile had a mixture of membership in both the
307 yellow and purple clusters.

308

309 **Identification of source-sink dynamics and gene flow pathways**

310 The transmission networks produced by StrainHub showed that *P. vivax* from Duffy-
311 positive individuals in Ethiopia served as the primary source, and *P. vivax* from the
312 Sudanese Duffy-negative hosts was the sink population (Figure 4a). The Ethiopian Duffy-
313 negative population also had a high Source Hub Ratio, but the weight of the arrows
314 leading from this population indicate that it was involved in fewer transitions than *P. vivax*
315 from Ethiopian Duffy-positives. At a site level, *P. vivax* from Arjo (AJ) was a source, and
316 *P. vivax* from infections in New Halfa (HA) was a sink (Figure 4b). The weight of the
317 arrows suggest that Arjo had more gene flow with the other populations in Ethiopia (GJ
318 and JM) than those in Sudan.

319

320 **Influence of geographic and landscape factors**

321 The Mantel test indicated a pattern of isolation-by-distance in the population F_{ST}-values
322 (r_M-value = 0.325; *p*-value = 0.017). Using a confidence threshold of 0.05, the null
323 hypothesis of no relationship between genetic and geographic distance should be
324 rejected. Based on the AICc score and ranking, the resistance surface based on a single
325 covariate that best fit the F_{ST}-values was land cover (Figure 5a), while road density and
326 elevation comprised the best-fitting combination surface (Figure 5b). All resistance
327 surfaces were significant (*p*-value < 0.05; Suppl. Table 2), and the analysis showed that
328 geographic distance alone does not explain the genetic distances in our dataset (*p*-value
329 > 0.025 for a two-tailed test). ResistanceGA weighted elevation as 92% of the
330 combination surface (Figure 5b), constituting most of the variation in resistance. Visual
331 inspection suggests the exception is in the vicinity of Khartoum, where high resistance
332 values were associated with high road density. On the other hand, land cover had the

333 lowest AICc of any of the individual surfaces and all composite surfaces containing land
334 cover were ranked close behind, with slightly higher AICc scores. This suggested that
335 land cover played an important role in explaining genetic distances, even though it does
336 not appear in the highest-ranked surface. Of the different land cover classes, grasslands
337 and croplands were associated with low resistance, while savannas and open shrublands
338 were modeled to have higher resistance. Taken together, the Mantel test and
339 ResistanceGA analyses indicated that elevation and land cover were key factors in
340 explaining *P. vivax* gene flow among our study sites in Ethiopia and Sudan, while
341 geographic distance alone is less important.

342

343 Discussion

344 Vivax malaria was previously thought to be rare or absent in African populations who lack
345 the Duffy blood group antigen expression [2]. However, recent studies reported several
346 cases of *P. vivax* infection in Duffy-negative people in different parts of Africa [4], including
347 countries where Duffy-negatives are predominant [5,52]. Such a distribution raises
348 important questions of whether *P. vivax* infections transmit among Duffy-negative
349 individuals and between Duffy-negative and Duffy-positive populations, and what factors
350 govern *P. vivax* transmission. The ability of *P. vivax* infections to transmit and spread in
351 Duffy-negative populations would pose a significant public health danger in Africa [4-7].
352 This study provides critical insights into the genetic relationship, transmission pattern, and
353 environmental determinants of *P. vivax* infections in Duffy-negative and Duffy-positive
354 individuals from Ethiopia and Sudan, where a large number of *P. vivax* cases have been

355 reported [16,17,53] and Duffy-negative and Duffy-positive individuals live together in the
356 same areas [3,15].

357 Our analyses indicated that *P. vivax* from Duffy-positives and Duffy-negatives
358 shared similar gene pools, and parasite populations did not cluster according to
359 geographical locations. Low levels of linkage disequilibrium were observed among
360 samples in most of our study sites, with the exception of Khartoum that showed a high
361 I_A^S-value, likely due to low parasite transmission in the capitol of Sudan, an urban setting
362 with adequate local public health infrastructure [54,55]. High genetic diversity observed
363 in most study sites could be related to higher transmission and frequent gene flow
364 between *P. vivax* in Duffy-negative and Duffy-positive individuals and across the Ethiopia-
365 Sudan border. The transmission network indicated that gene flow primarily occurred from
366 *P. vivax* infecting Duffy-positives to that infecting Duffy-negatives, but gene flow can also
367 go the other way to a lesser extent. This suggests that Duffy-negative hosts may not be
368 a dead-end for *P. vivax* that only results in infection and/or clinical symptoms in an
369 individual, but that the parasite can also spread from one Duffy-negative individual to
370 another. The StrainHub results also show that while most transmission is occurring within
371 Sudan and Ethiopia, substantial gene flow also occurred over the border.

372 Interestingly, our findings showed that the MOI was similar between the Duffy-
373 negative and Duffy-positive infections. One possible explanation is that 'older' parasite
374 genotypes may re-enter the bloodstream from a reservoir of hypnozoites in the liver,
375 giving rise to relapse infection in *P. vivax* patients regardless of the Duffy status [56].
376 Another factor could be submicroscopic infections, especially those in the Duffy negative
377 individuals with low parasitemia [57], may allow different parasite clones to persist in the

378 host bloodstream for a long-time without detection and treatment [58]. Additionally,
379 transmission intensity, demographic factors, and mosquito density and vectoral capacity
380 may contribute to the differences in MOI among the study populations. Previous studies
381 on *P. falciparum* have shown that areas with increased susceptibility to mosquitoes,
382 higher population density, and a younger age group are associated with a higher
383 entomological inoculation rate. These factors may apply equally to Duffy negative and
384 Duffy positive populations, and consequently increase MOI [59,60].

385 The lower genetic diversity observed in Duffy-negative infections compared to
386 Duffy-positive infections suggested a lower transmission frequency amongst Duffy-
387 negative Africans. This is consistent with the transmission network by StrainHub showing
388 that *P. vivax* infections were mostly originated from the Duffy positive individuals and
389 lends support to the hypothesis that *P. vivax* was recently adapted to and spread in Duffy-
390 negative individuals [52]. It is noteworthy that the low number of Duffy-negative infections
391 in Gojeb, Khartoum, and New Halfa might also influence genetic diversity measures. To
392 allow transmission in Duffy-negative individuals, merozoites and sporozoites of *P. vivax*
393 could have well adapted to the human and mosquito hosts. *P. vivax* merozoites may
394 have used non-Duffy receptors to invade human erythrocytes in *P. vivax* invasions.
395 Genes such as glycophosphatidylinositol-anchored micronemal antigen (GAMA) and
396 tryptophan-rich antigens have been shown to be expressed in patients with low parasite
397 density. Also, CD71 (Transferrin Receptor 1, TfR1) has been shown to bind readily with
398 the reticulocyte binding proteins based on *in-vitro* experiments [61,62]. These alternative
399 ligand/receptor proteins may play a key role in erythrocyte invasion [52]. *P. vivax*
400 merozoites in Duffy-negatives likely developed from rings to mature schizonts, of which

401 some converted into gametocytes (sexual stages) and transmitted by mosquitoes. Further
402 investigations are needed to uncover the Duffy-independent invasion pathway and
403 transmission mechanism in Duffy-negative individuals [63,64].

404 Mantel test indicated that gene flow is inhibited by geographical distance, though
405 the resistance surface analysis suggested that landscape factors including land cover and
406 elevation play a more important role in determining gene flow. The only exception was in
407 the vicinity of Khartoum, where road density has been fit to a high resistance value, likely
408 to explain the considerable genetic distance between Khartoum and the other
409 populations. The results suggest that *P. vivax* gene flow occurred more readily in the rural
410 parts of Ethiopia and Sudan than in the more urban setting of Khartoum consistent with
411 other studies that showed higher diversity of *P. vivax* in areas close to the country border
412 of Sudan and in agriculture areas outside the capitol [65-67]. Given that Mantel test
413 provides low analytical power especially with the small number of sampling sites as in the
414 present study [68], further investigations with samples from expanded study sites and
415 with higher coverage of urban areas will help confirm this hypothesis. In addition, it might
416 be informative to explore other datasets that serve as proxies for either human activities
417 or mosquito habitats, beyond the landscape data examined here.

418 In conclusion, this study is the first to show that *P. vivax* can transmit to *and from*
419 Duffy-negative individuals and shed light on the transmission patterns of Duffy-negative
420 *P. vivax* in endemic-regions of East Africa. We integrated genetic with landscape data
421 and found that urban environments pose a greater hindrance to the transmission of *P.*
422 *vivax* than rural and less-developed areas. Future investigations into the *P. vivax*
423 transmission mechanism involving gametocyte infectivity experiments and *in-vitro*

424 invasion assays in Duffy-negative individuals would offer deeper insights into this new 425 phenomenon and bring attention to the urge in diagnosing and controlling *P. vivax* in 426 Africa.

427

428 **Notes**

429 ***Acknowledgements.*** We thank the field team from Jimma University and University 430 of Khartoum for their technical assistance; the communities and hospitals, for their 431 support and willingness to participate in this research; Colton and Greyson Cantley, Haley 432 Wheeler, and Miriam Rodriguez for assistance with DNA extraction; and Dr. Colby Ford 433 for providing guidance on StrainHub.

434 ***Disclaimer.*** The findings and conclusions in this report are those of the authors and do 435 not necessarily represent the official position of the National Institutes of Health.

436 ***Financial support.*** This research was funded by National Institutes of Health (R15 437 AI138002, R01 AI050243, U19 AI129326, and D43 TW001505).

438 ***Potential conflicts of interest.*** All authors: No reported conflicts of interest. All authors 439 have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts 440 that the editors consider relevant to the content of the manuscript have been disclosed.

441

442 **References**

443 1. World Health Organization. World Malaria Report 2018. WHO, Geneva.

444 2. Price RN, Anstey NM, Guerra CA, Yeung S, Tjitra E, White NJ. Vivax Malaria:
445 Neglected and Not Benign. The American Journal of Tropical Medicine and Hygiene.
446 2007;77(6_Suppl):79–87.

447 3. Howes RE, Patil AP, Piel FB, Nyangiri OA, Kabaria CW, Gething PW, et al. The global
448 distribution of the Duffy blood group. Nature Communications. 2011;2(1).

449 4. Howes RE, Jr. RCR, Battle KE, Longbottom J, Mappin B, Ordanovich D, et al.
450 Plasmodium vivax Transmission in Africa. PLOS Neglected Tropical Diseases.
451 2015;9(11).

452 5. Zimmerman PA. Plasmodium vivax Infection in Duffy-Negative People in Africa. The
453 American Journal of Tropical Medicine and Hygiene. 2017;97(3):636–8.

454 6. Anantabotla VM, Antony HA, Parija SC, Rajkumari N, Kini JR, Manipura R, et al.
455 Polymorphisms in genes associated with drug resistance of Plasmodium vivax in India.
456 Parasitology International. 2019;70:92–7.

457 7. Snounou G. Improving Plasmodium vivax malaria treatment: a little more chloroquine.
458 The Lancet Infectious Diseases. 2018;18(9):934–5.

459 8. Yeshiwondim AK, Tekle AH, Dengela DO, Yohannes AM, Teklehaimanot A.
460 Therapeutic efficacy of chloroquine and chloroquine plus primaquine for the treatment
461 of Plasmodium vivax in Ethiopia. Acta Tropica. 2010;113(2):105–13.

462 9. Yohannes AM, Ringwald P, Teklehaimanot A, Bergqvist Y. Confirmed Vivax
463 Resistance to Chloroquine and Effectiveness of Artemether-Lumefantrine for the

464 Treatment of Vivax Malaria in Ethiopia. The American Journal of Tropical Medicine and
465 Hygiene. 2011;84(1):137–40.

466 10. Auburn S, Benavente ED, Miotto O, Pearson RD, Amato R, Grigg MJ, et al. Genomic
467 analysis of a pre-elimination Malaysian Plasmodium vivax population reveals selective
468 pressures and changing transmission dynamics. Nature Communications. 2018;9(1).

469 11. Markus MB. The hypnozoite concept, with particular reference to malaria.
470 Parasitology Research. 2010;108(1):247–52.

471 12. Lo E, Hemming-Schroeder E, Yewhalaw D, Nguyen J, Kebede E, Zemene E, et al.
472 Transmission dynamics of co-endemic Plasmodium vivax and P. falciparum in Ethiopia
473 and prevalence of antimalarial resistant genotypes. PLOS Neglected Tropical
474 Diseases. 2017;11(7).

475 13. Elgoraish AG, Elzaki SEG, Ahmed RT, Ahmed AI, Fadlalmula HA, Mohamed SA, et
476 al. Epidemiology and distribution of Plasmodium vivax malaria in Sudan. Transactions
477 of The Royal Society of Tropical Medicine and Hygiene. 2019;113(9):517–24.

478 14. White MT, Shirreff G, Karl S, Ghani AC, Mueller I. Variation in relapse frequency and
479 the transmission potential of Plasmodium vivax malaria. Proceedings of the Royal
480 Society B: Biological Sciences. 2016;283(1827):20160048.

481 15. Sibley CH. A Solid Beginning to Understanding Plasmodium vivax in Africa. The
482 Journal of Infectious Diseases. 2019;220(11):1716–8.

16. Lo E, Yewhalaw D, Zhong D, Zemene E, Degefa T, Tushune K, et al. Molecular epidemiology of Plasmodium vivax and Plasmodium falciparum malaria among Duffy-positive and Duffy-negative populations in Ethiopia. Malaria Journal. 2015;14(1):84.

17. Albsheer MM, Pestana K, Ahmed S, Elfaki M, Gamil E, Ahmed SM, et al. Distribution of Duffy Phenotypes among Plasmodium vivax Infections in Sudan. Genes. 2019;10(6):437.

18. World Health Organization. World Malaria Report (2015) WHO, Geneva.

19. Lo E, Lam N, Hemming-Schroeder E, Nguyen J, Zhou G, Lee M-C, et al. Frequent Spread of Plasmodium vivax Malaria Maintains High Genetic Diversity at the Myanmar-China Border, Without Distance and Landscape Barriers. The Journal of Infectious Diseases. 2017;216(10):1254–63.

20. Mahgoub H, Gasim GI, Musa IR, Adam I. Severe Plasmodium vivax malaria among sudanese children at New Halfa Hospital, Eastern Sudan. Parasites & Vectors. 2012;5(1):154.

21. Hupalo DN, Luo Z, Melnikov A, Sutton PL, Rogov P, Escalante A, et al. Population genomics studies identify signatures of global dispersal and drug resistance in Plasmodium vivax. Nature Genetics. 2016;48(8):953–8.

22. Parobek CM, Lin JT, Saunders DL, Barnett EJ, Lon C, Lanteri CA, et al. Selective sweep suggests transcriptional regulation may underlie Plasmodium vivax resilience to malaria control measures in Cambodia. Proceedings of the National Academy of Sciences. 2016;113(50).

504 23. Benavente ED, Ward Z, Chan W, Mohareb FR, Sutherland CJ, Roper C, et al.
505 Genomic variation in Plasmodium vivax malaria reveals regions under selective
506 pressure. Plos One. 2017;12(5).

507 24. Lima-Junior JDC, Pratt-Riccio LR. Major Histocompatibility Complex and Malaria:
508 Focus on Plasmodium vivax Infection. Frontiers in Immunology. 2016;7.

509 25. Kano FS, Souza AMD, Torres LDM, Costa MA, Souza-Silva FA, Sanchez BAM, et al.
510 Susceptibility to Plasmodium vivax malaria associated with DARC (Duffy antigen)
511 polymorphisms is influenced by the time of exposure to malaria. Scientific Reports.
512 2018;8(1).

513 26. Fola AA, Harrison GLA, Hazairin MH, Barnadas C, Hetzel MW, Iga J, et al. Higher
514 Complexity of Infection and Genetic Diversity of Plasmodium vivax Than
515 Plasmodium falciparum across all Malaria Transmission Zones of Papua New Guinea.
516 The American Journal of Tropical Medicine and Hygiene. 2017;:16–0716.

517 27. Orjuela-Sánchez P, Sá JM, Brandi MC, Rodrigues PT, Bastos MS, Amaratunga C, et
518 al. Higher microsatellite diversity in Plasmodium vivax than in sympatric Plasmodium
519 falciparum populations in Pursat, Western Cambodia. Experimental Parasitology.
520 2013;134(3):318–26.

521 28. Gray K-A, Dowd S, Bain L, Bobogare A, Wini L, Shanks G, et al. Population genetics
522 of Plasmodium falciparum and Plasmodium vivax and asymptomatic malaria in Temotu
523 Province, Solomon Islands. Malaria Journal. 2013;12(1):429.

524 29. Noviyanti R, Coutrier F, Utami RAS, Trimarsanto H, Tirta YK, Trianty L, et al.
525 Contrasting Transmission Dynamics of Co-endemic Plasmodium vivax and P.

526 falciparum: Implications for Malaria Control and Elimination. PLOS Neglected Tropical
527 Diseases. 2015;9(5).

528 30. Jennison C, Arnott A, Tessier N, Tavul L, Koepfli C, Felger I, et al. Plasmodium vivax
529 Populations Are More Genetically Diverse and Less Structured than Sympatric
530 Plasmodium falciparum Populations. PLOS Neglected Tropical Diseases. 2015;9(4).

531 31. Ferreira MU, Karunaweera ND, Silva-Nunes MD, Silva NSD, Wirth DF, Hartl DL.
532 Population Structure and Transmission Dynamics of Plasmodium vivax in Rural
533 Amazonia. The Journal of Infectious Diseases. 2007;195(8):1218–26.

534 32. Battle KE, Lucas TCD, Nguyen M, Howes RE, Nandi AK, Twohig KA, et al. Mapping
535 the global endemicity and clinical burden of Plasmodium vivax, 2000–17: a spatial and
536 temporal modelling study. The Lancet. 2019;394(10195):332–43.

537 33. Bereczky S, Färnert A, Mårtensson A, Gil JP. Short Report: Rapid Dna Extraction
538 From Archive Blood Spots On Filter Paper For Genotyping Of Plasmodium Falciparum.
539 The American Journal of Tropical Medicine and Hygiene. 2005;72(3):249–51.

540 34. Anderson TJC, Su X-Z, Bockarie M, Lagog M, Day KP. Twelve microsatellite markers
541 for characterization of Plasmodium falciparum from finger-prick blood samples.
542 Parasitology. 1999;119(2):113–25.

543 35. Karunaweera ND, Ferreira MU, Hartl DL, Wirth DF. Fourteen polymorphic
544 microsatellite DNA markers for the human malaria parasite Plasmodium vivax.
545 Molecular Ecology Notes. 2006;7(1):172–5.

546 36. Koepfli C, Mueller I, Marfurt J, Goroti M, Sie A, Oa O, et al. Evaluation ofPlasmodium
547 vivaxGenotyping Markers for Molecular Monitoring in Clinical Trials. The Journal of
548 Infectious Diseases. 2009;199(7):1074–80.

549 37. Haubold B, Hudson RR. LIAN 3.0: detecting linkage disequilibrium in multilocus data.
550 Bioinformatics. 2000;16(9):847–9.

551 38. Meirmans PG, Tienderen PHV. genotype and genodive: two programs for the analysis
552 of genetic diversity of asexual organisms. Molecular Ecology Notes. 2004;4(4):792–4.

553 39. Smouse PE, Peakall R. Spatial autocorrelation analysis of individual multiallele and
554 multilocus genetic structure. Heredity. 1999;82(5):561–73.

555 40. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using
556 multilocus genotype data. Genetics. 2000:155(2):945–59.

557 41. Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using
558 the software structure: a simulation study. Molecular Ecology. 2005;14(8):2611–20.

559 42. Rosenberg NA. distruct: a program for the graphical display of population structure.
560 Molecular Ecology Notes. 2003;4(1):137–8.

561 43. Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F, et al. Phylogeny.fr:
562 robust phylogenetic analysis for the non-specialist. Nucleic Acids Research.
563 2008;36(Web Server).

564 44. Rambaut, A. (2014) FigTree v1.4.2. Available at:
565 http://tree.bio.ed.ac.uk/software/figtree/

566 45. Schneider ADB, Ford CT, Hostager R, Williams J, Cioce M, Çatalyürek ÜV, et al.
567 StrainHub: A phylogenetic tool to construct pathogen transmission networks. 2019

568 46. Dray, S., Dufour, A., & Chessel, D. (2007). The ade4 Package – II: Two-Table and K-
569 Table Methods. R News, 7(2), 47–52.

570 47. Meijer JR, Huijbregts MAJ, Schotten KCGJ, Schipper AM. Global patterns of current
571 and future road infrastructure. Environmental Research Letters. 2018;13(6):064006.

572 48. Friedl, M., & Sulla-Menashe, D. MCD12Q1 MODIS/Terra+Aqua Land Cover Type
573 Yearly L3 Global 500m SIN Grid V006. NASA EOSDIS Land Processes DAAC. 2019
574 https://doi.org/10.5067/MODIS/MCD12Q1.006

575 49. NASA JPL. NASA Shuttle Radar Topography Mission Global 1 arc second. NASA
576 EOSDIS Land Processes DAAC. 2013.
577 https://doi.org/10.5067/MEaSUREs/SRTM/SRTMGL1.003

578 50. GDAL/OGR contributors. GDAL/OGR Geospatial Data Abstraction software Library.
579 Open Source Geospatial Foundation. 2020. https://gdal.org

580 51. Peterman WE. ResistanceGA: An R package for the optimization of resistance
581 surfaces using genetic algorithms. Methods in Ecology and Evolution. 2018;9(6):1638–
582 47.

583 52. Gunalan K, Niangaly A, Thera MA, Doumbo OK, Miller LH. Plasmodium vivax
584 Infections of Duffy-Negative Erythrocytes: Historically Undetected or a Recent
585 Adaptation? Trends in Parasitology. 2018;34(5):420–9.

586 53. Abdelraheem MH, Albsheer MMA, Mohamed HS, Amin M, Hamid MMA. Transmission
587 ofPlasmodium vivaxin Duffy-negative individuals in central Sudan. Transactions of The
588 Royal Society of Tropical Medicine and Hygiene. 2016;110(4):258–60.

589 54. Byrne N. Urban malaria risk in sub-Saharan Africa: Where is the evidence? Travel
590 Medicine and Infectious Disease. 2007;5(2):135–7.

591 55. Robert V, Duchemin J-B, Macintyre K, Warren M, Keating J, Beier JC, et al. Malaria
592 Transmission In Urban Sub-Saharan Africa. The American Journal of Tropical
593 Medicine and Hygiene. 2003;68(2):169–76.

594 56. Pacheco MA, Lopez-Perez M, Vallejo AF, Herrera S, Arévalo-Herrera M, Escalante
595 AA. Multiplicity of Infection and Disease Severity in Plasmodium vivax. PLOS
596 Neglected Tropical Diseases. 2016;10(1).

597 57. Kasehagen LJ, Mueller I, Kiniboro B, Bockarie MJ, Reeder JC, Kazura JW, et al.
598 Reduced Plasmodium vivax Erythrocyte Infection in PNG Duffy-Negative
599 Heterozygotes. PLoS ONE. 2007;2(3).

600 58. Vallejo AF, Martínez NL, González IJ, Arévalo-Herrera M, Herrera S. Evaluation of
601 the Loop Mediated Isothermal DNA Amplification (LAMP) Kit for Malaria Diagnosis in
602 P. vivax Endemic Settings of Colombia. PLoS Neglected Tropical Diseases. 2015;9(1).

603 59. Bendixen M, Msangeni H, Pedersen B, Shayo D, Bedker R. Diversity of Plasmodium
604 falciparum populations and complexity of infections in relation to transmission intensity
605 and host age: a study from the Usambara Mountains, Tanzania. Transactions of the
606 Royal Society of Tropical Medicine and Hygiene. 2001;95(2):143–8.

607 60. Vafa M, Troye-Blomberg M, Anchang J, Garcia A, Migot-Nabias F. Multiplicity of
608 Plasmodium falciparum infection in asymptomatic children in Senegal: relation to
609 transmission, age and erythrocyte variants. Malaria Journal. 2008;7(1).

610 61. Chan LJ, Dietrich MH, Nguitragool W, Tham WH. Plasmodium vivax Reticulocyte
611 Binding Proteins for invasion into reticulocytes. Cellular Microbiology. 2019;22(1).

612 62. Ovchynnikova E, Aglialoro F, Bentlage AEH, Vidarsson G, Salinas ND, Lindern MV,
613 et al. DARC extracellular domain remodeling in maturating reticulocytes explains
614 Plasmodium vivax tropism. Blood. 2017;130(12):1441–4.

615 63. Cheng Y, Lu F, Wang B, Li J, Han J-H, Ito D, et al. Plasmodium vivax GPI-anchored
616 micronemal antigen (PvGAMA) binds human erythrocytes independent of Duffy
617 antigen status. Scientific Reports. 2016;6(1).

618 64. Gupta S, Singh S, Popovici J, Roesch C, Shakri AR, Guillotte-Blisnick M, et al.
619 Targeting a Reticulocyte Binding Protein and Duffy Binding Protein to Inhibit
620 Reticulocyte Invasion by Plasmodium vivax. Scientific Reports. 2018;8(1).

621 65. Talha AA, Pirahmadi S, Mehrizi AA, Djadid ND, Nour BY, Zakeri S. Molecular genetic
622 analysis of Plasmodium vivax isolates from Eastern and Central Sudan using pvcsp
623 and pvmsp-3α genes as molecular markers. Infection, Genetics and Evolution.
624 2015;32:12–22.

625 66. Talha A. Low Prevalence of Plasmodium vivax - Plasmodium falciparum Mixed -
626 Infection in Patients from Central and Eastern Part of Sudan: Implication for Case
627 Management in Sudan. International Journal of TROPICAL DISEASE & Health.
628 2014;4(8):887–95.

629 67. Menegon M, Durand P, Menard D, Legrand E, Picot S, Nour B, et al. Genetic diversity
630 and population structure of Plasmodium vivax isolates from Sudan, Madagascar,
631 French Guiana and Armenia. Infection, Genetics and Evolution. 2014;27:244–9.

632 68. Legendre P, Fortin MJ, Borcard D. Should the Mantel test be used in spatial analysis?
633 Methods in Ecology and Evolution. 2015;6(11):1239–47.

634

635 **Tables**

636 **Table 1. Duffy status, linkage disequilibrium, and complexity of infection among *P.*
637 *vivax* samples by study site.** Asterisk denotes a significant level at $P<0.05$.

638 **Table 2. Comparison of genetic diversity based on microsatellite markers in Duffy-**
639 **negative and Duffy-positive *P. vivax* infections.**

640

641 **Figures**

642 **Figure 1. Study sites with *P. vivax* incidence heatmap for reference.** Each point
643 corresponds to one of the six sample locations, and the gray lines represent the borders
644 of Ethiopia and Sudan. The incidence layer was modeled by Battle *et al.* [32] and
645 represents the estimated cases per 1000 people in 2017.

646 **Figure 2. A) Neighbor-joining tree of genetic distance between populations; B) Two-**
647 **dimensional MDS of genetic distance between populations.** Shape of the icons
648 indicates host Duffy status, and color indicates country.

649 **Figure 3. Bayesian inferences of the *K* cluster estimated by STRUCTURE for the *P. vivax* samples, according to Duffy status and location.** The most probable clusters are labeled by different colors, and individuals are represented as columns.

652 **Figure 4. Source Hub Ratio (SHR) networks generated from A) samples grouped according to country and host Duffy status and B) samples grouped according to site.** The size of each node is proportional to the node's SHR value, and the number on each node is the number of samples in that group. Node colors were randomly assigned to each unique SHR value present in the plot to improve readability. Similarly, arrow colors are paired with the color of the node from which they begin to aid with interpretation. The weight of each arrow is proportional to the number of transitions between those two nodes. The position of nodes is arbitrary and is not equivalent to the position of sample sites in geographic space.

661 **Figure 5. Resistance surfaces fit to genetic distances based on A) land cover alone and B) a combination of road density and elevation.** Higher values indicate greater obstruction to gene flow. Study sites are denoted by points, and the lines indicate the borders of Ethiopia and Sudan.

665

666 **Supplementary files**

667 **Supplementary Table 1.** Microsatellite genotypes of *Plasmodium vivax* samples included in the present study.

669 **Supplementary Table 2.** Fit criteria used to rank the resistance surfaces, and the slope
670 and significance of the relationship between resistance and genetic distance for each
671 surface.

672 **Supplementary Figure 1.** The environmental datasets included in the study: A) road
673 density, as estimated by Meijer *et al*. [47]; B) elevation above sea level, as measured by
674 the Shuttle Radar Topography Mission [48]; and C) IGBP land cover classification,
675 obtained from the MCD12Q1 dataset [49]. Study sites are denoted by points, and the
676 lines indicate the borders of Ethiopia and Sudan.

677 **Supplementary Figure 2.** (A) Genetic relatedness among *P. vivax* populations with a
678 large sample size (i.e., *P. vivax* from Duffy-negatives at sites GJ, NH, and KH was
679 excluded) organized by country of origin, shown by color, and Duffy status, shown by
680 shape. (B) Two-dimensional MDS of genetic distance between the same populations (i.e.,
681 *P. vivax* from Duffy-negatives at sites GJ, NH, and KH was excluded).