

Clinical malaria incidence following an outbreak in Ecuador was predominantly associated with *Plasmodium falciparum* with recombinant variant antigen gene repertoires

Shazia Ruybal-Pesántez^{1§}, Fabian E. Sáenz², Samantha Deed^{1,3}, Erik K. Johnson⁴, Daniel B. Larremore^{5,6},
Claudia A. Vera-Arias^{2#}, Kathryn E. Tiedje^{1,3}, Karen P. Day^{1,3*}

¹ School of BioSciences/Bio21 Institute, University of Melbourne, Australia

² Centro de Investigación para la Salud en América Latina, Facultad de Ciencias Exactas y Naturales, Pontificia Universidad Católica del Ecuador, Av. 12 de octubre 1076, Apartado: 17-01-2184, Quito, Ecuador

³ Department of Microbiology and Immunology, University of Melbourne, Bio21 Institute, Melbourne, Australia

⁴ Department of Applied Mathematics, University of Colorado Boulder, Boulder, Colorado, USA

⁵ Department of Computer Science, University of Colorado Boulder, Boulder, Colorado, USA

⁶ BioFrontiers Institute, University of Colorado Boulder, Boulder, Colorado, USA

[§] *Current affiliations*: Population Health and Immunity Division, Walter and Eliza Hall Institute of Medical Research, Melbourne, Australia; Department of Medical Biology, University of Melbourne, Australia; Burnet Institute, Australia

[#] *Current affiliation*: Eck Institute for Global Health, University of Notre Dame, Notre Dame, IN, 46556, USA

**Corresponding author*: Karen P. Day; karen.day@unimelb.edu.au

Abstract

To better understand the factors underlying the continued incidence of clinical episodes of falciparum malaria in E-2020 countries targeting elimination, we have characterised *Plasmodium falciparum* disease transmission dynamics after a clonal outbreak on the northwest coast of Ecuador over a period of two years. We apply a novel, high-resolution genotyping method, the “varcode” based on a single PCR to fingerprint the DBL α region of the 40-60 members of the variant surface antigen-encoding *var* multigene family. *Var* genes are highly polymorphic within and between genomes, with *var* repertoires rapidly evolving by outcrossing during the obligatory sexual phase of *P. falciparum* in the mosquito. The continued incidence of clinical malaria after the outbreak in Ecuador provided a unique opportunity to use varcodes to document parasite microevolution and explore signatures of local disease transmission on the time scale of months to two years post-outbreak. We identified nine genetic varcodes circulating locally with spatiotemporal parasite genetic relatedness networks revealing that diversification of the

36 clonal outbreak parasites by sexual recombination was associated with increased incidence of clinical
37 episodes of malaria. Whether this was due to chance, immune selection or sexual recombination per se is
38 discussed. Comparative analyses to other South American parasite populations where *P. falciparum*
39 transmission remains endemic elucidated the possible origins of Ecuadorian *varcodes*. This analysis
40 demonstrated that the majority of clinical cases were due to local transmission and not importation.
41 Nonetheless, some of the *varcodes* that were unrelated to the outbreak *varcode* were found to be
42 genetically related to other South American parasites. Our findings demonstrate the utility of the *varcode*
43 as a high-resolution surveillance tool to spatiotemporally track disease outbreaks using variant surface
44 antigen genes and resolve signatures of recombination in an E-2020 setting nearing elimination.

45 Introduction

46 In 2016 the WHO identified 21 countries, known as the E-2020 countries, having the potential to
47 eliminate local transmission of malaria by 2020¹, with seven of them in Latin America at the time
48 (Paraguay and El Salvador have since been declared malaria-free). Malaria transmission in many of these
49 E-2020 countries is epidemic/unstable with risks of outbreaks and resurgent malaria²⁻⁴. The elimination
50 target requires a local reduction to zero incidence of indigenous cases. As progress is made towards
51 elimination in these countries, it becomes critical to understand risk factors for disease transmission while
52 maximizing limited funds and resources to examine individual cases. Ecuador is one of the five remaining
53 E-2020 countries in Latin America. National malaria elimination efforts have largely focused on tropical
54 areas, specifically the northwest coast and the Amazon region. These areas border non E-2020 countries,
55 Colombia and Peru, that still have endemic transmission. Thus, rather than low transmission, Ecuador has
56 epidemic/unstable *P. falciparum* transmission with *P. vivax* the dominant species causing clinical malaria
57 infections.

58
59 Clinical cases caused by *P. falciparum* in Ecuador are mostly concentrated in the northwest coast, where a
60 steady decrease in annual reported clinical cases was observed from 2009 (161 cases) until late 2012 (44
61 cases)⁵⁻⁷. A localized outbreak from November 2012 until November 2013 in Esmeraldas City
62 documented 151 *P. falciparum* cases, the majority of cases (76%) in the northwest coast and Amazon
63 region during this time period⁸. Population genetic analyses using microsatellite markers demonstrated
64 that this outbreak was caused by the clonal expansion (i.e., single-source) of a residual parasite lineage
65 previously reported on the north coast of Peru and on the coast of Ecuador⁸. Immediately after the
66 outbreak, 31 cases were reported in 2014 and 189 cases in 2015^{5,9,10}. It was unclear whether the clonal

67 outbreak had contributed to *P. falciparum* clinical cases since 2013, with the factors underlying sustained
68 disease transmission in Ecuador poorly understood. Moreover, Ecuador was categorized as “off-track” in
69 the most recent E-2020 progress report and has not met the elimination target due to malaria
70 resurgence^{11,12}. This points to an urgent need to better understand disease transmission patterns and
71 whether clinical cases are due to imported or locally-acquired parasites.

72
73 Molecular surveillance of malaria in Latin America has focused on non E-2020 countries, e.g. Colombia,
74 Brazil, Peru and Panama where malaria is endemic and transmission is moderate to low^{8,13–19}. This
75 surveillance has relied on genotyping neutral molecular markers such as microsatellites or single-
76 nucleotide polymorphisms. However, neutral markers, considered relatively slowly evolving, may not
77 provide enough discriminatory resolution to define parasite evolution in relevant epidemiological time (1-
78 2 years) in areas of low or epidemic and unstable transmission^{20,21}. Genes under selection can provide an
79 alternative but complementary view of microevolution and they are key to examine population
80 adaptation²².

81
82 In this context, the microbiological paradigm for the surveillance of diverse pathogens is to study
83 transmission dynamics using the genes encoding the major surface antigen to reflect pathogen
84 transmission dynamics in relation to immune selection^{23,24}. The failure to observe this paradigm in the
85 malaria field has stemmed from both the misconception that neutral markers track transmission
86 dynamics in relation to immunity and the fact that the genetics of the *var* genes encoding the major
87 surface antigen of the blood stages (PfEMP1) are considered “too complex”. Single copy antigen genes
88 under much less selection have become an alternative²⁵. Instead, we have embraced the complexity of
89 the *var* system to understand the population genetics of *var* genes^{23,26–31} in multiple locations in Africa,
90 South America, and globally to present here a way to explore transmission dynamics through the lens of
91 these immune evasion genes.

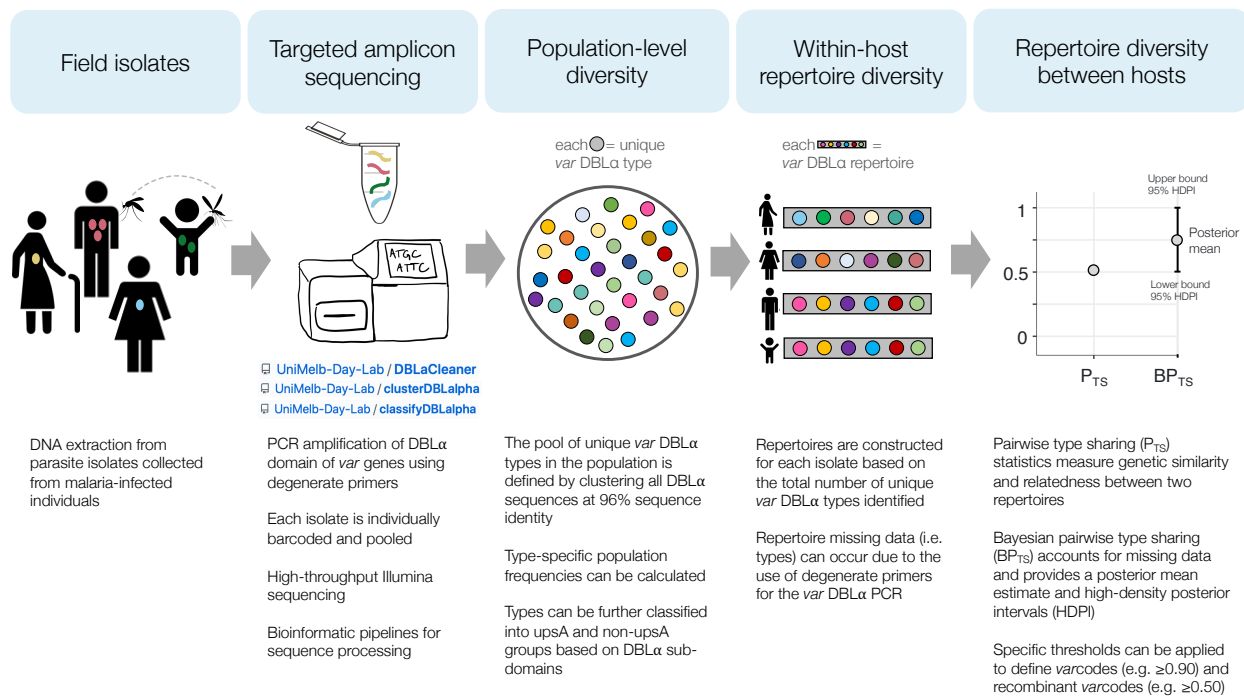
92
93 Each genomic *var* repertoire of *P. falciparum* consists of 40 to 60 members of the *var* multigene family
94 with evidence that South American parasites may have fewer *var* genes, e.g. 42 *var* genes (excluding
95 *var2csa*) reported in the Honduran laboratory strain HB3 after whole genome sequencing and assembly
96^{32,33}. Our analysis of a South American dataset with a global dataset also showed fewer *var* genes in South
97 American isolates as did a Brazilian data set^{31,34}. As *var* genes lie on almost all of the 14 chromosomes of
98 the haploid *P. falciparum* genome³⁵ they undergo independent assortment during conventional meiosis

99 to create new *var* repertoires as recombinants of parental *var* repertoires. During infection, differential
100 switching between these genes in a repertoire enables immune evasion in the human host^{36,37} thereby
101 regulating parasite density and risk of clinical infection³⁸. Indeed, several serologic studies examining anti-
102 PfEMP1 antibody responses in naturally-infected children have shown that the acquisition of anti-PfEMP1
103 protective immunity is variant-specific and susceptibility to clinical episodes is dependent on gaps in the
104 individual's pre-existing antibody repertoire³⁸⁻⁴⁷.

105
106 A body of work defining the population genomics of the *var* multigene family by ourselves and others⁴⁸
107 has demonstrated that *var* genes provide a promising biomarker for surveillance over epidemiological
108 time scales (1-10 years). *Var* DBL α sequences can resolve population structure at global, continental,
109 regional and local scales in areas of varying endemicity in Africa, Oceania, and South America^{23,26-31,49}.
110 Characterization of these major surface antigen genes also provides information about circulating
111 repertoires of variants to enable measurement of immune responses. Translation of *var* population
112 genetics in different epidemiological settings leads us to propose a new surveillance approach we call
113 *var*coding. This is essentially a fingerprinting method to describe the diversity of these genes within and
114 between human hosts as well as repertoire diversification by sexual recombination, with Bayesian
115 statistics to account for missing data (Fig. 1).

116
117 Defining the 2012-2013 clinical disease outbreak in Ecuador as a baseline, we applied the *var*code
118 methodology to characterize the transmission dynamics of clinical *P. falciparum* cases during and up to
119 two years after the outbreak. Reconstruction of parasite genetic relatedness networks in space and time
120 revealed that parasites with recombinant *var* repertoires were predominantly associated with the
121 continued incidence of clinical cases after the outbreak. Further comparative analyses to published data
122 from South American isolates²⁸ elucidated possible origins of Ecuadorian parasites, demonstrating that
123 the majority of clinical cases were due to local transmission and not importation. Our findings also
124 demonstrate the translational application of the *var*code as a high-resolution tool to detect,
125 spatiotemporally track immune evasion genes as well as resolve signatures of recombination in E-2020 or
126 low transmission settings.

127



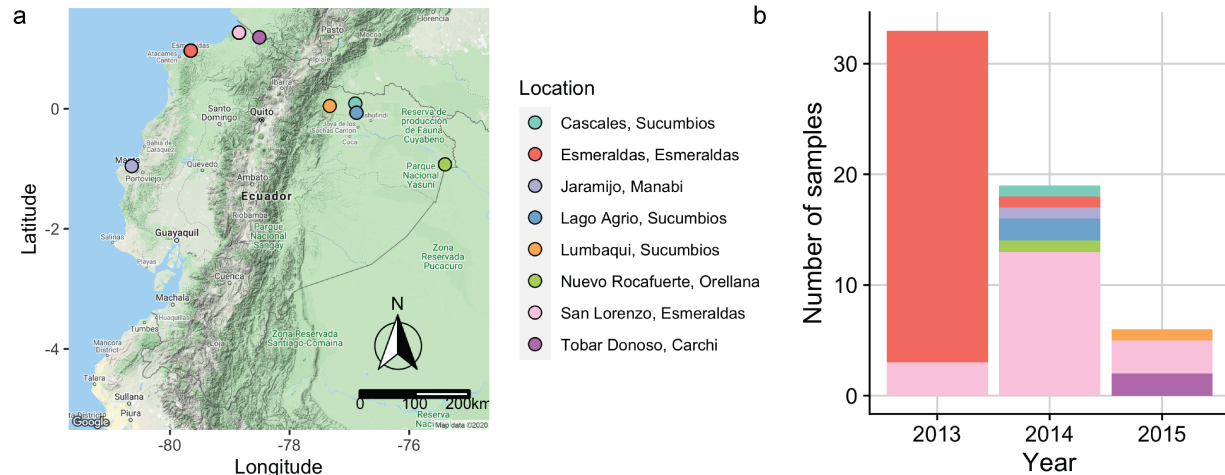
128

129 **Figure 1. Schematic diagram of the varcoding approach.** For more details about each step, see Methods.

130 Results

131 *Varcoding* was completed for 58 *P. falciparum* isolates that were collected between 2013 and 2015 in
 132 endemic areas of the northwest coast and Amazon region of Ecuador (Fig. 2) from individuals of all ages
 133 presenting with malaria symptoms and confirmed *P. falciparum*-positive by microscopy or rapid
 134 diagnostic tests. These isolates represent 21% of the total cases reported in 2013, 61% in 2014 and 3% in
 135 2015 (60% of the cases reported in January, 7% in May and 8% in November). For more details on the
 136 sampling scheme and bioinformatics workflow see Methods and Supplementary Fig. S1. Overall, we
 137 identified 195 unique *var* DBL α variants or types from 2,141 DBL α sequences from the 58 isolates,
 138 representative of the diversity circulating in Ecuadorian *P. falciparum* populations, as indicated by
 139 accumulation curves approaching a plateau (Fig. S2).

140



141
 142 **Figure 2. Sampling of *P. falciparum* isolates across endemic areas of Ecuador and over a period of three years (2013 – 2015).** (a) Map of Ecuador
 143 depicting the sampling locations during the study. (b) Bar plot showing the number of *P. falciparum* positive samples collected in each year and
 144 their respective sampling locations. Every location is indicated with a different color.
 145

146 Defining *var*codes circulating in Ecuador

147 To define distinct *var*codes circulating locally in Ecuador, we estimated parasite genetic relatedness by
 148 calculating the similarity index, Pairwise Type Sharing (P_{TS}) and a Bayesian version of the same (Bayesian
 149 pairwise type sharing; BP_{TS} , see Methods). BP_{TS} estimates include statistical uncertainty due to differences
 150 in isolate repertoire size (i.e. the number of DBL α types identified in each isolate repertoire), enabling a
 151 rigorous examination of relatedness. The median repertoire size in Ecuadorian isolates was 37 DBL α types
 152 (range: 11-43 types, Table S1), which is in line with the expected number of *var* genes in the Honduran
 153 laboratory reference strain HB3 ($n=42$ *var* genes excluding *var2csa*,³³). Nonetheless, we used a uniform
 154 prior distribution for repertoire sizes between 40 and 50 types, which accounted for the possibility of
 155 missing data in total isolate repertoire sizes using degenerate primers for *var* DBL α PCR. For a further
 156 analysis to validate our PCR and sequencing methodology, including a further sub-classification of DBL α
 157 types into upsA and non-upsA groups and their respective isolate genomic proportions, see Methods and
 158 Supplementary Text 1.

159
 160 We constructed genetic relatedness networks at a threshold of $P_{TS} \geq 0.90$ to identify putatively identical
 161 genomes within the margin of error of detection of all DBL α types in an isolate. We confirmed this
 162 definition by comparing *var*codes derived from P_{TS} to those derived from posterior mean BP_{TS} estimates
 163 (Supplementary Fig. S3), and by examining the lower and upper bounds of the 95% highest density
 164 posterior intervals (HDPIs). This revealed nine genetically distinct *var*codes in this study with *var*code4,
 165 *var*code5, *var*code9 seen only once and at the other extreme 36 isolates had *var*code1 (Fig. 3a and

166 Supplementary Fig. S4a). We were unable to confirm whether any of the Ecuadorian *varcodes* were
167 circulating pre-outbreak as samples were not available. As expected, the outbreak *varcode1* was clonal
168 (indicated by HDPIs that included 1; Fig. S5d), as previously demonstrated by microsatellite genotyping⁸. It
169 is worth noting that even though we only sampled 30 *P. falciparum* isolates in Esmeraldas City from 140
170 (21%) outbreak cases in 2013, the fact that the disease outbreak was clonal and represented 91% of the
171 154 total cases on the northwest coast and Amazon region during 2013, means that we characterized a
172 representative sample of the cases from the outbreak and during that year.

173
174 For each pairwise comparison, we calculated the HDPI width as an estimate of uncertainty (analogous to
175 calculating the width of a frequentist confidence interval) and found there was high confidence (HDPI
176 width ≤ 0.2) for 94.6% (632/668) of pairwise BP_{TS} estimates (Supplementary Fig. S4b). Low confidence
177 pairwise BP_{TS} estimates (HDPI width > 0.2 ; max = 0.33) were found only between the two *varcode1* isolates
178 with the smallest repertoire sizes (11 and 19 DBL α types; 36/36 comparisons). This lower statistical
179 confidence follows naturally from under sampling of those isolates' DBL α types. After aggregating the
180 within-*varcode* pairwise comparisons, there was high confidence in the definition of all *varcodes*
181 (Supplementary Fig. S5a-b).

182
183 For comparison of relatedness by microsatellite genotyping, we calculated Pairwise Allele Sharing (P_{AS})
184 using previously published microsatellite data for the same isolates^{8,50}. Not surprisingly, 7 microsatellites
185 did not resolve the nine *varcodes* but only identified four, three of which were single parasite isolates, at
186 the threshold of $P_{AS} \geq 0.90$ (Supplementary Fig. S6a). Because two of the seven microsatellite markers
187 were fixed in the population^{8,50}, we also generated the relatedness networks at the threshold of $P_{AS} \geq$
188 0.80, which allowed for an error of detection of 1 microsatellite allele. However, the majority of isolates
189 clustered at this threshold and only three *varcodes* were resolved (Supplementary Fig. S6b). There were
190 several pairwise comparisons (14%, 72/541 by P_{TS} and 6%, 31/541 by BP_{TS}) where *P. falciparum* isolates
191 were found to be identical by microsatellites ($P_{AS} = 1$) but much less related by *var* (P_{TS} or $BP_{TS} \leq 0.50$,
192 Supplementary Fig. S7). Nevertheless, the relationship between P_{TS} and P_{AS} scores was positively
193 correlated (Pearson's correlation coefficient = 0.757, $p < 0.001$, Supplementary Fig. S7a). Similarly, the
194 relationship between BP_{TS} and P_{AS} estimates was also positively correlated (Pearson's correlation
195 coefficient = 0.779, $p < 0.001$, Supplementary Fig. S7b).

196

197 Persistent association of clinical cases with the same *varcodes* over time and large
198 distances

199 We next explored whether parasites with the same *varcodes* caused clinical episodes after the outbreak.
200 Persistent disease transmission over time (Fig. 3b) and large distances (Fig. 3e-f) was found as the same
201 *varcodes* were identified post outbreak and sometimes in different sampling locations. The median time
202 between first and last identification of the same *varcodes* with any clinical case was 216 days or
203 approximately 7 months (range = 190 – 823 days) during the study period. For example, the outbreak
204 *varcode*1 identified in Esmeraldas City was also identified in San Lorenzo, Esmeraldas (~150km from
205 Esmeraldas city) in 2013, then Cascales, Sucumbios (>300km away) in 2014, and then in Tobar Donoso,
206 Carchi (~150km away) in 2015 (Fig. 3b). This is in line with an earlier study using microsatellites that
207 showed that *P. falciparum* isolates from the outbreak were genetically related to an Ecuadorian parasite
208 isolate collected around 1990⁸. This result was also consistent with reports in Peru⁵¹ and Colombia^{18,52}
209 where identical parasites were shown to be circulating up to five and eight years after first identification,
210 respectively.

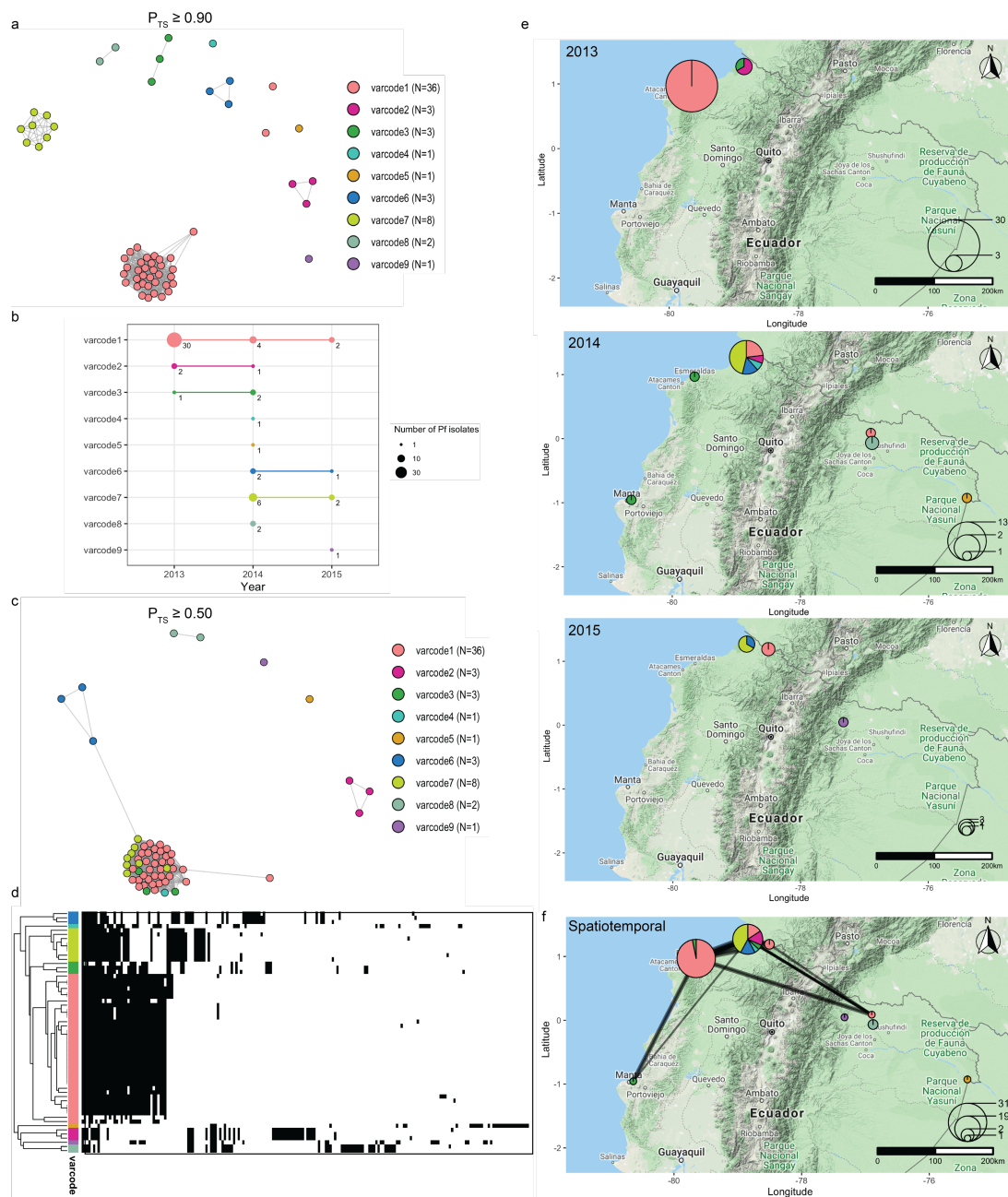


Figure 3. Reconstructing the disease transmission dynamics of *P. falciparum* reveals signatures of local transmission and sexual recombination after the outbreak. (a) A network visualization of the genetic relatedness of *P. falciparum* isolates at the threshold of $P_{TS} \geq 0.90$ to define *varcodes* (see Methods). Every node represents a *P. falciparum* isolate and an edge represents the P_{TS} value between two particular nodes/isolates. Isolates that cluster together (i.e., connected by edges) are considered to have the same *varcode*. Each color represents a different *varcode*. Two *P. falciparum* isolates belonging to *varcode1* appear as outliers in the network due to undersampling of their DBL α types. N refers to the number of isolates. For comparison to BP_{TS} estimates see Supplementary Fig. S3-5 and S8. (b) The number of *P. falciparum* isolates within each *varcode* (i.e. size of circle) in each year and their persistence over time. The three *varcodes* (*varcode1*, *varcode2*, *varcode3*) identified in 2013 were identified again in 2014, and *varcode1* was identified again in 2015. Two *varcodes* identified in 2014 (*varcode6*, *varcode7*) were also identified in 2015. (c) A network visualization of the genetic relatedness of *varcodes* at the threshold of $P_{TS} \geq 0.50$ to discriminate between recombinant *varcodes* and genetically distinct *varcodes*. Isolates/*varcodes* that cluster together (i.e., connected by edges) represent recombinants. (d) A clustered heatmap showing the genetic profiles of each *P. falciparum* isolate with rows representing each isolate and columns representing each DBL α type. Black and white denote the presence and absence of each type, respectively. Isolates that clustered together were more genetically similar (i.e., the same DBL α types were present). Similarly, *varcodes* that clustered together were more genetically similar (e.g. recombinants). (e) The spatial distribution of *varcodes* in each year, with the size of the circle representing the number of *P. falciparum* isolates sampled in a given location in each year and the pie chart depicting the proportion of each *varcode* identified. (f) Spatiotemporal genetic relatedness network of *varcodes* between 2013-2015. Every node represents a sampling location, the size of the circle represents the number of *P. falciparum* isolates

211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227

228 sampled in each location, the pie chart depicts the proportion of each *varcode* identified, and the weighted edges show genetically-related
229 *varcodes* ($P_{TS} \geq 0.50$). (Online version: GIF animation depicts the spatiotemporal transmission dynamics chronologically during and after the
230 outbreak).
231

232 Spatiotemporal genetic relatedness networks reveal signatures of recombination

233 We next reconstructed spatiotemporal genetic relatedness networks to explore signatures of local
234 disease transmission and parasite microevolution after the outbreak. We applied a P_{TS} threshold of ≥ 0.50
235 to detect recombinant and/or genetically-related *varcodes*. By using the date and location where each
236 isolate was collected, we animated the network chronologically to examine spatiotemporal disease
237 transmission dynamics during the outbreak in 2013 and post-outbreak in 2014 and 2015 (Fig. 3c, 3f, GIF).
238 This analysis resolved parasite microevolution relevant to disease transmission after the outbreak and
239 revealed that four *varcodes* (*varcode3*, *varcode4*, *varcode6*, *varcode7*) were recombinants of the
240 outbreak *varcode1* (Fig. 3c). The remaining *varcodes* did not cluster in the network and were genetically
241 distinct. These patterns were confirmed using BP_{TS} estimates based on posterior means (Supplementary
242 Fig. S8a) and their corresponding 95% HDPIs (Supplementary Fig. S5c-f and S8b-c). For the *varcodes* that
243 were recombinants and/or genetically-related (i.e., cluster in the networks in Fig. 3c and Supplementary
244 Fig. S8a), 70.9% (433/611) of the HDPIs were confidently above the threshold of $BP_{TS} \geq 0.50$
245 (Supplementary Fig. S5d). For *varcodes* that were genetically-distinct (i.e., do not cluster in the networks),
246 98.4% (368/374) of the HDPIs were confidently below the threshold of $BP_{TS} \leq 0.50$ (Supplementary Fig.
247 S5f). As expected, the relatedness networks based on the 7 microsatellites did not resolve these patterns
248 since all isolates were connected at the threshold of $P_{AS} \geq 0.50$ and genetic relatedness was
249 overestimated (Supplementary Fig. S9a). Indeed, we were not always able to discriminate recombinants
250 from genetically distinct parasites since several pairwise comparisons (25%, 307/1232 by P_{TS} and 16%,
251 196/1232 by BP_{TS}) that resulted in $P_{AS} \geq 0.50$ were found to be less related by *var* (P_{TS} or $BP_{TS} \leq 0.50$)
252 (Supplementary Fig. S7). Furthermore, the majority of recombinants were classified as essentially clonal
253 since they clustered together at the threshold of $P_{AS} \geq 0.80$ (allowing for differences in one microsatellite
254 allele, Supplementary Fig. S6b).

255
256 We were able to detect that *varcode3* was in fact a recombinant of the outbreak *varcode1*, even though
257 it was sampled in San Lorenzo during the same time period as the ongoing outbreak in 2013 (Fig. 3c, 3e).
258 With the available data we were unable to confirm whether *varcode3* was ‘imported’ to San Lorenzo
259 from Esmeraldas City. The most plausible explanation is that the recombination event occurred in San
260 Lorenzo during the outbreak. Parasites with recombinant *varcode3* caused clinical episodes in multiple

261 locations in 2014 (Fig. 3e). The other recombinant *var*codes (*varcode4*, *varcode6*, *varcode7*) were
262 identified after the outbreak in 2014 or 2015. Whether the recombinant *var*codes resulted from sexual
263 recombination events between outbreak *varcode1* and genetically distinct parasites that were already
264 circulating at low levels and/or in asymptomatic reservoirs in Ecuador (e.g. ⁵) or those that were
265 previously imported could not be ascertained from the current study population. Interestingly, all the
266 recombinant *var*codes were identified in San Lorenzo indicating this area is a transmission hot spot and a
267 reservoir of parasites with diverse *var* repertoires (Fig. 3e).

268
269 To further visualize the genetic profiles of *var*codes and confirm the observed recombination signatures,
270 we constructed a clustered heatmap based on the presence and absence of the 195 DBL α types across all
271 isolates, such that isolates with similar genetic profiles cluster together (Fig. 3d). This analysis confirmed
272 that in the case of isolates identified as recombinants of the outbreak *varcode1* (i.e., *varcode3*, *varcode4*,
273 *varcode6*, *varcode7*), a proportion of outbreak DBL α types as well as different DBL α types were present.
274 This is consistent with the generation of a new *varcode* through outcrossing between two genomes with
275 different *var* repertoires by conventional meiosis (i.e., approximately $\geq 50\%$ sharing of DBL α types). By
276 contrast, the genetic profiles of *varcodes* 2, 5, 8, and 9 had different DBL α types, leading to the definition
277 as parasites with genetically distinct *var*codes. When using 7 microsatellites, these recombination
278 signatures in the genetic profiles were not identified (Supplementary Fig. S9b).

279
280 *Parasites with recombinant varcodes most frequently caused P. falciparum clinical*
281 *episodes following the 2012-13 outbreak*

282 Prior to the outbreak there had been a steady decline in clinical cases, however, increased incidence of
283 disease occurred after the outbreak. Therefore, we analyzed trends in the epidemiology of *P. falciparum*
284 cases that occurred post-outbreak to understand if there were any key risk factors. Of the 25 individuals
285 in our study that had a clinical *P. falciparum* episode after the outbreak in 2014 or 2015, we had age data
286 for 18 individuals (72%, Table 1). There was no significant difference (Kruskal-Wallis test, $p = 0.65$) in the
287 median age of individuals experiencing clinical episodes caused by parasites with the outbreak *varcode1*
288 (19 years, range = 19 – 57 years, N = 3 patients), a recombinant of *varcode1* (i.e., *varcodes* 3, 4, 6, or 7)
289 (25 years, range = 17 – 58 years, N = 13 patients), or by a different *varcode* (34.5 years, range = 32 – 37
290 years, N = 2 patients). A greater diversity of *varcodes* (1-9 inclusive) was associated with incidence of
291 clinical malaria post-outbreak than during the 2013 outbreak (*varcodes* 1,2,3). Indeed in 2014, 79% of the

292 cases sampled were caused by either parasites with the outbreak *varcode1* (21%) or with any of the four
293 recombinants of *varcode1* (58%). The trend was similar in 2015 with 83% of the cases sampled caused by
294 either parasites with the outbreak *varcode1* (33%) or with any of the four recombinants of *varcode1*
295 (50%), although our sampling of reported cases in 2015 after January was limited. Importantly, overall we
296 found that 80% of the cases we sampled after the outbreak were caused by either parasites with the
297 outbreak *varcode1* (24%) or with any of the four recombinants of *varcode1* (56%), especially with
298 *varcode7*. Thus, disease transmission after the outbreak in 2014 and into 2015 was predominantly
299 associated with parasites with recombinant *var* repertoires of the outbreak clone.

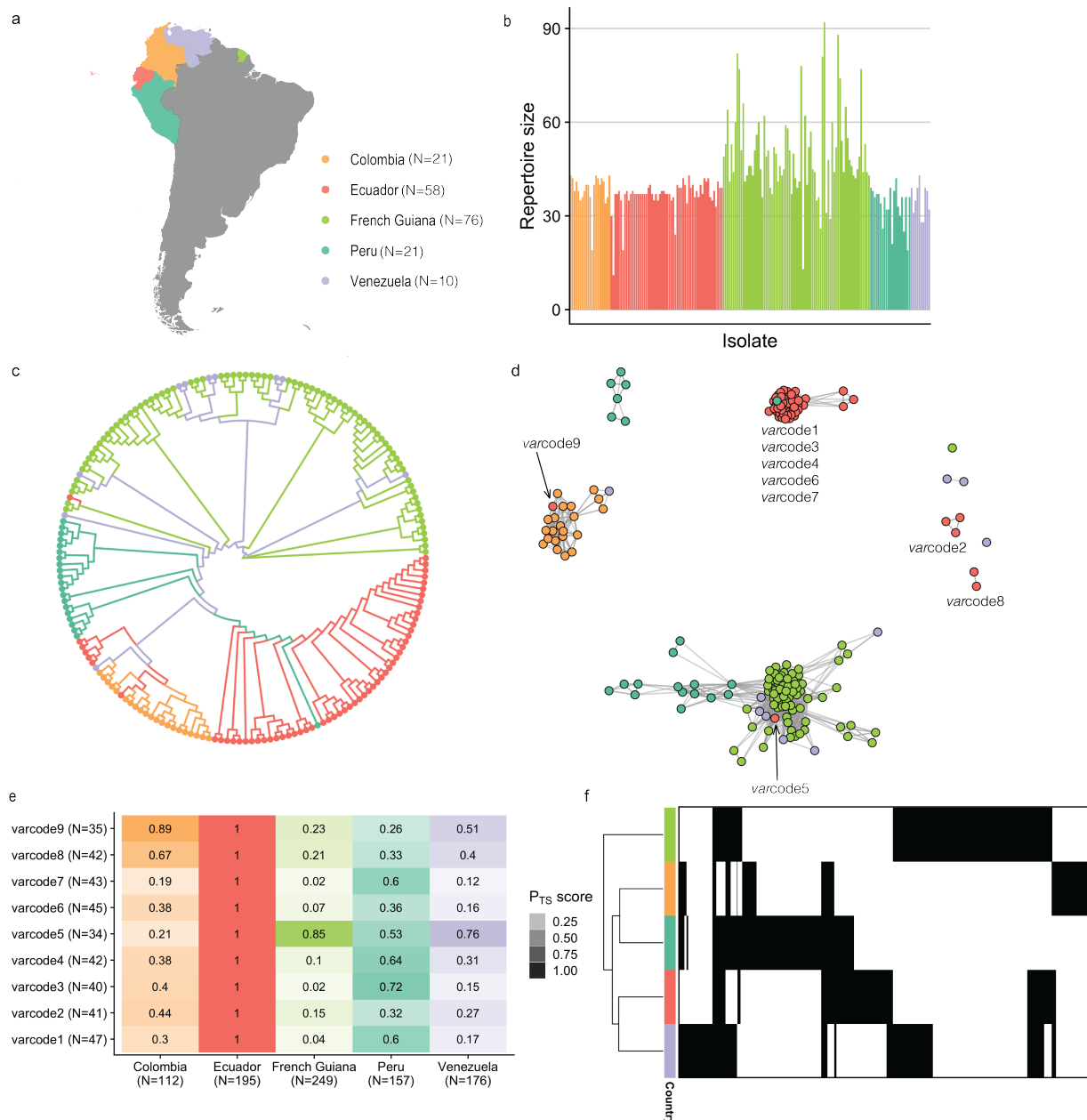
300

301 [Comparative analyses to South America: elucidating possible origins of Ecuadorian](#) 302 [varcodes](#)

303 We next examined the possible origins of the *varcodes* circulating locally in Ecuador by comparing to
304 published data²⁸ from 128 *P. falciparum* isolates collected from neighboring South American *P. falciparum*
305 populations (Fig. 4a). When we combined the sequences from Ecuadorian isolates with those previously
306 sequenced, we identified 543 unique *var* DBL α types in South America (compared to 458 in ²⁸). This is
307 largely representative of the *var* diversity circulating in South American *P. falciparum* populations, as
308 indicated by sampling accumulation curves approaching saturation (Supplementary Fig. S10). *Varcoding*
309 resolved a total of 97 *varcodes* in South America ($P_{TS} \geq 0.90$), with those identified in each country ranging
310 from 9 in Ecuador and Venezuela, to 56 *varcodes* in French Guiana (Supplementary Fig. S11). The number
311 of *var* DBL α types per isolate repertoire was low in all countries and indicative of only one *P. falciparum*
312 genome infecting an individual (i.e. repertoire size ≤ 60 , Fig. 4b, Supplementary Table S1). We found that
313 the median number of *var* DBL α types in each isolate ranged from 36.5 in Venezuela to 48.0 in French
314 Guiana and the maximum number of *var* DBL α types per isolate was 42 or 43 in all countries except
315 French Guiana that appeared to have more multi-genome infections (max = 92). The median repertoire
316 size was in the range of the number of *var* genes seen in the genome of the Honduran laboratory
317 reference strain HB3 (N=42 excluding *var2csa*,^{32,33}). An unrooted phylogenetic neighbor-joining tree
318 revealed distinct clusters of genetically-related isolates in South America generally clustering by country
319 (Fig. 4c), consistent with previous analyses in the region demonstrating geographic population
320 structure²⁸.

321

322 To examine whether any of the parasites with Ecuadorian *varcodes* were genetically related to other
323 South American parasites, we constructed regional spatiotemporal relatedness networks and applied our
324 threshold of $P_{TS} \geq 0.50$ (Fig. 4d). We identified a genetically-related Peruvian *P. falciparum* isolate that
325 clustered with the outbreak and recombinant *varcodes* ($P_{TS} = 0.66-0.75$ with *varcode1* isolates). This is
326 consistent with previous analyses using microsatellites showing the outbreak source was possibly a
327 residual parasite lineage circulating in Peru in 1999-2000 and in Ecuador in the early 1990s^{8,13}. The fact
328 that these isolates did not cluster with anything else suggests the outbreak may have been caused locally
329 due to circulating parasites in unidentified reservoirs. *Varcode5*, which was sampled in the Amazon had a
330 very different genetic profile to the other *varcodes* ($\geq 83\%$ unique types, Fig. 3d). It is worth noting that
331 the parasite isolate with *varcode5* also had a different genetic background with respect to its
332 microsatellite haplotype as compared to the rest of the Ecuadorian isolates (3 of 7 unique microsatellite
333 alleles). The epidemiological data for the putative infection location was recorded as possibly imported
334 from Peru (EC40, Table 1). However, our analysis would suggest that this *varcode* is more related to *P.*
335 *falciparum* populations from French Guiana and Venezuela. Nonetheless, Peruvian isolates also clustered
336 with *P. falciparum* isolates from French Guiana even though they did not cluster with *varcode5* and we
337 found that 53% of the types identified in *varcode5* were also identified in Peruvian parasites (Fig. 4e). We
338 were able to determine a putative importation of a Colombian parasite (*varcode9*, EC53, Table 1, Fig. 4d).
339 The *varcodes* 2 and 8 did not cluster with any other South American isolate suggesting parasites with
340 these *varcodes* may be local to Ecuador. The epidemiological data for the putative infection location for
341 *varcode8* was recorded as Colombia (EC49 and EC50, Table 1). Although it did not cluster with Colombian
342 *P. falciparum* isolates in our network analysis, we found that 67% of the types identified in *varcode8*
343 (collapsing all 3 isolates) were also identified in Colombian parasites (Fig. 4e). For *varcode2*, we found
344 that 44% of the types identified in *varcode2* were also identified in Colombia (Fig. 4e), providing evidence
345 that parasites with *varcode2* and *varcode8* may represent residual parasites previously or recently
346 imported from Colombia.



347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363

Figure 4. Relatedness networks in South America provide insights into the origins of the Ecuadorian varcodes and diversity patterns in South America. (a) A map showing the study sites across South America. The dates of sample collection in these countries occurred from 2002 to 2008, around five to thirteen years prior to the sample collection of the Ecuadorian isolates. N refers to the number of isolates. (b) Repertoire sizes are shown for each *P. falciparum* isolate, from which multiplicity of infection can be inferred (i.e., ~40-60 var DBL α types per *P. falciparum* genome). (c) An unrooted neighbor-joining tree shows the relatedness patterns among South American *P. falciparum* isolates. (d) A network visualization of the genetic relatedness of South American varcodes at the threshold of $P_{TS} \geq 0.50$. Due to the differences in time points and sample collections, for this analysis a genetically-related parasite did not necessarily reflect a recent transmission event. Every node represents a *P. falciparum* isolate and an edge represents the P_{TS} value between two particular nodes/isolates. Isolates/varcodes that cluster together (i.e., connected by edges) are genetically-related. (e) Relatedness of Ecuadorian varcodes to *P. falciparum* populations from South America. The relatedness of each varcode was measured by first concatenating all possible var DBL α types that were identified in the *P. falciparum* isolates comprising each varcode as well as concatenating all var DBL α types that were identified in each country. Then P_{TS} was calculated between each varcode and each country. The color gradient denotes the P_{TS} value for a particular comparison, with darker shades showing higher relatedness and the different colors corresponding to the different country comparisons. N refers to the number of var DBL α types. (f) A clustered heatmap showing the patterns of diversity across South America, with a row representing all the types identified in a particular country, and columns representing each of the 543 DBL α types. Black and white denote the presence and absence of each type, respectively. The total number of unique var DBL α types identified in each country ranged from 112 in Colombia to 249 in French Guiana. Countries that cluster together share more var DBL α types.

364 Conservation of individual DBL α types in space and time

365 Of significance, we found the same *var* DBL α types (56% of types identified in Ecuador, N=110/195, Fig.
366 4f) and *var*codes that were genetically related to other South American parasites and sampled up to 13
367 years before (Fig. 4d-e) were contributing to local transmission in Ecuador. Indeed, conservation of the
368 DBL α types in all *var*codes was observed in space and time to varying degrees (Fig. 4e). We also found
369 conservation of DBL α types on a continental scale with the number of the same types identified in any
370 two or three countries being 124 (23%) and 59 (11%), respectively (Fig. 4f). Moreover, even DBL α types
371 that were identified in relatively low frequencies in Ecuador (i.e., identified in a few *P. falciparum* isolates)
372 were also identified in the South America dataset, often at relatively high frequencies in other countries
373 (Supplementary Fig. S12). These patterns are consistent with some parasite *var*codes being putative
374 importations and/or highly-related to parasites in other countries, as described previously.

375

376 No evidence of recombination in any DBL α type over the course of the outbreak

377 The fact that the outbreak *var*code1 was clonal provided a unique opportunity to examine the
378 conservation of its 47 individual DBL α types after transmission to multiple human hosts over the course
379 of the outbreak and up to two years after (N=36 infections). DBL α types are very variable, with the
380 average pairwise identity of the sequences encoding this domain being approximately 42%³³. The cut-off
381 of 96% sequence identity in our pipeline has been routinely used to identify the same DBL α types within
382 the limits of sequencing errors and PCR artifacts (^{27,53}, Methods). Our data show that the same DBL α types
383 were conserved in isolates with *var*code1 over the course of the outbreak (N=30 infections) and up to
384 two years after the outbreak in the other six cases with *var*code1 (Supplementary Fig. S13a). In total, we
385 report 1,283 independent observations of the same 41 DBL α types identified in 36 cases with *var*code1
386 assessed over 2 to 3 years (any type was identified in a median of 34 infections, ranging from 1 to all 36
387 infections) (Supplementary Fig. S13a). Given our threshold of $P_{TS} \geq 0.90$ to identify putatively identical *var*
388 DBL α repertoires, only 10 of the 47 DBL α types found in the outbreak clone were not found in the
389 majority of the isolates with *var*code1 (median frequency = 1, max = 6, Supplementary Fig. S13a). Four of
390 these were identified in the cases with recombinant *var*codes (median frequency = 17.5, range = 6-19),
391 two of which were also identified in the Peruvian isolate that clustered with the outbreak/recombinants
392 in Fig. 4d, suggesting they are also conserved. The other 6/10 were singletons in the entire dataset
393 (including the South America data) and could be a consequence of the PCR with degenerate primers
394 and/or low-quality DNA.

395

396 Among the 15 cases with recombinant *var*codes (*var*codes 3,4,6, and 7), we report 339 independent
397 observations of 39 of the 47 outbreak DBL α types. There were 26-27 outbreak types shared between
398 *var*code1 and the recombinant *var*codes 3,4, and 7, and 17 types between *var*code1 and *var*code6
399 (Supplementary Fig. S13b). These findings highlight the conservation of DBL α types of the outbreak clone
400 in space and time, a pattern further supported by the fact that 31 of 47 outbreak DBL α types were found
401 among previously recorded South American isolates (in 2002 to 2008²⁸; see Supplementary Fig. S13c).
402 This evidence points to conservation of these types despite the many transmission events that must have
403 occurred between 2002 and 2013-15 across South America. We found that the highest number of
404 outbreak DBL α types were identified in Peru (N=28, 60%) followed by Colombia (N=14, 30%) and
405 Venezuela (N=8, 17%), and the lowest in French Guiana (N= 2, 4%, Supplementary Fig. S9c).

406 Discussion

407 Our investigation of the spatiotemporal incidence of clinical episodes of *P. falciparum* in Ecuador by
408 *var*coding supports the view that disease transmission after the 2012-2013 outbreak was sustained
409 predominantly by Ecuadorian parasites. We observed persistence of the outbreak clonal lineage
410 (identified with the same *var*codes) and outcrossing of this lineage with other locally circulating parasites
411 causing clinical cases, rather than recent importation of clonally transmitted parasites. Our results point
412 to the need for resources to be focused locally in Ecuador to uncover the circulating asymptomatic
413 reservoirs of infection identified by *var*code recombination events. A role for human mobility must also
414 be considered in the spread of *P. falciparum* as parasites with the same *var*codes were also observed
415 across large distances (~150-300km).

416
417 San Lorenzo was found to be a transmission hotspot likely due to mining activities and occupation-related
418 travel in these areas, as well as its proximity to the Ecuador-Colombia border^{3,50,54-56}. These findings point
419 to San Lorenzo for both genetic surveillance and targeted interventions. Importantly, our results provide
420 strong evidence for ongoing transmission in Ecuador and provide the first baseline characterization of *P.*
421 *falciparum* antigenic diversity and parasite *var*codes circulating in the northwest coast and Amazon
422 regions of Ecuador. It is worth noting that several recent outbreaks have occurred in the same areas of
423 Ecuador (in 2016¹⁰, 2018⁵⁷, 2019⁶ and 2020⁵⁸) with malaria transmission remaining unstable rather than
424 endemic.

425

426 We demonstrate that 58% of clinical cases sampled in 2014 immediately after the outbreak in Ecuador
427 were caused by parasites with recombinant *var*codes. The same trend was observed in 2015, although
428 our sampling was limited relative to the number of reported cases for the months sampled. We ask is this
429 a chance finding or a consequence of immune evasion? We hypothesize that these parasites with
430 recombinant *var*codes have new combinations of *var* DBL α types that may provide an immunological
431 advantage in a population with variable levels of immunity, as would be expected during elimination
432 campaigns⁵⁹.

433
434 Network analyses and computational models combining evolution and epidemiology point to variant-
435 specific immune selection defining *var* population structure^{60,61}. In scenarios where transmission is low
436 and there is limited *var* diversity, the chance of recombination is also relatively low and thus immune
437 selection is weaker. Selection in the context of an E-2020 country like Ecuador is likely variable due to the
438 recent transition from moderate to epidemic transmission over the last 15 years¹ with declining
439 population immunity as a consequence of approaching elimination. Low population immunity would
440 select for circulating clones leading to high similarity between *var* repertoires. Here, the signatures of
441 recently recombined genomes with the outbreak clone that we detect in four *var*codes were clearly
442 associated with the continued incidence of clinical disease after the outbreak. Thus parasites with
443 recombinant *var*codes may present a risk factor for disease. We hypothesize that they may be better able
444 to evade variant-specific immunity to PfEMP1 variants expressed by the outbreak clone by sharing 50% or
445 less of the DBL α types of the outbreak clone. The availability of *var* DBL α sequences provides the
446 potential to test this hypothesis by serological methods that detect variant-specific immunity. In this
447 context, inferences based on neutral markers do not give information about immune evasion genes and
448 variant-specific immunity, whereas *var*codes identify circulating variants to which individuals can become
449 immune. e.g. as was done in serologic studies in Papua New Guinea⁶² and the Brazilian Amazon⁶³.

450
451 A previous study that included some of the same *P. falciparum* isolates from Esmeraldas City and San
452 Lorenzo in the northwest coast described three main genetic clusters based on microsatellite
453 genotyping⁵⁰. In contrast, when considering the same isolates, *var*coding resolved six *var*codes circulating
454 in the northwest coast. In this study, as expected comparative analyses to “gold-standard” microsatellite
455 genotyping demonstrated that *var*coding provided higher discriminatory resolution to detect fine-scale
456 genetic variation and describe parasite microevolution related to disease transmission in epidemiological
457 time. These findings were in accordance with the rapidly evolving *var* multigene family due to immune

458 selection^{30,60,64}, the overall number of microsatellite loci examined and the fact that two loci were fixed in
459 the population^{8,50} decreasing the denominator of polymorphic microsatellite markers from seven to five.
460 Similar observations have been described in an earlier study in Venezuela where sympatric parasites
461 identical at neutral microsatellite loci were shown to have very different *var* repertoires⁶⁵. These South
462 American genetic epidemiology studies can be compared to those of a peri-urban area in Senegal where
463 reduced transmission has resulted from intense vector and antimalarial control. In Senegal, highly related
464 genotypic clusters of *P. falciparum* isolates were resolved by both *var* and use of a 24-SNP barcode as
465 neutral loci rather than microsatellite loci⁶⁶. Genotyping *var* loci under immune selection in conjunction
466 with neutral loci proves to be informative to validate clonality and avoids the need for whole genome
467 sequencing as a potential rapid response to an outbreak investigation. Even in areas where WGS is
468 routinely conducted, *var*coding has the potential to identify relevant samples for downstream WGS as
469 part of a larger and complementary genomic surveillance framework.

470
471 In contrast to the microevolution of parasite *var*codes, we documented conservation of individual *var*
472 DBL α types over space and time. This is similar to Africa^{27,29} where we see highly variable but many
473 conserved DBL α types yet rapidly evolving *var* repertoires or *var*codes. Here we observed conservation of
474 the outbreak *var*code1 clone DBL α types after transmission within and between multiple human hosts on
475 the scale of months to two years in Ecuador (1,283 independent observations of the same types when
476 examining 36 cases with *var*code1). Moreover, 110 of the 195 (56.4%) Ecuadorian DBL α types were
477 identified across South America over 5 to 13 years (2,130 independent observations of the Ecuadorian
478 DBL α types in 128 different infections from²⁸). These observations are consistent with balancing
479 selection⁶⁷ and in line with other studies demonstrating high levels of *var* DBL α sequence conservation on
480 the time scale of up to 23 years in Brazil^{34,68}. Another independent example of conservation of DBL α
481 sequences over time is found in a peri-urban area of Senegal over 7 years⁶⁶. This pattern of conservation
482 of *var* DBL α types observed in diverse sites confirms the potential of this genetic marker for malaria
483 surveillance.

484
485 Of note, the observed conservation of DBL α types in natural infections in humans in Africa^{27,29,69} and
486 South America would not be predicted by data from *in vitro* passage of long-term cloned lines of *P.*
487 *falciparum* showing high rates of mitotic recombination within *var* genes⁷⁰⁻⁷². Indeed, Claessens et al⁷¹
488 reported a very high rate of $2-8 \times 10^{-3}$ mitotic recombinants created per erythrocytic cycle for three of
489 four long-term *in vitro* cultured lines, “with DBL α domains showing the most recombinations”. The clone

490 tree of HB3, selected for passage through mosquitos, did not show any recombinants suggesting that the
491 high rates observed may relate to passage *in vitro*. While we identified conservation rather than
492 diversification of *var* DBL α types in space and time, it is possible that mitotic recombinants are generated
493 on the observed time scale but not transmitted. Such is the case in HIV (e.g. reviewed in ⁷³). Further
494 exploration of within and between host diversification of *var* genes in natural infections requires a more
495 detailed genomic approach analyzing recently adapted parasite lines.

496
497 This is the first case study demonstrating the translational application of the *var*code for outbreak
498 surveillance in epidemic/unstable malaria transmission such as E-2020 settings. Specifically, data
499 presented here have shown that *var*coding was able to resolve the population genetics of sustained
500 disease transmission after an outbreak in Ecuador. We conclude that *var*coding requiring a single PCR
501 with degenerate primers has great potential as a simple, cost effective and high-resolution approach to
502 examine *P. falciparum* antigenic diversity in relation to population immunity as well as genome
503 diversification by recombination. This will prove particularly useful in the context of changing patterns of
504 human mobility and gene flow in the Americas where there is high demand for such a diagnostic
505 surveillance method. The tool will also be useful in other epidemic or low transmission settings targeting
506 elimination across the globe.

507
508 Here we show that *var* repertoire evolution via sexual recombination was associated with sustained
509 transmission resulting in clinical infections. This is in line with our previous predictions of epidemic
510 disease transmission in South America with parasites with novel *var* repertoires, or *var*codes²⁸, and serves
511 as a warning of what may come in terms of monitoring the complexity of malaria resurgence in Latin
512 America with both importation and microevolution of *P. falciparum*. Surveillance with neutral markers
513 alone cannot investigate this potential risk related to recombination of variant antigen gene repertoires.
514 The widespread migration of Venezuelans has already led to massive increases in *P. falciparum* cases in
515 Brazil and Colombia⁷⁴, as well as in the Peru-Ecuador border³. Such migration points to the possibility of
516 importation of diverse parasites with antigenically novel *var* genes as we predicted ²⁸, and is now also
517 worrisome in the context of increased transmission of malaria in tropical regions due to the diversion of
518 healthcare resources to the COVID-19 pandemic, as recently warned by the WHO⁷⁵. Going the distance to
519 elimination must be supported by appropriate molecular surveillance.

520

521 Methods

522 Ethics statement

523 The *P. falciparum* isolates collected from individuals of all ages presenting with uncomplicated malaria
524 cases were obtained from the malaria surveillance protocol approved by the Ethical Review Committee of
525 Pontificia Universidad Católica del Ecuador. Ethical approval for this study was obtained from Pontificia
526 Universidad Católica del Ecuador (Quito, Ecuador) and University of Melbourne (Melbourne, Australia).
527 Written informed consent was provided by study participants and/or their legal guardians.

528

529 Study design, study area and sample collection

530 In this molecular epidemiological study, we examined samples that were collected passively from 2013 to
531 2015 in endemic areas of Ecuador, namely the northwest coast and the Amazon region, as part of
532 ongoing malaria surveillance by the Ecuadorian Ministry of Health and patients showing at the local
533 clinics. During 2013, the majority of samples were collected from an ongoing outbreak that occurred in
534 Esmeraldas city, Esmeraldas province. Details of the sample collection are described in Sáenz et al⁸. In
535 2014 and 2015, the “post-outbreak” samples were mostly collected from across the northwest coast with
536 few from the Amazon region, reflective of the epidemiology of clinical *P. falciparum* cases in Ecuador
537 (Northwest coast: 20 and 173 cases; Amazon: 11 and 16 cases in 2014 and 2015, respectively). Details of
538 the sample collection are described in Vera-Arias et al⁵⁰. It is worth noting that although cases were
539 reported throughout 2015, samples were not available for many of them due to logistical challenges.
540 Briefly, subjects who were diagnosed with falciparum malaria either by light microscopy or a rapid
541 diagnostic test were asked if they wanted to participate in the study and signed an informed consent
542 form. The participants had to: 1) live in the areas where the samples were taken, 2) be over 2 years old, 3)
543 agree to participate, 4) sign an informed consent form, 5) give a blood sample, 6) be able to understand
544 the informed consent. From each participant a blood sample (either venous blood or dried blood spot)
545 was collected as well as their responses to a basic demographic questionnaire (including places recently
546 travelled and their address). The positivity of all samples was confirmed in the laboratory by microscopy
547 and PET-PCR⁷⁶.

548

549 DNA extraction

550 Genomic DNA was extracted from venous blood or dried blood spot samples using a QIAamp DNA MINI
551 KIT (QIAGEN, Germantown, MD, USA) using the protocol recommended by the manufacturer.

552

553 [Microsatellite genotyping](#)

554 The same *P. falciparum* isolates in this study were previously genotyped for seven putatively neutral
555 microsatellite markers (TA1, POLY α , P α PK2, TA109, 2490, C2M34, C3M39) as described in^{8,50}. These
556 previously published microsatellite genotyping data^{8,48,75} were used in the present analyses.

557

558 [Var genotyping and related data processing](#)

559 For *var* genotyping, the DBL α domains of antigen-encoding *var* genes were amplified and sequenced on a
560 MiSeq 2x300bp paired-end Illumina platform as described in^{29,60}. We obtained 6,929,470 raw sequence
561 reads, which we cleaned and processed using the DBL α Cleaner pipeline^{(60, [http://github.com/Unimelb-](http://github.com/Unimelb-Day-Lab/DBLalphaCleaner)}
562 [Day-Lab/DBLalphaCleaner](http://github.com/Unimelb-Day-Lab/DBLalphaCleaner)). Our customized bioinformatic pipeline has been described in detail in⁶⁰. Briefly,
563 we de-multiplexed and merged the reads as well as removed low-quality reads and chimeras using
564 several filtering parameters (see Supplementary Fig. S1 for details). This resulted in 2,141 cleaned DBL α
565 sequences (Supplementary Fig. S1). To identify distinct or unique DBL α types (i.e., unique genetic
566 variants), we clustered the DBL α sequences from Ecuadorian *P. falciparum* isolates with 5,699 previously
567 published²⁸ DBL α sequences from other South American countries (Colombia (N=21 isolates), French
568 Guiana (N=76 isolates), Peru (N=21 isolates), and Venezuela (N=10 isolates) at the standard 96%
569 sequence identity²⁷ using the clusterDBLalpha pipeline ([http://github.com/Unimelb-Day-](http://github.com/Unimelb-Day-Lab/clusterDBLalpha)
570 [Lab/clusterDBLalpha](http://github.com/Unimelb-Day-Lab/clusterDBLalpha)). All the cleaned DBL α sequences in this study have been submitted to the
571 DDBJ/ENA/GenBank (BioProject Number: PRJNA642683).

572

573 We further curated our dataset by translating the DBL α types into amino acid sequences using the
574 classifyDBLalpha pipeline (<http://github.com/Unimelb-Day-Lab/classifyDBLalpha>) and removing any DBL α
575 types that were non-translatable (N=4). In addition, any *P. falciparum* isolate with low sequencing quality
576 (< 10 DBL α types) was removed from the analysis. *Var* coding was completed for a total of 70 *P.*
577 *falciparum* Ecuadorian isolates, however, 12 *P. falciparum* isolates with low DNA quality and/or poor
578 sequencing quality were removed and we obtained *var* DBL α data for 58 isolates (82.9%). A total of 543
579 unique DBL α types identified in the 186 South American *P. falciparum* isolates (N = 58 Ecuadorian *P.*
580 *falciparum* isolates from this study, N = 128 South American *P. falciparum* isolates previously published in
581 ²⁸) were used for subsequent *var* analyses at the regional-level, and only the 195 unique DBL α types
582 identified in Ecuador were used for Ecuador-specific analyses.

583

584 A further validation of our sequencing and genotyping methodology was undertaken by analyzing the
585 isolate genomic proportions of DBL α types classified as upsA and non-upsA using our classifyDBLalpha
586 pipeline²⁹ and are described in Supplementary Text 1.

587

588 Genetic relatedness analyses

589 *Measuring pairwise type or allele sharing*

590 To estimate genetic relatedness among isolates, we calculated the similarity index Pairwise Type Sharing
591 (P_{TS})²⁶, as adapted by He et al⁶⁰ to account for differences in *var* DBL α sampling across isolates in the case
592 of *var*, and Pairwise Allele Sharing (P_{AS})⁷⁷ in the case of microsatellites and constructed similarity matrices.
593 Because *P. falciparum* is haploid we refer to *var* repertoires and microsatellite haplotypes in the same
594 manner, i.e. a *var* repertoire of DBL α types or a multilocus haplotype of microsatellite alleles transmitted
595 together. Every parasite isolate was compared to every other parasite isolate in the population to
596 determine the proportion of shared loci (i.e. *var* types or microsatellites), but only “retrospective”
597 pairwise comparisons were analyzed as a conservative approach to account for time order for the
598 Ecuador-specific analyses. In other words, since we were interested in using the outbreak as a baseline
599 this approach allowed us to look at genetic relatedness of an isolate with everything else circulating
600 before its identification. For the comparative analyses with South American data we included all pairwise
601 comparisons and did not correct for time order due to the differences in sampling time points (five to
602 thirteen years before the Ecuadorian isolates were collected).

603

604 *Measuring Bayesian pairwise type sharing*

605 Genetic relatedness (or repertoire overlap) can also be explored more rigorously with unbiased Bayesian
606 pairwise type sharing (BP_{TS} , Johnson et al *in preparation*), which accounts for uncertainty in P_{TS} estimates
607 due to differences in isolate repertoire size. This approach uses Bayesian inference methods, which
608 estimate repertoire overlap and uncertainty⁷⁸, and uses them in a subsequent P_{TS} calculation, carrying
609 that uncertainty forward. The prior distribution for repertoire size, used in inference, was informed by
610 observations as follows. First, the median observed repertoire size in Ecuadorian isolates was 37 types,
611 ranging from 11 to 43 (Table S1). Second, the number of expected *var* genes with DBL α domains from
612 whole genome sequencing data of the Honduran laboratory reference strain HB3 was 42³³ and 50 *var*
613 genes based on long-read PacBio sequencing of HB3⁷⁹. And third, based on our sequencing data of 37

614 technical replicates of HB3, the median repertoire size or number of DBL α types per isolate was 39 (range
615 36-41 types), with 40 types consistently identified in the majority of replicates (range 21-37 replicates)
616 (Supplementary Text 1). We therefore used a uniform prior on repertoire sizes between 40 and 50 types,
617 combined with the general Bayesian repertoire overlap framework⁷⁸ to produce unbiased estimates
618 (posterior means). These were used to confirm our P_{TS} estimates. As expected, the P_{TS} and BP_{TS} estimates
619 were positively correlated (Pearson's correlation coefficient = 0.919, $p < 0.001$, Supplementary Fig. S3). To
620 measure uncertainty in central estimates, we computed a 95% highest density posterior interval (HDPI), a
621 Bayesian version of confidence intervals, for each pairwise estimate. Like a frequentist confidence
622 interval, the width of the HDPI provides a measure of uncertainty of each pairwise comparison. All
623 posteriors were sampled using MCMC.

624

625 *Interpretation of genetic relatedness measures*

626 Since *P. falciparum* undergoes sexual recombination (i.e. meiosis) in the mosquito, conventional genetic
627 interpretations can be applied to P_{TS} and P_{AS}. Therefore, pairwise isolate comparisons resulting in a P_{TS} or
628 P_{AS} of 0 (0% shared loci), 0.5 (50% shared loci), and 1 (100% shared loci) would indicate genetically distinct
629 isolates, recombinant isolates, and clones or genetically-identical isolates, respectively. We applied this
630 approach to define “*var*codes” as groups of isolates sharing $\geq 90\%$ of their DBL α types (P_{TS} ≥ 0.90) to
631 identify putatively identical genomes within the margin of error of detection of 1-5 DBL α types in an
632 isolate (i.e., within the margin of error of detection of all DBL α sequences in an isolate using degenerate
633 primers for *var* DBL α PCR). We confirmed our interpretations of *var*codes, recombinants and genetically
634 distinct isolates by comparing to unbiased BP_{TS} estimates (posterior means) and examining the lower and
635 upper bounds of the 95% HDPIs as a measure of the statistical uncertainty of each pairwise comparison.

636

637 In the case of microsatellite genotyping, we applied a threshold of P_{AS} ≥ 0.90 (sharing $\geq 90\%$ of their
638 microsatellite alleles) to define clones and also applied a threshold of P_{AS} ≥ 0.80 to identify putatively
639 identical genomes within the margin of error of detection of 1 microsatellite allele in an isolate (for
640 comparative purposes to the threshold for *var*, Supplementary Fig. S2). It is important to note that 2 out
641 of the 7 microsatellite markers genotyped were fixed in the population⁸.

642

643 *Visualization of genetic relatedness networks*

644 To visualize the genetic relatedness among isolates as determined by P_{TS} or P_{AS}, we constructed networks
645 using the R packages *ggraph*⁸⁰ and *tidygraph*⁸¹ where isolates are depicted as nodes and edges as the P_{TS}

646 or P_{AS} values at a given threshold. The R package *ggspatial*⁸² was used to plot spatial networks using
647 latitude/longitude coordinates for sampling location. To visualize genetic relatedness of parasites over
648 time we used the R package *gganimate*⁸³ to construct spatiotemporal relatedness networks. We
649 generated a clustered heatmap based on the presence/absence matrix of DBL α types to visualize the
650 genetic profiles of each isolate in Ecuador and each country in South America using the R package
651 *pheatmap*⁸⁴ and the “complete” clustering method. Unrooted neighbor-joining phylogenetic trees based
652 on pairwise genetic distance (calculated as $1-P_{TS}$) were constructed using the R packages *ape*⁸⁵ and
653 *ggtree*^{86,87}.

654

655 Statistical analysis

656 We used R version 3.5.2 for all analyses⁸⁸. The package *dplyr*⁸⁹ was used for further data curating. We
657 used base R and the R package *epiR*⁹⁰ to perform chi-squared tests for univariate analyses of categorical
658 variables to compare proportions and for non-parametric tests to compare distributions of continuous
659 variables between two groups (Mann-Whitney U test) or among k groups (Kruskal-Wallis test) with a
660 Bonferroni correction for multiple comparisons. To evaluate how well we sampled the true pool of *var*
661 DBL α diversity (i.e. the true number of genetic variants circulating) in Ecuador and in South America, we
662 used the R package *vegan*⁹¹ to generate species accumulation curves.

663

664 Data availability

665 The cleaned DBL α sequences generated in this study from Ecuadorian *P. falciparum* isolates have been
666 submitted to the DDBJ/ENA/GenBank (BioProject Number: PRJNA642683). The python code for the
667 DBL α Cleaner pipeline is available on GitHub at <https://github.com/Unimelb-Day-Lab/DBLaCleaner>. The
668 python code for the clusterDBLalpha pipeline is available on GitHub at <https://github.com/Unimelb-Day-Lab/DBLalpha>. The python code for the classifyDBLalpha pipeline is available on GitHub at
669 <https://github.com/Unimelb-Day-Lab/classifyDBLalpha>. All other deidentified data and analysis code are
670 available on the open-source GitHub repository at <https://github.com/shaziaruybal/varcode-ecuador>.

671

672 References

673

674
675 1. Organization, W. H. *Eliminating Malaria*. [https://www.who.int/malaria/publications/atoz/eliminating-](https://www.who.int/malaria/publications/atoz/eliminating-malaria/en/)
676 [malaria/en/](https://www.who.int/malaria/publications/atoz/eliminating-malaria/en/) (2016).

- 677 2. Espinoza, J. Malaria Resurgence in the Americas: An Underestimated Threat. *Pathogens* **8**, 11 (2019).
- 678 3. Jaramillo-Ochoa, R. *et al.* Effects of Political Instability in Venezuela on Malaria Resurgence at Ecuador–
679 Peru Border, 2018. *Emerg Infect Dis* **25**, 834–836 (2019).
- 680 4. Ferreira, M. U. & Castro, M. C. Malaria Situation in Latin America and the Caribbean: Residual and
681 Resurgent Transmission and Challenges for Control and Elimination. *Methods Mol Biology Clifton NJ*
682 **2013**, 57–70 (2019).
- 683 5. Sáenz, F. E. *et al.* Malaria epidemiology in low-endemicity areas of the northern coast of Ecuador: high
684 prevalence of asymptomatic infections. *Malaria J* **16**, 300 (2017).
- 685 6. MSP. *Gaceta Epidemiológica Semanal No. 52.* (2019).
- 686 7. PAHO. *Report on the Situation of Malaria in the Americas 2012.* (2013).
- 687 8. Sáenz, F. E. *et al.* Clonal population expansion in an outbreak of Plasmodium falciparum on the
688 northwest coast of Ecuador. *Malaria J* **14**, 497 (2015).
- 689 9. MSP. *Gaceta Epidemiológica Semanal No. 52.* (2016).
- 690 10. MSP. *Gaceta Epidemiológica Semanal No. 52.* (2017).
- 691 11. Organization, W. H. *World Malaria Report 2019.* (2019).
- 692 12. Organization, W. H. THE E-2020 INITIATIVE OF 21 MALARIA-ELIMINATING COUNTRIES: 2019 progress
693 report. (2019).
- 694 13. Griffing, S. M. *et al.* South American Plasmodium falciparum after the Malaria Eradication Era: Clonal
695 Population Expansion and Survival of the Fittest Hybrids. *Plos One* **6**, e23486 (2011).
- 696 14. Larrañaga, N. *et al.* Genetic structure of Plasmodium falciparum populations across the Honduras–
697 Nicaragua border. *Malaria J* **12**, 354 (2013).
- 698 15. Obaldia, N. *et al.* Clonal outbreak of Plasmodium falciparum infection in eastern Panama. *J Infect Dis*
699 **211**, 1087–96 (2014).
- 700 16. Okoth, S. A. *et al.* Molecular Investigation into a Malaria Outbreak in Cusco, Peru: Plasmodium
701 falciparum BV1 Lineage is Linked to a Second Outbreak in Recent Times. *Am J Tropical Medicine Hyg* **94**,
702 128–31 (2015).
- 703 17. Baldeviano, G. C. *et al.* Molecular Epidemiology of Plasmodium falciparum Malaria Outbreak, Tumbes,
704 Peru, 2010–2012. *Emerg Infect Dis* **21**, 797–803 (2015).
- 705 18. Chenet, S. M., Taylor, J. E., Blair, S., Zuluaga, L. & Escalante, A. A. Longitudinal analysis of Plasmodium
706 falciparum genetic variation in Turbo, Colombia: implications for malaria control and elimination. *Malaria*
707 *J* **14**, 363 (2015).

- 708 19. Carter, T. E. *et al.* Genetic Diversity of Plasmodium falciparum in Haiti: Insights from Microsatellite
709 Markers. *Plos One* **10**, e0140416 (2015).
- 710 20. Redmond, S. N. *et al.* De Novo Mutations Resolve Disease Transmission Pathways in Clonal Malaria.
711 *Mol Biol Evol* **35**, 1678–1689 (2018).
- 712 21. Chenet, S. M., Schneider, K. A., Villegas, L. & Escalante, A. A. Local population structure of
713 Plasmodium: impact on malaria control and elimination. *Malaria J* **11**, 412 (2012).
- 714 22. Barton, N. Understanding Adaptation in Large Populations. *Plos Genet* **6**, e1000987 (2010).
- 715 23. Chen, D. S. *et al.* A Molecular Epidemiological Study of var Gene Diversity to Characterize the
716 Reservoir of Plasmodium falciparum in Humans in Africa. *Plos One* **6**, e16629 (2011).
- 717 24. Kirk, H. & Freeland, J. R. Applications and Implications of Neutral versus Non-neutral Markers in
718 Molecular Ecology. *Int J Mol Sci* **12**, 3966–3988 (2011).
- 719 25. Nelson, C. S. *et al.* High-resolution micro-epidemiology of parasite spatial and temporal dynamics in a
720 high malaria transmission setting in Kenya. *Nat Commun* **10**, 5615 (2019).
- 721 26. Barry, A. E. *et al.* Population Genomics of the Immune Evasion (var) Genes of Plasmodium falciparum.
722 *Plos Pathog* **3**, e34 (2007).
- 723 27. Day, K. P. *et al.* Evidence of strain structure in Plasmodium falciparum var gene repertoires in children
724 from Gabon, West Africa. *Proc National Acad Sci* **114**, E4103–E4111 (2017).
- 725 28. Rougeron, V. *et al.* Evolutionary structure of Plasmodium falciparum major variant surface antigen
726 genes in South America: Implications for epidemic transmission and surveillance. *Ecol Evol* **7**, 9376–9390
727 (2017).
- 728 29. Ruybal-Pesántez, S. *et al.* Population genomics of virulence genes of Plasmodium falciparum in clinical
729 isolates from Uganda. *Sci Rep-uk* **7**, 11810 (2017).
- 730 30. Rorick, M. M. *et al.* Signatures of competition and strain structure within the major blood-stage
731 antigen of Plasmodium falciparum in a local community in Ghana. *Ecol Evol* **8**, 3574–3588 (2018).
- 732 31. Tonkin-Hill, G. *et al.* Evolutionary analyses of the major variant surface antigen-encoding genes reveal
733 population structure of Plasmodium falciparum within and between continents. *Plos Genet* **17**, e1009269
734 (2021).
- 735 32. Kraemer, S. M. *et al.* Patterns of gene recombination shape var gene repertoires in Plasmodium
736 falciparum: comparisons of geographically diverse isolates. *Bmc Genomics* **8**, 45 (2007).
- 737 33. Rask, T. S., Hansen, D. A., Theander, T. G., Pedersen, A. G. & Lavstsen, T. Plasmodium falciparum
738 Erythrocyte Membrane Protein 1 Diversity in Seven Genomes – Divide and Conquer. *Plos Comput Biol* **6**,
739 e1000933 (2010).

- 740 34. Albrecht, L. *et al.* The South American Plasmodium falciparum var gene repertoire is limited, highly
741 shared and possibly lacks several antigenic types. *Gene* **453**, 37–44 (2010).
- 742 35. Gardner, M. J. *et al.* Genome sequence of the human malaria parasite Plasmodium falciparum. *Nature*
743 **419**, 498–511 (2002).
- 744 36. Biggs, B. A. *et al.* Antigenic variation in Plasmodium falciparum. *Proc National Acad Sci* **88**, 9171–9174
745 (1991).
- 746 37. Scherf, A., Lopez-Rubio, J. J. & Riviere, L. Antigenic variation in Plasmodium falciparum. *Annu Rev*
747 *Microbiol* **62**, 445–70 (2008).
- 748 38. Bull, P. C. *et al.* Parasite antigens on the infected red cell surface are targets for naturally acquired
749 immunity to malaria. *Nat Med* **4**, 358–360 (1998).
- 750 39. Marsh, K. & Howard, R. Antigens induced on erythrocytes by P. falciparum: expression of diverse and
751 conserved determinants. *Science* **231**, 150–153 (1986).
- 752 40. Forsyth, K. P. *et al.* Diversity of Antigens Expressed on the Surface of Erythrocytes Infected with
753 Mature Plasmodium Falciparum Parasites in Papua New Guinea. *Am J Tropical Medicine Hyg* **41**, 259–265
754 (1989).
- 755 41. Newbold, C. I., Pinches, R., Roberts, D. J. & Marsh, K. Plasmodium falciparum: The human
756 agglutinating antibody response to the infected red cell surface is predominantly variant specific. *Exp*
757 *Parasitol* **75**, 281–292 (1992).
- 758 42. IQBAL, J., PERLMANN, P., GREENWOOD, B. M. & BERZINS, K. Seroreactivity with the Plasmodium
759 falciparum blood stage antigen Pf332 in adults and children from malaria-endemic regions. *Clin Exp*
760 *Immunol* **94**, 68–74 (1993).
- 761 43. Bull, P. C., Lowe, B. S., Kortok, M. & Marsh, K. Antibody Recognition of Plasmodium falciparum
762 Erythrocyte Surface Antigens in Kenya: Evidence for Rare and Prevalent Variants. *Infect Immun* **67**, 733–
763 739 (1999).
- 764 44. Giha, H. A. *et al.* Antibodies to variable Plasmodium falciparum-infected erythrocyte surface antigens
765 are associated with protection from novel malaria infections. *Immunol Lett* **71**, 117–126 (2000).
- 766 45. Ofori, M. F. *et al.* Malaria-Induced Acquisition of Antibodies to Plasmodium falciparum Variant Surface
767 Antigens. *Infect Immun* **70**, 2982–2988 (2002).
- 768 46. Chattopadhyay, R. *et al.* Plasmodium falciparum Infection Elicits Both Variant-Specific and Cross-
769 Reactive Antibodies against Variant Surface Antigens. *Infect Immun* **71**, 597–604 (2003).
- 770 47. Mackintosh, C. L. *et al.* Failure to respond to the surface of Plasmodium falciparum infected
771 erythrocytes predicts susceptibility to clinical malaria amongst African children. *Int J Parasitol* **38**, 1445–
772 1454 (2008).

- 773 48. Otto, T. D. *et al.* Evolutionary analysis of the most polymorphic gene family in falciparum malaria.
774 *Wellcome Open Res* **4**, 193 (2019).
- 775 49. Tessema, S. K. *et al.* Phylogeography of var gene repertoires reveals fine-scale geospatial clustering of
776 *Plasmodium falciparum* populations in a highly endemic area. *Mol Ecol* **24**, 484–97 (2015).
- 777 50. Vera-Arias, C. A., Castro, L. E., Gómez-Obando, J. & Sáenz, F. E. Diverse origin of *Plasmodium*
778 *falciparum* in northwest Ecuador. *Malaria J* **18**, 251 (2019).
- 779 51. Branch, O. H. *et al.* *Plasmodium falciparum* genetic diversity maintained and amplified over 5 years of
780 a low transmission endemic in the Peruvian Amazon. *Mol Biol Evol* **28**, 1973–86 (2010).
- 781 52. Echeverry, D. F. *et al.* Long term persistence of clonal malaria parasite *Plasmodium falciparum*
782 lineages in the Colombian Pacific region. *Bmc Genet* **14**, 2 (2013).
- 783 53. Rask, T. S., Petersen, B., Chen, D. S., Day, K. P. & Pedersen, A. G. Using expected sequence features to
784 improve basecalling accuracy of amplicon pyrosequencing data. *Bmc Bioinformatics* **17**, 176 (2016).
- 785 54. Douine, M. *et al.* Illegal gold miners in French Guiana: a neglected population with poor health. *Bmc*
786 *Public Health* **18**, 23 (2017).
- 787 55. Douine, M. *et al.* Investigation of a possible malaria epidemic in an illegal gold mine in French Guiana:
788 an original approach in the remote Amazonian forest. *Malaria J* **18**, 91 (2019).
- 789 56. Mosquera-Romero, M., Zuluaga-Idárraga, L. & Tobón-Castaño, A. Challenges for the diagnosis and
790 treatment of malaria in low transmission settings in San Lorenzo, Esmeraldas, Ecuador. *Malaria J* **17**, 440
791 (2018).
- 792 57. MSP. *Gaceta Epidemiológica Semanal No. 52.* (2018).
- 793 58. MSP. Personal communication. (2020).
- 794 59. Deloron, P. & Chougnet, C. Is Immunity to malaria really short-lived? *Parasitol Today* **8**, 375–378
795 (1992).
- 796 60. He, Q. *et al.* Networks of genetic similarity reveal non-neutral processes shape strain structure in
797 *Plasmodium falciparum*. *Nat Commun* **9**, 1817 (2018).
- 798 61. Pilosof, S. *et al.* Competition for hosts modulates vast antigenic diversity to generate persistent strain
799 structure in *Plasmodium falciparum*. *Plos Biol* **17**, e3000336 (2019).
- 800 62. Barry, A. E. *et al.* The Stability and Complexity of Antibody Responses to the Major Surface Antigen of
801 *Plasmodium falciparum* Are Associated with Age in a Malaria Endemic Area. *Mol Cell Proteomics* **10**,
802 M111.008326 (2011).
- 803 63. Carlos, B. C. *et al.* Expressed var gene repertoire and variant surface antigen diversity in a shrinking
804 *Plasmodium falciparum* population. *Exp Parasitol* **170**, 90–99 (2016).

- 805 64. Artzy-Randrup, Y. *et al.* Population structuring of multi-copy, antigen-encoding genes in *Plasmodium*
806 *falciparum*. *Elife* **1**, e00093 (2012).
- 807 65. Tami, A., Ord, R., Targett, G. A. & Sutherland, C. J. Sympatric *Plasmodium falciparum* isolates from
808 Venezuela have structured var gene repertoires. *Malaria J* **2**, 7 (2003).
- 809 66. Bei, A. K. *et al.* *Plasmodium falciparum* population genetic complexity influences transcriptional profile
810 and immune recognition of highly related genotypic clusters. *Biorxiv* 2020.01.03.894220 (2020)
811 doi:10.1101/2020.01.03.894220.
- 812 67. Zilversmit, M. M. *et al.* Hypervariable antigen genes in malaria have ancient roots. *Bmc Evol Biol* **13**,
813 110 (2013).
- 814 68. Albrecht, L. *et al.* Extense variant gene family repertoire overlap in Western Amazon *Plasmodium*
815 *falciparum* isolates. *Mol Biochem Parasit* **150**, 157–165 (2006).
- 816 69. Bei, A. K. *et al.* Immune Characterization of *Plasmodium falciparum* Parasites with a Shared Genetic
817 Signature in a Region of Decreasing Transmission. *Infect Immun* **83**, 276–285 (2014).
- 818 70. Duffy, M. F., Byrne, T. J., Carret, C., Ivens, A. & Brown, G. V. Ectopic Recombination of a Malaria var
819 Gene during Mitosis Associated with an Altered var Switch Rate. *J Mol Biol* **389**, 453–469 (2009).
- 820 71. Claessens, A. *et al.* Generation of Antigenic Diversity in *Plasmodium falciparum* by Structured
821 Rearrangement of Var Genes During Mitosis. *Plos Genet* **10**, e1004812 (2014).
- 822 72. Zhang, X. *et al.* Rapid antigen diversification through mitotic recombination in the human malaria
823 parasite *Plasmodium falciparum*. *Plos Biol* **17**, e3000271 (2019).
- 824 73. Theys, K. *et al.* The impact of HIV-1 within-host evolution on transmission dynamics. *Curr Opin Virol*
825 **28**, 92–101 (2018).
- 826 74. Grillet, M. E. *et al.* Venezuela's humanitarian crisis, resurgence of vector-borne diseases, and
827 implications for spillover in the region. *Lancet Infect Dis* **19**, e149–e161 (2019).
- 828 75. Organization, W. H. The potential impact of health service disruptions on the burden of malaria:
829 (2020).
- 830 76. Lucchi, N. W. *et al.* Molecular Diagnosis of Malaria by Photo-Induced Electron Transfer Fluorogenic
831 Primers: PET-PCR. *Plos One* **8**, e56677 (2013).
- 832 77. Ruybal-Pesántez, S. *et al.* Lack of Geospatial Population Structure Yet Significant Linkage
833 Disequilibrium in the Reservoir of *Plasmodium falciparum* in Bongo District, Ghana. *Am J Tropical*
834 *Medicine Hyg* **97**, 1180–1189 (2017).
- 835 78. Larremore, D. B. Bayes-optimal estimation of overlap between populations of fixed size. *Plos Comput*
836 *Biol* **15**, e1006898 (2019).

- 837 79. Otto, T. D. *et al.* Long read assemblies of geographically dispersed Plasmodium falciparum isolates
838 reveal highly structured subtelomeres. *Wellcome Open Res* **3**, 52 (2018).
- 839 80. Pedersen, T. L. *ggraph: An Implementation of Grammar of Graphics for Graphs and Networks*. (2020).
- 840 81. Pedersen, T. L. *tidygraph: A Tidy API for Graph Manipulation*. (2019).
- 841 82. Dunnington, D. *ggspatial: Spatial Data Framework for ggplot2*. (2018).
- 842 83. Pedersen, T. L. & Robinson, D. *gganimate: A Grammar of Animated Graphics*. (2020).
- 843 84. Kolde, R. *pheatmap: Pretty Heatmaps*. (2019).
- 844 85. Paradis, E. & Schliep, K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses
845 in R. *Bioinformatics* **35**, 526–528 (2018).
- 846 86. Yu, G., Smith, D. K., Zhu, H., Guan, Y. & Lam, T. T. ggtree : an r package for visualization and annotation
847 of phylogenetic trees with their covariates and other associated data. *Methods Ecol Evol* **8**, 28–36 (2016).
- 848 87. Yu, G., Lam, T. T.-Y., Zhu, H. & Guan, Y. Two Methods for Mapping and Visualizing Associated Data on
849 Phylogeny Using Ggtree. *Mol Biol Evol* **35**, 3041–3043 (2018).
- 850 88. Team, R. C. *R: A language and environment for statistical computing*. (R Foundation for Statistical
851 Computing, 2019).
- 852 89. Wickham, H., François, R., Henry, L. & Müller, K. *dplyr: A Grammar of Data Manipulation*. (2020).
- 853 90. Stevenson, M. *epiR: Tools for the Analysis of Epidemiological Data*. (2020).
- 854 91. Oksanen, J. *et al.* *vegan: Community Ecology Package*. (2019).
- 855

856 Acknowledgements

857 The authors would like to thank the patients who contributed samples and the field teams who were
858 involved in sample collections. We thank the Ecuadorian Ministry of Health, especially Luis E Castro, Julio
859 Valencia and Javier Gomez Obando. We appreciate the support of the Australian Genome Research
860 Facility for illumina sequencing of the samples. This research was financially supported by the Pontificia
861 Universidad Católica del Ecuador (grant numbers: L13058, L13248, M13416, N13416, O13087[QINV0084]
862 to F.E.S), the Fogarty International Center at the National Institutes of Health [Program on the Ecology
863 and Evolution of Infectious Diseases (EEID), Grant number: R01-TW009670 to K.P.D.], and the National
864 Institutes of Allergy and Infectious Diseases, National Institutes of Health (grant number: R01-AI084156 to
865 K.P.D.). S.R-P. was supported by a Melbourne International Engagement Award from the University of
866 Melbourne and gratefully acknowledges the J.D. Smyth Travel Award from the Australian Society for
867 Parasitology that enabled her to travel to Ecuador to establish this collaborative project.

868

869 Author contributions

870 S.R-P. and K.P.D. conceived and designed the study; F.E.S. and C.A.V-A. carried out field work to obtain *P.*
871 *falciparum* isolates and epidemiological metadata; S.R-P. and K.E.T. developed the *var*coding
872 methodology; S.R-P. and S.D. performed the molecular experiments and sequencing; S.R-P., F.E.S. and
873 C.A.V-A. curated the clinical and epidemiological metadata; S.R-P. performed the formal data analysis and
874 visualization; E.K.J. and D.B.L. designed and conducted the Bayesian statistical analysis; S.R-P., F.E.S.,
875 K.E.T. and K.P.D. contributed to the interpretation of the data; K.P.D. supervised the work; S.R-P., F.E.S.
876 and K.P.D. acquired funding; S.R-P. and K.P.D. wrote the paper with contributions from all authors.

877 Table 1. Epidemiological characteristics of study participants.

Case	Month collected		Age range, years	Sex	Sampling location (County, Province)	Self-reported infection location*	varcode	Putative origin**
EC1	Jan 2013	<i>Outbreak</i>	NA	NA	Esmeraldas, Esmeraldas		varcode1	Local
EC2	Jan 2013	<i>Outbreak</i>	NA	NA	Esmeraldas, Esmeraldas		varcode1	Local
EC3	May 2013	<i>Outbreak</i>	>30-40	M	Esmeraldas, Esmeraldas		varcode1	Local
EC4	May 2013	<i>Outbreak</i>	>30-40	M	Esmeraldas, Esmeraldas		varcode1	Local
EC5	May 2013	<i>Outbreak</i>	>50-60	F	Esmeraldas, Esmeraldas		varcode1	Local
EC6	May 2013	<i>Outbreak</i>	>30-40	M	Esmeraldas, Esmeraldas		varcode1	Local
EC7	May 2013	<i>Outbreak</i>	>50-60	F	Esmeraldas, Esmeraldas		varcode1	Local
EC8	May 2013	<i>Outbreak</i>	>10-20	F	Esmeraldas, Esmeraldas		varcode1	Local
EC9	May 2013	<i>Outbreak</i>	>10-20	F	Esmeraldas, Esmeraldas		varcode1	Local
EC10	May 2013	<i>Outbreak</i>	>10-20	M	Esmeraldas, Esmeraldas		varcode1	Local
EC11	May 2013	<i>Outbreak</i>	>20-30	F	Esmeraldas, Esmeraldas		varcode1	Local
EC12	Jun 2013	<i>Outbreak</i>	>10-20	F	Esmeraldas, Esmeraldas		varcode1	Local
EC13	Jun 2013	<i>Outbreak</i>	>10-20	M	Esmeraldas, Esmeraldas		varcode1	Local
EC14	Jun 2013	<i>Outbreak</i>	>40-50	M	Esmeraldas, Esmeraldas		varcode1	Local
EC15	Jun 2013	<i>Outbreak</i>	>40-50	M	Esmeraldas, Esmeraldas		varcode1	Local
EC16	Jun 2013	<i>Outbreak</i>	>10-20	F	Esmeraldas, Esmeraldas		varcode1	Local
EC17	Jun 2013	<i>Outbreak</i>	>10-20	F	Esmeraldas, Esmeraldas		varcode1	Local
EC18	Jun 2013	<i>Outbreak</i>	>20-30	M	Esmeraldas, Esmeraldas		varcode1	Local
EC19	Jun 2013	<i>Outbreak</i>	>40-50	M	Esmeraldas, Esmeraldas		varcode1	Local
EC20	Jul 2013		>20-30	M	San Lorenzo, Esmeraldas		varcode2	Local
EC21	Jul 2013		>10-20	F	San Lorenzo, Esmeraldas		varcode2	Local
EC22	Jul 2013	<i>Outbreak</i>	>10-20	M	Esmeraldas, Esmeraldas		varcode1	Local
EC23	Jul 2013	<i>Outbreak</i>	>0-10	M	Esmeraldas, Esmeraldas		varcode1	Local
EC24	Jul 2013	<i>Outbreak</i>	NA	F	Esmeraldas, Esmeraldas		varcode1	Local
EC25	Sep 2013	<i>Outbreak</i>	>30-40	M	Esmeraldas, Esmeraldas		varcode1	Local
EC26	Oct 2013	<i>Outbreak</i>	>20-30	F	Esmeraldas, Esmeraldas		varcode1	Local
EC27	Oct 2013		>10-20	F	San Lorenzo, Esmeraldas		varcode3	Local
EC28	Oct 2013	<i>Outbreak</i>	>20-30	M	Esmeraldas, Esmeraldas		varcode1	Local
EC29	Nov 2013	<i>Outbreak</i>	>10-20	M	Esmeraldas, Esmeraldas		varcode1	Local
EC30	Nov 2013	<i>Outbreak</i>	>10-20	M	Esmeraldas, Esmeraldas		varcode1	Local
EC31	Nov 2013	<i>Outbreak</i>	>10-20	F	Esmeraldas, Esmeraldas		varcode1	Local
EC32	Nov 2013	<i>Outbreak</i>	>30-40	M	Esmeraldas, Esmeraldas		varcode1	Local
EC33	Nov 2013	<i>Outbreak</i>	>30-40	M	Esmeraldas, Esmeraldas		varcode1	Local
EC34	Jan 2014		NA	NA	Jaramijo, Manabi		varcode3	Local
EC35	Jan 2014		>50-60	F	San Lorenzo, Esmeraldas		varcode1	Local
EC36	Jan 2014		>10-20	M	San Lorenzo, Esmeraldas		varcode1	Local
EC37	Jan 2014		>30-40	M	San Lorenzo, Esmeraldas		varcode2	Local
EC38	Jan 2014		>10-20	M	San Lorenzo, Esmeraldas		varcode4	Local
EC39	Jan 2014		>10-20	M	San Lorenzo, Esmeraldas		varcode1	Local
EC40	Jan 2014		>30-40	F	Aguarico, Orellana	Peru	varcode5	French Guiana/Peru/Venezuela
EC41	Mar 2014		>20-30	M	San Lorenzo, Esmeraldas		varcode6	Local
EC42	May 2014		>50-60	F	Esmeraldas, Esmeraldas	Colombia	varcode3	Local
EC43	May 2014		NA	NA	Cascales, Sucumbios	San Lorenzo, Esmeraldas	varcode1	Local
EC44	Jul 2014		>30-40	F	San Lorenzo, Esmeraldas		varcode7	Local
EC45	Aug 2014		>20-30	F	San Lorenzo, Esmeraldas		varcode7	Local
EC46	Aug 2014		>20-30	M	San Lorenzo, Esmeraldas		varcode7	Local
EC47	Aug 2014		>50-60	M	San Lorenzo, Esmeraldas		varcode7	Local
EC48	Aug 2014		>20-30	M	San Lorenzo, Esmeraldas		varcode7	Local
EC49	Sep 2014		NA	F	Lago Agrio, Sucumbios	Colombia	varcode8	Local
EC50	Oct 2014		NA	NA	Lago Agrio, Sucumbios	Colombia	varcode8	Local
EC51	Oct 2014		>30-40	M	San Lorenzo, Esmeraldas	Colombia	varcode6	Local
EC52	Dec 2014		>20-30	M	San Lorenzo, Esmeraldas		varcode7	Local
EC53	Jan 2015		NA	NA	Lumbaqui, Sucumbios	Colombia	varcode9	Colombia
EC54	Jan 2015		>20-30	M	San Lorenzo, Esmeraldas		varcode7	Local
EC55	Jan 2015		>10-20	M	San Lorenzo, Esmeraldas		varcode7	Local
EC56	Nov 2015		>10-20	F	San Lorenzo, Esmeraldas		varcode6	Local
EC57	May 2015		NA	NA	Tobar Donoso, Carchi		varcode1	Local
EC58	May 2015		NA	NA	Tobar Donoso, Carchi		varcode1	Local

878 NA refers to no data collected for that particular epidemiological variable.

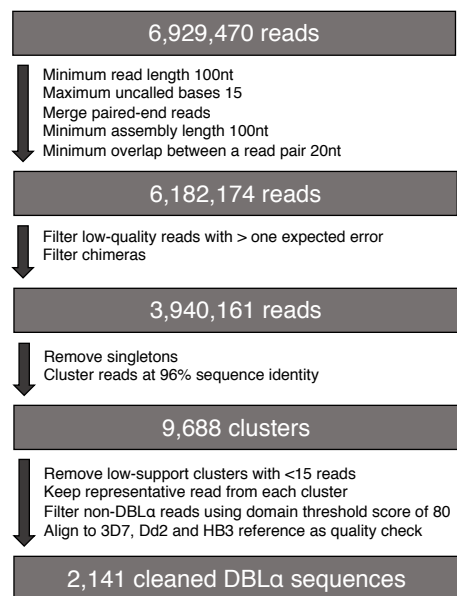
879 *The self-reported infection location was recorded at the time of sample collection based on the answers provided by the study participants.

880 **The putative origin as determined by varcoding.

881 Supplementary Information

882 Supplementary Figures

883



884

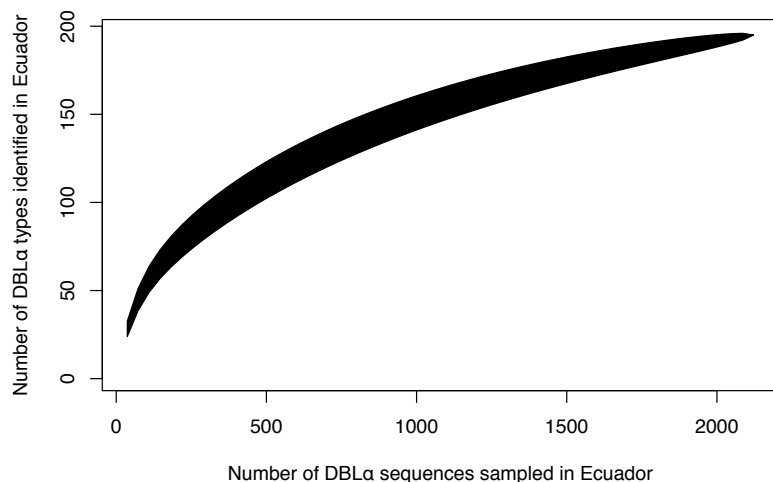
885

886

887

888

Figure S1. Bioinformatic sequence data processing flowchart. The flowchart shows the bioinformatic process to clean the raw illumina sequence reads with details on filter parameters for each step. This custom bioinformatic pipeline was developed to de-multiplex and remove PCR and sequencing artefacts and is described in detail in He et al ⁶⁰.



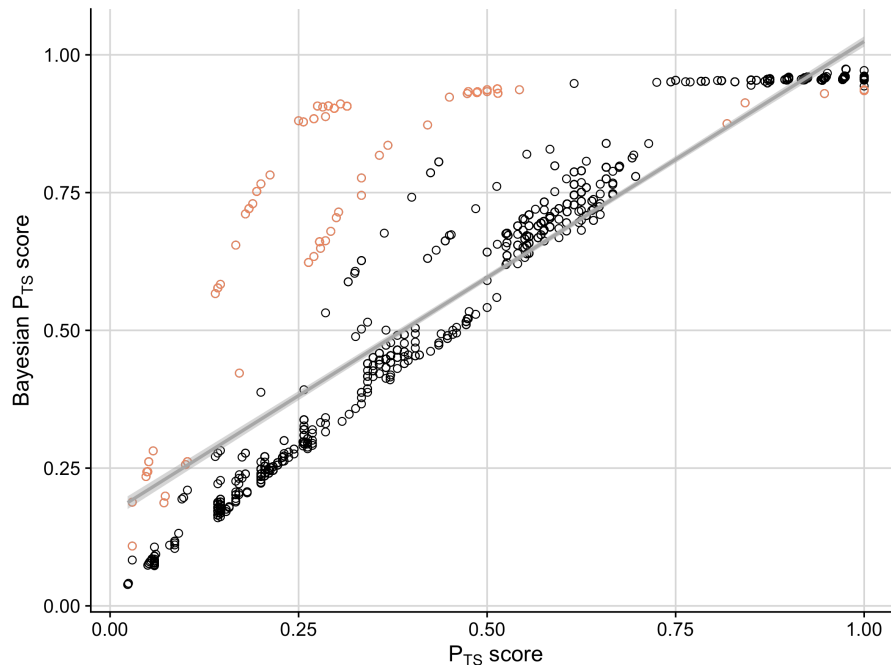
889

890

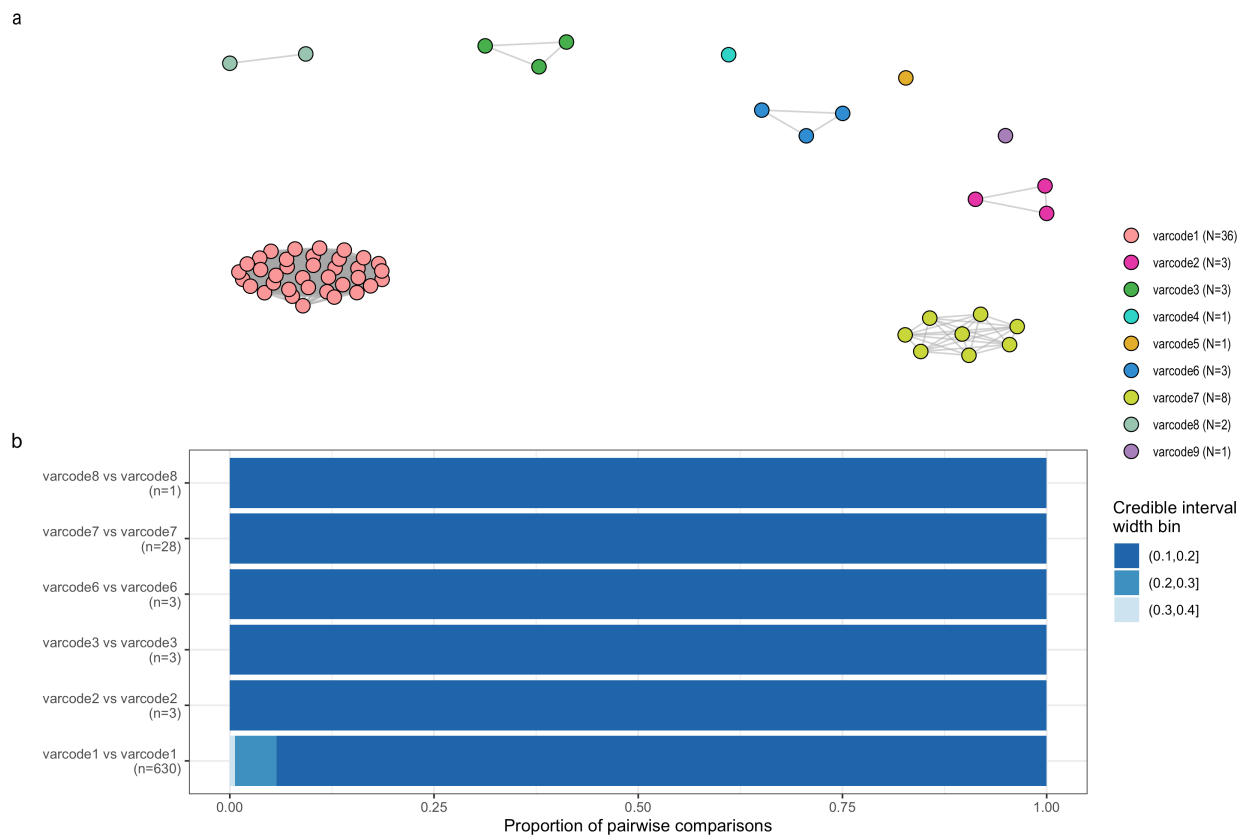
891

892

Figure S2. Sampling depth of *var* DBL α types in Ecuador. Accumulation curves depict the number of observed DBL α types plotted as a function of the number of DBL α sequences sampled.

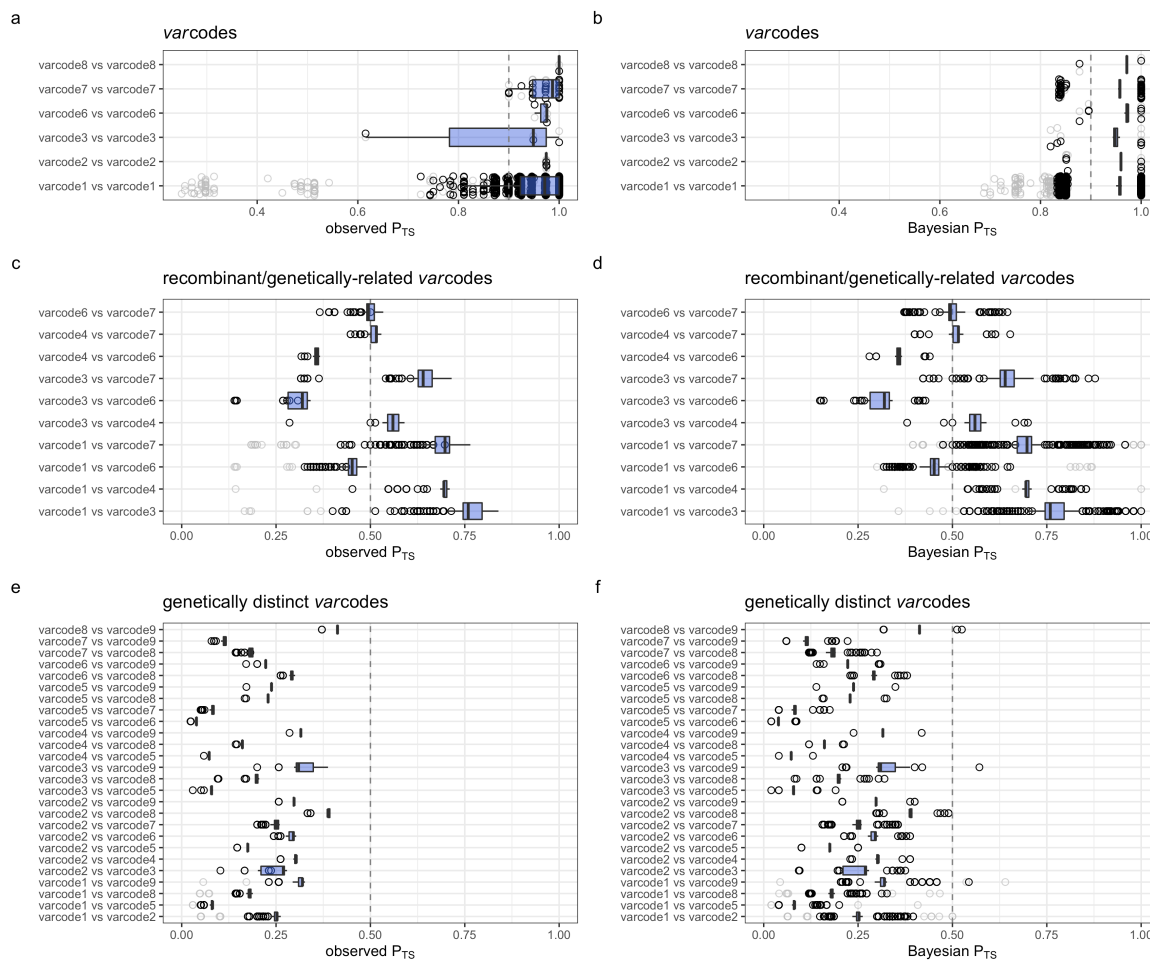


893
894 **Figure S3. Relatedness of isolates based on observed and unbiased Bayesian pairwise type sharing estimates.** There was a
895 positive correlation between the two relatedness measures (correlation coefficient = 0.919, $p < 0.001$). Circles colored in light red
896 correspond to pairwise comparisons involving the two *P. falciparum* isolates with the smallest repertoire sizes (11 and 19 DBL α
897 types), many of which had lower statistical confidence and the largest discrepancies between observed and unbiased estimates.
898

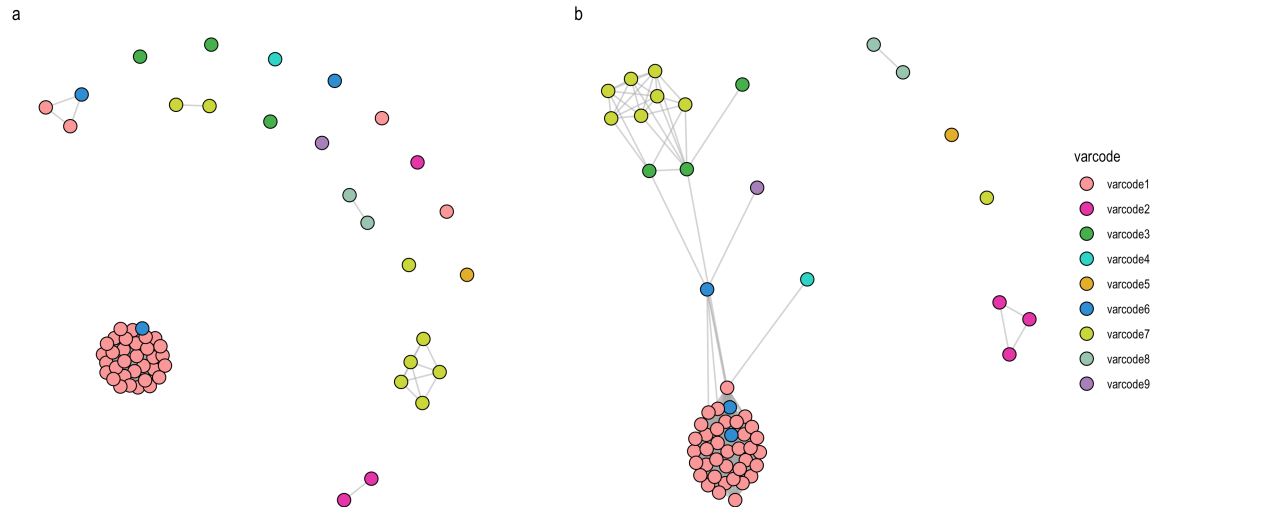


899

900 **Figure S4. Genetic relatedness networks based on Bayesian pairwise type sharing estimates to define varcodes.** A network
 901 visualization of the genetic relatedness of *P. falciparum* isolates at the threshold of (a) $BP_{TS} \geq 0.90$ to define clusters of genetically-
 902 related isolates based the mean posterior Bayesian PTS (BP_{TS}). Every node represents a *P. falciparum* isolate and an edge
 903 represents the BP_{TS} value between two particular nodes/isolates. Isolates that cluster together (i.e., connected by edges) are
 904 considered to be genetically identical (i.e., clones). Every color represents a different varcode. (b) The proportion of pairwise
 905 comparisons within a given 95% highest density posterior interval width, which provides a measure of the uncertainty of each
 906 pairwise estimate. Only varcodes with >1 *P. falciparum* isolate are shown. For the full distribution of within-varcode BP_{TS}
 907 estimates see Fig. S4b.
 908

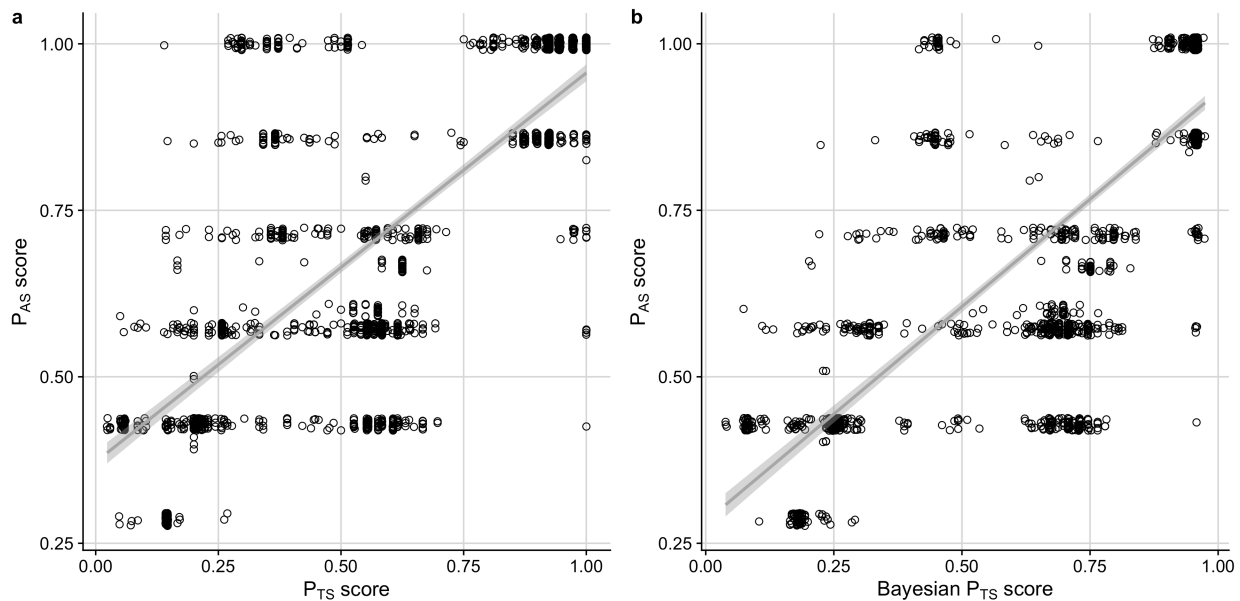


909 **Figure S5. Distribution of observed and Bayesian pairwise type sharing estimates used to define varcodes, as well as**
 910 **recombinant, genetically-related, and genetically distinct varcodes.** For observed P_{TS} estimates (a, c, e), the boxplots (in blue)
 911 show the median and interquartile range of the observed P_{TS} estimates (circles) for all pairwise comparisons among *P. falciparum*
 912 isolates for a given within-varcode or between-varcode comparison. For Bayesian P_{TS} estimates (b, d, f), the boxplots (in
 913 blue) show the median and interquartile range of the posterior mean BP_{TS} estimates. The circles show the BP_{TS} estimates for the
 914 lower and upper bound of the 95% highest density posterior intervals (HDPIs). Circles showing all of the posterior mean BP_{TS}
 915 estimates are omitted for clarity. Dashed lines indicate the threshold of (a-b) P_{TS} or $BP_{TS} \geq 0.90$ and (c-f) P_{TS} or $BP_{TS} \geq 0.50$. For
 916 comparison to the genetic relatedness network edges see Figures 2a, 2c, S3a, S7a. In all panels, light grey circles correspond to
 917 pairwise comparisons involving the two *P. falciparum* isolates with the smallest repertoire sizes (11 and 19 DBL α types), many of
 918 which had lower statistical confidence and the largest discrepancies between observed and unbiased estimates. Black circles
 919 denote all other estimates. Only varcodes with >1 *P. falciparum* isolate are shown.
 920
 921
 922
 923



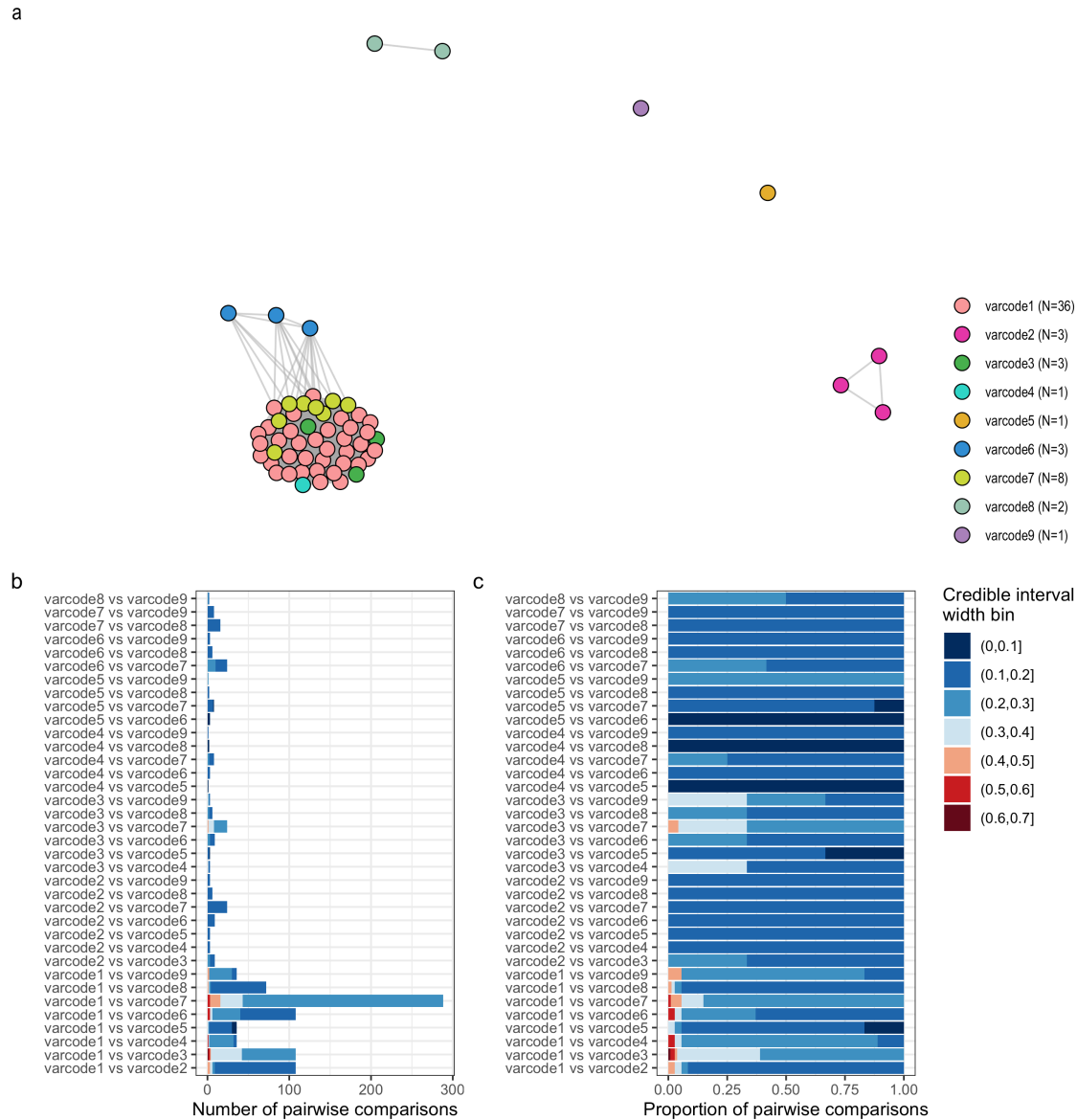
924
 925 **Figure S6. Genetic relatedness networks by microsatellite genotyping.** A network visualization of the genetic relatedness of *P. falciparum* isolates at the threshold of (a) $P_{AS} \geq 0.90$ and (b) $P_{AS} \geq 0.80$ to define clusters of genetically-related isolates based on microsatellite genotyping. Every node represents a *P. falciparum* isolate and an edge represents the P_{AS} value between two particular nodes/isolates. Isolates that cluster together (i.e., connected by edges) are considered to be genetically identical (i.e., clones) at the threshold of $P_{AS} \geq 0.90$ in (a) but can differ by 1 allele at the threshold of $P_{AS} \geq 0.80$ in (b). Every color represents a different varcode and are colored based on the definition of varcodes by var for comparative purposes. It is important to note that 2 out of the 7 microsatellite markers genotyped were fixed in the population⁸.

926
 927
 928
 929
 930
 931
 932
 933



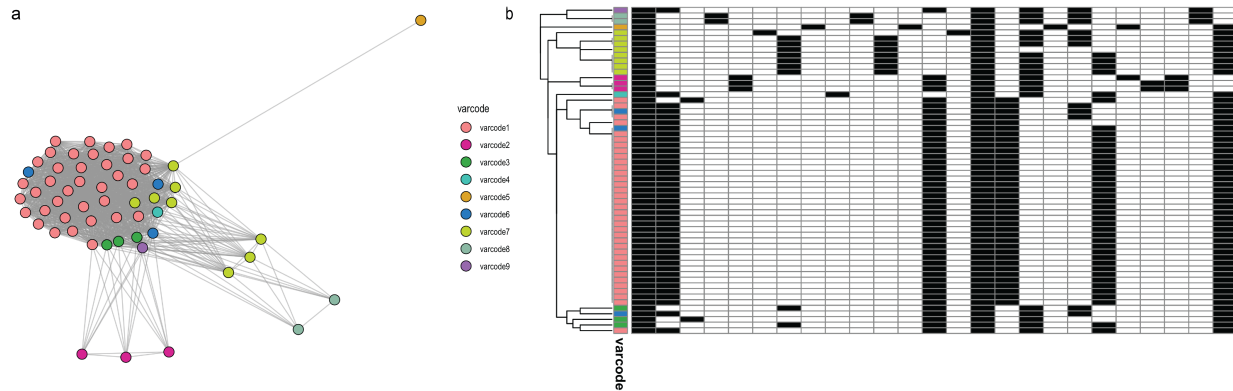
934
 935 **Figure S7. Relatedness of isolates based on var and microsatellite genotyping.** There was a positive correlation between (a) P_{TS} and P_{AS} relatedness measures (correlation coefficient = 0.757, $p < 0.001$) as well as (b) unbiased Bayesian P_{TS} and P_{AS} relatedness measures (correlation coefficient = 0.779, $p < 0.001$).

936
 937
 938

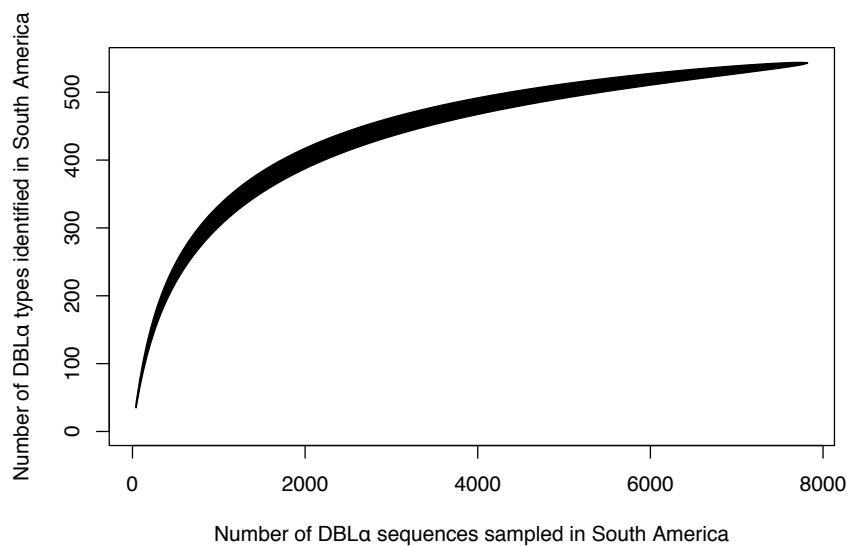


939
940
941
942
943
944
945
946
947
948

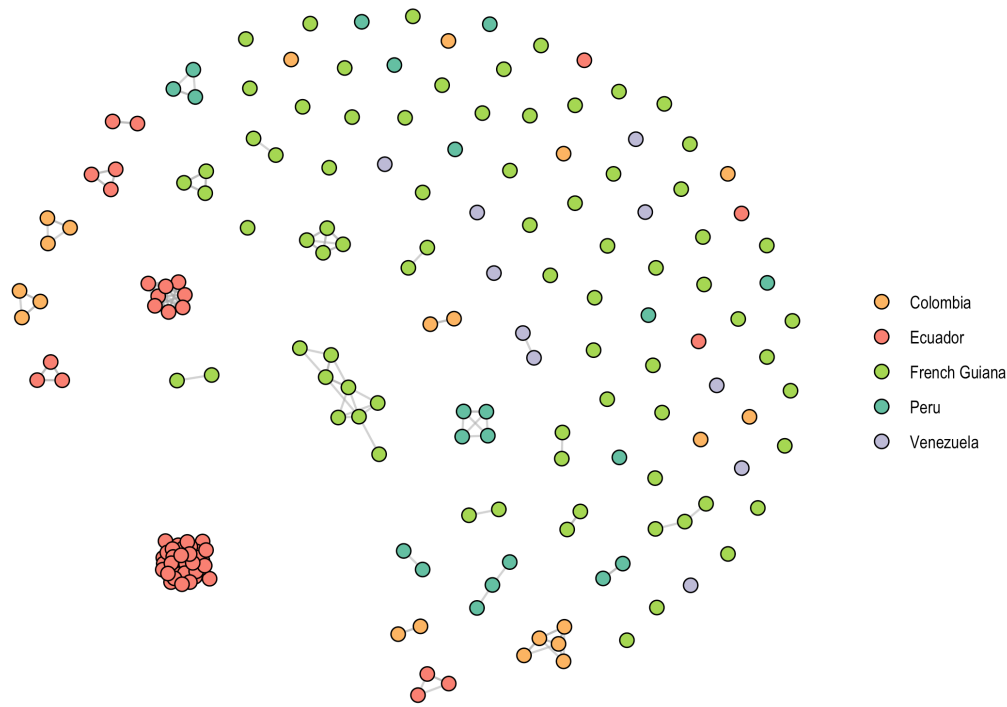
Figure S8. Genetic relatedness networks based on unbiased Bayesian pairwise type sharing estimates to identify recombinants. A network visualization of the genetic relatedness of *P. falciparum* isolates at the threshold of (a) $BP_{TS} \geq 0.50$ to define clusters of genetically-related isolates based on the posterior mean BP_{TS} estimates. Every node represents a *P. falciparum* isolate and an edge represents the BP_{TS} value between two particular nodes/isolates. Isolates that cluster together (i.e., connected by edges) are considered to be genetically identical (i.e., clones). Every color represents a different varcode. The (b) number and (c) proportion of pairwise comparisons within a given 95% highest density posterior interval width, which provides a measure of the uncertainty of each pairwise estimate. For the full distribution of within-varcode BP_{TS} estimates see Fig. S4d-f.



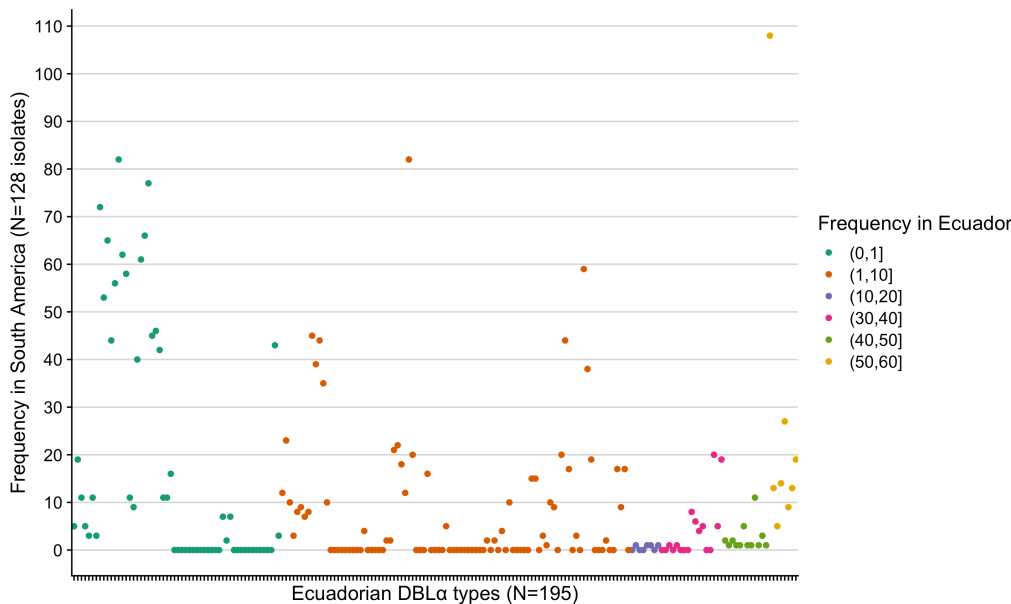
949
950 **Figure S9. Relatedness networks based on microsatellite genotyping.** (a) A network visualization of the genetic relatedness of *P.*
951 *falciparum* isolates at the threshold of $P_{AS} \geq 0.50$ colored by *varcode* for comparative purposes. Every node represents a *P.*
952 *falciparum* isolate and an edge represents the P_{AS} value between two particular nodes/isolates. (b) A clustered heatmap showing
953 the genetic profiles of each *P. falciparum* isolate with rows representing each isolate and columns representing each
954 microsatellite allele. Black and white denote the presence and absence of each allele, respectively. Isolates that clustered
955 together were more genetically similar (i.e., the same types were present).
956



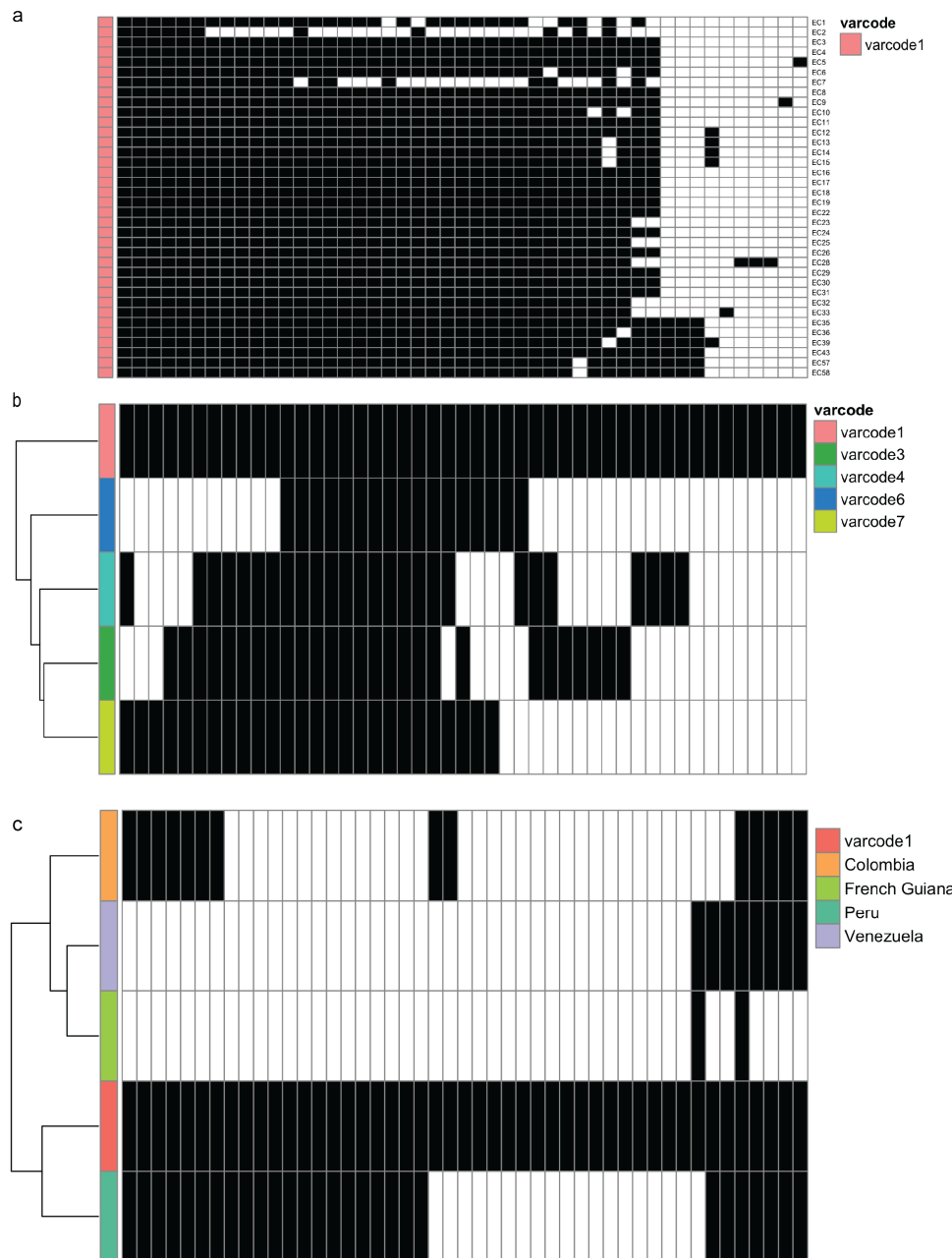
957
958 **Figure S10. Sampling depth of *var* DBL α types in South America.** Accumulation curves depict the number of observed DBL α types
959 plotted as a function of the number of DBL α sequences sampled.
960



961
 962 **Figure S11. South American varcodes.** A network visualization of the genetic relatedness of *P. falciparum* isolates at the threshold
 963 of $P_{TS} \geq 0.90$ to define clusters of genetically-related isolates, or *varcodes* in South America. Every node represents a *P. falciparum*
 964 isolate and an edge represents the P_{TS} value between two particular nodes/isolates. Isolates that cluster together (i.e., connected
 965 by edges) are considered to have the same *varcode*. The colors depict the country of origin/sampling location of each *P.*
 966 *falciparum* isolate.
 967



968
 969 **Figure S12. Frequency distribution of the 195 DBLα types identified in Ecuador.** Each point corresponds to an individual DBLα
 970 type and is colored by its frequency (i.e., the number of Ecuadorian *P. falciparum* isolates it was identified in). The y-axis depicts
 971 the frequency at which each Ecuadorian DBLα type was identified in the 128 *P. falciparum* isolates from South America. Of the 56
 972 “singleton” DBLα types (i.e., frequency of 1) in the Ecuadorian dataset, 32 (57.1%) were identified in other South American
 973 countries and 24 (42.9%) were not identified anywhere else in South America.



974
 975 **Figure S13. Identification of the outbreak varcode1 DBLα types.** Each column represents one of the 47 DBLα types and black and
 976 white denote the presence and absence of each type, respectively, in (a) the 36 infections with varcode1, (b) in the 15 infections
 977 with recombinant varcodes and (c) in South American *P. falciparum* populations. (a) The rows represent case numbers and are
 978 not clustered such that from top to bottom they are representative of sampling date (see Table 1 for details on sampling dates
 979 and whether they correspond to during or post-outbreak), and the columns are ordered from left to right in decreasing
 980 frequency of each DBLα type. *Note:* EC2 and EC7 were under sampled as compared to the other *P. falciparum* isolates (11 and 19
 981 DBLα types per isolate, respectively). (b-c) Clustered heatmaps show the identification of the 47 DBLα types from the outbreak
 982 varcode1 *P. falciparum* isolates in the other *P. falciparum* isolates from South America. A total of 31 (66%) of the outbreak types
 983 were identified in one or more countries. The highest number of types identified in another country was 28 (60%) in isolates
 984 collected in 2003-2004 from Peru, followed by 14 (30%) in isolates collected from 2002-2006 from Colombia, followed by 8 (17%)
 985 in isolates collected from 2003-2007 from Venezuela and only 2 (4%) in isolates collected from 2006-2008 from French Guiana.
 986 The isolate collection dates are reported in²⁸.

987 Supplementary Tables

988
989 Table S1. Summary statistics for total number of *var* DBL α types per isolate (i.e., repertoire size) in South American *P. falciparum*
990 populations.

Location	min*	med	mean	max
Colombia	19.0	40.0	38.4	43.0
Ecuador	11.0	37.0	36.6	43.0
French Guiana	13.0	48.0	50.5	92.0
Peru	19.0	36.0	33.4	42.0
Venezuela	28.0	36.5	35.2	43.0

991 *Note: all isolates with <10 types were removed due to low sequencing quality (see Methods)

Supplementary Text 1

1

2

3 As described in the Methods, the degenerate primers we use amplify the DBL α domain of *var* genes. We
4 performed an additional validation of our PCR (using these degenerate DBL α primers) and sequencing
5 methodology to understand the margin of error of detection of all DBL α types in an isolate. DBL α types were
6 translated into amino acid sequences and classified as upsA or non-upsA, using the classifyDBLalpha pipeline
7 (Ruybal-Pesántez et al. 2017) to examine whether the expected genomic proportions of upsA/non-upsA were
8 obtained in each isolate.

9 Given our study was conducted in South America, we used the Honduran laboratory reference strain HB3 as
10 a benchmark for the expected number of total DBL α types (i.e. repertoire size), as well as the proportion of
11 upsA/non-upsA. Whole genome sequencing of HB3 has identified 44 *var* genes, with 8 upsA DBL α types
12 (defined by DBL α domain 1), 34 non-upsA DBL α types (DBL α domains 0 and 2) and 2 upsE (*var2csa*
13 genes, defined by DBLpam domains and do not have DBL α domains) (Rask et al. 2010). We independently
14 verified this using 37 technical replicates of *var* DBL α PCR amplification and illumina sequencing of our
15 HB3 laboratory isolate. It is worth noting that since HB3 has two *var2csa* genes that do not have DBL α
16 domains, these will not be amplified by our degenerate primers, so we expect that only 42 of the *var* genes of
17 HB3 could be amplified.

18 HB3 technical replicates

19 From the data obtained from 37 HB3 technical replicates, we identified the expected repertoire sizes with
20 a median of 39 DBL α types (range: 36-41). A median of 7 upsA DBL α types (range: 6-8) and a median
21 of 33 non-upsA DBL α types (range: 30-34) were identified (Figure 1). The median genomic proportion of
22 upsA DBL α types was 17.5% (range: 15.4-19.5%) and 82.5% (range: 80.5-84.6%) for non-upsA DBL α types
23 (Figure 2). We found that of the 46 identified in the 37 technical replicates, 40 of them were consistently
24 identified in the majority of replicate isolates (range 21 to 37 replicates). Of these 40 types, 7 were ups-A
25 and 33 were non-upsA. All of these findings are in line with what is expected from whole genome sequencing
26 data (Rask et al. 2010).

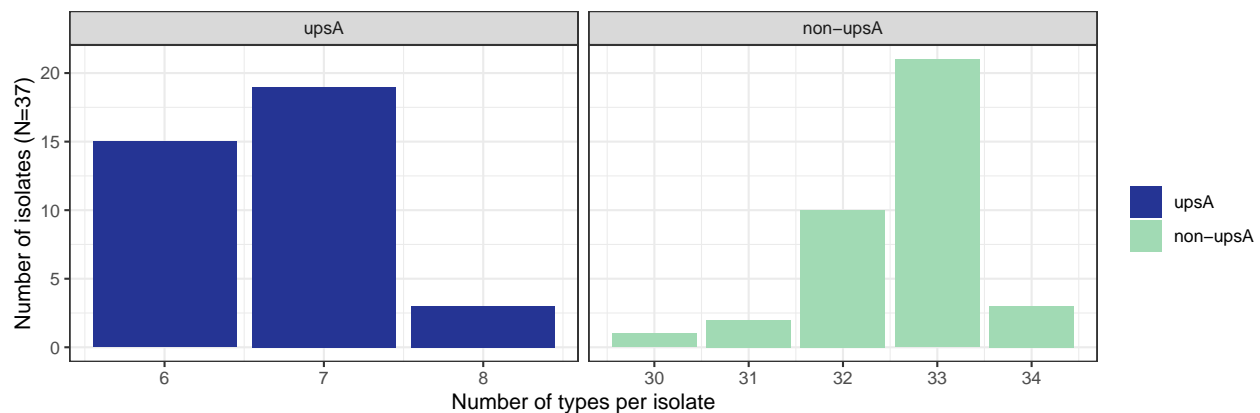


Figure 1: The distribution of the number of upsA and non-upsA types identified in each HB3 isolate repertoire.

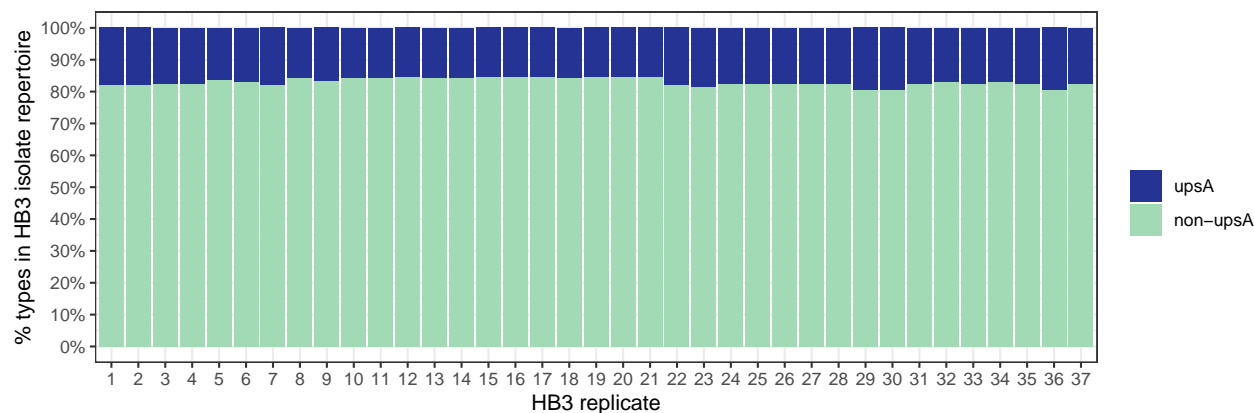


Figure 2: The proportion of upsA and non-upsA types identified in each HB3 isolate repertoire.

27 Field isolate genomic proportions of upsA and non-upsA DBL α types

28 The 543 unique DBL α types identified in the 186 South American *P. falciparum* isolates (N = 58 Ecuadorian
 29 *P. falciparum* isolates from this study, N = 128 previously published *P. falciparum* isolates from Colombia,
 30 French Guiana, Peru and Venezuela) were translated into amino acid sequences and classified as upsA or
 31 non-upsA, using the classifyDBLalpha pipeline (Ruybal-Pesántez et al. 2017). There were 79 upsA and 464
 32 non-upsA types.

33 Looking first at the 195 types (26 upsA and 169 non-upsA) identified in Ecuadorian isolates, we obtained a
 34 median genomic proportion of upsA of 10.8% (range: 5.1-18.2%) and 89.2% (range: 81.8-94.9%) of non-upsA
 35 types for all isolate *varcodes* (Figure 3).

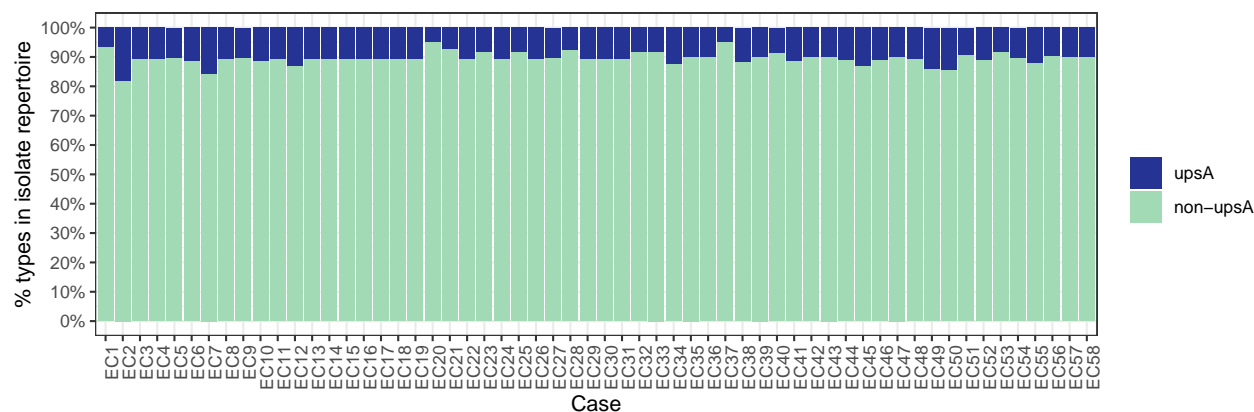


Figure 3: The proportion of upsA and non-upsA types in each isolate repertoire. The case numbers in the x-axis can be used to identify the clinical information of the participant in Table 1.

36 To confirm the patterns we observed in Ecuadorian isolates, we also compared them to the other South
 37 American isolates. In the other South American isolate *varcodes* the genomic proportions of upsA/non-upsA
 38 were similar, with a median proportion of upsA of 9-14% and 86-91% for non-upsA types (Table 1).

39 With regards to the number of upsA identified in all the South American isolate *varcodes*, we identified a
 40 median of 4-5 upsA types and 31-42.5 non-upsA types (Table 2).

Table 1: The proportion of upsA types in South American isolates

country	min	median	mean	max
Colombia	0.00	0.09	0.09	0.14
Ecuador	0.05	0.11	0.11	0.18
French Guiana	0.04	0.11	0.11	0.18
Peru	0.08	0.14	0.14	0.20
Venezuela	0.05	0.11	0.12	0.21

Table 2: The number of upsA types identified in South American isolates

country	min	median	mean	max
Colombia	0	4	4	6
Ecuador	2	4	4	6
French Guiana	2	5	6	14
Peru	2	5	5	7
Venezuela	2	4	4	9

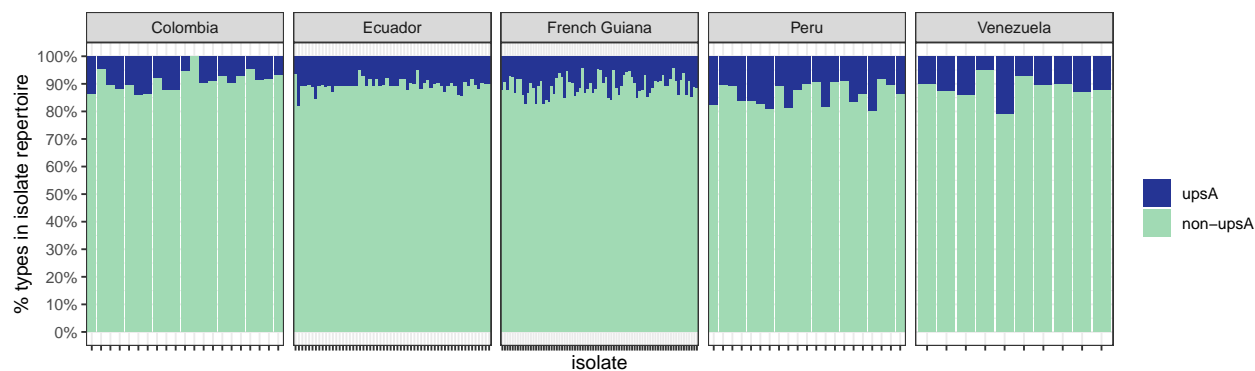


Figure 4: The proportion of upsA and non-upsA types in each isolate repertoire stratified by country.

41 **Inheritance of DBL_a types in the recombinant varcodes**

42 For the outbreak *varcode1*, its 47 DBL_a types were classified as 5 upsA and 42 non-upsA. In Figure 5 we
 43 look specifically at the “inheritance” of upsA (blues) vs non-upsA (greens) types from the parental outbreak
 44 *varcode1* in the case of the parasites with recombinant *varcodes* (*varcodes*3,4,6,7). This provides a proxy to
 45 examine inheritance of types with regards to their chromosomal location. The proportion of the outbreak
 46 types that were inherited in the recombinant parasite *varcodes* is indicated in the darker shades of blue or
 47 green, showing that the proportion of inherited types varied both by *varcode* and upsA/non-upsA. Overall,
 48 the DBL_a type sharing patterns in parasites with recombinant *varcodes* are consistent with inheritance of
 49 50% types, with a higher proportion of non-upsA inherited types (~40-70%, i.e. 17 to 30 of the 42 types) vs
 50 upsA (~20-50%, i.e. 1 to 3 of the 5 upsA types). The exception was *varcode7* where 60-80% of upsA were
 51 inherited, i.e. 3-4 of the 5 upsA types. The lighter shades correspond to those types that were not inherited
 52 from the outbreak clone but from the other parent.

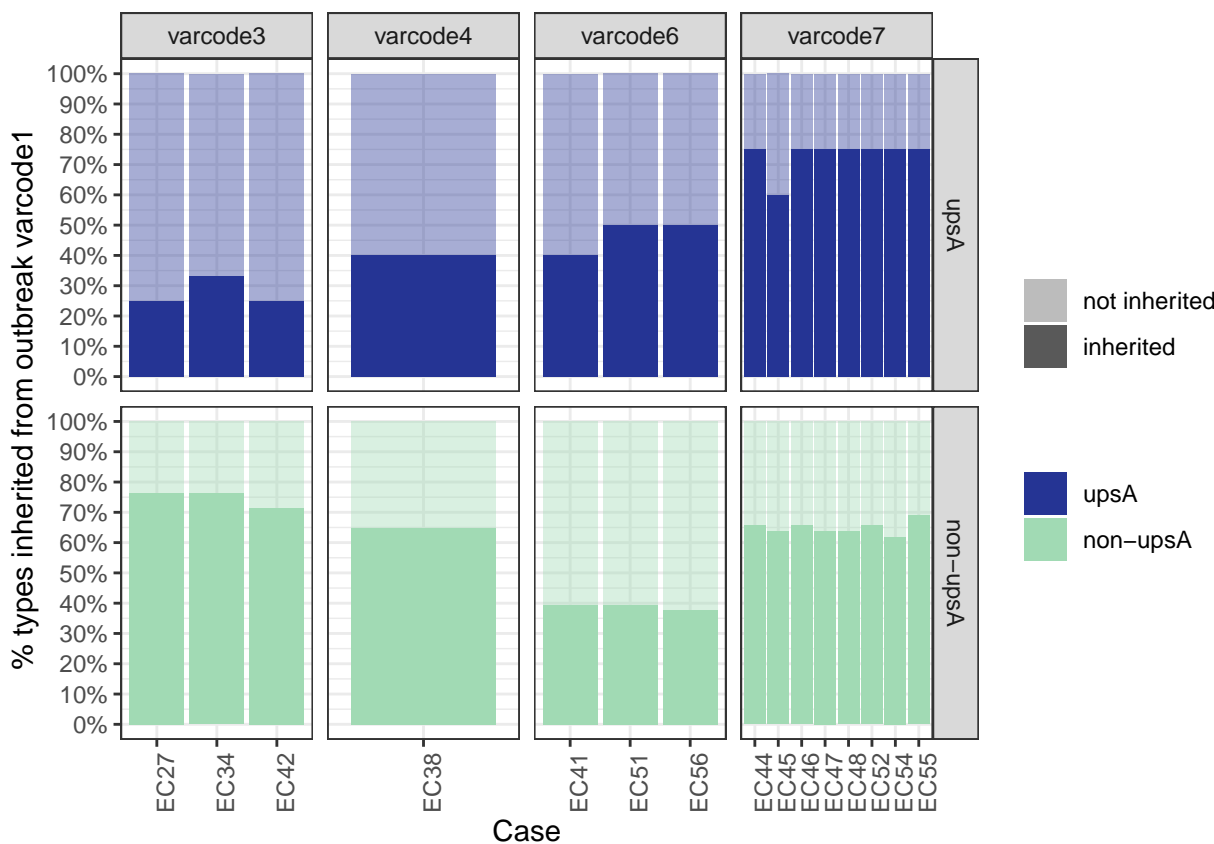


Figure 5: The proportion of the outbreak types that were inherited in the recombinant parasite *varcodes* is indicated in the darker shades of blue or green. The proportion of the outbreak types that were inherited in the recombinant parasite *varcodes* is indicated in the darker shades of blue or green and the lighter shades correspond to those types that were not inherited from the outbreak clone but from the other parent.

53 **References**

54 Rask, Thomas S., Daniel A. Hansen, Thor G. Theander, Anders Gorm Pedersen, and Thomas Lavstsen.
 55 2010. “Plasmodium Falciparum Erythrocyte Membrane Protein 1 Diversity in Seven Genomes Divide
 56 and Conquer.” Edited by Jonathan A. Eisen. *PLoS Computational Biology* 6 (9): e1000933. <https://doi.org/10.1371/journal.pcbi.1000933>.
 57

- 58 Ruybal-Pesántez, Shazia, Kathryn E. Tiedje, Gerry Tonkin-Hill, Thomas S. Rask, Moses R. Kamya, Bryan
59 Greenhouse, Grant Dorsey, Michael F. Duffy, and Karen P. Day. 2017. “Population Genomics of
60 Virulence Genes of Plasmodium Falciparum in Clinical Isolates from Uganda.” *Scientific Reports* 7 (1).
61 <https://doi.org/10.1038/s41598-017-11814-9>.