Measuring changes in *Plasmodium falciparum var* census population size and structure in response to sequential malaria control interventions

Kathryn E. Tiedje[1,2§], Qi Zhan[3,4§], Shazia Ruybal-Pésantez[2†], Gerry Tonkin-Hill[2,5], Qixin He[4‡], Mun Hua Tan[1], Dionne C. Argyropoulos[1], Samantha L. Deed[1,2], Anita Ghansah[6], Oscar Bangre[7], Abraham R. Oduro[7], Kwadwo A. Koram[8], Mercedes Pascual[4,9], and Karen P. Day[1,2]

[1] Department of Microbiology and Immunology, Bio21 Institute and Peter Doherty Institute, The University of Melbourne; Melbourne, Australia
[2] School of BioSciences, Bio21 Institute, The University of Melbourne; Melbourne, Australia
[3] Committee on Genetics, Genomics and Systems Biology, The University of Chicago; Chicago, Illinois, USA
[4] Department of Ecology and Evolution, The University of Chicago; Chicago, Illinois, USA
[5] Bioinformatics Division, Walter and Eliza Hall Institute; Melbourne, Australia
[6] Department of Parasitology, Noguchi Memorial Institute for Medical Research, University of Ghana; Legon, Ghana
[7] Navrongo Health Research Centre, Ghana Health Service; Navrongo, Ghana
[8] Epidemiology Department, Noguchi Memorial Institute for Medical Research, University of Ghana; Legon, Ghana
[9] Santa Fe Institute, Santa Fe, New Mexico, USA

† *Current affiliations:* Department of Infectious Disease Epidemiology and MRC Centre for Global Infectious Disease Analysis, School of Public Health, Imperial College London, United Kingdom

‡ *Current affiliation:* Department of Biological Sciences, Purdue University; West Lafayette, United States

§ These authors share first authorship

* Corresponding Author: Karen P. Day, Department of Microbiology and Immunology, Bio21 Institute and Peter Doherty Institute, The University of Melbourne, 30 Flemington Road, Parkville, VIC, 3011, Australia. E-mail: Karen.Day@unimelb.edu.au, Phone: +613 8344 2377

RUNNING TITLE: Impact of IRS and SMC on *P. falciparum var* diversity

ABSTRACT

The diversity of *Plasmodium falciparum* within human hosts requires parasite population size be defined in terms of parasite variation rather than the number of infected hosts. To calculate census population size, we rely on a definition of parasite variation known as multiplicity of infection (MOI_{var}), defined by the hyper-diversity of the *var* multigene family. We present a Bayesian approach to estimate MOI_{var} based on sequencing and counting the number of unique DBLα tags (or DBLα types) of *var* genes, taking into consideration measurement error, and derive from it census population size or the total number of distinct infections of relevance to transmission events. We track changes in parasite population size and structure, using MOI_{var}, from baseline and through sequential malaria interventions by indoor residual spraying (IRS) and seasonal malaria chemoprevention (SMC) in an area characterized by high-seasonal malaria transmission in northern Ghana. *Var* DBLα tag sequencing was completed on asymptomatic *P. falciparum* isolates at baseline (2012), during IRS (2014), post-IRS (2015) and during SMC (2017) from ~2,000 individuals of all ages surveyed at each time point. Following IRS, which reduced transmission intensity by > 90% and decreased parasite prevalence by ~40-50%, significant reductions in *var* diversity, MOI_{var}, and population size were observed across all ages. These changes, consistent with the loss of diverse parasite genomes, were short lived and 32-months after IRS was discontinued and SMC was introduced, *var* diversity and population size rebounded in all age groups except for the younger children (1-5 years) targeted by SMC. By measuring population size in this way, we show that despite major perturbations, the parasite population remained very large and retained the *var* population genetic characteristics of a high-transmission system (high *var* diversity; low *var* repertoire similarity) demonstrating the resilience of *P. falciparum* to short-term interventions in high-burden countries of sub-Saharan Africa.

61 # INTRODUCTION

62 Malaria in high-transmission endemic areas of sub-Saharan Africa (SSA) is characterised by vast diversity
63 of the *Plasmodium falciparum* parasites from the perspective of antigenic variation (Chen et al., 2011; Day
64 et al., 2017; Otto et al., 2019; Ruybal-Pesántez et al., 2022, 2017). As for other hosts of hyper-variable
65 pathogens (Futse, Brayton, Dark, Knowles, & Palmer, 2008), children experiencing clinical episodes of
66 malaria, eventually become immune to disease but not to infection. This results in a large reservoir of
67 chronic asymptomatic infections, in hosts of all ages, which sustain transmission to mosquitos. Given the
68 goal of malaria eradication by 2050, it is therefore of interest to examine how parasite population size
69 changes following perturbation by major intervention efforts, both in terms of its size and underlying
70 population genetics.
71
72 So, what do we mean by the parasite population size in the case of *P. falciparum* and how do we measure
73 it? Parasite prevalence, detected by microscopy or more sensitive molecular diagnostics (e.g., PCR),
74 describes the proportion of infected hosts. Studies of *P. falciparum* genetic diversity have shown that the
75 majority of people in high-transmission endemic areas harbour diverse multiclonal infections measured
76 as the complexity or multiplicity of infection (MOI) (e.g., (Anderson et al., 2000; Paul et al., 1995; Smith,
77 Felger, Fraser-Hurt, & Beck, 1999; Sumner et al., 2021)) with complex population dynamics (Bruce et al.,
78 2000; Farnert, Snounou, Rooth, & Bjorkman, 1997). These genetic data indicate much larger population
79 sizes than observed by prevalence alone. Thus, from an ecological perspective we can consider a human
80 host as a patch carrying a number of distinct "antigenic strains" of *P. falciparum*. The sum of these distinct
81 infections over all hosts provides us with a census of the parasite count of relevance to transmission
82 events in the host population. We refer to this census population size hereafter simply as population size
83 but make clear that this measure is distinct from effective population size (N_e) as measured by neutral
84 variation.
85
86 *P. falciparum* erythrocyte membrane protein 1 (PfEMP1), encoded by the *var* multigene family, is a strong
87 contender for the strain-defining antigen of the blood stages of this malaria parasite due its role in
88 establishing chronic infection (Zhang & Deitsch, 2022). The molecule mediates sequestration of *P.*
89 *falciparum* asexual stage parasites by cytoadherence in specific tissues to allow parasites to mature while

avoiding splenic mechanisms of clearance. PfEMP1 also facilitates immune evasion by a process known as clonal antigenic variation involving differential expression of ~50-60 genes of the *var* multigene family per haploid genome (i.e., *var* repertoire) (Gardner et al., 2002; Otto et al., 2018; Rask et al., 2010; Zhang & Deitsch, 2022). This results in the establishment of chronic infections within humans to ensure onward transmission. These features of *P. falciparum* biology are particularly important in SSA where malaria transmission can be highly seasonal and the parasite must survive a long dry season until rains bring an increase in anopheline mosquitos. PfEMP1 is highly immunogenic (Chan, Fowkes, & Beeson, 2014) and likely under strong immune selection as it is exposed on the surface of the infected erythrocyte from trophozoite to schizont stage for ~24 hours (Leech, Barnwell, Miller, & Howard, 1984), compared to seconds to minutes for other blood stage surface antigens, which are less diverse (Naung et al., 2022).

We have used a ~450bp region of a *var* gene, known as a DBLα tag encoding the immunogenic Duffy-binding-like alpha (DBLα) domain of *var* genes, to characterize *var* diversity (Barry et al., 2007) with 454 (Day et al., 2017; Rorick et al., 2018; Ruybal-Pesántez et al., 2017) and Illumina (Ruybal-Pesántez et al., 2022) high-throughput protocols to sequence *var* genes from *P. falciparum* isolates in SSA. Bioinformatic analyses of a large database of exon 1 sequences of *var* genes showed a predominantly 1-to-1 DBLα-*var* relationship, such that each DBLα tag typically represents a unique *var* gene (Tan, Shim, Chan, & Day, 2023). Population investigations in high-transmission areas of SSA have demonstrated that DBLα tags are highly diverse in local endemic areas (Chen et al., 2011; Day et al., 2017; Ruybal-Pesántez et al., 2022, 2017). Bioinformatic analyses show that recombination drives this high diversity (Bopp et al., 2013; Claessens et al., 2014; Feng et al., 2022) although we have found that many of these DBLα tags persist in parasite populations in space and time, even in high-transmission endemic areas in SSA (Day et al., 2017; Ruybal-Pesántez et al., 2022, 2017).

Var genes lie on all 14 chromosomes of the haploid genome and will undergo independent assortment during meiosis, an obligatory part of the life cycle in the mosquito. This means each time there is a cross between two genetically distinct parasites (i.e., outcrossing) or “strains”, *var* repertoires will recombine and share ~50% of the parental genes. Yet, to our surprise, we found *var* repertoires in chronic, asymptomatic infections in parasite populations of SSA to be composed of predominantly non-

119 overlapping sets of DBLα tags (Day et al., 2017; Rorick et al., 2018; Ruybal-Pesántez et al., 2022, 2017).
120 This observed pattern of repertoire diversity is not random, as shown with comparisons to randomizations
121 that preserve gene frequencies (Day et al., 2017) and also not neutral, as shown by comparison of
122 empirical patterns of repertoire similarity to those generated by a computational model (He et al., 2018).
123 This aspect of *var* population structure facilitates the accumulation of multiple diverse parasite repertoires
124 detectable in the blood of the chronically infected semi-immune human hosts in an age-specific manner
125 (Ruybal-Pesántez et al., 2022). As a consequence of the non-overlapping structure, there is a direct
126 relationship between the number of *var* repertoires in a host and population size in that host or patch.
127 Adding up the patches or MOI_{var} (i.e., the number of *var* repertoires based on *var*coding (Ruybal-Pesántez
128 et al., 2022; Tiedje et al., 2022)) of individual hosts gives a measure of overall population size and a metric
129 to evaluate the impact of interventions which reflects not only changes in transmission intensity but decay
130 of chronic infections to shorten what is known as the duration of infection (Macdonald, 1957).
131
132 Considering the part of the genome under the strongest immune selection due to variant-specific
133 immunity is akin to what has been recognized as central in pathogen phylodynamics, which examines the
134 interplay between epidemiology, immunology, and evolutionary processes (Grenfell et al., 2004; Volz,
135 Koelle, & Bedford, 2013). We are very familiar with this emphasis in relation to influenza strains and
136 COVID-19 causing strains. With respect to pathogens with large population sizes, as is likely the case for
137 *P. falciparum* in SSA, a view to measure diversity in adaptive genes rather than focus on neutral theory
138 to understand population size has been presented (Barton, 2010). Here we examine the response to
139 sequential malaria interventions with a definition of *P. falciparum* "strains" in a manner consistent with
140 this view. Defining "strains by *var* repertoires of unlinked genes" diversifying by sexual recombination also
141 allows us to take into consideration the dynamic nature of *P. falciparum* genomes caused by the high
142 frequency of outcrossing expected in moderate- to high-transmission endemic areas (Babiker et al., 1994;
143 Paul et al., 1995).
144
145 We specifically report an investigation of changes in parasite population size and structure through two
146 sequential interventions between 2012 and 2017 in Bongo District located in the Upper East Region of
147 Ghana, one of the 12 highest burden countries in Africa (World Health Organization, 2022). We present a

148 novel Bayesian approach to estimate MOI and therefore population size, which is based on counting the
149 number of DBLα tags (or DBLα types) (Ruybal-Pesántez et al., 2022; Tiedje et al., 2022) but takes into
150 consideration under-sampling of DBLα types in an isolate. We document *P. falciparum* prevalence as well
151 as *var* diversity and population structure from baseline in 2012 through a major perturbation by a short-
152 term indoor residual spraying (IRS) campaign managed under operational conditions, which reduced
153 transmission intensity by > 90% as measured by the entomological inoculation rate (EIR) and decreased
154 parasite prevalence by ~40-50% (Tiedje et al., 2022). Next, we followed what happened to parasite
155 population size more than two years after the IRS intervention was discontinued and seasonal malaria
156 chemoprevention (SMC) was introduced for children between the ages of 3-59 months (i.e., < 5 years)
157 (Wagman et al., 2018). Detectable changes in parasite population size were seen as a consequence of the
158 IRS intervention but this quantity rapidly rebounded 32-months after the intervention ceased. Overall,
159 throughout the IRS, SMC, and subsequent rebound, the parasite population in humans remained large in
160 size and retained the *var* population genetic characteristics of high transmission (i.e., high *var* diversity,
161 low *var* repertoire overlap), demonstrating the overall resilience of the species to survive significant short-
162 term perturbations.

METHODS

Human subject ethical approval

The study was reviewed/approved by the ethics committees at the Navrongo Health Research Centre (Ghana), Noguchi Memorial Institute for Medical Research (Ghana), The University of Melbourne (Australia), and The University of Chicago (United States). Details on the study area, study population, inclusion/exclusion criteria, and data collection procedures have been previously described (Tiedje et al., 2017, 2022).

Study design and sample collection

Using an interrupted time-series study design, four age-stratified cross-sectional surveys of ~2,000 participants per survey were undertaken to investigate the impacts of IRS and SMC in combination with long-lasting insecticide-treated nets (LLINs) under operational conditions on the asymptomatic *P. falciparum* reservoir from two proximal catchment areas (i.e., Vea/Gowrie and Soe, with a sampling area of ~60 km^2) in Bongo District, Ghana (Table S1). Bongo District, located in the Upper East Region, is categorized as high-seasonal malaria transmission based on the World Health Organization's (WHO) "A Framework for Malaria Elimination" (WHO/GMP, 2017) where *P. falciparum* prevalence was ≥ 35% at baseline in 2012 (Tiedje et al., 2017, 2022). These ~2,000 participants of all ages (1-97 years) represent ~15% of the total population that reside in these two catchment areas in Bongo (Tiedje et al., 2017). The four cross-sectional surveys were completed at the end of the wet season (i.e., high-transmission season) and the study can be separated into four distinct study time points: (1) October 2012 (Survey 1) prior to the IRS and SMC (i.e., baseline), (2) October 2014 (Survey 2) two months after the second round of IRS (Actellic 50EC), (3) October 2015 (Survey 3) seven months after the third round of IRS using a long-acting non-pyrethroid insecticide (Actellic 300CS) (Gogue et al., 2020; US President's Malaria Initiative Africa IRS (AIRS) Project, 2016), and finally (4) October 2017 (Survey 4) 32-months after the discontinuation of IRS, but during the deployment of SMC to all children 3-59 months (i.e., < 5 years) (Figure 1). Details on the study area, study population, and data collection procedures have been previously described (Tiedje et al., 2017, 2022).

Details of the IRS and SMC interventions

192 Starting in 2013, the AngloGold Ashanti Malaria Control Programme (AGAMal) in a public-private
193 partnership with the Global Fund, scaled up IRS across all of the Upper East Region of northern Ghana
194 (Gogue et al., 2020). As part of this initiative, three rounds of IRS with organophosphate formulations (i.e.,
195 non-pyrethroid) were rolled out prior to the start of the wet season between 2013 and 2015 (Figure 1A)
196 in Bongo District (Tiedje et al., 2022). Based on AGAMal's operational reports, IRS coverage in Bongo
197 District was 91.8% in Round 1, 95.6% in Round 2, and finally 96.6% in Round 3 (AGAMal, personal
198 communication). To monitor the impact of the IRS on the local vector population, monthly entomological
199 surveys were undertaken between February 2013 and September 2015 (Tiedje et al., 2022). Using these
200 surveys we observed that the monthly entomological inoculation rate (EIR) (infective bites/person/month
201 (ib/p/m)), a measure of local transmission intensity, declined by > 90% at the peak of the wet season
202 between August 2013 (pre-IRS) (EIR = 5.3 ib/p/m) and August 2015 (post-IRS) (EIR = 0.4 ib/p/m) (Tiedje et
203 al., 2022). Following the IRS, SMC was rolled out in the Upper East Region by the National Malaria
204 Elimination Programme (NMEP)/ Ghana Health Service (GHS) starting in 2016 (Gogue et al., 2020) (Figure
205 1A). SMC is the intermittent administration of full treatment courses of an antimalarial to children
206 between the ages of 3-59 months (i.e., < 5 years) (WHO, 2012). Like other countries, the SMC drug of
207 choice in Ghana is amodiaquine plus sulfadoxine-pyrimethamine, which is administered at monthly
208 intervals during the high-transmission season (i.e., wet season) (WHO, 2012). The goal of this age-targeted
209 intervention is to both clear current malaria infections and prevent malarial illness by maintaining a
210 therapeutic concentration of an antimalarial in blood over the period of greatest risk (i.e., high-
211 transmission season). Reported SMC coverage in Bongo District was 92.6% in 2016 (two rounds between
212 August – September 2016) and 94.6% in 2017 (four rounds between September – December 2017)
213 (Gogue et al., 2020).
214

215 ***Var* genotyping and sequence analysis**

216 Genomic DNA was extracted from the dried blood spots for all participants with a confirmed microscopic
217 asymptomatic *P. falciparum* infection (i.e., isolate) (2,572 isolates, Table S2) using the QIAamp DNA mini
218 kit (QIAGEN, USA) as previously described (Tiedje et al., 2017). For *var* genotyping or *var*coding, the
219 sequences encoding the DBLα domains of the antigen-encoding *P. falciparum var* genes were amplified
220 by PCR, pooled, and sequenced on a MiSeq 2x300 paired-end cycle protocol as previously published (New

221 York University Genome Technology Center, New York, NY, USA; Australian Genome Research Facility,
222 Melbourne, Australia) (Figure S1) (Day et al., 2017; He et al., 2018; Ruybal-Pesántez et al., 2017). The raw
223 sequence data was then cleaned and processed using our previously published customized bioinformatic
224 pipeline (https://github.com/UniMelb-Day-Lab/DBLaCleaner) (He et al., 2018). This pipeline was used to
225 de-multiplex and merge the paired reads as well as remove low-quality sequences and chimeras using
226 several filtering parameters (see Figure S2 for additional details). These steps resulted in a total of 291,498
227 cleaned DBLα sequences for all surveys (Figure S2). To identify the unique DBLα types, we then clustered
228 these cleaned DBLα sequences with 237,910 DBLα sequences previously published from Bongo (He et al.,
229 2018; Pilosof et al., 2019; Rorick et al., 2018) at the standard 96% sequence identity
230 (https://github.com/UniMelb-Day-Lab/clusterDBLalpha) (Day et al., 2017). Our dataset was then further
231 curated by translating the DBLα types (N = 68,507) into amino acid sequences and removing any DBLα
232 types that could not be translated (i.e., contained a stop codon) (N = 137; 0.2%). The remaining DBLα
233 types were then assigned to their most likely DBLα domain class (i.e., DBLα0, DBLα1, or DBLα2) using a
234 hidden Markov model, and further classified based on the association of specific domain classes with
235 semi-conserved upstream promoter sequences (ups) as either upsA or non-upsA DBLα types
236 (https://github.com/UniMelb-Day-Lab/classifyDBLalpha) (Ruybal-Pesántez et al., 2017). For additional
237 information on the use of these bioinformatic pipelines a detailed tutorial is available
238 (https://github.com/UniMelb-Day-Lab/tutorialDBLalpha). Using a cut off of ≥ 20 DBLα types, DBLα
239 sequencing data was obtained from 2,099 *P. falciparum* isolates (81.6%) with a total of 53,241 unique
240 DBLα types and 289,064 DBLα sequences being identified in the study population (Table S2). This
241 genotyping success was acceptable given we were working with low-density asymptomatic infections
242 (Table S1). The median *P. falciparum* density was ~4 times higher for isolates with ≥ 20 DBLα types
243 compared those that gave no or < 20 DBLα types (520 [IQR: 200-1880] parasites/µL vs. 120 [IQR: 40-200]
244 parasites/µL, respectively).
245

246 DBLα type diversity

247 We monitored the impacts of the sequential interventions (i.e., IRS and SMC) on diversity by measuring
248 changes in the population genetics of DBLα types at the population level (i.e., *P. falciparum* reservoir).
249 Diversity was monitored using two measures, DBLα type richness and DBLα type frequency. Richness was

250 defined as the number of unique DBLα types observed (i.e., DBLα type pool size) in 2012, 2014, 2015, and
251 2017. DBLα type richness, however, does not provide any information about the relative frequencies of
252 the different DBLα types in the population, as they are all weighted equally whether they are observed
253 once or more frequently (e.g., observed in > 20 isolates per survey). To further examine the impacts of
254 the interventions on DBLα type diversity we also assessed the frequency of each unique DBLα type in
255 2012, 2014, 2015, and 2017. Both upsA and non-upsA DBLα type diversity was measured due to their
256 different biological features, chromosomal positions (i.e., subtelomeric regions vs. internal or central
257 regions), as well as population genetics (Falk et al., 2009; Gardner et al., 2002; Jensen et al., 2004; Kaestli
258 et al., 2006; Kalmbach et al., 2010; Kraemer et al., 2007; Kraemer & Smith, 2006; Kyriacou et al., 2006;
259 Lavstsen, Salanti, Jensen, Arnot, & Theander, 2003; Normark et al., 2007; Rottmann et al., 2006; Warimwe
260 et al., 2012, 2009; Zhang & Deitsch, 2022). The proportion of upsA and non-upsA *var* genes in a repertoire
261 or single genome has been defined as ~15-20% and ~80-85%, respectively, based on whole genome
262 sequencing (Rask et al., 2010). The upsA and non-upsA DBLα type proportions were partitioned as
263 expected in our analyses, with the median proportions at the repertoire level being comparable in 2012
264 (19% upsA and 81% non-upsA), 2014 (22% upsA and 78% non-upsA), 2015 (21% upsA and 79% non-upsA),
265 and 2017 (20% upsA and 80% non-upsA).
266
267 Repertoire similarity as defined by pairwise type sharing
268 To estimate genetic similarity between the DBLα repertoires (i.e., unique DBLα types identified in each
269 isolate) identified from two isolates, pairwise type sharing (PTS) was calculated between all pairs of
270 isolates within in each survey as previously described (Barry et al., 2007). PTS, analogous to the Sørensen
271 Index, is a similarity statistic to evaluate the proportion of DBLα types shared between two isolate
272 repertoires (i.e., DBLα repertoire overlap) and ranges from 0 (i.e., no DBLα repertoire overlap) to 1 (i.e.,
273 identical DBLα isolate repertoires), where < 0.50 = unrelated, 0.5 = recent recombinants/siblings, > 0.5 =
274 related, and 1 = clones. PTS is a measure of identity-by-state (IBS) used to assess repertoire similarity (or
275 relatedness) between isolates and is not used to infer inheritance from a recent common ancestor (i.e.,
276 identity-by-decent (IBD)) (Speed & Balding, 2015).
277
278 DBLα isolate repertoire size and number of *P. falciparum var* repertories

279 For this study we have exploited the unique population structure of non-overlapping DBLα isolate
280 repertoires to estimate isolate MOI_{var}. To calculate MOI_{var} the non-upsA DBLα types were chosen since
281 not only are they more diverse and less conserved between isolate repertoires (i.e., low median PTS_{non-}
282 $_{upsA} \leq 0.020$) compared to the upsA DBLα types, but they have also been shown to have a more specific 1-
283 to-1 relationship with a single *var* gene than upsA (Tan et al., 2023).
284
285 **Bayesian estimation of MOI_{var} and associated population size**
286 The low to non-existent overlap of repertoires enables an estimation of MOI that relies on the number of
287 (non-upsA) DBLα types sequenced from an individual's isolate (Ruybal-Pesántez et al., 2022; Tiedje et al.,
288 2022). A constant repertoire size or number of DBLα types in a parasite genome can be used to convert
289 the number of types sequenced in an isolate to its estimated MOI (Ruybal-Pesántez et al., 2022; Tiedje et
290 al., 2022). As such, the approach neglects the measurement error in this size introduced by targeted PCR
291 and amplicon sequencing of *var* genes in an infection. Here, we extend the method to a Bayesian
292 formulation and estimate the posterior distribution for each sampled individual for the probability of
293 different MOI values. From individual posterior distributions, we can then obtain the estimated MOI
294 frequency distribution for the population as a whole.
295
296 The two pieces of information required for our approach are the measurement error and the prior
297 distribution of MOI. The measurement error is simply the repertoire size distribution, that is, the
298 distribution of the number of non-upsA DBLα types sequenced given MOI = 1, which is empirically
299 available (Figure S3). We refer to it as P(s | MOI=1) where *s* denotes repertoire size. In the following, s
300 denotes the number of non-upsA DBLα types sequenced when MOI > 1. The prior distribution of MOI
301 refers to the belief we have for what the actual MOI distribution might look like at the population level
302 before empirical evidence is taken into consideration. For example, the prior distribution of MOI is likely
303 to center around a higher value in high-transmission endemic areas than in low-transmission ones.
304
305 We can obtain P(s | MOI = m) from the serial convolution of the repertoire size distribution P(s | MOI =
306 1) and P(s | MOI = m - 1). Starting with the repertoire size distribution given a single infection, we can

307 derive Pr(s | MOI = m) for *m* equal to 2,3,..., up until a maximum value of 20 (empirically determined), as
308 follows:

$$P(s \mid MOI = m) = \sum_{x=L}^{U} P(x \mid MOI = 1) \times P(s - x \mid MOI = m - 1)$$

309 1.1

310 where *L* and *U* are the lower and upper limit for the repertoire size, 10 and 45 respectively from the
311 empirical repertoire size distribution.
312
313 For simplicity, we begin with a uniform prior. We use Bayes' rule to derive a posterior distribution of
314 MOI given a certain number of non-upsA DBLα types sequenced from an individual:

$$P(MOI = j \mid s) = \frac{P(s \mid MOI = j) \times P(MOI = j)}{\sum_{i=1}^{k} P(s \mid MOI = i) \times P(MOI = i)}$$

315 1.2

316 where *k* is the maximum value of MOI, here 20, as empirically determined.
317
318 To obtain the MOI distribution at the population level, we could either simply pool the maximum *a*
319 *posteriori* MOI estimate for each sampled individual, or use a technique called mixture distribution. For
320 the latter, we weighed each posterior MOI distribution for each sampled individual equally and sum over
321 all posterior distributions at the individual level to derive the MOI distribution at the population level:

$$f(MOI = m) = \sum_{i=1}^{n} \frac{1}{n} P(MOI = m \mid s_i)$$

322 1.3

323 where *n* is the number of sampled individuals. These two approaches gave similar results for our empirical
324 survey data as determined by the Kolmogorov-Smirnov Test. The obtained distance statistic is close to 0
325 and the corresponding *p-value* is non-significant across all surveys, indicating that the two estimates were
326 drawn from the same distribution (Table S3). Given this similarity, we present hereafter only the result of
327 pooling the maximum a posteriori MOI estimates. Note that we focused on individuals who had a
328 confirmed microscopic asymptomatic P. *falciparum* infection for our MOI estimation.
329
330 To examine alternative priors, we considered empirical MOI distributions described in the literature
331 incusing the Poisson, hyper-Poisson, and negative binomial distributions (Dietz, 1988; Henry, 2020). The

332 hyper-Poisson and negative binomial distributions can capture the overdispersion seen in the empirical
333 distribution of MOI for certain areas and caused by factors such as heterogeneous biting. We investigated
334 changing the parameters of a negative binomial distribution to generate priors with different means
335 spanning a wide range of MOI values (mean MOI within [~3, ~9]), including those seen in high-transmission
336 endemic areas. We found that the final MOI distribution at the population level is fairly robust against the
337 specification of those parameters. In other words, the estimated MOI distributions at the population level
338 for different parameters do not significantly diverge from each other. The pairwise distance between
339 these distributions measured by KL-divergence centres around 0, and the difference between their mean
340 MOI values across different priors is close to 0 (Figure S3 and S4). Additionally, we performed Pearson
341 correlation tests. We found that the correlation coefficients between the estimated MOI distributions at
342 the population level across different parameters were close to 1 with a highly significant *p-value* (Figure
343 S5 and S6). Thus, we provide in our analyses the estimated population MOI distribution using a uniform
344 prior.
345
346 Statistical analysis
347 We used the R v4.0.2 for all data analyses with the collection of R packages in *tidyverse* being used for
348 data curation along with *base*, *stats, gtsummary,* and *epiR* for the statistical analyses (R Core Team, 2018;
349 Sjoberg, D, Whiting, Curry, Lavery, & Larmarange, 2021; Stevenson, 2020; Wickham et al., 2019).
350 Continuous variables are presented as medians with interquartile ranges (IQRs) and discrete variables are
351 presented using the observed proportions or calculated values with 95% confidence intervals (CIs). Chi-
352 square (X^2) tests were utilized for univariate analyses of discrete variables to compare proportions (for
353 multiple testing *p*-values were adjusted using the Bonferroni correction); non-parametric Mann-Whitney
354 U and Kruskal-Wallis tests were used for comparing distributions between groups of continuous variables.
355 Kaplan-Meier survival curves were generated for the time (i.e., number of surveys) to first event (i.e.,
356 when the DBLα type was no longer observed/detected) comparing the upsA and non-upsA DBLα types; *p*-
357 values were determined using the log-rank test using the R packages *survival* and *survminer* (Kassambara,
358 Kosinski, Biecek, & Scheipl, 2021; Therneau, 2023). The time interval to first event considered for all
359 survival curves was the number of surveys or year (i.e., 2012, 2014, 2015, and 2017) that each DBLα type

360 was observed and only includes those upsA (N = 2,215) and non-upsA (N = 33,151) DBLα types that were
361 seen at baseline in 2012 (i.e., those DBLα types observed prior to the IRS).
362

363 DATA AVAILABILITY

364 The sequences utilized in this study are publicly available in GenBank under BioProject Number: PRJNA
365 396962. All data associated with this study are available in the main text or the supplementary
366 information.

367 # RESULTS

368 Between 2013 and 2015, three-rounds of IRS with non-pyrethroid insecticides were implemented across
369 all of Bongo District. Coincident with the > 90% decrease in transmission following IRS (Tiedje et al., 2022),
370 the prevalence of microscopic *P. falciparum* infections compared to the 2012 baseline survey (pre-IRS)
371 declined by 45.2% and 35.7% following the second (2012 to 2014) and the third (2012 to 2015) round of
372 IRS (Figure 1B, Table S1). These declines in parasite prevalence were observed across all ages, with the
373 greatest impacts being observed on the younger children (i.e., 1-5 years) who were ~3 times less likely to
374 have an infection in 2015 (post-IRS) compared to 2012 (pre-IRS) (Figure 1C, Table S1). These reductions
375 were however short lived and in 2017, 32-months after the discontinuation of IRS, but during SMC, overall
376 *P. falciparum* prevalence rebounded to 41.2%, or near pre-IRS levels (Figure 1B, Table S1). Importantly,
377 this increase in the prevalence of infection in 2017 was only observed among the older age groups (i.e., ≥
378 6 years) (Figure 1C, Table S1). This difference by age group in 2017 can be attributed to SMC, which only
379 targets children between 3-59 months (i.e., < 5 years). A notable increase in parasite prevalence for
380 adolescents (11-20 years) and adults (>20 years) was found in 2017 relative to 2012 (Figure 1C, Table S1).
381

382 Next, we wanted to explore changes in population size measured by MOI_{var}. As this metric is based on
383 non-overlap of *var* repertoire diversity of individual isolates, specifically non-upsA DBLα types, we
384 investigated whether DBLα isolate repertoire similarity (or overlap), as measured by PTS, increased
385 following the sequential interventions (i.e., IRS and SMC). Figure 2 shows that median PTS values for both
386 upsA and non-upsA DBLα types remained low in all surveys although the PTS distributions for both groups
387 changed significantly at each of the study time points relative to the 2012 baseline survey (pre-IRS) (*p-*
388 *values* < 0.001, Kruskal-Wallis test) (Figure 2). However, the change was in the direction of reduced
389 similarity with lower median PTS scores and a larger number of isolates sharing no DBLα types (i.e., PTS =
390 0) in 2014, 2015, and 2017 compared to 2012. Relevant to measurement of MOI_{var}, the median PTS scores
391 for non-upsA DBLα types were lower following the IRS intervention ($PTS_{non\text{-}upsA}$: 2014 = 0.013 and 2015 =
392 0.013 vs. $PTS_{non\text{-}upsA}$: 2012 = 0.020). In 2017, the non-upsA PTS frequency distributions shifted back
393 towards higher median PTS scores ($PTS_{non\text{-}upsA}$ = 0.016) and fewer isolates shared no DBLα types relative
394 to 2014 and 2015 (Figure 2). To verify this pattern was not influenced by multiclonal infections (MOI_{var} >
395 1) we also examined isolates with monoclonal infections (MOI_{var} = 1) and found that this non-overlapping

396 structure, particularly for the non-upsA DBLα types, persisted regardless of infection complexity (Figure
397 S7).
398
399 The raw data of non-upsA DBLα isolate repertoire sizes (Figure S8) were used to estimate MOI_{var} as
400 adjusted using the Bayesian approach (Figure 3). We observed that at baseline in 2012, the majority
401 (89.2%) of the population across all ages carried multiclonal infections (median MOI_{var} = 4 [IQR: 2 – 6])
402 (Figure 3A). Following the IRS intervention, the estimated MOI_{var} distributions became more positively
403 skewed, indicating that a lower proportion of participants harboured multiclonal infections in 2014
404 (64.5%; median MOI_{var} = 2 [IQR: 1 – 3]) and 2015 (71.4%; median MOI_{var} = 2 [IQR: 1 – 3]) compared to
405 2012 (Figure 3A). These reductions in median MOI_{var} and the proportion of multiclonal infections, which
406 were observed across all age groups (Figure 3B), are consistent with the > 90% decrease in transmission
407 intensity following the IRS in turn reducing exposure to new parasite genomes. However, in 2017, both
408 median MOI_{var} (3 [IQR: 2 – 4]) and the proportion of multiclonal infections (78.9%) rebounded in all age
409 groups, even among the younger children (1-5 years) targeted by SMC (Figure 3). While the prevalence
410 of infection in 2017 remained low for the younger children (1-5 years) targeted by SMC, those infected
411 still carried multiclonal infections (84.1%) (Figure 3B). Although the MOI_{var} distributions across all age
412 groups started to rebound in 2017 (i.e., less positively skewed compared to 2014 and 2015) they had not
413 fully recovered to the 2012 baseline patterns (Figure 3). This was most apparent among the younger
414 children (1-5 years), as a larger proportion of isolates in 2017, compared to 2012, had MOI_{var} values equal
415 to one or two, while a smaller proportion had MOI_{var} values ≥ 5 (Figure 3B).
416
417 From isolate MOI_{var} we also estimated population size or the number of *P. falciparum var* repertoires
418 circulating in the population during each survey. In 2014 during IRS, this number decreased by 72.4%
419 relative to the 2012 baseline survey (pre-IRS) (Figure 4A), whereas prevalence decreased by 54.5% (Figure
420 4B). Although the number of *var* repertoires increased slightly in 2015 relative to 2014, there were still
421 64.5% fewer *var* repertoires in the population compared to 2012 (Figure 4A) in comparison to a 42.6%
422 decrease in prevalence (Figure 4B). Importantly this loss of *var* repertoires in 2014 and 2015 following the
423 IRS intervention was seen for all age groups, with the greatest overall reductions (> 85%) being observed
424 for the younger children (1-5 years) (Figure 4CD). However, in 2017, the number of diverse *var* repertoires

425 in the population rebounded, more than doubling between 2015 and 2017 (Figure 4AB). This increase in
426 the number of *var* repertoires was seen for all age groups in 2017, except for the younger children (1-5
427 years) who were directly targeted by SMC (Figure 4CD). In fact, the greatest overall increase was observed
428 for the adolescents (11-20 years), where the number of *var* repertoires in 2017 was ~1.4 times higher
429 compared to 2012 (Figure 4CD). Similar trends in the number *var* repertoires were also observed for the
430 older children (6-10 years) and adults (>20 years) in 2017, although the rebound was not as striking as
431 that detected for the adolescents.
432
433 As parasite population size changed considerably during the sequential interventions, we investigated
434 how the removal or loss of *P. falciparum var* repertoires and subsequent rebound altered *var* richness,
435 measured as the number of unique upsA and non-upsA DBLα types in the parasite population in each
436 survey. Richness at baseline in 2012 (pre-IRS) was high with a large number of unique DBLα types (upsA =
437 2,215; non-upsA = 33,151) (Figure 5, Table S2) and limited overlap of *var* repertoires (i.e., median PTS_{non-}
438 $_{upsA} \leq 0.020$) seen in a relatively small study population of 685 microscopically positive individuals (Figure
439 2). In 2014, as *P. falciparum* prevalence and population size declined (Figure 4) so too did the number of
440 DBLα types, resulting in a 32.1% and 55.3% reduction in richness for the upsA and non-upsA DBLα types,
441 respectively compared to 2012 (Figure 5, Table S2). Again in 2015, as *P. falciparum* prevalence and
442 population size remined low (Figure 4) DBLα type richness was still less than that observed in 2012 (24.6%
443 and 45.9% reduction for upsA and non-upsA DBLα types, respectively) (Figure 5, Table S2). Finally, in 2017,
444 we found that upsA and non-upsA DBLα type richness rebounded relative to 2014 and 2015, coincident
445 with the increase in *P. falciparum* prevalence and parasite population size (Figure 4, Figure 5).
446
447 Given this reduction in *var* richness following the IRS intervention and subsequent rebound in 2017, we
448 next investigated the relative frequency of the DBLα types in each survey (Figure 6, Figure S9). By
449 examining frequency, we discovered that not only were individual upsA and non-upsA DBLα types
450 observed more than once within a survey (i.e., observed in > 1 isolate), but that they were also found
451 longitudinally (i.e., observed in > 1 survey) (Figure 6AB). DBLα types that were frequent (i.e., observed in
452 11-20 or >20 isolates), less frequent (i.e., observed in 2-10 isolates) or only seen once, typically maintained
453 a similar frequency count regardless of ups grouping (i.e., upsA and non-upsA) or survey (i.e., 2012, 2014,

454 2015, and 2017) (Figure 6AB). In 2014 and 2015, following the IRS, there was a significant increase in the
455 proportion of upsA and non-upsA DBLα types in the lower frequency categories (*p-value* < 0.001, Mann-
456 Whitney U test), with all DBLα types becoming rarer in the population (Figure 6CD). This change can be
457 attributed to the removal of *P. falciparum var* repertoires (Figure 4) and loss of upsA and non-upsA DBLα
458 types richness (Figure 5), which disproportionally affected those DBLα types seen once. This shift to all
459 DBLα types becoming rarer following IRS changed in 2017, where the proportion of DBLα types in the
460 more frequent categories (i.e., 2-10, 11-20, or > 20 isolates) significantly increased while the proportion
461 seen once decreased (*p-values* < 0.001, Mann-Whitney U tests) (Figure 6CD).
462
463 Finally, to explore these changes in DBLα type diversity (i.e., richness and frequency), we compared the
464 probability of survival for the DBLα types identified at baseline in 2012 (pre-IRS) and found that the upsA
465 DBLα types persisted significantly longer in the population relative to the non-upsA ($p < 0.001$, log-rank
466 test). Despite this difference, the upsA and non-upsA DBLα types that survived and were observed at
467 multiple study time points (i.e., 2012, 2014, 2015, and 2017) were those that were most frequent in the
468 population at baseline in 2012 (i.e., observed in 11-20 and >20 isolates) (Figure 6EF). As expected, the
469 DBLα types that were only observed once in 2012, were significantly less likely to be seen longitudinally
470 (*p-value* < 0.001, log-rank test) (Figure 6EF).

471 # DISCUSSION

472 *P. falciparum* populations in high-transmission endemic areas in SSA, are characterised by extensive
473 diversity, high rates of recombination, as well as frequent multiclonal infections. Here we defined census
474 population size of *P. falciparum* as a new metric to explore the efficacy of malaria interventions in such
475 areas typified by Bongo, Ghana. We captured age-specific changes in census population size in response
476 sequential malaria interventions. IRS reduced the MOI_{var} parasite population size substantially with the
477 greatest reductions (85%) seen in the younger children (i.e., 1-5 years). More than two years after
478 cessation of IRS, rebound in 2017 was rapid in all age groups with the exception of younger children
479 targeted by SMC. Population sizes in adolescents and adults showed they carried more infections in 2017
480 than at baseline in 2012. This is indicative of a loss of immunity during IRS and warrants further
481 investigation. During and following the IRS and SMC interventions, *var* diversity still remained high and
482 *var* repertoire overlap remained low, reflecting characteristic properties of high transmission and
483 demonstrating the overall resilience of the species to survive significant short-term perturbations.
484
485 What was striking about the Bongo study was the speed with which rebound in MOI_{var} per person and
486 overall population size occurred, once the short-term IRS was discontinued. We looked for a potential
487 explanation in our genetic data. PTS and population frequency data showed that many of the DBLα types
488 occurred in multiple repertoires or genomes. This enabled the survival of these more frequent DBLα types
489 through the interventions, facilitating rebound by maintenance of this diversity. The other notable
490 population genetic result of our study was the failure to increase similarity (or relatedness by state) of *var*
491 repertoires by reducing transmission by > 90% via IRS. From a baseline of very low overlap in repertoires
492 you need outcrossing to create relatedness. However, this was less likely to happen due to reduced
493 transmission as a result of IRS. Rebound, with associated increases in transmission, led to a small increase
494 in *var* repertoire similarity.
495
496 Our molecular approach to measure population size has been to sum MOI_{var} in individual hosts with
497 microscopically detectable infections. Like any diagnostic method there are limits to sensitivity and
498 specificity, which can be more or less tolerated dependent upon the purpose of the study. Here we have
499 looked at relative changes in population size with sequential interventions using an interrupted time series

500 study design and observed changes by measuring MOI_{var}. We have accounted for missing DBLα type data
501 where complete *var* repertoires may not have been sequenced using a Bayesian method based on
502 empirical knowledge of the measurement error. This approach has conceptual relation to the Bayesian
503 approach by Johnson and Larremore (Johnson & Larremore, 2022) to estimate complete repertoire size
504 of, and overlap between, monoclonal infections from incomplete sampling of DBLα types. It would be
505 valuable to consider how to jointly infer the number of DBLα types, isolate overlap, and MOI_{var} from
506 sequencing data of multiclonal infections. Our measurement of population size based on MOI_{var} will be
507 subject to other sampling errors which may in the end be more significant (discussed in detail in Labbé et
508 al. (Labbé et al., 2023)). Low parasitemia typical of asymptomatic infections, small blood volumes, clinical
509 status, and/or within host dynamics, including synchronicity, will all create sampling problems but these
510 are common to all measures of MOI.
511
512 Previously, we have drawn attention to underestimation of the number of DBLα types of related parasites
513 from a cross by *var* genotyping (Labbé et al., 2023). Such related parasites have been observed in
514 experimental crosses (Otto et al., 2018) and must be created frequently in high transmission due to
515 extensive outcrossing (Babiker et al., 1994; Paul et al., 1995). Single clone genomics experiments using
516 biallelic single nucleotide polymorphism (SNPs) from whole genome sequencing data have also detected
517 related parasites using IBD in clinical infections from humans in a high-transmission area of Malawi
518 (Nkhoma et al., 2020). Here we have analysed only low density, chronic asymptomatic infections under
519 strong immune selection in semi-immune hosts whom we have shown select against related parasites (He
520 et al., 2018; Ruybal-Pesántez et al., 2022). When considering the importance of possible exclusion of
521 parasites related by descent, the only sure way to detect such parasites in high transmission is by single
522 cell genomics, a methodology of limited application to malaria surveillance due to practicality and cost of
523 scale up. Again, the error from failure to sample related infections must be weighed up against the issues
524 of under sampling as described above.
525
526 Of interest to those working on malaria surveillance in high-transmission areas is whether measuring MOI,
527 and in turn population size, using *var* repertoires or single copy antigen genes or by neutral variation with
528 biallelic SNPs, is more relevant to the elimination agenda in high burden countries. Single copy antigen

529 genes such as *csp, msp2,* and *ama1* have each been used to measure MOI (e.g., (Falk et al., 2006; Lerch
530 et al., 2017; Nelson et al., 2019)) and have become part of newer genetic panels (e.g., Paragon v1
531 (Tessema et al., 2020) and AMPLseq v1 (LaVerriere et al., 2022)). These genes are under balancing
532 selection with geographically common alleles. Alone, they are much less diverse than the potential
533 combinations of ~45 non-upsA *var* genes so will underestimate population size. *Var* genotyping has also
534 proven more sensitive in high transmission to measure MOI due to the high prevalence of multiclonal
535 infections, which cannot be phased with either biallelic SNP panels (Labbé et al., 2023; Tessema et al.,
536 2020; Watson et al., 2021) or combinations of single copy antigen genes (Sumner et al., 2021).
537
538 A recent paper from Senegal using a 24-SNP barcode observed highly-related parasites circulating in
539 human hosts with symptomatic malaria (Wong et al., 2022). Based on modelling SNP barcodes from
540 polyclonal infections analysed by THE REAL McCOIL to calculate MOI and a new genetic metric R_H to
541 quantify genetic relatedness, they go on to propose these infections were a result of cotransmission (i.e.,
542 single infectious bite) rather than superinfection (i.e., multiple infectious bites). They claim from R_H
543 relatedness estimates that cotransmission was "ubiquitous" in symptomatic infections in three sites,
544 which would be designated by the WHO classification as low- to very low-transmission areas according to
545 *P. falciparum* prevalence (WHO/GMP, 2017). Instead, we are working in a high-transmission area (see
546 Methods) with chronic asymptomatic infections under strong immune selection as discussed above. We
547 found little empirical evidence of related parasites within hosts using polymorphic microsatellites markers
548 and *var*coding when the same 24-SNP barcode made monoclonal infections look highly related or similar
549 (Ghansah et al., 2023). We agree that the issue of cotransmission vs. superinfection is important and must
550 be considered when thinking about MOI in relation to transmission intensity.
551
552 In summary, our findings provide parasite population insights into why rebound is the inevitable
553 consequence of such short-term IRS interventions unless you simultaneously target the highly diverse,
554 long-lived parasite population in humans, not just children < 5 years by SMC. Of potential translational
555 significance for malaria molecular surveillance, we identify new metrics, especially MOI_{var} and census
556 population size as well as *var* frequency category, informative to monitor and evaluate interventions in
557 high-transmission areas with multiclonal infections and high rates of outcrossing. Such metrics could be

558 used longitudinally to detect incremental gains of transmission-reducing interventions, including IRS,
559 LLINs, and vaccines to perturb the high-transmission characteristics of the parasite population in humans
560 in high-burden countries in SSA.

REFERENCES

Anderson, T. J. C., Haubold, B., Williams, J. T., Estrada-Franco, J. G., Richardson, L., Mollinedo, R., … Day, K. P. (2000). Microsatellite markers reveal a spectrum of population structures in the malaria parasite Plasmodium falciparum. *Molecular Biology and Evolution*, *17*(10), 1467–1482. doi: 10.1093/oxfordjournals.molbev.a026247

Babiker, H. A., Ranford-Cartwright, L. C., Currie, D., Charlwood, J. D., Billingsley, P., Teuscher, T., & Walliker, D. (1994). Random mating in a natural population of the malaria parasite Plasmodium falciparum. *Parasitology*, *109*(4), 413–421. doi: 10.1017/S0031182000080665

Barry, A. E., Leliwa-Sytek, A., Tavul, L., Imrie, H., Migot-Nabias, F., Brown, S. M., … Day, K. P. (2007). Population genomics of the immune evasion (var) genes of Plasmodium falciparum. *PLoS Pathogens*, *3*(3), e34. doi: 10.1371/journal.ppat.0030034

Barton, N. (2010). Understanding adaptation in large populations. *PLoS Genetics*, *6*(6), 1–3. doi: 10.1371/journal.pgen.1000987

Bopp, S. E. R., Manary, M. J., Bright, a. T., Johnston, G. L., Dharia, N. V., Luna, F. L., … Winzeler, E. a. (2013). Mitotic Evolution of Plasmodium falciparum Shows a Stable Core Genome but Recombination in Antigen Families. *PLoS Genet.*, *9*(2), 1–15. doi: 10.1371/journal.pgen.1003293

Bruce, M. C., Galinski, M. R., Barnwell, J. W., Donnelly, C. A., Walmsley, M., Alpers, M. P., … Day, K. P. (2000). Genetic diversity and dynamics of plasmodium falciparum and P. vivax populations in multiply infected children with asymptomatic malaria infections in Papua New Guinea. *Parasitology*, *121 (Pt 3*, 257–272. doi: 10.1017/S0031182099006356

Chan, J.-A., Fowkes, F. J. I., & Beeson, J. G. (2014). Surface antigens of Plasmodium falciparum-infected erythrocytes as immune targets and malaria vaccine candidates. *Cellular and Molecular Life Sciences : CMLS*, 3633–3657. doi: 10.1007/s00018-014-1614-3

Chen, D. S., Barry, A. E., Leliwa-Sytek, A., Smith, T.-A. A., Peterson, I., Brown, S. M., … Day, K. P. (2011). A molecular epidemiological study of var gene diversity to characterize the reservoir of Plasmodium falciparum in humans in Africa. *PLoS ONE*, *6*(2), e16629. doi: 10.1371/journal.pone.0016629

Claessens, A., Hamilton, W. W. L., Kekre, M., Otto, T. D., Faizullabhoy, A., Rayner, J. C., & Kwiatkowski, D. (2014). Generation of antigenic diversity in Plasmodium falciparum by structured rearrangement of var genes during mitosis. *PLoS Genetics*, *10*(12), e1004812. doi: 10.1371/journal.pgen.1004812

Day, K. P., Artzy-Randrup, Y., Tiedje, K. E., Rougeron, V., Chen, D. S., Rask, T. S., … Pascual, M. (2017). Evidence of Strain Structure in Plasmodium falciparum Var Gene Repertoires in Children from Gabon, West Africa. *PNAS*, *114*(20), E4103–E4111. doi: 10.1073/pnas.1613018114

Dietz, K. (1988). Mathematical models for transmission and control of malaria. In *Malaria: principles and practice of malariology*.

Falk, N., Kaestli, M., Qi, W., Ott, M., Baea, K., Cortés, A., & Beck, H.-P. (2009). Analysis of Plasmodium falciparum var genes expressed in children from Papua New Guinea. *Journal of Infectious Diseases*, *200*, 347–356. doi: 10.1086/600071

Falk, N., Maire, N., Sama, W., Owusu-Agyei, S., Smith, T., Beck, H. P., & Felger, I. (2006). Comparison of PCR-RFLP and genescan-based genotyping for analyzing infection dynamics of Plasmodium falciparum. *American Journal of Tropical Medicine and Hygiene*, *74*, 944–950.

Farnert, A., Snounou, G., Rooth, I., & Bjorkman, A. (1997). Daily dynamics of Plasmodium falciparum subpopulations in asymptomatic children in a holoendemic area. *American Journal of Tropical Medicine and Hygiene*, *56*(5), 538–547. doi: 10.4269/ajtmh.1997.56.538

Feng, Q., Tiedje, K. E., Ruybal-Pesántez, S., Tonkin-Hill, G., Duffy, M. F., Day, K. P., … Chan, Y.-B. (2022). An accurate method for identifying recent recombinants from unaligned sequences. *Bioinformatics*, (January), 1–7. doi: 10.1093/bioinformatics/btac012

Futse, J. E., Brayton, K. A., Dark, M. J., Knowles, D. P., & Palmer, G. H. (2008). Superinfection as a driver of genomic

608 diversification in antigenically variant pathogens. *Proceedings of the National Academy of Sciences of the*
609 *United States of America*, *105*(6), 2123–2127. doi: 10.1073/pnas.0710333105
610 Gardner, M. J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R. W., … Barrell, B. (2002). Genome sequence of
611 the human malaria parasite Plasmodium falciparum. *Nature*, *419*(6906), 498–511. doi:
612 10.1038/nature01097
613 Ghansah, A., Tiedje, K. E., Argyropoulos, D. C., Onwona, C. O., Deed, S. L., Labbé, F., … Day, K. P. (2023).
614 Comparison of molecular surveillance methods to assess changes in the population genetics of Plasmodium
615 falciparum in high transmission. *Frontiers in Parasitology*, *2*, 1067966.
616 Gogue, C., Wagman, J., Tynuv, K., Saibu, A., Yihdego, Y., Malm, K., … Robertson, M. (2020). An observational
617 analysis of the impact of indoor residual spraying in Northern, Upper East, and Upper West Regions of
618 Ghana: 2014 through 2017. *Malaria Journal*, *19*(1), 1–13. doi: 10.1186/s12936-020-03318-1
619 Grenfell, B. T., Pybus, O. G., Gog, J. R., Wood, J. L. N., Daly, J. M., Mumford, J. A., & Holmes, E. C. (2004). Unifying
620 the Epidemiological and Evolutionary Dynamics of Pathogens. *Science*, *303*(5656), 327–332. doi:
621 10.1126/science.1090727
622 He, Q., Pilosof, S., Tiedje, K. E., Ruybal-Pesántez, S., Artzy-Randrup, Y., Baskerville, E. B., … Pascual, M. (2018).
623 Networks of genetic similarity reveal non-neutral processes shape strain structure in Plasmodium
624 falciparum. *Nature Communications*, *9*(1), 1817. doi: 10.1038/s41467-018-04219-3
625 Henry, J. M. (2020). A hybrid model for the effects of treatment and demography on malaria superinfection. *J.*
626 *Theor. Biol.*, *491*, 110194.
627 Jensen, A. T. R., Magistrado, P., Sharp, S., Joergensen, L., Lavstsen, T., Chiucchiuini, A., … Theander, T. G. (2004).
628 Plasmodium falciparum associated with severe childhood malaria preferentially expresses PfEMP1 encoded
629 by group A var genes. *The Journal of Experimental Medicine*, *199*, 1179–1190. doi: 10.1084/jem.20040274
630 Johnson, E. K., & Larremore, D. B. (2022). Bayesian estimation of community size and overlap from random
631 subsamples. *PLoS Computational Biology*, *18*(9), 1–16. doi: 10.1371/journal.pcbi.1010451
632 Kaestli, M., Cockburn, I. a, Cortés, A., Baea, K., Rowe, J. A., & Beck, H.-P. (2006). Virulence of malaria is associated
633 with differential expression of Plasmodium falciparum var gene subgroups in a case-control study. *The*
634 *Journal of Infectious Diseases*, *193*(11), 1567–1574. doi: JID35650 [pii]\n10.1086/503776
635 Kalmbach, Y., Rottmann, M., Kombila, M., Kremsner, P. G., Beck, H.-P., & Kun, J. F. J. (2010). Differential var gene
636 expression in children with malaria and antidromic effects on host gene expression. *Journal of Infectious*
637 *Diseases*, *202*(2), 313–317. doi: 10.1086/653586
638 Kassambara, A., Kosinski, M., Biecek, P., & Scheipl, F. (2021). *survminer: Drawing Survival Curves using ggplot2* (p.
639 R package version 0.4.9, https://cran.r-project.or). p. R package version 0.4.9, https://cran.r-project.or.
640 Kraemer, S. M., Kyes, S. A., Aggarwal, G., Springer, A. L., Nelson, S. O., Christodoulou, Z., … Smith, J. D. (2007).
641 Patterns of gene recombination shape var gene repertoires in Plasmodium falciparum: comparisons of
642 geographically diverse isolates. *BMC Genomics*, *8*, 45. doi: 10.1186/1471-2164-8-45
643 Kraemer, S. M., & Smith, J. D. (2006). A family affair: var genes, PfEMP1 binding, and malaria disease. *Current*
644 *Opinion in Microbiology*, Vol. 9, pp. 374–380. doi: 10.1016/j.mib.2006.06.006
645 Kyriacou, H. M., Stone, G. N., Challis, R. J., Raza, A., Lyke, K. E., Thera, M. A., … Rowe, J. A. (2006). Differential var
646 gene transcription in Plasmodium falciparum isolates from patients with cerebral malaria compared to
647 hyperparasitaemia. *Molecular and Biochemical Parasitology*, *150*, 211–218. doi:
648 10.1016/j.molbiopara.2006.08.005
649 Labbé, F., He, Q., Zhan, Q., Tiedje, K. E., Argyropoulos, D. C., Tan, M. H., … Pascual, M. (2023). Neutral vs . non-
650 neutral genetic footprints of Plasmodium falciparum multiclonal infections. *PLoS Computational Biology*,
651 *19*(1), e1010816. doi: doi.org/10.1101/2022.06.27.497801
652 LaVerriere, E., Schwabl, P., Carrasquilla, M., Taylor, A. R., Johnson, Z. M., Shieh, M., … Neafsey, D. E. (2022). Design
653 and implementation of multiplexed amplicon sequencing panels to serve genomic epidemiology of
654 infectious disease: A malaria case study. *Molecular Ecology Resources*, (March), 2285–2303. doi:
655 10.1111/1755-0998.13622

656 Lavstsen, T., Salanti, A., Jensen, A. T. R., Arnot, D. E., & Theander, T. G. (2003). Sub-grouping of Plasmodium
657 falciparum 3D7 var genes based on sequence analysis of coding and non-coding regions. *Malaria Journal*, *2*,
658 27. doi: 10.1186/1475-2875-2-27
659 Leech, J. H., Barnwell, J. W., Miller, L. H., & Howard, R. J. (1984). Indentification of a strain-specific malarial antigen
660 exposed on the surface of Plasmodium falciparum-infected erythrocytes. *Journal of Experimental Medicine*,
661 *159*(June), 1567–1575.
662 Lerch, A., Koepfli, C., Hofmann, N. E., Messerli, C., Wilcox, S., Kattenberg, J. H., … Felger, I. (2017). Development of
663 amplicon deep sequencing markers and data analysis pipeline for genotyping multi-clonal malaria infections.
664 *BMC Genomics*, *18*(1), 1–13. doi: 10.1186/s12864-017-4260-y
665 Macdonald, G. (1957). *The epidemiology and control of malaria.* London: Oxford University Press. Retrieved from
666 https://www.cabdirect.org/cabdirect/abstract/19581000237#
667 Naung, M. T., Martin, E., Munro, J., Mehra, S., Guy, A. J., Laman, M., … Barry, A. E. (2022). Global diversity and
668 balancing selection of 23 leading Plasmodium falciparum candidate vaccine antigens. *PLoS Computational*
669 *Biology*, *18*(2), 1–30. doi: 10.1371/journal.pcbi.1009801
670 Nelson, C. S., Sumner, K. M., Freedman, E., Saelens, J. W., Obala, A. A., Mangeni, J. N., … O'Meara, W. P. (2019).
671 High-resolution micro-epidemiology of parasite spatial and temporal dynamics in a high malaria transmission
672 setting in Kenya. *Nature Communications*, *10*(1). doi: 10.1038/s41467-019-13578-4
673 Nkhoma, S. C., Trevino, S. G., Gorena, K. M., Nair, S., Khoswe, S., Jett, C., … Cheeseman, I. H. (2020). Co-
674 transmission of Related Malaria Parasite Lineages Shapes Within-Host Parasite Diversity. *Cell Host and*
675 *Microbe*, *27*(1), 93-103.e4. doi: 10.1016/j.chom.2019.12.001
676 Normark, J., Nilsson, D., Ribacke, U., Winter, G., Moll, K., Wheelock, C. E., … Wahlgren, M. (2007). PfEMP1-
677 DBL1alpha amino acid motifs in severe disease states of Plasmodium falciparum malaria. *Proceedings of the*
678 *National Academy of Sciences of the United States of America*, *104*(40), 15835–15840. doi:
679 10.1073/pnas.0610485104
680 Otto, T. D., Assefa, S. A., Böhme, U., Sanders, M. J., Kwiatkowski, D. P., Berriman, M., & Newbold, C. (2019).
681 Evolutionary analysis of the most polymorphic gene family in falciparum malaria. *Wellcome Open Research*,
682 *4*, 1–29. doi: 10.12688/wellcomeopenres.15590.1
683 Otto, T. D., Böhme, U., Sanders, M., Reid, A., Bruske, E. I., Duffy, C. W., … Berriman, M. (2018). Long read
684 assemblies of geographically dispersed Plasmodium falciparum isolates reveal highly structured
685 subtelomeres [version 1; referees: 3 approved]. *Wellcome Open Research*, *3*(May), 1–24. doi:
686 10.12688/wellcomeopenres.14571.1
687 Paul, R. E., Packer, M. J., Walmsley, M., Lagog, M., Ranford-Cartwright, L. C., Paru, R., & Day, K. P. (1995). Mating
688 patterns in malaria parasite populations of Papua New Guinea. *Science*, *269*(September), 1709–1711.
689 Pilosof, S., He, Q., Tiedje, K. E., Ruybal-Pesántez, S., Day, K. P., & Pascual, M. (2019). Competition for hosts
690 modulates vast antigenic diversity to generate persistent strain structure in Plasmodium falciparum. *PLOS*
691 *Biology*, *17*(6), e3000336. doi: 10.1371/journal.pbio.3000336
692 R Core Team. (2018). R: A Language and Environment for Statistical Computing. *R Foundation for Statistical*
693 *Computing.* Vienna, Austria: R Foundation for Statistical Computing.
694 Rask, T. S., Hansen, D., Theander, T. G., Pedersen, A. G., Lavstsen, T., Gorm Pedersen, A., … Lavstsen, T. (2010).
695 Plasmodium falciparum erythrocyte membrane protein 1 diversity in seven genomes - divide and conquer.
696 *PLoS Comput. Biol.*, *6*(9), e1000933. doi: 10.1371/journal.pcbi.1000933
697 Rorick, M. M., Artzy-Randrup, Y., Ruybal-Pesántez, S., Tiedje, K. E., Rask, T. S., Oduro, A., … Pascual, M. (2018).
698 Signatures of competition and strain structure within the major blood-stage antigen of P. falciparum in a
699 local community in Ghana. *Ecology and Evolution*, *8*(7), 3574–3588.
700 Rottmann, M., Lavstsen, T., Mugasa, J. P., Kaestli, M., Jensen, A. T. R., Müller, D., … Beck, H.-P. P. (2006).
701 Differential expression of var gene groups is associated with morbidity caused by Plasmodium falciparum
702 infection in Tanzanian children. *Infect. Immun.*, *74*(7), 3904–3911. doi: 10.1128/IAI.02073-05
703 Ruybal-Pesántez, S., Tiedje, K. E., Pilosof, S., Tonkin-Hill, G., He, Q., Rask, T. S., … Day, K. P. (2022). Age-specific

patterns of DBLa var diversity can explain why residents of high malaria transmission areas remain susceptible to Plasmodium falciparum blood stage infection throughout life. *International Journal for Parasitology*, *20*, 721–731.

Ruybal-Pesántez, S., Tiedje, K. E., Tonkin-Hill, G., Rask, T. S., Kamya, M. R., Greenhouse, B., … Day, K. P. (2017). Population genomics of virulence genes of Plasmodium falciparum in clinical isolates from Uganda. *Scientific Reports*, *7*(1), 11810. doi: 10.1038/s41598-017-11814-9

Sjoberg, D, Whiting, K., Curry, M., Lavery, J., & Larmarange, J. (2021). Reproducible Summary Tables with the gtsummary Package. *The R Journal.*, *13*, 570–580.

Smith, T., Felger, I., Fraser-Hurt, N., & Beck, H.-P. P. (1999). Effect of insecticide-treated falciparum infections bed nets on the dynamics of multiple Plasmodium falciparum infections. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, *93 (Supp)*, S1/53-S1/57. doi: 10.1016/S0035-9203(99)90328-0

Speed, D., & Balding, D. J. (2015). Relatedness in the post-genomic era: Is it still useful? *Nature Reviews Genetics*, *16*(1), 33–44. doi: 10.1038/nrg3821

Stevenson, M. (2020). *epiR: Tools for the analysis of epidemiological data*. Retrieved from https://cran.r-project.org/package=epiR

Sumner, K. M., Freedman, E., Abel, L., Obala, A., Pence, B. W., Wesolowski, A., … Taylor, S. M. (2021). Genotyping cognate Plasmodium falciparum in humans and mosquitoes to estimate onward transmission of asymptomatic infections. *Nature Communications*, *12*(1), 1–12. doi: 10.1038/s41467-021-21269-2

Tan, M. H., Shim, H., Chan, Y., & Day, K. P. (2023). Unravelling chaos for malaria surveillance : Relationship between DBLα types and var genes in Plasmodium falciparum. *Frontiers in Parasitology*, *1*(1006341). doi: https://doi.org/10.3389/fpara.2022.1006341

Tessema, S. K., Hathaway, N. J., Teyssier, N. B., Murphy, M., Chen, A., Aydemir, O., … Greenhouse, B. (2020). Sensitive, highly multiplexed sequencing of microhaplotypes from the Plasmodium falciparum heterozygome. *Journal of Infectious Diseases*, *225*(7), 1227–1237.

Therneau, T. (2023). *A Package for Survival Analysis in R* (p. R package version 3.5-5, https://CRAN.R-project.or). p. R package version 3.5-5, https://CRAN.R-project.or.

Tiedje, K. E., Oduro, A. R., Agongo, G., Anyorigiya, T., Azongo, D., Awine, T., … Day, K. P. (2017). Seasonal Variation in the Epidemiology of Asymptomatic Plasmodium falciparum Infections Across Two Catchment Areas in Bongo District, Ghana. *Am. J. Trop. Med. Hyg.*, *97*(1), 199–212. doi: 10.4269/ajtmh.16-0959

Tiedje, K. E., Oduro, A. R., Bangre, O., Amenga-Etego, L., Dadzie, S. K., Appawu, M. A., … Day, K. P. (2022). Indoor residual spraying with a non-pyrethroid insecticide reduces the reservoir of Plasmodium falciparum in a high-transmission area in northern Ghana. *PLOS Global Public Health*, *2*(25), e0000285. doi: 10.1371/journal.pgph.0000285

US Agency for International Development (USAID) Global Health Supply Chain Program. (2020). *Technical Breif: Data visability makes all the difference in Ghana's 2018 LLIN mass distribution campaign.*

US President's Malaria Initiative Africa IRS (AIRS) Project. (2016). *Entomological monitoring of the PMI AIRS program in northern Ghana: 2016 annual report.* Bethesda, Marland, USA.

Volz, E. M., Koelle, K., & Bedford, T. (2013). Viral Phylodynamics. *PLoS Computational Biology*, *9*(3). doi: 10.1371/journal.pcbi.1002947

Wagman, J., Gogue, C., Tynuv, K., Mihigo, J., Bankineza, E., Bah, M., … Robertson, M. (2018). An observational analysis of the impact of indoor residual spraying with non-pyrethroid insecticides on the incidence of malaria in Ségou Region, Mali: 2012–2015. *Malaria Journal*, *17*(1), 19. doi: 10.1186/s12936-017-2168-2

Warimwe, G. M., Fegan, G., Musyoki, J. N., Newton, C. R. J. C., Opiyo, M., Githinji, G., … Bull, P. C. (2012). Prognostic indicators of life-threatening malaria are associated with distinct parasite variant antigen profiles. *Science Translational Medicine*, *4*(129), 129ra45. doi: 10.1126/scitranslmed.3003247

Warimwe, G. M., Keane, T. M., Fegan, G., Musyoki, J. N., Newton, C. R. J. C., Pain, A., … Bull, P. C. (2009). Plasmodium falciparum var gene expression is modified by host immunity. *Proc. Natl. Acad. Sci. U. S. A.*, *106*(51), 21801–21806.

752 Watson, O. J., Okell, L. C., Hellewell, J., Slater, H. C., Unwin, H. J. T., Omedo, I., … Verity, R. (2021). Evaluating the
753 Performance of Malaria Genetics for Inferring Changes in Transmission Intensity Using Transmission
754 Modeling. *Molecular Biology and Evolution*, *38*(1), 274–289. doi: 10.1093/molbev/msaa225
755 WHO/GMP. (2017). A Framework for Malaria Elimination. In *Geneva World Health Organization*. Retrieved from
756 http://apps.who.int/iris/bitstream/handle/10665/254761/9789241511988-eng.pdf?sequence=1
757 WHO. (2012). *WHO Policy Recommendation: Seasonal Malaria Chemoprevention (SMC) for Plasmodium*
758 *falciparum malaria control in highly seasonal transmission areas of the Sahel sub-region in Africa*.
759 Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., … Yutani, H. (2019). Welcome to the
760 Tidyverse. *Journal of Open Source Software*, *4*(43), 1686. doi: 10.21105/joss.01686
761 Wong, W., Volkman, S., Daniels, R., Schaffner, S., Sy, M., Ndiaye, Y. D., … Hartl, D. L. (2022). R H: a genetic metric
762 for measuring intrahost Plasmodium falciparum relatedness and distinguishing cotransmission from
763 superinfection. *PNAS Nexus*, *1*(4), 1–11. doi: 10.1093/pnasnexus/pgac187
764 World Health Organization. (2022). World Malaria Report 2022. In *World Health Organization*. doi: ISBN 978 92 4
765 1564403
766 Zhang, X., & Deitsch, K. W. (2022). The mystery of persistent, asymptomatic Plasmodium falciparum infections.
767 *Current Opinion in Microbiology*, *70*, 102231. doi: 10.1016/j.mib.2022.102231
768

769 **ACKNOWLEDGMENTS**. This research was supported by Fogarty International Center at the National
770 Institutes of Health through the joint NIH-NSF-NIFA Ecology and Evolution of Infectious Disease award
771 R01-TW009670 to KAK, MP, and KPD; and the National Institute of Allergy and Infectious Diseases,
772 National Institutes of Health through the joint NIH-NSF-NIFA Ecology and Evolution of Infectious Disease
773 award R01-AI149779 to ARO, KAK, MP, and KPD. We wish to thank the participants, communities, and the
774 Ghana Health Service in Bongo District, Ghana for their willingness to participate in this study. We would
775 like to thank the field teams in Bongo for their technical assistance in the field, as well as the laboratory
776 personnel at the Navrongo Health Research Centre for their expertise and for undertaking the sample
777 collections and parasitological assessments. We are also grateful to three anonymous referees for their
778 insightful comments.
779

780 **DATA AVAILABILITY STATEMENT**. The sequences utilized in this study are publicly available in GenBank
781 under BioProject Number: PRJNA 396962. All data associated with this study are available in the main text
782 or in the Supporting Information. All custom code is available in an open source repository: (1)
783 DBLαCleaner pipeline is available at, (2) clusterDBLalpha pipeline is available at
784 https://github.com/Unimelb-Day-Lab/clusterDBLalpha, and the (3) classifyDBLalpha pipeline is available
785 at https://github.com/Unimelb-Day-Lab/classifyDBLalpha. A dataset to demo this custom code is available
786 at https://github.com/UniMelb-Day-Lab/tutorialDBLalpha**.** For additional information on the use of the
787 Bayesian approach to estimate MOI_{var} please see https://github.com/qzhan321/ME-Bayesian-MOI-
788 estimation.
789

790 **BENEFIT-SHARING STATEMENT**. A research collaboration was developed with scientists from Ghana based
791 at the Navrongo Health Research Centre and the Noguchi Memorial Institute for Medical Research. All
792 collaborators are included as co-authors and the relevant results from the research has been shared with
793 key stakeholders and the local community (i.e., Paramount Chief of Bongo, divisional Chiefs, Queen
794 Mothers, and community members) as well as the Bongo District Health Directorate and the Regional
795 Health Directorate (Upper East Region). Before this research was undertaken, informed consent was
796 sought and obtained from the key stakeholders and the local community in Bongo District. In addition,
797 members of the local community were trained as field workers and were directly involved in liaising with

the local community and in the collection of the study data. The contribution of these individuals to this research is described in the Acknowledgements. This research addresses a priority concern regarding malaria control and the impact of interventions. These concerns are relevant for both the local community as well as for the National Malaria Elimination Programme in Ghana.

AUTHOR CONTRIBUTIONS. K.P.D and M.P conceptualized the study of population size using MOI_{var}. K.A.K and K.P.D conceived and designed the field study with input from A.R.O and K.E.T who together with O.B coordinated the field studies. K.E.T, S.R-P, and S.L.D processed the samples and performed the genotyping experiments. G.T-H developed and validated the bioinformatic pipelines. K.E.T, S.R-P, and D.C.A processed, cleaned, and curated the datasets for analysis. Q.Z developed and implemented the Bayesian estimations. K.E.T, Q.Z, M.P, and K.P.D analysed the data with contributions from S.R-P, Q.H, and M-H.T. K.E.T and Q.Z visualized the data. K.P.D, M.P, and K.E.T wrote the original draft of the manuscript. Q.Z., S.R-P, G.T-H, Q.H, M-H.T, D.C.A, S.L.D, O. B, A.R.O, and K.AK reviewed and edited manuscript. M.P and K.P.D supervised the research. A.R.O, K.A.K, M.P, and K.P.D acquired the funding.

COMPETING INTERESTS. The authors declare no competing interests.

FIGURES

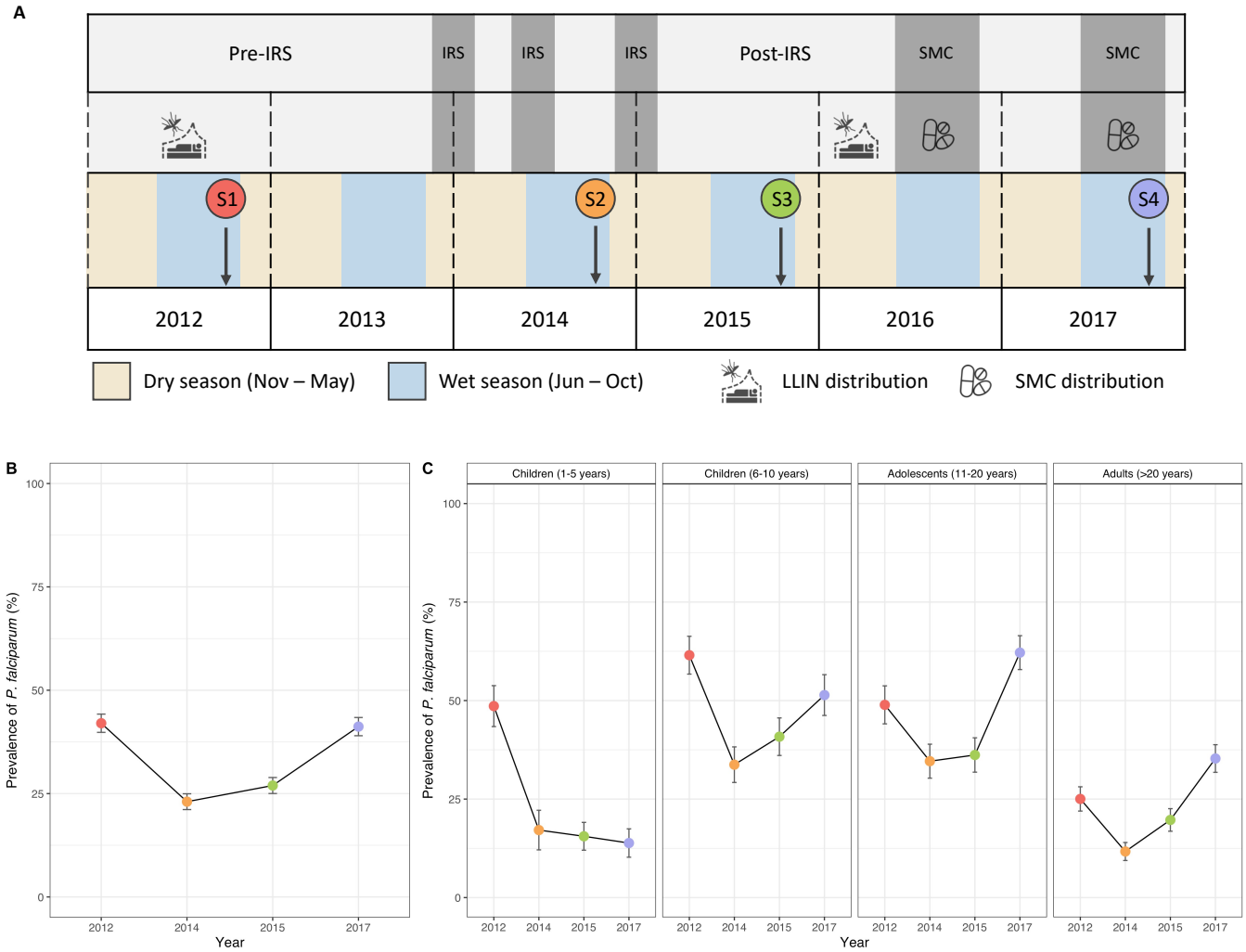


Figure 1. Study design and changes in the prevalence of microscopic *P. falciparum* infection following the IRS and SMC interventions in Bongo, Ghana. (**A**) Four age-stratified cross-sectional surveys of ~2,000 participants per survey were conducted in Bongo, Ghana at the end of the wet seasons in October 2012 (Survey 1, baseline pre-IRS, red), October 2014 (Survey 2, during IRS, orange), October 2015 (Survey 3, post-IRS, green), and October 2017 (Survey 4, SMC, purple) (Table S1) (see Methods). The three rounds of IRS (grey areas) were implemented between 2013 and 2015 (Tiedje et al., 2022). SMC was distributed to all children < 5 years of age during the wet seasons in 2016 (two rounds between August – September 2016) and 2017 (four rounds between September – December 2017) (Gogue et al., 2020). Both IRS and SMC were implemented against a background of widespread LLIN usage which were distributed across Bongo District by the NMEP/GHS between 2010 – 2012 and again in 2016 (Gogue et al., 2020; Tiedje et al., 2022; US Agency for International Development (USAID) Global Health Supply Chain Program, 2020). Prevalence of microscopic *P. falciparum* infections in the (**B**) study population and (**C**) for all age groups (years) in each survey. Error bars represent the upper and lower limits of the 95% confidence interval.

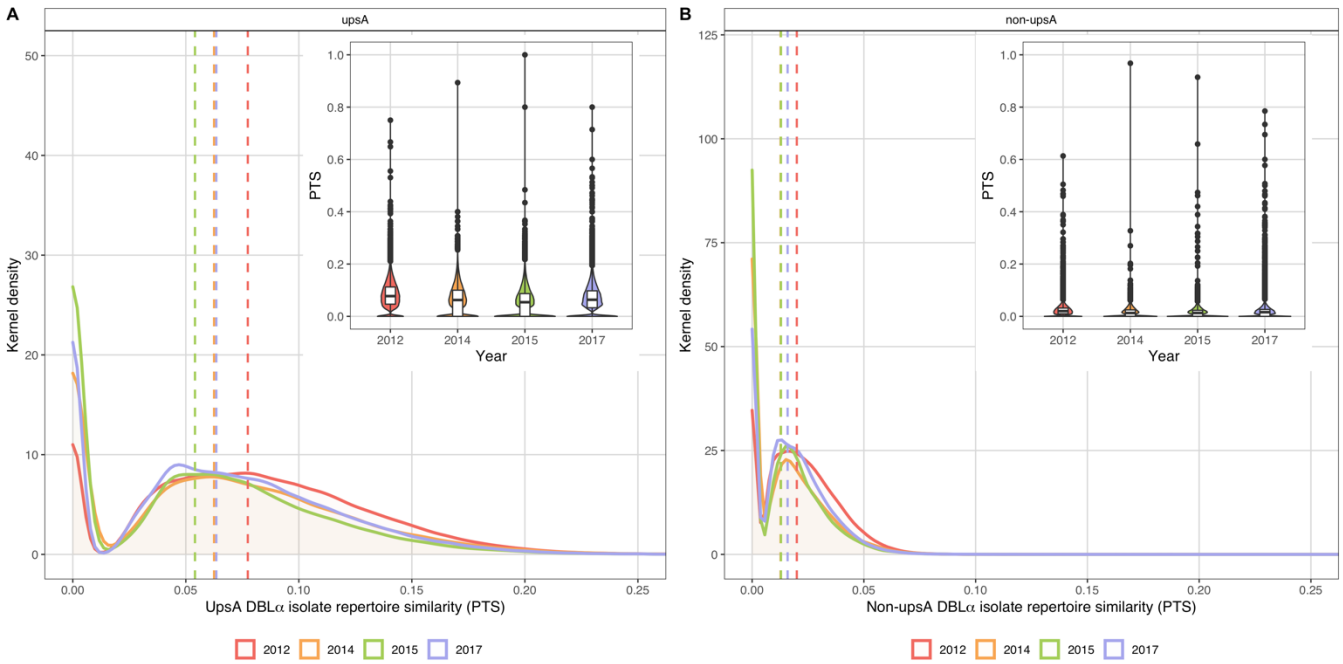


Figure 2. Sharing of upsA and non-upsA DBLα types among the DBLα isolate repertoires in 2012 (pre-IRS, red), 2014 (during IRS, orange), 2015 (post-IRS, green), and 2017 (SMC, purple). The overlapping density and violin plots (upper right-hand corners) show the distribution of PTS scores (i.e., DBLα isolate repertoire similarity) between the (**A**) upsA and (**B**) non-upsA DBLα isolate repertoires with DBLα sequencing data (Table S2) in each survey. The PTS scales in the density plots have been zoomed-in to provide better visualization of the upsA and non-upsA DBLα types PTS distributions. The colored dashed lines in the density plots indicate the median PTS scores in each survey for the upsA (2012 (red) = 0.078, 2014 (orange) = 0.063, 2015 (green) = 0.054, and 2017 (purple) = 0.064) and non-upsA (2012 (red) = 0.020, 2014 (orange) = 0.013, 2015 (green) = 0.013, and 2017 (purple) = 0.016) DBLα types. The non-upsA median PTS values in 2014 (orange) and 2015 (green) were both 0.013, and overlap in the figure. *Note:* In the PTS violin plots the central box plots indicate the medians (centre line), interquartile ranges (IQR, upper and lower quartiles), whiskers (1.5x IQR), and outliers (points).

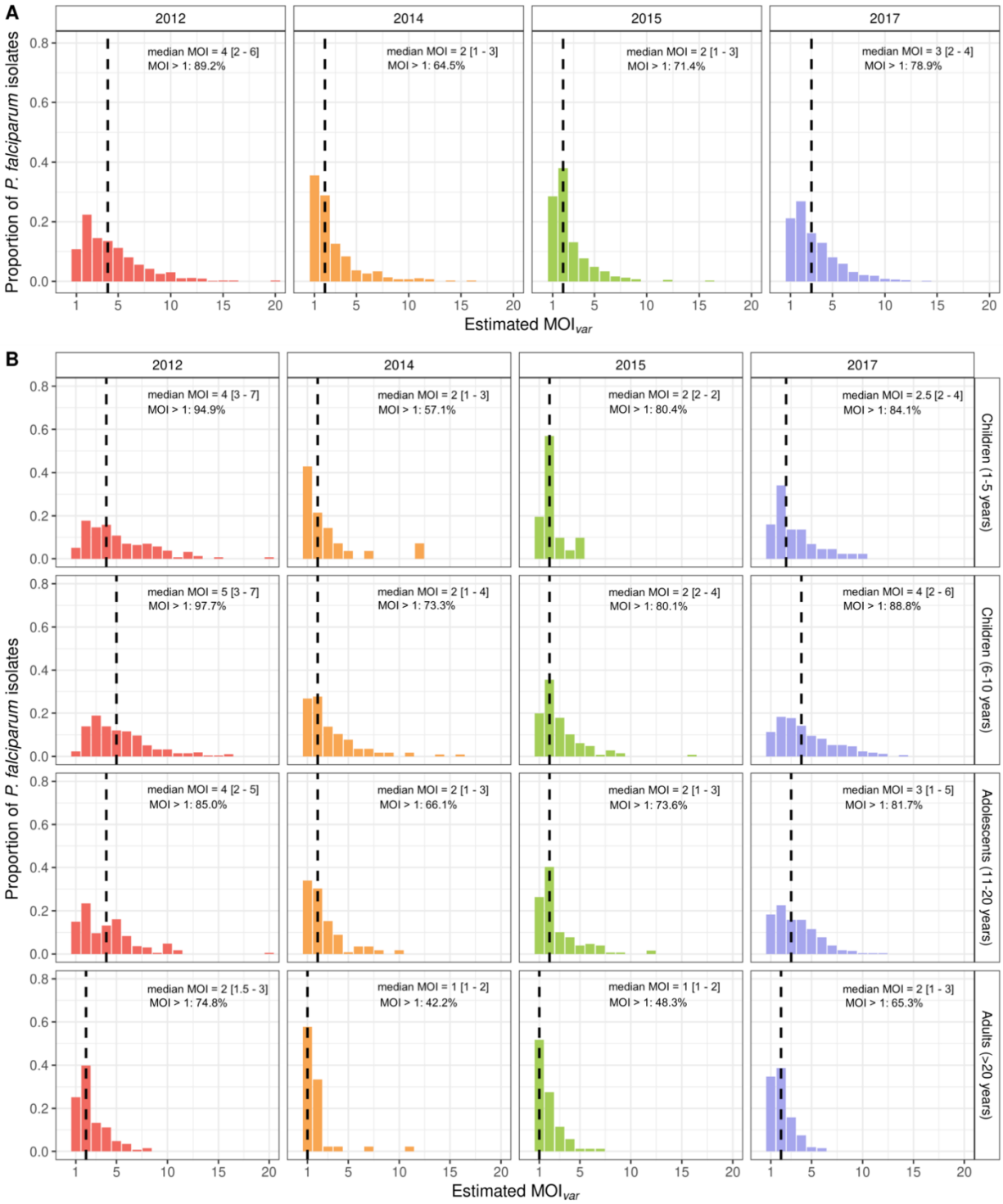
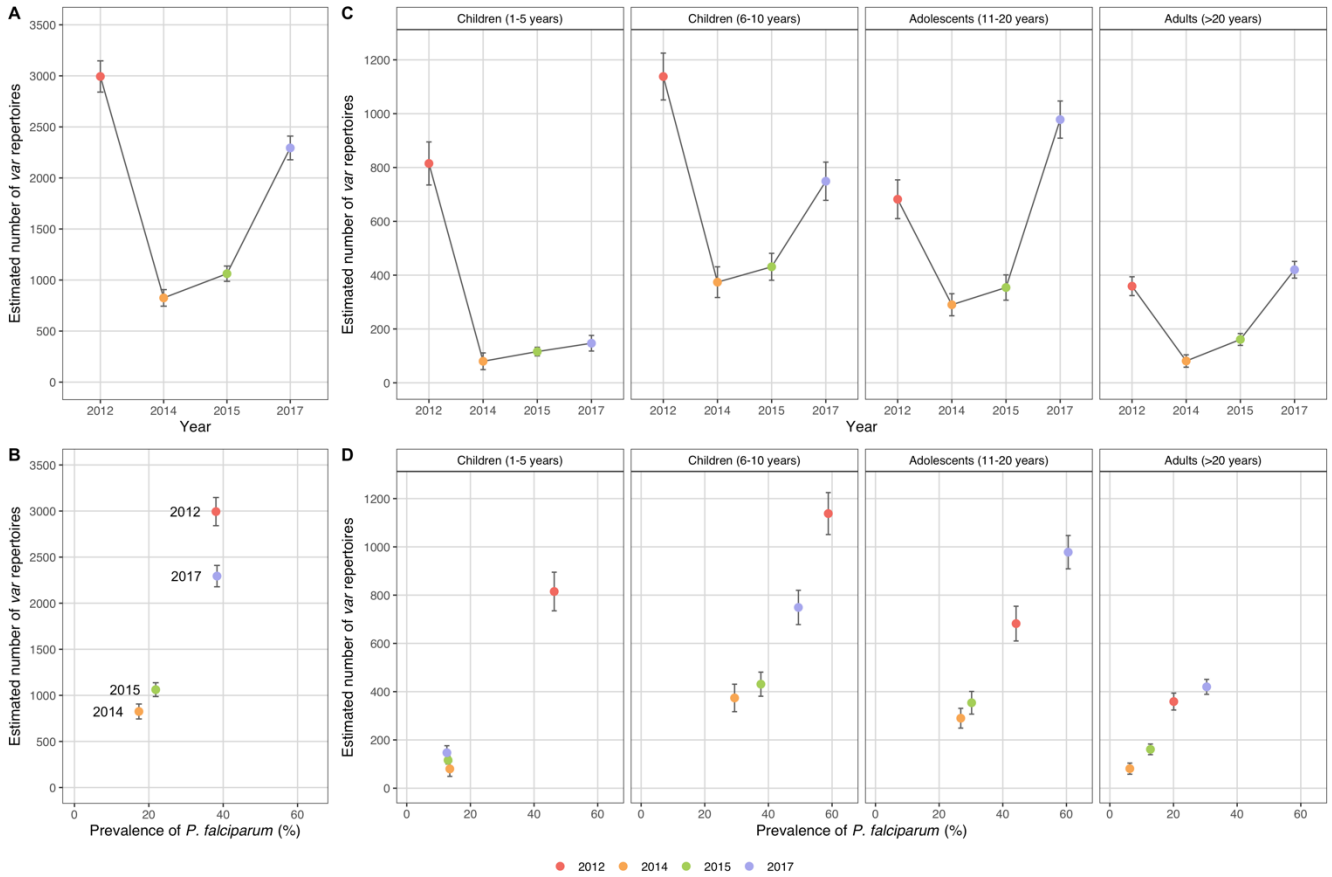


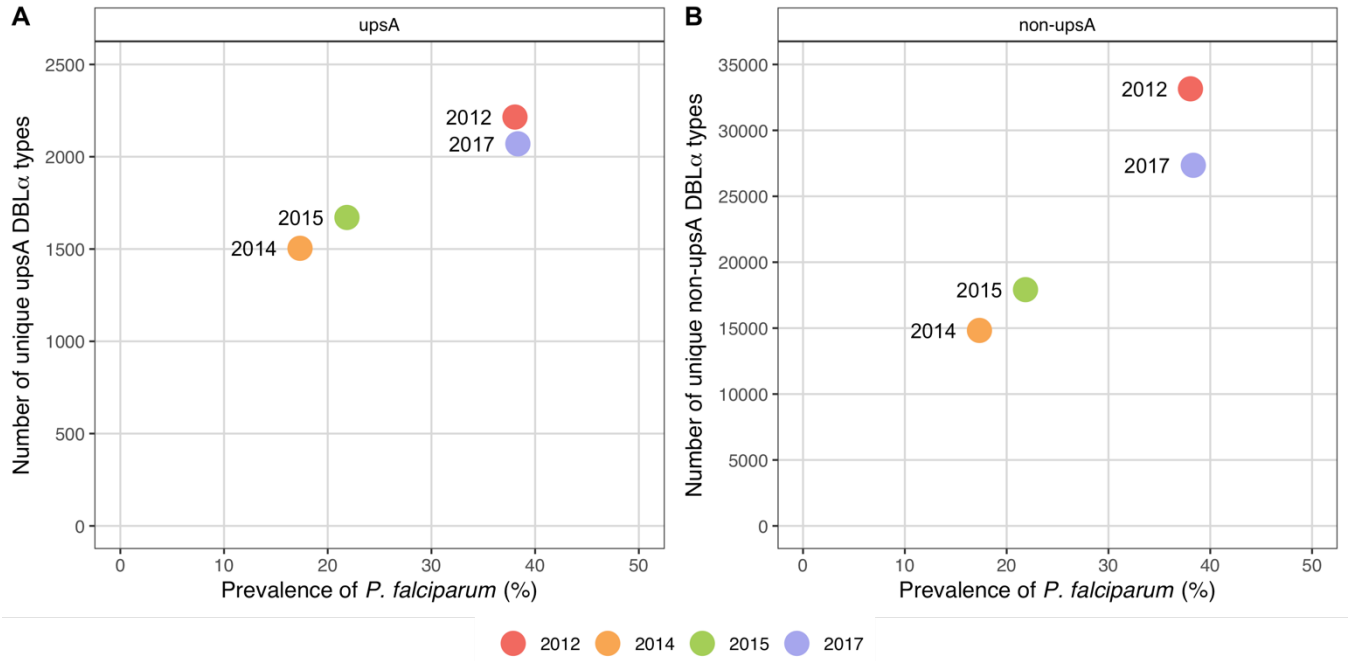
Figure 3. MOI_{var} distributions in 2012 (pre-IRS, red), 2014 (during IRS, orange), 2015 (post-IRS, green), and 2017 (SMC, purple). Estimated MOI_{var} distributions for the (**A**) study population and (**B**) for all age groups (years) in each survey for those isolates with DBLα sequencing data (Table S2). The median MOI_{var} values are indicated with the black dashed lines and have been provided in the top right corner (median MOI_{var} value [interquartile range, upper and lower quartiles]) along with the percentage of *P. falciparum* infections that were multiclonal (MOI_{var} > 1) in each survey and age group (years).



847

848 **Figure 4. Estimated number of *P. falciparum var* repertoires in 2012 (pre-IRS, red), 2014 (during IRS, orange), 2015**
849 **(post-IRS, green), and 2017 (SMC, purple).** The estimated number of *var* repertoires for the study population in
850 2012 (N = 2,994), 2014 (N = 825), 2015 (N = 1062), and 2017 (N = 2,294) (**A**) in each survey and (**B**) vs. *P. falciparum*
851 prevalence for those isolates with DBLα sequencing data (Table S2). To explore this further we also estimated the
852 number of *var* repertoires in all age groups (years) (**C**) in each survey and (**D**) vs. *P. falciparum* prevalence for those
853 isolates with DBLα sequencing data. Error bars represent the upper and lower limits of the 95% confidence interval
854 for the estimated number of *var* repertoires.

855
856



857
858 **Figure 5. UpsA and non-upsA DBLα type richness in 2012 (pre-IRS, red), 2014 (during IRS, orange), 2015 (post-IRS,**
859 **green), and 2017 (SMC, purple).** Number of unique (**A**) upsA and (**B**) non-upsA DBLα types (i.e., richness) observed
860 in each survey vs. *P. falciparum* prevalence based on those isolates with DBLα sequencing data (Table S2).
861

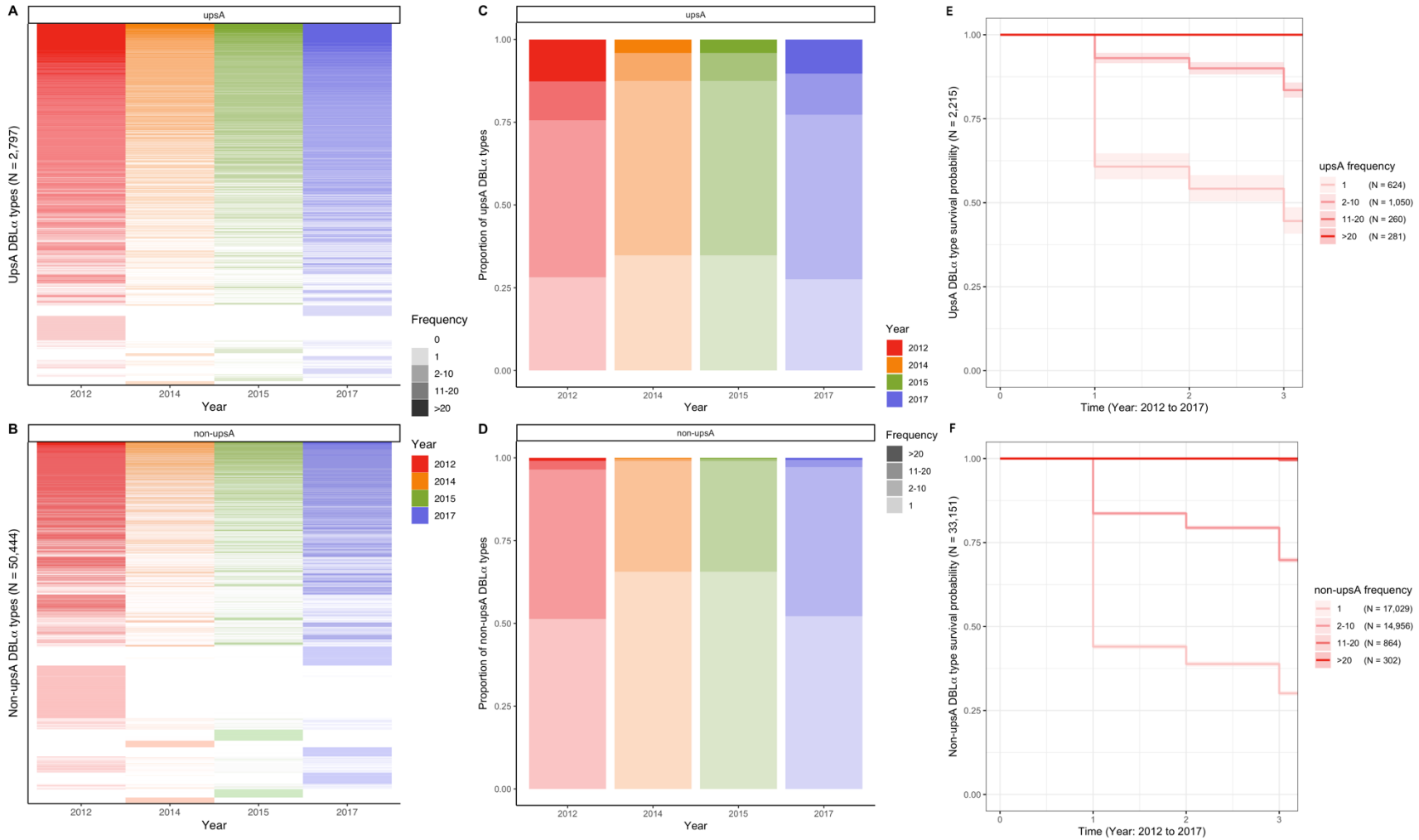


Figure 6. UpsA and non-upsA DBLα type frequencies and survival in 2012 (pre-IRS, red), 2014 (during IRS, orange), 2015 (post-IRS, green), and 2017 (SMC, purple). Heatmaps showing the patterns of diversity for the (**A**) upsA and (**B**) non-upsA DBLα types. The columns represent all the upsA and non-upsA DBLα types observed in the four surveys, and the rows represent each of the 2,797 upsA DBLα types and the 50,444 non-upsA DBLα types. White rows are used to denote the absence of a specific DBLα type, while the presence of a DBLα type is indicated in colour and further categorised (colour gradations) based on the frequency or the number of times (i.e., number of isolates) a DBLα type was observed within each survey (Frequency categories: 1, 2-10, 11-20, >20 isolates; Note the frequency category cut-offs were chosen based on the frequency distributions in Figure S9). The proportions of (**C**) upsA and (**D**) non-upsA DBLα types in each survey based on the number of times (i.e., number of isolates) they were observed within each survey. Kaplan-Meier survival curves for the (**E**) upsA and (**F**) non-upsA DBLα types across time (2012 to 2017) categorised based on

871 their frequency at baseline in 2012 (pre-IRS, red). The colored shaded areas represent the 95% confidence intervals and the number (N) of upsA and
872 non-upsA DBLα types in each frequency category are provided in parenthesis. These survival curves include only those upsA (N = 2,215) and non-
873 upsA (N = 33,151) DBLα types that were seen at baseline in 2012 (pre-IRS) as indicated in red. The x-axis indicates time where time "0" denotes
874 2012 (pre-IRS), "1" denotes 2014 (during IRS), "2" denotes 2015 (post-IRS), and finally "3" denotes 2017 (SMC).

875