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

Kathryn E. Tiedje1,2, Qi Zhan3,4, Shazia Ruybal-Pésantez2†, Gerry Tonkin-Hill2,5, Qixin He4, Mun Hua Tan1, Dionne C. Argyropoulos1, Samantha L. Deed1,2, Anita Ghansah6, Oscar Bangre7, Abraham R. Oduro7, Kwadwo A. Koram8, Mercedes Pascual4,9, and Karen P. Day1,2

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

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
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.
INTRODUCTION

Malaria in high-transmission endemic areas of sub-Saharan Africa (SSA) is characterised by vast diversity of the Plasmodium falciparum parasites from the perspective of antigenic variation (Chen et al., 2011; Day et al., 2017; Otto et al., 2019; Ruybal-Pesántez et al., 2022, 2017). As for other hosts of hyper-variable pathogens (Futse, Brayton, Dark, Knowles, & Palmer, 2008), children experiencing clinical episodes of malaria, eventually become immune to disease but not to infection. This results in a large reservoir of chronic asymptomatic infections, in hosts of all ages, which sustain transmission to mosquitos. Given the goal of malaria eradication by 2050, it is therefore of interest to examine how parasite population size changes following perturbation by major intervention efforts, both in terms of its size and underlying population genetics.

So, what do we mean by the parasite population size in the case of P. falciparum and how do we measure it? Parasite prevalence, detected by microscopy or more sensitive molecular diagnostics (e.g., PCR), describes the proportion of infected hosts. Studies of P. falciparum genetic diversity have shown that the majority of people in high-transmission endemic areas harbour diverse multiclonal infections measured as the complexity or multiplicity of infection (MOI) (e.g., (Anderson et al., 2000; Paul et al., 1995; Smith, Felger, Fraser-Hurt, & Beck, 1999; Sumner et al., 2021)) with complex population dynamics (Bruce et al., 2000; Farnert, Snounou, Rooth, & Bjorkman, 1997). These genetic data indicate much larger population sizes than observed by prevalence alone. Thus, from an ecological perspective we can consider a human host as a patch carrying a number of distinct “antigenic strains” of P. falciparum. The sum of these distinct infections over all hosts provides us with a census of the parasite count of relevance to transmission events in the host population. We refer to this census population size hereafter simply as population size but make clear that this measure is distinct from effective population size ($N_e$) as measured by neutral variation.

P. falciparum erythrocyte membrane protein 1 (PfEMP1), encoded by the var multigene family, is a strong contender for the strain-defining antigen of the blood stages of this malaria parasite due its role in establishing chronic infection (Zhang & Deitsch, 2022). The molecule mediates sequestration of P. 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 mosquitoes. 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-
overlapping sets of DBLα tags (Day et al., 2017; Rorick et al., 2018; Ruybal-Pesántez et al., 2022, 2017).

This observed pattern of repertoire diversity is not random, as shown with comparisons to randomizations that preserve gene frequencies (Day et al., 2017) and also not neutral, as shown by comparison of empirical patterns of repertoire similarity to those generated by a computational model (He et al., 2018).

This aspect of var population structure facilitates the accumulation of multiple diverse parasite repertoires detectable in the blood of the chronically infected semi-immune human hosts in an age-specific manner (Ruybal-Pesántez et al., 2022). As a consequence of the non-overlapping structure, there is a direct relationship between the number of var repertoires in a host and population size in that host or patch.

Adding up the patches or MOI_var (i.e., the number of var repertoires based on varcoding (Ruybal-Pesántez et al., 2022; Tiedje et al., 2022)) of individual hosts gives a measure of overall population size and a metric to evaluate the impact of interventions which reflects not only changes in transmission intensity but decay of chronic infections to shorten what is known as the duration of infection (Macdonald, 1957).

Considering the part of the genome under the strongest immune selection due to variant-specific immunity is akin to what has been recognized as central in pathogen phylodynamics, which examines the interplay between epidemiology, immunology, and evolutionary processes (Grenfell et al., 2004; Volz, Koelle, & Bedford, 2013). We are very familiar with this emphasis in relation to influenza strains and COVID-19 causing strains. With respect to pathogens with large population sizes, as is likely the case for P. falciparum in SSA, a view to measure diversity in adaptive genes rather than focus on neutral theory to understand population size has been presented (Barton, 2010). Here we examine the response to sequential malaria interventions with a definition of P. falciparum “strains” in a manner consistent with this view. Defining “strains by var repertoires of unlinked genes” diversifying by sexual recombination also allows us to take into consideration the dynamic nature of P. falciparum genomes caused by the high frequency of outcrossing expected in moderate- to high-transmission endemic areas (Babiker et al., 1994; Paul et al., 1995).

We specifically report an investigation of changes in parasite population size and structure through two sequential interventions between 2012 and 2017 in Bongo District located in the Upper East Region of Ghana, one of the 12 highest burden countries in Africa (World Health Organization, 2022). We present a
novel Bayesian approach to estimate MOI and therefore population size, which is based on counting the number of DBLα tags (or DBLα types) (Ruybal-Pesántez et al., 2022; Tiedje et al., 2022) but takes into consideration under-sampling of DBLα types in an isolate. We document *P. falciparum* prevalence as well as *var* diversity and population structure from baseline in 2012 through a major perturbation by a short-term indoor residual spraying (IRS) campaign managed under operational conditions, which reduced transmission intensity by > 90% as measured by the entomological inoculation rate (EIR) and decreased parasite prevalence by ~40-50% (Tiedje et al., 2022). Next, we followed what happened to parasite population size more than two years after the IRS intervention was discontinued and seasonal malaria chemoprevention (SMC) was introduced for children between the ages of 3-59 months (i.e., < 5 years) (Wagman et al., 2018). Detectable changes in parasite population size were seen as a consequence of the IRS intervention but this quantity rapidly rebounded 32-months after the intervention ceased. Overall, throughout the IRS, SMC, and subsequent rebound, the parasite population in humans remained large in size and retained the *var* population genetic characteristics of high transmission (i.e., high *var* diversity, low *var* repertoire overlap), demonstrating the overall resilience of the species to survive significant short-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 km²) 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
Starting in 2013, the AngloGold Ashanti Malaria Control Programme (AGAMal) in a public-private partnership with the Global Fund, scaled up IRS across all of the Upper East Region of northern Ghana (Gogue et al., 2020). As part of this initiative, three rounds of IRS with organophosphate formulations (i.e., non-pyrethroid) were rolled out prior to the start of the wet season between 2013 and 2015 (Figure 1A) in Bongo District (Tiedje et al., 2022). Based on AGAMal’s operational reports, IRS coverage in Bongo District was 91.8% in Round 1, 95.6% in Round 2, and finally 96.6% in Round 3 (AGAMal, personal communication). To monitor the impact of the IRS on the local vector population, monthly entomological surveys were undertaken between February 2013 and September 2015 (Tiedje et al., 2022). Using these surveys we observed that the monthly entomological inoculation rate (EIR) (infective bites/person/month (ib/p/m)), a measure of local transmission intensity, declined by > 90% at the peak of the wet season between August 2013 (pre-IRS) (EIR = 5.3 ib/p/m) and August 2015 (post-IRS) (EIR = 0.4 ib/p/m) (Tiedje et al., 2022). Following the IRS, SMC was rolled out in the Upper East Region by the National Malaria Elimination Programme (NMEP)/ Ghana Health Service (GHS) starting in 2016 (Gogue et al., 2020) (Figure 1A). SMC is the intermittent administration of full treatment courses of an antimalarial to children between the ages of 3-59 months (i.e., < 5 years) (WHO, 2012). Like other countries, the SMC drug of choice in Ghana is amodiaquine plus sulfadoxine-pyrimethamine, which is administered at monthly intervals during the high-transmission season (i.e., wet season) (WHO, 2012). The goal of this age-targeted intervention is to both clear current malaria infections and prevent malarial illness by maintaining a therapeutic concentration of an antimalarial in blood over the period of greatest risk (i.e., high-transmission season). Reported SMC coverage in Bongo District was 92.6% in 2016 (two rounds between August – September 2016) and 94.6% in 2017 (four rounds between September – December 2017) (Gogue et al., 2020).

Var genotyping and sequence analysis
Genomic DNA was extracted from the dried blood spots for all participants with a confirmed microscopic asymptomatic *P. falciparum* infection (i.e., isolate) (2,572 isolates, Table S2) using the QIAamp DNA mini kit (QIAGEN, USA) as previously described (Tiedje et al., 2017). For *var* genotyping or *var*coding, the sequences encoding the DBLa domains of the antigen-encoding *P. falciparum var* genes were amplified by PCR, pooled, and sequenced on a MiSeq 2x300 paired-end cycle protocol as previously published (New
York University Genome Technology Center, New York, NY, USA; Australian Genome Research Facility, Melbourne, Australia) (Figure S1) (Day et al., 2017; He et al., 2018; Ruybal-Pesántez et al., 2017). The raw sequence data was then cleaned and processed using our previously published customized bioinformatic pipeline (https://github.com/UniMelb-Day-Lab/DBLaCleaner) (He et al., 2018). This pipeline was used to de-multiplex and merge the paired reads as well as remove low-quality sequences and chimeras using several filtering parameters (see Figure S2 for additional details). These steps resulted in a total of 291,498 cleaned DBLa sequences for all surveys (Figure S2). To identify the unique DBLa types, we then clustered these cleaned DBLa sequences with 237,910 DBLa sequences previously published from Bongo (He et al., 2018; Pilosof et al., 2019; Rorick et al., 2018) at the standard 96% sequence identity (https://github.com/UniMelb-Day-Lab/clusterDBLaalpha) (Day et al., 2017). Our dataset was then further curated by translating the DBLa types (N = 68,507) into amino acid sequences and removing any DBLa types that could not be translated (i.e., contained a stop codon) (N = 137; 0.2%). The remaining DBLa types were then assigned to their most likely DBLa domain class (i.e., DBLa0, DBLa1, or DBLa2) using a hidden Markov model, and further classified based on the association of specific domain classes with semi-conserved upstream promoter sequences (ups) as either upsA or non-upsA DBLa types (https://github.com/UniMelb-Day-Lab/classifyDBLaalpha) (Ruybal-Pesántez et al., 2017). For additional information on the use of these bioinformatic pipelines a detailed tutorial is available (https://github.com/UniMelb-Day-Lab/tutorialDBLaalpha). Using a cut off of ≥ 20 DBLa types, DBLa sequencing data was obtained from 2,099 *P. falciparum* isolates (81.6%) with a total of 53,241 unique DBLa types and 289,064 DBLa sequences being identified in the study population (Table S2). This genotyping success was acceptable given we were working with low-density asymptomatic infections (Table S1). The median *P. falciparum* density was ~4 times higher for isolates with ≥ 20 DBLa types compared those that gave no or < 20 DBLa types (520 [IQR: 200-1880] parasites/μL vs. 120 [IQR: 40-200] parasites/μL, respectively).

**DBLa type diversity**

We monitored the impacts of the sequential interventions (i.e., IRS and SMC) on diversity by measuring changes in the population genetics of DBLa types at the population level (i.e., *P. falciparum* reservoir). Diversity was monitored using two measures, DBLa type richness and DBLa type frequency. Richness was
defined as the number of unique DBLα types observed (i.e., DBLα type pool size) in 2012, 2014, 2015, and 2017. DBLα type richness, however, does not provide any information about the relative frequencies of the different DBLα types in the population, as they are all weighted equally whether they are observed once or more frequently (e.g., observed in > 20 isolates per survey). To further examine the impacts of the interventions on DBLα type diversity we also assessed the frequency of each unique DBLα type in 2012, 2014, 2015, and 2017. Both upsA and non-upsA DBLα type diversity was measured due to their different biological features, chromosomal positions (i.e., subtelomeric regions vs. internal or central regions), as well as population genetics (Falk et al., 2009; Gardner et al., 2002; Jensen et al., 2004; Kaestli et al., 2006; Kalmbach et al., 2010; Kraemer et al., 2007; Kraemer & Smith, 2006; Kyriacou et al., 2006; Lavstsen, Salanti, Jensen, Arnot, & Theander, 2003; Normark et al., 2007; Rottmann et al., 2006; Warimwe et al., 2012, 2009; Zhang & Deitsch, 2022). The proportion of upsA and non-upsA var genes in a repertoire or single genome has been defined as ~15-20% and ~80-85%, respectively, based on whole genome sequencing (Rask et al., 2010). The upsA and non-upsA DBLα type proportions were partitioned as expected in our analyses, with the median proportions at the repertoire level being comparable in 2012 (19% upsA and 81% non-upsA), 2014 (22% upsA and 78% non-upsA), 2015 (21% upsA and 79% non-upsA), and 2017 (20% upsA and 80% non-upsA).

**Repertoire similarity as defined by pairwise type sharing**

To estimate genetic similarity between the DBLα repertoires (i.e., unique DBLα types identified in each isolate) identified from two isolates, pairwise type sharing (PTS) was calculated between all pairs of isolates within in each survey as previously described (Barry et al., 2007). PTS, analogous to the Sørensen Index, is a similarity statistic to evaluate the proportion of DBLα types shared between two isolate repertoires (i.e., DBLα repertoire overlap) and ranges from 0 (i.e., no DBLα repertoire overlap) to 1 (i.e., identical DBLα isolate repertoires), where < 0.50 = unrelated, 0.5 = recent recombinants/siblings, > 0.5 = related, and 1 = clones. PTS is a measure of identity-by-state (IBS) used to assess repertoire similarity (or relatedness) between isolates and is not used to infer inheritance from a recent common ancestor (i.e., identity-by-decent (IBD)) (Speed & Balding, 2015).

**DBLα isolate repertoire size and number of *P. falciparum* var repertoires**
For this study we have exploited the unique population structure of non-overlapping DBLα isolate repertoires to estimate isolate MOI\textsubscript{var}. To calculate MOI\textsubscript{var} the non-upsA DBLα types were chosen since not only are they more diverse and less conserved between isolate repertoires (i.e., low median PTS\textsubscript{non-upsA} ≤ 0.020) compared to the upsA DBLα types, but they have also been shown to have a more specific 1-to-1 relationship with a single var gene than upsA (Tan et al., 2023).

**Bayesian estimation of MOI\textsubscript{var} and associated population size**

The low to non-existent overlap of repertoires enables an estimation of MOI that relies on the number of (non-upsA) DBLα types sequenced from an individual’s isolate (Ruybal-Pesántez et al., 2022; Tiedje et al., 2022). A constant repertoire size or number of DBLα types in a parasite genome can be used to convert the number of types sequenced in an isolate to its estimated MOI (Ruybal-Pesántez et al., 2022; Tiedje et al., 2022). As such, the approach neglects the measurement error in this size introduced by targeted PCR and amplicon sequencing of var genes in an infection. Here, we extend the method to a Bayesian formulation and estimate the posterior distribution for each sampled individual for the probability of different MOI values. From individual posterior distributions, we can then obtain the estimated MOI frequency distribution for the population as a whole.

The two pieces of information required for our approach are the measurement error and the prior distribution of MOI. The measurement error is simply the repertoire size distribution, that is, the distribution of the number of non-upsA DBLα types sequenced given MOI = 1, which is empirically available (Figure S3). We refer to it as P(s | MOI=1) where s denotes repertoire size. In the following, s denotes the number of non-upsA DBLα types sequenced when MOI > 1. The prior distribution of MOI refers to the belief we have for what the actual MOI distribution might look like at the population level before empirical evidence is taken into consideration. For example, the prior distribution of MOI is likely to center around a higher value in high-transmission endemic areas than in low-transmission ones.

We can obtain P(s | MOI = m) from the serial convolution of the repertoire size distribution P(s | MOI = 1) and P(s | MOI = m - 1). Starting with the repertoire size distribution given a single infection, we can...
derive Pr(s | MOI = m) for m equal to 2,3,..., up until a maximum value of 20 (empirically determined), as follows:

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

where L and U are the lower and upper limit for the repertoire size, 10 and 45 respectively from the empirical repertoire size distribution.

For simplicity, we begin with a uniform prior. We use Bayes’ rule to derive a posterior distribution of MOI given a certain number of non-upsA DBLα types sequenced from an individual:

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

where k is the maximum value of MOI, here 20, as empirically determined.

To obtain the MOI distribution at the population level, we could either simply pool the maximum a posteriori MOI estimate for each sampled individual, or use a technique called mixture distribution. For the latter, we weighed each posterior MOI distribution for each sampled individual equally and sum over all posterior distributions at the individual level to derive the MOI distribution at the population level:

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

where n is the number of sampled individuals. These two approaches gave similar results for our empirical survey data as determined by the Kolmogorov-Smirnov Test. The obtained distance statistic is close to 0 and the corresponding p-value is non-significant across all surveys, indicating that the two estimates were drawn from the same distribution (Table S3). Given this similarity, we present hereafter only the result of pooling the maximum a posteriori MOI estimates. Note that we focused on individuals who had a confirmed microscopic asymptomatic P. falciparum infection for our MOI estimation.

To examine alternative priors, we considered empirical MOI distributions described in the literature incusing the Poisson, hyper-Poisson, and negative binomial distributions (Dietz, 1988; Henry, 2020). The
hyper-Poisson and negative binomial distributions can capture the overdispersion seen in the empirical distribution of MOI for certain areas and caused by factors such as heterogeneous biting. We investigated changing the parameters of a negative binomial distribution to generate priors with different means spanning a wide range of MOI values (mean MOI within \(\sim 3, \sim 9\)), including those seen in high-transmission endemic areas. We found that the final MOI distribution at the population level is fairly robust against the specification of those parameters. In other words, the estimated MOI distributions at the population level for different parameters do not significantly diverge from each other. The pairwise distance between these distributions measured by KL-divergence centres around 0, and the difference between their mean MOI values across different priors is close to 0 (Figure S3 and S4). Additionally, we performed Pearson correlation tests. We found that the correlation coefficients between the estimated MOI distributions at the population level across different parameters were close to 1 with a highly significant \(p\)-value (Figure S5 and S6). Thus, we provide in our analyses the estimated population MOI distribution using a uniform prior.

**Statistical analysis**

We used the R v4.0.2 for all data analyses with the collection of R packages in *tidyverse* being used for data curation along with *base, stats, gtsummary, and epiR* for the statistical analyses (R Core Team, 2018; Sjoberg, D, Whiting, Curry, Lavery, & Larmarange, 2021; Stevenson, 2020; Wickham et al., 2019). Continuous variables are presented as medians with interquartile ranges (IQRs) and discrete variables are presented using the observed proportions or calculated values with 95% confidence intervals (CIs). Chi-square \(X^2\) tests were utilized for univariate analyses of discrete variables to compare proportions (for multiple testing \(p\)-values were adjusted using the Bonferroni correction); non-parametric Mann-Whitney U and Kruskal-Wallis tests were used for comparing distributions between groups of continuous variables. Kaplan-Meier survival curves were generated for the time (i.e., number of surveys) to first event (i.e., when the DBL\(\alpha\) type was no longer observed/detected) comparing the ups\(A\) and non-ups\(A\) DBL\(\alpha\) types; \(p\)-values were determined using the log-rank test using the R packages *survival* and *survminer* (Kassambara, Kosinski, Biecek, & Scheipl, 2021; Therneau, 2023). The time interval to first event considered for all survival curves was the number of surveys or year (i.e., 2012, 2014, 2015, and 2017) that each DBL\(\alpha\) type
was observed and only includes those upsA (N = 2,215) and non-upsA (N = 33,151) DBLα types that were seen at baseline in 2012 (i.e., those DBLα types observed prior to the IRS).

DATA AVAILABILITY

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

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

Next, we wanted to explore changes in population size measured by MOIvar. As this metric is based on non-overlap of var repertoire diversity of individual isolates, specifically non-upsA DBLα types, we investigated whether DBLα isolate repertoire similarity (or overlap), as measured by PTS, increased following the sequential interventions (i.e., IRS and SMC). Figure 2 shows that median PTS values for both upsA and non-upsA DBLα types remained low in all surveys although the PTS distributions for both groups changed significantly at each of the study time points relative to the 2012 baseline survey (pre-IRS) (p-values < 0.001, Kruskal-Wallis test) (Figure 2). However, the change was in the direction of reduced similarity with lower median PTS scores and a larger number of isolates sharing no DBLα types (i.e., PTS = 0) in 2014, 2015, and 2017 compared to 2012. Relevant to measurement of MOIvar, the median PTS scores for non-upsA DBLα types were lower following the IRS intervention (PTSnon-upSA: 2014 = 0.013 and 2015 = 0.013 vs. PTSnon-upSA: 2012 = 0.020). In 2017, the non-upsA PTS frequency distributions shifted back towards higher median PTS scores (PTSnon-upSA = 0.016) and fewer isolates shared no DBLα types relative to 2014 and 2015 (Figure 2). To verify this pattern was not influenced by multiclonal infections (MOIvar > 1) we also examined isolates with monoclonal infections (MOIvar = 1) and found that this non-overlapping
structure, particularly for the non-upsA DBLα types, persisted regardless of infection complexity (Figure S7).

The raw data of non-upsA DBLα isolate repertoire sizes (Figure S8) were used to estimate MOI\textsubscript{var} as adjusted using the Bayesian approach (Figure 3). We observed that at baseline in 2012, the majority (89.2%) of the population across all ages carried multiclonal infections (median MOI\textsubscript{var} = 4 [IQR: 2 – 6]) (Figure 3A). Following the IRS intervention, the estimated MOI\textsubscript{var} distributions became more positively skewed, indicating that a lower proportion of participants harboured multiclonal infections in 2014 (64.5%; median MOI\textsubscript{var} = 2 [IQR: 1 – 3]) and 2015 (71.4%; median MOI\textsubscript{var} = 2 [IQR: 1 – 3]) compared to 2012 (Figure 3A). These reductions in median MOI\textsubscript{var} and the proportion of multiclonal infections, which were observed across all age groups (Figure 3B), are consistent with the > 90% decrease in transmission intensity following the IRS in turn reducing exposure to new parasite genomes. However, in 2017, both median MOI\textsubscript{var} (3 [IQR: 2 – 4]) and the proportion of multiclonal infections (78.9%) rebounded in all age groups, even among the younger children (1-5 years) targeted by SMC (Figure 3). While the prevalence of infection in 2017 remained low for the younger children (1-5 years) targeted by SMC, those infected still carried multiclonal infections (84.1%) (Figure 3B). Although the MOI\textsubscript{var} distributions across all age groups started to rebound in 2017 (i.e., less positively skewed compared to 2014 and 2015) they had not fully recovered to the 2012 baseline patterns (Figure 3). This was most apparent among the younger children (1-5 years), as a larger proportion of isolates in 2017, compared to 2012, had MOI\textsubscript{var} values equal to one or two, while a smaller proportion had MOI\textsubscript{var} values ≥ 5 (Figure 3B).

From isolate MOI\textsubscript{var} we also estimated population size or the number of \textit{P. falciparum var} repertoires circulating in the population during each survey. In 2014 during IRS, this number decreased by 72.4% relative to the 2012 baseline survey (pre-IRS) (Figure 4A), whereas prevalence decreased by 54.5% (Figure 4B). Although the number of \textit{var} repertoires increased slightly in 2015 relative to 2014, there were still 64.5% fewer \textit{var} repertoires in the population compared to 2012 (Figure 4A) in comparison to a 42.6% decrease in prevalence (Figure 4B). Importantly this loss of \textit{var} repertoires in 2014 and 2015 following the IRS intervention was seen for all age groups, with the greatest overall reductions (> 85%) being observed for the younger children (1-5 years) (Figure 4CD). However, in 2017, the number of diverse \textit{var} repertoires...
in the population rebounded, more than doubling between 2015 and 2017 (Figure 4AB). This increase in the number of var repertoires was seen for all age groups in 2017, except for the younger children (1-5 years) who were directly targeted by SMC (Figure 4CD). In fact, the greatest overall increase was observed for the adolescents (11-20 years), where the number of var repertoires in 2017 was ~1.4 times higher compared to 2012 (Figure 4CD). Similar trends in the number var repertoires were also observed for the older children (6-10 years) and adults (>20 years) in 2017, although the rebound was not as striking as that detected for the adolescents.

As parasite population size changed considerably during the sequential interventions, we investigated how the removal or loss of P. falciparum var repertoires and subsequent rebound altered var richness, measured as the number of unique upsA and non-upsA DBLα types in the parasite population in each survey. Richness at baseline in 2012 (pre-IRS) was high with a large number of unique DBLα types (upsA = 2,215; non-upsA = 33,151) (Figure 5, Table S2) and limited overlap of var repertoires (i.e., median PTSnon-upsA ≤ 0.020) seen in a relatively small study population of 685 microscopically positive individuals (Figure 2). In 2014, as P. falciparum prevalence and population size declined (Figure 4) so too did the number of DBLα types, resulting in a 32.1% and 55.3% reduction in richness for the upsA and non-upsA DBLα types, respectively compared to 2012 (Figure 5, Table S2). Again in 2015, as P. falciparum prevalence and population size remined low (Figure 4) DBLα type richness was still less than that observed in 2012 (24.6% and 45.9% reduction for upsA and non-upsA DBLα types, respectively) (Figure 5, Table S2). Finally, in 2017, we found that upsA and non-upsA DBLα type richness rebounded relative to 2014 and 2015, coincident with the increase in P. falciparum prevalence and parasite population size (Figure 4, Figure 5).

Given this reduction in var richness following the IRS intervention and subsequent rebound in 2017, we next investigated the relative frequency of the DBLα types in each survey (Figure 6, Figure S9). By examining frequency, we discovered that not only were individual upsA and non-upsA DBLα types observed more than once within a survey (i.e., observed in > 1 isolate), but that they were also found longitudinally (i.e., observed in > 1 survey) (Figure 6AB). DBLα types that were frequent (i.e., observed in 11-20 or >20 isolates), less frequent (i.e., observed in 2-10 isolates) or only seen once, typically maintained a similar frequency count regardless of ups grouping (i.e., upsA and non-upsA) or survey (i.e., 2012, 2014,
In 2014 and 2015, following the IRS, there was a significant increase in the proportion of upsA and non-upsA DBLα types in the lower frequency categories (*p*-value < 0.001, Mann-Whitney U test), with all DBLα types becoming rarer in the population (Figure 6CD). This change can be attributed to the removal of *P. falciparum var* repertoires (Figure 4) and loss of upsA and non-upsA DBLα types richness (Figure 5), which disproportionally affected those DBLα types seen once. This shift to all DBLα types becoming rarer following IRS changed in 2017, where the proportion of DBLα types in the more frequent categories (i.e., 2-10, 11-20, or > 20 isolates) significantly increased while the proportion seen once decreased (*p*-values < 0.001, Mann-Whitney U tests) (Figure 6CD).

Finally, to explore these changes in DBLα type diversity (i.e., richness and frequency), we compared the probability of survival for the DBLα types identified at baseline in 2012 (pre-IRS) and found that the upsA DBLα types persisted significantly longer in the population relative to the non-upsA (*p* < 0.001, log-rank test). Despite this difference, the upsA and non-upsA DBLα types that survived and were observed at multiple study time points (i.e., 2012, 2014, 2015, and 2017) were those that were most frequent in the population at baseline in 2012 (i.e., observed in 11-20 and >20 isolates) (Figure 6EF). As expected, the DBLα types that were only observed once in 2012, were significantly less likely to be seen longitudinally (*p*-value < 0.001, log-rank test) (Figure 6EF).
DISCUSSION

*P. falciparum* populations in high-transmission endemic areas in SSA, are characterised by extensive diversity, high rates of recombination, as well as frequent multiclonal infections. Here we defined census population size of *P. falciparum* as a new metric to explore the efficacy of malaria interventions in such areas typified by Bongo, Ghana. We captured age-specific changes in census population size in response sequential malaria interventions. IRS reduced the MOI$_{var}$ parasite population size substantially with the greatest reductions (85%) seen in the younger children (i.e., 1-5 years). More than two years after cessation of IRS, rebound in 2017 was rapid in all age groups with the exception of younger children targeted by SMC. Population sizes in adolescents and adults showed they carried more infections in 2017 than at baseline in 2012. This is indicative of a loss of immunity during IRS and warrants further investigation. During and following the IRS and SMC interventions, *var* diversity still remained high and *var* repertoire overlap remained low, reflecting characteristic properties of high transmission and demonstrating the overall resilience of the species to survive significant short-term perturbations.

What was striking about the Bongo study was the speed with which rebound in MOI$_{var}$ per person and overall population size occurred, once the short-term IRS was discontinued. We looked for a potential explanation in our genetic data. PTS and population frequency data showed that many of the DBLα types occurred in multiple repertoires or genomes. This enabled the survival of these more frequent DBLα types through the interventions, facilitating rebound by maintenance of this diversity. The other notable population genetic result of our study was the failure to increase similarity (or relatedness by state) of *var* repertoires by reducing transmission by > 90% via IRS. From a baseline of very low overlap in repertoires you need outcrossing to create relatedness. However, this was less likely to happen due to reduced transmission as a result of IRS. Rebound, with associated increases in transmission, led to a small increase in *var* repertoire similarity.

Our molecular approach to measure population size has been to sum MOI$_{var}$ in individual hosts with microscopically detectable infections. Like any diagnostic method there are limits to sensitivity and specificity, which can be more or less tolerated dependent upon the purpose of the study. Here we have looked at relative changes in population size with sequential interventions using an interrupted time series.
study design and observed changes by measuring $\text{MOI}_{\text{var}}$. We have accounted for missing DBL$\alpha$ type data where complete $\text{var}$ repertoires may not have been sequenced using a Bayesian method based on empirical knowledge of the measurement error. This approach has conceptual relation to the Bayesian approach by Johnson and Larremore (Johnson & Larremore, 2022) to estimate complete repertoire size of, and overlap between, monoclonal infections from incomplete sampling of DBL$\alpha$ types. It would be valuable to consider how to jointly infer the number of DBL$\alpha$ types, isolate overlap, and $\text{MOI}_{\text{var}}$ from sequencing data of multiclonal infections. Our measurement of population size based on $\text{MOI}_{\text{var}}$ will be subject to other sampling errors which may in the end be more significant (discussed in detail in Labbé et al. (Labbé et al., 2023)). Low parasitemia typical of asymptomatic infections, small blood volumes, clinical status, and/or within host dynamics, including synchronicity, will all create sampling problems but these are common to all measures of MOI.

Previously, we have drawn attention to underestimation of the number of DBL$\alpha$ types of related parasites from a cross by $\text{var}$ genotyping (Labbé et al., 2023). Such related parasites have been observed in experimental crosses (Otto et al., 2018) and must be created frequently in high transmission due to extensive outcrossing (Babiker et al., 1994; Paul et al., 1995). Single clone genomics experiments using biallelic single nucleotide polymorphism (SNPs) from whole genome sequencing data have also detected related parasites using IBD in clinical infections from humans in a high-transmission area of Malawi (Nkhoma et al., 2020). Here we have analysed only low density, chronic asymptomatic infections under strong immune selection in semi-immune hosts whom we have shown select against related parasites (He et al., 2018; Ruybal-Pesantez et al., 2022). When considering the importance of possible exclusion of parasites related by descent, the only sure way to detect such parasites in high transmission is by single cell genomics, a methodology of limited application to malaria surveillance due to practicality and cost of scale up. Again, the error from failure to sample related infections must be weighed up against the issues of under sampling as described above.

Of interest to those working on malaria surveillance in high-transmission areas is whether measuring MOI, and in turn population size, using $\text{var}$ repertoires or single copy antigen genes or by neutral variation with biallelic SNPs, is more relevant to the elimination agenda in high burden countries. Single copy antigen
genes such as *csp*, *msp2*, and *ama1* have each been used to measure MOI (e.g., (Falk et al., 2006; Lerch et al., 2017; Nelson et al., 2019)) and have become part of newer genetic panels (e.g., Paragon v1 (Tessema et al., 2020) and AMPLseq v1 (LaVerriere et al., 2022)). These genes are under balancing selection with geographically common alleles. Alone, they are much less diverse than the potential combinations of ~45 non-upsA var genes so will underestimate population size. Var genotyping has also proven more sensitive in high transmission to measure MOI due to the high prevalence of multiclonal infections, which cannot be phased with either biallelic SNP panels (Labbé et al., 2023; Tessema et al., 2020; Watson et al., 2021) or combinations of single copy antigen genes (Sumner et al., 2021).

A recent paper from Senegal using a 24-SNP barcode observed highly-related parasites circulating in human hosts with symptomatic malaria (Wong et al., 2022). Based on modelling SNP barcodes from polyclonal infections analysed by THE REAL McCOIL to calculate MOI and a new genetic metric $R_H$ to quantify genetic relatedness, they go on to propose these infections were a result of cotransmission (i.e., single infectious bite) rather than superinfection (i.e., multiple infectious bites). They claim from $R_H$ relatedness estimates that cotransmission was “ubiquitous” in symptomatic infections in three sites, which would be designated by the WHO classification as low- to very low-transmission areas according to *P. falciparum* prevalence (WHO/GMP, 2017). Instead, we are working in a high-transmission area (see Methods) with chronic asymptomatic infections under strong immune selection as discussed above. We found little empirical evidence of related parasites within hosts using polymorphic microsatellites markers and varcoding when the same 24-SNP barcode made monoclonal infections look highly related or similar (Ghansah et al., 2023). We agree that the issue of cotransmission vs. superinfection is important and must be considered when thinking about MOI in relation to transmission intensity.

In summary, our findings provide parasite population insights into why rebound is the inevitable consequence of such short-term IRS interventions unless you simultaneously target the highly diverse, long-lived parasite population in humans, not just children < 5 years by SMC. Of potential translational significance for malaria molecular surveillance, we identify new metrics, especially MOI$_{var}$ and census population size as well as var frequency category, informative to monitor and evaluate interventions in high-transmission areas with multiclonal infections and high rates of outcrossing. Such metrics could be
used longitudinally to detect incremental gains of transmission-reducing interventions, including IRS, LLINs, and vaccines to perturb the high-transmission characteristics of the parasite population in humans in high-burden countries in SSA.
REFERENCES


Labbé, F., He, Q., Zhan, Q., Tiedje, K. E., Argyropoulos, D. C., Tan, M. H., … Pascual, M. (2023). Neutral vs. non
Kraemer, S. M.
gene transcription in Plasmodium falciparum isolates from patients with cerebral malaria compared to
hyperparasitaemia. Molecular and Biochemical Parasitology, 150, 211–218. doi:
10.1016/j.molbiopara.2006.08.005
Labbé, F., He, Q., Zhan, Q., Tiedje, K. E., Argyropoulos, D. C., Tan, M. H., … Pascual, M. (2023). Neutral vs. non-
eutral genetic footprints of Plasmodium falciparum multiclonal infections. PLoS Computational Biology, 19(1), e1010816. doi: doi.org/10.1101/2022.06.27.497801
and implementation of multiplexed amplicon sequencing panels to serve genomic epidemiology of
infectious disease: A malaria case study. Molecular Ecology Resources, (March), 2285–2303. doi:
10.1111/1755-0998.13622


patterns of DBLα 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.


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

DATA AVAILABILITY STATEMENT. The sequences utilized in this study are publicly available in GenBank under BioProject Number: PRJNA 396962. All data associated with this study are available in the main text or in the Supporting Information. All custom code is available in an open source repository: (1) DBLaCleaner pipeline is available at https://github.com/Unimelb-Day-Lab/DBLaCleaner, and the (2) clusterDBLa pipeline is available at https://github.com/Unimelb-Day-Lab/clusterDBLa, and the (3) classifyDBLa pipeline is available at https://github.com/Unimelb-Day-Lab/classifyDBLa. A dataset to demo this custom code is available at https://github.com/UniMelb-Day-Lab/tutorialDBLa. For additional information on the use of the Bayesian approach to estimate MOIvar please see https://github.com/qzhan321/ME-Bayesian-MOI-estimation.

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


COMPETING INTERESTS. The authors declare no competing interests.
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\textsubscript{var} distributions in 2012 (pre-IRS, red), 2014 (during IRS, orange), 2015 (post-IRS, green), and 2017 (SMC, purple). Estimated MOI\textsubscript{var} distributions for the (A) study population and (B) for all age groups (years) in each survey for those isolates with DBL\textalpha{} sequencing data (Table S2). The median MOI\textsubscript{var} values are indicated with the black dashed lines and have been provided in the top right corner (median MOI\textsubscript{var} value [interquartile range, upper and lower quartiles]) along with the percentage of \textit{P. falciparum} infections that were multiclonal (MOI\textsubscript{var} > 1) in each survey and age group (years).
Figure 4. Estimated number of *P. falciparum* var repertoires in 2012 (pre-IRS, red), 2014 (during IRS, orange), 2015 (post-IRS, green), and 2017 (SMC, purple). The estimated number of var repertoires for the study population in 2012 (N = 2,994), 2014 (N = 825), 2015 (N = 1062), and 2017 (N = 2,294) (A) in each survey and (B) vs. *P. falciparum* prevalence for those isolates with DBLα sequencing data (Table S2). To explore this further we also estimated the number of var repertoires in all age groups (years) (C) in each survey and (D) vs. *P. falciparum* prevalence for those isolates with DBLα sequencing data. Error bars represent the upper and lower limits of the 95% confidence interval for the estimated number of var repertoires.
Figure 5. UpsA and non-upsA DBLα type richness in 2012 (pre-IRS, red), 2014 (during IRS, orange), 2015 (post-IRS, green), and 2017 (SMC, purple). Number of unique (A) upsA and (B) non-upsA DBLα types (i.e., richness) observed in each survey vs. *P. falciparum* prevalence based on those isolates with DBLα sequencing data (Table S2).
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
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 non-upsA DBLα types in each frequency category are provided in parenthesis. These survival curves include only those upsA (N = 2,215) and non-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 2012 (pre-IRS), “1” denotes 2014 (during IRS), “2” denotes 2015 (post-IRS), and finally “3” denotes 2017 (SMC).