



OPEN ACCESS

EDITED BY

Amy Wesolowski,
Bloomberg School of Public Health, Johns
Hopkins University, United States

REVIEWED BY

Jessica Briggs,
University of California, San Francisco,
United States
Christine Markwalter,
Duke University, United States

*CORRESPONDENCE

Karen P. Day
✉ karen.day@unimelb.edu.au

†These authors have contributed
equally to this work and share
first authorship

SPECIALTY SECTION

This article was submitted to
Epidemiology and Ecology,
a section of the journal
Frontiers in Parasitology

RECEIVED 12 October 2022

ACCEPTED 14 March 2023

PUBLISHED 03 April 2023

CITATION

Ghansah A, Tiedje KE, Argyropoulos DC,
Onwona CO, Deed SL, Labbé F, Oduro AR,
Koram KA, Pascual M and Day KP (2023)
Comparison of molecular surveillance
methods to assess changes in the
population genetics of *Plasmodium
falciparum* in high transmission.
Front. Parasitol. 2:1067966.
doi: 10.3389/fpara.2023.1067966

COPYRIGHT

© 2023 Ghansah, Tiedje, Argyropoulos,
Onwona, Deed, Labbé, Oduro, Koram,
Pascual and Day. This is an open-access
article distributed under the terms of the
[Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that
the original publication in this journal is
cited, in accordance with accepted
academic practice. No use, distribution or
reproduction is permitted which does not
comply with these terms.

Comparison of molecular surveillance methods to assess changes in the population genetics of *Plasmodium falciparum* in high transmission

Anita Ghansah^{1†}, Kathryn E. Tiedje^{2†}, Dionne C. Argyropoulos^{2†},
Christiana O. Onwona¹, Samantha L. Deed², Frédéric Labbé³,
Abraham R. Oduro⁴, Kwadwo A. Koram⁵, Mercedes Pascual^{3,6}
and Karen P. Day^{2*}

¹Department of Parasitology, Noguchi Memorial Institute for Medical Research, University of Ghana, Legon, Ghana, ²Department of Microbiology and Immunology, The University of Melbourne, Bio21 Institute and Peter Doherty Institute, Melbourne, VIC, Australia, ³Department Ecology and Evolution, The University of Chicago, Chicago, IL, United States, ⁴Navrongo Health Research Centre, Ghana Health Service, Navrongo, Ghana, ⁵Epidemiology Department, Noguchi Memorial Institute for Medical Research, University of Ghana, Legon, Ghana, ⁶Santa Fe Institute, Santa Fe, NM, United States

A major motivation for developing molecular methods for malaria surveillance is to measure the impact of control interventions on the population genetics of *Plasmodium falciparum* as a potential marker of progress towards elimination. Here we assess three established methods (i) single nucleotide polymorphism (SNP) barcoding (panel of 24-biallelic loci), (ii) microsatellite genotyping (panel of 12-multiallelic loci), and (iii) *var* coding (fingerprinting *var* gene diversity, akin to microhaplotyping) to identify changes in parasite population genetics in response to a short-term indoor residual spraying (IRS) intervention. Typical of high seasonal transmission in Africa, multiclonal infections were found in 82.3% (median 3; range 1-18) and 57.8% (median 2; range 1-12) of asymptomatic individuals pre- and post-IRS, respectively, in Bongo District, Ghana. Since directly phasing multilocus haplotypes for population genetic analysis is not possible for biallelic SNPs and microsatellites, we chose ~200 low-complexity infections biased to single and double clone infections for analysis. Each genotyping method presented a different pattern of change in diversity and population structure as a consequence of variability in usable data and the relative polymorphism of the molecular markers (i.e., SNPs < microsatellites < *var*). *Var* coding and microsatellite genotyping showed the overall failure of the IRS intervention to significantly change the population structure from pre-IRS characteristics (i.e., many diverse genomes of low genetic similarity). The 24-SNP barcode provided limited information for analysis, largely due to the biallelic nature of SNPs leading to a

high proportion of double-allele calls and a view of more isolate relatedness compared to microsatellites and *var* coding. Relative performance, suitability, and cost-effectiveness of the methods relevant to sample size and local malaria elimination in high-transmission endemic areas are discussed.

KEYWORDS

population genetics, high transmission, SNPs (single nucleotide polymorphisms), microsatellites, *var* genes, molecular markers, malaria control interventions, *Plasmodium falciparum*

1 Introduction

The World Health Organization's malaria elimination strategy recommends the use of molecular methods for surveillance to measure the impact of malaria control interventions on the population genetics of *Plasmodium falciparum* (World Health Organization, 2022). High-transmission endemic areas present a specific challenge for molecular surveillance, especially in sub-Saharan Africa (SSA) where ~95% of the global malaria cases occurred in 2021 (World Health Organization, 2022). Here there is extensive genomic diversity of the parasite population (e.g., Pf6 database (MalariaGEN Consortium, 2021)) with many infected individuals carrying multiclonal *P. falciparum* infections (i.e., complexity or multiplicity of infection (MOI) > 1) (Anderson et al., 2000; Mobegi et al., 2012; Auburn and Barry, 2017) with a genetic structure consistent with frequent sexual recombination (meiosis) (Babiker et al., 1994; Paul et al., 1995). This contrasts with the genetic diversity seen in low-transmission regions of South America and Southeast Asia, as well as areas of intense malaria control in SSA such as in Senegal and Zambia, where parasite populations are largely clonal (i.e., MOI = 1) and highly related (Anderson et al., 2000; Anthony et al., 2005; Pumpaibool et al., 2009; Daniels et al., 2013; Daniels et al., 2015; Noviyanti et al., 2015; Auburn and Barry, 2017).

Panels of multiallelic microsatellites have been widely used to look at *P. falciparum* population genetics in a range of transmission settings to define linkage disequilibrium (Anderson et al., 2000; Machado et al., 2004; Anthony et al., 2005; Mobegi et al., 2012; Yalcindag et al., 2012; Barry et al., 2013; Vera-Arias et al., 2019; Kattenberg et al., 2020). Most recently, biallelic single nucleotide polymorphisms (SNPs) have been used by malariologists working in low- to moderate-transmission settings to look at diversity and population structure of clinical infections in response to interventions (Daniels et al., 2013; Nkhoma et al., 2013; Daniels et al., 2015; Bei et al., 2018). Data from both SNPs and microsatellites were analyzed by neutral theory (Anderson et al., 2000; Daniels et al., 2008). However, when used in high transmission, SNP barcoding and microsatellite genotyping have limitations for genetic diversity inferences as multiclonal infections are common, resulting in multilocus haplotypes that cannot be accurately reconstructed or phased from genotyping data. Two

empirical solutions have been proposed to analyze SNP and/or microsatellite data for population genetics from these complex infections. The simplest solution is to only use monoclonal *P. falciparum* infections, thereby reducing the sample size. Consequently many samples are collected to analyze the few (Daniels et al., 2013; Daniels et al., 2015; Verity et al., 2020). The alternative is to use multiclonal infections to identify the major allele at each of the SNP or microsatellite loci to infer or construct a "dominant infection haplotype" dataset (Anderson et al., 2000; Verity et al., 2020). More recently, various computational methods have been developed to infer haplotypes (Zhu et al., 2018; Zhu et al., 2019). These remain largely untested on datasets from high transmission and do not account for the rate of sexual recombination seen in moderate- to high-transmission endemic areas (Babiker et al., 1994; Paul and Day, 1998).

To characterize *P. falciparum* population structure in response to interventions, we have developed an empirical approach known as *var* genotyping or *var* coding based on *var* genes (~40-60 *var* genes per genome) which encode for the major variant surface antigen of asexual blood stages known as *P. falciparum* erythrocyte membrane protein 1 (PfEMP1) (Bull et al., 1998; Gardner et al., 2002; Rask et al., 2010; Otto et al., 2019). *Var* coding is a fingerprinting method by amplicon sequencing, akin to microhaplotyping, which identifies the diversity of the *var* genes per *P. falciparum* infection (i.e., isolate) using sequences encoding the immunogenic Duffy-binding-like α domain (DBL α) of PfEMP1 variants, defined as DBL α types. Analysis of the relationship between DBL α types and exon 1 of *var* genes in Malawi and Ghana has shown that each DBL α type, especially upsB and upsC types (i.e., non-upsA), is predominantly associated with a single *var* gene, and therefore DBL α type diversity acts as a suitable surrogate for *var* diversity per host and in the population (Tan et al., 2023). Prior population investigations based on DBL α types in high-transmission settings of SSA have demonstrated that sequences of this marker are highly diverse in local endemic areas with thousands of variants described (Barry et al., 2007; Chen et al., 2011; Day et al., 2017; Ruybal-Pesántez et al., 2017b; Rorick et al., 2018; Tonkin-Hill et al., 2021; Ruybal-Pesántez et al., 2022). Parasite genomes in natural populations in SSA are typically composed of distinct sets of *var* genes (i.e., *var* repertoires), which are largely non-overlapping (Chen et al., 2011; Day et al., 2017; Ruybal-Pesántez et al., 2017b;

Rorick et al., 2018; Ruybal-Pesántez et al., 2022) likely due to immune selection (He et al., 2018). This makes it possible to count the number of diverse *var* repertoires (termed MOI_{var}) present in an isolate by simply counting the number of DBL α types in an isolate (repertoire size) and then dividing by the median number of DBL α types amplified per genome (Ruybal-Pesántez et al., 2022; Tiedje et al., 2022). The method has been shown to work well across MOI ranges from 1 to > 20 (Labbé et al., 2023). *Varcoding* does not require the construction of multilocus haplotypes for each clone in an isolate (i.e., phasing) due to the non-overlapping population structure of *var* repertoires allowing multiple *P. falciparum* clones to accumulate within a human host in high transmission (Barry et al., 2007; Chen et al., 2011; Day et al., 2017; Ruybal-Pesántez et al., 2017b; Ruybal-Pesántez et al., 2022). Using this method, measures of heterozygosity and similarity of isolate repertoires are easily calculated by the pairwise type sharing (PTS) statistic, a measure of identity-by-state (IBS) (Barry et al., 2007; Speed and Balding, 2015).

For molecular surveillance to be used routinely in Africa, it should efficiently report changes in diversity and population structure of *P. falciparum* locally and be easily deployed in endemic areas, especially in high transmission, in a cost-effective way. This requires us to identify genotyping techniques that provide information with few genetic markers to analyze relatively small sample sizes in regional laboratories. Here we assess the performance of three field applicable genotyping methods, i.e., (i) SNP barcoding (panel of 24-biallelic loci), (ii) microsatellite genotyping (panel of 12-multiallelic loci), and (iii) *varcoding* under the conditions of high seasonal transmission in Ghana, one of the 11 highest burden countries for malaria globally (World Health Organization & Roll Back Malaria Partnership to End Malaria, 2019; World Health Organization, 2021). The effectiveness of these methods to describe changes in the diversity and similarity of *P. falciparum* at the end of the wet season before (October 2012) and after (October 2015) the implementation of three rounds of indoor residual spraying (IRS) using non-pyrethroid insecticides, managed under operational conditions (Tiedje et al., 2022) is documented. The IRS reduced transmission intensity in Bongo by $> 90\%$ at the peak of the wet season as measured by the monthly entomological inoculation rate (EIR) (infective bites/person/month (ib/p/m)) 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). Coincident with this decrease in transmission, declines in microscopic *P. falciparum* prevalence (42.0% to 27.0%) and median densities (520 parasites/ μ L to 320 parasites/ μ L) were also observed pre- to post-IRS (Tiedje et al., 2022).

Each genotyping method presented a different pattern of change of population diversity and structure when sampling monoclonal or low-complexity infections, consistent with marker variability. The 24-SNP barcode was least informative, largely due to the biallelic nature of SNPs leading to a high proportion of double-allele calls (DACs), whereas microsatellites showed high haplotype diversity with ten markers but no measurable change in population structure after the IRS. For this sample size of 200 isolates, *varcoding* and microsatellite genotyping provided the most informative analysis

showing the overall failure of the IRS intervention to significantly change the population structure from pre-IRS (i.e., many diverse genomes of low genetic similarity). Although we note that *varcoding* did detect a subtle shift towards less similarity of *var* repertoires after IRS suggesting the method is sensitive to decreased transmission creating fewer recombinant genomes as a consequence of less outcrossing. The results of this study provide useful information for high-burden countries in Africa looking at strategies to deploy molecular surveillance to assess changes in parasite diversity and population structure in relatively small sample sizes from sentinel sites in local endemic areas.

2 Materials and methods

2.1 Study population and ethical approvals

The *P. falciparum* data utilized in this study was collected from participants with microscopically confirmed asymptomatic *P. falciparum* infections (i.e., isolates) at the end of the wet season (i.e., high-transmission season) from two proximal catchment areas (i.e., Ve/Gowrie and Soe, with a sampling area of ~ 60 km²) in Bongo District, Ghana (hereinafter referred to collectively as “Bongo”) (Tiedje et al., 2022). Using an interrupted time-series study design, two age-stratified surveys of $\sim 2,000$ participants per survey were undertaken pre-IRS (October 2012) and post-IRS (October 2015) against a backdrop of widely distributed long-lasting insecticidal nets (LLIN) (Tiedje et al., 2022). Bongo, located in the Upper East Region, is categorized as a high-transmission area based on the WHO “A Framework for Malaria Elimination” (WHO/GMP, 2017) where *P. falciparum* prevalence was $\geq 35\%$ in 2012 (73.8%) and 2015 (41.6%) (Tiedje et al., 2022). Detailed information on the study site, the study population, inclusion/exclusion criteria, data collection procedures, etc. have been described previously (Tiedje et al., 2017; Tiedje et al., 2022). The study was reviewed/approved by the ethics committees at the Navrongo Health Research Centre (Navrongo, Ghana), Noguchi Memorial Institute for Medical Research (Legon, Ghana), The University of Chicago (Chicago, United States), and The University of Melbourne (Melbourne, Australia).

2.2 Genotyping methods

2.2.1 *Varcoding*

For *varcoding*, the sequences encoding the DBL α tags of *P. falciparum var* genes were amplified by PCR, pooled, and sequenced on the Illumina MiSeq platform (2x300bp paired-end configuration) (New York University Genome Technology Center, New York, NY, United States; Australian Genome Research Facility, Melbourne, Australia) (Day et al., 2017; Ruybal-Pesántez et al., 2017b; He et al., 2018; Ruybal-Pesántez et al., 2022). This high-throughput sequencing method was utilized for all participants with microscopically confirmed asymptomatic *P. falciparum* infections in the pre-IRS (N = 808) and post-IRS (N = 545) surveys (Figure S1, Table S1). The DBL α sequence tags were then cleaned, clustered,

and classified using a suite of custom bioinformatic pipelines (Ruybal-Pesántez et al., 2017b; He et al., 2018). For a detailed description of these pipelines please see the tutorial: <https://github.com/UniMelb-Day-Lab/tutorialDBLalpha>.

2.2.2 24-SNP molecular barcoding

The 24-SNP molecular barcoding was undertaken for the isolates selected for this comparative analysis (Table S2) using a 384-well format using the method described by Daniels et al. (2008). Briefly for each reaction, template and water in a total volume of 2.5 μ L was added to a 2.5 μ L mix made up of 0.125 μ L 40 \times SNP assay and 2.5 μ L Master Mix in a 384-well optical PCR plate and mixed, for a total reaction volume of 5 μ L. The plate was covered with an optical plate seal and amplified in an ABI 7900 HT (Department of Parasitology, Noguchi Memorial Institute for Medical Research, Legon, Ghana). Following the amplification, all isolates were analyzed using the Applied Biosystem's proprietary Allelic Discrimination and Absolute Quantitation software.

2.2.3 Microsatellite genotyping

Microsatellite genotyping and sequencing methods utilized for the isolates included in this analysis (Table S2) have been previously genotyped for ten putatively neutral microsatellite markers (2490, TA81, TA87, TA109, TA60, POLYA, ARA2, Pfg377, PfpK2, and TA40) (Argyropoulos et al., 2021).

2.3 Multiplicity of infection

Multiplicity of infection by *var*coding (i.e., MOI_{var}) was calculated based on the number of DBL α types in an isolate due to the limited overlap between DBL α isolate repertoires as previously shown, especially in high transmission (Day et al., 2017; Ruybal-Pesántez et al., 2017b; Ruybal-Pesántez et al., 2022; Tiedje et al., 2022). To calculate MOI_{var} the non-upsA DBL α types were chosen since not only are they more diverse and less conserved than the upsA DBL α types, but they have also been shown to have a more specific 1-to-1 relationship with a single *var* gene compared to upsA (Tan et al., 2023). Based on each isolate's non-upsA DBL α repertoire size, MOI_{var} was estimated using a cut-off value of 45 non-upsA DBL α types per *P. falciparum* genome. This cut-off was selected based on the median number of non-upsA DBL α types identified for the 3D7 laboratory isolate included as a control during *var*coding (Figure S2). Using this cut-off, MOI_{var} bins were defined as follows: 1 to 45 non-upsA DBL α types were estimated to be monoclonal infections (i.e., $MOI_{var} = 1$), isolates with 46 to 90 non-upsA DBL α types were estimated to be carrying two *P. falciparum* clones (i.e., $MOI_{var} = 2$, multiclonal infections), isolates with 91 to 135 non-upsA DBL α types were estimated to be carrying three *P. falciparum* clones (i.e., $MOI_{var} = 3$, multiclonal infections), and so on.

MOI_{var} could only be estimated for those isolates with DBL α sequence data, resulting in 742 (91.8%) and 510 (93.6%) isolates in the pre- and post-IRS surveys, respectively (Figure S1 and Table S1).

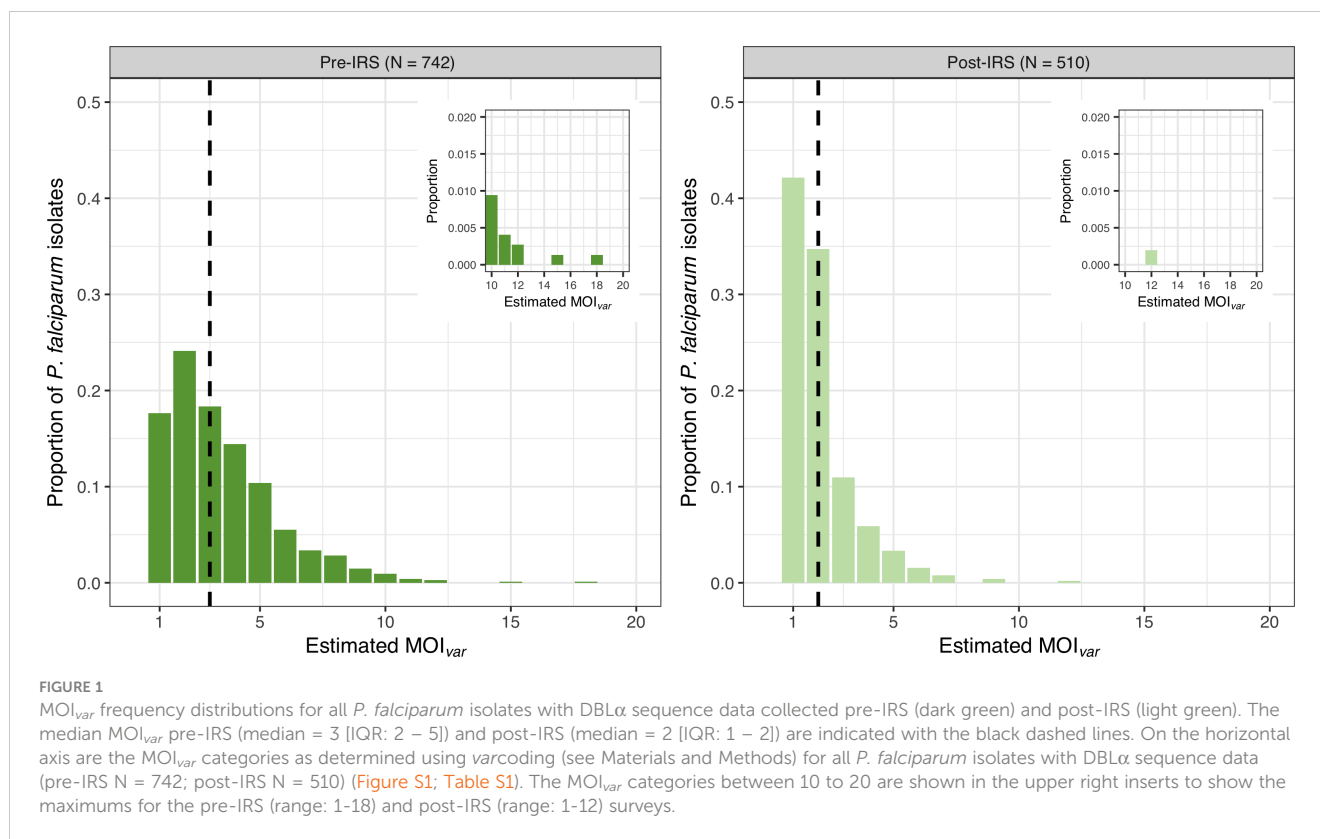
For these isolates with DBL α sequencing data, the median asymptomatic *P. falciparum* densities in the pre-IRS (560 parasites/ μ L) and post-IRS (360 parasites/ μ L) surveys were of similar magnitudes and ~5-9 times higher compared to those isolates with no DBL α sequencing data pre-IRS (160 parasites/ μ L) and post-IRS (40 parasites/ μ L). Consistent with other high-transmission areas in SSA, we found that the majority of *P. falciparum* isolates were composed of multiclonal infections, i.e., 82.3% and 57.8% for the pre- and post-IRS surveys, respectively (Figure 1). Whilst the median MOI_{var} values were 3 [IQR: 2 - 5] pre-IRS and 2 [IQR: 1 - 2] post-IRS, the MOI_{var} frequency distributions show many individuals had multiclonal infections greater than three (39.9% and 12.1% pre- and post-IRS, respectively). At the extreme of the range, MOI_{var} detected a maximum of 18 and 12 *P. falciparum* coinfections per isolate for the pre- and post-IRS, respectively (Figure 1).

2.4 Isolate filtering

In order to compare the performance of the three genotyping methods to describe diversity and population structure, a subset of 215 and 200 *P. falciparum* isolates with the lowest complexity were selected pre- and post-IRS, respectively. These isolates had previously been chosen for microsatellite genotyping using *var*coding to estimate MOI (Argyropoulos et al., 2021). They were then used for SNP genotyping. To improve the likelihood of obtaining SNP and microsatellite genotyping data, isolates with monoclonal infections ($MOI_{var} = 1$) and low quality DBL α sequencing data (22% and 49% pre- and post-IRS, respectively) were not included in the subset of low-complexity isolates selected pre-IRS (N = 215, median $MOI_{var} = 2$ [IQR: 1 - 2]) and post-IRS (N = 200, median $MOI_{var} = 1$ [IQR: 1 - 2]) (Figure S1; Table S2). Note that these isolates selected pre- and post-IRS were not statistically different than those isolates in the original study population for any of the key variables, except age pre-IRS (p -value < 0.001, Chi-square test) and parasitemia post-IRS (p -value < 0.01, Mann Whitney U test) (Tiedje et al., 2022).

2.5 Measures of genetic diversity

To estimate genetic diversity, the number of unique multilocus haplotypes (h), the number of alleles (A), and expected heterozygosity (H_e) were calculated for the SNP and microsatellite markers using the R package *poppr* 2.9.3 (Kamvar et al., 2014; Kamvar et al., 2015). For the DBL α types, diversity was measured using richness, defined as the number of unique DBL α types observed (i.e., DBL α type pool size). In addition, we also assessed *var* (or DBL α type) expected heterozygosity (H_v) (Rorick et al., 2018). If each isolate had a repertoire of exactly one DBL α type, pairwise type sharing (PTS) (see section 2.6 below for more information) would be roughly equivalent to *var* expected homozygosity ($1 - H_v$), and thus by calculating the pairwise type



difference (PTD = 1 - PTS) statistic we can obtain *var* expected heterozygosity (H_v).

2.6 Measures of genetic similarity

Genetic similarity among the isolates in the pre- and post-IRS surveys, i.e., the number of shared loci (SNPs, microsatellites, and DBL α types), was assessed by comparing every isolate to every other isolate. To undertake the comparisons for the SNPs and microsatellites, we used the pairwise allele sharing (P_{AS}) statistic, using only isolates with the “monoclonal infections” and complete multilocus infection haplotypes (i.e., no missing genotyping data) (Ruybal-Pesántez et al., 2017a; Argyropoulos et al., 2021). These complete multilocus haplotypes (i.e., phased isolates) for the SNPs and microsatellites were necessary to ensure that the denominator in the P_{AS} calculations would be consistent for the SNPs (i.e., 20 loci) and microsatellites (i.e., ten loci). The P_{AS} scores were calculated using $P_{AS} = n_{ab}/n_b$ where n_{ab} is the number of alleles shared between the isolate haplotypes and n_l is the number of loci examined (i.e., 20 SNPs and ten microsatellites). To measure similarity for the DBL α types we used the PTS statistic, calculated as $PTS = 2n_{ab}/(n_a + n_b)$, where n_a and n_b are the number of unique DBL α types in the repertoires of isolate *a* and isolate *b*, and n_{ab} are the number of DBL α types shared between isolate *a* and isolate *b* (Barry et al., 2007). The advantages of P_{AS} and PTS is that they are convenient statistics that can be quickly and easily calculated to

evaluate the number of SNP/microsatellite alleles or DBL α types shared between two different isolates. Both the P_{AS} and PTS scores were calculated between all isolate pairs in each survey (pre-IRS and post-IRS) and represent the proportion of sharing (or relatedness) between two isolates with scores ranging from 0 (i.e., dissimilar or unrelated) to 1 (i.e., identical or clones). Both the P_{AS} and PTS statistics are measures of identity-by-state (IBS) used to assess similarity or relatedness between isolates and were not used to infer inheritance from a recent common ancestor (i.e., identity-by-descent (IBD)) (Speed and Balding, 2015).

2.7 Statistical analysis

All statistical analyses were carried out in R 4.0.5 (Core Team, 2018) implemented in RStudio 1.4.1106 (RStudio, 2020) using the R package *tidyverse* 1.3.1 (Wickham et al., 2019) for data curation, analysis, and visualization.

3 Results

3.1 Genotyping of asymptomatic *Plasmodium falciparum* infections

Selection of isolates from asymptomatic individuals for genotyping was biased towards low-complexity infections as

outlined in the Materials and Methods. The amount of usable data for each marker from these low-density asymptomatic infections was assessed for each marker as follows.

3.1.1 SNPs

From the 200 isolates selected pre- and post-IRS, SNP barcoding data was successfully obtained from 161 (80.5%) and 200 (100%) isolates in the pre- and post-IRS surveys, respectively. Using these successfully genotyped isolates, the SNP calls were then aggregated and all SNP loci with a call rate of at least 80% in both surveys were included (Daniels et al., 2008; Daniels et al., 2013; Daniels et al., 2015). Based on this call rate, four loci (i.e., A4, A10, B10, and B12) were removed, resulting in 20-SNP loci being used for the molecular analyses (Table S3). Finally, only those isolates with ambiguous or missing calls at four or fewer of the 20-SNP loci ($\leq 20\%$) were included in downstream analyses (Daniels et al., 2008; Daniels et al., 2013; Daniels et al., 2015), resulting in 157 (78.5%) and 200 (100%) isolates in the pre- and post-IRS surveys, respectively (Figures 2, 3). These isolates were defined as the “cleaned infections” dataset (Figure 2). All isolates included in the “cleaned infections” dataset were then evaluated to determine if they were monoclonal or multiclonal infections following the 24-SNP barcode data exclusion criteria of Daniels et al (Daniels et al., 2008; Daniels et al., 2013; Daniels et al., 2015). All isolates with more than one of the 20-SNP loci ($> 5\%$) showing a double-allele call (DAC) were considered as multiclonal infections (i.e., $MOI > 1$) (Figure S3) and were excluded from the population genetic analyses (Figure 3, Tables S4, S5). We used this threshold to allow for the fact

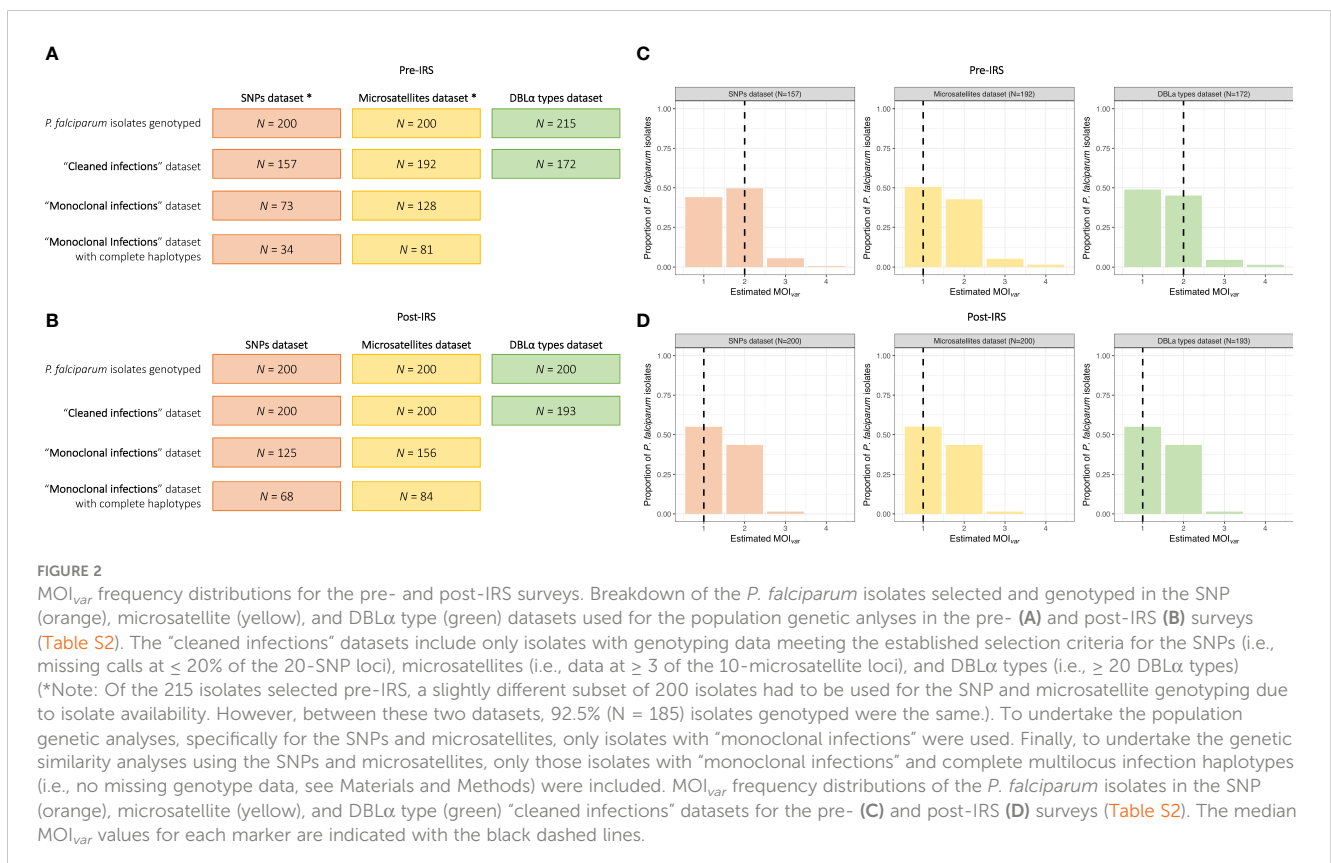
that one SNP may be miscalled as a DAC (i.e., low-level genotyping error) even in a monoclonal infection (Daniels et al., 2008; Nkhoma et al., 2013; Daniels et al., 2015). Based on this cut-off, only 73 (46.5%) and 125 (62.5%) isolates in the pre- and post-IRS surveys, respectively, were classified as monoclonal infections and included in the “monoclonal infections” dataset for analysis (Figure 2).

3.1.2 Microsatellites

All details for the microsatellite genotyping and data cleaning for the 200 isolates selected pre- and post-IRS have been previously published (Argyropoulos et al., 2021). Briefly, using ten microsatellite loci, 192 (96.0%) and 200 (100%) isolates with genotyping data at ≥ 3 microsatellite loci in the pre- and post-IRS surveys, respectively, were included in the “cleaned infections” dataset (Figure 2). All isolates with (i) “true” monoclonal infections (i.e., one allele at all ten loci) or (ii) a maximum of two alleles at any locus (i.e., $MOI = 2$) where a dominant multilocus haplotype could be constructed (or phased) (Anderson et al., 2000; Ruybal-Pesántez et al., 2017a; Argyropoulos et al., 2021), were included in the “monoclonal infections” dataset, since phasing is possible for up to two microsatellite haplotypes. This resulted in 128 (66.7%) and 156 (78.0%) isolates from the pre- and post-IRS surveys, respectively (Figure 2).

3.1.3 DBL α types

High-quality sequencing data (i.e., ≥ 20 DBL α types) was obtained from 172 (80.0%) and 193 (96.5%) isolates in the pre- and post-IRS surveys, respectively (i.e., “cleaned infections” dataset)



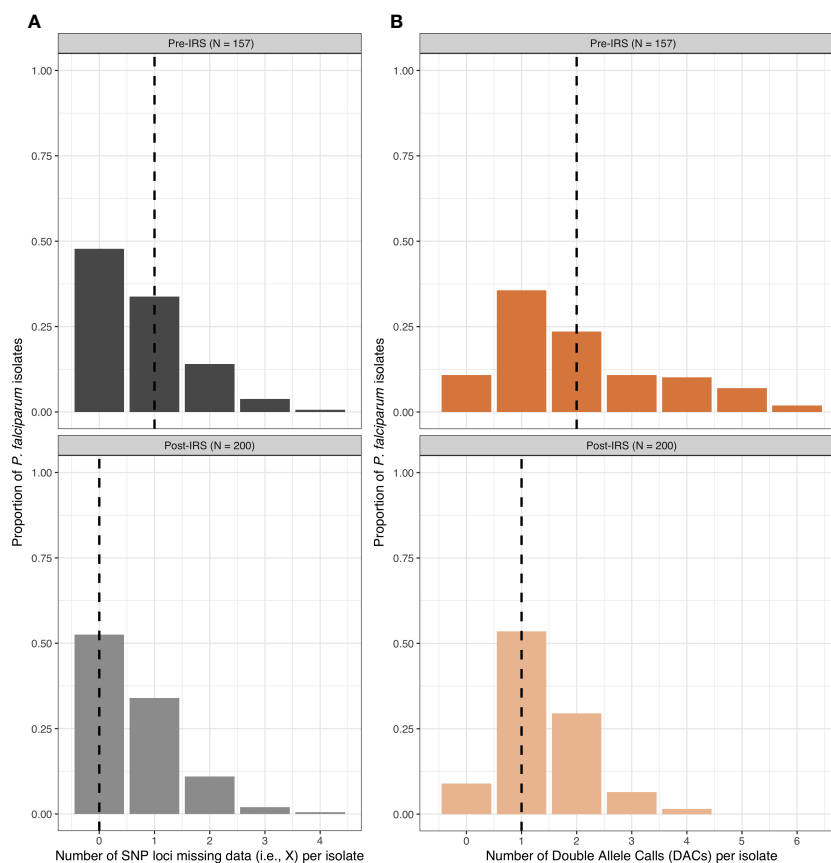


FIGURE 3

The frequency distributions for the number of SNP loci missing data (grey) (A) and the number of double-allele calls (DACs) (orange) (B) for the *P. falciparum* isolates included in the “cleaned infections” datasets pre- and post-IRS. The black dashed lines in each plot indicate the median number of SNP loci missing data per isolate (pre-IRS median = 1 [IQR: 0 – 1]; post-IRS median = 0 [IQR: 0 – 1]) and the median number DACs per isolate (pre-IRS median = 2 [IQR: 1 – 3]; post-IRS (median = 1 [IQR: 1 – 2]) are indicated with the black dashed lines (Tables S4, S5). For those isolates included in the pre- (N = 157) and post-IRS (N = 200) “cleaned infections” datasets for the SNPs, see Data Sheets 1 and 2, respectively.

(Figure 2; Table S1). Using *varcoding* we determined the number of unique DBL α types in each isolate. Since the underlying population structure of the *var* multigene family in Bongo pre-IRS was characterized by non-overlapping DBL α isolate repertoires (Ruybal-Pesántez et al., 2022), we were able to use this limited similarity of repertoires in an isolate in high transmission to estimate clonality of infections (i.e., mono- or multiclonal). Therefore, in comparison to the SNPs and microsatellites, the *varcoding* approach allowed for the inclusion of all isolates successfully genotyped for the data analyses, regardless of their MOI as phasing (i.e., haplotype construction) was unnecessary. As a result, all isolates successfully genotyped pre-IRS (N = 172) and post-IRS (N = 193) were included (Figure 2).

3.2 Genetic diversity

Even in isolates selected for low complexity, we found that 19 and 20 of the SNP loci were biallelic in the pre- and post-IRS surveys, respectively (Tables S4, S5), with the mean number of alleles per locus (A) being 1.9 and 2.0 in the pre- and post-IRS

surveys, respectively (Table 1). All the microsatellite loci genotyped were polymorphic pre-IRS (from 5 to 20 alleles per locus) and post-IRS (from 5 to 22 alleles per locus) with the mean number of alleles per locus (A) being similar in both surveys, despite the IRS intervention (Table 1) (Argyropoulos et al., 2021). Using the DBL α types, we found that diversity, as measured using richness, was considerable, with 7,736 and 7,251 unique DBL α types being identified pre- and post-IRS, respectively (Table 1).

For the pre-IRS survey, we found the number of haplotypes (*h*) matched the number of isolates (N) at each marker, except for the SNPs where two isolates shared the same infection haplotype (Table 1). We found that all haplotypes in the post-IRS survey were unique for the microsatellites (except for two participants that shared the same haplotype) (Argyropoulos et al., 2021) and the DBL α types. Conversely, using the biallelic SNPs only 106 unique multilocus haplotypes were observed from the 125 isolates, indicating that there were repeated 20-SNP barcodes in the population, thereby underestimating the genome diversity compared to the microsatellites and DBL α types. Finally, H_e (H_v for the DBL α types) remained relatively stable pre- to post-IRS across all three genetic markers (Table 1). However, H_e based on the

TABLE 1 Patterns of *P. falciparum* genetic diversity using the “monoclonal infections” datasets for the SNPs, microsatellites, and DBL α types pre- and post-IRS (Figure 2).

Genetic markers	Pre-IRS (October 2012)				Post-IRS (October 2015)			
	N	<i>h</i>	<i>A</i>	<i>H_e</i>	N	<i>h</i>	<i>A</i>	<i>H_e</i>
SNPs	73	71	1.9	0.25	125	106	2.0	0.24
Microsatellites	128	128	11.5	0.79	156	155	12.4	0.81
DBL α types	172	172	7,736 *	0.98 **	193	193	7,251 *	0.98**

N = number of isolates; *h* = number of unique haplotypes; *A* = mean number of alleles per locus (SNPs, microsatellites) (*) Note for DBL α types this number reflects the total number of unique DBL α types (i.e., richness); *H_e* = expected heterozygosity (**) Note for the DBL α types the expected heterozygosity was measured using *var* expected heterozygosity (*H_e*) as described in the Materials and Methods.

SNPs was lower than those based on the microsatellites and DBL α types for both surveys (i.e., pre-IRS and post-IRS). Thus, by using less diverse markers that require the selection of monoclonal infections (or those with the lowest complexity) we were underestimating parasite diversity in this high-transmission setting in SSA.

3.3 Genetic similarity

To undertake the comparisons for the SNPs and microsatellites using the P_{AS} statistic, we only used isolates with “monoclonal infections” and complete multilocus infection haplotypes (i.e., no missing genotype data) resulting in 34 (21.7%) and 68 (34.0%) isolates for the pre- and post-IRS SNPs, and 81 (42.2%) and 84 (42.0%) isolates for the pre- and post-IRS microsatellites (Figure 2). The advantage of using DBL α types and the PTS statistic to assess genetic similarity, compared to the SNPs and microsatellites, is that we can compare all isolates genotyped irrespective of their MOI_{var} (i.e., repertoire size) or missing data because phasing is not necessary. Thus, all of the isolates successfully *var*coded during the pre- (N = 172) and post-IRS (N = 193) surveys were included for this analysis (Figure 2).

Using the SNP P_{AS} comparisons, we observed that the majority of isolate infection haplotypes in both the pre- and post-IRS surveys were highly similar as they shared ≥ 0.75 alleles in their SNP barcodes (i.e., identical ≥ 15 out of the 20 loci genotyped) (Figures 4A, B; Table S6). Although the median P_{AS} score was the same in both surveys, the P_{AS} distributions were significantly different with a shift towards higher P_{AS} values post-IRS compared to pre-IRS (p -value < 0.001, Mann Whitney U test) (Figures 4A, B; Table S6). This result based on SNPs implies that isolate haplotypes in the population became more similar (i.e., increased sharing of alleles) following the IRS intervention. Although these high P_{AS} scores indicate that these isolates are highly similar and could be related, this must be interpreted with caution since P_{AS} is sensitive to the local minor allele frequencies (MAF) of the SNP panel and thus higher P_{AS} scores may be observed when a greater number of low-heterozygosity loci are included (Speed and Balding, 2015; Taylor et al., 2019). Since ~45–50% of the loci had MAF ≤ 0.1 (10%) (i.e., one highly dominant allele at these loci) (Tables S4, S5), this 20-SNP barcode is expected

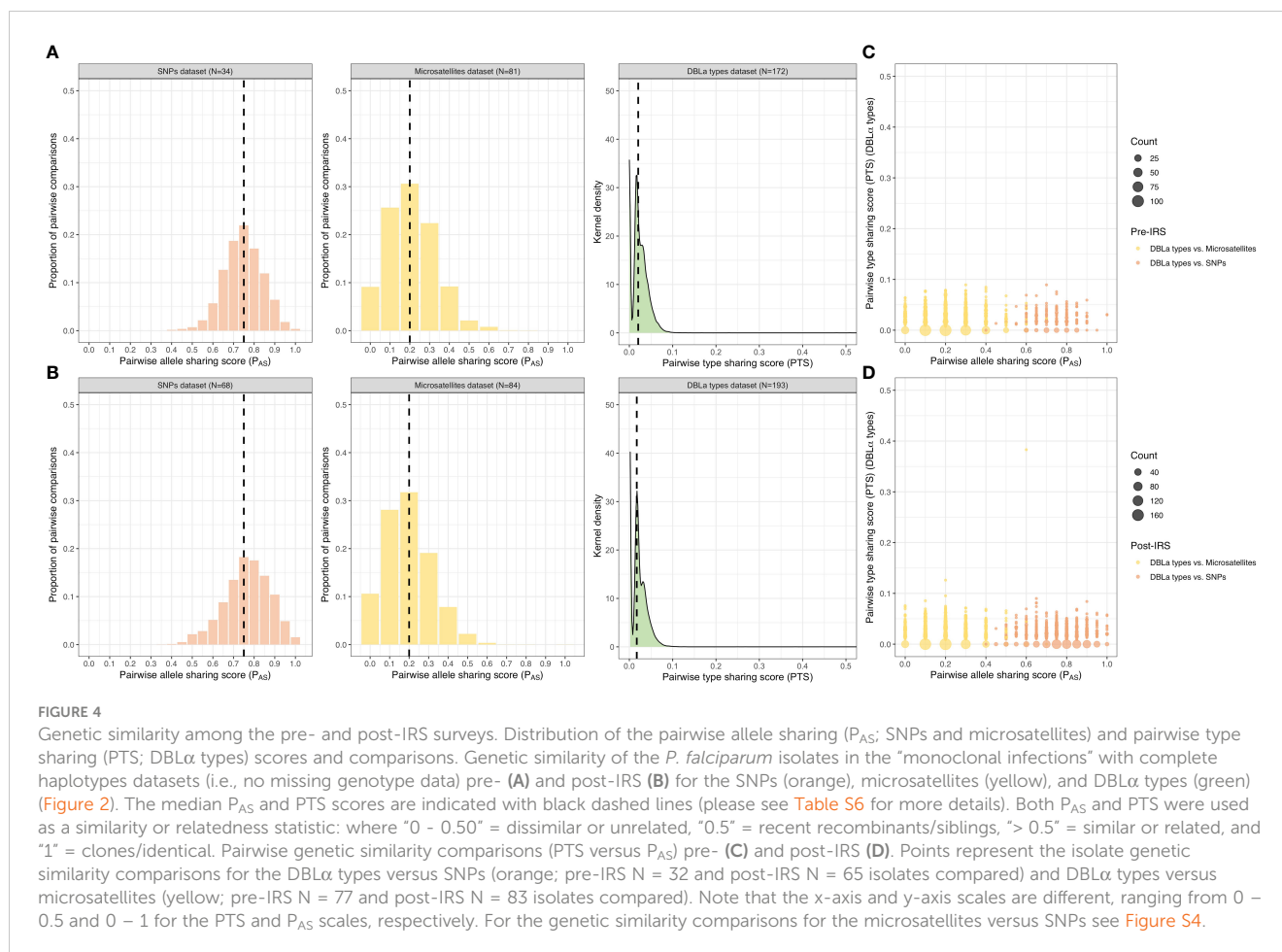
to be less informative for assessing genetic similarity in this population compared to the more polymorphic molecular markers such as the microsatellites and DBL α types (Figures 4C, D, S4).

When we evaluated similarity using the multiallelic microsatellites pre- and post-IRS data, the majority of isolate infection haplotypes were found to be dissimilar or unrelated as they only shared ≤ 0.2 of their alleles (i.e., identical at two or fewer loci out of ten) (Figures 4A, B; Table S6), making them more informative than SNPs to determine genetic similarity in this population. P_{AS} distributions using microsatellites showed that isolates were significantly less similar (i.e., lower P_{AS} values) post-IRS than pre-IRS (p -value < 0.001, Mann Whitney U test) despite the median P_{AS} scores being the same (Figures 4A, B; Table S6). Finally using the DBL α type data, we found that 99.9% of the PTS comparisons were ≤ 0.1 (i.e., shared $\leq 10\%$ of their DBL α types), indicating that DBL α isolate repertoires in this population were highly dissimilar and composed of diverse DBL α types both pre- and post-IRS (Figures 4A, B; Table S5). Using this analysis, we found that the PTS distributions were significantly different (p -value < 0.001, Mann Whitney U test) and that there was a shift towards a lower median PTS value post-IRS (i.e., less similar) (Figures 4A, B; Table S6).

Since DBL α types were the most diverse marker, they provided higher resolution to distinguish between similar and dissimilar infection haplotypes compared to the biallelic SNPs and multiallelic microsatellites (Figures 4C, D, S4). In fact, when we compared the same set of isolates with both P_{AS} (SNPs or microsatellites) and PTS scores, we observed that isolates that were identical or highly similar using their SNP barcodes, were found to be dissimilar (or unrelated) when assessed using their DBL α isolate repertoires (Figures 4C, D).

3.4 Varcoding analyses for all infections

Given that the *var*coding approach allows for the analysis of all isolates regardless of clonality, we were able to further analyze the population structure of all isolates successfully *var*coded in the larger initial dataset pre- (84.8%, N = 685) and post-IRS (75.8%, N = 413) (Table S1) without the need to subsample for low-complexity infections. Using this larger dataset, we found that the parasite



reservoir in Bongo was still composed of highly dissimilar DBL α isolate repertoires (median PTS [IQR]: pre-IRS = 0.033 [0.021-0.046] and post-IRS = 0.023 [0.013 - 0.035]) that became less similar following the IRS intervention (p -value < 0.001, Mann Whitney U test). Thus, the reduced similarity reported using the sample of low-complexity infections in this analysis was confirmed with observations based on the larger dataset.

As we were able to varcode all isolates in the larger dataset, additional analyses of population structure by age (i.e., children: 1–10 years, adolescents: 11–20 years, and adults: > 20 years) were possible. Although multiclonal infections were observed across all age groups in Bongo, children (1–10 years) and adolescents (11–20 years) carried the largest proportion of these multiclonal infections pre- and post-IRS, as previously published (Tiedje et al., 2022). Using these age stratifications, we found that not only was there limited overlap of DBL α isolate repertoires in each age group both pre- and post-IRS (Figure 5; Table S7) but that repertoire similarity was significantly higher in the children compared with the adolescents and adults pre-IRS (for further discussion see Ruybal-Pesántez et al. (2022)). This age-specific pattern was maintained post-IRS (p -value < 0.001 for all comparisons, Mann Whitney U test). Finally, by examining the PTS distributions, we observed that DBL α isolate repertoires were significantly less similar in each age group pre- to post-IRS (p -value < 0.001 for all comparisons, Mann

Whitney U test) (Figure 5; Table S7). Again, this confirmed the detection of the subtle but significant decrease in similarity following the IRS intervention. This result shows that we can get a snapshot of changes in population structure due to the intervention by varcoding with samples taken from any of these three age classes without the need to limit analyses to only monoclonal infections.

4 Discussion

Our study highlights the challenge of doing *P. falciparum* population genetics in the highest burden countries of SSA where ~550 million people live at risk of infection (World Health Organization & Roll Back Malaria Partnership to End Malaria, 2019; World Health Organization, 2021). We have addressed several interconnected issues related to defining appropriate methods for molecular surveillance of *P. falciparum* using low-density asymptomatic infections to capture changes in population genetics at local or regional levels as a result of vector control interventions. Namely, the importance that marker choice has in relation to prevalence of multiclonal infections to resolve diversity and population structure estimates in relatively small sample sizes. When assessing the impact of an IRS intervention by three different

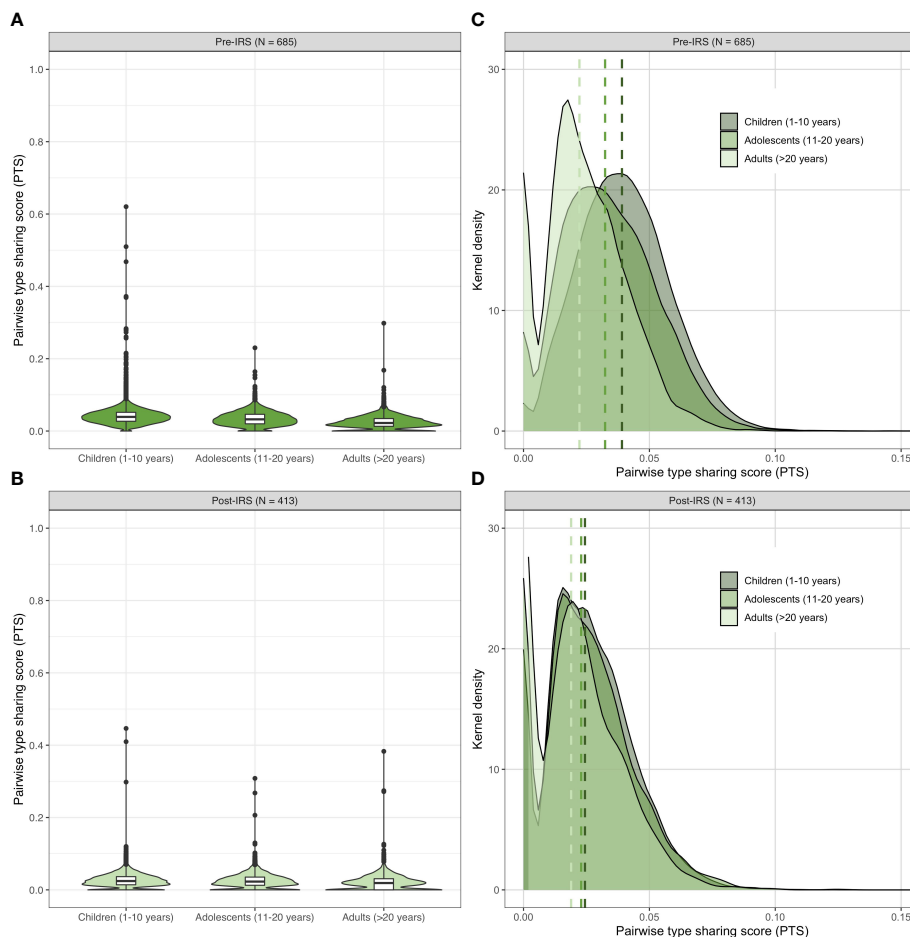


FIGURE 5

Age-specific patterns of genetic similarity in the pre- and post-IRS surveys. Violin plots showing the pairwise type sharing (PTS) distributions of all *P. falciparum* isolates with DBL α sequence data (Table S1) in each age group pre- (A) and post-IRS (B). The box plots for each age group show the median and interquartile ranges and the black dots denote outliers. Kernel density plots showing the lower end of the PTS distributions for each age group pre- (C) and post-IRS (D). The PTS scales in the density plots have been zoomed-in to provide better visualization of the DBL α types PTS distributions. The dashed lines indicate the median PTS values for the children (1-10 years, dark green), adolescents (11-20 years, medium green), and adults (> 20 years, light green) (please see Table S7 for more details).

markers, the resolution of population genetics estimates increased with marker polymorphism. The extent of multiclonal infections was particularly an issue related to data availability for biallelic SNP and microsatellite analyses due to phasing issues. In contrast, *var* coding worked relatively independent of MOI in high transmission (Day et al., 2017; Ruybal-Pesántez et al., 2017b; Ruybal-Pesántez et al., 2022) and provided informative data with relatively small sample sizes. Microsatellite genotyping provided similar population structure data, but required the selection of low-complexity infections from the larger initial sample.

An important result of our study is the failure to increase *var* repertoire or microsatellite haplotype similarity (or relatedness) by the IRS intervention. We start our intervention with very high *var* repertoire diversity in the asymptomatic parasite population. We show IRS reduced transmission intensity by > 90% and so we expect to have greatly reduced outcrossing by both reducing the population of biting mosquitos and the lifespan of blood fed

mosquitos thereby minimizing exposure of humans to new infections. As a consequence, we are largely looking at the decay of the parasite population in the human host (asymptomatic reservoir) with few new transmission events. As this parasite population at baseline pre-IRS has low *var* similarity, it maintains this feature or becomes less similar as the number of *var* repertoires to be compared declines post-IRS. Microsatellite genotyping shows the same pattern of low similarity for genome diversity per se, pre- and post-IRS. The extent of diversity post-IRS may also be contributed to via migration of diverse parasites from uncontrolled areas as our study site shares an immediate border with Burkina Faso. Some evidence for such migration came from our earlier microsatellite work where we observed significant spatiotemporal differentiation in Bongo. In this analysis we showed that not only were the catchment areas (i.e., Veá/Gowrie and Soe) genetically different post-IRS, but that the parasite population in Soe (proximal to Burkina Faso) was genetically

differentiated pre- to post-IRS as measured using Jost's D and G_{ST} (Argyropoulos et al., 2021).

A study from Thiès, Senegal a peri-urban region, showed the opposite result where they saw a shift towards more similarity and clonality following interventions targeting clinical infections (i.e., rapid diagnostic tests and artemisinin-based combination therapies) and transmission intensity (i.e., LLINs and IRS) simultaneously (Daniels et al., 2015). So, what is different about the Thiès, Senegal and Bongo, Ghana studies with opposite outcomes? We point to the overall diversity of the parasite populations in low vs high transmission. Firstly Thiès, prior to intervention scale-up was a low-transmission area as defined by the WHO (i.e., prevalence $\leq 10\%$ and an annual incidence of 100-250 cases/1000) (Mouzin et al., 2010; Daniels et al., 2015; WHO/GMP, 2017). Whereas Bongo remained classified as a high-transmission area pre- and post-IRS (see Materials and Methods) maintaining a very large reservoir of asymptomatic infections with a similar incidence of clinical cases to Thiès in 2006 (Ghana Health Service; Tiedje et al., 2017; Tiedje et al., 2022). Secondly in Thiès, they sampled and analyzed clinical cases of malaria and saw a $> 95\%$ reduction between 2006 and 2009. While in Bongo we undertook longitudinal sampling of the asymptomatic reservoir across all ages in a ~ 60 km² area and saw a 37.5% decline in *P. falciparum* prevalence (Tiedje et al., 2022). Thus, the size of the parasite populations both pre- and post-intervention differs substantially in these studies. From population genetic theory, small amounts of outcrossing will have a greater impact in increasing parasite similarity (or relatedness) in a small parasite population as seen in low transmission.

By measuring the genetic diversity of loci, haplotypes, and P_{AS}/PTS for three established molecular assays, we get different pictures of both the extent of genetic diversity and similarity of the *P. falciparum* reservoir in Bongo pre- and post-IRS. Given the same initial sample size of ~ 200 isolates with low-complexity infections, the variable resolution of the methods relates to the relative polymorphism of the markers (i.e., SNPs $<$ microsatellites $<$ DBL α types). Biallelic SNPs showed a less diverse parasite reservoir both pre- and post-IRS, underestimating the greater diversity of genomes seen with microsatellites and *var*coding. The reduced diversity and higher genetic similarity observed using the SNP barcode was largely due to the biallelic nature of SNPs and the necessary removal of multiclonal infections, reducing sample size for analysis. Multiallelic microsatellites showed a parasite reservoir that was diverse and genetically dissimilar during both the pre- and post-IRS surveys. When considering measuring changes in neutral variation, it is clear that the more polymorphic microsatellite markers have greater resolution than the 20-SNP barcode. Genotyping a larger panel of SNP or microsatellite markers (e.g., ≥ 200 biallelic or 100 polymorphic loci to achieve low error rates) to undertake relatedness estimates using IBD has been recommended (Taylor et al., 2019; Han et al., 2022), but the same problems will occur as the number of markers is not the issue in high-transmission areas. Instead, it is the relative polymorphism of the marker.

DBL α types showed that the parasite isolates were highly diverse and composed of dissimilar *var* repertoires pre- and post-IRS. In fact, the lack of the need for phasing with DBL α types provided more usable data for *var*coding than the SNP or microsatellite markers. Most significantly, *var*coding picked up a subtle change in *var* repertoire similarity of the parasite population post-IRS. As the IRS intervention reduced transmission intensity, it makes sense that isolate *var* repertoire similarity would reduce as a consequence of less outcrossing in the mosquito, limiting the possibility to create more similar or related recombinant genomes. *Var*coding was able to detect this change in any age group and any infection complexity in sample sizes of up to 200 isolates. This observed reduction in similarity post-IRS was further confirmed by the analysis of all data from *var*coding of microscopy-positive infections sampled from the larger cohort of $\sim 2,000$ participants (see Figure 5 and Table S7) as all of the high MOI_{*var*} data were usable.

Cost estimates of molecular genotyping assays for surveillance usually focus on the price of reagents per sample, but the sample size required to get informative population genetic data with an assay also contributes significantly to dollars spent. This can vary tenfold as shown in this study where pre-screening $\sim 2,000$ participants of all ages had to be performed to successfully identify less than 200 monoclonal infections per survey suitable for SNP barcoding and microsatellite genotyping. Isolates collected from residents with low-density asymptomatic infections were genotyped in this study, resulting in isolates being excluded due to low-quality genotyping data or missing data in the "cleaned infections" datasets for all three markers analyzed. While a pre-amplification step with selective whole-genome amplification (sWGA) is required by other panels to improve performance when amplifying DNA from low-density infections from dried blood spots (e.g., 10 parasites/ μ l of blood), it adds a considerable cost to the genotyping assays (Oyola et al., 2016; Jacob et al., 2021; LaVerriere et al., 2022). Although sWGA could increase the number of isolates available to assess diversity and genetic relatedness for all markers, it does not resolve the issue of multiclonal infections; this is what actually limits the use of SNPs in high transmission due to their biallelic nature and the high proportion of DACs, leading to reduced numbers of usable multilocus haplotypes (Figure S3).

More recently with the development of newer genetic panels composed of a larger number of biallelic SNPs (Jacob et al., 2021; LaVerriere et al., 2022) or multiallelic microhaplotypes (Tesemma et al., 2020) (≥ 2 SNPs within a DNA segment unbroken by recombination (Baetscher et al., 2018)), statistical packages (e.g., DEploid (Zhu et al., 2018), DEploidIBD (Zhu et al., 2019), and Dcifer (Gerlovina et al., 2022)) have been developed using identity-by-descent (IBD) based methods to estimate relatedness for phased or unphased multiclonal infections. Although promising, they have been developed on limited datasets from high transmission showing greater error for isolates with greater than five clones and are yet to be tested in the field. Such methods do not consider the extent of outcrossing in natural populations, especially in high transmission

(Babiker et al., 1994; Paul and Day, 1998), where haplotypes are not stable in epidemiologic time.

Besides the markers compared in this study, additional genetic panels have been developed for molecular surveillance of malaria parasites in the field. These newer panels, including SpotMalaria v2 (Jacob et al., 2021), Paragon v1 (Tessema et al., 2020), and AMPLseq v1 (LaVerriere et al., 2022), have been designed for multiplexed PCR amplicon sequencing and incorporate single copy antigenic loci under selection, known antimalarial drug resistance markers, biallelic SNP loci, and/or microhaplotypes. *In silico* validation of these panels from countries with low or high parasite diversity have shown they are more accurate to assess genetic relatedness compared to the 24-SNP barcode. Nonetheless, simulated monoclonal infections were needed for this analysis using AMPLseq v1 (LaVerriere et al., 2022). Until there is sufficient local population genetic data to train computational approaches to accurately phase in the MOI ranges typical of high-transmission settings, even these newer panels with deeper coverage at a larger number of loci (i.e., > 100 SNPs and/or microhaplotypes) are not sufficient to overcome the issue of phasing of multiclonal infections.

If the primary goal for these molecular surveillance methods is to be used in countries to directly inform National Malaria Control/Elimination Programmes, cost-effective and scalable platforms will be necessary. Although microsatellites have been informative in a variety of malaria transmission settings, they have unfortunately proven to be technically difficult to standardize across laboratories due to issues with allele calling and errors in the assessment of clonality (Figan et al., 2018). New platforms are emerging to use these markers which should reduce cost and be easier to use (Han et al., 2022). The advantages of *var*coding and the newer panels (i.e., SpotMalaria v2, Paragon v1, and AMPLseq v1), using amplicon sequencing, is that large number of isolates can be multiplexed into a single-pool, thus overall costs are mainly driven by the number of PCRs/clean-up, the number of samples indexed per sequencing run, and the next-generation sequencing (NGS) technology used (\$20 to \$40 USD per isolate) (Tessema et al., 2020; LaVerriere et al., 2022). However, several key features set *var*coding apart from these other methods. First, *var*coding can amplify DNA collected as dried blood spots and stored for more than 5-years from both clinical or low-density asymptomatic infections (Day et al., 2017; Ruybal-Pesántez et al., 2017b; Ruybal-Pesántez et al., 2021; Ruybal-Pesántez et al., 2022), without the added cost and additional step of sWGA (\$8 to \$32 USD per isolate, for costing estimates see (Tessema et al., 2020; LaVerriere et al., 2022)). Second, since all *P. falciparum* genomes possess ~40 to 60 *var* genes (Bull et al., 1998; Gardner et al., 2002; Rask et al., 2010; Otto et al., 2019) and there is extensive repertoire diversity (Day et al., 2017; Ruybal-Pesántez et al., 2017b; He et al., 2018; Ruybal-Pesántez et al., 2022), *var*coding can be used for molecular surveillance both locally and globally without the need to be customized or updated (Chen et al., 2011; Day et al., 2017; Rougeron et al., 2017; Ruybal-Pesántez et al., 2021; Tonkin-Hill et al., 2021; Ruybal-Pesántez et al., 2022). Thus, with a simple PCR using degenerate primers, *var*coding can be easily deployed in malaria endemic regions (Tonkin-Hill et al., 2021).

In conclusion, we have exploited the fact that *var* genes and *var* repertoires diversify by recombination to create *var*coding. Here we show that this method can be used in high-transmission settings to measure diversity and population structure even in multiclonal infections. This is achieved more easily than SNP barcoding and microsatellite genotyping as there is no need for pre-selection of isolates and phasing. Measuring PTS or IBS complements other previously published uses of the method to measure MOI (Ruybal-Pesántez et al., 2022; Tiedje et al., 2022; Labbé et al., 2023) as well as identify geographic signatures at the country level within Africa to assess importation of parasite *var*codes (Ruybal-Pesántez et al., 2017b; Tonkin-Hill et al., 2021; Ruybal-Pesántez et al., 2022). These three measurements can be made robustly in relatively small sample sizes. This triple output of the method was not seen with either SNP barcoding or microsatellite genotyping.

Data availability statement

The SNP and microsatellite datasets used for this analysis are available in Dryad at <https://doi.org/10.5061/dryad.jsxksn0bp> and <https://doi.org/10.5061/dryad.kh189324z>, respectively. The DBL α sequences utilized in this study are publicly available in GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>) under BioProject Number: PRJNA 396962. All custom code is available in an open source repository: (i) DBL α Cleaner pipeline is available at <https://github.com/UniMelb-Day-Lab/DBLaCleaner>, (ii) clusterDBLalpha pipeline is available at <https://github.com/Unimelb-Day-Lab/clusterDBLalpha>, and the (iii) classifyDBLalpha pipeline is available at <https://github.com/Unimelb-Day-Lab/classifyDBLalpha>. A tutorial and dataset to demo this custom code is available at <https://github.com/UniMelb-Day-Lab/tutorialDBLalpha>.

Ethics statement

The studies involving human participants were reviewed and approved by Navrongo Health Research Centre, Noguchi Memorial Institute for Medical Research, The University of Chicago, and The University of Melbourne. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

Author contributions

KPD, KAK, MP conceived and designed the study. AG, KET, DCA, COO, SLD processed the samples and performed the genotyping experiments. AG, KET, DCA processed, cleaned, and curated the datasets for analysis. AG, KET, DCA, FL analyzed the data. KET, DCA visualized the data. KPD, KET, DCA wrote the original draft of the manuscript. AG, COO, SLD, FL, ARO, KAK, MP reviewed and edited manuscript. KPD, ARO, KAK supervised the research. ARO, KAK, MP, KPD acquired the funding. All authors contributed to the article and approved the submitted version.

Funding

Funding was provided by the 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.

Acknowledgments

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 thank Dr. Mun Hua Tan for helpful comments with the manuscript.

References

- Anderson, T. J. C., Haubold, B., Williams, J. T., Estrada-Franco, J. G., Richardson, L., Mollinedo, R., et al. (2000). Microsatellite markers reveal a spectrum of population structures in the malaria parasite *Plasmodium falciparum*. *Mol. Biol. Evol.* 17 (10), 1467–1482. doi: 10.1093/oxfordjournals.molbev.a026247
- Anthony, T. G., Conway, D. J., Cox-Singh, J., Matusop, A., Ratnam, S., Shamsul, S., et al. (2005). Fragmented population structure of *Plasmodium falciparum* in a region of declining endemicity. *J. Infect. Dis.* 191 (9), 1558–1564. doi: 10.1086/429338
- Argyropoulos, D. C., Ruybal-Pesántez, S., Deed, S. L., Oduro, A. R., Dadzie, S. K., Appawu, M. A., et al. (2021). The impact of indoor residual spraying on *Plasmodium falciparum* microsatellite variation in an area of high seasonal malaria transmission in Ghana, West Africa. *Mol. Ecol.* 30 (16), 3974–3992. doi: 10.1111/mec.16029
- Auburn, S., and Barry, A. E. (2017). Dissecting malaria biology and epidemiology using population genetics and genomics. *Int. J. Parasitol.* 47 (2–3), 77–85. doi: 10.1016/j.ijpara.2016.08.006
- Babiker, H. A., Ranford-Cartwright, L. C., Currie, D., Charlwood, J. D., Billingsley, P., Teuscher, T., et al. (1994). Random mating in a natural population of the malaria parasite *Plasmodium falciparum*. *Parasitology* 109 (4), 413–421. doi: 10.1017/S0031182000080665
- Baetscher, D. S., Clemento, A. J., Ng, T. C., Anderson, E. C., and Garza, J. C. (2018). Microhaplotypes provide increased power from short-read DNA sequences for relationship inference. *Mol. Ecol. Resour.* 18 (2), 296–305. doi: 10.1111/1755-0998.12737
- Barry, A. E., Leliwa-Sytek, A., Tavul, L., Imrie, H., Migot-Nabias, F., Brown, S. M., et al. (2007). Population genomics of the immune evasion (*var*) genes of *Plasmodium falciparum*. *PLoS Pathog.* 3 (3), e34. doi: 10.1371/journal.ppat.0030034
- Barry, A. E., Schultz, L., Senn, N., Nale, J., Kiniboro, B., Siba, P. M., et al. (2013). High levels of genetic diversity of *Plasmodium falciparum* populations in Papua New Guinea despite variable infection prevalence. *Am. J. Trop. Med. Hyg.* 88 (4), 718–725. doi: 10.4269/ajtmh.12-0056
- Bei, A. K., Niang, M., Deme, A. B., Daniels, R. F., Sarr, F. D., Sokhna, C., et al. (2018). Dramatic changes in malaria population genetic complexity in dielmo and ndiopi, Senegal, revealed using genomic surveillance. *J. Infect. Dis.* , 1–6. doi: 10.1093/infdis/jix580
- Bull, P. C., Lowe, B. S., Kortok, M., Molyneux, C. S., Newbold, C. I., and Marsh, K. (1998). Parasite antigens on the infected red cell surface are targets for naturally acquired immunity to malaria. *Nat. Med.* 4 (3), 358–360. doi: 10.1038/nm0398-358
- Chen, D. S., Barry, A. E., Leliwa-Sytek, A., Smith, T.-A. A., Peterson, I., Brown, S. M., et al. (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

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpara.2023.1067966/full#supplementary-material>

Core Team, R. (2018). “R: A language and environment for statistical computing.” in *R foundation for statistical computing* (Vienna, Austria: R Foundation for Statistical Computing).

Daniels, R. F., Chang, H.-H. H., Séne, P. D., Park, D. C., Neafsey, D. E., Schaffner, S. F., et al. (2013). Genetic surveillance detects both clonal and epidemic transmission of malaria following enhanced intervention in Senegal. *PLoS One* 8 (4), 4–10. doi: 10.1371/journal.pone.0060780

Daniels, R. F., Schaffner, S. F., Wenger, E. A., Proctor, J. L., Chang, H.-H., Wong, W., et al. (2015). Modeling malaria genomics reveals transmission decline and rebound in Senegal. *PNAS* 112 (22), 7067–7072. doi: 10.1073/pnas.1505691112

Daniels, R., Volkman, S. K., Milner, D. A., Mahesh, N., Neafsey, D. E., Park, D. J., et al. (2008). A general SNP-based molecular barcode for *Plasmodium falciparum* identification and tracking. *Malar J.* 7, 223. doi: 10.1186/1475-2875-7-223

Day, K. P., Artzy-Randrup, Y., Tiedje, K. E., Rougeron, V., Chen, D. S., Rask, T. S., et al. (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

Figan, C. E., Sá, J. M., Mu, J., Melendez-Muniz, V. A., Liu, C. H., and Wellems, T. E. (2018). A set of microsatellite markers to differentiate *Plasmodium falciparum* progeny of four genetic crosses. *Malar J.* 17 (1), 1–6. doi: 10.1186/s12936-018-2210-z

Gardner, M. J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R. W., et al. (2002). Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature* 419 (6906), 498–511. doi: 10.1038/nature01097

Gerlovina, I., Gerlovin, B., Rodríguez-Barraquer, I., and Greenhouse, B. (2022). Dcifer: An IBD-based method to calculate genetic distance between polyclonal infections. *Genetics* 222 (2), iyac126. doi: 10.1093/genetics/iyac126

Ghana Health Service. *District health information management system (DHIMS2)*, Ghana.

Han, J., Munro, J. E., Kocoski, A., Barry, A. E., and Bahlo, M. (2022). Population-level genome-wide STR discovery and validation for population structure and genetic diversity assessment of *Plasmodium* species. *PLoS Genet.* 18 (1), e1009604. doi: 10.1371/journal.pgen.1009604

He, Q., Pilosof, S., Tiedje, K. E., Ruybal-Pesántez, S., Artzy-Randrup, Y., Baskerville, E. B., et al. (2018). Networks of genetic similarity reveal non-neutral processes shape strain structure in *Plasmodium falciparum*. *Nat. Commun.* 9 (1), 1817. doi: 10.1038/s41467-018-04219-3

Jacob, C. G., Thuy-nhien, N., Mayxay, M., Maude, R. J., Quang, H. H., Hongvanthong, B., et al. (2021). Genetic surveillance in the greater Mekong

- subregion and south Asia to support malaria control and elimination. *Elife* 10 (e62997), 1–22. doi: 10.7554/eLife.62997.sa2
- Kamvar, Z. N., Brooks, J. C., and Grünwald, N. J. (2015). Novel r tools for analysis of genome-wide population genetic data with emphasis on clonality. *Front. Genet.* 6 (JUN), 1–10. doi: 10.3389/fgene.2015.00208
- Kamvar, Z. N., Tabima, J. F., and Grünwald, N. J. (2014). Poppr: An r package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ* 2, e281. doi: 10.7717/peerj.281
- Kattenberg, J. H., Razook, Z., Keo, R., Koepfli, C., Jennison, C., Lautu-Gumal, D., et al. (2020). Monitoring plasmodium falciparum and plasmodium vivax using microsatellite markers indicates limited changes in population structure after substantial transmission decline in Papua new Guinea. *Mol. Ecol.* 29, 1–17. doi: 10.22541/au.159015516.68850426
- Labbé, F., He, Q., Zhan, Q., Tiedje, K. E., Argyropoulos, D. C., Tan, M. H., et al. (2023). Neutral vs. non-neutral genetic footprints of plasmodium falciparum multiclinal infections. *PLoS Comput. Biol.* 19 (1), e1010816. doi: 10.1371/journal.pcbi.1010816
- LaVerriere, E., Schwabl, P., Carrasquilla, M., Taylor, A. R., Johnson, Z. M., Shieh, M., et al. (2022). Design and implementation of multiplexed amplicon sequencing panels to serve genomic epidemiology of infectious disease: A malaria case study. *Mol. Ecol. Resour.* 22, 2285–2303. doi: 10.1111/1755-0998.13622
- Machado, R. L. D., Póvoa, M. M., Calvosa, V. S. P., Ferreira, M. U., Rossit, A. R. B., Dos Santos, E. J. M., et al. (2004). Genetic structure of plasmodium falciparum populations in the Brazilian Amazon region. *J. Infect. Dis.* 190 (9), 1547–1555. doi: 10.1086/424601
- MalariaGEN Consortium (2021). An open dataset of plasmodium falciparum genome variation in 7,000 worldwide samples. *Wellcome Open Res.* 6, 42. doi: 10.12688/wellcomeopenres.16168.2
- Mobegi, V. A., Loua, K. M., Ahouidi, A. D., Satoguina, J., Nwakanma, D. C., Amambua-Ngwa, A., et al. (2012). Population genetic structure of plasmodium falciparum across a region of diverse endemicity in West Africa. *Malar J.* 11, 203. doi: 10.1186/1475-2875-11-223
- Mouzin, E., Thior, P. M., Diouf, M. B., and Sambou, B. (2010). *Focus on Senegal roll back malaria: Progress and impact series* Vol. 4 (Geneva: World Heal Organ), 1–54.
- Nkhoma, S. C., Nair, S., Al-Saai, S., Ashley, E., McGready, R., Phyo, A. P., et al. (2013). Population genetic correlates of declining transmission in a human pathogen. *Mol. Ecol.* 22 (2), 273–285. doi: 10.1111/mec.12099
- Noviyanti, R., Coutrier, F., Utami, R. A. S., Trimarsanto, H., Tirta, Y. K., Trianty, L., et al. (2015). Contrasting transmission dynamics of Co-endemic plasmodium vivax and p. falciparum: Implications for malaria control and elimination. *PLoS Negl. Trop. Dis.* 9 (5), 1–19. doi: 10.1371/journal.pntd.0003739
- Otto, T. D., Assefa, S. A., Böhme, U., Sanders, M. J., Kwiatkowski, D. P., Berriman, M., et al. (2019). Evolutionary analysis of the most polymorphic gene family in falciparum malaria. *Wellcome Open Res.* 4, 1–29. doi: 10.12688/wellcomeopenres.15590.1
- Oyola, S. O., Ariani, C. V., Hamilton, W. L., Kekre, M., Amenga-Etego, L. N., Ghansah, A., et al. (2016). Whole genome sequencing of plasmodium falciparum from dried blood spots using selective whole genome amplification. *Malar J.* 15 (1), 1–12. doi: 10.1186/s12936-016-1641-7
- Paul, R. E., and Day, K. P. (1998). Mating patterns of plasmodium falciparum. *Parasitol. Today* 14 (5), 197–202. doi: 10.1016/S0169-4758(98)01226-5
- Paul, R. E., Packer, M. J., Walmsley, M., Lagog, M., Ranford-Cartwright, L. C., Paru, R., et al. (1995). Mating patterns in malaria parasite populations of Papua new Guinea. *Science* 269 (September), 1709–1711. doi: 10.1126/science.7569897
- Pumpaibool, T., Arnathau, C., Durand, P., Kanchanachan, N., Siripoon, N., Suegorn, A., et al. (2009). Genetic diversity and population structure of plasmodium falciparum in Thailand, a low transmission country. *Malar J.* 8, 155. doi: 10.1186/1475-2875-8-155
- Rask, T. S., Hansen, D., Theander, T. G., Pedersen, A. G., Lavstsen, T., Gorm Pedersen, A., et al. (2010). Plasmodium falciparum erythrocyte membrane protein 1 diversity in seven genomes - divide and conquer. *PLoS Comput. Biol.* 6 (9), e1000933. doi: 10.1371/journal.pcbi.1000933
- Rorick, M. M., Artzy-Randrup, Y., Ruybal-Pesántez, S., Tiedje, K. E., Rask, T. S., Oduro, A., et al. (2018). Signatures of competition and strain structure within the major blood-stage antigen of p. falciparum in a local community in Ghana. *Ecol. Evol.* 8 (7), 3574–3588. doi: 10.1002/ece3.3803
- Rougeron, V., Tiedje, K. E., Chen, D. S., Rask, T. S., Gamboa, D., Maestre, A., et al. (2017). Evolutionary structure of plasmodium falciparum major variant surface antigen genes in south America: Implications for epidemic transmission and surveillance. *Ecol. Evol.* 7, 9376–9390. doi: 10.1002/ece3.3425
- RStudio (2020). “RStudio: Integrated development for R”. RStudio, PBC, Boston, MA. Available at: <http://www.rstudio.org/>.
- Ruybal-Pesántez, S., Sáenz, F. E., Deed, S. L., Johnson, E. K., Larremore, D. B., Vera-Arias, C. A., et al. (2023). Molecular epidemiology of continued Plasmodium falciparum disease transmission after an outbreak in Ecuador. *Fron. Trop. Dis.* 4, 1085862. doi: 10.3389/ftd.2023.1085862
- Ruybal-Pesántez, S., Tiedje, K. E., Pilosof, S., Tonkin-Hill, G., He, Q., Rask, T. S., et al. (2022). Age-specific patterns of DBL α var diversity can explain why residents of high malaria transmission areas remain susceptible to plasmodium falciparum blood stage infection throughout life. *Int. J. Parasitol.* 20, 721–731. doi: 10.1016/j.ijpara.2021.12.001
- Ruybal-Pesántez, S., Tiedje, K. E., Rorick, M. M., Amenga-Etego, L., Ghansah, A., R Oduro, A., et al. (2017a). Lack of geospatial population structure yet significant linkage disequilibrium in the reservoir of plasmodium falciparum in bongo district, Ghana. *Am. J. Trop. Med. Hyg.* 97 (4), 1180–1189. doi: 10.4269/ajtmh.17-0119
- Ruybal-Pesántez, S., Tiedje, K. E., Tonkin-Hill, G., Rask, T. S., Kamya, M. R., Greenhouse, B., et al. (2017b). Population genomics of virulence genes of plasmodium falciparum in clinical isolates from Uganda. *Sci. Rep.* 7 (1), 11810. doi: 10.1038/s41598-017-11814-9
- Speed, D., and Balding, D. J. (2015). Relatedness in the post-genomic era: Is it still useful? *Nat. Rev. Genet.* 16 (1), 33–44. doi: 10.1038/nrg3821
- Tan, M. H., Chan, H., Chan, Y., and Day, K. P. (2023). Unravelling chaos for malaria surveillance: Relationship between DBL α types and var genes in plasmodium falciparum. *Front. Parasitol.* 1. doi: 10.3389/fpara.2022.1006341
- Taylor, A. R., Jacob, P. E., Neafsey, D. E., and Buckee, C. O. (2019). Estimating relatedness between malaria parasites. *Genetics* 212 (4), 1337–1351. doi: 10.1534/genetics.119.302120
- Tessem, S. K., Hathaway, N. J., Teyssier, N. B., Murphy, M., Chen, A., Aydemir, O., et al. (2020). Sensitive, highly multiplexed sequencing of microhaplotypes from the plasmodium falciparum heterozygote. *J. Infect. Dis.* 225 (7), 1227–1237. doi: 10.1101/2020.02.25.964536
- Tiedje, K. E., Oduro, A. R., Agongo, G., Anyorigiya, T., Azongo, D., Awine, T., et al. (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., et al. (2022). Indoor residual spraying with a non-pyrethroid insecticide reduces the reservoir of plasmodium falciparum in a high-transmission area in northern Ghana. *PLoS Glob Public Heal.* 2 (25), e0000285. doi: 10.1371/journal.pgph.0000285
- Tonkin-Hill, G., Ruybal-Pesántez, S., Tiedje, K. E., Rougeron, V., Duffy, M. F., Zakeri, S., et al. (2021). Evolutionary analyses of the major variant surface antigen-encoding genes reveal population structure of plasmodium falciparum within and between continents. *PLoS Genet.* 7 (2), e1009269. doi: 10.1371/journal.pgen.1009269
- Vera-Arias, C. A., Castro, L. E., Gómez-Obando, J., and Sáenz, F. E. (2019). Diverse origin of plasmodium falciparum in northwest Ecuador. *Malar J.* 18 (1), 1–11. doi: 10.1186/s12936-019-2891-y
- Verity, R., Aydemir, O., Brazeau, N. F., Watson, O. J., Hathaway, N. J., Mwandagaliwa, M. K., et al. (2020). The impact of antimalarial resistance on the genetic structure of plasmodium falciparum in the DRC. *Nat. Commun.* 11 (1), 2107. doi: 10.1038/s41467-020-15779-8
- WHO/GMP (2017). *A framework for malaria elimination* (Geneva: World Health Organization), 100. Available at: <http://apps.who.int/iris/bitstream/handle/10665/254761/9789241511988-eng.pdf?sequence=1>.
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., et al. (2019). Welcome to the tidyverse. *J. Open Source Software* 4 (43), 1686. doi: 10.21105/joss.01686
- World Health Organization (2021). *World malaria report 2021* (World Health Organization).
- World Health Organization (2022). *World malaria report 2022* (World Health Organization).
- World Health Organization & Roll Back Malaria Partnership to End Malaria. (2019). *High burden to high impact: A targeted malaria response* (Geneva: World Health Organization).
- Yalcindag, E., Elguero, E., Arnathau, C., Durand, P., Akiana, J., Anderson, T. J., et al. (2012). Multiple independent introductions of plasmodium falciparum in south America. *PNAS* 109 (2), 511–516. doi: 10.1073/pnas.1119058109
- Zhu, S. J., Almagro-Garcia, J., and McVean, G. (2018). Deconvolution of multiple infections in plasmodium falciparum from high throughput sequencing data. *Bioinformatics* 34 (1), 9–15. doi: 10.1093/bioinformatics/btx530
- Zhu, S. J., Hendry, J., Almagro-Garcia, J., Pearson, R., Amato, R., Miles, A., et al. (2019). The origins and relatedness structure of mixed infections vary with local prevalence of p. falciparum malaria. *Elife* 8, e40845. doi: 10.7554/eLife.40845