HHS Public Access

Author manuscript

Int J Parasitol. Author manuscript; available in PMC 2022 December 06.

Published in final edited form as:

Int J Parasitol. 2022 October ; 52(11): 721–731. doi:10.1016/j.ijpara.2021.12.001.

Age-specific patterns of DBLα *var* diversity can explain why residents of high malaria transmission areas remain susceptible to *Plasmodium falciparum* blood stage infection throughout life

Shazia Ruybal-Pesántez[a,1,2,3], **Kathryn E. Tiedje**[a,b], **Shai Pilosof**[c,d], **Gerry Tonkin-Hill**[a,e], **Qixin He**[c], **Thomas S. Rask**[a], **Lucas Amenga-Etego**[f,g], **Abraham R. Oduro**[g], **Kwadwo A. Koram**[h], **Mercedes Pascual**[c], **Karen P. Day**[a,b,*]

[a]School of BioSciences, Bio21 Institute, The University of Melbourne, Australia

[b]Department of Microbiology and Immunology, Bio21 Institute and Peter Doherty Institute, The University of Melbourne, Australia

[c]Department of Ecology and Evolution, University of Chicago, USA

[d]Department of Life Sciences, Ben-Gurion University, Be'er-Sheva, Israel

[e]Bioinformatics Division, Walter and Eliza Hall Institute of Medial Research, Australia

[f]West African Centre for Cell Biology and Infectious Pathogens, Department of Biochemistry, Cell and Molecular Biology, University of Ghana, Ghana

[g]Navrongo Health Research Centre, Ghana Health Service, Ghana

[h]Epidemiology Department, Noguchi Memorial Institute for Medical Research, University of Ghana, Ghana

Abstract

Immunity to *Plasmodium falciparum* is non-sterilising, thus individuals residing in malaria-endemic areas are at risk of infection throughout their lifetime. Here we seek to find a genomic epidemiological explanation for why residents of all ages harbour blood stage infections despite lifelong exposure to *P. falciparum* in areas of high transmission. We do this by exploring, for the first known time, the age-specific patterns of diversity of variant antigen encoding *(var)* genes in the reservoir of infection. Microscopic and submicroscopic *P. falciparum* infections were analysed at the end of the wet and dry seasons in 2012–2013 for a cohort of 1541 residents aged from 1 to 91 years in an area characterised by high seasonal malaria transmission in Ghana. By sequencing the near ubiquitous Duffy-binding-like alpha domain (DBLα) that encodes immunogenic domains, we defined *var* gene diversity in an estimated 1096 genomes detected in sequential wet and dry season sampling of this cohort. Unprecedented *var* (DBLα) diversity

*Corresponding author. karen.day@unimelb.edu.au (K.P. Day).
[1]Current address: Population Health and Immunity Division, Walter and Eliza Hall Institute of Medical Research, Australia.
[2]Burnet Institute, Australia.
[3]Bio21 Institute/Department of Medical Biology, University of Melbourne, Australia.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijpara.2021.12.001.

was observed in all ages with 42,399 unique *var* types detected. There was a high degree of maintenance of types between seasons (>40% seen more than once), with many of the same types, especially upsA, appearing multiple times in isolates from different individuals. Children and adolescents were found to be significant reservoirs of *var* DBLα diversity compared with adults. *Var* repertoires within individuals were highly variable, with children having more related *var* repertoires compared to adolescents and adults. Individuals of all ages harboured multiple genomes with *var* repertoires unrelated to those infecting other hosts. High turnover of parasites with diverse isolate *var* repertoires was also observed in all ages. These age-specific patterns are best explained by variant-specific immune selection. The observed level of *var* diversity for the population was then used to simulate the development of variant-specific immunity to the diverse *var* types under conservative assumptions. Simulations showed that the extent of observed *var* diversity with limited repertoire relatedness was sufficient to explain why adolescents and adults in this community remain susceptible to blood stage infection, even with multiple genomes.

Keywords

Plasmodium falciparum ; Genomic epidemiology; Parasite diversity; Malaria; Variant-specific immunity; Asymptomatic infection; *var* genes; PfEMP1

1. Introduction

Unlike measles where a single infection leads to sterilising immunity, individuals of all ages continue to be infected with *Plasmodium falciparum.* This is particularly apparent in high-transmission settings in Africa where individuals develop immunity that protects against clinical disease during childhood but remain susceptible to sporozoite and blood stage infections throughout life (Marsh, 1992; Owusu-Agyei et al., 2001; Tran et al., 2013). Antigenic diversity has been the accepted explanation for why immunity does not protect against infection by the blood stages of *P. falciparum.* A contemporary understanding of *P. falciparum* genomics has identified both diverse single copy antigen encoding genes as well as multigene families that contribute to immune evasion of these life cycle stages.

The major surface antigen of the blood stages, known as *P. falciparum* Erythrocyte Membrane Protein 1 (PfEMP1), is encoded by the variant antigen encoding *(var)* multigene family and is the major target of naturally acquired immunity (Chan et al., 2012). Clonal antigenic variation involving differential expression of up to 60 *var* genes per genome enables chronic infection and facilitates transmission to mosquitoes (Biggs et al., 1991). Variation in PfEMP1 is considered a key driver of transmission dynamics within and between hosts due to immune evasion (Gupta and Day, 1994). Expression of different groups of *var* genes identified by chromosomal locations has also been associated with virulence in case-control studies and transcriptomic analyses, with a conserved subset of PfEMP1 associated with severe disease (Lavstsen et al., 2012; Bengtsson et al., 2013; Rorick et al., 2013; Lau et al., 2015; Bernabeu et al., 2016; Jespersen et al., 2016; Magallón-Tejada et al., 2016; Lennartz et al., 2017; Tonkin-Hill et al., 2018). Our previous work has explored *var* gene diversity in infected children by focusing on the near ubiquitous Duffy-binding-like alpha domain (DBLα) region of *var* genes as a population genetic marker (Chen et al., 2011;

Day et al., 2017; Ruybal-Pesántez et al., 2017). The DBLα domain encodes the N-terminal surface exposed antigenic region of PfEMP1, which is immunogenic (Barry et al., 2011; Tessema et al., 2019), and acquired immunity to this domain is age-dependent (Barry et al., 2011). This work showed that *var* genes are highly diverse in children, but to date we have no understanding of *var* diversity and population structure in adults compared with children.

Previously, we defined a non-random population structure of non-overlapping *var* DBLα repertoires in children in several African sites characterised by high transmission and sexual recombination (Chen et al., 2011; Day et al., 2017; Ruybal-Pesántez et al., 2017). This result, revealing an absence of recombinant repertoires (i.e., limited relatedness) in a sexually reproducing organism such as *P. falciparum,* was striking. Network analyses and stochastic simulations that consider both epidemiological and evolutionary processes confirmed that frequency-dependent variant-specific immune selection rather than exposure-dependent generalised immunity can structure parasite *var* DBLα repertoires to be non-overlapping to maximise immune evasion potential (He et al., 2018). Given variant-specific immunity, rather than generalised immunity to single copy antigen genes, is shown to be the dominant force structuring the parasite population in the blood stages, this leads us to propose that there must be an extraordinary number of PfEMP1 variants circulating locally to sustain infection in all ages and that age-specific signatures of *var* diversity and population structure could be detected.

Here we test these two hypotheses by describing *var* DBLα diversity and structure after deep sampling of *P. falciparum* infections in an age-stratified cohort of 1541 individuals (aged 1–91 years) experiencing intense, seasonal malaria transmission in Ghana, ranked the seventh high-burden country globally by the World Health Organization (WHO, 2019). Sequential wet and dry season patterns of *var* DBLα diversity in both microscopy-positive, as well as submicroscopic infections, were analysed by age to measure the size of the *var* reservoir of diversity in infections from children, adolescents and adults to look for signatures of selection by variant-specific immunity. Distinct age-specific patterns of diversity revealed children and adolescents having parasites with the most *var* diversity compared with those of adults. Significant differences in *var* DBLα population structure were also detected in parasites from children compared with those found in adolescents and adults. The total measure of the reservoir of *var* DBLα diversity in all ages was then used in a simulation to model the lifetime acquisition of variant-specific immunity to all observed variants. These simulations, based on conservative assumptions, showed that both the size and structure of the reservoir of *var* DBLα diversity is sufficient to explain why residents of high transmission areas remain susceptible to blood stage infection, even after a lifetime of exposure to *P. falciparum.*

2. Materials and methods

2.1. Human subjects and ethical approval

This study was reviewed and approved by the ethics committees at the Navrongo Health Research Centre (Navrongo, Ghana), Noguchi Memorial Institute for Medical Research (Legon, Ghana), New York University (New York, USA), The University of Melbourne (Melbourne, Australia), and University of Chicago (Chicago, USA). Study details and

procedures were explained to participants in their local language and witnessed informed consent (and assent where applicable) was obtained from each individual. For those under the age of 18 years, witnessed informed consent was obtained from a parent or guardian, with all children between the ages of 12–17 years also providing assent. Individuals who did not meet the inclusion criteria were excluded from the study; this included if they were pregnant, had a disability, or presented with a serious or acute disease (including symptomatic malaria as defined by a temperature ≥37.5 °C and a positive rapid diagnostic test). Individuals requiring treatment for malaria were referred to their respective local health centre (Vea Health Centre, Soe Health Centre or Bongo District Hospital) for appropriate care.

2.2. Study design

This study was designed to evaluate the seasonal and age-specific diversity patterns in the *P. falciparum* reservoir in Bongo District, located in the Upper East Region of Ghana where malaria transmission is high and seasonal. The age-stratified serial cross-sectional surveys were conducted over consecutive wet and dry seasons with each survey lasting approximately 4 weeks. The first survey was conducted at the end of the wet season (EWS) during October 2012, after the highest point of malaria transmission during the year. The second survey was conducted at the end of the dry season (EDS) during mid-May to June 2013, with an 83% retention rate of the same participants enrolled during October 2012. A detailed description of the study area and study design has been published elsewhere (Tiedje et al., 2017). Briefly, a total of 1900 and 1868 healthy participants between the ages of 1–92 years were surveyed at the EWS and EDS, respectively. Blood samples (thick/thin blood films, rapid diagnostic tests, and dried blood spots) as well as demographic and malaria-related questionnaires were collected after obtaining informed consent. Here we focus on the 83% of participants who were retained in the second survey, totaling 1541 individuals surveyed at both time points.

2.3. Parasitological measurements

Parasitological measurements were previously published in Tiedje et al. (2017). Briefly, parasite densities were counted per 200 white blood cells (WBCs) on 10% Giemsa-stained thick film blood smears and examined under oil immersion of 100-fold magnification. Parasite densities were calculated by averaging two independent readings completed by two experienced technicians and recorded as parasites per μL of blood, assuming the average WBC count was 8000 per μL of blood. Parasite species were also identified using a 100-fold magnification of the thin film smears and categorized based on morphology.

2.4. DNA extraction

Peripheral blood from survey participants was blotted onto 3MM Whatman filter paper and dried at room temperature. Two 5 mm × 5 mm sections were cut from each dried blood spot (DBS) and placed in a 1.5 mL centrifuge tube. Genomic DNA was extracted from the DBS for all participants surveyed as per the manufacturer's instructions using a Qiagen DNA Mini kit (QIAGEN, California, USA), with one modification in the final step where isolates were eluted in 50 μL of AE buffer instead of 150 μL. The extracted genomic DNA from each isolate was stored at −20 °C prior to PCR analysis, and at −80 °C for long-term storage.

2.5. 18S rRNA genotyping

A species-specific nested PCR targeting the 18S rRNA gene was performed previously by our group on all participant samples that were negative for *P. falciparum* by microscopy in order to detect the presence of submicroscopic *P. falciparum* infections (Tiedje et al., 2017). The PCR methodology is described in detail in Tiedje et al. (2017). Note that all participants who were positive for *P. falciparum* (microscopic or submicroscopic), were afebrile (temperature <37.5 °C) on the day the survey was conducted and did not report a history of fever in the 24 h prior to being surveyed, thus were defined as having an asymptomatic *P. falciparum* infection.

2.6. Targeted amplicon PCR and var DBLα sequencing

A targeted amplicon Illumina sequencing approach was used to sequence the ubiquitous DBLα domain of the var genes for all *P. falciparum* isolates (i.e., microscopic and submicroscopic). The DBLα domains of *P. falciparum var* genes were amplified from genomic DNA, as previously described (Ruybal-Pesántez et al., 2017) and the individually-tagged amplicons were pooled equimolarly. Barcoded libraries were prepared using the KAPA Low-Throughput Library Preparation Kit Standard (Kapa Biosystems, Woburn, MA, USA) and amplified using the KAPA HiFi Library Amplification kit (Kapa Biosystems, Woburn, MA, USA) (eight cycles). An equimolar pool of the barcoded libraries was sequenced on an Illumina MiSeq sequencer using the 2×300 paired-end cycle protocol, MiSeq Reagent kit v3 chemistry (New York University Genome Technology Center, New York, NY, USA; Australian Genome Research Facility, Melbourne, Australia). Sequence data was successfully obtained from 664 isolates at the EWS (out of 1,151 infected residents) and 435 isolates (out of 669 infected residents) at the EDS. As expected, a significant failure rate was observed in both seasons for low-density submicroscopic infections but not microscopic infections (Chi-square test, $P < 0.001$).

2.7. Illumina DBLα sequence data processing

A full tutorial of each data processing step is available on GitHub at https://github.com/UniMelb-Day-Lab/tutorialDBLalpha. See below details of each step.

2.7.1. Illumina DBLα sequence cleaning pipeline—A pipeline tailored to the analysis of DBLα sequence tags was developed to de-multiplex and remove PCR and sequencing artefacts from the DBLα reads. Briefly, flexbar v2.5 (Dodt et al., 2012) was used to demultiplex each pooled paired-end fastq file into individual files for each isolate based on both the forward and reverse MID tags. PEAR v.0.9.10 (Zhang et al., 2014) was then used to merge the overlapping paired-end reads. The minimum assembly length required was set to 100 nucleotides (nt) and at least 10 overlapping bases were required. The resulting merged reads were then filtered for quality, and those reads with more than one expected error were removed using the *fastq_filter* command (Edgar, 2010). Chimeras were filtered using *Uchime denovo* (Edgar et al., 2011) and then the remaining reads were clustered using the *cluster_fast* function of Usearch (Edgar, 2010) after first removing singletons to reduce the impact of errors. The standard pairwise sequencing identity threshold of 96% was used for clustering (Day et al., 2017). To remove clusters with low support, those that

contained less than 15 reads were discarded. The centroid from each cluster was kept as a representative of each cluster for the remaining stages of the pipeline. Finally, non-DBLα sequences were removed using Hmmer (Rask et al., 2016) with a domain score threshold of 80. The code for the sequence cleaning pipeline is available on GitHub at https://github.com/UniMelb-Day-Lab/DBLaCleaner.

2.7.2. Illumina DBLα sequence clustering pipeline—To match DBLα types between isolates, the cleaned DBLα reads were clustered using a pipeline optimised for the analysis of DBLα sequence tags (Ruybal-Pesántez et al., 2017). In short, reads were sorted based upon their frequency in the dataset and duplicates removed. The reads were then clustered at 96% pairwise sequence identity using the Usearch *cluster_fast* command. The original unfiltered reads were aligned back to the centroids of the clusters and an operational taxonomic unit (OTU) table was generated using the *usearch_global* command. The code for the pipeline is available on GitHub at https://github.com/UniMelb-Day-Lab/clusterDBLalpha.

2.7.3. Translation and classification of DBLα sequences to upsA/non-upsA var gene groups—DBLα types were assigned a DBLα domain class using a hidden Markov model (HMM) to classify each DBLα type into either upsA or non-upsA groups, as described in Ruybal-Pesántez et al. (2017). The reads were first translated into all six reading frames and protein HMMs of 150 *var* domains (Rask et al., 2010) were aligned to the translated sequences using HMMER v3.1b1 with an e-value cutoff of 1e-8. A read was then classified as upsA if their most significant match was to a DBLα1 domain and as non-upsA if it matched otherwise (i.e., either DBLα0 or DBLα2 domains). Any DBLα types that were “non-translatable” or were unable to be matched to a DBLα domain were excluded. The code for the pipeline is available on GitHub at https://github.com/UniMelb-Day-Lab/classifyDBLalpha.

2.8. Pairwise type sharing statistics

To quantify the relatedness (or overlap) between the DBLα repertoires identified from two isolates, pairwise type sharing (PTS) (Barry et al., 2007; He et al., 2018) statistics were utilised. This statistic is analogous to the Sørensen Index (Chao et al., 2005) and is a useful similarity index to determine the number of DBLα types shared between two isolates. Briefly, a PTS score is the proportion of shared DBLα types between two isolates and ranges between 0 and 1, where a PTS score of 0 signifies no shared DBLα types, a low PTS score indicates little overlap and a high PTS score indicates a high degree of overlap between the two isolates.

2.9. Accumulation curves and simulation model

We implemented a computational experiment to simulate the time it would take to acquire immunity to 95% of (i) upsA DBLα types, (ii) non-upsA DBLα types and (iii) all DBLα types circulating in a human population based on an annual entomological inoculation rate (EIR) in Bongo of 25 (Tiedje et al., 2017). The accumulation curves were only run up to 95% (instead of 100%) due to computational limitations, as reaching 100% requires extremely intensive and long computing times. We assumed that isolate DBLα repertoires

are composed of *n* DBLα types and that each DBLα type is immunogenic and would present a unique variant-specific epitope and consequently elicit a variant-specific host immune response. The isolate repertoires in our simulation were sampled from our empirical data at random. In this simulation, we also assume exposure to a particular isolate repertoire results in acquired immunity to all the types present in the repertoire. Therefore, the accumulation of variant-specific immunity to all DBLα types circulating in the population would lead to the acquisition of sterilising immunity in the host. We envisage two scenarios. In scenario A, mosquitoes can transmit only single-genome infections (i.e., multiplicity of infection (MOI) = 1), whereas in scenario B, mosquitoes can transmit one or more than one infection (MOI ≥ 1). Therefore, in scenario A (but not in B) we sub-sampled our dataset to include only isolates with MOI = 1. We ran both scenarios with subsampling and without subsampling of MOI = 1 infections separately for upsA, non-upsA and all DBLα types for 100 iterations. The code used for this simulation is available on GitHub at https://github.com/pascualgroup/dbla_curves.

2.10. Statistical analysis

Statistical analyses were carried out using R v3.3.3 (https://cran.r-project.org). We used *base R* and the R packages *tidyverse* (Wickham et al., 2019) for data curation and visualisation. We used the *R* package *vegan* (Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P.R., O'Hara, R.B., Simpson, G. L., Solymos, P., Stevens, M.H.H., Szoecs, E., Wagner, H., 2019. vegan: Community Ecology Package. R package version 2.5–6. https://cran.r-project.org/web/packages/vegan/index.html) to generate species accumulation curves. For age-stratified analyses the study participants were categorized into five age groups (1–5, 6–10, 11–20, 21–39 and ≥40 years) based on the age-stratified study design, as well as into three broader age groups (children: 1–10 years, adolescents: 11–20 years, and adults: ≥20 years) for additional analyses. Chi-squared tests ($\chi 2$) were used for univariate analyses of categorical variables to compare proportions. Nonparametric tests were used to compare distributions of continuous variables between two groups (Mann-Whitney test) and among *k* groups (Kruskal-Wallis test) with a Bonferroni correction for multiple comparisons. A test was deemed to be statistically significant if the *P*-value was less than 0.05.

2.11. Data accessibility

The sequences for this Targeted Locus Study project have been deposited at DDBJ/ENA/GenBank under the Bio-Project Number: PRJNA 396962. The open-source tutorial detailing the data processing steps is available at https://github.com/UniMelb-Day-Lab/tutorialDBLalpha. The python code for the sequence cleaning pipeline is available at https://github.com/UniMelb-Day-Lab/DBLaCleaner. The python code to determine DBLα types is available at https://github.com/UniMelb-Day-Lab/clusterDBLalpha. The python code to translate and classify DBLα types is available at https://github.com/UniMelb-Day-Lab/classifyDBLalpha. The R script used for the simulations is available at https://github.com/pascualgroup/dbla_curves. All other analysis code is available at: https://github.com/shaziaruybal/ghana-var-age-immunity.

3. Results

3.1. Individuals of all ages harbour high var DBLα diversity at the end of sequential wet and dry seasons

To test the hypothesis that high var diversity explains infection in all ages, we characterised age-specific DBLα type diversity in this cohort at two time points. As expected for a high transmission setting, the prevalence of infection detected by microscopy and PCR varied by age in the wet and the sequential dry season (Supplementary Fig. S1A–B). A high proportion of adults harboured asymptomatic infections, particularly submicroscopic infections (>50%) (Supplementary Fig. S1A–B), and they were more likely to become parasite-negative at the EDS than children or adolescents (Chi-square test, $P < 0.001$, Supplementary Fig. 1C). Isolates from all individuals harbouring asymptomatic *P. falciparum* infections at either the EWS or the EDS were then typed for var DBLα diversity ($n = 1820$ *P. falciparum* isolates).

Sequence data were successfully obtained from 664 isolates at the EWS (out of 1151 infected residents) and 435 isolates (out of 669 infected residents) at the subsequent EDS. There were totals of 33,517 ($n = 2138$ upsA and 31,379 non-upsA) unique DBLα types at the EWS and 26,078 ($n = 1801$ upsA and 24,277 non-upsA) unique DBLα types at the EDS (Fig. 1A). Cumulative diversity curves indicate that we sampled the majority of the upsA types but not the more diverse non-upsA types in this local parasite population (Supplementary Fig. S2).

This study allowed us to document the extent of var diversity in the population over two seasons where the diversity at the end of the dry season can be considered as surviving from the prior wet season. Fig. 1 shows that 40.8% ($n = 17,296$) of the DBLα types (upsA and non-upsA) were found in both seasons (Fig. 1A) across all ages (Fig. 1B). The upsA were ~3× more likely to be found in both seasons than non-upsA types (Chi-square test, $P < 0.001$). There was a high proportion of non-upsA types seen only once in each season, i.e. rare types (50.6% EWS, 56.2% EDS, Supplementary Fig. S3). These rare types may have also persisted between seasons but were not detected by the seasonal sampling. Overall, we observed a large pool of types at a point in time and >40% of the types were seen in different individuals at two time points.

A key result was that individuals of all ages harboured diverse DBLα types at the EWS and EDS, albeit with children and adolescents harbouring more of the DBLα diversity than adults, i.e. a higher proportion of the total number of unique types was observed in children (Fig. 1B). Children harboured 55% of all the 42,399 unique DBLα types sampled in the population in both seasons compared with 45% in adolescents and 15% in adults. A similar age-specific decline in the proportion of total unique types identified was observed for both upsA and non-upsA types, but the proportion of conserved upsA types identified was higher compared with non-upsA, regardless of age (Fig. 1B). This was expected due to the smaller pool of these types in the population (2383 unique upsA types in both seasons compared with 40,016 non-upsA types). Overall, these age-specific diversity patterns show exposure to a large population of variants, with children and adolescents having greater carriage of var DBLα diversity.

3.2. var DBLα isolate population structure

Previously we showed limited relatedness and restricted networks of var DBLα repertoires in single infections in this population with combined data for EWS and EDS. The absence of recombinants was explained by variant-specific immune selection (He et al., 2018). As >80% of infections in this population were multi-genome ($MOI_{var} > 1$)–see below in Section 3.3, we repeated the repertoire relatedness analysis among isolates with DBLα data from all 1099 isolates using the similarity index PTS (Section 2.8). Overall, our results demonstrate a pattern of minimal relatedness (i.e., low median PTS ≤0.10) in the majority of isolates and very few highly-related parasites identified at the end of each season (Supplementary Fig. S4A–B) and temporally across seasons (Supplementary Fig. S4C). For all population-level stratifications, we found significantly higher sharing of upsA compared with the non-upsA types (Mann-Whitney test, $P < 0.001$, Supplementary Fig. S4). This pattern was expected due to the lower number of upsA types in the population. This lower number is also consistent with approximately 18–24% of the *var* genes per repertoire being upsA on average (Supplementary Table S1), as is the case for whole genome sequencing of laboratory strains (Rask et al., 2010).

Given the large pool of *var* DBLα types in the population, under conditions of variant-specific immune selection, we would expect the DBLα repertoires of isolates infecting children to show more relatedness than those associated with adults. To test this hypothesis, we examined whether age-related exposure would also result in differences in DBLα repertoire relatedness in children (i.e. limited exposure to the pool of variants) compared with adolescents and/or adults (i.e., semi-immune) within and between seasons. There was minimal repertoire relatedness (Fig. 2) and significantly higher sharing of upsA compared with the non-upsA types (Mann-Whitney test, $P < 0.001$) in each age group regardless of season (Supplementary Fig. S5). As predicted, the age-related repertoire relatedness was significantly higher in the children compared with adolescents and adults within and between seasons (Mann-Whitney test, $P \leq 0.001$ for all comparisons, Fig. 2) and more unrelated repertoires (i.e., PTS = 0) were identified in adults (Fig. 2).

3.3. Age-specific multiplicity of infection

Next, we examined age-dependent patterns of infection within individuals in relation to DBLα repertoire diversity. The unique structure of limited relatedness of DBLα repertoires among genomes in this transmission system (described in Section 3.2) allowed us to gain an understanding of the reservoir of diversity of DBLα types within an individual and to define isolate-specific DBLα repertoires regardless of multiplicity of infection (MOI). An estimate of MOI was inferred by *var* (MOI_{var} – i.e., assuming 45 non-upsA DBLα types per parasite genome) to describe the number of genetically distinct parasite genomes per isolate repertoire. Our data show that a high proportion of multi-genome ($MOI_{var} > 1$) infections were seen in all age groups at the EWS (maximum MOI of 14, 18, and 6 in children, adolescents, and adults, respectively) and at the EDS (maximum MOI of 11, 8, and 4 in children, adolescents, and adults, respectively) (Fig. 3). Even in adults, we found that more than half of their infections were $MOI_{var} > 1$ despite repeated exposure to potentially hundreds to thousands of distinct parasite genomes during their lifetime. Children had significantly higher MOI compared with adolescents or adults, and adolescents

had significantly higher MOI compared with adults (Mann-Whitney test, $P < 0.001$ for all comparisons, Fig. 3) in both seasons.

3.4. Within-host var diversity

Since a proportion of the cohort participants ($n = 575$) had infections (microscopic or submicroscopic) at both time points, we also measured DBLα repertoire relatedness "retrospectively" in the parasite isolates collected from the same individuals to examine temporal within-host diversity turnover patterns. For our analysis, we examined 296 of these individuals ($n = 592$ *P. falciparum* isolates) with DBLα data at both time points. Regardless of MOI_{var} or age, DBLα repertoire relatedness in these paired samples was very low (median PTS = 0.05), indicating within-host diversity turnover (Fig. 4). Consistent with our other analyses at the population level, there was also significantly higher sharing of upsA compared with the non-upsA types (Mann-Whitney test, $P < 0.001$, Supplementary Figs. S4D, S5D). We found evidence of a small number of possible chronic infections, since four individuals (1.4%, age range = 4–15 years) harboured infections sharing the same 43–50 non-upsA DBLα types between seasons, possibly indicating chronicity of one *P. falciparum* genome between seasons. There was one definitive case of a chronic infection where a 15 year old male harboured a chronic infection (PTS = 0.98) for at least 7 months between the EWS and EDS (i.e., throughout the dry season). This chronic infection was easily detected due to the PTS being almost 1, however, in most cases MOI_{var} was high and PTS relatively low, making it more difficult to confirm chronicity in the context of multi-genome infections. As expected, chronic infections were not readily observed in adults since they were more likely to turnover their infections between time points. The overall pattern of high turnover within hosts was again consistent with variant-specific immune selection.

3.5. Acquisition of variant-specific immunity: A simulation exercise

Having defined the extent of DBLα diversity and population structure in residents of all ages in both seasons, and for the first known time in adolescents and adults, we then explored what this diversity, degree of DBLα repertoire relatedness, and prevalence of MOI > 1 meant for the development of immunity to the major variant surface antigen of *P. falciparum* blood stages. The existence of high DBLα diversity (a minimum of 42,399 DBLα types) in 1099 isolates with unrelated DBLα repertoires implies that it would take a long time for an individual in Bongo to be exposed, and acquire variant-specific immunity, to all the currently circulating DBLα types if we assume that each type represents a unique major antigenic epitope (e.g. Recker et al., 2004). In addition, theory predicts that the acquisition of anti-PfEMP1 DBLα immunity is dependent on the degree of DBLα repertoire relatedness in the population and not just on prevalence of infection per se (Artzy-Randrup et al., 2012).

Given the enormous task of measuring variant-specific immunity to 42,399 variants in the cohort of 1541 individuals, we implemented a computational experiment to simulate the time it would take an individual to acquire immunity (i.e., via exposure) to 95% of the circulating types in Bongo. The key genomic epidemiological features described in our empirical data were used to refine our simulations to present a conservative exploration of acquisition of immunity in this population (see Section 2.9). We assume that isolate DBLα repertoires are composed of n DBLα types (i.e., repertoire size) and that each DBLα type is

immunogenic and would present a unique variant-specific epitope and consequently elicit a variant-specific host immune response. This assumption is reasonable as pairwise identity of types encoding DBLα domains are approximately 42% similar (Tonkin-Hill et al., 2021) and diversify by recombination, but not at the high mitotic rates described in vitro (Claessens et al., 2014) as seen by the levels of conservation in this study.

The isolate repertoires in our simulation were sampled from our empirical data at random and from our observed distribution of minimal DBLα repertoire relatedness. In this simulation, we also assume exposure to a particular isolate repertoire results in acquired immunity to all the types present in the repertoire. Therefore, the accumulation of variant-specific immunity to all DBLα types circulating in the population would lead to the acquisition of sterilising immunity in the host. We also envisaged two scenarios: scenario A, where mosquitoes can transmit only one infection with each infectious bite assuming no co-transmission of genomes (i.e., subsampling only MOI = 1, green in Fig. 5) or scenario B, assuming co-transmission of ≥1 genomes can occur with each infectious bite (i.e., MOI = 1 and MOI > 1, magenta in Fig. 5).

Fig. 5 shows the minimum number of infective bites needed to acquire immunity to 50% (Fig. 5A) and 95% of DBLα types (Fig. 5B), with 301 and 1353 infectious bites needed to acquire immunity to 50% of all types when allowing for co-transmission (i.e., MOI ≥ 1 infections) and no co-transmission (i.e., transmission of only MOI = 1 infections), respectively. We found the minimum number of infective bites needed to acquire immunity to 95% of DBLα types was 2185 and 13,774 infectious bites when allowing for co-transmission and no co-transmission, respectively. To translate these observations into a simulation of the accumulation of immunity within an individual host, we used an annual EIR of 25 infective bites per person per year for Bongo (Tiedje et al., 2017) to generate accumulation curves by age (Fig. 5C). As expected, acquisition of immunity to the upsA would take less time compared with the non-upsA or all DBLα types (Fig. 5C) due to the smaller pool of variants in the population. This is consistent with children acquiring immunity to severe disease earlier in the case of the disease-associated upsA types since there are fewer of these types in the population. There were marked differences when we allowed for co-transmission (i.e., MOI ≥ 1), with faster acquisition of immunity compared with transmission of MOI = 1, regardless of whether we stratified by upsA/non-upsA/all types (Fig. 5C). This is due to increased exposure to a higher number of types in the case of co-transmission. Importantly, the simulations showed that it would take more than 100 years to develop immunity to 95% of the circulating DBLα types, even with co-transmission.

4. Discussion

The motivation for this study was to understand why residents of high-transmission settings in Africa remain susceptible to blood stage infection. We did this by examining age-specific patterns of diversity of the genes encoding the major variant surface antigen that facilitate persistence of *P. falciparum* in the blood of infected humans. We were able to make a conservative estimate of the size of the reservoir of var DBLα diversity by deep sampling of microscopic and submicroscopic infections in all ages in an area of high seasonal transmission in Ghana. Overall diversity in these genes, as assessed by DBLα types in

all ages combining the EWS and the subsequent EDS samples, was in the order of tens of thousands of variants. This DBLα diversity at the end of two seasons with all ages combined was structured as predominantly non-overlapping isolate repertoires, as previously reported for children in other African sites (Chen et al., 2011; Day et al., 2017; Ruybal-Pesántez et al., 2017). This observed low relatedness of isolate var DBLα repertoires in all ages would not be expected under conditions of short-lived immune memory to PfEMP1 variants or generalised exposure-dependent immune responses (He et al., 2018).

Infections in children, adolescents, and adults were characterised by highly diverse *var* DBLα isolate repertoires, demonstrating that risk of infection by parasites with diverse *var* DBLα repertoires may occur at any age. Moreover, we discover the extent of multiple diverse *var* repertoires per infection ($MOI_{var} > 1$) with the finding that these multi-genome infections with diverse *var* repertoires occur in children as well as adolescents and adults. This is in line with earlier studies conducted in Ghana and Mali, where adults were also found to harbour multiclonal infections with diverse parasite clones, albeit based on genotyping of *msp2* polymorphisms, an antigen-encoding gene orders of magnitude less diverse than *var* genes (Owusu-Agyei et al., 2002; Sama et al., 2005; Felger et al., 2012; Sondén et al., 2015).

The observed age-specific decline in MOI inferred by *var* DBLα types is consistent with immune memory to PfEMP1 variants being long-lived as otherwise adults would keep accumulating parasites with diverse repertoires. Importantly, the age-specific patterns we describe support frequency-dependent immune selection as described by He et al. (2018), but interpreted here as age differences in host immune space. Indeed, our findings show that limited exposure in children is consistent with them having a higher carrying capacity to harbour $MOI_{var} > 1$ infections with diverse *var* DBLα repertoires as well as more related repertoires since they have more “gaps” in immune space than adults. In adults, however, related repertoires would be cleared more readily due to acquired variant-specific immunity, with only parasites with unrelated repertoires able to establish infections in these hosts. Existing infection and clinical data from West Africa are in support of our findings (Owusu-Agyei et al., 2001; Tran et al., 2013), such as reports from Mali where individuals followed up longitudinally were at risk of *P. falciparum* infections (based on PCR positivity) regardless of age, but risk of clinical malaria was inversely related with age (Tran et al., 2013). Our findings provide a genetic explanation for such observations since the high parasite diversity that exists in a high-transmission setting such as Bongo in Ghana is sufficient to allow for re-infection even in adults, despite the acquisition of broad exposure-dependent immunity over time. Based on the observed patterns we describe here, a role for cross-immunity or exposure-dependent generalised immunity is not ruled out but does not appear to be dominant in adolescents and adults.

Individuals infected at both time points (~7 months apart) were infected by unrelated var DBLα repertoires except for one adolescent and possibly a few children harbouring highly-related chronic infections. This apparent within-host diversity “turnover” at the EDS is not surprising since individuals may naturally clear their infections throughout the dry season when transmission is negligible. However, our observations of high turnover between two time points in this high transmission seasonal setting may appear to be at odds with

other longitudinal studies from Africa describing persistent, chronic infections in several African settings and in travellers returning from malaria-endemic areas (Staalsoe et al., 2002; Giobbia et al., 2005; Bachmann et al., 2009; Berry et al., 2018; Ndam et al., 2018). For example, in an unstable, seasonal transmission setting of Sudan, a longitudinal study reported an individual that harboured a chronic, monoclonal, and genetically identical infection (based on *msp2* genotyping) for over 7 months (Staalsoe et al., 2002). Given the limited resolution when genotyping only *msp2* alleles and the fact that parasite clones can harbour identical *msp2* alleles but different var repertoires as the latter fingerprint is more diverse than *msp2*, it is possible that these two seemingly genotypically identical isolates could indeed have different var repertoires. Several earlier longitudinal studies exploring infection dynamics in Africa and Papua New Guinea using 1–3 antigen-encoding genes have described fluctuations in parasite densities as well as periodicity in detection of *P. falciparum* due to synchronous replication, i.e., expansion of major and minor parasite populations throughout the course of an infection (Farnert et al., 1997; Babiker et al., 1998; Bruce et al., 2000; Koepfli et al., 2011; Felger et al., 2012). From these data it has been estimated that on average only 47–82% of the infection parasite clones are detected with a single finger-prick blood sample (Sama et al., 2005; Koepfli et al., 2011). When looking at the individuals with paired samples in our cohort, 53.7% of the infections would have been "asynchronous" at the time of the EDS survey (based on calculations of 48 h synchronous replication cycles) and possibly below detection limits. Nonetheless, it is worth noting that only 37.3% (575/1541) of individuals in our cohort experienced either microscopic or submicroscopic infections at both time points, highlighting the fact that recurrent infections were not a common epidemiological feature and most individuals who were infected at the EWS became slide- or PCR-negative by the EDS. Transmission intensity also plays a role in the apparent complexity of these chronic infections.

Importantly, we have discovered a local transmission system with tens of thousands of diverse *var* DBLα types organised into *var* DBLα repertoires that were not significantly related to those infecting other hosts nor the same host over time with only rare exceptions. Of further significance, we report maintenance of a large pool of *var* DBLα types over short time scales (3–6 months). Our longitudinal data indicate the stability of the *var* DBLα types in the reservoir over time despite potential for high sexual recombination rates, where *var* types but not repertoires are maintained temporally. The pattern of conservation of many of the individual *var* DBLα types over time is noteworthy given the high rates of mitotic recombination predicted by in vitro studies (Claessens et al., 2014). Overall this stability is consistent with the important role of balancing selection reported for *var* genes (Zilversmit et al., 2013; Larremore et al., 2015).

We undertook *in silico* simulations of acquisition of variant-specific immunity as a practical alternative to measuring immunity to 42,399 variants over a lifetime. Using the observed diversity and denominator for the number of DBLα repertoires detected in the study population as well as realistic transmission parameters, simulations of these empirical observations illustrated that even adults will remain susceptible to blood stage infection after a lifetime of repeated exposure to *P. falciparum* as a consequence of the extent and structure of diversity of *var* genes in highly diverse, non-overlapping repertoires. By making the distinction between upsA and non-upsA *var* DBLα types, we demonstrated that

acquisition of immunity to the less diverse and smaller pool of upsA types (i.e., severe disease-related types) would occur faster than to the more diverse non-upsA types. By the age of ~5 years an individual will have acquired immunity to ~50% of the upsA types circulating in Bongo. These patterns are consistent with theoretical work demonstrating immunity to severe disease is acquired early in life (Gupta et al., 1999) and with serological network studies (Buckee et al., 2009) where it was shown that different levels of immune selection occur upon different *var* gene groups (i.e., upsA and non-upsA). In addition, our findings support the existing empirical evidence demonstrating faster antibody acquisition to upsA-like PfEMP1 variants and their implication in protection from severe disease (Cham et al., 2010, 2009; Barry et al., 2011; Tessema et al., 2019).

Our simulation model presents a conservative case since we assume that exposure to a repertoire leads to lifelong immunity to all its *var* types, which in turn will elicit variant-specific host immune responses. In reality, it is conceivable that not all *var* types in a repertoire are expressed during an infection and that each antigenic *var* type may encode several antigenic epitopes (Recker et al., 2004; He et al., 2018), which would consequently lengthen the time to develop immunity to all epitopes and types. We have also underestimated the actual circulating diversity in the population since we assume a closed population i.e., no new diversity is generated, which does not consider genetic processes such as mutation, meiotic and mitotic recombination (He et al., 2018). In fact, even after identifying 42,399 unique DBLα types, sampling of the diversity still did not reach saturation as estimated by cumulative diversity curves. However, the upsA cumulative diversity curves appeared to level off, indicating the upsA immunity patterns we describe are based on exposure to the majority of upsA types in Bongo. They are consistent with the overall higher temporal stability compared with the non-upsA types. Moreover, the immunity patterns we describe by examining only the DBLα region of *var* genes would be more prolonged when considering other diverse regions of the entire gene (Otto et al., 2019). Various factors such as seasonality, differences in transmission intensity, crossreactivity among antigenic epitopes, and the efficiency of within-host immune responses (including to other surface antigens encoded by single copy genes, e.g. merozoite surface proteins and antigens encoded by other multi-gene families such as *rifins* and *stevors*), would make this process of developing immunity even longer.

Here we show that *var* DBLα diversity in this local community is orders of magnitude greater than the variation in the targets of current allele-specific blood-stage vaccines (Early et al., 2018). Whilst this high *var* diversity makes PfEMP1 variants unsuitable targets for blood-stage vaccines against infection in Africa, it points to the need for pre-erythrocytic vaccines and allele-specific blood stage vaccines, although targeted at different molecules, to be totally efficacious against all parasites to avoid maintaining sufficient *var* diversity in the parasite population.

In conclusion, the extent and structure of *var* DBLα diversity that we discovered is sufficient to explain why “gaps” still exist in the variant-specific or strain-specific PfEMP1 antibody responses as well as the age-specific signatures of infection in high transmission African settings. These diversity patterns can explain why immunity to *P. falciparum* blood stages is non-sterilising and why adults remain susceptible to infection, often with multiple

genomes, even after a lifetime of repeated exposure to this parasite. Importantly, examining *var* repertoires highlights within-host diversity of genomes in relation to key antigenic diversity driving transmission dynamics within and between hosts. Consequently, examining changes in *var* diversity and structure will provide important insights into the efficacy of transmission-blocking interventions with insecticides, antimalarial drugs, and/or vaccines on parasite antigenic fitness to persist in the host. The ability of such interventions to minimise the reservoir of *var* diversity will be key to achieving malaria elimination.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank the participants, communities, and the Ghana Health Service in Bongo District, Ghana, for their willingness to participate in this study. We also thank the personnel at the Navrongo Health Research Centre, Ghana, for sample collection and parasitological assessment/expertise. We thank Michael F. Duffy for his valuable input and discussions on an earlier version of the manuscript. We appreciate the support of the sequencing facilities at New York University Genome Technology Center, USA, and Australian Genome Research Facility, as well as the University of Melbourne/Bio21 Institute (Australia), University of Chicago (USA), and Walter and Eliza Hall Institute of Medical Research (Australia) for computational resources. This research was supported by the Fogarty International Center at the National Institutes of Health (USA) [Program on the Ecology and Evolution of Infectious Diseases (EEID), Grant number: R01-TW009670 to K.A.K., M.P. and K.P.D.] and the National Institute of Allergy and Infectious Disease, National Institutes of Health [Grant number: R01-AI084156 to K.P.D.]. S.R-P. was supported by a Melbourne International Engagement Award from the University of Melbourne. S.P. was supported by a James S. McDonnell, Foundation, United States 21st Century Science Initiative (USA) –Postdoctoral Program in Complexity Science Complex Systems Fellowship Award and by a Fulbright Fellowship from the U.S. Department of State.

References

Artzy-Randrup Y, Rorick MM, Day K, Chen D, Dobson AP, Pascual M, 2012. Population structuring of multi-copy, antigen-encoding genes in *Plasmodium falciparum*. eLife 1,. 10.7554/eLife.00093e00093.

Babiker HA, Abdel-Muhsin AM, Ranford-Cartwright LC, Satti G, Walliker D, 1998. Characteristics of *Plasmodium falciparum* parasites that survive the lengthy dry season in eastern Sudan where malaria transmission is markedly seasonal. Am. J. Trop. Med. Hyg 59, 582–590. 10.4269/ajtmh.1998.59.582. [PubMed: 9790434]

Bachmann A, Esser C, Petter M, Predehl S, von Kalckreuth V, Schmiedel S, Bruchhaus I, Tannich E, Heussler VT, 2009. Absence of erythrocyte sequestration and lack of multicopy gene family expression in *Plasmodium falciparum* from a splenectomized malaria patient. PLoS ONE 4 (10), e7459. 10.1371/journal.pone.0007459. [PubMed: 19826486]

Barry AE, Leliwa-Sytek A, Tavul L, Imrie H, Migot-Nabias F, Brown SM, McVean GAV, Day KP, Mackinnon MJ, 2007. Population Genomics of the Immune Evasion (*var*) Genes of *Plasmodium falciparum*. PLoS Pathog. 3 (3), e34. 10.1371/journal.ppat.0030034. [PubMed: 17367208]

Barry AE, Trieu A, Fowkes FJI, Pablo J, Kalantari-Dehaghi M, Jasinskas A, Tan X, Kayala MA, Tavul L, Siba PM, Day KP, Baldi P, Felgner PL, Doolan DL, 2011. The Stability and Complexity of Antibody Responses to the Major Surface Antigen of *Plasmodium falciparum* Are Associated with Age in a Malaria Endemic Area. Mol Cell Proteom 10, M111.008326. 10.1074/mcp.m111.008326

Bengtsson A, Joergensen L, Rask TS, Olsen RW, Andersen MA, Turner L, Theander TG, Hviid L, Higgins MK, Craig A, Brown A, Jensen ATR, 2013. A novel domain cassette identifies *Plasmodium falciparum* PfEMP1 proteins binding ICAM-1 and is a target of cross-reactive, adhesion-inhibitory antibodies. J. Immunol 190 (1), 240–249. 10.4049/jimmunol.1202578. [PubMed: 23209327]

Bernabeu M, Danziger SA, Avril M, Vaz M, Babar PH, Brazier AJ, Herricks T, Maki JN, Pereira L, Mascarenhas A, Gomes E, Chery L, Aitchison JD, Rathod PK, Smith JD, 2016. Severe adult

malaria is associated with specific PfEMP1 adhesion types and high parasite biomass. Proc. Nat. Acad. Sci 113 (23), E3270–E3279. 10.1073/pnas.1524294113. [PubMed: 27185931]

Berry I, Walker P, Tagbor H, Bojang K, Coulibaly SO, Kayentao K, Williams J, Oduro A, Milligan P, Chandramohan D, Greenwood B, Cairns M, 2018. Seasonal dynamics of malaria in pregnancy in West Africa: Evidence for carriage of infections acquired before pregnancy until first contact with antenatal Care. Am. J. Trop. Med. Hyg 98, 534–542. 10.4269/ajtmh.17-0620. [PubMed: 29210351]

Biggs BA, Gooze L, Wycherley K, Wollish W, Southwell B, Leech JH, Brown GV, 1991. Antigenic variation in Plasmodium falciparum. Proc. Natl. Acad. Sci 88 (20), 9171–9174. 10.1073/pnas.88.20.9171. [PubMed: 1924380]

Bruce MC, Donnelly CA, Packer M, Lagog M, Gibson N, Narara A, Walliker D, Alpers MP, Day KP, 2000. Age- and species-specific duration of infection in asymptomatic malaria infections in Papua New Guinea. Parasitology 121 (3), 247–256. 10.1017/s0031182099006344. [PubMed: 11085245]

Buckee CO, Bull PC, Gupta S, 2009. Inferring malaria parasite population structure from serological networks. Proc. Royal Soc. B Biol. Sci 276 (1656), 477–485. 10.1098/rspb.2008.1122.

Cham GKK, Turner L, Kurtis JD, Mutabingwa T, Fried M, Jensen ATR, Lavstsen T, Hviid L, Duffy PE, Theander TG, 2010. Hierarchical, domain type-specific acquisition of antibodies to *Plasmodium falciparum* erythrocyte membrane protein 1 in Tanzanian children. Infect. Immun 78 (11), 4653–4659. 10.1128/IAI.00593-10. [PubMed: 20823214]

Cham GKK, Turner L, Lusingu J, Vestergaard L, Mmbando BP, Kurtis JD, Jensen ATR, Salanti A, Lavstsen T, Theander TG, 2009. Sequential, ordered acquisition of antibodies to *Plasmodium falciparum* erythrocyte membrane protein 1 domains. J. Immunol 183 (5), 3356–3363. 10.4049/jimmunol.0901331. [PubMed: 19675168]

Chan J-A, Howell KB, Reiling L, Ataide R, Mackintosh CL, Fowkes FJI, Petter M, Chesson JM, Langer C, Warimwe GM, Duffy MF, Rogerson SJ, Bull PC, Cowman AF, Marsh K, Beeson JG, 2012. Targets of antibodies against *Plasmodium falciparum*–infected erythrocytes in malaria immunity. J. Clin. Invest 122 (9), 3227–3238. 10.1172/JCI62182. [PubMed: 22850879]

Chao A, Chazdon RL, Colwell RK, Shen T, 2005. A new statistical approach for assessing similarity of species composition with incidence and abundance data. Ecol. Lett 8, 148–159. 10.1111/j.1461-0248.2004.00707.x.

Chen DS, Barry AE, Leliwa-Sytek A, Smith T-A, Peterson I, Brown SM, Migot-Nabias F, Deloron P, Kortok MM, Marsh K, Daily JP, Ndiaye D, Sarr O, Mboup S, Day KP, Gruner AC, 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. 10.1371/journal.pone.0016629. [PubMed: 21347415]

Claessens A, Hamilton WL, Kekre M, Otto TD, Faizullabhoy A, Rayner JC, Kwiatkowski D, Deitsch K, 2014. Generation of antigenic diversity in *Plasmodium falciparum* by structured rearrangement of *var* genes during mitosis. PLoS Genet. 10 (12), e1004812. 10.1371/journal.pgen.1004812. [PubMed: 25521112]

Day KP, Artzy-Randrup Y, Tiedje KE, Rougeron V, Chen DS, Rask TS, Rorick MM, Migot-Nabias F, Deloron P, Luty AJF, Pascual M, 2017. Evidence of strain structure in *Plasmodium falciparum var* gene repertoires in children from Gabon, West Africa. Proc. Nat. Acad. Sci 114 (20), E4103–E4111. 10.1073/pnas.1613018114. [PubMed: 28461509]

Dodt M, Roehr JT, Ahmed R, Dieterich C, 2012. FLEXBAR—Flexible Barcode and Adapter Processing for Next-Generation Sequencing Platforms. Biology 1, 895–905. 10.3390/biology1030895. [PubMed: 24832523]

Early AM, Lievens M, MacInnis BL, Ockenhouse CF, Volkman SK, Adjei S, Agbenyega T, Ansong D, Gondi S, Greenwood B, Hamel M, Odero C, Otieno K, Otieno W, Owusu-Agyei S, Asante KP, Sorgho H, Tina L, Tinto H, Valea I, Wirth DF, Neafsey DE, 2018. Host-mediated selection impacts the diversity of *Plasmodium falciparum* antigens within infections. Nat. Commun 9, 1381. 10.1038/s41467-018-03807-7. [PubMed: 29643376]

Edgar RC, 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26, 2460–2461. 10.1093/bioinformatics/btq461. [PubMed: 20709691]

Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R, 2011. UCHIME improves sensitivity and speed of chimera detection. Bioinformatics 27, 2194–2200. 10.1093/bioinformatics/btr381. [PubMed: 21700674]

Farnert A, Rooth I, Snounou G, Bjorkman A, 1997. Daily dynamics of Plasmodium falciparum subpopulations in asymptomatic children in a holoendemic area. Am. J. Trop. Med. Hyg 56, 538–547. 10.4269/ajtmh.1997.56.538. [PubMed: 9180605]

Felger I, Maire M, Bretscher MT, Falk N, Tiaden A, Sama W, Beck H-P, Owusu-Agyei S, Smith TA, Gosling RD, 2012. The dynamics of natural *Plasmodium falciparum* infections. PLoS ONE 7 (9), e45542. 10.1371/journal.pone.0045542. [PubMed: 23029082]

Giobbia M, Tonon E, Zanatta A, Cesaris L, Bisoffi Z, Vaglia A, 2005. Late recrudescence of *Plasmodium falciparum* malaria in a pregnant woman: a case report. Int. J. Infect. Dis 9 (4), 234–235. 10.1016/j.ijid.2004.08.002.

Gupta S, Day KP, 1994. A strain theory of malaria transmission. Parasitol. Today 10 (12), 476–481. 10.1016/0169-4758(94)90160-0. [PubMed: 15275516]

Gupta S, Snow RW, Donnelly CA, Marsh K, Newbold C, 1999. Immunity to non-cerebral severe malaria is acquired after one or two infections. Nat. Med 5 (3), 340–343. 10.1038/6560. [PubMed: 10086393]

He Q, Pilosof S, Tiedje KE, Ruybal-Pesántez S, Artzy-Randrup Y, Baskerville EB, Day KP, Pascual M, 2018. Networks of genetic similarity reveal nonneutral processes shape strain structure in *Plasmodium falciparum*. Nat. Commun 9, 1817. 10.1038/s41467-018-04219-3. [PubMed: 29739937]

Jespersen JS, Wang CW, Mkumbaye SI, Minja DT, Petersen B, Turner L, Petersen JE, Lusingu JP, Theander TG, Lavstsen T, 2016. *Plasmodium falciparum var* genes expressed in children with severe malaria encode CIDRα1 domains. EMBO Mol. Med 8, 839–850. 10.15252/emmm.201606188. [PubMed: 27354391]

Koepfli C, Schoepflin S, Bretscher M, Lin E, Kiniboro B, Zimmerman PA, Siba P, Smith TA, Mueller I, Felger I, Borrmann S, 2011. How much remains undetected? Probability of molecular detection of human Plasmodia in the field. PLoS ONE 6 (4), e19010. 10.1371/journal.pone.0019010. [PubMed: 21552561]

Larremore DB, Sundararaman SA, Liu W, Proto WR, Clauset A, Loy DE, Speede S, Plenderleith LJ, Sharp PM, Hahn BH, Rayner JC, Buckee CO, 2015. Ape parasite origins of human malaria virulence genes. Nat. Commun 6, 8368. 10.1038/ncomms9368. [PubMed: 26456841]

Lau CY, Turner L, Jespersen J, Lowe E, Petersen B, Wang C, Petersen JV, Lusingu J, Theander T, Lavstsen T, Higgins M, 2015. Structural Conservation Despite Huge Sequence Diversity Allows EPCR Binding by the PfEMP1 Family Implicated in Severe Childhood Malaria. Cell Host Microbe 17 (1), 118–129. 10.1016/j.chom.2014.11.007. [PubMed: 25482433]

Lavstsen T, Turner L, Saguti F, Magistrado P, Rask TS, Jespersen JS, Wang CW, Berger SS, Baraka V, Marquard AM, Seguin-Orlando A, Willerslev E, Gilbert MTP, Lusingu J, Theander TG, 2012. *Plasmodium falciparum* erythrocyte membrane protein 1 domain cassettes 8 and 13 are associated with severe malaria in children. Proc Nat Acad Sci 109 (26), E1791–E1800. 10.1073/pnas.1120455109. [PubMed: 22619319]

Lennartz F, Lavstsen T, Higgins MK, 2017. Towards an anti-disease malaria vaccine. Emerg Top Life Sci 1 (6), 539–545. 10.1042/ETLS20170091. [PubMed: 33525843]

Magallón-Tejada A, Machevo S, Cisteró P, Lavstsen T, Aide P, Rubio M, Jiménez A, Turner L, Valmaseda A, Gupta H, De Las Salas B, Mandomando I, Wang CW, Petersen JEV, Muñoz J, Gascón J, Macete E, Alonso PL, Chitnis CE, Bassat Q, Mayor A, Blackman MJ, 2016. Cytoadhesion to gC1qR through *Plasmodium falciparum* Erythrocyte Membrane Protein 1 in Severe Malaria. PLoS Pathog 12 (11), e1006011. 10.1371/journal.ppat.1006011. [PubMed: 27835682]

Marsh K, 1992. Malaria-a neglected disease? Parasitology 104 (S1), S53–S69. 10.1017/S0031182000075247. [PubMed: 1589300]

Ndam NT, Tornyigah B, Dossou AY, Escriou G, Nielsen MA, Salanti A, Issifou S, Massougbodji A, Chippaux J-P, Deloron P, 2018. Persistent *Plasmodium falciparum* Infection in Women With an Intent to Become Pregnant as a Risk Factor for Pregnancy-associated Malaria. Clin. Infect. Dis 10.1093/cid/ciy380.

Otto TD, Assefa SA, Böhme U, Sanders MJ, Kwiatkowski D, Berriman M, Newbold C, 2019. Evolutionary analysis of the most polymorphic gene family in falciparum malaria. Wellcome Open Res 4, 193. 10.12688/wellcomeopenres. [PubMed: 32055709]

Owusu-Agyei S, Koram KA, Baird JK, Utz GC, Binka FN, Nkrumah FK, Fryauff DJ, Hoffman SL, 2001. Incidence of symptomatic and asymptomatic *Plasmodium falciparum* infection following curative therapy in adult residents of northern Ghana. Am. J. Trop. Med. Hyg 65, 197–203. 10.4269/ajtmh.2001.65.197. [PubMed: 11561704]

Owusu-Agyei S, Smith T, Beck H-P, Amenga-Etego L, Felger I, 2002. Molecular epidemiology of *Plasmodium falciparum* infections among asymptomatic inhabitants of a holoendemic malarious area in northern Ghana. Trop. Med. Int. Health 7 (5), 421–428. 10.1046/j.1365-3156.2002.00881.x. [PubMed: 12000651]

Rask TS, Hansen DA, Theander TG, Gorm Pedersen A, Lavstsen T, Eisen JA, 2010. *Plasmodium falciparum* erythrocyte membrane protein 1 diversity in seven genomes – Divide and conquer. PLoS Comput. Biol 6 (9), e1000933. 10.1371/journal.pcbi.1000933. [PubMed: 20862303]

Rask TS, Petersen B, Chen DS, Day KP, Pedersen AG, 2016. Using expected sequence features to improve basecalling accuracy of amplicon pyrosequencing data. Bmc Bioinform 17, 176. 10.1186/s12859-016-1032-7.

Recker M, Nee S, Bull PC, Kinyanjui S, Marsh K, Newbold C, Gupta S, 2004. Transient cross-reactive immune responses can orchestrate antigenic variation in malaria. Nature 429 (6991), 555–558. 10.1038/nature02486. [PubMed: 15175751]

Rorick MM, Rask TS, Baskerville EB, Day KP, Pascual M, 2013. Homology blocks of *Plasmodium falciparum var* genes and clinically distinct forms of severe malaria in a local population. BMC Microbiol. 13 (1), 244. 10.1186/1471-2180-13-244. [PubMed: 24192078]

Ruybal-Pesántez S, Tiedje KE, Tonkin-Hill G, Rask TS, Kamya MR, Greenhouse B, Dorsey G, Duffy MF, Day KP, 2017. Population genomics of virulence genes of *Plasmodium falciparum* in clinical isolates from Uganda. Sci. Rep 7, 11810. 10.1038/s41598-017-11814-9. [PubMed: 28924231]

Sama W, Owusu-Agyei S, Felger I, Vounatsou P, Smith T, 2005. An immigration–death model to estimate the duration of malaria infection when detectability of the parasite is imperfect. Stat. Med 24 (21), 3269–3288. 10.1002/sim.2189. [PubMed: 16143990]

Sondén K, Doumbo S, Hammar U, Vafa Homann M, Ongoiba A, Traoré B, Bottai M, Crompton PD, Färnert A, 2015. Asymptomatic multiclonal *Plasmodium falciparum* infections carried through the dry season predict protection against subsequent clinical malaria. J. Infect. Dis 212 (4), 608–616. 10.1093/infdis/jiv088. [PubMed: 25712968]

Staalsoe T, Hamad A, Hviid L, Elhassan I, Arnot D, Theander T, 2002. In vivo switching between variant surface antigens in human *Plasmodium falciparum* infection. J. Infect. Dis 186 (5), 719–722. 10.1086/342390. [PubMed: 12195363]

Tessema SK, Nakajima R, Jasinskas A, Monk SL, Lekieffre L, Lin E, Kiniboro B, Proietti C, Siba P, Felgner PL, Doolan DL, Mueller I, Barry AE, 2019. Protective immunity against severe malaria in children is associated with a limited repertoire of antibodies to conserved PfEMP1 variants. Cell Host Microbe 26 (5), 579–590.e5. 10.1016/j.chom.2019.10.012. [PubMed: 31726028]

Tiedje KE, Oduro AR, Agongo G, Anyorigiya T, Azongo D, Awine T, Ghansah A, Pascual M, Koram KA, Day KP, 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, 199–212. 10.4269/ajtmh.16-0959. [PubMed: 28719306]

Tonkin-Hill G, Ruybal-Pesántez S, Tiedje KE, Rougeron V, Duffy MF, Zakeri S, Pumpaibool T, Harnyuttanakorn P, Branch OH, Ruiz-Mesía L, Rask TS, Prugnolle F, Papenfuss AT, Chan Y-B, Day KP, Buchrieser C, 2021. Evolutionary analyses of the major variant surface antigen-encoding genes reveal population structure of *Plasmodium falciparum* within and between continents. PLoS Genet. 17 (2), e1009269. 10.1371/journal.pgen.1009269. [PubMed: 33630855]

Tonkin-Hill GQ, Trianty L, Noviyanti R, Nguyen HHT, Sebayang BF, Lampah DA, Marfurt J, Cobbold SA, Rambhatla JS, McConville MJ, Rogerson SJ, Brown GV, Day KP, Price RN, Anstey NM, Papenfuss AT, Duffy MF, Schneider D, 2018. The *Plasmodium falciparum* transcriptome in severe malaria reveals altered expression of genes involved in important processes including surface antigen–encoding *var* genes. PLoS Biol 16 (3), e2004328. 10.1371/journal.pbio.2004328. [PubMed: 29529020]

Tran TM, Li S, Doumbo S, Doumtabe D, Huang C-Y, Dia S, Bathily A, Sangala J, Kone Y, Traore A, Niangaly M, Dara C, Kayentao K, Ongoiba A, Doumbo OK, Traore B, Crompton PD, 2013. An Intensive Longitudinal Cohort Study of Malian Children and Adults Reveals No Evidence of

Acquired Immunity to *Plasmodium falciparum* Infection. Clin. Infect. Dis 57, 40–47. 10.1093/cid/cit174. [PubMed: 23487390]

Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, Grolemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen T, Miller E, Bache S, Müller K, Ooms J, Robinson D, Seidel D, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H, 2019. Welcome to the Tidyverse. J Open Source Softw 4, 1686. 10.21105/joss.01686.

World Health Organization, 2019. World Malaria Report 2019. World Health Organization, Geneva.

Zhang J, Kobert K, Flouri T, Stamatakis A, 2014. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics 30 (5), 614–620. 10.1093/bioinformatics/btt593. [PubMed: 24142950]

Zilversmit MM, Chase EK, Chen DS, Awadalla P, Day KP, McVean G, 2013. Hypervariable antigen genes in malaria have ancient roots. BMC Evol. Biol 13 (1), 110. 10.1186/1471-2148-13-110. [PubMed: 23725540]

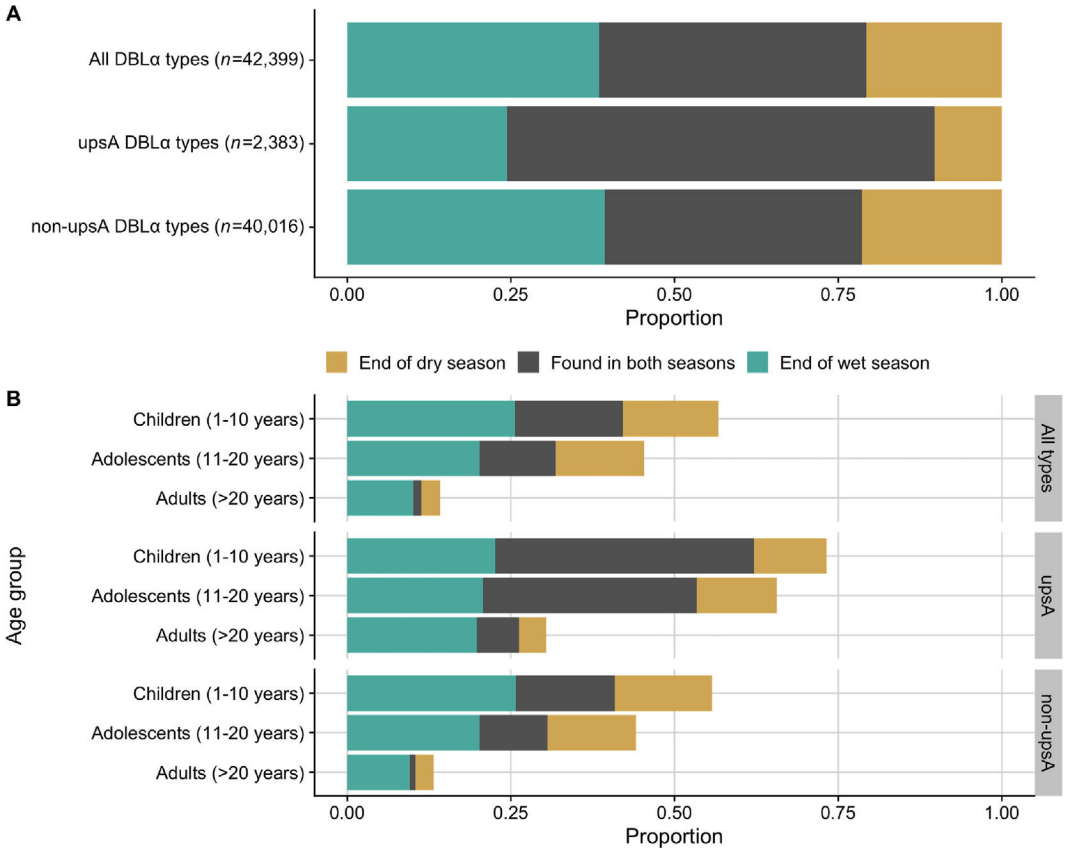

Fig. 1.

Individuals of all ages in Bongo, Ghana, harbour high *Plasmodium falciparum var* Duffy-binding-like alpha domain (DBLα) diversity at the end of both seasons. (A) Proportion of unique DBLα types identified at the end of the wet season, (EWS, turquoise), at the end of the dry season, (EDS, gold), and found in both seasons (grey). *n* refers to the total number of unique DBLα types. A total of 17,296/42,399 DBLα types were found at both time points. (B) Proportion of the total number of unique DBLα types identified in the study as seen in each age group at the EWS (turquoise), at the EDS (gold) and found in both seasons (grey). The proportion not filled (i.e., no colour) corresponds to types not identified in that particular age group. Note that a DBLα type could be identified in more than one age group. The DBLα types are also stratified by ups grouping (i.e. upsA and non-upsA) since expression of genes from these groups has been associated with different disease and infection outcomes.

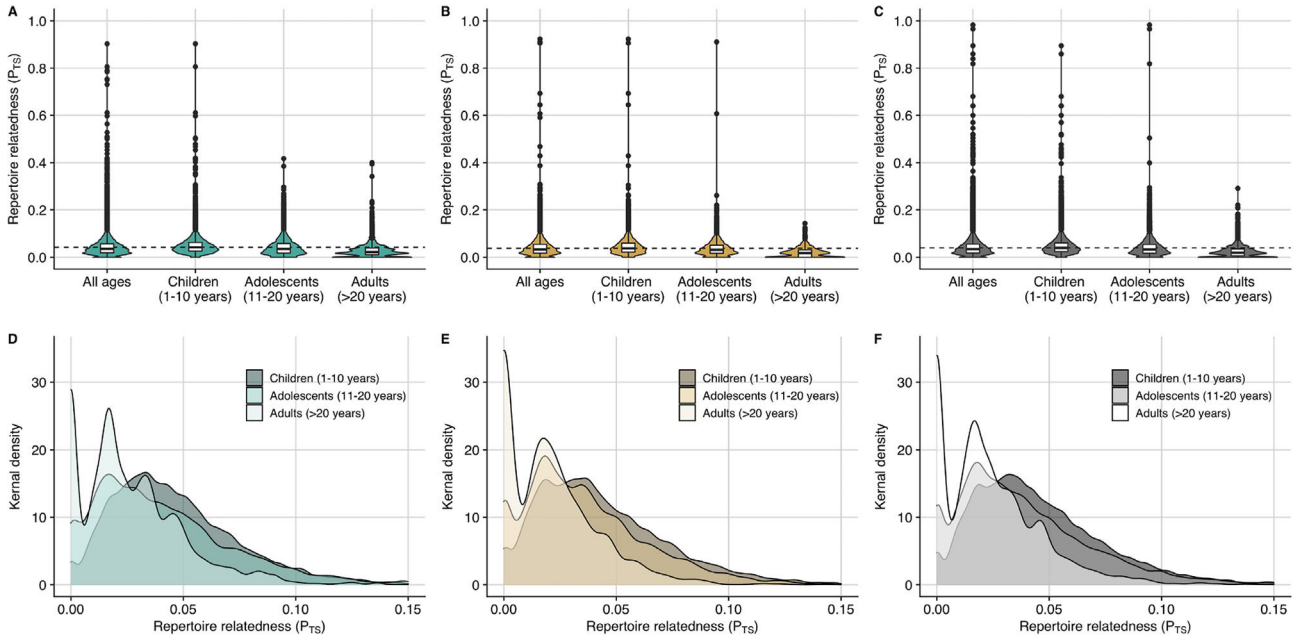

Fig. 2.

Age-specific patterns of *Plasmodium falciparum* isolate repertoire relatedness. (A–C) Violin plots show the distribution of pairwise type sharing (PTS) scores among isolate repertoires in each age group (A) at the end of the wet season (EWS), (B) at the end of the dry season (EDS), (C) comparing between seasons. The dashed line in (A–C) indicates the median PTS for children and box plots show the median and interquartile ranges for each age group stratification. (D–F) Kernal density plots show the lower end of the distribution of PTS scores among isolate repertoires in children between 1–10 years, adolescents between 11–20 years and adults >20 years at the (D) EWS, (E) EDS, and (F) between seasons. The total number of pairwise comparisons at the EWS were as follows: n = 440,232 for all comparisons (not age-stratified), n = 133,590 among children, n = 23,562 among adolescents, and n = 20,592 among adults. The total number of pairwise comparisons at the EDS were as follows: n = 188,790 for all comparisons (not age-stratified), n = 62,250 among children, n = 16,770 among adolescents, and n = 2970 among adults. The total number of pairwise comparisons between seasons were as follows: n = 577,680 for all comparisons (not age-stratified), n = 183,000 among children, n = 40,040 among adolescents, and n = 15,840 among adults.

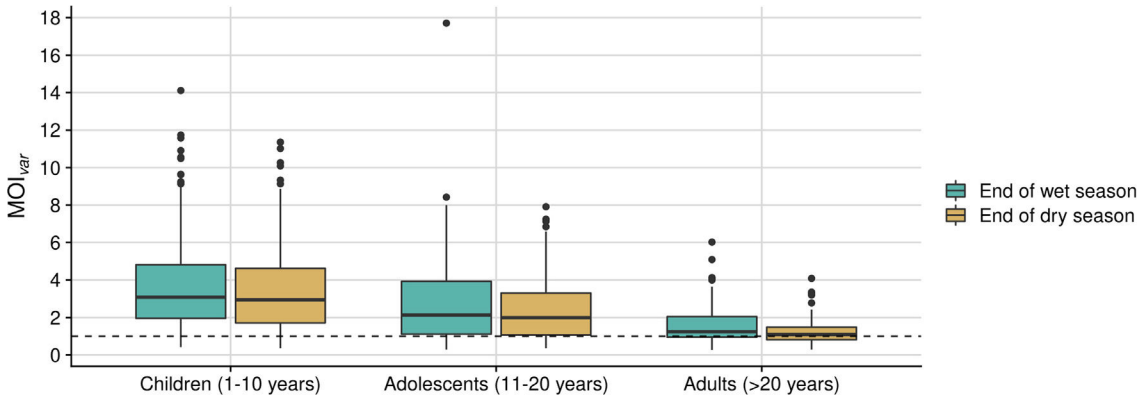

Fig. 3.

Age-specific differences in *Plasmodium falciparum* multiplicity of infection inferred by *var* (MOI_{var}) at the end of the wet season (EWS, turquoise) and at the end of the dry season (EDS, gold). Values above the dashed line indicate $MOI_{var} > 1$. There were significant differences in the distribution of MOI_{var} between children, adolescents and adults in each season (Mann-Whitney test, $P < 0.001$ for all comparisons at EWS or EDS). There were no significant differences in MOI_{var} between seasons for any age group (Mann-Whitney test, $P > 0.05$) except for adults (Mann-Whitney test, $P = 0.02$).

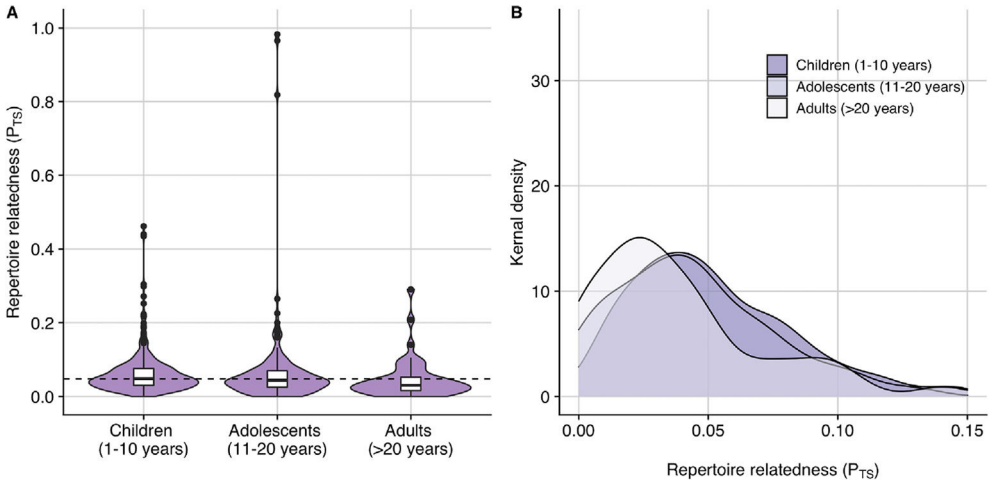

Fig. 4.

Age-specific patterns of *Plasmodium falciparum* isolate repertoire relatedness in paired samples. (A) Violin plots show the distribution of pairwise type sharing (PTS) scores among paired isolate repertoires in each age group. The dashed line indicates the median PTS for children and box plots show the median and interquartile ranges. (B) Density plots show the lower end of the distribution of PTS scores among isolate repertoires in paired samples from children between 1–10 years, adolescents between 11–20 years and adults >20 years. The total number of pairwise comparisons were as follows: $n = 385$ among children, $n = 162$ among adolescents, and $n = 45$ among adults.

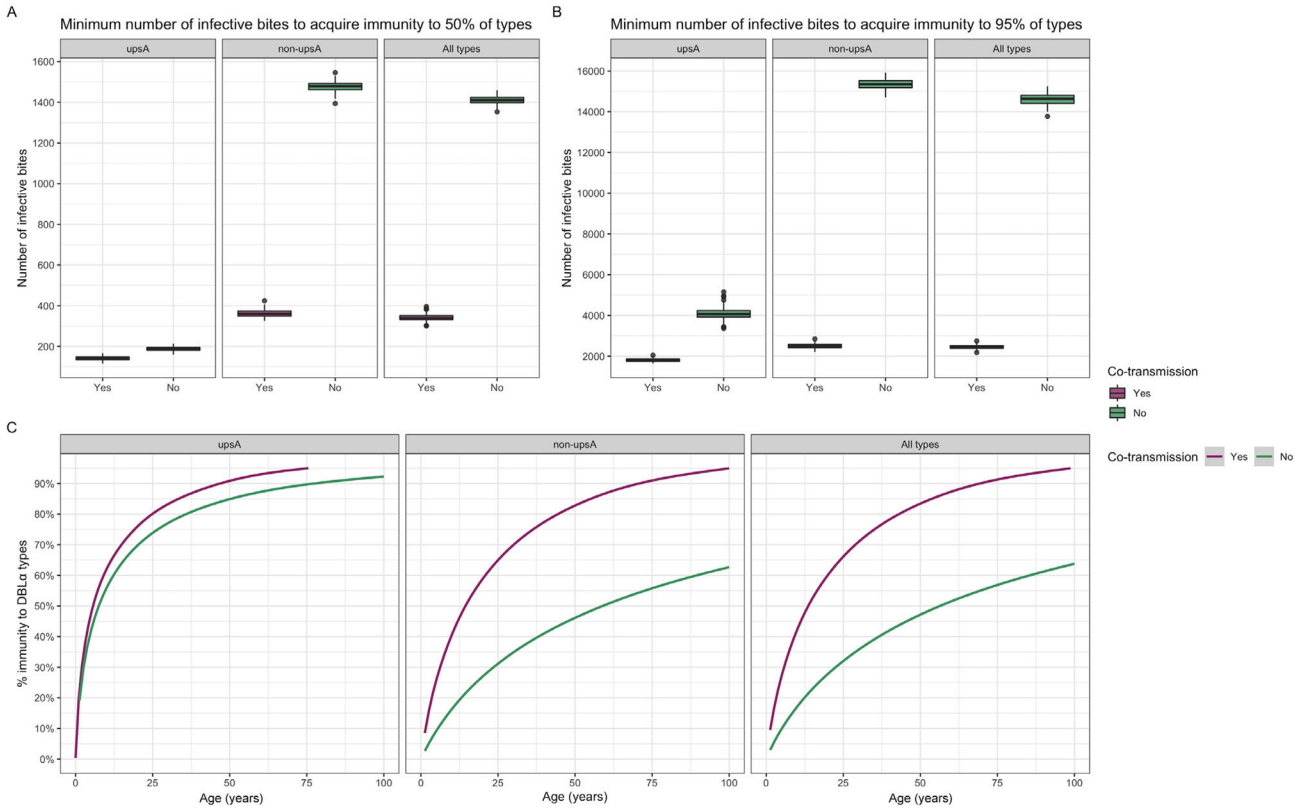

Fig. 5.

Simulation of patterns of acquisition of variant-specific immunity to *Plasmodium falciparum* var Duffy-binding-like alpha domain (DBLα) types in Bongo, Ghana. Scenario A (green) was calculated by assuming a mosquito can transmit only one infection (i.e., multiplicity of infection or MOI = 1, no co-transmission) and scenario B (magenta) was calculated by assuming a mosquito can transmit ≥1 infection with each infectious bite (i.e., MOI ≥ 1, allowing for co-transmission). Simulations were carried out for the different DBLα type ups groupings (i.e. upsA and non-upsA) since expression of genes from these groups has been associated with different disease and infection outcomes. (A-B) The minimum number of infective bites necessary to acquire immunity to (A) 50% of the DBLα types and (B) 95% of the DBLα types based on 100 simulations. (C) Accumulation curves showing the time it takes an individual to acquire immunity to 95% of upsA, non-upsA, and all DBLα types based on an annual entomological inoculation rate (EIR) of 25 (see Section 2.9). As an illustrative example, it would take ~5, ~12 and ~12 years for an individual to acquire immunity to 50% of the upsA, non-upsA, and all DBLα types, respectively, in the case of co-transmission.