**Extreme diversity and population structure of *var* genes explains why immunity to the blood stages of *Plasmodium falciparum* is non-sterilizing**

**Opening sentence…**

Epidemiologic theory provides a framework for the implementation of malaria elimination. Specifically, the ability of a parasite to transmit from one human to another is defined mathematically as R0 , the basic reproductive number. The R0 must be less than one to achieve local elimination, and must be greater than one for malaria parasites to persist in an endemic area. R0 for *Plasmodium spp* is equal to the *vectorial capacity multiplied by* the *duration* of infection in the human host. The daily rate at which new humans will become infected from an infective human case by mosquitoes defines vectorial capacity and can be measured experimentally. *Duration* of infection captures the complexity of host-parasite interactions in humans to allow successful transmission to the mosquito.

*Plasmodium falciparum*, the most virulent of human malarias, has evolved a specific immune evasion mechanism called clonal antigenic variation to allow chronic infection to establish within the human host. Prolonging the duration of infection within the human host facilitates transmission to mosquitoes and is critical in areas where malaria transmission is highly seasonal. Clonal antigenic variation involves the differential and sequential expression of up to 60 highly polymorphic *var* genes within a parasite genome. These genes encode the major variant surface antigen of the blood stages known as *P. falciparum* erythrocyte membrane protein 1.

*Var* genes have been shown to be highly diverse among the limited number of sequenced genomes (Freitas-Junior 2000, Rask 2010). Diversification of *var* genes occurs by both meiotic and mitotic recombination (Freitas-Junior 2000, Duffy 2009, Claessens 2014), where the obligatory sex in the mosquito diversifies repertoires of up to 60 *var* genes as a result of independent assortment among the 14 *P. falciparum* chromosomes during conventional meiosis. Mitotic recombination can generate within-repertoire diversity. Based on their chromosomal location and semi-conserved upstream promoter sequences (ups) *var* genes can be classified into three groups: upsA, upsB, and upsC (Gardner 2002, Lavstsen 2003). Even with limited sampling, we (Barry 2007, Chen 2011, Day 2017, Ruybal-Pesántez 2017, Rougeron 2017, Rorick 2018) and others (Tessema 2015, Dara 2017 etc) have shown that the DBLα regions of *var* genes are highly variable in nature. Intriguingly, we have shown a non-random population structure of *var* (DBLα)repertoires in areas where outcrossing is frequent and across areas that varied dramatically in transmission (Day 2017, Ruybal-Pesántez 2017). Strikingly, there was minimal overlap among DBLα repertoires within the populations and modeling using similarity networks pointed to a strain structure of the least related repertoires of *Pf*EMP1 DBLα variants generated by immune selection (He 2018).

Here we extend our previous work to explore the key features of the epidemiology of malaria infection in a high transmission setting of West Africa through the lens of *var* genomics. Specifically, we have examined the genetic complexity of the *var* system to answer three intriguing questions: (i) how does transmission reestablish after a dry season (ii) why are adults infected?, and (iii) how can superinfection/co-infection occur during blood-stage infections even in immune adults? By scaling up molecular surveillance of the *P. falciparum* reservoir of infection to include infections in all ages, not just children as was the case for our previous work (REFS), we addressed the age-specific and temporal changes in *var* diversity and population structure in this area of highly seasonal transmission in West Africa. Here we also explore what *var* complexity at the individual and population level means for the development of immunity to the *P. falciparum* blood stages.

Using a longitudinal cohort, we surveyed 1,541 residents of all ages (range = 1- 91 years) across two broad community catchment areas from Bongo District, Ghana over two sequential transmission seasons. Using both microscopy and species-specific PCR diagnosis, the prevalence of malaria parasites was 75.0% at the end of the wet season (EWS) and 43.4% at the end of the dry season (EDS) (Figure S1, Tiedje 2017). Individuals of all ages harbored asymptomatic *P. falciparum* infections at the EWS (range = 1-86 years) and EDS (range = 1-82 years). The demographic characteristics of the individuals in the cohort are presented in Supplementary Table X. To explore seasonal, age-specific and temporal patterns of *var* diversity we sequenced the conserved DBLα domain of the *var* genes of *P. falciparum* isolates collected from 1,099 infected residents of Bongo (see Supp for sequencing details). *Var* DBLα diversity in this cohort was characterized at both the population level and by the *P. falciparum* isolate from an individual in each season and over time.

**Genetic complexity of the *var* system in Bongo**

We aimed to understand whether there were significant changes in the population genetics of *var* genes at the EDS compared to the preceding EWS since the *P. falciparum* reservoir that survives through the 7 month dry season among asymptomatic carriers will reinitiate transmission when mosquito populations re-emerge. Our findings showed that despite a significant reduction in the prevalence of asymptomatic infection from the EWS to the EDS (from 75.0% to 43.4%, *p* < 0.001), total *var* DBLαdiversity remained strikingly high (33,617 unique DBLα types EWS to 26,078 unique DBLα types at the EDS) and infection complexity was not greatly reduced (83.4% and 79.5% complex asymptomatic infections at the EWS and EDS, respectively, *p* = 0.052) even after approximately seven months of limited to no malaria transmission between seasons. We next examined whether there was a bottleneck in diversity as evidenced by a change in the prevalence of rare and abundant DBLα types between seasons. At both seasonal time points, approximately half of the DBLα types were rare and <10% of the DBLα types were abundant (Supp).

The expression of upsA *var* genes has been associated with uncomplicated and/or severe malaria (including specific severe disease phenotypes, e.g. cerebral malaria) providing compelling evidence to support antigenic and functional differences between upsA and non-upsA *var* gene groups (REFS). There were no significant differences in the proportion of upsA and non-upsA DBLα types identified in each season (*p* = 0.204) (Figure 1A, 1B). Importantly, we showed that the upsA DBLα types were significantly more likely to be abundant compared to non-upsA DBLα types in both seasons (*p* < 0.001 EWS, EDS).

Interestingly, despite the extensive seasonal DBLα diversity and over 42,300 unique DBLα typescirculating in the population, we report considerable maintenance of 41% of these types between seasons (Figure 1C). The upsA DBLα types were ~3x more likely to be maintained between seasons than non-upsA DBLα types (*p* < 0.001). These findings are consistent with upsA DBLα types being more abundant in the population (Supp) and generally more conserved relative to non-upsA types (Ruybal-Pesántez 2017). Overall, the seasonal changes in the prevalence of DBLα types were directly related to their frequency category at the EWS (Figure 1D, Supp). Accordingly, there were 16,321 DBLα types (48.5%) sampled at the EWS that were not sampled again at the EDS, of which 72.6% were rare types. In contrast, all of the abundant upsA DBLα types and all but four of the abundant non-upsA DBLα types sampled at the EWS were maintained between seasons (Figure 1D, Supp). Indeed the median percent attributable risk (%AR) (i.e., percent change in DBLα type prevalence between seasons) was close to 0 for the intermediary and abundant DBLα types that were maintained between seasons (Figure 1D). Given the extremely high rates of meiotic and mitotic recombination that the parasite could undergo in this short time scale our results are striking. Here, we have clearly shown that DBLα types (both upsA and non-upsA) are maintained between seasons, with abundant types being more likely to be maintained (Supp), and a clear demonstration of the extent of diversity to start the next transmission season. This also provides the first evidence of temporal *var* sequence conservation and stability in the *P. falciparum* transmission system over short periods of time (i.e. across two sequential transmission seasons).

We then assessed whether parasite genetic structure was impacted by the seasonal changes in transmission. We described *var* population structure, as defined by the organization and relatedness (i.e. overlap) of DBLα repertoires among isolates, within each season and the temporal *var* dynamics between seasons by calculating pairwise type sharing (PTS) (see Supp). Importantly, the extent of overlap of *var* DBLα repertoires among isolates will determine cross-immunity in the host population to distinct parasite genomes. Our PTS comparisons revealed minimal overlap of repertoires as evidenced by strikingly low PTS scores (i.e., PTS ≤ 0.10, or ≤10% DBLα type sharing) (Figure 2). Overall there was an extremely low prevalence of highly related isolate repertoire pairs with higher overlap (i.e., PTS ≥ 0.50) (Figure 2). When we examined the type-specific PTS patterns, there was significantly higher sharing of upsA DBLα types compared to the non-upsA types (*p* < 0.001). These patterns were independent of seasonality (Figure 2) and infection complexity (Supp). We interpret these data as indicating high parasite fitness in the reservoir to evade the host immune response even under conditions of seasonal transmission. Perhaps most interestingly, when we examined the turnover of infections on a population-level and within the same individuals, we observed a high turnover of repertoires between seasons, with evidence of only one chronic infection that persisted in the same host from the EWS through to the subsequent EDS for at least 7 months (Figure 2). Intriguingly, we show that there is maintenance of DBLα sequence types but not repertoires between seasons.

Our population genetic data have revealed the complexity of the *P. falciparum* transmission system, which, even at a local scale (5 - 40 km), is composed of thousands of genetically distinct *P. falciparum* genomes organized into distinct and minimally overlapping isolate *var* repertoires. Based on our data we estimated at least 2,986 *P. falciparum* genomes circulating in this population. It is noteworthy that this is an underestimation of the actual number of infecting *P. falciparum* genomes. Our findings revealed that the 664 asymptomatic *P. falciparum* infections surveyed at the EWS were highly unrelated to the 434 asymptomatic infections at the EDS, with the exception of one chronic infection (Figure 2). This implies extensive population turnover and is consistent with immunity to *Pf*EMP1 variants creating transmission dynamics (e.g. (Artzy-Randrup et al. 2012)) and the periodic turnover of the immunodominant epitopes of *Pf*EMP1 previously shown through expression studies (Warimwe et al. 2016; Abdi et al. 2016). Furthermore our results indicate that the asymptomatic reservoir in Bongo consists of an exceptionally large effective parasite population that is extremely diverse, as defined by DBLα diversity, and exists within a relatively small human host population (1,099 infected individuals surveyed).

**How can superinfection/co-infection occur during the blood stages? What are the age-specific patterns of *var* dynamics? Why can adults still carry infections?**

Superinfection or co-infection (i.e., complexity of infection) is a key feature of the epidemiology of *P. falciparum* in areas of high transmission. The prevalence of complex infections at the EWS and EDS was 83.4% and 79.5%, respectively (Figure 3A). Overall, we describe an inverse relationship between age and infection complexity in both seasons (Figure 3A), consistent with the age- and exposure-dependent acquisition of immunity, which manifests as a decrease in parasite prevalence with age as well as asymptomatic parasitaemia through to adulthood (REFS). Interestingly, our data show that adults are infected even after repeated exposure to thousands of distinct parasite genomes during their lifetime, and more than half of the infections in adults >20 years were complex (Figure 3A). Remarkably, even in the adults ≥40 years, 65.5% and 50.0% of the asymptomatic infections were complex at the EWS and EDS, respectively (Figure 3A). This is likely due to the vast DBLα diversity that exists in this area (i.e., the “diversity hypothesis”) and the potential for diverse parasites expressing new variants to avoid variant-specific immune responses in these hosts, which in turn facilitates the establishment of new infections, even in adults.

We next described, the age-specific patterns of *var* repertoire overlap in detail. A key result was that immune selection appeared to act against recombinant repertoires (i.e., with higher PTS overlap) in an age-specific manner due to differences in host immune space, whereby increased exposure and acquired immunity in older children and adults selects against variants already seen by the host (Figure 3B-E). In turn, recombinant repertoires are cleared more readily in adults. Higher immune space in children ≤10 years is also consistent with significantly higher complexity of infection (i.e., superinfection/co-infection) compared to adolescents and adults and leads to higher overlap among repertoires in children with complex infections but not adolescents or adults, even though >50% of adults carried complex infections (Figure 3B-E). These observations are in strong agreement with the age-specific patterns of acquisition of immunity to *P. falciparum* (REF).

**What are the implications of *var* diversity for the development of immunity to *P. falciparum* blood stages?**

The development of naturally acquired immunity to malaria is highly dependent on the antigenic variants that circulate in a given population. The existence of extremely high DBLα diversity (minimum of 42,399 DBLα types) in a parasite population with essentially non-overlapping DBLα repertoires observed implies that it would take a long time for an individual to be exposed to and acquire specific immunity to all the currently circulating DBLα types, if we assume that each DBLα represents a unique variant–specific epitope (e.g. (Recker et al. 2004)). In fact, theory predicts that the acquisition of anti-*Pf*EMP1 DBLα immunity is dependent on the *degree* of *var* repertoire overlap in the population and not just on prevalence of infection per se (Artzy-Randrup et al. 2012). Thus, an important motivation of our work was to explore what *var* complexity, in terms of the extent of diversity, infection complexity and degree of repertoire overlap, means for the development of immunity to malaria blood stages and perhaps explain why it is non-sterilizing. We therefore set out to answer the following question: how long would it take an individual in Bongo to be exposed to and develop sterilizing immunity to all the DBLα types circulating in the population? A simple simulation experiment allowed us to gain some insights into this question. Using an approach similar to those used in ecology, we generated accumulation curves to simulate the time it would take an individual to accumulate, or acquire immunity (i.e., via exposure) to all the circulating DBLα types using the key epidemiological features described in our empirical data to refine our simulation (see Suppfor further details on the simulation experiment). We implemented these simulations separately for upsA, non-upsA and all DBLα types to examine whether there were any differences in the accumulation of the different ups DBLα type groupings. The more limited pool of upsA DBLα types compared to the pool of non-upsA DBLα types maintained between seasons led us to hypothesize that immunity to the more conserved upsA DBLα types would be acquired more rapidly than to the non-upsA DBLα types. Additionally, we also tested the effect of varying DBLα pool sizes on the acquisition of immunity.

Two clear patterns emerged: (i) firstly, regardless of DBLα pool size it was clear that the acquisition of immunity to the upsA DBLα types would take the least amount of time compared to the non-upsA DBLα types and all DBLα types, (ii) secondly, there were marked differences when we allowed for co-transmission of DBLα repertoires (MOI ≥ 1), with faster overall acquisition of immunity compared to mosquito transmission strictly of single-genome infections (MOI = 1), regardless of whether we stratified by ups groupings (Figure 4). This is likely due to the potential of increased exposure to a higher number of DBLα types in the case of co-transmission. Interestingly, the simulations showed that it would take an extraordinary amount of time to develop immunity to 95% of the “more conserved” upsA DBLα types in the population, considerably longer than an individual’s average lifespan and this is assuming that the population is closed (i.e., no new diversity is generated or imported), which of course would lead to an underestimate of the actual diversity in a population. The assumption that each DBLα type would present a unique variant–specific epitope and consequently elicit a variant-specific host immune response is perhaps overly simplistic since modeling has shown that both major and minor epitopes exist (Recker et al. 2004), which in turn may be recognized differently by the host (Warimwe et al. 2013; Abdi et al. 2016). Therefore, we defined DBLα types at various sequence identity thresholds to account for the possibility that epitopes that are antigenically cross-reactive can be shared between sequences with a lower identity (e.g. (Bengtsson et al. 2013)). Overall, our *in silico* simulation experiment coupled with our empirical observations can explain why immunity to *P. falciparum* blood stages is non-sterilizing since residents of Bongo will continually be exposed to genetically distinct *P. falciparum* genomes with diverseDBLα repertoires. Importantly, we demonstrate that it would take an extremely long time for them to be exposed to and acquire specific immunity to all the circulating DBLα types, even when we corrected for potential cross-reactivity between the DBLα types defined in our study.

**Concluding Remarks-**

This investigation represents the most comprehensive study to date of the asymptomatic *P. falciparum* reservoir across all ages, including but not limited to submicroscopic infections, and the first study to describe the age-specific, seasonal trends, and temporal *var* transmission dynamics. We found no evidence of a population bottleneck with regards to *var* diversity at the EDS despite a drop in transmission between seasons, since a large pool of DBLαdiversity and overall a highly diverse parasite reservoir existed at the EDS to seed transmission during the subsequent wet season. On the time scale we examined, a highly diverse asymptomatic parasite reservoir would be available to start the next transmission season. Furthermore, the presence of complex asymptomatic infections across all ages will facilitate the transmission of parasites to mosquitos even when vectors appear seasonally because of the long duration of infection within the host. Our findings revealed that every asymptomatic carrier in Bongo was infected by parasites with diverse repertoires and on a population-level these repertoires were highly unrelated within and between seasons. The high population turnover of repertoires could be explained by residual transmission at the EWS into the dry season (e.g. areas with standing water) resulting in a small number of recombination events or also recent liver-stage infections that were not yet in circulation thus undetectable at the time of survey. Fluctuations in parasite densities (i.e., expansion of major and minor parasite populations) have been shown to occur in individuals harboring complex *P. falciparum* infections (Bruce 2000). This could explain the high within-host turnover of repertoires we observed between seasons since 53.7% of the infections in the paired samples would have been “asynchronous” at the time of survey.

Through the lens of *var* genomics, we demonstrate that the *P. falciparum* transmission system in Bongo is extremely complex. Despite this complexity, three key features of the molecular epidemiology emerge: (i) extremely high sequence diversity of *var* DBLα, (ii) limited overlap of DBLα repertoires in both seasons, and (iii) rapid turnover of isolate repertoires but maintenance of DBLα types between seasons. Our findings provide a unified explanation for several intriguing features of the molecular epidemiology of *P. falciparum* malaria in endemic areas. Firstly, the long-standing given explanation for the failure to develop sterilizing immunity in endemic areas is antigenic diversity, however, until now we have lacked population genetic and longitudinal field data to describe the key diversity that may drive these transmission dynamics of *P. falciparum* within and between hosts. The extensive *var* DBLα sequence diversity, high proportion of complex infections and unique structure of non-overlapping DBLαrepertoires explains why acquisition of immunity to malaria is non-sterilizing, since individuals will be continually exposed to genetically distinct *P. falciparum* genomes with extremely diverse and non-overlapping *var* repertoires throughout their lifetime. Furthermore, simulations of our empirical data showed that individuals in Bongo would never acquire sterilizing immunity during their lifetime due to the vast DBLα diversity that exists in this area and the potential for diverse parasites expressing new variants to avoid variant-specific immune responses in these hosts and establish new infections, even in adults. Even when the overallpool size of *var* antigenic variants circulating in Bongo was reduced there was a minimal effect on the pattern of acquisition of immunity to malaria, with the development of sterilizing immunity taking considerably longer than an individuals average lifespan. Superinfection/co-infection with multiple distinct genomes has also been shown to be an age-dependent feature of the epidemiology of *P. falciparum* in humans. Again the high diversity and limited overlap among *var* repertoires explains how superinfection can occur in the blood of an individual and how these individuals can be continually reinfected at any age by the same parasite population. This is why adults in endemic areas continue to carry blood stage infections, even complex infections. Importantly, these findings can also explain how high endemicity exists on such a local scale in a relatively small human population.

Strain theory – R0 concluding sentences.