

RESEARCH ARTICLE

# Competition for hosts modulates vast antigenic diversity to generate persistent strain structure in *Plasmodium falciparum*

Shai Pilosof<sup>1\*</sup>, Qixin He<sup>1</sup>, Kathryn E. Tiedje<sup>2</sup>, Shazia Ruybal-Pesántez<sup>1,3</sup>, Karen P. Day<sup>2</sup>, Mercedes Pascual<sup>1,3</sup>

**1** Department of Ecology and Evolution, University of Chicago, Chicago, Illinois, United States of America, **2** School of BioSciences, Bio21 Institute/University of Melbourne, Melbourne, Australia, **3** Santa Fe Institute, Santa Fe, New Mexico, United States of America

\* [shainova@gmail.com](mailto:shainova@gmail.com)



**OPEN ACCESS**

**Citation:** Pilosof S, He Q, Tiedje KE, Ruybal-Pesántez S, Day KP, Pascual M (2019) Competition for hosts modulates vast antigenic diversity to generate persistent strain structure in *Plasmodium falciparum*. PLoS Biol 17(6): e3000336. <https://doi.org/10.1371/journal.pbio.3000336>

**Academic Editor:** Daniel B. Larremore, University of Colorado, Boulder, UNITED STATES

**Received:** September 1, 2018

**Accepted:** June 5, 2019

**Published:** June 24, 2019

**Copyright:** © 2019 Pilosof et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** The sequences for this project have been deposited at DDBJ/ENA/GenBank under the BioProject Number: PRJNA 396962. Data used to produce figures are deposited in [https://figshare.com/projects/Competition\\_for\\_hosts\\_modulates\\_vast\\_antigenic\\_diversity\\_to\\_generate\\_persistent\\_strain\\_structure\\_in\\_Plasmodium\\_falciparum/63218](https://figshare.com/projects/Competition_for_hosts_modulates_vast_antigenic_diversity_to_generate_persistent_strain_structure_in_Plasmodium_falciparum/63218) The following code is available at GitHub: The original C++ code for the ABM: <https://github.com/pascualgroup/VarModel2>. R and Mathematica code for analysis: [https://github.com/pascualgroup/Pf\\_](https://github.com/pascualgroup/Pf_)

## Abstract

In their competition for hosts, parasites with antigens that are novel to the host immune system will be at a competitive advantage. The resulting frequency-dependent selection can structure parasite populations into strains of limited genetic overlap. For the causative agent of malaria, *Plasmodium falciparum*, the high recombination rates and associated vast diversity of its highly antigenic and multicopy *var* genes preclude such clear clustering in endemic regions. This undermines the definition of strains as specific, temporally persisting gene variant combinations. We use temporal multilayer networks to analyze the genetic similarity of parasites in both simulated data and in an extensively and longitudinally sampled population in Ghana. When viewed over time, populations are structured into modules (i.e., groups) of parasite genomes whose *var* gene combinations are more similar within than between the modules and whose persistence is much longer than that of the individual genomes that compose them. Comparison to neutral models that retain parasite population dynamics but lack competition reveals that the selection imposed by host immunity promotes the persistence of these modules. The modular structure is, in turn, associated with a slower acquisition of immunity by individual hosts. Modules thus represent dynamically generated niches in host immune space, which can be interpreted as strains. Negative frequency-dependent selection therefore shapes the organization of the *var* diversity into parasite genomes, leaving a persistence signature over ecological time scales. Multilayer networks extend the scope of phylodynamics analyses by allowing quantification of temporal genetic structure in organisms that generate variation via recombination or other non-bifurcating processes. A strain structure similar to the one described here should apply to other pathogens with large antigenic spaces that evolve via recombination. For malaria, the temporal modular structure should enable the formulation of tractable epidemiological models that account for parasite antigenic diversity and its influence on intervention outcomes.

[temporal\\_networks](#). Python code for the sequence cleaning pipeline: <https://github.com/UniMelb-Day-Lab/DBLaCleaner>. Python code to determine DBL-alpha types: <https://github.com/UniMelb-Day-Lab/clusterDBLalpha>.

**Funding:** The study was supported by a grant from the Fogarty International Center at the National Institutes of Health (Program on the Ecology and Evolution of Infectious Diseases), Grant number: R01-TW009670 to MP and KPD. SP was supported by a James S. McDonnell Foundation 21st Century Science Initiative – Postdoctoral Program in Complexity Science-Complex Systems Fellowship Award (grant number 220020423) and by a Fulbright Fellowship from the U.S. Department of State. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

**Abbreviations:** ABM, agent-based model; BD, Bongo District; CR1, complement receptor 1; DBL $\alpha$ , Duffy binding-like-alpha; EDS, end of dry season; EIR, entomological inoculation rate; EWS, end of wet season; IRS, indoor residual spraying; MOI, multiplicity of infection; PfEMP1, *P. falciparum* erythrocyte membrane protein 1; Sv, survey; ups, upstream promoter.

## Introduction

The dynamic arms race between hosts and pathogens sets the stage for a selective advantage of rare variants that confer either immune protection to hosts or immune escape to pathogens and a corresponding disadvantage of common ones. This frequency-dependent effect can act as a form of balancing selection and is a powerful force promoting high antigenic diversity and maintaining polymorphisms significantly longer than neutral drift. High diversity and temporal persistence of diversity at the gene level have been shown in numerous host-pathogen systems. Genes that encode pathogen resistance in hosts, such as the major histocompatibility complex [1], and those that underlie antigenic variation in parasites, such as the *var* multigene complex in the malaria parasite *P. falciparum*, display exceptional polymorphism compared to other functional genes. Several components of *var* genes are known to have originated millions of years ago and to be shared with closely related species that infect apes [2].

In transmission systems, diversity and persistence are also relevant at a higher level of organization than that of individual genes. In particular, pathogen strains concern temporally persistent combinations of genes related to infection, including those encoding antigens. The role of immune selection at this higher level of organization remains, however, poorly understood and documented, especially in pathogens whose antigenic variation involves vast diversity generated via recombination within the genome or between different genomes, as is the case in several bacteria, protozoa, and fungi [3]. Can strains exist and persist in such vast antigenic spaces? What is a strain in dynamic systems undergoing recombination at the level of both the genes themselves and the genomes they compose? A key characteristic signature of immune selection would involve the persistence of gene combinations over longer time scales than expected under neutrality [4], a hypothesis that remains to be examined despite its relevance for the existence and definition of strains themselves.

We address this temporal dimension by examining the role frequency dependence plays in maintaining gene combinations over time in the highly diverse multicopy *var* gene family of *P. falciparum*. Frequency-dependent selection in pathogen systems is analogous to stabilizing competition in ecological communities [5, 6]. Competition can drive coexisting species to self-organize into clusters with limiting similarity along a niche or trait axis [7–10]. In epidemiology, strain theory had proposed that competition for hosts through cross-immunity, the cross-protection conferred by previous acquisition of immunity, can structure pathogen populations into temporally stable sets of genetically distinct strains with limited overlap of antigenic repertoires [11]. Theoretical studies for low-to-medium genetic diversity predicted parasite strains with no or limited genetic overlap [12, 13], including in the case of multicopy genes [14]. For *P. falciparum*, recent work allowing for realistic levels of genetic diversity comparable to that found in endemic high transmission regions of sub-Saharan Africa resulted in a more complex similarity structure clearly distinguishable from patterns generated under neutrality but that can no longer be described by distinct clusters [15]. Deep sampling and sequencing of *var* gene isolates from asymptomatic human populations within a given time window or transmission season have confirmed these nonrandom patterns that are also non-neutral, in the sense that they cannot be simply explained by stochastic extinction, immigration, and transmission in the absence of acquired immune memory and therefore competition of parasites for hosts [15–17]. The lack of explicit consideration of the temporal dimension in the analysis of structure, despite the highly dynamic nature of the *P. falciparum* system, may be the reason why no apparent distinct clustering was identified, and it is the motivation behind this work.

Analyzing the temporal dimension requires simultaneously tracking the population dynamics of genomes and their genetic relationship in a hyperdiverse system. An additional challenge is allowing for evolution via recombination. Here, we apply multilayer networks to

characterize patterns of genetic similarity in *var* gene repertoires through time for both a theoretical model and an empirical data set from Ghana, unique in its depth of sampling and coverage over multiple seasons. The patterns of modularity we describe should be relevant to other pathogen systems, whether possessing multicopy gene families [3] or multiple independent loci encoding different antigens.

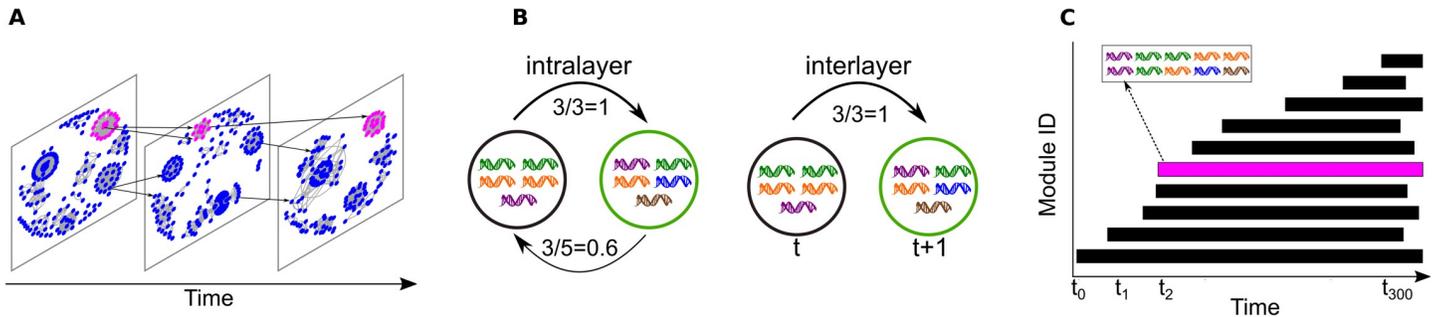
## Results

The *var* genes encode the major antigen of the blood stage of infection, *P. falciparum* erythrocyte membrane protein 1 (PfEMP1) [18]. Besides immune evasion, PfEMP1 promotes adherence of erythrocytes to blood vessels, leading to disease manifestations. Hence, *var* genes play an important role in malaria epidemiology. Each parasite genome has a “repertoire” of 50–60 unique *var* gene copies sequentially expressed to produce different variants of PfEMP1 during an infection. In endemic regions of high transmission, *var* genes exhibit enormous diversity [16, 17, 19], resulting from evolutionary innovation at two levels of organization. At the gene level, *var* gene variants can be generated through both mutation and ectopic recombination [20–23], with tens of thousands of variants documented in local populations [16]. At the repertoire level, variation in gene composition is obtained through sexual (meiotic) recombination in the mosquito vector [20, 24, 25].

Transmission and *var* diversity are positively correlated in the evolutionary history of malaria; i.e., areas of low and high transmission have evolved towards low (for example, South America) and high (for example, sub-Saharan Africa) diversity, respectively. Because our questions target systems of high diversity and transmission, we focus on this regime in the main text and present results for low and medium diversity in [S1 Text](#). We simulated populations using an agent-based model (ABM) and sampled them monthly for 25 years with a total of 300 time points, where each time point is a snapshot of changes accumulated in the parasite population during a 30-day period. We used a temporal multilayer network to analyze population structure in these time series. Competition between repertoires, as well as their evolution and persistence, are intimately interlinked processes emanating from genetic differences. Hence, each layer (time point) was constructed to represent a network of genetic similarity, in which nodes are repertoires and intralayer edges encode genetic similarity between repertoire pairs in terms of their *var* gene composition ([Methods](#); [15]). Using the same index, we connected layers to each other via unidirectional interlayer edges depicting the genetic similarity between a repertoire in time  $t$  to those in time  $t + 1$  ([Fig 1](#); [Methods](#)). This definition of interlayer edges maps the genetic relationship among repertoires, which underlies competition—the process determining persistence—into the network, explicitly creating temporal dependency between layers. We characterized network structure by looking for modules using Infomap, an algorithm that explicitly considers the multilayer structure and the temporal “flow” associated with interlayer edges [26–28]. These modules (i) contain repertoires that are genetically more similar to each other than to other repertoires in the network and (ii) are defined within and across layers (across time). Under immune selection, two repertoires that are clustered to the same module can be thought of as sharing a similar niche in human immune space. I.e., they are able to infect hosts with more similar immune profiles than parasites from other modules.

### The temporal and non-neutral population structure

Under both immune selection and the two neutral scenarios, we find a dynamic modular structure, in which modules are constantly generated and die out. Interestingly, the structure generated in the presence of immune selection exhibits long-lived coexisting modules, whereas the one generated in the two neutral models is characterized by short-lived ones ([Fig 2A, 2B](#)



**Fig 1. A toy example for a temporal multilayer network of repertoire genetic similarity and its associated modular structure.** In (A), each layer represents a time point. The network in each time point depicts the genetic similarity between pairs of repertoires, where each repertoire is a combination of alleles (see Methods for the reason why we used alleles instead of *var* genes). The measure of genetic similarity is asymmetric to take into account the asymmetric competition resulting from different numbers of distinct allele variants, as shown in the example in (B): the two repertoires share 3 alleles, but the black one has 3 unique alleles and the green has 5, and so the later will outcompete the former (alleles are depicted as DNA symbols with different colors). Interlayer edges between repertoires leading from any time point  $t$  to time point  $t + 1$  were defined in the same way, but these edges only point in one direction to represent temporal flow. For clarity, we present only a few interlayer edges in (A). (C) We used an algorithm for community detection in networks to identify “modules” (depicted as rectangles), which contain repertoires that are more genetically similar to each other than to repertoires from other modules. Modules persist in time, but the number and identity of the repertoires that compose them can change (for example, repertoires can be removed by host immunity and can appear as a result of recombination). New modules appear in time, while others die out. Pink repertoires in (A) are part of the same module illustrated by the same color in (C).

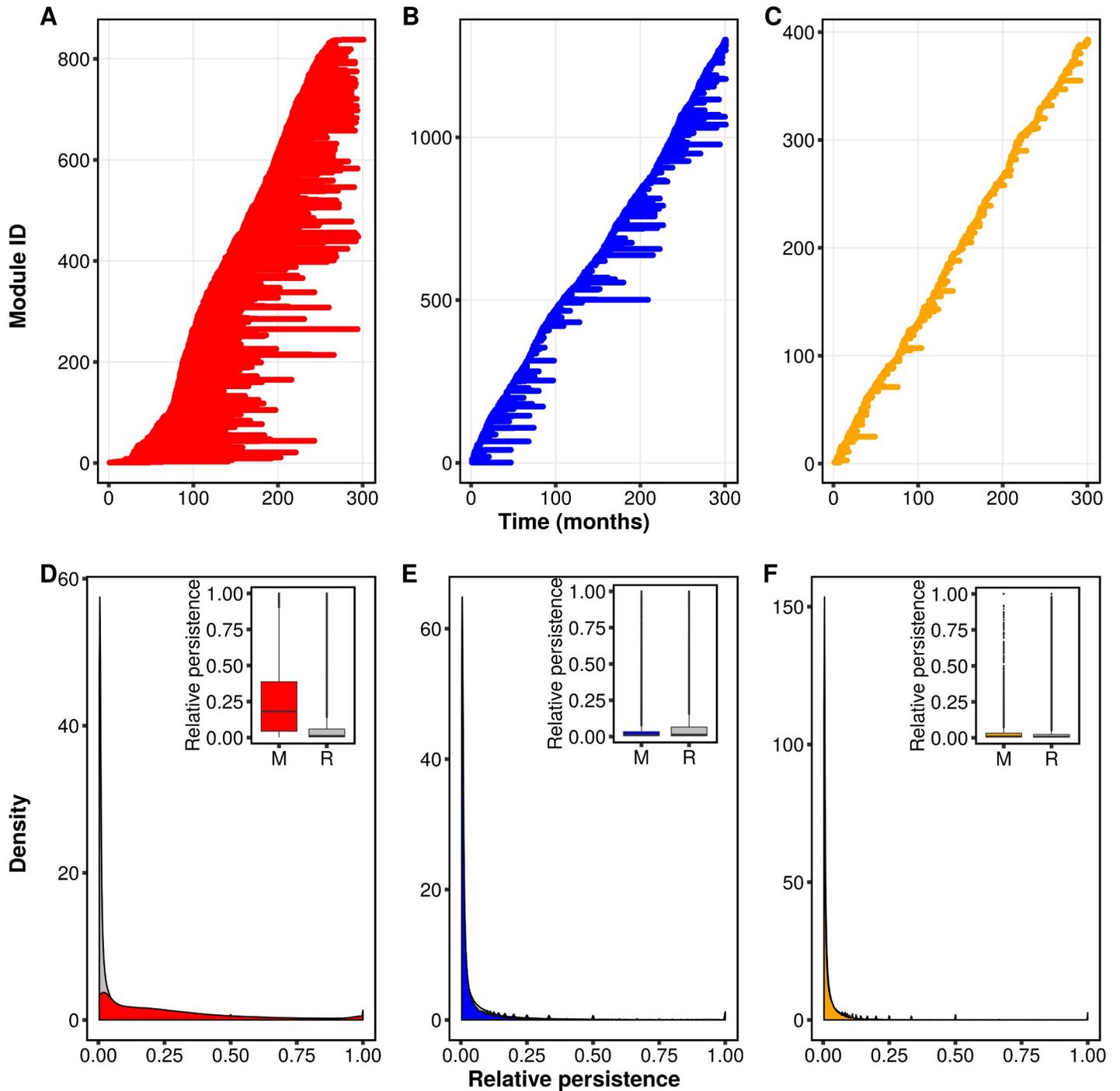
<https://doi.org/10.1371/journal.pbio.3000336.g001>

and 2C). We can examine the role of immune selection by comparing the persistence of repertoires to that of modules. When immune selection is at play, we find that modules persist longer than the repertoires that compose them (Fig 2D). This mismatch in persistence indicates that modules are not necessarily composed of the same repertoires across time and that repertoires do not necessarily have to persist as long as the module they occupy. By contrast, in the neutral scenarios, modules are short-lived compared to their repertoires (Fig 2E and 2F). This contrast suggests that the process of module formation is inherently different between the scenarios: under neutrality, genetic changes accumulate due to antigenic drift until a point is reached at which the repertoire population is sufficiently different to create a new set of modules. Conversely, a modular structure under selection is a result of the stabilizing competition between repertoires, which pushes them to be as dissimilar as possible. Indeed, limiting similarity and module coexistence in the selection scenario leads to a more even distribution of the number of repertoires in modules compared to neutrality (S6 Fig). The uneven distribution of module sizes in the neutral scenarios is a result of (i) short-lived modules with low tendency to coexist and (ii) the high number of recombinant repertoires, which in the absence of immune selection are not purged and thus fill the antigenic space and blur the clear niche separation observed in the selection scenario. Taken together, these results indicate that immune selection acts as a form of balancing selection, maintaining long-lived modules.

### Comparison to empirical data

We tested our theoretical findings using parasite isolate data collected in an age-stratified longitudinal study in the Bongo District (BD), Ghana [29]. This area is characterized by high seasonality, with a prolonged dry season (November–May) and a short wet season (June–October). Our data set contains 6 surveys (Svs) carried out at the end of both the wet and dry seasons (EWS and EDS) (S1 Table); it is the only longitudinal data set available of an asymptomatic *P. falciparum* reservoir population in a given location using deep sequencing.

To validate our theoretical results with empirical data requires generating an expectation benchmark for the population structure under the empirical sampling scheme with the ABM. To this end, we first need to corroborate our theoretical predictions when the seasonality of

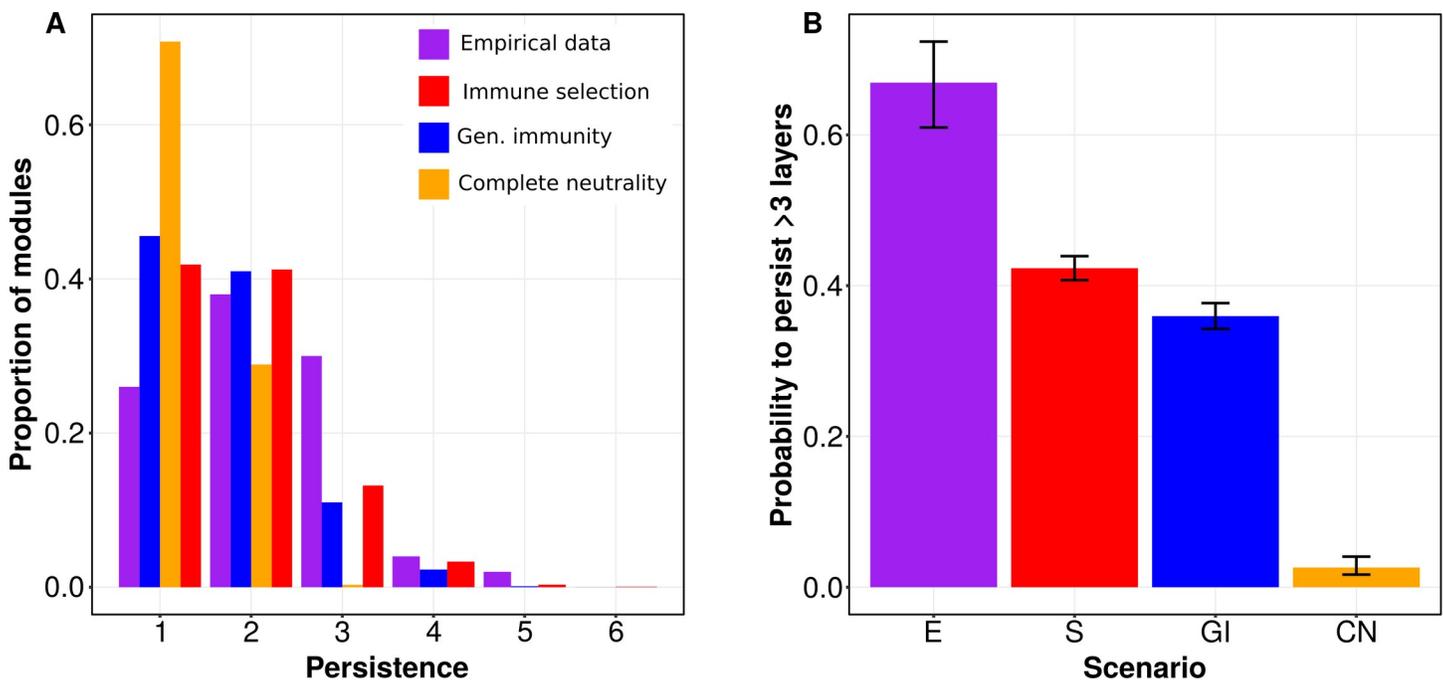


**Fig 2. Temporal population structure in the high-diversity regime.** The left, middle, and right columns represent the selection (red), generalized immunity (blue), and complete neutrality (orange) scenarios, respectively. (A,B,C) Example of population structure from one run of the ABM. Each line on the y-axis corresponds to a different module, and its horizontal length depicts its occurrence across layers. Modules are generated and die out. (D) The selection scenario is characterized by modules (red) that persist for much longer than the repertoires (gray) that compose them. The inset is a comparison between the values in the red and gray density plots using box plots. (E,F) In the neutral scenarios, modules and repertoires have similar relative persistence. See [Methods](#) for details on how we calculated relative persistence. Results in (D)–(E) are for data pooled across 50 runs of the ABM. Data from model output used to produce the figure can be found in [https://figshare.com/articles/Figure\\_data/8079713](https://figshare.com/articles/Figure_data/8079713). ABM, agent-based model; M, module; R, repertoire.

<https://doi.org/10.1371/journal.pbio.3000336.g002>

the sampling area is included (see [Methods](#) for details). We find that, although seasonality does leave a signature in the formation of modules ([S9 Fig](#)), the differences between the immune selection and the neutral scenarios remain qualitatively similar ([S10 Fig](#)). Next, we ran 100 ABM benchmark simulations with parameter values that capture our uncertainty of the exact local parameters in the sampling area, as was done in [15]. Beyond seasonality, these benchmark simulations consider the specific characteristics of our empirical data: only 6 layers (as compared to 300 in the theoretical results) and three sessions of indoor residual spraying (IRS) conducted as interventions during the sampling period. Our goal is not to fit the benchmark simulations to the data, but rather to show that when we imitate the empirical data with our ABM, the immune selection scenario still produces modules with higher persistence than the two neutral scenarios and is also more similar to the empirical data.

We find that module persistence is indeed longer under immune selection than under neutrality ([Fig 3A](#)). Results of logistic regression show that the probability that a module persists for at least 3 layers is higher under immune selection than under either of the two neutral scenarios and that this difference is statistically significant. This probability is also closer to that of the empirical data ([Fig 3B](#); we chose here 3 layers because no module persisted for longer under complete neutrality). Because we matched by construction the distribution of infection duration in the generalized immunity model to that of the immune selection one (see [Methods](#)), module persistence probabilities are more similar to each other between these two scenarios than to those generated under complete neutrality (yet still significantly different; Wald test,  $\chi^2 = 28.1$ ,  $P < 0.0001$ ). Therefore, even with a short time window and one that includes an intervention, we are able to show differences consistent with those expected from our



**Fig 3. Comparison of empirical data to ABM simulations.** (A) Proportion of modules that persisted for a given number of layers in the empirical data (purple) compared to ABM simulations of the selection (red), generalized immunity (blue), and complete neutrality (orange) scenarios. (B) Predicted probability (by logistic regression) that a module in each of the scenarios will persist for at least 3 layers. Note that we can only include modules that were born in layer 3 or before. All scenarios are different from each other, and these differences are statistically significant. In particular, there is a statistically significant difference between selection and generalized immunity (Wald test,  $\chi^2 = 28.1$ ,  $P < 0.0001$ ). Data from model output and empirical data used to produce the figure can be found in [https://figshare.com/articles/figure\\_data/8079713](https://figshare.com/articles/figure_data/8079713). ABM, agent-based model; CN, complete neutrality; E, empirical data; GI, generalized immunity; S, immune selection.

<https://doi.org/10.1371/journal.pbio.3000336.g003>

theoretical results. This analysis provides evidence for the role of immune selection in maintaining niche coexistence and persistence in an empirical setting.

### Epidemiological consequences of strain structure

The modular structure we have uncovered may affect epidemiological parameters because immunity gained to a repertoire from one module facilitates immunity to other repertoires from the same module, and infection with a repertoire from a different module adds exposure largely to alleles that have not been seen before. Also, despite modularity, genetic similarity between any two repertoires is generally low because of the very large diversity of the gene pool to begin with. Thus, to examine whether the modular structure does have an epidemiological meaning, we used duration of infection as a measure. Specifically, we generated curves describing the decline in infection duration with accumulated infections in an individual host. By sampling from the parasite population structure generated by the ABM under the three scenarios, we simulated the accumulation of infections in naive hosts for 1 year, explicitly incorporating the epidemiological force of infection of the system.

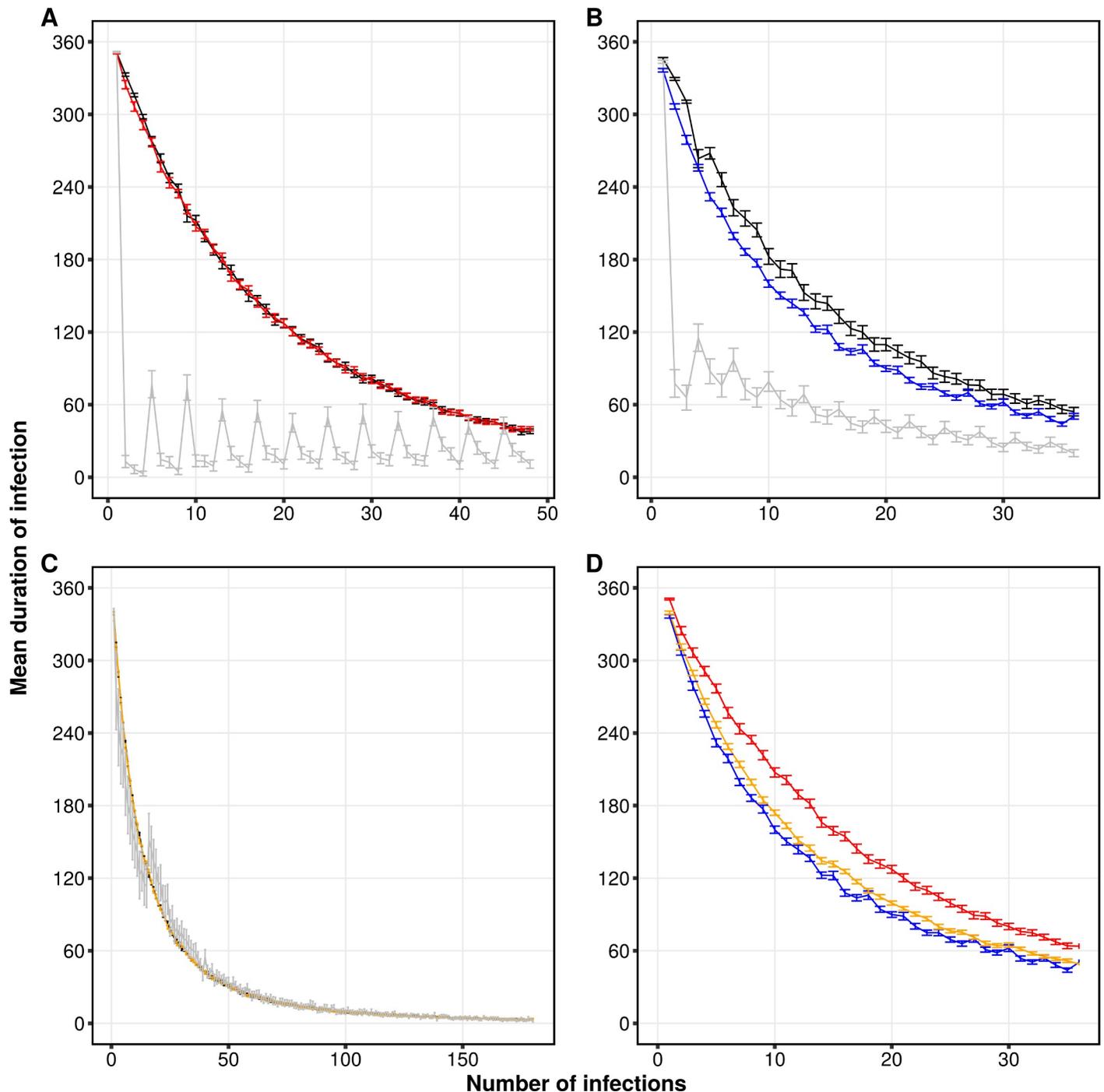
We find that the duration of infection declines at a faster rate when hosts are infected with repertoires originating from the same module compared to infections with repertoires from different modules (Fig 4A). In generalized immunity, we still observe this difference between the curves, though to a lesser extent (Fig 4B). Under complete neutrality, however, population structure has no effect on duration of infection (Fig 4C) because modules are extremely short-lived, and the difference in genetic diversity between repertoires that belong to either the same or different modules is low. This analysis indicates that the modular population structure generated by immune selection can exhibit epidemiological consequences. It also suggests that modules can be viewed as “strains” that are epidemiologically different from each other from the perspective of host immunity.

With this result in mind, we then explored the epidemiological consequence of structure resulting from immune selection in a more realistic way. In nature, infections can occur with repertoires belonging or not to the same module. We therefore additionally sampled repertoires uniformly at random, regardless of their module affiliation (colored curves in Fig 4A, 4B and 4C). Under selection, this generates a curve that overlaps with that of the between-module infections because under high diversity, limiting similarity pushes modules to be as different as possible from each other. A direct comparison of these curves (Fig 4D) shows that in populations structured by immune selection, the duration of infection is generally higher than in populations assembled by stochastic processes alone (the red curve is above the other two). This is because competition drives repertoires to be as different from each other as possible, making the genetic diversity within a repertoire higher on average under immune selection.

An exponential fit of the form  $d = ae^{-bi}$  (where  $d$  is the duration of infection and  $i$  is the number of accumulated infections) to these curves for the first 50 infections indicates an approximately 28% and 15% faster decline in  $d$  under generalized immunity and complete neutrality, respectively, compared to immune selection ( $b_{selection} = 0.0502 \pm 0.000159$  SE;  $b_{generalized} = 0.0647 \pm 0.000234$  SE;  $b_{neutral} = 0.0577 \pm 0.000129$  SE). These model fits, which consider the entomological inoculation rate (EIR) of each of the scenarios, show a consistently slower reduction in duration of infection under selection than under neutrality. For example, the predicted duration of a new infection in a 5-month-old infant would be 163, 131, and 5 days under selection, generalized immunity, and complete neutrality, respectively (S13 Fig).

### Discussion

We find that competitive interactions between genomes for hosts act as a stabilizing force that dynamically organizes the hyperdiversity of *var* gene repertoires into temporally persistent



**Fig 4. Epidemiological consequences of repertoire population structure under high *var* gene diversity.** The figure shows the decline in duration of infection as a function of the number of infections accumulated in a naive host during a 1-year period. Panels (A), (B), and (C) correspond respectively to selection, generalized immunity, and complete neutrality. In any given simulation, a naive host was infected with repertoires originating either from the same module (within-module infections; gray), from different modules (between-module infections; black), or randomly (colored; red: immune selection, blue: generalized immunity; orange: complete neutrality). Each point is the average duration of infection across 50 runs of the ABM for 10 hosts with 5 random starting layers in 10 different modules (see [Methods](#) and [S1 Text](#) for details on experimental design). We are interested in the comparison of the decreasing trends for the duration of infection across scenarios. The small fluctuations overlaid on these trends reflect intermittent increases in this duration (they are the result of the way infections were sampled over discrete time steps, by advancing from one layer to the next, in the design of the simulations; see [S1 Text](#) for details). The number of infections in each layer was determined as a function of the EIR of the ABM. (D) A direct comparison of the curves for random infections from panels (A)–(C). The curve of immune selection is above that of the neutral scenarios because under immune selection, repertoires consist of a lower proportion of identical alleles than under neutrality. In all panels, error bars are 95% confidence intervals. Data from model output used to produce the figure can be found in [https://figshare.com/articles/Figure\\_data/8079713](https://figshare.com/articles/Figure_data/8079713). ABM, agent-based model; EIR, entomological inoculation rate.

<https://doi.org/10.1371/journal.pbio.3000336.g004>

modules. These clusters reflect dynamic “niches” emerging in the immune space of the human host population. By considering the temporal dimension, we therefore recover the clustered nature characteristic of stabilizing competition/balancing selection, which was not evident in static analyses at high gene diversity [15]. This clustering is also evident on a shorter time scale for field data. Thus, immune selection acts to promote persistence of gene combination clusters over epidemiological time, in addition to persistence of the genes themselves over evolutionary time [2, 30].

In ecology, stabilizing competition between species in ecosystems and operational taxonomic units in microbiomes was also shown to promote coexistence and the formation of clusters, albeit for much lower-dimensional trait spaces [7–10]. Here, modules are effectively niches in human immune space, formed by competition for a substitutable resource (naive immune systems). Hence, competition acts to create and fill available niches dynamically [31]. Conceptual similarities and differences in deviations from neutral expectations and in the formation and detection of niches should enrich our understanding of the role of interactions in ecological and epidemiological systems.

Earlier strain theory proposed the stable coexistence of genetically discordant antigenic sets motivated by several pathogens, including malaria, H3N2 influenza, and meningococcus [32–34]. This predicted pattern proved robust in a variety of models depending on the strength of competition (and also the speed of evolution [34]). Unlike earlier computational implementations of strain theory (for example, MANTIS [35]), our ABM is stochastic, incorporates evolutionary innovation, and considers sequential presentation of antigenic variants to the immune system. This formulation was needed to reach diversity comparable to that observed in nature, especially under high endemism. For the large diversity of multigene families such as *var*, neither the complete nor the partial discordance of antigenic repertoires from simpler models is evident.

While for pathogens such as influenza, genetically discordant antigenic sets can represent different strains [33, 34], for malaria, it remains unclear what constitutes a “strain” in regions where antigenic diversity, transmission, and recombination are all high. The genetic coherence apparent in the emergence of persistent niches in antigenic space suggests that the modules we have identified can play the role of strains. As such, strains are dynamic entities arising from the interaction between parasite genomes and the human immune system. Although generated by very different mechanisms, these modules are conceptually similar to viral quasispecies in which high mutation rates and negative selection by host immunity create continuously changing repositories of viral variants, which are the source of virus adaptability [36].

The modular structure we find in the immune selection scenario implies a slower acquisition of immunity by individual hosts (compared to neutrality), demonstrating the epidemiological and evolutionary advantages of immune selection for the parasite. Duration of infection is a key epidemiological parameter, which together with transmission rate, ultimately determines  $R_0$ . There is clear empirical evidence for the relationship between PfEMP1 expression and duration of infection. As the major antigen of blood stage, PfEMP1 elicits variant-specific immune responses that increases with age and exposure [37]. In vivo studies [38, 39] show that the whole population of parasites inside a host usually expresses only one or a few dominant *var* genes, with a moderate switching rate per generation. Although the exact genotype–phenotype map of the Duffy binding-like-alpha (DBL $\alpha$ ) domain analyzed in our empirical data is unknown, this domain is known to help bind to complement receptor (CR)1 and AB blood group antigens on other erythrocytes and form rosetting [40]. Thus, variation of this domain within PfEMP1 is relevant to understanding the chronic nature of malaria infections. The relationship between sequential expression and duration of infection is one mechanism resulting in frequency-dependent selection. We conjecture that regardless of the specific

mechanism, a temporal modular structure would emerge when frequency-dependent selection is operating, even in models in which all antigens are seen by the host at once.

Existing epidemiological models of malaria are able to incorporate aspects of repeated infection but neither explicit strain diversity per se nor its overlap structure [41]. How to account for these aspects in transmission models that remain sufficiently parsimonious for epidemiological application remains an open question. Consideration of dynamic strains (or modules) provides one way forward even though the definition of modules may be context-dependent (for example, depend on model parameters). In Senegal, where extensive control efforts have shrunk the parasite population considerably, parasites from the same genetic clone (defined using neutral markers) express similar dominant *var* genes and have similar responses to growth-inhibitory antibodies [42]. These findings appear consistent with our theoretical results, although follow-up work should consider responses to intervention.

Incorporating diversity and its structure may prove particularly relevant to modeling interventions. Control efforts can lower diversity either directly by clearing infections (for example, antimalarial drugs) or through reduction in transmission by targeting the vector (for example, bed nets or IRS). Our simulations and data clearly show that bottlenecks to transmission (and associated generation of *var* diversity), such as seasonality, limit the persistence of modules, while the replenishment and the restructuring of diversity during the following wet season create new ones (S9 Fig). Immune selection can counter the effects of seasonality and/or interventions by promoting persistence of modules across these bottlenecks (S10 Fig). This kind of resilience implies that upon lifting the intervention, the remaining diversity should be able to quickly restructure itself to occupy available niches in immune space. A recent study in Northern Ghana following 7 years of IRS has shown a consistent reduction in EIR and sporozoite rates compared to a nearby control site [43]. Yet, the same study also showed that upon removing the IRS intervention, EIR increased from 30 to 90 infectious bites per person per month in only 2 years. Monitoring *var* gene diversity and its structure may provide further insight, as a reduction in epidemiological parameters per se may not be a sufficient condition for elimination. Previous arguments on multiple fixed strains and  $R_0$ , coupled with our findings on structure and infection duration, suggest the existence and raise the open question of a threshold in antigenic (and associated genetic) diversity in the transition to elimination.

One possible limitation of the approach taken here is the assumption of functional equivalence of *var* genes in the model. Previous studies have shown that variation in upstream sequences [40] or homology blocks [17] result in functional differences among *var* genes. This could affect the signature of immune selection. We therefore limited our analyses of empirical data to groups of upstream promoter (ups)B/C to ensure most *var* variants are by and large functionally equivalent. In the future, we will expand our theoretical and computational approach to include variation in virulence or other functions across *var* genes. Another limitation is the potential sensitivity to the community detection method we used. While this can, in principle, be true, the patterns we obtain are sensible from a theoretical perspective. Also, on the technical side, there is currently no other method available to perform community detection in directed and weighted networks with directed interlayer edges in large networks. Because quantitative network properties such as the number of modules can be sensitive to the exact parameterization of the algorithm's implementation or to the cutoff imposed on network edge weights, we emphasize that we are primarily interested in the qualitative differences between the immune selection and neutral scenarios.

From this qualitative perspective, temporal multilayer networks present an opportunity to extend the scope of current phylodynamic theory, which relates the epidemiology and transmission of pathogens to phylogenetic structure [44, 45]. Because trees are a particular case of networks, the multilayer network approach we devise here can extend traditional

phylodynamic analysis to pathogens evolving through recombination. As in phylodynamics, the structure of our network reveals information about the interplay between selective forces and transmission dynamics. There is, in particular, a clear difference in structure between the immune selection and neutral scenarios. We also find differences in network structure between regimes with different diversity, whereby low diversity is characterized by a replacement regime instead of module coexistence (S1 Text)—patterns that reflect differences in transmission dynamics and evolutionary rates between these regimes.

Our findings are generally relevant for pathogens characterized by large antigenic spaces. One extreme example is found in *Trypanosoma brucei*, the causal agent of African sleeping sickness. Its genome contains about 1,000 copies of the variant surface glycoprotein gene, the vast majority of which are pseudogenes. This enormous within-genome diversity provides a clear advantage to the parasite, allowing it to assemble new active genes from pseudogenes via gene conversion once it has exhausted its functional copies during an infection [46]. Other examples are found in the fungus *Pneumocystis carinii* and in the bacterium *Neisseria gonorrhoeae* [3]. Beyond multicopy family genes, the theory and analyses we have developed should also apply to sets of genes encoding different antigens.

Competition for hosts is a powerful feature of parasites' life history, enabling the emergence and maintenance of a dynamical strain structure across bottlenecks of transmission and affecting the rate at which host immunity is gained. For malaria, understanding the consequences of parasite population structure for elimination constitutes an important direction for future research.

## Methods

### Ethics statement

The study was reviewed and approved by the ethics committees at the Navrongo Health Research Centre, Ghana (approval #NHRC IRB-131); Noguchi Memorial Institute for Medical Research, Ghana (approval #NMIMR-IRB CPN 089/11-12); University of Melbourne, Australia (approval #HREC 144-1986); and the University of Chicago, United States (approval #IRB 14-1495).

### ABM

The ABM is flexible enough to include seasonality and methods for malaria control to compare simulated dynamics to empirical data (see below). We used an improved implementation of the ABM developed by [15]. A complete description of the model can be found there. Here, we briefly describe the main elements of the model.

**Structure of human and parasite populations.** The human population in the ABM has a fixed size of  $H = 10,000$ , with a given age structure described by an exponential distribution with a mean life span of 30 years. Hence, whenever a host dies, another is born. The genome of one individual *P. falciparum* parasite was considered as a repertoire consisting of a set of 60 not necessarily unique *var* genes,  $R = \{g_1, g_2, \dots, g_l\}$  where  $1 \leq l \leq 60$  is the index of the gene within the repertoire and  $g \in U[1, G]$  is the ID of the gene in a pool of  $G$  genes. Each *var* gene was composed of two loci, corresponding to two epitopes. We refer to the variants of each locus as alleles hereafter. This model of *var* genes is supported by empirical evidence for two hypervariable regions bounded by three semiconserved regions [47, 48]. Hence, each gene is defined as  $g = \{a_1^k, a_2^m\}$ , where the superscripts  $k, m \in U[1, A]$  are the ID of the allele in a pool of  $A$  alleles (the two epitopes have separate pools that are equal in size) and the subscript defines the epitope. A focal population of the parasite is initialized from an external gene pool

representing regional diversity. New genomes with randomly selected genes from the regional gene pool enter the system with an immigration rate of 1 genome per day, and a random host is selected and infected with the new genome. Each simulation starts with 20 infected hosts, each with one repertoire selected at random from the pool.

**Transmission dynamics and seasonality.** We set a biting rate  $b$  so that the average waiting time to the next biting event is equal to  $1/(b \times H)$ . When a biting event occurs, two hosts are randomly selected, one donor and one recipient. If the donor has infectious parasite repertoires, the receiver will be infected with a probability  $p$  (the transmission probability). If the donor is infected with multiple repertoires in the blood stage, then the transmission probability of each strain is reduced to  $p/I$ , where  $I$  is the number of repertoires.

We implemented seasonality by multiplying  $b$  by a temporal vector  $v$  of length 360 (days), containing the daily number of mosquitoes, resulting in a temporal vector of daily biting rates  $v = v \times b$ . To obtain  $v$ , we used the deterministic model for mosquito population dynamics from [49]. The model was originally developed for *Anopheles gambiae* and consists of a set of ODEs describing the dynamics of 4 mosquito stages: eggs, larvae, pupae, and adults. Seasonality is implemented via density dependence at the egg and larva stages as a function of rainfall (availability of breeding sites). The model can also handle IRS interventions for comparison to empirical data. We emphasize that the values in the biting rate vector  $v$  rather than the absolute number of daily mosquitoes are the key parameter for the purposes of this work, used to match the EIR observed in the study area to that of the simulations. Because implementing seasonality requires some knowledge on the dynamics of mosquitoes and on the resulting EIR, we only implemented seasonality for the high-diversity regime. Details on the model can be found in [49], and our implementation (in Mathematica) at the GitHub repository.

**Mechanisms of genetic change.** During the sexual stage of the parasite (within mosquitoes), different parasites can exchange *var* repertoires through meiotic recombination. The receiver host can receive either recombinant repertoires, original repertoires, or a combination. During the asexual stage of the parasite (blood stage of infection), *var* genes within the same repertoire can exchange epitope alleles through mitotic (ectopic) recombination (at a rate of  $\rho = 1.87 \times 10^{-7}$  per gene pair per day [23]). Also, new alleles of each epitope can be generated through mutations at a rate of  $\mu = 1.42 \times 10^{-8}$  per day. Mitotic recombination, mutation, and immigration generate new *var* genes [23].

**Within-host dynamics.** Each repertoire is individually tracked through its entire life cycle, encompassing the liver stage, asexual blood stage, and the transmission and sexual stages explicitly or implicitly. Because we do not explicitly model transmission via mosquitoes, we delay the expression of each strain in the receiver host by 14 days to account for the time required for the sexual stage (in the mosquito) and the liver stage (in the host). Specifically, the infection of the host is delayed 7 days to account for the time required for gametocytes to develop into sporozoites in mosquitoes. When a host is infected, the parasite remains in the liver stage for an additional time before being released as merozoites into the bloodstream, invading red blood cells, and starting the expression of the *var* repertoire.

During the expression of the repertoire, the host is considered infectious with the active repertoire. The expression order of the repertoires is randomized for each infection, while the deactivation rates of the *var* genes is controlled by the host immunity. When one gene is actively expressed, host immunity “checks” whether it has seen any of its epitope alleles in the infection history. In the immune selection model, the deactivation rate changes so that the duration of active period of a gene is proportional to the number of unseen alleles. A gene deactivates at a rate of 1/6 per day in a naive host. After the gene is deactivated, the host adds its deactivated alleles to the immunity memory. A new gene from the repertoire then becomes immediately active. The repertoire is cleared from the host when the whole repertoire of *var*

genes is depleted. The immunity towards a certain epitope allele wanes at a rate  $w = 1/1,000$  per day [50].

## Scenarios

In the immune selection scenario, the duration of infection in a host depends on the history of infection with given *var* genes (or their alleles) as described above. The model of “complete neutrality” retains the process of transmission but does not consider any aspect of infection history. Hence, individuals clear infection after a given amount of time (matched to the average duration of infection from counterpart simulations of the selection scenario). In the neutral model of “generalized immunity,” the duration of infection depends on the number of previous infections but not on their specific genetic composition. For a meaningful comparison, we parameterized this model to match the curve for duration of infection with previous number of infections of the immune selection scenario [15].

In both neutral models, we ran the exact same ABM implementation of the immune selection scenario, conserving all its parameters except for the duration of infection. Specifically, we used  $G = 12,000$  and  $b = 0.5$  in all simulations in the high-diversity regime (medium diversity:  $G = 1,200$ ,  $b = 0.2$ ; low diversity:  $G = 120$ ,  $b = 0.1$ ). The duration of infection of a repertoire in a naive host was  $D = 360$  days (i.e., 1 year). Because we know that in nature, hosts build specific immunity to malaria infections, the neutral models can be regarded as process-based null models against which to test the structure obtained with the full model (with immune selection) and that of the empirical data. These null models lack immune selection/competition for hosts via specific immunity and have either no competition in the complete neutrality formulation or general competition for the overall host population in the generalized immunity one. For each combination of diversity regime (low, medium, high) and immune scenario (immune selection, complete neutrality, and generalized immunity), we analyzed data generated by 50 simulations of the ABM.

## Construction of temporal networks

We calculated genetic similarity of repertoire  $i$  to repertoire  $j$  as  $S_{ij} = (N_i \cap N_j) / N_i$ , where  $N_i$  and  $N_j$  are the number of unique alleles for repertoires  $i$  and  $j$ , respectively (the genetic similarity of repertoire  $j$  to repertoire  $i$  was calculated as  $S_{ji} = (N_i \cap N_j) / N_j$ ). We used a directional index because of the asymmetric competition resulting from different numbers of unique alleles in a repertoire [15]. We used the same index ( $S$ ) for both intra- and interlayers because it situates the inter- and intralayer edges on the same scale, which is crucial when looking for optimal partitioning in multilayer networks [51, 52]. To optimize the signal-to-noise ratio in our analysis, we imposed a cutoff on edge weights (S1 Text and S3 Fig).

## Community detection

To capture the organization of the population into groups of highly similar repertoires, we used the map equation as an objective function to calculate the optimal partition of the network. Briefly, the map equation is a flow-based and information-theoretic method (implemented with Infomap) recently extended to multilayer networks that calculates network partitioning based on the movement of a random walker on the network (see [26–28] for details). In any given partition of the network, the random walker moves across nodes in proportion to the direction and weight of the edges. Hence, it will tend to stay longer in dense areas where there are many repertoires similar to each other. These areas can be defined as “modules.” Note that for convenience, we use the term “modules” because it is commonly used to refer to partitions of networks, but Infomap does not calculate a modularity function

(sensu [53]). The time spent in each module can be converted to an information-theoretic currency using an objective function called the map equation. The best network partition corresponds to that with the minimum value of the map equation [26, 27]. This method has been applied to describe temporal flows in networks that do not have interlayer edges [54]. In our particular network, once the random walker moves along an interlayer edge, it cannot go back, capturing the temporal flow in the network.

### Quantification of structure

We calculated the relative persistence of repertoires and modules as  $\mathcal{P} = (l_d - l_b)/(l_{max} - l_b)$ , where  $l_b$  and  $l_d$  refer to the layers in which a module (or repertoire) first appeared or died, respectively and  $l_{max}$  is the maximum number of layers ( $l_{max} = 300$  in our simulations). This index has a maximum theoretical value of 1 when a module (repertoire) persists from the layer when it was generated till  $l_{max}$  and a theoretical minimum of  $1/(l_{max} - l_b)$  when a module (repertoire) persists for a single layer. We used the relative persistence instead of the plain number of layers in which a module (repertoire) persisted because persistence can be truncated artificially by reaching the end of the simulation. For example, a repertoire could have persisted for 60 layers because it appeared in layer 100 and died in layer 160 ( $\mathcal{P} = 0.3$ ) or because it appeared in layer 240 and died only because it reached the end of the simulation at  $l_{max} = 300$  ( $\mathcal{P} = 1$ ). In the former case, the repertoire did not live to its full potential (200 layers) and thus has lower relative persistence than the second one. We also examined the results for absolute persistence (i.e., the number of layers), and these do not qualitatively change the conclusions.

We calculated  $J$ , the evenness in module size in each layer. We chose evenness instead of simply considering the number of repertoires in a module because the number of repertoires is not comparable between scenarios and because from a theoretical perspective, limiting similarity resulting from selection should promote an even distribution of repertoires across modules. We calculated  $J$  as  $J = H'/H_{max}$  [55].  $H' = -\sum_{i=1}^m p_i \ln(p_i)$  is Shannon's diversity, where  $m$  is the total number of modules in the layer and  $p_i$  is the proportion of repertoires in module  $i$  in that layer.  $H_{max} = \ln(m)$  is the maximum diversity possible, obtained when all modules have the same size [55].  $J$  varies between 0 and 1, where 1 indicates a perfect distribution of module sizes such that each repertoire is as different from any other as possible.

### Empirical data

The empirical data were collected from a study of two catchment areas in BD, Ghana, located in the Upper East Region near the Burkina Faso border. This age-stratified serial cross-sectional study was done across 6 Svs (Sv1–Sv6), conducted at either the EWS or EDS: October 2012 (EWS, Sv1), May–June 2013 (EDS, Sv2), May–June 2014 (EDS, Sv3), October 2014 (EWS, Sv4), October 2015 (EWS, Sv5), and May–June 2016 (EDS, Sv6). Each survey lasted approximately 3–4 weeks. Details on surveys are in [S1 Table](#). During the sampling period, three rounds of IRS intervention were performed: between October–December 2013, between May–July 2014, and between December 2014 and Feb 2015, using organophosphates. We sequenced and genotyped part of the *var* DBL $\alpha$  domain, which is a reliable marker for differentiating *var* genes.

*Var* genes can be divided into subgroups based on their chromosomal location and semi-conserved ups sequences: groups A (upsA), B (upsB), C (upsC), and E (ups E). [56]. Different ups *var* gene groups are associated with differences in disease virulence and sequence conservation; in particular, upsA *var* genes are more conserved (i.e., have lower genetic variation at the population level) and are preferentially expressed in children with severe and/or clinical malaria compared to upsB/upsC *var* genes. Hence, DBL $\alpha$  reads were clustered at 96% pairwise

identity, translated into all 6 reading frames, and classified into either upsA or upsB/upsC (i.e., non-upsA) groups [19]. Details on the study population, data collection procedures, molecular and genetic work, and epidemiology have been published elsewhere [15, 29]. While the *var* gene data from Sv1 and Sv2 were used in [15], data from Sv3–Sv6 are used here for the first time and are summarized in S1 Table.

### Comparison of empirical and simulated benchmark networks

The DBL $\alpha$  tag region consists of two hypervariable regions and one conserved region between the S2 and S3 domains [48]. Therefore, we split each tag sequence before and after the HB5 block into two hypervariable regions using USEARCH blast and considered them as two epitopes. We clustered each epitope sequence into different alleles using 80% similarity to match the mean number of alleles in a layer in the ABM simulations. As in our simulations, we constructed each layer in the empirical network based on pairwise similarities of repertoires consisting of different alleles of the two epitopes. Only upsB/upsC DBL $\alpha$  tags were included because upsA genes tend to be more conserved and thus have disproportionately higher sharing rates among repertoires [19]. Since infections with multiple parasite genomes (multiplicity of infection [MOI] > 1) are very common in malaria-endemic regions, we selected isolates with a total number of upsB/upsC DBL $\alpha$  types ranging from 40–55 copies to maximize the probability of selecting hosts with single-genome infections (MOI = 1). This resulted in 90, 68, 69, 52, 115, and 44 isolates in layers 1–6, respectively. We defined edge weights using the same index as in the theoretical work and used a cutoff of 97.8% (S11 Fig and S12 Fig).

We note, based on our preliminary analyses, that constructing an empirical network is not as straightforward as for the benchmark theoretical networks because of the uneven sampling interval between the surveys. One way to account for uneven sampling is to rescale the interlayer edges. Any rescaling entails, however, assumptions about processes underlying repertoire persistence. One approach based on genetic drift would require knowledge on the population demography of the parasite, which we do not have. Another would rely on the time scale of the processes influencing similarities between repertoires that include the consecutive sampling of the same repertoires across time, the generation of a new copy of a repertoire, and the similarity between recombinant repertoires. The relative importance of these processes is difficult to estimate. On the more technical side, rescaling may place the intra- and interlayer edges on different scales, which could impose different weights in how the clustering algorithm (Infomap) treats the relative importance of similarity between and within layers [52]. Finally, rescaling is applied only to interlayer edges, giving them a different interpretation than that of intralayer edges. Given these limitations, we opted for not rescaling the interlayer edges, especially because the relatively long persistence of epitope alleles (66% persist for 3 or more layers) reduces the need to account for uneven sampling periods, making assumption-dependent corrections unnecessary.

We ensured that our results of the theoretical part were qualitatively the same when seasonality was considered. To imitate the empirical setting, we ran 50 ABM simulations with  $G = 35,000$  and a vector of biting rates  $\nu$  that generated EIR values similar to those observed in the field (S2 Fig and S8 Fig). Because of the uncertainty in realistic parameter values of the ABM for the given field sites, we based our comparisons to empirical data on 100 simulations that span plausible variation across the two main parameters:

1. Average biting rate,  $b$ . We tested 10 sequential values of  $b$ : 0.00010, 0.00012, 0.00014, 0.00017, 0.00019, 0.00021, 0.00023, 0.00026, 0.00028, and 0.00030. In our simulations in the main text,  $b = 0.00002$ . Remember that effective biting rate is obtained by  $\nu = \nu \times b$  (see description of the ABM above); we kept  $\nu$  constant.

2. Size of the *var* gene pool,  $G$ . We tested 10 sequential values of  $G$ : 30,000, 31,111, 32,222, 33,333, 34,444, 35,556, 36,667, 37,778, 38,889 and 40,000. In our simulations in the main text,  $G = 35,000$ .

Beyond seasonality, we included the IRS by matching the relative differences in biting rates to those obtained in the field. We accounted for the three rounds of IRS by reducing the biting rates in the mosquito model (see description of the ABM above) to obtain EIR values of approximately 0, as was recorded in the field, without collapsing the mosquito population. In each of the 100 ABM runs, we selected 6 layers corresponding to the 6 months in which surveys were conducted in the field (encompassing the interventions). We then subsampled these layers randomly to obtain the same number of repertoires as in each layer of the empirical network. We used *var* epitopes as the unit of comparison to define edge weights, as was done for the empirical networks and the theoretical work. In these 100 networks, we imposed the same 97.8% cutoff that was used in the empirical network (S11 Fig). Given the sensitivity of the analysis to the particular way one builds the empirical network (and consequently, the counterpart theoretical networks), one may fail to find differences in the temporal modular structure between generalized immunity and immune selection given short and uneven longitudinal sampling.

For any module that was “born” in layer 3 or before, we calculated the probability that it persisted for at least 3 layers using a logistic regression implemented in the function `glm` with a binomial family in R. Because our explanatory variable is categorical with 4 levels (empirical data and three scenarios), the model has a set of “dummy variables” and can be written as  $\ln(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$ , where  $y$  is 1 if the module persisted or 0 otherwise,  $\beta_0$  is the coefficient for the empirical data, and  $\beta_1, \beta_2, \beta_3$  are the coefficients for immune selection, generalized immunity, and complete neutrality, respectively.

## Simulations of host infection

We simulated repeated infections of naive hosts by sampling repertoires originating in the same module, in different modules, or regardless of modules. We compared the resulting curves depicting the decrease in the duration of infection as a function of the number of previous infections for each of these dynamics. In these simulations, infection takes place within a layer for a given number of bites determined by a given force of infection and then moves on to the next layer. In this way, we explicitly consider different times. Details on the simulations are in S1 Text.

## Supporting information

**S1 Text. Additional details on methods and results.** The text includes methodological details on selection of cutoff values and the epidemiological simulations, results for low- and medium-diversity regimes, and a sensitivity analysis.  
(PDF)

**S1 Fig. Characterization of diversity regimes.** The left, middle, and right columns represent the low-, medium-, and high-diversity regimes, respectively. Each row depicts one parameter relevant for epidemiology or genetic diversity. Whereas parameters that are set a priori in the model (number of bites, number of alleles) have similar values between scenarios, emerging parameters (for example, the number of repertoires, prevalence, or MOI) can differ. These regimes apply to the simulations in the theoretical analyses. Data are for 50 ABM simulations. Data from model output used to produce the figure can be found in [https://figshare.com/articles/Figure\\_data/8079713](https://figshare.com/articles/Figure_data/8079713). ABM, agent-based model;  $G$ , generalized immunity; MOI, multiplicity of infection;  $N$ , complete neutrality;  $S$ , selection.  
(PDF)

**S2 Fig. EIR in ABM simulations.** Panels (A) and (B) show the EIR for simulations without and with seasonality, respectively, for the 3 scenarios: immune selection (red), generalized immunity (blue), and complete neutrality (orange). In (A), each panel depicts a different transmission regime. Because there is no seasonality, EIR does not vary across months. In (B), values are for simulations with seasonality in the high transmission regime only, and each panel depicts a different scenario. Data from model output used to produce the figure can be found in [https://figshare.com/articles/Figure\\_data/8079713](https://figshare.com/articles/Figure_data/8079713). ABM, agent-based model; EIR, entomological inoculation rate.

(PDF)

**S3 Fig. Determination of cutoff values for networks.** The left, middle, and right columns correspond to the low (teal), medium (purple), and high (green) diversity regimes, respectively, in the immune selection scenario. (A–I) Change in values of three network properties as a function of cutoff percentile (see [Methods](#) for details on how we calculated these properties). In each cutoff value, we draw a box plot and mean (red point) for values across 10 simulations of the ABM. We selected cutoff percentiles of 0.3, 0.6, and 0.85 in the low, medium, and high regimes, respectively. The same percentiles were used in the counterpart neutral and generalized immunity model scenarios. Data from model output used to produce the figure can be found in [https://figshare.com/articles/Figure\\_data/8079713](https://figshare.com/articles/Figure_data/8079713). ABM, agent-based model.

(PDF)

**S4 Fig. Temporal population structure in the low-diversity regime.** The left, middle, and right columns represent the selection (red), generalized immunity (blue), and neutral (orange) scenarios, respectively. (A,B,C) Example of population structure from one run of the ABM. Each line on the *y*-axis corresponds to a different module, and its horizontal length depicts its occurrence across layers. Modules are generated and die out. All scenarios are characterized by replacement dynamics whereby one module persists until it is replaced by another, possibly following a short transition period. (D,E,F) These dynamics result in module persistence (colored density plots), which is shorter than that of repertoires (gray density plots) in all scenarios. The insets depict a comparison between the values in the colored and gray density plots using box plots. See [Methods](#) for details on how we calculated relative persistence. Results in (D)–(E) are for data pooled across 50 runs of the ABM. Data from model output used to produce the figure can be found in [https://figshare.com/articles/Figure\\_data/8079713](https://figshare.com/articles/Figure_data/8079713). ABM, agent-based model.

(PDF)

**S5 Fig. Temporal population structure in the medium-diversity regime.** The left, middle, and right columns represent the selection (red), generalized immunity (blue), and neutral (orange) scenarios, respectively. (A,B,C) Example of population structure from one run of the ABM. Each line on the *y*-axis corresponds to a different module, and its horizontal length depicts its occurrence across layers. Modules are generated and die out. Under complete neutrality, the persistence of repertoires is on the same scale as that of modules, as expected (F). However, there is no apparent difference in population structure and patterns of relative persistence between immune selection and generalized immunity (D,E) (see results for low- and medium-diversity regimes in [S1 Text](#) for details). In panels (D)–(F), the insets depict a comparison between the values in the colored and gray density plots using box plots. See [Methods](#) for details on how we calculated relative persistence. Results in (D)–(E) are for data pooled across 50 runs of the ABM. Data from model output used to produce the figure can be found in [https://figshare.com/articles/Figure\\_data/8079713](https://figshare.com/articles/Figure_data/8079713). ABM, agent-based model.

(PDF)

**S6 Fig. Evenness in module size.** Limiting similarity resulting from selection should promote an even distribution of repertoires across modules. We calculated the evenness  $J$  in each layer as  $J = H'/H_{max}$  [55], where  $H'$  is Shannon's diversity and  $H_{max}$  is the maximum diversity possible, obtained when all modules have the same size [55].  $J$  varies between 0 and 1, where 1 indicates a perfect distribution of module sizes such that repertoires are at the "limit" of their limiting similarity because each repertoire is as different from any other as possible. We compare between scenarios using box plots for the low- (A), medium- (B), and high-diversity regimes (C). Because immune selection does not operate in the low-diversity regime, the evenness is comparable across scenarios. By contrast to high diversity, competition under medium diversity is not strong enough to generate patterns that distinguish between immune selection and generalized immunity. (D) For high diversity, evenness is higher when there are more modules in a layer. Results are for data pooled across 50 runs of the ABM. Data from model output used to produce the figure can be found in [https://figshare.com/articles/Figure\\_data/8079713](https://figshare.com/articles/Figure_data/8079713). ABM, agent-based model; G, generalized immunity; N, complete neutrality; S, Immune selection. (PDF)

**S7 Fig. Sensitivity analysis for high-diversity regime.** We repeat the analysis of the theoretical part of the main text using 112 simulations for the immune selection (red) and generalized immunity (blue) scenarios. These 112 simulations encompass variation in three key parameters of the ABM. We find the same qualitative results for the differences between the two scenarios as in the main text. Specifically, the relative persistence of modules ( $\mathcal{P}$ ) is longer (A), and the evenness in module size is higher (B) under immune selection compared to generalized immunity. Data from model output used to produce the figure can be found in [https://figshare.com/articles/Figure\\_data/8079713](https://figshare.com/articles/Figure_data/8079713). ABM, agent-based model. (PDF)

**S8 Fig. Characterization of the seasonality diversity regime.** The left, middle, and right columns represent the selection, generalized immunity, and complete neutrality scenarios, respectively. Each row depicts one parameter relevant for epidemiology or genetic diversity. This regime is for the baseline simulations of seasonality, with  $G = 35,000$ , used to confirm that our nonseasonal theoretical results hold. Data are for 50 ABM simulations. Data from model output used to produce the figure can be found in [https://figshare.com/articles/Figure\\_data/8079713](https://figshare.com/articles/Figure_data/8079713). ABM, agent-based model. (PDF)

**S9 Fig. Seasonal signature in the formation of modules.** Seasonality promotes the extinction of repertoires in the dry season and the generation of new ones in the wet season. This bottleneck causes a decrease in the creation of new niches, translating to a decrease in the number of modules in the dry season. Data are pooled across 50 runs of the ABM. Error bars represent 95% confidence intervals. Red: immune selection; blue: generalized immunity; orange: complete neutrality. Data from model output used to produce the figure can be found in [https://figshare.com/articles/Figure\\_data/8079713](https://figshare.com/articles/Figure_data/8079713). ABM, agent-based model. (PDF)

**S10 Fig. Temporal population structure in the high-diversity regime with seasonality.** The left, middle, and right columns represent the selection (red), generalized immunity (blue), and neutral (orange) scenarios, respectively. (A,B,C) Example of representative population structure from one run of the ABM. Each line on the  $y$ -axis corresponds to a different module, and its horizontal length depicts its occurrence across layers. Modules are generated and die out. (D) The selection scenario is characterized by modules (red) that persist for much longer than

the repertoires (gray) that compose them. The inset is a comparison between the values in the red and gray density plots using box plots. (E,F) In the neutral scenarios, modules and repertoires have similar relative persistence. See [Methods](#) for details on how we implemented seasonality and calculated relative persistence. Results in (D)–(E) are for data pooled across 50 runs of the ABM. Data from model output used to produce the figure can be found in [https://figshare.com/articles/Figure\\_data/8079713](https://figshare.com/articles/Figure_data/8079713). ABM, agent-based model.

(PDF)

**S11 Fig. Exploration of cutoff for empirical networks.** Because empirical networks are not necessarily identical to those produced in our theoretical work (for example, in size and genetic diversity), we explored the effect of thresholding. It is clear from the figure that cutoff does not have a major effect on the distribution of module persistence. We chose 97.8% because this was the point of transition at which modules started to persist for 5 layers. Data used to produce the figure can be found in [https://figshare.com/articles/Figure\\_data/8079713](https://figshare.com/articles/Figure_data/8079713).

(PDF)

**S12 Fig. Comparison of edge weight distributions between empirical and simulated networks.** We assess how well our simulations for the immune selection scenario (red) matched the empirical data (purple) by comparing the distributions of edge weights. The lower and upper panels depict the distributions for the intra- and interlayer edges, respectively. Vertical lines represent the values where a cutoff of 97.8% was applied. Data from model output used to produce the figure can be found in [https://figshare.com/articles/Figure\\_data/8079713](https://figshare.com/articles/Figure_data/8079713).

(PDF)

**S13 Fig. Epidemiological model curve fits.** The lines depict the predicted values for duration of infection as a function of host age. The values were obtained for the exponential model  $d = ae^{-bi}$  (where  $d$  is the duration of infection and  $i$  is the number of previous infections) that was fitted to the data from [Fig 4D](#) in the Main text. We used the EIR of each scenario and the number of accumulated infections to calculate host age. Curves of the standard error overlap with the curves of predicted values and are not presented (the standard error of model coefficient is very small). Red: Immune selection; blue: generalized immunity; orange: complete neutrality. Data used to produce the figure can be found in [https://figshare.com/articles/Figure\\_data/8079713](https://figshare.com/articles/Figure_data/8079713). EIR, entomological inoculation rate.

(PDF)

**S1 Table. Summary of empirical data relevant for this study.** Extended details are in [15, 29]. Isolates are the number of people in which the parasite was detected by microscopy. Unique *var* types is the total number of genes obtained from the DBL $\alpha$  marker with the standard 96% similarity cutoff. Alleles is the number of alleles from the UpsBC *var* gene group that were detected in isolates with a single infection (MOI1; defined as isolates from which 40–55 *var* UpsBC genes were sampled). The last two columns describe the data used in the empirical analysis in the main text, with MOI1 being the number of nodes. DBL $\alpha$ , Duffy binding-like-alpha; MOI, multiplicity of infection; ups, upstream promoter.

(PDF)

## Acknowledgments

We thank Martin Rosvall and Daniel Edler for advice and support with Infomap. We thank the participants, communities, and the Ghana Health Service in BD, Ghana for their willingness to participate in this study. We would also like to thank the personnel at the Navrongo Health Research Centre for sample collection and parasitological assessment/expertise. We are

grateful to Abraham R. Oduro, Anita Ghansah, and Kwadwo Koram for their helpful input related to the field study and to Michael F. Duffy for insightful comments on an earlier version of the manuscript. We appreciate the support of the University of Chicago through computational resources of the Midway high-performance computing cluster.

## Author Contributions

**Conceptualization:** Shai Pilosof, Qixin He, Karen P. Day, Mercedes Pascual.

**Data curation:** Shai Pilosof, Qixin He, Kathryn E. Tiedje, Shazia Ruybal-Pesántez.

**Formal analysis:** Shai Pilosof, Qixin He.

**Funding acquisition:** Kathryn E. Tiedje, Karen P. Day, Mercedes Pascual.

**Investigation:** Shai Pilosof, Qixin He.

**Methodology:** Shai Pilosof, Qixin He, Kathryn E. Tiedje, Shazia Ruybal-Pesántez, Mercedes Pascual.

**Project administration:** Karen P. Day, Mercedes Pascual.

**Resources:** Kathryn E. Tiedje, Karen P. Day, Mercedes Pascual.

**Software:** Shai Pilosof, Qixin He.

**Supervision:** Mercedes Pascual.

**Validation:** Shai Pilosof, Karen P. Day.

**Visualization:** Shai Pilosof.

**Writing – original draft:** Shai Pilosof, Qixin He, Mercedes Pascual.

**Writing – review & editing:** Shai Pilosof, Qixin He, Kathryn E. Tiedje, Shazia Ruybal-Pesántez, Karen P. Day, Mercedes Pascual.

## References

1. Charlesworth D. Balancing selection and its effects on sequences in nearby genome regions. *PLoS Genet.* 2006; 2(4):e64. <https://doi.org/10.1371/journal.pgen.0020064> PMID: 16683038
2. Larremore DB, Sundararaman SA, Liu W, Proto WR, Clauset A, Loy DE, et al. Ape parasite origins of human malaria virulence genes. *Nat Commun.* 2015; 6:8368. <https://doi.org/10.1038/ncomms9368> PMID: 26456841
3. Deitsch KW, Lukehart SA, Stringer JR. Common strategies for antigenic variation by bacterial, fungal and protozoan pathogens. *Nat Rev Microbiol.* 2009; 7(7):493–503. <https://doi.org/10.1038/nrmicro2145> PMID: 19503065
4. Figueroa F, Günther E, Klein J. MHC polymorphism pre-dating speciation. *Nature.* 1988; 335(6187):265–267. <https://doi.org/10.1038/335265a0> PMID: 3137477
5. Chesson P. Mechanisms of maintenance of species diversity. *Annu Rev Ecol Syst.* 2000; 31(1):343–366. <https://doi.org/10.1146/annurev.ecolsys.31.1.343>
6. Adler PB, Hillerislambers J, Levine JM. A niche for neutrality. *Ecol Lett.* 2007; 10(2):95–104. <https://doi.org/10.1111/j.1461-0248.2006.00996.x> PMID: 17257097
7. Scheffer M, van Nes EH. Self-organized similarity, the evolutionary emergence of groups of similar species. *Proc Natl Acad Sci U S A.* 2006; 103(16):6230–6235. <https://doi.org/10.1073/pnas.0508024103> PMID: 16585519
8. Jeraldo P, Sipos M, Chia N, Brulc JM, Dhillon AS, Konkel ME, et al. Quantification of the relative roles of niche and neutral processes in structuring gastrointestinal microbiomes. *Proc Natl Acad Sci U S A.* 2012; 109(25):9692–9698. <https://doi.org/10.1073/pnas.1206721109> PMID: 22615407
9. D'Andrea R, Ostling A. Can Clustering in Genotype Space Reveal "Niches"? *Am Nat.* 2016; 187(1):130–135. <https://doi.org/10.1086/684116> PMID: 27277409

10. D'Andrea R, Riolo M, Ostling AM. Generalizing clusters of similar species as a signature of coexistence under competition. *PLoS Comput Biol*. 2019; 15:e1006688. <https://doi.org/10.1371/journal.pcbi.1006688> PMID: 30668562
11. Gupta S, Maiden MCJ, Feavers IM, Nee S, May RM, Anderson RM. The maintenance of strain structure in populations of recombining infectious agents. *Nat Med*. 1996; 2(4):437–442. <https://doi.org/10.1038/nm0496-437> PMID: 8597954
12. Gupta S, Day KP. A theoretical framework for the immunoepidemiology of *Plasmodium falciparum* malaria. *Parasite Immunol*. 1994; 16(7):361–370. <https://doi.org/10.1111/j.1365-3024.1994.tb00361.x> PMID: 7970875
13. Gupta S, Anderson RM. Population structure of pathogens: The role of immune selection. *Parasitol Today*. 1999; 4758(1994):497–501. [https://doi.org/10.1016/S0169-4758\(99\)01559-8](https://doi.org/10.1016/S0169-4758(99)01559-8)
14. Artzy-Randrup Y, Rorick MM, Day KP, Chen D, Dobson AP, Pascual M. Population structuring of multi-copy, antigen-encoding genes in *Plasmodium falciparum*. *Elife*. 2012; 1:e00093. <https://doi.org/10.7554/eLife.00093> PMID: 23251784
15. He Q, Pilosof S, Tiedje KE, Ruybal-Pesántez S, Artzy-Randrup Y, Baskerville EB, et al. Networks of genetic similarity reveal non-neutral processes shape strain structure in *Plasmodium falciparum*. *Nat Commun*. 2018; 9(1):1817. <https://doi.org/10.1038/s41467-018-04219-3> PMID: 29739937
16. Day KP, Artzy-Randrup Y, Tiedje KE, Rougeron V, Chen DS, Rask TS, et al. Evidence of strain structure in *Plasmodium falciparum* var gene repertoires in children from Gabon, West Africa. *Proc Natl Acad Sci U S A*. 2017; 114(20):E4103–E4111. <https://doi.org/10.1073/pnas.1613018114> PMID: 28461509
17. Rorick MM, Artzy-Randrup Y, Ruybal-Pesántez S, Tiedje KE, Rask TS, Oduro A, et al. Signatures of competition and strain structure within the major blood-stage antigen of *Plasmodium falciparum* in a local community in Ghana. *Ecol Evol*. 2018; 8(7):3574–3588. <https://doi.org/10.1002/ece3.3803> PMID: 29686839
18. Kirkman LA, Deitsch KW. Antigenic variation and the generation of diversity in malaria parasites. *Curr Opin Microbiol*. 2012; 15(4):456–462. <https://doi.org/10.1016/j.mib.2012.03.003> PMID: 22503815
19. Ruybal-Pesántez S, Tiedje KE, Tonkin-Hill G, Rask TS, Kanya MR, Greenhouse B, et al. Population genomics of virulence genes of *Plasmodium falciparum* in clinical isolates from Uganda. *Sci Rep*. 2017; 7(1):11810. <https://doi.org/10.1038/s41598-017-11814-9> PMID: 28924231
20. Freitas-Junior LH, Bottius E, Pirrit LA, Deitsch KW, Scheidig C, Guinet F, et al. Frequent ectopic recombination of virulence factor genes in telomeric chromosome clusters of *P. falciparum*. *Nature*. 2000; 407(6807):1018–1022. <https://doi.org/10.1038/35039531> PMID: 11069183
21. Duffy MF, Byrne TJ, Carret C, Ivens A, Brown GV. Ectopic recombination of a malaria var gene during mitosis associated with an altered var switch rate. *J Mol Biol*. 2009; 389:453–469. <https://doi.org/10.1016/j.jmb.2009.04.032> PMID: 19389407
22. Bopp SER, Manary MJ, Bright AT, Johnston GL, Dharia NV, Luna FL, et al. Mitotic Evolution of *Plasmodium falciparum* Shows a Stable Core Genome but Recombination in Antigen Families. *PLoS Genet*. 2013; 9:1–15. <https://doi.org/10.1371/journal.pgen.1003293> PMID: 23408914
23. Claessens A, Hamilton WL, Kekre M, Otto TD, Faizullahoy A, Rayner JC, et al. Generation of Antigenic Diversity in *Plasmodium falciparum* by Structured Rearrangement of Var Genes During Mitosis. *PLoS Genet*. 2014; 10(12):e1004812. <https://doi.org/10.1371/journal.pgen.1004812> PMID: 25521112
24. Paul RE, Packer MJ, Walmsley M, Lagog M, Ranford-Cartwright LC, Paru R, et al. Mating patterns in malaria parasite populations of Papua New Guinea. *Science*. 1995; 269(September):1709–1711. <https://doi.org/10.1126/science.7569897>
25. Su X, Ferdig MT, Huang Y, Huynh CQ, Liu A, You J, et al. A genetic map and recombination parameters of the human malaria parasite *Plasmodium falciparum*. *Science*. 1999; 286(5443):1351–1353. <https://doi.org/10.1126/science.286.5443.1351> PMID: 10558988
26. Rosvall M, Bergstrom CT. Maps of random walks on complex networks reveal community structure. *Proc Natl Acad Sci U S A*. 2008; 105(4):1118–1123. <https://doi.org/10.1073/pnas.0706851105> PMID: 18216267
27. Rosvall M, Axelsson D, Bergstrom CT. The map equation. *Eur Phys J Spec Top*. 2010; 178(1):13–23. <https://doi.org/10.1140/epjst/e2010-01179-1>
28. De Domenico M, Lancichinetti A, Arenas A, Rosvall M. Identifying modular flows on multilayer networks reveals highly overlapping organization in interconnected systems. *Phys Rev X*. 2015; 5(1):011027. <https://doi.org/10.1103/PhysRevX.5.011027>
29. Tiedje KE, Oduro A, Agongo G, Anyorigiya T, Azongo D, Awine T, et al. Seasonal Variation in the Epidemiology of Asymptomatic *Plasmodium falciparum* Infections Across Two Catchment Areas in Bongo District, Ghana. *Am J Trop Med Hyg*. 2017; 97(1):199–212. <https://doi.org/10.4269/ajtmh.16-0959> PMID: 28719306

30. Zilversmit MM, Chase EK, Chen DS, Awadalla P, Day KP, McVean G. Hypervariable antigen genes in malaria have ancient roots. *BMC Evol Biol.* 2013; 13(1):110. <https://doi.org/10.1186/1471-2148-13-110> PMID: 23725540
31. Ashby B, Watkins E, Lourenço J, Gupta S, Foster KR. Competing species leave many potential niches unfilled. *Nature Ecol Evol.* 2017. <https://doi.org/10.1038/s41559-017-0295-3> PMID: 28983517
32. Gupta S, Ferguson N, Anderson R. Chaos, Persistence, and Evolution of Strain Structure in Antigenically Diverse Infectious Agents. *Science.* 1998; 280(5365):912–915. <https://doi.org/10.1126/science.280.5365.912> PMID: 9572737
33. Buckee CO, Recker M, Watkins ER, Gupta S. Role of stochastic processes in maintaining discrete strain structure in antigenically diverse pathogen populations. *Proc Natl Acad Sci U S A.* 2011; 108(37):15504–15509. <https://doi.org/10.1073/pnas.1102445108> PMID: 21876129
34. Zinder D, Bedford T, Gupta S, Pascual M. The roles of competition and mutation in shaping antigenic and genetic diversity in influenza. *PLoS Pathog.* 2013; 9(1):e1003104. <https://doi.org/10.1371/journal.ppat.1003104> PMID: 23300455
35. Lourenço J, Wikramaratna PS, Gupta S. MANTIS: an R package that simulates multilocus models of pathogen evolution. *BMC Bioinformatics.* 2015; 16:176. <https://doi.org/10.1186/s12859-015-0598-9> PMID: 26017358
36. Domingo E, Sheldon J, Perales C. Viral quasispecies evolution. *Microbiol Mol Biol Rev.* 2012; 76(2):159–216. <https://doi.org/10.1128/MMBR.05023-11> PMID: 22688811
37. Barry AE, Trieu A, Fowkes FJI, Pablo J, Kalantari-Dehaghi M, Jasinskas A, et al. 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 Proteomics.* 2011; 10(11):M111.008326–M111.008326. <https://doi.org/10.1074/mcp.M111.008326> PMID: 21825279
38. Bachmann A, Predehl S, May J, Harder S, Burchard GD, Gilberger TW, et al. Highly co-ordinated *var* gene expression and switching in clinical *Plasmodium falciparum* isolates from non-immune malaria patients. *Cell Microbiol.* 2011; 13(9):1397–1409. <https://doi.org/10.1111/j.1462-5822.2011.01629.x> PMID: 21740496
39. Bachmann A, Petter M, Krümkamp R, Esen M, Held J, Scholz JAM, et al. Mosquito Passage Dramatically Changes *var* Gene Expression in Controlled Human *Plasmodium falciparum* Infections. *PLoS Pathog.* 2016; 12(4):e1005538. <https://doi.org/10.1371/journal.ppat.1005538> PMID: 27070311
40. Rowe JA, Opi DH, Williams TN. Blood groups and malaria: fresh insights into pathogenesis and identification of targets for intervention. *Curr Opin Hematol.* 2009; 16(6):480–487. <https://doi.org/10.1097/MOH.0b013e3283313de0> PMID: 19812491
41. Mandal S, Sarkar RR, Sinha S. Mathematical models of malaria—a review. *Malar J.* 2011; 10:202. <https://doi.org/10.1186/1475-2875-10-202> PMID: 21777413
42. Bei AK, Diouf A, Miura K, Larremore DB, Ribacke U, Tullo G, et al. Immune characterization of *P. falciparum* parasites with a shared genetic signature in a region of decreasing transmission. *Infect Immun.* 2014; 83(1):276. <https://doi.org/10.1128/IAI.01979-14> PMID: 25368109
43. Coleman S, Dadzie SK, Seyoum A, Yihdego Y, Mumba P, Dengela D, et al. A reduction in malaria transmission intensity in Northern Ghana after 7 years of indoor residual spraying. *Malar J.* 2017; 16(1):324. <https://doi.org/10.1186/s12936-017-1971-0> PMID: 28797269
44. Grenfell BT, Pybus OG, Gog JR, Wood JLN, Daly JM, Mumford JA, et al. Unifying the Epidemiological and Evolutionary Dynamics of Pathogens. *Science.* 2004; 303(5656):327–332. <https://doi.org/10.1126/science.1090727> PMID: 14726583
45. Volz EM, Koelle K, Bedford T. Viral phylodynamics. *PLoS Comput Biol.* 2013; 9(3):e1002947. <https://doi.org/10.1371/journal.pcbi.1002947> PMID: 23555203
46. Taylor JE, Rudenko G. Switching trypanosome coats: what's in the wardrobe? *Trends Genet.* 2006; 22(11):614–620. <https://doi.org/10.1016/j.tig.2006.08.003> PMID: 16908087
47. Bull PC, Buckee CO, Kyes S, Kortok MM, Thathy V, Guyah B, et al. *Plasmodium falciparum* antigenic variation. Mapping mosaic *var* gene sequences onto a network of shared, highly polymorphic sequence blocks. *Mol Microbiol.* 2008; 68(6):1519–1534. <https://doi.org/10.1111/j.1365-2958.2008.06248.x> PMID: 18433451
48. Rask TS, Hansen Da, Theander TG, Pedersen AG, Lavstsen T. *Plasmodium falciparum* Erythrocyte Membrane Protein 1 Diversity in Seven Genomes—Divide and Conquer. *PLoS Comput Biol.* 2010; 6(9):e1000933. <https://doi.org/10.1371/journal.pcbi.1000933> PMID: 20862307
49. White MT, Griffin JT, Churcher TS, Ferguson NM, Basáñez MG, Ghani AC. Modelling the impact of vector control interventions on *Anopheles gambiae* population dynamics. *Parasit Vectors.* 2011; 4:153. <https://doi.org/10.1186/1756-3305-4-153> PMID: 21798055

50. Collins WE, Warren M, Skinner JC, Fredericks HJ. Studies on the relationship between fluorescent antibody response and ecology of malaria in Malaysia. *Bull World Health Organ.* 1968; 39(3):451–463. PMID: [4882987](#)
51. Mucha PJ, Richardson T, Macon K, Porter MA, Onnela JP. Community structure in time-dependent, multiscale, and multiplex networks. *Science.* 2010; 328(5980):876–878. <https://doi.org/10.1126/science.1184819> PMID: [20466926](#)
52. Pilosof S, Porter MA, Pascual M, Kéfi S. The multilayer nature of ecological networks. *Nature Ecology & Evolution.* 2017; 1:0101. <https://doi.org/10.1038/s41559-017-0101> PMID: [28812678](#)
53. Newman MEJ, Girvan M. Finding and evaluating community structure in networks. *Phys Rev E Stat Nonlin Soft Matter Phys.* 2004; 69(2 Pt 2):026113. <https://doi.org/10.1103/PhysRevE.69.026113> PMID: [14995526](#)
54. Rosvall M, Bergstrom CT. Mapping change in large networks. *PLoS ONE.* 2010; 5(1):e8694. <https://doi.org/10.1371/journal.pone.0008694> PMID: [20111700](#)
55. Pielou EC. The measurement of diversity in different types of biological collections. *J Theor Biol.* 1966; 13:131–144. [https://doi.org/10.1016/0022-5193\(66\)90013-0](https://doi.org/10.1016/0022-5193(66)90013-0)
56. Kraemer SM, Kyes SA, Aggarwal G, Springer AL, Nelson SO, Christodoulou Z, et al. Patterns of gene recombination shape var gene repertoires in *Plasmodium falciparum*: comparisons of geographically diverse isolates. *BMC Genomics.* 2007; 8(1):45. <https://doi.org/10.1186/1471-2164-8-45> PMID: [17286864](#)