nature communications

Article

https://doi.org/10.1038/s41467-024-51468-6

Hyper-diverse antigenic variation and resilience to transmission-reducing intervention in falciparum malaria

Received: 16 February 2024

Accepted: 7 August 2024

Published online: 26 August 2024

Check for updates

Qi Zhan[1], Qixin He[2], Kathryn E. Tiedje[3], Karen P. Day[3] & Mercedes Pascual[4,5,6]

Intervention efforts against falciparum malaria in high-transmission regions remain challenging, with rapid resurgence typically following their relaxation. Such resilience co-occurs with incomplete immunity and a large transmission reservoir from high asymptomatic prevalence. Incomplete immunity relates to the large antigenic variation of the parasite, with the major surface antigen of the blood stage of infection encoded by the multigene and recombinant family known as *var*. With a stochastic agent-based model, we investigate the existence of a sharp transition in resurgence ability with intervention intensity and identify molecular indicators informative of its proximity. Their application to survey data with deep sampling of *var* sequences from individual isolates in northern Ghana suggests that the transmission system was brought close to transition by intervention with indoor residual spraying. These results indicate that sustaining and intensifying intervention would have pushed malaria dynamics to a slow-rebound regime with an increased probability of local parasite extinction.

High-transmission endemic regions present an important challenge to controlling and eliminating falciparum malaria. Although malaria prevalence has declined considerably in many parts of the world in the last two decades, high-transmission regions, found primarily in sub-Saharan Africa, tend to be highly resilient to intervention efforts[1]. Here, malaria incidence often rebounds to pre-intervention levels once control is relaxed, and sustained efforts are limited not only by the availability of resources but also by the spread of resistance to insecticides and drug treatments[2,3]. Children experience clinical episodes, eventually becoming immune to severe disease but not infection[4]. This results in a large reservoir of asymptomatic infections in hosts of all ages, which contributes to sustained disease transmission[5,6]. A similar transmission reservoir is found in other vector-borne diseases in domestic and wildlife hosts that exhibit high prevalence of infection but incomplete immunity with no clinical symptoms[7–10].

In these pathogens, a high level of asymptomatic infection at high transmission rates and associated incomplete immunity are enabled by high antigenic variation encoded via multigene families[11]. The *var* multigene family in the malaria parasite *Plasmodium falciparum* provides an example; it encodes for the major variant surface antigen (VSA) during the blood stage of infection, the protein PfEMP1 (*Plasmodium falciparum* erythrocyte membrane protein 1)[12–15]. Anti-PfEMP1 immunity has been found to be crucial to prevent severe disease manifestations and to clear infection[16–19]. Each parasite carries 50-60 *var* genes across its chromosomes. They are expressed largely sequentially, producing and exporting different variants of this protein to the surface of red blood cells[12,20]. Specific immune memory of a given VSA allows the host immune system to clear parasitized red blood cells, shortening infection and reducing the likelihood of parasite

[1]Committee on Genetics, Genomics and Systems Biology, The University of Chicago, Chicago, IL 60637, USA. [2]Department of Biological Sciences, Purdue University, West Lafayette, IN 47907, USA. [3]Department of Microbiology and Immunology, Bio21 Institute and Peter Doherty Institute, The University of Melbourne, Melbourne, Australia. [4]Department of Biology, New York University, New York, NY 10003, USA. [5]Department of Environmental Studies, New York University, New York, NY 10003, USA. [6]Santa Fe Institute, Santa Fe, NM 87501, USA. e-mail: mp6774@nyu.edu

transmission. In contrast, when a strain expresses many VSAs a host has not previously encountered, long infectious periods ensue, enhancing its chance of transmission to another host[12,19]. Under high-transmission settings, local parasite populations exhibit a vast pool of *var* gene variants, documented to reach in the thousands to tens of thousands[21–23] and generated mostly through mitotic and meiotic recombination but also mutation, as well as from migration of the human host and the mosquito vectors[24–27]. Previous studies have shown that negative frequency-dependent selection (NFDS) mediated by the acquisition of specific immunity by hosts contributes to the limited overlap of the *var* genes of both individual repertoires (individual parasite genomes) and isolates (sets of individual parasite genomes co-infecting individual hosts), in a pattern that is both non-random and non-neutral[23,28]. Given the large number of possible *var* gene combinations, hosts are not likely to have seen all or many of the VSAs encoded by the different circulating genomes, and they therefore remain susceptible to re-infection throughout their lifetime despite previous repeated exposure[29,30]. Multi-genomic infections are common resulting from either successive mosquito bites (i.e., superinfection) or a single mosquito bite (i.e., co-transmission)[31–34].

A recent longitudinal field study in Bongo District, located in northern Ghana, implemented deep sampling and sequencing of *var* genes from individual isolates to monitor the response of the transmission system to a three-round transient intervention of indoor residual spraying (IRS) with insecticides[21,35]. Sequencing specifically concerns the DBLα tag, a small conserved ~450 bp region within *var* genes which encodes for the immunogenic Duffy-binding-like alpha domain of PfEMP1[21–23,29,36–38]. Bioinformatic analyses of a large database of exon 1 sequences of *var* genes showed a predominantly 1-to-1 DBLα-*var* relationship, such that each DBLα tag typically represents a unique *var* gene[39]. We use DBLα and *var* interchangeably hereafter. This surveillance through molecular epidemiology revealed persistent estimates of high *var* diversity and low parasite strain overlap (low repertoire similarity), characteristic features of high-transmission regions, despite a considerable decrease in prevalence and parasite population size[21,35]. The transmission system rebounded rapidly after the IRS was discontinued, with this transient intervention appearing to have had only a limited impact[21]. Was this indeed the case? We examine here whether deep sampling of sequences of *var* genes in a host population can help evaluate the impact of intervention further, by providing information on the proximity of a rapid change in resurgence ability with transmission intensity, so that sustaining and intensifying control efforts would avert the rapid rebound of prevalence.

To address these questions, we first investigate the response of the epidemiological system to transmission-reducing interventions with an extended stochastic agent-based model (ABM) that incorporates strain diversity from a multigene family and the interplay between parasite evolution and malaria population dynamics[28] (Methods). We show the existence in numerical simulations of the model of a threshold behavior with intervention intensity, whereby the perturbed system would lose its capacity to rebound rapidly, and identify molecular indicators of such a transition based on deep sampling of *var* gene sequences. The existence of such a transition is suggested by the positive feedback between transmission intensity and *var* diversity recently emphasized in another ABM[40], and by the alternative steady-states and associated tipping point arising from this feedback in a deterministic compartmental model whose formulation requires however a number of simplifications[41]. We apply the lessons learned from our computational model to the field data from northern Ghana across the transient IRS intervention. Our findings should be relevant to other pathogens of humans, wildlife, and livestock with similar immune evasion strategies based on hyper-diverse antigenic variation[7–10].

Results

The resurgence ability of the malaria system to sustained transmission-reducing intervention is non-linear

We investigate how different intervention schemes impact the population genomic structure and dynamics under different transmission regimes (Methods, see Fig. 1a–c and Supplementary Table 1 for the numerical simulation designs and scenarios considered). We start with sustained interventions of a decade to then consider shorter, more realistic, transient ones (Methods, Fig. 1b). Numerical simulations of the ABM demonstrate three qualitatively different rebound dynamics in prevalence (Fig. 2a, Supplementary Fig. 1a) across the series of sustained IRS interventions (Fig. 1b, c). Systems under low- or mid-coverage vector control rebound rapidly in prevalence back to levels similar to those pre-intervention, in what we refer to as a fast-rebound regime. We use "rebound" hereafter synonymously with "resurgence" from the malaria literature, both describing a return to a baseline. In contrast, systems under high-coverage interventions remain at low prevalence, entering what we describe as a slow-rebound regime. Lastly, systems under interventions with levels between these two extremes gradually settle into intermediate prevalence, exhibiting a "transition" regime.

The prevalence level reached after the rebound (a decade into sustained IRS, sampled at the end of the wet/high-transmission season for seasonal transmission) is highly nonlinear along the gradient of transmission intensity (Fig. 2b, Supplementary Fig. 1b). The fast-rebound regime occurs for a broad range within the gradient of transmission intensity, whereas the transition and slow-rebound regimes correspond instead to narrow ranges. Thus, the perturbed system experiences a sharp transition along the gradient of transmission intensities, from a rapid recovery down to a lingering low prevalence. Concomitantly, *var* diversity itself, whether measured as richness or the Shannon Diversity Index[42,43] (Methods), also exhibits the three general behaviors, from the maintenance of high values, through the establishment of intermediate ones, to the eventual persistence of low levels, across the series of sustained IRS interventions (Fig. 2c, Supplementary Fig. 2a, Supplementary Fig. 3a). Post-intervention *var* diversity levels (a decade into sustained IRS) are also highly nonlinear with intervention transmission intensity (Fig. 2d, Supplementary Fig. 2b, Supplementary Fig. 3b).

The nonlinearity of the rebound in prevalence and corresponding *var* diversity levels along the gradient of transmission intensity holds, regardless of whether the system is under constant or seasonal transmission, and whether it is closed, semi-open, or regionally-open, over wide ranges of parameters (Fig. 2, Supplementary Figs. 1–3).

Although to examine long-term rebound dynamics we sustain the vector reduction for up to a decade in the numerical simulations, we find that the perturbed system in the slow-rebound regime can be prone to extinction at a faster time scale. Replicate runs can go extinct as early as in the fourth year into intervention (Fig. 3a), a period that is more comparable to what can be realistically implemented in empirical settings.

Under low or mid-coverage vector control, transmission intensity is maintained at a somewhat high level (Fig. 1b, c). Immediately following intervention, the majority, or a high fraction, of infections can still get transmitted, resulting in a high surviving diversity (Fig. 2a, c, Supplementary Fig. 1-3a). Hosts' immune profiles are heterogeneous because they were exposed to different subsets of the large circulating diversity before intervention. The high surviving diversity, coupled with hosts' heterogeneous immune profiles, allows for a heterogeneous distribution of infection duration (Supplementary Fig. 4). Infections with longer duration are selected due to a reduction in transmission intensity. Thus, the distribution of infection duration immediately following intervention shifts to longer

https://doi.org/10.1038/s41467-024-51468-6

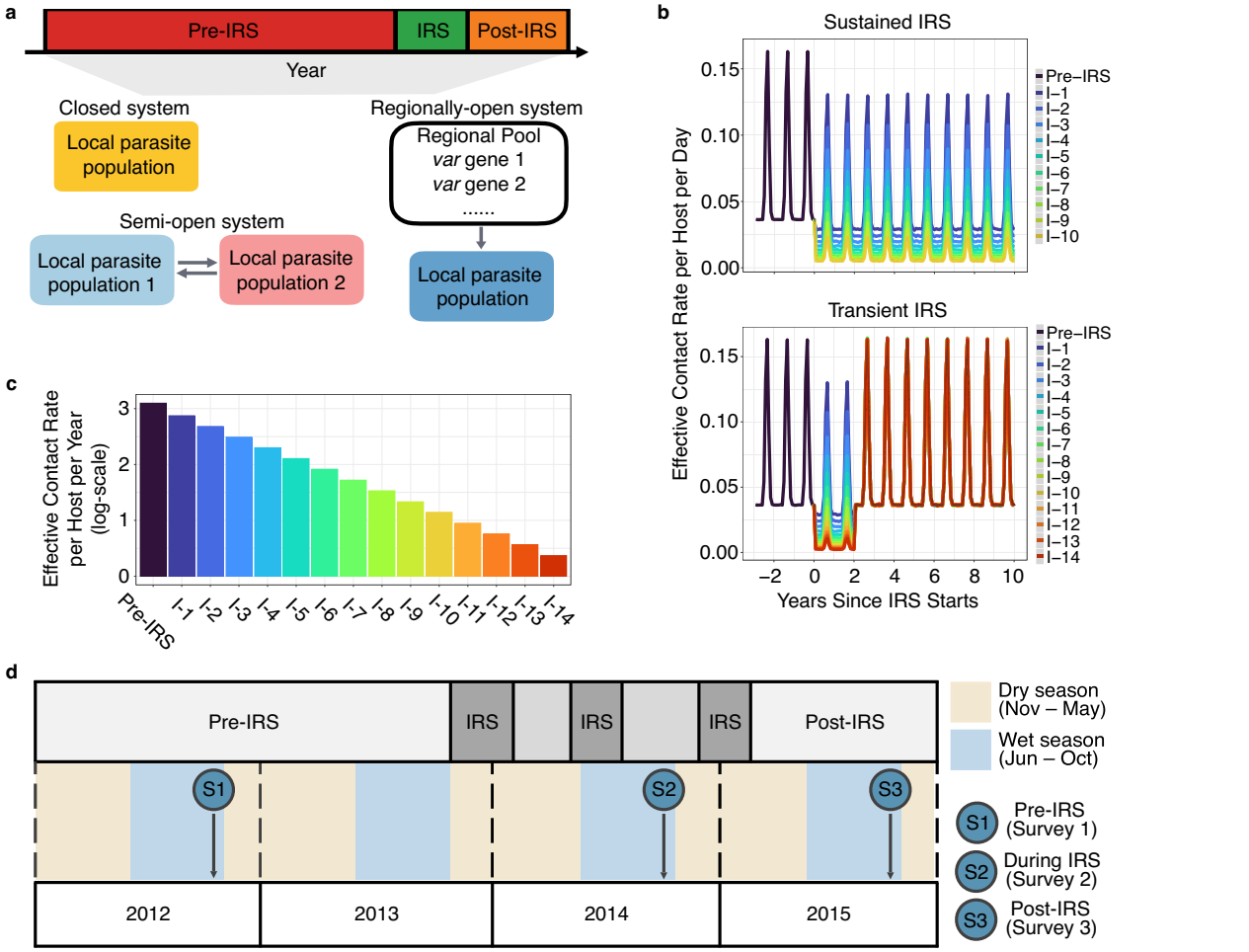

Fig. 1 | Diagram of the numerical simulation design mimicking different lengths of the transmission-reducing intervention (the application of indoor residual spraying or IRS) and the study design of field data sampled from Bongo District in northern Ghana. a Each simulation follows either two or three stages depending on intervention duration: a "pre-IRS" period during which the transmission in the local population reaches a stable state, followed by an "IRS" intervention period of ten or two years which reduces transmission rate (in what we call sustained vs. transient IRS, respectively), and a "post-IRS" period when transmission rates go back to their original levels only applicable to the transient IRS. Three different spatial configurations are implemented: a closed system with no migration, a semi-open system in which two individual parasite populations are connected via migration events, and a regionally-open system in which migrant genomes are sourced from a regional pool of *var* genes (Methods). **b** Transmission intensity or effective contact rate (Methods) varies as a function of time, from pre-, to during, to post-intervention (only applicable to the transient IRS) for a seasonal and closed system, with the two panels illustrating the sustained (top) vs. transient (bottom) implementation for the different reduction levels. **c** Different levels of intervention coverage are considered that reduce transmission rate log-linearly along the gradient of transmission intensity. We show here levels for a seasonal and closed system under sustained interventions. The lowest and highest reductions of the transmission rate correspond respectively to 20% and ~90%, with colors corresponding to those in the top panel of **b** the sustained interventions. The highest reduction for transient interventions can correspond to more than 90%, reaching ~96% for certain scenarios. **d** The study design of the three-round transient IRS in Bongo District in northern Ghana. It is adapted from Argyropoulos et al.[89] and Tiedje et al.[21].

values compared to those before intervention (Supplementary Fig. 4). The selection for infections of longer duration effectively compensates for the reduction in transmission, allowing prevalence to start rebounding. Further into intervention, the loss of immunity due to the reduction in transmission and exposure increases individual susceptibility and consequently infection duration. Prevalence rebounds further, eventually reaching comparable levels to those pre-intervention. In contrast, under high-coverage vector control, immediately following intervention, transmission intensity and circulating diversity both drop to very low levels with a low fraction of infections getting transmitted (Fig. 1b, c, Fig. 2a, c, Supplementary Fig. 1–3a). Few infections can survive longer than the average waiting time between consecutive transmission events. The reduction in transmission intensity cannot be effectively compensated for by the shift in infection duration distribution towards longer values. Thus, the transmission system remains at low prevalence and diversity.

Molecular indicators based on the structure of diversity and multiplicity of infection reveal the malaria system's proximity to a sharp nonlinear transition in its resurgence ability

Having documented a threshold behavior of the transmission system in its response to intervention, we address next the existence of indicators revealing its proximity to the transition regime. Because prevalence is a commonly tracked quantity in malaria surveillance, we first examine whether it can function as one such indicator. The short-term rebound prevalence levels following simulated sustained IRS (two years into intervention) show a much more linear behavior with intervention levels than those after longer sustained control (Fig. 3b, Supplementary Fig. 5a). The relationship between these two types of prevalence levels is non-linear and the degree of nonlinearity varies across different spatial configurations and parameter values (Fig. 3c, Supplementary Fig. 5b). Overall, short-term prevalence levels are only weakly correlated with, and are by themselves only poorly predictive of, long-term rebound prevalence levels. Moreover, high-quality

https://doi.org/10.1038/s41467-024-51468-6

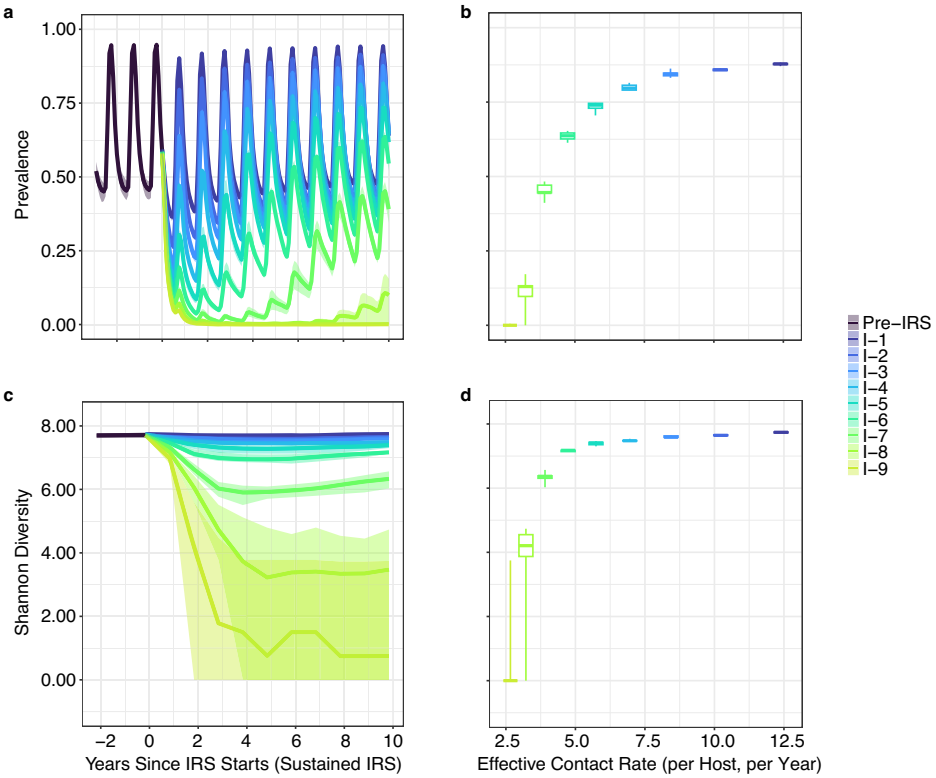

Fig. 2 | Responses of a seasonal regionally-open system to a series of sustained interventions (scenario VII in Supplementary Table 1). a Rebound dynamics of prevalence across time. The dynamics can be categorized into three regimes described respectively as fast rebound, transition, and slow rebound, depending on the behavior of prevalence over time. As the name indicates, for the fast-rebound regime, prevalence returns quickly to similar levels as those pre-intervention, whereas for the slow-rebound regime, prevalence does not recover within a decade and remain substantially lower, and the transition regime exhibits an intermediate recovery. **b** The long-term rebound prevalence levels at the end of the decade (sampled at the end of wet/high-transmission season) are shown against the transmission rates corresponding to the different intervention levels (with colors corresponding to those in Fig. 1c). **c** Rebound dynamics of *var* gene diversity measured by the Shannon Diversity Index across time. **d** Long-term rebound dynamics of *var* gene diversity against transmission rates (with values shown at the end of the decade). Both prevalence and *var* gene diversity exhibit a non-linear pattern against transmission rates. The fast-rebound regime corresponds to a wide interval, whereas the transition and slow-rebound regimes cover a narrow interval along the gradient of transmission rates. For the rebound dynamics of additional combinations of parameter values for closed, semi-open and regionally-open systems, see Supplementary Figs. 1 and 3. In (**b**) and (**d**), boxes represent the 25% quantile and 75% quantile, with a central line marking the median level. The upper and lower whiskers represent the maximum and minimum values. In (**a**–**d**), the range and boxplots are derived based on five replicates.

prevalence data is often unavailable in empirical settings, and efforts to collect these data can be hindered by low testing rates and imperfect detective power for asymptomatic infections[44]. Thus, prevalence by itself does not provide sufficient information on either the proximity of the perturbed system to the transition regime or the loss of resilience of the epidemiological system.

As complementary information, we therefore address indicators related to the structure of diversity from the perspective of *var* genes and isolates. As we seek to apply such indicators to field data, we also consider whether they are robust to measurement errors common in sampling schemes (Methods). We further focus on responsive indicators that rely on sampling immediately following the start of intervention, specifically two years into it (at the end of wet/high-transmission season for seasonal systems).

We first compared whether the distributions of pairwise type sharing (PTS)[38] differ between before and early into intervention. PTS, a similarity index analogous to the Sørensen or Jaccard Index[45,46], describes the fraction of shared DBLα types between any two isolates and hence quantifies the repertoire similarity of parasites circulating in host population (Methods). If each isolate consisted of a monoclonal infection by a repertoire of one *var* gene, sample PTS would be equivalent to homozygosity, and the comparison of 1 - mean(PTS) (i.e., heterozygosity) within and between populations would be equivalent to Wright's fixation indices. PTS can therefore be viewed as a simple extension of fixation indices to the case where multigene family members of unknown allelic status have been collected[38,47]. A second mode at complete overlap emerges in the PTS distribution when the perturbed system approaches the transition regime (Fig. 3d). Concomitant with the appearance of the high overlap mode, the low end of the PTS distribution shifts towards zero or extremely low values (Fig. 3d, Supplementary Figs. 6, 7) due to reduced outcrossing. Outcrossing creates relatedness and increases similarity between parental and offspring strains. Because the frequency of outcrossing events scales with transmission intensity, intervention results in a decrease in outcrossing rates and an increase in clonal transmission of non-related strains. Therefore, the PTS distribution is pulled into the two opposite extremes of complete and no, or extremely low, overlap. The emergence of this second mode at complete overlap is however sensitive to parameter specification and measurement error. It specifically becomes less apparent when the circulating diversity is high among all the simulated scenarios (Methods), effectively disappearing when we do not completely sample all the *var* genes constituting infections (Supplementary Figs. 6, 7). This happens because this incipient appearance of clonality involves too few parasites in the population, a result of intervening in a high-transmission endemic system where the original diversity is high. Any subset of strains from this original diversity which survived intervention, will likely be dissimilar from each other. Unless transmission remains sufficiently low with

https://doi.org/10.1038/s41467-024-51468-6

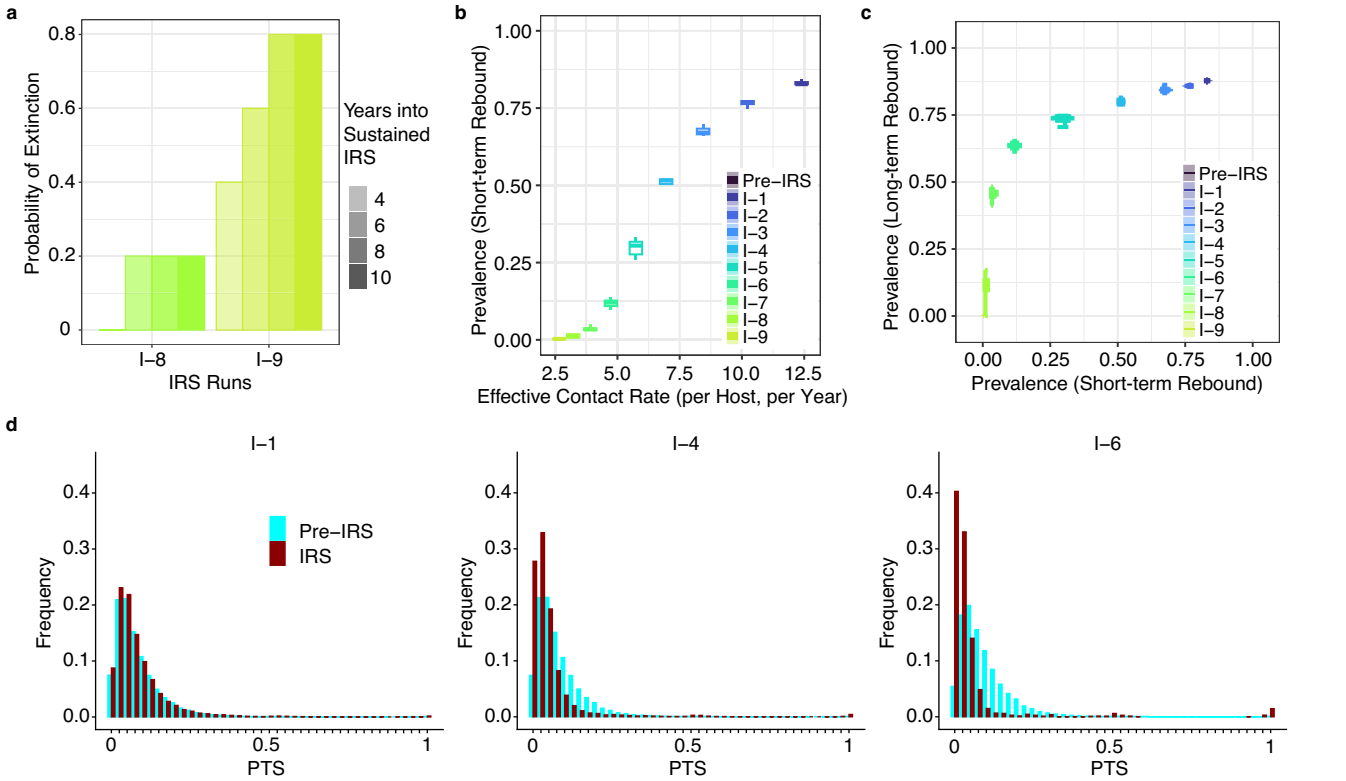

Fig. 3 | The probability of extinction in the slow-rebound regime, short-term vs. long-term rebound prevalence levels, and the population-wise PTS distribution before and early into sustained interventions for a seasonal regionally-open system (scenario VII in Supplementary Table 1). a Probability of parasite extinction calculated from replicates for intervention intensities representative of those in the slow-rebound regime. Extinction occurs as early as in the fourth year into sustained interventions. The two intervention levels correspond to those in Fig. 1c. **b** The short-term rebound prevalence levels at the end of the 2nd year of sustained intervention are shown against the transmission rates corresponding to the different intervention levels. For additional scenarios simulated, see Supplementary Fig. 5a. **c** Short-term rebound prevalence levels (in the second year of intervention) are not good predictors of those in the longer term (at the end of the decade). For additional scenarios simulated, see Supplementary Fig. 5b. **d** The PTS distribution during sustained intervention without measurement error. With increased intervention intensities, and as the system approaches the transition regime, the PTS distribution exhibits the emergence of a second mode at complete overlap due to clonal expansion, and a shift toward lower values (of zero) at the lowest end due to reduction in out-crossing and hence the circulation of non-related strains. The emergence of the second mode is sensitive to the specification of parameters and disappears under measurement error (Supplementary Figs. 6, 7). Thus, it is not a robust indicator of approach to the transition regime, and is unsuitable for practical surveillance purposes. In (**b**) and (**c**), Boxes represent the 25% quantile and 75% quantile, with a central line marking the median level. The upper and lower whiskers represent the maximum and minimum values. In (**a**–**c**), results are obtained from five replicates. In (**b**–**d**), sampling happens at the end of wet/high-transmission season.

importation of *var* gene diversity curtailed so that clonal transmission follows a bottleneck in strain diversity, the second mode in the PTS distribution will be insignificant. In addition, this second mode of complete overlap becomes more diffused under measurement error: two identical strains will have a high but uncertain PTS value within some range, depending on which subset of their *var* genes are sequenced and typed. Thus, the appearance of the second mode cannot be reliably used for empirical surveillance.

Instead of the population-wise PTS, we then examine the age group differences at the low-end of the PTS distributions. A previous study has shown age-dependent patterns of PTS in empirical data, with infections in children (1–10 years old) exhibiting more related *var* repertoires compared to adults (>20 years old), consistent with the effect of immune selection (negative frequency-dependent selection)[29]. We therefore compare these distributions in children and adults from the simulation output across the series of sustained transmission-reducing interventions. We find a reduced difference as the system moves from the fast-rebound regime toward the transition one. Specifically, the PTS distribution within children manifests a mode around zero or extremely low overlap, which was originally present in the PTS distribution of adults. In other words, the PTS distribution between the two age groups becomes more similar, overlapping at the lowest mode when the system approaches and enters the transition regime (Fig. 4a, Supplementary Figs. 8–21a). This behavior is robust to sampling schemes simulating the collection of field data (Methods).

A second candidate for an indicator quantity is the distribution of the multiplicity of infection (MOI), defined as the number of genetically distinct parasite strains co-infecting a single human host[33,48,49]. MOI, also referred to as complexity of infection (COI), is one of the most frequently reported proxies for parasite transmission[50,51]. Although a variety of statistical approaches have been developed that rely on various sources of genotypic data, the estimation of MOI remains challenging for high-transmission endemic settings where individuals typically carry multiple co-occurring infections (i.e., MOI > 1)[52]. The low to non-existent overlap of *var* repertoires enables estimation of MOI on the basis of the number of *var* genes sequenced from an individual's isolate[29,35]. This so-called "*var*coding" method has been shown to provide accurate estimation in numerically simulated infections, and to perform better when transmission is high than a previous method based on a collection of neutral single nucleotide polymorphism loci (SNPs)[52]. Having extended this approach to a Bayesian framework to account for measurement error in the sequencing[21] (Methods), we apply the approach here to derive MOI estimates at the population level across the series of IRS interventions. We find that the MOI distribution becomes increasingly centered at low values around one and two, as the system approaches the

https://doi.org/10.1038/s41467-024-51468-6

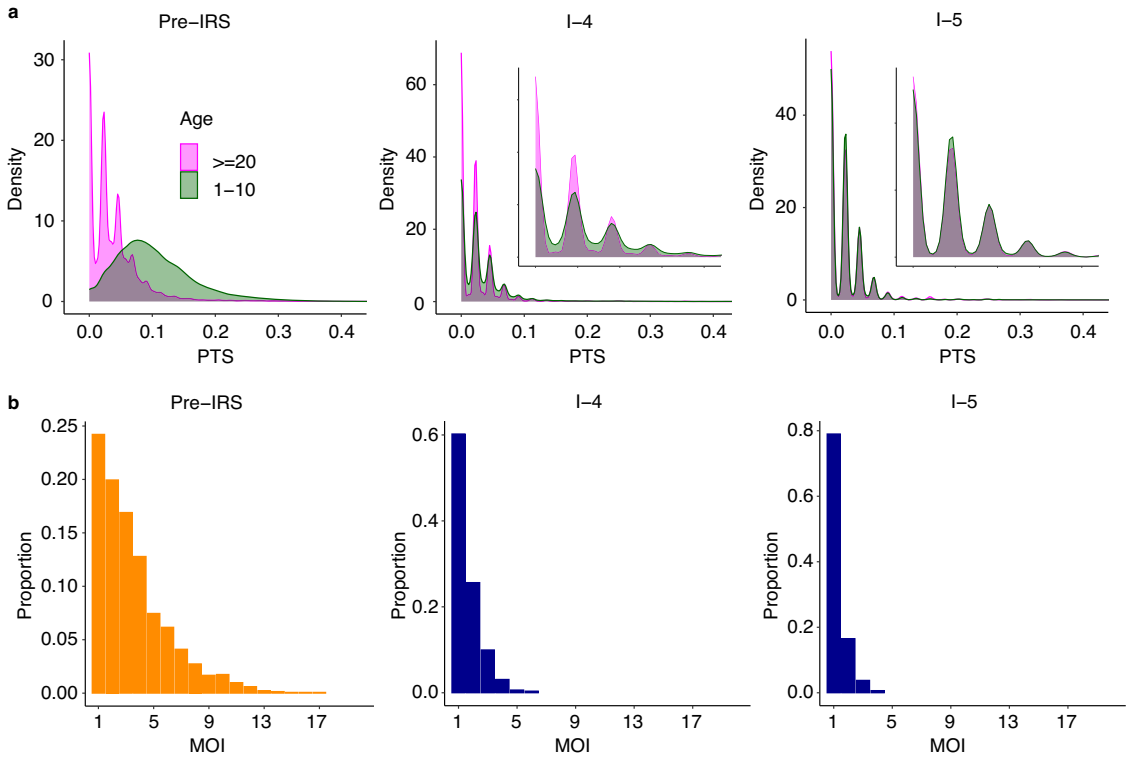

Fig. 4 | Molecular indicators of the proximity of the perturbed local system to the transition regime for a seasonal regionally-open system with a medium-size pool and a baseline migration rate (scenario VII in Supplementary Table 1). For illustration purposes, we show the pre-IRS case, an example IRS under which the system approaches the transition regime, and another example IRS under which the system falls into the transition or slow-rebound regime. **a** As the system approaches the transition regime, the difference between PTS distributions across age groups is significantly reduced and even disappears. In particular, the two distributions almost completely overlap at the lowest mode around 0. The x-axis range for inset figures is 0–0.1. **b** MOI distribution starts to center around 1 and 2 with the majority of infections being either mono-clonal or multi-genomic with two genetically distinct parasites. These two molecular indicators are robust under implemented sampling schemes and sampling limitations representative of those encountered in the collection of field data (Supplementary Fig. 8). Molecular indicators for regionally-open systems with other setups or closed and semi-open systems can be found in Supplementary Figs. 9–21.

transition regime (Fig. 4b, Supplementary Figs. 8–21b). Specifically, whereas for the fast-rebound regime the majority of infections remain multi-genomic (MOI > 1), the fraction of multi-genomic infections decreases significantly when the intervened system approaches the transition regime, and the majority of infections become either mono-clonal or multi-genomic with two genetically distinct parasites. For the transition and slow-rebound regime, the majority of infections become mono-genomic (MOI = 1). This indicator is also robust under sampling schemes simulating the collection of field data (Methods).

The application of identified molecular indicators to an empirical dataset indicates that the local transmission system was brought close to transition by intervention

We apply next the above indicators to an interrupted time-series study involving three age-stratified cross-sectional surveys undertaken at the end of the wet/high-transmission season from Bongo District in northern Ghana, respectively before (Survey 1, October 2012), during (Survey 2, October 2014), and immediately after (Survey 3, October 2015) the transient IRS intervention (Fig. 1d)[21,35]. Following IRS, the difference in the empirical PTS distribution across the two age groups is reduced, almost completely overlapping at the lowest mode (Fig. 5a). To further evaluate this difference in the corresponding PTS values of the two age groups, we considered consecutive quantiles within the range of 0-0.40 for both the simulated output and the Ghana surveys. We chose this range because it encompasses the lower end of PTS, whose change reflects the rate of outcrossing. The difference for those quantiles exhibits two distinct patterns in the simulated system. At high transmission before IRS, or under intervention entering the fast-rebound regime during IRS, the difference is significantly above zero for higher quantiles (Supplementary Fig. 22a, b). In contrast, when the simulated system approaches or enters the transition regime, the difference becomes null (Supplementary Fig. 22a, b). When calculated for the Ghana surveys, its value goes from significantly above zero for higher quantiles for the pre-IRS period, to equal to zero at all quantiles during IRS (Supplementary Fig. 22c). Concomitantly, the empirical MOI distribution shifts towards a high proportion of mono-genomic infections or multi-genomic ones with only two genetically distinct parasites (Fig. 5b), within the range of proportions for simulated systems approaching the transition regime (Supplementary Fig. 22d). Although a given threshold value cannot be identified that indicates this approach, for all scenarios and parameters considered this threshold is above 50%. Thus, we can consider this value a necessary but not sufficient condition, to be considered together with the PTS-based difference described above. Both quantitative patterns taken together suggest that intervention had pushed the empirical system near the transition regime. Similar patterns are found for the survey immediately after the IRS intervention (termed as the post-IRS phase, Fig. 1d, Supplementary Fig. 22c, d), suggesting the impact of IRS persisted through the wet/high-transmission season after it was discontinued.

Individual hosts can seek antimalarial curative treatments in response to their symptoms or their perception of transmission risk. Such curative drug treatments can potentially impact the infection status of treated individuals, as well as their MOI and pairwise-type sharing score or PTS between isolates, although the actual impact on the indicators is difficult to disentangle from the expected changes,

https://doi.org/10.1038/s41467-024-51468-6

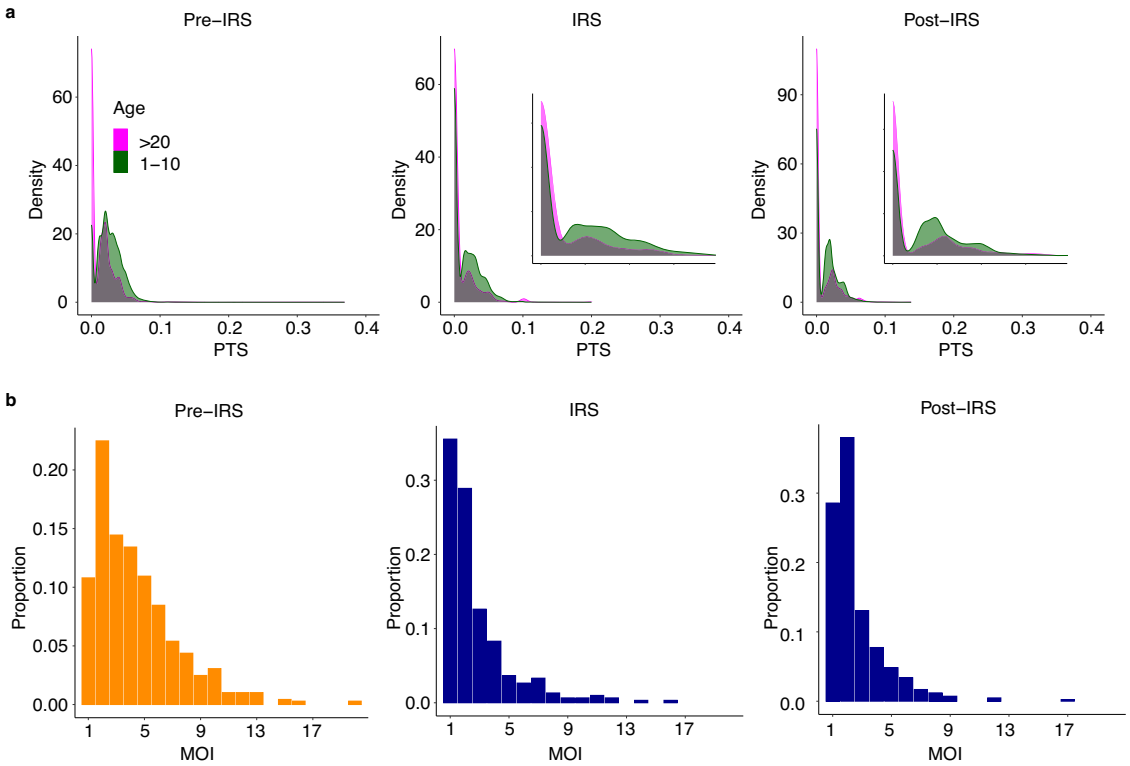

Fig. 5 | The two molecular indicators applied to the field data from northern Ghana before, during, and immediately after the three-round transient IRS intervention, indicate proximity to the transition region where rebound prevalence decreases sharply with intervention intensity. a The PTS distribution between the two age groups becomes more similar, especially for its lowest mode, which is almost completely overlapping. The x-axis range for inset figures is 0–0.075. **b** The MOI distribution exhibits its main mode at one, with the majority of infections becoming either mono-clonal or multi-genomic with two genetically distinct parasites. Those patterns persisted during the year right after when the IRS was discontinued (the post-IRS panel), although to a slightly lesser degree.

given the complex biology of the parasite. We thus repeat the analysis based on samples with treated individuals removed, whose status we know from the field questionnaire that accompanied the surveys[35] (Methods). Similar conclusions hold on the impact of the three-round transient IRS (Supplementary Figs. 23, 24).

Effective restoration of diversity enhances resurgence ability of the malaria system to transient transmission-reducing intervention

Because the three-round IRS intervention was only implemented over a two-year window, we return to the computational model to examine a series of transient interventions of this length across the different spatial configurations (Fig. 1a–c, Methods). Overall, the simulated systems are highly resilient to transient interventions, especially when embedded within a regional pool. After interventions are discontinued, regionally-open systems always rebound rapidly in prevalence back to pre-IRS levels, regardless of IRS coverage and configuration of their regional pool (Fig. 6a, Supplementary Fig. 25). The majority of closed and semi-open systems rebound rapidly as well. However, in a few instances, prevalence remains confined to intermediate or low prevalence levels requiring more than a decade to fully rebound (Fig. 6a, Supplementary Fig. 25). In these, IRS coverage had been at the highest end and sufficiently strong to bring prevalence levels and parasite population sizes close to zero (Fig. 6a, Supplementary Fig. 25), with a low *var* diversity during IRS (Fig. 6b).

We further investigate why the regionally-open systems are highly resilient to transient intervention. Because antigenic diversity facilitates immune evasion and further infection, we address the question from the perspective of change in *var* diversity and associated strain structure by examining how the diversity of migrant genomes and the degree of novelty they introduce to the local parasite population impact the resilience of a regionally-open system. We specifically implement the following two scenarios: a high migration rate per se, and one that is additionally accompanied by a highly diverse gene composition of migrant genomes. We consider for this purpose a static regional pool of medium size that does not update its genes, versus a dynamic one of the same size that does so continuously (Methods). Migrant genomes drawn from this static regional pool necessarily import the same set of *var* genes repeatedly to the local parasite population, and therefore hosts typically develop more complete immunity against their products and the system rebounds in prevalence, although not rapidly. In contrast, migrant genomes drawn from the dynamic regional pool import new *var* genes to which hosts are susceptible, and the system rebounds in prevalence rapidly (Fig. 6c, Supplementary Fig. 26). The transient intervention considered here is among the highest-coverage ones, reducing prevalence and parasite population size close to zero, and *var* diversity to very low values. We do not repeat this analysis for regional pools of large size because they necessarily entail migrant genomes introducing a high degree of novelty to the local parasite population, regardless of the rate of update of their gene composition. Regionally-open systems are highly resilient to transient intervention because migrant genomes effectively introduce novelty and help the recovery in *var* diversity once control is discontinued.

In contrast, closed and semi-open systems can lose their resilience under transient high-coverage intervention. We examine how the disruption of parasite genetic structure typical of high transmission underlies this loss of resilience. For closed and semi-open systems that exhibit a fast rebound, their parasite populations remain large and retain *var* gene diversity characteristic of high-transmission regions during transient intervention (Figs. 6a,b, 6d, Supplementary Fig. 25). After control is discontinued, surviving parasite populations expand,

https://doi.org/10.1038/s41467-024-51468-6

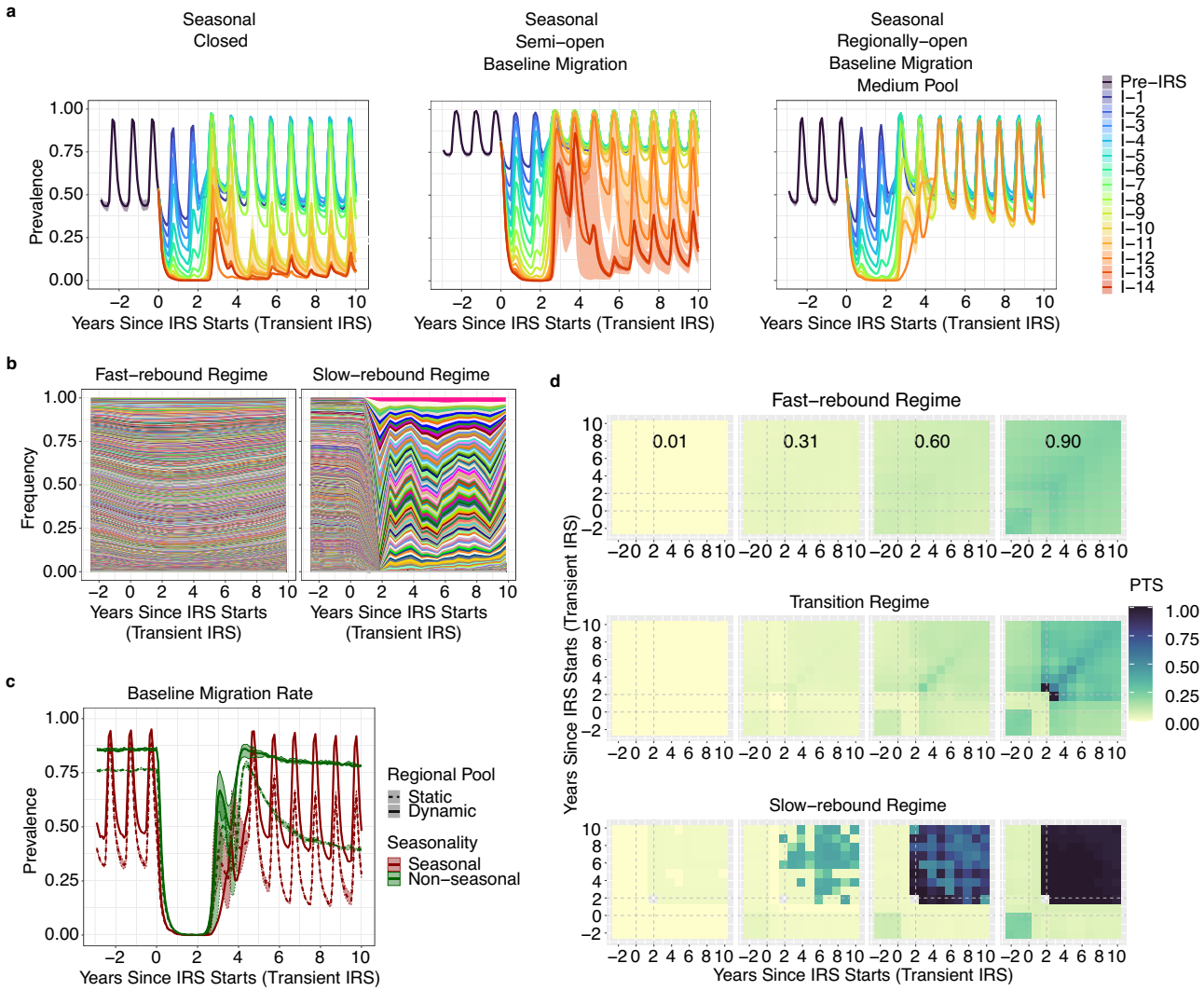

Fig. 6 | Response of the transmission system to a series of transient interventions (2 years), with an effective recovery of *var* diversity and the associated strain structure of limiting similarity underlying the rapid resurgence. a For closed and semi-open systems, the dynamics can be categorized into three regimes described respectively as fast rebound, transition, and slow rebound. For regionally-open systems, the dynamics are always those of fast rebound kind. Different colors represent different intervention levels (see Fig. 1c). Other scenarios are shown in Supplementary Fig. 25. **b** The frequency of genes circulating in the host population varies as a function of time, from pre-, to during, to post-intervention. Each color in the stacked area plots refers to a different gene. The left and right panels correspond to a perturbed seasonal closed system which demonstrates either a fast rebound dynamics (circulating genes are plentiful and typically exhibit low frequency) or a slow-rebound one (circulating genes are low in numbers with each at high frequency), respectively. **c** The rebound dynamics of regionally-open systems with a medium pool and a baseline migration rate to transient and highest-coverage interventions. These systems receive migration from either a dynamic pool or a static one (Methods). Rapid rebound occurs when the pool is dynamic. Same qualitative conclusions hold for the scenario which assumes a migration rate that is an order higher than the baseline (Supplementary Fig. 26). **d** When the system rebounds from a population bottleneck induced by transient interventions, in the fast-rebound regime, clonal expansions are limited and circulating strains are maintained to be dissimilar from each other, measured by PTS distributions. The representative four quantiles of these distributions are at low values. In contrast, circulating strains are highly similar in the slow-rebound regime. Clonal expansion dominates. In (**a**) and (**c**), ranges are derived from three replicates.

and the degree of clonal transmission is low. Circulating genes are plentiful and typically exhibit low frequency (Fig. 6b), and strains are dissimilar from one another as reflected in PTS distributions. These distributions capture the degree of overlap of isolates that are circulating within the same time period as well as those in the past and present (diagonal and off-diagonal values, respectively, in Fig. 6d). Their four quantiles systematically remain low. Hosts' exposure to dissimilar strains facilitates immune evasion and further infection, which in turn promotes the generation and maintenance of *var* gene diversity.

For closed and semi-open systems that exhibit a transition or slow rebound, their parasite populations approach zero with low *var* diversity during transient intervention (Fig. 6a, b, Supplementary Fig. 25). After control is discontinued, small surviving parasite populations expand, resulting in a high degree of clonal transmission. Circulating genes are low in numbers with each at high frequency (Fig. 6b) and strains are similar to one another (Fig. 6d). Hosts' exposure to highly similar strains throughout time compromises immune evasion and further infection, which in turn reduces the generation and maintenance of *var* gene diversity. In our simulations, such systems require more than a decade to recover their prevalence fully.

Discussion

We have described a non-linear and sharp response to the intensity of a sustained transmission-reducing intervention, with an agent-based model that explicitly tracks specific immune memory to strain

variation encoded by a hyper-variable multigene family. The perturbed transmission system experiences a threshold behavior, losing its resilience when pushed into a narrow transition region toward a regime where prevalence rebounds only slowly with an associated extinction risk of the parasite. By examining the relationship between the loss of *var* diversity and the disruption in the strain structure of limiting overlap, we have identified molecular indicators providing information on the approach of the perturbed transmission system to the transition between a fast and slow rebound to high prevalence. These indicators rely on sampling following the intervention, effectively complementing the monitoring of the transmission system through epidemiological quantities over time. They can inform timely adjustment of intervention efforts, to push the system into the transition regime and beyond, into the slow-rebound behavior enhancing local parasite extinction, to do so more rapidly than by observing the temporal trajectory of prevalence itself.

The response behavior to the series of sustained transmission-reducing interventions we have described involves a clear coupling between the rebound dynamics in prevalence and those of *var* gene diversity. In the fast-rebound regime, the heterogeneity of hosts' immunity profiles and large surviving strain diversity underlie a heterogeneous distribution of infection duration. Immunity loss increases hosts' susceptibility to infection and contributes to further lengthening infection duration. Selection for infections with longer duration in response to reduced transmission, effectively compensates for this reduction. In contrast, in the slow-rebound regime, few infections can survive for sufficiently long times to withstand the waiting time between consecutive bite events, resulting in a low surviving diversity and population size of the parasite. Selection for infections with longer duration can no longer effectively compensate for the reduction in transmission.

We illustrated the application of the identified indicators to field data from Bongo District in northern Ghana[21]. Their patterns suggest that the transmission system had been brought close to transition by the two-year transient IRS intervention, and that therefore sustaining and intensifying these efforts would have been worthwhile.

Similar patterns for the two molecular indicators of approaching the transition phase persisted through the wet/high-transmission season after the three-round transient IRS was discontinued. However, prevalence and the population genetics of the *var* multigene family rebounded rapidly, already returning to their pre-IRS characteristics when surveyed 32-month after IRS was discontinued[21]. During the transient intervention, despite a considerable reduction in parasite prevalence and parasite population size (by ~50% and ~70–80%, respectively compared to the pre-IRS level), estimates of parasite population size were not close to zero and *var* population genetics showed persistent characteristics of high transmission regions[21]. These characteristics would provide a large enough founding population to effectively maintain a low degree of clonal transmission after IRS was discontinued. The local parasite population was likely embedded in a large regional pool, which enables the maintenance and rapid recovery of *var* diversity. Resulting immune evasion and associated transmission would have further promoted the generation and maintenance of diversity, enabling a positive feedback between transmission and diversity which has been previously described in two other malaria models[40,41].

The relationship between the response of the transmission system described here in a stochastic individual-based model and the tipping point behavior identified in a deterministic compartmental population model of malaria[41] should be investigated. The latter model explicitly incorporates the temporal dynamics of gene diversity underlying antigenic variation, and the population structure of the host population in both age and immune memory. It is not able however to track parasite population structure, as it samples infections at random from the pool of genetic diversity. Taken together, these previous models and ours underscore that regardless of assumptions and implementations, the positive feedback between antigen-encoding genetic diversity and transmission intensity can influence in important ways both response and resilience to intervention in high-transmission falciparum malaria, and in other infections whose pathogens exhibit large genetic and antigenic diversity at high transmission.

Although we relied initially on relatively long interventions (i.e., a decade), we also examined more transient ones, which underscored the fundamental importance of spatial context through the importation of *var* gene diversity. Although regionally-open systems not surprisingly always exhibit a fast-rebound dynamics in prevalence regardless of intervention coverage, semi-open and closed systems can experience the full range of possible temporal trajectories. The positive feedback described above should remain highly efficient for regionally-open systems, or closed and semi-open systems with fast resurgence once transient intervention is discontinued. In this case, surviving parasite populations can rapidly expand, yet the degree of clonal transmission remains low. Circulating *var* diversity remains high and strains are dissimilar from one another despite a population bottleneck. In contrast, the positive feedback is weak for closed and semi-open systems which enter the transition and slow-rebound regimes. After control is discontinued, clonal expansion dominates, resulting in the circulation of highly similar strains. Thus, as for intervention in other transmission settings, coordinated regional efforts and information on sources and sinks are clearly important. Ways to infer spatial context and the importation of *var* gene diversity from genomic data should be investigated. Although the evolutionary rate of *var* genes measured in vitro is high[25], local innovation events (ectopic recombination and mutation) give rise to single copies of new genes, which have a low probability of surviving the genetic drift barrier. In contrast, migrant genes can be repeatedly imported and are therefore more likely to rise to higher frequencies. Additionally, negative frequency-dependent selection, as a form of balancing selection, maintains genetic variation for longer than expected by random chance, including genes imported either recently or a long time ago[53]. Thus, the overall diversity represents a long-term malaria burden for different regions, and genomic data for the *var* genes should contain signatures of both local innovation events, and recent or more ancient migration ones.

Our computational model did not consider generalized (VSA-transcending) immunity or functional divergence between genes[4,54–65]. Future work will examine the performance of the identified indicators under further aspects of immunity and incorporate different *var* gene groups. We do not expect the addition of generalized immunity to change the qualitative and quantitative aspects of the transmission system's response to transient intervention. But it could potentially change the quantitative aspects of responses to sustained intervention. Given the different mechanisms underlying variant-specific and generalized immunity, the rate at which they wane can differ. Accordingly, the rate at which duration of infection increases due to reduced exposure and associated immunity loss could vary, resulting in potentially different rates of prevalence rebound under sustained interventions. Qualitative aspects of the system's response to sustained intervention, shall remain the same. While strain-transcending immunity from exposure to various conserved antigens reduces parasitemia and clinical symptoms (e.g., merozoite opsonization)[66], it does not necessarily prevent re-infections[29,35,67]. We have focused here on high-transmission endemic region, where the asymptomatic pool of infection contributes significantly to local transmission[5,6]. Despite the majority of hosts having acquired some generalized immunity, they harbor chronic *P. falciparum* infections detectable as microscopic- or PCR-positive. Therefore, we assumed in our model the duration of infection is mainly determined by variant-specific immunity. We also used genetic variation as a proxy for antigenic variation[38] because a genotype to phenotype map is not available for *var* genes. Future work

should examine ways to derive aspects of the map from a variety of perspectives, including bioinformatics, the fitting of ABMs, and protein structure prediction based on deep learning. For regionally-open systems, our simulation considers a local parasite population embedded within a regional pool providing the source of migrant genomes and does not represent metapopulation dynamics explicitly. An explicit simulation of multiple parasite populations in our ABM is computationally expensive. An open population with continuous migration from a regional pool is an alternative typical setting of many assembly models in community ecology.

Our stochastic agent-based model depends on a relatively large number of parameters. The population dynamics of the simulation are most sensitive to a subset of key parameters, however (whose chosen values are provided in the Methods, Supplementary Methods, and Supplementary Data 1). These key parameters include in particular those specifying the configuration of the regional pool for regionally-open systems. To take into account this parameter sensitivity, we conducted simulations with extreme sets of values, with a wide range encompassing from a few to more than ten local parasite populations composing a regionally-open system. Our results hold for the two extreme scenarios considered. Transmission intensity is also key to population dynamics. Because we focus on control efforts in high-transmission endemic regions, we considered simulations whose pre-intervention contact rate parameter led to an emergent force of infection consistent with the empirical range, as determined either from direct measurements or indirect inference for high-transmission endemic regions in sub-Saharan Africa (Methods). The immunity loss rate is another key parameter. Our results hold for the estimated value based on the historical datasets of the treatment of neurosyphilis patients with malaria infections[68,69].

The application of the molecular indicators involves measuring their relative change from pre- to during intervention periods. These quantities could also potentially be used for calibration purposes with the ABM. In particular, the difference in the PTS distribution across the two age groups is associated with transmission intensity. Analysis of empirical datasets from high-transmission regions reveals an age-dependent pattern in the PTS distribution for the two age groups with infections in children (1–10 years old) exhibiting more related *var* repertoires compared to adults (>20 years old), consistent with immune selection being at play[29]. This contrasts with low-transmission regions where all age groups experience low exposure to infections and are similarly susceptible to the circulating diversity[70,71]. Here, the age-dependent pattern is weak, and the difference becomes much less significant or does not exist. Moreover, changes in the MOI distribution, including those of the mean or the proportion of monoclonal infections, also reflect transmission intensity. Therefore, quantitative information, based on moments or quantiles, for the population-wise PTS distribution, the difference in the PTS distribution across age groups, and the MOI distribution, can potentially be used for calibration purposes when fitting the ABM to empirical datasets, including for estimating transmission rates and diversity-related quantities.

The World Health Organization has recently underscored the importance of surveillance for malaria intervention[72,73]. Molecular epidemiology provides approaches to complement traditional surveillance, with methods best suited for moderate- to low-transmission intensities[74–76]. Deep sampling of sequences of *var* genes in host population provides a basis to complement these efforts for high transmission, with a focus on the high diversity that characterizes the *var* multigene family in these regions. Our results establish a link between the population dynamics of the disease and observations of parasite population genomics from the perspective of hyper-variable antigen-encoding genes. They indicate that the structure of diversity revealed via molecular surveillance can inform intervention against malaria in high-transmission endemic settings. These findings should enhance our understanding of transmission dynamics where large standing pathogen diversity represents a major challenge to control efforts.

Methods

Stochastic agent-based model

We extend a previous computational model[28] whose key assumptions and processes are summarized below (see the supplementary Methods for further description). Model parameters and symbols are listed and explained in Supplementary Data 1. The model is an agent-based (individual-based), discrete-event, and continuous-time stochastic system in which all known possible future events are stored in a single event queue along with their putative times, which may be fixed or drawn from a probability distribution with a certain rate. We choose values for the rates of these events based on the literature of malaria epidemiology, related field studies, and in vitro or in vivo values (see supplementary Methods and Supplementary Data 1 for further details). When an event happens, it can cause the addition or removal of future events in the queue, or the modification of their rates, resulting in a recalculation of putative times. This approach is implemented with the next-reaction method[77], which optimizes the Gillespie first-reaction method[78].

Individual human hosts die and are replaced with infants with no immunity. The age structure of the human host population follows a truncated exponential distribution with a mean age of 30 years and a maximum age of 80 years. Individual infections and the immune history of individual human hosts are tracked, and evolutionary mechanisms including mitotic/ectopic recombination and mutation are modeled explicitly. We simulate a local host population of a given size which can assume different spatial configurations (see section "Seasonality and spatial configuration of the local transmission system" below). At the beginning of each simulation, a small number of hosts are randomly selected and infected with distinct parasite genomes from the random assembly of *var* genes from a "regional" pool to initialize local transmission.

Each parasite genome is represented as a specific combination of 45 (non-upsA) *var* genes, and each *var* gene is considered a linear combination of two epitopes (alleles) based on the empirical description of two hypervariable regions in the *var* tag region of the DBLα domain[57]. Mosquito vectors are not explicitly represented as agents in the model. Instead, we consider an effective contact rate (hereafter, the transmission rate, which under some assumptions is effectively equivalent to vectorial capacity) which determines the times of local transmission events. Donor and recipient hosts are selected at random, and transmission occurs if the former carries active blood-stage infections, and the latter has not reached the carrying capacity of its liver-stage. We model ectopic recombination among genes within the same genome during the asexual stage inside the human host. We model meiotic recombination between genomes as occurring at the time of a transmission event, as this process happens in nature during sexual replication within the mosquito vector. See section "Meiotic recombination (or outcrossing) during the sexual stage of infection" in the supplementary Methods for further details. The expression of genes in the repertoire is sequential and the infection ends when the whole repertoire is depleted. Hosts acquire transitory immune memory toward the product of expressed epitopes, and such memory precludes expression of such epitopes in future infections. The total duration of infection of a particular repertoire is therefore proportional to the number of unseen epitopes by the infected host across all individual *var* genes of the given repertoire. See supplementary Methods for further details of the agent-based model (ABM).

Extensions of the original computational model[28] implemented for this work concern the regionally-open spatial configuration, ectopic recombination events, and within-host dynamics. In particular, the regional pool for the regionally-open system now updates its genes at a

rate reflecting the substitution rate of genes in individual local parasite populations within the region. Ectopic recombination events can now generate truly novel alleles instead of simply shuffling the alleles of the two parental genes. The current formulation of the model further allows ectopic recombination events to be associated with a load such that there is a certain probability of generating non-functional offspring genes which replace functional parental genes. These non-functional genes will not be expressed and hence reduce infection duration of repertoires within infected hosts.

Seasonality and spatial configuration of the local transmission system

We simulate both constant and seasonal transmission dynamics. We implemented seasonality by multiplying a scaling constant by a temporal vector of 360 days, containing the daily number of mosquitoes over a full year. In other words, the temporal vector represents the mosquito abundance, and the scaling constant encapsulates all other parameters involved in vectorial capacity. The product of both gives the effective contact rate. To obtain this temporal vector, Pilosof et al. used[79] a deterministic model[80] for mosquito population dynamics. The model[80] 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). We adopt either constant effective contact rates or seasonal ones, which result in EIR or FOI values typical of high-transmission endemic regions[81] (see section "Repertoire transmission" and "Within-host dynamics" in supplementary Methods).

We simulate transmission dynamics for closed, semi-open, and regionally-open systems. For closed systems, after the initial seeding of local transmission from a pool of *var* genes, migration is discontinued. Thus, *var* diversity is generated intrinsically by the dynamics of the ABM for the local population. For semi-open systems, two individual populations coupled via migration are explicitly simulated. They represent empirical transmission systems which are mainly coupled to one or a few neighboring sites. Regionally-open systems represent a local transmission system that is connected with a few to many individual neighboring parasite populations within a broader regional scale. Instead of simulating those individual parasite populations explicitly, we let the focal population receive migration from a regional pool of *var* genes of a certain size. This regional pool acts as a proxy for regional parasite diversity, i.e., diversity of the aggregate of individual local parasite populations in the region. Because each parasite genome is a repertoire of a given number of *var* genes, migrant genomes are assembled from random sampling of *var* genes from the regional pool. Estimation of key parameters concerning the regional pool is described in the following sections.

The temporal course of a simulation and the series of IRS interventions

Each simulation follows either two or three stages (Fig. 1a, b): (i) a pre-IRS period during which the local parasite community is assembled and the transmission system reaches a (semi-) stationary state before the intervention; (ii) an IRS period of either 10 or 2 years during which transmission is decreased (referred to as sustained and transient IRS, respectively); (iii) a post-IRS period when transmission rates recover to pre-IRS levels (only applicable to the transient IRS simulation, since the sustained one extends for the full decade).

We simulate a series of IRS intervention efforts, which reduce transmission rates to levels that are log-linear along the gradient of transmission intensity (Fig. 1c). The lowest and highest coverage of the sustained IRS intervention efforts correspond to a reduction of 20% and slightly more than 90%, respectively. The highest reduction for transient IRS interventions can reach ~96% for certain scenarios.

The IRS intervention efforts are assumed to be regional for open systems. In empirical settings, regional intervention efforts imply that similar efforts are applied to both the local parasite population of interest and the neighboring parasite populations which exchange genomes with the local population via either short-distance or long-range dispersal of mosquitoes or human hosts. All these parasite populations then experience a similar level of reduction in transmission and consequently in their prevalence. As a result, during IRS, the importation of migrant genomes from these neighboring locations also decreases. To implement regional interventions for regionally-open systems, we let migration rates be proportional to the local transmission rate and local prevalence level, both being surrogates of transmission rates and prevalence levels of neighboring populations.

We assume the size of the regional pool remains unaffected during and post-intervention because the aggregate of *var* genes across individual parasite populations should remain high, even when intervention is regional and of high coverage. We do so because this work uses the simulated "regionally-open systems" to represent high-transmission endemic regions in empirical settings, where immigration is known to represent an important challenge to control and elimination efforts. Here, regional interventions rarely eliminate the majority of individual parasite populations.

Estimation of the baseline migration rate for the implementation of open systems

We rely on empirical data on *var* genes from Bongo District, a high-transmission region in northern Ghana to infer the rate of gene exchange between populations, and to obtain a reasonable estimate of migration rates to implement the open transmission systems. The fast-evolving *var* genes under immune selection provide a higher resolution than neutral loci for recent migration events. Following the approach by He et al.[82], we calculate Jost's D[83], a measure of population divergence for highly diverse genetic markers, to quantify *var* gene differentiation within and between two proximal catchment areas (i.e., Vea/Gowrie and Soe) in Bongo District for which molecular sequences were previously obtained at the end of the wet/high-transmission season prior to the IRS intervention[21,29]. Furthermore, under the finite-island model with infinite alleles, the divergence between two populations, D, is proportional to the local innovation (primarily ectopic recombination) rate, and inversely proportional to the migration rate between the two populations[83]. These relationships are intuitive because local innovation events generate new genes unique to the specific local population and hence promote genetic differentiation and isolation between populations, whereas migration events promote genetic exchange and hence reduce divergence between populations. We estimate m, the percentage of contacts leading to infection due to migration relative to local transmission events, by dividing the local innovation rate, i.e., the ectopic recombination rate of the *var* genes, by D. This approach was initially proposed for single-locus genes with multiple alleles[83]. Here we apply it to the *var* multigene family, effectively treating it as if it were single-locus, and different variants of *var* as if they were different alleles of this locus. We consider a range of ectopic recombination rates representative of those in nature, although values of the ectopic recombination rate have been measured only in vitro[25]. We consider the daily rate of ectopic recombination, and the percentage of migration rate per day relative to daily local transmission events. The estimated baseline migration rate is around 0.0026, which translates to a percentage of migrant contacts relative to local ones of around 0.26%. We simulate transmission dynamics for both semi-open and regionally-open systems with migration rates equal to, or an order of magnitude higher than, this estimated baseline.

Estimation of the baseline rate at which genes are substituted in the regional pool

We estimate a value for the rate at which genes in the regional pool are substituted. We refer to this rate as the substitution rate hereafter. As the regional pool represents an aggregate of diversity circulating in individual local populations within the region, this rate can be approximated by the substitution rate of genes in individual local populations. Specifically, we can view a regional pool as the aggregate of diversity of a certain number of equally sized individual parasite populations. Note that we consider regional pools consisting of diversity in the form of the effective number of equifrequent genes and we do not specify a frequency distribution of genes in these regional pools. The collection of this effective number of equifrequent genes, in the ideal scenario, would exploit hosts' immune space equivalently relative to the collection of genes at their original "true" frequency. We use the inverse Simpson diversity index (or, one over the expected homozygosity)[84], $\frac{1}{\sum_i x_i^2}$, to determine the effective number of genes in individual local parasite populations of our simulated closed and semi-open systems, which is at the order of one to two thousand. Moreover, connected individual local parasite populations of these simulated semi-open systems share a significant fraction of their genes when ranked by frequency, around 50% or more. Therefore, we let the number of genes in the regional pool span from several thousand to over tens of thousands, to mimic from a few to about ten to twenty individual local parasite populations. We specifically consider two setups for the regional pool representing roughly the two extremes, referred to as "medium" versus "large" pool scenarios, respectively.

The substitution rate of common genes in a local population depends on the local innovation rate, i.e., the rate at which new variants are generated from ectopic recombination and mutation, and the fixation probability, or the invasion probability, of these new variants[85,86]. Fixation of a new gene variant in this context refers to a single copy of the variant present at a given time point having descendants in the population after a very large amount of time has passed, which typically entails that the frequency of the new variant reaches a certain threshold and becomes sufficiently common so that it is much less susceptible to genetic drift. When the system is at, or close to, a stationary state, the fixation of a new gene variant would imply the loss or replacement of an older common gene variant, hence we term this process and its rate as substitution of genes and substitution rate respectively. The rate of ectopic recombination and mutation has been measured in vitro[25]. The magnitude of the fixation or invasion probability can be derived on the basis of population genetic models and the simulation output from our ABM. From the two together, we can obtain an expectation for the substitution rate as follows.

We write the invasion probability of new variants, or genes, or alleles in the system. At the stationary state, malaria transmission can be described by birth-death processes according to a Moran model with selection. The invasion probability of a low-frequency variant is determined by its fitness advantage relative to that of others[85].

$$p_{inv} = \frac{1-(\frac{W}{W_{new}})^n}{1-(\frac{W}{W_{new}})^N} = \frac{1-(\frac{1}{1+s})^n}{1-(\frac{1}{1+s})^N} \quad (1)$$

where n denotes the number of copies of a new gene, N, the parasite population size, W_{new} and W, respectively the fitness of the low-frequency gene versus that of other genes circulating in the population. Conventionally, the fitness of other genes is normalized to be 1, and that of the new gene is denoted as $1+s$. We assume that there is a large enough number of alleles so that any mutation or ectopic recombination would lead to a different allele and the probability of back mutation to the original allele would be low enough to be negligible. Hence, new variants, generated via mutation or ectopic recombination events, always start with a single copy and $n=1$. We consider parasite populations in high-transmission endemic regions, hence $N \gg 1$. The above equation simplifies to:

$$p_{inv} \approx 1 - \frac{1}{1+s} = \frac{s}{1+s} \quad (2)$$

As described by He et al.[87], in the context of malaria transmission, the fitness of a gene is essentially given by its reproductive number, i.e., the number of offspring genes it produces, R. The reproductive number depends on the transmission rate (b), the transmissibility of the given gene (T), and the typical infection duration (τ) of parasites that carry the given gene.

$$s = \frac{R_{new}}{R} - 1 = \frac{b_{new} T_{new} \tau_{new}}{bT\tau} - 1 = \frac{\tau_{new}}{\tau} - 1 \quad (3)$$

$$p_{inv} = \frac{\frac{\tau_{new}}{\tau} - 1}{\frac{\tau_{new}}{\tau}} \quad (4)$$

We consider the infection duration of a strain constituted by "average" genes versus the infection duration of a strain constituted by 1 new gene and g-1 average genes where g is the genome size (i.e., the number of *var* genes per individual parasite or repertoire). An average gene is a conceptual entity, and the proportion of susceptible hosts to an average gene ($\bar{S}$) is the average of the proportion of susceptible hosts to all circulating genes. Because transmission success of a new gene depends on the total infection duration of the strain that carries it, its invasion probability depends on the overall ensemble of genes. We write here the expected invasion probability for a new gene across the different genome backgrounds it may arise in, by considering that this background consists of average genes. Let d be the duration of expression of a gene in a naïve host. Since the proportion of susceptible hosts to any new gene is 1, we have the following:

$$p_{inv} = \frac{\frac{S_{new} d + (g-1)\bar{S}d}{g\bar{S}d} - 1}{\frac{S_{new} d + (g-1)\bar{S}d}{g\bar{S}d}} = \frac{\frac{1-\bar{S}}{g\bar{S}}}{\frac{1+(g-1)\bar{S}}{g\bar{S}}} = \frac{1-\bar{S}}{1+(g-1)\bar{S}} \quad (5)$$

Note that the expression of p_{inv} derived here is slightly different from, but more rigorous than, the one in He et al.[87] The two expressions are close under certain conditions, for example, when g is large.

The values for $\bar{S}$ typical of high-transmission endemic regions are ~0.4 for the low end and around ~0.8 for the high end, based on our ABM simulation outputs. We can use these ranges of $\bar{S}$ to calculate the invasion probability of a new gene in an individual local parasite population, which together with the innovation rate (primarily through ectopic recombination) and the number of individual local parasite populations (a few for "medium" pools and more than ten for "large" pools) gives the baseline rate at which common genes are substituted in a regional pool with its specific configuration.

For details on the parameterization of the ABM, see supplementary Methods (the section "Selection of values for other parameters") and Supplementary Data 1.

Sampling schemes, measurement error, and curative drug treatment information of empirical data

Empirical surveys are carried out at a specific sampling depth and under sparse sampling schemes. When sampling individual hosts and their *var* types in our simulations, we follow the sampling depth implemented in the empirical surveys from Bongo District, namely around 15% of the total human population size[21]. We further set the sampling period to be once a year for both constant and seasonal transmission (with the former occurring at the end of the year and the

latter occurring at the end of the wet/high-transmission season), to emulate the field sampling of hosts primarily at that time of the year, although a few additional surveys were also conducted at the end of the dry (low transmission) season as a baseline for comparison.

The collection of empirical data includes a measurement error, specifically concerning the distribution of the number of non-upsA DBLα sequenced and typed per monoclonal infection, accounting for certain potential sampling errors or imperfect detection of *var* genes[21,52]. We incorporate this measurement error when sampling from the simulation output by sub-sampling the number of *var* genes per strain based on this empirical distribution.

Moreover, only individuals with microscopy-positive *P. falciparum* infections (i.e., isolate) were included for sequencing of *var* genes in the empirical surveys from Bongo District. A subset of the longitudinal surveys from Bongo District includes in addition the submicroscopic infections detected by the more sensitive method of PCR (the *18S rRNA* PCR), which resulted in a considerably higher fraction[35]. Using surveys for which both microscopy and PCR detection were used, we estimated a conversion factor between the proportion of hosts that are microscopy-positive and those that are PCR-positive of 57%. To address the robustness of our results on molecular indicators to such missing data, we performed the analysis for 57% of all infected hosts originally sampled at the 15% depth.

Individuals can also seek and get antimalarial treatment in response to symptoms or the perception of transmission risk. Field questionnaires were conducted along with each Ghana survey, and participants were asked whether they had received an antimalarial treatment in the previous two weeks (i.e., participants that reported they were sick, sought treatment, and were provided with an antimalarial treatment) prior to their blood samples being collected[35]. Such curative drug treatment can potentially impact the infection status of treated individuals, as well as their MOI and pairwise-type sharing score or PTS between isolates. The precise effects are difficult to disentangle from the expected changes, given the complex biology of the parasite. These treated individuals can be excluded from the analysis when calculating the distributions for molecular indicators, which reduces sample size. The fraction of treated individuals can be high, reaching ~25–50% across different age groups with children exhibiting the highest fraction for the pre-IRS phase[21,35]. However, our relatively large sample size still leaves us with enough individuals pre-IRS. During and immediately after the IRS intervention, this fraction can be much lower, reaching ~5–20% across different age groups, again with children exhibiting the highest fraction for our site[21,35].

Shannon Diversity Index

We calculate and plot *var* gene diversity with both richness and the Shannon Diversity Index[42,43]. The Shannon Diversity Index is defined as follows:

$$H' = -\sum_{i=1}^{R} p_i \ln(p_i) \tag{6}$$

where p_i is the proportion of the ith type and R is the total number of types. Its expression takes the evenness of different gene types into account. It increases either by more unique gene types, or by a greater gene type evenness. Both simulated and empirical *var* gene frequency distributions demonstrate that most genes are rare with a few copies circulating in the host population, and a small fraction of genes are more common with more copies[21]. The Shannon Diversity Index considers this feature when quantifying diversity levels, whereas richness plainly counts the number of genes.

Molecular indicators

A 96% DNA sequence identity cutoff was used to define different *var* gene types[38]. Sequence reads from either the same or different isolates with ≤4% sequence differences were grouped into a single consensus or a single type[21]. We consider the following quantities computed on the basis of *var* gene types and evaluate how these relate to the rebound dynamics of prevalence across the series of simulations for the sustained interventions.

Pairwise type sharing (PTS). The pairwise type sharing (PTS) index describes the degree of shared DBLα types between any two isolates, and is analogous to the Jaccard and Sørensen indices in Ecology[38,45,46]. We use a directional version of this statistic[28] defined as the following:

$$PTS_{ij} = \frac{n_{ij}}{n_i} \tag{7}$$

$$PTS_{ji} = \frac{n_{ij}}{n_j} \tag{8}$$

where n_{ij} is the total number of types shared by isolate i and isolate j, and n_i and n_j are the number of unique types in isolate i and j respectively. This directionality in the formula means that the sharing of a given number of types is relative to the length of the isolate under consideration. A PTS score ranges between 0 and 1, where a score of 0 signifies no shared DBLα types, and a score of 1 indicates complete overlap between the two isolates. We consider several representative quantiles of the PTS distribution (0.01, 0.31, 0.6, and 0.9) to examine the similarity of isolates that are circulating in the population at the same time or across different times. For comparison of PTS distributions across age groups, we limit the calculation to within the same MOI groups and then aggregate across all MOI values to obtain the final distribution. In other words, within each age group, we compute the PTS between individuals of MOI equal to 1, 2, and so on, and then pool the resulting PTS values together to obtain the final PTS distribution. We do so to minimize the effect of different MOI values on PTS.

Multiplicity of infection (MOI). Because *Plasmodium* parasites reproduce asexually as haploid stages within human hosts, signatures of polymorphic genotypes are evidence of multiclonal infection. While any polymorphic marker should thus be suitable in theory for estimating MOI, the typically high number of multiclonal infections in high-transmission endemic regions limits our ability to accurately do so in practice. This is mainly because the diversity of the polymorphic markers can be limited, with two parasites sharing the same marker but carrying distinct sets of antigen-encoding genes. The *var* multigene family provides one solution. Because of the large number of different genes in local populations and the effect of immune selection (negative frequency-dependent selection), repertoires of *var* genes in individual genomes exhibit very limited overlap[23,28]. As a result, a constant repertoire size or number of non-upsA DBLα types in a parasite genome can be used to convert the number of types sequenced in an isolate to an estimate of MOI[29,35]. In its simplest form of considering a typical repertoire length per parasite, the approach neglects the measurement error introduced by targeted PCR and amplicon sequencing of *var* genes in an infection. Recently, the method was extended to a Bayesian formulation that does consider this error and provides a posterior distribution of different MOI values for each sampled individual[21]. We have previously documented the major steps of the Bayesian approach, compared two ways of obtaining a MOI distribution at the population level (by either pooling the maximum a posteriori MOI estimates or calculating a mixture distribution)[21], and examined the impact of different priors on the final MOI distribution at the population level. Here, we obtain the estimated MOI distribution at the population level from pooling the maximum a posteriori MOI estimates and from using a uniform prior for individuals.

Due to the commonness of multi-genomic infections in high-transmission regions, host prevalence levels are not a good indicator of parasite population size. Summing MOI estimates from this improved "*var*coding" approach over sampled individuals with infections, enables the calculation of census population size for the parasite or the total number of infections of relevance to transmission events[21].

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The sequences utilized in this study are publicly available in GenBank under BioProject Number: PRJNA396962. Requests for the data on individual age classes corresponding to *P. falciparum* isolates should be made by contacting the Malaria Reservoir Study Team represented by Prof. Karen Day (karen.day@unimelb.edu.au; Response timeframe: ~1 month), in order to discuss how these data will be utilized for academic or research purposes and, if appropriate, to identify opportunities for collaboration. The individual age data are not publicly available due to ethical reasons. All additional information associated with this study, on the output of numerical simulations and parameters of the agent-based model, is available in the main text and supplementary information.

Code availability

For information on the simulation code and analysis scripts, please see the GitHub repository at: https://github.com/qzhan321/Intervention, digital resource identifier: https://doi.org/10.5281/zenodo.12775297, Zhan[88]. R version 4.0.3 was utilized for all analyses. The base R, along with the R packages RSQLite, dplyr, tidyverse, and stringr were used.

References

1. World Health Organization. *World Malaria Report 2022* (World Health Organization, 2022).
2. Cohen, J. M. et al. Malaria resurgence: a systematic review and assessment of its causes. *Malar. J.* **11**, 122 (2012).
3. Gogue, C. et al. An observational analysis of the impact of indoor residual spraying in Northern, Upper East, and Upper West Regions of Ghana: 2014 through 2017. *Malar. J.* **19**, 242 (2020).
4. Doolan, D. L., Dobaño, C. & Baird, J. K. Acquired immunity to malaria. *Clin. Microbiol. Rev.* **22**, 13–36 (2009).
5. Andolina, C. et al. Sources of persistent malaria transmission in a setting with effective malaria control in eastern Uganda: a longitudinal, observational cohort study. *Lancet Infect. Dis.* **21**, 1568–1578 (2021).
6. Lindblade, K. A., Steinhardt, L., Samuels, A., Kachur, S. P. & Slutsker, L. The silent threat: asymptomatic parasitemia and malaria transmission. *Expert Rev. Anti Infect. Ther.* **11**, 623–639 (2013).
7. Kennedy, P. G. E. & Rodgers, J. Clinical and neuropathogenetic aspects of human African Trypanosomiasis. *Front. Immunol.* **10**, 39 (2019).
8. Gabutti, G., Stefanati, A. & Kuhdari, P. Epidemiology of *Neisseria meningitidis* infections: case distribution by age and relevance of carriage. *J. Prev. Med. Hyg.* **56**, E116–E120 (2015).
9. Guglielmone, A. A. Epidemiology of babesiosis and anaplasmosis in South and Central America. *Vet. Parasitol.* **57**, 109–119 (1995).
10. Kocan, K. M., de la Fuente, J., Guglielmone, A. A. & Meléndez, R. D. Antigens and alternatives for control of *Anaplasma marginale* infection in cattle. *Clin. Microbiol. Rev.* **16**, 698–712 (2003).
11. Deitsch, K. W., Lukehart, S. A. & Stringer, J. R. Common strategies for antigenic variation by bacterial, fungal and protozoan pathogens. *Nat. Rev. Microbiol.* **7**, 493–503 (2009).
12. Zhang, X. & Deitsch, K. W. The mystery of persistent, asymptomatic *Plasmodium falciparum* infections. *Curr. Opin. Microbiol* **70**, 102231 (2022).
13. Baruch, D. I. et al. Cloning the *P. falciparum* gene encoding PfEMP1, a malarial variant antigen and adherence receptor on the surface of parasitized human erythrocytes. *Cell* **82**, 77–87 (1995).
14. Smith, J. D. et al. Switches in expression of *Plasmodium falciparum var* genes correlate with changes in antigenic and cytoadherent phenotypes of infected erythrocytes. *Cell* **82**, 101–110 (1995).
15. Su, X. Z. et al. The large diverse gene family *var* encodes proteins involved in cytoadherence and antigenic variation of *Plasmodium falciparum*-infected erythrocytes. *Cell* **82**, 89–100 (1995).
16. Chan, J. et al. Targets of antibodies against *Plasmodium falciparum*-infected erythrocytes in malaria immunity. *Clin. Invest.* **122**, 3227–3238 (2012).
17. Chan, J. et al. Antibody targets on the surface of *Plasmodium falciparum*-infected erythrocytes that are associated with immunity to severe malaria in young children. *J. Infect. Dis.* **219**, 819–828 (2019).
18. Travassos, M. A. et al. Children with cerebral malaria or severe malarial anaemia lack immunity to distinct variant surface antigen subsets. *Sci. Rep.* **8**, 6281 (2018).
19. Chew, M. et al. Selective expression of variant surface antigens enables *Plasmodium falciparum* to evade immune clearance in vivo. *Nat. Commun.* **13**, 4067 (2022).
20. Deitsch, K. W. & Dzikowski, R. Variant gene expression and antigenic variation by malaria parasites. *Annu. Rev. Microbiol.* **71**, 625–641 (2017).
21. Tiedje, K. E. et al. Measuring changes in *Plasmodium falciparum* census population size in response to sequential malaria control interventions. *eLife* **12** (2023).
22. Ruybal-Pesántez, S. et al. Population genomics of virulence genes of *Plasmodium falciparum* in clinical isolates from Uganda. *Sci. Rep.* **7**, 11810 (2017).
23. Day, K. P. et al. Evidence of strain structure in *Plasmodium falciparum var* gene repertoires in children from Gabon, West Africa. *Proc. Natl Acad. Sci. USA.* **114**, 4103–4111 (2017).
24. Frank, M. et al. Frequent recombination events generate diversity within the multi-copy variant antigen gene families of *Plasmodium falciparum*. *Int. J. Parasitol.* **38**, 1099–1109 (2008).
25. Claessens, A. et al. Generation of antigenic diversity in *Plasmodium falciparum* by structured rearrangement of *Var* Genes during mitosis. *PLoS. Genet.* **10**, e1004812 (2014).
26. Freitas-Junior, L. H. et al. Frequent ectopic recombination of virulence factor genes in telomeric chromosome clusters of *P. falciparum*. *Nat.* **407**, 1018–1022 (2000).
27. Bopp, S. E. R. et al. Mitotic evolution of *Plasmodium falciparum* shows a stable core genome but recombination in antigen families. *PLoS Genet* **9**, e1003293 (2013).
28. He, Q. et al. Networks of genetic similarity reveal non-neutral processes shape strain structure in *Plasmodium falciparum*. *Nat. Commun.* **9**, 1817 (2018).
29. Ruybal-Pesántez, S. et al. Age-specific patterns of DBLα *var* diversity can explain why residents of high malaria transmission areas remain susceptible to *Plasmodium falciparum* blood stage infection throughout life. *Int. J. Parasitol.* **52**, 721–731 (2022).
30. Tran, T. M. et al. An intensive longitudinal cohort study of Malian children and adults reveals no evidence of acquired immunity to *Plasmodium falciparum* infection. *Clin. Infect. Dis.: Off. Publ. Infect. Dis. Soc. Am.* **57**, 40–47 (2013).
31. Portugal, S., Drakesmith, H. & Mota, M. M. Superinfection in malaria: Plasmodium shows its iron will. *EMBO Rep.* **12**, 1233–1242 (2011).
32. Nkhoma, S. C. et al. Co-transmission of related malaria parasite lineages shapes within-host parasite diversity. *Cell Host Microbe* **27**, 93–103 (2020).

33. Anderson, T. J. C. et al. Microsatellite markers reveal a spectrum of population structures in the malaria parasite *Plasmodium falciparum*. *Mol. Biol. Evol.* **17**, 1467–1482 (2000).
34. Volkman, S. et al. Harnessing genomics and genome biology to understand malaria biology. *Nat. Rev. Genet.* **13**, 315–328 (2012).
35. Tiedje, K. E. et al. Indoor residual spraying with a non-pyrethroid insecticide reduces the reservoir of *Plasmodium falciparum* in a high-transmission area in northern Ghana. *PLOS Glob. Public Health* **2**, e0000285 (2022).
36. Tessema, S. K. et al. Phylogeography of *var* gene repertoires reveals fine-scale geospatial clustering of *Plasmodium falciparum* populations in a highly endemic area. *Mol. Ecol.* **24**, 484–497 (2015).
37. Chen, D. S. et al. A molecular epidemiological study of *var* gene diversity to characterize the reservoir of *Plasmodium falciparum* in humans in Africa. *PLoS ONE* **6**, e16629 (2011).
38. Barry, A. E. et al. Population genomics of the immune evasion (*var*) genes of *Plasmodium falciparum*. *PLoS Pathog.* **3**, 34 (2007).
39. Tan, M. H., Shim, H., Chan, Y. & Day, K. P. Unravelling *var* complexity: relationship between DBLα types and *var* genes in *Plasmodium falciparum*. *Front. Parasitol.* **1** (2023).
40. Holding, T., Valletta, J. J. & Recker, M. Multiscale immune selection and the transmission-diversity feedback in antigenically diverse pathogen systems. *Am. naturalist* **192**, 189–201 (2018).
41. de Roos, A. M., He, Q. & Pascual, M. An immune memory-structured SIS epidemiological model for hyperdiverse pathogens. *Proc. Natl Acad. Sci. USA* **120**, e2218499120 (2023).
42. Shannon, C. E. & Weaver, W. *The Mathematical Theory of Communication* (University of Illinois Press,1949).
43. Magurran, A. *Measuring Biological Diversity* (Blackwell, 2004).
44. Bousema, T., Okell, L., Felger, I. & Drakeley, C. Asymptomatic malaria infections: detectability, transmissibility and public health relevance. *Nat. Rev. Microbiol.* **12**, 833–840 (2014).
45. Sørensen, T. A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on Danish commons. *K. Dan. Videnskabernes Selsk.* **5**, 1–34 (1948).
46. Dice, L. R. Measures of the amount of ecologic association between species. *Ecology* **26**, 297–302 (1945).
47. Wright, S. Genetical structure of populations. *Nature* **166**, 247–249 (1950).
48. Paul, R. E. et al. Mating patterns in malaria parasite populations of Papua New Guinea. *Science* **269**, 1709–1711 (1995).
49. Sumner, K. M. et al. Genotyping cognate *Plasmodium falciparum* in humans and mosquitoes to estimate onward transmission of asymptomatic infections. *Nat. Commun.* **12**, 909 (2021).
50. Sondo, P. et al. Determinants of *Plasmodium falciparum* multiplicity of infection and genetic diversity in Burkina Faso. *Parasites Vectors* **13**, 427 (2020).
51. Arnot, D. Clone multiplicity of *Plasmodium falciparum* infections in individuals exposed to variable levels of disease transmission. *Trans. R. Soc. Tropical Med. Hyg.* **92**, 580–585 (1998).
52. Labbé, F. et al. Neutral vs. non-neutral genetic footprints of *Plasmodium falciparum* multiclonal infections. *PLoS Comput. Biol.* **19**, e1010816 (2023).
53. King, R. C., Stansfield, W. D., & Mulligan, P. K. *A Dictionary of Genetics* (Oxford University Press, 2006).
54. Geffery, G. M. Epidemiological significance of repeated infections with homologous and heterologous strains and species of Plasmodium. *Bull. World Health Organ.* **35**, 873–882 (1966).
55. Boyd, M. F. & Kitchen, S. F. On the heterologous value of acquired immunity to *Plasmodium falciparum*. *Natl Malar. Soc. (U.S.)* **4**, 301–306 (1945).
56. Ghumra, A. et al. Induction of Strain-Transcending Antibodies Against Group A PfEMP1 surface antigens from virulent malaria parasites. *PLoS. Pathog.* **8**, e1002665 (2012).
57. Larremore, D. B., Clauset, A. & Buckee, C. O. A network approach to analyzing highly recombinant malaria parasite genes. *PLoS Comput. Biol.* **9**, e1003268 (2013).
58. Buckee, C. O. & Recker, M. Evolution of the multi-domain structures of virulence genes in the human malaria parasite, *Plasmodium falciparum*. *PLoS Comput. Biol.* **8**, e1002451 (2012).
59. Bull, P. C. et al. *Plasmodium falciparum* antigenic variation. Mapping mosaic *var* gene sequences onto a network of shared, highly polymorphic sequence blocks. *Mol. Microbiol.* **68**, 1519–1534 (2008).
60. Claessens, A. et al. A subset of group A-like *var* genes encodes the malaria parasite ligands for binding to human brain endothelial cells. *Proc. Natl Acad. Sci. USA* **109**, 1772–1781 (2012).
61. Rottmann, M. et al. Differential expression of var gene groups is associated with morbidity caused by *Plasmodium falciparum* infection in Tanzanian children. *Infect. Immun.* **74**, 3904–3911 (2006).
62. Kaestli, M. et al. Virulence of malaria is associated with differential expression of *Plasmodium falciparum var* gene subgroups in a case-control study. *J. Infect. Dis.* **193**, 1567–1574 (2006).
63. Rask, T. S., Hansen, D. A., Theander, T. G., Pedersen, A. G. & Lavstsen, T. *Plasmodium falciparum* erythrocyte membrane protein 1 diversity in seven genomes—divide and conquer. *PLoS Comput. Biol.* **6**, e1000933 (2010).
64. Lavstsen, T. et al. Sub-grouping of *Plasmodium falciparum* 3D7 *var* genes based on sequence analysis of coding and non-coding regions. *Malar. J.* **2** (2003).
65. Kraemer, S. M. & Smith, J. D. Evidence for the importance of genetic structuring to the structural and functional specialization of the *Plasmodium falciparum var* gene family. *Mol. Microbiol.* **50**, 1527–1538 (2003).
66. Hill, D. L. et al. Merozoite antigens of *Plasmodium falciparum* elicit strain-transcending opsonizing immunity. *Infect. Immun.* **84**, 2175–2184 (2016).
67. Tiedje, K. E. 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.* **97**, 199–212 (2017).
68. Collins, W. E., Skinner, J. C. & Jeffery, G. M. Studies on the persistence of malarial antibody response. *Am. J. Epidemiol.* **87**, 592–598 (1968).
69. Collins, W. E., Jeffery, G. M. & Skinner, J. C. Fluorescent antibody studies in human malaria. II. Development and persistence of antibodies to *Plasmodium falciparum*. *Am. J. Tropical Med. Hyg.* **13**, 256–260 (1964).
70. Teressa, M. et al. Severe malaria in an adult patient from low-endemic area in Flores Island, East Nusa Tenggara. *Case. Rep. Med.* **2023**, 1239318 (2023).
71. White, N. J. Severe malaria. *Malar. J.* **21**, 284 (2022).
72. Lourenço, C. et al. Strengthening surveillance systems for malaria elimination: a global landscaping of system performance, 2015–2017. *Malar. J.* **18**, 315 (2019).
73. Fountain, A. et al. Surveillance as a core intervention to strengthen malaria control programs in moderate to high transmission settings. *Am. J. Trop. Med Hyg.* **108**, 8–13 (2022).
74. Taylor, A. R., Jacob, P. E., Neafsey, D. E. & Buckee, C. O. Estimating relatedness between malaria parasites. *Genetics* **212**, 1337–1351 (2019).
75. Chang, H.-H. et al. THE REAL McCOIL: a method for the concurrent estimation of the complexity of infection and SNP allele frequency for malaria parasites. *PLoS Comput. Biol.* **13**, e1005348 (2017).
76. Wong, W. et al. RH: a genetic metric for measuring intrahost *Plasmodium falciparum* relatedness and distinguishing cotransmission from superinfection, *PNAS Nexus*, **1** (2022).

77. Gibson, M. A. & Bruck, J. Efficient exact stochastic simulation of chemical systems with many species and many channels. *J. Phys. Chem. A* **104**, 1876–1889 (2000).
78. Gillespie, D. T. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comput. Phys.* **22**, 403–434 (1976).
79. Pilosof, S. et al. Competition for hosts modulates vast antigenic diversity to generate persistent strain structure in *Plasmodium falciparum*. *PLoS. Biol.* **17**, e3000336 (2019).
80. White, M. T. et al. Modelling the impact of vector control interventions on *Anopheles gambiae* population dynamics. *Parasites Vectors* **4**, 153 (2011).
81. Smith, D. L., Drakeley, C. J., Chiyaka, C. & Hay, S. I. A quantitative analysis of transmission efficiency versus intensity for malaria *Nat. Commun.* **1**, 108 (2010).
82. He, Q., Pilosof, S., Tiedje, K. E., Day, K. P. & Pascual, M. Frequency-dependent competition between strains imparts persistence to perturbations in a model of *Plasmodium falciparum* malaria transmission. *Front. Ecol. Evol.* **9**, 633263 (2021).
83. Jost, L. G(ST) and its relatives do not measure differentiation. *Mol. Ecol.* **17**, 4015–4026 (2008).
84. Simpson, E. H. Measurement of diversity. *Nature* **163**, 688 (1949).
85. Maruyama, T. A Markov process of gene frequency change in a geographically structured population. *Genetics* **76**, 367–377 (1974).
86. Otto, S. P. & Whitlock, M. C. The probability of fixation in populations of changing size. *Genetics* **146**, 723–733 (1997).
87. He, Q. & Pascual, M. An antigenic diversification threshold for falciparum malaria transmission at high endemicity. *PLoS. Comput. Biol.* **17**, e1008729 (2021).
88. Zhan, Q. Intervention. Repository for simulation code and analysis script. *Github* https://github.com/qzhan321/Intervention (2024).
89. Argyropoulos, D. C. et al. The impact of indoor residual spraying on *Plasmodium falciparum* microsatellite variation in an area of high seasonal malaria transmission in Ghana, West Africa. *Mol. Ecol.* **30**, 3974–3992 (2021).

Acknowledgements

This research was supported by Fogarty International Center at the National Institutes of Health through the joint NIH-NSF-NIFA Ecology and Evolution of Infectious Diseases award R01-TW009670 to K.P.D. and M.P.; and the National Institute of Allergy and Infectious Diseases, National Institutes of Health through the joint NIH-NSF-NIFA Ecology and Evolution of Infectious Diseases award R01-AI149779 to K.P.D. and M.P. We wish to thank the participants, communities, and the Ghana Health Service in Bongo District, Ghana for their willingness to participate in the study of empirical data. We would like to thank the field teams in Bongo for their technical assistance in the field, as well as the laboratory personnel at the Navrongo Health Research Centre for their expertise and for undertaking the sample collections and parasitological assessments. We appreciate the support of the Research Computing Center at the University of Chicago through the computational resources of the Midway cluster.

Author contributions

Q.Z., K.P.D., and M.P. conceived and designed the study. Q.Z. worked on the model and extended the simulation code. Q.Z. analyzed and visualized the data. Q.Z. wrote the original draft of the manuscript. Q.Z., Q.H., K.E.T., K.P.D., and M.P. contribute to the final version of the manuscript. K.P.D. and M.P. supervised the research. K.P.D. and M.P. acquired the funding.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-024-51468-6.

Correspondence and requests for materials should be addressed to Mercedes Pascual.

Peer review information *Nature Communications* thanks Roberto Amato and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.