

Supplementary Material for “Hyper-diverse antigenic variation and resilience to transmission-reducing intervention in falciparum malaria”

This PDF file includes Supplementary Methods, Supplementary Figs. 1-26, and Supplementary Table 1.

Methods

We describe details of key assumptions and processes of the ABM, including extensions implemented to the model for this work.

Var genes and repertoire structure

Each parasite genome consists of a specific combination of 45 non-*upsA* *var* genes in the simulation. The size of the repertoire is based on the median number of non-*upsA* DBL α types identified in the previously reported 3D7 laboratory data^{1,2}. The grouping of *var* genes (i.e., *upsA* and non-*upsA* (*upsB* and *upsC*)) is based on their structural information (i.e., the semi-conserved upstream promoter sequences or *ups*)^{3,4,5}. Each empirically sampled parasite carries both types of *var* genes in a fairly constant proportion, with the majority (~80%) being non-*upsA* groups^{6,7,8,9,3}. Although there is evidence indicating functional differences between the two groups, the groupings do not necessarily correlate with function. Genes of the same group can differ in their functional properties^{10,11}. Genes of different groups may both be significantly associated with a particular type of infection symptom^{12,11}. Within-group functional heterogeneity and variation, and cross-group functional similarity exist. In addition, the mechanism underlying the fairly constant proportion of two structural groups in empirical samples across time and geographical locations remains to be properly understood. It is therefore unclear how to model functional divergence of *var* groups in a rigorous and parsimonious way, and whether assumptions made on the particular coarse-graining apply to the population dynamics of the disease. We consider here only non-*upsA* types in our simulation and discuss extensions introducing the more conserved *upsA* types. The non-*upsA* DBL α sequences are in general ~20 times more diverse and less conserved among repertoires than the *upsA* DBL α sequences⁶, making them suitable and sufficient for MOI estimation based on deep sampling of sequences of *var* genes, as described in the section “Molecular indicators” of Methods.

Repertoire transmission

Mosquito vectors are not explicitly represented as agents in the model. Instead, we consider an effective contact rate (hereafter, the transmission rate, which determines the times of local transmission events (exponentially distributed according to a Poisson process). At these times, a donor and a recipient host are selected randomly, and successful transmission occurs only if the donor carries active blood-stage infections, and the recipient has not reached a carrying capacity of the liver stage of infection (see the following section “Within-host dynamics”). Co-infection reduces transmission success of each individual parasite strain. When the donor is infected with multiple strains in the blood stage, then the transmission probability of each strain is reduced by a factor equal to the number of co-infecting strains^{13,14,15}. We apply a delay between a successful transmission event and the transmitted strains in the recipient host starting to express their *var* genes and becoming infectious, to account for stages of the life cycle in the vector and in the human host that are not modeled explicitly (see the following section “Within-host dynamics”).

Note that for simplicity, when a transmission event occurs, our model considers that all infectious “bites” (i.e., for a donor host that carries active blood-stage infections and a recipient host that has liver capacity), result in infection of the recipient host. The probability of infection given infectious bites can be considerably less than 1 in nature. For example, a bite with an adequate volume of sporozoites has been shown to have a probability of infecting human hosts of around 10% under certain constraints¹⁶. In general, when transmission is high and multi-genomic infections are common, this probability can be even much lower than 10%¹⁷. This transmission probability functions as a general parameter encapsulating the effect of a variety of processes (including immunity, within-host dynamics, measurement bias, and heterogeneous transmission), all of which may cause infectious bites not to result in detectable blood-stage infections. Because for our dynamics, this parameter represents a rescaling, it does not need to be incorporated explicitly, as long as for comparison of entomological inoculation rates (the number of infectious bites received by an individual over a given time period or EIR) in our model and field estimates, the former is divided by the transmissibility probability. Additionally, our model generates values of the force of infection (the number of new infections acquired by an individual host over a given time interval or FOI). For pre-intervention dynamics, the emergent annual FOI values are within the range of 9-15, directly comparable to the measures and estimates from the field for high-transmission endemic regions in sub-Saharan Africa¹⁷. Because FOI measures the effective infections which advance to the blood stage, it defines the time scale of malaria transmission. To achieve comparable FOI values, we would need to increase the values of the effective contact rate in the simulation should we explicitly include a transmission probability. The size of the increment would depend on the specific choice of this transmission probability factor.

Ectopic recombination during the asexual blood stage of infection

Ectopic recombination is a major mechanism of *var* genes’ diversification, occurring during both the sexual and asexual stages. For simplicity, we only model ectopic recombination among genes within the same genome during the asexual stage as follows. Two genes are first selected randomly from a given strain. Then, the location of the recombination breakpoint is randomly chosen. Under normal recombination, both genes have their alleles swapped and there is a certain probability that new alleles are created in the process. Under conversion, the second gene remains unchanged. In the current implementation, we assume all ectopic recombination events result in normal recombination rather than gene conversion. Newly recombined genes have a probability of being functional (i.e., viable), dependent on the similarity of their parental genes and the locations of the breakpoints¹⁸, and they will replace their parental ones. Non-functional genes will not express and get deactivated immediately and hence do not increase infection duration.

Meiotic recombination (or outcrossing) during the sexual stage of infection

Meiotic recombination occurs between strains during sexual replication inside the mosquito vector. Because we do not explicitly model mosquitoes, we represent meiotic recombination between genomes at the time of a transmission event by the following process. When multiple strains whose number we denote by m ($m > 1$) co-infect a donor host, a Bernoulli trial is conducted for each of these m strains to determine whether or not it will be transmitted via the contact event, with the probability of success parameter equal to each strain’s transmission probability. Co-infection reduces the transmission success of each individual strain by a factor equal to m (See section “Repertoire transmission” above). Only a subset n of the strains, smaller

or equal to m , will be chosen. In nature, this subset of strains would co-infect a mosquito vector. To include the possibility of recombination, each of the n strains which are to be transmitted to a recipient human host, are obtained by drawing two parental strains from the pool of n with replacement. If the two parental strains are the same, then the original one is transmitted to the recipient human host. If the two parental strains are different, they recombine, and the recombinant strain is transmitted to the recipient human host (see below the description of how the recombinant strain is formed). Both the original and recombinant strains can be transmitted to a recipient human host, and the respective probability is $\frac{1}{n}$ and $1 - \frac{1}{n}$.

Although the association of physical locations and major groups of *var* genes is established, orthologous gene pairs between two strains are often unknown. Therefore, we implement recombination between strains as a process in which genes are randomly selected out of all the original genes from the two strains pooled together. As physical locations of *var* genes can be mobile through ectopic recombination and gene conversions, this assumption is a reasonable simplification of the meiotic recombination process. The resulting offspring strains share some fraction of their *var* genes with the two parental strains. Outcrossing creates relatedness and increases similarity between parental and offspring strains.

Within-host dynamics

Each strain is individually tracked through its entire life cycle, encompassing the liver stage and asexual blood stage in the human host, and the sexual stage in the mosquito. As we do not explicitly model mosquitoes, we delay the expression of each strain in the recipient host by 14 days to account for the time required for the sexual and liver stages. Specifically, the infection of the host is delayed 7 days to account for the time required for gametocytes to develop into sporozoites within mosquitoes. When a host is infected, the parasite remains in the liver stage for an additional 7 days before being released as merozoites into the bloodstream, invading red blood cells and starting the expression of the *var* repertoire^{19,20}. The expression of genes in the repertoire is sequential and the infection ends when the whole repertoire is depleted. During repertoire expression, the host is considered infectious with the active strain. Only these active “blood-stage” strains are transmissible to another host. The deactivation rates of the genes are controlled by host specific immunity. When one gene is actively expressed, host immunity “checks” whether it has seen any of the two epitopes in its infection history. If both epitopes have been previously encountered, the gene gets deactivated immediately. The duration of the active period of a gene is thus proportional to the number of unseen epitopes of the gene. After the gene is deactivated, the host adds the deactivated two epitopes to its immunity memory. The next new gene from the repertoire then becomes immediately active and the strain is cleared from the host when the whole repertoire of *var* genes is depleted. The total duration of infection of a particular repertoire is therefore proportional to the number of unseen epitopes by the infected host aggregated across all individual *var* genes of that repertoire. The immunity toward a certain epitope wanes at a certain rate.

In addition, we consider a carrying capacity for both the liver- and blood-stage infection respectively. We adopt 20 for both quantities because this has been the maximum MOI observed in the empirical data from Bongo⁶. When the number of strains in the liver stage reaches the specified carrying capacity, the host no longer receives additional infections when it is selected as the recipient host for transmission events. When the number of strains in the blood stage reaches the carrying capacity, strains in the liver stage are not released into the bloodstream and

fail to “transition to the blood-stage”. In this implementation, we assume these strains are lost instead of queuing indefinitely²¹.

As previously described in the section “Repertoire transmission”, we make sure that the chosen values for transmission rates or associated FOI in the simulation are high enough to capture disease transmission in high-transmission endemic regions but still sufficiently low to preclude significant void bites (because of exceeding carrying capacities).

Selection of values for other parameters

Immunity loss rate

Values of this parameter can be obtained from the data on the treatment of neurosyphilis patients with malaria infections in the 1950s to 1960s^{22,23}. Their antibody titers were tracked through observation periods of up to decades following the termination of their very first inoculation. During these periods, these individuals usually did not receive repeated, either homologous or heterologous, infections, which can boost immunity and confound the immunity loss rate. The persistence of the antibody response against malaria was studied by Collins et al.²² Sera from 95 patients having had induced infections with *P. falciparum*, *P. malariae*, *P. vivax*, or *P. ovale* were tested for presence of specific antibody using the indirect fluorescent antibody methods. The intervals between termination of the infections and acquisition of the sera ranged from approximately 6 months to 26 years. Among these patients, 14 of them had a past history of *P. falciparum* infection. A specific serum dilution high enough to be considered positive was found as late as 8 years after the termination of *P. falciparum* infection. In general, the median response decreased sharply when three years had elapsed from the termination of infections.

A different set of 7 patients was followed to track persistence of the antibody response only up to 20 months post inoculation²³. During the 8-month period (between 12 months and 20 months post inoculation), 3 of the titers remained unchanged, 2 decreased by one dilution, 1 decreased by two dilutions, and 1 decreased by three dilutions. Roughly titers decrease by 1 dilution for every 8 months.

Combining information from both studies^{22,23}, we set the mean immunity duration to be 3 years in the simulation.

Number of epitopes (alleles) per locus

Larremore et al.²⁴ presented a new method that identifies nine highly variable regions (HVRs) and studied 307 amino acid sequences from the DBL α domain of the *var* genes of seven *P. falciparum* isolates. Two HVRs (5 and 6) correspond to the DBL α tag sequences that were amplified in our field isolates from Bongo District. For each HVR, the study constructed a network in which each sequence is a node, and two nodes are linked if they share an exact match of significant length of the HVR. Using a community partition inference algorithm, the authors found that the number of components ranges from 1 to 106 across different HVRs, with a mean around 30. Specifically, 10 and 8 components were found for HVRs 5 and 6. We therefore assume a reduction of an order of 10 between the number of genes and the number of alleles (epitopes) at each of the two loci.

Ectopic recombination rate

In the study by Claessens et al.²⁵, the average total number of *var* recombination events per life cycle (2 days) was estimated as 2×10^{-3} . Given that there are around 60 *var* genes per parasite

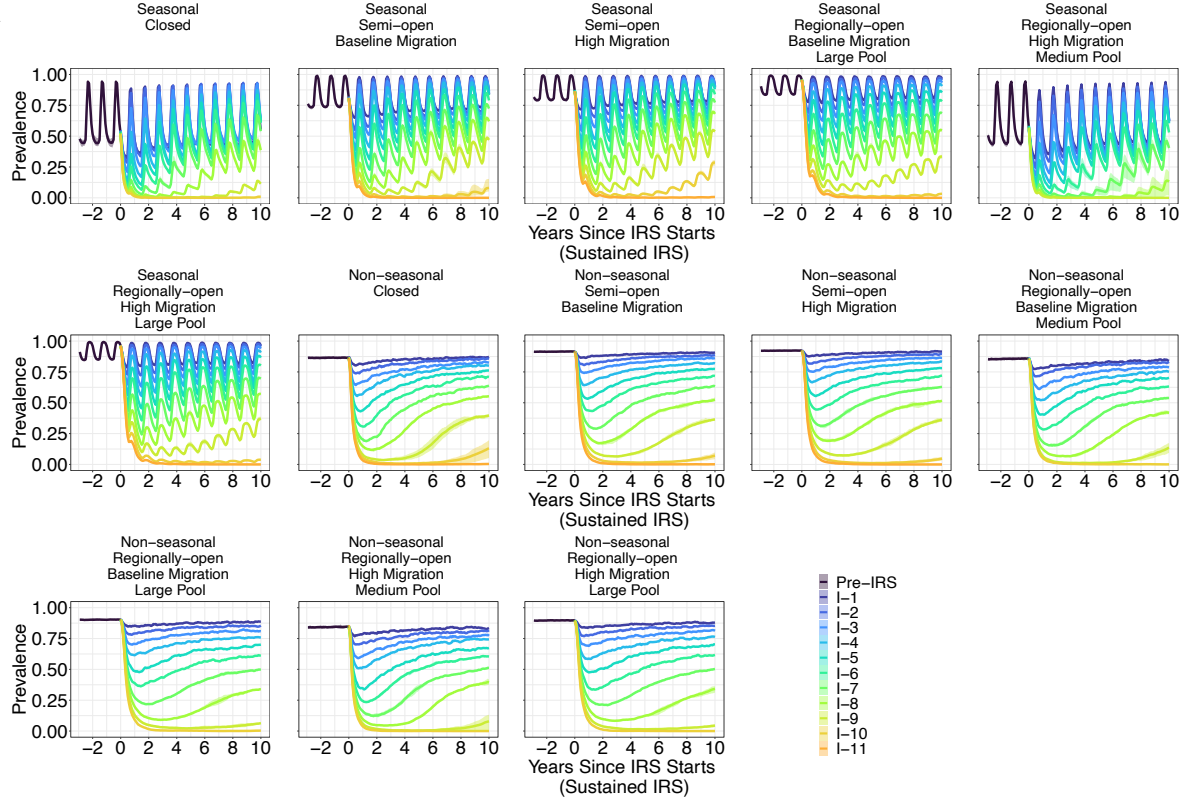
175 genome and hence $\frac{60 \times 59}{2}$ different combinations of *var* gene pairs for a recombination event to
 176 happen, the previous value amounts to about $\frac{\frac{2 \times 10^{-3}}{2}}{\frac{60 \times 59}{2}} = 5.64971751 \times 10^{-7}$ recombination
 177 events per *var* gene pair per day. Note that there was some heterogeneity and hierarchy observed
 178 in the study so that certain pairs of *var* genes from a specific structural group recombined more
 179 frequently than others and recombination events are more likely to be observed between genes
 180 from the same structural group than the ones from different structural groups. We do not
 181 consider such a variation across all gene pairs but use the above-calculated mean value in our
 182 simulation output.

183 Ectopic recombination events do not necessarily generate new alleles, and they happen more
 184 frequently between similar homology blocks, the exchange of which is less likely to result in
 185 antigenically new offspring alleles. We test a high and a low probability of generating
 186 antigenically new alleles per ectopic recombination event, i.e., 0.5 (the upper bound) and 0.16, in
 187 other words, 1 in 2 versus 1 in 6 ectopic recombination events. The results and patterns are
 188 robust across the two probability values. We thus report the results and patterns based on
 189 simulation output which assume that a 0.16 probability (1 in 6 ectopic recombination events) will
 190 result in a new allele.

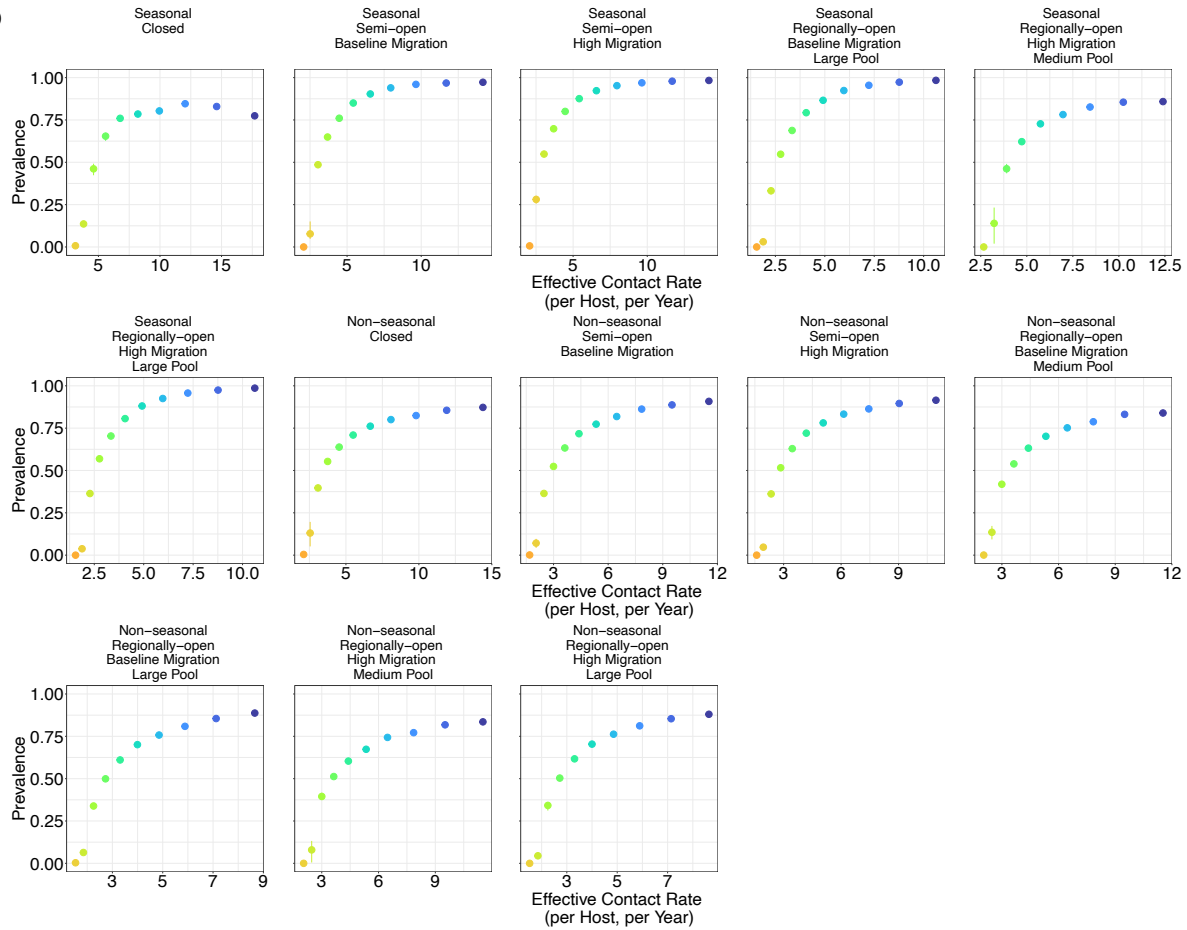
191 Mutation rate

192 The same study²⁵ estimated that a weighted average mutation rate of 4.07×10^{-10} ,
 193 3.63×10^{-10} , 3.78×10^{-10} SNPs per erythrocytic life cycle per nucleotide for 3D7, Dd2, and
 194 HB3 respectively. Based on the study by Childs et al.²⁶, we assume 5AA mutations will result in
 195 a new type. Each allele can represent either one HVR (25-57 bp)²⁴ or half of the exon 1 of *var*
 196 genes (around 2500 bp)²⁷. The average mutation rate per allele per day is of the order of 10^{-9} to
 197 10^{-7} . We assume it to be 10^{-8} in our simulation.

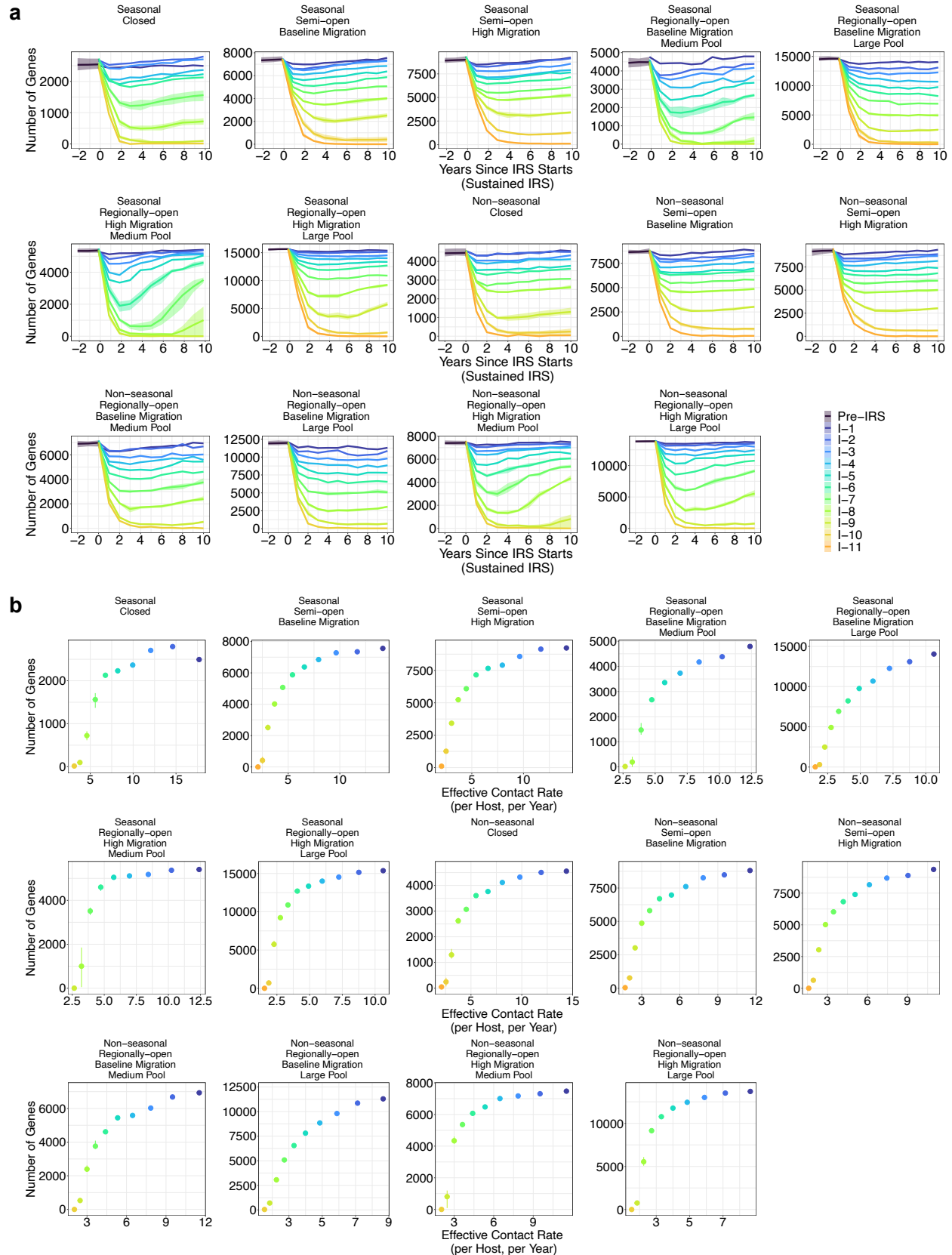
a



b

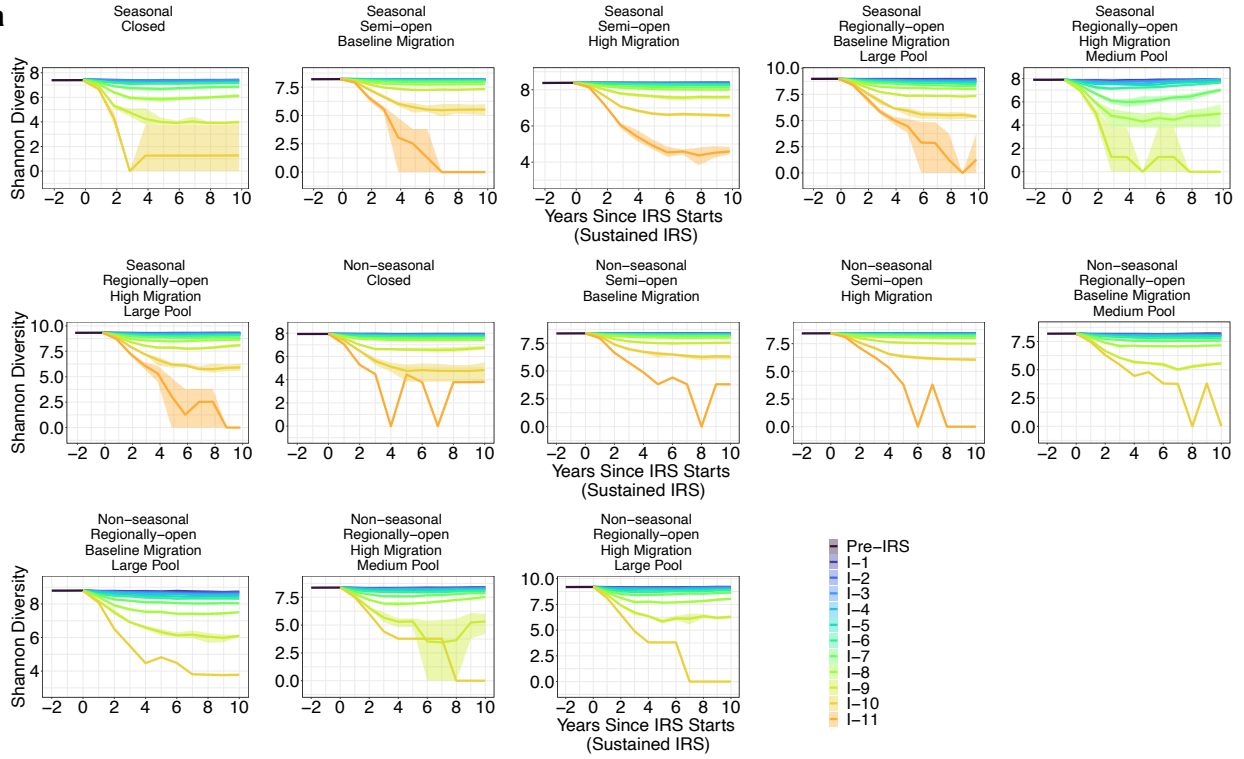


Supplementary Fig. 1. Systems' response in terms of prevalence levels to a series of sustained interventions. **a** Rebound dynamics of prevalence across time under a series of sustained IRS interventions, for closed, semi-open, and regionally-open systems simulated with additional combinations of parameter values than the one shown in Fig. 2a-b (Supplementary Table 1). The dynamics can be categorized into three regimes: the fast-rebound, the transition, and the slow-rebound regime. **b** Long-term rebound prevalence levels against transmission rates during sustained IRS interventions. Specifically, we plot prevalence levels at the 10th year into the sustained IRS intervention, sampled either at the end of wet/high-transmission season for seasonal transmission, or at the end of the year for constant transmission. In **a** and **b**, different colors indicate varying levels of intervention intensity (corresponding to those in Fig. 1c.), and ranges are derived based on three replicate runs.

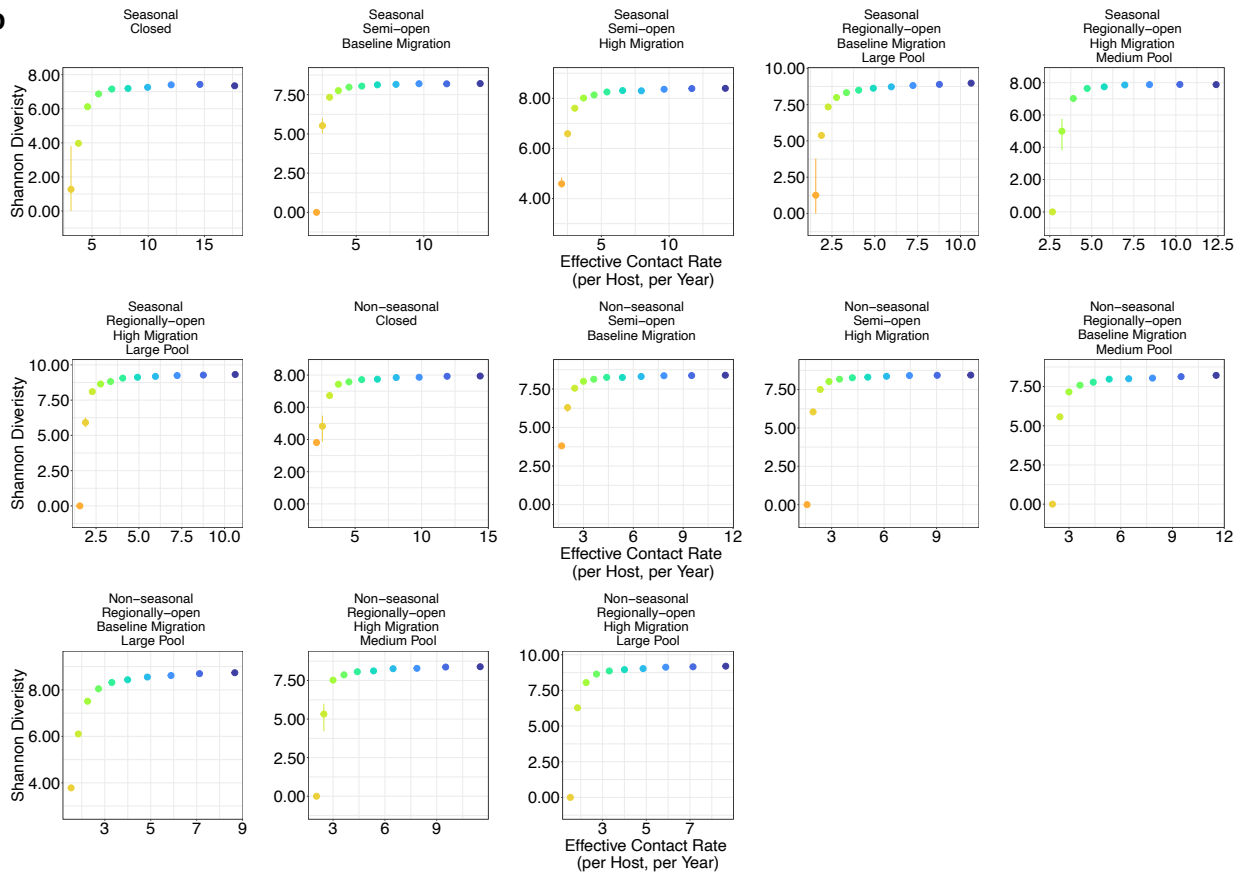


Supplementary Fig. 2. Systems' response in terms of *var* gene diversity measured as richness to a series of sustained interventions. **a** Rebound dynamics of *var* gene diversity measured as richness across time under a series of sustained IRS interventions, in various simulated scenarios (Supplementary Table 1). **b** Long-term rebound dynamics of richness against transmission rates. Specifically, we plot richness levels at the 10th year into the sustained IRS intervention, sampled either at the end of wet/high-transmission season for seasonal transmission, or at the end of the year for constant transmission. For semi-open systems, we plot the total richness of the two explicitly coupled local parasite populations as we sample both populations. For closed and regionally-open systems, we plot the richness of the focal parasite population only. In **a** and **b**, different colors indicate varying levels of intervention intensity (corresponding to those in Fig. 1c.), and ranges are derived based on three replicate runs.

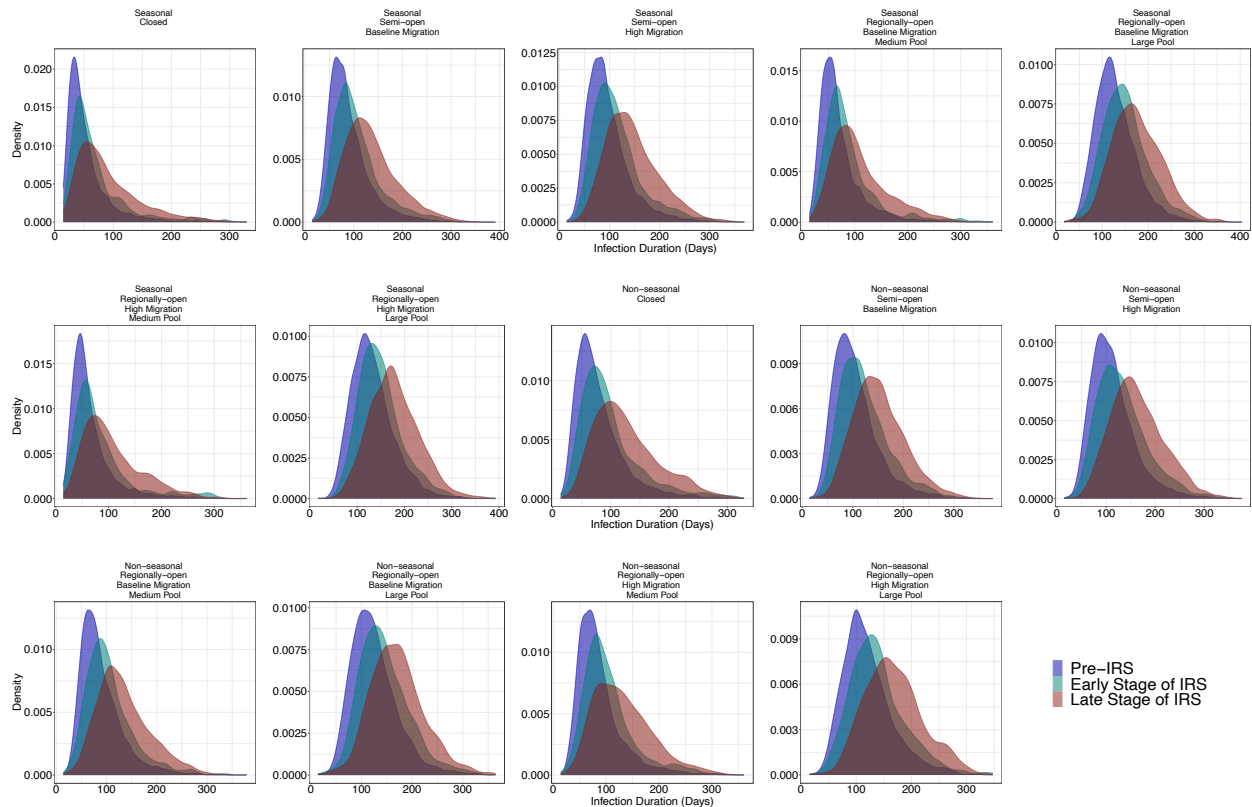
a



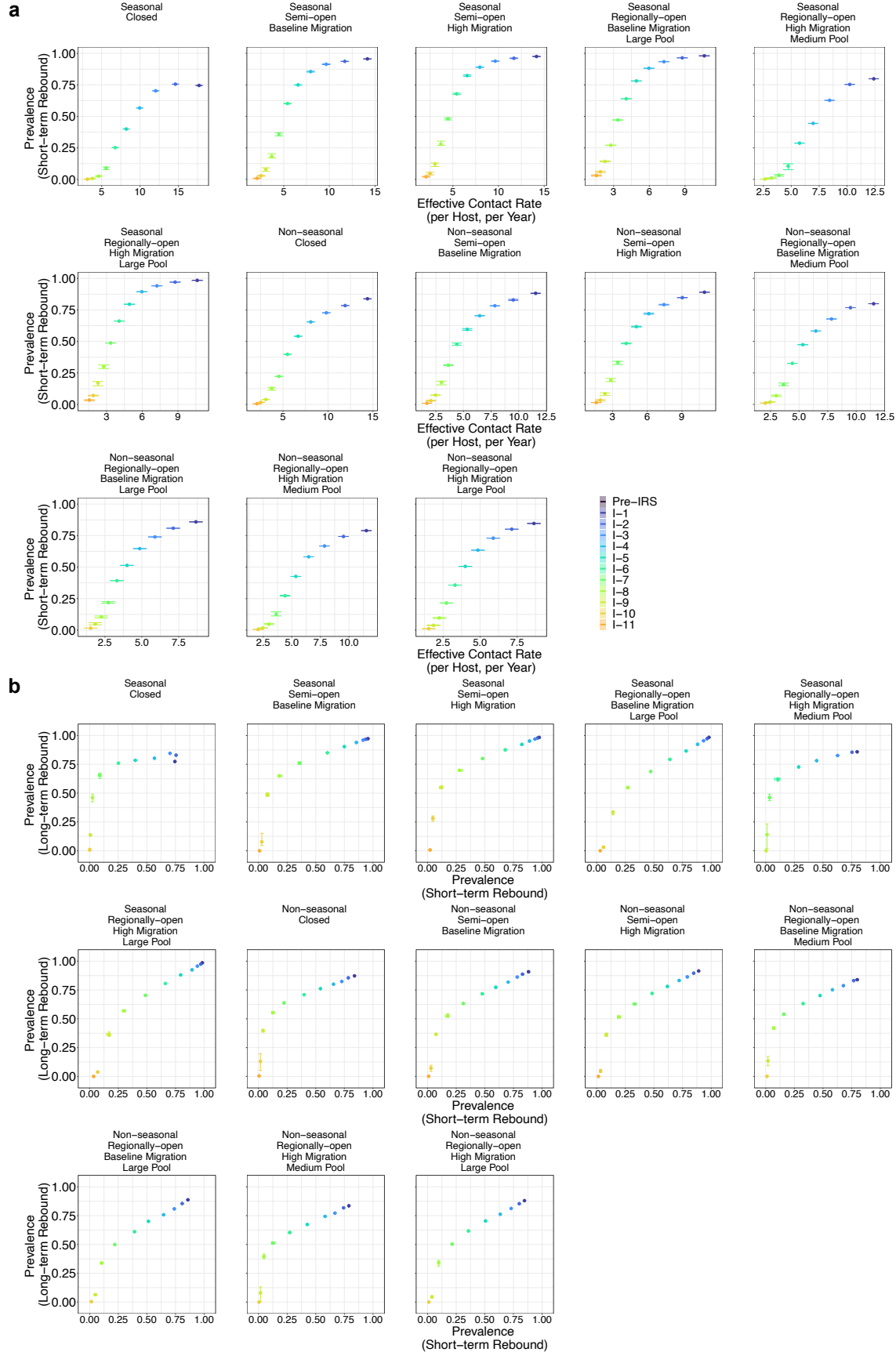
b



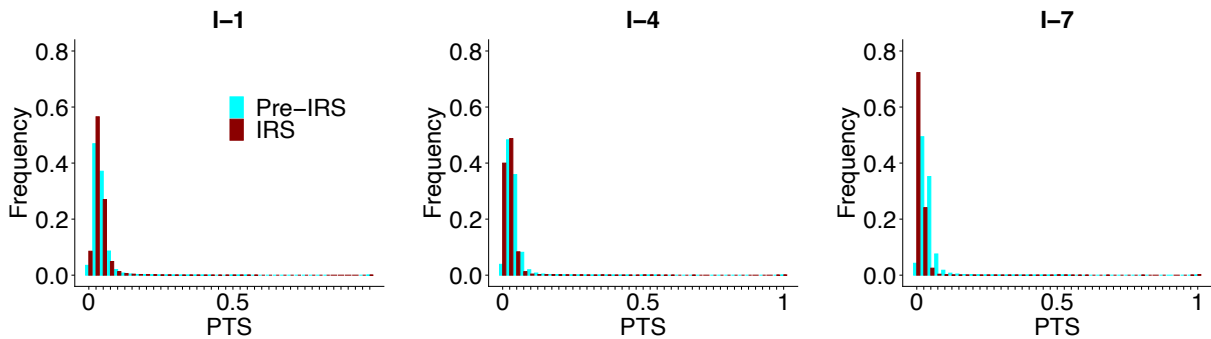
Supplementary Fig. 3. Systems' response in terms of *var* gene diversity measured by the Shannon Diversity Index to a series of sustained interventions. a Rebound dynamics of *var* gene diversity measured by the Shannon Diversity Index across time under a series of sustained IRS interventions, for closed, semi-open, and regionally-open systems simulated with additional combinations of parameter values than the one shown in Fig. 2c-d (Supplementary Table 1). **b** Long-term rebound dynamics of the Shannon Diversity Index against transmission rates. Specifically, we plot diversity levels at the 10th year into the sustained IRS intervention, sampled either at the end of wet/high-transmission season for seasonal transmission, or at the end of the year for constant transmission. For semi-open systems, we plot the Shannon Diversity Index of the two explicitly coupled local parasite populations as we sample both populations. For closed and regionally-open systems, we plot the Shannon Diversity Index of the focal parasite population only. Prevalence in Supplementary Fig. 1, *var* gene diversity measured as richness in Supplementary Fig. 2, and *var* gene diversity measured by the Shannon Diversity Index here all demonstrate a non-linear behavior as a function of the transmission rates. The fast-rebound regime occurs for a wide range of transmission rates during intervention, whereas the transition and slow-rebound regimes occur for a narrow region along the gradient of transmission rates. In **a** and **b**, different colors indicate varying levels of intervention intensity (corresponding to those in Fig. 1c.), and ranges are derived based on three replicate runs.



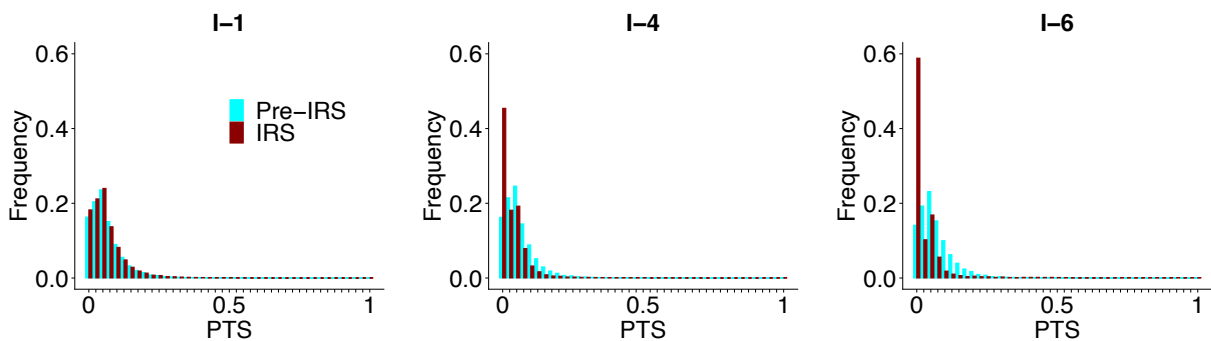
Supplementary Fig. 4. The shift in distributions of infection duration from before to immediately following to towards the end of a particular sustained IRS intervention, in various simulated scenarios (Supplementary Table 1). For illustration purposes, we present here I-4 (for its level of transmission intensity, see Fig. 1c). Specifically, we consider infections which start and end within three years before intervention, those which start and end within three years after intervention is applied, and those which start and end within the last three years of the decade-long intervention phase. The distribution keeps shifting to the right towards longer values throughout the intervention phase relative to that before intervention. This pattern also holds for intervention runs of different coverages than I-4 shown here. The degree of the shift increases as the coverage of the intervention increases. Note that in general duration infection is longer for systems with a higher circulating diversity, for example, regionally-open systems with a large pool, than systems with a lower circulating diversity, for example, closed systems. This is intuitive as high circulating diversity promotes immune evasion of parasites and lengthens infection duration.



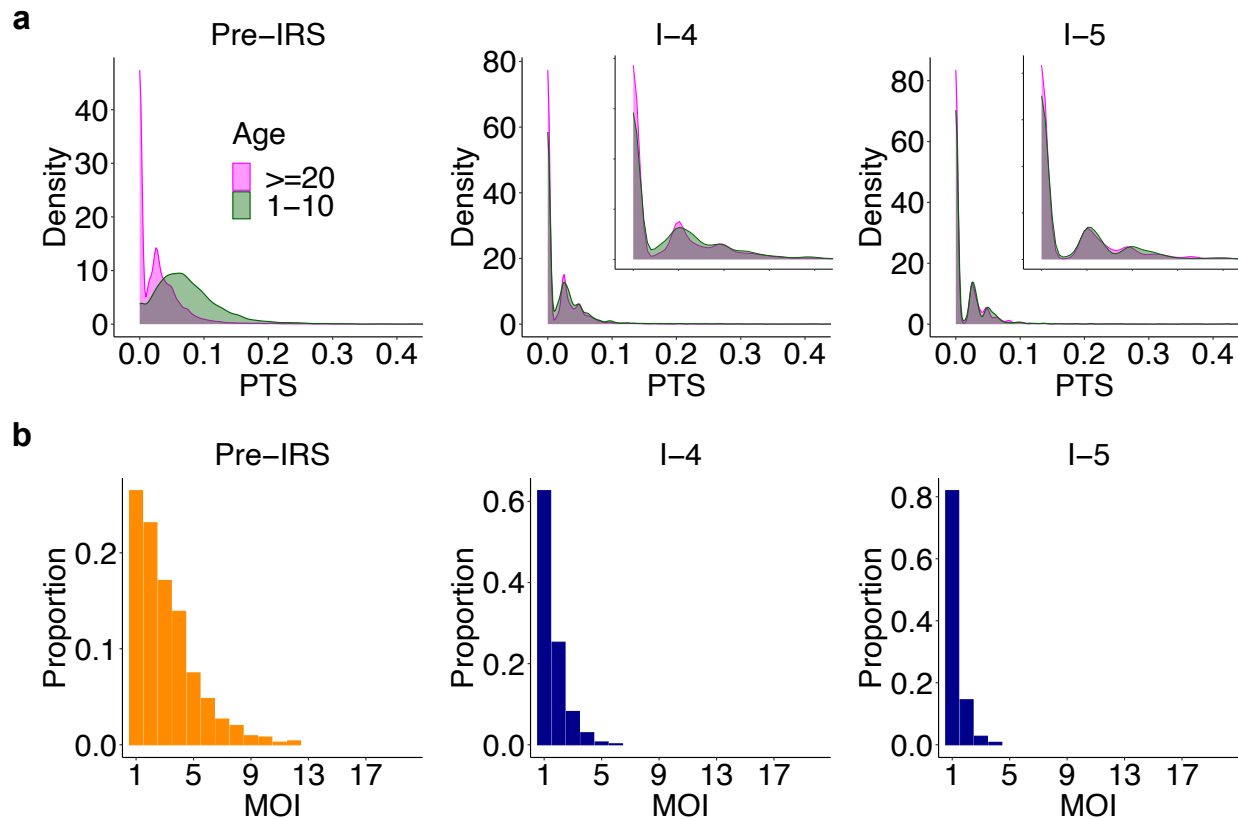
Supplementary Fig. 5. Systems' short-term response, and the relationship between short-term and long-term responses to a series of sustained interventions. **a** Short-term rebound prevalence levels against transmission rates of the series of sustained intervention in various simulated scenarios (Supplementary Table 1). We specifically plot prevalence levels at the 2nd year into the series of sustained intervention (sampled at the end of wet/high-transmission season for seasonal transmission and at the end of the 2nd year for constant transmission). **b** The relationship between short-term and long-term (at the 10th year into intervention) rebound prevalence levels. 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. In **a** and **b**, different colors indicate varying levels of intervention intensity (corresponding to those in Fig. 1c.), and ranges are derived based on three replicate runs.



Supplementary Fig. 6. The emergence of the second mode in PTS distribution is sensitive to the specification of parameters. The emergence of the second mode at complete overlap becomes insignificant when intervening systems with a high circulating diversity. For example, seasonal and regionally-open ones with a high migration rate and a large regional pool (scenario X in Supplementary Table 1). Plotted here are PTS values calculated based on true samples, i.e., the ones without sampling limitations (missing data issue and measurement error) representative of those encountered in the collection of field data.

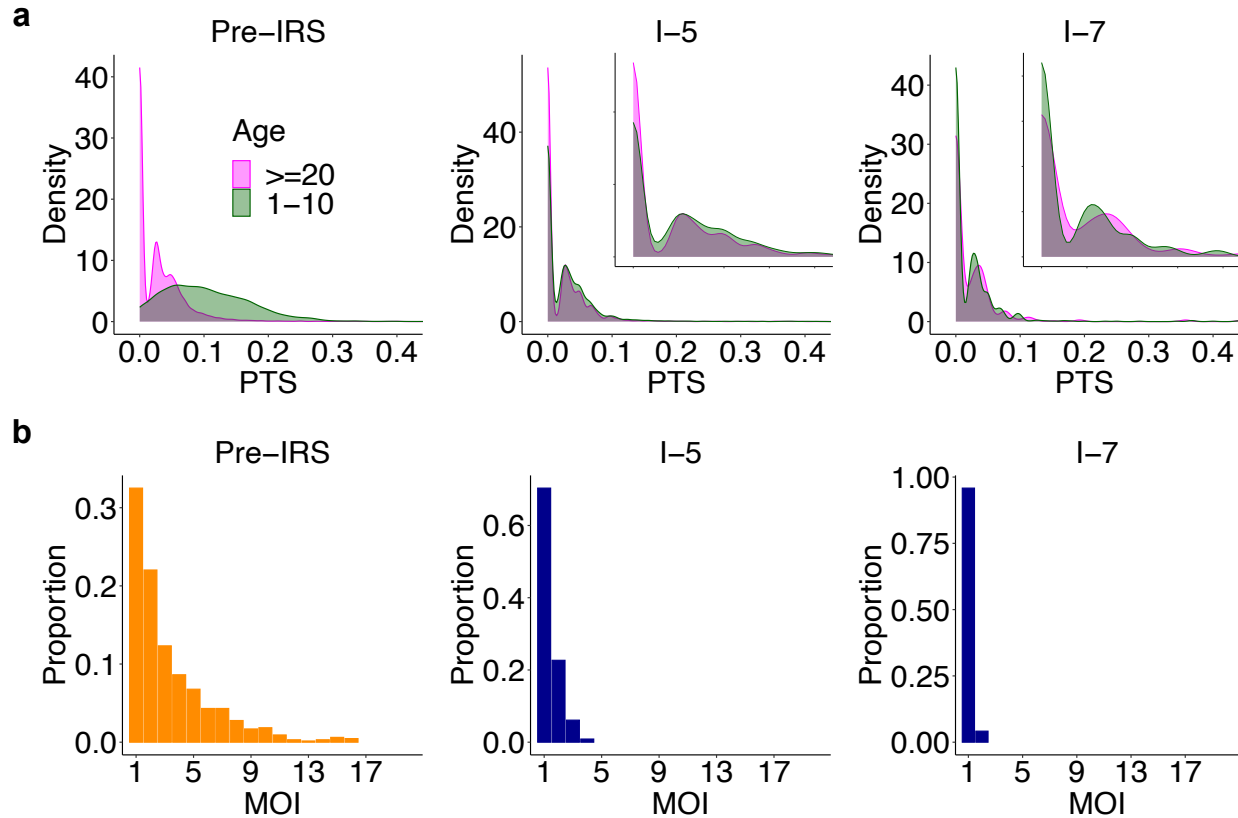


Supplementary Fig. 7. The emergence of the second mode in PTS distribution is not robust against sampling limitations. The pairwise type sharing or PTS distribution during sustained intervention for regionally-open systems with a baseline migration rate and a medium pool (scenario VII in Supplementary Table 1). When sampling the simulation output for PTS calculation, we implemented sampling schemes and sampling limitations representative of those encountered in the collection of field data. The emergence of the second mode disappears (see section “Sampling schemes and measurement error of empirical data” in Methods). The case without sampling limitations (missing data issue and measurement error) is shown in Fig. 3d.

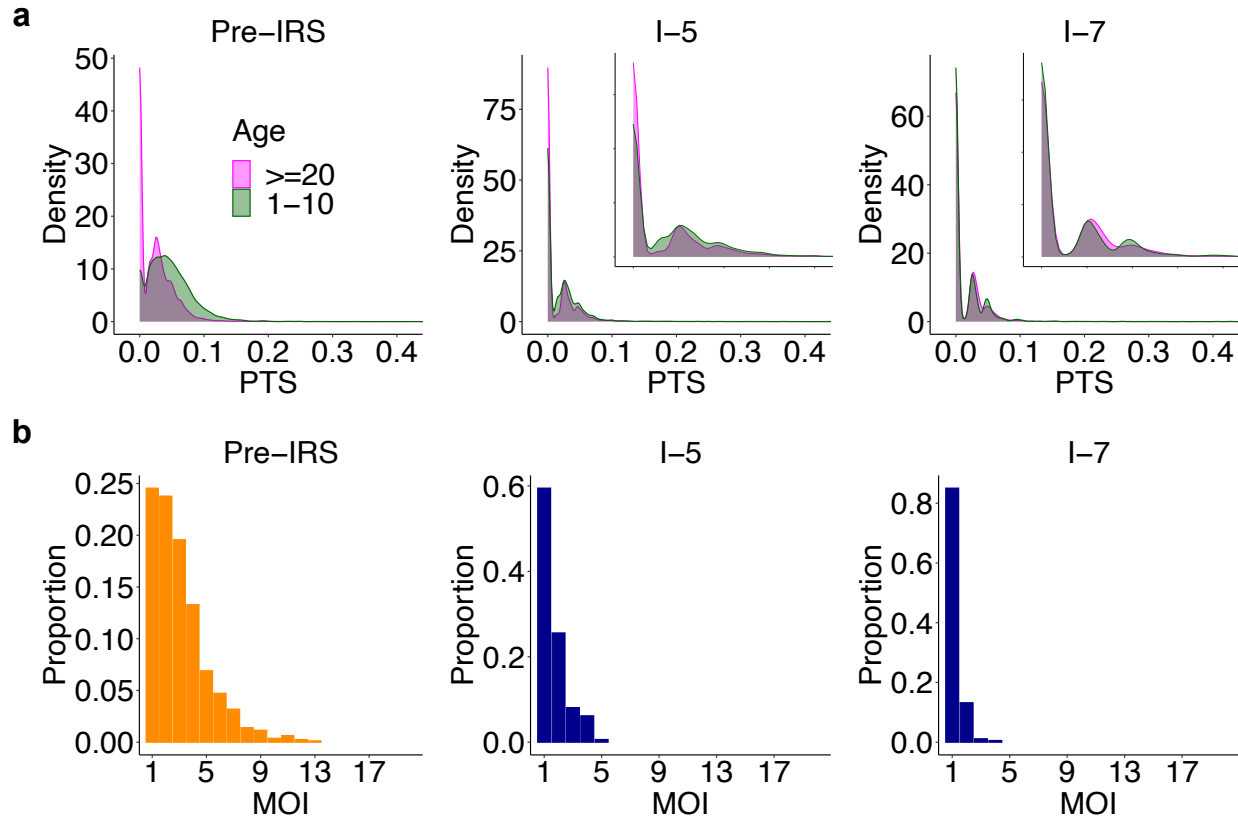


Supplementary Fig. 8. Molecular indicators of the proximity of the perturbed local system to the transition regime for seasonal regionally-open systems with a medium pool and a baseline migration rate (scenario VII in Supplementary Table 1). For illustration purposes, we show the pre-IRS case, an example 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-genomics with two genetically distinct parasites. These two molecular indicators are robust under implemented sampling schemes (sampling frequency and depth) and sampling limitations representative of those encountered in the

collection of field data (the missing data issue and measurement error). The case without sampling limitations (the missing data issue and measurement error) is shown in Fig. 4.

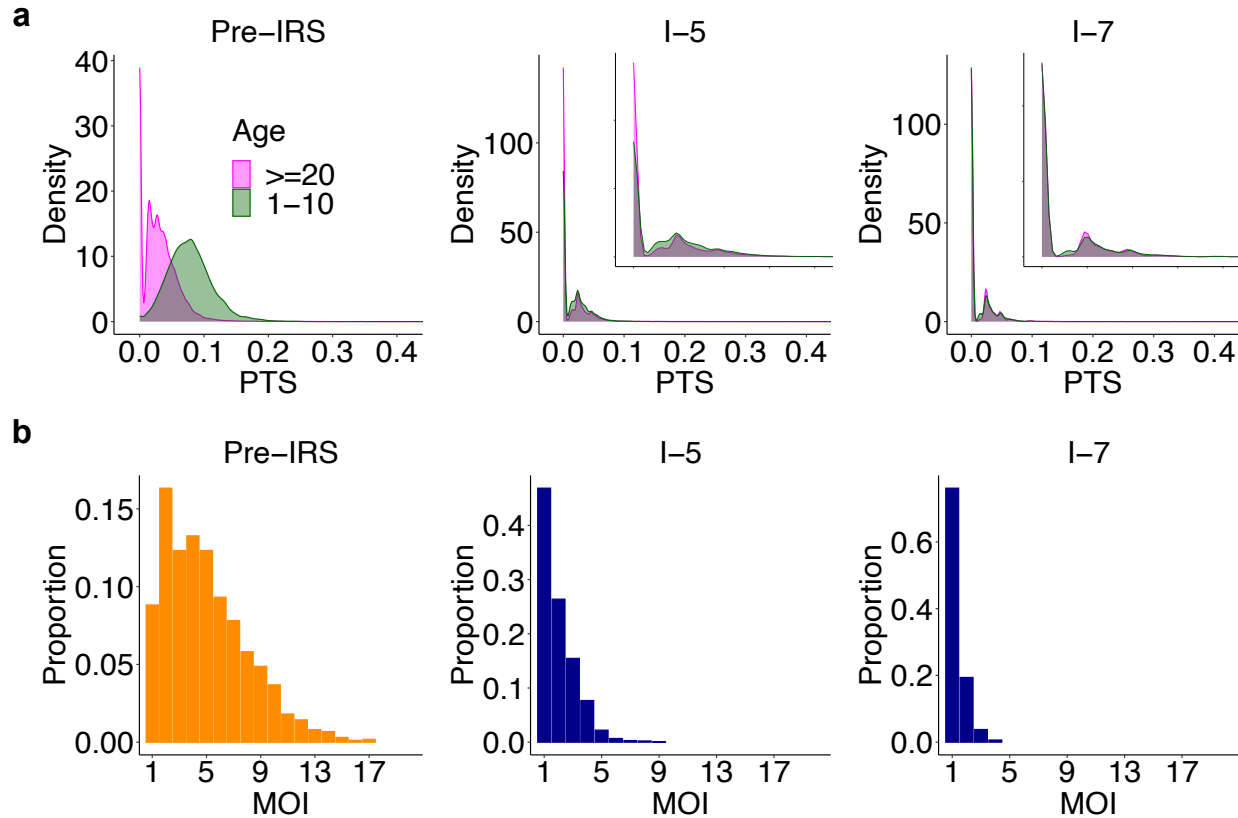


Supplementary Fig. 9. The two molecular indicators for seasonal closed systems (scenario I in Supplementary Table 1). For illustration purposes, we show the pre-IRS case, an example IRS under which the system approaches the transition regime, and another example IRS under which the system falls into 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 (sampling frequency and depth) and sampling limitations representative of those encountered in the collection of field data (the missing data issue and measurement error).

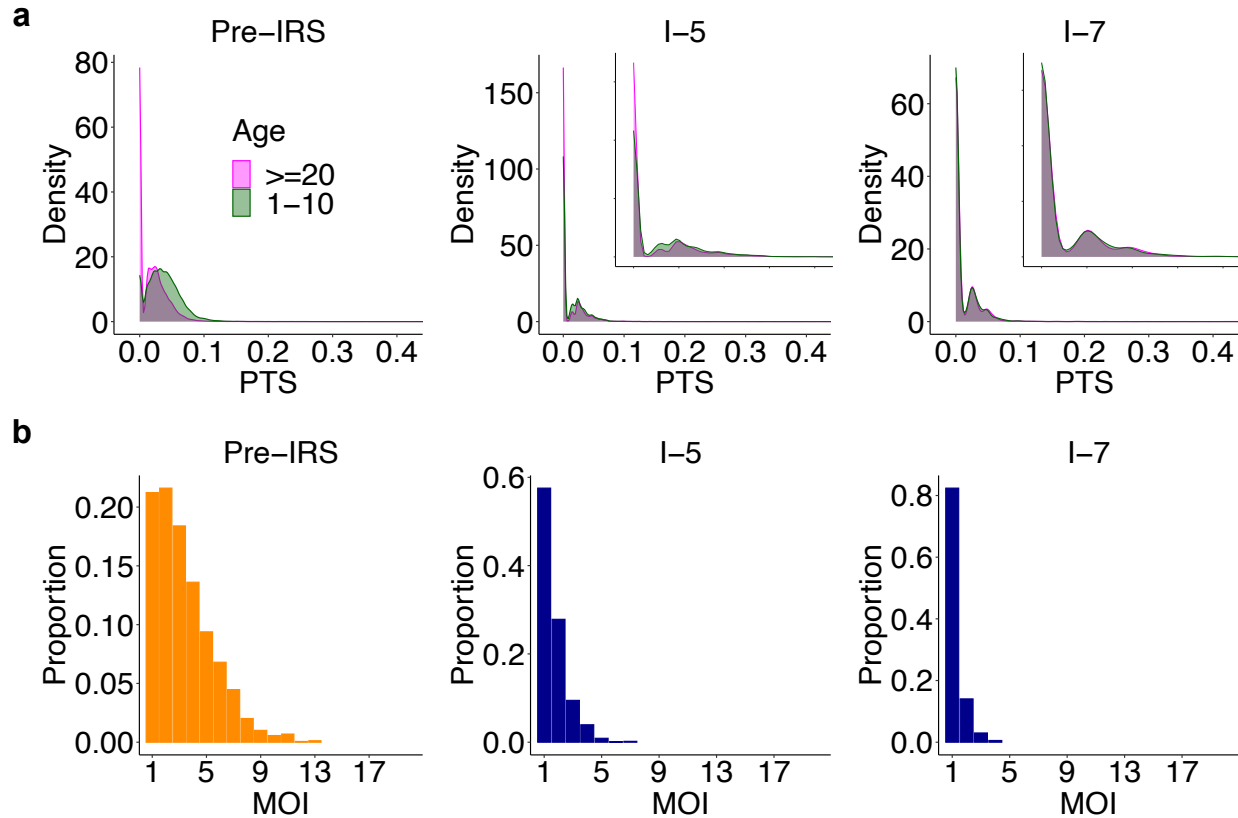


Supplementary Fig. 10. The two molecular indicators for non-seasonal closed systems

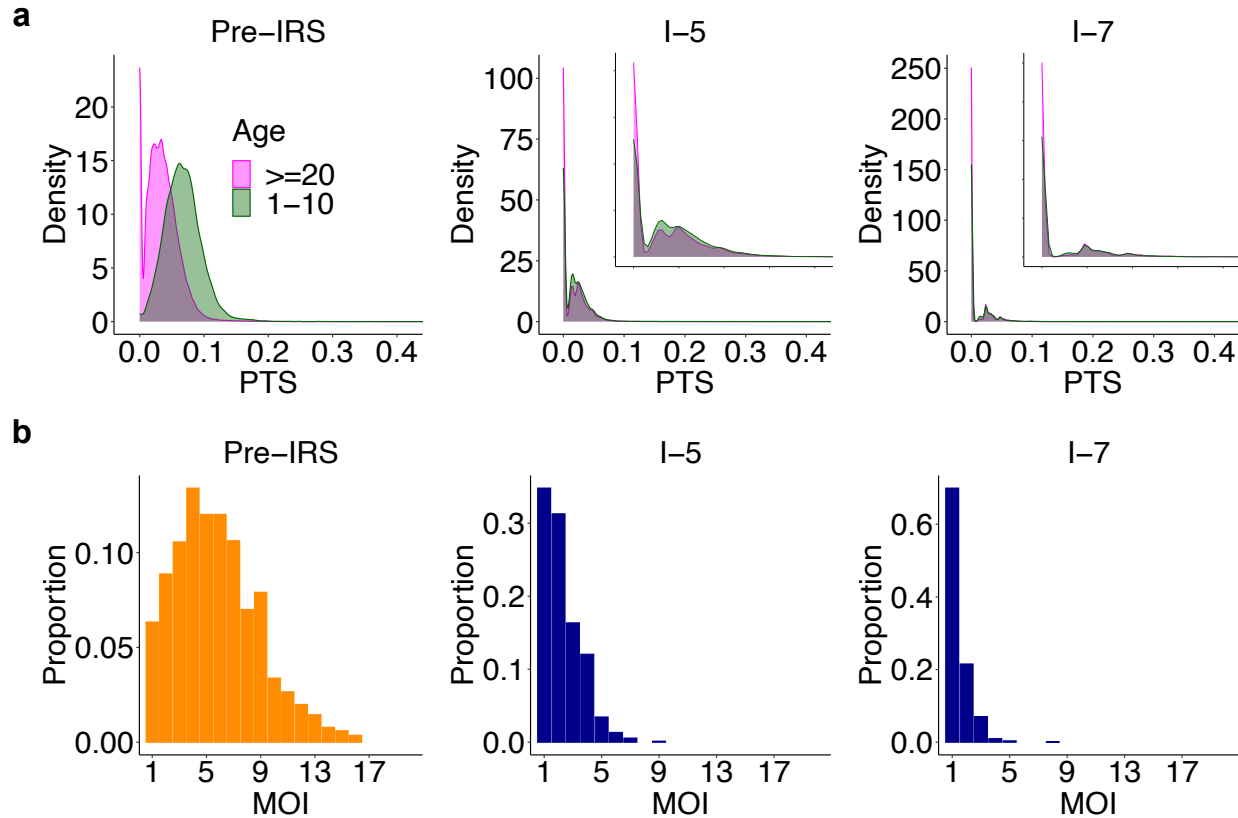
(scenario II in Supplementary Table 1). For illustration purposes, we show the pre-IRS case, an example IRS under which the system approaches the transition regime, and another example IRS under which the system falls into the transition or slow-rebound regime. **a** As the system approaches the transition 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 (sampling frequency and depth) and sampling limitations representative of those encountered in the collection of field data (the missing data issue and measurement error).



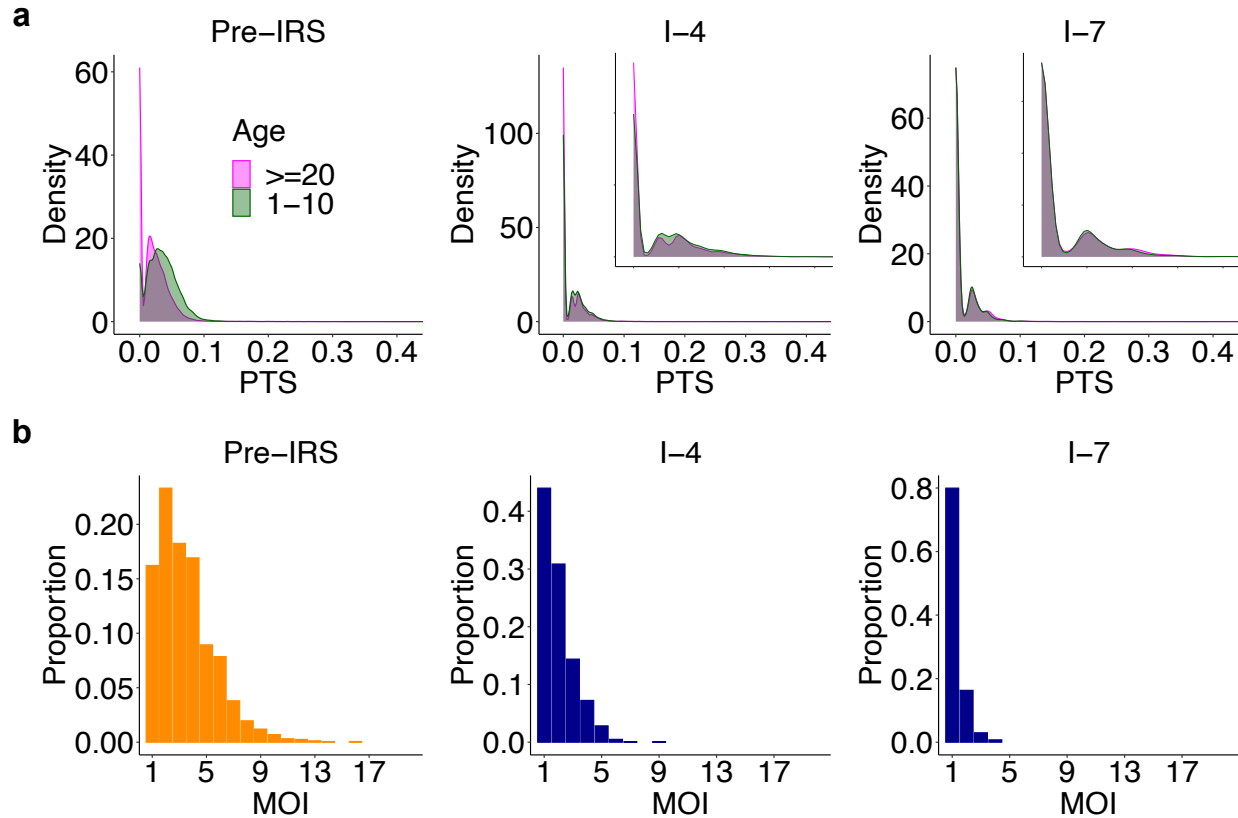
Supplementary Fig. 11. The two molecular indicators for seasonal semi-open systems with a baseline migration rate (scenario III in Supplementary Table 1). For illustration purposes, we show the pre-IRS case, an example IRS under which the system approaches the transition regime, and another example IRS 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 (sampling frequency and depth) and sampling limitations representative of those encountered in the collection of field data (the missing data issue and measurement error).



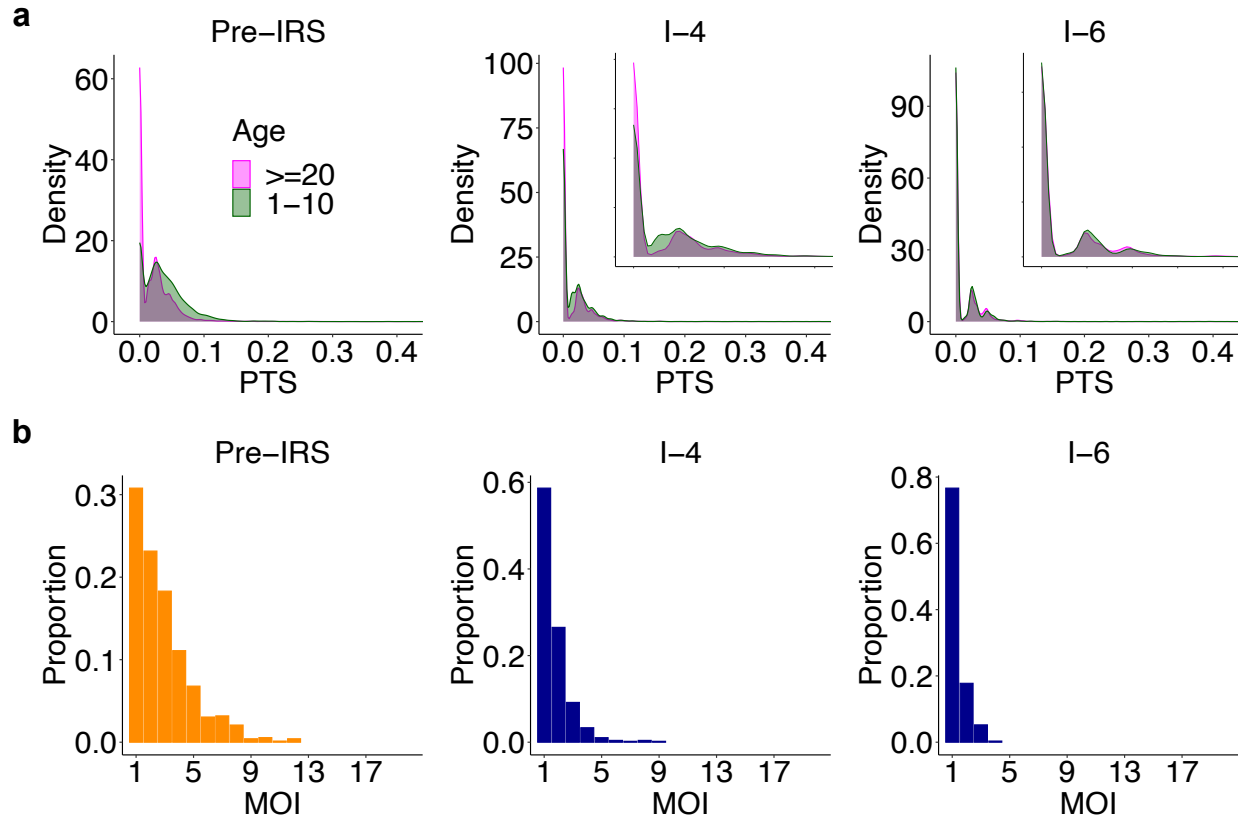
Supplementary Fig. 12. The two molecular indicators for non-seasonal semi-open systems with a baseline migration rate (scenario V in Supplementary Table 1). For illustration purposes, we show the pre-IRS case, an example IRS under which the system approaches the transition regime, and another example 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 parasite. These two molecular indicators are robust under implemented sampling schemes (sampling frequency and depth) and sampling limitations representative of those encountered in the collection of field data (the missing data issue and measurement error).



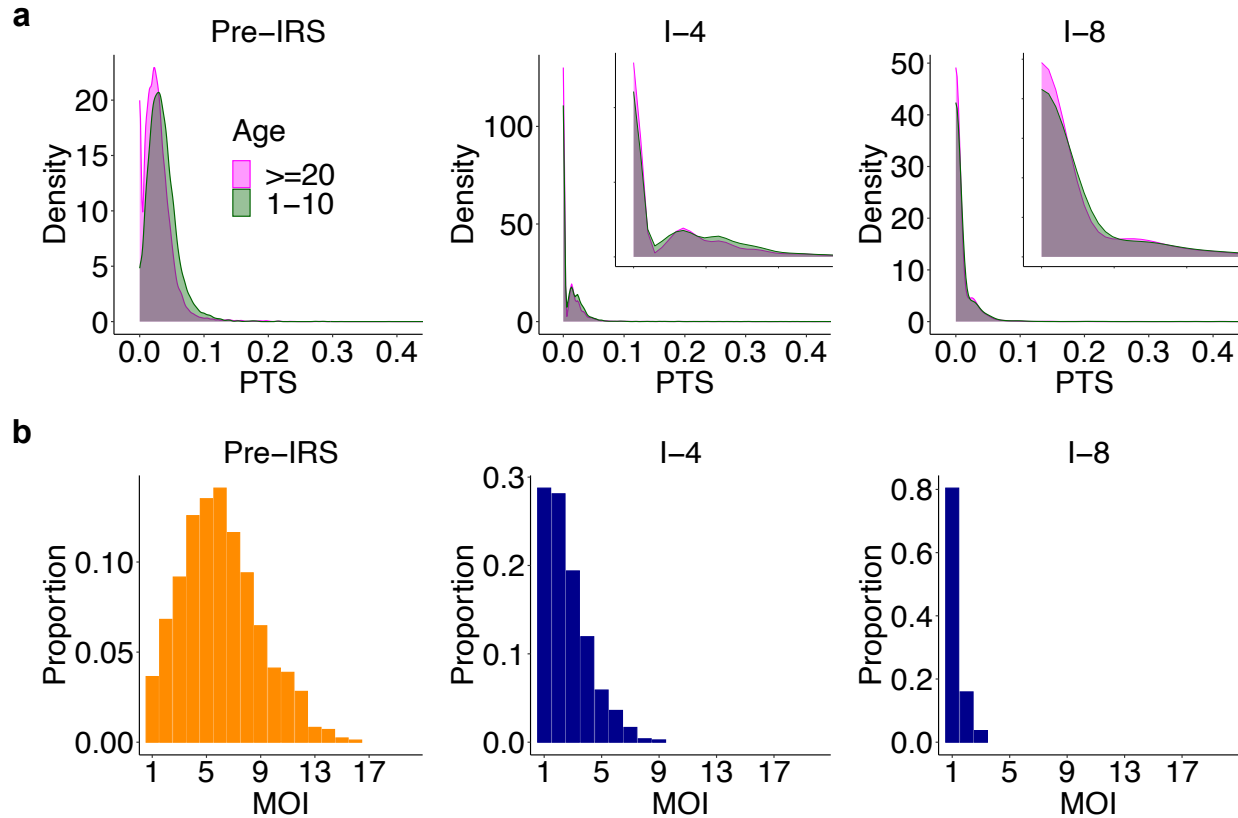
Supplementary Fig. 13. The two molecular indicators for seasonal semi-open systems with a high migration rate (scenario IV in Supplementary Table 1). For illustration purposes, we show the pre-IRS case, an example IRS under which the system approaches the transition regime, and another example IRS under 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 (sampling frequency and depth) and sampling limitations representative of those encountered in the collection of field data (the missing data issue and measurement error).



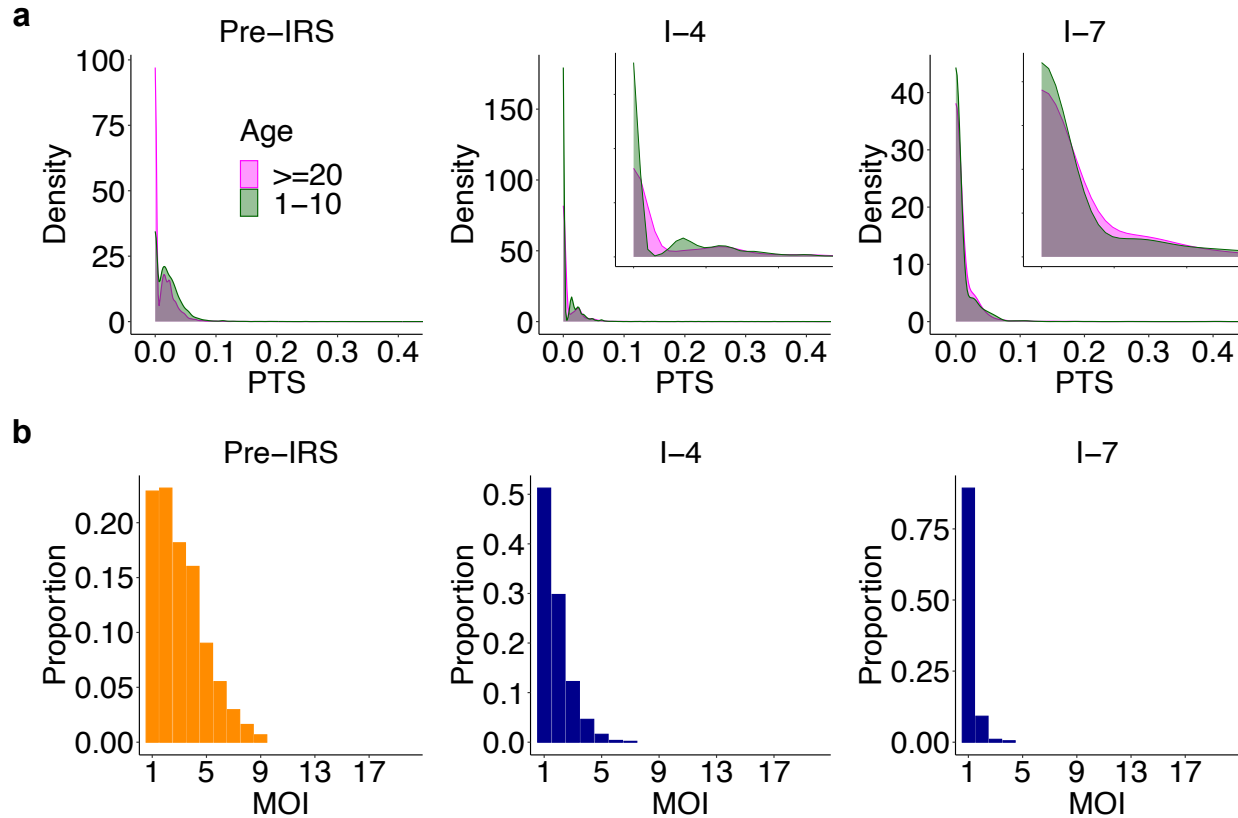
Supplementary Fig. 14. The two molecular indicators for non-seasonal semi-open system with a high migration rate (scenario VI in Supplementary Table 1). For illustration purposes, we show the pre-IRS case, an example IRS under which the system approaches the transition regime, and another example IRS 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 (sampling frequency and depth) and sampling limitations representative of those encountered in the collection of field data (the missing data issue and measurement error).



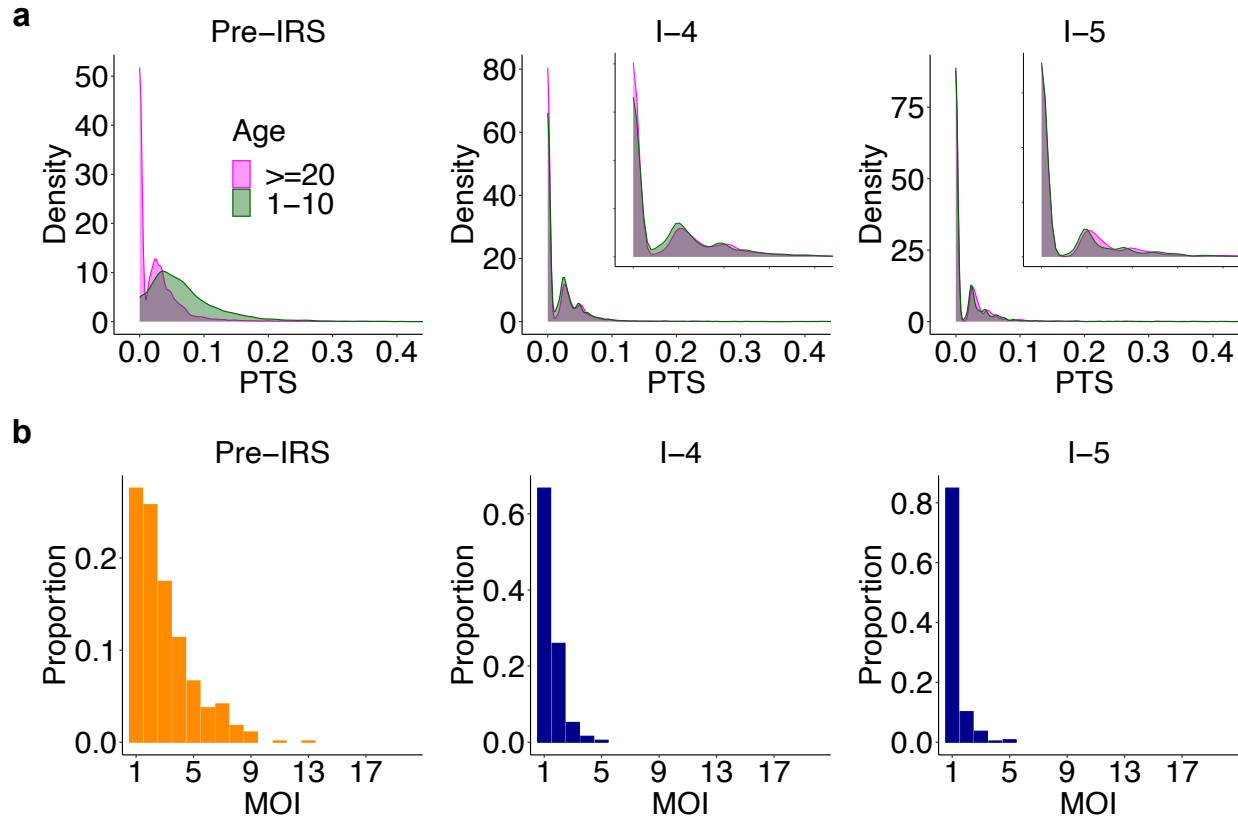
Supplementary Fig. 15. The two molecular indicators for non-seasonal regionally-open systems with a baseline migration rate and a medium regional pool (scenario XI in Supplementary Table 1). For illustration purposes, we show the pre-IRS case, an example IRS under which the system approaches the transition 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 (sampling frequency and depth) and sampling limitations representative of those encountered in the collection of field data (the missing data issue and measurement error).



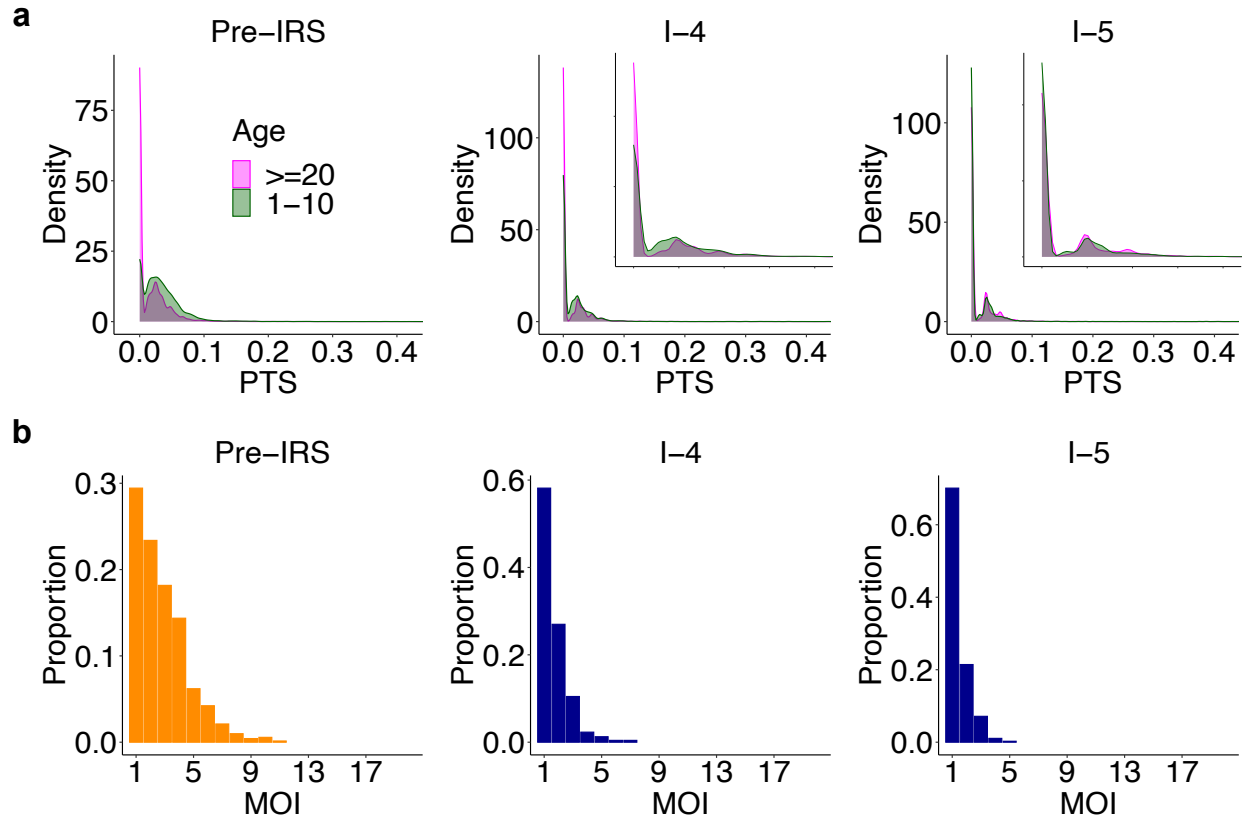
Supplementary Fig. 16. The two molecular indicators for seasonal regionally-open systems with a baseline migration rate and a large regional pool (scenario VIII in Supplementary Table 1). For illustration purposes, we show the pre-IRS case, an example IRS under which the system approaches the transition 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 0-0.05. **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 (sampling frequency and depth) and sampling limitations representative of those encountered in the collection of field data (the missing data issue and measurement error).



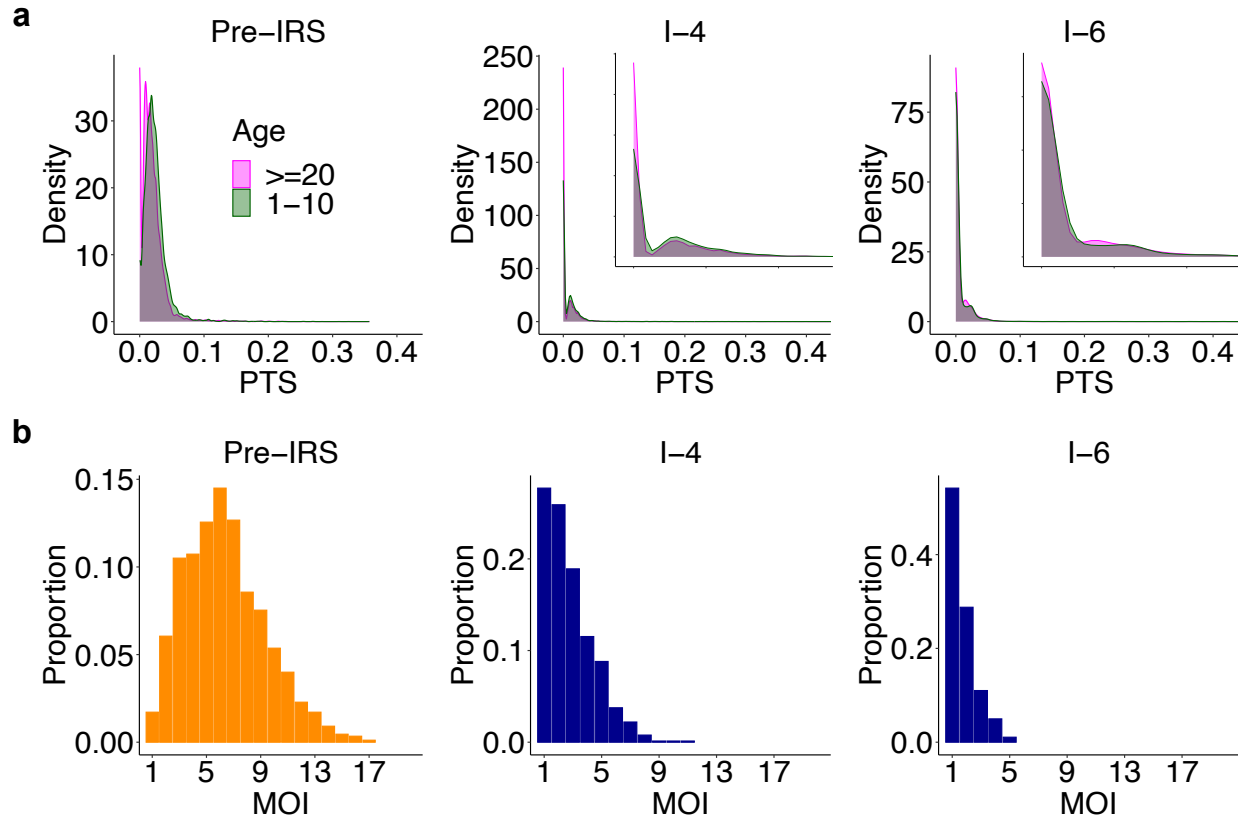
Supplementary Fig. 17. The two molecular indicators for non-seasonal regionally-open systems with a baseline migration rate and a large regional pool (scenario XII in Supplementary Table 1). For illustration purposes, we show the pre-IRS case, an example IRS under which the system approaches the transition 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 reversed. In particular, the PTS distribution of children develops an even higher peak around 0 than that of adults, the opposite of pre-IRS. The x-axis range for inset figures is 0-0.05. **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 (sampling frequency and depth) and sampling limitations representative of those encountered in the collection of field data (the missing data issue and measurement error).



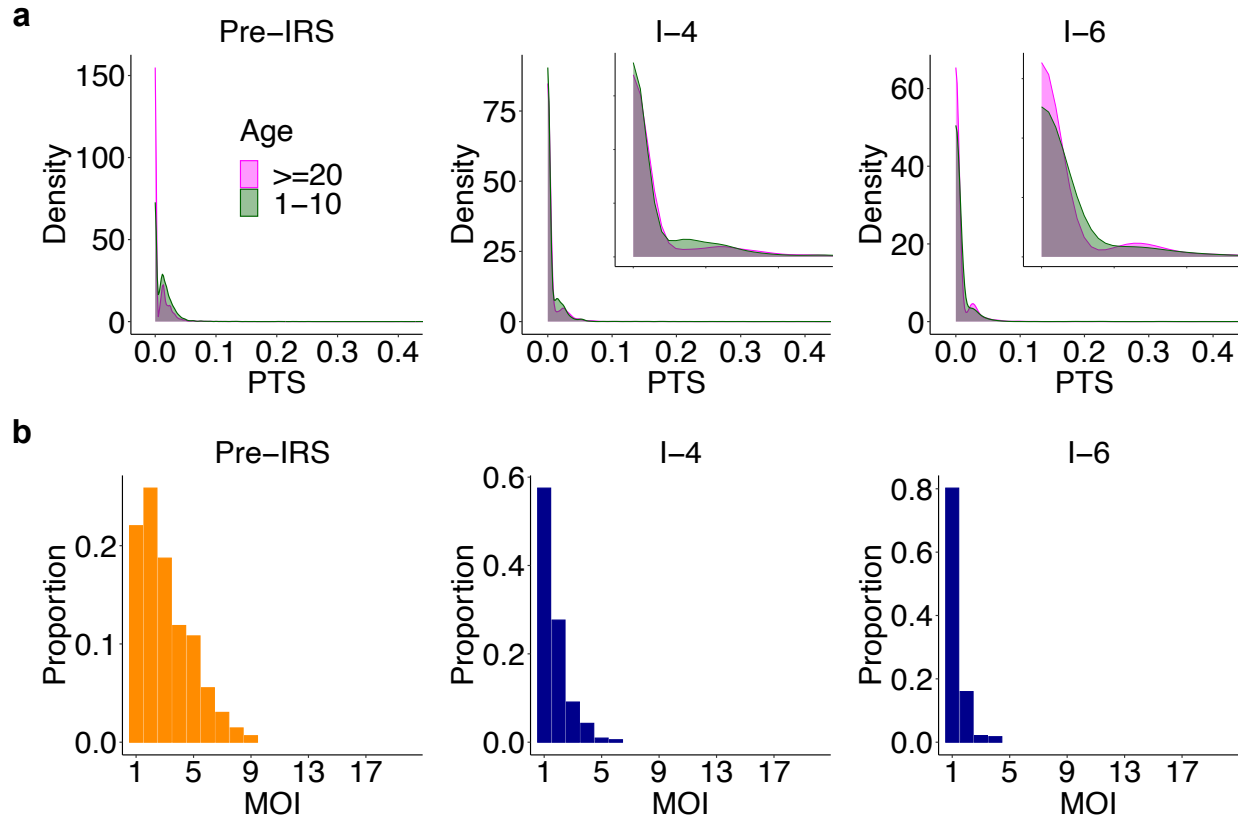
Supplementary Fig. 18. The two molecular indicators for seasonal regionally-open systems with a high migration rate and a medium regional pool (scenario IX in Supplementary Table 1). For illustration purposes, we show the pre-IRS case, an example IRS under which the system approaches the transition regime, 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 (sampling frequency and depth) and sampling limitations representative of those encountered in the collection of field data (the missing data issue and measurement error).



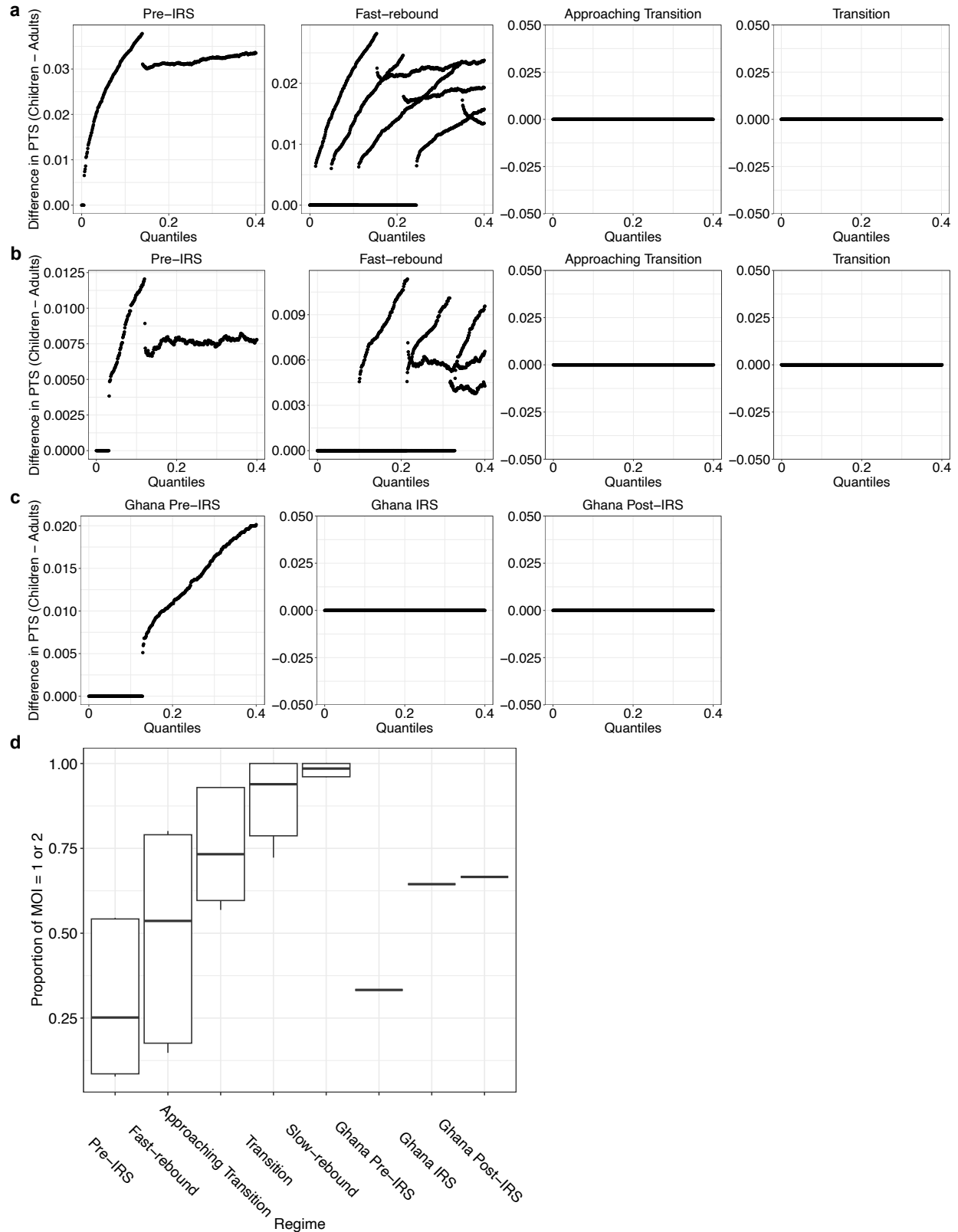
Supplementary Fig. 19. The two molecular indicators for non-seasonal regionally-open systems with a high migration rate and a medium regional pool (scenario XIII in Supplementary Table 1). For illustration purposes, we show the pre-IRS case, an example IRS under which the system approaches the transition 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 (sampling frequency and depth) and sampling limitations representative of those encountered in the collection of field data (the missing data issue and measurement error).



Supplementary Fig. 20. The two molecular indicators for seasonal regionally-open systems with a high migration rate and a large regional pool (scenario X 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 0-0.05. **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 (sampling frequency and depth) and sampling limitations representative of those encountered in the collection of field data (the missing data issue and measurement error).

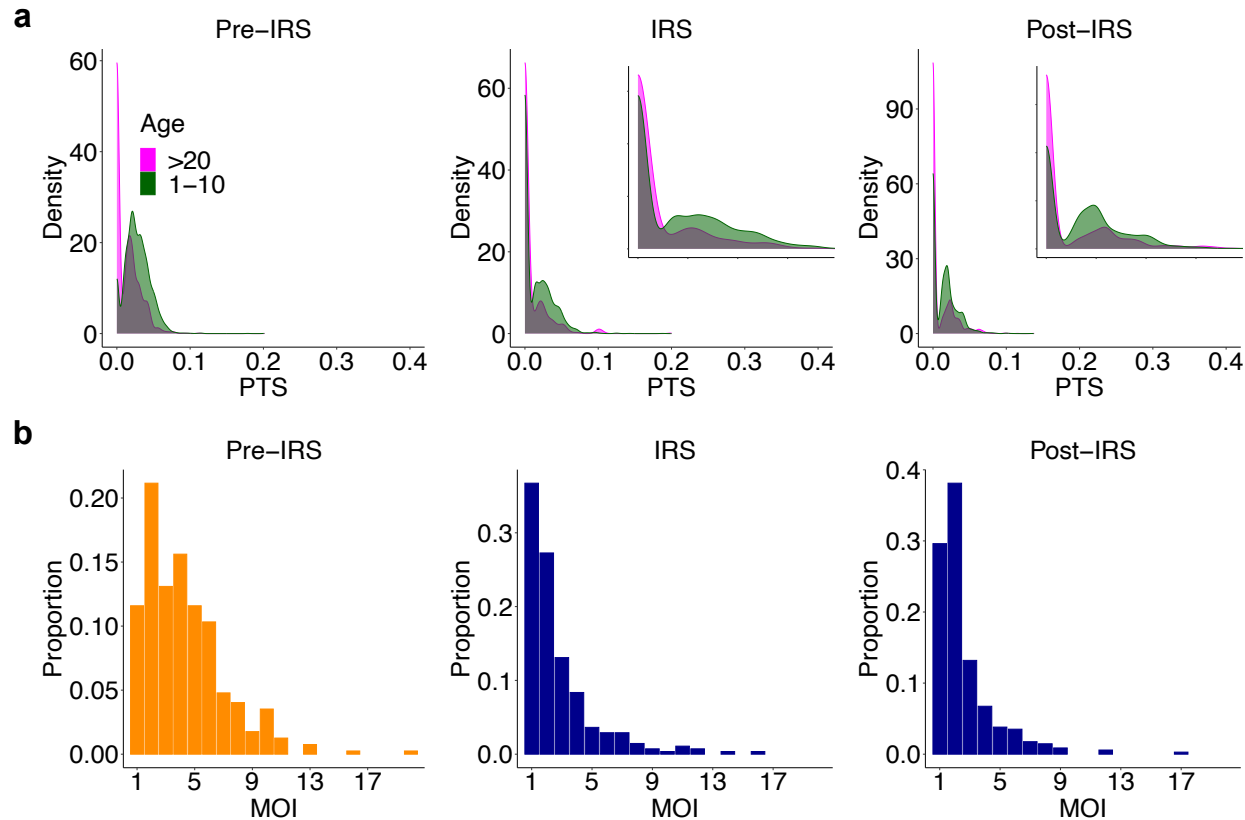


Supplementary Fig. 21. The two molecular indicators for non-seasonal regionally-open systems with a high migration rate and a large regional pool (scenario XIV in Supplementary Table 1). For illustration purposes, we show the pre-IRS case, an example IRS under which the system approaches the transition 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 0-0.05. **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 (sampling frequency and depth) and sampling limitations representative of those encountered in the collection of field data (the missing data issue and measurement error).



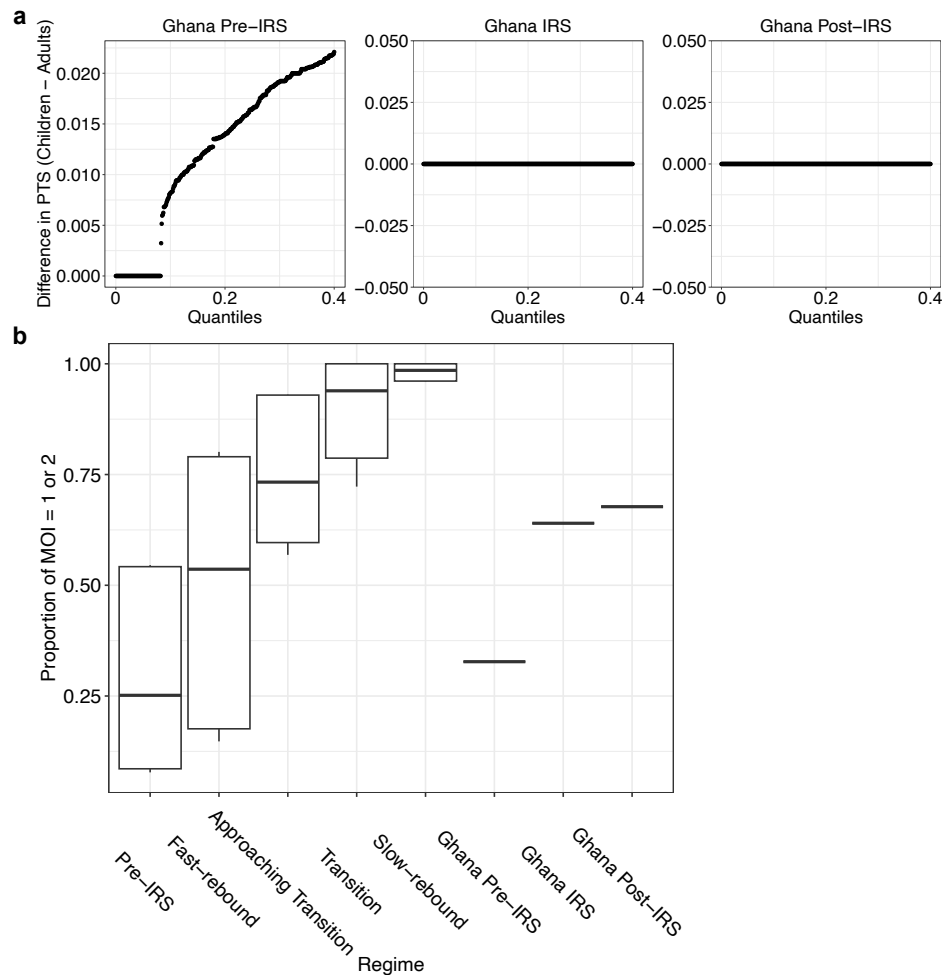
Supplementary Fig. 22. The quantitative patterns of the two indicators, i.e., the difference in the PTS distributions across the two age groups, and the proportion of monoclonal

infections or multi-genomic infections with two distinct strains ($\text{MOI} = 1$ or 2), when the system is at high-transmission pre-IRS, and pushed into the fast-rebound regime, approaching or entering the transition regime during IRS. Specifically, we calculate the difference in PTS values across the two age groups at individual, consecutive quantiles within the range of 0-0.40. **a** The difference in PTS for a seasonal and semi-open system with a high migration rate (an order of magnitude higher than the estimated baseline value) (Methods, scenario IV in Supplementary Table 1). **b** The difference in PTS for a seasonal and regionally-open system with a baseline migration rate and a large regional pool (Methods, scenario VIII in Supplementary Table 1). **c** The difference in PTS for Ghana surveys from pre-, to during, to post-IRS. **d** The proportion of monoclonal infections or multi-genomic infections with two distinct strains ($\text{MOI} = 1$ or 2) is shown for both simulated systems at the high-transmission pre-IRS phase, the ones which are pushed by IRS into the fast-rebound, approaching or entering the transition, or slow-rebound regime, and empirical transmission system in Ghana across the three-round transient IRS.

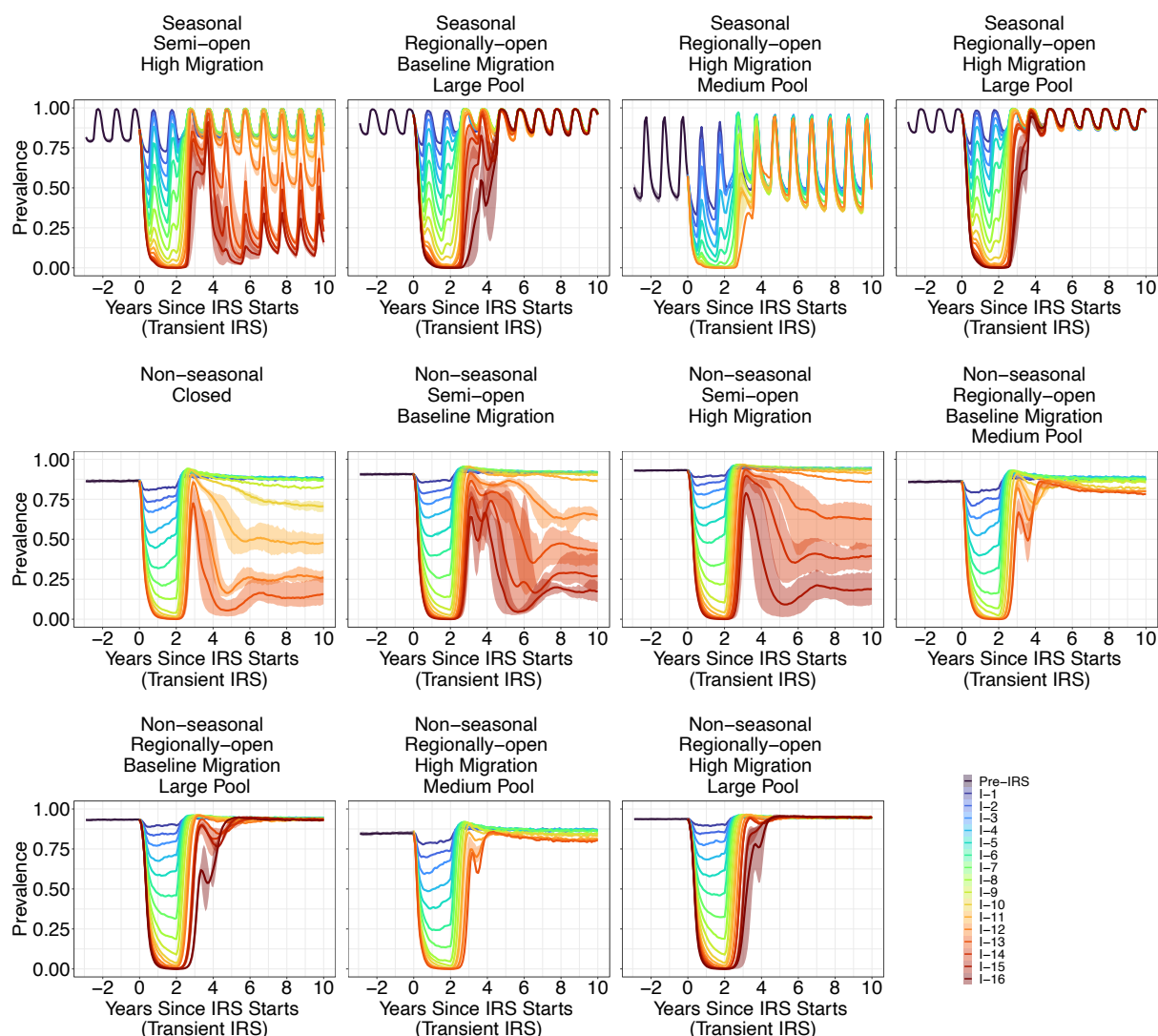


Supplementary Fig. 23. The two molecular indicators applied to the field data from northern Ghana with curative drug treated individuals removed before, during, and immediately after the three-round transient IRS intervention, indicate proximity to the transition region where rebound prevalence decreases sharply with intervention intensity.

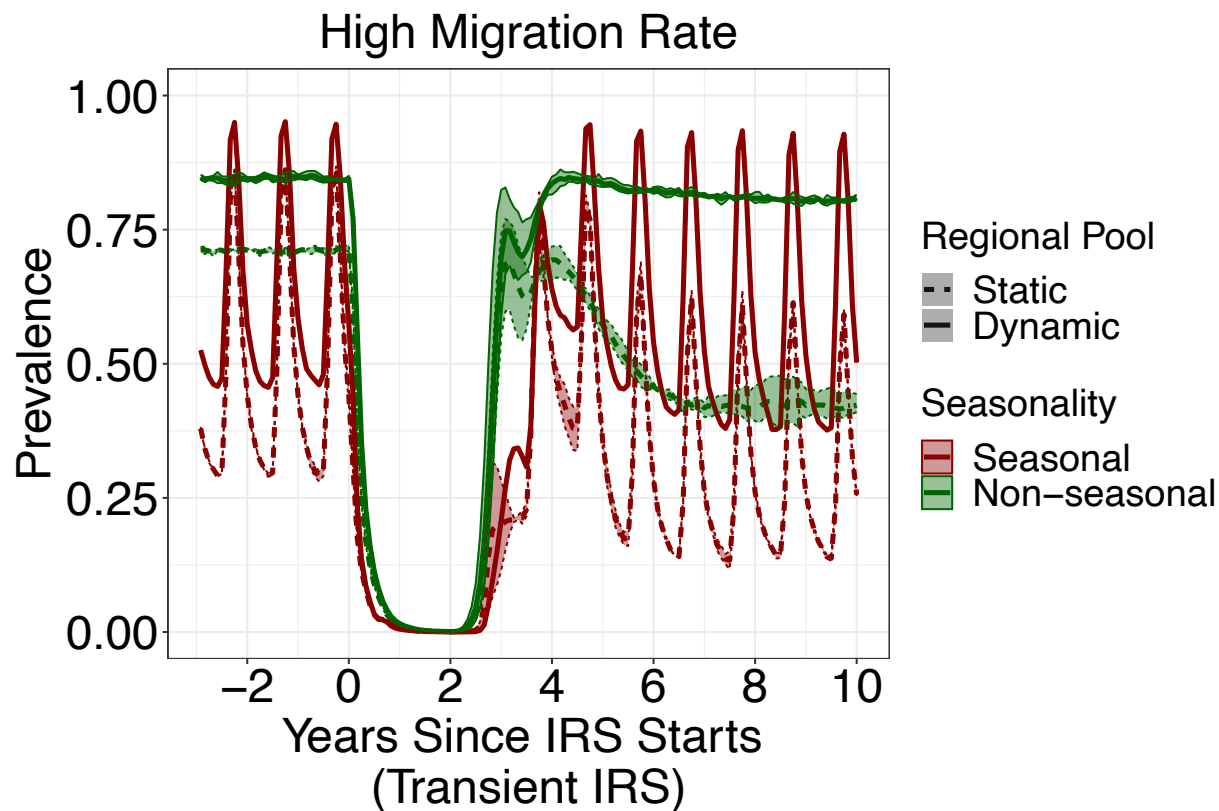
a The PTS distribution between the two age groups becomes more similar, especially for its lowest mode, which is almost completely overlapping. The x-axis range for inset figures is 0-0.075. **b** The MOI distribution exhibits its main mode at one, with the majority of infections becoming either mono-clonal 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.



Supplementary Fig. 24. The quantitative patterns of the two indicators for the field data from northern Ghana with curative drug treated individuals removed. **a** The difference in PTS values across the two age groups at individual, consecutive quantiles within the range of 0-0.40 for Ghana surveys from pre-, to during, to post-IRS. **b** The proportion of monoclonal infections or multi-genomic infections with two distinct strains (MOI = 1 or 2) is shown for both simulated systems at the high-transmission pre-IRS phase, the ones which are pushed by IRS into the fast-rebound, approaching or entering the transition, or slow-rebound regime, and empirical transmission system in Ghana across the three-round transient IRS.



Supplementary Fig. 25. Rebound dynamics of prevalence across time, under a series of transient IRS interventions (2 years). The dynamics for closed and semi-open systems can be categorized into three regimes described respectively as fast rebound, transition, and slow rebound. However, regionally-open systems always rebound rapidly. Simulation runs with other parameter values are shown in Fig. 6a. Different colors indicate varying levels of intervention intensity (corresponding to those in Fig. 1b-c.). Ranges are derived based on three replicate runs.



Supplementary Fig. 26. The rebound dynamics of regionally-open systems with a medium pool to transient and highest-coverage interventions, assuming a high migration rate (an order higher than the baseline migration rate, Methods). These systems receive migration from either a dynamic pool or a static one (Methods). Rapid rebound occurs when the pool is dynamic, and the importation of novelty is effective. The scenario which assumes a baseline migration rate is shown in Fig. 6c. Ranges are derived based on three replicate runs.

Scenario	Openness	Seasonality	Migration Rate	Regional Pool Size
I	Closed	Seasonal	Not Applicable (NA)	NA
II	Closed	Non-seasonal	NA	NA
III	Semi-open	Seasonal	Baseline	NA
IV	Semi-open	Seasonal	High	NA
V	Semi-open	Non-seasonal	Baseline	NA
VI	Semi-open	Non-seasonal	High	NA
VII	Regionally-open	Seasonal	Baseline	Medium
VIII	Regionally-open	Seasonal	Baseline	Large
IX	Regionally-open	Seasonal	High	Medium
X	Regionally-open	Seasonal	High	Large
XI	Regionally-open	Non-seasonal	Baseline	Medium
XII	Regionally-open	Non-seasonal	Baseline	Large
XIII	Regionally-open	Non-seasonal	High	Medium
XIV	Regionally-open	Non-seasonal	High	Large

Supplementary Table 1: Numerical simulation scenarios considered.

References

1. 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**(5) (2022).
2. Ghansah, A. et al. Comparison of molecular surveillance methods to assess changes in the population genetics of *Plasmodium falciparum* in high transmission. *Front. Parasitol.*, **2** (2023).
3. 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**(9) (2010).
4. 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**(27) (2003).
5. 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**(5), 1527-1538 (2003).
6. Tiedje, K. E. et al. Measuring changes in *Plasmodium falciparum* census population size in response to sequential malaria control interventions. *eLife*, **12** (2023).
7. Ruybal-Pesántez, S. et al. Population genomics of virulence genes of *Plasmodium falciparum* in clinical isolates from Uganda. *Sci. Rep.*, **7**(11810) (2017).
8. Ruybal-Pesántez, S. et al. Age-specific patterns of DBL α *var* diversity can explain why residents of high malaria transmission areas remain susceptible to *Plasmodium falciparum* blood stage infection throughout life. *International Journal for Parasitology*, **52**(11), 721-731 (2022).

- 674 9. Buckee, C. O. & Recker, M. Evolution of the multi-domain structures of virulence genes in
675 the human malaria parasite, *Plasmodium falciparum*. *PLoS computational biology*, **8**(4)
676 (2012).
- 677 10. Claessens, A. et al. A subset of group A-like *var* genes encodes the malaria parasite ligands
678 for binding to human brain endothelial cells, *Proc. Natl. Acad. Sci. U.S.A.*, **109**(26), 1772–
679 1781 (2012).
- 680 11. Kaestli, M. et al. Virulence of malaria is associated with differential expression of
681 *Plasmodium falciparum* *var* gene subgroups in a case-control study. *The Journal of*
682 *infectious diseases*, **193**(11), 1567–1574 (2006).
- 683 12. Rottmann, M. et al. Differential Expression of *var* Gene Groups Is Associated with
684 Morbidity Caused by *Plasmodium falciparum* Infection in Tanzanian Children. *Infection and*
685 *immunity*, **74**(7), 3904–3911 (2006).
- 686 13. Sondo, P. et al. Genetically diverse *Plasmodium falciparum* infections, within-host
687 competition and symptomatic malaria in humans. *Sci. Rep.*, **9**, 127 (2019).
- 688 14. Nkhoma, S. C. et al. Dynamics of parasite growth in genetically diverse *Plasmodium*
689 *falciparum* isolates. *Molecular and Biochemical Parasitology*, **254**, (2023).
- 690 15. Camponovo, F. et al. Mechanistic within-host models of the asexual *Plasmodium falciparum*
691 infection: a review and analytical assessment. *Malar. J.*, **20**(309) (2021).
- 692 16. Churcher, T. S. et al. Probability of Transmission of Malaria from Mosquito to Human Is
693 Regulated by Mosquito Parasite Density in Naïve and Vaccinated Hosts. *PLoS. Pathog.*,
694 **13**(1) (2017).
- 695 17. Smith, D. L., Drakeley, C. J., Chiyaka, C. & Hay, S. I. A quantitative analysis of
696 transmission efficiency versus intensity for malaria. *Nat. Commun.*, **1**(108) (2010).
- 697 18. Drummond, A., Silberg, J. J., Meyer, M. M., Wilke, C. O. & Arnold, F. H. On the
698 conservative nature of intragenic recombination. *Proc. Natl. Acad. Sci. U.S.A.*, **102**(15),
699 5380–5385 (2005).
- 700 19. Hermesen, C. C. et al. Detection of *Plasmodium falciparum* malaria parasites in vivo by real-
701 time quantitative PCR. *Molecular and Biochemical Parasitology*, **118**(2), 247–251 (2001).
- 702 20. Venugopal, K. et al. Plasmodium asexual growth and sexual development in the
703 haematopoietic niche of the host. *Nat Rev Microbiol*, **18**, 177–189 (2020).
- 704 21. Escalante, A. A., Cepeda, A. S. & Pacheco, M. A. Why *Plasmodium vivax* and *Plasmodium*
705 *falciparum* are so different? A tale of two clades and their species diversities. *Malar. J.*,
706 **21**(139) (2022).
- 707 22. Collins, W. E., Skinner, J. C. & Jeffery, G. M. Studies on the persistence of malarial antibody
708 response. *American journal of epidemiology*, **87**(3), 592–598 (1968).
- 709 23. Collins, W. E., Jeffery, G. M. & Skinner, J. C. Fluorescent Antibody Studies in Human
710 Malaria. II. Development and Persistence of Antibodies to *Plasmodium falciparum*. *The*
711 *American journal of tropical medicine and hygiene*, **13**, 256–260 (1964).
- 712 24. Larremore, D. B., Clauset, A. & Buckee, C. O. A network approach to analyzing highly
713 recombinant malaria parasite genes. *PLoS computational biology*, **9**(10) (2013).

- 714 25. Claessens, A. et al. Generation of Antigenic Diversity in *Plasmodium falciparum* by
715 Structured Rearrangement of *Var* Genes During Mitosis. *PLoS. Genet.*, **10**(12) (2014).
- 716 26. Childs, L. M., Baskerville, E. B. & Cobey, S. Trade-offs in antibody repertoires to complex
717 antigens. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **370** (1676)
718 (2015).
- 719 27. Gardner, M. et al. Genome sequence of the human malaria parasite *Plasmodium falciparum*.
720 *Nature* **419**, 498–511 (2002).