PROCEEDINGS B

royalsocietypublishing.org/journal/rspb

Research

Check for updates

Article submitted to journal

Subject Areas:
Ecology, Evolution, Health and Disease and Epidemiology

Keywords:
Eco-evolutionary dynamics, multigene families, malaria

Author for correspondence:
Qixin He
e-mail: heqixin@gmail.com

Evolutionary trade-offs between functional and immune selection shape multigene families in pathogens

Qi Zhan[1], Mercedes Pascual[2,3] and Qixin He[4]

[1]Committee on Genetics, Genomics and Systems Biology, The University of Chicago, Chicago, IL, 60637, USA.
[2]Department of Biology and Department of Environmental Studies, New York University, New York, NY, 10003, USA.
[3]Santa Fe Institute, Santa Fe, NM, 87501, USA.
[4]Department of Biological Sciences, Purdue University, West Lafayette, IN, 47907, USA.

Many infectious pathogens employ rapidly diversifying multigene families to encode surface proteins that influence fitness in both a frequency-dependent and an absolute manner, through respective variation in antigenic determinants and the basic reproductive number (R_0). How evolution shapes gene composition as pathogens compete for hosts remains largely unexplored, even though these two trait axes map to those of Modern Coexistence Theory. We address this question with a stochastic computational model for the transmission of the *Plasmodium falciparum* parasite and the birth-death gene dynamics of the *var* multigene family encoding for its major surface antigen. Selection alone cannot maintain the stable ratio observed for two gene groups within parasite genomes, suggesting that group-based classifications do not clearly reflect transmission strategy and virulence. When a trade-off exists between the two axes, strong immune selection attenuates functional selection for traits associated with absolute fitness and favors fast-recombining genes for antigenic diversification. Thus, the proportion of high-R_0 genes in individual genomes (and in the population) inverts along the transmission intensity gradient, with differences between genes becoming increasingly frequency-dependent. Overall, strong immune selection increases invasion probability of novel antigens and niche differences between parasite genomes, while reducing gene variance in transmissibility, virulence and infection duration.

© The Authors. Published by the Royal Society under the terms of the Creative Commons Attribution License http://creativecommons.org/licenses/by/4.0/, which permits unrestricted use, provided the original author and source are credited.

THE ROYAL SOCIETY PUBLISHING

1. Background

The evolutionary arms race between infectious pathogens and their hosts reflects the constant struggle of pathogens to proliferate and the need for infected hosts to curb their growth [1,2]. Multigene families play key roles in the adaptation of both sides. The major histocompatibility complex (MHC) genes in vertebrates encode highly polymorphic cell surface proteins that are essential for adaptive immune responses [3]. Similarly, many pathogens employ multigene families to encode diverse surface antigens, which can undergo rapid innovation, generating extensive antigenic variation at both within-host and population levels to facilitate immune evasion [4]. Although evolutionary models for MHC have been extensively explored [5–7], the birth-death processes of individual member genes within antigen-encoding multigene families of parasite genomes, have been less studied. Understanding the evolution of these member genes and the associated traits of their strains, as well as the drivers of this evolution, is crucial for predicting changes in a pathogen's ability to cause severe disease and for identifying strategies to reduce disease burden in human populations [8].

Within theoretical ecology, modern coexistence theory (MCT) for competitive systems helps explain how species interactions shape ecological communities. It does so by organizing trait variation along two major axes. The first axis concerns traits that influence competitive ability and consequently fitness in an "absolute" manner, independent of their frequencies, which typically oppose coexistence under conditions of high competition. Traits of this kind are related to demography and intrinsic growth rates. The second axis concerns traits whose competitive effects are purely frequency-dependent, which therefore have a "stabilizing" impact, promoting coexistence through niche differentiation and resource partitioning [9]. Coexistence is favored by "equalizing" differences in absolute fitness and by enhancing niche differences. This framework has been recently extended to host-pathogen systems [10]. As for previous studies, the focus has been on pairwise interactions and therefore on low-diversity systems. Evolutionary dynamics are not explicitly considered or are assumed to be slow relative to ecological dynamics, with pre-existing diversity and no emergence of novelty. Extending MCT to high-dimensional systems and phenotypes composed of multiple explicit traits, with ecological and evolutionary processes occurring on similar timescales, remains a major challenge [11]. Nevertheless, this conceptual framework allows us to categorize here trait variation among individual members of a given gene family along two axes: one for absolute fitness differences, due to virulence, transmissibility, and duration of expression, which all influence the basic reproductive number R_0, and another for niche differences, associated with the rate of novel gene generation which facilitates immune evasion. We can then ask how variation along these axes influences the persistence and turnover of individual member genes, and therefore, the coexistence and composition of gene properties, within parasite genomes and populations.

The human malaria parasite *Plasmodium falciparum* is one major human pathogen whose extensive within-host and population-level antigenic diversity underlies high asymptomatic prevalence and resilience against intervention efforts at high transmission [12–15]. The *var* multigene family encodes *Plasmodium falciparum* erythrocyte membrane protein 1 (*Pf*EMP1), the major variant surface antigen of the blood-stage infection. This pathogen provides a good model system for addressing how both functional selection for high absolute fitness and immune selection for antigenic novelty shape the coexistence of individual member genes within a multigene family. Such a system involves a high-dimensional trait space and experiences evolutionary change at timescales comparable to those of ecological processes, in both the frequency of genes and genotypes within population and the emergence of new genes through recombination and immigration.

Each parasite carries around 50-60 *var* genes, each encoding a distinct antigenic form of *Pf*EMP1 [16]. Typically, only one is expressed at a time, enabling sequential switching of functionally similar yet antigenically distinct surface antigens. *Var* genes and their corresponding *Pf*EMP1 proteins exhibit trait variation along the two above-mentioned axes [17–20]. Specifically,

functionally (or regarding the axis of absolute fitness), *Pf*EMP1 binds to the host's endothelial receptors, preventing parasites from splenic clearance. Thus, higher binding affinity confers higher absolute fitness of parasites [17], influencing pathogen virulence [21], transmissibility, and expression duration, or their combined effect as measured by R_0. Regarding the frequency-dependent axis, *var* genes are known to exhibit different rates of mitotic recombination, which generates novel offspring genes encoding new surface proteins that enhance immune evasion [22]. The major groups of *var* genes, upsA and upsB/C, are classified by their 5′-flanking regions (ups), which regulate gene expression [16]. These two groups are thought to occupy distinct regions of trait space along these two axes, reflecting diversification adapted to varying conditions of transmission and host immune response encountered by the parasite [23]. UpsA genes are thought to encode highly virulent *Pf*EMP1s, which underlie severe disease outcomes. In contrast, upsB/C genes are considered to have evolved to facilitate fast mitotic recombination for immune evasion, potentially compromising their ability to effectively bind to host endothelial receptors. Another important observation is that repertoires of genes (individual parasite genomes) and isolates (sets of individual parasite genomes co-infecting individual hosts) in the field consistently show an approximately constant proportion of the two ups groups across different geographical locations and transmission intensity, about 20% upsA and 80% upsB/C [12,14,16,24,25]. The mechanism behind this stable proportion remains unclear.

Using a stochastic computational agent-based model of the transmission system, we examine here the eco-evolutionary dynamics of the *var* multigene family to address how transmission intensity and associated immune selection shape the composition of multigene families within individual parasite genomes. We ask under what conditions pathogens will evolve to optimize virulence, transmissibility, and duration of expression, leading to the dominance of genes associated with optimal values for these traits within parasite genomes. We further investigate the effects of trade-offs between traits on the birth-death processes of individual member genes within these families. Can trade-offs between the two axes reflecting different transmission strategies (i.e., an enhanced basic reproductive number vs. immune evasion, respectively) promote the stable coexistence of genes adapted to each strategy within parasite genomes? Understanding the evolutionary outcomes of these trade-offs can help elucidate the relationship between the ups groups and transmission strategies. This relationship has important implications for the feasibility of identifying pathologically and clinically relevant subsets of *var* genes solely based on sequence information, an ongoing challenge due to the hyperdiversity of *var* genes.

To address these questions, we consider four different scenarios with an extension of the stochastic agent-based model (ABM) in Zhan et al. [15]. The first scenario allows traits to vary freely along both axes without any trade-off constraint. The second and the third ones introduce trade-offs within, but not between, the two axes. These consist of respective associations between transmissibility and duration of expression within the axis for absolute fitness differences, and between innovation rate and load of novelty within the axis for niche differences. The final scenario includes a trade-off between the two axes. We find that no scenario can generate the emergence of stable proportions of gene groups. Furthermore, high transmission rates promote genes with high mitotic recombination rates and reduced functional differences, resulting in high diversity of both genes and antigenic genomes, with the latter exhibiting low overlap in their gene composition. This "equalizing" of differences in reproductive number with reduced variance across genes at high niche differences, opposes the expectation of MCT for purely ecological dynamics. We discuss the implications of both results.

2. Methods

(a) Trait variation, trade-offs, and ups groups of *var* genes

(i) Two different selection pressures from hosts and the resulting transmission strategies of *var* genes

*Pf*EMP1s, encoded by *var* genes, are transported to and expressed on the surface of infected erythrocytes (IEs). These *Pf*EMP1s are responsible for IE adhesion, including cytoadherence to endothelial cells (often referred to simply as cytoadherence or cytoadhesion) and rosetting with uninfected erythrocytes, which prevent splenic clearance and enhance parasite reproduction rates [17]. Furthermore, *Pf*EMP1s are the major blood-stage antigens that mediate IEs' binding to various immune system cells, leading to significant immunological consequences [17]. Thus, *var* genes and more generally many other infectious pathogen antigen-encoding genes [26] experience two different selection pressures from hosts during infections. On the one hand, gene variants with enhanced cytoadhesion allow parasites carrying a higher proportion of these variants to grow faster and more efficiently in naive hosts, increasing transmissibility, prolonging duration of expression, and ultimately increasing the basic reproductive number (R_0). The result is a positive, directional selection favoring genes that maximize absolute fitness (R_0), regardless of their frequency, opposing coexistence. We refer to this as functional selection acting along the axis of absolute fitness differences. On the other hand, gene variants with novel epitopes, unrecognized by preexisting antibody responses, allow parasites carrying a higher proportion of these variants to be selected in immune hosts. The result is a negative frequency-dependent selection where common genes face a fitness disadvantage due to widespread immune memory, whereas rare genes are favored due to the absence of prior immune recognition. Since this frequency-dependent selection promotes coexistence through niche differentiation and resource partitioning, we refer to it as immune selection acting along the axis of niche differences. Consequently, *var* genes can adopt either one or both of two different transmission strategies: (1) intrinsic growth advantage within naive human hosts for higher virulence, transmissibility, longer duration of expression, or collectively a higher R_0; (2) fast innovation to produce novel genes effectively evading the immune system of hosts and expanding the pool of susceptible hosts, i.e., niche differentiation and resource partitioning.

(ii) Potential trade-offs between traits of *var* genes

Trade-offs [27] can potentially operate along each of the two axes, as well as between the two axes, as parasites adopt either or both transmission strategies. Due to the lack of understanding regarding genotype-phenotype mapping, we examine four scenarios (Table 1): (1) genes are allowed to exhibit trait variation along both axes and evolve their trait values without any trade-off constraints, (2) genes are allowed to exhibit trait variation only along the axis of absolute fitness differences and evolve their trait values under trade-off constraints between traits related to this axis, (3) genes are allowed to exhibit trait variation only along the axis of niche differences and evolve their trait values under trade-off constraints between traits related to this axis, and (4) genes are allowed to exhibit trait variation along both axes and evolve their trait values under trade-off constraints between traits related to the two axes. This final scenario takes into account the tug-of-war between functional selection and immune selection.

The absolute fitness of genes and their pathogens is determined by the combined effects of multiple traits, such as transmissibility and duration of expression. Meanwhile, the niche differentiation and resource partitioning capacity of genes and their pathogens is quantified by gene innovation rates primarily through mitotic recombination. In scenario (1), newly created genes can choose between high and low transmissibility, long and short duration of expression, and hence high or low R_0, as well as fast or slow mitotic recombination rate. Genes with higher transmissibility may require higher parasitemia, thereby provoking a stronger immune response from the infected host and consequently reducing their duration of expression. This trade-off between transmissibility and duration of expression can vary in strength [28–31], either partially (weak trade-off) or fully (strong trade-off) equalizing the absolute fitness (R_0) of different genes. In scenario (2), we consider both cases. In the case of a strong trade-off, an effectively

neutral scenario arises where no deterministic competitive winner exists, allowing genes with different transmissibility and duration of expression to coexist for extended periods before some genes randomly drift to extinction. Mitotic recombination events can produce functional and novel offspring genes that no longer compete with existing genes for the same immune space. However, these events also come with a cost (termed recombination load in genetics): there is a probability of creating non-functional offspring genes that replace their parental genes and are cleared immediately, thereby reducing the total duration of infection and fitness of their strains. In scenario (3), we consider whether the existence of such a load would undermine the selection for a fast mitotic recombination rate. As gene variants, particularly their immunodominant regions, diversify, they may deteriorate in their binding affinities, resulting in a trade-off between niche differentiation/resource partitioning capacity and absolute fitness. In scenario (4), we consider the selection dynamics between genes with higher R_0 but slow mitotic recombination versus genes with lower R_0 but fast mitotic recombination. Further details can be found in the section below titled "Setup of simulations for the four scenarios".

(iii) Ups groups of *var* genes

Var genes diversify rapidly, primarily through mitotic recombination, but also via meiotic recombination and mutation [22]. The number of circulating *var* genes in a local host population can reach thousands to tens of thousands [12–14]. Despite this immense diversity, major groups of *var* genes exist based on their 5′-flanking region, called ups, which controls gene expression: upsA and upsB/C [16]. Field repertoires (individual parasite genomes) and isolates (sets of genomes co-infecting the same individual) consistently contain a stable proportion of the two gene groups across different geographical locations and transmission intensity: approximately 20% upsA variants and 80% upsB/C variants [12,14,16,24,25].

Ups groups determine many gene features [12,14,16,25]. UpsA genes are generally much more conserved among field repertoires and isolates compared to upsB/C genes [12]. These repertoires and isolates contain a mix of genes from different frequency classes, based on sample-specific frequencies, ranging from rare types with only one or a few copies circulating in the population to common types with many more copies. Consistent with their higher level of conservation, upsA genes include more common types, most of which occur at stable frequencies across geographical locations, whereas upsB/C genes include fewer common types, most of which tend to be more localized [32]. Additionally, upsA genes form a distinct recombination group, sharing few homologous segments with upsB/C genes [33,34]. Recombination between the two ups groups is restricted, leading to a recombination hierarchy of *var* genes. Moreover, the two ups groups exhibit different mitotic recombination rates: upsB/C genes recombine more frequently with one another compared to upsA genes [22]. These distinct features of the two ups groups have led researchers to hypothesize that upsA and upsB/C genes serve different roles [23]. UpsA genes encode functionally constrained and highly virulent *Pf*EMP1s, which are associated with severe disease outcomes. In contrast, upsB/C genes are thought to adapt for fast recombination to facilitate immune evasion, albeit with less effective receptor binding. However, the empirical evidence to date remains mixed, with findings from different studies being inconsistent with one another [19,35]. Investigating the evolution of gene composition within multigene families in individual parasite genomes across the transmission gradient will shed light on the correspondence between the two transmission strategies and the two ups groups. If a stable proportion of genes adapted for the two transmission strategies emerges dynamically across the transmission gradient, it will provide additional support for the correspondence between transmission strategies and ups categories. If not, it will suggest that the ups groups do not map well to the two transmission strategies, and hence one cannot reliably identify pathologically important subsets based solely on sequence information, i.e., the ups categories.

(b) The stochastic agent-based model of malaria transmission

(i) Overview

We extend a previous computational model [15], of which key assumptions and processes are outlined in the electronic supplementary material and illustrated in supplementary figure S1. 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 an exponential distribution with a certain rate. We choose values for the rates of these events based on the literature on malaria epidemiology, related field studies, and in vitro or in vivo values. 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 [36], which optimizes the Gillespie first-reaction method [37].

(ii) Setup of simulations for the four scenarios

We consider two discrete values for each of the aforementioned traits: a high/fast value versus a low/slow value (Table 1, SimParams.xlsx) [22,38–40]. In scenario (1), newly generated genes can be assigned with a high or low transmissibility, a long or short duration of expression, and collectively a high or low absolute fitness (R_0), as well as a fast or slow mitotic recombination rate. In scenario (2), we assume two groups of genes with either a weak or strong trade-off between transmissibility and duration of expression. For the weak trade-off case, more transmissible genes (group 1) are associated with a slightly shorter duration of expression, yet they have a greater R_0 than less transmissible genes (group 2). In other words, the advantage in transmissibility is only partially compromised by the disadvantage in duration of expression. For the strong trade-off case, more transmissible genes (group 1) are associated with a much shorter duration of expression than less transmissible ones (group 2); and the two groups have the same R_0. In other words, the advantage in transmissibility is fully compromised by the disadvantage in duration of expression. We turn off gene innovation via mitotic recombination and mutation events. In scenario (3), we assume a fast-recombining group (group 1) and a slow-recombining group (group 2) of genes. Fast-recombining genes tend to be more diverse and less similar in their sequences. As a result, when they recombine, they are more likely to produce non-functional offspring genes. They are thus associated with a higher recombination load. Further details on the specification and calculation of recombination load can be found in the section entitled "Mitotic recombination" in the electronic supplementary material. We assume that the two groups of genes share the same R_0, specifically the same transmissibility and duration of expression. In scenario (4), we assume two groups of genes. Group 1 genes have a higher transmissibility and a longer duration of expression than group 2 ones. Therefore, group 1 genes' R_0 is higher than that of group 2 genes. But group 1 genes have a lower recombination rate than group 2 ones.

In scenario (1), mitotic recombination is allowed to occur between any pairwise genes. But in the last three scenarios, mitotic recombination occurs between pairs of genes within the same group and different groups of genes do not share alleles or epitopes with one another. Offspring genes inherit their properties from parental genes. We assume the expression order of the 60 genes within individual parasite genomes to be random [41]. In other words, each gene, regardless of its properties, has an equal probability of expressing early.

(iii) Initial conditions and the maximum transmission intensity considered

We consider the following three different initial conditions: strains used to initialize transmission contain either primarily group 1 genes (80% vs. 20%, or a ratio of 8:2), an equal proportion of the two groups of genes (50% vs. 50%, or a ratio of 5:5), or primarily group 2 genes (20% vs. 80%, or a ratio of 2:8).

We increase the transmission intensity log-linearly along the axis of transmission intensity measured by effective contact rate (supplementary material). We cap the highest effective contact rate at values that produce epidemiological and diversity measurements comparable to those

Table 1. The four scenarios considered. G1 refers to group 1 genes and G2 refers to group 2 genes. When assuming a recombination load associated with mitotic recombination events, non-functional offspring genes will replace their parental ones. Since these genes do not express and are deactivated immediately, they shorten the infection duration of the strain they constitute, thereby reducing its fitness. In contrast, when assuming no recombination load, non-functional offspring genes do not replace the parental ones. As a result, they do not reduce the total duration of infection of the strain.

	Scenario (1)	Scenario (2)	Scenario (3)	Scenario (4)
Transmissibility (Trans)	G1 = 2×G2	G1 = 2×G2	G1 = G2	G1 = 2× G2
Duration (Dur)	G1 = $\frac{4}{3}$×G2	G1= $\frac{12}{13}$ (or $\frac{1}{2}$) × G2	G1 = G2	G1 = $\frac{4}{3}$× G2
R_0 (Trans × Dur)	G1 > G2	G1 > (or =) G2	G1 = G2	G1 > G2
Recombination Rate	G1 & G2 ∈ (Low, High)	G1 = G2 = 0	G1 < G2	G1 < G2
Recombination Load	G1 = G2 = 0	Not Applicable	G1 < G2	G1 < G2

observed in high-transmission endemic regions in the field [42]. Each simulation runs for 200 years to allow the transmission system to reach a semi-stationary state. At the end of this period, we sample 15% of the host population at the end of the high-transmission season.

(c) Pairwise type sharing (PTS) between repertoires and isolates

The pairwise type sharing (PTS) index describes the degree of shared gene types between any two repertoires or isolates [43]. It is analogous to the Jaccard and Sørensen indices in Ecology [44,45]. We use a directional version of this statistic [46] defined as the following:

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

$$PTS_{ji} = \frac{n_{ji}}{n_j}$$

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 indicates that the sharing of a given number of types is relative to the length of the isolate being considered. A PTS score ranges between 0 and 1, where a score of 0 signifies no shared gene types, and a score of 1 indicates complete overlap between the two isolates.

(d) The invasion probability of new genes

Below, we summarize the invasion probability of new genes as derived in [15]. Invasion or fixation of new genes [47,48] occurs when a single copy of the variant present at a given time point leaves descendants in the population after a sufficiently long period. This typically requires the variant to surpass a frequency threshold, becoming common enough to resist genetic drift. We estimate invasion probabilities using population genetic models and simulations from our ABM. At the stationary state, malaria transmission dynamics can be modeled as a birth-death process under a Moran model with selection. The invasion probability of a rare variant depends on its fitness advantage relative to that of competing ones [47].

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

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. By convention, the fitness of other genes is normalized to be 1, and that of the new gene is expressed as $1+s$, where s represents the selective advantage. We assume that allele pool is sufficiently large such that any mutation or mitotic recombination event would generate

a unique allele, and back-mutation to the original allele is vanishingly rare. Consequently, all new variants, whether arising through mutation or mitotic recombination initially exist as single copies ($n = 1$). Our analysis focuses on parasite populations with effective population sizes N satisfying $N \gg 1$. The above equation simplifies to:

$$p_{inv} \approx 1 - \frac{1}{1+s} = \frac{s}{1+s}$$

As described by He et al. [49] for malaria transmission, a gene's fitness is determined by its reproductive number (R), defined as the number of offspring genes it produces. This reproductive number depends on the transmission rate (b), the gene-specific transmissibility (T), and the typical infection duration τ of parasites carrying the gene.

$$s = \frac{R_{new}}{R} - 1 = \frac{b_{new} T_{new} \tau_{new}}{bT\tau} - 1 = \frac{T_{new}\tau_{new}}{T\tau} - 1$$

We consider the infection duration of a strain composed entirely of 'average' genes versus that of a strain with one 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 theoretical construct whose susceptible host proportion ($\bar{S}$) equals the mean susceptibility averaged across all circulating genes. Because a novel gene's transmission success depends on the total infection duration of its carrier strain, its invasion probability is influenced by its genomic context. We derive the expected invasion probability for a new gene across possible genome backgrounds, assuming the remaining genes are the average ones. Let d_1 and d_2 denote the duration of expression of a gene in naïve hosts. Let T_1 and T_2 represent the transmissibility of a gene. Let n_1 and n_2 denote the average number of genes in parasite genomes that are associated with expression duration d_1 and transmissibility T_1, versus those associated with d_2 and T_2, respectively. Since the proportion of susceptible hosts to any new gene is 1, for any novel gene with d_1 and T_1 as its duration of expression and transmissibility, we obtain:

$$s = \frac{(\frac{T_1 n_1}{g} + \frac{T_2 n_2}{g})(\frac{d_1}{2} + ((n_1 - 1 + 0.5)d_1 + n_2 d_2)\bar{S})}{(\frac{T_1 n_1}{g} + \frac{T_2 n_2}{g})(n_1 d_1 + n_2 d_2)\bar{S}} - 1$$

The values of $\bar{S}$ can be directly computed from our ABM simulation outputs.

3. Results

(a) No trade-off within or between the two axes (Scenario 1): functional selection optimizes absolute fitness of strains and immune selection increases divergence in *var* gene composition between them.

The intensity of immune-mediated competition and immune selection is measured by the proportion of hosts immune to circulating genes, which increases with transmission intensity (Figure 1a). Low transmission results in low competition between co-circulating genes and strains, which allows for the coexistence of those diverging in their absolute fitness (R_0), and niche differentiation or resource partitioning capacity measured by mitotic recombination rate. Genes with high transmissibility, prolonged duration of expression, collectively high R_0, and fast mitotic recombination rate do not exclude their inferior counterparts. As a result, they fail to dominate or become fixed within individual genomes or in the population. This is evident from the observation that the mean values of these traits across circulating genes remain below their respective optima (Figure 1b-c). Consistently, these traits exhibit relatively high variance among circulating genes (Figure 1b-c). Furthermore, strains and isolates often share a significant fraction of genes (Figure 1b). These patterns hold regardless of the initial ratio of genes with different fitness profiles used to initialize transmission (see colors in Figure 1).

In contrast, high transmission leads to high competition between co-circulating genes and strains, which eliminates inferior genes and purges strains carrying them at high frequency. Consequently, parasite populations evolve towards optimal transmissibility, duration of expression, R_0, and recombination rate (Figure 1b-c). Furthermore, strains and isolates exhibit limited gene sharing, as evidenced by a decrease in the mean PTS with increasing competition (Figure 1b). Accordingly, trait variances across genes approach zero (Figure 1b-c). Therefore, in the no-trade-off scenario, directional functional selection acts independently from immune selection. Both forces are more intense under high transmission settings.

(b) A trade-off within the absolute fitness axis (Scenario 2): genes with a high R_0 are favored under intermediate but not high immune-mediated competition.

The growth rate of parasites and their parasitemia levels are determined by *Pf*EMP1 binding affinities to host endothelial receptors and other infection-induced within-host environmental changes. While elevated parasitemia can enhance both transmissibility and virulence, it can also trigger stronger immune responses that accelerate parasite clearance and reduce duration of expression. We therefore model a trade-off between transmissibility and duration of expression, considering both weak and strong cases (Methods). To specifically examine how trade-offs between traits governing absolute fitness (R_0) influence gene competition and persistence within multigene families, we disable gene innovation in our model. Under this condition, pathogens exploit host susceptibility maintained by immunity loss and the continual replenishment of susceptible individuals through births.

In the weak trade-off case, more transmissible genes are associated with slightly reduced duration of expression and thus maintain higher R_0 values than less transmissible ones. At low transmission, the system largely preserves the composition of parasite genomes used to initialize transmission. As transmission increases from low to intermediate levels, functional selection favors more transmissible genes with high R_0, leading to their progressive accumulation within individual genomes and in the population (Figure 2a-b). This characterizes a transmission-limited regime when parasites have not yet exhausted the hosts' immune space (supplementary figure S4a). Although these high-R_0 genes and strains enriched with them get transmitted more, they continue to encounter adequate numbers of susceptible hosts.

Conversely, when transmission continues to increase from intermediate to high levels, the proportion of high-R_0 genes declines within individual genomes and in the population (Figure 2a-b). This characterizes a niche-limited regime marked by parasites' intense competition for hosts (supplementary figure S4a). This reversal occurs because high-R_0 genes rapidly deplete available immune space, subjecting them to stronger negative immune selection compared to less transmissible genes. High-R_0 genes consequently experience an inversion pattern in selection, transitioning from being favored to disfavored as transmission intensity increases.

Beyond analyzing equilibrium patterns across transmission intensities, we also track the temporal dynamics of high-R_0 gene proportion in the population through two simulations representing the two distinct regimes: one at intermediate transmission (transmission-limited regime) and one at high transmission (niche-limited regime), across all three initial conditions (Figure 2c). Under intermediate transmission, high-R_0 genes exhibit continuous expansion in the population throughout the simulation period without reaching saturation. In contrast, under high transmission, we observe an initial proliferation followed by rapid plateauing of high-R_0 gene proportion.

In the strong trade-off case, the increased transmissibility of genes is precisely offset by their shorter duration of expression, resulting in same R_0 values between more and less transmissible genes. Both gene groups consequently maintain proportions near their initial levels used to initialize transmission (supplementary figure S4b). This case effectively represents a neutral

system in which no deterministic competitive winner exists, allowing sustained coexistence until stochastic drift ultimately causes random extinction of one gene group.

(c) A trade-off within the niche difference axis (Scenario 3): recombination load is insufficient to fully counteract immune selection for fast-recombining genes.

Although fast-recombining genes innovate more novel antigens, they also produce more non-functional offspring genes as byproducts, resulting in significantly higher recombination load compared to their slow-recombining counterparts (Methods, Figure 3a). This recombination load exhibits a positive relationship with transmission intensity, eventually saturating at high transmission levels (Methods, Figure 3a).

At low transmission, the system preserves the composition of parasite genomes used to initialize transmission. As transmission increases from low to intermediate levels, novel functional offspring genes generated through mitotic recombination experience weak selection, as most hosts remain susceptible to existing diversity (supplementary figure S5). As transmission continues to increase from intermediate to high levels, selection for novel variants intensifies with the declining proportion of susceptible hosts to existing diversity (supplementary figure S5).

Fast-recombining genes face opposing selective forces. Positive selection favors functional novel offspring genes that evade host immunity, prolonging infection duration and enhancing transmission of the strains carrying them. Negative selection eliminates non-functional variants that shorten infection duration of the strains carrying them and reduce their onward transmission.

Despite this trade-off, the proportion of fast-recombining genes increases within individual genomes and in the population as transmission intensifies, ultimately approaching fixation (Figure 3b). Therefore, recombination load alone is insufficient to counteract the selective advantage of fast-recombining genes under conditions of strong immune-mediated competition for hosts.

The system is thus resilient to mitotic recombination load at high transmission and competition. To further illustrate this resilience, we analyze age distributions of functional and non-functional offspring genes across the full transmission gradient (Figure 3c). At low transmission, non-functional genes persist extensively in the population (Figure 3c). At intermediate transmission, non-functional genes show decreased longevity (loss of older-age mode), while functional genes exhibit increased persistence (loss of younger-age mode) (Figure 3c). At high transmission, non-functional genes become rapidly eliminated (approaching zero age), whereas functional genes predominantly reach intermediate-to-old ages (Figure 3c). Therefore, at intermediate to high transmission, the system efficiently purges non-functional offspring genes while retaining functional ones.

What explains the system's resilience against recombination load and its efficient purifying selection against non-functional offspring genes? As transmission intensity increases, mixed infections involving multiple strains, arising from either superinfection or co-transmission, become increasingly prevalent in diseases with high strain diversity, such as falciparum malaria. These mixed infections promote frequent meiotic recombination during the sexual stage in mosquito vectors. Elevated meiotic recombination has two consequences: enhanced purifying selection and amplified positive selection. By breaking the linkage between functional and non-functional genes, which normally constrains adaptation, meiotic recombination combines non-functional genes from different genomic backgrounds into the same genome, enhancing selection against these genes and their carrier strains. Additionally, meiotic recombination assembles functional genes into the same genome, increasing both the intensity and efficiency of selection for these genes and their carrier strains.

(d) Under a trade-off between the two axes (Scenario 4), genes with high R_0/slow recombination rates are selected at intermediate transmission intensity, whereas those with low R_0/fast recombination rates are selected at higher transmission intensity .

In scenario 4 (Table 1), we introduce a trade-off between traits governing absolute fitness and those mediating niche differentiation. This configuration creates a tug-of-war between functional selection and immune selection.

Similar to previous scenarios, at low transmission with small parasite populations, the system largely preserves the composition of parasite genomes used to initialize transmission (Figure 4a). As transmission reaches intermediate levels, a relatively high proportion of hosts still remain susceptible to circulating diversity (supplementary figure S6), indicating limited immune-mediated competition for hosts. Slow-recombining (high-R_0) genes are more advantageous due to their higher transmission in these susceptible hosts. Novel epitopes that can evade host immunity confer minimal selective advantage. Consequently, selection for fast-recombining genes remains weak. However, at high transmission with large parasite populations, host susceptibility to existing diversity drops significantly (supplementary figure S6), reflecting intense competition for hosts. Here, slow-recombining (high-R_0) genes face transmission bottlenecks due to the depletion of vacant immune spaces. Gene variants with novel, immune-evading epitopes gain a strong selective advantage. Strains enriched with these variants dominate transmission (Figure 4a). The observed inversion in slow-recombining, high-R_0 genes proportion reflects a strategic transition across the transmission gradient, from a high-R_0, low-mitotic-recombination strategy at intermediate transmission to a low-R_0, high-mitotic-recombination strategy at high transmission (Figure 4b).

In addition to analyzing equilibrium patterns across transmission intensities, we examine the temporal dynamics of the proportion of slow-recombining (high-R_0) genes in the parasite population, which revealed distinct trajectories across transmission regimes (Figure 4c). In the transmission-limited regime (intermediate transmission), the proportion of slow-recombining genes increases progressively throughout simulation, reflecting them replacing fast-recombining (low-R_0) ones. Conversely, in the niche-limited regime (high transmission), the proportion of slow-recombining genes initially increases but shortly thereafter starts to decrease rapidly. This reversal occurs as the depletion of susceptible hosts favors immune-evading novel variants and thus fast-recombining genes over their high-R_0 counterparts.

(e) The invasion probability of new genes increases, while the variance of traits along the two axes decreases, with increasing intensity of frequency-dependent selection.

Overall, the invasion probability of new genes increases with niche differences, quantified as $1-$ the mean PTS of sampled strains, which is consistent with previous findings under the MCT framework. Both the variance of traits related to absolute fitness and the variance in absolute fitness itself tend to decline with increasing niche differences (Figure 5, supplementary figure S2-3). With a trade-off between absolute fitness and mitotic recombination rate, high variance can still persist but only as an effect of initial conditions when starting with predominantly high- or low-absolute-fitness genes (Figure 4a, c). Each simulation runs for 200 years, long enough to approach a semi-stationary state and reveal selection trends. Although the system does not fully reach equilibrium and gene proportions continue to evolve, there is a trend toward the fixation of one gene group with the corresponding decline in variance approaching zero (Figure 4c). The observed tendency of decreasing trait variance with increasing niche differences stands in contrast to previous findings under the MCT framework ([10,50]), as discussed below.

4. Figures & Tables

5. Conclusion

The persistence and turnover of individual member genes within multigene families encoding important surface proteins are shaped by two distinct selective pressures. Directional functional selection favors genes with high absolute fitness, while immune-mediated negative frequency-dependent selection favors genes that enhance antigenic diversification and immune evasion. As parasites evolve to adopt one or both transmission strategies, an enhanced reproductive number vs. immune evasion, trade-offs can potentially operate. To examine long-term evolutionary trends of individual member genes within multigene families, we consider here the *var* multigene family of the human malaria parasite *Plasmodium falciparum*, which encodes the major surface antigen of the blood-stage of infection. The gene composition and gene properties within individual genomes are shaped by the intensity of immune-mediated competition. As expected, low competition allows coexistence of genes with diverse traits, whereas high competition leads to the dominance of genes with optimal traits. We find that when a trade-off exists between absolute fitness and immune evasion, strong competition can weaken directional functional selection for optimal absolute fitness, making differences among all circulating genes more purely frequency-dependent. In this context, traits that promote immune evasion are favored and the proportion of high absolute fitness genes exhibits an inversion pattern across a transmission gradient, shifting from being favored to disfavored as transmission intensity increases. We also find that, regardless of the existence of a trade-off, selection pressures alone cannot precisely maintain a stable proportion of genes adapted to the two transmission strategies within individual genomes across a transmission gradient.

The dynamic variation of the proportion of different groups in our model contrasts with the empirical observation of an approximately constant proportion of upsA and upsB/C genes in field repertoires and isolates across diverse geographical locations and transmission intensities. One possible explanation for this discrepancy is that the ups groups do not map directly onto the two transmission strategies shaped by functional and immune selection. If this is the case, these groups would be unreliable markers of pathogenicity. An alternative explanation is heterogeneity in trait values within each ups group. For instance, some upsA genes may be more prone to mitotic recombination than others, while certain upsB/C genes could exhibit greater virulence. Such within-group variation could help maintain stable proportions of each group in individual repertoires and isolates across transmission intensities, even as selection pressures and the relative abundance of genes adapted to each transmission strategy shift dynamically.

The inversion pattern in the proportion of high absolute fitness/high virulence genes along a transmission gradient may help explain the paradoxical relationship observed in the field between malaria transmission intensity and severe malaria admissions. Regions with high transmission often show a lower apparent risk of malaria compared to areas with intermediate transmission [51]. Similar sign inversions, influenced by population size, have been reported in other systems where trade-offs exist between innovation rates and absolute fitness [52].

Pathogens with multigene families encoding important surface proteins are resilient to the load associated with gene innovation. A high copy number reduces the dependence of a pathogen's fitness on individual member genes, thereby relaxing selective constraints. Meiotic recombination between co-infecting strains can disrupt the positive linkage between functional and non-functional genes, allowing functional novel genes to be integrated into the same genomic backgrounds, which enhances positive selection for these genes and their strains. At the same time, meiotic recombination helps purge non-functional genes by bringing them into the same genomic backgrounds where they are more effectively eliminated. Both processes reflect the Hill-Robertson [53] and the Fisher-Muller effects [54]. Together, gene copy number and meiotic recombination allow pathogens with antigen-encoding multigene families to harness the benefits of accelerated evolutionary innovation while minimizing the potential load incurred in this process.

Our study investigates coexistence and diversity in a host-pathogen system with a high-dimensional trait space while explicitly incorporating feedbacks between evolutionary processes and ecological dynamics and accounting for the generation of novelty. The observed reduction in the variance of traits as niche differences grow, contrasts with previous findings under MCT, where variance in absolute fitness can typically exhibit higher values at higher niche differences. A key distinction is that in our system evolution is explicit and interacts with population dynamics of the disease and the associated competition for hosts. Despite this difference, we find that the invasion probability of new genes increases with niche differences, consistent with previous MCT findings. The "equalizing" of traits associated with (gene) demography would make the system appear ecologically neutral, even though it is under strong selection from specific antigenic differences underlying competition for hosts.

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 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 M.P. We appreciate the support of the Research Computing Center at the University of Chicago through the computational resources of the Midway cluster.

References

1. Sironi M, Cagliani R, Forni D, Clerici M. 2015 Evolutionary insights into host–pathogen interactions from mammalian sequence data. *Nature Reviews Genetics* **16**, 224–236. (10.1038/nrg3905)
2. Woolhouse MEJ, Webster JP, Domingo E, Charlesworth B, Levin BR. 2002 Biological and biomedical implications of the co-evolution of pathogens and their hosts. *Nature Genetics* **32**, 569–577. (10.1038/ng1202-569)
3. Piertney SB, Oliver MK. 2006 The evolutionary ecology of the major histocompatibility complex. *Heredity* **96**, 7–21. (10.1038/sj.hdy.6800724)
4. Deitsch KW, Lukehart SA, Stringer JR. 2009 Common strategies for antigenic variation by bacterial, fungal and protozoan pathogens. *Nature Reviews Microbiology* **7**, 493–503. (10.1038/nrmicro2145)
5. Kaufman J. 2018 Unfinished Business: Evolution of the MHC and the Adaptive Immune System of Jawed Vertebrates. *Annual Review of Immunology*. (10.1146/annurev-immunol-051116-052450)
6. Nei M, Rooney AP. 2005 Concerted and Birth-and-Death Evolution of Multigene Families. *Annual Review of Genetics* **39**, 121–152. (10.1146/annurev.genet.39.073003.112240)
7. Maeda N, Smithies O. 1986 THE EVOLUTION OF MULTIGENE FAMILIES: Human Haptoglobin Genes. *Annual Review of Genetics*. (10.1146/annurev.ge.20.120186.000501)
8. Gupta S. 2024 Darwin review: the evolution of virulence in human pathogens. *Proceedings of the Royal Society B: Biological Sciences* **291**, 20232043. (10.1098/rspb.2023.2043)
9. Chesson P. 2000 Mechanisms of Maintenance of Species Diversity. *Annual Review of Ecology and Systematics* **31**, 343–366. (10.1146/annurev.ecolsys.31.1.343)
10. Park SW, Cobey S, Metcalf CJE, Levine JM, Grenfell BT. 2024 Predicting pathogen mutual invasibility and co-circulation. *Science* **386**, 175–179. (10.1126/science.adq0072)
11. Song C, Barabás G, Saavedra S. 2019 On the Consequences of the Interdependence of Stabilizing and Equalizing Mechanisms. *The American Naturalist* **194**, 627–639. (10.1086/705347)
12. Tiedje KE, Zhan Q, Ruybal-Pésantez S, Tonkin-Hill G, He Q, Tan MH, Argyropoulos DC, Deed SL, Ghansah A, Bangre O, Oduro AR, Koram KA, Pascual M, Day KP. 2024 Measuring changes in Plasmodium falciparum census population size in response to sequential malaria control interventions. (10.7554/eLife.91411.2)
13. Day KP, Artzy-Randrup Y, Tiedje KE, Rougeron V, Chen DS, Rask TS, Rorick MM, Migot-Nabias F, Deloron P, Luty AJF, Pascual M. 2017 Evidence of strain structure in *Plasmodium falciparum var* gene repertoires in children from Gabon, West Africa. *Proceedings of the National Academy of Sciences* **114**. (10.1073/pnas.1613018114)
14. Ruybal-Pesántez S, Tiedje KE, Tonkin-Hill G, Rask TS, Kamya MR, Greenhouse B, Dorsey G, Duffy MF, Day KP. 2017 Population genomics of virulence genes of Plasmodium falciparum

in clinical isolates from Uganda. *Scientific Reports* **7**, 11810. (10.1038/s41598-017-11814-9)
15. Zhan Q, He Q, Tiedje KE, Day KP, Pascual M. 2024 Hyper-diverse antigenic variation and resilience to transmission-reducing intervention in falciparum malaria. *Nature Communications* **15**, 7343. (10.1038/s41467-024-51468-6)
16. Rask TS, Hansen DA, Theander TG, Gorm Pedersen A, Lavstsen T. 2010 Plasmodium falciparum Erythrocyte Membrane Protein 1 Diversity in Seven Genomes – Divide and Conquer. *PLoS Computational Biology* **6**, e1000933. (10.1371/journal.pcbi.1000933)
17. Rowe JA, Claessens A, Corrigan RA, Arman M. 2009 Adhesion of *Plasmodium falciparum* -infected erythrocytes to human cells: molecular mechanisms and therapeutic implications. *Expert Reviews in Molecular Medicine* **11**, e16. (10.1017/S1462399409001082)
18. Kyriacou HM, Stone GN, Challis RJ, Raza A, Lyke KE, Thera MA, Koné AK, Doumbo OK, Plowe CV, Rowe JA. 2006 Differential var gene transcription in Plasmodium falciparum isolates from patients with cerebral malaria compared to hyperparasitaemia. *Molecular and Biochemical Parasitology* **150**, 211–218. (https://doi.org/10.1016/j.molbiopara.2006.08.005)
19. Kaestli M, Cockburn I, Cortés A, Baea K, Rowe J, Beck H. 2006 Virulence of Malaria Is Associated with Differential Expression of *Plasmodium falciparum var* Gene Subgroups in a Case-Control Study. *The Journal of Infectious Diseases* **193**, 1567–1574. (10.1086/503776)
20. Robinson BA, Welch TL, Smith JD. 2003 Widespread functional specialization of Plasmodium falciparum erythrocyte membrane protein 1 family members to bind CD36 analysed across a parasite genome. *Molecular Microbiology* **47**, 1265–1278. (https://doi.org/10.1046/j.1365-2958.2003.03378.x)
21. Abdi AI, Warimwe GM, Muthui MK, Kivisi CA, Kiragu EW, Fegan GW, Bull PC. 2016 Global selection of Plasmodium falciparum virulence antigen expression by host antibodies. *Scientific Reports* **6**, 19882. (10.1038/srep19882)
22. Claessens A, Hamilton WL, Kekre M, Otto TD, Faizullabhoy A, Rayner JC, Kwiatkowski D. 2014 Generation of Antigenic Diversity in Plasmodium falciparum by Structured Rearrangement of Var Genes During Mitosis. *PLoS Genetics* **10**, e1004812. (10.1371/journal.pgen.1004812)
23. Lavstsen T, Salanti A, Jensen AT, Arnot DE, Theander TG. 2003 Sub-grouping of Plasmodium falciparum 3D7 var genes based on sequence analysis of coding and non-coding regions. *Malaria Journal*.
24. Ruybal-Pesántez S, Tiedje KE, Pilosof S, Tonkin-Hill G, He Q, Rask TS, Amenga-Etego L, Oduro AR, Koram KA, Pascual M, Day KP. 2022 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. *International Journal for Parasitology* **52**, 721–731. (10.1016/j.ijpara.2021.12.001)
25. Buckee CO, Recker M. 2012 Evolution of the Multi-Domain Structures of Virulence Genes in the Human Malaria Parasite, Plasmodium falciparum. *PLoS Computational Biology* **8**, e1002451. (10.1371/journal.pcbi.1002451)
26. Wang Y, Tang CY, Wan XF. 2022 Antigenic characterization of influenza and SARS-CoV-2 viruses. *Analytical and Bioanalytical Chemistry* **414**, 2841–2881. (10.1007/s00216-021-03806-6)
27. Li Y, Petrov DA, Sherlock G. 2019 Single nucleotide mapping of trait space reveals Pareto fronts that constrain adaptation. *Nature Ecology & Evolution* **3**, 1539–1551. (10.1038/s41559-019-0993-0)
28. Ben-Shachar R, Koelle K. 2018 Transmission-clearance trade-offs indicate that dengue virulence evolution depends on epidemiological context. *Nature Communications* **9**, 2355. (10.1038/s41467-018-04595-w)
29. Blanquart F, Grabowski MK, Herbeck J, Nalugoda F, Serwadda D, Eller MA, Robb ML, Gray R, Kigozi G, Laeyendecker O, Lythgoe KA, Nakigozi G, Quinn TC, Reynolds SJ, Wawer MJ, Fraser C. 2016 A transmission-virulence evolutionary trade-off explains attenuation of HIV-1 in Uganda. *eLife* **5**, e20492. (10.7554/eLife.20492)
30. Fraser C, Hollingsworth TD, Chapman R, De Wolf F, Hanage WP. 2007 Variation in HIV-1 set-point viral load: Epidemiological analysis and an evolutionary hypothesis. *Proceedings of the National Academy of Sciences* **104**, 17441–17446. (10.1073/pnas.0708559104)
31. Mackinnon MJ, Read AF. 2004 Virulence in malaria: an evolutionary viewpoint. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **359**, 965–986. (10.1098/rstb.2003.1414)
32. Tan MH, Tiedje KE FQZQPMSHea. 2025 A paradoxical population structure of var DBL types in Africa. *PLoS Pathog* **21**. (10.1371/journal.ppat.1012813)

33. Larremore DB, Clauset A, Buckee CO. 2013 A Network Approach to Analyzing Highly Recombinant Malaria Parasite Genes. *PLoS Computational Biology* **9**, e1003268. (10.1371/journal.pcbi.1003268)
34. Bull PC, Buckee CO, Kyes S, Kortok MM, Thathy V, Guyah B, Stoute JA, Newbold CI, Marsh K. 2008 *Plasmodium falciparum* antigenic variation. Mapping mosaic *var* gene sequences onto a network of shared, highly polymorphic sequence blocks. *Molecular Microbiology* **68**, 1519–1534. (10.1111/j.1365-2958.2008.06248.x)
35. Claessens A, Adams Y, Ghumra A, Lindergard G, Buchan CC, Andisi C, Bull PC, Mok S, Gupta AP, Wang CW, Turner L, Arman M, Raza A, Bozdech Z, Rowe JA. 2012 A subset of group A-like *var* genes encodes the malaria parasite ligands for binding to human brain endothelial cells. *Proceedings of the National Academy of Sciences* **109**. (10.1073/pnas.1120461109)
36. Gibson MA, Bruck J. 2000 Efficient Exact Stochastic Simulation of Chemical Systems with Many Species and Many Channels. *The Journal of Physical Chemistry A* **104**, 1876–1889. (10.1021/jp993732q)
37. Gillespie DT. 1976 A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics* **22**, 403–434. (10.1016/0021-9991(76)90041-3)
38. Eichner M, Diebner HH, Molineaux L, Collins WE, Jeffery GM, Dietz K. 2001 Genesis, sequestration and survival of Plasmodium falciparum gametocytes: parameter estimates from fitting a model to malariatherapy data. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **95**, 497–501. (10.1016/S0035-9203(01)90016-1)
39. Molineaux L, Diebner HH, Eichner M, Collins WE, Jeffery GM, Dietz K. 2001 *Plasmodium falciparum* parasitaemia described by a new mathematical model. *Parasitology* **122**, 379–391. (10.1017/S0031182001007533)
40. Diebner HH, Eichner M, Molineaux L, Collins WE, Jeffery GM, Dietz K. 2000 Modelling the Transition of Asexual Blood Stages of Plasmodium falciparum to Gametocytes. *Journal of Theoretical Biology* **202**, 113–127. (10.1006/jtbi.1999.1041)
41. Milne K, Ivens A, Reid AJ, Lotkowska ME, O'Toole A, Sankaranarayanan G, Munoz Sandoval D, Nahrendorf W, Regnault C, Edwards NJ, Silk SE, Payne RO, Minassian AM, Venkatraman N, Sanders MJ, Hill AV, Barrett M, Berriman M, Draper SJ, Rowe JA, Spence PJ. 2021 Mapping immune variation and var gene switching in naive hosts infected with Plasmodium falciparum. *eLife* **10**, e62800. (10.7554/eLife.62800)
42. Smith DL, Drakeley CJ, Chiyaka C, Hay SI. 2010 A quantitative analysis of transmission efficiency versus intensity for malaria. *Nature Communications* **1**, 108. (10.1038/ncomms1107)
43. Barry AE, Leliwa-Sytek A, Tavul L, Imrie H, Migot-Nabias F, Brown SM, McVean GAV, Day KP. 2007 Population Genomics of the Immune Evasion (var) Genes of Plasmodium falciparum. *PLoS Pathogens* **3**, e34. (10.1371/journal.ppat.0030034)
44. Sørensen T. 1948 *A Method of Establishing Groups of Equal Amplitude in Plant Sociology Based on Similarity of Species Content and Its Application to Analyses of the Vegetation on Danish Commons*. Biologiske skrifter. Munksgaard in Komm.
45. Dice LR. 1945 Measures of the Amount of Ecologic Association Between Species. *Ecology* **26**, 297–302.
46. He Q, Pilosof S, Tiedje KE, Ruybal-Pesántez S, Artzy-Randrup Y, Baskerville EB, Day KP, Pascual M. 2018 Networks of genetic similarity reveal non-neutral processes shape strain structure in Plasmodium falciparum. *Nature Communications* **9**, 1817. (10.1038/s41467-018-04219-3)
47. Maruyama T. 1974 A MARKOV PROCESS OF GENE FREQUENCY CHANGE IN A GEOGRAPHICALLY STRUCTURED POPULATION. *Genetics* **76**, 367–377. (10.1093/genetics/76.2.367)
48. Otto SP, Whitlock MC. 1997 The Probability of Fixation in Populations of Changing Size. *Genetics* **146**, 723–733. (10.1093/genetics/146.2.723)
49. He Q, Pascual M. 2021 An antigenic diversification threshold for falciparum malaria transmission at high endemicity. *PLOS Computational Biology* **17**, e1008729. (10.1371/journal.pcbi.1008729)
50. MacDougall AS, Gilbert B, Levine JM. 2009 Plant invasions and the niche. *Journal of Ecology* **97**, 609–615. (10.1111/j.1365-2745.2009.01514.x)
51. Idro R, Aloyo J, Mayende L, Bitarakwate E, John CC, Kivumbi GW. 2006 Severe malaria in children in areas with low, moderate and high transmission intensity in Uganda. *Tropical Medicine & International Health* **11**, 115–124. (10.1111/j.1365-3156.2005.01518.x)

52. Raynes Y, Wylie CS, Sniegowski PD, Weinreich DM. 2018 Sign of selection on mutation rate modifiers depends on population size. *Proceedings of the National Academy of Sciences* **115**, 3422–3427. (10.1073/pnas.1715996115)
53. Hill WG, Robertson A. 1966 The effect of linkage on limits to artificial selection. *Genetical Research* **8**, 269–294. (10.1017/S0016672300010156)
54. Kim Y, Orr HA. 2005 Adaptation in Sexuals *vs.* Asexuals: Clonal Interference and the Fisher-Muller Model. *Genetics* **171**, 1377–1386. (10.1534/genetics.105.045252)

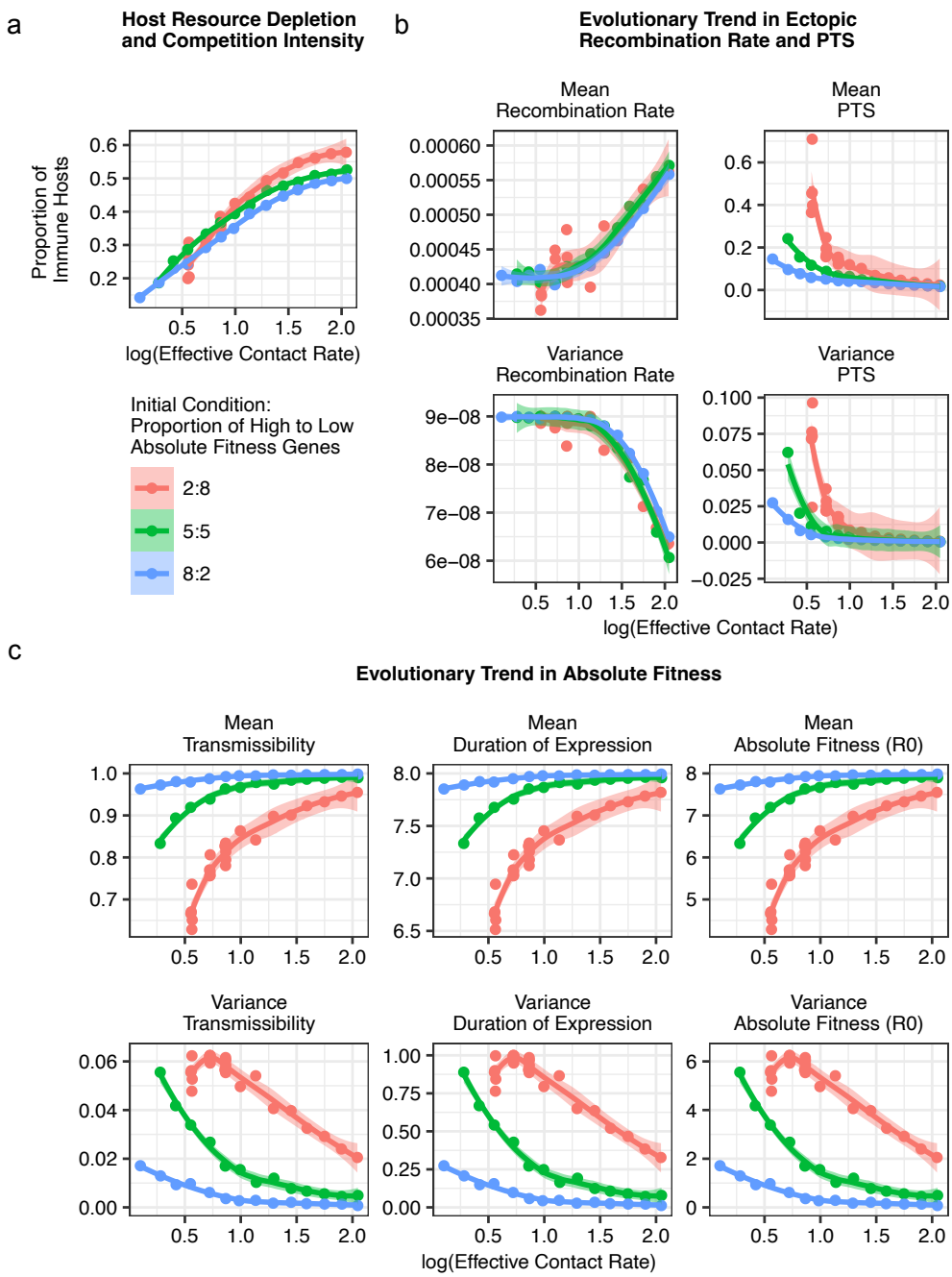

Figure 1. Scenario (1) with no trade-offs. **a** The intensity of immune-mediated competition and immune selection, quantified as the proportion of hosts immune to circulating genes, scales positively with transmission intensity (effective contact rate, see Methods). Under heightened competition, parasite populations evolve towards **b** faster recombination rate and limited gene sharing between strains measured by the pairwise type sharing index or PTS (Methods), as well as **c** enhanced transmissibility, prolonged duration of expression, and consequently higher R_0. Genes with lower transmissibility, shorter duration of expression, or collectively, smaller R_0, and slower recombination rate are progressively eliminated, whereas their superior counterparts are reaching fixation within individual genomes and in the population. Consistently, population-wide variances decrease for **b** mitotic recombination rate and gene sharing of strains, as well as **c** transmissibility, duration of expression, and R_0. All metrics (y axis) are calculated from end-of-high-season samples in the final simulation year (Year 200). Each point represents an individual simulation run at a specified effective contact rate, with trend lines generated using ggplot2's geom_smooth. stats::loess() is used for less than 1,000 observations; otherwise mgcv::gam() is used with formula = $y \sim s(x$, bs = "cs") with method = "REML") function.

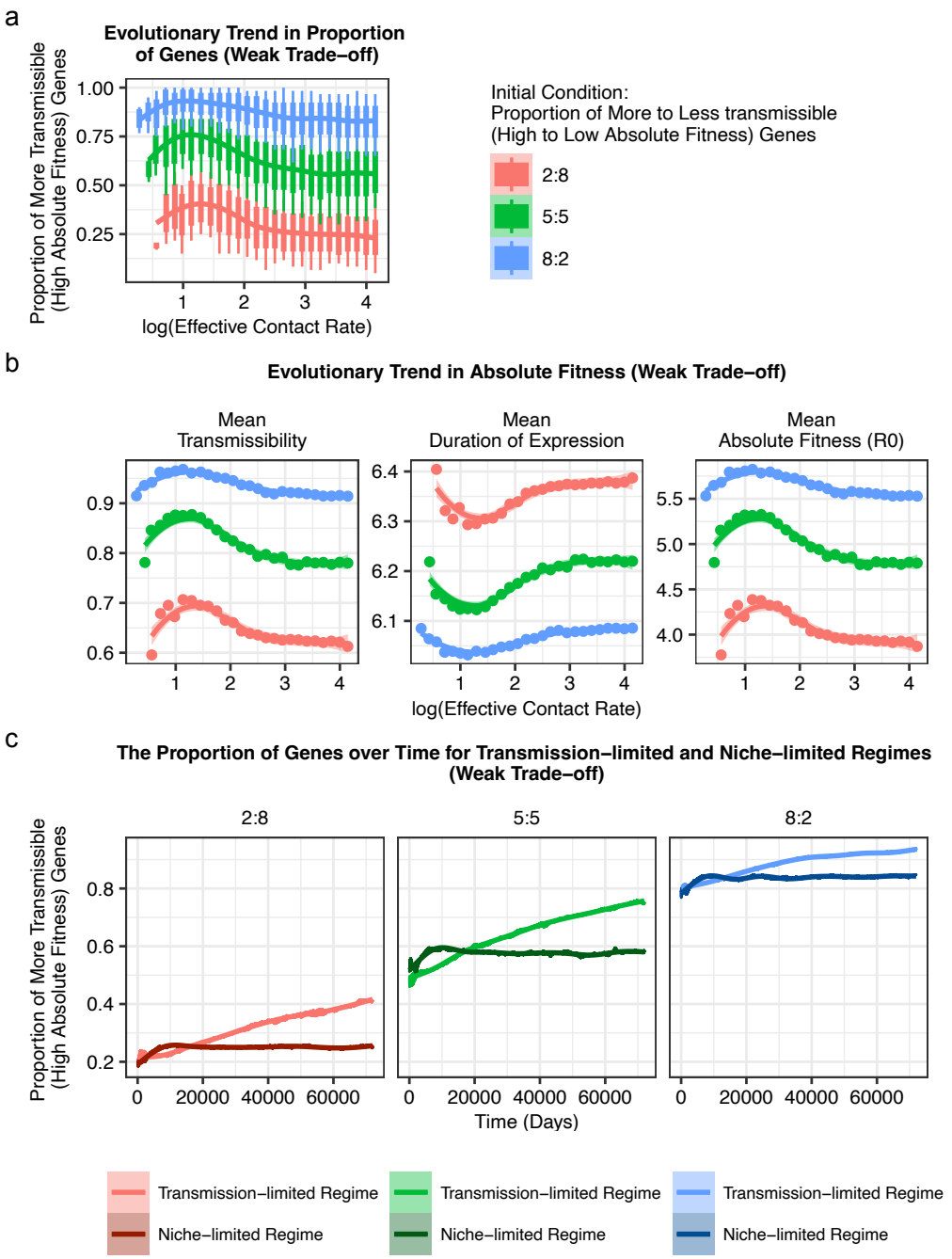

Figure 2. Trade-off between transmissibility and duration of expression. In the weak trade-off case, **a** with increasing transmission intensity, the proportion of more transmissible (high-R_0) genes initially increases in the transmission-limited regime before declining in the niche-limited regime. **b** Correspondingly, population-wide mean transmissibility and R_0 values first increase and then decrease, while mean duration of expression shows an inversion trend (initial decrease followed by increase due to trade-off). **c** Temporal dynamics of the proportion of high-R_0 genes reveal distinct trajectories. The proportion shows a sustained increase in a simulation run from the transmission-limited regime (reflecting available immune space), whereas it quickly reaches a plateau in a simulation run from the niche-limited regime (indicating depletion of host immune space). The three panels correspond to the three initial conditions with the shades of colors matching those in **a**. In **a** and **b**, all metrics on the y axis are calculated from end-of-high-season samples in the last year of simulation (Year 200). Points or boxplots show individual simulation runs with specific effective contact rates. Trends are fitted using ggplot2's geom_smooth function. stats::loess() is used for less than 1,000 observations; otherwise mgcv::gam() is used with formula = $y \sim s(x$, bs = "cs") with method = "REML") function. In **a**, each boxplot shows the minimum, 5% quantile, median, 95% quantile, and maximum of values across all sampled strains in each simulation.

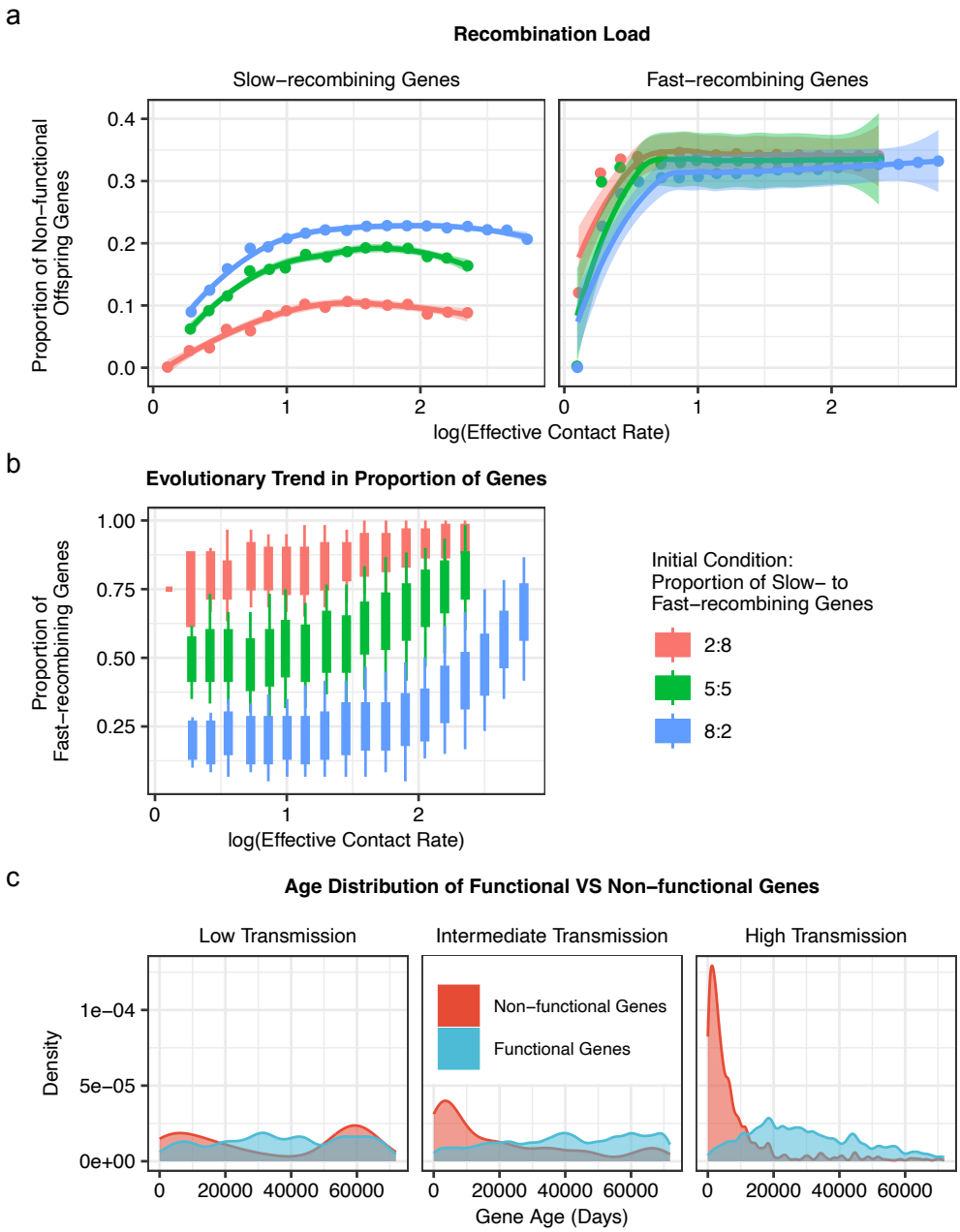

Figure 3. Trade-off between mitotic recombination rate and load associated with mitotic recombination events. **a** Mitotic recombination between fast-recombining genes generates non-functional offspring genes more frequently than that between slow-recombining genes. The former are thus associated with a higher recombination load. **b** The proportion of fast-recombining genes increases as transmission and immune-mediated competition increases. **c** Age distributions of functional versus non-functional genes using three simulation runs at low, intermediate, and high transmission, respectively. At low transmission, a significant fraction of non-functional genes shows long-term persistence in the population (secondary mode at high ages). At intermediate or high transmission, non-functional genes become more rapidly and efficiently purged (age distribution peaks near zero), while functional genes are long-lived. In **a** and **b**, all metrics on the y axis are calculated from end-of-high-season samples in the last simulation year (Year 200). Points or boxplots represent individual simulation runs with specific effective contact rates. Trends are fitted using ggplot2's geom_smooth function. stats::loess() is used for less than 1,000 observations; otherwise mgcv::gam() is used with formula = $y \sim s(x$, bs = "cs") with method = "REML") function. In **b**, each boxplot shows the minimum, 5% quantile, median, 95% quantile, and maximum of values across all sampled strains.

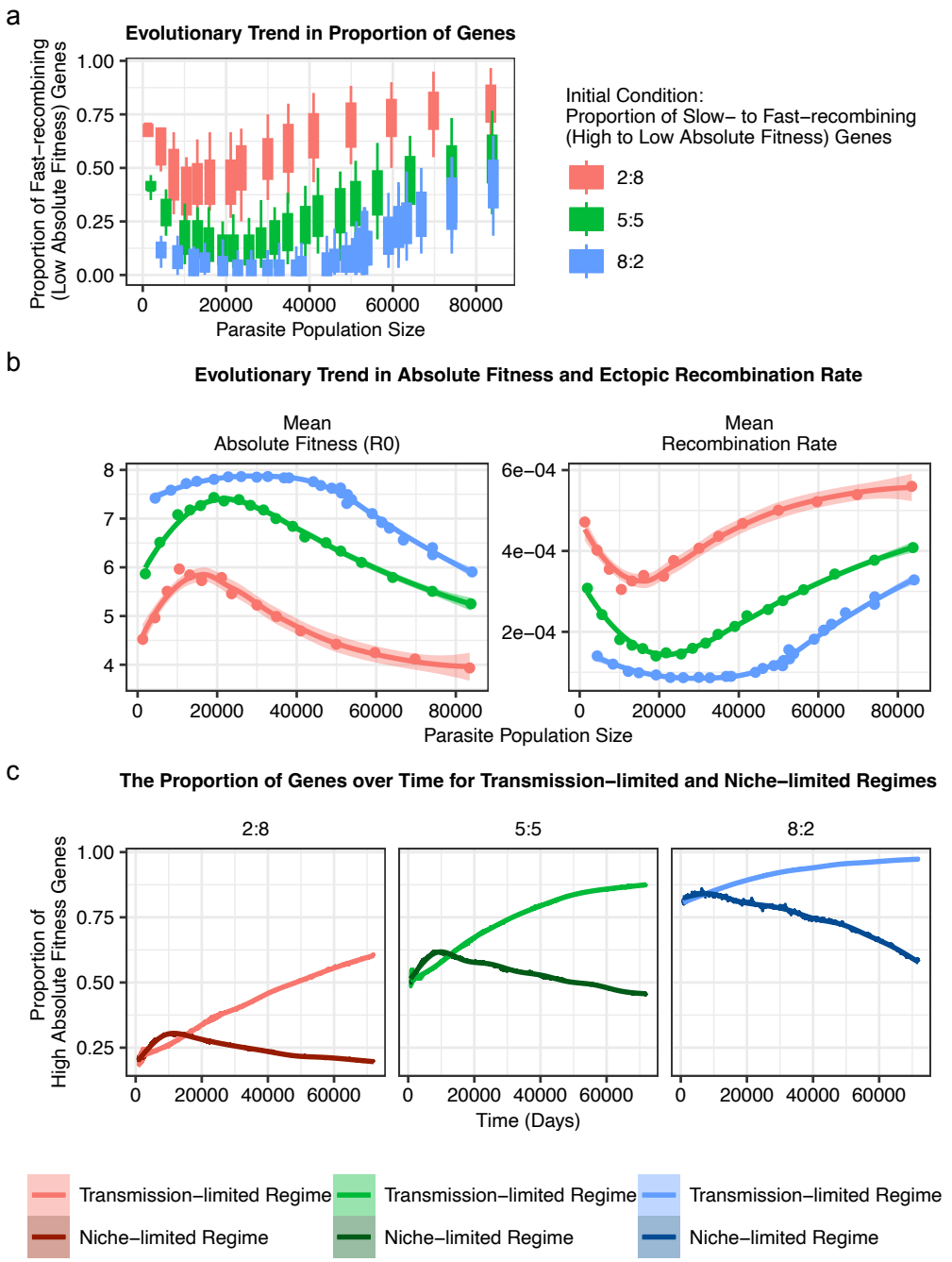

Figure 4. Trade-off between absolute fitness (R_0) and mitotic recombination rate. **a** Slow-recombining (high-R_0) genes are selected at intermediate transmission, but fast-recombining (low-R_0) genes are selected at high transmission. Consistently, **b** Population-level mean absolute fitness first increases and then decreases with transmission (peak at intermediate transmission), while population-level mean recombination rate first decreases and then increases with transmission due to trade-off. **c** Temporal dynamics of the proportion of slow-recombining (high-R_0) genes in the population reveal distinct trajectories. In the simulation from the transmission-limited regime, the proportion accumulates progressively in the population. In the simulation from the niche-limited regime, the proportion initially increases and then decreases. The three panels correspond to the three initial conditions with the shades of colors matching those in **a**. In **a** and **b**, all metrics on the y axis are calculated from end-of-high-season samples in the last simulation year (Year 200). Points or boxplots represent individual simulation runs with specific effective contact rates. Trends are fitted using ggplot2's geom_smooth function. stats::loess() is used for less than 1,000 observations; otherwise mgcv::gam() is used with formula = $y \sim s(x$, bs = "cs") with method = "REML") function. In **a**, each boxplot shows the minimum, 5% quantile, median, 95% quantile, and maximum of values across all sampled strains.

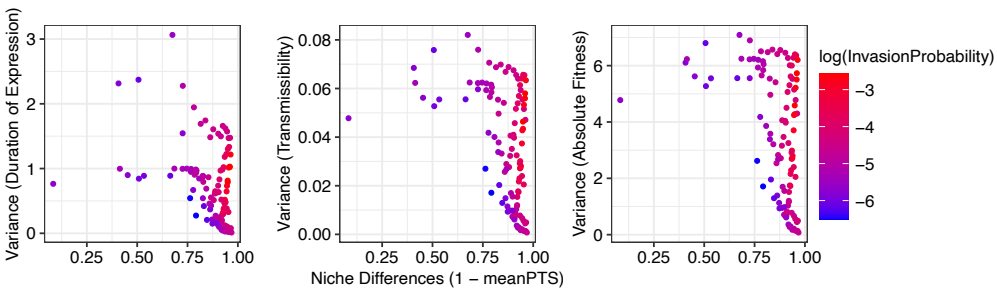

Figure 5. Quantifying invasion probabilities of new genes. Each point represents a simulation run. The variance of traits related to absolute fitness, including duration of expression and transmissibility, as well as absolute fitness, is calculated across all genes. Niche differences are measured as the divergence in *var* genes between strain pairs, or equivalently as 1 minus the mean PTS across all sampled strain pairs. As niche differences increase, the system shifts towards a state with higher invasion probabilities, where new genes are more likely to invade and become established in the population. Meanwhile, the variance of absolute fitness and absolute fitness-related traits tends to decline. Some runs at high niche differences still show elevated variances due to initial conditions (see Results for details).