# A multilocus perspective on the evolutionary dynamics of multistrain pathogens

David V. McLeod*[1,2,3], Claudia Bank†[2,3,5], and Sylvain Gandon‡[4,5]

[1]Département de mathématiques et statistique, Université de Montréal, Montréal, Canada
[2]Institute of Ecology and Evolution, Universität Bern, Bern, Switzerland
[3]Swiss Institute of Bioinformatics, Lausanne, Switzerland
[4]CEFE, CNRS, Univ Montpellier, EPHE, IRD, Montpellier, France
[5]equal contribution

June 18, 2024

**Abstract**

Many human pathogens, including malaria, dengue, influenza, *Streptococcus pneumoniae*, and cytomegalovirus, coexist as multiple genetically distinct strains. Understanding how these multistrain pathogens evolve is of critical importance for forecasting epidemics and predicting the consequences of vaccination. One factor believed to play an important role is naturally acquired immunity. Consequently, a large body of research has sought to predict how acquired immunity molds the genomics of pathogen populations (i.e., what shapes pathogen strain structure). The diversity of existing models has resulted in conflicting evolutionary predictions, and has sparked an ongoing debate about which predictions are most broadly applicable. Here, we adopt a multilocus population genetics perspective that unifies the predictions of existing models. We identify three key factors that determine the role of naturally acquired immunity in the evolution of pathogen strain structure: (i) the strength and specificity of immune protections, (ii) the dynamic immunological landscape, and (iii) the number of loci coding for the antigens of the pathogen. Isolating and discussing these three factors clarifies the relationship among previous models of multistrain dynamics, and establishes a solid theoretical foundation for the study of the evolutionary epidemiology of multistrain pathogens.

**Keywords**: evolutionary epidemiology, multistrain pathogens, adaptive immune response, pathogen strain structure

*david.mcleod@umontreal.ca
†claudia.bank@unibe.ch
‡sylvain.gandon@cefe.cnrs.fr

# 1 Introduction

Genomic data of many pathogens of importance for human health, including malaria [1], dengue [2], *Streptococcus pneumoniae* [3], and cytomegalovirus [4], reveal species genetically subdivided into distinct genotypes, or strains, characterized by statistical associations between alleles at different loci [5–8]. Disentangling what produces these genomic patterns (henceforth termed pathogen strain structure or PSS) is a major focus of evolutionary epidemiology, owing to its relevance for a variety of human health concerns, including forecasting epidemics [9] and predicting the epidemiological consequences of vaccination campaigns [10, 11]. One factor believed to play an important role is naturally acquired host immunity, and a large body of work has sought to discern how host immunity affects the population genomics of pathogens [6, 9, 12–20].

The aspect of naturally acquired immunity that has received the most attention is the role of cross-protection, broadly defined as any immune protection an infection provides against genetically similar strains [e.g., 6, 9, 12, 14–16, 21–25]. Roughly speaking, the predictions are as follows [15, 26]: if cross-protection is weak, there is no PSS; if it is of intermediate strength, evolutionary oscillations are common; if it is strong, a set of strains sharing no antigens in common is overrepresented in the population [6, 12]. Yet, each of these predictions comes with exceptions, which has caused debate about the generality of the results. For example, it is not well-understood why only some models produce evolutionary oscillations [27], how predictions of PSS scale with an increasing number of immunogenic loci and alleles [18–20], or what patterns of PSS are either possible or likely [28]. One reason for this debate is that models which combine naturally acquired immunity with pathogen genetic diversity require a large number of variables [29]. Consequently, existing work has taken different approaches to simplify the dynamics, with poorly understood consequences [27, 29, 30].

Here we take a population genetics perspective to review, generalize, and provide evolutionary intuition about when, and why, existing predictions hold (or not). We do so in three parts, each of which emphasizes biological intuition [see e.g., 27, 30, for a more quantitative perspective]. In Section 2, we define PSS. In Section 3, we use this definition to identify the evolutionary forces producing PSS. In Section 4, we show how the identified evolutionary forces are shaped by the epidemiology of naturally acquired immunity. This results in the identification of three crucial factors that determine whether and which type of PSS occurs: the strength and specificity of the immune protections, the dynamic immunological landscape, and the amount of pathogen genetic diversity. We conclude by discussing future directions. Notation and acronyms used are detailed in Table 1.

# 2 What is pathogen strain structure (PSS)?

Pathogen strain structure (PSS) refers to the statistical association between alleles at different loci which are neither explainable by chance or by selection on individual alleles. Thus, PSS requires some strains (i.e., some multilocus combination of alleles) to be *overrepresented* relative to their expectation under a random shuffling of alleles, that is, some strains are more frequent than the product of their allele frequencies. Hence, identification of PSS necessitates specifying the multilocus pathogen genetics. For a pathogen genome consisting of two diallelic loci, $\{a, A\}$ and $\{b, B\}$, corresponding to four strains $\{ab, Ab, aB, AB\}$, PSS

requires

$$D(t) \equiv p_{AB}(t) - p_A(t)p_B(t) \neq 0, \tag{1}$$

where $p_i(t)$ and $p_{ij}(t)$ denote the frequency of infections by strains carrying allele(s) $i$ and $ij$, respectively, at time $t$. In classical population genetics, the quantity $D(t)$ is pairwise linkage disequilibrium (LD) [31], and is expected to be removed by recombination. The concept of LD generalizes to multiple loci and alleles [31], and since LD captures any nonrandom associations between alleles, the presence of LD is necessary for some strains to be overrepresented. Hence, the minimal definition of PSS is the presence of LD (pairwise or higher-order).

A more stringent form of PSS, *discordant strain structure* (DSS), was first proposed by Gupta et al. [6]. DSS occurs when a set of strains whose members share no alleles (e.g., $\{ab, AB\}$ or $\{Ab, aB\}$) is overrepresented in the population [6, 12, 32]. The presence of pairwise LD is a necessary but not sufficient condition for DSS [32, 33].

Finally, although PSS requires the presence of LD, in the long term, this could correspond to either LD reaching a stable equilibrium with unchanging strain frequencies or to LD undergoing sustained (potentially chaotic) oscillations, where strain frequencies temporally fluctuate. If LD reaches a non-zero stable equilibrium, we will refer to this as *stable PSS*, whereas if LD undergoes sustained oscillations, we will refer to this as *oscillatory PSS*.

# 3 Which evolutionary forces produce strain structure?

As PSS requires the presence of LD, the dynamics of LD are key to understanding its evolution. To describe these dynamics, we initially focus on two diallelic loci. Moreover, we initially assume that each allele has the same intrinsic fitness, that is, each allele induces the same degree of immune protections. For multistrain models of adaptive immunity without intrinsic fitness differences between alleles, two timescales often emerge [27, 34]. On the fast timescale, strong negative frequency-dependent selection (NFDS) operates through the additive selection coefficients at each locus, $w_A \equiv r_{Ab} - r_{ab}$ and $w_B \equiv r_{aB} - r_{ab}$, to push the alleles to an equal frequency, $p_A(t) \approx p_a(t) \approx p_B(t) \approx p_b(t)$ (here, $r_{ij}$ is the per-capita growth rate of strain $ij$; see Box 1). At near-equal allele frequencies, the strains necessarily self-organize into two sets $\{ab, AB\}$ and $\{Ab, aB\}$ with $p_{ab}(t) \approx p_{AB}(t)$ and $p_{Ab}(t) \approx p_{aB}(t)$. Following this self-organization, strains belonging to the same set behave similarly, that is, they have the same per-capita growth rates (i.e., $r_{AB} \approx r_{ab}$ and $r_{Ab} \approx r_{aB}$). As a consequence of $r_{Ab} \approx r_{aB}$, it follows that $w_A \approx w_B$. Moreover, if we let $w_{AB}$ denote epistasis in fitness between alleles $A$ and $B$, that is,

$$w_{AB} \equiv r_{AB} + r_{ab} - r_{Ab} - r_{aB},$$

then, using the relations between the per-capita growth rates, we obtain

$$w_{AB} \approx 2(r_{ab} - r_{Ab}) \approx -2w_A$$

and so $w_A \approx w_B \approx -w_{AB}/2$. Consequently, while the allele frequencies are in quasi-equilibrium, in the long term, LD evolves according to the equation

$$\frac{\mathrm{d}D}{\mathrm{d}t} \approx w_{AB}\left(\frac{1}{4} + D\right)\left(\frac{1}{4} - D\right) + \Delta_D, \tag{2}$$

where $\Delta_D$ is the net effect of recombination on LD (Sup. Info. S3 and equations (S22)). Equation (1) and (2) reveal the evolutionary forces that generate or erode PSS.

First, equation (1) implies that the presence of PSS requires both loci to be polymorphic. Maintaining genetic polymorphisms is rarely a problem in models of adaptive immunity, as NFDS will tend to maintain allelic diversity. In fact, most models tend to maximize this diversity; for example, in the case of two diallelic loci without intrinsic fitness differences between alleles, allele frequencies tend to 1/2. As our interest is what favours or disfavours some strains given allelic variation observable in genomic data, we will not consider further what favours or disfavours allelic diversity [e.g., see 8, 35, 36].

Second, assuming that both loci are polymorphic, the long-term presence of PSS requires a force that generates LD. Equation (2) yields two possibilities: epistasis and (biased) recombination. Because virtually all multistrain models implicitly revolve around epistasis [but see 37, 38], we will assume that recombination is weak and unbiased, i.e., $\Delta_D \approx -\rho D$, where $\rho$ is a (small) positive constant. Whenever the rate of recombination is small relative to the magnitude of epistasis, that is, $\rho \ll |w_{AB}|$, the dynamics of (2) are driven by epistasis, and its sign indicates whether the combination of alleles $A$ and $B$ is favored, i.e., $w_{AB} > 0$, or disfavoured, i.e., $w_{AB} < 0$.

# 4 Epistasis in an epidemiological context

How is epistasis in fitness shaped by epidemiology? To answer this requires an epidemiological model. When possible, we will use the model of Box 1. Briefly, in this model, we assume that all strains have the same transmission rate $\beta$ and exposure to an allele (or antigen) induces an adaptive immune response in the host that protects against future infections. In particular, the probability of infection of a host with immunity to $n = 0, 1, 2$ antigens of the transmitted strain is $\sigma_n$, and the probability a host is re-infected by the same strain is $\sigma_h$ (homologous immunity). Then epistasis is

$$\begin{aligned} w_{AB} = \beta(\overbrace{\sigma_0 - 2\sigma_1 + \sigma_2}^{\sigma_{\mathcal{A}}})(\overbrace{\mathcal{A}_{\{a,b\}}(t) + \mathcal{A}_{\{A,B\}}(t) - \mathcal{A}_{\{A,b\}}(t) - \mathcal{A}_{\{a,B\}}(t)}^{\mathcal{A}_\Delta(t)}) \\ - \beta(\underbrace{\sigma_2 - \sigma_h}_{\sigma_{\mathcal{S}}})(\underbrace{\mathcal{S}_{ab}(t) + \mathcal{S}_{AB}(t) - \mathcal{S}_{Ab}(t) - \mathcal{S}_{aB}(t)}_{\mathcal{S}_\Delta(t)}), \end{aligned} \quad (3)$$

The quantity $\mathcal{A}_{\{i,j\}}(t)$ denotes the density of uninfected hosts with immunity to *at least* the set of antigens $i$ and $j$ ($\mathcal{A}$ for *antigen*). In contrast, the quantity $\mathcal{S}_{ij}(t)$ denotes the density of uninfected hosts that have been previously infected by strain $ij$ ($\mathcal{S}$ for *strain*; Box 1). Equation (3) shows that two factors contribute to the dynamics of epistasis. The first is the strength and specificity of the immune protection, captured by the composite parameters $\sigma_{\mathcal{A}}$ and $\sigma_{\mathcal{S}}$ (Fig. 2). The second is the dynamic immunological landscape, captured by the composite variables $\mathcal{A}_\Delta(t)$ and $\mathcal{S}_\Delta(t)$.

**Box 1: A population model of naturally acquired immunity**

Models of adaptive immunity largely differ regarding three types of assumptions (Sup. Info. S2): the type of immune protection considered (e.g., does immunity reduce infection or transmissibility?); how immunity is gained and lost; and whether hosts can be simultaneously infected by multiple strains (co-infection). Typically, the choice of assumptions is made to simplify the mathematics. The model we focus on here prioritizes conceptual clarity; extensions and links to other models can be found in the Supplementary Information.

We consider a pathogen whose genome consists of some number of multiallelic loci; each allele, or antigen, induces a (host) adaptive immune response. Transmission occurs exclusively to uninfected hosts (so we do not permit co-infection) via mass-action with the same rate constant $\beta$ for all strains, whereas the probability of successful infection, $\sigma_x^g$, depends upon the host's immune history, $x$, and the genotype, $g$, of the pathogen strain. In principle, $x$ could be any combination of antigens and/or strains and could also account for the temporal order of infections (e.g., more recent infections induce a more protective response). However, here we assume that $x$ is an unordered set of prior infections. Thus for $n$ pathogen strains, there are $2^n$ possible infection histories. For a host with immunity to $n = 0, 1, ...$ antigens of strain $g$ but no prior infection by that strain, $\sigma_x^g = \sigma_n$, whereas for a host with prior infection by strain $g$, $\sigma_x^g = \sigma_h$ (homologous immunity). For example, for two diallelic loci, a host with immune history $x' = \{ab, Ab\}$ has $\sigma_{x'}^{ab} = \sigma_{x'}^{Ab} = \sigma_h$ and $\sigma_{x'}^{aB} = \sigma_{x'}^{AB} = \sigma_1$. For simplicity, we will assume that $\sigma_0 \geq \sigma_1 \geq \cdots \geq \sigma_n \geq \sigma_h$. Hosts clear infections and die from natural and virulence-related mortality at per-capita rates $\mu$, $d$, and $\alpha$, respectively. Immunologically naive hosts ($x = \{\varnothing\}$) enter the population at rate $b$; immunity is life-long and acquired at clearance. For hosts with immune history $x$, let $S_x(t)$ denote the density of uninfected individuals and $I_x^g(t)$ denote the density of individuals infected by strain $g$ at time $t$. Then

$$\begin{aligned} \frac{\mathrm{d}S_x}{\mathrm{d}t} &= b\mathbf{1}_0(x) - dS_x - \beta \sum_g \sigma_x^g \sum_z I_z^g S_x + \sum_g \sum_{z:g\cup z=x} \mu I_z^g \\ \frac{\mathrm{d}I_x^g}{\mathrm{d}t} &= \beta \sum_z I_z^g \sigma_x^g S_x - (d+\mu+\alpha) I_x^g \end{aligned} \tag{4}$$

where $\mathbf{1}_0(x) = 1$ if $x = \{\varnothing\}$ and 0 otherwise.

The per-capita growth rates of the different strains link the epidemiological and evolutionary dynamics through the (additive) selection coefficients and epistasis. To extract the selection coefficients and epistasis from model (4), we focus upon two diallelic loci (i.e., $g = ij$, where $i \in \{a, A\}$ and $j \in \{b, B\}$). Let $\mathcal{A}_{\{i,j\}}(t)$ denote the density of uninfected hosts with immunity to *at least* the set of antigens $\{i, j\}$ and $\mathcal{S}_{ij}(t)$ denote the density of uninfected hosts with prior infection by strain $ij$ (Fig. 1). For example, uninfected hosts with infection history $\{ab, AB\}$ would contribute to all four of the variables $\mathcal{A}_{\{a,b\}}(t)$, $\mathcal{A}_{\{A,b\}}(t)$, $\mathcal{A}_{\{a,B\}}(t)$ and $\mathcal{A}_{\{A,B\}}(t)$ but only to the two variables $\mathcal{S}_{ab}(t)$ and $\mathcal{S}_{AB}(t)$. Using this notation, the per-capita growth rate of strain $ij$ is

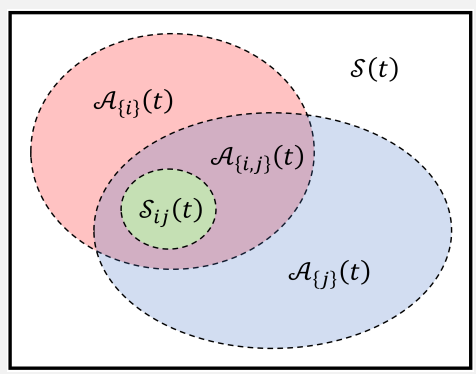


Figure 1: Relationship between variables describing the distribution of immunity among uninfected hosts, where $\mathcal{S} \equiv \sum_x S_x$ is the total density of uninfected hosts.

$$r_{ij} \equiv \sum_x \frac{\mathrm{d}I_x^{ij}/\mathrm{d}t}{\sum_y I_y^{ij}} = \beta \left( \sigma_0 \sum_x S_x - (\sigma_0 - \sigma_1)(\mathcal{A}_{\{i\}} + \mathcal{A}_{\{j\}}) + \sigma_{\mathcal{A}} \mathcal{A}_{\{i,j\}} - \sigma_{\mathcal{S}} \mathcal{S}_{ij} \right) - d - \mu - \alpha.$$

where $\sigma_{\mathcal{A}} \equiv \sigma_0 - 2\sigma_1 + \sigma_2$ and $\sigma_{\mathcal{S}} \equiv \sigma_2 - \sigma_h$.

## 4.1 The strength and specificity of immune protections

Existing models can be partitioned into two groups based on the specificity and strength of immune protections. Hosts acquire *antigen-specific* immunity in the first group [e.g., 18–20, 39]: here, immunity to strains $Ab$ and $aB$ translates to immunity to the set of antigens $\{a, A, b, B\}$ and so full protection against strain $AB$. In this case, $\sigma_2 = \sigma_h$ and so $\sigma_{\mathcal{S}} = 0$, that is, homologous immunity is no more protective than full antigenic immunity. Consequently, the evolutionary dynamics hinge on $\sigma_{\mathcal{A}}$, which controls whether the rate of increase of antigen-specific cross-protection is accelerating ($\sigma_{\mathcal{A}} < 0$) or saturating ($\sigma_{\mathcal{A}} > 0$) as the number of antigens recognized by the immune system of the host increases (Fig. 2). Models of antigen-specific immunity may apply to, for example, pathogens that temporally express different antigens to which the host develops an immune response (e.g., malaria [18–20], pathogenic *Neisseria* [40]).

Hosts acquire *strain-specific* immunity in the second group [e.g., 6, 9, 12, 14–16, 26, 41]: the level of cross-protection against a particular strain is determined by the most antigenically similar strain that has infected the host, but this is less protective than homologous immunity, that is, $\sigma_2 > \sigma_h$ and so $\sigma_{\mathcal{S}} > 0$. Thus $\sigma_{\mathcal{S}}$ controls the degree to which homologous immunity is more protective than full antigenic immunity (Fig. 2). For example, although 'strains' might be defined based on the antigenic variation at a few immunodominant epitopes, strains might also vary at other weakly antigenic sites. If this variation is more similar within a strain than between strains, homologous immunity may be more protective than full antigenic immunity.

More generally, if the adaptive immune system recognizes, and targets, the pathogen at the level of its constitutive antigens, rather than as a unit (i.e., strain), then $\sigma_{\mathcal{S}} = 0$. If, instead, the pathogen is targeted as a unit and cross-protection arises through single-strain similarity, then $\sigma_{\mathcal{S}} > 0$. These two perspectives yield very different evolutionary dynamics, as we will see below.

### Rapidly increasing antigen-specific cross-protection produces stable PSS

When full antigenic immunity is as protective as homologous immunity (i.e. $\sigma_2 = \sigma_h$ and $\sigma_{\mathcal{S}} = 0$), epistasis reduces to $w_{AB} = \beta\sigma_{\mathcal{A}}\mathcal{A}_{\Delta}(t)$, where the variable $\mathcal{A}_{\Delta}(t)$ indicates whether more uninfected hosts have full antigenic immunity to strains belonging to the set $\{ab, AB\}$ (if positive) or to the set $\{Ab, aB\}$ (if negative). There are three possible outcomes, determined by the sign of $\sigma_{\mathcal{A}}$. First, if $\sigma_{\mathcal{A}} < 0$, all strains stably coexist at equal frequencies. Second, if $\sigma_{\mathcal{A}} = 0$, epistasis vanishes. Hence, any observed PSS is neutral and contingent upon the initial conditions. Third, if $\sigma_{\mathcal{A}} > 0$, then stable PSS occurs; in this case, because there is a single epidemiological feedback, $\mathcal{A}_{\Delta}(t)$, one set of strains is competitively excluded. Thus, the sign of $\sigma_{\mathcal{A}}$ drives the formation of PSS. In the Supplementary Information S4, we support these predictions with extensive simulations over a range of $R_0 \in [1.2, 20]$ and $\theta \in [5 \times 10^{-5}, 0.5]$, where

$$R_0 \equiv \frac{\sigma_0\beta}{\mu + \alpha + d}\frac{b}{d} \quad \text{and} \quad \theta \equiv \frac{d}{\mu + \alpha + d} \tag{5}$$

are the basic reproductive number and the ratio of the infectious period to host lifespan [27, 42], respectively. In addition, we analytically show that $\sigma_{\mathcal{A}} > 0$ is the necessary condition for stable PSS.

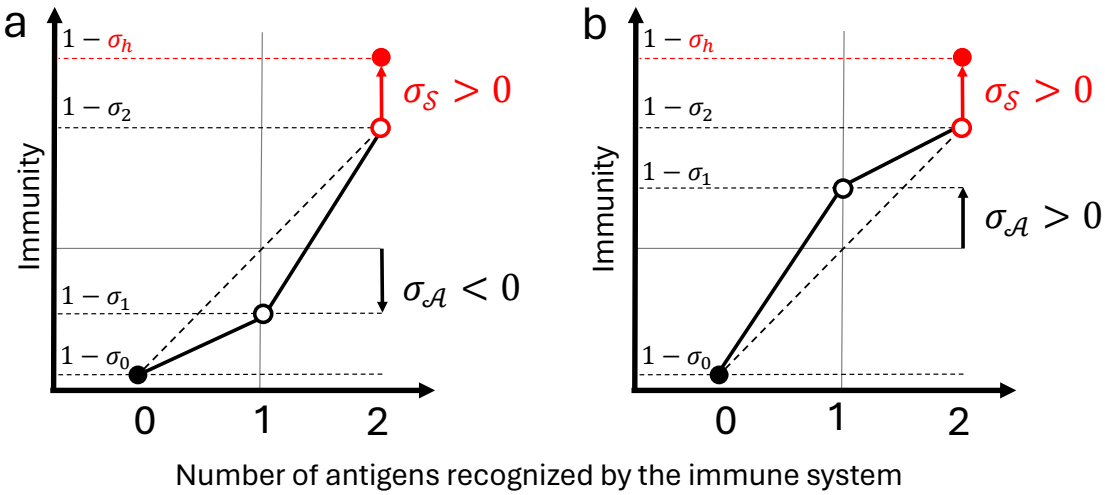


Figure 2: **Schematic of cross-protection**. Specification of cross-protective immunity translates to two composite parameters controlling how cross-protection affects epistasis. The first composite parameter, $\sigma_{\mathcal{A}}$, captures the degree to which the increase in cross-protection is accelerating ($\sigma_{\mathcal{A}} < 0$, panel **a**) or saturating ($\sigma_{\mathcal{A}} > 0$, panel **b**) as the number of antigens recognized by the immune system increases. The second composite parameter, $\sigma_{\mathcal{S}}$, captures whether homologous immunity provides equivalent ($\sigma_{\mathcal{S}} = 0$) or more ($\sigma_{\mathcal{S}} > 0$) protection than full antigenic immunity.

To understand the impact of the sign of $\sigma_{\mathcal{A}}$, suppose that, starting from a state in which each strain is equally abundant ($D(t) = 0$ and $\mathcal{A}_{\Delta}(t) = 0$), strains belonging to the set $\{ab, AB\}$ (say) slightly increase in frequency by infecting more hosts (so $D(t) > 0$, that is, these strains become overrepresented). As these infected hosts clear the infections, this will produce an excess of uninfected hosts with immunity to strains belonging to the overrepresented set of strains $\{ab, AB\}$, that is, $\mathcal{A}_{\Delta}(t) > 0$. These uninfected hosts have immunity to one of the antigens of the underrepresented strains, $\{Ab, aB\}$ and immunity to zero or two antigens of the overrepresented strains. If $\sigma_{\mathcal{A}} < 0$, then $\sigma_1 > (\sigma_0 + \sigma_2)/2$, and so these uninfected hosts are more susceptible to infection by a randomly chosen strain from the set of underrepresented strains than from the set of overrepresented strains. Consequently, the underrepresented strains will increase in frequency, returning $D(t)$ to 0. If $\sigma_{\mathcal{A}} = 0$, then $\sigma_1 = (\sigma_0 + \sigma_2)/2$, and so these uninfected hosts are equally likely to be infected by any strain, and therefore $D(t)$ will not change any further. Finally, if $\sigma_{\mathcal{A}} > 0$, then $\sigma_1 < (\sigma_0 + \sigma_2)/2$, and so these uninfected hosts are more susceptible to infection by a randomly chosen overrepresented strain. Consequently, the overrepresented strains will further increase in frequency, creating a positive feedback loop that will push $D(t) \to 1/4$.

This analysis assumes that there are no intrinsic fitness differences between alleles or strains. However, some intrinsic variation may be expected in the ability of the host immune system to target specific antigens or strains [43, 44]. If we suppose alleles and/or combinations of alleles vary in the strength of induced immune protections (i.e., $\sigma_1$ and $\sigma_2$ depend on the identity of the allele(s)), then each allelic combination $\{i, j\}$ will have a specific value of $\sigma_{\mathcal{A}}$, i.e., $\sigma_{\mathcal{A}}^{\{i,j\}}$. If we define $\bar{\sigma}_{\mathcal{A}}$ as the average of $\sigma_{\mathcal{A}}^{\{i,j\}}$, numerical simulations indicate that $\bar{\sigma}_{\mathcal{A}}$ remains a strong predictor of whether PSS occurs or not, provided allelic variation is not too

large (Sup. Info. S7). Unsurprisingly, the accuracy of $\bar{\sigma}_{\mathcal{A}}$ as an evolutionary predictor varies with its value: $\bar{\sigma}_{\mathcal{A}}$ is most accurate when strongly positive (predicting stable PSS), and is least accurate when $\bar{\sigma}_{\mathcal{A}} = 0$. The reason why the prediction breaks down when $\bar{\sigma}_{\mathcal{A}} = 0$ is because epistasis is the sum of the product of the allele combination specific $\sigma_{\mathcal{A}}^{\{i,j\}}$ parameters and the dynamic immunological landscape, $\mathcal{A}_{\{i,j\}}$ (equation (S49)). Thus epistasis need not be zero even if $\bar{\sigma}_{\mathcal{A}}$ is.

## Strongly protective homologous immunity generates oscillatory PSS and maintains strain diversity

Importantly, the sign of $\sigma_{\mathcal{A}}$ alone cannot sustain evolutionary oscillations (oscillatory PSS), nor can it stably maintain an intermediate level of stable PSS (i.e., $0 < |D(t)| < 1/4$); two common predictions of multistrain models. This is not surprising: when $\sigma_{\mathcal{S}} = 0$, there is a single epidemiological feedback on epistasis, whether more hosts have *antigenic immunity* to the combination of antigens of strains belonging to one set of strains or the other, $\mathcal{A}_{\Delta}(t)$. Consequently, there is no reason to expect the sign of epistasis to change based on the amount of LD. However, if homologous immunity is more protective than full antigenic immunity, then $\sigma_{\mathcal{S}} > 0$. This creates a second epidemiological feedback, whether more hosts have *homologous immunity* to strains belonging to one set or the other, $\mathcal{S}_{\Delta}(t)$. The second feedback promotes strain diversity because overrepresented strains will have caused more infections, putting them at a competitive disadvantage. If $\sigma_{\mathcal{A}} < 0$, both feedbacks prevent strains from becoming overrepresented, leading to a stable coexistence of all strains at an equal frequency (Fig. 3). However, if $\sigma_{\mathcal{A}} > 0$, in principle promoting PSS, this will be counteracted by the feedback created by $\sigma_{\mathcal{S}} > 0$, which inhibits PSS (Fig. 3).

The counteracting forces of $\sigma_{\mathcal{A}} > 0$ and $\sigma_{\mathcal{S}} > 0$ can produce oscillatory PSS, where the amount of PSS undergoes sustained oscillations (potentially chaotic), and numerical simulations indicate $\sigma_{\mathcal{S}} > 0$ remains a necessary condition for oscillations even when there is intrinsic variation in fitness of alleles (Sup. Info. S7). Indeed, the divergence between homologous immunity and full antigenic immunity is not only the source of the complex dynamics seen in many models [6, 9, 12, 16, 45], it also explains why oscillations are absent from so-called 'status-based' models but not from 'history-based' models [13, 27, 30, 34]. History-based models, such as the model in Box 1, assume that following infection, hosts may be partially protected against genetically similar strains (i.e., they have a reduced, but non-zero probability of infection). On the other hand, status-based models assume that following infection, hosts become either completely protected or unprotected against each strain, with the probabilities dependent on immune history (Sup. Info. S2). This is sometimes referred to as *polarized immunity*, as for each strain, the population is divided into either naive hosts or hosts with sterilizing immunity (i.e., if the host recognizes *any* antigens, it cannot be infected, so $\sigma_n = 0$ for $n \geq 1$ and $\sigma_h = 0$). Consequently, for status-based models, there is no difference between full antigenic or homologous immunity (so $\sigma_{\mathcal{S}} = 0$), and so no potential for sustained evolutionary oscillations.

## Strain structure depends on the life-history effects of immune protection

Of course, immunity need not only protect against infection. In the model of Box 1, any of transmissibility, $\beta$, clearance rate, $\mu$, and virulence mortality, $\alpha$, could depend upon the immune history of the host. Although in this case epistasis can still be written as a time-

varying quantity as in equation (3), the effective reproductive number of an infection will depend on its current environment (i.e., the immune history of the host). Consequently, the per-capita growth rate of strain $ij$ is the weighted average across possible environments, where the weights are the frequency of strain $ij$ infections by environment (Sup. Info. S5). As such, time-varying epistasis is less amenable to interpretation, and so here we focus on epistasis at equilibrium. Suppose homologous immunity is no more protective than full antigenic immunity. Then, the population will evolve to a stable equilibrium, whether it be stable PSS or no strain structure. At the stable PSS equilibrium, epistasis and LD have the same sign if

$$\epsilon_{\mathcal{A}} \equiv f_0 - 2f_1 + f_2 \tag{6}$$

is positive (Sup. Info. S5). In equation (6), $f_n \equiv \frac{\sigma_n \beta_n}{\mu_n + \alpha_n + d}$ is proportional to the effective reproductive number of an infection in a host with immunity to $n$ of its antigens (Sup. Info. S5). In other words, the quantity $\epsilon_{\mathcal{A}}$ measures how cross-immunity increases with the number of antigens recognised by the immune system when immunity affects multiple traits of the pathogen ($\epsilon_{\mathcal{A}}$ is proportional to $\sigma_{\mathcal{A}}$ when immunity affects only $\sigma$). Moreover, the sign of $\epsilon_{\mathcal{A}}$ captures whether the increase in cross-protection is accelerating ($\epsilon_{\mathcal{A}} < 0$, Fig. 2**a**) or saturating ($\epsilon_{\mathcal{A}} > 0$, Fig. 2**b**) as the number of antigens recognized by the immune system increases.

Equation (6) reveals that the life-history effects of immune protection alter the evolutionary predictions in two ways. First, although increasing the protection offered by immunity to one antigen promotes PSS if immunity reduces susceptibility and/or transmissibility [6, 12, 15, 46] or increases clearance rate [18–20, 39], it inhibits PSS if immunity reduces virulence mortality [47, 48]. Note, however, that immune protections against virulence mortality alone are insufficient to maintain genetic polymorphisms (Sup. Info. S5). Second, if each antigen produces an additive increase in the strength of immunity [8], and and we suppose only one type of immune protection is present, then if immunity reduces susceptibility to infection or transmissibility, $\epsilon_{\mathcal{A}} = 0$, whereas if immunity increases clearance rate (a common assumption for malaria models [18–20]), $\epsilon_{\mathcal{A}} > 0$. Thus, immunity that increases the clearance rate is more likely to produce PSS. This is particularly true when pathogens produce acute infections (smaller $\theta$) and/or do not cause host death (smaller $\alpha$ for fixed $\theta$) (Sup. Info. S5).

## 4.2 The dynamic immunological landscape

In addition to the strength and specificity of immune protections, epistasis also depends upon the dynamic immunological landscape controlled by $\mathcal{A}_\Delta(t)$ and $\mathcal{S}_\Delta(t)$. When models become more complex than that of Box 1, this landscape plays an increasingly prominent role. However, because these complexities typically alter the dynamics of uninfected hosts, the analysis of equation (3) helps understand the influence of the immunological landscape on the presence of PSS. In particular, this analysis illustrates the impact of re-infections and how immunity is gained and lost.

### The frequency of re-infections shapes the evolutionary dynamics

When both epidemiological feedbacks, $\beta\sigma_{\mathcal{A}}\mathcal{A}_\Delta(t)$ and $\beta\sigma_{\mathcal{S}}\mathcal{S}_\Delta(t)$, are acting on (3), the sign and dynamical behaviour of epistasis depend upon the difference between the distribution of

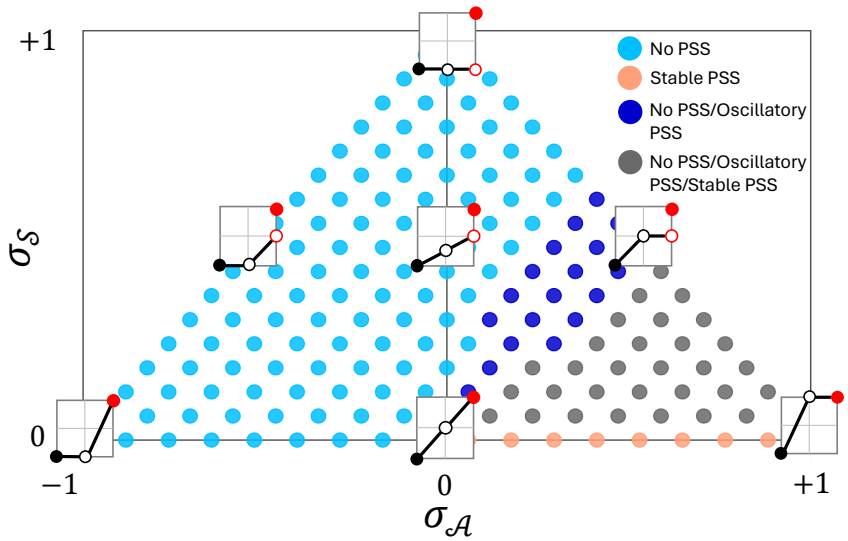


Figure 3: **The strength and specificity of cross-protection determines when PSS can form.** The strength and specificity of cross-protection is controlled by $\sigma_{\mathcal{S}}$ and $\sigma_{\mathcal{A}}$ (see Fig. 2); as they vary, this permits the long-term existence of stable or oscillatory PSS or some combination of both. If $\sigma_{\mathcal{A}} < 0$, no PSS is expected, regardless of the value of $\sigma_{\mathcal{S}}$ (light blue dots). If $\sigma_{\mathcal{S}} = 0$ and $\sigma_{\mathcal{A}} > 0$, stable PSS is expected (light red dots). In between these extremes, different outcomes are possible depending upon the epidemiological parameters (i.e., $R_0$ and $\theta$). In particular, if $0 < \sigma_{\mathcal{A}} < \sigma_{\mathcal{S}}$, both oscillatory PSS and no PSS are possible, but not stable PSS (dark blue dots), whereas if $0 < \sigma_{\mathcal{S}} < \sigma_{\mathcal{A}}$, no PSS, stable PSS, and oscillatory PSS are possible (gray dots). Each circle represents the long-term outcome of 100 simulations of equation (4), where we loop over 10 evenly-spaced values of $R_0 \in [1.2, 18]$ and 10 evenly-spaced values (on log-scale) of $\theta \in [10^{-4}, 0.39]$. Each square inset shows how immunity varies in $(\sigma_{\mathcal{A}}, \sigma_{\mathcal{S}})$ space using the same format as Figure 2. Note that the values of $\sigma_{\mathcal{A}}$ and $\sigma_{\mathcal{S}}$ are constrained due to the assumption that $\sigma_0 \geq \sigma_1 \geq \sigma_2 \geq \sigma_h$. Representative dynamical behaviour through time for the different cases is shown in Figure S2. All simulations use parameters $b = d = 0.01$, $\alpha = 0$, $\sigma_0 = 1$, $\sigma_h = 0$, and $\rho = 10^{-9}$.

immunity to different antigens, $\mathcal{A}_\Delta(t)$, and different strains, $\mathcal{S}_\Delta(t)$. We show in Sup. Info. S3.2 that only hosts with immunity to the antigens defining a single strain contribute to $\mathcal{A}_\Delta(t)$, whereas $\mathcal{S}_\Delta(t)$ depends upon hosts with immunity to more than one strain. Thus, the 'length' of the immune history, that is, the number of strains to which a host currently has immunity, plays an important role. If most hosts have immunity to only one strain, then $\mathcal{A}_\Delta(t) \approx \mathcal{S}_\Delta(t)$. In this case, equation (3) shows that stable PSS occurs if $\sigma_{\mathcal{A}} > \sigma_{\mathcal{S}}$; otherwise no PSS occurs. If almost all hosts have immunity to more than one strain, then $\mathcal{A}_\Delta(t) \to 0$ much faster than $\mathcal{S}_\Delta(t) \to 0$. In this case, the epidemiological feedback $\sigma_{\mathcal{S}}\mathcal{S}_\Delta(t)$ dominates, and as this favours rarer strains, no PSS occurs. In between these extremes, stable PSS, no PSS, or oscillatory PSS are all possible outcomes. Thus, a necessary but not sufficient condition for evolutionary oscillations is that reinfection is possible but not too frequent.

In addition to the strength and specificity of immune protections, two generic pathogen life-history quantities affect the relationship between $\mathcal{A}_\Delta(t)$ and $\mathcal{S}_\Delta(t)$. The first is the basic

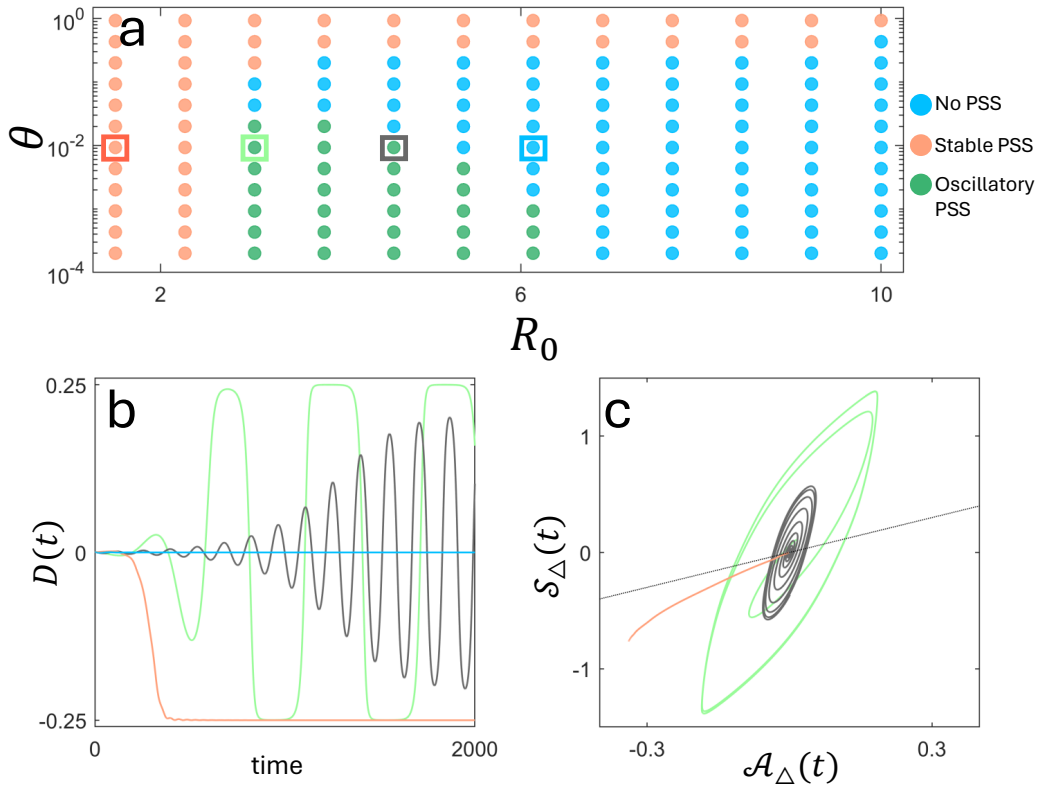


Figure 4: **The dynamic immunological landscape**. Panel **a** shows the long-term evolutionary outcome as $R_0$ and the ratio of infectious period to host lifespan, $\theta$, vary for fixed $\sigma_0 = 1$, $\sigma_1 = 0.35$, $\sigma_2 = 0.25$, $\sigma_h = 0.15$ (so $\sigma_{\mathcal{A}} = 0.55$ and $\sigma_{\mathcal{S}} = 0.1$). When $R_0$ is small and/or infections are increasingly chronic (large $\theta$), $\mathcal{A}_\Delta(t) \approx \mathcal{S}_\Delta(t)$, and since (for this example) $\sigma_{\mathcal{A}} > \sigma_{\mathcal{S}}$, stable PSS occurs (red circles). For a given $\theta$, as $R_0$ increases, it becomes likely that hosts are infected multiple times during their lifespan; here, $\mathcal{S}_\Delta(t)$ becomes larger than $\mathcal{A}_\Delta(t)$, which destabilizes stable PSS. For acute infections (small $\theta$), this can lead to oscillatory PSS (green circles) before sufficiently high $R_0$ leads to no PSS (blue circles), whereas for more chronic infections, the increase in $R_0$ leads to a transition from stable PSS to no PSS without oscillations. Panels **b** and **c** show example dynamical behavior for four different values of $R_0$ and fixed $\theta$ (indicated by the coloured squares in panel **a**). Note that in panel **c** the initial condition is $\mathcal{A}_\Delta(0) = \mathcal{S}_\Delta(0) = 0$. All simulations used parameters $b = d = 0.01$, $\alpha = 0$, $\rho = 10^{-8}$, with $\beta$ and $\mu$ set by the values of $R_0$ and $\theta$.

reproductive number, $R_0$, and the second is the ratio of the infectious period to host lifespan, $\theta$ [27, 42] (see equation (5)). Generally speaking, the 'length' of immune history is expected to increase with $R_0$ and decrease with $\theta$ (see equations (S5) and (S6)). Thus, when $R_0$ is small and/or infections are chronic ($\theta \approx 1$) and co-infection is not possible, hosts are unlikely to be infected more than once during their lifespan, so $\mathcal{A}_\Delta(t) \approx \mathcal{S}_\Delta(t)$. Consequently, if $\sigma_{\mathcal{A}} > \sigma_{\mathcal{S}}$, stable PSS will occur (Fig. 4**a**), otherwise no PSS is expected. As $R_0$ becomes large for a given $\theta$, hosts are commonly infected multiple times during their lifespan, thus $\mathcal{A}_\Delta(t) \to 0$ faster than $\mathcal{S}_\Delta(t) \to 0$, and so no PSS is expected (Fig. 4**a**). Finally, although $\sigma_{\mathcal{S}} > 0$ and an 'intermediate' $R_0$ are necessary conditions for evolutionary oscillations, these conditions

are not sufficient. In addition, infections need to be highly acute (small $\theta$; Fig. 4). This is because oscillations arise through a rapid divergence between $\mathcal{S}_\Delta(t)$ and $\mathcal{A}_\Delta(t)$ (Fig. 4**c**). As both of these quantities involve uninfected hosts, it is not sufficient for reinfections to occur in order for evolutionary oscillations, the infections must also be quickly cleared.

Additional factors may affect the 'length' of the immune history and may thus act on the dynamic immunological landscape. For example, if immunity wanes over time [49] the length of the immune history will be shortened as hosts 'forget' previous infections. Likewise, if there is short-term strain-transcending immunity [17, 41, 50–54] (analogous to no co-infection), the length of immune history will be shortened as the likelihood of becoming re-infected is reduced. If $R_0$ is very high, shortening the immune history will increase the density of hosts contributing to $\mathcal{A}_\Delta(t)$. This increases the likelihood of both oscillatory and stable PSS. If $R_0$ is intermediate, the likelihood of oscillations will be reduced. This dependence upon $R_0$ explains why shortening immune history through waning immunity [49], conserved immune responses [55], or strain-transcending immunity [41] has been observed to both decrease [49, 55] and increase [41] the likelihood of oscillations.

## The impact of cross-protection depends on the gain and loss of immunity

The predicted consequences of the sign of $\sigma_\mathcal{A}$ hinge upon the expectation that the distribution of infections is coupled with the immunological landscape. That is, when hosts acquire immunity, they acquire immunity to all the antigens of the infecting strain, and when hosts lose immunity, they lose immunity to all the antigens of the previously infecting strain. This tends to create a positive correlation between the sign of $D(t)$ and the immunological landscape, $\mathcal{A}_\Delta(t)$ (Fig. S3). However, this relationship can be disrupted if a host infected by strain $ij$ does not always simultaneously acquire and/or lose immunity to the combination of antigens $\{i, j\}$. For example, suppose host heterogeneity in antigen-presenting cells means that hosts are more likely to acquire immunity to only a single, randomly chosen antigen of an infecting strain rather than both. Moreover, suppose that the set of overrepresented strains is $\{ab, AB\}$, that is, $D(t) > 0$. Naive hosts are then most likely to become infected by one of the overrepresented strains (say $AB$), after which they are most likely to become infected by the other overrepresented strain, $ab$, as it is more abundant. But if each infection provides immunity to a single antigen, following the second infection, the host will have immunity to either the set of antigens $\{A, b\}$ or $\{a, B\}$. Therefore, the host will have full antigenic immunity to one of the underrepresented strains, $\{Ab, aB\}$. Consequently, this will produce an excess of hosts with immunity to the antigens of one of the underrepresented strains, that is, $\mathcal{A}_\Delta(t)$ will have the opposite sign to $D(t)$. Thus, from equation (3), PSS can only occur if $\sigma_\mathcal{A} < 0$, i.e., if the increase in cross-protection is accelerating (Sup. Info. S7).

Indeed, there exist a variety of immunological reasons why hosts may not simultaneously gain and/or lose immunity to the combination of antigens $\{i, j\}$. For example, host variation in the adaptive immune response might mean that different hosts target and/or lose immunity to different antigens, based upon host genetics (e.g., MHC alleles [8]) or randomly (e.g., which $B$-cell encounters which epitope first; see [56] and Sup. Info. S7 for a longer discussion). Importantly, this heterogeneity can disrupt the positive correlation between the distribution of infections and the immunological landscape, reversing predictions about the consequences of the sign of $\sigma_\mathcal{A}$.

## 4.3 The number of loci and the possible patterns of PSS

Many of the key predictions from our model of two diallelic loci hold as the number of alleles and number of loci increase: PSS is promoted if the increase in cross-protection saturates as the number of antigens recognized by the host increases; homologous immunity being more protective than full antigenic immunity promotes strain diversity and is necessary for evolutionary oscillations; and oscillations are more likely when hosts tend to be infected sequentially by more than one strain during their lifespan. What changes is that as the number of loci increases, the population need not self-organize into sets of strains whose members have no antigens in common (DSS). As a consequence, different types of PSS are possible [9, 16, 28, 41]. This is because LD is no longer restricted to non-random associations spanning two loci. Instead, LD can involve as many loci as are segregating, with each order of LD corresponding to a different type of PSS. Here, order of LD refers to the number of loci across which the non-random allelic association is measured; see Sup. Info. S9. For example, triple LD is an overrepresentation of strains whose members share at most one antigen (e.g., in the case of three segregating loci, one such set of strains is $\{abC, aBc, Abc, ABC\}$). Generally, the number of overrepresented strains increases with the order of LD. Although the evolutionary dynamics are more complicated, in the long-term, at most one order of LD will be nonzero and evolve. This evolution is still dictated by epistasis, whether it be pairwise or higher-order (Sup. Info. S9).

When full antigenic immunity provides as much protection as homologous immunity, each order of epistasis hinges upon the rate of increase of cross-protection. For $N$ loci, this rate is captured by the $N-1$ composite parameters (Fig. 5a) ,

$$\sigma_{\mathcal{A}}^{(n)} \equiv \sigma_{n-1} - 2\sigma_n + \sigma_{n+1}, \quad n = 1, ..., N-1. \tag{7}$$

As before, the sign of $\sigma_{\mathcal{A}}^{(n)}$ determines whether the rate of increase of cross-protection against $n$ antigens is accelerating, $\sigma_{\mathcal{A}}^{(n)} < 0$, or saturating, $\sigma_{\mathcal{A}}^{(n)} > 0$. Thus, these parameters control how the strength of competition between strains changes based on genetic similarity, shaping epistasis (Fig. 5). For example, for three diallelic loci, triple epistasis in fitness, $w_{ABC}$, can maintain triple LD if $\sigma_{\mathcal{A}}^{(2)} > \sigma_{\mathcal{A}}^{(1)}$. In contrast, pairwise LD between alleles $i$ and $j$ can be maintained if pairwise LD and $w_{ij} + w_{ABC}/2$ share sign, which can be approximated as $\sigma_{\mathcal{A}}^{(2)} > -\sigma_{\mathcal{A}}^{(1)}$ (Sup. Info. S9). One important consequence of multiple possible types of PSS is that the conditions permitting different orders of LD to evolve can be simultaneously satisfied. When this occurs, an evolutionary multistability emerges such that the type of PSS present depends on the initial conditions (Fig. 5; [see also 28]).

Given $N > 2$ loci and the possibility of multiple types of PSS, what genomic patterns are expected? This hinges upon the number of antigens sufficient to induce a strongly protective response. For example, if immunity to two antigens is strongly protective, $\sigma_{\mathcal{A}}^{(1)} \to -1$ and $\sigma_{\mathcal{A}}^{(2)} \to 1$, triple LD is more likely than pairwise LD (Fig. 5). If instead immunity to one antigen is strongly protective, $\sigma_{\mathcal{A}}^{(1)} \to 1$ and $\sigma_{\mathcal{A}}^{(2)} \to 0$, then pairwise LD is more likely (Fig. 5). This explains why models assuming that immunity to one antigen is strongly cross-protective (i.e., $\sigma_1 = \sigma_2 = \cdots = \sigma_n$, and $\sigma_n \to \sigma_h$) predict DSS and pairwise LD [6, 12].

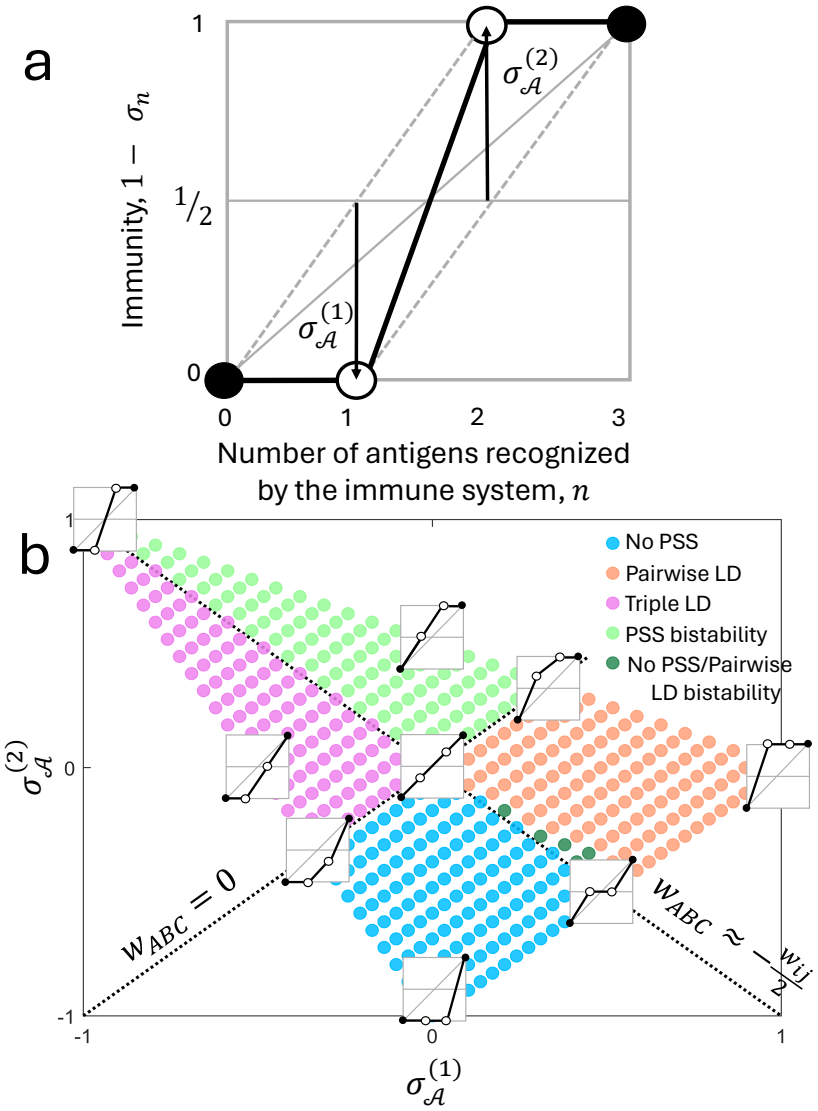


Figure 5: **As the number of loci increases, so do the possible types of PSS**. The number of loci determines the number of parameters capturing the shape of the function that determines how cross-protections increase with the number of antigens recognized by the immune system (panel **a**). As these parameters vary, different long-term outcomes are possible (panel **b**; see Fig. S11 for the representative dynamics through time). When $\sigma_{\mathcal{A}}^{(2)} > \sigma_{\mathcal{A}}^{(1)}$, triple LD has the same sign as $w_{ABC}$ and so can be stably maintained. When $\sigma_{\mathcal{A}}^{(2)} > -\sigma_{\mathcal{A}}^{(1)}$, pairwise LD between alleles $i$ and $j$ will generally have the same sign as $w_{ABC} + w_{ij}/2$ (Sup. Info. S9) and so can be stably maintained. As both conditions can be simultaneously met, evolutionary bistability can occur, where the type of PSS that evolves depends upon the initial conditions. Our assumption of $\sigma_0 \geq \sigma_1 \geq \sigma_2 \geq \sigma_3$ constrains the possible values that $\sigma_{\mathcal{A}}^{(i)}$ can take in panel **b**. Because $\sigma_{\mathcal{S}} = 0$, when PSS occurs, it is stable. Parameter values used in panel **b**: $\beta = 1$, $d = b = 0.01$, $\alpha = 0$, $\mu = 0.4$, $\rho = 10^{-7}$, $\sigma_0 = 1$, $\sigma_3 = \sigma_h = 0$, whereas $\sigma_1$ and $\sigma_2$ vary.

# 5 Discussion

Knowledge of when, and why, naturally acquired immunity produces non-random genomic patterns in pathogen populations (pathogen strain structure or PSS) is critical for understanding and predicting pathogen dynamics. Here we review and generalize predictions for the evolution of PSS by focusing on linkage disequilibrium (LD) and the forces that drive its dynamics. In doing so, we establish the evolutionary connections between the multitude of existing models. In particular, our analysis illustrates the factors which determine (i) the long-term presence of PSS as a stable equilibrium (stable PSS) and (ii) the long-term presence of PSS as evolutionary oscillations (oscillatory PSS).

The long-term presence of stable PSS hinges upon two factors. First, stable PSS is more likely when the increase in cross-protection is saturating ($\sigma_{\mathcal{A}} > 0$) and prior exposure to *all* antigens of a given strain provides similar protection to homologous immunity ($\sigma_{\mathcal{S}} \approx 0$). Second, stable PSS is more likely if hosts have immunity to only a few strains. This occurs for pathogens with a low-to-intermediate $R_0$, pathogens causing chronic infections with limited co-infection, or when hosts display either waning or strain-transcending immunity. The most likely type of stable PSS, expressed through the non-zero order of LD, depends on the number of antigens required to induce a strongly protective immune response. As this number increases, so will the expected non-zero order of LD, and thus the fraction of possible strains present in the population. This implies that, although each type of PSS is in principle detectable from genomic data, it is necessary to identify the appropriate order of LD. For example, a highly structured population dominated by triple LD could appear unstructured based on the assessment of pairwise LD.

We showed that the shape of the function specifying cross-protection (see Fig. 2 and Fig. 5**a**) controls the amount of genotypic diversity which provides testable predictions. For example, vaccine challenge studies of influenza A virus (IAV) in humans and equine influenza virus (EIV) in horses indicate that cross-protection increases roughly linearly with antigenic distance for IAV, but the increase is saturating for EIV (see Fig. 1a in [57]). Consequently, our model would predict that, all else being equal, EIV should be less *genotypically* diverse than IAV. Consistent with our prediction, EIV can be roughly divided into three antigenic clusters [58], whereas IAV consists of 11 antigenic clusters [59] (see also [60]). Notably, EIV and IAV differ with respect to other properties (e.g., IAV undergoes more rapid antigenic drift [60] and the human hosts of IAV have a longer lifespan than the equine hosts of EIV). Quantitatively evaluating how the shape of the function specifying cross-protection limits genotypic diversity thus warrants further work.

Evolutionary oscillations, or oscillatory PSS, have been a focal point of research for multistrain pathogens [e.g., 6, 9, 12, 16, 45]. Our multilocus perspective reveals that oscillatory PSS is most likely when three conditions are met (Fig. 3). First, immunity to each additional antigen provides a larger-than-additive gain in protection ($\sigma_{\mathcal{A}}$ is positive), but only homologous immunity triggers a strongly protective immune response (so $\sigma_{\mathcal{S}}$ is positive). Second, hosts are likely to be infected by, and so have homologous immunity to, multiple strains over their lifespan. This is most likely for pathogens with either an intermediate $R_0$, or those with a high $R_0$ in the presence of waning or strain-transcending immunity (i.e., something shortening the 'length' of immune history). Third, infections are acute ($\theta$ is small), creating a rapid divergence between the distribution of hosts with immunity to certain antigens and the distribution of hosts with immunity to certain strains. Whether these conditions are

met or not will depend upon the pathogen species. We propose that identifying these key biological quantities may have greater utility and be more empirically feasible than identifying regions of parameter space for a given model that permit oscillations. For example, Recker et al. [9] hypothesized that rather than being a mutation-driven process [13, 14, 52], influenza evolution may be better explained by the periodic recycling (or oscillations) of a finite set of antigens (but see [36]). This argument was supported by oscillations occurring over a wide range of parameter space. Future work could evaluate whether influenza meets the biological conditions for displaying such behaviour.

## Co-infection and within-host interactions

The model of Box 1 assumes that transmission only occurs to uninfected hosts. It thus ignores within-host interactions. Multistrain models frequently assume that hosts can be simultaneously infected by any number of *different* strains (co-infection; see Table S2), each of which is unaffected by the presence of other co-infecting strains [e.g., 6, 12, 15, 41, 54, 61, 62]. Under these assumptions, and by redefining certain variables in the model of Box 1, all our key predictions hold (Sup. Info. S2.3). However, assumptions maintain very high levels of genetic diversity, including otherwise selectively disadvantageous trait variation [e.g., hypervirulence 22]. Although a previous simulation study found limited effects of within-host interactions [18], this may not hold generally. For example, co-infection in *Streptococcus pneumonia* can lead to the simultaneous clearance of all infecting strains [8, 63, 64]. This clearance will disproportionately (negatively) affect the more abundant strains, potentially removing PSS which may partially explain why *S. pneumoniae* genomic data show low amounts of strain structure [65]. Thus, the impact of within-host interactions on PSS warrants further work.

## Demographic stochasticity

Our analysis is based on a deterministic model that neglects demographic stochasticity, which was previously shown to affect multistrain dynamics in two ways. First, demographic stochasticity can produce random fluctuations in allele and strain frequencies not seen in a deterministic model [18, 66]. Demographic stochasticity can generate LD, cause LD to drift in a quasi-neutral fashion, and/or trigger infrequent oscillations between strains, thus affecting PSS, especially when epistasis is weak. Indeed, demographic stochasticity coupled with weak epistasis is likely the source of the quasi-stable PSS observed in some stochastic models, in which the identity of the overrepresented strains infrequently changes [18]. For two diallelic loci, epistasis is weak if: $\sigma_{\mathcal{S}} = 0$ and hosts are likely to be infected multiple times ($R_0$ is large); $\sigma_{\mathcal{S}} \approx 0$ and $\sigma_{\mathcal{A}} \approx 0$; and/or hosts are likely to be infected once and $\sigma_{\mathcal{S}} \approx \sigma_{\mathcal{A}}$.

Second, demographic stochasticity causes the loss of strains or antigens by chance [21, 23, 50, 54, 67, 68]. This is especially relevant if the segregating loci have an unequal number of alleles. Here, antigens at the more diverse locus are predicted to be more easily lost in finite populations, particularly when cross-protection is high [23]. This can be explained as follows (see also Sup. Info. S8). If one locus is more antigenically diverse than the others, e.g., if two loci have alleles $\{a, A\}$ and $\{B_1, B_2, B_3\}$, it is impossible to maintain each antigen at its expected frequency, $p_A(t) = 1/2$ and $p_{B_i}(t) = 1/3$, and simultaneously maximize LD between each $\{A, B_i\}$. This creates tension between selection on antigens through the addi-

tive selection coefficients and selection on strains through epistasis in fitness. This tension is modulated by the relative strength of $\sigma_{\mathcal{A}}$ and $\sigma_{\mathcal{S}}$. As cross-protection increases, selection maintaining antigens at their expected frequencies becomes weaker (i.e., the strength of $\sigma_{\mathcal{S}}$ decreases). At the same time, epistasis becomes stronger, favouring some strains (i.e., $\sigma_{\mathcal{A}}$ becomes increasingly positive). Consequently, some antigens at the more diverse loci are forced to lower frequencies and are thus more likely to be stochastically lost.

## The multilocus perspective clarifies the source of PSS and the drivers of antigenic variation

The multilocus perspective identifies epistasis as the source of PSS, which explains the previously contentious role played by negative frequency-dependent selection (NFDS) for PSS [18–20, 69]. Although NFDS is required to maintain genetic polymorphism, it need not yield epistasis (e.g., if $\sigma_{\mathcal{A}} = \sigma_{\mathcal{S}} = 0$). Our multilocus perspective demonstrates that epistasis is necessary to produce strain overrepresentation. Indeed, past work has compared time series of the genomic patterns produced by models of PSS against a null model in which antigenic variation is neutral [19, 20]. The multilocus perspective suggests an intermediate null model, in which there is NFDS but no epistasis (e.g., $\sigma_{\mathcal{A}} = 0$, $\sigma_{\mathcal{S}} = 0$). In such a model, NFDS would maintain antigenic variation but no directional force would maintain distinct allelic combinations, such that demographic stochasticity would be the only cause of LD.

Moreover, the multilocus perspective synthesizes different explanations for observed patterns of antigenic diversity, summarized in a recent review [8]. In addition to host heterogeneity and functional constraints, the authors discussed three hypotheses to explain patterns of antigenic diversity. First, each antigen contributes to host immunity additively (so $\sigma_{\mathcal{A}} = 0$); second, immunity to a single antigen provides strong cross-protection [i.e., $\sigma_1 = \sigma_{i+1}$, $i > 0$; 6, 12]; and third, there is an interaction between antigens at different loci. The multilocus perspective reveals considerable overlap between these hypotheses; for example, although the first and second predict, respectively, no PSS and stable pairwise LD (DSS), the third simply means that there is epistasis (e.g., $\sigma_{\mathcal{A}} \neq 0$), which can lead to either no PSS or stable PSS, as well as PSS consisting of higher-order LD. Indeed, as the number of antigens required to induce a strongly protective immune response determines the probable order of LD, the likelihood a novel antigen can evade population immunity increases with the dominant order of LD. Thus, we would predict a positive correlation between the dominant order of LD and the amount of allelic variation at each locus.

## Multilocus models and vaccination

Our analysis provides new insight into 'evolutionarily optimal' vaccine design, complementing a body of work trying to understand how genetically diverse pathogen populations respond to vaccination [70–73]. Our work shows that the dominant order LD is generally indicative of how many antigens are required to generate a strongly protective immune response. However, our analysis also emphasizes challenges associated with predicting the consequences of vaccination. First, the input of vaccinal immunity to the population may disrupt the relationship between strain frequency and population immunity with complex dynamical consequences [e.g., 10, 11, 74, 75]. Second, the predicted optimal vaccine type depends on whether homologous immunity is more protective than full antigenic immunity ($\sigma_{\mathcal{S}} > 0$), and if so, whether vaccination provides equivalent protection to full antigenic or homologous

immunity. Third, the pathogen population response will hinge upon whether evolutionary bistability is possible, as the influx of vaccinal immunity may alter the dominant order of PSS, increasing or decreasing genotypic diversity. Fourth, vaccines are often chosen to target a subset of strains, such as those linked to virulence factors [76]. The inclusion of additional pathogen life-history variation, such as virulence factors, will induce fitness asymmetries among strains that are difficult to address with existing models. However, by emphasizing key evolutionary quantities, the multilocus perspective provides a useful frame of reference to tackle these questions in the future.

## Conclusion

Many key human pathogen species, including malaria, dengue, *S. pneumoniae*, *N. meningitidis* and cytomegalovirus, consist of multiple coexisting strains that make medical interventions, such as vaccination, challenging. Thus, understanding the dynamics of multistrain pathogens is important not only to unravel what maintains and structures genetic diversity, but as an urgent question of public health. Yet, combining models of pathogen genetic diversity with realistic epidemiological feedbacks has remained a daunting task. One way forward is by adopting a population genetics perspective, and focusing upon evolutionary quantities like linkage disequilibrium. As we have shown here, taking such an approach can disentangle the factors driving the evolutionary dynamics of pathogen populations. It illustrates when, and why, pathogen strain structure occurs, and carries the promise to predict pathogen dynamics and identify pathogen-specific optimal interventions.

## Acknowledgements

SG thanks Julia Gog and Sunetra Gupta for many inspiring discussions.

| Notation | Description |
|---|---|
| $p_i(t)$, $p_{ij}(t)$ | Frequency of infections by strains carrying allele(s) $i$ and $i$, $j$, respectively, at time $t$. |
| $D(t)$ | Pairwise linkage disequilibrium (LD) between alleles $A$ and $B$. |
| PSS | Pathogen strain structure; indicates the presence of LD of some order, that is, some strains are overrepresented. |
| DSS | Discordant strain structure; is a specific form of PSS in which a set of strains sharing no alleles is overrepresented. |
| $w_k$ | Additive selection coefficient for allele $k$. Indicates whether allele $k$ is favoured ($w_k > 0$) or disfavoured ($w_k < 0$). |
| $w_{ij}$, $w_{ijk}$ | Epistasis between alleles $i$, $j$ and $i$, $j$, $k$. If positive (resp. negative), the combination of alleles are more (resp. less) favoured than expected from adding the independent effects of each allele. |
| $\sigma_n$, $\sigma_h$ | Probability of infection of a host with immunity to $n = 0, 1, 2, ...$ antigens, or homologous immunity $h$, of the transmitted strain. |
| $\sigma_{\mathcal{A}}$ | Controls the rate of increase of cross-protection as the number of antigens recognized by the immune system increases. If positive (resp. negative), cross protections are saturating (resp. accelerating). |
| $\sigma_{\mathcal{S}}$ | Controls whether homologous immunity is equivalent to (i.e., $\sigma_{\mathcal{S}} = 0$), or more protective than (i.e., $\sigma_{\mathcal{S}} > 0$), full antigenic immunity. |
| $\mathcal{A}_\Delta(t)$ | Indicates whether more uninfected hosts have full antigenic immunity to strains belonging to the set $\{ab, AB\}$ (i.e., $\mathcal{A}_\Delta(t) > 0$) or to the set $\{aB, Ab\}$ (i.e., $\mathcal{A}_\Delta(t) < 0$). |
| $\mathcal{S}_\Delta(t)$ | Indicates whether more uninfected hosts have homologous immunity to strains belonging to the set $\{ab, AB\}$ (i.e., $\mathcal{S}_\Delta(t) > 0$) or to the set $\{aB, Ab\}$ (i.e., $\mathcal{S}_\Delta(t) < 0$). |
| $f_n$ | Proportional to the effective reproductive number of an infection in a host with immunity to $n$ of its antigens. |
| $\epsilon_{\mathcal{A}}$ | Indicates whether an infection in a host with immunity to one of its antigens has a smaller ($\epsilon_{\mathcal{A}} > 0$) or larger ($\epsilon_{\mathcal{A}} < 0$) effective reproductive number than expected. |
| $R_0$ | Basic reproductive number. |
| $\theta$ | Ratio of infectious period to host lifespan. If small, infections are acute; if large, infections are chronic. |

Table 1: Key notation used in main text.

# References


[1] Gupta, S., Trenholme, K., Anderson, R. M. & Day, K. P., 1994 Antigenic diversity and the transmission dynamics of *Plasmodium falciparum*. *Science* **263**, 961–963.

[2] Holmes, E. C. & Burch, S. S., 2000 The causes and consequences of genetic variation in dengue virus. *Trends Microbiol.* **8**, 74–77.

[3] Donati, C., Hiller, N. L., Tettelin, H., Muzzi, A., Croucher, N. J., Angiuoli, S. V., Oggioni, M., Hotopp, J. C. D., Hu, F. Z., Riley, D. R. *et al.*, 2010 *Genome Biology* **11**, R107.

[4] Lehner, R., Stamminger, T. & Mach, M., 1991 Comparative sequence analysis of human cytomegalovirus strains. *J. Clinical Microbiol.* **29**, 2494–2502.

[5] Maynard Smith, J., Smith, N. H., O'Rourke, M. & Spratt, B. G., 1993 How clonal are bacteria? *Proc. Natl. Acad. Sci.* **90**, 4384–4388.

[6] Gupta, S., Maiden, M. C. J., Feavers, I. M., Nee, S., May, R. M. & Anderson, R. M., 1996 The maintenance of strain structure in populations of recombining infectious agents. *Nature Medicine* **2**, 437–442.

[7] Lipsitch, M. & O'Hagan, J. J., 2007 Patterns of antigenic diversity and the mechanisms that maintain them. *J. R. Soc. Interface* **4**, 787–802.

[8] Georgieva, M., Buckee, C. O. & Lipsitch, M., 2019 Models of immune selection for multi-locus antigenic diversity of pathogens. *Nature Reviews Immunology* **19**, 55–62.

[9] Recker, M., Pybus, O. G., Nee, S. & Gupta, S., 2007 The generation of influenza outbreaks by a network of host immune responses against a limited set of antigenic types. *Proc. Natl. Acad. Sci.* **104**, 7711–7716.

[10] Lipsitch, M., 1997 Vaccination against colonizing bacteria with multiple serotypes. *Proc. Natl. Acad. Sci.* **94**, 6571–6576.

[11] Gupta, S., Ferguson, N. M. & Anderson, R. M., 1997 Vaccination and the population structure of antigenically diverse pathogens that exchange genetic material. *Proc. R. Soc. B* **264**, 1435–1443.

[12] Gupta, S., Ferguson, N. & Anderson, R., 1998 Chaos, persistence, and evolution of strain structure in antigenically diverse infectious agents. *Science* **280**, 912–915.

[13] Gog, J. & Grenfell, B. T., 2002 Dynamics and selection of many-strain pathogens. *Proc. Natl. Acad. Sci.* **99**, 17209–17214.

[14] Gomes, M. G. M., Medley, G. F. & Nokes, D. J., 2002 On the determinants of population structure in antigenically diverse pathogens. *Proc. Roy. Soc. B* **269**, 227–233.

[15] Ferguson, N. M. & Andreasen, V., 2002 The influence of different forms of cross-protective immunity on the population dynamics of antigenically diverse pathogens. In *Mathematical Approaches for Emerging and Re-emerging Infectious Diseases: Models, Methods and Theory* (eds. C. Castillo-Chavez, S. Blower, P. van den Driessche, D. Kirschner & A. Yakubu), volume 125, chapter 9, pp. 157–171. Springer.

[16] Recker, M. & Gupta, S., 2005 A model for pathogen population structure with cross-protection depending on the extent of overlap in antigenic variant repertoires. *J. Theor. Biol.* **232**, 363–373.

[17] Cobey, S. & Lipsitch, M., 2012 Niche and neutral effects of acquired immunity permit coexistence of pneumococcal serotypes. *Science* **335**, 1376–1380.

[18] Artzy-Randrup, Y., Rorick, M. M., Day, K., Chen, D., Dobson, A. P. & Pascual, M., 2012 Population structuring of multi-copy, antigen-encoding genes in *Plasmodium falciparum*. *eLife* **1**, e00093.

[19] He, Q., Pilosof, S., Tiedje, K. E., Ruybal-Pesantez, S., Artzy-Randrup, Y., Baskerville, E. B., Day, K. P. & Pascual, M., 2018 Networks of genetic similarity reveal non-neutral processes shape strain structure in *Plasmodium falciparum*. *Nat. Comm.* **9**, 1817.

[20] Pilosof, S., He, Q., Tiedje, K. E., Ruybal-Pesantez, S., Day, K. P. & Pascual, M., 2019 Competition for hosts modulates vast antigenic diversity to generate persistent strain structure in *Plasmodium falciparum*. *PLoS Biology* **17**, e3000336.

[21] Buckee, C. O., Koelle, K., Mustard, M. J. & Gupta, S., 2004 The effects of host contact network structure on pathogen diversity and strain structure. *Proc. Natl. Acad. Sci.* **101**, 10839–10844.

[22] Buckee, C. O., Jolley, K. A., Recker, M., Penman, B., Kriz, P., Gupta, S. & Maiden, M. C. J., 2008 Role of selection in the emergence of lineages and the evolution of virulence in *Neisseria meningitidis*. *Proc. Natl. Acad. Sci.* **105**, 15082–15087.

[23] Buckee, C. O., Recker, M., Watkins, E. R. & Gupta, S., 2011 Role of stochastic processes in maintaining discrete strain structure in antigenically diverse pathogen populations. *Proc. Natl. Acad. Sci.* **108**, 15504–15509.

[24] Adams, B. & Sasaki, A., 2007 Cross-immunity, invasion and coexistence of pathogen strains in epidemiological models with one-dimensional antigenic space. *Math. Biosci.* **210**, 680–699.

[25] Adams, B. & Sasaki, A., 2009 Antigenic distance and cross-immunity, invasibility and coexistence of pathogen strains in an epidemiological model with discrete antigenic space. *Theor. Pop. Biol.* **76**, 157–167.

[26] Ferguson, N. M. & Galvani, A. P., 2003 The impact of antigenic variation on pathogen population structure, fitness and dynamics. In *Antigenic variation* (eds. A. Craig & A. Scherf), chapter 18, pp. 403–432. Academic Press.

[27] Dawes, J. H. P. & Gog, J. R., 2002 The onset of oscillatory dynamics in models of multiple disease strains. *J. Math. Biol.* **45**, 471–510.

[28] Calvez, V., Korobeinikov, A. & Maini, P. K., 2005 Cluster formation for multi-strain infections with cross-immunity. *J. Theor. Biol.* **233**, 75–83.

[29] Wikramaratna, P. S., Kucharski, A., Gupta, S., Andreasen, V., McLean, A. R. & Gog, J. R., 2015 Five challenges in modelling interacting strain dynamics. *Epidemics* **10**, 31–34.

[30] Kucharski, A. J., Andreasen, V. & Gog, J. R., 2016 Capturing the dynamics of pathogens with many strains. *J. Math. Biol.* **72**, 1–24.

[31] Slatkin, M., 2008 Linkage disequilibrium - understanding the evolutionary past and mapping the medical future. *Nature Reviews Genetics* **9**, 477–485.

[32] Gupta, S. & Anderson, R. M., 1999 Population structure of pathogens: the role of immune selection. *Parasitology Today* **15**, 497–501.

[33] Gupta, S. & Maiden, M., 2001 Exploring the evolutionary diversity in pathogen populations. *Trends in Microbiology* **9**, 181–185.

[34] Gog, J. R. & Swinton, J., 2002 A status-based approach to multiple strain dynamics. *J. Math. Biol.* **44**, 169–184.

[35] Lange, A. & Ferguson, N. M., 2009 Antigenic diversity, transmission mechanisms, and the evolution of pathogens. *PLoS Comp. Biol.* **5**, e1000536.

[36] Zinder, D., Bedford, T., Gupta, S. & Pascual, M., 2013 The roles of competition and mutation in shaping antigenic and genetic diversity in influenza. *PLoS Pathogens* **9**, e1003104.

[37] Hastings, I. M. & Wedgwood-Oppenheim, B., 1997 Sex, strains and virulence. *Parasitology Today* **13**, 375–383.

[38] Recker, M., Arinaminpathy, N. & Buckee, C. O., 2008 The effects of a partitioned *var* gene repertoire of *Plasmodium falciparum* on antigenic diversity and the acquisition of clinical immunity. *Malaria Journal* **7**, 18.

[39] Severins, M., Klinkenberg, D. & Heesterbeek, H., 2012 How selection forces dictate the variant surface antigens used by malaria parasites. *J. Roy. Soc. Interface* **9**, 246–260.

[40] Palmer, G. H., Bankhead, T. & Seifert, H. S., 2016 Antigenic variation in bacterial pathogens. *Microbiology Spectrum* **4**, VMBF–0005–2015.

[41] Minayev, P. & Ferguson, N., 2009 Improving the realism of deterministic multi-strain models: implications for modelling influenza A. *J. Roy. Soc. Interface* **6**, 509–518.

[42] Andreasen, V., Lin, J. & Levin, S. A., 1997 The dynamics of cocirculating influenza strains conferring partial cross-immunity. *J. Math. Biol.* **35**, 825–842.

[43] Skarlupka, A. L., Handel, A. & M., R. T., 2020 Influenza hemagglutinin antigenic distance measures capture trends in HAI, differences and infection outcomes, but are not suitable predictive tools. *Vaccine* **38**, 5822–5830.

[44] Anderson, C. S., McCall, P. R., Stern, H. A., Yang, H. & Topham, D. J., 2018 Antigenic cartography of H1N1 influenza viruses using sequence-based antigenic distance calculations. *BMC Bioinformatics* **19**.

[45] Ferguson, N., Anderson, R. & Gupta, S., 1999 The effect of antibody-dependent enhancement of the transmission dynamics and persistence of multiple-strain pathogens. *Proc. Natl. Acad. Sci.* **96**, 790–794.

[46] White, L. J., Cox, M. J. & Medley, G. F., 1998 Cross immunity and vaccination against multiple microparasite strains. *IMA Journal of Math. Applied in Med. Biology* **15**, 211–233.

[47] Adams, B. & Boots, M., 2006 Modelling the relationship between antibody-dependent enhancement and immunological distance with application to dengue. *J. Theor. Biol.* **242**, 337–346.

[48] Kawaguchi, I., Sasaski, A. & Boots, M., 2003 Why are dengue viruse serotypes so distantly related? Enhancement and limiting serotype similarity between dengue virus strains. *Proc. R. Soc. B* **270**, 2241–2247.

[49] Watkins, E. R., Grad, Y. H., Gupta, S. & Buckee, C. O., 2014 Contrasting within- and between-host immune selection shapes *Neisseria* opa repertoires. *Sci. Reports* **4**, 6554.

[50] Abu-Raddad, L. J. & Ferguson, N. M., 2004 The impact of cross-immunity, mutation and stochastic extinction on pathogen diversity. *Proc. Roy. Soc. B* **271**, 2431–2438.

[51] Wearing, H. J. & Rohani, P., 2006 Ecological and immunological determinants of dengue epidemics. *Proc. Natl. Acad. Sci.* **103**, 11802–11807.

[52] Ferguson, N. M., Galvani, A. P. & Bush, R. M., 2003 Ecological and immunological determinants of influenza evolution. *Nature* **422**, 428–433.

[53] Nagao, Y. & Koelle, K., 2008 Decreases in dengue transmission may act to increase the incidence of dengue hemorrhagic fever. *Proc. Natl. Acad. Sci.* **105**, 2238–2243.

[54] Minayev, P. & Ferguson, N., 2009 Incorporating demographic stochasticity into multi-strain epidemic models: application to influenza A. *J. R. Soc. Interface* **6**, 989–996.

[55] Gupta, S. & Galvani, A., 1999 The effects of host heterogeneity on pathogen population structure. *Proc. R. Soc. B* **354**, 711–719.

[56] Cobey, S. & Pascual, M., 2011 Consequences of host heterogeneity, epitope immunodominance, and immune breadth for strain competition. *J. Theor. Biol.* **270**, 80–87.

[57] Park, A. W., Daly, J. M., Lewis, N. S., Smith, D. J., Wood, J. L. N. & Grenfell, B. T., 2009 Quantifying the impact of immune escape on transmission dynamics of influenza. *Science* **326**, 726–728.

[58] Lewis, N. S., Daly, J. M., Russell, C. A., Horton, D. L., Skepner, E., Bryant, N. A., Burke, D. F., Rash, A. S., Wood, J. L., Chambers, T. M. *et al.*, 2011 Antigenic and genetic evolution of equine influenza a (h3n8) virus from 1968 to 2007. *J Virol* **85**, 12742–12749.

[59] Smith, D. J., Lapedes, A. S., de Jong, J. C., Bestebroer, T. M., Rimmelzwaan, G. F., Osterhaus, A. D. M. E. & Fouchier, R. A. M., 2004 Mapping the antigenic and genetic evolution of influenza virus. *Scinece* **305**, 371–376.

[60] Whitlock, F., Murcia, P. R. & Newton, J. R., 2022 A review on equine influenza from a human influenza perspective. *Viruses* **14**, 1312.

[61] Kucharski, A. J. & Gog, J. R., 2012 Age profile of immunity to influenza: effect of original antigenic sin. *Theor. Pop. Biol.* **81**, 102–112.

[62] Hamelin, F. M., Allen, L. J. S., Bokil, V. A., Gross, L. J., Hilker, F. M., Jeger, M. J., Manore, C. A., Power, A. G., Rua, M. A. & Cunniffe, N. J., 2019 Coinfections by noninteracting pathogens are not independent and require new tests of interaction. *PLoS Biol.* **17**, e30000551.

[63] Trzcinski, K., Thompson, C. M., Srivastava, A., Basset, A., Malley, R. & Lipsitch, M., 2008 Protection against nasopharyngeal colonization by *Streptococcus pneumoniae* is mediated by antigen-specific CD4+ T cells. *Infect. Immun.* **76**, 2678–2684.

[64] Li, Y., Gierahn, T., Thompson, C. M., Trzcinski, K., Ford, C. B., Croucher, N., Gouveia, P., Flechtner, J. B., Malley, R. & Lipsitch, M., 2012 Distinct effects on diversifying selection by two mechanisms of immunity against *Streptococcus pneumoniae*. *PLoS Pathogens* **8**, e1002989.

[65] Croucher, N. J., Campo, J. J., Le, T. Q., Liang, X., Bentley, S., Hanage, W. P. & Lipsitch, M., 2017 Diverse evolutionary patterns of pneumococcal antigens identified by pangenome-wide immunological screening. *Proc. Natl. Acad. Sci.* pp. E357–E366.

[66] Koelle, K., Pascual, M. & Yunus, M., 2006 Serotype cycles in cholera dynamics. *Proc. Roy. Soc. B* **273**, 2879–2886.

[67] Cummings, D. A. T., Schwartz, I. B., Billings, L. B., Shaw, L. B. & Burke, D. S., 2005 Dynamic effects of antibody-dependent enhancement on the fitness of viruses. *Proc. Natl. Acad. Sci.* **102**, 15259–15264.

[68] Restif, O. & Grenfell, B. T., 2007 Vaccination and the dynamics of immune evasion. *J. R. Soc. Interface* **4**, 143–153.

[69] He, Q., Pilosof, S., Tiedje, K. E., Day, K. P. & Pascual, M., 2021 Frequency-dependent competition between strains imparts persistence to perturbations in a model of *Plasmodium falciparum* malaria transmission. *Frontiers in Ecology and Evolution* **9**, 633263.

[70] Watkins, E. R., Penman, B. S., Lourenco, J., Buckee, C. O., Maiden, M. C. J. & Gupta, S., 2015 Vaccination drives changes in metabolic and virulence profiles of *Streptococcus pneumoniae*. *PLoS Pathogens* **11**, e1005034.

[71] Azarian, T., Martinez, P. P., Arnold, B. J., Qiu, X., Grant, L. R., Corander, J., Fraser, C., Croucher, N. J., Hammitt, L. L., Reid, R. *et al.*, 2020 Frequency-dependent selection can forecast evolution in *Streptococcus pneumonaie*. *PLoS Biology* **18**, e3000878.

[72] Lourenco, J., Daon, Y., Gori, A. & Obolski, U., 2021 Pneumococcal competition modulates antibiotic resistance in the pre-vaccination era: a modelling study. *Vaccines* **9**, 265.

[73] Obolski, U., Lourenco, J., Thompson, C., Thompson, R., Gori, A. & Gupta, S., 2018 Vaccination can drive an increase in frequencies of antibiotic resistance among nonvaccine serotypes of *Streptococcus pneumoniae*. *Proc. Natl. Acad. Sci.* **115**, 3102–3107.

[74] McLean, A. R., 1995 Vaccination, evolution and changes in the efficacy of vaccines: a theoretical framework. *Proc. Roy. Soc. B* **261**, 389–393.

[75] Pitzer, V. E., Patel, M. M., Lopman, B. A., Viboud, C., Parashar, U. D. & Grenfell, B. T., 2011 Modeling rotavirus strain dynamics in developed countries to understand the potential impact of vaccination on genotype distributions. *Proc. Natl. Acad. Sci.* **108**, 19353–19358.

[76] Colijn, C., Corander, J. & Croucher, N. J., 2020 Designing ecologically optimized pneumococcal vaccines using population genomics. *Nature Microbiology* **5**, 473–485.

[77] Andreasen, V. & Sasaki, A., 2006 Shaping the phylogenetic tree of influenza by cross-immunity. *Theor. Pop. Biol.* **70**, 164–173.

[78] Kryazhimskiy, S., Dieckmann, U., Levin, S. A. & Dushoff, J., 2007 On state-space reduction in multi-strain pathogen models, with an application to antigenic drift in influenza A. *PLoS Comp. Biol.* **3**, 1513–1525.

[79] Day, T. & Gandon, S., 2012 The evolutionary epidemiology of multilocus drug resistance. *Evolution* **66**, 1582–1597.

[80] McLeod, D. V. & Gandon, S., 2021 Understanding the evolution of multiple drug resistance in structured populations. *eLife* **10**, e65645.

[81] Thomson, G. & Baur, M. P., 1984 Third order linkage disequilibrium. *Tissue Antigens* **24**, 250–255.

[82] Robinson, W. P., Asmussen, M. A. & Thomson, G., 1991 Three-locus systems impose additional constraints on pairwise disequilibria. *Genetics* **129**, 925–930.

[83] Hurford, A., Cownden, D. & Day, T., 2010 Next-generation tools for evolutionary invasion analyses. *J. Roy. Soc. Interface* **7**, 561–571.

[84] van den Driessche, P. & Watmough, J., 2002 Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. *Math. Biosci.* **180**, 29–48.

[85] Diekmann, O., Heesterbeek, J. A. P. & Roberts, M. G., 2010 The construction of next-generation matrices for compartmental epidemic models. *J. Roy. Soc. Interface* **7**, 873–885.

[86] Kouyos, R. D., Fouchet, D. & Bonhoeffer, S., 2009 Recombination and drug resistance in HIV: population dynamics and stochasticity. *Epidemics* **1**, 58–69.

[87] Mani, R., St. Onge, R. P., Hartman, J. L., Giaever, G. & Roth, F. P., 2008 Defining genetic interaction. *Proc. Natl. Acad. Sci.* **105**, 3461–3466.

[88] Lewontin, R. C., 1964 The interaction of selection and linkage. i. general considerations; heterotic models. *Genetics* **49**, 49–67.

# Supplementary Information

**Contents**


**S1 Epidemiological model** **S1**
S1.1 Basic model without adaptive immunity . . . S1
S1.1.1 'Length' of immune history . . . S2
S1.2 Specification of adaptive immunity . . . S3
S1.3 Epidemiological dynamics and per-capita growth rates . . . S4

**S2 Biological connections between existing models** **S5**
S2.1 Immune protections . . . S5
S2.2 Acquisition of immunity . . . S6
S2.3 Co-infection . . . S6

**S3 Separation of evolutionary and epidemiological dynamics** **S7**
S3.1 The additive selection coefficients and epistasis in fitness . . . S9
S3.2 Relationship between $\mathcal{A}_\Delta(t)$ and $\mathcal{S}_\Delta(t)$ . . . S10
S3.3 Equations describing the evolutionary dynamics . . . S11
S3.4 Fast and slow timescales in the evolutionary dynamics . . . S12

**S4 Evolutionary dynamics on the slow timescale** **S14**
S4.1 Intuition behind the dynamics . . . S14
S4.1.1 Sign of $\sigma_\mathcal{A}$ . . . S15
S4.1.2 Conflicting evolutionary pressures induced by $\sigma_\mathcal{S} > 0$ . . . S17
S4.2 Epistasis and LD must share sign for local asymptotic stability when $\sigma_\mathcal{S} = 0$ S18
S4.3 Numerical simulations of key predictions . . . S18

**S5 Different types of immune protections** **S19**

**S6 Variation in immune protection between alleles and and/or strains** **S22**

**S7 The link between strain frequencies and the immunological landscape** **S25**

**S8 Variation in allelic diversity across loci** **S30**

**S9 More than two loci** **S31**
S9.1 Fast and slow timescales, and different types of strain structure . . . S35
S9.2 Shape of the function controlling cross-protection . . . S36
S9.3 $\sigma_\mathcal{S} > 0$ promotes oscillations and maintains diversity . . . S37


# S1 Epidemiological model

## S1.1 Basic model without adaptive immunity

In the absence of immunity, the host-pathogen model considered here is straightforward: the pathogen is transmitted to uninfected hosts by mass-action with rate constant $\beta$. If trans-

mission occurs, an infection is successfully established with probability $\sigma$, while infections are naturally cleared by the host at a per-capita rate $\mu$. Hosts enter the population at a fixed rate $b$ and are removed from the population due to natural- and virulence-related mortality (if infected) at per-capita rates $d$ and $\alpha$, respectively. Hosts cannot be simultaneously infected by multiple strains (no co-infection); we later discuss how co-infection affects the dynamics. At time $t$, let $S(t)$ denote the density of uninfected hosts and and $I(t)$ denote the total density of infected hosts. Then, the epidemiological dynamics are

$$\begin{aligned} \frac{\mathrm{d}S}{\mathrm{d}t} &= b - dS - \beta\sigma SI + \mu I \\ \frac{\mathrm{d}I}{\mathrm{d}t} &= (\beta\sigma S - d - \mu - \alpha)I. \end{aligned} \tag{S1}$$

The equilibrium density of uninfected hosts in the absence of the pathogen is $b/d$. Therefore, the basic reproduction number of the pathogen, $R_0$, is

$$R_0 \equiv \frac{\beta\sigma}{\mu+\alpha+d}\frac{b}{d}, \tag{S2}$$

and the endemic equilibrium is

$$\left(\bar{S}, \bar{I}\right) = \left(\frac{1}{R_0}\frac{b}{d}, \frac{b}{\alpha+d}\frac{R_0-1}{R_0}\right). \tag{S3}$$

The endemic equilibrium is globally asymptotically stable whenever $R_0 > 1$. We denote the ratio between the expected duration of infection, $1/(\mu+\alpha+d)$, and the expected lifespan of the host, $1/d$, as

$$\theta \equiv \frac{d}{\mu+\alpha+d}. \tag{S4}$$

For small $\theta$, infections are acute, whereas as $\theta \to 1$, infections become increasingly chronic.

#### S1.1.1 'Length' of immune history

Both $R_0$ and $\theta$ drive the dynamics of the pathogen and affect the 'length' of the immune history of an individual host. In particular, they determine (i) the expected number of infections a host experiences through its lifespan, and (ii) the number of prior infections of a randomly selected uninfected host.

To calculate the expected number of infections a host experiences through its lifespan, suppose we are at the endemic equilibrium, that $b = d$, and, for simplicity, $\alpha = 0$. The probability, $P_n$, that a host is infected $n > 0$ times during its life, is the product of the probability that a host is infected, and clears, $n-1$ infections

$$\left(\frac{\beta\sigma\bar{I}}{d+\beta\sigma\bar{I}}\frac{\mu}{\mu+d}\right)^{n-1} = \left(\frac{R_0\bar{I}(1-\theta)}{\theta+R_0\bar{I}}\right)^{n-1},$$

and the probability a host is infected an $n$-th time, after which the host dies before becoming re-infected:

$$\frac{\beta\sigma\bar{I}}{d+\beta\sigma\bar{I}}\left(\frac{d}{\mu+d} + \frac{\mu}{\mu+d}\frac{d}{\beta\sigma\bar{I}+d}\right) = \frac{R_0\bar{I}\theta}{\theta+R_0\bar{I}}\left(1+\frac{1-\theta}{\theta+R_0\bar{I}}\right).$$

The expected number of infections during the lifespan of an average host, which we denote $T$, is

$$T \equiv \sum_{n=0}^{\infty} nP_n = \frac{R_0 - 1}{R_0\theta} \tag{S5}$$

where to compute the sum we have used the relation $\sum_{n=0}^{\infty} nx^{n-1} = 1/(1-x)^2$ for $0 < x < 1$.

The expected number of prior infections of a randomly selected uninfected host also depends upon $R_0$ and $\theta$. To calculate this, let $\bar{S}_i$ denote the equilibrium density of uninfected hosts with $i$ prior infections, and $\bar{I}_i$ the density of infected hosts with a history of $i$ infections (including the current infection). Assuming $\alpha = 0$ and $b = d$ as before, at the endemic equilibrium

$$\bar{S}_0 = \frac{d}{d + \beta\sigma\bar{I}} = \frac{\theta}{\theta + R_0\bar{I}}$$

and

$$\bar{S}_i = \frac{\mu\bar{I}_i}{d + \beta\bar{I}} = \frac{(1-\theta)\bar{I}_i}{\theta + R_0\bar{I}} \quad \text{and} \quad \bar{I}_i = \frac{\beta\sigma I}{\mu + d} = R_0\bar{I}\bar{S}_{i-1}.$$

Therefore,

$$\bar{S}_n = \left(\frac{R_0\bar{I}(1-\theta)}{\theta + R_0\bar{I}}\right)^n \bar{S}_0.$$

Hence, the expected number of prior infections of a randomly selected uninfected host, which we denote $M$, is

$$M \equiv \sum_{n=0}^{\infty} n\frac{\bar{S}_n}{\bar{S}} = \frac{(R_0 - 1)(1-\theta)}{R_0\theta}. \tag{S6}$$

Notice that both $M$ and $T$ increase with $R_0$ and decrease with $\theta$ (i.e., infections become increasingly chronic). Although these calculations do not account for immune protections, immunity will generally act to reduce $M$ and $T$ , and not change their qualitative dependence upon $R_0$ and $\theta$. Thus, the 'length' of immune history, which is discussed in the main text, is expected to depend on $M$ and $T$, and therefore to increase with $R_0$ and decrease with $\theta$.

## S1.2 Specification of adaptive immunity

Adaptive immunity is built into the basic model of equations (S1) as follows. We consider a pathogen whose genotype consists of some number of multiallelic loci. The host can develop an adaptive immune response to the alleles (i.e., antigens) at each of these loci. This immune response can potentially reduce:

(i) The susceptibility to infection; we let $\sigma_x^g$ denote the probability of infection of a host with immune history $x$ by strain (genotype) $g$.

(ii) The transmissibility of an infection; we let $\beta_x^g$ denote the transmissibility of an infection by strain $g$ from a host with immune history $x$.

(iii) The duration of infection; we let $\mu_x^g$ denote the per-capita rate at which a host with immune history $x$ clears an infection by strain $g$.

(iv) The virulence-related mortality; we let $\alpha_x^g$ denote the per-capita rate at which a host with immune history $x$ dies due to an infection by strain $g$.

For each of these protections, we distinguish between immunity to antigens and immunity to previously-infecting strains (i.e., homologous immunity). In particular, for a host with immunity to $n = 0, 1, 2, ...$ antigens of strain $g$, the immune protections are

$$(\sigma_x^g, \beta_x^g, \mu_x^g, \alpha_x^g) = (\sigma_n, \beta_n, \mu_n, \alpha_n),$$

whereas for a host with homologous immunity to strain $g$, the immune protections are

$$(\sigma_x^g, \beta_x^g, \mu_x^g, \alpha_x^g) = (\sigma_h, \beta_h, \mu_h, \alpha_h).$$

We assume that homologous immunity is at least as strong as full antigenic immunity, and that the host immune response is always the most protective available to the host, based upon its immune history. For example, for a pathogen genome with two diallelic loci, $\{a, A\}$ and $\{b, B\}$, a host with immune history $x' = \{ab, Ab\}$ has immune protections

$$(c_{x'}^{ab},\, c_{x'}^{Ab},\, c_{x'}^{aB},\, c_{x'}^{AB}) = (c_h,\, c_h,\, c_1,\, c_1), \quad c \in \{\sigma, \beta, \mu, \alpha\}.$$

## S1.3 Epidemiological dynamics and per-capita growth rates

At time $t$, let $S_x(t)$ denote the density of uninfected hosts with prior infection by strains belonging to the set $x$ (so $x$ is the immune history), and let $I_x^g(t)$ denote the density of strain $g$ infections in hosts with immune history $x$. The resulting epidemiological dynamics are

$$\begin{aligned} \frac{\mathrm{d}S_x}{\mathrm{d}t} &= b\mathbf{1}_0(x) - dS_x - \sum_g \sigma_x^g \sum_z \beta_z^g I_z^g S_x + \Gamma_x \\ \frac{\mathrm{d}I_x^g}{\mathrm{d}t} &= \sum_z \beta_z^g I_z^g \sigma_x^g S_x - (d + \mu_x^g + \alpha_x^g) I_x^g + \Delta_{g,x} \end{aligned} \tag{S7}$$

where $\mathbf{1}_0(x) = 1$ if $x = 0$ and 0 otherwise, $\Delta_{g,x}$ is the net change due to recombination of strain $g$ infections in hosts with immune history $x$ (for now, left arbitrary), and $\Gamma_x$ controls the rate at which hosts gain (or lose) the immune history $x$. If immunity is life-long and is updated when infections are cleared,

$$\Gamma_x = \sum_g \sum_{z:g\cup z=x} \mu_z^g I_z^g. \tag{S8}$$

Next, let $r_g$ denote the per-capita growth rate of strain $g$; this is the growth rate of strain $g$ averaged across the different immune environments $x$. It follows that

$$r_g = \sum_x \frac{\mathrm{d}I_x^g/\mathrm{d}t}{I^g} = \sum_x \sum_z \beta_z^g \frac{I_z^g}{I^g} \sigma_x^g S_x - d - \sum_x (\mu_x^g + \alpha_x^g) \frac{I_x^g}{I^g}, \tag{S9}$$

where $I^g = \sum_x I_x^g$ is the total density of strain $g$. In Box 1 of the main text we present a version of this model where we have restricted immune protections to reducing $\sigma_x^g$. That is, immunity only reduces host susceptibility to infection.

# S2 Biological connections between existing models

Next, we discuss the relationship between model (S7) and existing models. Multistrain models typically differ from one another with respect to three biological properties:

(i) the immune protections, and how these are affected by antigenic distance;

(ii) the specification of how hosts acquire immunity, that is, the specification of $\Gamma_x$; and

(iii) whether co-infection is permitted or not.

Here, we discuss each of these assumptions in more detail, and in Table S2 we summarize how the model (S7) relates to existing models. It is important to note that the combination of assumptions employed by previous models is typically made because they reduce the number of equations. Doing so makes the problem more tractable for mathematical analysis and simulation. As our purpose here is to understand and, where possible, generalize the biological predictions of these models, we will simply state the assumptions made by each model without showing the mathematical benefits of these assumptions. We refer interested readers to Kucharski & Gog [30], who provide a detailed treatment of the mathematics behind some of the commonly used models.

## S2.1 Immune protections

Multistrain models vary in what type of protection immunity provides and how prior infections combine to determine the strength of immunity.

First, most models assume that immunity reduces either transmissibility or susceptibility; this choice is made typically for model-specific mathematical reasons. For example, 'status'-based models often assume reduced susceptibility [13, 27, 34], whereas 'history'-based models often assume reduced transmissibility [6, 9, 12, 16] (but not always, [14, 42]). More recently, simulation models have assumed that immune protections reduce the duration of carriage, as this is more biologically realistic for malaria [18–20]. Finally, some models of antibody-dependent enhancement have allowed immunity to reduce virulence-related mortality [47, 48]. Although immune protections reducing virulence-related mortality can have a qualitatively different effect [e.g., 47], it is generally believed that there is no qualitative difference between protections reducing susceptibility, transmissibility, or clearance [15, 26]. At least one consequence of the differences in immune protections is that if immunity reduces susceptibility, hosts are less likely to acquire immunity to different strains, provided that immunity is only acquired after successful infection [15].

Second, models differ in how multiple prior infections combine to determine the strength of immunity. We will illustrate these differences using the example immune history $x'' = \{Ab, aB\}$. Thus, we focus on a pathogen genome consisting of two diallelic loci. Moreover, we focus on the case in which immune protections reduce susceptibility to infection; the other cases follow similarly. The most common approach is to assume that immunity against a particular strain is determined by the *single* most antigenically similar previously infecting strain [6, 9, 12, 16, 25]. For the immune history $x''$, we obtain

$$(\sigma_{x''}^{ab}, \sigma_{x''}^{Ab}, \sigma_{x''}^{aB}, \sigma_{x''}^{AB}) = (\sigma_1, \sigma_h, \sigma_h, \sigma_1).$$

A less common model assumption is that immunity is the *product* of the individual cross-protections against previously infecting strains [14, 24]. For the immune history $x''$, this yields

$$(\sigma_{x''}^{ab}, \sigma_{x''}^{Ab}, \sigma_{x''}^{aB}, \sigma_{x''}^{AB}) = \left((\sigma_1)^2, \sigma_h, \sigma_h, (\sigma_1)^2\right).$$

Finally, immunity to each antigen can be treated independently. Here, the strength of immunity is *antigen-specific* and depends upon the total number of antigens against which the host has protection [18–20]. For the immune history $x''$, it follows that

$$(\sigma_{x''}^{ab}, \sigma_{x''}^{Ab}, \sigma_{x''}^{aB}, \sigma_{x''}^{AB}) = (\sigma_2, \sigma_h, \sigma_h, \sigma_2).$$

Note that although some models restrict immune memory to only the most recent infections [67, 77], we do not consider this further here, as it can be considered a form of waning immunity.

## S2.2 Acquisition of immunity

The model of (S7) assumes that after hosts have naturally cleared an infection, they have homologous immunity to the infecting strain, as well as some degree of protection against antigenically similar strains. Such a framework has been referred to as a 'history-based' model [13, 34], and is a commonly taken modeling approach [6, 12, 14, 42]. An alternative to history-based models are so-called 'status-based' models [13, 27, 34]. Status-based models assume that each host is either completely protected or completely susceptible to a given strain (so immunity reduces susceptibility); there is no 'partial' protection at the level of the individual host. Thus, following clearance of an infection (or at time of infection, depending upon when immunity is acquired), each host becomes either fully-protected or fully-sensitive to each strain. The probability of acquiring protection against a particular strain then depends upon the antigenic similarity between that strain and the other strains against which the host has protection [13, 27, 34].

To convert model (S7) into a status-based model, we first restrict immune protections to reducing susceptibility, such that

$$\sigma_x^g = \begin{cases} 0 & \text{if } g \in x \\ 1 & \text{otherwise} \end{cases}, \quad \text{while} \quad (\beta_x^g, \mu_x^g, \alpha_x^g) = (\beta, \mu, \alpha).$$

Then, we suppose that upon clearing an infection by strain $g$, a host with immune history $z$ becomes a host with immune history $x$ with probability $p_{z\to x}^g$ (so $\sum_x p_{z\to x}^g = 1$), and thus

$$\Gamma_x = \sum_g \sum_z \mu_z^g p_{z\to x}^g I_z^g. \tag{S10}$$

Under these assumptions, model (S7) is a status-based model.

## S2.3 Co-infection

The model of (S7) assumes that transmission only occurs to uninfected hosts, and thus hosts cannot be simultaneously infected by different strains (i.e., there is no co-infection). This assumption was made primarily for modelling clarity. It is biologically reasonable if,

for example, the pathogen is circulating at a low prevalence, infections are acute, and/or infection induces short-term strain-transcending immunity. Yet, these conditions need not apply generally. Hence, co-infection is a common attribute of multistrain models. Most analytically tractable models that include co-infection make all of the following assumptions [6, 12, 15, 41, 54]:

(i) Hosts can be simultaneously infected by any number of *different* strains and these strains do not interact. Thus, transmission and clearance rates are unaffected by the presence, or number, of co-infecting strains. Here, the immune environment a strain experiences is set at the time of infection and is unaffected by subsequent infections.

(ii) There is no virulence mortality ($\alpha_x^g = 0$) to ensure independence of co-infecting strains [61, 62].

(iii) Homologous immunity is sterilizing and life-long ($\sigma_h = 0$).

Incorporating these assumptions into our model alters the specification of $\Gamma_x$ and the biological interpretation of $S_x(t)$ in equation (S7). In particular, $S_x(t)$ now denotes the density of hosts with immune history $x$, irrespective of their current infection status, and

$$\Gamma_x = \sum_g \sum_{z_1 : z_1 \cup g = x} \sigma_{z_1}^g S_{z_1} \sum_{z_2} \beta_{z_2}^g I_{z_2}^g. \tag{S11}$$

Importantly, because homologous immunity is sterilizing (i.e., $\sigma_h = 0$), and acquired at the time of infection, a host that is infected by strain $g'$ cannot be reinfected by strain $g'$. As a consequence, and because co-infections are assumed to be non-interacting, it is not necessary to keep track of how many (and which) strains currently infect a given host.

Notably, treating co-infecting strains as non-interacting is biologically unsatisfying, given that a host simultaneously infected with $n$ strains is $n$-fold as infectious as a host infected with one strain. Specifically, genetic diversity is unreasonably easily maintained under this assumption. For example, Buckee et al [22] studied the evolution of hypervirulent lineages of *Neisseria meningitidis* using a model in which each strain is defined by its sequence type and level of virulence (which causes host mortality). Since hypervirulence shortens the duration of infection, it should be eliminated from the population by selection. But as hosts can be simultaneously infected by any number of genetically distinct strains, each level of virulence creates 'new' genetically distinct strains that can therefore co-infect the same host (even if they share the same sequence type as existing infecting strains), thus maintaining variation in virulence. Incorporating virulence in models of co-infection is not trivial and requires co-infecting strains to interact [61, 62].

# S3 Separation of evolutionary and epidemiological dynamics

The number of equations associated with model (S7) makes the full model challenging to analyze and to interpret. Thus, a variety of clever mathematical approximations have been

| Model | Immune protections | Time immunity updated | History or status? | Effect of multiple prior infections | Coinfection | Sign of $\epsilon_{\mathcal{A}}$ | Is $\epsilon_{\mathcal{S}} > 0$? |
|---|---|---|---|---|---|---|---|
| [6, 12] | $\beta_{n+1} = \beta_0(1-\gamma)$<br>$\sigma_h = 0$ | infection | history | single | yes | positive | yes |
| [9, 16] | $\beta_1 = \beta_0(1-\gamma)$<br>$\beta_{n+2} = \beta_0(1-\delta)$<br>$\sigma_h = 0$ | infection | history | single | yes | positive | yes |
| [13, 27, 34] | $\sigma_{n+1} = 0$<br>$\sigma_h = 0$ | infection | status | single | yes | positive | no |
| [18–20] | $\mu_n = \mu_0(n+1)$ | clearance | history | antigen-specific | yes | positive | no |
| [25] | $\sigma_n = \left(\frac{N-n}{N}\right)^{\gamma}$<br>$\sigma_h = 0$ | infection | history | single | yes | positive | yes |
| [12, 21] | $\sigma_n = \left(1 - \left(\frac{n}{N}\right)^{\frac{1}{\gamma}}\right)^{\gamma}$<br>$\sigma_h = 0$ | infection | history | single | yes | positive | yes |
| [14] | $\sigma_0 = 1$<br>$\sigma_x^g = \prod_{g_0 \in x} \sigma_{g_0}^g$<br>$\sigma_h = 0$ | infection | history | product | yes | positive | yes |
| [42] | $\sigma_0 = 1$<br>$\sigma_1 = \sigma_2 = \sigma$<br>$\sigma_h = 0$ | infection | history | single | no | positive | non-negative |
| Model (S1) | variable | either | either | variable | no | variable | variable |

Table S2: **Summary of key models and their relationship to the framework used here.** Here, $\gamma$, $\sigma$, and $\delta$ are positive parameters and $N$ is the total number of loci. We define $\epsilon_{\mathcal{A}} \equiv f_0 - 2f_1 + f_2$ where $f_i \equiv \sigma_i\beta_i/(\mu_i + \alpha_i + d)$, and $\epsilon_{\mathcal{S}} \equiv f_2 - f_h$. Note that both $\epsilon_{\mathcal{A}}$ and $\epsilon_{\mathcal{S}}$ are calculated assuming there are only two diallelic loci. In this case if $\sigma_h = 0$, each of the different ways in which immunity to multiple prior strains combines (single, antigen-specific, and product) is captured by model (S7). There also exist various models whose specification of cross-immunity assumes a one-dimensional antigenic space [e.g., 24, 42], which we do not consider here (the model of [42] considered in the table is the version with four strains).

employed to simplify the dynamics of multistrain models [e.g., 6, 13, 14, 42, 61, 78]. Although these techniques are effective at reducing the (mathematical) dimensionality of the problem, they struggle to provide an evolutionary context for the predictions; for example, they do not provide biological intuition why oscillations are possible in some models but not others [see discussion in 27]. Thus, many important results lack a clear biological interpretation. A promising way to provide greater clarity on the evolutionary process is to separate the variables describing the epidemiological dynamics from those describing the evolutionary dynamics [79, 80]. Doing so allows for the identification and interpretation of the evolutionary forces at play. The evolutionary variables are the allele (i.e., antigen) frequencies and the linkage disequilibrium (LD; i.e., nonrandom associations between alleles) [31]. The epidemio-

logical variables include the total density of infections, $I(t)$, and the population distribution of immune histories across the uninfected hosts, $S_x(t)$. Here, we focus on extracting the evolutionary variables for a pathogen genome consisting of two diallelic loci; in Sections S8 and S9 we extend our model and analysis to the study of greater antigenic variation.

## S3.1 The additive selection coefficients and epistasis in fitness

For two diallelic loci, there are four possible pathogen genotypes (strains),

$$g \in \{ab, Ab, aB, AB\}.$$

Without loss of generality, we consider strain $ab$ to be the reference genotype. Then, the additive selection coefficients of alleles $A$ and $B$ are

$$w_A \equiv r_{Ab} - r_{ab} \quad \text{and} \quad w_B \equiv r_{aB} - r_{ab}.$$

The selection coefficients indicate whether allele $A$ (resp. $B$) is favoured, i.e., $w_A > 0$ (resp. $w_B > 0$), or disfavoured, i.e., $w_A < 0$ (resp. $w_B < 0$), over allele $a$, when $A$ (resp. $B$) occurs on the genetic background $b$ (resp. $a$). We also account for whether the fitness of an allele depends upon its genetic background, i.e., whether there is *epistasis in fitness*. Epistasis in fitness is defined as

$$w_{AB} \equiv r_{AB} + r_{ab} - r_{Ab} - r_{aB},$$

and indicates whether the combination of alleles $A$ and $B$ is favoured ($w_{AB} > 0$) or disfavoured ($w_{AB} < 0$) over what is expected based upon the additive selection coefficients. Thus, the additive selection coefficients can be thought of as controlling selection on individual alleles, whereas epistasis can be thought of as controlling selection on particular combinations of alleles (in the case of two loci, these combinations are equivalent to genotypes or strains).

Although we can write down the additive selection coefficients and epistasis for the general model, here we assume that immune protections only affect susceptibility. That is because if immune protections only affect susceptibility, the per-capita growth rates of equation (S9) can be written in a particularly simple and illustrative form. In Section S5, we consider other immune protections using an equilibrium analysis.

To write down $w_A$, $w_B$, and $w_{AB}$, we define

$$\sigma_{\mathcal{A}} \equiv \sigma_0 - 2\sigma_1 + \sigma_2 \quad \text{and} \quad \sigma_{\mathcal{S}} \equiv \sigma_2 - \sigma_h.$$

Conceptually, if we consider the function mapping antigenic immunity to the strength of immune protections, $h : n \mapsto \sigma_n$, then $\sigma_{\mathcal{A}}$ is the central difference approximation of $\mathrm{d}^2h/\mathrm{d}n^2$, evaluated at $n = 1$. Consequently, $\sigma_{\mathcal{A}}$ can be thought of as controlling whether the function relating antigen-specific cross-protection to the number of antigens recognized by the immune system has an accelerating or saturating shape. The function has an accelerating shape if $\sigma_{\mathcal{A}} < 0$, and a saturating shape if $\sigma_{\mathcal{A}} > 0$ (Fig. 2). On the other hand, $\sigma_{\mathcal{S}}$ measures how much more protective homologous immunity is than full antigenic immunity. If $\sigma_{\mathcal{S}} = 0$, full antigenic immunity is equally strong as homologous immunity.

Using these composite parameters, the per-capita growth rate can be written as

$$r_{ij} = \beta(\sigma_0 \mathcal{S} - (\sigma_0 - \sigma_1)(\mathcal{A}_{\{i\}} + \mathcal{A}_{\{j\}}) + \sigma_{\mathcal{A}} \mathcal{A}_{\{i,j\}} - \sigma_{\mathcal{S}} \mathcal{S}_{ij}) - d - \mu - \alpha, \tag{S12}$$

where $\mathcal{A}_{\{i\}}(t)$ and $\mathcal{A}_{\{i,j\}}(t)$ are the densities of uninfected hosts that have immunity to *at least* antigen $i$ or antigens $i$ and $j$, $\mathcal{S}(t) = \sum_x S_x(t)$ is the total density of uninfected hosts, and $\mathcal{S}_{ij}(t)$ is the density of uninfected hosts that have homologous immunity (prior infection) to *at least* strain $ij$.

Then, the additive selection coefficients can be written as

$$w_A = \beta(\sigma_0 - \sigma_1)(\mathcal{A}_{\{a\}} - \mathcal{A}_{\{A\}}) + \beta\sigma_{\mathcal{A}}(\mathcal{A}_{\{A,b\}} - \mathcal{A}_{\{a,b\}}) - \beta\sigma_{\mathcal{S}}(\mathcal{S}_{Ab} - \mathcal{S}_{ab}) \tag{S13}$$

$$w_B = \beta(\sigma_0 - \sigma_1)(\mathcal{A}_{\{b\}} - \mathcal{A}_{\{B\}}) + \beta\sigma_{\mathcal{A}}(\mathcal{A}_{\{a,B\}} - \mathcal{A}_{\{a,b\}}) - \beta\sigma_{\mathcal{S}}(\mathcal{S}_{aB} - \mathcal{S}_{ab}), \tag{S14}$$

and epistasis in fitness is

$$w_{AB} = \beta\sigma_{\mathcal{A}}(\mathcal{A}_{\{A,B\}} + \mathcal{A}_{\{a,b\}} - \mathcal{A}_{\{A,b\}} - \mathcal{A}_{\{a,B\}}) - \beta\sigma_{\mathcal{S}}(\mathcal{S}_{AB} + \mathcal{S}_{ab} - \mathcal{S}_{Ab} - \mathcal{S}_{aB}). \tag{S15}$$

## S3.2 Relationship between $\mathcal{A}_\Delta(t)$ and $\mathcal{S}_\Delta(t)$

Define

$$\mathcal{A}_\Delta(t) = \mathcal{A}_{\{A,B\}}(t) + \mathcal{A}_{\{a,b\}}(t) - \mathcal{A}_{\{A,b\}}(t) - \mathcal{A}_{\{a,B\}}(t)$$
$$\text{and} \quad \mathcal{S}_\Delta(t) = \mathcal{S}_{AB}(t) + \mathcal{S}_{ab}(t) - \mathcal{S}_{Ab}(t) - \mathcal{S}_{aB}(t). \tag{S16}$$

The sign of $\mathcal{A}_\Delta(t)$ indicates whether more uninfected hosts have full antigenic immunity to strains belonging to the set $\{ab, AB\}$ ($\mathcal{A}_\Delta(t) > 0$) or to the set $\{Ab, aB\}$ ($\mathcal{A}_\Delta(t) < 0$). The sign of $\mathcal{S}_\Delta(t)$ indicates whether more uninfected hosts have homologous immunity to strains belonging to the set $\{ab, AB\}$ ($\mathcal{S}_\Delta(t) > 0$) or to the set $\{Ab, aB\}$ ($\mathcal{S}_\Delta(t) < 0$). Using these definitions, we can rewrite epistasis in fitness (equation (S15)) as

$$w_{AB} = \beta\sigma_{\mathcal{A}}\mathcal{A}_\Delta(t) - \beta\sigma_{\mathcal{S}}\mathcal{S}_\Delta(t). \tag{S17}$$

Next, notice that if we expand the sums in $\mathcal{A}_\Delta(t)$, we obtain after cancellation

$$\mathcal{A}_\Delta(t) = S_{AB}(t) + S_{ab}(t) - S_{Ab}(t) - S_{aB}(t). \tag{S18}$$

Thus, only hosts with immunity to the antigens defining a single strain contribute to $\mathcal{A}_\Delta(t)$. Unlike $\mathcal{A}_\Delta(t)$, however, $\mathcal{S}_\Delta(t)$ depends upon hosts with immunity to more than one strain, since, after cancellation,

$$\mathcal{S}_\Delta(t) = \mathcal{A}_\Delta(t) + S_{ab,AB}(t) - S_{aB,Ab}(t) + \mathcal{S}_{ab,AB}(t) - \mathcal{S}_{aB,Ab}(t) \tag{S19}$$

where $S_{ij,k\ell}(t)$ denotes the density of uninfected hosts that have homologous immunity to *only* strains $ij$ and $k\ell$, and $\mathcal{S}_{ij,k\ell}(t)$ denotes the density of uninfected hosts that have homologous immunity to *at least* strains $ij$ and $k\ell$. A graphical representation of this relationship is provided in Figure S1. As discussed in the main text, in order for $\mathcal{S}_\Delta(t)$ and $\mathcal{A}_\Delta(t)$ to diverge, hosts must be infected multiple times during their lifespan (and retain memory of these infections).

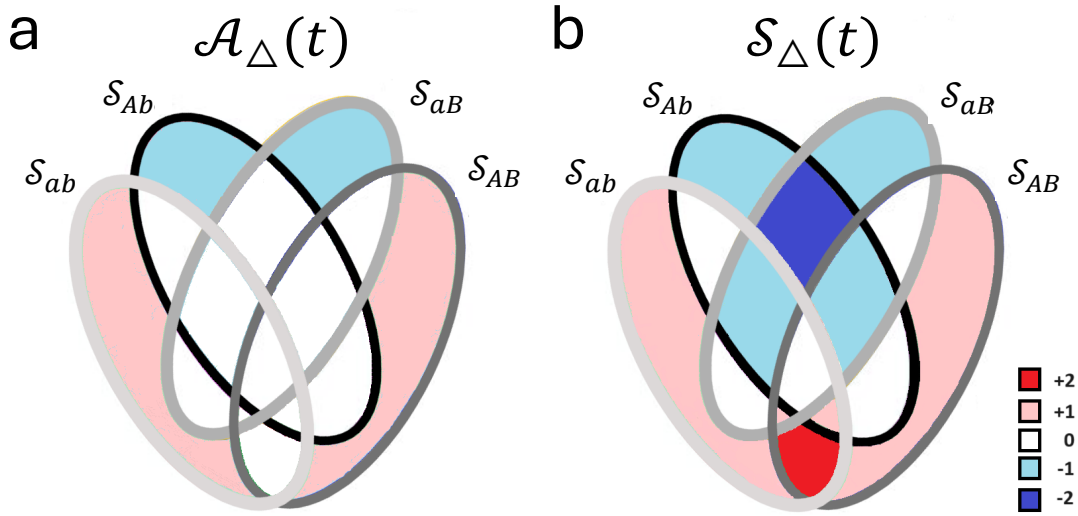


Figure S1: **The dynamic immunological landscape**. Panel **a** is a Venn diagram of $\mathcal{A}_\Delta(t)$, that is, whether more uninfected hosts have full antigenic immunity to strains belonging to the set $\{ab, AB\}$ (i.e., $\mathcal{A}_\Delta(t) > 0$) or to the set $\{aB, Ab\}$ (i.e., $\mathcal{A}_\Delta(t) < 0$). Panel **b** is a Venn diagram of $\mathcal{S}_\Delta(t)$, that is, whether more uninfected hosts have homologous immunity to strains belonging to the set $\{ab, AB\}$ ($\mathcal{S}_\Delta(t) > 0$) or to the set $\{aB, Ab\}$ ($\mathcal{S}_\Delta(t) < 0$). Each ellipse corresponds to a different density of hosts with homologous immunity to *at least* strain $ij$, $\mathcal{S}_{ij}(t)$. Then, $\mathcal{A}_\Delta(t)$ and $\mathcal{S}_\Delta(t)$ are computed as the weighted sum of the coloured regions, where the colors indicate the weights. Whereas $\mathcal{A}_\Delta(t)$ only includes uninfected hosts with homologous immunity to a single strain, $\mathcal{S}_\Delta(t)$ includes uninfected hosts with homologous immunity to multiple strains.

## S3.3 Equations describing the evolutionary dynamics

To write out the evolutionary dynamics, we first define the evolutionary variables. The allele frequencies are

$$p_A(t) = \frac{I^{Ab}(t) + I^{AB}(t)}{I(t)} \quad \text{and} \quad p_B(t) = \frac{I^{aB}(t) + I^{AB}(t)}{I(t)}, \tag{S20}$$

and the linkage disequilibrium between alleles $A$ and $B$ is

$$D(t) = \frac{I^{AB}(t)}{I(t)} - p_A(t)p_B(t) = \frac{I^{AB}(t)}{I(t)}\frac{I^{ab}(t)}{I(t)} - \frac{I^{Ab}(t)}{I(t)}\frac{I^{aB}(t)}{I(t)}. \tag{S21}$$

Then, the evolutionary dynamics can be written as

$$\frac{\mathrm{d}p_A}{\mathrm{d}t} = w_A p_A(1-p_A) + w_B D + w_{AB}(1-p_A)(p_A p_B + D) + \Delta_{p_B} \tag{S22a}$$

$$\frac{\mathrm{d}p_B}{\mathrm{d}t} = w_B p_B(1-p_B) + w_A D + w_{AB}(1-p_B)(p_A p_B + D) + \Delta_{p_B} \tag{S22b}$$

$$\frac{\mathrm{d}D}{\mathrm{d}t} = \sum_{k=A,B} w_k(1-2p_k)D + w_{AB}(p_A p_B + D)((1-p_A)(1-p_B) - D) + \Delta_D, \tag{S22c}$$

where $\Delta_q$ indicates the effect of recombination on the variable $q$. Note that the general form of these equations follows standard population genetics and is not particular to the epidemiological system we consider here. Epidemiology enters this system through the additive selection coefficients, recombination, and epistasis in fitness, which can vary through time due to the epidemiological dynamics. By writing out the evolutionary dynamics as we have done in (S22), we can identify the evolutionary forces at play and study how they impact the evolutionary process. We can then link each of these forces to the epidemiological model of interest.

## S3.4 Fast and slow timescales in the evolutionary dynamics

Unbiased recombination does not affect antigen frequencies, $\Delta_{p_B} = \Delta_{p_A} = 0$ and acts only to remove LD from the population. That is, $\Delta_D \approx -\rho D(t)$, where $0 < \rho \ll 1$ is a positive (but small) constant controlling the rate of recombination. Because we assume that there are no intrinsic biases favouring one antigen over the other at a given locus, and because immunity induces strong negative frequency-dependent selection (NFDS) at each locus, the antigen frequencies will quickly reach a quasi-equilibrium state at which each antigen is present at equal frequency within the population, that is, $p_k(t) \to 1/2$ for $k \in \{a, A, b, B\}$. Once this quasi-equilibrium state is reached, since $p_A(t) = p_B(t) = p_b(t) = p_a(t)$, we also obtain $p_{Ab}(t) = p_{aB}(t)$ and $p_{AB}(t) = p_{ab}(t)$. That is, the inherent symmetry within the system drives the population to self-organize into two sets of strains, $\{ab, AB\}$ and $\{Ab, aB\}$. At this point, the members of each set behave identically dynamically. Consequently, strains within each set have the same per-capita growth rates, that is, $r_{ab} = r_{AB}$ and $r_{Ab} = r_{aB}$.

Thus the evolutionary dynamics occur on two timescales. On the fast timescale, strong NFDS operating through the additive selection coefficients will push the antigen and strain frequencies to a quasi-equilibrium state, $p_A(t) \approx 1/2$, $p_B(t) \approx 1/2$, $p_{AB}(t) \approx p_{ab}(t)$, and $p_{Ab}(t) \approx p_{aB}(t)$. At this quasi-equilibrium state, since $r_{AB} = r_{ab}$ and $r_{Ab} = r_{aB}$, we have

$$w_A = r_{Ab} - r_{ab} = r_{aB} - r_{ab} = w_B, \tag{S23}$$

and moreover,

$$w_{AB} = r_{AB} + r_{ab} - r_{Ab} - r_{aB} = 2(r_{ab} - r_{Ab}) = -2w_A. \tag{S24}$$

Therefore at quasi-equilibrium we have the relations $w_A \approx w_B \approx -w_{AB}/2$, and so on the slow timescale, the system of equations (S22) reduces to the single equation

$$\frac{\mathrm{d}D}{\mathrm{d}t} \approx w_{AB}\left(\frac{1}{4} + D\right)\left(\frac{1}{4} - D\right) - \rho D. \tag{S25}$$

Because LD quantifies whether certain strains are more, or less, abundant than predicted based upon allele frequencies, the presence of LD in the population is necessary for the existence of pathogen strain structure or PSS (see the longer discussion in Section 2 of the main text). Thus, on the slow timescale, equation (S25) captures the evolutionary dynamics of PSS. Moreover, since $p_{AB}(t) = p_{ab}(t)$ and $p_{Ab}(t) = p_{aB}(t)$, if $D(t) > 0$, then the set of strains $\{ab, AB\}$ will be overrepresented in the population, whereas if $D(t) < 0$, the set of strains $\{Ab, aB\}$ will be overrepresented. Note that in this particular case, an overrepresented strain is also more abundant.

Assuming $\rho = 0$, equation (S25) has four potential equilibria: (1,2) when either set of strains has competitively excluded the other ($D(t) = \pm 1/4$); (3) when one set of strains is overrepresented, but all strains are present ($w_{AB} = 0$ for $0 < |D| < 1/4$); and (4) when there is no structure ($w_{AB} = 0$ for $D = 0$). Finally, if the sign of epistasis fluctuates, evolutionary oscillations, here defined to be any sustained dynamical behaviour of LD (including cycles and chaos) may occur. In all cases, however, the dynamical behaviour of equation (S25) hinges upon epistasis. The asymptotic behaviour of LD, arising from numerical simulation of the full epidemiological model of Box 1 is illustrated in Figure 3, and Figure S2 shows representative examples of the dynamical behaviour.

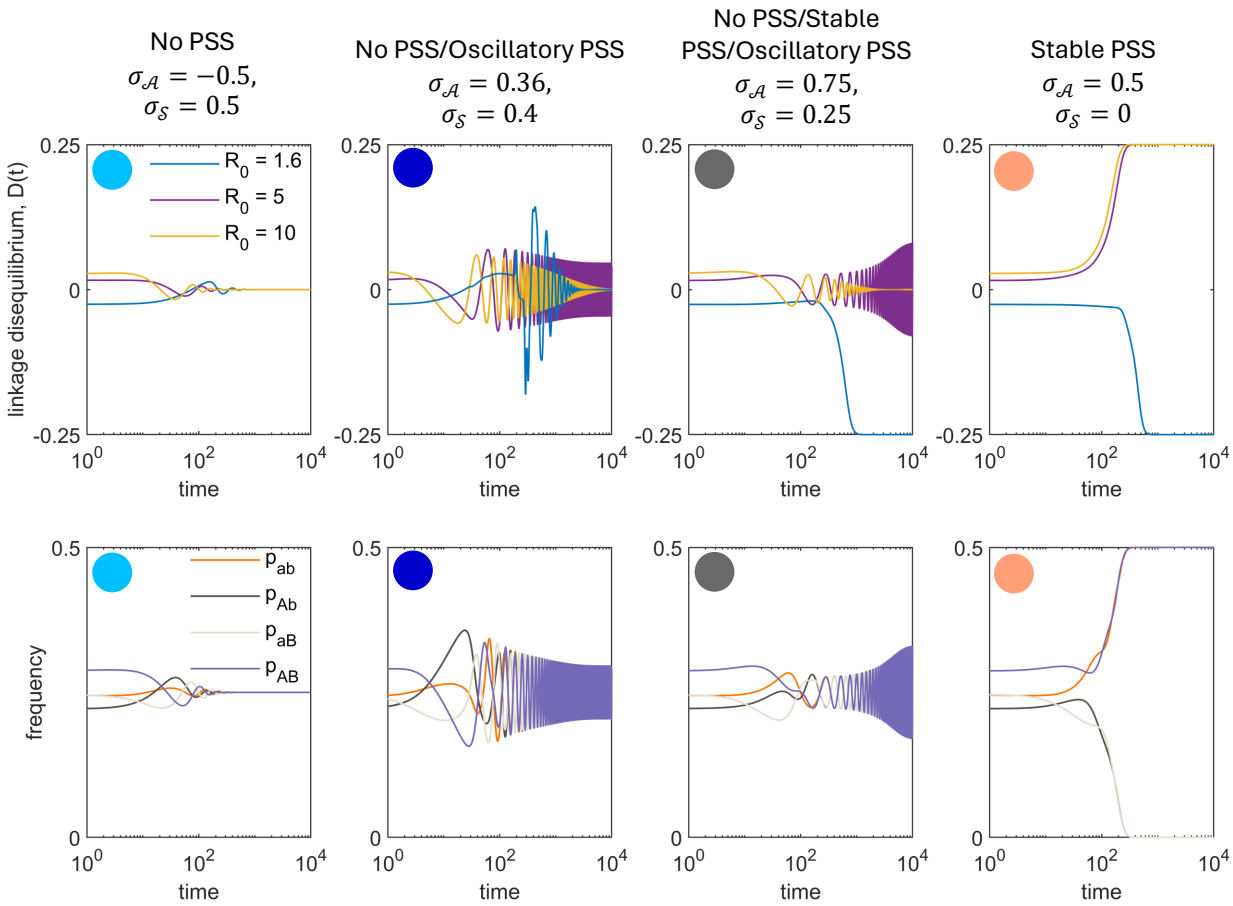


Figure S2: **Dynamical behaviour of LD and strain frequencies** as $\sigma_{\mathcal{A}}$, $\sigma_{\mathcal{S}}$, $\theta$ and $R_0$ vary. The dynamics of LD depend upon the values of $\sigma_{\mathcal{A}}$ and $\sigma_{\mathcal{S}}$, and also on the dynamic immunological landscape (controlled by $R_0$ and $\theta$). Each column corresponds to the indicated combination of $\sigma_{\mathcal{A}}$ and $\sigma_{\mathcal{S}}$. Coloured circles indicate the approximate position of the shown scenario in Figure 3. In the first row, the different trajectories correspond to the indicated value of $R_0$. The second row shows the strain frequency dynamics for a fixed value of $R_0 = 4$. The second column assumes $\theta = 0.01$, whereas the other columns assume $\theta = 0.05$. All simulations assume $b = d = 0.01$, $\alpha = 0$, $\rho = 10^{-8}$, $\sigma_0 = 1$ and $\sigma_h = 0$, whereas $\sigma_1$ and $\sigma_2$ vary by panel.

Finally, we note that given an observed distribution of allele frequencies, there is a maximum and minimum value of LD. For example, for two diallelic loci

$$\max(D(t)) = \min\big(p_A(t)(1-p_B(t)),\, p_B(t)(1-p_A(t))\big) \tag{S26}$$

and

$$\min(D(t)) = -\min\big(p_A(t)p_B(t), (1-p_A(t))(1-p_B(t))\big) \tag{S27}$$

[see 31]. Importantly, in order for pairwise LD to achieve this maximum or minimum, at least one strain must be absent from the population; in the special case when $p_A(t) = p_B(t)$, two strains are absent. Similar results hold for higher-order LD among more than two loci [81, 82], and when LD reaches either its maximum or minimum, one or more strains are absent from the population. We will use $D^*$ to indicate that the specified order of LD attains either its maximum or minimum, that is, one or more strains are competitively excluded.

# S4 Evolutionary dynamics on the slow timescale

In this section, we expand on our discussion in the main text about the evolutionary dynamics, focusing in particular on the role of $\sigma_{\mathcal{A}}$ and $\sigma_{\mathcal{S}}$. We do so in three sections. First, we attempt to provide additional intuition for our predictions. Second, we show the necessary conditions for stable PSS when $\sigma_{\mathcal{S}} = 0$ is $\sigma_{\mathcal{A}} > 0$. Finally, we provide numerical results supporting the claim that when $\sigma_{\mathcal{S}} = 0$, evolutionary oscillations do not occur and the long-term outcome is determined by the sign of $\sigma_{\mathcal{A}}$.

## S4.1 Intuition behind the dynamics

First suppose $\sigma_{\mathcal{S}} = 0$, that is, homologous immunity is no more protective than full antigenic immunity. In this case,

$$w_{AB} = \beta\sigma_{\mathcal{A}}\mathcal{A}_{\Delta}(t). \tag{S28}$$

As was demonstrated in Section S3.2, we can write

$$\mathcal{A}_{\Delta}(t) = S_{AB}(t) + S_{ab}(t) - S_{Ab}(t) - S_{aB}(t). \tag{S29}$$

In the absence of waning immunity, $\mathcal{A}_{\Delta}(t)$ indicates whether more uninfected hosts with a single prior infection were infected by one of the strains belonging to one set of strains or the other, i.e., $\{ab, AB\}$ or $\{Ab, aB\}$. If the set of strains $\{ab, AB\}$ was responsible for more of these infections, $\mathcal{A}_{\Delta}(t) > 0$; if the set of strains $\{Ab, aB\}$ was responsible for more of these infections, $\mathcal{A}_{\Delta}(t) < 0$. Because there is a single epidemiological feedback in (S28), $\mathcal{A}_{\Delta}(t)$, we would not expect epistasis to undergo sustained oscillations, nor for it to change sign at a particular amount of LD. Therefore, we eventually expect LD to reach an equilibrium: either stable PSS, $|D(t)| \to 1/4$, or no PSS, $D(t) \to 0$. The sign of $\sigma_{\mathcal{A}}$ determines which of these outcomes occurs: if $\sigma_{\mathcal{A}} > 0$, then stable PSS occurs, whereas if $\sigma_{\mathcal{A}} < 0$, all strains coexist at equal frequencies. Finally, if $\sigma_{\mathcal{A}} = 0$, then epistasis is identically zero. Here, the equilibrium amount of LD is contingent on the initial conditions. Note that this prediction for $\sigma_{\mathcal{A}} = 0$ assumes $\rho = 0$; otherwise, for any non-zero $\rho$, if $\sigma_{\mathcal{A}} = 0$, then no PSS is the only stable state. Below we provide the intuition for these predictions.

### S4.1.1 Sign of $\sigma_{\mathcal{A}}$

Suppose we are on the slow timescale, such that $p_{AB}(t) = p_{ab}(t)$ and $p_{Ab}(t) = p_{aB}(t)$. Consider a population in which all strains are present at equal frequencies and there is an equal number of uninfected hosts with immunity to each strain, that is, $D(t) = 0$ and $\mathcal{A}_\Delta(t) = 0$. Moreover, suppose one set of strains, say $\{ab, AB\}$, increases in frequency relative to the set of strains $\{Ab, aB\}$, due to some new infections, at least some of which occur in naive hosts. This results in $D(t) > 0$, that is, the set of strains $\{ab, AB\}$ is overrepresented. As these infections are cleared, this produces an excess of uninfected hosts with immunity to the overrepresented strains $\{ab, AB\}$, that is, $S_{AB}(t)+S_{ab}(t) > S_{Ab}(t)+S_{aB}(t)$, or $\mathcal{A}_\Delta(t) > 0$. This excess of uninfected hosts will have immunity to a single antigen of a randomly chosen strain belonging to the set of underrepresented strains, $\{Ab, aB\}$, and immunity to either 0 or 2 antigens of a randomly chosen strain belonging to the set of overrepresented strains $\{ab, AB\}$. The sign of $\sigma_{\mathcal{A}}$ then determines three possible cases:

(1) If $\sigma_{\mathcal{A}} < 0$, then $\sigma_1 > (\sigma_0 + \sigma_2)/2$. Here, the excess of uninfected hosts is, on average, more susceptible to infection by a randomly chosen strain belonging to the set of underrepresented strains, $\{Ab, aB\}$, than a randomly chosen strain belonging to the set of overrepresented strains, $\{ab, AB\}$. Consequently, the underrepresented strains will cause more infections per capita (i.e., have higher fitness) and so increase in frequency. As a result, the strains will return to an equal frequency, that is, $D(t) \to 0$.

(2) If $\sigma_{\mathcal{A}} > 0$, then $\sigma_1 < (\sigma_0+\sigma_2)/2$. Here, the excess of uninfected hosts is more susceptible to infection by a randomly chosen strain belonging to the set of overrepresented strains, $\{ab, AB\}$, than the set of underrepresented strains, $\{Ab, aB\}$. Thus, the overrepresented strains will further increase in frequency. This creates a positive feedback loop that eventually drives the set of overrepresented strains to fixation, that is, $D(t) \to 1/4$.

(3) Finally, if $\sigma_{\mathcal{A}} = 0$, then $\sigma_1 = (\sigma_0 + \sigma_2)/2$. Here, the excess of uninfected hosts is equally susceptible to infection by any strain. Consequently, neither set of strains will change in frequency. However, if there is any recombination in the population, that is, $\rho > 0$, recombination will systematically remove LD. Recombination will therefore reduce the frequency of the set of overrepresented strains, returning the population to a state at which all strains have equal frequency.

To show this mathematically, note that in the absence of waning immunity and virulence mortality,

$$\frac{dS_{ij}}{dt} = \mu I_0^{ij} - \beta(\sigma_0 I^{\hat{i}\hat{j}} + \sigma_1(I^{\hat{i}j} + I^{i\hat{j}}) + \sigma_2 I^{ij})S_{ij} - dS_{ij}$$

where $\hat{k}$ means the allele opposite to $k$ at the indicated locus, e.g., $\hat{a} = A$. Therefore, we can write

$$\begin{aligned}\frac{1}{\mu+d}\frac{d\mathcal{A}_\Delta}{dt} &= (1-\theta)(I_0^{AB} + I_0^{ab} - I_0^{Ab} - I_0^{aB}) - \theta\mathcal{A}_\Delta \\ &\quad - R_0(\sigma_0(S_{ab}I^{AB} + S_{AB}I^{ab}) + \sigma_1(S_{ab} + S_{AB})(I^{Ab} + I^{aB}) + \sigma_2(S_{ab}I^{ab} + S_{AB}I^{AB})) \\ &\quad + R_0(\sigma_0(S_{aB}I^{Ab} + S_{Ab}I^{aB}) + \sigma_1(S_{aB} + S_{Ab})(I^{ab} + I^{AB}) + \sigma_2(S_{Ab}I^{Ab} + S_{aB}I^{aB})), \end{aligned} \tag{S30}$$

which, on the slow timescale, reduces to

$$\frac{1}{\mu+d}\frac{d\mathcal{A}_\Delta}{dt} = 2(1-\theta)(I_0^{AB} - I_0^{Ab}) - \theta\mathcal{A}_\Delta \\ - 2R_0\left(I^{AB}([\sigma_0+\sigma_2]S_{AB} - 2\sigma_1 S_{Ab}) - I^{Ab}([\sigma_0+\sigma_2]S_{Ab} - 2\sigma_1 S_{AB})\right). \quad \text{(S31)}$$

Now, suppose we are at the equilibrium state in which there is no strain structure. In this case, $D(t) = 0$ and $S_{AB}(t) = S_{Ab}(t) = S_2(t)$. Imagine that infections by the set of strains $\{ab, AB\}$ slightly increase in density,

$$I_0^{AB} \approx \frac{I_0}{4} + \delta I_0^{(1)} + \mathcal{O}(\delta^2) \quad \text{(S32)}$$

$$I^{AB} \approx \frac{I}{4} + \delta I^{(1)} + \mathcal{O}(\delta^2), \quad \text{(S33)}$$

where $I_0$ is the total density of strains (irrespective of genotype) infecting naive hosts, $I_0^{(1)}$ and $I^{(1)}$ are the (relative) perturbed values of $I_0^{AB}$ and $I^{AB}$, respectively, and $0 < \delta \ll 1$ is a small quantity. In this case,

$$D(t) \approx \frac{1}{2}\frac{I^{(1)}(t)}{I(t)}\delta + \mathcal{O}(\delta^2).$$

Using these relations, equation (S31) reduces to

$$\frac{1}{\mu+d}\frac{d\mathcal{A}_\Delta}{dt} \approx \left(4(1-\theta)\frac{I_0^{(1)}}{I^{(1)}} - R_0\sigma_{\mathcal{A}}S_2\right)DI + \mathcal{O}(\delta^2). \quad \text{(S34)}$$

Next, note that $S_2$ is the number of hosts with immunity to a single strain; in the vicinity of the no strain structure equilibrium, the dynamics of $S_2$ are given by

$$\frac{dS_2}{dt} \approx \mu I_0 - dS_2 - \beta\frac{(\sigma_0+2\sigma_1+\sigma_2)}{4}IS_2 + \mathcal{O}(\delta) \\ = (\mu+d)\left((1-\theta)I_0 - \theta S_2 - R_0\frac{(\sigma_0+2\sigma_1+\sigma_2)}{4}IS_2\right) + \mathcal{O}(\delta),$$

and so

$$S_2 \approx \frac{4(1-\theta)I_0}{4\theta + R_0(\sigma_0+2\sigma_1+\sigma_2)I} + \mathcal{O}(\delta).$$

Hence,

$$\frac{1}{\mu+d}\frac{d\mathcal{A}_\Delta}{dt} \approx 4(1-\theta)\left(\frac{I_0^{(1)}}{I^{(1)}} - \frac{R_0\sigma_{\mathcal{A}}I_0}{4\theta + R_0(\sigma_0+2\sigma_1+\sigma_2)I}\right)DI + \mathcal{O}(\delta^2). \quad \text{(S35)}$$

It therefore follows that if

$$\frac{I_0^{(1)}}{I^{(1)}} > \frac{R_0\sigma_{\mathcal{A}}I_0}{4\theta + R_0(\sigma_0+2\sigma_1+\sigma_2)I} \quad \text{(S36)}$$

then the increase in strains $\{ab, AB\}$ due to the perturbation results in $\mathcal{A}_\Delta(t)$ and $D(t)$ sharing sign, in which case, if $\sigma_\mathcal{A} > 0$, $D(t)$ will increase, whereas if $\sigma_\mathcal{A} < 0$, $D(t)$ will decrease. Inequality (S36) always holds if $\sigma_\mathcal{A} < 0$. If $\sigma_\mathcal{A} > 0$, inequality (S36) will also hold for biologically 'reasonable' perturbations. Here, a biologically reasonable perturbation should be roughly proportional to the existing distribution of infections, that is,

$$\frac{\mathcal{I}_0^{(1)}}{\mathcal{I}^{(1)}} \approx \frac{I_0}{I}.$$

That is because the existing distribution of infections (at equilibrium) indicates the relative likelihood of infection based upon immune history for a given set of parameters. Hence, $I_0/I$ measures the relative infection risk of a naive host.

#### S4.1.2 Conflicting evolutionary pressures induced by $\sigma_\mathcal{S} > 0$

Next, suppose $\sigma_\mathcal{S} > 0$. Then

$$w_{AB} = \beta\sigma_\mathcal{A}\mathcal{A}_\Delta(t) - \beta\sigma_\mathcal{S}\mathcal{S}_\Delta(t). \tag{S37}$$

Since $\mathcal{S}_\Delta(t)$ consists of uninfected hosts that have been infected only once and uninfected hosts that have been infected multiple times, it can diverge from $\mathcal{A}_\Delta(t)$. This can create a second epidemiological feedback on epistasis. In general, this second epidemiological feedback removes LD. To see why, assume that $\sigma_\mathcal{A} = 0$. As before, suppose that we start from a state in which all strains are at equal frequency and there is an equal number of uninfected hosts with immunity to each strain, i.e., $D(t) = 0$ and $\mathcal{A}_\Delta(t) = \mathcal{S}_\Delta(t) = 0$. Then suppose that one set of strains, say $\{ab, AB\}$, increases in frequency relative to the other through additional infections. This leads to $D(t) > 0$, that is, the set of strains $\{ab, AB\}$ is overrepresented. As hosts clear these infections, an excess of uninfected hosts with homologous immunity to the strains $\{ab, AB\}$ appears, and so $\mathcal{S}_\Delta(t) > 0$. This excess of uninfected hosts has immunity to a single antigen of the underrepresented strains, $\{Ab, aB\}$, and either no immunity or homologous immunity to the overrepresented strains, $\{ab, AB\}$. Thus, for this excess of uninfected hosts, the per-capita susceptibility to infection by a randomly chosen member of the set of underrepresented strains is $\sigma_1$, and the per-capita susceptibility to infection by a randomly chosen member of the set of overrepresented strains is $(\sigma_0 + \sigma_h)/2$. Since $\sigma_\mathcal{A} = 0$, this implies $\sigma_1 = (\sigma_0 + \sigma_2)/2$, and therefore

$$\sigma_1 = \frac{\sigma_0 + \sigma_2}{2} > \frac{\sigma_0 + \sigma_h}{2}. \tag{S38}$$

Thus, the excess of uninfected hosts is more susceptible to infection by the underrepresented strains, $\{Ab, aB\}$, and so the underrepresented strains will increase in frequency. This will return $D(t) \to 0$. If $\sigma_\mathcal{A} < 0$ in addition to $\sigma_\mathcal{S} > 0$, this implies $\sigma_1 > (\sigma_0 + \sigma_2)/2$, and so inequality (S38) will be more pronounced. Thus, the per-capita advantage of the underrepresented strains increases. Consequently, we would again expect $D(t) \to 0$. However, if $\sigma_\mathcal{A} > 0$, inequality (S38) will not hold. In this case, the two epidemiological feedbacks become opposed. This can either produce evolutionary oscillations or result in a stable nonzero value of LD that is not equal to $D^*$ (i.e., at equilibrium, $w_{AB} = 0$ and $0 < |D(t)| < 1/4$).

Due to the high dimensionality of the model, analytical results are difficult to obtain. However, in the next section we improve the mathematical prediction of the stability of the PSS equilibrium when $\sigma_\mathcal{S} = 0$.

## S4.2 Epistasis and LD must share sign for local asymptotic stability when $\sigma_{\mathcal{S}} = 0$

Taking the derivative of the differential equation for LD, $\frac{\mathrm{d}D}{\mathrm{d}t}$, given in system (S22), with respect to any of the population variables (arbitrarily indicated as $X$) under the assumption that recombination is negligible ($\rho \approx 0$), we obtain:

$$\begin{aligned}\frac{\partial}{\partial X}\frac{\mathrm{d}D}{\mathrm{d}t} = D\left((1-2p_A)\frac{\partial w_A}{\partial X} + (1-2p_B)\frac{\partial w_B}{\partial X}\right) \\ + (p_A p_B + D)((1-p_A)(1-p_B) - D)\frac{\partial w_{AB}}{\partial X} + w_A\frac{\partial[1-2p_A]D}{\partial X} \\ + w_B\frac{\partial[1-2p_B]D}{\partial X} + w_{AB}\frac{\partial}{\partial X}[(p_A p_B + D)((1-p_A)(1-p_B) - D)]. \quad \text{(S39)}\end{aligned}$$

We evaluate (S39) at $p_A = p_B = 1/2$ and $D = 1/4$ or $D = -1/4$; these are the stable PSS equilibria that occur when $\sigma_{\mathcal{S}} = 0$. Since $r_{AB} = r_{ab}$ and $r_{Ab} = r_{aB}$, and so $w_{AB} = -2w_A = -2w_B$, equation (S39) reduces to

$$\frac{\partial}{\partial X}\frac{\mathrm{d}D}{\mathrm{d}t} = \begin{cases} -2w_{AB}D & \text{if } X = D \\ 0 & \text{otherwise} \end{cases} \quad \text{(S40)}$$

Here, the Jacobian of the full system that combines both the evolutionary and epidemiological variables, evaluated at either of the stable PSS equilibria, becomes block triangular, with one of the 'blocks' on the diagonal given by equation (S40) (when $X = D$). Thus, one of the eigenvalues is $-w_{AB}/2$ (if $D = 1/4$) or $w_{AB}/2$ (if $D = -1/4$). It follows that a necessary condition for PSS to be locally asymptotically stable when $\sigma_{\mathcal{S}} = 0$ is that the sign of epistasis matches the sign of LD.

Now, on the slow timescale, the sign of $D(t)$ indicates whether there are more hosts currently infected by members of the set of strains $\{AB, ab\}$ (if $D(t) > 0$) or by members of the set of strains $\{Ab, aB\}$ (if $D(t) < 0$). Therefore $D(t)$ and $\mathcal{A}_\Delta(t)$ share sign whenever more uninfected hosts have full antigenic immunity to the set of overrepresented strains. Indeed, as $|D(t)| \to 1/4$, only one set of strains will be present in the population. Therefore, if hosts gain or lose immunity to a set of antigens that define an infecting (or previously infecting) strain, at the stable PSS equilibrium

$$\mathcal{A}_\Delta(t) = \begin{cases} S_{AB}(t) + S_{ab}(t) & \text{if } D(t) = 1/4 \\ S_{Ab}(t) + S_{aB}(t) & \text{if } D(t) = -1/4 \end{cases} \quad \text{(S41)}$$

Applying this result to equation (S40), it follows that when $\sigma_{\mathcal{A}} > 0$, epistasis and LD share sign at the PSS equilibrium, otherwise they do not. Thus, a *necessary* condition for local asymptotic stability of the PSS equilibrium when $\sigma_{\mathcal{S}} = 0$ is that $\sigma_{\mathcal{A}} > 0$. The high dimensionality of the system impeded the derivation of the necessary and sufficient conditions. In the next section, we therefore numerically explore the predictions.

## S4.3 Numerical simulations of key predictions

In the previous sections, we have provided intuition for our key predictions: (i) if $\sigma_{\mathcal{S}} = 0$ and $\sigma_{\mathcal{A}} > 0$, stable PSS occurs (we also proved this is the necessary condition); (ii) if $\sigma_{\mathcal{S}} = 0$

and $\sigma_{\mathcal{A}} < 0$, no PSS occurs; (iii) evolutionary oscillations can only occur if both $\sigma_{\mathcal{S}} > 0$ and $\sigma_{\mathcal{A}} > 0$. Here, we provide numerical simulations to confirm these predictions. In the main text, we show (numerically) that evolutionary oscillations are possible for $\sigma_{\mathcal{S}} > 0$ and $\sigma_{\mathcal{A}} > 0$ (see Fig. 3 and Fig. 4); furthermore, we show that when $\sigma_{\mathcal{S}} = 0$, the sign of $\sigma_{\mathcal{A}}$ predicts the (stable) outcome of either no PSS ($\sigma_{\mathcal{A}} < 0$) or stable PSS ($\sigma_{\mathcal{A}} > 0$). Here, we expand the simulations for the case when $\sigma_{\mathcal{S}} = 0$ to also show the dynamics of $\mathcal{A}_{\Delta}(t)$ in relation to $D(t)$.

For our simulations, we used the model specified by equations (S7), and assumed that: (i) immunity is lifelong and acquired at time of clearance, (ii) immune protections only affect susceptibility to infection (so $\beta_z^g = \beta$, $\mu_z^g = \mu$, $\alpha_z^g = \alpha$), and (iii) we fixed $b = d = 0.01$, $\sigma_0 = 1$, $\sigma_2 = \sigma_h = 0$, $\alpha = 0$, and $\rho = 10^{-8}$. We then looped over 40 evenly spaced values of $R_0 \in [1.3, 20]$, 40 evenly spaced values of $\theta \in [5 \times 10^{-5}, 0.5]$, and 20 evenly spaced values of $\sigma_1 \in [0, 1]$. For each parameter combination, we generated a random initial condition at $t = 0$, subject to the constraint that the total population density is 1. Thus, the initial condition is not on the slow timescale. We then numerically integrated the system of ODEs until an equilibrium is reached; we determined this numerically by integrating forward in time until

$$|\max_{t \in t_x} D(t) - \min_{t \in t_x} D(t)| < 10^{-7},$$

where $t_x$ is the last $10^4$ time units. Figure S3 shows the outcome of these $3.2 \times 10^4$ simulations. Three notable implications arise from the simulations. First, $\mathcal{A}_{\Delta}(t)$ and $D(t)$ share sign (Fig. S3**a**, **b**), except during the fast timescale transient dynamics (these fast timescale dynamics are the horizontal trajectories in panels **a**,**b**; here, $\mathcal{A}_{\Delta}(t)$ is rapidly changing while $D(t)$ is not). Second, in all $3.2 \times 10^4$ simulations, the sign of $\sigma_{\mathcal{A}}$ predicts the long-term outcome (Fig. S3**c**): when $\sigma_{\mathcal{A}} > 0$, $|D(t)| \to 1/4$, whereas when $\sigma_{\mathcal{A}} < 0$, $|D(t)| \to 0$. Finally, sustained evolutionary oscillations were not observed in any of the simulations, in agreement with our prediction that these can only occur if both $\sigma_{\mathcal{A}} > 0$ and $\sigma_{\mathcal{S}} > 0$.

# S5 Different types of immune protections

Next, we consider a more general case in which we assume that immune protections may affect transmissibility, clearance, and virulence mortality, but that homologous immunity is equivalent to full antigenic immunity. We obtain an expression for epistasis in fitness, which is more complicated than when immunity only affects susceptibility. That is because if immunity affects transmissibility, clearance, and/or virulence-mortality, the fitness (per-capita growth rate) of an infection will depend upon the immune environment the infection is found in. As a consequence, the per-capita growth rate of strain $ij$ depends upon the frequency of strain $ij$ infections found in each particular environment. Therefore, we focus upon deriving epistasis in fitness at equilibrium. In this case, there is one epidemiological feedback, which is whether more hosts have antigenic immunity to members of one set of strains or the other (e.g., $\{ab, Ab\}$ or $\{Ab, aB\}$). Hence, if recombination is weak, there is no reason to expect epistasis to become zero and/or change sign at some non-zero amount of LD. Consequently, we expect that if PSS occurs, it will be the largest in magnitude possible given the allele frequencies, that is, $D(t) \to D^*$. Specifically, if $p_A(t) = p_B(t) = 1/2$, and if PSS occurs, then $|D(t)| \to 1/4$ (this means two strains will be excluded from the population). We will focus upon these two equilibria here.

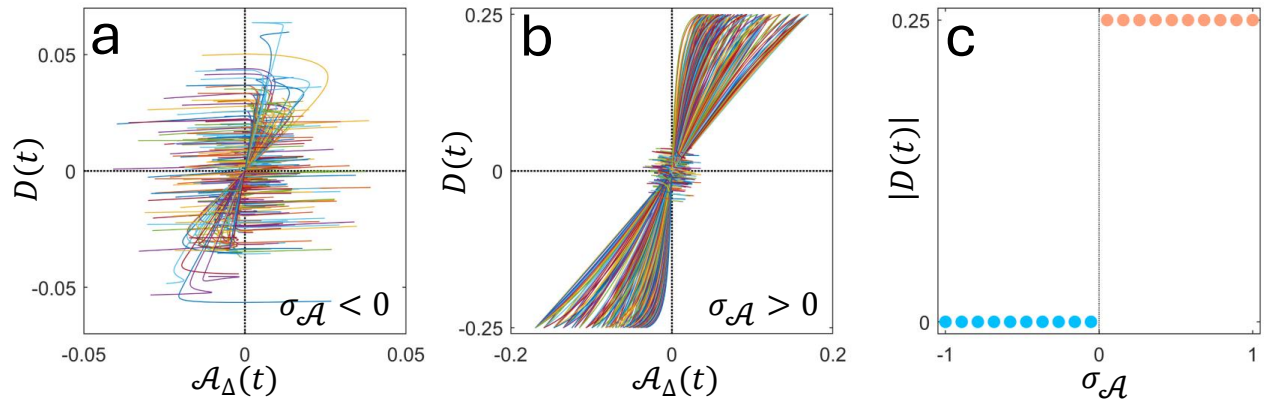


Figure S3: **Relationship between LD and $\mathcal{A}_\Delta(t)$ for different values of $R_0$ and $\theta$.** Numerical simulations show that $\mathcal{A}_\Delta(t)$ and $D(t)$ generally share sign, regardless of whether $\sigma_\mathcal{A} > 0$ (panel **a**) or $\sigma_\mathcal{A} < 0$ (panel **b**); any divergence is transient, occurring on the fast timescale (horizontal trajectories in panels **a**,**b**). In the long term, the evolutionary outcome is determined by the sign of $\sigma_\mathcal{A}$: if $\sigma_\mathcal{A} < 0$, $D(t) \to 0$, whereas if $\sigma_\mathcal{A} > 0$, $|D(t)| \to 1/4$ (panel **c**). Shown are the results of $3.2 \times 10^4$ simulations, in which we looped over 40 evenly spaced values of $R_0 \in [1.3, 20]$, 40 evenly spaced (on log-scale) values of $\theta \in [5 \times 10^{-5}, 0.5]$ and 20 evenly spaced values of $\sigma_1 \in [0, 1]$. All simulations assume $b = d = 0.01$, $\rho = 10^{-8}$, $\sigma_0 = 1$, $\sigma_2 = \sigma_h = 0$, and $\alpha = 0$.

To calculate epistasis at equilibrium we define

$$f_n \equiv \frac{\sigma_n \beta_n}{\mu_n + \alpha_n + d}, \tag{S42}$$

which is proportional to the effective reproductive number of an infection in a host with immunity to $n$ antigens of the infecting strain. This is because, an infection is established in a host with immunity to $n$ of its antigens with probability $\sigma_n$; such an infection has an expected lifetime of $(\mu_n + \alpha_n + d)^{-1}$, during which it will produce new infections at a rate proportional to $\beta_n$.

Next, at equilibrium, the reproductive number of strain $ij$ [e.g., see 83], which we denote $\bar{R}_{ij}$, is

$$\begin{aligned}\bar{R}_{ij} &= f_0(\bar{\mathcal{S}} - \bar{\mathcal{A}}_{\{i\}} - \bar{\mathcal{A}}_{\{j\}} + \bar{\mathcal{A}}_{\{i,j\}}) + f_1(\bar{\mathcal{A}}_{\{i\}} + \bar{\mathcal{A}}_{\{j\}} - 2\bar{\mathcal{A}}_{\{i,j\}}) + f_2\bar{\mathcal{A}}_{\{i,j\}} \\ &= f_0\bar{\mathcal{S}} - (f_0 - f_1)(\bar{\mathcal{A}}_{\{i\}} + \bar{\mathcal{A}}_{\{j\}}) + (f_0 - 2f_1 + f_2)\bar{\mathcal{A}}_{\{i,j\}}.\end{aligned}$$

Now, the per-capita growth rate, $r_{ij}$, is a measure of fitness in continuous time, whereas the reproductive number, $\bar{R}_{ij}$, is a measure of fitness in discrete time [83]. This can be explicitly shown using, for example, the Next Generation Theorem [84, 85], a technique which converts a continuous-time system into a discrete time system [83]. Whereas in continuous time, epistasis in fitness is defined as

$$w_{AB} \equiv r_{ab} + r_{AB} - r_{Ab} - r_{aB},$$

in discrete time, epistasis in fitness becomes

$$\bar{w}_{AB} \equiv \bar{R}_{AB}\bar{R}_{ab} - \bar{R}_{Ab}\bar{R}_{aB}$$

[see e.g., 86, 87]. Without loss of generality, we focus upon the equilibrium in which the set of overrepresented strains is $\{ab, AB\}$. At this equilibrium, $\bar{R}_{AB} = \bar{R}_{ab} = 1$; moreover, $\bar{R}_{Ab} = \bar{R}_{aB}$. Therefore, when $(p_A, p_B, D) = (1/2, 1/2, 1/4)$,

$$\bar{w}_{AB} = \bar{R}_{AB}\bar{R}_{ab} - \bar{R}_{Ab}\bar{R}_{aB} = 1 - \bar{R}_{Ab}^2 = (1 - \bar{R}_{Ab})(1 + \bar{R}_{Ab}).$$

Since $\bar{R}_{ij} \geq 0$,

$$\bar{w}_{AB} \propto 2(1 - \bar{R}_{Ab}) = 2(\bar{R}_{ab} - \bar{R}_{Ab}) = 2\left(f_0 - 2f_1 + f_2\right)\left(\bar{\mathcal{A}}_{\{a,b\}} - \bar{\mathcal{A}}_{\{A,b\}}\right) = \epsilon_{\mathcal{A}}\bar{\mathcal{A}}_{\Delta}, \qquad \text{(S43)}$$

where proportionality is up to a positive constant, $\epsilon_{\mathcal{A}} = f_0 - 2f_1 + f_2$, and $\bar{\mathcal{A}}_{\Delta} = \bar{\mathcal{A}}_{\{a,b\}} + \bar{\mathcal{A}}_{\{A,B\}} - \bar{\mathcal{A}}_{\{A,b\}} - \bar{\mathcal{A}}_{\{a,B\}}$ (note that $\bar{\mathcal{A}}_{\{a,b\}} = \bar{\mathcal{A}}_{\{A,B\}}$ and $\bar{\mathcal{A}}_{\{A,b\}} = \bar{\mathcal{A}}_{\{a,B\}}$ at equilibrium). Hence, if $\epsilon_{\mathcal{A}} > 0$, epistasis and LD will share sign.

There are four important consequences of the inclusion of different types of immune protections.

(1) First, $\epsilon_{\mathcal{A}}$ is a decreasing function of $\sigma_1$, $\beta_1$, and $\alpha_1$, and is an increasing function of $\mu_1$. Thus, increasing the strength of immune protections targeting transmission (smaller $\beta_1$), susceptibility to infection (smaller $\sigma_1$), and clearance rate (larger $\mu_1$) promotes PSS by making it more likely $\epsilon_{\mathcal{A}}$ is positive. On the other hand, increasing the strength of immune protections targeting virulence mortality (smaller $\alpha_1$) inhibits PSS by making it more likely $\epsilon_{\mathcal{A}}$ is negative.

(2) Second, if immunity only reduces virulence mortality, then genetic polymorphisms cannot be stably maintained. This can be seen by supposing only one strain is present, say $AB$, and that the population is at its endemic equilibrium. If, in addition, homologous immunity is as protective as full antigenic immunity and there is no waning immunity, uninfected hosts can be either naive or have immunity to the antigens $AB$. Denote the equilibrium density of naive hosts as $\bar{S}_0$, and the equilibrium density of uninfected hosts with immunity to antigens $AB$ as $\bar{S}_{AB}$. Then strain $Ab$ and/or strain $aB$ can increase in frequency when rare if they have a higher basic reproductive number than strain $AB$, that is,

$$f_1\bar{S}_{AB} + f_0\bar{S}_0 > f_2\bar{S}_{AB} + f_0\bar{S}_0.$$

Likewise, strain $ab$ can increase in frequency when rare if

$$f_0(\bar{S}_{AB} + \bar{S}_0) > f_2\bar{S}_{AB} + f_0\bar{S}_0.$$

If immunity only reduces virulence mortality, that is, $\alpha_2 < \alpha_1 < \alpha_0$, then $f_0 < f_1 < f_2$ and so neither inequality is satisfied. Thus if one strain is dominant in the population, it can out compete all the other strains. Consequently, if immunity only acts to reduce virulence mortality, it cannot maintain genetic polymorphisms.

(3) Third, suppose only one type of immune protection is operating. In particular, suppose immunity either reduces susceptibility to infection, reduces transmissibility, or increases clearance rate. Furthermore, suppose that whatever type of immune protection is present, its strength additively increases with the number of antigens that a host has been exposed to, that is,

$$\sigma_1 = \frac{\sigma_0 + \sigma_2}{2}, \quad \beta_1 = \frac{\beta_0 + \beta_2}{2}, \quad \text{and} \quad \mu_1 = \frac{\mu_0 + \mu_2}{2}.$$

If immunity affects only transmissibility or susceptibility to infection, then $\epsilon_{\mathcal{A}} = 0$. Thus PSS can only occur for transmissibility or susceptibility if immune protections are saturating, that is, the increase in immunity is more than additive (this is also shown in Figure 2). However, if immunity affects only clearance rate, then an additive increase in immune protections yields

$$\epsilon_{\mathcal{A}} = \frac{f_0 f_2}{\beta_0 \sigma_0} \frac{(\mu_0 - \mu_2)^2}{\mu_0 + \mu_2 + 2(\alpha_0 + d)} \tag{S44}$$

As this is positive, immunity affecting clearance rate is more effective at producing PSS (Fig. S4). Besides, we would generally expect the opposite prediction for immune protections that reduce virulence mortality, provided there is another type of immune protection maintaining genetic polymorphisms. For example, suppose the strength of immune protections additively increases, and affects both susceptibility and virulence mortality. Further, assume that $f_0 > f_1$ and $f_0 > f_2$, that is, genetic polymorphisms at both loci are stably maintained. Then

$$\epsilon_{\mathcal{A}} = -(f_0 - f_1) \frac{\alpha_0 - \alpha_2}{\mu_0 + \alpha_2 + d}. \tag{S45}$$

Since $f_0 > f_1$ and $\alpha_0 > \alpha_2$, it follows that $\epsilon_{\mathcal{A}} < 0$, that is, no strain structure is favoured when immune protections against virulence mortality are additively increasing. Thus immunity affecting virulence mortality is less effective at producing strain structure (Fig. S4).

Indeed, suppose immunity affects only clearance rate and we solve for the value of $\mu_1$, say $\mu_1^*$, at which $\epsilon_{\mathcal{A}} = 0$. Thus if $\mu_1 > \mu_1^*$ PSS occurs. Solving the relation $\theta = d/(\mu_0 + \alpha_0 + d)$ for $d$, and using this in $\mu_1^*$, we obtain

$$\mu_1^* = \frac{(\mu_2 - \mu_0)(\mu_0 + \alpha_0)}{\mu_2(1 - \theta) + \mu_0(1 + \theta) + 2\alpha_0} + \mu_0. \tag{S46}$$

This is an increasing function of $\theta$ and $\alpha_0$. Therefore, PSS is increasingly likely (smaller $\mu_1^*$) when infections become more acute ($\theta \to 0$) and infections become less virulent ($\alpha_0 \to 0$ for fixed $\theta$), as claimed in the main text.

(4) Finally, the derivation of equation (S43) assumes that homologous immunity is equivalent to full antigenic immunity. The analogous quantity to $\sigma_{\mathcal{S}}$ is $\epsilon_{\mathcal{S}} \equiv f_2 - f_h$. We conjecture that $\epsilon_{\mathcal{S}} > 0$ is a necessary condition for oscillatory PSS.

# S6 Variation in immune protection between alleles and and/or strains

The key assumption facilitating the separation of timescales in Section S3.4 is that the alleles at each locus, and the combinations of alleles, are of equal intrinsic fitness, i.e., they induce identical immune protections and do not differ with respect to any other life-history characteristic (e.g., transmissibility [13, 42]). Thus, all hosts with immunity to (for example)

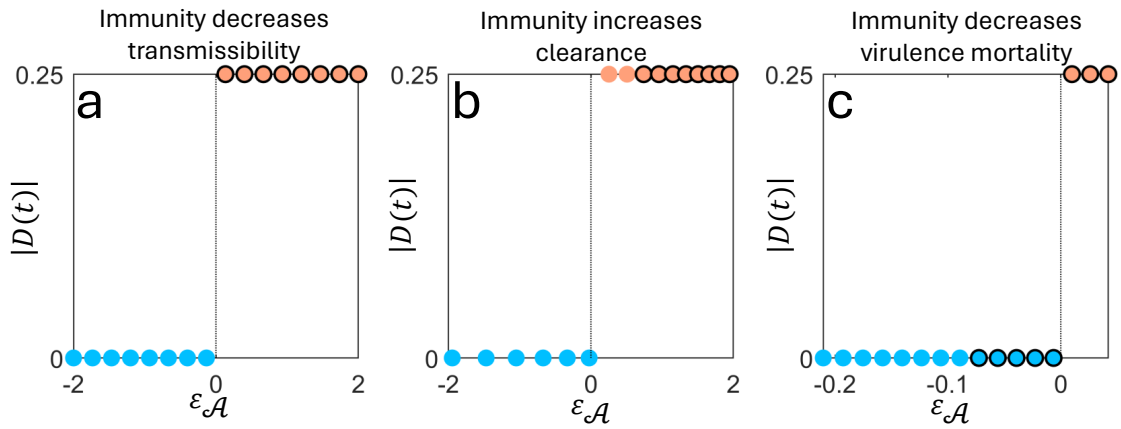


Figure S4: **The ability of different types of immune protections to produce PSS can be predicted by the sign of $\epsilon_{\mathcal{A}}$.** In panel **a**, immunity only reduces transmissibility (i.e., $\sigma_n = \sigma_0$, $\mu_n = \mu_0$, and $\alpha_n = \alpha_0$), and we vary $\beta_1$ over 16 evenly-spaced values between $\beta_0$ and $\beta_2 = \beta_0/2$ inclusive. Here, $\epsilon_{\mathcal{A}} = 0$ when $\beta_1 = (\beta_0+\beta_2)/2$, and so an equal number of circles are to either side of the dashed black line (circles with black edges indicate $\beta_1 > (\beta_0 + \beta_2)/2$). In panel **b**, we assume immunity only effects clearance rate (i.e., $\sigma_n = \sigma_0$, $\beta_n = \beta_0$, and $\alpha_n = \alpha_0$), and vary $\mu_1$ over 16 evenly-spaced values between $\mu_0$ and $\mu_2 = 2\mu_0$ inclusive. Because immune protections affecting clearance rate are more likely to produce PSS, $\epsilon_{\mathcal{A}} = 0$ for $\mu_1 < (\mu_0 + \mu_2)/2$, and so more circles are to the right of the dashed black line than the left (circles with black edges indicate $\mu_1 > (\mu_0 + \mu_2)/2$). Finally, in panel **c**, immunity only reduces virulence mortality and susceptibility (i.e., $\beta_n = \beta_0$ and $\mu_n = \mu_0$). Here, we assume that immunity additively reduces susceptibility, $\sigma_1 = (\sigma_0 + \sigma_2)/2$, and vary $\alpha_1$ over 16 evenly-spaced values between $\alpha_2 = 0$ and $\alpha_0$ inclusive. Because immune protections affecting virulence mortality are less likely to produce PSS, $\epsilon_{\mathcal{A}} = 0$ for $\alpha_1 > (\alpha_0 + \alpha_2)/2$. Thus, more circles are to the left of the dashed black line than to the right (circles with black edges indicate $\alpha_1 < (\alpha_0 + \alpha_2)/2$). Unless otherwise noted, all simulations assume $\mu_0 = 9.49$, $\beta_0 = 40$, $\alpha_0 = 0.5$, $b = d = 0.01$, $\sigma_0 = 1$, and $\rho = 10^{-9}$; thus $R_0 = 4$, $\theta = 10^{-3}$, and the probability a naive host dies from an infection, $\alpha_0/(\mu_0 + \alpha_0 + d)$, is 0.05.

one antigen of a strain have the same level of protection, irrespective of the identity of the antigen. As a result, allele frequencies rapidly converge to 1/2. After that, the LD, which determines the amount of strain structure, evolves on the slow timescale.

The symmetry between alleles (and strains) could be disrupted in at least two ways: (i) if there is an unequal number of alleles at the different loci, preventing all alleles from coexisting at equal frequencies, or (ii) if the different alleles and/or combinations of alleles induce different levels of immune protection. In Section S8, we explore what happens when the number of alleles at each loci is not identical. Here, we examine how variation in the immune protections against each allele, and each combination of alleles, affects our predictions. We investigate two key predictions: (1) whether homologous immunity being more protective than full antigenic immunity is necessary for sustained evolutionary oscillations, and (2) whether, if homologous immunity is as protective as full antigenic immunity and so oscillations are not possible, then $\sigma_{\mathcal{A}} > 0$ implies that LD stably persists at either its maximum (equation (S26)) or minimum (equation (S27)) value, whereas if $\sigma_{\mathcal{A}} < 0$ no strain

structure occurs.

To test these predictions, we define $\sigma_1^{\{i\}}$, $\sigma_2^{\{i,j\}}$, and $\sigma_h^{\{ij\}}$, as the reduction in susceptibility to an infection by strain $ij$ of an uninfected host with immunity to antigen $i$ but not $j$, immunity to both antigens $i$ and $j$, and homologous immunity, respectively. We assume that full antigenic immunity is as protective as homologous immunity for each combination of antigens, that is, $\sigma_2^{\{i,j\}} = \sigma_h^{\{ij\}}$. Then the per-capita growth rate of strain $ij$ is

$$r_{ij} = \beta\left(\sigma_0\mathcal{S} - (\sigma_0 - \sigma_1^{\{i\}})\mathcal{A}_{\{i\}} - (\sigma_0 - \sigma_1^{\{j\}})\mathcal{A}_{\{j\}} + \sigma_{\mathcal{A}}^{\{i,j\}}\mathcal{A}_{\{i,j\}}\right) - d - \mu - \alpha, \tag{S47}$$

where

$$\sigma_{\mathcal{A}}^{\{i,j\}} = \sigma_0 - \sigma_1^{\{i\}} - \sigma_1^{\{j\}} + \sigma_2^{\{i,j\}}. \tag{S48}$$

Consequently, the additive selection coefficients and epistasis are

$$\begin{aligned} w_A &= \beta\left((\sigma_0 - \sigma_1^{\{a\}})\mathcal{A}_{\{a\}} - (\sigma_0 - \sigma_1^{\{A\}})\mathcal{A}_{\{A\}} + \sigma_{\mathcal{A}}^{\{A,b\}}\mathcal{A}_{\{A,b\}} - \sigma_{\mathcal{A}}^{\{a,b\}}\mathcal{A}_{\{a,b\}}\right) \\ w_B &= \beta\left((\sigma_0 - \sigma_1^{\{b\}})\mathcal{A}_{\{b\}} - (\sigma_0 - \sigma_1^{\{B\}})\mathcal{A}_{\{B\}} + \sigma_{\mathcal{A}}^{\{a,B\}}\mathcal{A}_{\{a,B\}} - \sigma_{\mathcal{A}}^{\{a,b\}}\mathcal{A}_{\{a,b\}}\right) \\ w_{AB} &= \beta\left(\sigma_{\mathcal{A}}^{\{A,B\}}\mathcal{A}_{\{A,B\}} + \sigma_{\mathcal{A}}^{\{a,b\}}\mathcal{A}_{\{a,b\}} - \sigma_{\mathcal{A}}^{\{A,b\}}\mathcal{A}_{\{A,b\}} - \sigma_{\mathcal{A}}^{\{a,B\}}\mathcal{A}_{\{a,B\}}\right). \end{aligned} \tag{S49}$$

Because the amount of LD possible depends upon the allele frequencies, we test our predictions using a normalized version of LD [88], $D'$, defined as

$$D' = \begin{cases} \frac{-D}{\min(p_A p_B, (1-p_A)(1-p_B))} & \text{if } D < 0 \\ \frac{D}{\min(p_A(1-p_B), (1-p_A)p_B)} & \text{if } D > 0 \end{cases} \tag{S50}$$

Here, if $D' = 0$, there is no PSS. If $D' = 1$, there is the maximum possible amount of PSS, given allele frequencies. Further, we test our predictions for the sign of $\sigma_{\mathcal{A}}$ using the sign of the value of $\sigma_{\mathcal{A}}$ averaged across allelic combinations:

$$\bar{\sigma}_{\mathcal{A}} = \frac{\sum_{ij}\sigma_{\mathcal{A}}^{\{i,j\}}}{4}. \tag{S51}$$

Here, epistasis is no longer the product of a composite parameter, $\sigma_{\mathcal{A}}$, and a composite variable $\mathcal{A}_{\Delta}(t)$, which makes analytic progress difficult. Therefore, we rely on simulations. To do so, we let the protection against alleles $i$ and $j$ be

$$\sigma_2^{\{i,j\}} = \sigma_2 + \delta^{\{i,j\}}$$

where $\delta^{\{i,j\}} \sim \mathcal{N}(0, v_2)$ is a normally-distributed random variable with mean 0 and variance $v_2$, and we let the protection against the alleles $k \in \{a, b\}$ and $K \in \{A, B\}$ be

$$\sigma_1^{\{k\}} = \sigma_1 + \delta^{\{k\}} \quad \text{and} \quad \sigma_1^{\{K\}} = \sigma_1 - \delta^{\{k\}}$$

where $\delta^{\{k\}} \sim \mathcal{N}(0, v_1)$ is a normally-distributed random variable with mean 0 and variance $v_1$. We adjust the immune protection of allele $k$ and $K$ by $\pm$ the same amount so that fitness variation between alleles at the same locus does not cancel on average.

If there is no variation in protection against combinations of alleles, that is, $\sigma_2^{\{i,j\}} = \sigma_2$, then we can measure the total deviation from symmetry for a randomly generated set of $\sigma_1^{\{k\}}$ as

$$\sum_{K=A,B} |\sigma_1^{\{K\}} - \sigma_1^{\{k\}}|.$$

The logic here is that when $\sigma_1^{\{K\}} = \sigma_1^{\{k\}}$, there is symmetry between alleles; as the difference grows, so too will the deviation. Likewise, if there is no variation in protection against single alleles, that is, $\sigma_1^{\{i\}} = \sigma_1$, then we can measure the total deviation from symmetry for a randomly generated set of $\sigma_2^{\{i,j\}}$ using strain $ab$ as reference,

$$\sum_{ij} |\sigma_2^{\{i,j\}} - \sigma_2^{\{a,b\}}|.$$

Figures S5 and S6 show the outcome of numerical simulations using randomly chosen values of $R_0$, $\theta$, $\sigma_1^{\{i\}}$, and $\sigma_2^{\{i,j\}}$. Each simulation was run until an equilibrium was reached, which was determined by the condition

$$|\max_{t \in t_x} D(t) - \min_{t \in t_x} D(t)| < 10^{-5},$$

where $t_x$ is the last $10^4$ time units. Two notable implications emerge from the simulations. First, for all simulations, the system eventually converges to an equilibrium; sustained evolutionary oscillations were not possible. This suggests that our prediction that the necessary condition for evolutionary oscillations is a divergence between homologous immunity and full antigenic is robust to variation in immune protections against alleles and strains. Note, however, due to the variation between strains in intrinsic fitness, stable PSS no longer needs to correspond to $D' = 1$; it is possible that $w_{AB} = 0$ for some intermediate value of $D'$. Second, the sign of $\bar{\sigma}_{\mathcal{A}}$ becomes an increasingly worse predictor of whether PSS occurs as the variation in immune protections against alleles and strains increases. This is particularly true when (1) $\bar{\sigma}_{\mathcal{A}} < 0$ as compared to $\bar{\sigma}_{\mathcal{A}} > 0$ or (2) $\bar{\sigma}_{\mathcal{A}} = 0$. In case (1), individual variation in the fitness of alleles and strains creates additional epistatic pressure favouring individual strains, which is capable of building up LD. In case (2), even if $\bar{\sigma}_{\mathcal{A}} = 0$, this need not mean that epistasis is zero.

# S7 The link between strain frequencies and the immunological landscape

As discussed in the main text, various epidemiological assumptions can disrupt the correlation between which set of strains is overrepresented in the population, and which set of strains has more hosts with immunity to it (so $D(t)$ and $\mathcal{A}_\Delta(t)$, $\mathcal{S}_\Delta(t)$ do not share sign).

Here we explore two ways in which this disruption can occur; in both we assume for simplicity that full antigenic immunity is equivalent to homologous immunity, $\sigma_h = \sigma_2$. In the first example, suppose uninfected hosts lose immunity to each antigen at a per-capita rate of $\omega$. For example, an uninfected host with immunity to antigens $\{a, A, b\}$ becomes immune to antigens $\{a, A\}$ at a per-capita rate $\omega$. In the second example, suppose that following clearance of infection, hosts gain immunity to a single, randomly selected antigen with probability $p$, and both antigens with probability $1 - p$.

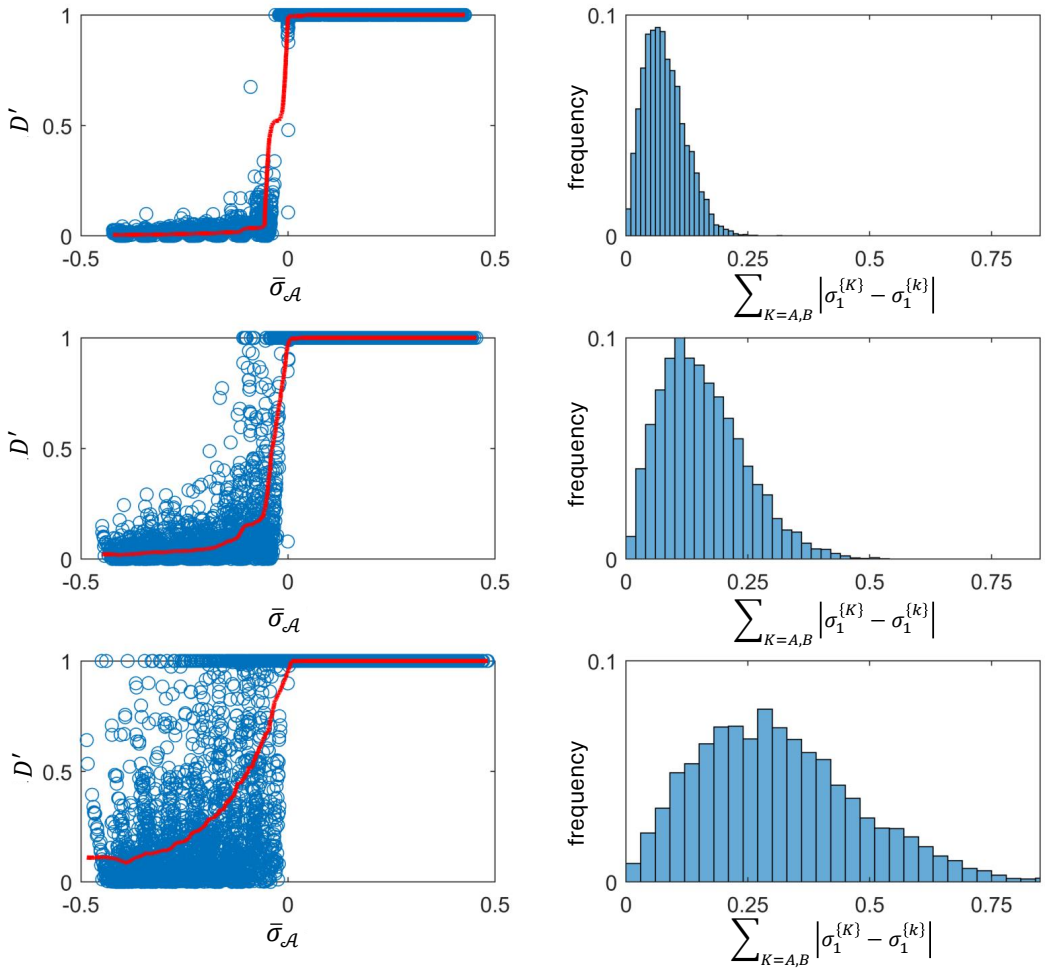


Figure S5: **Variation in immune protections against alleles and the predictive ability of $\sigma_{\mathcal{A}}$.** Here, we vary the value of $\sigma_1$ for each allele, specifically, $\sigma_1^{\{k\}} = \sigma_1 + \delta^{\{k\}}$ and $\sigma_1^{\{K\}} = \sigma_1 - \delta^{\{k\}}$, where $\delta^{\{k\}} \sim \mathcal{N}(0, v_1)$, that is, $\delta^{\{k\}}$ is a normally-distributed random variable with mean 0 and variance $v_1$. For each row we vary the value of $v_1$: for the first row, $v_1 = 0.025$; for the second row, $v_1 = 0.05$; for the third row, $v_1 = 0.1$. We then loop over 15 evenly-spaced values of $\sigma_1 \in [0.4, 0.8]$. For each value of $\sigma_1$, we perform 400 simulations (so each panel is $6 \times 10^3$ simulations), where in each simulation we randomly set allelic fitness variation (i.e., the values of $\delta^{\{k\}}$), as well as the values of $\theta$ and $R_0$; these are uniform random variables in the intervals $\theta \in [10^{-4}, 0.5]$ and $R_0 \in [1.5, 18]$. The first column shows how the sign of $\bar{\sigma}_{\mathcal{A}}$ affects the equilibrium value of $D'$. The second column shows the distribution of fitness variation used in the simulations. The red line is the rolling average of $D'$ using a window of size 400. All simulations assume $\sigma_0 = 1$, $\sigma_2 = 0.2$, $b = d = 0.01$, $\alpha = 0$ and $\rho = 10^{-8}$.

Since homologous immunity is of the same strength as full antigenic immunity, suppose the immune history $x$ for these examples only tracks the unique antigens against which a host

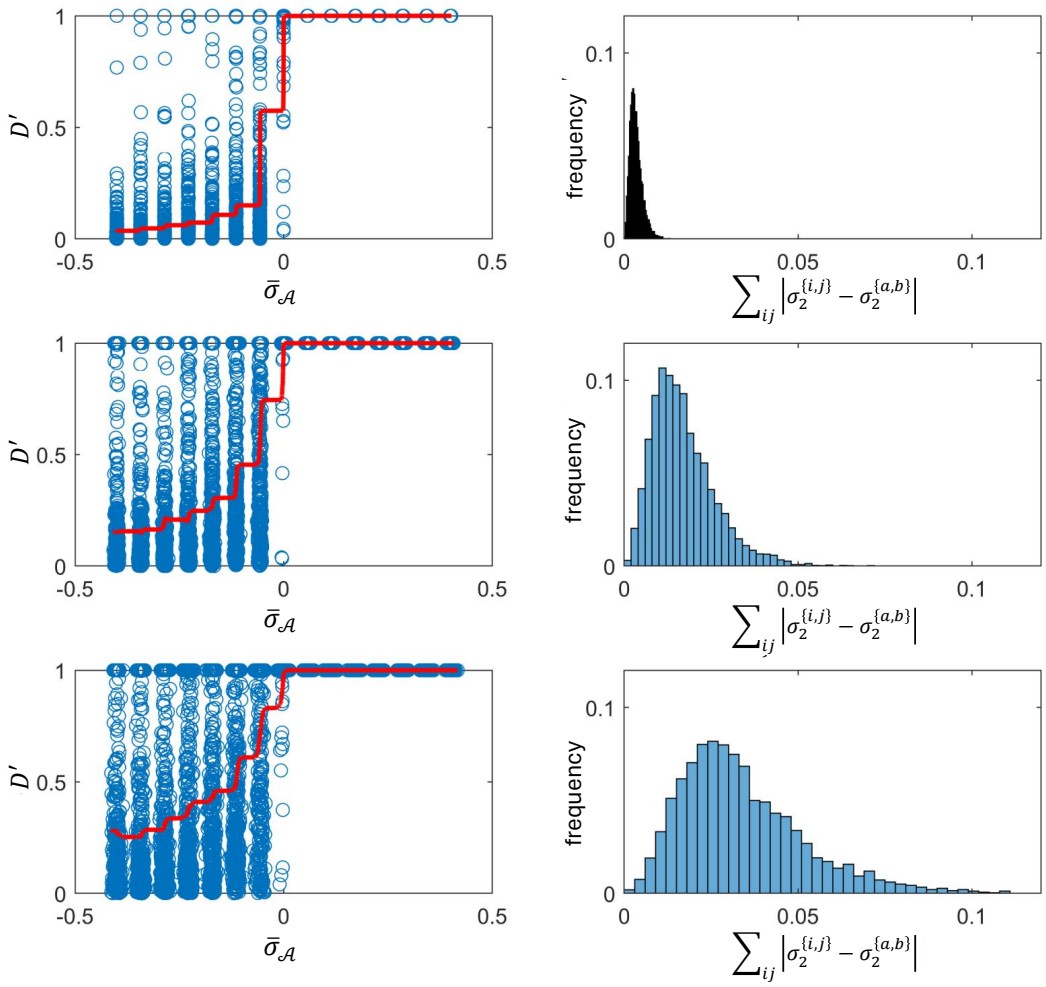


Figure S6: **Variation in immune protections against combinations of alleles and the predictive ability of $\sigma_{\mathcal{A}}$.** Here we vary the value of $\sigma_2$ for each allele, specifically, $\sigma_2^{\{i,j\}} = \sigma_2 + \delta^{\{i,j\}}$, where $\delta^{\{i,j\}} \sim \mathcal{N}(0, v_2)$. For each row we vary the value of $v_2$: for the first row, $v_2 = 0.001$; for the second row, $v_2 = 0.005$; for the third row, $v_2 = 0.01$. We then loop over 15 evenly-spaced values of $\sigma_1 \in [0.4, 0.8]$. For each value of $\sigma_1$, we perform 400 simulations (so each panel is $6 \times 10^3$ simulations). For each simulation we randomly set the fitness variation for combinations of alleles (i.e., the values of $\delta^{\{i,j\}}$), as well as the values of $\theta$ and $R_0$; these are uniform random variables in the intervals $\theta \in [10^{-4}, 0.5]$ and $R_0 \in [1.5, 18]$. The first column shows how the sign of $\bar{\sigma}_{\mathcal{A}}$ affects the equilibrium value of $D'$, and the second column shows the distribution of fitness variation used in the simulations. The red line is the rolling average of $D'$ using a window of size 400. All simulations assume $\sigma_0 = 1$, $\sigma_2 = 0.2$, $b = d = 0.01$, $\alpha = 0$ and $\rho = 10^{-8}$.

has immunity (e.g., in the previous formulation, $x = \{ab, Ab\}$, would become $x = \{a, A, b\}$. The two assumptions above leave the equations for the $I_x^{ij}$ unchanged, but the net rate at

which immune history $x$ is gained and lost is

$$\Gamma_x = \frac{1}{2}\sum_g \Big(2\sum_{z:g\cup z=x}(1-p)\mu_z^g I_z^g + \sum_{z:g\cup(z\setminus z_1)=x} p\mu_z^g I_z^g + \sum_{z:g\cup(z\setminus z_2)=x} p\mu_z^g I_z^g\Big) + \omega\Big(\sum_{z:(z\setminus z_i)=x} S_z - \sum_{i\in x} S_x\Big) \tag{S52}$$

where $(z\setminus z_i)$ indicates the genotype $z$ without the $i$-th locus; e.g., if $z = ab$, then $(z\setminus z_1) = b$.

When we simulate the dynamics (Fig. S7 and Fig. S8), we observe two differences from the previous predictions.

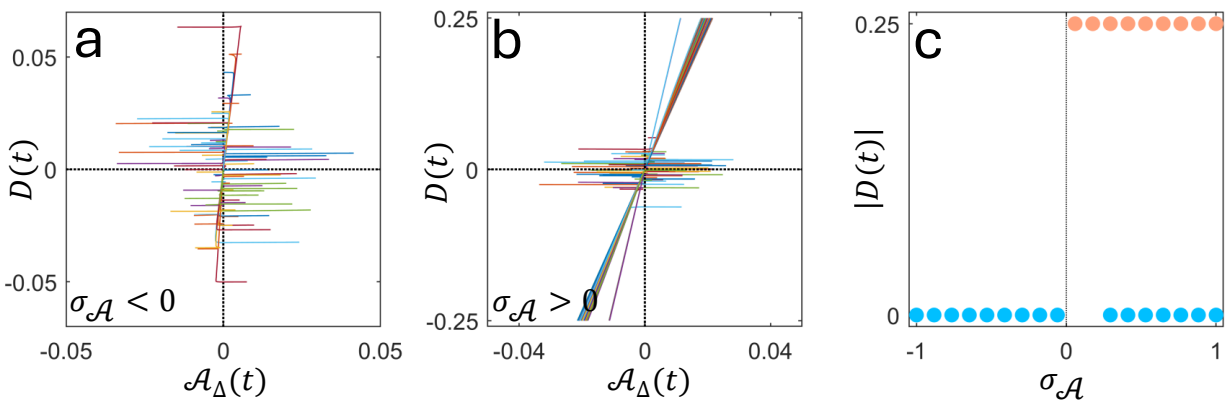


Figure S7: **Disrupting the link between strain frequencies and the immunological landscape through the loss of immunity**. In this example, host lose immunity to different antigens at independent rates $\omega = 0.2$. This can disrupt the expected coupling between $D(t)$ and $\mathcal{A}_\Delta(t)$ (panel **a**,**b**), reversing the predictions about the effects of the sign of $\sigma_{\mathcal{A}}$ for the formation of PSS or no PSS (panel **c**), depending upon the value of $\theta$ and $R_0$. Each circle in panel **c** represents $10^2$ simulations in which we loop over 10 evenly spaced values of $R_0 \in [1.5, 10]$ and 10 evenly spaced values (on log-scale) of $\theta \in [10^{-4}, 0.39]$. All simulations assume $b = d = 0.01$, $\rho = 10^{-8}$, $\sigma_0 = 1$, $\sigma_2 = \sigma_h = 0$, and $\alpha = 0$.

(1) If immunity wanes in an antigen-by-antigen fashion (i.e., $\omega > 0$), then strain structure may not form when $\sigma_{\mathcal{A}}$ becomes large (Fig. S7).

The logic here is as follows. When $\sigma_{\mathcal{A}}$ is large, $\sigma_1, \sigma_2 \to 0$. Consequently, hosts with immunity to one strain (say $AB$) can be infected by, at most, the other strain in the set of strains $\{ab, AB\}$, that is, strain $ab$. When this occurs, this produces hosts with immunity to all four antigens and thus full protection from infection. In order for these hosts to then become susceptible again, they have to lose immunity to one antigen at each locus. Since immunity to each antigen is lost randomly, such hosts are equally likely to produce hosts that are susceptible to members of one set of strains or the other, irrespective of which set of strains is overrepresented. However, as these hosts become susceptible to infection, the set of overrepresented strains will infect and deplete its available hosts faster. This creates an excess of hosts with immunity to the set of underrepresented strains, meaning that $\mathcal{A}_\Delta(t)$ can have the opposite sign to LD. Since $\sigma_{\mathcal{A}} \to \sigma_0 > 0$, epistasis in fitness will then also have the opposite sign to population LD. Thus, strain structure may not form, even if $\sigma_{\mathcal{A}}$ is large. Note that in Figure S7, these predictions are sensitive to the epidemiological parameters $\theta$ and $R_0$, because these

parameters shape the magnitude of the impact of a fixed rate of waning immunity (in all simulations in Figure S7, $\omega = 0.2$).

(2) Suppose that following infection, hosts acquire immunity to both antigens of a strain with probability $1-p$, or only a single antigen with probability $p$ (each antigen is chosen with probability 1/2). If $p$ is sufficiently large, strain structure occurs only when $\sigma_{\mathcal{A}} < 0$ (Fig. S8).

The logic here is as follows. Suppose that the set of strains $\{ab, AB\}$ is overrepresented in the population ($D(t) > 0$). Here, we would expect naive hosts to be more likely to become infected by either strains $ab$ or $AB$. Following infection by strain $AB$ (say), the host is then most likely to become infected by strain $ab$ (since it is more abundant). However, each infection can provide immunity to at most a single antigen (since $p \to 1$). Thus, when the two randomly selected antigens are at different loci, following a secondary infection, the host will then have immunity to either the antigens $A, b$ or $a, B$ from the set of underrepresented strains, $\{Ab, aB\}$. Consequently, $\mathcal{A}_{\Delta}(t)$ will have the opposite sign to $D(t)$ and more hosts will have immunity to the set of underrepresented strains than to the set of overrepresented strains. If $\sigma_{\mathcal{A}} < 0$, this provides an advantage to the set of overrepresented strains, such that PSS can occur.

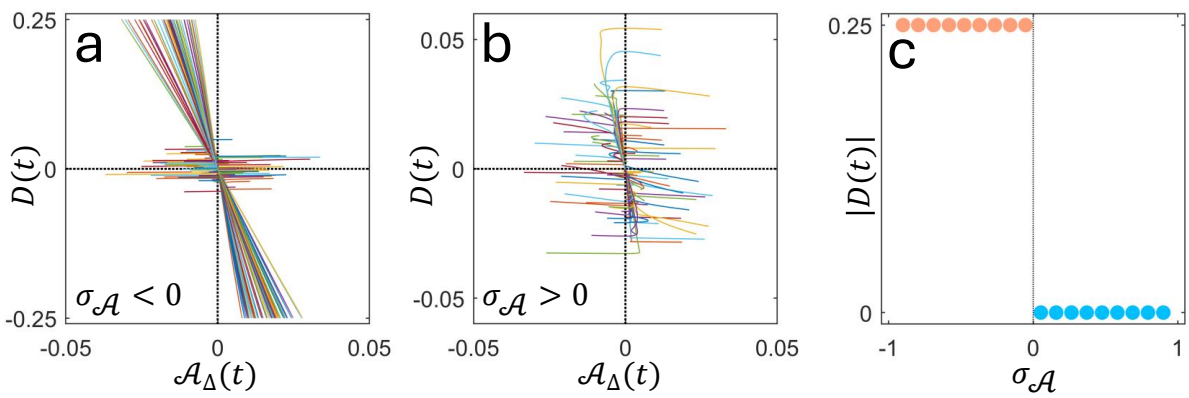


Figure S8: **Disrupting the link between strain frequencies and the immunological landscape through the gain of immunity**. In this example, following infection, hosts acquire immunity to a single, randomly chosen antigen (i.e., $p = 1$). This disrupts the expected coupling between $D(t)$ and $\mathcal{A}_{\Delta}(t)$ (panel **a**,**b**), reversing the predictions about the effects of the sign of $\sigma_{\mathcal{A}}$ for the formation of PSS or no PSS (panel **c**). Each circle in panel **c** represents $10^2$ simulations in which we loop over 10 evenly spaced values of $R_0 \in [1.5, 10]$ and 10 evenly spaced values (on log-scale) of $\theta \in [10^{-4}, 0.39]$. All simulations assume $b = d = 0.01$, $\rho = 10^{-8}$, $\sigma_0 = 1$, $\sigma_2 = \sigma_h = 0$, and $\alpha = 0$.

Note that these predictions are sensitive to how immune protections are modelled. In particular, when immunity reduces susceptibility, a host will typically only gain immunity to the infecting strain when it is successfully infected [15]. If, instead, immune protections affect, for example, transmissibility, even an 'unsuccessful' infection (i.e., an infection that is incapable of forward transmission) will update the host's immunity. These differences are

not expected to have a large effect, at least in models with simple epidemiological assumptions [15]. However, as the underlying epidemiological model becomes more complex, the evolutionary dynamics can become more sensitive to the type of immune protection, as the above examples illustrate.

# S8 Variation in allelic diversity across loci

The model featured in Section S3 assumes that there is an equal number of alleles at each locus. This symmetry allowed for the separation of fast and slow timescales, as all alleles rapidly converged to the same frequency, at which point we could trace the evolution of LD without needing to account for further allele frequency change. Here, we consider what happens if at one locus there are two alleles $i \in \{A_1, A_2\}$, and at the other there are three, $j \in \{B_1, B_2, B_3\}$. We obtain six possible strains,

$$ij \in \{A_1B_1, A_1B_2, A_2B_3, A_2B_1, A_2B_2, A_2B_3\}.$$

Without loss of generality, we choose strain $A_1B_1$ as the reference strain. Here, the evolutionary variables are the frequency of alleles $A_2$ and $B_j$ at time $t$, which are

$$p_{A_2}(t) = \frac{\sum_j I^{A_2B_j}(t)}{I(t)}, \quad p_{B_j}(t) = \frac{I^{A_1B_j}(t) + I^{A_2B_j}(t)}{I(t)} \; j = 2, 3,$$

and the pairwise LD between alleles $A_2$ and $B_j$,

$$D_{A_2B_j}(t) = \frac{I^{A_2B_j}(t)}{I(t)} - p_{A_2}(t)p_{B_j}(t), \quad j = 2, 3.$$

Next, the three additive selection coefficients are defined as

$$w_A \equiv r_{A_2B_1} - r_{A_1B_1} \quad \text{and} \quad w_{B_j} \equiv r_{A_1B_j} - r_{A_1B_1}, \; j = 2, 3, \tag{S53}$$

and the two terms specifying epistasis in fitness are

$$w_{A_2B_j} \equiv r_{A_2B_j} + r_{A_1B_1} - r_{A_1B_j} - r_{A_2B_1}, \quad j = 2, 3,$$

that is,

$$\begin{aligned} w_{A_2B_j} = \beta\sigma_{\mathcal{A}}(\mathcal{A}_{\{A_2,B_j\}} + \mathcal{A}_{\{A_1,B_1\}} - \mathcal{A}_{\{A_2,B_1\}} - \mathcal{A}_{\{A_1,B_j\}}) \\ - \beta\sigma_{\mathcal{S}}(\mathcal{S}_{A_2B_j} + \mathcal{S}_{A_1B_1} - \mathcal{S}_{A_2B_1} - \mathcal{S}_{A_1B_j}). \end{aligned} \tag{S54}$$

Here, as before, $\mathcal{A}_{\{i\}}$ and $\mathcal{A}_{\{i,j\}}$ denote the density of all uninfected hosts with immunity to at least the antigen $i$ and antigens $i$ and $j$, respectively, and $\mathcal{S}_{ij}$ denotes the density of uninfected hosts with homologous immunity to at least strain $ij$.

It is then straightforward to write out the equations describing the evolutionary dynamics. Here we focus on interpreting the simulated dynamics. The key is to recognize that additive selection favours the frequency of each allele at a given locus to be equal, that is, $p_{A_2} = 1/2$, $p_{B_j} = 1/3$, whereas epistatic selection either favours LD to increase or decrease in magnitude.

This depends upon whether $D_{ij}(t)$ and $w_{ij}$ are of the same or opposite signs. In the two-allele, two-locus model of Section S3, allele frequencies could be equal *and* LD could assume any value (depending upon epistatic selection). As a consequence, two timescales emerged: a fast timescale on which allele frequencies were driven to equilibrium ($p_k \to 1/2$), and a slow timescale on which LD evolved (see equation (S25)). When there are more alleles at one locus than the other, this need not be true. In this case, competition between epistatic and allelic selection can induce complex dynamics.

To illustrate this, we first consider the case in which $\sigma_{\mathcal{S}} = 0$. As was true for the model of Section S3, there is a single epidemiological feedback on epistasis. Thus, oscillatory evolutionary dynamics are not possible and the population will converge to an evolutionary equilibrium. If $\sigma_{\mathcal{A}} < 0$, then epistasis favours linkage equilibrium, $D_{A_2B_i} = 0$, whereas additive selection favours $p_{A_i} = 1/2$ and $p_{B_j} = 1/3$. Since these can be simultaneously achieved, there is no conflict between epistasis and additive selection (Fig. S9). However, when $\sigma_{\mathcal{A}} > 0$, epistasis favours pairwise LD evolving towards $D^*$. However, this cannot be achieved for all three $\{A_2, B_j\}$ pairs if the population is in a state favoured by additive selection (i.e., $p_{A_i} = 1/2$, $p_{B_j} = 1/3$). This creates a conflict between epistasis and additive selection, and so the evolutionary outcome depends upon their relative strengths. If $\sigma_{\mathcal{A}}$ is positive but small, epistasis will be relatively weak compared to additive selection. This will cause the allele frequencies to diverge from their expectation, but $D_{A_2B_j}(t) = D^*$ for two of the three $\{A_2, B_j\}$ pairs; the third pair will be in linkage equilibrium (so two genotypes will have been eliminated from the population). As $\sigma_{\mathcal{A}}$ increases, the allele frequencies diverge further from their allele-frequency based optimum, that is, the distance (measured here using the Euclidean norm)

$$\xi_p = \sqrt{\sum_i (p_{A_i} - 1/2)^2 + \sum_j (p_{B_j} - 1/3)^2}$$

becomes increasingly large (Fig. S9), while the pairwise LD becomes closer to $D^*$, that is,

$$\xi_D = \sqrt{\sum_j (D_{A_2B_j} - D^*)^2}$$

becomes increasingly small (Fig. S9). Eventually, when $\sigma_{\mathcal{A}}$ is sufficiently large, $p_{A_i} = 1/2$, $p_{B_j} = p_{B_k} = 1/4$, $p_{B_\ell} = 1/2$ for $j \neq k \neq \ell = 1, 2, 3$, and $D_{A_2B_j}(t) = D^*$ for each pairwise combination (so three genotypes will have been eliminated from the population) (Fig. S9). Thus, the sign of $\sigma_{\mathcal{A}}$ still determines whether PSS forms or not, but now its strength additionally affects the amount of PSS observed. Importantly, however, even if $\sigma_{\mathcal{A}}$ is maximized, the PSS observed will not be equivalent to the DSS of Gupta et al [6, 12].

Second, suppose that $\sigma_{\mathcal{S}} > 0$. As was the case for two diallelic loci, evolutionary oscillations are possible. But there is an important difference: because there is no longer a fast and a slow timescale, the allele frequencies will *not* reach an equilibrium. Both LD and the allele frequencies will oscillate (Fig. S10).

# S9 More than two loci

What happens as the number of loci increases? To illustrate this, we suppose there are three diallelic loci, $\{a, A\}$, $\{b, B\}$ and $\{c, C\}$. Without loss of generality, choose strain *abc*

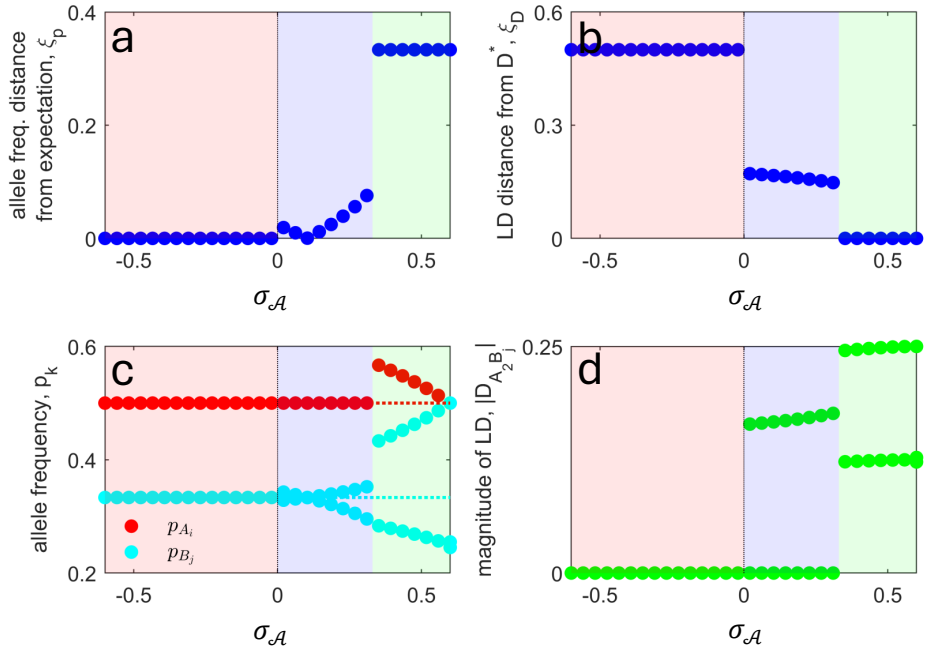


Figure S9: **Consequences of variation in allelic diversity across loci**. An unequal number of alleles across loci can produce conflict between selection on alleles (additive selection coefficients) and on strains (epistasis) because allele frequencies cannot be at their expected values ($p_{A_i} = 1/2$, $p_{B_j} = 1/3$) *and* the LD between each $\{A_2, B_j\}$ pair satisfy $D_{A_2B_j} = D^*$. If $\sigma_{\mathcal{A}} < 0$ (red shaded region), there is no conflict because epistasis does not favour strain structure. If $\sigma_{\mathcal{A}} > 0$ but weak (blue shaded region), although PSS is favoured, selection on alleles is stronger than on strains and so there is no LD between one of the $\{A_2, B_j\}$ pairs. As $\sigma_{\mathcal{A}}$ increases further (green shaded region), so does the strength of epistasis; this pushes all three $D_{A_2B_j} \to D^*$ (panel **b**) and allele frequencies away from their expectation (panel **a**). Parameter values: $\beta = 1$, $\mu = 0.4$, $b = d = 0.01$, $\rho = 10^{-8}$, $\sigma_0 = 1$, $\sigma_2 = \sigma_h = 0.4$, and $\sigma_1 \in [\sigma_0, \sigma_2]$.

as the reference strain. Here, the evolutionary dynamics consist of the dynamics of the allele frequencies, $p_k(t)$, $k \in \{A, B, C\}$, e.g.,

$$p_A(t) \equiv \frac{I^{Abc}(t) + I^{ABc}(t) + I^{AbC}(t) + I^{ABC}(t)}{I(t)},$$

three pairwise LD terms, $D_{ij}(t)$, $ij \in \{AB, AC, BC\}$, e.g.,

$$D_{AB}(t) \equiv \frac{I^{ABc}(t) + I^{ABC}(t)}{I(t)} - p_A(t)p_B(t),$$

and triple LD,

$$D_{ABC}(t) \equiv p_{ABC}(t) - p_A(t)p_B(t)p_C(t) - p_A(t)D_{BC}(t) - p_B(t)D_{AC}(t) - p_C(t)D_{AB}(t).$$

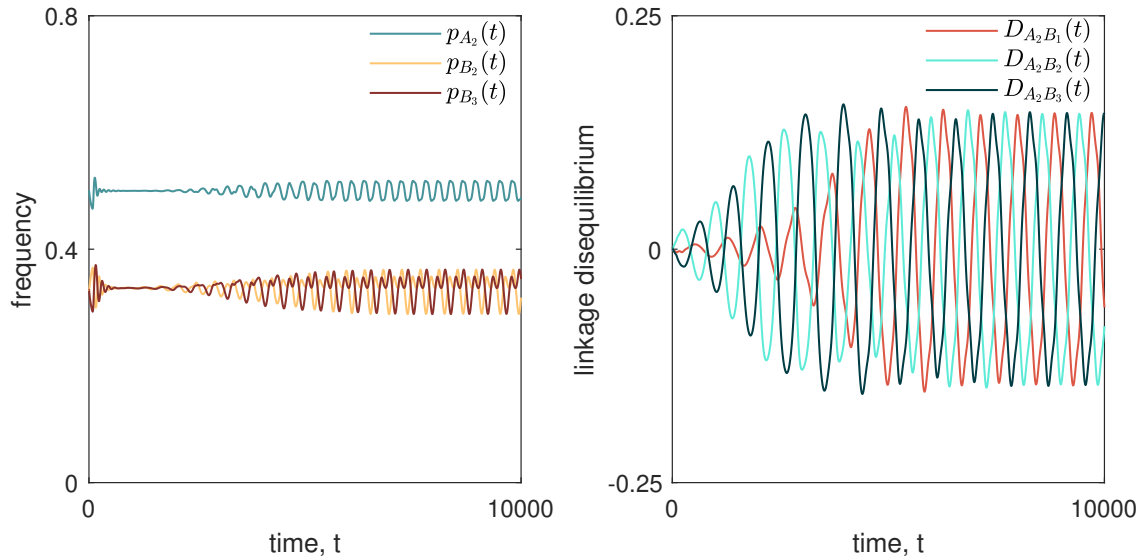


Figure S10: **Conflict between selection on alleles and strains generates complex evolutionary oscillations**. With variation in allelic diversity across loci, there is no longer a separation between fast and slow timescales. This is because allele frequencies cannot be at their expectation *and* all pairwise LD terms reach $D^*$. Consequently, when evolutionary oscillations occur, these involve both allele frequencies and LD. Parameter values: $\beta = 1$, $\mu = 0.4$, $\alpha = 0$, $b = d = 0.01$, $\rho = 0$, $\sigma_0 = 1$, $\sigma_1 = 0.3$, $\sigma_2 = 0.15$, and $\sigma_h = 0$.

Limiting our attention to immune protections that reduce susceptibility, from the model of Section S1, the per-capita growth rate of strain $ijk$ is

$$r_{ijk} = \beta\big(\sigma_0\mathcal{S} - (\sigma_0 - \sigma_1)\sum_{\ell=i,j,k}\mathcal{A}_{\{\ell\}} + \sigma_{\mathcal{A}}^{(1)}\sum_{\ell_1\ell_2\in\{ij,ik,jk\}}\mathcal{A}_{\{\ell_1,\ell_2\}}\big) - \beta\big((\sigma_{\mathcal{A}}^{(1)} - \sigma_{\mathcal{A}}^{(2)})\mathcal{A}_{\{i,j,k\}} + \sigma_{\mathcal{S}}\mathcal{S}_{ijk}\big) - \mu - \alpha - d, \quad \text{(S55)}$$

where $\sigma_{\mathcal{S}} \equiv \sigma_3 - \sigma_h$, and

$$\sigma_{\mathcal{A}}^{(n)} \equiv \sigma_{n-1} - 2\sigma_n + \sigma_{n+1}, \quad n = 1, 2.$$

Thus, $\sigma_{\mathcal{A}}^{(n)}$ measures whether having antigenic immunity to $n$ antigens of the infecting strain provides more, $\sigma_{\mathcal{A}}^{(n)} > 0$, or less, $\sigma_{\mathcal{A}}^{(n)} < 0$, protection than would be expected based upon the protection offered by having immunity to $n-1$ versus $n+1$ antigens.

Using these per-capita growth rates, the additive selection coefficient for allele $A$ (and similarly for alleles $B$ and $C$) is

$$w_A = r_{Abc} - r_{abc} = -\beta(\sigma_0 - \sigma_1)(\mathcal{A}_{\{A\}} - \mathcal{A}_{\{a\}}) + \beta\sigma_{\mathcal{A}}^{(1)}(\mathcal{A}_{\{A,b\}} + \mathcal{A}_{\{A,c\}} - \mathcal{A}_{\{a,b\}} - \mathcal{A}_{\{a,c\}}) + \beta(\sigma_{\mathcal{A}}^{(2)} - \sigma_{\mathcal{A}}^{(1)})(\mathcal{A}_{\{A,b,c\}} - \mathcal{A}_{\{a,b,c\}}) - \beta\sigma_{\mathcal{S}}(\mathcal{S}_{Abc} - \mathcal{S}_{abc}). \quad \text{(S56)}$$

Pairwise epistasis for alleles $ij \in \{AB, AC, BC\}$ is defined as before when using the allele at the third locus as the reference allele, $\ell \in \{a, b, c\}$; for example,

$$w_{AB} = r_{ABc} + r_{abc} - r_{Abc} - r_{aBc}.$$

Using our per-capita growth rates, this can be written as

$$w_{ij} = \beta\sigma_{\mathcal{A}}^{(1)}\mathcal{A}_{\Delta}^{(ij)} + \beta(\sigma_{\mathcal{A}}^{(2)} - \sigma_{\mathcal{A}}^{(1)})\mathcal{A}_{\Delta}^{(ij)/\ell} - \beta\sigma_{\mathcal{S}}\mathcal{S}_{\Delta}^{(ij)/\ell}. \tag{S57}$$

$w_{ij}$ includes the following composite variables:

- $\mathcal{A}_{\Delta}^{(ij)}$ indicates whether more uninfected hosts have immunity to the antigen combinations $ij$ and/or $\hat{i}\hat{j}$ (i.e., $\mathcal{A}_{\Delta}^{(ij)} > 0$), or immunity to the antigen combinations $\hat{i}j$ and/or $i\hat{j}$ (i.e., $\mathcal{A}_{\Delta}^{(ij)} < 0$). Note that this is irrespective of genetic background, and that $\hat{i}$ indicates the other allele at the indicated locus, e.g., if $i = A$, then $\hat{i} = a$. For example,

$$\mathcal{A}_{\Delta}^{(AB)} = \mathcal{A}_{\{A,B\}} + \mathcal{A}_{\{a,b\}} - \mathcal{A}_{\{A,b\}} - \mathcal{A}_{\{a,B\}}. \tag{S58}$$

- $\mathcal{A}_{\Delta}^{(ij)/\ell}$ indicates whether more uninfected hosts with immunity to antigen $\ell$ also have immunity to the antigen combinations $ij$ and/or $\hat{i}\hat{j}$ (i.e., $\mathcal{A}_{\Delta}^{(ij)/\ell} > 0$), or immunity to the antigen combinations $\hat{i}j$ and/or $i\hat{j}$ (i.e., $\mathcal{A}_{\Delta}^{(ij)/\ell} < 0$). For example,

$$\mathcal{A}_{\Delta}^{(AB)/c} = \mathcal{A}_{\{A,B,c\}} + \mathcal{A}_{\{a,b,c\}} - \mathcal{A}_{\{A,b,c\}} - \mathcal{A}_{\{a,B,c\}}. \tag{S59}$$

- $\mathcal{S}_{\Delta}^{(ij)/\ell}$ indicates whether more uninfected hosts have homologous immunity to members of the set of strains $\{ij\ell, \hat{i}\hat{j}\ell\}$ (i.e., $\mathcal{S}_{\Delta}^{(ij)/\ell} > 0$), or homologous immunity to members of the set of strains $\{i\hat{j}\ell, \hat{i}j\ell\}$ (i.e., $\mathcal{S}_{\Delta}^{(ij)/\ell} < 0$). For example,

$$\mathcal{S}_{\Delta}^{(AB)/c} = \mathcal{S}_{ABc} + \mathcal{S}_{abc} - \mathcal{S}_{Abc} - \mathcal{S}_{aBc}. \tag{S60}$$

Finally, when there are three loci, we need to account for triple epistasis in fitness, denoted $w_{ABC}$. Since by definition, $r_{ABC} = r_{abc} + \sum_k w_k + \sum_{ij} w_{ij} + w_{ABC}$, this is

$$w_{ABC} = r_{ABC} + r_{Abc} + r_{aBc} + r_{abC} - r_{abc} - r_{ABc} - r_{AbC} - r_{aBC} \tag{S61}$$

$$= \beta(\sigma_{\mathcal{A}}^{(2)} - \sigma_{\mathcal{A}}^{(1)})\mathcal{A}_{\Delta}^{(2)} - \beta\sigma_{\mathcal{S}}\mathcal{S}_{\Delta}^{(2)}. \tag{S62}$$

$w_{ABC}$ includes the following composite variables:

- $\mathcal{A}_{\Delta}^{(2)}$ indicates if more hosts have full antigenic immunity to strains belonging to the set $\{ABC, Abc, aBc, abC\}$ (i.e., $\mathcal{A}_{\Delta}^{(2)} > 0$), or strains belonging to the set $\{abc, ABc, AbC, aBC\}$ (i.e., $\mathcal{A}_{\Delta}^{(2)} < 0$). That is,

$$\mathcal{A}_{\Delta}^{(2)} = \mathcal{A}_{\{A,B,C\}} + \mathcal{A}_{\{A,b,c\}} + \mathcal{A}_{\{a,B,c\}} + \mathcal{A}_{\{a,b,C\}} - \mathcal{A}_{\{A,B,c\}} - \mathcal{A}_{\{A,b,C\}} - \mathcal{A}_{\{a,B,C\}} - \mathcal{A}_{\{a,b,c\}}. \tag{S63}$$

- $\mathcal{S}_{\Delta}^{(2)}$ indicates if more hosts have homologous immunity to strains belonging to the set $\{ABC, Abc, aBc, abC\}$ (i.e., $\mathcal{S}_{\Delta}^{(2)} > 0$), or strains belonging to the set $\{abc, ABc, AbC, aBC\}$ (i.e., $\mathcal{A}_{\Delta}^{(2)} < 0$). That is,

$$\mathcal{S}_{\Delta}^{(2)} = \mathcal{S}_{ABC} + \mathcal{S}_{Abc} + \mathcal{S}_{aBc} + \mathcal{S}_{abC} - \mathcal{S}_{ABc} - \mathcal{S}_{AbC} - \mathcal{S}_{aBC} - \mathcal{S}_{abc}. \tag{S64}$$

Note that the sets of strains $\{ABC, Abc, aBc, abC\}$ and $\{abc, ABc, AbC, aBC\}$ are characterized by each member having no more than one antigen in common with the other members.

This allows us to write down the equations describing the evolutionary dynamics. In what follows, we use these equations to extract general qualitative predictions. Note that our emphasis is on biological interpretability, so we do not attempt to rigorously characterize the dynamics.

## S9.1 Fast and slow timescales, and different types of strain structure

We first observe that in the three-diallelic-locus model, there is also a fast timescale on which $p_k(t) \to 1/2$, and a slow timescale on which pairwise and triple-LD (strain structure) evolve. After the fast transient dynamics, on the slow timescale pairwise LD becomes

$$\begin{aligned} D_{AB}(t) &\approx \frac{p_{ABc}(t) + p_{abc}(t) - p_{aBC}(t) - p_{AbC}(t)}{2}, \\ D_{AC}(t) &\approx \frac{p_{AbC}(t) + p_{abc}(t) - p_{aBC}(t) - p_{ABc}(t)}{2}, \\ D_{BC}(t) &\approx \frac{p_{aBC}(t) + p_{abc}(t) - p_{AbC}(t) - p_{ABc}(t)}{2}, \end{aligned} \tag{S65}$$

and triple LD becomes

$$D_{ABC}(t) \approx \frac{p_{ABC}(t) - p_{abc}(t)}{2}. \tag{S66}$$

Inspection of these quantities allows us to identify two distinct ways in which strain structure can emerge:

(1) When a set of strains whose members share no more than one allele with any other is overrepresented, that is, either set

$$\{ABC, Abc, aBc, abC\} \quad \text{or} \quad \{abc, aBC, ABc, AbC\}$$

is overrepresented. This type of strain structure is driven by the build-up and maintenance of triple LD, and the removal of pairwise LD. Here, when strain structure is maximized (one set is competitively excluded), $D_{ABC}(t) = \pm 1/8$ while $D_{ij}(t) = 0$.

If we suppose there is no pairwise LD ($D_{ij}(t) = 0$), then, on the slow timescale

$$\frac{\mathrm{d}D_{ABC}}{\mathrm{d}t} \approx w_{ABC}\left(\frac{1}{8} - D_{ABC}\right)\left(\frac{1}{8} + D_{ABC}\right). \tag{S67}$$

By inspection of equation (S67) we observe that in order for triple LD to build up and be maintained, triple LD ($D_{ABC}$) and triple epistasis ($w_{ABC}$) must share sign.

(2) When one of the four sets of strains whose members have no alleles in common is overrepresented, that is, one of the sets

$$\{ABC, abc\}, \quad \{Abc, aBC\}, \quad \{aBc, AbC\}, \quad \text{or} \quad \{abC, ABc\}, \tag{S68}$$

is overrepresented. This type of strain structure is driven by the build-up and maintenance of pairwise LD, and the removal of triple LD. Here, when strain structure is maximized (i.e., three of the sets of strains in (S68) are competitively excluded), then $D_{ij}(t) = \pm 1/4$, and $D_{ABC}(t) = 0$.

If we suppose that there is no triple LD ($D_{ABC}(t) = 0$), then on the slow timescale

$$\frac{\mathrm{d}D_{ij}}{\mathrm{d}t} \approx \left(w_{ij} + \frac{w_{ABC}}{2}\right)\left(\frac{1}{4} - D_{ij}\right)\left(\frac{1}{4} + D_{ij}\right) + \sum_{k\ell}\left(w_{k\ell} + \frac{w_{ABC}}{2}\right)\left(\frac{D_{\hat{k}\hat{\ell}}}{4} - D_{ij}D_{k\ell}\right). \quad \text{(S69)}$$

From equation (S69) we observe that it is no longer necessary that $w_{ij}$ and $D_{ij}$ must share sign for pairwise LD to be maintained. Instead, $w_{ij} + w_{ABC}/2$ and $D_{ij}$ need to share sign. Note that

$$w_{ij} + \frac{w_{ABC}}{2} = \beta\sigma_{\mathcal{A}}^{(1)}\mathcal{A}_{\Delta}^{(ij)} + \frac{\beta}{2}(\sigma_{\mathcal{A}}^{(2)} - \sigma_{\mathcal{A}}^{(1)})(\mathcal{A}_{\Delta}^{(ij)/\ell} + \mathcal{A}_{\Delta}^{(ij)/\hat{\ell}}) - \frac{\beta}{2}\sigma_{\mathcal{S}}(\mathcal{S}_{\Delta}^{(ij)/\ell} + \mathcal{S}_{\Delta}^{(ij)/\hat{\ell}}). \quad \text{(S70)}$$

## S9.2 Shape of the function controlling cross-protection

Next, suppose $\sigma_{\mathcal{S}} = 0$. Here, we study how the $\sigma_{\mathcal{A}}^{(n)}$ affect the evolutionary dynamics. Recall that $\sigma_{\mathcal{A}}^{(n)}$ measures whether the rate of increase of cross-protection is locally saturating ($\sigma_{\mathcal{A}}^{(n)} > 0$) or accelerating ($\sigma_{\mathcal{A}}^{(n)} < 0$) at $n$.

Two curves divide up the qualitatively different evolutionary behaviour in the $(\sigma_{\mathcal{A}}^{(1)}, \sigma_{\mathcal{A}}^{(2)})$ space. In what follows, we determine these two curves; the results are presented in Figure 5 in the main text. Throughout, we assume that the set of overrepresented strains also has more hosts with antigenic immunity to its members, that is, $\mathcal{A}_{\Delta}^{(2)}$ and $D_{ABC}$ share sign, and $\mathcal{A}_{\Delta}^{(ij)}$ and $D_{ij}$ share sign. We will not consider ways in which these relationships could break down as in Section S7.

**Curve** 1: $w_{ABC} = 0$. This occurs when $\sigma_{\mathcal{A}}^{(2)} = \sigma_{\mathcal{A}}^{(1)}$, that is, the rate of increase of cross-protection is the same regardless of the number of antigens the host has immunity to. This partitions the parameter space between when triple LD can be maintained and when it cannot. In particular, since triple LD requires the signs of $w_{ABC}$ and $D_{ABC}$ to match, it can only be maintained if $\sigma_{\mathcal{A}}^{(2)} > \sigma_{\mathcal{A}}^{(1)}$.

**Curve** 2: $w_{ij} + w_{ABC}/2 = 0$. From equation (S69), this curve specifies when pairwise LD can or cannot form. Since this curve depends upon the epidemiological dynamics in addition to the parameters $\sigma_{\mathcal{A}}^{(n)}$, it is not straightforward to precisely specify the curve in $(\sigma_{\mathcal{A}}^{(1)}, \sigma_{\mathcal{A}}^{(2)})$ space. However, note that in equation (S70), we can write $\mathcal{A}_{\Delta}^{(ij)} = \mathcal{A}_{\Delta}^{(ij)/r} + \mathcal{A}_{\Delta}^{(ij)/\hat{r}} - \mathcal{A}_{\Delta}^{(ij)/r\hat{r}}$, where

$$\mathcal{A}_{\Delta}^{(ij)/r\hat{r}} = \mathcal{A}_{\{i,j,r,\hat{r}\}} + \mathcal{A}_{\{\hat{i},\hat{j},r,\hat{r}\}} - \mathcal{A}_{\{\hat{i},j,r,\hat{r}\}} - \mathcal{A}_{\{i,\hat{j},r,\hat{r}\}}.$$

Each of these terms requires hosts to be infected by different pathogen strains. Hence, if we suppose that the frequency of hosts being infected multiple times is low compared to hosts being infected once, e.g., $\mathcal{A}_{\{i,j,r,\hat{r}\}} \ll \mathcal{A}_{\{i,j,r\}}$ then $\mathcal{A}_{\Delta}^{(ij)/r\hat{r}} \ll \mathcal{S}_{\Delta}^{(ij)/r} + \mathcal{A}_{\Delta}^{(ij)/\hat{r}}$. Here, equation (S70) becomes

$$w_{ij} + \frac{w_{ABC}}{2} \approx \frac{\beta}{2}\big(\sigma_{\mathcal{A}}^{(1)} + \sigma_{\mathcal{A}}^{(2)}\big)\big(\mathcal{A}_{\Delta}^{(ij)/r} + \mathcal{A}_{\Delta}^{(ij)/\hat{r}}\big). \quad \text{(S71)}$$

It follows that $w_{ij} + \frac{w_{ABC}}{2} \approx 0$ when $\sigma_{\mathcal{A}}^{(1)} = -\sigma_{\mathcal{A}}^{(2)}$, or $\sigma_0 - \sigma_1 - \sigma_2 + \sigma_3 = 0$. Since the maintenance of pairwise LD requires the signs of $w_{ij} + w_{ABC}/2$ and $D_{ij}$ to match, invoking the approximation of (S71), pairwise LD can be maintained if $\sigma_{\mathcal{A}}^{(2)} > -\sigma_{\mathcal{A}}^{(1)}$, otherwise it cannot.

From consideration of these curves and Figure 5, notice that if both

$$\sigma_{\mathcal{A}}^{(2)} > -\sigma_{\mathcal{A}}^{(1)} \quad \text{and} \quad \sigma_{\mathcal{A}}^{(2)} > \sigma_{\mathcal{A}}^{(1)} \tag{S72}$$

are satisfied, it is possible that pairwise LD can be maintained in the absence of triple LD. Moreover, triple LD can be maintained in the absence of pairwise LD. This creates the possibility of an evolutionary bistability: depending upon the initial conditions, the population can stably exhibit different types of structure, that is, either $|D_{ABC}|$ is maximized and $D_{ij} = 0$, $|D_{ij}|$ is maximized and $D_{ABC} = 0$.

From our analysis above and Figures 5 and S11, increasing $\sigma_{\mathcal{A}}^{(1)}$ and/or $\sigma_{\mathcal{A}}^{(2)}$, produces strain structure, but depending on which of these increases more rapidly, different types of structure can occur. Large $\sigma_{\mathcal{A}}^{(2)}$ relative to $\sigma_{\mathcal{A}}^{(1)}$ tends to favour triple LD, resulting in a population with four strains present, which share no more than one antigen with any other. If instead $\sigma_{\mathcal{A}}^{(1)}$ is large relative to $\sigma_{\mathcal{A}}^{(2)}$, this favours pairwise LD, resulting in a population in which only two strains are present that share no antigens. The intuition behind this is as follows: if $\sigma_{\mathcal{A}}^{(1)}$ is positive and larger than $\sigma_{\mathcal{A}}^{(2)}$, immune protections are much more effective against a single antigen than expected, and so the minimum amount of strain diversity that preserves antigenic variation is favoured. If $\sigma_{\mathcal{A}}^{(2)}$ is instead large, then immune protections are more effective when strains share two antigens and relatively ineffective against strains sharing a single antigen. Consequently, greater strain (genotype) diversity is expected to be maintained in the population.

## S9.3 $\sigma_{\mathcal{S}} > 0$ promotes oscillations and maintains diversity

When $\sigma_{\mathcal{S}} > 0$, this has a similar effect for three loci than for two. In particular, $\sigma_{\mathcal{S}} > 0$ increases the likelihood that $w_{ABC}$ and $w_{ij} + w_{ABC}/2$ do not share sign with $D_{ABC}$ and $D_{ij}$, respectively. Thus, $\sigma_{\mathcal{S}} > 0$ breaks up strain structure and promotes genotypic diversity. Likewise, by introducing a second epidemiological feedback through hosts with homologous immunity (hosts belonging to either $\mathcal{S}_{\Delta}^{(ij)/r}$ (for $w_{ij}$) or $\mathcal{S}_{\Delta}^{(2)}$ (for $w_{ABC}$)), $\sigma_{\mathcal{S}} > 0$ can produce evolutionary oscillations with any level of strain structure, during which the set of overrepresented strains varies (Fig. S12). Simulations suggest that these oscillations do not occur between different types of structure in the deterministic system, even when bistability is possible. However, in the presence of demographic stochasticity, it seems likely that this can occur, which would further complicate the dynamics.

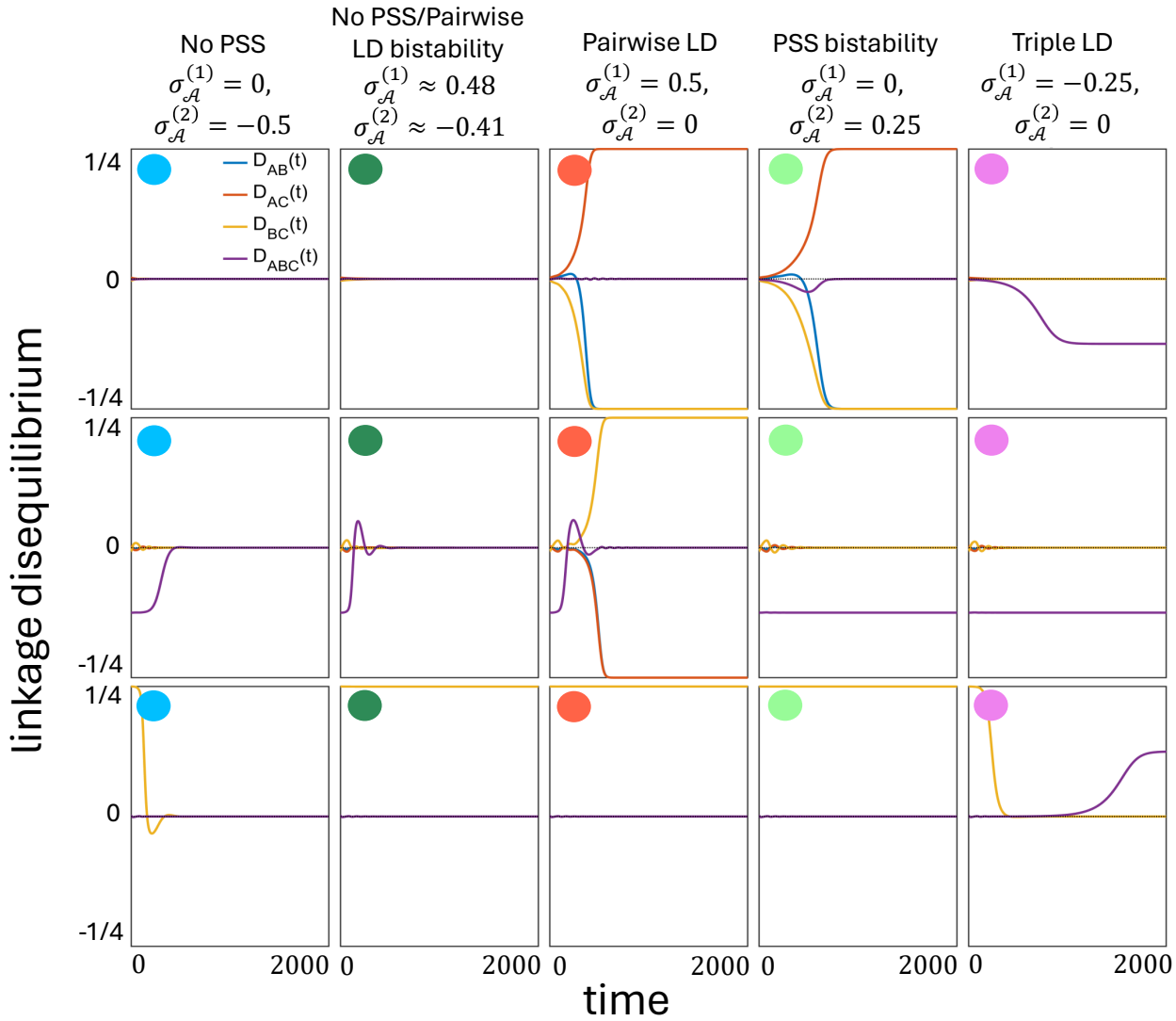


Figure S11: **Dynamical behaviour of LD as $\sigma_{\mathcal{A}}^{(1)}$, $\sigma_{\mathcal{A}}^{(2)}$ and the initial conditions vary.** The dynamics of pairwise- and triple LD ($D_{ij}(t)$ and $D_{ABC}(t)$) depend upon the values of $\sigma_{\mathcal{A}}^{(1)}$ and $\sigma_{\mathcal{A}}^{(2)}$ (different columns) and the initial conditions (different rows). Each column corresponds to the indicated values of $\sigma_{\mathcal{A}}^{(1)}$ and $\sigma_{\mathcal{A}}^{(2)}$, with the coloured circle indicating the approximate position of this scenario in Figure 5. Each row corresponds to a different initial condition. In the first row, all simulations start near a state of no PSS; in the second row, all simulations start near a state of triple LD PSS; in the third row, all simulations start near a state of pairwise LD PSS. All simulations assume $b = d = 0.01$, $\beta = 1$, $\mu = 0.4$, $\rho = 10^{-6}$, $\sigma_0 = 1$ and $\sigma_3 = \sigma_h = 0$; $\sigma_1$ and $\sigma_2$ vary as indicated by column.

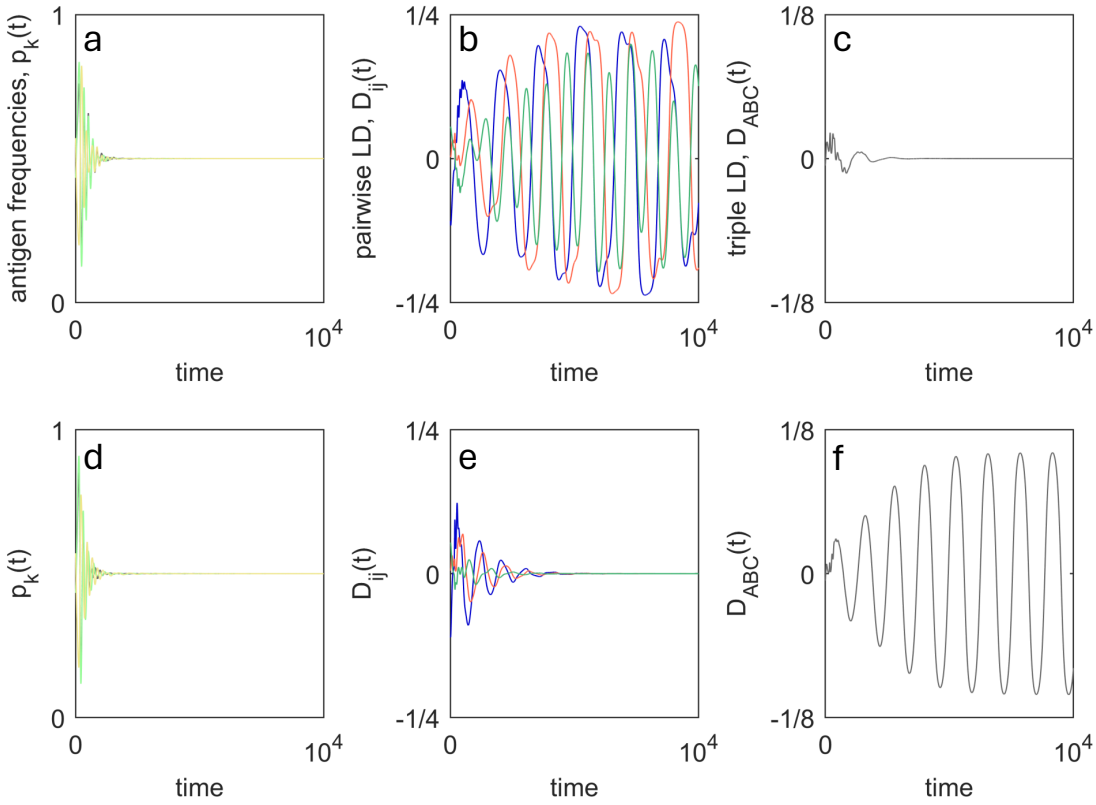


Figure S12: **Evolutionary oscillations for more than two loci.** For more than two loci there is more than one possible order of LD, and consequently, there are different 'types' of PSS. As a result, evolutionary oscillations can occur for any order of LD. All simulations assume $b = d = 0.005$, $\beta = 1$, $\mu = 0.5$, $\rho = 10^{-8}$, $\sigma_0 = 1$ and $\sigma_h = 0.05$. In panels **a-c**, $\sigma_1 = 0.6$, $\sigma_2 = 0.25$, $\sigma_3 = 0.2$; in panels **d-f**, $\sigma_1 = 0.7$, $\sigma_2 = 0.3$, $\sigma_3 = 0.2$ .