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

October 27, 2023

Abstract

Understanding the evolution of multistrain pathogens 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, and this has motivated a large-body of research seeking to understand how acquired immunity can mold the genomics of pathogen populations (pathogen strain structure). However, the diversity of existing models makes understanding the evolutionary connections between predictions difficult, leading to confusion about those that are broadly applicable and those that are not. Here we adopt a multilocus population genetics perspective to unify the evolutionary predictions of existing models, and provide insight on the role of naturally-acquired immunity in the evolution of pathogen strain structure. We identify three key factors: (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. We use these three factors to clarify the relationship among previous models of multistrain dynamics and to establish solid theoretical foundations for the study of the evolutionary epidemiology of multitrain 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

The genomic data of many pathogens reveals species genetically subdivided into distinct genotypes, or strains, characterized by genomic associations between alleles spanning different loci [1–4]. Disentangling what produces these genomic patterns (henceforth, pathogen strain structure or PSS) is a major focus of evolutionary epidemiology, owing to its relevance for a variety of human health issues, including forecasting epidemics [5] and predicting the epidemiological consequences of vaccination campaigns [6, 7]. 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 dynamics of the genomics of pathogen populations [2, 5, 8–16].

The aspect of naturally-acquired immunity receiving the most attention is the role of cross-protection, broadly defined as any immune protections an infection provides against genetically similar strains [e.g., 2, 5, 8, 10–12, 17–21]. Roughly speaking, the predictions are as follows [11, 22]: 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 over-represented in the population [2, 8]. Yet for each prediction there are exceptions, and this has led to confusion about the generality of the results. For example, it is not well-understood why only some models produce evolutionary oscillations [23], how predictions about PSS scale with an increasing number of immunogenic loci and alleles [14–16], or what patterns of PSS are either possible or likely [24]. One reason for this confusion is that models combining naturally-acquired immunity with pathogen genetic diversity produce an explosion in the number of variables [25]. Consequently, existing work has taken subtly different approaches to simplify the dynamics, with poorly understood consequences [23, 25, 26].

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. 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 these evolutionary forces are shaped by the epidemiology of naturally-acquired immunity. This allows the identification of three factors: 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. Throughout, our aim is to extract as much biological intuition as possible, and so we do not make rigorous mathematical arguments [see e.g., 23, 26, for a more quantitative perspective]. Notation and acronyms used throughout are detailed in Table 1.

2 What is pathogen strain structure (PSS)?

As a strain consists of multiple alleles, any definition of pathogen strain structure (PSS) requires genomic patterns unexplainable by selection for individual alleles. It is therefore insufficient for some strains to be more abundant; this could arise solely through variation in allele frequencies. Instead, PSS minimally requires some strains to be *over-represented* relative to their allelic expectation, and so identification of PSS necessitates specifying the multilocus pathogen genetics. For a pathogen genome consisting of two diallelic loci, $\{a, A\}$

and $\{b, B\}$, and so four strains $\{ab, Ab, aB, AB\}$, PSS requires

$$D(t) \equiv p_{ij}(t) - p_i(t)p_j(t) \neq 0, \tag{1}$$

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) [27]. LD can be generalized to a larger number of loci and alleles [27], and since LD captures any nonrandom associations between alleles, the presence of LD is necessary for some strains to be over-represented. Hence, the minimal definition of PSS is the presence of some pairwise (or higher-order) LD.

There also exists a more stringent form of PSS. *Discordant strain structure* (DSS), first proposed by Gupta et al. [2], occurs when a set of strains whose members share no alleles (e.g., $\{ab, AB\}$ and $\{Ab, aB\}$) is over-represented in the population [2, 8, 28]. Generally speaking, the presence of pairwise LD is necessary, but not sufficient, for DSS [28, 29].

Given an observed distribution of allele frequencies, there is a maximum and minimum value LD can possibly take. For example, for two diallelic loci

$$\max(D(t)) = \min\big(p_A(t)(1 - p_B(t)),\, p_B(t)(1 - p_A(t))\big)$$

[see 27]. Importantly, in order for pairwise LD to achieve this maximum, 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 [30, 31], and so if LD assumes 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.

3 What evolutionary forces produce strain structure?

As PSS requires the presence of LD, the dynamics of LD are key to understanding it's evolution. To describe these dynamics, we will initially focus upon two diallelic loci. For multistrain models of adaptive immunity, two timescales often emerge [23, 32]. Over the short timescale, strong negative frequency-dependent selection (NFDS) operating through the additive selection coefficients, $w_A \equiv r_{Ab} - r_{ab}$ and $w_B \equiv r_{aB} - r_{ab}$, pushes the alleles to an equal frequency (here, r_{ij} is the per-capita growth rate of strain ij; see Box 1). As this occurs, the strains self-organize into two sets, $\{ab, AB\}$ and $\{Ab, aB\}$, so that over the long-term, LD evolves according to the equation

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

where w_{AB} denotes epistasis in fitness between alleles A and B, that is,

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

and $\Delta_D\rho$ is the net change in LD due to recombination (Sup. Info. S3). Equation (1) and (2) reveal the evolutionary forces generating and maintaining PSS.

Box 1: A population model of naturally-acquired immunity

Models of adaptive immunity largely differ in three ways (Sup. Info. S2): the type of immune protections (e.g., does immunity reduce infection or transmissibility?); how immunity is gained and lost; and if hosts can be simultaneously infected by multiple strains (co-infection). Typically, the choice of assumptions is made to simplify the mathematics. The model we introduce here prioritizes conceptual simplicity and is similar to that of [33]. Whenever possible, we will use this model to explain existing results; required extensions, as well as more detailed links to existing 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 an (host) adaptive immune response. Transmission occurs exclusively to uninfected hosts (so we do not permit co-infection) via mass-action with rate constant β, while the probability of successful infection, σ_x^g, depends upon the host's immune history, x, and the genotype, g, of the 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 will assume 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 $\sigma_{x'}^{aB} = \sigma_{x'}^{AB} = \sigma_1$. Hosts clear infections and die from natural and virulence-related mortality at per-capita rates μ, d, and α, respectively. Immunologically naive hosts ($x = \{\varnothing\}$) enter the population at rate b, while immunity is life-long and acquired at clearance. Let $S_x(t)$ and $I_x^g(t)$ denote, respectively, the density of uninfected, and infected with strain g, hosts with immune history x at time t. Then

$$\begin{aligned} \frac{\mathrm{d}S_x}{\mathrm{d}t} &= b\mathbf{1}_0(x) - dS_x - \beta \sum_g \sigma_x^g \sum_z I_z^g S_x + \sum_g \sum_{z:g\cup z=x} \mu I_z^g \\ \frac{\mathrm{d}I_x^g}{\mathrm{d}t} &= \beta \sum_z I_z^g \sigma_x^g S_x - (d+\mu+\alpha) I_x^g \end{aligned} \tag{3}$$

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 them from model (3), we focus upon two diallelic loci (i.e., $g = ij$ where $i \in \{a, A\}$ and $j \in \{b, B\}$). Let $\mathcal{A}_{\{i,j\}}(t)$ and $\mathcal{S}_{ij}(t)$ denote the density of uninfected hosts with immunity to *at least* the set of antigens $\{i, j\}$ and prior infection by strain ij, respectively (Fig. 1). For example, uninfected hosts with infection history $\{ab, AB\}$ would contribute to $\mathcal{A}_{\{a,b\}}(t)$, $\mathcal{A}_{\{A,b\}}(t)$, $\mathcal{A}_{\{a,B\}}(t)$ and $\mathcal{A}_{\{A,B\}}(t)$ but only to $\mathcal{S}_{ab}(t)$ and $\mathcal{S}_{AB}(t)$. Using this notation, the per-capita growth rate of strain ij is

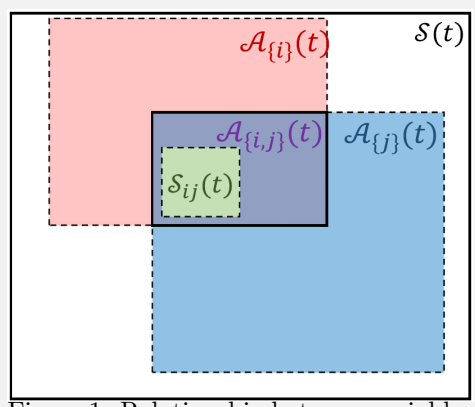

Figure 1: Relationship between variables describing the distribution of immunity.

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

where $\sigma_{\mathrm{PI}} \equiv \sigma_0 - 2\sigma_1 + \sigma_2$ and $\sigma_{\mathrm{HI}} \equiv \sigma_2 - \sigma_h$, and $\mathcal{S} = \sum_x S_x$ is the total density of uninfected hosts, irrespective of immune history.

First, from equation (1) 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. As our interest is what favours or disfavours some strains given allelic variation observable from genomic data, we will not consider further what favours or disfavours allelic diversity [e.g., see 4, 34, 35].

Second, assuming both loci are polymorphic, the long-term presence of PSS requires some force to generate and maintain LD. From equation (2) there are two possibilities: epistasis and (biased) recombination. Because virtually all multistrain models implicitly revolve around epistasis [but see 36, 37], we will assume recombination is weak, $\Delta_D \rho \approx 0$. Then the dynamics of (2) hinge on epistasis, w_{AB}, which indicates whether the combination of alleles A and B are more, $w_{AB} > 0$, or less, $w_{AB} < 0$, favoured than expected.

4 Epistasis in an epidemiological context

Next, 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 exposure to an allele (or antigen) induces an adaptive immune response protecting against future infection. In particular, the probability of infection of a host with immunity to $n = 0, 1, 2$ antigens of the transmitted strain is σ_n, while the probability a host is re-infected by the same strain is σ_h (homologous immunity). Then epistasis is

$$\begin{aligned} w_{AB} = \beta(\overbrace{\sigma_0 - 2\sigma_1 + \sigma_2}^{\sigma_{\mathrm{PI}}})(\overbrace{\mathcal{A}_{\{a,b\}}(t) + \mathcal{A}_{\{A,B\}}(t) - \mathcal{A}_{\{A,b\}}(t) - \mathcal{A}_{\{a,B\}}(t)}^{\mathcal{A}_\Delta(t)}) \\ - \beta(\underbrace{\sigma_2 - \sigma_h}_{\sigma_{\mathrm{HI}}})(\underbrace{\mathcal{S}_{ab}(t) + \mathcal{S}_{AB}(t) - \mathcal{S}_{Ab}(t) - \mathcal{S}_{aB}(t)}_{\mathcal{S}_\Delta(t)}), \end{aligned} \quad (4)$$

where β is a transmissibility rate constant and $\mathcal{A}_{\{i,j\}}(t)$ and $\mathcal{S}_{ij}(t)$ denote the density of uninfected hosts with immunity to *at least* the set of antigens i and j and prior infection by strain ij, respectively, at time t (Box 1). There are two factors contributing to equation (4). The first is the strength and specificity of the immune protections, captured by the composite parameters σ_{PI} and σ_{HI} (Fig. 2). The second is the dynamic immunological landscape, captured by the composite variables $\mathcal{A}_\Delta(t)$ and $\mathcal{S}_\Delta(t)$.

4.1 The specificity and strength of immune protections

The specificity and strength of immune protections partitions existing models into two groups. In the first group [e.g., 14–16, 38], hosts acquire *antigen-specific* immunity: 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, $\sigma_1 \geq \sigma_2$, so σ_{PI} can be positive or negative, but $\sigma_2 = \sigma_h$ and so $\sigma_{\mathrm{HI}} = 0$, that is, homologous immunity is no more protective then full-antigenic immunity. Consequently, we will say that σ_{PI} controls the strength of *antigen-specific cross-protection* (ASCP). These models apply to, for example, pathogens which temporally express different antigens to which the host develops an immune response (e.g., malaria [14–16], pathogenic *Neisseria* [39]).

In the second group [e.g., 2, 5, 8, 10–12, 22, 40], hosts acquire *strain-specific* immunity: the level of cross-protection against a particular strain is determined by the most antigeni-

cally similar strain that has infected the host, but this is less protective than homologous immunity, that is, $\sigma_2 > \sigma_h$ and so $\sigma_{\mathrm{HI}} > 0$. Thus σ_{HI} controls the strength of *strain-specific cross-protection* (SSCP). For example, although 'strains' might be defined based upon 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 then between strains, homologous immunity may be more protective then full-antigenic immunity.

More generally, if the adaptive immune system recognizes, and targets, the pathogen at the level of it's constitutive antigens, rather than as a unit (i.e., strain) then $\sigma_{\mathrm{HI}} = 0$, whereas if the pathogen is targeted as a unit and cross-protection arises through single-strain similarity, then $\sigma_{\mathrm{HI}} > 0$. These two perspectives yield very different evolutionary dynamics.

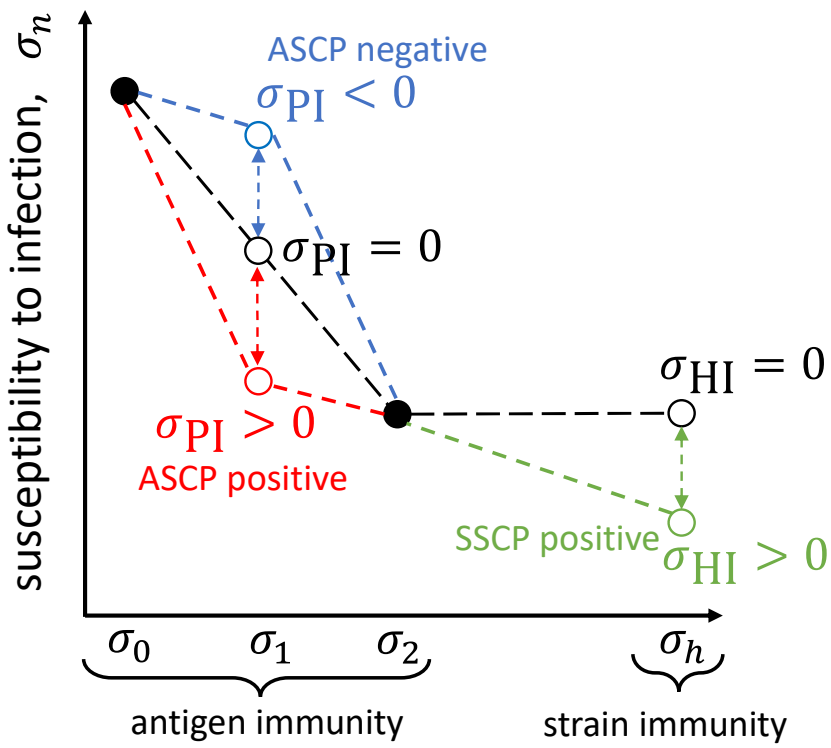

Figure 2: **Schematic of cross-protection differences**. Specification of susceptibility to infection translates to different values of antigen-specific cross-protection (ASCP), captured by σ_{PI}, and strain-specific cross-protection (SSCP), captured by σ_{HI}.

Strong ASCP produces stable strain structure

In the absence of SSCP, $w_{AB} = \sigma_{\mathrm{PI}}\mathcal{A}_\Delta(t)$. The variable $\mathcal{A}_\Delta(t)$ indicates if more uninfected hosts have full antigenic immunity to the set of strains $\{ab, AB\}$ (if positive) or $\{Ab, aB\}$ (if negative). Thus if one strain set, say $\{ab, AB\}$, is over-represented in the population, our intuitive expectation would be that more hosts have full antigenic immunity to strains belonging to that set, that is, $\mathcal{A}_\Delta(t)$ and $D(t)$ share the same sign. If this holds, there are three possible outcomes (Fig. 3**a**). First, if $\sigma_{\mathrm{PI}} < 0$, ASCP is weak and so all strains

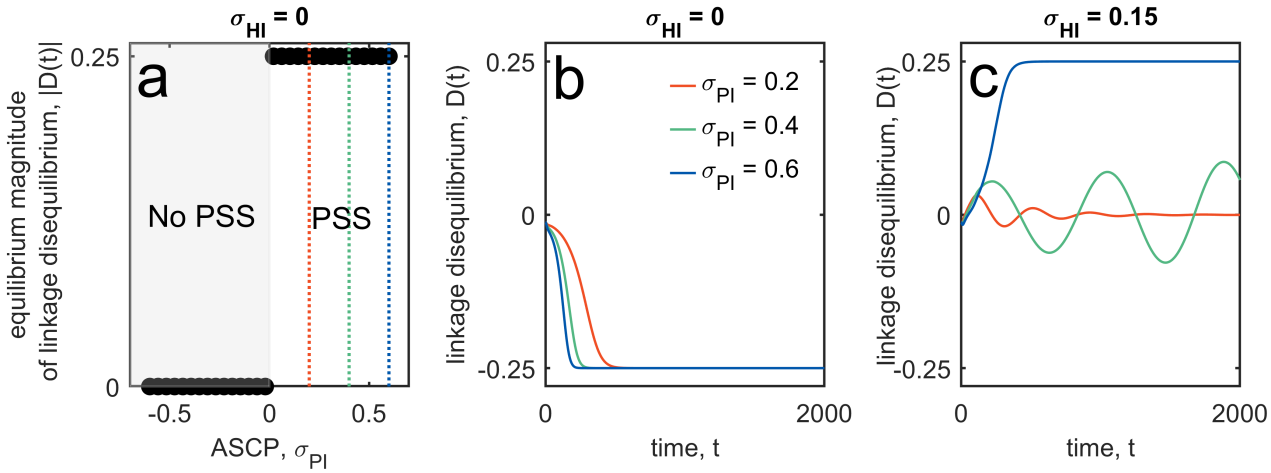

Figure 3: **Consequences of ASCP and SSCP**. In the absence of SSCP (i.e. $\sigma_{\mathrm{HI}} = 0$), the long-term outcome (panel **a**) and dynamics (panel **b**) are straightforward: if ASCP is weak, $\sigma_{\mathrm{PI}} < 0$, no PSS occurs ($|D(t)| = 0$) whereas if it is strong, $\sigma_{\mathrm{PI}} > 0$, PSS occurs ($|D(t)| = D^*$). The inclusion of SSCP ($\sigma_{\mathrm{HI}} > 0$), however, can counteract the effects of ASCP when $\sigma_{\mathrm{PI}} > 0$, and this tension can sustain oscillations (panel **c**). The dotted lines on panel **a** indicate the value of σ_{PI} used in the simulations on panels **b**,**c**. All simulations start from same randomly chosen initial condition, while parameter values used are $\beta = 1$, $\mu = 0.58$, $b = d = 0.01$, $\alpha = 0$, $\sigma_0 = 1$, $\sigma_2 = 0.4$, σ_1 varies, and $\sigma_h = 0.4$ (panels **a**,**b**) or $\sigma_h = 0.25$ (panel **c**).

coexist. Second, if $\sigma_{\mathrm{PI}} = 0$, epistasis vanishes and any observed PSS is neutral and is contingent upon the initial conditions. Third, if $\sigma_{\mathrm{PI}} > 0$, ASCP is strong and so PSS will occur; in this case because there is a single epidemiological feedback, $\mathcal{A}_\Delta(t)$, one strain set is competitively excluded, $D(t) \to D^*$. Thus the sign of ASCP drives the formation of PSS by controlling whether competition between strains sharing one antigen is stronger or weaker than expected if the strength of protection offered by the immune response to multiple antigens was additive. Indeed, if we consider the function mapping antigenic immunity to the strength of immune protections, $h : n \mapsto \sigma_n$, then σ_{PI} is the central difference approximation of $\mathrm{d}^2h/\mathrm{d}n^2$ evaluated at $n = 1$. Consequently, if $\sigma_{\mathrm{PI}} > 0$ (resp. $\sigma_{\mathrm{PI}} < 0$), the rate of increase of immune protections at $n = 1$ is increasing (resp. decreasing).

Strong SSCP generates oscillations and maintains strain diversity

Importantly, ASCP alone cannot sustain evolutionary oscillations, nor can it stably maintain an intermediate level of PSS (i.e., LD is not equal to D^*); two common predictions of multistrain models. This is not surprising: in the absence of SSCP, 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 or the other, $\mathcal{A}_\Delta(t)$, and so no reason to expect the sign of epistasis to change based upon the amount of LD. However, SSCP provides a second epidemiological feedback, whether more hosts have *homologous immunity* to strains belonging to one set or the other, $\mathcal{S}_\Delta(t)$. This promotes strain diversity because the more abundant strains will have caused more infections and so be at a competitive dis-

advantage. If ASCP is weaker than expected (i.e. $\sigma_{\mathrm{PI}} < 0$), both feedbacks prevent strains from becoming over-represented (Fig. 3**a**). However, if ASCP is stronger than expected (i.e. $\sigma_{\mathrm{PI}} > 0$), it promotes PSS whereas SSCP inhibits it (Fig. 3**c**). This has two consequences.

First, these counteracting forces can sustain evolutionary oscillations. Indeed, not only is this the source of the complex dynamics seen in many models [2, 5, 8, 12, 41], but it also explains why oscillations are absent in so-called 'status-based' models but not 'history-based' models [9, 23, 26, 32]. History-based models, such as the model of Box 1 assume that following infection, hosts may be partially-protected against genetically similar strains (i.e., 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 and hosts with sterilizing immunity. Consequently, for status-based models there is no difference between full-antigenic or homologous immunity (so $\sigma_{\mathrm{HI}} = 0$), and so no potential for sustained evolutionary oscillations.

Second, the presence of SSCP maintains greater strain diversity than expected based upon consideration of the sign of ASCP. For example, multistrain models commonly assume that immunity to a single antigen is as protective as full antigenic-immunity [i.e., $\sigma_1 = \sigma_2$ and so $\sigma_{\mathrm{PI}} = \sigma_0 - \sigma_2$ 2, 8, 17–19]. Despite $\sigma_{\mathrm{PI}} > 0$, these models often predict no strain structure is a common outcome because homologous immunity is assumed to be sterilizing and so SSCP is very strong (Sup. Info. S2). Moreover, unlike the case in which there is no SSCP, in the presence of SSCP a stable amount of PSS can be maintained in the population (so $D(t) \neq 0$), but this need not lead to the exclusion of some strains (so $D(t) \neq D^*$).

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

Of course, immunity need not only protect against infection. Suppose in the model of Box 1 any of transmissibility, β, clearance rate, μ, and virulence mortality, α, depend upon the immune history of the host. Although in this case epistasis can be written as a time-varying quantity as in equation (4), since the fitness of an infection will depend on it's current environment (immune history of the host), 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. S4). As such, time-varying epistasis is less amenable to interpretation, and so here we focus on epistasis at equilibrium. To do so, suppose homologous immunity is no more protective than full-antigenic immunity. Then the population will evolve to a stable equilibrium, whether it be PSS or no PSS, and at the PSS equilibrium epistasis shares the same sign as LD if

$$\epsilon_{\mathrm{PI}} \equiv f_0 - 2f_1 + f_2 \tag{5}$$

is positive (Sup. Info. S4). In equation (5), $f_n \equiv \frac{\sigma_n \beta_n}{\mu_n + \alpha_n + d}$ is the relative fitness of an infection in a host with immunity to n of it's antigens (Sup. Info. S4). Thus whether an infection in a host with immunity to one of it's antigens is more ($\epsilon_{\mathrm{PI}} > 0$) or less ($\epsilon_{\mathrm{PI}} < 0$) fit than expected determines the evolutionary outcome: if it is more fit than expected, PSS occurs.

Equation (5) reveals that the type of immune protections affect the predictions for PSS in two ways. First, although increasing the protection offered by immunity to one antigen

promotes PSS if immunity reduces susceptibility and/or transmissibility [2, 8, 11, 42] or increases clearance rate [14–16, 38], it inhibits PSS if immunity reduces virulence-mortality [43, 44]. Second, if each antigen produces an additive increase in the strength of immunity [4], then when immunity reduces susceptibility to infection or transmissibility, $\epsilon_{\mathrm{PI}} = 0$, whereas if immunity increases clearance rate [a common assumption for malaria models 14–16], $\epsilon_{\mathrm{PI}} > 0$. Thus immunity increasing clearance rate is more likely to produce PSS, and this is particularly true if pathogens cause chronic infections (larger d) and/or cause host death (larger α).

As the derivation of equation (5) assumes homologous immunity is equivalent to full-antigenic immunity, this suggests that the analogous quantity to σ_{HI} is $\epsilon_{\mathrm{HI}} \equiv f_2 - f_h$, that is, ϵ_{HI} is the generalization of SSCP. Consequently, SSCP is necessary for evolutionary oscillations.

4.2 The dynamic immunological landscape

In addition to the specificity and strength of immune protections, epistasis also depends upon the dynamic immunological landscape controlled by $\mathcal{A}_\Delta(t)$ and $\mathcal{S}_\Delta(t)$. As 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 (4) can help understand the influence of the immunological landscape on the buildup of strain structure. In particular, this analysis reveals the impact of re-infections and how immunity is gained and lost.

The frequency of re-infections shapes the likelihood of evolutionary oscillations

Although evolutionary oscillations require SSCP to produce a second epidemiological feedback acting on epistasis ($\mathcal{S}_\Delta(t)$), from equation (4) this alone is not sufficient: the two epidemiological feedbacks must sufficiently diverge from one another to produce counteracting effects. In particular, there needs to be a difference between the distribution of immunity to different antigens, $\mathcal{A}_\Delta(t)$, and different strains, $\mathcal{S}_\Delta(t)$. It can be shown 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 (Fig. 4; Sup. Info. S3.2). Consequently, the 'length' of the immune history x, is critical (Fig. 4). Generally speaking, the 'length' of the immune history is simply the number of strains to which a host currently has immunity to. However, in the absence of waning immunity, it is the number of prior infections by genetically distinct strains. If hosts have immunity to only one strain, then $\mathcal{A}_\Delta(t) = \mathcal{S}_\Delta(t)$ and so from equation (4) PSS forms if $\sigma_{\mathrm{PI}} > \sigma_{\mathrm{HI}}$, otherwise it will not; sustained oscillations are not possible. If instead almost all hosts have immunity to more than one strain, then $\mathcal{A}_\Delta(t) \to 0$ much faster than $\mathcal{S}_\Delta(t) \to 0$, and so neither PSS nor oscillations are possible. Only between these two extremes can oscillations occur.

There are two generic pathogen life-history quantities affecting the relationship between $\mathcal{A}_\Delta(t)$ and $\mathcal{S}_\Delta(t)$. The first is the basic reproductive number, $R_0 \equiv \beta\sigma_0/(\mu + \alpha + d)$ (Fig. 4). When R_0 is small, hosts are unlikely to be infected more than once, so $\mathcal{A}_\Delta(t) \approx \mathcal{S}_\Delta(t)$, whereas when R_0 is large, hosts being infected multiple times is common, so $\mathcal{A}_\Delta(t) \to 0$. Thus oscillations are more likely for pathogens with intermediate R_0 while PSS is unlikely for pathogens with high R_0. The second quantity is the ratio of the infectious period to

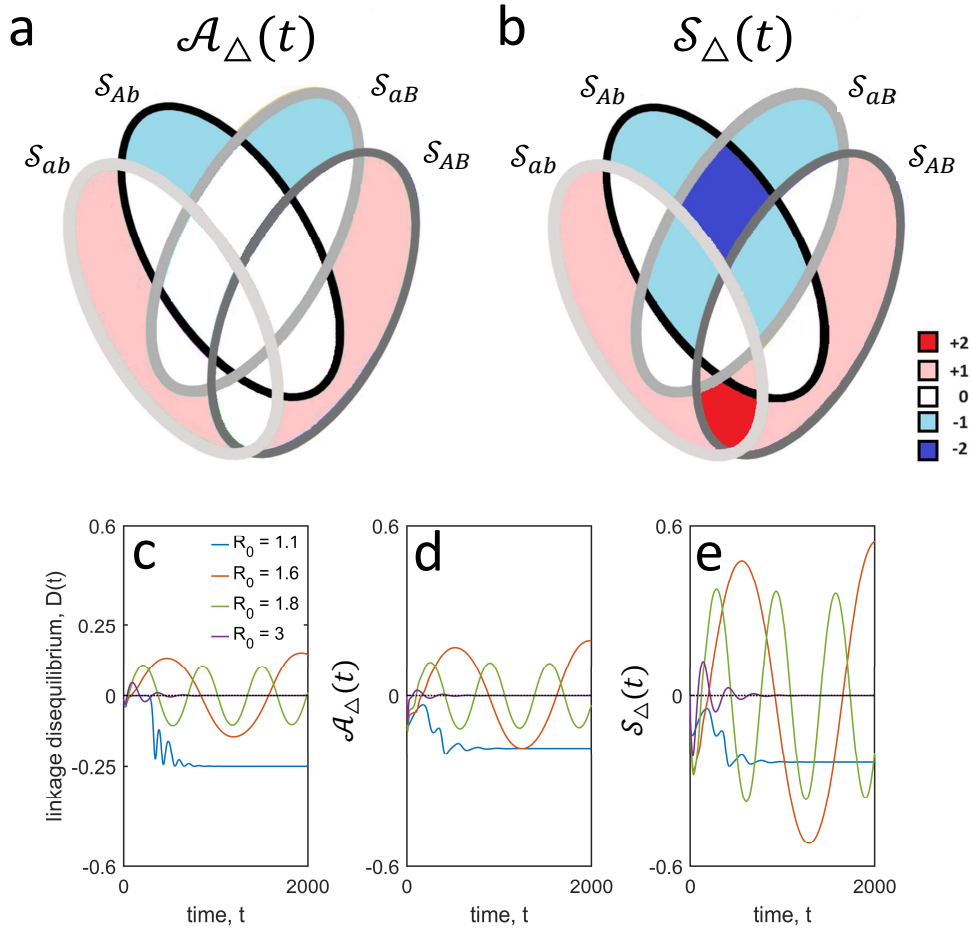

Figure 4: **The dynamic immunological landscape**. Panels **a** and **b** are Venn diagrams of $\mathcal{A}_\Delta(t)$ (**a**) and $\mathcal{S}_\Delta(t)$ (**b**). Each ellipse corresponds to a different $\mathcal{Z}_{ij}(t)$, which is the density of hosts with immunity to *at least* strain ij. Then $\mathcal{A}_\Delta(t)$ and $\mathcal{S}_\Delta(t)$ are computed as the weighted sum of the coloured regions, where the weights are colour dependent. Although $\mathcal{A}_\Delta(t)$ only involves hosts with homologous immunity to a single strain, $\mathcal{S}_\Delta(t)$ involves hosts with immunity to multiple strains. Panels **c-e** show how the dynamic immunological landscape is shaped by the 'length' of the immune history, which in this case is controlled by the basic reproductive number, R_0. For small R_0 (blue curve), most hosts are infected a single time during their lives and so $\mathcal{A}_\Delta(t) \approx \mathcal{S}_\Delta(t)$. For intermediate R_0 (orange and green) hosts are infected multiple times and so $\mathcal{A}_\Delta(t)$ and $\mathcal{S}_\Delta(t)$ diverge, sustaining evolutionary oscillations. For large R_0 (purple), hosts are always infected multiple times and so $\mathcal{A}_\Delta(t) \to 0$ faster than $\mathcal{S}_\Delta(t) \to 0$ and thus no PSS is possible. Parameter values used for the simulations of panels **c-e** are $\beta = 1$, $b = d = 0.01$, $\alpha = 0$, $\sigma_0 = 1$, $\sigma_1 = 0.5$, $\sigma_2 = 0.4$ (i.e. $\sigma_{\text{PI}} = 0.2$), and $\sigma_h = 0.25$ (i.e. $\sigma_{\text{HI}} = 0.15$), while μ is chosen so that R_0 is equal to indicated value.

host lifespan, which is $\theta \equiv d/(\mu + d + \alpha)$ [23, 33]. This ratio controls whether infections are acute (small θ) or chronic (large θ). The importance of θ hinges upon whether co-infections are possible. If co-infection is not possible, chronic infections reduce the likelihood hosts are

infected more than once, so $\mathcal{A}_\Delta(t) \approx \mathcal{S}_\Delta(t)$. If co-infection is possible, whether an infection is acute or chronic is less important as infected hosts remain susceptible to infection.

Because R_0 and θ are present in virtually every epidemiological model, they will influence the consequences of other factors altering the 'length' of immune memory. For example, if immunity wanes over time [45], or there is short-term strain-transcending immunity [13, 40, 46–50] (analogous to no co-infection), the length of immune memory will be shortened either by causing the host to 'forget' previous infections, or by reducing the likelihood of becoming re-infected, respectively. If R_0 is very high, shortening the immune memory will increase the density of hosts contributing to $\mathcal{A}_\Delta(t)$, and so increase the likelihood of oscillations and strain structure; if R_0 is intermediate, however, the likelihood of oscillations will be reduced. This dependence upon R_0 can help explain why shortening immune memory through waning immunity [45], conserved immune responses [51], or strain-transcending immunity [40] has been observed to both decrease [45, 51] and increase [40] the likelihood of oscillations.

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

The predicted consequences of ASCP based upon the sign of σ_{PI} hinge upon the expectation that the distribution of infections is coupled to the immunological landscape, that is, there is a positive correlation between $D(t)$ and $\mathcal{A}_\Delta(t)$. The intuition behind this expectation is that the more abundant set of strains (sign of LD) will have caused more infections, thereby tightly coupling the sign of $D(t)$ and the immunological landscape, $\mathcal{A}_\Delta(t)$. 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 hosts are more likely to acquire immunity to only a single, randomly chosen antigen of an infecting strain rather than both. Then suppose strain set $\{ab, AB\}$ is over-represented, that is, $D(t) > 0$. Naive hosts are then most likely to become infected by one of the over-represented strains (say AB), after which they are most likely to become infected by the other over-represented strain (say 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 $\{a, B\}$, and therefore have antigenic immunity to the under-represented strain set. Consequently, $\mathcal{A}_\Delta(t)$ will have the opposite sign to $D(t)$, and therefore from equation (4) PSS will only occur if ASCP is *weaker* than expected, $\sigma_{\mathrm{PI}} < 0$ (Sup. Info. S5).

Indeed, there are a variety of immunological reasons for 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 different hosts target and/or lose immunity to different antigens, based upon host genetics (e.g., MHC alleles [4]) or randomly (e.g., which B-cell encounters which epitope first; see [52] and Sup. Info. S5 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 effect of ASCP.

4.3 The number of loci and the possible patterns of PSS

Many of the key predictions for two diallelic loci hold as the number of alleles and number of loci increase: ASCP still promotes PSS, SSCP still promotes strain diversity and is necessary

for evolutionary oscillations, and oscillations are still more likely when hosts tend to be infected by more than one strain. 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); different types of PSS are possible [5, 12, 24, 40]. This is because LD is no longer restricted to non-random associations spanning two loci and instead involves as many loci as are present, with each order of LD corresponding to a different type of PSS. For example, triple LD is an over-representation of strains sharing one antigen, and more generally, the number of over-represented 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, and this evolution is still dictated by epistasis, whether it be pairwise or higher-order (Sup. Info. S7).

Indeed, in the absence of SSCP, each order of epistasis hinges upon ASCP, which for N loci is captured by the $N-1$ composite parameters (Fig. 5a) ,

$$\sigma_{\mathrm{PI}}^{(n)} \equiv \sigma_{n-1} - 2\sigma_n + \sigma_{n+1}, \quad n = 1, ..., N-1. \tag{6}$$

As before the sign of $\sigma_{\mathrm{PI}}^{(n)}$ determines whether the rate of increase of the strength of immune protections against n antigens is increasing, $\sigma_{\mathrm{PI}}^{(n)} > 0$, or decreasing, $\sigma_{\mathrm{PI}}^{(n)} < 0$. Thus these parameters control how the strength of competition between strains changes based on genetic similarity, and this in turn shapes epistasis (Fig. 5). For example, for three diallelic loci, triple epistasis in fitness, w_{ABC}, can maintain triple LD if $\sigma_{\mathrm{PI}}^{(2)} > \sigma_{\mathrm{PI}}^{(1)}$. On the other hand, pairwise LD between alleles i and j can be maintained if it shares the same sign as $w_{ij} + w_{ABC}/2$ which can be approximated as $\sigma_{\mathrm{PI}}^{(2)} > -\sigma_{\mathrm{PI}}^{(1)}$ (Sup. Info. S7). 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 bistability emerges such that the type of PSS present is contingent upon the initial conditions (Fig. 5; [see also 24]).

Given $N > 2$ loci and so multiple types of PSS are possible, what genomic patterns are expected? Roughly speaking, this hinges upon the number of antigens sufficient to induce a strongly protective response. For example, if immunity to two antigens is strongly protective, $\sigma_{\mathrm{PI}}^{(1)} \to -1$ and $\sigma_{\mathrm{PI}}^{(2)} \to 1$, triple LD is more likely (Fig. 5). If instead immunity to one antigen is strongly protective, $\sigma_{\mathrm{PI}}^{(1)} \to 1$ and $\sigma_{\mathrm{PI}}^{(2)} \to 0$, then pairwise LD is more likely (Fig. 5). This explains why models which assume immunity to one antigen is strongly cross-protective, that is, $\sigma_1 = \sigma_2 = \cdots = \sigma_n$, and $\sigma_n \to \sigma_h$, predict DSS or pairwise LD [2, 8].

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 LD and the forces that drive its dynamics via the production of epistasis in fitness. By doing so, we established the evolutionary connections between the multitude of existing models, providing an unifying perspective for understanding the evolution of PSS. In particular, our analysis allows us to better understand the factors shaping (i) the longterm presence or absence of PSS and (ii) evolutionary oscillations.

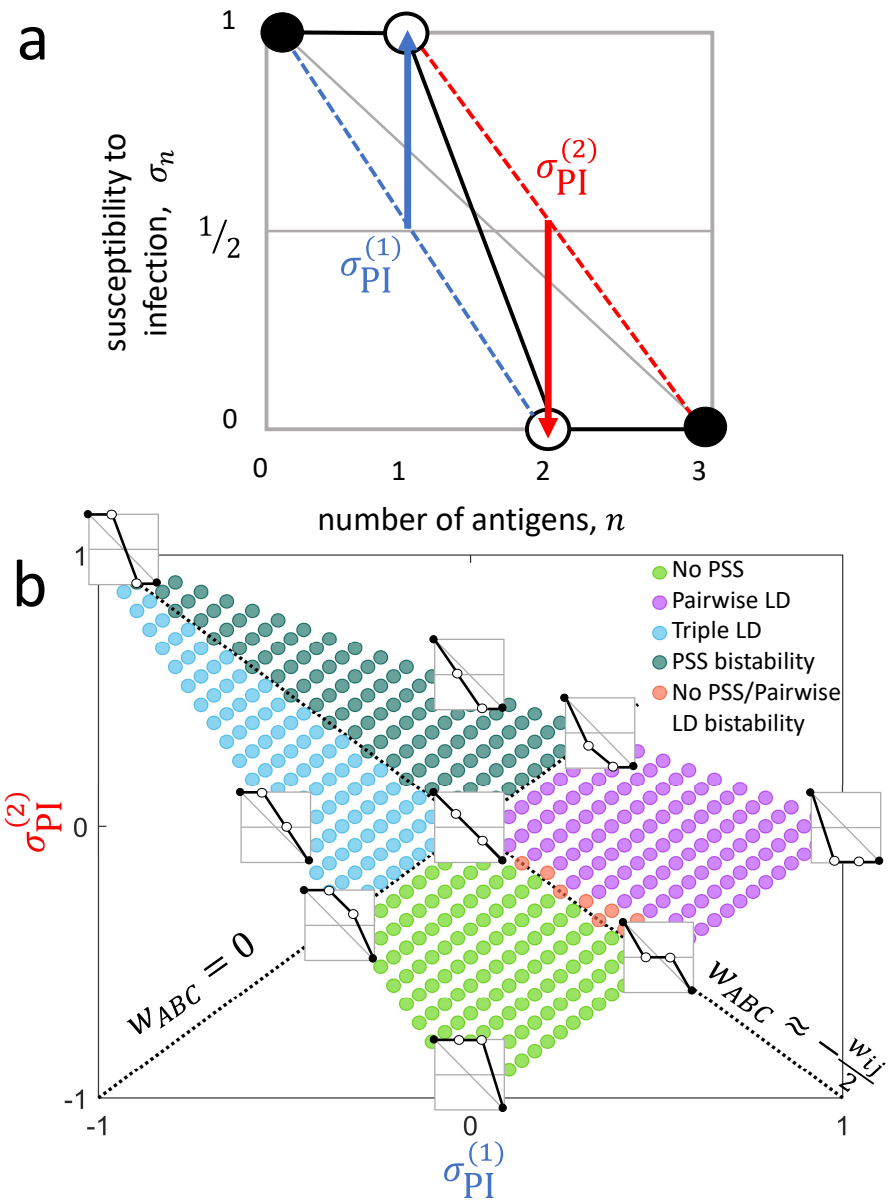

Figure 5: **As the number of loci increases so do the possible types of PSS**. The number of loci determines the number of parameters capturing ASCP (panel **a**). As these parameters vary, different dynamical outcomes are possible (panel **b**). When $\sigma_{\mathrm{PI}}^{(2)} > \sigma_{\mathrm{PI}}^{(1)}$, triple LD shares the same sign as w_{ABC}, and so can be stably maintained, while when $\sigma_{\mathrm{PI}}^{(2)} > -\sigma_{\mathrm{PI}}^{(1)}$, pairwise LD between alleles i and j will generally share the same sign as $w_{ABC} + w_{ij}/2$ (Sup. Info. S7), and so can be stably maintained. As both conditions can be simultaneously met, there is an evolutionary bistability where the type of PSS that evolves depends upon the initial conditions. Because we assume $\sigma_0 \geq \sigma_1 \geq \sigma_2 \geq \sigma_3$, this constrains the possible values the $\sigma_{\mathrm{PI}}^{(i)}$ can take in panel **b**. Parameter values used in panel **b**: $\beta = 1$, $d = b = 0.01$, $\mu = 0.4$, $\rho = 10^{-7}$, $\sigma_0 = 1$, $\sigma_3 = \sigma_h = 0$, while σ_1 and σ_2 vary.

The long-term presence of PSS hinges upon two factors. First, PSS is more likely for pathogens in which prior exposure to *some* antigens of a given strain provides much stronger immune protection than expected (strong ASCP), while prior exposure to *all* antigens of a given strain provides similar protection to homologous immunity (weak SSCP). Second, PSS is more likely if hosts have immunity to no more than a few strains. This occurs for pathogens with a low-to-intermediate R_0 or those causing chronic infections (if co-infection is limited), or when hosts display either waning or strain-transcending immunity. The most likely type of PSS, or order of LD, hinges upon the number of antigens required to induce a strongly protective immune response. As the number of antigens needed increases, so will the expected order of LD, and thus the fraction of possible strains present in the population. More broadly, although each type of PSS is in principle detectable from genomic data, it is necessary to test for the correct order of LD. For example, a highly-structured population dominated by triple LD would appear unstructured if PSS is assessed based on the presence of pairwise LD.

Evolutionary oscillations have been a focal research point for multistrain pathogens [e.g., 2, 5, 8, 12, 41]. The multilocus perspective reveals such oscillations are most likely when two conditions are met. First, the gain in protection from immunity to each additional antigen is slightly better than expected (slightly strong ASCP), but only homologous immunity triggers a strongly protective immune response (strong SSCP). Second, hosts are likely to be infected by, and so have homologous immunity to, multiple strains over the course of their lives. This is most likely for pathogens with either an intermediate R_0, or a high R_0 in the presence of waning or strain-transcending immunity (i.e., something shortening the 'length' of immune memory). Whether these conditions are met or not will depend upon the pathogen species, but identifying these key biological quantities may have greater utility than identifying regions of parameter space for a given model that permit oscillations. For example, Recker et al. [5] hypothesized that rather than being a mutation-driven process [9, 10, 48], influenza evolution may be better explained by the periodic recycling (or oscillations) of a finite set of antigens (but see [35]). Although this argument was supported by identifying that oscillations occurred over a wide range of parameter space, there is value in considering whether influenza meets the biological conditions for displaying such behaviour. More broadly, as these oscillations involve fluctuations in LD and epistasis, they may have important implications for the evolution of recombination [e.g., 53], that warrants future work.

Co-infection and within-host interactions

The model of Box 1 assumes that transmission only occurs to uninfected hosts, and thus ignored within-host interactions. Multistrain models typically assume hosts can be simultaneously infected by any number of *different* strains (co-infection), each of which are unaffected by the presence of other co-infecting strains [e.g., 2, 8, 11, 40, 50, 54, 55]. Under these assumptions, and by redefining certain variables in the model of Box 1, all our key predictions hold (Sup. Info. S2.3). However, this model of co-infection is biologically unsatisfying as it maintains high levels of genetic diversity, including otherwise selectively disadvantageous trait variation [e.g., hypervirulence 18]. Although simulations comparing different within-host interactions found limited effect [14], there are reasons to believe this will not hold generally. For example, co-infection in *Streptococcus pneumonia* can lead to

the simultaneous clearance of all infecting strains [4, 56, 57]. Since this will disproportionately (negatively) affect the more abundant strains, it will remove PSS. This may partially explain why *S. pneumoniae* genomic data shows low amounts of strain structure [58]. Thus the impact of within-host interactions on PSS warrants further work.

Demographic stochasticity

Our analysis is based on a deterministic model. but previous studies have shown that demographic stochasticity can have important effects on multistrain dynamics. First, demographic stochasticity can produce random oscillations not seen in the deterministic model [14, 59]. Because PSS is inhibited or promoted by epistasis, whenever epistasis is weak demographic stochasticity can cause LD to drift in a quasi-neutral fashion and/or trigger infrequent oscillations between strains. Indeed, this is likely the source of the quasi-stable PSS observed in some stochastic models, in which the over-represented strain set infrequently changes [14]. For two diallelic loci, epistasis is weak if any of the following hold: there is no SSCP and hosts are likely to be infected multiple times (R_0 is large), both SSCP and ASCP are weak, and/or hosts are likely to be infected once and ASCP and SSCP are of similar strength.

Second, demographic stochasticity is expected to increase the chance loss of strains or antigens [17, 19, 46, 50, 60, 61]. This is especially relevant if there are an unequal number of alleles at the different loci: here, antigens at the more diverse locus are predicted to be lost in finite populations, particularly when cross-protection is high [19]. This can be explained as follows (Sup. Info. S6). If one locus is more antigenically diverse than the others, e.g., the two loci are $\{a, A\}$ and $\{B_1, B_2, B_3\}$, it is not possible for each antigen to be present at it's expected frequency, $p_A(t) = 1/2$ and $p_{B_i}(t) = 1/3$, *and* LD between each $\{A, B_i\}$ pair to be equal to D^*. This creates tension between selection on antigens (through the additive selection coefficients) and selection on strains (through epistasis in fitness), and this tension is modulated by cross-protection through the relative values of ASCP and SSCP. As cross-protection increases, selection maintaining antigens at their expected frequencies is reduced (SSCP decreases), while at the same time, epistasis becomes increasingly strong, favouring some strains (ASCP increases). Consequently, at the more diverse loci some antigens are forced to lower frequencies than expected and so are more likely to be stochastically lost. More generally, models incorporating an unequal number of alleles across loci disrupt the symmetry permitting the separation of the short- and long-term timescales as in equation (2); for example, when evolutionary oscillations occur, both allele frequencies and LD will fluctuate (rather than just LD). Understanding the resulting dynamics warrants future work.

The multilocus perspective clarifies the source of PSS and the explanations for antigenic variation

One of the strengths of the multilocus perspective is that it provides an evolutionary lens that clarifies the factors shaping the evolutionary dynamics of multistrain pathogens. For example, that epistasis is the source of PSS runs counter to the view the source is NFDS [14–16, 62]. Although NFDS is required to maintain genetic polymorphisms, it need not yield epistasis (e.g., if $\sigma_{\mathrm{PI}} = \sigma_{\mathrm{HI}} = 0$), and it is epistasis that produces strain over-representation. 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 [15, 16]. The multilocus perspective suggests an intermediate null model, in which there is NFDS but epistasis is

zero (e.g., $\sigma_{\mathrm{PI}} = 0$, $\sigma_{\mathrm{HI}} = 0$). In such a model, NFDS would maintain antigenic variation but nothing would maintain distinct allelic combinations, leading to stochastic fluctuations in LD.

Second, a recent review considered different explanations for the abundance of antigenic variation [4]; in addition to host heterogeneity and functional constraints, they discussed three hypotheses. First, each antigen additively contributes to host immunity (so $\sigma_{\mathrm{PI}} = 0$); second, immunity to a single antigen provides strong cross-protection [i.e., $\sigma_1 = \sigma_{i+1}$, $i > 0$; 2, 8]; 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 pairwise LD (DSS), the third simply means there is epistasis (e.g., $\sigma_{\mathrm{PI}} \neq 0$), which can lead to either situation, as well as 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 order of LD. Thus we would predict a positive correlation between the order of LD and the amount of allelic variation at each loci.

Multilocus models and vaccination

Understanding how genetically-diverse pathogen populations respond to vaccination has been a focus of recent work [63–66]. Our analysis provides some insight into 'evolutionarily optimal' vaccine design, as the type of LD present is generally indicative of how many antigens are required to generate a strongly protective immune response. However, our analysis has also emphasized some of the 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., 6, 7, 67, 68]. Second, different conclusions will be reached based upon whether SSCP is present, and if present, whether vaccination provides equivalent protection to full-antigenic or homologous immunity. Third, the pathogen population response will hinge upon whether an evolutionary bistability is possible, as the influx of vaccinal immunity may alter the type of PSS. Fourth, often vaccines are chosen to target a subset of strains, such as those linked to virulence factors [69]. 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, and should be a starting point for future work.

Conclusion

Understanding the dynamics of multistrain pathogens is important not only as an application of evolutionary biology, but as a question of public health. Yet combining pathogen genetic diversity with realistic epidemiological feedbacks remains a daunting task and existing approaches often obscure the evolutionary predictions. One way forward is adopting a population genetics perspective. As we have shown here, this allows us to disentangle the factors driving the evolutionary dynamics, providing greater insight into when, and why, pathogen strain structure occurs.

Symbol or acronym	Description
$p_i(t)$, $p_{ij}(t)$	Frequency of infections by strains carrying allele(s) i and ij, respectively, at time t.
$D(t)$	Pairwise linkage disequilibrium (LD) between alleles A and B.
D^*	Maximum or minimum value of the indicated order of LD, given the observed allele frequencies.
PSS	Pathogen strain structure; indicates the presence of LD of some order.
DSS	Discordant strain structure; is a specific form of PSS in which a set of strains sharing no alleles is over-represented in the long-term.
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.
σ_n, σ_h	Probability of infection of a host with immunity to $n = 0, 1, 2, ...$ antigens, or homologous immunity h, of the transmitted strain
σ_{PI}	Antigen-specific cross-protection (ASCP). If positive (resp. negative) immunity to one antigen is more (resp. less) protective than expected if the effects of having multiple antigens targeted by host immunity are additive.
σ_{HI}	Strain-specific cross-protections (SSCP). If positive, homologous immunity is stronger than full antigenic immunity. We assume this quantity is always non-negative.
$\mathcal{A}_\Delta(t)$	Indicates whether more uninfected hosts have full antigenic immunity to the set of strains $\{ab, AB\}$ ($\mathcal{A}_\Delta(t) > 0$) or $\{aB, Ab\}$ ($\mathcal{A}_\Delta(t) < 0$).
$\mathcal{S}_\Delta(t)$	Indicates whether more uninfected hosts have homologous immunity to the set of strains $\{ab, AB\}$ ($\mathcal{S}_\Delta(t) > 0$) or $\{aB, Ab\}$ ($\mathcal{S}_\Delta(t) < 0$).
f_n	Relative fitness of infection in host with immunity to n of it's antigens.
ϵ_{PI}	Indicates whether infection in host with immunity to one of it's antigens is more ($\epsilon_{\mathrm{PI}} > 0$) or less ($\epsilon_{\mathrm{PI}} < 0$) fit than expected.
R_0	Basic reproductive number.
θ	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] Maynard Smith, J., Smith, N. H., O'Rourke, M. & Spratt, B. G., 1993 How clonal are bacteria? *Proc. Natl. Acad. Sci.* **90**, 4384–4388.

[2] 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.

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

[4] 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.

[5] 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.

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

[7] 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.

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

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

[10] 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.

[11] 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.

[12] 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.

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

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] 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.

[20] 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.

[21] 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.

[22] 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.

[23] 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.

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

[25] 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.

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

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

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

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

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

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

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

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

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

[35] 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.

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

[37] 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.

[38] 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.

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

[40] 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.

[41] 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.

[42] 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.

[43] 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.

[44] 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.

[45] 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.

[46] 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.

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

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

[49] 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.

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

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

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

[53] Gandon, S. & Otto, S. P., 2007 The evolution of sex and recombination in response to abiotic or coevolutionary fluctuations in epistasis. *Genetics* **175**, 1835–1853.

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

[55] 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.

[56] 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.

[57] 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.

[58] 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.

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

[60] 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.

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

[62] 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.

[63] 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.

[64] 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.

[65] 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.

[66] 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.

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

[68] 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.

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

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

[71] 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.

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

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

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

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

[76] 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.

Supplementary Information

Contents

S1 Epidemiological model **S1**
S1.1 Basic model without adaptive immunity S1
S1.2 Specification of adaptive immunity S2
S1.3 Epidemiological dynamics and per-capita growth rates S2

S2 Biological connections between existing models **S3**
S2.1 Immune protections S3
S2.2 Acquisition of immunity S4
S2.3 Co-infection S5

S3 Separation of evolutionary and epidemiological dynamics **S6**
S3.1 The additive selection coefficients and epistasis in fitness S7
S3.2 Relationship between $\mathcal{A}_\Delta(t)$ and $\mathcal{S}_\Delta(t)$ S8
S3.3 Equations for the evolutionary dynamics S8
S3.4 Fast- and slow-timescales in the evolutionary dynamics S9

S4 Different types of immune protections **S10**
S4.1 Epistasis must share same sign as LD for LAS S10
S4.2 Epistasis at equilibrium S11

S5 The link between strain frequencies and the immunological landscape **S12**

S6 Variation in allelic diversity across loci **S14**

S7 More than two loci **S15**
S7.1 Fast- and slow-timescales, and different types of strain structure S18
S7.2 Strength of antigen-specific cross-protections S19
S7.3 SSCP promotes oscillations and maintains diversity S21

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 β while infections are naturally cleared by the host at a per-capita rate μ. 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 α, respectively. Hosts cannot be simultaneously infected by multiple strains (no co-infection); we later discuss how co-infection affects the dynamics. Let $S(t)$ and $I(t)$ denote the density of uninfected hosts, and total density of

infected hosts, respectively, at time t. Then the epidemiological dynamics are

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

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 (antigens) at each of these loci which can potentially reduce:

(i) The susceptibility to infection, and we let σ_x^g denote the probability of infection of a host with immune history x by strain (genotype) g.

(ii) The transmissibility of an infection, and we let β_x^g denote the transmissibility of an infection by strain g in a host with immune history x.

(iii) The duration of infection, and we let μ_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, and we let α_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 (homologous immunity). In particular, for a host with immunity to $n = 0, 1, 2, ...$ antigens of strain g, the immune protections are

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

while 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 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 it's immune history. For example, for a two diallelic loci pathogen genome, $\{a, A\}$ and $\{b, B\}$, then 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

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

$$\begin{aligned}\frac{\mathrm{d}S_x}{\mathrm{d}t} &= b\mathbf{1}_0(x) - dS_x - \sum_g \sigma_x^g \sum_z \beta_z^g I_z^g S_x + \Gamma_x \\ \frac{\mathrm{d}I_x^g}{\mathrm{d}t} &= \sum_z \beta_z^g I_z^g \sigma_x^g S_x - (d + \mu_x^g + \alpha_x^g) I_x^g + \Delta_{g,x}\rho\end{aligned} \tag{S2}$$

where $\mathbf{1}_0(x) = 1$ if $x = 0$ and 0 otherwise, $\Delta_{g,x}\rho$ is the net change due to recombination of strain g infections in hosts with immune history x (for now, left arbitrary), and Γ_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,

$$\Gamma_x = \sum_g \sum_{z:g\cup z=x} \mu_z^g I_z^g. \tag{S3}$$

Using this model, we 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:

$$r_g = \sum_x \frac{\mathrm{d}I_x^g/\mathrm{d}t}{I^g} = \sum_x \sum_z \beta_z^g \frac{I_z^g}{I^g} \sigma_x^g S_x - d - \sum_x (\mu_x^g + \alpha_x^g) \frac{I_x^g}{I^g}, \tag{S4}$$

where $I^g = \sum_x I_x^g$ is the total density of strain g. In Box 1 we present a version of this model where we have restricted immune protections to reducing σ_x^g, the susceptibility of host with immune history x to infection by strain g.

S2 Biological connections between existing models

Next, we consider the relationship between model (S2) and existing models. Multistrain models typically diverge 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 form of Γ_x; and

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

In what follows we discuss each of these in more detail, and in Table S2 we summarize how the model (S2) relates to existing models. Before doing so, it is important to note that the combination of assumptions employed by previous models are typically made because they reduce the number of equations; this then makes the problem more tractable for 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 [26], who provide a more 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, as well as how prior infections combine to determine the strength of immunity.

First, most models assume immunity reduces transmissibility or susceptibility; the choice of which is typically for model-specific mathematical reasons. For example, 'status'-based models often assume reduced susceptibility [9, 23, 32], while 'history'-based models often assume reduced transmissibility [2, 5, 8, 12] (but not always, [10, 33]). More recently, simulation models have assumed immune protections reduce duration of carriage, as this is more

biologically realistic for malaria [14–16]. Finally, some models of antibody-dependent enhancement have allowed immunity to reduce virulence-related mortality [43, 44]. Although immune protections reducing virulence-related mortality can have a qualitatively different effect [e.g., 43], it is generally believed that there is not a qualitative difference between protections reducing susceptibility, transmissibility, or clearance [11, 22]. Perhaps the most meaningful difference is that if immunity reduces susceptibility, hosts are less likely to acquire immunity to different strains, provided immunity is only acquired after successful infection [11].

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 $x'' = \{Ab, aB\}$, so we are focusing upon a pathogen genome consisting of two diallelic loci. The most common approach is to assume immunity against a particular strain is determined by the *single* most antigenically-similar previously infecting strain [2, 5, 8, 12, 21]. Using the immune history x'',

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

Less commonly, some models assume that immunity is the *product* of the individual cross-protections against previously infecting strains [10, 20]. For the immune history x'',

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

Finally, immunity to each antigen can also be treated independently, and so the strength of immunity is *antigen-specific* and depends upon the total number of antigens to which the host has protection against [14–16]. For the immune history x'',

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

Note that although some models restrict immune memory to only the most recent infections [60, 70], we do not consider this further here as this is more similar to a form of waning immunity.

S2.2 Acquisition of immunity

The model of (S2) 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 [9, 32], and is a commonly taken approach [2, 8, 10, 33]. An alternative to history-based models are so-called 'status-based' models [9, 23, 32]. Status-based models assume 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 antigenically similar strain, where the probability of acquiring protection against a particular strain depends upon the antigenic similarity between that strain and the other strains to which the host has protection against [9, 23, 32].

To convert model (S2) to 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

$$\Gamma_x = \sum_g \sum_z \mu_z^g p_{z\to x}^g I_z^g. \tag{S5}$$

Under these assumptions, model (S2) is a status-based model.

S2.3 Co-infection

The model of (S2) assumes that transmission only occurs to uninfected hosts, and thus hosts cannot be simultaneously infected by different strains (co-infection). This assumption was made primarily for modelling clarity, however, it is biologically reasonable if, for example, the pathogen is not circulating at a high prevalence, infections are acute, and/or infection induces short-term strain-transcending immunity. Yet these conditions need not apply generally, and in fact, co-infection is a common attribute of multistrain models. Virtually all non-simulation models that include co-infection make the following assumptions [2, 8, 11, 40, 50]:

(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, while 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 [54, 55].

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

These assumptions alter the specification of Γ_x, as well as the biological interpretation of $S_x(t)$ in (S2). In particular, $S_x(t)$ now denotes the density of hosts with immune history x, irrespective of their current infection status, while

$$\Gamma_x = \sum_g \sum_{z_1 : z_1 \cup g = x} \sigma_{z_1}^g S_{z_1} \sum_{z_2} \beta_{z_2}^g I_{z_2}^g. \tag{S6}$$

Importantly, because homologous immunity is sterilizing ($\sigma_h = 0$), and acquired at 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.

Treating co-infecting strains as non-interacting is somewhat biologically unsatisfying, given a host simultaneously infected with n strains is n-fold as infectious as a host infected with one. This means genetic diversity is unreasonably easily maintained. For example, Buckee et al [18] studied the evolution of hypervirulent lineages of *Neisseria meningitidis* using a model in which each strain is defined by it's sequence type (ST) 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' strains that can co-infect the same host (even if they share the same ST), thus maintaining variation in virulence. Note that incorporating virulence in models of co-infection is not trivial, and will mean co-infecting strains do interact [54, 55].

Model	Immune protections	Time immunity updated	History or status?	Effect of multiple prior infections	Coinfection	Sign of ASCP	Presence of SSCP
[2, 8]	$\beta_{n+1} = \beta_0(1-\gamma)$ $\sigma_h = 0$	infection	history	single	yes	positive	yes
[5, 12]	$\beta_1 = \beta_0(1-\gamma)$ $\beta_{n+2} = \beta_0(1-\delta)$ $\sigma_h = 0$	infection	history	single	yes	positive	yes
[9, 23, 32]	$\sigma_{n+1} = 0$ $\sigma_h = 0$	infection	status	single	yes	positive	no
[14–16]	$\mu_n = \mu_0(n+1)$	clearance	history	antigen-specific	yes	positive	no
[21]	$\sigma_n = \left(\frac{N-n}{N}\right)^\gamma$ $\sigma_h = 0$	infection	history	single	yes	positive	yes
[8, 17]	$\sigma_n = \left(1 - \left(\frac{n}{N}\right)^{\frac{1}{\gamma}}\right)^\gamma$ $\sigma_h = 0$	infection	history	single	yes	positive	yes
[10]	$\sigma_0 = 1$ $\sigma_h = 0$ $\sigma_x^g = \prod_{g_0 \in x} \sigma_{g_0}^g$	infection	history	product	yes	positive	yes
Model (S1)	variable	either	either	variable	no	variable	variable

Table S2: **Summary of key models and their relationship to the framework used here.** Here, γ and δ are positive parameters, N is the total number of loci. ASCP is defined as $f_0 - 2f_1 + f_2$ and SSCP is defined as $f_2 - f_h$ where $f_i = \sigma_i\beta_i/(\mu_i + \alpha_i + d)$. Note that ASCP and SSCP 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 (S2). There have also been various models whose specification of cross-immunity assumes one-dimensional antigenic space [e.g., 20, 33] and so we do not consider them here.

S3 Separation of evolutionary and epidemiological dynamics

The number of equations associated with model (S2) make it both challenging to analyze and to interpret. Thus a variety of clever mathematical approximations have been employed to simplify the dynamics [e.g., 2, 9, 10, 33, 54, 71]. While these techniques are effective at reducing the (mathematical) dimensionality of the problem, they struggle to provide an

evolutionary context for the predictions [e.g., see discussion in 23]. Thus many important results lack a clear biological interpretation. One way to provide greater clarity on the evolutionary process is to first separate the variables describing the epidemiological dynamics from those describing the evolutionary dynamics [72, 73]; doing so allows identification, and interpretation, of the various evolutionary forces. The evolutionary variables are the allele (antigen) frequencies, and the linkage disequilibrium (LD), or nonrandom associations, between them [27], while the epidemiological variables include the total density of infections, $I(t)$, and the population distribution of immune histories across the susceptible hosts, $S_x(t)$. Here, we focus on extracting the evolutionary variables for a pathogen genome consisting of two diallelic loci; in Sections S6 and S7 we extend this to 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, take strain ab to be the reference genotype. Then the additive selection coefficients for allele A and B are, respectively,

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

The selection coefficients indicate whether allele A (resp. B) is favoured, $w_A > 0$ (resp. $w_B > 0$) or disfavoured, $w_A < 0$ (resp. $w_B < 0$) on the genetic background b (resp. a). We also need to account for whether the fitness of an allele depends upon its genetic background, or *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 are favoured ($w_{AB} > 0$) or disfavoured ($w_{AB} < 0$) more than would be expected based upon the additive selection coefficients. Thus while the additive selection coefficients can be roughly thought of as controlling selection on individual alleles, epistasis can be thought of as controlling selection on particular combinations of alleles (in the case of two loci, genotypes or strains).

Although we can write down the additive selection coefficients and epistasis for the general model, here we assume immune protections only affect susceptibility. The reason for doing so is because if immune protections only affect susceptibility, the per-capita growth rates of equation (S4) can be written in a particularly simple form. However, in Section S4 we consider other immune protections using an equilibrium analysis. To write down w_A, w_B, and w_{AB}, we define

$$\sigma_{\text{PI}} \equiv \sigma_0 - 2\sigma_1 + \sigma_2 \quad \text{and} \quad \sigma_{\text{HI}} \equiv \sigma_2 - \sigma_h.$$

Conceptually, σ_{PI} can be thought of as measuring whether immunity to a single antigen provides more ($\sigma_{\text{PI}} > 0$) or less ($\sigma_{\text{PI}} < 0$) protection than expected, while σ_{HI} measures how much more protective homologous immunity is as compared to full-antigenic immunity. Therefore we refer to σ_{PI} as controlling *antigen-specific cross-protection* (ASCP) and σ_{HI} as controlling *strain-specific cross-protection* (SSCP). Using these composite parameters, the per-capita growth rate can be written

$$r_{ij} = \beta(\sigma_0 \mathcal{S} - (\sigma_0 - \sigma_1)(\mathcal{A}_{\{i\}} + \mathcal{A}_{\{j\}}) + \sigma_{\text{PI}} \mathcal{A}_{\{i,j\}} - \sigma_{\text{HI}} \mathcal{S}_{ij}) - d - \mu - \alpha, \tag{S7}$$

where $\mathcal{A}_{\{i\}}(t)$ and $\mathcal{A}_{\{i,j\}}(t)$ are the density 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

$$w_A = \beta(\sigma_0 - \sigma_1)(\mathcal{A}_{\{a\}} - \mathcal{A}_{\{A\}}) + \beta\sigma_{\mathrm{PI}}(\mathcal{A}_{\{A,b\}} - \mathcal{A}_{\{a,b\}}) - \beta\sigma_{\mathrm{HI}}(\mathcal{S}_{Ab} - \mathcal{S}_{ab}) \tag{S8}$$

$$w_B = \beta(\sigma_0 - \sigma_1)(\mathcal{A}_{\{b\}} - \mathcal{A}_{\{B\}}) + \beta\sigma_{\mathrm{PI}}(\mathcal{A}_{\{a,B\}} - \mathcal{A}_{\{a,b\}}) - \beta\sigma_{\mathrm{HI}}(\mathcal{S}_{aB} - \mathcal{S}_{ab}) \tag{S9}$$

while epistasis in fitness is

$$w_{AB} = \beta\sigma_{\mathrm{PI}}(\mathcal{A}_{\{A,B\}} + \mathcal{A}_{\{a,b\}} - \mathcal{A}_{\{A,b\}} - \mathcal{A}_{\{a,B\}}) - \beta\sigma_{\mathrm{HI}}(\mathcal{S}_{AB} + \mathcal{S}_{ab} - \mathcal{S}_{Ab} - \mathcal{S}_{aB}). \tag{S10}$$

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

If we define

$$\mathcal{A}_\Delta(t) = \mathcal{A}_{\{A,B\}} + \mathcal{A}_{\{a,b\}} - \mathcal{A}_{\{A,b\}} - \mathcal{A}_{\{a,B\}} \quad \text{and} \quad \mathcal{S}_\Delta(t) = \mathcal{S}_{AB} + \mathcal{S}_{ab} - \mathcal{S}_{Ab} - \mathcal{S}_{aB}, \tag{S11}$$

that is, $\mathcal{A}_\Delta(t)$ and $\mathcal{S}_\Delta(t)$ indicates whether more susceptible hosts have full antigenic-immunity or homologous immunity, respectively, to the set of strains $\{ab, AB\}$ (if positive) or $\{Ab, aB\}$ (if negative), we can rewrite epistasis in fitness (equation (S10)) as

$$w_{AB} = \beta\sigma_{\mathrm{PI}}\mathcal{A}_\Delta(t) - \beta\sigma_{\mathrm{HI}}\mathcal{S}_\Delta(t). \tag{S12}$$

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

$$\mathcal{A}_\Delta(t) = S_{AB}(t) + S_{ab}(t) - S_{Ab}(t) - S_{aB}(t). \tag{S13}$$

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 some cancellation:

$$\mathcal{S}_\Delta(t) = \mathcal{A}_\Delta(t) + S_{ab,AB}(t) - S_{aB,Ab}(t) + \mathcal{S}_{ab,AB}(t) - \mathcal{S}_{aB,Ab}(t) \tag{S14}$$

where $S_{ij,k\ell}(t)$ and $\mathcal{S}_{ij,k\ell}(t)$ indicate the density of uninfected hosts that have homologous immunity to *only*, and *at least*, strains ij and $k\ell$, respectively. Thus as discussed in the main text, in order for $\mathcal{S}_\Delta(t)$ and $\mathcal{A}_\Delta(t)$ to diverge in value, hosts must be infected multiple times during the course of their life (and retain memory of these infections).

S3.3 Equations for the evolutionary dynamics

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

$$p_A(t) = \frac{I^{Ab}(t) + I^{AB}(t)}{I(t)} \quad \text{and} \quad p_B(t) = \frac{I^{aB}(t) + I^{AB}(t)}{I(t)}, \tag{S15}$$

while the linkage disequilibrium between alleles A and B is

$$D(t) = \frac{I^{AB}(t)}{I(t)} - p_A(t)p_B(t) = \frac{I^{AB}(t)}{I(t)}\frac{I^{ab}(t)}{I(t)} - \frac{I^{Ab}(t)}{I(t)}\frac{I^{aB}(t)}{I(t)}. \tag{S16}$$

Then the evolutionary dynamics can be written

$$\frac{\mathrm{d}p_A}{\mathrm{d}t} = w_A p_A(1-p_A) + w_B D + w_{AB}(1-p_A)p_{AB} + \Delta_{p_B}\rho \tag{S17a}$$

$$\frac{\mathrm{d}p_B}{\mathrm{d}t} = w_B p_B(1-p_B) + w_A D + w_{AB}(1-p_B)p_{AB} + \Delta_{p_B}\rho \tag{S17b}$$

$$\frac{\mathrm{d}D}{\mathrm{d}t} = \sum_{k=A,B} w_k(1-2p_k)D + w_{AB}p_{AB}(p_{ab} - 2D) + \Delta_D\rho, \tag{S17c}$$

where $\Delta_x\rho$ indicates the effect of recombination on the variable x. Note that $p_{AB} = p_A p_B + D$ and $p_{ab} = (1-p_A)(1-p_B) + D$.

The additive selection coefficients, recombination, and epistasis in fitness need not be constant and instead vary with the epidemiological dynamics. But by writing out the evolutionary dynamics as we have done in (S17), we can identify the evolutionary forces at play and see 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

If recombination is unbiased, then it will not affect allele frequencies, $\Delta_{p_B}\rho = \Delta_{p_A}\rho = 0$ and will only act to remove LD from the population, that is, $\Delta_D\rho = -\rho D(t)$ for $\rho \geq 0$. In this case, there are no intrinsic biases favouring one allele over the other at a given loci, and so by symmetry we would expect the population will self-organize so that the per-capita growth rates of the set of strains $\{ab, AB\}$ are similar, and likewise the per-capita growth rates of the set of strains $\{Ab, aB\}$ [23]. Indeed, after some short-term transient dynamics owing to the initial conditions, strong NFDS operating through the additive selection coefficients will push $r_{AB} \to r_{ab}$ and $r_{Ab} \to r_{aB}$. At this point, the antigen frequencies are in quasi-equilibrium, $p_A(t), p_B(t) \approx 1/2$, and moreover, $w_A = w_B = -w_{AB}/2$. Consequently, over the long-term the set of equations (S17) reduces to the single equation

$$\frac{\mathrm{d}D}{\mathrm{d}t} \approx w_{AB}\left(\frac{1}{4} + D\right)\left(\frac{1}{4} - D\right) - \rho D. \tag{S18}$$

Because LD quantifies whether certain strains are more, or less, abundant than predicted based upon allelic frequencies, the presence of LD in the population is necessary for there to be pathogen strain structure or PSS (see the longer discussion in Section 2 of the main text). Thus over the long timescale, equation (S18) 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 strain set $\{ab, AB\}$ will be over-represented in the population, whereas if $D(t) < 0$, strain set $\{Ab, aB\}$ will be over-represented.

Assuming $\rho = 0$, equation (S18) has four potential equilibria: (1,2) when either strain set has competitively excluded the other ($D(t) = \pm 1/4$); (3) when one set of strains is over-represented, 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, defined to be any sustained dynamical behaviour of LD (so including cycles and chaos) may occur. In all cases, however, the dynamical behaviour of equation (S18) hinges on epistasis.

S4 Different types of immune protections

Next we consider the more general case in which immune protections may affect transmissibility, clearance, and virulence mortality, but that homologous immunity is equivalent to full antigenic immunity (so there is no SSCP). Although epistasis in fitness can be written out in full, it is more complicated than when immunity only affects susceptibility 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 rates will depend upon the frequency of strain ij infections found in a particular environment. Thus we instead focus upon deriving epistasis in fitness at equilibrium. Since there is effectively one epidemiological feedback, which is whether more hosts have antigenic immunity to one set of strains or the other (e.g., $\{ab, Ab\}$ or $\{Ab, aB\}$), if recombination is weak then there is no reason to expect epistasis to become zero and/or change sign at some non-zero amount of LD. Consequently, we should expect that if PSS occurs, it will be the maximum amount possible given the allele frequencies (i.e., $D(t) = D^*$); for example, if $p_A(t) = p_B(t) = 1/2$, then if PSS occurs, $|D(t)| \to 1/4$. It is these two equilibria that we will focus upon here.

S4.1 Epistasis must share same sign as LD for LAS

Here we focus upon one of the equilibria $(p_A, p_B, D) = (1/2, 1/2, \pm 1/4)$; our objective is to show that a necessary condition for it to be locally asymptotically stable (LAS) is that epistasis in fitness shares the same sign as LD.

To see this, suppose we take the derivative of the differential equation for LD, $\frac{\mathrm{d}D}{\mathrm{d}t}$, with respect to any of the population variables (arbitrarily indicated as X) under the assumption recombination is negligible ($\rho \approx 0$). Then we obtain

$$\begin{aligned}\frac{\partial}{\partial X}\frac{\mathrm{d}D}{\mathrm{d}t} = D\left((1-2p_A)\frac{\partial w_A}{\partial X} + (1-2p_B)\frac{\partial w_B}{\partial X}\right) \\ + (p_A p_B + D)((1-p_A)(1-p_B) - D)\frac{\partial w_{AB}}{\partial X} + w_A\frac{\partial[1-2p_A]D}{\partial X} \\ + w_B\frac{\partial[1-2p_B]D}{\partial X} + w_{AB}\frac{\partial}{\partial X}[(p_A p_B + D)((1-p_A)(1-p_B) - D)]. \quad \text{(S19)}\end{aligned}$$

Now consider evaluating (S19) when $p_A = p_B = 1/2$ and $D = 1/4$ or $D = -1/4$. In this case, $r_{AB} = r_{ab}$ and $r_{Ab} = r_{aB}$; thus $w_{AB} = -2w_A = -2w_B$. Therefore

$$\frac{\partial}{\partial X}\frac{\mathrm{d}D}{\mathrm{d}t} = \begin{cases} -2w_{AB}D & \text{if } X = D \\ 0 & \text{otherwise} \end{cases} \quad \text{(S20)}$$

Hence if we were to compute the Jacobian of the full system (that combines both the evolutionary and epidemiological variables) evaluated at either of these equilibria, the Jacobian will be block triangular, and one of the entries on the diagonal is given by equation (S20) (when $X = D$). Thus one of the eigenvalues will be $-w_{AB}/2$ (if $D = 1/4$) or $w_{AB}/2$ (if $D = -1/4$). It follows that a necessary condition for PSS to be LAS is that the sign of epistasis matches the sign of LD.

S4.2 Epistasis at equilibrium

Based upon the preceding results, we therefore seek to calculate epistasis at equilibrium. To do so, define

$$f_n \equiv \frac{\sigma_n \beta_n}{\mu_n + \alpha_n + d}; \tag{S21}$$

this can be thought of as the relative fitness of infections in hosts with immunity to n antigens of the infecting strain. This is because, all else being equal, with probability σ_n an infection is established in a host with immunity to n of it's antigens; 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 β_n.

Next, at equilibrium, the basic reproductive number of strain ij [e.g., see 74], which we will denote $\bar{R}_{ij}$, can be written

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

Now, the per-capita growth rate, r_{ij} is a measure of fitness in continuous time, whereas the basic reproductive number, $\bar{R}_{ij}$, is a measure of fitness in discrete time [74]. Although 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., 75, 76]. Without loss of generality, focus upon the equilibrium in which the over-represented stain set 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 $\bar{R}_{ij} \geq 0$,

$$\bar{w}_{AB} \propto 2(1 - \bar{R}_{Ab}) = 2(\bar{R}_{ab} - \bar{R}_{Ab}) = 2\left(f_0 - 2f_1 + f_2\right)\left(\bar{\mathcal{A}}_{\{a,b\}} - \bar{\mathcal{A}}_{\{A,b\}}\right) = \epsilon_{\mathrm{PI}}\bar{\mathcal{A}}_{\Delta}, \tag{S22}$$

where proportionality is up to a positive constant, $\epsilon_{\mathrm{PI}} = f_0 - 2f_1 + f_2$, 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 $\epsilon_{\mathrm{PI}} > 0$, epistasis will share the same sign as LD.

S5 The link between strain frequencies and the immunological landscape

As discussed in the main text, various epidemiological assumptions can disrupt the correlation between which strain set is most abundant in the population, and which strain set has more hosts with immunity to it (so $D(t)$ and $\mathcal{A}_\Delta(t)$, $\mathcal{S}_\Delta(t)$ do not share the same 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, $\sigma_h = \sigma_2$. First, we assume that uninfected hosts lose immunity to each antigen at a per-capita rate of δ. For example, an uninfected host with immunity to antigens $\{a, A, b\}$ becomes immune to antigens $\{a, A\}$ at a per-capita rate δ. Second, we assume that following clearance of infection, hosts gain immunity to a single, randomly selected antigen with probability p, and both antigens with probability $1 - p$.

Since we assume homologous immunity is of the same strength as full-antigenic immunity, let us assume that x in this case only tracks the unique antigens to which a host has immunity to (e.g., in the previous formulation, $x = \{ab, Ab\}$, would become $x = \{a, A, b\}$. Then the two assumptions above leave the equations for the I_x^{ij} unchanged, but mean the net rate at which immune history x is gained and lost is

$$\Gamma_x = \frac{1}{2}\sum_g \Big(2\sum_{z:g\cup z=x}(1-p)\mu_z^g I_z^g + \sum_{z:g\cup(z\setminus z_1)=x} p\mu_z^g I_z^g + \sum_{z:g\cup(z\setminus z_2)=x} p\mu_z^g I_z^g\Big) + \delta\Big(\sum_{z:(z\setminus z_i)=x} S_z - \sum_{i\in x} S_x\Big) \tag{S23}$$

where $(z \setminus z_i)$ means 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. S1), we observe two differences.

(1) If immunity wanes in an antigen-by-antigen fashion (i.e., $\delta > 0$), then when ASCP is maximized ($\sigma_1, \sigma_2 \to 0$), strain structure will not form (Fig. S1**a**).

The logic here is as follows. When $\sigma_1, \sigma_2 \to 0$, hosts with immunity to one strain (say AB) can be infected by at most the other strain in the strain set (ab). When this occurs, this will produce 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 one strain set or the other, irrespective of which strain set is more abundant. However, as these hosts become susceptible to infection, the more abundant strain set will infect, and deplete, it's available hosts faster. On balance, this will create an excess of hosts with immunity to the less abundant strain set, meaning that $\mathcal{A}_\Delta(t)$ will have the opposite sign to LD. Since $\sigma_{\mathrm{PI}} = \sigma_0 > 0$, epistasis in fitness will also have the opposite sign to population LD. Thus strain structure cannot form, despite ASCP being maximized.

(2) Second, 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 ASCP is weaker than expected, $\sigma_{\mathrm{PI}} < 0$ (Fig. S1**b**).

The logic here is as follows. Suppose strain set $\{ab, AB\}$ is over-represented in the population ($D(t) > 0$). Therefore, 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). But because each infection can at most provide immunity to a single antigen (since $p \to 1$), when the two randomly selected antigens are at different loci, following a secondary infection the host will then have immunity to either the antigens Ab or aB, and so will have immunity to a strain belonging to the under-represented strain set, $\{Ab, aB\}$. But this will mean that $\mathcal{A}_\Delta(t)$ will have opposite sign to $D(t)$, that is, more hosts will have immunity to the less-abundant strain set, and so if $\sigma_{\mathrm{PI}} < 0$, this provides an advantage for the more abundant strain set. Thus strain structure will occur when ASCP is weaker then expected.

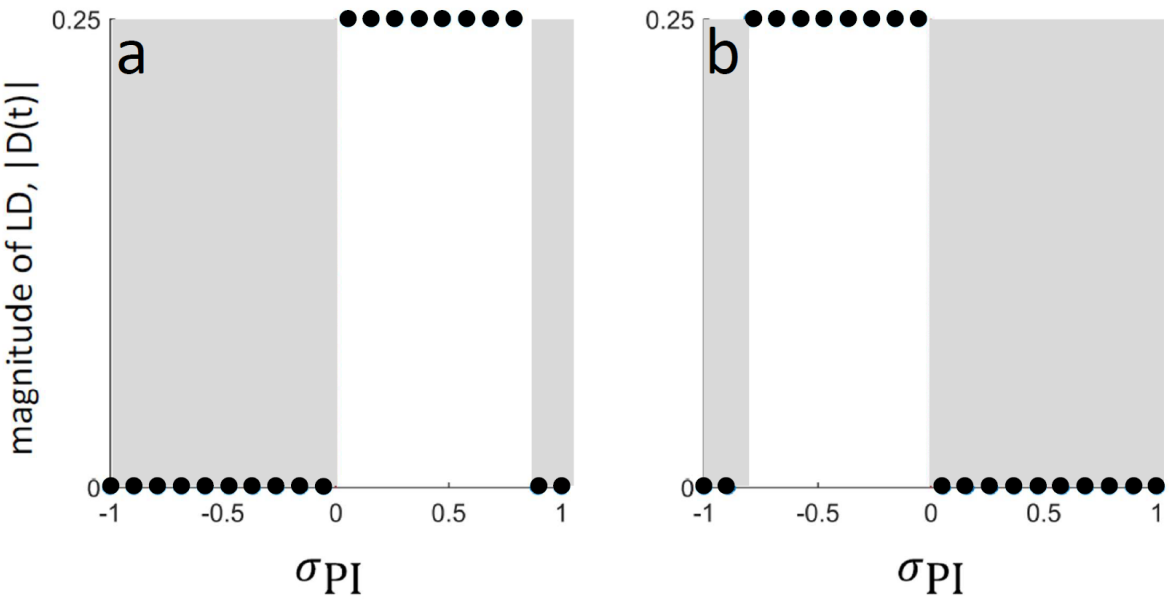

Figure S1: **Disrupting the link between strain frequencies and the immunological landscape**. Here we show two ways in which LD need not share the same sign as $\mathcal{A}_\Delta(t)$. In panel **a**, immunity wanes in an antigen-by-antigen fashion. In panel **b**, hosts acquire immunity to both antigens with probability $1 - p$, and a single, randomly chosen antigen with probability p.

Note that these predictions are sensitive to how immune protections are modelled. In particular, when immunity reduces susceptibility typically only when a host is successfully infected will it potentially gain immunity to the infecting strain [11]. 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. Generally speaking, these differences are not expected to have a large effect, at least in models with very simple epidemiological assumptions [11]. However, as the underlying epidemiological model becomes more complex, the evolutionary dynamics can become more sensitive to the type of immune protections as the above examples illustrate.

S6 Variation in allelic diversity across loci

The model featured in Section S3 assumes that there are an equal number of alleles at each locus. Here we consider what happens if at one locus there are two alleles $i \in \{A_1, A_2\}$, while at the other there are three, $j \in \{B_1, B_2, B_3\}$, and so 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. Then the evolutionary variables are the frequency of alleles A_2 and B_j at time t, which are

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

as well as 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

$$w_A \equiv r_{A_2B_1} - r_{A_1B_1} \quad \text{and} \quad w_{B_j} \equiv r_{A_1B_j} - r_{A_1B_1}, \; j = 2, 3, \tag{S24}$$

while the two epistasis in fitness terms 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,

$$\begin{aligned} w_{A_2B_j} = \beta\sigma_{\mathrm{PI}}(\mathcal{A}_{\{A_2,B_j\}} + \mathcal{A}_{\{A_1,B_1\}} - \mathcal{A}_{\{A_2,B_1\}} - \mathcal{A}_{\{A_1,B_j\}}) \\ - \beta\sigma_{\mathrm{HI}}(\mathcal{S}_{A_2B_j} + \mathcal{S}_{A_1B_1} - \mathcal{S}_{A_2B_1} - \mathcal{S}_{A_1B_j}) \end{aligned} \tag{S25}$$

where as before, $\mathcal{A}_{\{i\}}$ and $\mathcal{A}_{\{i,j\}}$ denotes the density of all the uninfected hosts with immunity to at least the antigen i and antigens i and j, respectively, while $\mathcal{Z}_{ij}$ 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 skip this, and simply interpret the simulated dynamics. To do so, the key is to recognize that additive selection favours the frequency of each allele at a given locus to be equal, that is, $p_{A_2} = 1/2$, $p_{B_j} = 1/3$, while epistatic selection either favours LD to increase or decrease in magnitude, depending upon whether $D_{ij}(t)$ and w_{ij} are of the same or opposite sign, respectively, and an intermediate level of LD only when $w_{ij} = 0$ for some non-zero D_{ij}. In the two-allele, two-locus model of Section S3, both of these could be simultaneously achieved. One consequence of this was that there were two timescales: a short timescale over which allele frequencies were driven to equilibrium $p_k \to 1/2$, and a longer timescale along which LD evolved (see equation (S18)). When there are more alleles at one locus then the other, this need not be true, and thus there is competition between epistatic and allelic selection, which can induce more complex dynamics and long-term outcomes.

To understand this, first consider the case in which there is no SSCP, $\sigma_{\mathrm{HI}} = 0$. As was true for the model of Section S3, there is a single epidemiological feedback on epistasis and

so oscillatory evolutionary dynamics are not possible and the population will converge to an evolutionary equilibrium. If $\sigma_{\rm PI} < 0$, then epistasis favours linkage equilibrium, $D_{A_2B_i} = 0$, while 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. S2). However, when $\sigma_{\rm PI} > 0$ epistasis favours pairwise LD attaining D^*, but 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 $\sigma_{\rm PI}$ 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 other will be in linkage equilibrium (so two genotypes will have been eliminated from the population). As $\sigma_{\rm PI}$ increases, the allele frequencies diverge further from their unlinked 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. S2), 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. S2). Eventually when $\sigma_{\rm PI}$ is sufficiently large, $p_{A_i} = 1/2$, $p_{B_j} = p_{B_k} = 1/4$, $p_{B_\ell} = 1/2$ for $j \neq k \neq \ell = 1, 2, 3$ while $D_{A_2B_j}(t) = D^*$ for each pairwise combination (so three genotypes will have been eliminated from the population) (Fig. S2). Thus the sign of $\sigma_{\rm PI}$ still determines whether PSS forms or not, but now additionally its strength can affect the amount of PSS observed. Importantly, however, even if $\sigma_{\rm PI}$ is maximized, the PSS observed will not be the DSS of Gupta et al [2, 8].

Second, suppose there is SSCP, $\sigma_{\rm HI} > 0$. As was the case of the two-locus, two-allele model, evolutionary oscillations are possible. But here there is a important difference: because there is no longer a short- and long-timescale, the allele frequencies will *not* reach an equilibrium. Thus both LD and the allele frequencies will oscillate (Fig. S3).

S7 More than two loci

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

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

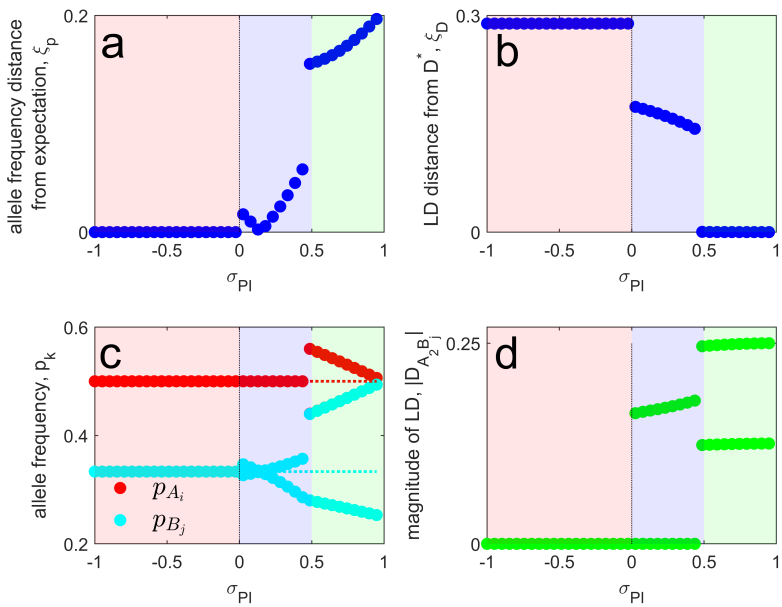

Figure S2: **Consequences of variation in allelic diversity across loci**. An unequal number of alleles across loci can produce conflict between selection on alleles (additive selection coefficients) and on strains (epistasis) because allele frequencies cannot be at their expected values ($p_{A_i} = 1/2$, $p_{B_j} = 1/3$) *and* the LD between each $\{A_2, B_j\}$ pair satisfy $D_{A_2B_j} = D^*$. If $\sigma_{\rm PI} < 0$ (red shaded region), there is no conflict because epistasis does not favour strain structure. If $\sigma_{\rm PI} > 0$ 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 $\sigma_{\rm PI}$ 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: $\beta = 1$, $\mu = 0.5$, $b = d = 0.01$, $\rho = 10^{-5}$, $\sigma_0 = 1$, $\sigma_2 = \sigma_h = 0$, and $\sigma_1 \in [\sigma_0, \sigma_2]$.

and the triple LD,

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

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

$$r_{ijk} = \beta\big(\sigma_0\mathcal{S} - (\sigma_0 - \sigma_1)\sum_{\ell=i,j,k}\mathcal{A}_{\{\ell\}} + \sigma_{\rm PI}^{(1)}\sum_{\ell_1\ell_2\in\{ij,ik,jk\}}\mathcal{A}_{\{\ell_1,\ell_2\}}\big) - \beta\big((\sigma_{\rm PI}^{(1)} - \sigma_{\rm PI}^{(2)})\mathcal{A}_{\{i,j,k\}} + \sigma_{\rm HI}\mathcal{S}_{ijk}\big) - \mu - \alpha - d \quad \text{(S26)}$$

where $\sigma_{\rm HI} \equiv \sigma_3 - \sigma_h$, and

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

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

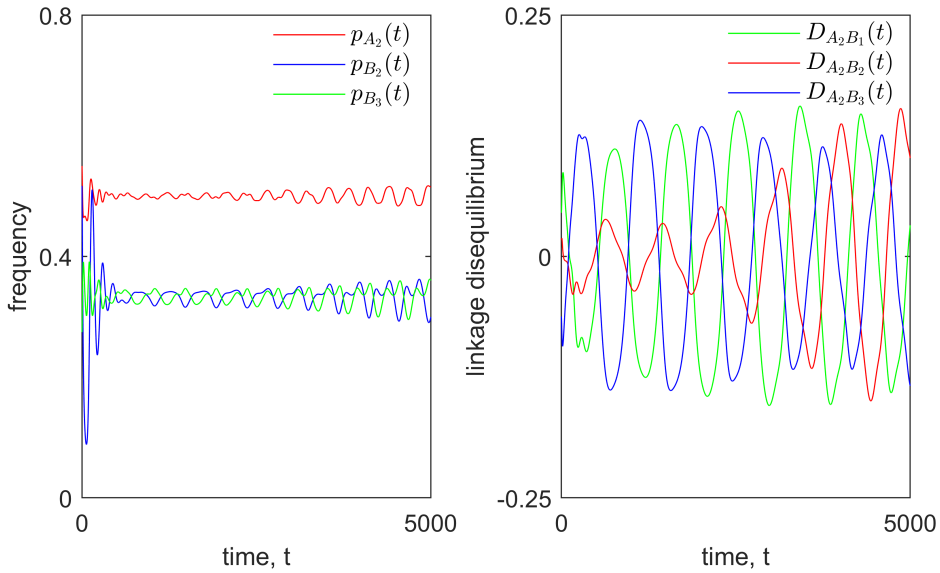

Figure S3: **Conflict between selection on alleles and strains generates more complex evolutionary oscillations**. When there is variation in allelic diversity across loci, there is no longer a fast- and slow-timescale. This is because allele frequencies cannot be at their expectation **and** all pairwise 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 $\sigma_h = 0$.

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

$$w_A = r_{Abc} - r_{abc} = -\beta(\sigma_0 - \sigma_1)(\mathcal{A}_{\{A\}} - \mathcal{A}_{\{a\}}) + \beta\sigma_{\mathrm{PI}}^{(1)}(\mathcal{A}_{\{A,b\}} + \mathcal{A}_{\{A,c\}} - \mathcal{A}_{\{a,b\}} - \mathcal{A}_{\{a,c\}}) + \beta(\sigma_{\mathrm{PI}}^{(2)} - \sigma_{\mathrm{PI}}^{(1)})(\mathcal{A}_{\{A,b,c\}} - \mathcal{A}_{\{a,b,c\}}) - \beta\sigma_{\mathrm{HI}}(\mathcal{S}_{Abc} - \mathcal{S}_{abc}). \quad \text{(S27)}$$

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

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

which using our per-capita growth rates can be written as

$$w_{ij} = \beta\sigma_{\mathrm{PI}}^{(1)}\mathcal{A}_{\Delta}^{(ij)} + \beta(\sigma_{\mathrm{PI}}^{(2)} - \sigma_{\mathrm{PI}}^{(1)})\mathcal{A}_{\Delta}^{(ij)/\ell} - \beta\sigma_{\mathrm{HI}}\mathcal{S}_{\Delta}^{(ij)/\ell}. \quad \text{(S28)}$$

Here $\mathcal{A}_{\Delta}^{(ij)}$ measures whether more hosts have immunity to strains with antigens $\{ij, \hat{i}\hat{j}\}$ or $\{\hat{i}j, i\hat{j}\}$, irrespective of genetic background (note that $\hat{i}$ indicates the other allele at the indicated locus, e.g., if $i = A$, then $\hat{i} = a$). For example,

$$\mathcal{A}_{\Delta}^{(AB)} = \mathcal{A}_{\{A,B\}} + \mathcal{A}_{\{a,b\}} - \mathcal{A}_{\{A,b\}} - \mathcal{A}_{\{a,B\}}. \quad \text{(S29)}$$

On the other hand, $\mathcal{A}_{\Delta}^{(ij)/\ell}$ and $\mathcal{S}_{\Delta}^{(ij)/\ell}$ measures whether more uninfected hosts have immunity to antigens i, j and the reference allele $\ell \in \{a, b, c\}$ at the third locus, or homologous

immunity to the strain whose genotype consists of alleles i, j, ℓ. For example,

$$\begin{aligned}\mathcal{A}_{\Delta}^{(AB)/c} &= \mathcal{A}_{\{A,B,c\}} + \mathcal{A}_{\{a,b,c\}} - \mathcal{A}_{\{A,b,c\}} - \mathcal{A}_{\{a,B,c\}},\\ \mathcal{S}_{\Delta}^{(AB)/c} &= \mathcal{S}_{ABc} + \mathcal{S}_{abc} - \mathcal{S}_{Abc} - \mathcal{S}_{aBc}.\end{aligned}$$

Finally, when there are three loci, we need to also 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

$$\begin{aligned}w_{ABC} &= r_{ABC} + r_{Abc} + r_{aBc} + r_{abC} - r_{abc} - r_{ABc} - r_{AbC} - r_{aBC}\\ &= \beta(\sigma_{\mathrm{PI}}^{(2)} - \sigma_{\mathrm{PI}}^{(1)})\mathcal{A}_{\Delta}^{(2)} - \beta\sigma_{\mathrm{HI}}\mathcal{S}_{\Delta}^{(2)}\end{aligned}$$

where $\mathcal{X}_{\Delta}^{(2)}$ indicates if more hosts have immunity to the antigens ($\mathcal{X} = \mathcal{A}$) or homologous immunity to the strains ($\mathcal{X} = \mathcal{S}$) belonging to one of the two set of strains whose members share no more than one antigen in common, that is,

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

In particular,

$$\begin{aligned}\mathcal{A}_{\Delta}^{(2)} &= \mathcal{A}_{\{A,B,C\}} + \mathcal{A}_{\{A,b,c\}} + \mathcal{A}_{\{a,B,c\}} + \mathcal{A}_{\{a,b,C\}} - \mathcal{A}_{\{A,B,c\}} - \mathcal{A}_{\{A,b,C\}} - \mathcal{A}_{\{a,B,C\}} - \mathcal{A}_{\{a,b,c\}},\\ \mathcal{S}_{\Delta}^{(2)} &= \mathcal{S}_{ABC} + \mathcal{S}_{Abc} + \mathcal{S}_{aBc} + \mathcal{S}_{abC} - \mathcal{S}_{ABc} - \mathcal{S}_{AbC} - \mathcal{S}_{aBC} - \mathcal{S}_{abc},\end{aligned}$$

We are then in a position to write down the equations describing the evolutionary dynamics. As these are complicated, we will not do this here, but in what follows we use these equations to extract general qualitative predictions. Note that our emphasis is upon biological interpretability, and so we do not attempt to rigorously characterize the dynamics.

S7.1 Fast- and slow-timescales, and different types of strain structure

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

$$\begin{aligned}D_{AB}(t) &\approx \frac{p_{ABc}(t) + p_{abc}(t) - p_{aBC}(t) - p_{AbC}(t)}{2},\\ D_{AC}(t) &\approx \frac{p_{AbC}(t) + p_{abc}(t) - p_{aBC}(t) - p_{ABc}(t)}{2},\\ D_{BC}(t) &\approx \frac{p_{aBC}(t) + p_{abc}(t) - p_{AbC}(t) - p_{ABc}(t)}{2},\end{aligned} \tag{S30}$$

while triple LD becomes

$$D_{ABC}(t) \approx \frac{p_{ABC}(t) - p_{abc}(t)}{2}. \tag{S31}$$

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

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

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

is over-represented. 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 present ($D_{ij}(t) = 0$), then along the slow timescale

$$\frac{\mathrm{d}D_{ABC}}{\mathrm{d}t} \approx w_{ABC}\left(\frac{1}{8} - D_{ABC}\right)\left(\frac{1}{8} + D_{ABC}\right). \tag{S32}$$

By inspection of equation (S32), in order for triple LD to build up and be maintained, triple LD (D_{ABC}) and triple epistasis (w_{ABC}) must share the same sign.

(2) When one of the four sets of strains whose members share no alleles in common is over-represented, that is, one of

$$\{ABC, abc\}, \quad \{Abc, aBC\}, \quad \{aBc, AbC\}, \quad \text{or} \quad \{abC, ABc\}, \tag{S33}$$

is over-represented. 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 (three of the sets in (S33) are competitively excluded) then $D_{ij}(t) = \pm 1/4$, and $D_{ABC}(t) = 0$.

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

$$\frac{\mathrm{d}D_{ij}}{\mathrm{d}t} \approx \left(w_{ij} + \frac{w_{ABC}}{2}\right)\left(\frac{1}{4} - D_{ij}\right)\left(\frac{1}{4} + D_{ij}\right) + \sum_{k\ell}\left(w_{k\ell} + \frac{w_{ABC}}{2}\right)\left(\frac{D_{\hat{k}\hat{\ell}}}{4} - D_{ij}D_{k\ell}\right). \tag{S34}$$

From equation (S34) it is no longer necessary that w_{ij} must share the same sign as D_{ij} in order for pairwise LD to be maintained. Instead, $w_{ij} + w_{ABC}/2$ and D_{ij} need to share the same sign. Note that we can write

$$w_{ij} + \frac{w_{ABC}}{2} = \beta\sigma_{\mathrm{PI}}^{(1)}\mathcal{A}_{\Delta}^{(ij)} + \frac{\beta}{2}(\sigma_{\mathrm{PI}}^{(2)} - \sigma_{\mathrm{PI}}^{(1)})(\mathcal{A}_{\Delta}^{(ij)/\ell} + \mathcal{A}_{\Delta}^{(ij)/\hat{\ell}}) - \frac{\beta}{2}\sigma_{\mathrm{HI}}(\mathcal{S}_{\Delta}^{(ij)/\ell} + \mathcal{S}_{\Delta}^{(ij)/\hat{\ell}}). \tag{S35}$$

S7.2 Strength of antigen-specific cross-protections

Next, suppose there is no strain-specific cross-protection ($\sigma_{\mathrm{HI}} = 0$). Here we want to understand how the strength of antigen-specific cross-protections, $\sigma_{\mathrm{PI}}^{(n)}$, affects the evolutionary dynamics. Recall that $\sigma_{\mathrm{PI}}^{(n)}$ measures whether immunity to n antigens carried by a strain provides more or less protection than expected, based upon the protection offered against $n-1$ or $n+1$ antigens.

If we consider $(\sigma_{\mathrm{PI}}^{(1)}, \sigma_{\mathrm{PI}}^{(2)})$ space, there are two curves that divide up the qualitatively different evolutionary behaviour. In what follows we consider these two curves; the results are presented in the main text Figure 5. Throughout, we assume that the more abundant strain set also has more hosts with antigenic immunity to it, that is, $\mathcal{A}_{\Delta}^{(2)}$ shares the same sign as D_{ABC}, and $\mathcal{A}_{\Delta}^{(ij)}$ shares the same sign as D_{ij}. We will not consider ways in which these relationships could breakdown as in Section S5.

Curve 1: $w_{ABC} = 0$. This occurs when $\sigma_{\mathrm{PI}}^{(2)} = \sigma_{\mathrm{PI}}^{(1)}$, that is, the relative strength 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 sign of w_{ABC} and D_{ABC} to agree, it can only be maintained if $\sigma_{\mathrm{PI}}^{(2)} > \sigma_{\mathrm{PI}}^{(1)}$.

Curve 2: $w_{ij} + w_{ABC}/2 = 0$. From equation (S34), this curve species when pairwise LD can, and cannot, form. Since this curve depends upon the epidemiological dynamics in addition to the parameters $\sigma_{\mathrm{PI}}^{(n)}$, it is not straightforward to precisely specify the curve in $(\sigma_{\mathrm{PI}}^{(1)}, \sigma_{\mathrm{PI}}^{(2)})$ space. However, note that in equation (S35), 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

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

Since each of these terms requires hosts be infected by different pathogen strains, 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 $\mathcal{A}_{\Delta}^{(ij)/r\hat{r}} \ll \mathcal{S}_{\Delta}^{(ij)/r} + \mathcal{A}_{\Delta}^{(ij)/\hat{r}}$, and so equation (S35) becomes

$$w_{ij} + \frac{w_{ABC}}{2} \approx \frac{\beta}{2}\big(\sigma_{\mathrm{PI}}^{(1)} + \sigma_{\mathrm{PI}}^{(2)}\big)\big(\mathcal{A}_{\Delta}^{(ij)/r} + \mathcal{A}_{\Delta}^{(ij)/\hat{r}}\big), \tag{S36}$$

Hence $w_{ij} + \frac{w_{ABC}}{2} \approx 0$ when $\sigma_{\mathrm{PI}}^{(1)} = -\sigma_{\mathrm{PI}}^{(2)}$, or $\sigma_0 - \sigma_1 - \sigma_2 + \sigma_3 = 0$. Since the maintenance of pairwise LD requires the sign of $w_{ij} + w_{ABC}/2$ and D_{ij} to agree, if we are willing to accept the approximation of (S36), then pairwise LD can be maintained if $\sigma_{\mathrm{PI}}^{(2)} > -\sigma_{\mathrm{PI}}^{(1)}$, otherwise it cannot.

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

$$\sigma_{\mathrm{PI}}^{(2)} > -\sigma_{\mathrm{PI}}^{(1)} \quad \text{and} \quad \sigma_{\mathrm{PI}}^{(2)} > \sigma_{\mathrm{PI}}^{(1)} \tag{S37}$$

are satisfied, it is possible that pairwise LD can be maintained in the absence of triple LD, but also that 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 $D_{ij} = 0$, or $D_{ABC} = 0$ and $|D_{ij}|$ is maximized.

Biologically speaking, from our analysis above and Figure 5, increasing the strength of antigen-specific cross-protection, that is, increasing $\sigma_{\mathrm{PI}}^{(1)}$ and/or $\sigma_{\mathrm{PI}}^{(2)}$, produces strain structure, but depending on which of these is increased, different types of structure can occur. Increasing $\sigma_{\mathrm{PI}}^{(2)}$ relative to $\sigma_{\mathrm{PI}}^{(1)}$ tends to favour triple LD, and so a population in which there are four strains present, each of which share no more than one antigen with any other. If

instead $\sigma_{\mathrm{PI}}^{(1)}$ is increased relative to $\sigma_{\mathrm{PI}}^{(2)}$, this favours pairwise LD, and so a population in which only two strains are present, each of which share no antigens. This is logical: if $\sigma_{\mathrm{PI}}^{(1)}$ is positive and larger than $\sigma_{\mathrm{PI}}^{(2)}$ immune protections are much more effective against a single antigen then expected, and so the minimum amount of strain diversity that preserves antigenic variation is favoured. If $\sigma_{\mathrm{PI}}^{(2)}$ is instead large, then immune protections are much more effective when strains share two antigens, and is relatively ineffective against strains sharing a single antigen. Consequently, greater strain diversity can be maintained in the population.

S7.3 SSCP promotes oscillations and maintains diversity

The presence of strain-specific cross-protection, $\sigma_{\mathrm{HI}} > 0$, has a similar effect for three loci as it does when there are only two. In particular, it increases the likelihood that w_{ABC} and $w_{ij} + w_{ABC}/2$ do not share the same sign with D_{ABC} and D_{ij} respectively, and thus acts to break-up strain structure and promote genotypic diversity. Likewise, by introducing a second epidemiological feedback of hosts with homologous immunity, either $\mathcal{S}_{\Delta}^{(ij)/r}$ (for w_{ij}) or $\mathcal{S}_{\Delta}^{(2)}$ (for w_{ABC}), strain-specific cross-protection can produce evolutionary oscillations within either level of strain-structure, as the over-represented strain set varies (Fig. S4). Simulations suggest that these oscillations do not occur between different types of structure in the deterministic system, even when bistability is possible, but in the presence of demographic stochasticity it seems likely that this can occur, which would further complicate the dynamics.

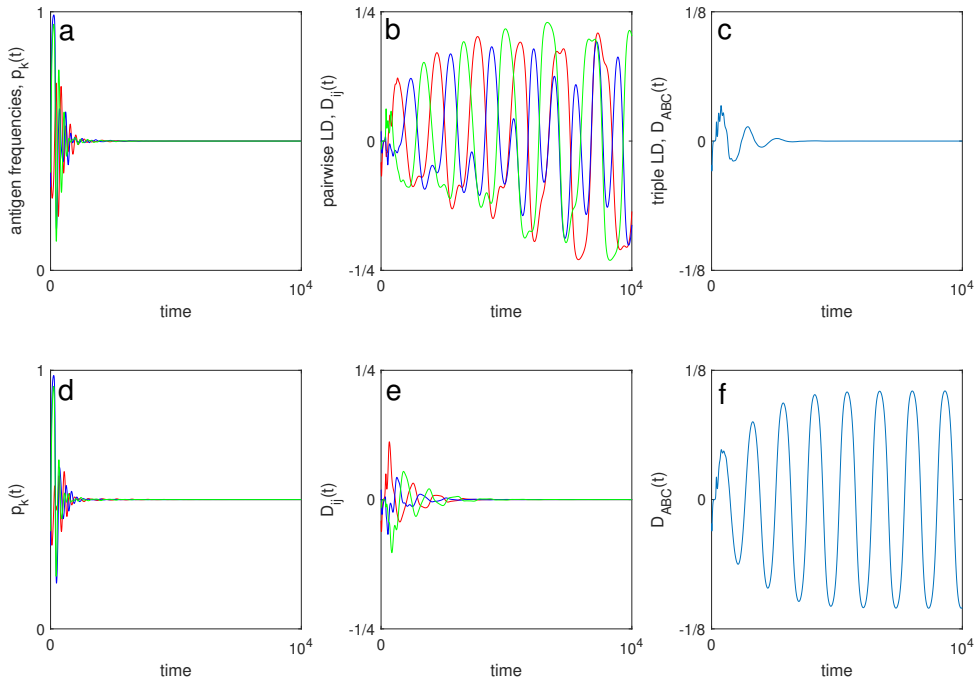

Figure S4: **Evolutionary oscillations for more than two loci.** For more then two loci there is more than one possible type of PSS. Consequently, evolutionary oscillations can occur for any order of LD. All simulations used $b = d = 0.005$, $\beta = 1$, $\mu = 0.5$, $\rho = 10^{-6}$, $\sigma_0 = 1$ and $\sigma_h = 0.05$, while panels **a-c** $\sigma_1 = 0.6$, $\sigma_2 = 0.25$, $\sigma_3 = 0.2$, whereas for panels **d-f**, $\sigma_1 = 0.7$, $\sigma_2 = 0.3$, $\sigma_3 = 0.2$.