Multilocus adaptation to vaccination

David V. McLeod*[1] and Sylvain Gandon†[1]

[1]Centre D'Ecologie Fonctionnelle & Evolutive, CNRS, Univ. Montpellier, EPHE, IRD, Univ. Paul Valéry Montpellier 3.

May 26, 2021

Abstract

Pathogen adaptation to public health interventions, such as vaccination, may take tortuous routes and involve a multitude of genetic mutations acting on distinct phenotypic traits. For example, pathogens can escape the vaccine-induced immune response, or adjust their virulence so as to increase transmission in vaccinated hosts. Despite its importance for public health and vaccine efficacy, how these two adaptations jointly evolve is poorly understood. Taking a trait-centered, rather than variant-centered perspective, here we elucidate the role played by epistasis and recombination, with an emphasis on the different protective effects of vaccination. Vaccines reducing transmission and/or increasing clearance generate positive epistasis between the vaccine-escape and virulence alleles, favouring strains that carry both mutations. Vaccines reducing virulence mortality generate negative epistasis, favouring strains that carry either mutation, but not both. If epistasis is positive, recombination can lead to the sequential fixation of the two mutations and prevent the transient build-up of more virulent escape strains. If epistasis is negative, recombination between loci can create an evolutionary bistability between alternative routes of adaptation, such that whichever adaptation is more accessible tends to be favoured in the long-term. More generally, our model provides a valuable framework for studying pathogen adaptation from a trait-centered view, shifting the focus away from variants.

Keywords: vaccine resistance, antigenic escape, virulence evolution, linkage disequilibrium, evolutionary epidemiology

*david.mcleod@cefe.cnrs.fr
†sylvain.gandon@cefe.cnrs.fr

Introduction

Unlike antimicrobials, vaccines typically provide evolutionarily robust protection [1–3], and have had some remarkable successes. However, although less common than antimicrobial resistance, pathogen evolution in response to vaccination does occur [1, 3–6]. There are two key adaptive routes along which pathogens evolve in response to vaccination.

First, pathogens can undergo rapid antigenic evolution, allowing them to escape the vaccine-induced immune response (vaccine-escape). This not only presents a barrier to developing effective vaccines [7], but it may require that vaccines be regularly updated to match with circulating antigens, a time consuming process [8], with substantial forecasting uncertainty [9, 10]. Rapid antigenic evolution has led to a focus upon developing 'universal' vaccines, which target conserved viral sequences and so are expected to provide more robust protection [11–13].

Second, by altering the selective balance between transmission and virulence, vaccination can select for more virulent pathogens [14–19]. Specifically, if vaccines are imperfect and so do not completely block infection, the pathogen population experiences two selective environments, vaccinated and unvaccinated hosts. Pathogen fitness in vaccinated hosts is typically reduced, and so increased pathogen transmissibility is selected for [14–20]. As increasing transmission often requires an increase in virulence [21–24], more virulent pathogens will be favoured. The evolution of higher virulence as a consequence of vaccination has been documented empirically in Marek's disease [6, 25, 26], experimentally in rodent malaria [27, 28], and has been implicated as a contributing factor to the re-emergence of pertussis [29].

Although both the vaccine-induced evolution of vaccine-escape [e.g., 30–34] and virulence [14–17, 34] are well-studied, these adaptations have typically been considered in isolation from one another (but see [20]). Yet, vaccination could potentially select pathogens that both escape immunity and be more virulent. Even though this would be the the worst possible public health outcome, this scenario has been overlooked in previous analyses. Here we develop a multilocus epidemiological model to study the joint evolution of vaccine-escape and virulence in response to vaccination. Our goal is to understand the role of (deterministic) selection and so we neglect the influence of stochasticity and chance mutations [e.g., 35–38]. We focus upon the protective effects of vaccination acting at different stages of the pathogen's life-cycle, and show that these have important consequences for the sign of epistasis; this is key to understanding the transient and long-term evolutionary dynamics. We then examine the role of recombination in shaping the evolutionary outcome.

Model

Here we provide only a brief summary; more details can be found in Sup. Info. 1.1. We consider a standard SIR model. Transmission of the pathogen occurs via mass-action with rate constant β, while pathogen infection causes virulence-related mortality at a per-capita rate α. We assume a virulence-transmission trade-off [21–24], and so let $\beta[\alpha] = \alpha^b$, where $b \in (0, 1)$ is a constant. Infections are cleared at rate γ, and recovered hosts are fully protected from future infections through naturally-acquired immunity. Hosts enter the population at a fixed rate δ (e.g., through birth or immigration) and are removed due to natural causes (e.g., death or emigration) at a per-capita rate δ.

A fraction p of the hosts entering the population are vaccinated. The vaccine has five protective effects that can alter distinct steps of the life-cycle of the pathogen:

(1) it reduces the risk of infection by a factor $0 \leq \rho_1 \leq 1$ (sterilizing immunity);

(2) it reduces within-host pathogen growth by a factor $0 \leq \rho_2 \leq 1$, affecting both virulence and transmission (due to the virulence-transmission trade-off);

(3) it reduces the infectiousness of vaccinated hosts (without affecting virulence) by a factor $0 \leq \rho_3 \leq 1$;

(4) it reduces virulence (without affecting transmission) by a factor $0 \leq \rho_4 \leq 1$; and,

(5) it reduces the duration of infection by increasing the recovery rate by a factor $0 \leq \rho_5$.

Vaccine protection is lifelong. Let S and $\hat{S}$ denote the density of unvaccinated and vaccinated susceptible hosts, and I and $\hat{I}$ denote the density of unvaccinated and vaccinated infected hosts at time t. The epidemiological dynamics highlighting the different vaccine protections are shown in Fig. 1.

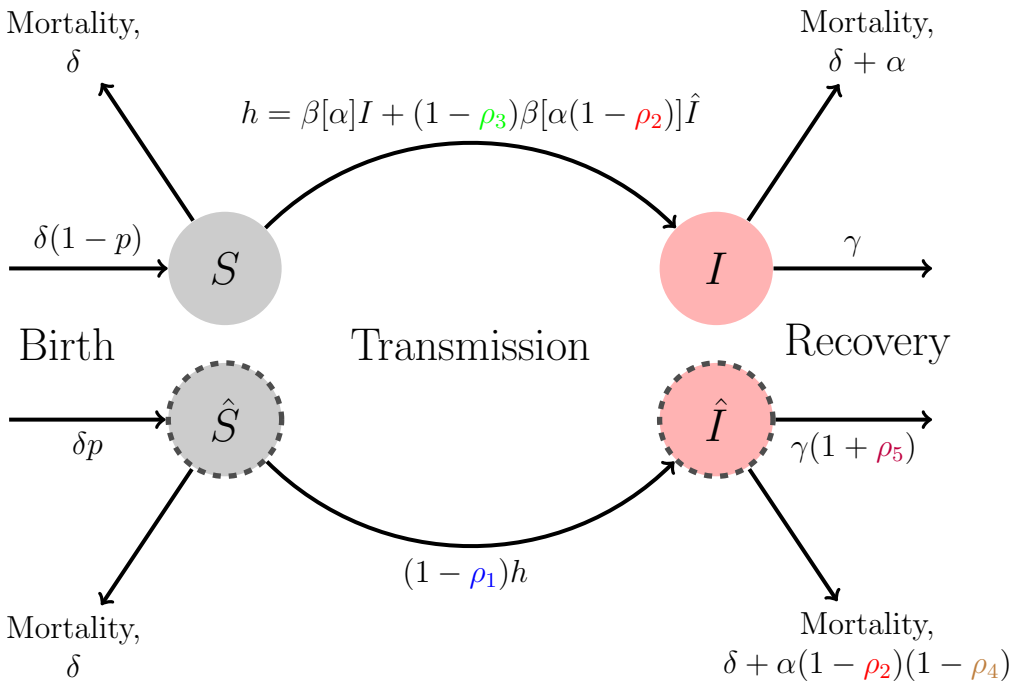

Figure 1: **Schematic of the epidemiological dynamics and the effects of vaccination**. Each of the possible vaccine protections are highlighted. In particular, vaccines may: reduce the risk of infection (ρ_1), reduce within-host pathogen growth (ρ_2), reduce infectiousness (ρ_3), reduce virulence (ρ_4), and/or hasten infection clearance (ρ_5). Note that these are the epidemiological dynamics of the non-escape strain; strains carrying the escape allele are unaffected by the vaccine, but pay a cost of reduced transmission and duration of carriage.

The pathogen has two diallelic loci of interest. The first loci controls the pathogens ability to escape (allele E) or not (allele N) vaccine protection. Escape of vaccine protection is complete. Although cross-protective immunity is known to affect the likelihood of

antigenic escape [e.g., 30, 31, 33], our primary focus here is the action of selection upon the evolutionary dynamics, and so we ignore cross-protection for simplicity. Carrying allele E comes at the cost of reduced transmission, by a factor $0 \leq c_\beta \leq 1$, and an increased rate of infection clearance, by a factor $c_\gamma \geq 0$, that is, the strains carrying allele E have transmissibility $(1 - c_\beta)\beta(\alpha)$ and clearance rate $\gamma(1 + c_\gamma)$. The second loci controls the pathogens virulence. Strains carrying allele A express virulence α_A, which is the optimal virulence in the unvaccinated population, while strains carrying allele V express virulence α_V, which is the optimal virulence in the entirely vaccinated population. Thus vaccines that more effectively block pathogen growth (ρ_2), reduce pathogen virulence (ρ_4), and more rapidly clear infections (ρ_5), select for higher α_V, whereas vaccine coverage, p, and the ability of vaccines to either prevent infection (ρ_1), and/or reduce transmission (ρ_3), will not affect α_V (Sup. Info. 1.1; [14–16]).

There are four possible pathogen strains, $ij \in \{NA, NV, EA, EV\}$, and we will denote the per-capita growth of strain ij as r_{ij}. Note that the specific choices of transmission function and the virulence alleles are used for the sake of argument. More generally, in a diallelic model (as used here) it suffices to assume that the more virulent allele is also more transmissible without specifying a functional relationship between virulence and transmission. In addition to per-capita growth, infections undergo unbiased mutations at the escape and virulence locus at per-capita rates μ_E and μ_V, respectively, and recombination (or genetic reassortment) at a per-capita rate σI, where I is the total density of infections and σ is a rate constant.

Let f_{ij} and f_k denote the frequency of strain ij and allele k, respectively, and D the linkage disequilibrium (LD) between alleles E and V, that is, $D = f_{EV} - f_E f_V$ [39]. Further, denote the additive selection coefficients for vaccine-escape and (increased) virulence as $s_E \equiv r_{EA} - r_{NA}$ and $s_V \equiv r_{NV} - r_{NA}$, respectively, and the epistasis in per-capita growth between alleles E and V as $s_{EV} \equiv r_{EV} - r_{EA} + r_{NA} - r_{NV}$ (see Fig. 2 for an illustration of the selection coefficients and epistasis). Then the evolutionary dynamics can be written

$$\begin{aligned}
\frac{\mathrm{d}f_E}{\mathrm{d}t} &= s_E f_E(1 - f_E) + s_V D + s_{EV}(1 - f_E) f_{EV} + \mu_E(1 - 2f_E), \\
\frac{\mathrm{d}f_V}{\mathrm{d}t} &= s_V f_V(1 - f_V) + s_E D + s_{EV}(1 - f_V) f_{EV} + \mu_V(1 - 2f_V), \\
\frac{\mathrm{d}D}{\mathrm{d}t} &= (s_E + s_V - 2s)\, D + s_{EV} f_{EV} f_{NA} - (2\mu_E + 2\mu_V + \sigma I) D,
\end{aligned} \tag{1}$$

where $s = s_E f_E + s_V f_V + s_{EV} f_{EV}$. System (1) has the standard interpretation [e.g., 40–42]. The dynamics of the frequency of allele E (*mutatis mutandis* allele V) depend upon: (i) the action of direct selection, s_E, proportional to the genetic variance, (ii) the influence of indirect selection, s_V, mediated through LD, (iii) epistasis, s_{EV}, and (iv) unbiased mutations. Thus the two key emergent quantities when considering the multilocus dynamics are the non-random assortment of alleles (LD; [39]) and any gene interactions (epistasis; [43]). Moreover, from (1), these two quantities are linked: LD is generated by epistasis ($s_{EV} \neq 0$), which produces same sign LD [44, 45]. Once LD is present in the population, it can be amplified by directional selection, while it is removed by mutations and recombination [46, 47].

In addition to the evolutionary dynamics described by (1), there is a system of differential equations describing the epidemiological dynamics. These equations can be found in Sup. Info. 1.2; here our focus is upon system (1). However, we note that the selection coefficients and epistasis depend upon the epidemiological processes, and so are dynamic quantities that can temporally vary in both magnitude and sign.

Results

Single locus evolution

First, consider evolution restricted to a single locus, in a population initially monomorphic for strain NA (the 'fittest' strain in an unvaccinated population). This occurs if the mutation rate at one locus is much smaller than the other due to the absence of accessible mutations. For example, smallpox was unable to escape vaccine immunity, leading to its eradication [48], while for other pathogens, evolutionary changes to virulence may not be possible due to various biological constraints.

If evolution is restricted to the escape locus (so $\mu_V = 0$ and $f_V = D = 0$), system (1) reduces to

$$\frac{\mathrm{d}f_E}{\mathrm{d}t} = s_E f_E(1 - f_E) + \mu_E(1 - 2f_E). \tag{2}$$

Thus the single locus dynamics are governed by unbiased mutations, μ_E, and selection on the escape allele, s_E. As mutation rate is expected to be small, the dynamics of f_E will be primarily governed by selection: if $s_E > 0$, the escape allele will be favoured. Hence, increasing the vaccine protection against infection (ρ_1), growth (ρ_2), transmission (ρ_3) and clearance (ρ_5), reduces strain NA fitness and so favours allele E. Likewise, decreasing the costs of escape increases transmissibility (c_β) and duration of carriage (c_γ) of strain EA, favouring allele E (Sup. Info. 1.3). Vaccine protection against virulence (ρ_4) is an exception as it increases the duration of carriage without a concomitant reduction in transmission, and so increases strain NA fitness. Generally speaking, however, if the costs of escape are not too high, and vaccine protection and coverage are not too low, allele E is competitively superior to allele N on the genetic background A ($s_E > 0$).

If instead evolution is restricted to the virulence locus (so $\mu_E = 0$ and $f_E = D = 0$), system (1) reduces to

$$\frac{\mathrm{d}f_V}{\mathrm{d}t} = s_V f_V(1 - f_V) + \mu_V(1 - 2f_V). \tag{3}$$

Thus the more virulent strain will be favoured by selection if $s_V > 0$. Selection for virulence on the genetic background N is a weighted average across two environments: unvaccinated hosts, in which allele A is favoured, and vaccinated hosts, in which allele V is favoured; the weights are the fraction of infections by a given strain that are found in unvaccinated and vaccinated hosts (Sup. Info. 1.4). Whether strain NV outcompetes strain NA depends on the relative importance of the two environments ([14–16, 18]; Sup. Info. 1.4). In particular, increasing coverage (p) and/or decreasing the vaccines ability to block infection (ρ_1) increases the frequency of infections in vaccinated hosts, while decreasing the vaccine's ability to block transmission (ρ_3) increases the value of infections in vaccinated hosts; in either case, this favours allele V. On the other hand, increasing the vaccines ability to block growth (ρ_2), to reduce virulence (ρ_4), and/or to clear infections (ρ_5) has two conflicting effects. First, it reduces the fitness of NA-infections in vaccinated hosts relative to NV-infections, favouring allele V, and second, it increases α_V, thereby reducing the fitness of NV-infections in unvaccinated hosts, favouring allele A. The balance between these two effects, and so whether allele A or V is favoured, depends upon the relative importance of infections in unvaccinated and vaccinated hosts, which is determined by p, ρ_1, and ρ_3 [14–16, 18].

More broadly, the key competitive difference between alleles E and N is that infections carrying allele N optimally exploit either unvaccinated hosts (strain NA) or vaccinated hosts (strain NV), but have reduced fitness in vaccinated hosts (strain NA) or unvaccinated hosts (strain NV), whereas infections carrying allele E pay a fixed cost to exploit both hosts equally. Thus strain Nj is a *specialist* whose fitness depends upon the fraction of infections in unvaccinated hosts, and fraction of vaccinated susceptible hosts, whereas strain Ej is a *generalist* whose fitness depends only upon the availability of hosts. Indeed, because alleles E and N, and on the genetic background N, alleles A and V, differentially exploit the host population, a protected polymorphism at the escape locus, and at the virulence locus (on the genetic background N) is possible.

Epistasis and the production of linkage disequilibrium

Next, suppose that segregation is occurring at both loci, that is, both adaptations are mutationally accessible ($\mu_E > 0$ and $\mu_V > 0$). The key emergent quantity when considering multilocus dynamics is the non-random assortment of alleles (LD; [39]). From (1), LD is produced by epistasis, s_{EV}; moreover, epistasis generates same sign LD [44, 45]. Thus, we are interested in two questions: what generates epistasis, and what predictions can we make about how vaccine protections affect the sign of epistasis?

First, epistasis is generated whenever the fitness of an allele depends upon its genetic background; there are two ways in which this occurs. The first is that vaccination creates two selective environments, vaccinated and unvaccinated hosts [14]. This has implications for the genetic background N: since strain NV is fitter in vaccinated hosts while strain NA is fitter in unvaccinated hosts, a greater proportion of strain NV infections are found in vaccinated hosts than strain NA infections (Sup. Info. 1.2). As a result, on the genetic background N allele V will be disproportionately affected by vaccination as compared to allele A. The second way epistasis is generated is through any multiplicative interactions between parameters associated with the virulence locus (β, α), and those associated with the escape locus (vaccine protection, costs of escape). Because $\alpha_V > \alpha_A$ and $\beta(\alpha_V) > \beta(\alpha_A)$, any such interactions will disproportionately affect infections carrying allele V over A for a given genetic background (E or N).

Second, what is the sign of epistasis produced by the different types of vaccine protection? Critically, manipulating vaccine protection will only directly affect the per-capita growth rate of strains NA and NV. Thus to predict whether vaccine protections produce positive or negative epistasis, it suffices to consider whether any nonlinear interactions between vaccine protections and virulence will disproportionately increase or decrease the fitness of strain NA relative to strain NV. If strain NA experiences an increase (resp. decrease) in fitness relative to strain NV, then positive (resp. negative) epistasis will be produced (Sup. Info. 1.5). With this in mind, the predictions are as follows:

(1) **Vaccines that reduce transmission (ρ_1, ρ_2, ρ_3) produce positive epistasis**. Since allele V is more transmissible than allele A, and as a larger fraction of NV-infections are in vaccinated hosts than NA-infections, on the genetic background N allele V will experience a disproportionately higher reduction in transmission due to vaccination than allele A. Thus strain NV is at a disadvantage relative to strain NA, and so vaccines reducing transmission produce positive epistasis.

(2) **Vaccines that reduce virulence mortality (ρ_2, ρ_4) produce negative epistasis**. Ignoring any concomitant effects to transmission, reducing virulence mortality increases pathogen fitness. Since allele V corresponds to more virulent infections, and as a larger fraction of NV-infections are in vaccinated hosts as compared to NA-infections, allele V will disproportionately benefit from any reductions in virulence. Thus strain NV will be at an advantage relative to strain NA, and so vaccines reducing virulence mortality produce negative epistasis.

(3) **Vaccines that increase clearance (ρ_5) produce positive epistasis**. Because a larger fraction of NV-infections are in vaccinated hosts than NA-infections, on the genetic background N allele V will be disproportionately negatively affected by increased clearance due to vaccination. Consequently, strain NV will be at a disadvantage relative to strain NA and so vaccines increasing clearance produce positive epistasis.

There are a few things to note. First, because costs of escape to transmission, c_β, interact multiplicatively with transmission [42], on the genetic background E allele V will pay disproportionately greater costs, and so c_β produces negative epistasis. However, costs of escape to duration of carriage, c_γ, do not interact with the virulence locus and so do not directly contribute to epistasis. Second, vaccines that block growth (ρ_2) have two opposing contributions: by reducing transmission, ρ_2 produces positive epistasis, and by reducing virulence mortality, ρ_2 produces negative epistasis. Thus, depending upon the epidemiology, the epistatic contribution of ρ_2 can be positive or negative; numerical results, however, suggest that typically the net epistatic contribution of ρ_2 is negative. This leads us to the final point: because s_{EV} depends upon the epidemiological dynamics, the relative importance of the different factors contributing to epistasis can temporally vary; for example, if susceptible hosts are abundant, the contribution of transmission to epistasis is likely to be more important.

Negative epistasis and evolutionary bistability

Next, we ask how the joint evolutionary dynamics influence the competition between adaptations when there is directional selection favouring alleles E and V ($s_E > 0$, $s_V > 0$), but epistasis is negative. As explained previously (Fig. 2**a**), negative epistasis is produced if the vaccine reduces virulence mortality (ρ_2, ρ_4), and/or escape is costly to transmission (c_β).

From the strain NA equilibrium, the initial increase in alleles E and V will be dictated by the relative magnitude of the additive selection coefficients. Since epistasis is negative, the increase in alleles E and V will be primarily due to an increase in strains EA and NV, and so negative LD will build up, leading to clonal interference between strains EA and NV [49]. Eventually alleles E and V will be sufficiently abundant such that $f_E + f_V \approx 1$, that is,

$$\Gamma(f_E) \equiv (f_E, f_V, D) = (f_E, 1 - f_E, -f_E(1 - f_E)); \tag{4}$$

along the curve $\Gamma(f_E)$, any further increase to either allele E or V will come at the expense of the other and so clonal interference is strong. Assuming mutations are sufficiently infrequent, then in the vicinity of the curve $\Gamma(f_E)$, system (1) reduces to

$$\begin{aligned} \frac{\mathrm{d}f_E}{\mathrm{d}t} &\approx (s_E - s_V) f_E(1 - f_E) + \mu_E(1 - 2f_E), \\ \frac{\mathrm{d}D}{\mathrm{d}t} &\approx (s_E - s_V)(1 - 2f_E)D - (2\mu_E + 2\mu_V + \sigma I)D. \end{aligned} \tag{5}$$

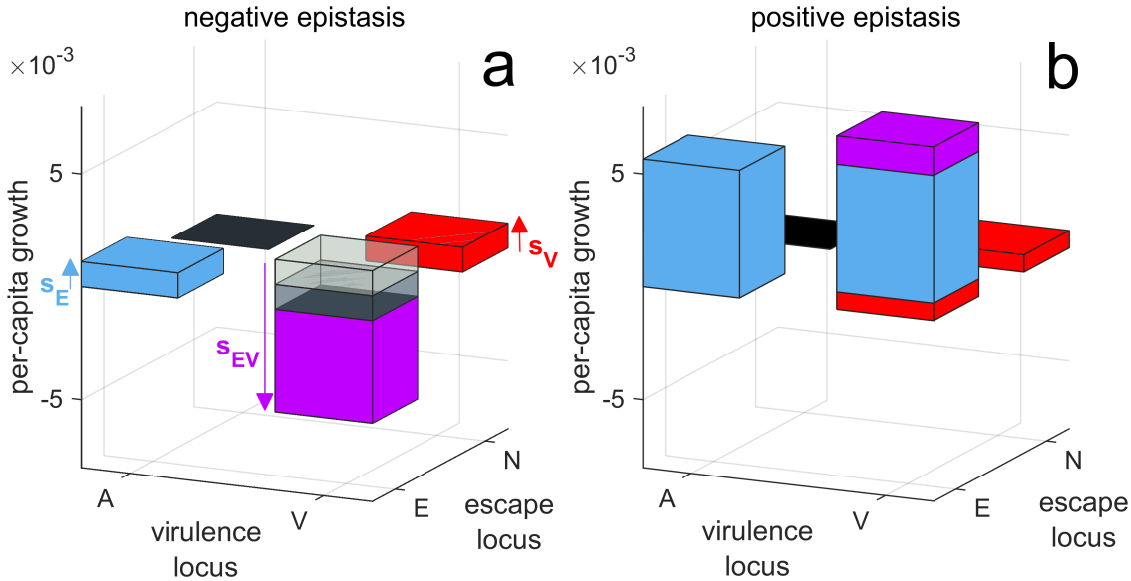

Figure 2: **Fitness landscape at the strain NA equilibrium**. Each bar represents the per-capita growth rate of strain ij at the strain NA equilibrium (so $r_{NA} = 0$); the colours indicate the contribution of allele E (blue), allele V (red), and epistasis (magenta). Panel **a** shows an example of negative epistasis, which occurs when vaccines reduce virulence mortality (ρ_2, ρ_4), while panel **b** shows an example of positive epistasis, which occurs when vaccines reduce transmission (ρ_1, ρ_3) and increase clearance (ρ_5). Parameter values used were $p = 0.6$, $b = 1/2$, $\delta = 0.05$, $\gamma = 0.1$; while for panel **a**, $\rho_1 = \rho_3 = \rho_5 = 0$, $\rho_2 = 0.3$, $\rho_4 = 0.01$, $c_\beta = 0$, $c_\gamma = 0.01$, and for panel **b**, $\rho_2 = \rho_4 = 0$, $\rho_1 = 0.025$, $\rho_3 = 0.02$, $\rho_5 = 1$, $c_\beta = 0.03$, $c_\gamma = 0.45$.

From (5), epistasis no longer directly impacts the dynamics of allele frequencies; instead whichever of alleles E or V is more strongly selected for (larger selection coefficient) will increase in frequency at the expense of the other as the population evolves along $\Gamma(f_E)$ (Fig. 3).

Recombination affects these dynamics: by removing LD from the population, recombination prevents LD from increasing in magnitude [46, 47]. Consequently, the combination of clonal interference and recombination mean that it no longer need be true that the 'fitter' strain ultimately reaches fixation if doing so requires the population to first pass through an intermediate state with a large amount of LD. As such, recombination can create a bistability between the NV- and EA-strain equilibria (Fig. 3**b**); in fact, in the limiting case of quasi-linkage equilibrium [50–52], recombination greatly increases the likelihood that the EA- and NV-equilibria are simultaneously locally stable attractors (Sup. Info. 1.8).

The evolutionary implications of this bistability is that historical contingency, or being 'first', is more important than being 'fitter'. Specifically, if one allele (either E or V) is more mutationally accessible, or is associated with more rapid growth from the strain NA equilibrium (i.e., has a larger selection coefficient), even if this allele is competitively inferior to the other in the long-term (in the absence of recombination), it is likely to reach fixation and exclude the other (Fig. 3**a**,**b**).

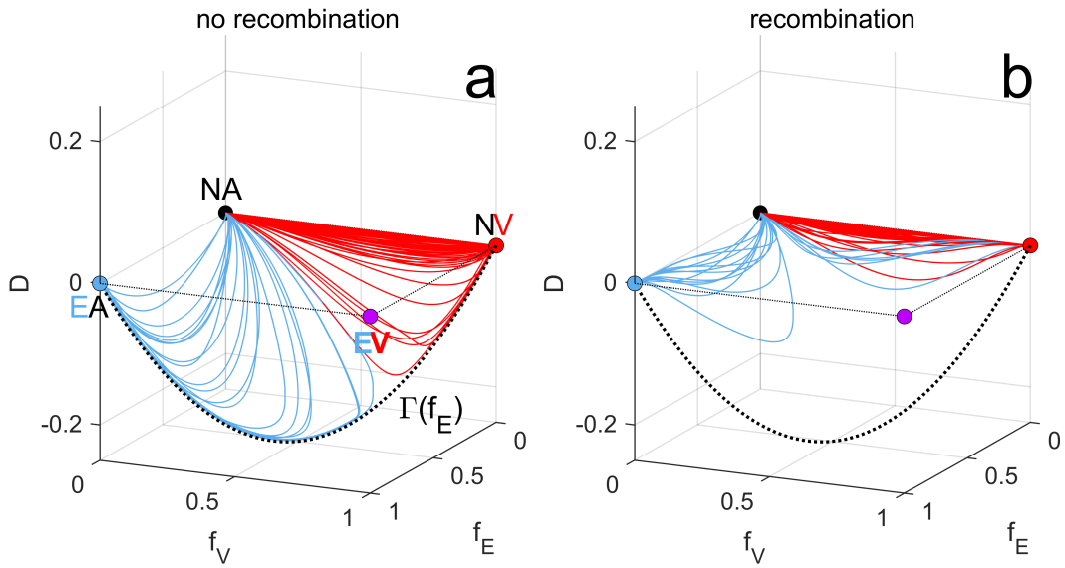

Figure 3: **Negative epistasis and recombination can lead to evolutionary bistability**. If both vaccine-escape and virulence are advantageous ($s_E > 0$ and $s_V > 0$), but epistasis is negative ($s_{EV} < 0$), alleles E and V are in direct competition. Starting from the strain NA equilibrium (solid black circle), strains NV and EA will increase in frequency, producing negative LD until the population is in the vicinity of the clonal interference curve, $\Gamma(f_E)$ (dashed black curve). Along $\Gamma(f_E)$, whichever allele has the larger selection coefficient will go to fixation (panel **a**, $\sigma = 0$); blue curves correspond to fixation of strain EA, red curves to fixation of strain NV. Recombination can create an evolutionary bistability such that 'faster' growing strains tend to reach fixation, even if they are competitively inferior (panel **b**, $\sigma = 0.05$); here the colours indicate which strain would fix in the absence of recombination. Panels **a**,**b** show 100 simulations starting from the strain NA equilibrium using the parameter set: $p = 0.6$, $b = 1/2$, $\delta = 0.05$, $\gamma = 0.1$, $\mu_E = 10^{-4}$, $\mu_V = 10^{-4}$, $\rho_1 = \rho_3 = \rho_5 = 0$, and with ρ_2, ρ_4, c_β, c_γ chosen uniformly at random on the intervals $[0, 0.95]$, $[0, 0.95]$, $[0, 0.05]$, $[0, 0.25]$, respectively, subject to the constraints that at the strain NA equilibrium, $s_E > 0$, $s_V > 0$, and $\alpha_V/\alpha_A > 1.25$.

Positive epistasis and the evolution of virulence

Next, suppose there is directional selection favouring alleles E and V but epistasis is positive. Recall that positive epistasis is produced if the vaccine reduces infection/transmission (ρ_1, ρ_3) and/or increases clearance (ρ_5) (Fig. 2**b**). Here the transient dynamics from the strain NA equilibrium are controlled by the costs of escape to duration of carriage, c_γ. If c_γ is small, escape will be virtually cost-free (since c_β is small), and so allele E is much more strongly selected for than allele V (i.e., $s_E \gg s_V$). Therefore allele E will rapidly increase to quasi-fixation, and so whether strain EA or EV transiently dominates depends upon their initial relative abundances. Specifically, if mutations at each loci are independent (and rare), strain EA will transiently dominate as it is more accessible initially (from the strain NA equilibrium; Fig. 4**a**). If instead strains EA and EV are initially equally abundant, or equivalently, double mutations (e.g., NA mutates to EV) occur with comparable frequency

to single mutations, strain EV will transiently dominate, as it benefits both from positive epistasis and the (relatively weak) directional selection on allele V. In this circumstance, significant positive LD will transiently build up in the population (Sup. Info. Fig. 6). Hence, if c_γ is small, the transient evolutionary dynamics are strongly influenced by the initial conditions, and so chance fluctuations can play an outsized role. If c_γ is large, selection on escape will be of comparable strength to selection on virulence. Therefore both alleles E and V increase in the population at a similar speed, and so the positive LD generated by epistasis will have time to be amplified (Fig. 4**a**). Finally, if c_γ is very large, the escape allele will be too costly and so strain NV will dominate.

In the long-term, and assuming c_γ is not too large, once the escape allele approaches quasi-fixation ($f_E \to 1$), LD will necessarily vanish and so the dynamics of (1) reduce to

$$\frac{\mathrm{d}f_V}{\mathrm{d}t} \approx (s_V + s_{EV})f_V(1 - f_V) + \mu_V(1 - 2f_V), \tag{6}$$

where $s_V + s_{EV} = r_{EV} - r_{EA}$ is the selection coefficient for allele V on the genetic background E. The key factor determining the sign of $s_V + s_{EV}$ are the available hosts: if hosts are abundant, more transmissible strains are favoured [22, 53–55]. As hosts are more abundant whenever total infections are low, and since increasing c_γ will decrease total infections, increasing c_γ will therefore favour a compensatory increase in virulence on the genetic background E and so favours allele V over A (Sup. Info. 1.7). Thus if c_γ is large, strain EV is favoured both transiently and in the long-term. If c_γ is small and strains EA and EV are initially equally abundant (or double-mutations common), the EV strain temporarily dominates before ultimately being outcompeted by strain EA (Fig. 4**c**). Note that this sensitivity to initial conditions (or likelihood of double mutations) seen with positive epistasis does not occur when epistasis is negative; in that case, strain EV is at a fitness disadvantage and so is immediately removed from the population by selection (Sup. Info. Fig. 6).

Recombination can affect these dynamics: by breaking up LD, recombination can prevent the transient selection for the EV strain (c_γ is small, strains EA and EV initially equally abundant). If instead strain EV is favoured both transiently and in the long-term (c_γ is large), frequent recombination will prevent positive LD from building up in the population. Therefore evolution occurs sequentially such that either allele E or V will fix first (depending upon which allele is more strongly selected for), before evolution proceeds at the other loci (Fig. 4**b**).

Discussion

Two primary adaptations have been identified by which pathogens resist vaccines [6]: by evading the vaccine-induced immune response, or by adjusting life history traits, specifically virulence and transmission. Although each of these adaptations have been well-studied in isolation [e.g., 14–17, 30, 32, 33], how the simultaneous evolution of these traits affect transient evolutionary dynamics and/or evolutionary outcomes is poorly understood. Here we considered the joint evolution of vaccine-escape and virulence, with a focus upon the role played by the different possible protections conferred by vaccination. We considered three classes of vaccine protections: (i) those that reduce pathogen transmission either by blocking infection or reducing infectiousness; (ii) those that increase clearance rate (e.g., by inducing a T-cell response [56]); and (iii) those that reduce virulence (e.g., by blocking growth or

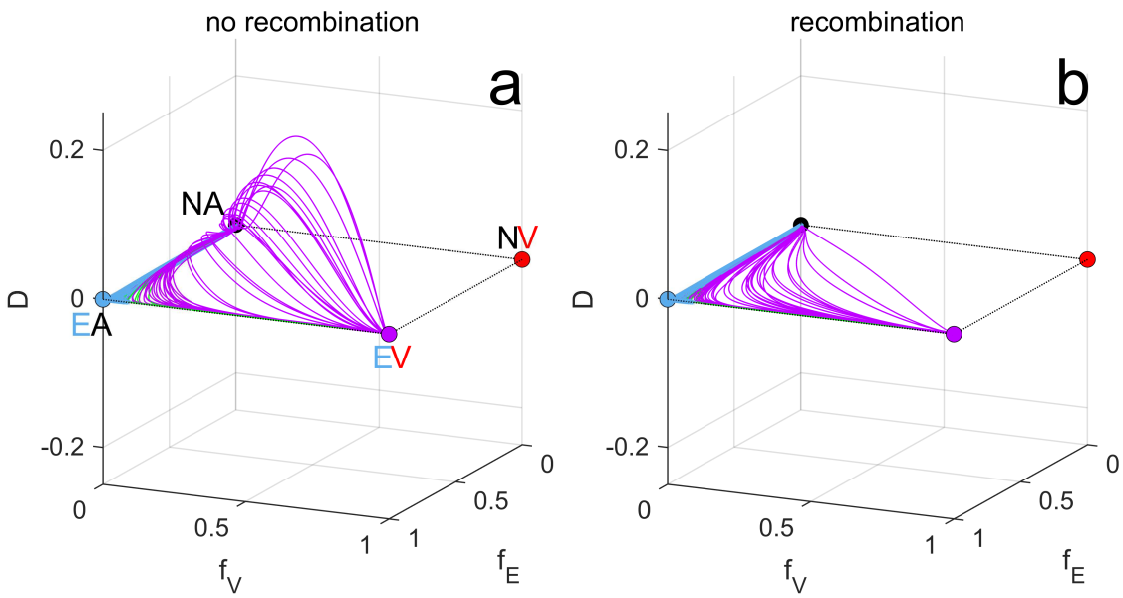

Figure 4: **Positive epistasis and the evolution of virulence**. Positive epistasis favours strains that are both more virulent and evasive in the short- and long-term leading to the build up of LD (panel **a**; $\sigma = 0$). Frequent recombination breaks up LD (panel **b**; $\sigma = 0.05$), favouring the sequential fixation of traits; in this case, allele E reaches quasi-fixation first followed by allele V. Colours indicate which strain fixes: blue is strain EA, magenta is strain EV. Panels **a**,**b** show 100 simulations starting from the strain NA equilibrium using the parameter set: $p = 0.6$, $b = 1/2$, $\delta = 0.05$, $\gamma = 0.1$, $\mu_E = 10^{-4}$, $\mu_V = 10^{-4}$, $\rho_2 = \rho_4 = c_\beta = 0$, and with ρ_1, ρ_3, ρ_5, c_γ chosen uniformly at random on the intervals $[0, 0.9]$, $[0, 0.9]$, $[0, 10]$, $[0, 5]$, respectively, subject to the constraints that at the strain NA equilibrium, $s_E > 0$, $s_V > 0$, and $\alpha_V/\alpha_A > 1.25$.

removing virulence factors [57]). The primary class of vaccine protection determines the evolutionary dynamics. Vaccines reducing virulence mortality generate negative epistasis, favouring one adaptative route or the other, but not both. Vaccines increasing clearance and/or reducing transmission generate positive epistasis, and so strains that carry both vaccine-escape and virulence mutations can be favoured transiently and in the long-term.

From a public health standpoint, the least desirable situation is the evolution of strains that escape vaccine immunity and are more virulent; this occurs when epistasis is positive. Our results indicate that when epistasis is positive, the dynamics are very sensitive to the initial conditions, and significant positive LD (strain EV over-representation) can build-up transiently (Sup. Info. Fig. 6). This suggests stochasticity and chance are far more likely to play important evolutionary roles if epistasis is positive. By breaking up LD, recombination can prevent the transient evolution of more virulent, vaccine-escape strains, while if such strains are favoured in the long-term, recombination leads to the sequential fixation of mutations. Our simulations suggest this typically involves vaccine-escape mutations reaching quasi-fixation first, followed by the evolution of higher virulence. In this circumstance, if vaccines can be 'tweaked' to restore coverage once their efficacy begins to wane due to vaccine-escape (e.g., as for seasonal influenza), this will force the pathogen to

continually evolve vaccine-escape, rather than becoming more virulent. This would suggest that proactive updates to vaccines may be advisable; although this may be impractical with conventional technology (e.g., updating/producing seasonal influenza vaccines takes months [58]), recent technological advances allow vaccines to be rapidly updated and manufactured (e.g., mRNA vaccines [59]). If instead escape strains are in competition with more virulent strains (negative LD), it may be preferable for vaccine-escape strains to outcompete the more virulent strains, particularly if it is possible to 'tweak' the vaccine to restore coverage against the escape strains. Once a more virulent strain has evolved, short of ceasing vaccination, it will be difficult to reverse its evolution.

More generally, in epidemiological models the frequency of recombination depends upon the density of infections through the rate of transmission [41]. This is because the rate of transmission will determine the likelihood of co-infection and so the opportunity for genetic reassortment. Non-pharmaceutical interventions (NPIs), such as social-distancing, travel restrictions, or mask-wearing should therefore reduce the evolutionary contributions of recombination. Previous work suggests that the usage of NPIs in response to the SARS-CoV-2 pandemic may affect the epidemiological dynamics of seasonal influenza and respiratory syncytial virus [60]; our work indicates there may also be evolutionary consequences.

The competition between vaccine-escape and virulent strains may have implications for 'universal' vaccines [11–13, 61]. Universal vaccines target more conserved viral components than 'conventional' vaccines, and so are expected to provide broader cross-protection against a range of pathogen subtypes, thereby preventing and/or slowing antigenic escape [11, 12]. However, the evolutionary consequences are not yet understood [13]. To date, models have primarily focused upon the potential of universal vaccines to reduce antigenic escape [11, 12], and have neglected other potential adaptive responses (e.g., virulence evolution). Reducing the likelihood of vaccine-escape evolution (by reducing μ_E) could induce collateral evolutionary damage by tilting the competitive balance in favour of virulence (assuming μ_V is also not small).

An additional consideration for universal vaccines is the mechanism of protection offered. For example, in influenza there are two broad categories of universal vaccine candidates: those that target haemagglutinin proteins [62, 63] and those that induce a broadly protective T-cell immune response [11, 13, 62]. Vaccines in the former category are expected to block infection [62], while vaccines in the latter are not, and instead are likely to reduce growth, disease severity, and duration of infection [11, 13, 64]. Although both categories of vaccines should reduce the likelihood of escape [11, 12], our analysis suggests that by only weakly blocking infection and by decreasing disease severity, vaccines inducing a broadly protective T-cell immune response will also favour selection for increased virulence. More broadly, models aiming to understand the evolutionary response to universal vaccines should not focus solely upon the consequences for antigenic drift (vaccine escape).

Previous work on the joint evolution of vaccine-escape and virulence in response to vaccination [20] assumed continuous traits and an adaptive dynamics-style model [20]. Adaptive dynamics assumes evolution proceeds according to a sequential-fixation process [65, 66], and so necessarily precludes the possibility of transient dynamics and epistasis, which we have shown can play an important role. Multilocus models, on the other hand, can easily handle these processes and so represent a richer description of the evolutionary dynamics. More generally, the framework we have developed here could equally be applied to other situations involving multiple interacting loci, such as other life history trade-offs (e.g., virulence-

clearance [14, 22, 67] or clearance-transmission [68]), or distinct loci controlling different aspects of immune-escape (e.g., antibody vs T-cell evasion).

From a more applied perspective, the sequencing of SARS-CoV-2 during the ongoing pandemic is unveiling the complex evolutionary dynamics taking place at multiple sites in the virus genome. Trying to understand the dynamics of the multitude of *variants* is overwhelming. Yet, these variants often share the same mutations. Some of the mutations of the spike protein (e.g. D614G) fall in the receptor binding domain (RBD) of the ACE2 receptor and are likely to be involved in cell penetration and ability to transmit effectively [69, 70]. Other mutations of the spike protein fall outside the RBD and are more likely to be involved in the ability to escape from immune recognition (e.g. E484K; [71, 72]). Still other mutations are acting on other proteins with unknown effects on the pathogen life-cycle (e.g., compensatory mutations?). Multilocus models constitute a particularly relevant framework to account for the joint evolution of multiple traits and to escape the variant-centered view that is currently being used.

References

[1] McLean, A. R., 1998 Vaccines and their impact on the control of disease. *British Medical Bulletin* **54**, 545–556.

[2] Kennedy, D. A. & Read, A. F., 2017 Why does drug resistance readily evolve but vaccine resistance does not? *Proc. R. Soc. B* **284**.

[3] Kennedy, D. A. & Read, A. F., 2018 Why the evolution of vaccine resistance is less of a concern than the evolution of drug resistance. *Proc. Natl. Acad. Sci.* **115**, 12878–12886.

[4] Harrison, T. J., Hopes, E. A., J., O. C., Zanetti, A. R. & Zuckerman, A. J., 1991 Independent emergence of a vaccine-induced escape mutant of hepatitis B virus. *J. Hepatology* **13**, S105–S107.

[5] Mooi, F. R., van Oirschot, H., Heuvelman, K., van der Heide, H. G. J., Gaastra, W. & Willems, R. J. L., 1998 Polymorphism in the *Bordetella pertussis* virulence factors P.69/Pertactin and pertussis toxin in the Netherlands: temporal trends and evidence for vaccine-driven evolution. *Infect. Immun.* **66**, 670–675.

[6] Read, A. F. & Mackinnon, M. J., 2008 Pathogen evolution in a vaccinated world. In *Evolution in health and disease*, pp. 139–52. Oxford University Press.

[7] Klausner, R. D., Fauci, A. S., Corey, L., Nabel, G. J., Gayle, H., Berkley, S., Haynes, B. F., Baltimore, D., Collins, C., Douglas, R. G. *et al.*, 2003 Enhanced: The need for a global HIV vaccine enterprise. *Science* **300**, 2036–2039.

[8] Morris, D. H., Gostic, K. M., Pompei, S., Bedford, T., Luksza, M., Neher, R. A., Grenfell, B. T., Lassig, M. & McCauley, J. W., 2018 Predictive modeling of influenza shows the promise of applied evolutionary biology. *Trends Microbiol.* **26**, 102–118.

[9] Luksza, M. & Lassig, M., 2014 A predictive fitness model for influenza. *Nature* **507**, 57–61.

[10] Barrat-Charlaix, P., Huddleston, J., Bedford, T. & Neher, R. A., 2021 Limited predictability of amino acid substitutions in seasonal influenza viruses. *Mol. Biol. Evol.* (doi:10.1093/molbev/msab065).

[11] Arinaminpathy, N., Ratmann, O., Koelle, K., Epstein, S. L., Price, G. E., Viboud, C., Miller, M. A. & Grenfell, B. T., 2012 Impact of cross-protective vaccines on epidemiological and evolutionary dynamics of influenza. *Proc. Natl. Acad. Sci.* **109**, 3173–3177.

[12] Subramanian, R., Graham, A. L., Grenfell, B. T. & Arinaminpathy, N., 2016 Universal or specific? A modeling-based comparison of broad-spectrum influenza vaccines against conventional, strain-matched vaccines. *PLoS Comp. Biol.* **12**, e1005204.

[13] Viboud, C., Gostic, K., Nelson, M. I., Price, G. E., Perofsky, A., Sun, K., S., T. N., Cowling, B. J., Epstein, S. L. & Spiro, D. J., 2020 Beyond clinical trials: evolutionary and epidemiological considerations for development of a universal influenza vaccine. *PLoS Pathogens* **16**, e1008583.

[14] Gandon, S., Mackinnon, M. J., Nee, S. & Read, A. F., 2001 Imperfect vaccines and the evolution of pathogen virulence. *Nature* **414**, 751–756.

[15] Gandon, S., Mackinnon, M., Nee, S. & Read, A. F., 2003 Imperfect vaccination: some epidemiological and evolutionary consequences. *Proc. R. Soc. B.* **270**, 1129–1136.

[16] Andre, J. & Gandon, S., 2006 Vaccination, within-host dynamics, and virulence evolution. *Evolution* **60**, 13–23.

[17] Atkins, K. E., Read, A. F., Savill, N. J., Renz, K. G., Fakhrul Islam, A. F. M., Walkden-Brown, S. W. & Woolhouse, M. E. J., 2012 Vaccination and reduced cohort duration can drive virulence evolution: Marek's disease virus and industrialized agriculture. *Evolution* **67**, 851–860.

[18] Ganusov, V. V. & Antia, R., 2006 Imperfect vaccines and the evolution of pathogens causing acute infections in vertebrates. *Evolution* **60**, 957–969.

[19] Williams, P. D. & Day, T., 2008 Epidemiological and evolutionary consequences of targeted vaccination. *Mol. Ecol.* **17**, 485–499.

[20] Bernhauerová, V., 2016 Vaccine-driven evolution of parasite virulence and immune evasion in age-structured population: the case of pertussis. *Theor. Ecol.* **9**, 431–442.

[21] Anderson, R. M. & May, R. M., 1982 Coevolution of hosts and parasites. *Parasitology* **85**, 411–426.

[22] Frank, S. A., 1996 Models of parasite virulence. *Quarterly Rev. Biol.* **71**, 37–78.

[23] Alizon, S., Hurford, A., Mideo, N. & van Baalen, M., 2009 Virulence evolution and the trade-off hypothesis: history, current state of affairs and the future. *J. Evol. Biol.* **22**, 245–259.

[24] Cressler, C. E., McLeod, D. V., Rozins, C., van den Hoogen, J. & Day, T., 2016 The adaptive evolution of virulence: a review of theoretical predictions and empirical tests. *Parasitology* **143**, 915–930.

[25] Witter, R. L., 1997 Increased virulence of Marek's disease virus field isolates. *Avian Diseases* **41**, 149–163.

[26] Read, A. F., Baigent, S. J., Powers, C., Kgosana, L. B., Blackwell, L., Smith, L. P., Kennedy, D. A., Walkden-Brown, S. W. & Nair, V. K., 2015 Imperfect vaccination can enhance the transmission of highly virulent pathogens. *PLoS Biol.* **13**, e1002198.

[27] Mackinnon, M. J. & Read, A. F., 2004 Immunity promotes virulence evolution in a malaria model. *PLoS Biol.* **2**, 1286–1292.

[28] Barclay, V. C., Sim, D., Chan, B. H. K., Nell, L. A., Rabaa, M. A., Bell, A. S., Anders, R. F. & Read, A. F., 2012 The evoutionary consequences of blood-stage vaccination on the rodent malaria *Plasmodium chabaudi*. *PLoS Biol.* **10**, e1001368.

[29] Mooi, F. R., van Loo, I. H. M., van Gent, M., He, Q., Bart, M. J., Heuvelman, K. J., de Greeff, S. C., Diavatopoulos, D., Teunis, P., Nagelkerke, N. *et al.*, 2009 *Bordetella pertussis* strains with increased toxin production associated with pertussis resurgence. *Emerg. Infect. Diseases* **15**, 1206–1213.

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

[31] 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.

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

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

[34] van Boven, M., Mooi, F. R., Schellekens, J. F. P., de Melker, H. E. & Kretzschmar, M., 2005 Pathogen adaptation under imperfect vaccination: implications for pertussis. *Proc. Roy. Soc. B.* **272**, 1617–1624.

[35] Grenfell, B. T., Pybus, O. G., Gog, J. R., Wood, J. L. N., Daly, J. M., Mumford, J. A. & Holmes, E. C., 2004 Unifying the epidemiological and evolutionary dynamics of pathogens. *Science* **303**, 327–323.

[36] Saad-Roy, C. M., Morris, S. E., Metcalf, C. J. E., Mina, M. J., Baker, R. E., Farrar, J., Holmes, E. C., Pybus, O. G., Graham, A. L., Levin, S. A. *et al.*, 2021 Epidemiological and evolutionary considerations of sars-cov-2 vaccine dosing regimes. *Science* **372**, 363–370. (doi:10.1126/science.abg8663).

[37] Cobey, S., Larremore, D. B., Grad, Y. H. & Lipsitch, M., 2021 Concerns about SARS-CoV-2 evolution should not hold back efforts to expand vaccination. *Nat. Rev. Immunol.* (doi:https://doi.org/10.1038/s41577-021-00544-9).

[38] Kouyos, R. D., Metcalf, C. J. E., Birger, R., Klein, E. Y., Abel zur Wiesch, P., Ankomah, P., Arinaminpathy, N., Bogich, T. L., Bonhoeffer, S., Brower, C. *et al.*, 2014 The path of least resistance: aggressive or moderate treatment? *Proc. Roy. Soc. B* **281**, 20140566. (doi:10.1098/rspb.2014.0566).

[39] Slatkin, M., 2008 Linkage disequilibrium – understanding the evolutionary past and mapping the medical future. *Nat. Rev. Genetics* **9**, 477–485.

[40] Rice, S. H., 2004 *Evolutionary Theory: Mathematical and Conceptual Foundations.* Sunderland, MA, USA: Sinauer Associates.

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

[42] McLeod, D. V. & Gandon, S., 2020 Understanding the evolution of multiple drug resistance in structured populations .

[43] de Visser, J. A. G. M., Cooper, T. F. & Elena, S. F., 2011 The causes of epistasis. *Proc. R. Soc. B* **278**, 3671–3624.

[44] Felsenstein, J., 1965 The effect of linkage on directional selection. *Genetics* **52**, 349–363.

[45] Lewontin, R. C. & Kojima, K., 1960 The evolutionary dynamics of complex polymorphisms. *Evolution* **14**, 458–472.

[46] de Visser, J. A. G. M. & Elena, S. F., 2007 The evolution of sex: empirical insights into the roles of epistasis and drift. *Nat. Rev. Genetics* **8**, 139–149.

[47] Otto, S. P. & Barton, N. H., 1997 The evolution of recombination: removing the limits to natural selection. *Genetics* **147**, 879–906.

[48] Fenner, F. J., 1983 The Florey Lecture, 1983 - Biological control, as exemplified by smallpox eradication and myxomatosis. *Proc. R. Soc. B.* **218**, 259–285.

[49] Gerrish, P. J. & Lenski, R. E., 1998 The fate of competing beneficial mutations in an asexual population. *Genetica* **102**, 127–144.

[50] Neher, R. A. & Shraiman, B. I., 2011 Statistical genetics and evolution of quantitative traits. *Rev. Mod. Phys.* **83**, 1283–1300.

[51] Kimura, M., 1965 Attainment of quasi-linkage equilibrium when gene frequencies are changing by natural selection. *Genetics* **52**, 875–890.

[52] Otto, S. P. & Day, T., 2007 *A Biologist's Guide to Mathematical Modeling in Ecology and Evolution*. Princeton University Press.

[53] Lenski, R. E. & May, R. M., 1994 Evolution of virulence in parasites and pathogens: reconciliation between two competing hypotheses. *J. Theor. Biol.* **169**, 253–265.

[54] Day, T. & Proulx, S. R., 2004 A general theory for the evolution of virulence. *Am. Nat.* **163**, E40–E63.

[55] Bolker, B. M., Nanda, A. & Shah, D., 2010 Transient virulence of emerging pathogens. *J. Roy. Sci. Interface* **7**, 811–822.

[56] Gilbert, S. C., 2011 T-cell-inducing vaccines - what's the future. *Immunology* **135**, 19–26.

[57] Schaffer, A. C. & Lee, J. C., 2008 Vaccination and passive immunisation against *Staphylococcus aureus*. *International Journal of Antimicrobial Agents* **32**, S71–S78.

[58] Pica, N. & Palese, P., 2013 Toward a universal influenza virus vaccine: prospects and challenges. *Annu. Rev. Med.* **64**, 189–202.

[59] Pardi, N., Hogan, M. J., Porter, F. W. & Weissman, D., 2018 mRNA vaccines – a new era in vaccinology. *Nat. Rev. Drug Discovery* **17**, 261–279.

[60] Baker, R. E., Park, S. W., Yang, W., Vecchi, G. A., Metcalf, C. J. E. & Grenfell, B. T., 2020 The impact of COVID-19 nonpharmaceutical interventions on the future dynamics of endemic infections. *Proc. Natl. Acad. Sci.* **117**, 30547–30553.

[61] Saad-Roy, C. M., McDermott, A. B. & Grenfell, B. T., 2019 Dynamic perspectives on the search for a universal influenza vaccine. *J. Infect. Dis.* **219**, S46–S56.

[62] Arinaminpathy, N., Riley, S., Barclay, W. S., Saad-Roy, C. & Grenfell, B., 2020 Population implications of the deployment of novel universal vaccines against epidemic and pandemic influenza. *J. R. Soc. Interface* **17**, 20190879.

[63] Ekiert, D. C., Bhabha, G., Elsliger, M., Friesen, R. H. E., Jongeneelen, M., Throsby, M., Goudsmit, J. & Wilson, I. A., 2009 Antibody recognition of a highly conserved influenza virus epitope. *Science* **324**, 246–251.

[64] Epstein, S. L. & Price, G. E., 2010 Cross-protective immunity to influenza A viruses. *Expert Rev. Vaccines* **9**, 1325–1341.

[65] Geritz, S. A. H., Kisdi, E., Meszena, G. & Metz, J. A. J., 1998 Evolutionarily singular strategies and the adaptive growth and branching of the evolutionary tree. *Evol. Ecol.* **12**, 35–57.

[66] Dieckmann, U. & Law, R., 1996 The dynamical theory of coevolution: a derivation from stochastic ecological processes. *J. Math. Biol.* **34**, 579–612.

[67] Antia, R., Levin, B. R. & May, R. M., 1994 Within-host population dynamics and the evolution and maintenance of microparasite virulence. *Am. Nat.* **144**, 457–472.

[68] Ben-Shachar, R. & Koelle, K., 2018 Transmission-clearance trade-offs indicate that dengue virulence evolution depends on epidemiological context. *Nat. Comm.* **9**, 1–11.

[69] Yurkovetskiy, L., Wang, X., Pascal, K. E., Tomkins-Tinch, C., Nyalile, T. P., Wang, Y., Baum, A., Diehl, W. E., Dauphin, A., Carbone, C. *et al.*, 2020 Structural and functional analysis of the D614G sars-cov-2 spike protein variant. *Cell* **183**, 739–751.e8. ISSN 0092-8674.

[70] Ozono, S., Zhang, Y., Ode, H., Sano, K., Tan, T. S., Imai, K., Miyoshi, K., Kishigami, S., Ueno, T., Iwatani, Y. *et al.*, 2021 SARS-CoV-2 D614G spike mutation increases entry efficiency with enhanced ACE2-binding affinity. *Nat. Comm.* **12**.

[71] Jangra, S., Ye, C., Rathnasinghe, R., Stadlbauer, D., Personalized Virology Initiative study group, Krammer, F., Simon, V., Martinez-Sobrido, L., Garcia-Sastre, A. & Schotsaert, M., 2021 SARS-CoV-2 spike E484K mutation reduces antibody neutralisation. *Lancet Microbe* .

[72] Greaney, A. J., Starr, T. N., Barnes, C. O., Weisblum, Y., Schmidt, F., Caskey, M., Gaebler, C., Cho, A., Agudelo, M., Finkin, S. *et al.*, 2021 Mutational escape from the polyclonal antibody response to SARS-CoV-2 infection is largely shaped by a single class of antibodies. *bioRxiv* (doi:10.1101/2021.03.17.435863).

1 Supplementary Information

1.1 Model

Consider a standard SIR model. Transmission of the pathogen occurs via mass-action with rate constant β, while infection causes virulence-related mortality at a per-capita rate α. We assume a virulence-transmission trade-off [23, 24], and so let $\beta[\alpha] = \alpha^b$, where $b \in (0, 1)$ is a constant. Infections are naturally cleared at rate γ; following clearance, hosts have long-lasting naturally-acquired immunity. Hosts enter the population at rate δ (e.g., birth/immigration), and are removed from the population at a per-capita rate δ (e.g., death/emigration). A fraction p of the hosts entering the population are vaccinated. Vaccination has five possible protective effects:

(1) It reduces the probability of infection by a factor $0 \leq \rho_1 \leq 1$.

(2) It reduces the growth of the pathogen within the host, and so reduces virulence by a factor $0 \leq \rho_2 \leq 1$. Since transmission is a function of virulence, this will also reduce transmission.

(3) It reduces transmissibility (without affecting virulence) by a factor $0 \leq \rho_3 \leq 1$.

(4) It reduces pathogen virulence (without affecting transmission) by a factor $0 \leq \rho_4 \leq 1$.

(5) It hastens the clearance of the pathogen by a factor $0 \leq \rho_5$.

Vaccine protection is life-long.

The pathogen has two diallelic loci of interest. The first locus controls vaccine escape: infections carrying the escape allele, E, can evade vaccine protection, while infections with the non-escape allele, N, cannot. Carriage of the escape allele comes at the cost of reduced transmissibility by a factor $0 \leq c_\beta \leq 1$, that is, $(1 - c_\beta)\beta(\alpha)$, and/or an increased rate of clearance by a factor $c_\gamma \geq 0$, that is, $\gamma(1+c_\gamma)$. The second locus is virulence, and we suppose there are two virulence alleles, A and V, corresponding to virulence α_A and α_V, respectively. We assume that allele A maximises pathogen fitness in an wholly unvaccinated population, that is, α_A maximizes

$$\mathcal{R}_A[\alpha] = \frac{\beta[\alpha]}{\delta + \alpha + \gamma}, \tag{7}$$

while allele V maximises pathogen fitness in an entirely vaccinated population, that is, α_V maximizes

$$\mathcal{R}_V[\alpha] = \frac{(1-\rho_1)(1-\rho_3)\beta[\alpha(1-\rho_2)]}{\delta + \alpha(1-\rho_2)(1-\rho_4) + \gamma(1+\rho_5)}. \tag{8}$$

For the transmission function $\beta[\alpha] = \alpha^b$,

$$\alpha_A = \frac{b}{1-b}(\delta + \gamma) \quad \text{and} \quad \alpha_V = \frac{b}{1-b}\frac{\delta + \gamma(1+\rho_5)}{(1-\rho_2)(1-\rho_4)}. \tag{9}$$

Thus vaccines that are more effective at clearing infection (ρ_5), blocking virulence (ρ_4), and blocking pathogen growth (ρ_2) select for higher virulence, α_V.

We stress that the choice of the function describing the transmission-virulence relationship and the choice of virulence alleles are arbitrary. For example, rather than supposing allele V maximises pathogen fitness in an entirely vaccinated population, we could instead suppose allele V maximises pathogen fitness on the genetic background N. In this case, α_V is the solution of

$$0 = \frac{\mathcal{R}_A[\alpha_V]S}{\beta[\alpha_V]}\left\{\frac{\mathrm{d}\beta}{\mathrm{d}\alpha} - \frac{\beta[\alpha]}{\delta+\alpha+\gamma}\right\}_{\alpha=\alpha_V} + \frac{\mathcal{R}_V[\alpha_V](1-\rho_2)\hat{S}}{\beta[\alpha_V(1-\rho_2)]}\left\{\frac{\mathrm{d}\beta[\alpha]}{\mathrm{d}\alpha} - \frac{\beta[\alpha]}{\delta+\alpha(1-\rho_4)+\gamma(1+\rho_5)}\right\}_{\alpha=\alpha_V(1-\rho_2)}, \quad (10)$$

(see e.g., [15]), where S and $\hat{S}$ are the unvaccinated and vaccinated susceptibles at the equilibrium in which the strain NV is monomorphic. The key difference is that the solution of (10) will depend upon vaccine coverage, p, and the ability of vaccines to block and/or reduce transmission (ρ_1, ρ_3) [see e.g., 14–16]. However, the qualitative predictions for the joint evolutionary dynamics will be similar. More generally, the key to our analysis is that there are two virulence alleles, and that the allele causing more virulent infections is also more transmissible.

There are thus four pathogen strains, $ij \in \{NA, NV, EA, EV\}$, and we will let I_{ij} and $\hat{I}_{ij}$ denote the density of ij-infections in unvaccinated and vaccinated hosts, respectively, at time t. Unbiased mutation occurs at the escape and virulence locus at per-capita rates μ_E and μ_V, respectively. Additionally, recombination (or genetic reassortment) between the two loci may also occur. Specifically, if σ is a rate constant controlling the incidence of recombination, then recombination between strains ij and $k\ell$ occurs at rate $2\sigma I_{ij}I_{k\ell}$. Given recombination has occurred, either strain is equally likely to be the 'recipient' or 'donor' of the allele at a loci chosen with equal probability. Thus the change in density of ij infections in unvaccinated and vaccinated hosts, respectively, due to recombination is

$$\begin{aligned}\sigma_{I_{ij}} &\equiv -\sigma\big(I_{ij}I_{k\ell} - I_{i\ell}I_{kj}\big) + \frac{\sigma}{2}\big(I_{kj}(\hat{I}_{ij}+\hat{I}_{i\ell}) + I_{i\ell}(\hat{I}_{ij}+\hat{I}_{kj}) - I_{ij}(\hat{I}_{kj}+\hat{I}_{i\ell}+2\hat{I}_{k\ell})\big)\\ \sigma_{\hat{I}_{ij}} &\equiv -\sigma\big(\hat{I}_{ij}\hat{I}_{k\ell} - \hat{I}_{i\ell}\hat{I}_{kj}\big) + \frac{\sigma}{2}\big(\hat{I}_{kj}(I_{ij}+I_{i\ell}) + \hat{I}_{i\ell}(I_{ij}+I_{kj}) - \hat{I}_{ij}(I_{kj}+I_{i\ell}+2I_{k\ell})\big)\end{aligned} \quad (11)$$

where $i \neq k$ and $j \neq \ell$.

Let S and $\hat{S}$ denote the density of uninfected unvaccinated and vaccinated hosts, respectively, and let $\mathcal{B}_{ij}$ denote the force of infection of strain ij, that is,

$$\mathcal{B}_{Ej} = (1-c_\beta)\beta[\alpha_j](I_{Ej}+\hat{I}_{Ej}) \quad (12)$$

and

$$\mathcal{B}_{Nj} = \beta[\alpha_j]I_{Nj} + (1-\rho_3)\beta[\alpha_j(1-\rho_2)]\hat{I}_{Nj}. \quad (13)$$

Then the dynamics of the different strains in vaccinated and unvaccinated hosts are

$$\begin{aligned}
\frac{\mathrm{d}I_{Ej}}{\mathrm{d}t} &= \mathcal{B}_{Ej}S - (\delta + \alpha_j + \gamma(1+c_\gamma))I_{Ej} + \mu_E(I_{Nj} - I_{Ej}) + \mu_V(I_{Ek} - I_{Ej}) + \sigma_{I_{Ej}} \\
\frac{\mathrm{d}\hat{I}_{Ej}}{\mathrm{d}t} &= \mathcal{B}_{Ej}\hat{S} - (\delta + \alpha_j + \gamma(1+c_\gamma))\hat{I}_{Ej} + \mu_E(\hat{I}_{Nj} - \hat{I}_{Ej}) + \mu_V(\hat{I}_{Ek} - \hat{I}_{Ej}) + \sigma_{\hat{I}_{Ej}} \\
\frac{\mathrm{d}I_{Nj}}{\mathrm{d}t} &= \mathcal{B}_{Nj}S - (\delta + \alpha_j + \gamma)I_{Nj} + \mu_E(I_{Ej} - I_{Nj}) + \mu_V(I_{Nk} - I_{Nj}) + \sigma_{I_{Nj}} \\
\frac{\mathrm{d}\hat{I}_{Nj}}{\mathrm{d}t} &= (1-\rho_1)\mathcal{B}_{Nj}\hat{S} - (\delta + \alpha_j(1-\rho_2)(1-\rho_4) + \gamma(1+\rho_5))\hat{I}_{Nj} \\
&\qquad + \mu_E(\hat{I}_{Ej} - \hat{I}_{Nj}) + \mu_V(\hat{I}_{Nk} - \hat{I}_{Nj}) + \sigma_{\hat{I}_{Nj}},
\end{aligned} \tag{14}$$

for $j \neq k$. Likewise, the dynamics of S and $\hat{S}$ are

$$\begin{aligned}
\frac{\mathrm{d}S}{\mathrm{d}t} &= \delta(1-p) - \delta S - \sum_{ij} \mathcal{B}_{ij}S \\
\frac{\mathrm{d}\hat{S}}{\mathrm{d}t} &= \delta p - \delta\hat{S} - \sum_j (\mathcal{B}_{Ej} + (1-\rho_1)\mathcal{B}_{Nj})\hat{S}.
\end{aligned} \tag{15}$$

From the preceding assumptions, the per-capita growth rate of Ej-infections, denoted r_{Ej}, is equal to

$$r_{Ej} = (1-c_\beta)\beta[\alpha_j](S + \hat{S}) - (\delta + \alpha_j + \gamma(1+c_\gamma)), \tag{16}$$

whereas if we let $q_{ij} = I_{ij}/(I_{ij} + \hat{I}_{ij})$ denote the fraction of ij-infections found in unvaccinated hosts, the per-capita growth rate of Nj-infections, r_{Nj}, is equal to

$$r_{Nj} = \bar{\beta}_{Nj}(S + (1-\rho_1)\hat{S}) - (\delta + \bar{\alpha}_{Nj} + \bar{\gamma}_{Nj}) \tag{17}$$

where

$$\begin{aligned}
\bar{\beta}_{Nj} &= \beta[\alpha_j]q_{Nj} + (1-\rho_3)\beta[\alpha_j(1-\rho_2)](1-q_{Nj}) \\
\bar{\alpha}_{Nj} &= \alpha_j q_{Nj} + (1-\rho_2)(1-\rho_4)\alpha_j(1-q_{Nj}) \\
\bar{\gamma}_{Nj} &= \gamma q_{Nj} + (1+\rho_5)\gamma(1-q_{Nj})
\end{aligned} \tag{18}$$

denote the average transmission, virulence, and clearance, respectively, of Nj-infections. The formulation of r_{Nj}, specifically the $\bar{\beta}_{Nj}$, $\bar{\alpha}_{Nj}$, and $\bar{\gamma}_{Nj}$ terms, emphasizes that the per-capita growth of Nj-infections is a weighted average across selective environments (vaccinated and unvaccinated hosts), where the weights are the fraction of Nj infections in unvaccinated and vaccinated hosts, q_{Nj} and $1 - q_{Nj}$, respectively. On the other hand, the per-capita growth rate of strain Ej is unaffected by the distribution of Ej infections across vaccinated and unvaccinated hosts.

1.2 Linkage disequilibrium formulation

We are interested in converting systems (14) and (15) into a form that separates the epidemiological and evolutionary processes. To do so, let $I = \sum_{ij}(I_{ij} + \hat{I}_{ij})$ denote the total density of infections in the population, and let $f_{ij} = \sum_{ij}(I_{ij} + \hat{I}_{ij})/I$ denote the frequency of strain ij infections; it follows that $f_E = f_{EA} + f_{EV}$ and $f_V = f_{EV} + f_{NV}$ are the frequency of alleles

E and V, respectively. Moreover, define the linkage disequilibrium (LD), or non-random association of alleles E and V [39], as

$$D \equiv f_{EV} f_{NA} - f_{EA} f_{NV} = f_{EV} - f_E f_V. \tag{19}$$

The additive selection coefficients for alleles E and V are defined as $s_E \equiv r_{EA} - r_{NA}$ and $s_V \equiv r_{NV} - r_{NA}$, respectively, while the epistasis in per-capita growth for alleles E and V, is defined as $s_{EV} \equiv r_{EV} - r_{EA} + r_{NA} - r_{NV}$ (Fig. 2). Using this notation, the evolutionary variables are (f_E, f_V, D), and the evolutionary dynamics are described by

$$\begin{aligned}
\frac{\mathrm{d}f_E}{\mathrm{d}t} &= s_E f_E(1-f_E) + s_V D + s_{EV}(1-f_E)(D + f_E f_V) + \mu_E(1-2f_E) \\
\frac{\mathrm{d}f_V}{\mathrm{d}t} &= s_V f_V(1-f_V) + s_E D + s_{EV}(1-f_V)(D + f_E f_V) + \mu_V(1-2f_V) \\
\frac{\mathrm{d}D}{\mathrm{d}t} &= (s_E + s_V - 2s)D + s_{EV} f_{EV} f_{NA} - (2\mu_E + 2\mu_V + \sigma I)D,
\end{aligned} \tag{20}$$

where $s = s_E f_E + s_V f_V + s_{EV} f_{EV}$.

In addition to the evolutionary dynamics we also have the set of epidemiological variables, $(S, \hat{S}, I, q_{EA}, q_{EV}, q_{NA}, q_{NV})$. The dynamical equations for susceptible hosts, S and $\hat{S}$ are in (15). The other epidemiological variables describe the total density of infections, I, and the fraction of ij-infections in unvaccinated hosts, q_{ij}. The equations describing these dynamics are

$$\begin{aligned}
\frac{\mathrm{d}I}{\mathrm{d}t} &= \sum_{ij} r_{ij} f_{ij} I, \\
\frac{\mathrm{d}q_{Ej}}{\mathrm{d}t} &= (1-c_\beta)\beta[\alpha_j]\big(S(1-q_{Ej}) - \hat{S} q_{Ej}\big) + \mu_{q_{Ej}} + \sigma_{q_{Ej}}
\end{aligned} \tag{21}$$

and

$$\begin{aligned}
\frac{\mathrm{d}q_{Nj}}{\mathrm{d}t} = \bar{\beta}_{Nj}\big(S(1-q_{Nj}) - (1-\rho_1)\hat{S} q_{Nj}\big) \\
- q_{Nj}(1-q_{Nj})(\alpha_j(\rho_2(1-\rho_4) + \rho_4) - \gamma\rho_5) + \mu_{q_{Nj}} + \sigma_{q_{Nj}},
\end{aligned} \tag{22}$$

where

$$\mu_{q_{ij}} \equiv \frac{\mu_V f_{i\ell}(q_{i\ell} - q_{ij}) + \mu_E f_{kj}(q_{kj} - q_{ij})}{f_{ij}} \tag{23}$$

and

$$\sigma_{q_{ij}} \equiv \frac{\sigma}{2} \frac{f_i f_{kj}(q_{kj} - q_{ij}) + f_j f_{i\ell}(q_{i\ell} - q_{ij})}{f_{ij}} I, \tag{24}$$

are the change in q_{ij} infections due to mutation and recombination, respectively. It is important to keep in mind that the additive selection coefficients and epistasis are dynamic quantities that depend upon the epidemiological variables.

In general, following any brief transient dynamics which may be under the influence of the initial conditions, we expect that $q_{NA} \geq q_{NV}$. The logic is that strain NA is fitter in unvaccinated hosts, while strain NV is fitter in vaccinated hosts. Numerical results suggest

the relation $q_{NA}(t) > q_{NV}(t)$ holds (Fig. 5); but to provide a rough mathematical argument for why, observe that Nj infections in unvaccinated hosts are produced at rate $\mathcal{B}_{Nj}S$, and on average will last $(\delta + \alpha_j + \gamma)^{-1}$ time units. Similarly, Nj infections in vaccinated hosts are produced at rate $(1-\rho_1)\mathcal{B}_{Nj}\hat{S}$, and on average will last $(\delta+\alpha_j(1-\rho_2)(1-\rho_4)+\gamma(1+\rho_5))^{-1}$ time units. Thus we may be motivated to approximate $q_{Nj}(t)$ as $q^*_{Nj}(t)$, where

$$q^*_{Nj}(t) \equiv \frac{\frac{\mathcal{B}_{Nj}S(t)}{\delta+\alpha_j+\gamma}}{\frac{\mathcal{B}_{Nj}S(t)}{\delta+\alpha_j+\gamma} + \frac{(1-\rho_1)\mathcal{B}_{Nj}\hat{S}(t)}{\delta+\alpha_j(1-\rho_2)(1-\rho_4)+\gamma(1+\rho_5)}} = \frac{\frac{S(t)}{\delta+\alpha_j+\gamma}}{\frac{S(t)}{\delta+\alpha_j+\gamma} + \frac{(1-\rho_1)\hat{S}(t)}{\delta+\alpha_j(1-\rho_2)(1-\rho_4)+\gamma(1+\rho_5)}}. \tag{25}$$

Over long timescales ($t \to \infty$), if strain Nj is maintained by selection and mutation is small then $q^*_{Nj}(t) \to q_{Nj}(t)$; if strain Nj is instead maintained by mutation, then although selection (and the approximation (25)) roughly apply, q_{Nj} will be closer to q_{Nk} then predicted by (25) (due to mutation). Over short-to-medium timescales, although $q^*_{Nj}(t)$ does not exactly match $q_{Nj}(t)$ (since this would require $q_{Nj}(t)$ to be in quasi-equilibrium), numerical results indicate that the relationship $q_{Nj}(t) > q_{Nk}(t)$ implies $q^*_{Nj}(t) > q^*_{Nk}(t)$. It is straightforward to compute

$$q^*_{NA}(t) - q^*_{NV}(t) = C(t)\frac{(\alpha_A(1-\rho_{24}) + \frac{b}{1-b}\frac{\gamma\rho_5}{\rho_{24}})(\gamma(1+\rho_5-\rho_{24}) + \delta(1-\rho_{24}))}{(\delta+\alpha_A\rho_{24}+\gamma(1+\rho_5))(\delta+\alpha_V\rho_{24}+\gamma(1+\rho_5))} \tag{26}$$

where $C(t) = q^*_{NA}(t)q^*_{NV}(t)\frac{(1-\rho_1)\hat{S}(t)}{S(t)(1-\rho_{24})}$ and $\rho_{24} \equiv (1-\rho_2)(1-\rho_4)$. Since $\rho_{24} \in (0,1]$, equation (26) is positive for all t, i.e., $q^*_{NA}(t) > q^*_{NV}(t)$.

1.3 Evolution of vaccine-escape

If evolution is restricted to the escape locus (so $\mu_V = 0$ and $f_V = D = 0$), system (1) reduces to

$$\frac{\mathrm{d}f_E}{\mathrm{d}t} = s_E f_E(1-f_E) + \mu_E(1-2f_E), \tag{27}$$

where the selection coefficient for vaccine-escape, s_E, is

$$\begin{aligned} s_E \equiv r_{EA} - r_{NA} = (1-c_\beta)\beta[\alpha_A](S+\hat{S}) - \bar{\beta}_{NA}(S+(1-\rho_1)\hat{S}) \\ - (\rho_2(1-\rho_4)+\rho_4)\alpha_A(1-q_{NA}) - \gamma(c_\gamma - \rho_5(1-q_{NA})). \end{aligned} \tag{28}$$

By inspection of s_E, decreasing the costs of escape (c_β, c_γ) increases transmissibility and duration of carriage of strain EA, favouring allele E. Likewise, increasing the vaccine protection against infection (ρ_1), growth (ρ_2), transmission (ρ_3) and clearance (ρ_5), reduces strain NA fitness and so favours allele E. Vaccine protection against virulence (ρ_4) is an exception as it increases the duration of carriage without a concomitant reduction in transmission, and so increases strain NA fitness relative to strain EA.

1.4 Evolution of virulence

If evolution is restricted to the virulence locus (so $\mu_E = 0$ and $f_E = D = 0$), then system (1) reduces to

$$\frac{\mathrm{d}f_V}{\mathrm{d}t} = s_V f_V(1-f_V) + \mu_V(1-2f_V), \tag{29}$$

where the selection coefficient for virulence, s_V, is

$$s_V \equiv r_{NV} - r_{NA} = (\bar{\beta}_{NV} - \bar{\beta}_{NA})(S + (1-\rho_1)\hat{S}) - (\bar{\alpha}_{NV} - \bar{\alpha}_{NA}) - (\bar{\gamma}_{NV} - \bar{\gamma}_{NA}). \quad (30)$$

Since the $\bar{\beta}_{Nj}$, $\bar{\alpha}_{Nj}$, and $\bar{\gamma}_{Nj}$ depend upon the q_{Nj}, selection for virulence on the genetic background N is a weighted average across two environments: unvaccinated hosts, in which allele A is favoured, and vaccinated hosts, in which allele V is favoured [14–16]. The weights are the fraction of Nj-infections in unvaccinated and vaccinated hosts, that is, q_{Nj} and $1 - q_{Nj}$, respectively (see equation (18)). Increasing coverage (p) and/or decreasing the vaccines ability to block infection (ρ_1), increases the frequency of infections in vaccinated hosts, $1 - q_{Nj}$, while decreasing the vaccines ability to block transmission (ρ_3), increases the reproductive value of infections in vaccinated hosts; in either case, this favours allele V. On the other hand, increasing the vaccines ability to block growth (ρ_2), to reduce virulence (ρ_4), and/or to clear infections (ρ_5) has two conflicting consequences as detailed in the main text: on the one hand, it reduces the fitness of NA-infections in vaccinated hosts, favouring allele V, while on the other hand it increases α_V, thereby reducing the fitness of NV-infections in unvaccinated hosts, favouring allele A. Thus whether increasing ρ_2, ρ_4, or ρ_5 favours allele V or A will depend upon the relative importance of infections in unvaccinated and vaccinated hosts, which is determined by p, ρ_1, and ρ_3. Generally speaking, if infections in vaccinated hosts are a more important contributor to fitness, increasing ρ_2, ρ_4, and/or ρ_5 will benefit allele V over allele A.

1.5 Epistasis

Epistasis is generated whenever the fitness of an allele depends upon its genetic background, and for continuous time models is defined as

$$s_{EV} \equiv r_{EV} - r_{EA} + r_{NA} - r_{NV}. \quad (31)$$

We are interested in how the parameters controlling vaccine protection and costs of escape contribute to epistasis, and in particular, whether their epistatic contribution is positive or negative. There are three distinct ways in which these contributions can occur: (i) directly, by the parameter of interest explicitly appearing in (31), (ii) indirectly, by creating or amplifying differences between α_V and α_A (and $\beta[\alpha_V]$ and $\beta[\alpha_A]$), and (iii) by affecting the population structure (q_{Nj}), and/or availability of susceptible hosts (S and $\hat{S}$).

From (31), the direct contributions (and the sign of the direct contributions) of the vaccine protections are as follows:

(1) **Vaccines that reduce transmission (ρ_1, ρ_2, ρ_3) produce positive epistasis**. Using the fact that $\beta[\alpha(1-\rho_2)] = \beta[\alpha]\beta[1-\rho_2]$ for $\beta[\alpha] = \alpha^b$, from (31) the epistatic contribution of vaccine protection against transmission is

$$\big(\beta[\alpha_V](1-q_{NV}) - \beta[\alpha_A](1-q_{NA})\big)\big(1-(1-\rho_3)\beta[1-\rho_2]\big)(S+\hat{S}) + \rho_1(\bar{\beta}_{NV} - \bar{\beta}_{NA})\hat{S}. \quad (32)$$

Since $\beta[\alpha_V] > \beta[\alpha_A]$, whereas we should generally expect that $q_{NA} \geq q_{NV}$ (see Fig. 5) and $\bar{\beta}_{NV} > \bar{\beta}_{NA}$ (i.e., strain NV is more transmissible then strain NA; we should expect

this to hold whenever $s_V > 0$), equation (32) will be positive, and so vaccines that reduce transmission contribute positive epistasis.

The logic here is that transmission reducing vaccines lower the fitness of allele N relative to allele E. On the genetic background N, this disproportionately affects allele V relative to allele A because: (i) allele V is more transmissible (and the parameters ρ_1, ρ_2, ρ_3 multiplicatively interact with the quantity $\beta[\alpha]$), and (ii) a greater proportion of strain NV infections are in vaccinated hosts ($q_{NA} \geq q_{NV}$). By disproportionately reducing the fitness (equivalently, per-capita growth) of strain NV, from (31) this creates positive epistasis.

(2) **Vaccines that reduce virulence mortality (ρ_2, ρ_4) produce negative epistasis**. From (31), the epistatic contribution of vaccine protection against virulence mortality is

$$-(\rho_2(1-\rho_4)+\rho_4)(\alpha_V(1-q_{NV})-\alpha_A(1-q_{NA})). \tag{33}$$

Since $\alpha_V > \alpha_A$ and $q_{NA} > q_{NV}$, equation (33) is negative and so vaccines reducing virulence mortality contributes negative epistasis.

The logic here is that, ignoring any concomitant effects to transmission (by ρ_2), vaccines which reduce virulence mortality increase the fitness of allele N relative to allele E by increasing infection duration. On the genetic background N, this increase in fitness disproportionately affects allele V relative to allele A because: (i) allele V is more virulent (and the parameters ρ_2, ρ_4 multiplicatively interact with α_j), and (ii) a greater proportion of strain NV infections are in vaccinated hosts ($q_{NA} > q_{NV}$). By disproportionately increasing the fitness (per-capita growth) of strain NV, from (31), this creates positive epistasis.

(3) **Vaccines that increase clearance (ρ_5) produce positive epistasis**. The epistatic contribution of vaccines hastening clearance is

$$\gamma\rho_5(q_{NA}-q_{NV}). \tag{34}$$

Since $q_{NA} > q_{NV}$, equation (34) is negative, and so vaccines increasing clearance contribute negative epistasis.

The logic here is that vaccines which increase clearance reduce the fitness of allele N relative to allele E by decreasing infection duration. On the genetic background N, this decrease in fitness disproportionately affects allele V relative to allele A because a greater proportion of strain NV infections are in vaccinated hosts ($q_{NA} \geq q_{NV}$). By disproportionately decreasing the fitness (per-capita growth) of strain NV, from (31), this creates negative epistasis.

In addition to the three categories of epistatic contributions of the vaccine protections, which were also highlighted in the main text, the costs of escape can also produce epistasis. Specifically:

(4) **Costs of escape to transmission (c_β) produces negative epistasis**. The epistatic contribution of costs of escape is

$$-c_\beta(\beta[\alpha_V]-\beta[\alpha_A])(S+\hat{S}). \tag{35}$$

Since $\beta[\alpha_V] > \beta[\alpha_A]$, equation (35) is negative, and so costs of escape to transmission contributes negative epistasis.

The logic here is that costs of escape reduce the fitness of allele E relative to allele N by blocking transmission. On the genetic background E, this decrease in fitness disproportionately affects allele V relative to allele A because allele V is more transmissible and the costs of transmission interact multiplicatively with $\beta[\alpha]$. By disproportionately decreasing the fitness (per-capita growth) of strain EV, from (31) this creates negative epistasis.

The costs of escape to duration of carriage, c_γ, do not directly contribute epistasis because they neither interact directly with the virulence locus (i.e., through multiplicative interactions with α_j and/or $\beta[\alpha]$) nor are they influenced by population structure (i.e., the difference between q_{NA} and q_{NV}). Therefore the sum of equations (32), (33), (34), and (35) is equal to (31), and so the above accounts for all possible direct epistatic contributions.

1.5.1 Influence of vaccine coverage on epistasis

Notice that vaccine coverage, p, does not directly contribute to epistasis. However, vaccine coverage is very important as it controls the influence of population structure. Specifically, when $p \to 0$ or $p \to 1$, population structure vanishes. If the population is fully unvaccinated, $p = 0$, then $\hat{S} = 0$ and $q_{Nj} = 1$, and so equation (31) reduces to

$$s_{EV} = -c_\beta(\beta[\alpha_V] - \beta[\alpha_A])S \tag{36}$$

which is negative provided $c_\beta > 0$. So in the unvaccinated population, epistasis can only be negative.

At the other extreme, when the entire population is vaccinated, $p = 1$, then $S = 0$ and $q_{Nj} = 0$, and so equation (31) reduces to

$$\begin{aligned} s_{EV} = &-(\rho_2(1-\rho_4)+\rho_4)(\alpha_V - \alpha_A) \\ &+ (\beta[\alpha_V] - \beta[\alpha_A])\left(1 - c_\beta - (1-\rho_1)(1-\rho_3)\beta[1-\rho_2]\right)\hat{S}. \end{aligned} \tag{37}$$

Unlike equation (36), equation (37) can be positive or negative, depending upon the protective effects of vaccines relative to the costs of escape. For example, if $\rho_2 = \rho_4 = 0$, then epistasis will be positive whenever $\rho_1(1-\rho_3) + \rho_3 > c_\beta$.

Most of the protective effects of vaccination will have the largest impact upon epistasis when population structured is maximised. Specifically, from inspection of (32), (33), (34), and (35), the direct epistatic contributions of the vaccine protections ρ_2, ρ_3, ρ_4, and ρ_5 will be the most substantial as the difference between q_{NA} and q_{NV} grows. Since for $p = 0$ or $p = 1$, $q_{NA} = q_{NV}$, it follows that the difference between q_{NA} and q_{NV} will be greatest for intermediate vaccination rates (intermediate p). Thus intermediate vaccination rates will maximise the direct epistatic contributions of the vaccine protections ρ_2, ρ_3, ρ_4, and ρ_5. On the other hand, the (positive) epistatic contribution of ρ_1 will generally be the greatest as vaccination rate increases, as although the impact of ρ_1 depends upon the difference between q_{NA} and q_{NV}, its effect is also directly proportional to $\hat{S}$, which increases with p.

1.6 Negative epistasis and evolutionary bistability

Negative epistasis occurs whenever the main vaccine protections are to reduce virulence mortality (ρ_2, ρ_4) and costs of escape reduce transmission (c_β). These results are generally robust if vaccine protections also include an effect on ρ_1, ρ_3, and ρ_5, provided these effects are smaller than the effects on ρ_2, ρ_4, and c_β. For simplicity, however, we focus here upon the case in which we manipulate ρ_2, ρ_4, c_β, and c_γ, with $\rho_1 = \rho_3 = \rho_5 = 0$.

Assume both traits are under directional selection ($s_E > 0$, $s_V > 0$) and epistasis is negative. Then negative LD will build up in the population; specifically, the population will tend to be composed of strains EA and NV. Thus the escape and virulence adaptations are in direct competition with one another. From the strain NA equilibrium, initially each adaptation exerts a limited influence upon one another, allowing both alleles E and V to increase in frequency. However, as $f_E + f_V \to 1$, clonal interference sets in and any further increase in allele E or V comes at the expense of the other. This leads to the population evolving along (or in the vicinity of) the curve

$$\Gamma(f_E) \equiv (f_E, f_V, D) = (f_E, 1 - f_E, -f_E(1 - f_E)). \tag{38}$$

In the vicinity of curve $\Gamma(f_E)$, system (20) reduces to

$$\begin{aligned} \frac{\mathrm{d}f_E}{\mathrm{d}t} &\approx (s_E - s_V)f_E(1 - f_E) + \mu_E(1 - 2f_E), \\ \frac{\mathrm{d}D}{\mathrm{d}t} &\approx (s_E - s_V)(1 - 2f_E)D - (2\mu_E + 2\mu_V + \sigma I)D. \end{aligned} \tag{39}$$

Therefore whichever allele has the larger additive selection coefficient, s_k, will outcompete the other, provided the mutation and recombination rates are sufficiently small (μ_E, μ_V, σ small).

When recombination is frequent, then LD will be continuously removed from the population. Consequently, once the system is in the vicinity of the curve $\Gamma(f_E)$, recombination will 'push' the population towards the equilibria along the curve that the population can reach without LD increasing. This creates an advantage for faster growth, as recombination will tend to 'lock-in' this advantage in the long-term, when in the absence of recombination, long-term selection would favour an alternate strain.

This bistability is distinct from the one documented in single-locus models [14, 15]. In those models, bistability occurred at the virulence locus because pathogens could either evolve to specialize in unvaccinated hosts (low virulence) or specialize in vaccinated hosts (high virulence); as virulence was modelled as a continuous trait [14, 15], the population evolved towards the more accessible of the two fitness 'peaks' from the wild-type virulence. In contrast, here the fitness 'peaks' correspond to the different adaptations (strain EA vs. strain NV), and the fitness 'valley' separating the peaks is generated by negative epistasis between alleles E and V. Recombination breaks up negative LD, preventing the population from crossing the fitness valley, and so creates an evolutionary bistability. The potential for recombination to create an evolutionary bistability has also been observed in models of multidrug resistance [41]. In that case, it prevented the loss of resistance in the absence of treatment, but the principle was similar.

1.7 Positive epistasis and the evolution of virulence

Positive epistasis occurs whenever the costs of escape to transmission (c_β) are small, and the primary vaccine protections are to reduce infection/transmission (ρ_1, ρ_3) and/or to increase clearance (ρ_5). Therefore for simplicity, assume $c_\beta = 0$ and $\rho_2 = \rho_4 = 0$, and so we are free to manipulate ρ_1, ρ_3, ρ_5, and c_γ.

Focus first upon the transient dynamics when both alleles E and V are under directional selection ($s_E > 0$ and $s_V > 0$). The key parameter is the costs of escape to duration of carriage, c_γ. If c_γ is small, then the escape allele will be virtually cost-free (since $c_\beta = 0$). Consequently we should expect that if vaccines offer any protection, selection favouring the escape allele will be much stronger than selection favouring the virulence allele (i.e., $s_E \gg s_V$). In this case, system (20) can be approximated by

$$\begin{aligned}
\frac{\mathrm{d}f_E}{\mathrm{d}t} &\approx s_E f_E(1-f_E) + s_{EV}(1-f_E)f_{EV} + \mu_E(1-2f_E) \\
\frac{\mathrm{d}f_V}{\mathrm{d}t} &\approx s_E D + s_{EV}(1-f_E)f_{EV} + \mu_V(1-2f_V) \\
\frac{\mathrm{d}D}{\mathrm{d}t} &\approx (s_E - 2(s_E f_E + s_{EV} f_{EV}))D + s_{EV} f_{EV} f_{NA} - (2\mu_E + 2\mu_V + \sigma I)D.
\end{aligned} \tag{40}$$

Thus from (40), the escape allele will rapidly increase in frequency in the population due to directional selection (s_E), irrespective of its genetic background (allele A or V), while the rate of increase in allele V (relative to allele E) will depend primarily upon the amount of LD present. As a consequence, whichever of strain EA or EV has the initial numerical advantage will tend to dominant transiently (Fig. 6):

- If the strains EA, EV, and NV are initially absent and are generated by independent mutations at either loci, then we would expect strain EA to initially have the numerical advantage over strain EV from the strain NA equilibrium, as it is accessible by a single mutation. This equally applies if the population is initially in a state at which LD is zero, $D(0) = 0$, in the vicinity of the strain NA equilibrium (and so strains EA, EV, and NV are present, but rare); in this circumstance, strain EA (although rare) must be initially much more abundant than strain EV (to ensure $D(0) = 0$ since $f_{NA} \approx 1$) (Fig. 6**a**).

- If instead strain EA and EV are initially equally abundant (but rare) or double-mutations (e.g., strain NA mutating to strain EV) occur at a comparable rate to single mutations, then because epistasis is positive, while (weak) directional selection favours allele V, strain EV will transiently increase and substantial positive LD will build up in the population (Fig. 6**c**).

Thus if c_γ is small and epistasis is positive, even though (positive) epistasis produces positive LD, since $s_E \gg s_V$, whichever of strains EA or EV has the initial numerical advantage will dominate the transient dynamics.

If c_γ is large, the escape allele will be costly. As a result, selection favouring the escape allele can be of equal strength to selection favouring the virulence allele (i.e. $s_E \approx s_V$). In this case, both allele E and V will increase at a similar rate in the population and so owing to the positive epistasis, substantial positive LD will transiently build up (Fig. 6**c**). Finally,

if c_γ is very large, the escape allele will be too costly and so strain NV will rapidly fix in the population; any positive LD produced will not have build-up in the population.

The costs of escape to duration of carriage, c_γ, also play a pivotal role in the long-term evolutionary dynamics. Specifically, if the escape allele is near quasi-fixation (i.e., $f_E \to 1$), then LD will necessarily disappear (i.e., $D \to 0$), and so the dynamics of system (20) reduces to

$$\frac{\mathrm{d}f_V}{\mathrm{d}t} \approx (s_V + s_{EV})f_V(1 - f_V) + \mu_V(1 - 2f_V). \tag{41}$$

Thus the long-term evolution of virulence on genetic background E is determined by the sign of $s_V + s_{EV} = r_{EV} - r_{EA}$. Since escape is complete, population structure induced by vaccination does not affect the virulence allele when $f_E \approx 1$. As a result, evolution of virulence on the genetic background E is a maximisation problem, that is, if we define

$$\mathcal{R}_E[\alpha] \equiv \frac{(1 - c_\beta)\beta[\alpha]}{\delta + \alpha + \gamma(1 + c_\gamma)}, \tag{42}$$

then if $\mathcal{R}_E[\alpha_j] > \mathcal{R}_E[\alpha_k]$, strain Ej is competitively superior to strain Ek. Solving $\mathcal{R}_E[\alpha_A] = \mathcal{R}_E[\alpha_V]$ for $c_\gamma = c_\gamma^\dagger$ gives

$$c_\gamma^\dagger = \frac{1}{\gamma}\left(\frac{\beta[\alpha_A]\alpha_V - \beta[\alpha_V]\alpha_A}{\beta[\alpha_V] - \beta[\alpha_A]} - \delta - \gamma\right). \tag{43}$$

If $c_\gamma > c_\gamma^\dagger$, higher virulence is selected for on the genetic background E. The logic is straightforward: increasing c_γ increases the clearance rate of Ej-infections, selecting for a compensatory increase in virulence.

In fact, higher virulence than α_V can be favoured on the genetic background E. Denote α_E as the α maximising $\mathcal{R}_E[\alpha]$, that is,

$$\alpha_E = \frac{b}{1 - b}(\delta + \gamma(1 + c_\gamma)). \tag{44}$$

Thus the critical parameter controlling the optimal virulence on the genetic background E is the costs of escape to duration of carriage, c_γ. If we set $\alpha_E = \alpha_V$ and solve for $c_\gamma = c_\gamma^*$, this gives

$$c_\gamma^* = \frac{1}{\gamma}\left(\frac{1 - b}{b}\alpha_V - \delta - \gamma\right). \tag{45}$$

Hence if $c_\gamma > c_\gamma^*$, selection favours higher virulence on the genetic background E (i.e., $\alpha_V < \alpha_E$). In general, the likelihood that $c_\gamma > c_\gamma^*$ decreases with more effective vaccines (higher ρ_2, ρ_4 and ρ_5) as this will increase α_V, reducing the likelihood that $\alpha_E > \alpha_V$. Thus if vaccine escape is sufficiently costly to duration of carriage, a higher level of virulence than that which maximises fitness in unvaccinated hosts can be selected for, that is, $\alpha_E > \alpha_V$.

1.8 Recombination

If recombination rate is large relative to the additive selection coefficients and epistasis, i.e., $\sigma \gg s_E, s_V, s_{EV}$, then the dynamics of LD occurs on a fast timescale relative to change in

allele frequencies, and so LD can be assumed to be in a quasi-steady state, say $D = D_{\text{QLE}}$. This quasi-steady state of LD is referred to as quasi-linkage equilibrium (QLE; [50–52]). If we set $\mathrm{d}D/\mathrm{d}t = 0$ and use a perturbative approximation to solve for $D = D_{\text{QLE}}$, to subleading order we obtain

$$D_{\text{QLE}} \approx \frac{s_{EV} f_E (1 - f_E) f_V (1 - f_V)}{\sigma I} \tag{46}$$

which in turn reduces system (20) to

$$\begin{aligned} \frac{\mathrm{d}f_E}{\mathrm{d}t} &= s_E f_E (1 - f_E) + s_V D_{\text{QLE}} + s_{EV}(f_E f_V + D_{\text{QLE}})(1 - f_E) \\ \frac{\mathrm{d}f_V}{\mathrm{d}t} &= s_V f_V (1 - f_V) + s_E D_{\text{QLE}} + s_{EV}(f_E f_V + D_{\text{QLE}})(1 - f_V) \end{aligned} \tag{47}$$

where we have assumed the per-capita mutation rate is very small and so neglected it. If we compute the Jacobian of system (47) and check the local stability of the EA and NV-equilibria, the strain EA equilibrium will be stable if $s_E > 0$ and $s_V + s_{EV} < 0$, while the strain NV equilibrium will be stable if $s_V > 0$ and $s_E + s_{EV} < 0$. Given directional selection favouring escape ($s_E > 0$) and virulence ($s_V > 0$), it is apparent that a bistability will occur if epistasis is negative and strong (i.e., $-s_{EV} > \max(s_E, s_V)$). The logic here is that recombination prevents double mutants from rising to abundance, and thus in the vicinity of their respective equilibria, the EA and NV infections are only competing against the EV and NA strains and not each other. Since epistasis is negative and directional selection favours alleles V and E, strain EV and strain NA will tend to be a selective disadvantage relative to either strain EA or NV.

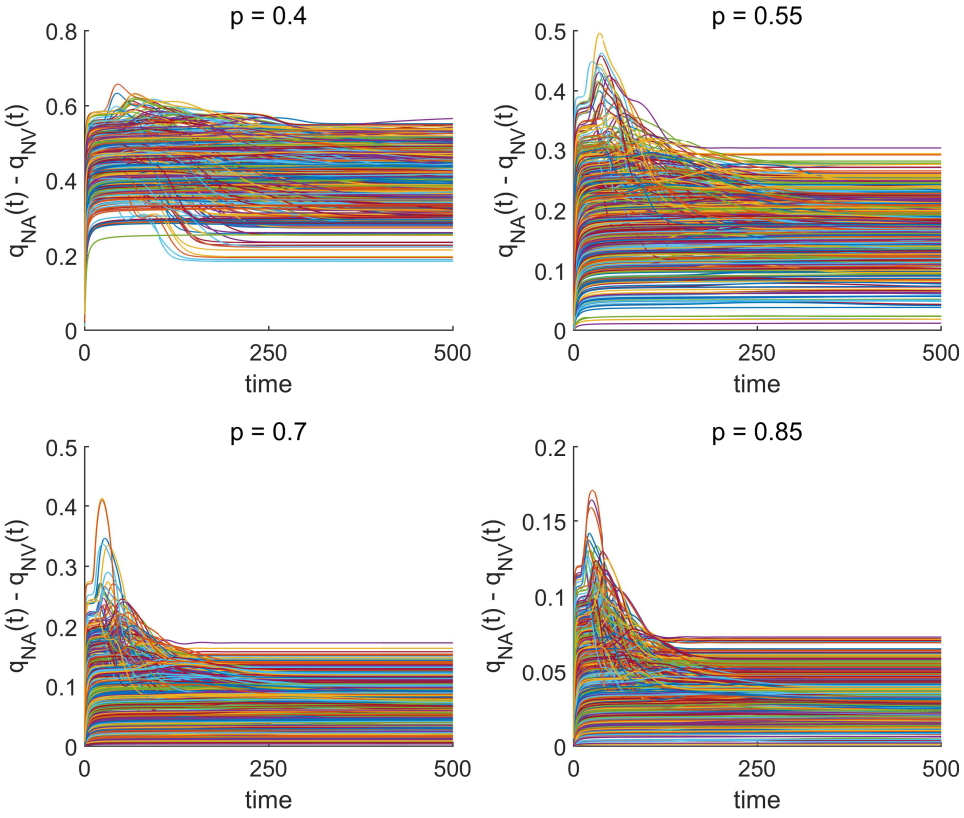

Figure 5: **A greater fraction of NA infections are in unvaccinated hosts than NV infections, $q_{NA}(t) \geq q_{NV}(t)$.** Each panel shows the difference between $q_{NA}(t)$ and $q_{NV}(t)$ for 500 randomly chosen parameter sets with the fraction vaccinated (p) given above each panel. Notice that in all cases, $q_{NA}(t) \geq q_{NV}(t)$. Each simulation starts from the strain NA equilibrium, using parameters $b = 1/2$, $\delta = 0.05$, $\gamma = 0.1$, $\mu_E = 10^{-4}$, $\mu_V = 10^{-4}$, $\sigma = 0$, and the remaining parameters chosen uniformly at random over the given intervals: $\rho_1, \rho_2, \rho_3, \rho_4, c_\beta \in [0, 0.95]$, $\rho_5 \in [0, 8]$, $c_\gamma \in [0, 2]$. All parameter sets satisfied the constraints that $s_E > 0$ and $s_V > 0$ at the strain NA equilibrium and $\alpha_V/\alpha_A > 1.2$.

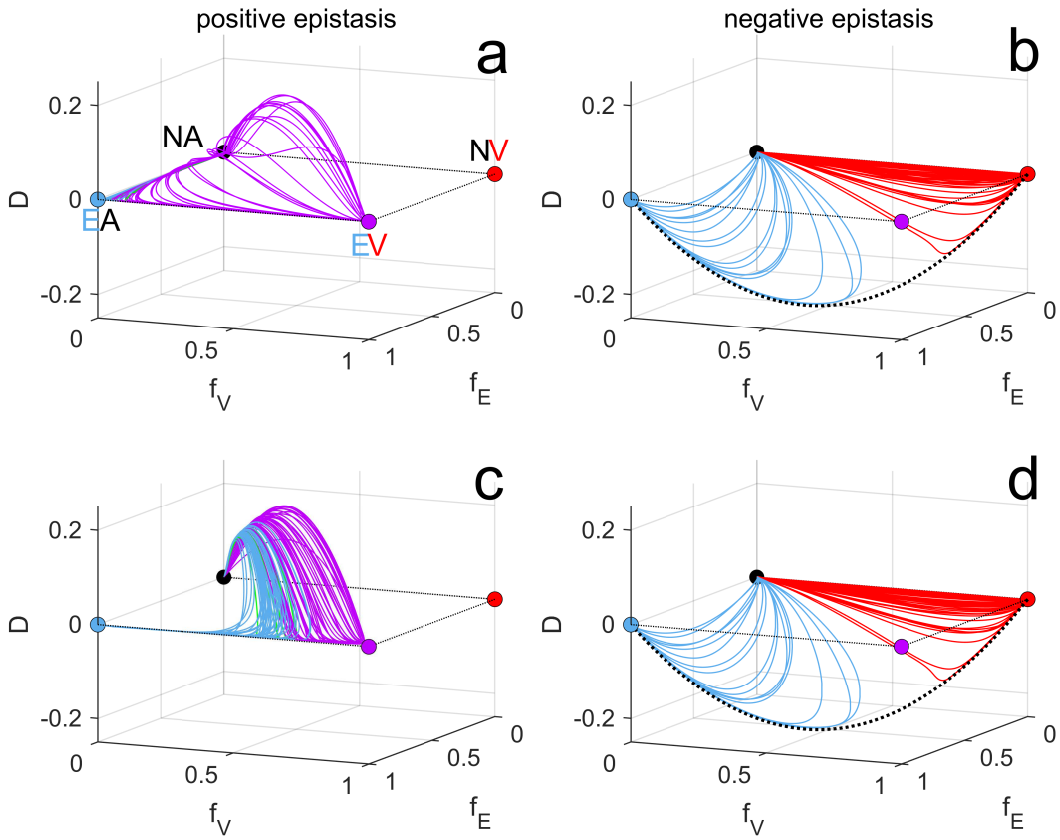

Figure 6: **When epistasis is positive, the evolutionary dynamics are sensitive to initial conditions.** Panels **a**, **c** are when epistasis is positive, while panels **b**, **d** are when epistasis is negative. In panels **a**, **b**, double-mutations are not possible (e.g., strain NA cannot mutate to strain EV and vice-versa), whereas in panels **a**, **d** double mutations occur at the same per-capita rate as single mutations. If epistasis is positive (panel **d**), this leads to a substantial, transient build-up of LD, whereas if epistasis is negative (panel **c**), this has negligible effect. Each panel represents 100 simulations starting from the strain NA equilibrium, using parameters $p = 0.6$, $b = 1/2$, $\delta = 0.05$, $\gamma = 0.1$, $\mu_E = 10^{-4}$, $\mu_V = 10^{-4}$, $\sigma = 0$. For panels **a**, **c**, $\rho_2 = \rho_4 = c_\beta = 0$, while ρ_1, ρ_3, ρ_5, c_γ were chosen uniformly at random from the intervals $[0, 0.95]$, $[0, 0.95]$, $[0, 15]$, $[0, 5]$, respectively. For panels **b**, **d**, $\rho_1 = \rho_3 = \rho_5 = 0$, while ρ_2, ρ_4, c_β, c_γ were chosen uniformly at random from the intervals $[0, 0.95]$, $[0, 0.95]$, $[0, 0.1]$, $[0, 0.2]$, respectively. All parameter sets satisfied the constraints that $s_E > 0$ and $s_V > 0$ at the strain NA equilibrium and $\alpha_V/\alpha_A > 1.25$.