No coexistence for free: Neutral null models for multistrain pathogens

Marc Lipsitch, a, b, *, Caroline Colijn, a, c, Ted Cohen, a, d, William P. Hanage, e, and Christophe Fraser, f

a Department of Epidemiology, Harvard School of Public Health, Boston, MA, USA
b Department of Immunology and Infectious Diseases, Harvard School of Public Health, USA
c Department of Engineering Mathematics, University of Bristol, Bristol, UK
d Division of Global Health Equity, Harvard Medical School, Boston, MA, USA
e Department of Infectious Disease Epidemiology, Imperial College London, London, UK
f Medical Research Council Centre for Outbreak Analysis and Modelling, Department of Infectious Disease Epidemiology, Imperial College London, London, UK

Abstract

In most pathogens, multiple strains are maintained within host populations. Quantifying the mechanisms underlying strain coexistence would aid public health planning and improve understanding of disease dynamics. We argue that mathematical models of strain coexistence, when applied to indistinguishable strains, should meet criteria for both ecological neutrality and population genetic neutrality. We show that closed clonal transmission models which can be written in an “ancestor-tracing” form that meets the former criterion will also satisfy the latter. Neutral models can be a parsimonious starting point for studying mechanisms of strain coexistence; implications for past and future studies are discussed.

Keywords

Mathematical models; Stain coexistence; Neutral models; Population genetics; Ecology; Infectious disease epidemiology

A central question in infectious disease epidemiology is: what creates and maintains patterns of strain diversity in pathogen populations? This question not only holds intrinsic scientific interest, but affects a number of practical decisions. How much, and how fast, will the prevalence of antimicrobial resistance in pathogen populations change in response to changes in antimicrobial use? For vaccines that target a subset of strains in polymorphic pathogen populations (e.g. vaccines against the malaria parasite Plasmodium falciparum, influenza virus, Streptococcus pneumoniae, human papillomavirus), how will vaccination change the patterns of pathogen diversity, and how will these changes affect the incidence of...
disease? Mathematical models are routinely used to try to address such questions, and any such model, implicitly or explicitly, includes assumptions about the mechanisms that created the pathogen diversity upon which a new selective pressure may act. Confidence in the predictions of such models depends on whether the mechanisms that underlie strain diversity are biologically and mechanistically plausible. The impetus for the work described in this paper was to describe conditions that guarantee that mechanisms in a model that promote stable coexistence are explicit (rather than hidden mechanisms that give “coexistence for free”), allowing these mechanisms to be tested against biological data.

Strains may differ by neutral markers, by drug resistance determinants, by their expression of various antigens that trigger immune responses, by factors affecting infectiousness or persistence, and often by a combination of these traits. Patterns of diversity may be static in space and time or may be dynamic, reflecting changing selective pressures and/or the effects of genetic and ecological drift (Gupta and Maiden, 2001; Grenfell et al., 2004; Rambaut et al., 2004; Lipsitch and O’Hagan, 2007). Indeed, the question of strain diversity in epidemiology closely parallels the problems of standing genetic polymorphism in population genetics (Kimura, 1985), and of species diversity in community ecology (Hubbell, 2001). As in these fields, both selective pressure and neutral drift undoubtedly play some role, and a key problem is to distinguish these roles. A central feature of problems concerning diversity is the coexistence of strains. This may be a consequence of “balancing” selective pressures maintaining each strain alongside one another, or it may be a feature of neutral drift.

Mathematical models are important tools both to understand the mechanisms underlying strain coexistence, and to predict the impact of public health interventions that exert selection on pathogen populations. In ecology, the classic model of competition between two species is the Lotka-Volterra model, which posits that the growth of each species is inhibited in a linear fashion by the number of individuals of its own species and of the other species. This model predicts that stable coexistence is possible if each species experiences intraspecific competition more strongly than it experiences competition from the other species. In epidemiology, similar models can track the prevalence of infection with each of two or more strains, and similar qualitative findings may apply, though often the “resource” – susceptible hosts – is explicit in epidemiologic models while being implicit in the classic Lotka-Volterra competition model. More generally, just as simple ecological models with a single resource tend to predict competitive exclusion by the species that can maintain equilibrium density at the lowest level of resource (Tilman et al., 1997), simple epidemiological models with a single host type tend to predict competitive exclusion by the strain with the largest basic reproductive number $R_0$ (Anderson and May, 1991); indeed these two findings are equivalent for the simplest transmission models since the minimum number of susceptible hosts that can support equilibrium of a particular strain is inversely proportional to $R_0$ (Anderson and May, 1991). On the other hand, such a resource may support stable coexistence if the competitors face tradeoffs between different modes of reproduction. Examples include tradeoffs between using an unused resource and displacing a competitor (Levin and Pimentel, 1981), between different modes of transmission (Lipsitch et al., 1996) or between using the primary resource and using metabolites of that resource (Levin, 1972).

Long-term coexistence is a problem only if one assumes that strains compete with one another. Ecological, host species, or other barriers may in some cases prevent or minimize such interaction, making the different strains essentially independent of each other, such that coexistence needs no further explanation. In some virus species, such as rabiesvirus, “coexistence” may be maintained by circulation in different host species or in different, isolated geographic areas, such that the prevalence of one strain has no impact on the prevalence of others (Holmes, 2004). In other systems, however, the evidence for
competition is compelling. In *Streptococcus pneumoniae*, for example, the use of a vaccine against seven serotypes has resulted in significant absolute increases in the prevalence of carriage of serotypes not included in the vaccine, such that total carriage has not changed significantly, but the serotype composition of carriage has been drastically altered (Millar et al., 2006; Huang et al., 2007). These findings are in accord with prior findings from randomized trials (Obaro et al., 1996; Mbelle et al., 1999; Lipsitch, 2001a, b; Dagan et al., 2002) and with animal experiments (Lipsitch et al., 2000a, b; Dawid et al., 2007) that indicate competition between strains in vivo. Another line of evidence for within-host competition in pneumococci (Feikin et al., 2000) as well as other colonizing bacteria (Scanvic-Hameg et al., 2002) is the rapid emergence of large populations of drug-resistant bacteria during antimicrobial treatment, which suggests that the resistant subgroup was present before treatment but held in check by the larger drug-sensitive population. In influenza, the appearance of new variants by antigenic drift or, in pandemics, by antigenic shift, is often associated with the disappearance of prior variants (Grenfell et al., 2004), though this process may take several years in the case of antigenic drift (Rambaut et al., 2008). Strikingly, host mortality (Rohani et al., 2003) or disease-associated behaviors (Rohani et al., 1998) may induce competitive interactions even between unrelated pathogens, such as measles and pertussis, though we do not consider such competitive interactions here.

The work described in this paper was stimulated by our efforts to understand the apparent long-term coexistence of drug-sensitive and drug-resistant strains of *S. pneumoniae*. In attempting to model this phenomenon, we first considered two parsimonious models, (a) a model in which hosts may be colonized with only one strain or the other, but not both, (Lipsitch, 2001a, b), and (b) a model in which dual colonization is possible, similar in structure to that described in (Lipsitch, 1997). Model (a) was unsatisfactory because it predicts competitive exclusion of either resistant or sensitive strains, hence is not capable of explaining stable coexistence of strains. Model (b) suffers from a different problem: it predicts stable coexistence of two strains in some circumstances, and in particular predicts a stable equilibrium frequency distribution of 50%–50% for two indistinguishable strains. This prediction is not plausible, because it implies that if one parameterized such a model with no antibiotic use and no difference between resistant and sensitive strains, and started resistant strains as only 1% of the population, they would rise to comprise 50%. A sensible model to understand the interaction between sensitive and resistant strains would not make the frequency of the resistant strain increase dramatically in the absence of any explicit mechanism promoting its rise. In the Discussion, we mention another problem for which such models may be relevant in *S. pneumoniae*, the problem of serotype coexistence.

If one wants to understand the maintenance of distinct strains within in a particular biological system in which competition is likely and neutral drift alone cannot account for coexistence, then there is a need to examine other potential mechanisms that may account for such coexistence. Biological knowledge in such cases will identify potential candidate mechanisms and inform model structure: for example current infection or colonization with one strain may partially or completely prevent acquisition of a second strain, or a history of exposure to one strain (now cleared) may affect the risk of subsequent infections with the same or other strains through acquired immunity. To evaluate such candidates for their ability to explain coexistence, it is crucial that any features of the model that affect the likelihood of stable coexistence are explicit. Put another way, when the model is stripped of specific mechanisms that may promote stable coexistence, it should not produce stable coexistence “for free,” as a side-effect of some implicit assumption in the model structure. The work in this paper represents an effort to formalize this intuition and identify conditions under which a model is a sensible starting point for understanding strain interactions. The
effort to understand the coexistence of resistant and sensitive pneumococci, which builds on the theory developed here, is deferred to a companion paper.

Because detailed data on strain interactions within an individual are often difficult to obtain for ethical and practical reasons, there may be a number of models consistent with existing data. We can constrain the set of candidate model structures by requiring that if two strains that are functionally indistinguishable (for example, they differ by a selectively neutral single nucleotide polymorphism), the model should treat them as such. Specifically, when applied to indistinguishable strains, a neutral null model for stable coexistence of strains should fulfill two criteria: (1, “ecological neutrality”) the dynamics of the ecological variables – the number of uninfected hosts and the number infected with 0, 1, 2, … strains – should depend on the ecological state variables but, given these, should be independent of the identity of the particular strains involved; and (2, “population genetic neutrality”) there should be no stable equilibrium frequency of the strains in the model; rather, it should be possible to choose initial conditions to guarantee an arbitrary frequency of strains that remains constant for all time \( t \geq 0 \). A neutral null model is one that has no intrinsic mechanism that promotes stable coexistence of indistinguishable strains. Throughout this paper, when we use the term “stable coexistence,” we mean that the model has an attracting steady state with a fixed proportion of various strains; we specifically exclude neutral stability from our definition of stable coexistence.

In this paper, we define a single set of conditions that guarantees a model will meet both criterion (1) and criterion (2). Specifically, any closed clonal transmission model that can be rewritten in a precisely specified sense which meets criterion 1 will itself meet both criterion 1 and criterion 2. We call this way of rewriting the model an “ancestor-tracing ecological model” and describe the meaning of this term in detail. In Section 1 we present one model that meets both criteria show that it meets criterion (1) and argue informally that it meets criterion (2) in order to provide some intuition for the structure of the more formal argument, which is given in Section 2. In Section 3 we discuss how these criteria relate to previously analyzed models. Up to this point, the discussion is framed in terms of models without strain-specific immunity and with state variables that reflect only the number and identities of strains in a given host. In Section 4 we discuss how the considerations described here generalize to other types of models, including those with more complex state variables, with strain-specific immunity, and with mutation or migration. We conclude with a discussion of how such models should be used to understand the mechanisms of coexistence.

A transmission dynamic model for two strains, illustrating how both criteria are met

Consider the model for transmission of two strains shown in Fig. 1A and in the following equations:

\[
\begin{align*}
\frac{dI_I}{dt} &= -\left(\lambda_1 + \lambda_2\right)I_0 + u_1I_1 + u_2I_2 + u_{11}I_{11} + u_{12}I_{12} + u_{22}I_{22} \\
\frac{dI_0}{dt} &= \lambda_1 I_0 - k_1 I_0 I_1 I_2 - k_2 I_0 I_{11} - u_1 I_1 \\
\frac{dI_1}{dt} &= \lambda_2 I_0 - k_1 I_1 I_2 - k_2 I_1 I_{11} - u_2 I_2 \\
\frac{dI_{11}}{dt} &= k_1 I_0 I_1 - u_{11} + 2k_2 I_{11} + c k_1 I_1 I_{12} \\
\frac{dI_{12}}{dt} &= k_1 I_0 I_2 + k_2 I_2 I_1 - u_{12} + 2c k_2 I_{11} + c k_1 I_1 I_{12} - c k_1 I_1 I_{12} - c k_2 I_2 I_{12} \\
\frac{dI_{22}}{dt} &= k_2 I_2 I_2 - u_{22} + 2c k_1 I_1 I_{12} + c k_2 I_2 I_{12} \\
\lambda_1 &= \beta_1 (I_1 + 2q I_{12} + 2q I_{22}) \\
\lambda_2 &= \beta_2 (I_2 + 2q I_{12} + 2q I_{22})
\end{align*}
\]
In this model, there are uninfected individuals \((I_0)\), individuals infected with either strain individually \((I_1\) and \(I_2)\), and individuals infected with both \((I_{12})\). In addition, there are now individuals dually infected with the same strain \((I_{11}\) and \(I_{22})\). The force of infection \(\lambda_1\) is (consistent with traditional usage) the hazard for an uninfected person to become infected with strain 1, and similarly for strain 2. For each strain \(j = 1, 2\), the force of infection is equal to an infectiousness rate constant \(\beta_j\) times a weighted sum of the numbers of infected persons in each category containing strain \(j\). Dually infected individuals are \(q\) times as infectious with each of their strains as singly infected individuals, hence the \(I_{12}\) individuals contribute twice as much as \(I_{12}\) individuals per capita, \(2\beta_1 q I_{11}\) because they have two “copies” of strain 1. Singly infected individuals may become infected by a second strain \(j\) at a rate equal to \(k \lambda_j\), making them become dually colonized. Dually colonized persons may have one strain “knocked out” by another at a rate \(c\) times that for secondary infection of individuals singly colonized. Note that this “knocking out” may or may not be biologically realistic in particular cases, and can be eliminated from the model (setting \(c = 0\)).

The model of Eq. (1) is intended simply as one example of a possible null model for two strains (and one that we will show is neutral in the senses that interest us), and is not put forth as the best model for all (or even any) pathogen. The assumption of a maximum of two strains is made for analytical simplicity, but clearly one could imagine more elaborate models with larger numbers of maximum strains (or perhaps even with no limit on the number of strains) in a host. Such a model is illustrated in Fig 1B, which shows that structurally the model for an arbitrary number of strains and multiplicity of infections can be viewed rather simply.

The construction of Fig 1B also sheds some light on the interpretation dually infected \(I_{11}\) and \(I_{22}\) states. These could correspond to the presence of different patches on a single host that can be colonized as a result of multiple infection. They must at least correspond to some state which is equally reached by re-exposure to any of the strains.

For indistinguishable strains, we set all parameters \((k_i, u_i, \beta_i)\) equal for \(i = 1, 2\) and set the duration for dual carriage equal to that for single carriage, \(u = u_1 = u_2\). (Note: a variant model formulation would be to allow each strain in an infected host to have clearance rate equal to that in a singly infected host, hence the total clearance rate from dual infections is twice that from single infections, but only one strain becomes cleared, e.g. \(I_{12}\) hosts clear to become \(I_1\) or \(I_2\) hosts, each at rate \(u\). We do not consider this further though it also meets our criteria). Then it is possible to write the ecological dynamics of this model (dynamics of the number of individuals infected with 0, 1 or 2 strains: \(N_0 = I_0, N_1 = I_1 + I_2, N_2 = I_{11} + I_{22} + I_{12}\)) in terms only of the state variables \(N_0, N_1, N_2\) without reference to the strains involved. In particular,

\[
\frac{dN_0}{dt} = \beta(N_1 + 2q N_2)N_0 + u(N_1 + N_2) \\
\frac{dN_1}{dt} = \beta(N_1 + 2q N_2)N_0 - k\beta(N_1 + 2q N_2)N_1 - uN_1 \\
\frac{dN_2}{dt} = k\beta(N_1 + 2q N_2)N_1 - uN_2
\]

Hence this model meets our first, “ecological” criterion. Fig. 2 shows how it may meet the second, “population genetic” criterion. In Fig. 2A, we initialize the model with \(I_0(0) = 0.8, I_1(0) = 0.19, I_2(0) = 0.01, I_{11}(0) = I_{12}(0) = I_{22}(0) = 0\). Fig. 2A tracks the value of the six \(I_X\) state variables, and Fig. 2B tracks the proportion of all infections that are strain 1, defined as \(f_1 = \frac{I_1 + 2q I_{11} + q I_{12}}{I_1 + I_2 + 2q (I_{11} + I_{12} + I_{22})} = \frac{I_1 + 2q I_{11} + q I_{12}}{N_1 + 2q N_2}\). This is a natural definition – weighting dually infected persons according to their relative infectiousness \(q\), though as we will see below.
this choice need not be made. Clearly, as the state variables change, the value of $f_1$ remains constant at its starting value, in this case 0.95. A similar result occurs in Fig. 2B, where we start with only dually infected individuals, here starting with strain 1 at a frequency $f_1 = 0.1$: $I_0 (0) = 0.8, I_1 (0) = I_2 (0) = 0, I_{11} (0) = 0.02, I_{12} (0) = 0, I_{22} (0) = 0.18$. Here, again, the frequency of strain 1 stays constant as the system evolves. If, on the other hand, we start with both singly and dually infected hosts ($I_0 (0) = 0.60, I_1 (0) = 0.19, I_2 (0) = 0.01, I_{11} (0) = 0.02, I_{12} (0) = 0, I_{22} (0) = 0.18$) with different frequencies of strain 1, we find that the value of $f_1$ does not remain constant with time, but changes modestly as the singly and dually infected individuals change their relative contributions to the overall strain composition (Fig. 2C).

These numerical results suggest that it is possible to start the model with an arbitrary frequency of strain 1 at time $t = 0$ and maintain that frequency for all $t > 0$ as the model evolves. This strictly constant value of $f_1$ appears to be possible when the starting conditions involve singly or dually infected hosts, but not when both singly and dually infected hosts are present at the start, with different starting frequencies of the strains. However, the modest change in $f_1$ with time when both types of hosts are present at the start suggests that somehow even in that case, the final state of the model depends on the initial conditions, and this is borne out when other combinations of strain frequencies in dual and singly infected hosts are chosen (results not shown).

In summary, we have constructed a model whose ecological dynamics are independent of strain identity, and we have shown informally that starting conditions in this model can be chosen to guarantee an arbitrary frequency of strain 1 that remains constant in the model over time. In the next section we use the intuition gained from this example to motivate a general proof that if the “ancestor-tracing” ecological dynamics of a closed clonal transmission model are independent of strain identity, it is always possible to choose initial conditions that will guarantee any desired frequency of the strains, at any time. This implies that any combination of frequencies of strains can be a long-term equilibrium of the model, and thus that the model, by virtue of its structure, does not encode for any selection on these strains. We will also show the converse, namely that some commonly used models only permit equilibria where all the strains are at equal frequencies, and thus implicitly encode for balancing selection acting on the strains.

**General conditions that assure a transmission model meets the criteria for a neutral null model**

**Informal overview of this section**

This section identifies a set of conditions for a transmission model that guarantee that the two criteria for a neutral null model – (a) ecological dynamics independent of strain identities and (b) no stable equilibrium of frequencies, but rather the ability to choose initial conditions that will maintain any arbitrary strain frequencies throughout the time evolution of the model. Although there is much notation involved, the basic concept of this section is that both conditions are met when one can follow a procedure like that in Section 1, starting with only uninfected and (say) singly infected hosts and initializing strain frequencies among the singly infected hosts at time 0 to the desired frequencies. This in turn is possible when one can trace every strain at any time $t$ back to a unique ancestor of the same type present at time 0, in a fashion that never involves terms identifying particular strain types. If this is possible, then one can start with singly infected hosts only, and descendants of the strains infecting these hosts, by symmetry, will have the same strain compositions as the starting population. Since all strains in the model will be descended from these hosts, all strains in the model will have the desired frequencies of strain types.
The argument involves introducing a scheme for tracking not only the ecological dynamics, but the “ancestor-tracing ecological dynamics,” meaning state variables track not only the multiplicity of the host carrying a given strain, but also the multiplicity of the host of a strain’s ancestor. These are the variables called $M_i$.

The argument is generalized slightly by following not only host multiplicity, but also host “classes,” denoted by $\theta$ – which may refer to gender, age, antibiotic treatment or other fixed or time-varying characteristics. This becomes useful later but merely adds to the notation in the proof. A first-time reader is advised to ignore all the $\theta$ in this section, equivalent to assuming only a single class of hosts.

Definitions and notation

We focus on a dynamical system

$$\frac{dI_{\theta x}}{dt} = g_{\theta x}(I_x, t), x=0, \ldots, X, \theta=1, \ldots, \Theta$$

in which the state variables $I_{\theta x}$ can be understood as referring to the number of hosts in a population of class $\theta$ with colonization status $x$, $x=1, \ldots, X$. The “class” of a host may be understood as a vector of host covariates (possibly including time-varying or constant characteristics such as age, sex, antibiotic treatment state, etc.), while the colonization status indicates which strains, at which multiplicity, currently infect the host. Let every host be capable of being infected by at most $n$ strains simultaneously, and let there be $Y$ strains in circulation. Let the $x$ index all possible ordered lists $\vec{h}^x = \{h_{1}^x, h_{2}^x, \ldots, h_{n}^x\}$, where $\vec{h}^x$ is a nondecreasing list of the strain identities of the strains infecting a host. Some of the entries will be zero in hosts not infected with the full number of strains, and some of the entries may be repeated if hosts are infected more than once with a given strain. The requirement that the list is nondecreasing defines a unique order for the strains infecting a given host.

Define the multiplicity of state $x$, or $m(x)$, as the number of strains infecting a host in state $x$.

For any dynamical system of the $\{I_{\theta x}\}$ we can define the ecological state variables

$$N_{\theta i} = \sum_{m \leq i} I_{\theta x}$$

as the number of hosts of class $\theta$ with multiplicity $i$. Thus the ecological state of the system, defined by $\vec{N}(t) \{N_{\theta i}(t)\}$ tells us how many hosts of each class exist infected by 0, 1, 2, ..., $n$ strains, without reference to which strains they are infected by.

In what follows we will want to trace the “ancestry” of strains infecting various hosts. In particular, for some dynamical systems it makes sense to define every strain infecting a host as descended from a strain that was present at time 0, which we call the ancestor. We call such a model a closed clonal transmission model. It is closed in the sense that no infecting strain can enter the model (though uninfected hosts may enter, and infected or uninfected hosts may die). It is a clonal transmission model in that all strains were either present at the start or result from transmission, and in such a model, transmission implies clonal descent from a unique ancestor.

For example, a classic SIR – type transmission model is a closed pure transmission model, since every infection traces back through a chain of infection events to an initial infective. A model with immigration of infectives, on the other hand, would not be closed because some strains would appear with no ancestor at time 0. A model where descent cannot be traced clonally (eg where strains appear by recombination from two ancestors) would not strictly meet our definition of a clonal transmission model, since no unique ancestor can be identified. It is likely that recombination could be readily incorporated into null models of the sort we are considering here, but we ignore this issue because of the notational difficulties it would create.
We define the *ancestor-tracing ecological state variables* $M_{\theta_i \rightarrow [\theta_k]} (t)$ to be the number of strains infecting individuals of class $\theta$ infected with multiplicity $i$ at time $t$ whose ancestor at time 0 was in an individual of class $\theta'$ infected with multiplicity $k$. In a closed model, note that $M_{\theta_i \rightarrow [\theta_k]} (0) = iN_{\theta_i} (0)$ if $i = k$ and $\theta = \theta'$ and 0 otherwise. Moreover,

$$\sum_{\theta' = 1}^{\Theta} \sum_{k = 1}^{n} M_{\theta_i \rightarrow [\theta_k]} (t) / iN_{\theta_i} (t)$$

We define the *ancestor frequencies* $\mu_{\theta_i \rightarrow [\theta_k]} (t) = M_{\theta_i \rightarrow [\theta_k]} (t) / iN_{\theta_i} (t)$ as the frequency of strains descended from $\theta'$, $k$ ancestors among all those strains currently infecting $\theta$, $i$ hosts. We will use the vector of such frequencies $\mu_{\theta_i} (t) = \{ \mu_{\theta_i \rightarrow [\theta][1,1]} \cdots \mu_{\theta_i \rightarrow [\Theta,Y]} \}$ which represents the frequencies of all possible ancestries among strains currently infecting $\theta$, $i$ hosts; the sum of the components in each vector $\mu_{\theta_i} (t)$ is 1 by definition.

For models broadly of the form of traditional SIR models, every term appearing in an equation in the full model reflects one of the following four types of events:

1. Changes in the demographics of uninfected hosts (births, deaths, etc.)
2. Changes in the covariates of individual hosts, in a way that may depend on their infection status (e.g. aging, acquiring immunity, etc.)
3. Clearance of one or more strains from a host, resulting in a reduction in the host’s multiplicity
4. Infection of a host with a new strain, resulting in an increase in its multiplicity and no change in the presence of other strains infecting that host. In this case, for a pure transmission model, there is a unique host (the infectee)-who becomes infected with a particular strain that is of the same type as a strain in a unique host (the infector).

In addition, a single event may encompass more than one of these changes. For example, in model (1), the terms including $c$ refer to events that combine infection with a new strain and loss of an existing strain. Table 1 shows in general terms, using general functional forms for these events, how such terms appear in both the ecological and ancestor-tracing ecological models corresponding to a particular full model.

As an example of the ancestor-tracing ecological dynamics, we consider the model of Eq. (1) and Fig. 1 (where there is only one class of host, so we ignore $\theta$ and $\theta'$ and drop their subscripts). Here, we have

$$\begin{align*}
M_{11} &= \beta (M_{11} + qM_{22}) N_0 - uM_{11} - k_1 \lambda_{total} M_{11} \\
M_{12} &= \beta (M_{12} + qM_{21}) N_0 - uM_{12} - k_2 \lambda_{total} M_{12} \\
M_{21} &= \beta k (M_{11} + qM_{22}) N_1 + k_1 \lambda_{total} M_{11} - uM_{12} - c\beta k (M_{12} + qM_{22}) M_{21} + c\beta k (M_{11} + qM_{21}) M_{22} \\
M_{22} &= \beta k (M_{12} + qM_{22}) N_1 + k_2 \lambda_{total} M_{12} - uM_{12} - c\beta k (M_{12} + qM_{22}) M_{21} - c\beta k (M_{11} + qM_{21}) M_{22} \\
\lambda_{total} &= \beta (M_{11} + M_{12} + qM_{21} + M_{22}) = \beta (N_1 + 2qN_2)
\end{align*}$$

Note that $M_{0,0} = 0$ since by definition hosts of multiplicity 0 have no strains infecting them, and $M_{0,i} = 0$ since hosts of multiplicity 0 at time 0 are not the ancestors of any infections. Also note that the terms involving $c$, which drop out of the ecological system because they do not change strain multiplicity, reappear in the ancestor-tracing ecological system because they change the identity of strains.
We can write the $M$ in matrix form as $M_{[0,i]}(t') = M_{[0,i]}(t)$, where the brackets in subscripts indicate a single number between 1 and $\Theta n$ chosen to indicate a particular combination of host class and multiplicity. Then $M(t)$ is a square matrix of dimension $\Theta n \times \Theta n$, with each row representing one of the $\Theta n$ possible combinations of strain multiplicity and host class (for strains at time $t$), and each column representing the same (for their ancestors).

Define $R_{jx}$ as the number of strains of type $j$ present in an individual of state $x$, that is, the number of entries equal to $j$ in the list of $\{h_1^x, h_2^x, \ldots, h_n^x\}$.

Define $S_{\theta ij}(t)$ as the number of strain $j$ infections in hosts of class $\theta$ infected with multiplicity $i$ at time $t$. Formally, we define

$$S_{\theta ij}(t) = \sum_{m, \gamma} R_{j\gamma} I_{\theta \gamma}(t).$$

It follows that

$$\sum_{k=1}^{\gamma} S_{\theta \gamma k}(t) = iN_{\theta i}(t).$$

(4)

Define a strain frequency for strain $j$ in the population as the total representation of strain $j$ among hosts with all multiplicities, where hosts of class $\theta$ with multiplicity $i$ are given arbitrary weight $w_{\theta i}$ in both numerator and denominator, such that

$$f_j(t) = \frac{\sum_{\theta=1}^{\Theta} \sum_{i=1}^{n} w_{\theta i} S_{\theta ij}(t)}{\sum_{\theta=1}^{\Theta} \sum_{\gamma=1}^{n} \sum_{i=1}^{n} w_{\theta i} S_{\theta \gamma i}(t)}.$$ 

(5)

Define

$$\phi_{\theta ij}(t) = \frac{S_{\theta ij}(t)}{iN_{\theta i}(t)} = \frac{S_{\theta ij}(t)}{\sum_{k=1}^{\gamma} S_{\theta \gamma k}(t)}$$

(6)

as the frequency of strain $j$ among hosts of multiplicity $i$ and class $\theta$ at time $t$. These frequencies add to 1 for each multiplicity $\theta$: $\sum_{j=1}^{\gamma} \phi_{\theta ij} = 1$.

**Formal statement of conditions**

CLAIM: Consider a dynamical system of $\{I_{\theta x}(t)\}$ which represents a closed clonal transmission model. Suppose that the ancestor-tracing dynamical system associated with this model for the case of indistinguishable strains, using variables $\{M_{[0,i]}(t')\}$, can be written in a form independent of strain identities:

$$\dot{M}_{[\theta i]}(t') = F_{[\theta i]}(t', N_0(t), M(t)) \text{ for all } \theta, \theta' \in [1, \Theta], i, k \in [0, n].$$

(7)

*Epidemics. Author manuscript; available in PMC 2011 May 23.*
Then the two conditions for a neutral null model are met. (a) The ecological dynamical system $\mathbf{N}(t) \{N_{\theta i}(t)\}$ is independent of strain identities; and (b) there is no locally asymptotically stable equilibrium strain frequency in the model, and it is possible to devise starting conditions $I(0)$ to produce a fixed vector of strain frequencies $f(t) = f(0)$ for all time $t > 0$.

**Proof:** Condition (a) follows trivially. Each state variable in the ecological model is proportional to a sum of variables in the ancestor-tracing ecological model,

$$\sum_{\theta' = 1}^{\Theta} \sum_{k=1}^{n} M_{\{\theta|\theta'\}}(t) = iN_{\theta}(t).$$

Thus if the ancestor-tracing ecological system is independent of strain frequencies, so is the ecological system. The converse is not true: it is possible to have a model with a strain-independent ecological system whose ancestor-tracing ecological system is not strain-independent. For example, a model similar to that in Fig. 1 and Eq. (1) but lacking the transitions from dual colonization with both strains to dual colonization with a single strain, the terms containing $c_i\lambda_i^jI_{12}$, $i = 1, 2$, would have the same ecological dynamics as Eq. (1), since these transitions do not change the multiplicity of strains in any host, but would no longer have strain-independent dynamics for their ancestor-tracing ecological dynamics, and indeed would for some parameters yield stable coexistence of indistinguishable strains.

To prove condition (b), the key point is to trace back every strain present in a host at time $t$ to its ancestor, and then use the fact that it has the same type as its ancestor to calculate the total number of strains at time $t$ that are of any given type $j$. From this total number we can calculate strain frequencies.

The tracing back of all strains is possible because we are considering a closed clonal transmission model, for which Eq. (3) holds, assigning every strain a unique ancestor.

The fact that the ancestor-tracing ecological dynamics can be written without reference to strain frequencies implies that per capita, any ancestral strain infecting a $[\theta'k]$ host will leave the same number of descendants as any other ancestral strain in a $[\theta'k]$ host, regardless of which strain type these hosts carry. Combining this with the assumption that the system is clonal (so the descendants are of the same type as the ancestor), it follows that the descendants of ancestral strains infecting $[\theta'k]$ hosts will have the same frequencies of strain types as the ancestors did. Putting these assumptions together to make the key step in the proof, we have:

$$S_{\theta j}(t) = \sum_{\theta' = 1}^{\Theta} \sum_{k=1}^{n} M_{\{\theta|\theta'\}}(t) \phi_{\theta'j}(0).$$

Eq. (8) states that at any time $t$, we can calculate the number of strain $j$ infections in hosts of multiplicity $i$ and type $\theta$ by tracing back the source of all strains in hosts of multiplicity $i$ and type $\theta$ to their origins in hosts of multiplicity $k$ and type $\theta'$ at the start of the model, through ancestor-tracing ecological calculations (with the $M_s$) and then weighting the contribution from such $[k\theta']$ hosts by the representation of strain $j$ in those original hosts.

Now, for any $t$ and any desired choice of strain frequencies $\tilde{f}_j(t)$, we can choose initial conditions for the full model that will produce $f_j(t)$. In particular, we can infect only hosts of a single host class (call it class $\theta = 1$) and choose initial conditions such that only hosts of multiplicity 0 (uninfected) and 1 (singly infected) are present at time 0, and such that a proportion of infected hosts $f_j(0) = \tilde{f}_j(t)$ are infected with strain $j$. Then because there were no
infections of multiplicity $>1$ present at the start, we have $S_{0ij}(0) = 0$ if $(i > 1 \text{ or } \theta > 1)$, and $M_{0i-11}(t) = iN_{0i}(t)$, and $M_{0i-\theta'k}(t) = 0$ if $(k > 1 \text{ or } \theta > 1)$. Then at time $t$ we will have

$$f_j(t) = \frac{\sum_{\theta=1}^{\Theta} \sum_{i=1}^{n} w_{\theta} S_{\theta ij}(t)}{\sum_{\theta=1}^{\Theta} \sum_{i=1}^{n} \sum_{k=1}^{n} \sum_{l=1}^{n} w_{\theta} M_{\theta'k-\theta'lj}(t) \phi_{\theta'lj}(0)} = \frac{\sum_{\theta=1}^{\Theta} \sum_{i=1}^{n} \sum_{k=1}^{n} \sum_{l=1}^{n} w_{\theta} M_{i(k-1)-(l-1)}(t) \phi_{i(l-1)}(0)}{\sum_{\theta=1}^{\Theta} \sum_{i=1}^{n} \sum_{k=1}^{n} \sum_{l=1}^{n} w_{\theta} M_{i(k-1)-(l-1)}(t) \phi_{i(l-1)}(0)} = \phi_{11j}(0) \sum_{\theta=1}^{\Theta} \sum_{i=1}^{n} w_{\theta} M_{\theta i-11j}(t) = \phi_{11j}(0) \sum_{\theta=1}^{\Theta} \sum_{i=1}^{n} w_{\theta} M_{\theta i-11j}(t) = f_j(t)$$

(10)

Here, the first equality is Eq. (5), the second follows from Eq. (8), the third from the starting conditions, and the fourth by rearrangement. Thus we have constructed a way to obtain any desired strain frequency for all $t > 0$ by choosing the correct initial conditions. This implies that there is no stable equilibrium for the strain frequencies, QED.

**Practicalities: how can one tell if a model meets the conditions for a neutral null model?**

In principle, it may be tedious to write out the full ancestor-tracing ecological system {
\[ M_{[\theta]}k-\theta'lj(t) \]} associated with a transmission model larger than the simple system used for illustration in Section 2.1. It is easier, however, to calculate the ecological system without ancestor-tracing, \{N_{[\theta]}\}, simply by adding up the differential equations for all terms with a given multiplicity according to the definition $N_{[\theta]} = \sum m(x) N_{[\theta]x}$, and substituting appropriately. If all strain-specific terms do not drop out of this formulation, then clearly the ecological criterion is violated, which implies that the conditions of Section 2.2 are violated, and the model is not a neutral null model. This strategy is followed in the next section, in Eqs. (12) and (13).

If the ecological system can be written without terms that identify particular strains, we still need to know whether the ancestor-tracing ecological system can be written in such a way. As noted in Section 2.2, a strain-independent ecological system may be associated with an ancestor-tracing system that cannot eliminate references to particular strains. Hence, if the ecological criterion is satisfied, it becomes necessary to write down and inspect the ancestor-tracing ecological system to confirm that the population genetic criterion is satisfied. Since this involves bookkeeping and arithmetic, rather than the sometimes complicated algebra of stability analysis, we suspect that this is easier in most real cases than performing the formal stability analysis directly.
Starting from mixed initial states

For proving our result, it is sufficient to show that any equilibrium can be reached. Eq. (10) shows that this can be done by starting from a ‘pure’ initial conditions with all infections in the singly infected \((i=1)\) first state \((\theta=1)\) in which case the strain frequencies stay constant for all time \(t>0\). This can trivially be generalized to any other ‘pure’ initial states \((\theta i)\), but as we saw in simulations in Section 1, starting from a mixed initial state will in general result in a transient initial change in strain frequencies before an equilibrium is reached.

In general, it can be shown by contradiction that the strain frequencies are equal for all the ecological states in the final equilibrium state, i.e.

\[
\lim_{t\to\infty} \phi_{\theta i}(t) = f_{i}^{\infty},
\]

as long as all the host types are “connected” by transmission, perhaps indirectly.

In this case the matrix \(M\) becomes degenerate at the equilibrium, and the final frequencies are given by

\[
lim_{t\to\infty} \sum_{\theta=1}^{\Theta} \sum_{i=1}^{\mu} M_{[\theta \mid i]}(\theta \phi_{\theta i}(0)) = \frac{\lim_{t\to\infty} \hat{N}_{\theta}(t)}{\lim_{t\to\infty} \hat{N}_{\theta}(t)}
\]

for any choice of \(i\) and \(\theta\). The matrix

\[
K = \frac{\lim_{t\to\infty} M_{[\theta \mid i]}(t)}{\lim_{t\to\infty} \hat{N}_{\theta}(t)}
\]

specifies the transition from the initial strain frequencies to the equilibrium. To understand how the model can simultaneously have any combination of equilibrium strain frequencies and still have dependence on initial conditions, one can draw an analogy with the Hardy-Weinberg model. Starting from a diploid state, a full generation of replication is required before the Hardy-Weinberg equilibrium is obtained. In this case, an equivalence can be drawn where the multiplicity of infection \(i\) is a generalization of polyploidy for infection states, and \(K\) is the next generation operator for strain frequencies.

Relationship to prior models

For pathogens without acquired immunity, two simple sorts of models have been used in the past to model strain competition and coexistence. The first simply assumes that a host may be infected at a given time with one strain or another, but not both, and only uninfected hosts may become infected. This model, shown in Fig. 3A, may be written as follows:

\[
\begin{align*}
\frac{dI_{0}}{dt} &= -(\lambda_{1} + \lambda_{2})I_{0} + u_{1}I_{1} + u_{2}I_{2} \\
\frac{dI_{1}}{dt} &= \lambda_{1}I_{0} - u_{1}I_{1} \\
\frac{dI_{2}}{dt} &= \lambda_{2}I_{0} - u_{2}I_{2} \\
\lambda_{1} &= \beta_{1}I_{1} \\
\lambda_{2} &= \beta_{2}I_{2}
\end{align*}
\]
and is a special case of model 1, with $k = 0$. As such, it by definition meets our criteria, and if the strains are indistinguishable ($\beta_1 = \beta_2$, $u_1 = u_2$), then the relative frequency of each strain will remain at its initial value indefinitely.

The simplest model that permits co-infection of a single host (Dietz, 1979; Gupta et al., 1994; Lipsitch, 1997) is shown in Fig. 3B, with equations

$$
\begin{align*}
\frac{dI_0}{dt} &= -(\lambda_1 + \lambda_2)I_0 + u_1 I_1 + u_2 I_2 \\
\frac{dI_1}{dt} &= \lambda_1 I_0 - u_1 I_1 - k_1 \lambda_2 I_1 - u_2 I_1 \\
\frac{dI_2}{dt} &= \lambda_2 I_0 - u_2 I_2 - k_1 \lambda_1 I_2 + u_1 I_1 \\
\frac{dI_1}{dt} &= k_2 \lambda_2 I_1 + k_1 \lambda_1 I_2 - (u_1 + u_2) I_1 \\
\lambda_1 &= \beta_1 (I_1 + q I_2) \\
\lambda_2 &= \beta_2 (I_2 + q I_1)
\end{align*}
$$

(12)

Moreover, this model has a single stable equilibrium including both strains whenever the strains are indistinguishable and secondary infection is possible ($k > 0$) (Dietz, 1979; Gupta et al., 1994; Lipsitch, 1997). Before considering how we may adjust this model to meet our conditions, it is worthwhile to consider why it violates them.

Note that in this model, an individual infectious with one strain may become more infectious in total (since $q = 1$ makes dually infected hosts equally infectious with each serotype as singly infected hosts) and $k_1 = k_2 = k > 0$ (so dual infection is possible in both directions), this model fails our criteria for an appropriate null model that does not “build in” stable coexistence. The ecological dynamics of this model cannot be written independent of strain composition, since here,

$$
\begin{align*}
\frac{dN_0}{dt} &= -\beta(N_1 + q N_2) N_0 + u N_1 \\
\frac{dN_1}{dt} &= \beta(N_1 + q N_2) N_0 - u N_1 - 2k\beta[I_1 I_2 + q N_2 N_1] \\
\frac{dN_2}{dt} &= -2u N_2 + 2k\beta[I_1 I_2 + q N_2 N_1]
\end{align*}
$$

(13)

Moreover, this model has a single stable equilibrium including both strains whenever the strains are indistinguishable and secondary infection is possible ($k > 0$) (Dietz, 1979; Gupta et al., 1994; Lipsitch, 1997). Before considering how we may adjust this model to meet our conditions, it is worthwhile to consider why it violates them.

Another equivalent way of stating this is that, in this model unlike the earlier model of Eq. (1), superinfection leading from a pure $I_1$ or $I_2$ state to a mixed $I_{12}$ state only arises due to within-host balancing competition between the two strains. The model of Eq. (12) encodes within-host balancing selection, and as a result the predicted strain dynamics show population-level balancing selection.

It is important to note at this point that the modeler has a choice about how to apply a given model, such as Model 13, to indistinguishable strains. In the discussion so far, we have assumed that the secondary infection coefficients $k_1$ and $k_2$ are fixed constants, which are nonzero whether or not the strains are different (because, as noted above, secondary infection arises due to balancing selection within a host). Suppose instead that secondary infection could only occur if the incoming strain was more “fit” in some sense (e.g. had higher in vivo growth rate) than the resident strain. For simplicity, let each strain have some defined level of within-host competitive ability $w_i$, and let $k_i = k(w_i - w_{-i})$, with the
restriction that \( k(v) = 0 \) for \( v \leq 0 \), and \( k(v) \geq 0 \) for \( v > 0 \). Under this assumption, the equations become as follows (suppose without loss of generality that strain 1 is always of greater or equal fitness than strain 2):

\[
\frac{dI_0}{dt} = -(\lambda_1 + \lambda_2)I_0 + u_1 I_1 + u_2 I_2 + u_3 I_{12}
\]

\[
\frac{dI_1}{dt} = \lambda_1 I_0 - u_1 I_1 + u_2 I_{12}
\]

\[
\frac{dI_2}{dt} = \lambda_2 I_0 - u_2 I_2 - k_1 I_1 I_2 + u_1 I_{12}
\]

\[
\frac{dI_{12}}{dt} = k_1 I_1 I_2 - (u_1 + u_2) I_{12}
\]

\[
\lambda_1 = \beta_1 (I_1 + q I_{12})
\]

\[
\lambda_2 = \beta_2 (I_2 + q I_{12})
\]

On their face, these equations do not look neutral, but when they are applied to indistinguishable strains, neither strain can establish secondary colonization, hence \( k_1 \to 0 \) and the model reduces to the neutral Model 11. However, when the two strains are distinguishable, strain 1 may be able to establish secondary colonization. This finding emphasizes that the idea of “applying a model to indistinguishable strains” depends on one’s assumption of how various parameters behave when the strains are indistinguishable.

Returning to the assumption that the \( k_i \) are always nonzero, one way to modify Model 12 to be neutral when applied to indistinguishable strains is to include two additional states of “dual” infection with the same strain, as in model 1. In this way, singly infected hosts can become infected with another strain, at equal rates whether it is the same strain they already have, or not.

A second, slightly simpler model would be one that keeps almost the structure of model Eq. (12), but adds the possibility that dually infected hosts may have one strain “knocked out” by contact with the other strain, returning to being singly infected. This occurs at a rate \( c \) times as great as second infection of a singly infected person. Also, we now allow dually infected hosts to clear both strains simultaneously; this does not change anything important but allows us to make a structure identical to that in Eq. (1). The modified system is shown in Fig. 3C and the following equations, written now with \( J \) to facilitate comparisons below:

\[
\frac{dJ_0}{dt} = -(\lambda_1 + \lambda_2)J_0 + u_1 J_1 + u_2 J_2 + u_3 J_{12}
\]

\[
\frac{dJ_1}{dt} = \lambda_1 J_0 - u_1 J_1 - k_2 J_2 J_1 + c k_1 J_1 J_{12}
\]

\[
\frac{dJ_2}{dt} = \lambda_2 J_0 - u_2 J_2 - k_1 J_2 J_1 + c k_2 J_2 J_{12}
\]

\[
\frac{dJ_{12}}{dt} = k_2 J_2 J_1 + k_1 J_1 J_2 - u_3 J_{12} - c(k_1 J_1 + k_2 J_2) J_{12}
\]

\[
\lambda_1 = \beta_1 (J_1 + q J_{12})
\]

\[
\lambda_2 = \beta_2 (J_2 + q J_{12})
\]

(14)

In this case, it is still not possible to write the ecological dynamics without including strain identifiers. However, note that for the particular parameter values \( c = q = 1/2 \), Eq. (14) become equivalent to Eq. (1), with \( J_1 = I_1 + I_{11} \) and \( J_2 = I_2 + I_{22} \), hence its ecological and ancestor-tracing ecological dynamics become strain-independent. Whichever way the model is written, it can be interpreted as follows: an individual is either infected or not. If infected, he exerts a certain force of infection, regardless of which strains infect him (since \( q = 1/2 \), dually infected individuals are equally infectious in total to singly infected individuals). If he is infected, he may lose half of his infectiousness to a new strain with which he has contact, at a rate \( k \) times that at which an uninfected person becomes infected given contact. In a two strain model, an individual with both strains who acquires a strain through subsequent infection will half the time lose the strain of the same type he acquired, hence \( c = 1/2 \).
To summarize the implications for models similar to that in Eq. (12): if dual infection in individual hosts is impossible (Eq. (11)), the model is neutral when applied to indistinguishable strains. If dual infection is possible (for any pair of strains), and is implemented as it is in Eq. (12), then the model is not neutral, and tends to promote coexistence of indistinguishable strains. Two fundamentally different solutions are possible to make a neutral null model. First, one can retain a model structure like that of Eq. (12), but with the proviso that indistinguishable strains cannot cause dual infection. This would be plausible if a within-host fitness advantage were required for the second strain to grow within an already colonized host. In this case, no secondary infections would occur for indistinguishable strains, and only one of the $k_i$ would be positive when two different strains were considered. Second, one can add two more states (dual colonization with two of the same strain), transforming Eq. (12) into Eq. (1). This is equivalent, as we showed, to maintaining the states named in Eq. (12) but allowing secondary infections of the dual state, returning hosts to the singly colonized state, for a particular choice of parameters.

**Generalizations**

**More complex state variables**

In the foregoing, we have considered only models in which a host’s state $x$ is determined by his multiplicity of infection and the identity of the strain(s) present. More generally, models might permit a host with, say, multiplicity 2, to be more infectious with one of the strains than with the other. As long as the determining factor for which strain is more infectious is not the identity of the strain, this should not violate our conditions for neutral null models. For example, perhaps the first strain to infect the host is more infectious than the second (or vice-versa), or the two strains infecting a host are always unequally infectious, but it is random which strain infecting the host is more infectious.

Such models may be easily accommodated within our proof, simply by expanding the possible number of states, such that (for example) $h \rightarrow x$ is now a list ordered by the relative infectiousness (or order of arrival) of the strains infecting that host (rather than in nondecreasing order of the strain identities as defined in Section 2). Then, the $i$ and $k$ indices of $M[θ_i] ← [θ'k]$ must index not multiplicities, but positions within hosts of a particular multiplicity. For example, if hosts may have up to three strains of unequal infectiousness, $i$ and $k$ would range from 1 to 6, indicating the first position in a singly infected host; the first and second position in a dually infected host; and the first, second and third position in a triply infected host.

This is an important point. A model need not be symmetric in all respects in order to meet our criteria for neutrality. Rather, it must be symmetric with respect to the identities of the strains.

**Models with immunity**

Models with strain-specific immunity – in which prior experience of strain $j$ makes a host less likely to be infected with strain $j$ to a degree greater than it protects against other strains – are well known for promoting stable coexistence of strains (Gupta et al., 1996; Gupta and Maiden, 2001). This is the standard mechanism of negative frequency-dependent selection often invoked to explain stable antigenic polymorphism (Lipsitch and O’Hagan, 2007). A simple model of that type is shown in Fig. 4A, and is given by the following equations:
Such models should fail our test for neutral null models, since they often predict stable coexistence of more than one strain. They fail the ecological criterion because, by definition, a state of immunity to one strain affects the transmission of that strain more than the transmissions of other strains, hence the ecological dynamics cannot be written independent of strain identity. It should be emphasized that this does not mean that there is anything intrinsically wrong with models including strain-specific immunity, only that the strain-specific immunity itself is a mechanism promoting stable coexistence.

On the other hand, it is possible to have a neutral model that includes immunity, as long as the immunity does not depend on the identity of the strains with which an individual has been infected. An example is shown in Fig. 4B and is given by the equations:

\[
\begin{align*}
\frac{dX}{dt} &= - \left( \lambda_1 + \lambda_2 \right) X + w'Z' + wZ \\
\frac{dY_1}{dt} &= \lambda_1 X - u_1 Y_1 \\
\frac{dY_2}{dt} &= \lambda_2 X - u_2 Y_2 \\
\frac{dZ}{dt} &= u_1 Y_1 - kZ - wZ \\
\frac{dY_1}{dt} &= kY_1 - u_1 Y_1 \\
\frac{dY_2}{dt} &= kY_2 - u_2 Y_2 \\
\frac{dY_{i+1}}{dt} &= u_i Y_i + u_{i+2} Y_{i+2} - wZ \\
\lambda_i &= \beta_i (Y_i + qY_{i-1}), \; i = 1, 2
\end{align*}
\]

This fits nicely into the framework developed in Section 2, where there are three host classes, “naïve”, “infected once” and “infected twice” (denoted by unprimed and primed variables respectively), though the memory of these prior infections can wane; thus here \( \Theta = 3 \). We can rewrite the dynamics now for identical strains as

\[
\begin{align*}
\frac{dN_{i\theta}}{dt} &= - \beta(N_{i+1} + qN_{i+1})N_{i+2} + wN_{i+1} \\
\frac{dN_{i\theta}}{dt} &= \beta(N_{i+1} + qN_{i+1})N_{i+2} - uN_{i+1} \\
\frac{dN_{i\theta}}{dt} &= uN_{i+1} - k\beta(N_{i+1} + qN_{i+1})N_{i+2} - wN_{i+1} \\
\frac{dN_{i\theta}}{dt} &= k\beta(N_{i+1} + qN_{i+1})N_{i+2} - uN_{i+1} \\
\frac{dN_{i\theta}}{dt} &= uN_{i+1} - wN_{i+1}
\end{align*}
\]

where the \( N_{i\theta} \) indicate hosts with \( \theta - 1 \) prior infections and \( i \) current infections. The key here is that we defined the host classes indexed by \( \theta \) so that they could be time-varying and even depend on the host’s history, as long as they do not depend on the identity of strains with which a host was infected.
In summary, models with immunity that is strain-specific are not neutral because the ecological dynamics depend on the identity of strains with which individuals have been previously infected, and because they tend in fact to promote stable coexistence of strains. Indeed, such models represent one class of mechanistic models for the stable coexistence of strains. More general models of strain immunity involve matrices of cross immunity between strains, and can exhibit complex regular or chaotic oscillations in strain frequencies over time (e.g. (Gupta et al., 1998)). Models with immunity that reflects previous exposure but is not strain-specific may be neutral, as exemplified by Eq. (15), and these may be understood within the framework of Section 2 by considering a host’s history of exposure as determining his class, indexed by $\theta$.

Models with immigration or mutation

Our main result was stated for “closed clonal transmission models,” namely those in which each infection is uniquely descended from a particular infection of the same type present at time 0. This property is used in the proof since it implies the existence of the $M_{\theta i \leftarrow \theta k}$ necessary for Eq. (7).

Models with immigration of hosts infected with particular strains violate the definition of “closed clonal transmission models,” since some infections are not descended from infections present at time 0 but rather from immigrant infections. Models with mutation of types violate this definition because, although a particular infection may descend from one present at time 0, it was of a different type. Neither type of model can be accommodated within our framework of neutral null models. This is understandable, because in a model with immigration, strain frequencies will reflect to some extent the influence of immigration, hence cannot be controlled for all times by the choice of starting conditions (Hubbell, 2001).

While models which meet our two tests for null neutrality cannot include mutation or immigration, they can form the basis for the exploration of these phenomena on strain diversity in a context where the effect of strain selection is understood and controlled for. Epidemiological models typically consider only a finite number of strains (or alleles), and thus mutations may cause all strains to be present at all times (in mutation-mutation balance) because of recurrent mutation from other types. Such models therefore cannot in general meet the population genetic criterion for a neutral null model, and immigration (Lipsitch et al., 2000a,b) or mutation (Bonhoeffer and Nowak, 1994) can maintain stable strain coexistence.

Population genetic models usually take a slightly different perspective, allowing for evolution at the level of genetic sequences and thus consider many or an infinite number of possible alleles. In that case, coexistence of strains is a result of the balance between mutation and extinction in a constant process of strain turnover (Fraser et al., 2005). Some models, notably several motivated by influenza dynamics, incorporate both population genetic and epidemiological features (Boni et al., 2004; Grenfell et al., 2004; Boni et al., 2006; Koelle et al., 2006; Day and Gandon, 2007).

Discussion

We have shown that any model whose ancestor-tracing ecological dynamics can be written without naming the strains involved will lack a stable coexistence equilibrium for indistinguishable strains. A model that satisfies these two criteria is thus an appropriate null model with which to attempt to understand coexistence. Thus, if a neutral null model for indistinguishable strains is identified, it may serve as a basis, upon which mechanisms that might promote stable coexistence may be layered by creating additional states or making
decisions that, for explicit biological reasons, the parameters for two or more biologically different strains should depart from those which satisfy the criteria. For example, if biological considerations suggest that distinct strains are able to colonize distinct niches within a host by attaching to different receptors, then one might postulate that Model 13 is appropriate with parameters that violate our criteria for null models: for example, with $k > 0$ (allowing dual infections) but with $0 \leq c < \frac{1}{2}$ (because exposure to one strain is less likely to displace another strain than to create a dual infection), and/or with $1 \geq q > \frac{1}{2}$ (because by using different receptors, the two strains can create more total infectiousness from a coinfected host than if only one strain were present). Alternatively or in addition, strains may experience tradeoffs between infectiousness $\beta$ and clearance rate $u$, between different modes of transmission, or between other properties that could promote coexistence. The key point is that such hypothetical mechanisms of coexistence should be explicit, i.e. layered on top of a model that does not predict stable coexistence of indistinguishable strains.

From a different perspective, the analysis here indicates that prior efforts to model the coexistence of strains have had to incorporate mechanisms to permit stable coexistence, often without explicit biological justification. Lipsitch (Lipsitch, 1997) modeled the coexistence of two serotypes of *Streptococcus pneumoniae* or other colonizing bacteria using Model 13 but with parameter values $q = 1, c = 0$, and $k > 0$. This parameterization violates our criteria, not surprising in retrospect since the purpose of the earlier model was to study perturbations to a stable state of coexistence among serotypes. The model used there is justified in a general sense, by the fact that in every population studied, one observes the coexistence of pneumococcal serotypes, often maintained over long periods of time (Bogaert et al., 2004; Lipsitch and O’Hagan, 2007), and the purpose of the study was to understand the effects of selectively inhibiting one serotype by vaccination. However, the mechanism permitting stable coexistence was not specified for that model. Different serotypes have many biological differences – including invasiveness (Hanage et al., 2005), duration of colonization (Sleeman et al., 2006; Hogberg et al., 2007), accessibility of surface determinants (Abeyta et al., 2003), linkage disequilibrium with antimicrobial resistance (McCormick et al., 2003) and the pilus operon (Basset et al., 2007), and possibly susceptibility to mechanisms of immunity directed at noncapsular antigens (Malley et al., 2005). Unfortunately, it is not currently clear which of these mechanisms (perhaps it is several) accounts for the maintenance of diversity in this species (Lipsitch and O’Hagan, 2007). In the absence of such certainty, a generic mechanism for promoting stable coexistence in a model – in this case, the existence of a “protected” class of dually infected persons – may be justifiable as a placeholder but does not represent a mechanistic explanation of stable coexistence. Moreover, predictions about how the population will change as a result of policy changes, such as vaccination or antimicrobial use, must be regarded as tentative when they rely on a model in which stable coexistence of strains prior to intervention is permitted through an unidentified mechanism. On the other hand, just as the classic Lotka-Volterra model with no explicit resource but with phenomenological competition terms for inter- and intraspecific competition provides considerable analytic insight into competitive dynamics, such “phenomenological” models of strain coexistence may be valuable tools until we have a clearer understanding of the mechanisms maintaining coexistence.

Such models have been used by a number of investigators besides those already cited, for example Zhang and colleagues (Zhang et al., 2007) to study host-pathogen coevolution and Turner and Garnett (Turner and Garnett, 2002) to study multiple strains of gonorrhea. Elbasha and Galvani (Elbasha and Galvani, 2005) used a similar model, though including immune states, to study vaccination against multiple HPV types. In each case, the model
structure builds in stable coexistence of strains “for free,” without stating an explicit mechanism.

Other authors have built in an explicit mechanism for stable coexistence of nonidentical strains, such as a tradeoff between transmission fitness to uninfected hosts vs. ability to “superinfect” a host already carrying a the other strain (Levin and Pimentel, 1981; Martcheva et al., 2008). In such cases, the model reduces for indistinguishable strains to Eq. (10), where individuals simply have one strain or the other and are impervious to superinfection; by allowing one strain to transmit better and the other to superinfect better, stable coexistence can be maintained (Levin and Pimentel, 1981). Mosquera generalized such models to include both “superinfection” (replacement of one strain by another) and “coinfection” (dual infection with both strains) (Mosquera and Adler, 1998).

A related problem was faced by Austin et al. (Austin et al., 1999), who wished to model the coexistence of drug-sensitive and drug-resistant strains of \textit{S. pneumoniae}. These authors faced the problem that a simple model of resistant and sensitive strains with no possibility for a host to carry both strains \((k = 0)\) leads to competitive exclusion of resistant or sensitive strains, depending on the balance between the use of antimicrobial drugs and the fitness deficit of resistant strains (Lipsitch, 2001a,b).

Austin et al (Austin et al., 1999) created a model that permitted stable coexistence of the two strains by two mechanisms. First, they allowed for subdivision of the host population into those currently on treatment and those currently not on treatment. This in principle is a form of “habitat heterogeneity” that in general ecological terms can lead to stable coexistence, though as we show in a companion paper (Cohen et al., 2008) this mechanism alone makes a minor contribution to stable coexistence of resistant and sensitive strains. In addition, the model was set up such that although a host could have only one strain at a time, each strain could displace the other by “superinfection”. Austin et al. used parameters such that, in line with that requirement, the sensitive strain was less transmissible than the resistant strain (even in the absence of antimicrobial use), but better at superinfection. To date, we are aware of no evidence that resistant strains are more transmissible than sensitive ones for any bacterium, and considerable evidence to the contrary (Andersson and Levin, 1999; Andersson, 2003). Moreover, the relative rates of superinfection of different strains are, we believe, entirely unstudied. Nonetheless, Austin et al (Austin et al., 1999) designed a model that could capture the observed phenomenon of stable coexistence of resistant and sensitive strains, albeit with a mechanism that was not well supported by empirical data.

A separate problem of ongoing interest is the characterization of the genetic diversity of pathogens, and what light this can shed on evolutionary and epidemiological processes. The models typically treat epidemiological effects as entirely implicit (Fraser et al., 2005; Rambaut et al., 2008), in some cases using discrete generations for tractability, although the dynamics clearly have overlapping generations (Fraser et al., 2005). In rare cases, the effect of strain immunity is explicitly included via a coupled epidemiological and population genetic model (Boni et al., 2006; Koelle et al., 2006). In this case, where the aim is to identify and characterize selective forces, it is important that the null models be neutral; our analysis partly addresses this, but further work is needed to examine whether stochastic fluctuations in higher dimensional epidemic models can be mapped onto neutral fluctuations in gene frequencies derived for simple population genetic models. This would be needed for example to model explicitly the effects of co-infection, recombination and epidemic spread on gene frequencies, all previously treated implicitly (Fraser et al., 2005). It is important to note that we have not explored the conditions for a model to generate neutral or nearly neutral drift dynamics amongst strains, which are quite general and parameter dependent,
but have rather attempted to elucidate conditions under which models may generate selection due to structural features even when all parameters are equal.

In conclusion, we have described two simple criteria that ensure that null models for strain interactions do not “build in” stable coexistence through implicit, unstated mechanisms, and have shown that if the “ancestor-tracing ecological” version of a model satisfies the ecological criterion (autonomous ecological dynamics for indistinguishable strains), then the model itself also satisfies both the ecological and population-genetic (no stable coexistence of identical strains) criteria. If the goal of a modeling exercise is to understand the biological mechanisms underlying strain coexistence, then the model that includes such mechanisms should reduce – in the case of identical strains – to one that satisfies this criterion. When the mechanism of coexistence is not the object of study, it may be necessary to build in such mechanisms in order to create a model that reproduces observed patterns of coexistence, but ultimately one’s confidence in predictions of models that rely on (but do not seek to explain) coexistence would be increased if one had confidence in the mechanisms of coexistence that are assumed. An exciting area of ongoing research is to understand the roles of selective (Gupta et al., 1997; Ferguson et al., 2003; Koelle et al., 2006) vs. neutral (Fraser et al., 2005) explanations of persistent pathogen strain diversity for a range of pathogens, and among the selective explanations to distinguish those based on acquired immunity to specific and conserved antigens, differential tropisms for tissues and receptors, and other mechanisms of niche differentiation.

Acknowledgments

ML was supported by cooperative agreement 5U01GM076497 (Models of Infectious Disease Agent Study) from the National Institutes of Health. WPH and CF were supported by Royal Society Research Fellowships. Edward Goldstein is thanked for helpful discussions.

References


Epidemics. Author manuscript; available in PMC 2011 May 23.


Fig. 1.
(A) structure of the model in Eq. (1). Hosts may become infected by one or two strains (which may be the same or different), but not more; second infections occur at a reduced rate compared to primary infections. Among hosts who have two strains already, it is possible (if $c>0$) for one strain to “knock out” another. See text for more details (B) General approach for a model with multiple infections. The figure shows compartments with different multiplicities of infection, starting from an uninfected shaded state. Infection from this state leads to an infected state with strain $i$. Superinfection leads to a dually infected state, a triply infected state, and so on, leading potentially but not necessarily to increased infectiousness. The ‘knock-out’ process is not shown explicitly. If the ‘ladder’ of superinfection is truncated to two co-infections, this model is equivalent to the model of A.
Fig. 2.
Evolution of the model in Eq. (1) under various starting conditions. Left ordinate shows the state variables; right ordinate shows $f_1$ (in black). Colors: $I_1$: red; $I_2$: green; $I_{11}$: blue; $I_{12}$: tan; $I_{22}$: magenta. Parameters: $c=0.14$; $k=0.4$; $q=0.5$; $\beta=2$; $u=1$. 
Fig. 3.
Alternative model structures. (A) Model equations 11; (B) Model equations 12; (C) Model equations 14.
Fig. 4.
Models with immunity. (A) A model with strain-specific immunity, Eq. (15). This model is not neutral, and it promotes stable coexistence of the two strains. (B) A neutral model with immunity, Eq. (16). This model has immune states that depend only on the number of times a host has been infected, but not on which strains have infected the host.
Corresponding terms in ecological and ancestor-tracing ecological models.

<table>
<thead>
<tr>
<th>Event</th>
<th>Term in equation for $\mathcal{N}_0$, $i \geq 1$</th>
<th>Event</th>
<th>Term in equation for $\mathcal{M}[\theta t^{-1}]$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Host demographics</td>
<td>none</td>
<td>Birth, death, etc.</td>
<td>none</td>
</tr>
<tr>
<td>2. Change in host covariates from $0$ to $0'$</td>
<td>$-\omega_0(\mathcal{N}_0)$</td>
<td>Aging, acquisition of immunity, etc., exit from class $0$</td>
<td>$-\psi_{0[0]^{-1}}[\theta']\omega_0(\mathcal{N}_0)$</td>
</tr>
<tr>
<td>3. Clearance:</td>
<td>$-u_0(\mathcal{N}_0)$</td>
<td>a. Clearance of $h$ strains, regardless of ancestry, from a $[0</td>
<td>\theta]$ host who harbors one or more strains descended from a $[0'</td>
</tr>
<tr>
<td>a. Clearance of one or more strains from a $[0</td>
<td>\theta]$ host</td>
<td></td>
<td>(corresponding, positive terms for entry from other classes into class $0'$)</td>
</tr>
<tr>
<td>b. Clearance of $h$ strains from a $[0(i+h)]$ host</td>
<td>$\nu_{0h,\theta h}(\mathcal{N}_{0h,\theta h})$</td>
<td>b. Clearance of $h$ strains, not descended from a $[0'</td>
<td>j]$ host, from a $[0(i+h)]$ host who also harbors one or more strains descended from a $[0'</td>
</tr>
<tr>
<td>4. New infections:</td>
<td>$\sum_{\theta',d,}\Lambda_{1<a href="i">i</a>}^{-1}<a href="%5Cmathcal%7BN%7D_%7B0(i+h)%7D,%5Cvec%7BN%7D">\theta',j</a>_i$</td>
<td>a. Infection of a $[0(i-1)]$ host by a strain descended from a $[0'</td>
<td>j]$ ancestor</td>
</tr>
<tr>
<td>a. Infection of a $[0(i-1)]$ host by a new strain, which may originate from any host class index by $0',d$</td>
<td>b. Infection of a $[0(i-1)]$ host already carrying a strain descended from a $[0</td>
<td>\theta]$ host by any new strain</td>
<td>(corresponding negative terms for loss from lower multiplicities)</td>
</tr>
</tbody>
</table>

NOTE: The force of infection of strains descended from $[0'|j]$ ancestors, $\lambda_{[0]|j}$, would be defined in a “classic” transmission model tracing ancestry as $\Lambda_{[0](l)}^{-1}[\theta',j]$; the force of infection of strains from $[0(i-1)]$ hosts (the infectees) and the number of all other types of hosts. In an ordinary transmission model like the one in Eq. (1), the term is much simpler, i.e. $\Lambda_{[0](i-1)]^{-1}[\theta'](\sum_{\theta'}d_i[N_{[0(i-1)],[\theta']})$ is the force of infection exerted on hosts of class $0,i-1$ by hosts of class $0',d$, by which they acquire an additional strain and move into class $0,i$. In general, this force of infection may be a complicated function of the number of $0,i-1$ hosts (the infectees) and the number of all other types of hosts. In an ordinary transmission model like the one in Eq. (1), the term is much simpler, i.e. $\Lambda_{[0](i-1)]^{-1}[\theta'](\sum_{\theta'}d_i[N_{[0(i-1)]})$. The corresponding vector $\bar{\Lambda}_{[0](i-1)](N_{[0(i-1)]})$ is a list of the forces of infection exerted on $0,i-1$ hosts by all types of hosts. The dot product in the equations for the ancestor-tracing model apportions this force of infection according to the ancestry of the strains currently infecting each of these types of infecting hosts.

*Epidemics.* Author manuscript; available in PMC 2011 May 23.